

AN EULERIAN APPROACH TO OPTIMAL TRANSPORT WITH
APPLICATIONS TO THE OTTO CALCULUS

by

Benjamin Schachter

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

© Copyright 2017 by Benjamin Schachter

Abstract

An Eulerian Approach to Optimal Transport with Applications to the Otto Calculus

Benjamin Schachter

Doctor of Philosophy

Graduate Department of Mathematics

University of Toronto

2017

This thesis studies the optimal transport problem with costs induced by Tonelli Lagrangians. The main result is an extension of the Otto calculus to higher order functionals, approached via the Eulerian formulation of the optimal transport problem. Open problems 15.11 and 15.12 from Villani's *Optimal Transport: Old and New* are resolved. A new class of displacement convex functionals is discovered that includes, as a special case, the functionals considered by Carrillo-Slepčev. Improved and simplified proofs of the relationships between the various formulations of the optimal transport problem, first seen in Bernard-Buffoni and Fathi-Figalli, are given. Progress is made towards developing a rigorous Otto calculus via the DiPerna-Lions theory of renormalized solutions. As well, progress is made towards understanding general Lagrangian analogues of various Riemannian structures.

In memory of my mother, Donna

Acknowledgements

I've had more luck in this journey than anyone could ever ask for. I've been lucky to have had patient and generous teachers, wonderful friends, and a loving and supportive family.

First and foremost, I am grateful to have had Almut Burchard and Wilfrid Gangbo as my advisors. Almut, with boundless generosity, has guided my development as a mathematician for the past five years. She provided me with invaluable mathematical, professional, and personal—as well as financial—support throughout my time as a graduate student, especially through times where I was discouraged. Wilfrid, in his short time visiting Toronto in 2014, initiated the problem which blossomed into this thesis, and continued to guide and support this project throughout the ensuing years.

I am grateful to have had Robert McCann and Peter Rosenthal on my supervisory committee; they gave me thoughtful and considerate advice and guidance over the years. Without Robert's exceptional insight into and understanding of the theory of optimal transport, I could not have completed this project.

Over the years, I have had excellent teachers. I am particularly grateful to Bob Jerrard, Fiona Murnaghan, and Stu Rankin.

I began my journey into mathematics because of two people: Andres del Junco, who sadly passed away in 2015, and Yiannis Sakellaridis. Without their mentorship and guidance I never would have followed this path. I am forever indebted to them.

I am grateful to have been in a department with excellent staff. I am proud to be able to add my name to the long list of people who have thanked Jemima Merisca and Ida Bulat for their help and support over the years.

Cody Schacter, my closest friend since kindergarten, has been with me through everything; I could not ask for more. I am grateful for the friendship of Cristina Diaz Borda and Ariel Riesenbach. The Riesenbachs' have made me feel like a member of their family. I'm grateful to have Lynda Jacobs in my life; she has been there for me every day.

I'm grateful to have been surrounded by smart and funny colleagues in the math department. I'm grateful for the friendships I've formed here, especially with Jeremy Lane and Craig Sinnamon. Thanks is owed to Kevin Luk for teaching me how to do case interviews.

The Ontario Ministry of Advanced Education and Skills Development provided me with financial support through Ontario Graduate Scholarships and a Queen Elizabeth II Graduate Scholarship in Science and Technology.

Last of all, I'm thankful to my family for their unwavering love, support, and guidance. The best advice anyone ever gave me about doing a PhD came from my father, Harry. "It's just a thing you have to do so you can move on with your life."

Contents

1	Introduction	1
1.1	The Optimal Transport Problem from Lagrangian and Eulerian Perspectives	3
1.2	Outline of this thesis	4
2	Lagrangian Induced Cost Functions	6
2.1	Introduction	6
2.2	Tonelli Lagrangians	7
2.2.1	Lagrangian Systems and Action Minimizing Curves	14
2.2.2	Properties of Cost Functions Induced by Lagrangians	16
2.3	Hamiltonians	20
2.4	Lagrangian and Hamiltonian Systems	23
2.4.1	Hamiltonian Systems Have Unique Local Solutions of Class C^2 . .	24
2.4.2	Equivalence of Hamiltonian and Lagrangian Systems	25
2.4.3	Conservation of Hamiltonians	27
2.4.4	Dependence on Initial Conditions	27
3	Five Formulations of the Optimal Transport Problem	30
3.1	Introduction	30
3.2	Some General Optimal Transport Theory	31
3.2.1	Equivalence of the Monge and Kantorovich Problems	34
3.2.2	Kantorovich Duality	36
3.3	The Lagrangian Optimal Transport Problem	37
3.3.1	Further Properties of Cost Functions Induced by Lagrangians . .	39
3.3.2	Regularity of Optimal Trajectories	42
3.4	The Eulerian Optimal Transport Problem	48
3.5	Five Equivalent Formulations of the Kantorovich Cost	51

4	The Otto Calculus for Higher Order Functionals	55
4.1	Setup	56
4.2	One derivative and one dimension	58
4.3	A General Formula for the Displacement Hessian	64
4.4	Epilogue: Issues of regularity	69
	Appendix A Otto Calculus Computations - Riemannian Setting	71
A.1	Preliminaries	71
A.2	Coordinate computations	73
A.2.1	My displacement Hessian	73
A.2.2	Villani's displacement Hessian	76
A.3	Comparison of Displacement Hessians	82
	Bibliography	83

Chapter 1

Introduction

This thesis is focused on the theory of optimal transport. In the optimal transport problem, one seeks to find the most efficient way to transfer mass from one place to another. More formally, the problem is to choose, among all maps which push forward a specified initial probability measure to a specified final probability measure, the map that minimizes a given cost function.

In this thesis, several improvements are made to the *Otto calculus*, first developed by Otto ([34], [35]). The Otto calculus is a formal calculus on Wasserstein space, the space of probability measures on a manifold endowed with a notion of distance derived from the optimal transport problem. Although Wasserstein space lacks the structure needed to define derivatives—for instance, it is not a Hilbert manifold—the Otto calculus allows one to formally compute tangent vectors to curves and Hessians of functionals along these curves, where the notion of a curve in Wasserstein space needs to be interpreted correctly.

The project which resulted in this thesis was first presented to me by Wilfrid Gangbo (who eventually became my co-advisor) while he was visiting the Fields Institute for a semester in the fall of 2014. The problem originally posed to me was to understand the relationships between the various formulations of the optimal transport problem ([16], [4]) and to try to use this understanding to develop improved versions of the inequalities established in [18] and [11]. Over the course of working on this problem, as I learned the optimal transport literature, I became interested in McCann’s *displacement convexity* ([28]) and understanding it via the Otto calculus.

The notion of displacement convexity of a functional on Wasserstein space is the analogue of geodesic convexity of a function on a Riemannian manifold. A functional on Wasserstein space is called displacement convex if it is convex along every Wasserstein geodesic, where McCann’s *displacement interpolants* are understood as geodesics. The

Otto calculus provides a natural way to test displacement convexity of functionals on Wasserstein space, by taking second derivatives—computing the *displacement Hessian* of the functional—and checking for positivity, analogous to checking whether a smooth function is geodesically convex on a Riemannian manifold.

The definition of a displacement Hessian of a functional is motivated by the definition of the Hessian of a smooth function on a Riemannian manifold. Let $\phi : M \rightarrow \mathbb{R}$ be a smooth function on a Riemannian manifold M . Its Hessian at a point x is the bilinear form defined by

$$\nabla\nabla\phi(v, v) = \frac{d^2}{ds^2}\phi(\gamma(s))\Big|_{s=0}$$

where γ is a geodesic curve with $\gamma(0) = x$ and $\gamma'(0) = v$. If $\nabla\nabla\phi$ is positive definite, then ϕ is geodesically convex. The situation in Wasserstein space is analogous: given a functional \mathcal{F} , the Otto calculus allows for the computation of its displacement Hessian:

$$\frac{d^2}{ds^2}\mathcal{F}(\rho(s))$$

where ρ is a Wasserstein geodesic. The positivity of the displacement Hessian then corresponds to the convexity of the functional.

The Otto calculus has yielded several results; most notably, the Ricci curvature non-negativity condition associated to the displacement convexity of the entropy functional was first predicted via the Otto calculus in [35]. In learning the literature, it became clear to me that with a careful understanding of the different forms of the optimal transport problem, especially of the regularity of optimal trajectories in the Lagrangian formulation of the problem, and the regularity of the corresponding optimal vector field in the Eulerian formulation of the problem, improvements could be made to the Otto calculus.

The main contributions of this thesis are:

- I compute a canonical form for the displacement Hessians of functionals involving arbitrary derivatives of densities. The problem of computing the displacement Hessian of a functional involving the only the first derivative of the density was posed by Villani as Open Problem 15.11 in [39]; this is now resolved.
- I use the extended Otto calculus to find a new class of displacement convex functionals involving derivatives on S^1 . This class contains, as a special case, the functionals considered by Carrillo-Slepčev ([8]), and thus provides an alternate proof of their displacement convexity.

- I extend the Otto calculus to the setting of the optimal transport problem with costs induced by Tonelli Lagrangians, which resolves Open Problem 15.12 in [39]. (Paul Lee has worked in a similar direction from a Hamiltonian perspective in [26]).

As well, this thesis contains a careful analysis of the regularity of the optimal trajectories in the Lagrangian formulation of the optimal transport problem. I believe the proofs of these results are simpler than what is currently in the literature ([4], [16], [39]), but are known to experts.

There are two ongoing avenues of investigation; partial results from these investigations will make it into the final version of this thesis.

- A long-standing open problem is that of developing a rigorous Otto calculus, which has so far been a purely formal method of computation. There are limited, and highly technical, results in this direction ([36], [14]). This problem is discussed in [39], where Villani espouses a pessimistic view on the matter.

I am working towards a rigorous Otto calculus: with an improved understanding of the regularity of the optimal vector field in the Eulerian formulation of the problem, together with results from the DiPerna-Lions theory of renormalized solutions ([15], [1], [10], [13]), I hope to develop a rigorous Otto calculus.

- I am working towards understanding the geometry induced by a Tonelli Lagrangian. The formulation of the Otto calculus in this thesis suggests that there should be general Lagrangian analogues of Riemannian structures, such as the Riemannian volume form, Ricci curvature, Christoffel symbols, etc.

Thus far, Villani's tome *Optimal Transport: Old and New* ([39]) has been cited more than any other reference. This will continue to be the case throughout the rest of this thesis; his book is an invaluable resource on Optimal Transport.

1.1 The Optimal Transport Problem from Lagrangian and Eulerian Perspectives

In keeping with the last remark from the previous section, my description of the historical background on the optimal transport problem draws heavily on chapter 3 from [39].

The optimal transport problem was first described in the late eighteenth century by Monge. In [32], he posed the problem of the finding the least costly way to transport material extracted from the earth to the construction of a building. In Monge's original

formulation, the cost of transporting a mass m from position x to position y is the mass multiplied by distance: $m|x - y|$. There is a natural Lagrangian “particle trajectory” interpretation of this cost, that of a particle of mass m moving in a straight line from the point x to the point y . Little progress was made on the problem until the 20th century when Kantorovich ([24], [23]) reformulated and solved the optimal transport problem in a more general setting, viewing it as an infinite dimensional linear programming problem.

In 1991, Brenier characterized the solutions to the optimal transportation problem on \mathbb{R}^d with cost given by the square of Euclidean distance; he showed that optimal transport maps are given by the gradient of convex functions ([5]). In 1994, McCann developed this work further, realizing that solutions to the Monge problem yielded *displacement interpolants*: geodesic curves in the space of probability measures ([28]). He further discovered the phenomenon of *displacement convexity*: several important functionals are convex along displacement interpolants ([28]). The interpretation of solutions to the optimal transport problem as corresponding to particle trajectories was present in McCann’s work at this time ([27]). He later, in 2001, extended Brenier’s theorem to Riemannian manifolds, characterizing optimal transport maps in this setting as particle trajectories defined by a potential function ([30]).

The particle trajectory view of optimal transport was made more explicit by Benamou and Brenier in [3], where they interpret the Monge optimal transport problem from an explicitly Lagrangian perspective, and, motivated by fluid mechanics, transform the problem into a corresponding Eulerian problem. The Lagrangian formulation was developed further by Figalli, Fathi-Figalli, and Bernard-Buffoni ([17], [16], [4]), all of whom consider the optimal transport problem with costs induced by Tonelli Lagrangians. In [16] and [4], they make explicit the relationships between the optimizers of the Lagrangian and Eulerian formulations of the optimal transport problem.

1.2 Outline of this thesis

In the second chapter, Tonelli Lagrangians are defined and the regularity properties of their induced costs are deduced. The Hamiltonian dual to the Tonelli Lagrangian is introduced as its Legendre transform, and the properties of the corresponding flows on the cotangent and tangent bundle are stated.

In the first half of chapter three (section 3.2), background on the optimal transport problem is presented. The Monge, Kantorovich, and Kantorovich dual optimal transportation problems are defined. Several standard results are recalled.

In the second half of chapter three, the Lagrangian optimal transport problem is

defined (section 3.3). A careful analysis of the regularity of the optimal trajectories in the Lagrangian formulation of the optimal transportation problem is performed (subsection 3.3.2). The Eulerian formulation of the optimal transport problem is defined (section 3.4). The equivalence of the five formulations of the optimal transport problem and the relationships between their minimizers is proven (theorems 3.5.1 and 3.5.2). The results here are, I believe, presented more simply than what is currently in the literature, and are perhaps slightly stronger, but will not be a surprise to experts.

In the fourth chapter, the Otto calculus is introduced within the framework of the Eulerian optimal transport problem. Here, new results are proven.

In theorem 4.3.2, a canonical computation of the displacement Hessian of functionals involving derivatives of densities is presented. This resolves Open Problem 15.11 in Villani [39], which asks whether one can compute the displacement Hessian of a functional involving only a single derivative of the density.

In theorem 4.2.3, a new class of displacement convex functionals is discovered. These include, as a special case, the functionals considered by Carrillo-Slepčev ([8]), and thus provides a new proof of their displacement convexity. A counterexample shows that certain aspects of the sufficient condition used to establish the displacement convexity of this new class of functionals is also necessary.

The formulation of the Otto calculus in the fourth chapter resolves Open Problem 15.12 in Villani, where he asks whether it is possible to extend the Otto calculus to the setting of general Lagrangian induced costs. In appendix A, the displacement Hessian of the entropy functional is computed in coordinates; Villani's formulation of the Otto calculus is realized as a special case, when the cost function is Riemannian distance squared.

I am investigating, in ongoing work (some of which will be in the final version of this thesis), a regularization scheme to make the Otto calculus rigorous, via the theory of renormalized solutions introduced by DiPerna-Lions ([15]). As well, the Otto calculus developed in chapter 4 suggests general Lagrangian analogues to various Riemannian structures (e.g. volume forms, Ricci curvature, etc.); I am working to make these notions precise.

Chapter 2

Lagrangian Induced Cost Functions

2.1 Introduction

As discussed in the introduction, there is a natural “particle trajectory” interpretation of the optimal transport problem. A good approach to the optimal transport problem from a particle trajectory perspective is achieved by working with costs induced by Tonelli Lagrangians. This is the most general framework where the cost of sending a particle from a point x to a point y corresponds to a smooth curve starting at x and ending at y . Both the classical optimal transport problem and the optimal transport problem with costs given by squared Riemannian distance occur as special cases of the optimal transport problem with costs induced by Tonelli Lagrangians. However, the approach via Tonelli Lagrangians is more general and—perhaps more importantly—strips away all but the essential structure of the problem. These cost functions have been considered by, among others, Fathi-Figalli [16], Bernard-Buffoni [4] and are presented in chapter 7 of Villani’s book [39].

In the first half of this chapter, Tonelli Lagrangians will be defined and a careful application of the direct method in the calculus of variations will yield the existence of action minimizing curves. Cost functions will be defined by the least action of Tonelli Lagrangians. Regularity properties of action minimizing curves and cost functions will be proven. A large amount of information about Tonelli Lagrangian induced cost functions can be understood directly from this framework, without invoking the additional structure of the optimal transport problem.

In the latter half of this chapter, the Hamiltonian corresponding to a Tonelli Lagrangian will be introduced as the Legendre transform the Lagrangian, from the perspective of convex analysis. The relationships between Lagrangians and Hamiltonians, as presented in classical mechanics, will be shown. Properties of the Lagrangian flow on

the tangent bundle of a manifold and of the Hamiltonian flow on the cotangent bundle will be proven.

Instead of treating the Lagrangian flow as a flow of a vector field on the tangent bundle of a manifold, it will be more fruitful to view the flow as families of curves on the manifold itself, each of which satisfy the Euler-Lagrange equation. This will be the perspective taken through most of this thesis. A technical issue that will arise (in a later chapter) will be to find a preferred family of these curves which themselves are the flow of a vector field on the base manifold. These preferred curves will be interpreted as a solution to the optimal transport problem, and will be an object of interest, rather than the entire Lagrangian flow on the tangent bundle of a manifold.

The material in this chapter is classical, but is also crucial to the understanding of the optimal transport problem.

2.2 Tonelli Lagrangians

For notational convenience, $W^{1,1}((0, 1); \mathbb{R}^d) = W^{1,1}$. The tangent bundle of a manifold M is denoted by TM . When $M = \mathbb{R}^d$, which will almost always be the case in this document, then $T\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d$. The cotangent bundle is denoted T^*M and $T^*\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d$.

Definition 2.2.1. A function $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ is called a **Tonelli Lagrangian** if it satisfies the following properties.

- (a) The function L is of class C^2 .
- (b) For every $x \in \mathbb{R}^d$, the function $L(x, \cdot) : T_x\mathbb{R}^d \cong \mathbb{R}^d \rightarrow \mathbb{R}$ is strictly convex.
- (c) There exists a constant c_0 and a non-negative function $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ with superlinear growth, i.e.

$$\lim_{|v| \rightarrow +\infty} \frac{\theta(v)}{|v|} = +\infty,$$

such that

$$L(x, v) \geq c_0 + \theta(v).$$

The first variable in a Lagrangian will often be referred to as **position** and the second variable will often be referred to as **velocity**. For example, if L is a Tonelli Lagrangian, then $L(x, v)$ would be the Lagrangian evaluated at position x and velocity

v . The derivatives and gradients of a Tonelli Lagrangian L in the position and velocity variables will be denoted by $D_x L$, $D_v L$, $\nabla_x L$, and $\nabla_v L$.

Definition 2.2.2. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ of class C^2 is called **strongly convex** if $\nabla^2 f(x)$ is positive definite for all $x \in \mathbb{R}^d$. Denote a point in $\mathbb{R}^d \times \mathbb{R}^n$ by (x, y) . A function $f : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}$ of class C^2 is called **strongly convex in y** if $D^2 f_x(y) = D_y^2 f(x, y)$ is positive definite for every $(x, y) \in \mathbb{R}^d \times \mathbb{R}^n$, where $D^2 g(y)$ denotes the matrix of second derivatives of g and $D_y^2 g(x, y)$ denotes the matrix of second derivatives of g in y .

Tonelli Lagrangians $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ will often be assumed to be strongly convex in the second argument. The reason strong convexity is not built into the definition of Tonelli Lagrangians is that definition is chosen to have the minimal assumptions needed to guarantee the existence of action minimizing curves.

Definition 2.2.3. Let σ be a curve in $W^{1,1}((0, 1); \mathbb{R}^d) = W^{1,1}$. The **velocity** of the curve σ at $t \in (0, 1)$ is

$$\dot{\sigma}(t) := \frac{d}{ds} \sigma(s) \Big|_{s=t}.$$

The Tonelli Lagrangian naturally induced a functional on the space of weakly differentiable curves. This functional, the action of the Lagrangian, will be interpreted as a cost function.

Definition 2.2.4. Let L be a Tonelli Lagrangian. Let $\sigma \in W^{1,1}$. The **action of L on σ** is

$$A_L(\sigma) = \int_0^1 L(\sigma(t), \dot{\sigma}(t)) dt.$$

Definition 2.2.5. Fix $x, y \in \mathbb{R}^d$. The **cost of transport from x to y induced by L** is

$$c(x, y) = \inf \{ A_L(\sigma) : \sigma \in W^{1,1}, \sigma(0) = x, \sigma(1) = y \}.$$

Since the Tonelli Lagrangian is bounded below, this infimum is guaranteed to be finite. A curve $\sigma \in W^{1,1}$ from x to y which minimizes the action of L is called an **action minimizing curve** or an **optimal trajectory** from x to y . It is not obvious that action minimizing curves exist for general Tonelli Lagrangians. For example, the minimal growth assumptions on L (for instance, $L(x, \cdot)$ may grow slower than $|\cdot|^p$ for any $p > 1$) present technical challenges to demonstrating the existence of action minimizers

(the potentially slow asymptotic growth prevents using the Banach-Alaoglu theorem when trying to prove existence of action minimizers). However, Tonelli Lagrangians do indeed have action minimizing curves.

Theorem 2.2.6 (Existence of Action Minimizers). *Let L be a Tonelli Lagrangian. Fix $x, y \in \mathbb{R}^d$. Let $c(x, y) = \inf\{A_L(\sigma) : \sigma \in W^{1,1}, \sigma(0) = x, \sigma(1) = y\}$. Then, the infimum is achieved. That is, there exists some curve $\bar{\sigma} \in W^{1,1}$ such that $c(x, y) = A_L(\bar{\sigma})$.*

The following characterization of weakly convergent sequences in L^1 from Buttazzo, Giaquinta, and Hildebrandt [7] will be used in the proof of the above theorem.

Theorem 2.2.7 (Theorem 2.11 from Buttazzo, Giaquinta, and Hildebrandt [7]). *Suppose that Ω is a bounded open set in \mathbb{R}^d . Suppose that $(u_k)_{k=1}^\infty$ is a sequence in $L^1(\Omega)$ with the following properties.*

(a) *The sequence $(u_k)_{k=1}^\infty$ is uniformly bounded in $L^1(\Omega)$. That is,*

$$\sup_k \|u_k\|_1 < +\infty.$$

(b) *The sequence of measures $(u_k dx)_{k=1}^\infty$ is equiabsolutely continuous with respect to Lebesgue measure. That is, for all $\epsilon > 0$, there exists $\delta > 0$ such that for every u_k , and for every $E \subset \Omega$ with $m(E) < \delta$,*

$$\int_E |u_k(x)| dx < \epsilon.$$

Then, the sequence $(u_k)_{k=1}^\infty$ has a subsequence which converges weakly in $L^1(\Omega)$. Further, if $(u_k)_{k=1}^\infty$ converges weakly in $L^1(\Omega)$, then properties (a) and (b) hold.

The proof of the existence of action minimizers will proceed by showing the following results. Let $(\sigma_n)_{n=1}^\infty$ be a minimizing sequence for $c(x, y)$.

Step 1: The sequence $(\dot{\sigma}_n)_{n=1}^\infty$ is weakly compact in $L^1((0, 1); \mathbb{R}^d)$.

Step 2: A subsequence $(\sigma_k)_{k=1}^\infty$ of the sequence $(\sigma_n)_{n=1}^\infty$ is uniformly bounded and equicontinuous.

Step 3: The subsequence $(\sigma_k)_{k=1}^\infty$ converges to a limit σ in L^1 (in fact, the convergence will be uniform) and $\dot{\sigma}_k \rightharpoonup \dot{\sigma}$ in L^1 .

Step 4: This limit is a minimizer of the action of L .

Proof. **Step 1:** It suffices to show that $(\dot{\sigma}_n)_{n=1}^\infty$ is bounded in L^1 and is equiabsolutely continuous. Since $(\sigma_n)_{n=1}^\infty$ is a minimizing sequence, by passing to a subsequence, the sequence $(A_L(\sigma_n))_{n=1}^\infty$ may be taken to be non-increasing. Therefore, for every n ,

$$A_L(\sigma_1) \geq A_L(\sigma_n) \geq \int_0^1 c_0 + \theta(\dot{\sigma}_n(t)) dt = c_0 + \int_0^1 \theta(\dot{\sigma}_n(t)) dt.$$

Since $\lim_{|v| \rightarrow \infty} \theta(v)/|v| = +\infty$ and for every n ,

$$\int_0^1 \theta(\dot{\sigma}_n(t)) dt \leq A_L(\sigma_1) - c_0 = K_0,$$

it follows that there exists a constant K_1 such that for every n ,

$$\int_0^1 |\dot{\sigma}_n(t)| dt < K_1 \quad \implies \quad \sup_n \|\dot{\sigma}_n\|_1 < +\infty. \quad (2.1)$$

Fix $\epsilon > 0$. Let M be sufficiently large that

$$\frac{\theta(v)}{|v|} > \frac{K_0}{\epsilon} \quad \text{for all } |v| > M.$$

Then,

$$|\dot{\sigma}_n(t)| < \frac{\epsilon}{K_0} \theta(\dot{\sigma}_n(t)) \quad \text{for all } t \in F_M^n = \{t \in (0, 1) : |\dot{\sigma}_n(t)| > M\}.$$

From this, it follows that

$$\int_{F_M^n} |\dot{\sigma}_n(t)| dt < \frac{\epsilon}{K_0} \int_0^1 \theta(\dot{\sigma}_n(t)) dt = \epsilon.$$

In particular, $m(F_M^n) < \frac{\epsilon}{M}$ for every n , for otherwise, if $m(F_M^n) \geq \frac{\epsilon}{M}$ for some n , then

$$\int_{F_M^n} |\dot{\sigma}_n(t)| dt \geq m(F_M^n) M \geq \epsilon,$$

contradicting

$$\int_{F_M^n} |\dot{\sigma}_n(t)| dt < \epsilon.$$

Therefore, if $m(E) < \frac{\epsilon}{M}$, then for every n ,

$$\begin{aligned} \int_E |\dot{\sigma}_n(t)| dt &= \int_{E \cap F_M^n} |\dot{\sigma}_n(t)| dt + \int_{E \setminus F_M^n} |\dot{\sigma}_n(t)| dt \\ &\leq \int_{F_M^n} |\dot{\sigma}_n(t)| dt + m(E)M \\ &< \epsilon + \frac{\epsilon}{M}M = 2\epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the sequence of measures $(\dot{\sigma}_n dx)_{n=1}^\infty$ is equiabsolutely continuous with respect to Lebesgue measure. By 2.2.7, the sequence $(\dot{\sigma}_n)_{n=1}^\infty$ has a weakly convergent subsequence.

Step 2: Fix $s \in (0, 1)$. Then,

$$\begin{aligned} |\sigma_n(s)| &= \left| \sigma(0) + \int_0^s \dot{\sigma}_n(t) dt \right| \\ &= \left| x + \int_0^s \dot{\sigma}_n(t) dt \right| \\ &\leq |x| + \|\dot{\sigma}_n\|_1 \\ &\leq |x| + K_1. \end{aligned}$$

Therefore,

$$\sup_n \|\sigma_n\|_u < +\infty, \tag{2.2}$$

where $\|\cdot\|_u$ denotes the uniform norm.

Fix $\epsilon > 0$. By the previous part, there exists $\delta > 0$ such that for every set E with $m(E) < \delta$, for all n ,

$$\int_E |\dot{\sigma}_n(t)| dt < \epsilon.$$

If $|r - s| < \delta$ (without loss of generality, assume $r \geq s$), then,

$$\begin{aligned} |\sigma_n(r) - \sigma_n(s)| &= \left| x + \int_0^r \dot{\sigma}_n(t) dt - x - \int_0^s \dot{\sigma}_n(t) dt \right| \\ &= \left| \int_s^r \dot{\sigma}_n(t) dt \right| \\ &< \epsilon. \end{aligned}$$

Since this is independent of n , the sequence $(\sigma_n)_{n=1}^\infty$ is equicontinuous on $[0, 1]$.

Step 3: By step 2, the sequence $(\sigma_n)_{n=1}^\infty$ is uniformly bounded and equicontinuous, so satisfies the hypotheses of the Arzelà-Ascoli theorem. Therefore the sequence has a uniformly convergent subsequence. After possibly passing to a further subsequence, by the results from step one, the uniformly convergent subsequence $(\sigma_{n_k})_{k=1}^\infty$ has derivatives $(\dot{\sigma}_{n_k})_{k=1}^\infty$ which converge weakly in L^1 .

The sequence $\dot{\sigma}_n \rightharpoonup \gamma$ weakly, and $\sigma_n \rightarrow \sigma$ uniformly. The function γ is in fact the velocity of σ . Fix $s \in (0, 1)$. Then,

$$\begin{aligned} \left| x + \int_0^s \gamma(t) dt - \sigma(s) \right| &\leq \left| x + \int_0^s \gamma(t) dt - \sigma_n(s) \right| + |\sigma_n(s) - \sigma(s)| \\ &= \left| \int_0^s \gamma(t) dt - \int_0^s \dot{\sigma}_n(t) dt \right| + |\sigma_n(s) - \sigma(s)|. \end{aligned}$$

Since $\dot{\sigma}_n \rightharpoonup \gamma$ weakly and $\sigma_n \rightarrow \sigma$ uniformly, both summands approach zero as $n \rightarrow \infty$. Therefore,

$$\left| x + \int_0^s \gamma(t) dt - \sigma(s) \right| = 0.$$

Hence, $\gamma = \dot{\sigma}$.

Step 4: Let $F_M = \{t \in (0, 1) : |\dot{\sigma}(t)| > M\}$. Let

$$A = \sup_n \left\{ \|\dot{\sigma}_n - \dot{\sigma}\|_{L^1(0,1)} \right\}.$$

By equation (2.1), together with the fact that $\dot{\sigma} \in L^1$, it follows that $A < +\infty$. Set $M > A$. Then, as $\nabla_v L$ is of class C^1 , it is Lipschitz on

$$\overline{B}_M \times \overline{B}_M,$$

where \overline{B}_M is the closed ball of radius M centred at the origin. Let ℓ_M be the Lipschitz constant.

Then,

$$\begin{aligned} \int_{F_M^c} |(\nabla_v L(\sigma_n, \dot{\sigma}) - \nabla_v L(\sigma, \dot{\sigma})) (\dot{\sigma}_n - \dot{\sigma})| dt &\leq \|\nabla_v L(\sigma_n, \dot{\sigma}) - \nabla_v L(\sigma, \dot{\sigma})\|_{L^\infty(F_M^c)} \|\dot{\sigma}_n - \dot{\sigma}\|_{L^1(0,1)} \\ &\leq \ell_M \|\sigma_n - \sigma\|_u A. \end{aligned} \tag{2.3}$$

Since $\sigma_n \rightarrow \sigma$ uniformly as $n \rightarrow \infty$,

$$Al_M \|\sigma_n - \sigma\|_u \rightarrow 0$$

as $n \rightarrow \infty$. For $t \in F_M^c$,

$$\begin{aligned} L(\sigma_n, \dot{\sigma}_n) &\geq L(\sigma_n, \dot{\sigma}) + \nabla_v L(\sigma_n, \dot{\sigma})(\dot{\sigma}_n - \dot{\sigma}) \\ &= L(\sigma_n, \dot{\sigma}) + (\nabla_v L(\sigma_n, \dot{\sigma}) - \nabla_v L(\sigma, \dot{\sigma}))(\dot{\sigma}_n - \dot{\sigma}) + \nabla_v L(\sigma, \dot{\sigma})(\dot{\sigma}_n - \dot{\sigma}). \end{aligned}$$

Then, using the estimate from equation (2.3),

$$\int_{F_M^c} L(\sigma_n, \dot{\sigma}_n) dt \geq \int_{F_M^c} L(\sigma_n, \dot{\sigma}) dt - Al_M \|\sigma_n - \sigma\|_u + \int_{F_M^c} \nabla_v L(\sigma, \dot{\sigma})(\dot{\sigma}_n - \dot{\sigma}) dt. \quad (2.4)$$

Without loss of generality, it may be assumed that $L \geq 0$. By Fatou's lemma,

$$\liminf_{n \rightarrow \infty} \int_{F_M^c} L(\sigma_n, \dot{\sigma}) dt \geq \int_{F_M^c} L(\sigma, \dot{\sigma}) dt. \quad (2.5)$$

Since $\dot{\sigma}_n \rightharpoonup \dot{\sigma}$ and

$$\nabla_v L(\sigma, \dot{\sigma}) \chi_{F_M^c} \in L^\infty(0, 1),$$

where χ_E denotes the characteristic function of the set E , it follows that for each M ,

$$\lim_{n \rightarrow \infty} \int_0^1 \nabla_v L(\sigma, \dot{\sigma}) \chi_{F_M^c}(\dot{\sigma}_n - \dot{\sigma}) dt = 0. \quad (2.6)$$

Combining the estimates from (2.4), (2.5), and (2.6), and using that $L \geq 0$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_0^1 L(\sigma_n, \dot{\sigma}_n) dt &\geq \liminf_{n \rightarrow \infty} \int_{F_M^c} L(\sigma_n, \dot{\sigma}) dt \\ &\quad + \lim_{n \rightarrow \infty} \left(-Al_M \|\sigma_n - \sigma\|_u + \int_{F_M^c} \nabla_v L(\sigma, \dot{\sigma})(\dot{\sigma}_n - \dot{\sigma}) dt \right) \\ &= \liminf_{n \rightarrow \infty} \int_{F_M^c} L(\sigma_n, \dot{\sigma}) dt \\ &\geq \liminf_{n \rightarrow \infty} \int_{F_M^c} L(\sigma, \dot{\sigma}) dt. \end{aligned}$$

Taking the limit as $M \rightarrow \infty$ yields

$$\liminf_{n \rightarrow \infty} \int_0^1 L(\sigma_n, \dot{\sigma}_n) dt \geq \int_0^1 L(\sigma, \dot{\sigma}) dt.$$

Since $(A_L(\sigma_n))_{n=1}^\infty$ is a minimizing sequence, it follows that σ minimizes the action of L . ■

2.2.1 Lagrangian Systems and Action Minimizing Curves

Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian and let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the induced cost function. Let $x, y \in \mathbb{R}^d$. By theorem 2.2.6, there exists an optimal trajectory $\sigma : [0, 1] \rightarrow \mathbb{R}^d$ with $\sigma(0) = x$ and $\sigma(1) = y$ which minimizes the action of L , so

$$c(x, y) = A_L(\sigma).$$

A priori, the curve σ is merely an element of $W^{1,1}$; it is not obvious that it possesses better regularity properties than this. The main result of this subsection is that σ does, in fact, possess good regularity properties: curves which minimize the action of L are of class C^2 and satisfy the Euler-Lagrange equation. This subsection consists mostly of results from chapters 15 and 16 of Clarke's *Functional Analysis, Calculus of Variations, and Optimal Control* [9].

First, it will be shown that optimal trajectories are Lipschitz continuous. This will be done by verifying that L satisfies a sufficient growth condition:

Definition 2.2.8. A Lagrangian L has **Nagumo growth** along the curve ϕ if there exists a function $\tilde{\theta} : [0, +\infty) \rightarrow \mathbb{R}$ with the properties

- (a) $\lim_{t \rightarrow +\infty} \tilde{\theta}(t)/t = +\infty$
- (b) $L(\phi(t), v) \geq \tilde{\theta}(|v|)$ for all $t \in [0, 1]$ and for all $v \in \mathbb{R}^d$.

Theorem 2.2.9 (Theorem 16.18 from Clarke). *Let $\sigma : [0, 1] \times \mathbb{R}^d$ be an absolutely continuous curve which minimizes the action of L , where the Lagrangian L is continuous, autonomous (no explicit dependence on t), convex in v , and has Nagumo growth along σ . Then, σ is Lipschitz continuous.*

Upon showing that the Tonelli Lagrangians have Nagumo growth along every optimal trajectory, it will follow that optimal trajectories are Lipschitz continuous.

Proposition 2.2.10. *Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian. Then L has Nagumo growth along every curve.*

Proof. By definition 2.2.1, there exists a function $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that

$$L(x, v) \geq c + \theta(v) \quad \text{for all } x, v \in \mathbb{R}^d \quad \text{and} \quad \lim_{|v| \rightarrow +\infty} \frac{\theta(v)}{|v|} = +\infty.$$

Therefore, the function

$$\tilde{\theta}(t) = \inf_{|v| \geq t} \{\theta(v) + c\}$$

satisfies

$$\lim_{t \rightarrow +\infty} \frac{\tilde{\theta}(t)}{t} = +\infty,$$

and for all $x, v \in \mathbb{R}^d$, $L(x, v) \geq \tilde{\theta}(|v|)$, so L has Nagumo growth along every curve. ■

Corollary 2.2.11. *Let $\sigma : [0, 1] \rightarrow \mathbb{R}^d$ be an action minimizing curve for the Tonelli Lagrangian $L : T\mathbb{R}^d \rightarrow \mathbb{R}$. Then σ is Lipschitz continuous.*

Having established that optimal trajectories are locally Lipschitz, it will be seen that they satisfy the Euler-Lagrange equation in a weak sense. This will be “bootstrapped” to the desired regularity.

Definition 2.2.12. A curve $\sigma : [0, 1] \rightarrow \mathbb{R}^d$ satisfies the **integral Euler equation** if there exists a constant $c \in \mathbb{R}^d$ such that for almost every $t \in [0, 1]$,

$$\nabla_v L(\sigma(t), \dot{\sigma}(t)) = c + \int_0^t \nabla_x L(\sigma(s), \dot{\sigma}(s)) ds. \quad (2.7)$$

Theorem 2.2.13 (Theorem 15.2 from Clarke). *Suppose that σ minimizes A_L and σ is Lipschitz continuous. Then, σ satisfies the integral Euler equation.*

Theorem 2.2.14 (Theorem 15.5 from Clarke). *Suppose that σ is Lipschitz continuous, satisfies the integral Euler equation, and for almost every $t \in [0, 1]$, the function $v \mapsto L(\sigma(t), v)$ is strictly convex. Then, σ is of class C^1 .*

Theorem 2.2.15 (Theorem 15.7 from Clarke). *Suppose that σ is Lipschitz continuous and satisfies the integral Euler equation. Suppose the Lagrangian L is of class C^m for $m \geq 2$ and for all $t \in [0, 1]$ and all $v \in \mathbb{R}^d$, the Hessian $\nabla_v^2 L(\sigma(t), v)$ is positive definite. Then, σ is of class C^m .*

Combining corollary 2.2.11 and theorems 2.2.13, 2.2.14, and 2.2.15 yields the desired regularity of optimal trajectories, which is recorded in the following proposition.

Proposition 2.2.16. *Let L be a Tonelli Lagrangian which is strongly convex in the second argument. Let σ be a minimizer for A_L . Then, σ is of class C^2 and satisfies the Euler equation.*

Proof. By corollary 2.2.11, the optimal trajectory σ is Lipschitz continuous.

By theorem 2.2.13, the curve σ satisfies the integral Euler equation. Then, since $L(x, \cdot)$ is strictly convex, by theorem 2.2.14, the curve σ is of class C^1 . Finally, since $\nabla_v^2 L(x, v)$ is positive definite for all $x, v \in \mathbb{R}^d$ and L is of class C^2 , by theorem 2.2.15, the curve σ is also of class C^2 .

Since the curve σ satisfies the integral Euler equation (2.7) and is of class C^2 , differentiating the integral Euler equation with respect to t shows that σ satisfies the Euler-Lagrange equation. ■

2.2.2 Properties of Cost Functions Induced by Lagrangians

Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex in the second argument. By the results thus far, for every $x, y \in \mathbb{R}^d$, there is an action minimizing curve σ_x^y with $\sigma_x^y(0) = x$ and $\sigma_x^y(1) = y$ which is of class C^2 and satisfies the Euler equation; the cost induced by L is

$$c(x, y) = A_L(\sigma_x^y).$$

The goal of this subsection is to establish regularity properties for cost functions induced by Tonelli Lagrangians: they are locally Lipschitz and superdifferentiable. Of particular interest will be the differentiability of the cost function: when the cost function c is differentiable at a point (x, y) , there will be a unique optimal trajectory from x to y . There are additional important properties of cost functions induced by Tonelli Lagrangians, but these will need to wait until the next chapter when more structure from the optimal transport problem has been introduced.

To proceed, some facts about sub- and superdifferentiability are stated.

Definition 2.2.17. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$. The **subdifferential** of f at x is the set

$$\partial f(x) = \{v \in \mathbb{R}^d : f(y) - f(x) \geq v \cdot (y - x) - o(x - y) \forall y\}.$$

If the subdifferential of a function f is non-empty at a point x , the function f is said to be **subdifferentiable** at x . The **superdifferential** of a function f is minus the subdif-

ferential of $-f$. If $-f$ is subdifferentiable at x , then f is said to be **superdifferentiable** at x .

A function is subdifferentiable at a point if it locally has a supporting hyperplane at that point. The following proposition (stated without proof) records a standard fact from convex analysis about the relationship between subdifferentiability, superdifferentiability, and differentiability.

Proposition 2.2.18. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$. The function f is differentiable at a point x if and only if it is both subdifferentiable and superdifferentiable at x . If f is differentiable at x , then both its sub- and superdifferentials at x are singletons and the unique element in its sub- and superdifferential is the derivative of f at x .*

With this in place, the Lipschitz regularity of cost functions induced by Tonelli Lagrangians can be proven.

Proposition 2.2.19. *The cost $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ induced by the Tonelli Lagrangian $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ is locally Lipschitz.*

Proof. Fix $x, y \in \mathbb{R}^d$. Let σ be an optimal path from x to y . Fix $u, v \in \mathbb{R}^d$. For $s > 0$, the curve

$$[0, 1] \ni t \mapsto \sigma(t) + tsv + (1-t)su \in \mathbb{R}^d \quad (2.8)$$

is a (potentially non-optimal) curve from $x + su$ to $y + sv$. Then,

$$\lim_{s \rightarrow 0} \frac{c(x + su, y + sv) - c(x, y)}{s} \quad (2.9)$$

$$\leq \lim_{s \rightarrow 0} \frac{1}{s} \int_0^1 [L(\sigma(t) + tsv + (1-t)su, \dot{\sigma}(t) + sv - su) - L(\sigma(t), \dot{\sigma}(t))] dt \quad (2.10)$$

$$= \int_0^1 \lim_{s \rightarrow 0} \frac{L(\sigma(t) + tsv + (1-t)su, \dot{\sigma}(t) + sv - su) - L(\sigma(t), \dot{\sigma}(t))}{s} dt. \quad (2.11)$$

The inequality in equation (2.10) holds because $t \mapsto \sigma(t) + tsv + (1-t)su$ is a (potentially non-optimal) curve from $x + su$ to $y + sv$. In equation (2.11) the limit may be brought inside the integral because the integrand is of class C^1 . Further, the integrand in (2.11) is a derivative:

$$\begin{aligned} & \lim_{s \rightarrow 0} \frac{L(\sigma(t) + tsv + (1-t)su, \dot{\sigma}(t) + sv - su) - L(\sigma(t), \dot{\sigma}(t))}{s} \\ &= \left. \frac{d}{ds} L(\sigma(t) + tsv + (1-t)su, \dot{\sigma}(t) + sv - su) \right|_{s=0} \end{aligned} \quad (2.12)$$

Substituting equation (2.12) into (2.11) yields

$$\begin{aligned} & \lim_{s \rightarrow 0} \frac{c(x + su, y + sv) - c(x, y)}{s} \\ & \leq \int_0^1 \left. \frac{d}{ds} L(\sigma(t) + tsv + (1-t)su, \dot{\sigma}(t) + sv - su) \right|_{s=0} dt \end{aligned} \quad (2.13)$$

$$= \int_0^1 [\nabla_x L(\sigma(t), \dot{\sigma}(t)) \cdot [tv + (1-t)u] + \nabla_v L(\sigma(t), \dot{\sigma}(t)) \cdot (v - u)] dt \quad (2.14)$$

Since σ solves the Euler-Lagrange equation, $\nabla_x L(\sigma(t), \dot{\sigma}(t)) = \frac{d}{dt} \nabla_v L(\sigma(t), \dot{\sigma}(t))$. Substituting this into (2.14) produces

$$\begin{aligned} & \lim_{s \rightarrow 0} \frac{c(x + su, y + sv) - c(x, y)}{s} \\ & = \int_0^1 \left[\left(\frac{d}{dt} \nabla_v L(\sigma(t), \dot{\sigma}(t)) \right) \cdot [tv + (1-t)u] + \nabla_v L(\sigma(t), \dot{\sigma}(t)) \cdot (v - u) \right] dt \end{aligned} \quad (2.15)$$

$$= \int_0^1 \frac{d}{dt} [\nabla_v L(\sigma(t), \dot{\sigma}(t)) \cdot [tv + (1-t)u]] dt \quad (2.16)$$

$$= \nabla_v L(y, \dot{\sigma}(1)) \cdot v - \nabla_v L(x, \dot{\sigma}(0)) \cdot u \quad (2.17)$$

By fixing u and v to have norm 1 and allowing x and y to vary over a large compact set B , it follows that for every vector v ,

$$\lim_{s \rightarrow 0} \frac{c(x + su, y + sv) - c(x, y)}{s} \leq 2 \|\nabla_v L(w, z)\|_{(w,z) \in B} =: K.$$

Then,

$$\begin{aligned} -K & \leq \lim_{s \rightarrow 0} \frac{c(x, y) - c(x + su, y + sv)}{s} \\ & = \lim_{t \rightarrow 0} \frac{c(x, y + t(-v)) - c(x, y)}{t} \quad (t = -s) \\ & \leq K \end{aligned}$$

Therefore, c is locally Lipschitz. \blacksquare

It follows that cost functions induced by Tonelli Lagrangians are superdifferentiable in each argument.

Corollary 2.2.20. *The cost c induced by the Tonelli Lagrangian L is superdifferentiable in each argument.*

Proof. From the equation (2.17), if σ is an action minimizing curve from x to y , then

$\nabla_v L(y, \dot{\sigma}(1))$ is a superderivative of $c(x, \cdot)$ at y and $-\nabla_v L(x, \dot{\sigma}(0))$ is a superderivative for $c(\cdot, y)$ at x . ■

Since cost functions induced by Tonelli Lagrangians are locally Lipschitz, by applying Rademacher's theorem, it immediately follows that the cost function c is differentiable everywhere.

Corollary 2.2.21. *The cost c induced by the Tonelli Lagrangian L is differentiable almost everywhere.*

Proof. By the previous proposition, the cost function is a locally Lipschitz function from $\mathbb{R}^d \times \mathbb{R}^d$ to \mathbb{R} . By Rademacher's theorem, a locally Lipschitz function from \mathbb{R}^n to \mathbb{R}^m is differentiable almost everywhere. ■

Corollary 2.2.22. *If the cost function c induced by the Tonelli Lagrangian L is differentiable in the second argument, its derivative is*

$$\frac{d}{dy}c(x, y) = \nabla_v L(y, \dot{\sigma}(1)),$$

where σ is an optimal curve from x to y . If c is differentiable in the first argument, its derivative is

$$\frac{d}{dx}c(x, y) = -\nabla_v L(x, \dot{\sigma}(0)).$$

Proof. Let σ be an optimal curve from x to y . By corollary 2.2.20, $\nabla_v L(y, \dot{\sigma}(1))$ is a superderivative for $c(x, \cdot)$ at y and $-\nabla_v L(x, \dot{\sigma}(0))$ is a superderivative for $c(\cdot, y)$ at x . By 2.2.21, c is differentiable almost everywhere. ■

The differentiability of the cost function c induced by the Tonelli Lagrangian L leads to a uniqueness result about action minimizing curves of L .

Proposition 2.2.23. *For any pair x, y such that either $\frac{d}{dx}c(x, y)$ or $\frac{d}{dy}c(x, y)$ exists, there exists a unique action minimizing curve from x to y .*

In particular, for each $x \in \mathbb{R}^d$, $c(x, \cdot)$ is differentiable for almost every $y \in \mathbb{R}^d$, and for each $y \in \mathbb{R}^d$, $c(\cdot, y)$ is differentiable for almost every $x \in \mathbb{R}^d$

Proof. Let (x, y) be a point at which $\frac{d}{dy}c(x, y)$ exists. Suppose that σ and γ are both action minimizing curves from x to y . Then, by proposition 2.2.22,

$$\nabla_y c(x, y) = \nabla_v L(y, \dot{\sigma}(1)) = \nabla_v L(y, \dot{\gamma}(1)).$$

By lemma 2.3.5 and proposition 2.4.6, the function $\nabla_v L(y, \cdot)$ is injective, so $\dot{\sigma}(1) = \dot{\gamma}(1)$. Hence, the curves γ and σ both satisfy the Euler-Lagrange equation with final condition

$$(\sigma(1), \dot{\sigma}(1)) = (\gamma(1), \dot{\gamma}(1)).$$

By uniqueness of solutions to the Lagrangian flow, $\sigma = \gamma$. An analogous argument holds for the case where $\frac{d}{dx}c(x, y)$ exists. ■

2.3 Hamiltonians

The dual object to a Lagrangian on the tangent bundle of a manifold is a function on the cotangent bundle, called a Hamiltonian. The Hamiltonian corresponding to a Tonelli Lagrangian will be defined via the Legendre transform, and the usual properties (for instance, those seen in a standard classical mechanics course) will be derived.

Let E be a normed vector space and let E^* be its dual.

Definition 2.3.1. Let $f : E \rightarrow \mathbb{R} \cup \{\infty\}$ which is not identically $+\infty$. The **Legendre transform** or **convex conjugate** of f is the map $f^* : E^* \rightarrow \mathbb{R} \cup \{\infty\}$ given by

$$f^*(x) = \sup_{y \in E} \{\langle y, x \rangle - f(y)\}.$$

An important fact about Legendre transforms is given by the Fenchel-Moreau theorem, quoted from [6].

Theorem 2.3.2 (Fenchel-Moreau). *Let E be a normed vector space and let $f : \mathbb{R} \cup \{\infty\}$. Suppose that f is convex, lower semi-continuous, and $f \not\equiv +\infty$. Then, $f^{**}|_E = f$.*

Let $L : T\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian. Assume that L is strongly convex in the second argument, so $\nabla_v^2 L(x, v)$ is positive definite at every point $(x, v) \in T\mathbb{R}^d$.

Definition 2.3.3. The **Hamiltonian** corresponding to L is the function $H : T^*\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$H(q, p) := \sup_{v \in \mathbb{R}^d} \{p \cdot v - L(q, v)\}.$$

That is, H is the Legendre transform of L with respect to velocity.

The first argument in a Hamiltonian will often be referred to as **position** and the second argument will often be referred to as **momentum**. For example, if H is a Hamiltonian, then $H(q, p)$ is the Hamiltonian evaluated at position q and momentum p . The derivatives and gradients of a Hamiltonian in the position and momentum variables will be denoted $D_x H$, $D_p H$, $\nabla_x H$, and $\nabla_p H$.

Proposition 2.3.4. *Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian. Let $H : T^*\mathbb{R}^d \rightarrow \mathbb{R}$ be the corresponding Hamiltonian. Then,*

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

Proof. Tonelli Lagrangians are convex with respect to velocity, are continuous, and are not identically infinite. For each x , the Hamiltonian $H(x, \cdot)$ is the Legendre transform of $L(x, \cdot)$. By the Fenchel-Moreau theorem, the Legendre transform of H in the second variable is therefore L . ■

Since L is of class C^2 , is strictly convex, and has superlinear growth, the maximization problem $\sup_{v \in \mathbb{R}^d} \{p \cdot v - L(q, v)\}$ can be solved with calculus:

$$H(q, p) = p \cdot r - L(q, r) \quad \text{where} \quad p = \nabla_v L(q, r). \quad (2.18)$$

An application of the implicit function theorem will show that the variable r in equation (2.18) is of class C^1 , which will be the first step in deriving the classical structure of the Hamiltonian corresponding to a Lagrangian; this will require a useful lemma about the injectivity of gradients of convex functions.

Lemma 2.3.5. *Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be function of class C^2 with a positive definite Hessian. Then, the gradient $\nabla\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is injective.*

Proof. Fix $x \neq y \in \mathbb{R}^d$. Define $f : \mathbb{R} \rightarrow \mathbb{R}$ by $f(t) = \varphi((1-t)x + ty)$. The function f is strictly convex:

$$f''(t) = (x - y)^T \nabla^2 \varphi((1-t)x + ty)(x - y),$$

and since $\nabla^2 \varphi$ is positive definite, the second derivative $f''(t) > 0$. Then, $f' : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing, so $f'(0) < f'(1)$, and

$$f'(0) = \nabla\varphi(x)(x - y) < \nabla\varphi(y)(x - y) = f'(1).$$

Hence, $\nabla\varphi(x) \neq \nabla\varphi(y)$. ■

Proposition 2.3.6. *The variable r in the formula for $H(q, p)$ in equation (2.18) is a C^1 function of q and p .*

Proof. Let $F(q, p, r) = p - \nabla_v L(q, r)$. The function F is of class C^1 , since L is of class C^2 . Then,

$$\nabla_r F(q, p, r) = -\nabla_v^2 L(q, r).$$

By assumption, $\nabla_v^2 L(q, r)$ is positive definite. Hence it is invertible. The hypotheses of the implicit function theorem are therefore satisfied. Therefore, r can be written locally as a C^1 function of q and p : $r = \phi(q, p)$.

By lemma 2.3.5, for every $x \in \mathbb{R}^d$, the function $\nabla_v L(x, \cdot)$ is injective. It follows that, given p and q , there is at most one vector r satisfying $p = \nabla_v L(q, r)$.

Taken together, it follows that r can be written as a C^1 function of q and p . ■

Proposition 2.3.7. *The Hamiltonian $H(q, p)$ is of class C^2 .*

Proof. The Hamiltonian can be written as

$$H(q, p) = p \cdot r - L(q, r) \quad \text{where} \quad r = \phi(q, p)$$

and ϕ is of class C^1 . Then,

$$\nabla_x H(q, p) = p \cdot \nabla_x \phi(q, p) - \nabla_x L(q, \phi(q, p)) - \nabla_v L(q, \phi(q, p)) \cdot \nabla_x \phi(q, p).$$

Since $p = \nabla_v L(q, r) = \nabla_v L(q, \phi(q, p))$,

$$\nabla_x H(q, p) = p \cdot \nabla_x \phi(q, p) - \nabla_x L(q, \phi(q, p)) - p \cdot \nabla_x \phi(q, p) = -\nabla_x L(q, \phi(q, p)).$$

Similarly,

$$\nabla_p H(q, p) = \phi(q, p) + p \cdot \nabla_x \phi(q, p) - \nabla_x L(q, \phi(q, p)) \cdot \nabla_x \phi(q, p) = \phi(q, p). \quad (2.19)$$

Since ϕ is of class C^1 and L is of class C^2 , it follows that $\nabla_x H$ and $\nabla_p H$ are both of class C^1 , so H is of class C^2 . ■

Corollary 2.3.8. *Let $q, r \in \mathbb{R}^d$. Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian and let H be the corresponding Hamiltonian. Then*

$$\nabla_p H(q, \nabla_v L(x, r)) = r.$$

That is, $\nabla_p H(q, \cdot) \circ \nabla_v L(q, \cdot) = id$.

Proof. From the proof of proposition 2.3.7, in equation (2.19), the variables p and r satisfy $p = \nabla_v L(q, \phi(q, p))$ and $r = \phi(q, p)$. Substituting these into equation (2.19) yields $\nabla_p H(q, \nabla_v L(x, r)) = r$. ■

By the Fenchel-Moreau theorem, the function $L(x, \cdot)$ is the Legendre transform of $H(x, \cdot)$. It can be checked that $\nabla_v L(x, \cdot)$ is also a left inverse to $\nabla_v L(x, \cdot)$, in addition to a right inverse. With this, the Legendre transform may be recast as a C^1 diffeomorphism:

Definition 2.3.9. Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian. The **global Legendre transform** for L is the map $\mathcal{L} : T\mathbb{R}^d \rightarrow T^*\mathbb{R}^d$ given by

$$\mathcal{L}(x, v) = (x, \nabla_v L(x, v)).$$

Proposition 2.3.10. *The global Legendre transform is a C^1 -diffeomorphism from $T\mathbb{R}^d$ to $T^*\mathbb{R}^d$.*

Proof. The map $\mathcal{L}^{-1} : T^*\mathbb{R}^d \rightarrow T\mathbb{R}^d$ given by

$$(x, p) \mapsto (x, \nabla_p H(x, p))$$

satisfies $\mathcal{L} \circ \mathcal{L}^{-1} = id$ and $\mathcal{L}^{-1} \circ \mathcal{L} = id$. From corollary 2.3.8 and the following remark, the Hamiltonian H is of class C^2 and Tonelli Lagrangians are, by definition, of class C^2 . Therefore, both \mathcal{L} and \mathcal{L}^{-1} are of class C^1 . ■

2.4 Lagrangian and Hamiltonian Systems

Let $L : T\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex in the second argument. Let $H : T^*\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be the corresponding Hamiltonian.

Definition 2.4.1. The **Lagrangian System** is the system of differential equations

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v). \end{cases} \quad (2.20)$$

The differential equation $\frac{d}{dt} \nabla_v L(x, \dot{x}) = \nabla_x L(x, \dot{x})$ is called the **Euler-Lagrange Equation**.

Definition 2.4.2. The **Lagrangian flow from time s to time t** is the map $\Phi_{s,t} : T\mathbb{R}^d \rightarrow T\mathbb{R}^d$ given by

$$\Phi_{s,t}(x, v) = (\sigma(t), \dot{\sigma}(t)),$$

where σ solves the Lagrangian system with time s initial conditions

$$(\sigma(s), \dot{\sigma}(s)) = (x, v).$$

Definition 2.4.3. The **Hamiltonian System** is the system of differential equations

$$\begin{cases} \dot{q} = \nabla_p H(q, p) \\ \dot{p} = -\nabla_x H(q, p) \end{cases}$$

Definition 2.4.4. The **Hamiltonian flow from time s to time t** is the map $\Psi_{s,t} : T^*\mathbb{R}^d \rightarrow T^*\mathbb{R}^d$ given by

$$\Psi_{s,t}(x, r) = (q(t), p(t)),$$

where (q, p) solves the Hamiltonian system with time s initial conditions

$$(q(s), p(s)) = (x, r).$$

The main goal of this section is to show that these systems are equivalent in a precise sense and they both have solutions of class C^2 .

2.4.1 Hamiltonian Systems Have Unique Local Solutions of Class C^2

The Hamiltonian system

$$\begin{cases} \dot{q} = \nabla_p H(q, p) \\ \dot{p} = -\nabla_x H(q, p) \\ (q, p)(0) = (q_0, p_0) \end{cases}$$

is a system of first order ordinary differential equations. Write $y = \begin{pmatrix} q \\ p \end{pmatrix}$ and $F(y) = \begin{pmatrix} \nabla_p H(q, p) \\ -\nabla_x H(q, p) \end{pmatrix}$. Since H is C^2 , the first order ordinary differential equation $\dot{y} = F(y)$ has F locally Lipschitz. Therefore, this system satisfies the hypotheses of the Picard-Lindelöf

theorem, so there exists a unique local solution (q, p) , which is of class C^1 (that the functions are of class C^1 is given by the theorem). It immediately follows that the solution (q, p) is in fact of class C^2 , since

$$\dot{q} = \nabla_p H(q, p) \quad \text{and} \quad \dot{p} = -\nabla_x H(q, p)$$

are both of class C^1 .

2.4.2 Equivalence of Hamiltonian and Lagrangian Systems

Fact. Suppose that $\varphi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex function on a normed vector space X . Let $\varphi^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ denote its Legendre-Fenchel transform. Then,

$$\varphi(a) + \varphi^*(b) = \langle a, b \rangle \iff b \in \partial\varphi(a) \iff a \in \partial\varphi^*(b),$$

where $\partial f(x)$ denotes the subdifferential of f at x . (If $X = \mathbb{R}^d$ and f is differentiable at x , then $y \in \partial f(x)$ means that $y = \nabla f(x)$.)

Proposition 2.4.5. *Suppose that (q, p) is a C^2 solution to the Hamiltonian system. Then, q is a C^2 solution to the Lagrangian system (with compatible initial conditions).*

Proof. Suppose that (q, p) is a C^2 solution to the Hamiltonian system. (Solutions are necessarily of class C^2 by the previous subsection). Then, by the results of the section on Hamiltonians,

$$\begin{aligned} H(q, p) &= p \cdot \nabla_p H(q, p) - L(q, \nabla_p H(q, p)) \\ &= p \cdot \dot{q} - L(q, \dot{q}). \end{aligned}$$

The second equality follows by the assumption that (q, p) is a solution to the Hamiltonian system, so $\dot{q} = \nabla_p H(q, p)$. Then,

$$H(q, p) + L(q, \dot{q}) = p \cdot \dot{q}.$$

By the above fact about Legendre transforms, it follows that

$$p = \nabla_v L(q, \dot{q}) \implies \dot{p} = \frac{d}{dt} \nabla_v L(q, \dot{q}).$$

And, from the computation of the gradients of H in the proof of proposition??,

$$\nabla_x H(q, p) = -\nabla_x L(q, \dot{q}).$$

Then, since (q, p) is a solution to the Hamiltonian system,

$$\dot{p} = -\nabla_x H(q, p) = \nabla_x L(q, \dot{q}).$$

Therefore, $\nabla_x L(q, \dot{q}) = \frac{d}{dt} \nabla_v L(q, \dot{q})$, so q is a solution to the Lagrangian system. ■

Proposition 2.4.6. *Suppose q is of class C^2 and is a solution to the Lagrangian system. Then, $(q, p = \nabla_v L(q, \dot{q}))$ is a solution to the Hamiltonian system with corresponding initial conditions.*

Proof. Suppose that q is of class C^2 and satisfies $\nabla_x L(q, \dot{q}) = \frac{d}{dt} \nabla_v L(q, \dot{q})$. Since q is of class C^2 , it ensures that the right hand side is defined, as

$$\frac{d}{dt} \nabla_v L(q, \dot{q}) = \nabla_x \nabla_v L(q, \dot{q}) \cdot \dot{q} + \nabla_v^2 L(q, \dot{q}) \cdot \ddot{q}.$$

The Hamiltonian H satisfies

$$H(q, p) = p \cdot r - L(q, r) \quad \text{where } p = \nabla_v L(q, r) \text{ and } r = \nabla_p H(q, p).$$

Set $p = \nabla_v L(q, \dot{q})$. We wish to show that $(q, \nabla_v L(q, \dot{q}))$ is a solution to the Hamiltonian system. Then,

$$\nabla_v L(q, r) = \nabla_v L(q, \dot{q}).$$

Since L is of class C^2 and $\nabla_v^2 L$ is positive definite, by the above lemma, $\nabla_v L(q, \cdot)$ is injective, and it follows that $\dot{q} = r$. From the previous proposition and the proof that H is of class C^2 ,

$$\nabla_x H(q, p) = -\nabla_x L(q, r = \dot{q}).$$

Therefore,

$$\begin{aligned} \dot{p} &= \frac{d}{dt} \nabla_v L(q, \dot{q}) && \text{(We set } p = \nabla_v L(q, \dot{q})) \\ &= \nabla_x L(q, \dot{q}) && \text{(} q \text{ solves the Lagrangian system)} \\ &= -\nabla_x H(q, p). \end{aligned}$$

And, $\dot{q} = r = \nabla_p H(q, p)$. Therefore, $(q, \nabla_v L(q, \dot{q}))$ is a solution to the Hamiltonian system. ■

2.4.3 Conservation of Hamiltonians

Proposition 2.4.7. *Let (q, p) be a solution to the Hamiltonian system. Then, for all t where the solution is defined,*

$$H(q(t), p(t)) = H(q(0), p(0)).$$

Proof. For all time t where the solution (q, p) is defined,

$$\begin{aligned} \frac{d}{dt}H(q, p) &= \nabla_x H(q, p) \cdot \dot{q} + \nabla_p H(q, p) \cdot \dot{p} \\ &= \nabla_x H(q, p) \cdot \nabla_p H(q, p) + \nabla_p H(q, p) \cdot (-\nabla_x H(q, p)) \\ &= 0. \end{aligned}$$

Therefore, $H(q(t), p(t))$ is constant. ■

Assume that there exists a constant c_1 and a function $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$, where $\lim_{|p| \rightarrow +\infty} \psi(p) = +\infty$, and

$$H(q, p) \geq c_1 + \psi(p).$$

If (q, p) is a solution to the Hamiltonian system, then H is constant along the solution, so, for all time t where a solution is defined, $p(t)$ is bounded. Otherwise, the above assumption would lead to a contradiction.

2.4.4 Dependence on Initial Conditions

Proposition 2.4.8. *The Lagrangian flow $\Phi_{s,t} : T\mathbb{R}^d \rightarrow T\mathbb{R}^d$ is of class C^1 .*

This does not follow from the standard ODE theory; Tonelli Lagrangians are only of class C^2 , so the left hand side of the Euler-Lagrange equation is, a priori, only continuous.

Proof. By standard results from the theory of ordinary differential equations, (see, for instance, theorem 17.9 in [25]), the Hamiltonian flow $\Psi_{s,t}$ is of class C^1 . Since the Lagrangian flow satisfies

$$\Phi_{s,t} = \mathcal{L}^{-1} \circ \Psi_{s,t} \circ \mathcal{L} : T\mathbb{R}^d \rightarrow T\mathbb{R}^d,$$

where \mathcal{L} is the global Legendre transform, which is a C^1 diffeomorphism by proposition 2.3.10, it follows that the Lagrangian flow is of class C^1 . ■

A computation of the spatial derivative of the Hamiltonian flow will later prove useful.

Proposition 2.4.9. *Let $\pi_q : T^*\mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\pi_p : T^*\mathbb{R}^d \rightarrow \mathbb{R}^d$ be projections given by $\pi_q(q, p) = q$ and $\pi_p(q, p) = p$. Then,*

$$\pi_q D_2 \Psi_{s,t}(q, p) = (t - s) \nabla_p^2 H(q, p) + o(t - s).$$

Proof. The Hamiltonian flow $\Psi_{s,t}$ solves the differential equation

$$\dot{\Psi}_{s,t}(q, p) = \begin{pmatrix} \nabla_p H(\Psi_{s,t}(q, p)) \\ -\nabla_x H(\Psi_{s,t}(q, p)) \end{pmatrix}$$

By standard results from ODE theory (see, for instance, [22]), the derivative $D\Psi_{s,t}$ satisfies the differential equation

$$\dot{z} = \begin{pmatrix} \nabla_p \nabla_x H(\Psi_{s,t}(q, p)) & \nabla_p^2 H(\Psi_{s,t}(q, p)) \\ -\nabla_x^2 H(\Psi_{s,t}(q, p)) & -\nabla_p \nabla_x H(\Psi_{s,t}(q, p)) \end{pmatrix} z \quad (2.21)$$

where z is a $2d \times 2d$ matrix. Let $(\xi, \eta) \in \mathbb{R}^d \times \mathbb{R}^d$. Then, the directional derivative of $\Psi_{s,t}$ in the direction (ξ, η) is given by

$$D\Psi_{s,t} \begin{pmatrix} \xi \\ \eta \end{pmatrix},$$

and these directional derivatives continue to solve the linearized differential equation (2.21).

Note that $\Psi_{s,s} = id$, so $D\Psi_{s,s} = I$. Substituting $z = D\Psi_{s,s}$ into (2.21) yields

$$D\dot{\Psi}_{s,s} = \begin{pmatrix} \nabla_p \nabla_x H(\Psi_{s,t}) & \nabla_p^2 H(\Psi_{s,t}) \\ -\nabla_x^2 H(\Psi_{s,t}) & -\nabla_p \nabla_x H(\Psi_{s,t}) \end{pmatrix} D\Psi_{s,s} \quad (2.22)$$

$$= \begin{pmatrix} \nabla_p \nabla_x H(\Psi_{s,t}) & \nabla_p^2 H(\Psi_{s,t}) \\ -\nabla_x^2 H(\Psi_{s,t}) & -\nabla_p \nabla_x H(\Psi_{s,t}) \end{pmatrix} \quad (2.23)$$

Taylor expanding $D\Psi_{s,t}$ in t around $t = s$ yields,

$$D\Psi_{s,t} = D\Psi_{s,s} + (t - s) D\dot{\Psi}_{s,s} + o(t - s). \quad (2.24)$$

Substituting equation (2.23) into equation (2.24):

$$D\Psi_{s,t} = D\Psi_{s,s} + (t-s) \begin{pmatrix} \nabla_p \nabla_x H(\Psi_{s,t}) & \nabla_p^2 H(\Psi_{s,t}) \\ -\nabla_x^2 H(\Psi_{s,t}) & -\nabla_p \nabla_x H(\Psi_{s,t}) \end{pmatrix} + o(t-s). \quad (2.25)$$

The above equation holds true when multiplied by any vector in $\mathbb{R}^d \times \mathbb{R}^d$. Let $\pi_x(x, p) = x$ and $\pi_p(x, p) = p$. Multiplying equation (2.25) by a vector $(0, \eta) \in \mathbb{R}^d \times \mathbb{R}^d$, remembering that $D\Psi_{s,s} = I$, yields

$$\begin{pmatrix} \pi_q D_q \Psi_{s,t} & \pi_q D_p \Psi_{s,t} \\ \pi_p D_q \Psi_{s,t} & \pi_p D_p \Psi_{s,t} \end{pmatrix} \begin{pmatrix} 0 \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ \eta \end{pmatrix} + (t-s) \begin{pmatrix} \nabla_p^2 H(\Psi_{s,t}) \cdot \eta \\ -\nabla_p \nabla_x H(\Psi_{s,t}) \cdot \eta \end{pmatrix} + o(t-s).$$

Restricting to the first d coordinates, it follows that

$$\pi_q D_p \Psi_{s,t}(p, q) = (t-s) \nabla_p^2 H(\Psi_{s,t}(p, q)) + o(t-s). \quad \blacksquare$$

Chapter 3

Five Formulations of the Optimal Transport Problem

3.1 Introduction

In this chapter, the optimal transport problem is introduced. First, the Monge, Kantorovich, and Kantorovich dual problems are presented, together with criteria for when their infima are equal. These results are standard, and are cited from various sources; some simple results are proven to give a feeling for the arguments involved in proving the equality of the infima of these three costs. In this initial section, many of the results are cited from [38],[31], and [39].

After recording these results, the Lagrangian formulation of the optimal transport problem is stated. A careful analysis of the relationships between the optimal transport problem relates the optimal trajectories in the Lagrangian problem to both the Monge problem and the Kantorovich dual problems. Following easily from the results of the previous chapter, optimal trajectories in the Lagrangian problem will be seen to be solutions of boundary value problem with boundary conditions given by the solution to the Monge problem. Complementary to this, but much more subtle, optimal trajectories in the Lagrangian problem will also be seen to be solutions of an initial value problem with initial conditions given by the Kantorovich dual problem.

A careful analysis of the relationships between the optimal trajectories and the corresponding potential functions from the Kantorovich dual problem establishes the best possible regularity results for optimal trajectories. The key tool in this analysis is Mather's shortening lemma, which gives a quantitative estimate on the maximum distance between optimal trajectories in terms of the distance between optimal trajectories at their

midpoints.

Once the regularity of the optimal trajectories in the Lagrangian formulation of the problem is established, the Eulerian formulation of the optimal transport problem is introduced. The regularity of the optimizers in the Eulerian problem will follow directly from the regularity of the optimal trajectories from the previous section.

The chapter concludes with a proof of the equivalence of the five forms of the optimal transport problem. In particular, the relationships between the minimizers is made explicit. These results are known to experts ([4],[16]), but I believe this is the first time these relationships have been presented this simply.

3.2 Some General Optimal Transport Theory

Background on the optimal transport problem is presented in this section. The Monge, Kantorovich, and Kantorovich dual optimal transport problems are defined, together with criteria for their infima to agree. This section mostly follows McCann-Guillen’s lecture notes [31], Villani’s *Topics in Optimal Transportation* [38], and Fathi-Figalli [16].

Let $\mathcal{P}(\mathbb{R}^d)$ denote the space of Borel probability measures on \mathbb{R}^d . Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$. The function c is called a **cost function**.

Definition 3.2.1. A Borel measurable map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is said to **push forward μ to ν** , denoted $T_{\#}\mu = \nu$, if for all Borel subsets $B \subset \mathbb{R}^d$,

$$\mu(T^{-1}(B)) = \nu(B).$$

Definition 3.2.2. The **Monge formulation of the optimal transport problem from μ to ν for the cost c** is the minimization problem

$$\min_T \{\text{cost}(T)\} := \min_T \left\{ \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\}.$$

Definition 3.2.3. A Borel measurable map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ which minimizes $\text{cost}(T)$ and pushes forward μ to ν is called an **optimal transport map from μ to ν for the cost c** or a **Monge map from μ to ν for the cost c** .

Definition 3.2.4. Let $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. The measure γ has **first marginal μ** and **second marginal ν** if for all Borel subsets $B \subset \mathbb{R}^d$,

$$\gamma(B \times \mathbb{R}^d) = \mu(B) \quad \text{and} \quad \gamma(\mathbb{R}^d \times B) = \nu(B).$$

The set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν is denoted $\Gamma(\mu, \nu)$. If μ and ν are absolutely continuous with respect to Lebesgue measure and have Radon-Nikodym derivatives ρ_0 and ρ_1 , respectively, then this set is denoted $\Gamma(\rho_0, \rho_1)$.

Definition 3.2.5. The **Kantorovich formulation of the optimal transport problem from μ to ν for the cost c** is the minimization problem

$$\min_{\gamma} \{\text{cost}(\gamma)\} := \min_{\gamma} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\}.$$

Definition 3.2.6. A measure $\gamma \in \Gamma(\mu, \nu)$ which minimizes $\text{cost}(\gamma)$ is called an **optimal transport plan from μ to ν for the cost c** .

Definition 3.2.7. The **(Kantorovich) cost from μ to ν for the cost c** is

$$W_c(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y) \right\}.$$

Definition 3.2.8. The **Monge cost from μ to ν for the cost c** is

$$W_c^M(\mu, \nu) = \inf_T \left\{ \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\}.$$

Unless otherwise specified, the cost from μ to ν for the cost c refers to the Kantorovich cost.

Proposition 3.2.9. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$. Then, the Kantorovich cost from μ to ν for the cost c is no larger than the Monge cost. That is,*

$$W_c(\mu, \nu) \leq W_c^M(\mu, \nu).$$

Proof. Let T be a Borel map pushing forward μ to ν . The corresponding measure $\gamma_T = (\text{id} \times T)_{\#}\mu$ has the following properties:

(a) $\gamma_T \in \Gamma(\mu, \nu)$.

(b) $\text{cost}(\gamma_T) = \text{cost}(T)$.

For (a), let $B \subset \mathbb{R}^d$ be a Borel subset. Then,

$$\begin{aligned}\gamma_T(B \times \mathbb{R}^d) &= \mu((\text{id} \times T)^{-1}(B \times \mathbb{R}^d)) \\ &= \mu(\text{id}^{-1}(B) \cap T^{-1}(\mathbb{R}^d)) \\ &= \mu(B),\end{aligned}$$

and,

$$\begin{aligned}\gamma_T(\mathbb{R}^d \times B) &= \mu((\text{id} \times T)^{-1}(\mathbb{R}^d \times B)) \\ &= \mu(\text{id}^{-1}(\mathbb{R}^d) \cap T^{-1}(B)) \\ &= \mu(T^{-1}(B)) \\ &= \nu(B),\end{aligned}$$

where the last equality holds since T is assumed to push forward μ to ν .

For (b), note that for Borel subset $A, B \subset \mathbb{R}^d$,

$$\begin{aligned}\gamma_T(A \times B) &= \mu((\text{id} \times T)^{-1}(A \times B)) \\ &= \mu(A \cap T^{-1}(B)) \\ &= \mu(\{x \in A \mid T(x) \in B\}).\end{aligned}$$

Therefore,

$$\begin{aligned}\text{cost}(\gamma_T) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma_T(x, y) \\ &= \int_{\{(x, y) \mid y=T(x)\}} c(x, y) d\gamma_T(x, y) + \int_{\{(x, y) \mid y=T(x)\}^c} c(x, y) d\gamma_T(x, y).\end{aligned}$$

Since $\{(x, y) \mid y = T(x)\}^c$ is γ_T -measure 0,

$$\begin{aligned}\text{cost}(\gamma_T) &= \int_{\{(x, y) \mid y=T(x)\}} c(x, y) d\gamma_T(x, y) \\ &= \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) \\ &= \text{cost}(T).\end{aligned}$$

Finally, let $(T_n)_{n=1}^\infty$ be a minimizing sequence in the Monge problem. Then,

$$W_c(\mu, \nu) \leq \lim_{n \rightarrow \infty} \text{cost}(\gamma_{T_n}) = \lim_{n \rightarrow \infty} \text{cost}(T_n) = W_c^M(\mu, \nu),$$

concluding the proof. ■

3.2.1 Equivalence of the Monge and Kantorovich Problems

Under appropriate hypotheses, the Monge and Kantorovich costs coincide.

Definition 3.2.10. A cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty)$ is said to be **left-twisted** or to satisfy the **left twist condition** if the map

$$(x, y) \mapsto (x, \nabla_x c(x, y))$$

defined on the domain $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \nabla_x c(x, y) \text{ exists}\}$ is injective.

When there is no ambiguity, the left twist condition will be referred to as the twist condition. The right twist condition refers to the injectivity of the map

$$(x, y) \mapsto (\nabla_y c(x, y), y).$$

If a cost function c satisfies the twist condition, then given $(x, p) \in T^*\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d$, there is at most one $y \in \mathbb{R}^d$ satisfying $\nabla_x c(x, y) + p = 0$. If such a y exists, it is denoted $Y(x, p)$.

The cost functions considered in the first chapter, those induced by Tonelli Lagrangians, satisfy the twist condition, as recorded in the following theorem from Fathi-Figalli [16].

Proposition 3.2.11 (Proposition 4.1 from Fathi-Figalli). *Let L be a Tonelli Lagrangian which is strongly convex with respect to velocity. Then, the cost function induced by L is continuous, bounded below, satisfies the twist condition, and the family of functions $x \mapsto c_y(x) = c(x, y)$ is locally semi-concave in x locally uniformly in y .*

The twist condition plays an important role in ensuring that optimizers to the Monge optimal transport problem exist and that the Monge and Kantorovich costs coincide. The following theorem from McCann-Guillen [31] states a sufficient condition for the Monge and Kantorovich costs and minimizers to coincide in the classical setting. (N.B. They cite this theorem from Gangbo’s habilitation thesis, but it is in French, so I cannot read it.)

Theorem 3.2.12 (Theorem 2.9 from McCann-Guillen). *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. Suppose that μ is absolutely continuous with respect to Lebesgue measure. Let $c \in C(\mathbb{R}^d \times \mathbb{R}^d)$ be differentiable in the first d coordinates and satisfy the twist condition. Suppose that*

$\nabla_x c(x, y)$ is locally bounded in x and globally bounded in y . Then, there exists a Lipschitz function $u : \mathbb{R}^d \rightarrow \mathbb{R}$ such that

- (a) The map $G(x) = Y(x, Du(x))$ pushes forward μ to ν . (The map $Du(x)$ is defined almost everywhere.)
- (b) The map G is the unique minimizer to Monge's formulation of the optimal transport problem.
- (c) The measure $\gamma_G = (id \times G)_\# \mu$ is the unique minimizer to Kantorovich's formulation of the optimal transport problem.

In the proof that $W_c(\mu, \nu) \leq W_c^M(\mu, \nu)$, it was shown that $\text{cost}(G) = \text{cost}(\gamma_G)$.

A more delicate sufficient condition is given in Fathi-Figalli [16].

Theorem 3.2.13 (Theorem 3.2 from Fathi-Figalli). *Let μ and ν be probability measures on \mathbb{R}^d . Suppose that the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is lower semicontinuous and bounded below. Suppose the following conditions hold:*

- (a) The family of maps $x \mapsto c(x, y) = c_y(x)$ is locally semi-concave in x locally uniformly in y ,
- (b) The cost c satisfies the twist condition,
- (c) The measure μ is absolutely continuous with respect to Lebesgue measure,
- (d) The cost from μ to ν for the cost c is finite.

Then, there exists a Borel optimal transport map T from μ to ν for the cost c , this map is unique μ -a.e., and any optimal transport plan $\gamma \in \Gamma(\mu, \nu)$ from μ to ν for the cost c is concentrated on the graph of T .

The definition for a function being locally semi-concave in x locally uniformly in y is given in appendix A of Fathi-Figalli (in particular, see definitions A.1, A.3, A.10, A.13, and A.15). Fathi-Figalli actually use a slightly more general hypothesis than (c), requiring only that the measure μ does not give mass to “small sets”. Finally, because any optimal transport plan has its cost concentrated on the graph of the optimal transport map, it follows that the Monge cost and Kantorovich cost coincide.

3.2.2 Kantorovich Duality

Let μ and ν be absolutely continuous probability measures on \mathbb{R}^d . Let c be a cost function. There is a dual optimization problem to the Kantorovich optimal transport problem from μ to ν for the cost c .

Definition 3.2.14. The **Kantorovich dual formulation of the optimal transport problem from μ to ν for the cost c** (or the **Kantorovich dual problem**) is the optimization problem

$$\max_{u,v} \{\text{cost}(u, v)\} := \max_{u,v} \left\{ \int_{\mathbb{R}^d} v(y) d\nu(y) - \int_{\mathbb{R}^d} u(x) d\mu(x) : \right. \\ \left. v(y) - u(x) \leq c(x, y) \text{ for all } x, y \in \mathbb{R}^d \right\}. \quad (3.1)$$

The supremum in this optimization problem is called the **Kantorovich dual cost from μ to ν for the cost c** .

Different authors have different conventions for the signs in front of the functions u and v in the Kantorovich dual problem. I am following the convention of Villani’s 2008 book [39]. (This disagrees with the convention in Villani’s 2003 book [38].)

Definition 3.2.15. A pair of functions u and v which are optimizers in the Kantorovich dual problem from μ to ν for the cost c are called **Kantorovich potential functions** or **Kantorovich potentials**.

The Kantorovich dual cost and the Kantorovich cost agree under weak hypotheses, given in the following theorem from Villani [38].

Theorem 3.2.16 (Theorem 1.3 from [38]). *Let X and Y be Polish spaces (complete separable metric spaces). Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Let $c : X \times Y \rightarrow [0, +\infty]$ be a lower semicontinuous cost function. Then,*

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \{\text{cost}(\gamma)\} = \sup_{\substack{(u,v) \in L^1(\mu) \times L^1(\nu) \\ v(y) - u(x) \leq c(x,y)}} \{\text{cost}(u, v)\} = \sup_{\substack{(u,v) \in C_b(X) \times C_b(Y) \\ v(y) - u(x) \leq c(x,y)}} \{\text{cost}(u, v)\},$$

where $C_b(X)$ and $C_b(Y)$ denote the spaces of bounded continuous functions on X and Y , respectively. Further, the infimum on the left hand side is achieved by a measure in $\Gamma(\mu, \nu)$.

More precise information about the structure of optimizers in the Kantorovich optimal transport problem and Kantorovich dual problem are given by Schachermayer-Teichmann [37] and Villani [39].

Theorem 3.2.17 (Part of theorem 5.10 from [39]). *Let μ and ν be absolutely continuous probability measures on \mathbb{R}^d and let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$ be a lower semicontinuous cost function. Then, there exists an optimizer $\bar{\gamma} \in \Gamma(\mu, \nu)$ for the Kantorovich optimal transport problem:*

$$\text{cost}(\bar{\gamma}) = \inf_{\gamma \in \Gamma(\mu, \nu)} \{\text{cost}(\gamma)\},$$

and there exists a pair of Kantorovich potentials $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ and $v : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ satisfying

$$v(y) - u(x) \leq c(x, y) \text{ for all } x, y \in \mathbb{R}^d$$

with equality holding $\bar{\gamma}$ -a.e. These functions are optimizers in the Kantorovich dual problem and satisfy

$$v(y) = \inf_{x \in \mathbb{R}^d} \{c(x, y) + u(x)\} \quad \text{and} \quad u(x) = \sup_{y \in \mathbb{R}^d} \{v(y) - c(x, y)\}.$$

If the functions u , v , and c are all differentiable, then their derivatives are related to each other, as given in the following proposition.

Proposition 3.2.18. *Suppose that the cost function c as well as the Kantorovich potentials u and v are differentiable. Let γ be an optimal transport plan for the corresponding Kantorovich optimal transport problem. Then, at each point in the support of γ ,*

$$\frac{\partial c}{\partial x} = -\frac{\partial u}{\partial x} \quad \text{and} \quad \frac{\partial c}{\partial y} = \frac{\partial v}{\partial y}$$

Proof. Since c , u , and v are differentiable, the function $c(x, y) + u(x) - v(y)$ is differentiable, so in particular is differentiable on $\text{supp}(\gamma)$. By construction, $c(x, y) + u(x) - v(y) \geq 0$ and, by proposition 3.2.17, $c(x, y) + u(x) - v(y) = 0$ almost everywhere on $\text{supp}(\gamma)$. Therefore, the function $c(x, y) + u(x) - v(y)$ has a minimum at almost every point of $\text{supp}(\gamma)$. Since a differentiable function has derivative 0 at its local extrema, the proposition holds. ■

3.3 The Lagrangian Optimal Transport Problem

Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex with respect to velocity. In the previous chapter, cost functions were defined in terms of the action of

Tonelli Lagrangians (definition 2.2.5) via

$$c(x, y) = \inf \left\{ \int_0^1 L(\sigma(t), \dot{\sigma}(t)) dt : \sigma \in W^{1,1}, \sigma(0) = x, \sigma(1) = y \right\}. \quad (3.2)$$

In theorem 2.2.6, it was shown that for any $x, y \in \mathbb{R}^d$, there is an optimal trajectory $\sigma : [0, 1] \rightarrow \mathbb{R}^d$ with endpoint x and y which yields the cost.

In subsection 2.2.1, it was shown that optimal trajectories are of class C^2 and satisfy the Euler-Lagrange equation. Furthermore, it was seen that c is locally lipschitz (proposition 2.2.19), superdifferentiable (corollary 2.2.20), and differentiable almost everywhere (corollary 2.2.21); the derivatives of c were computed explicitly (corollary 2.2.22). Finally, proposition 2.2.23 showed that if either $\partial_x c(x, y)$ or $\partial_y c(x, y)$ exist, then there exists a unique optimal trajectory from x to y . This motivates structuring the optimal transport problem explicitly in terms of a Tonelli Lagrangian.

Definition 3.3.1. Let ρ_0 and ρ_1 be probability densities on \mathbb{R}^d . The **Lagrangian optimal transport problem from ρ_0 to ρ_1 for the cost c induced by the Tonelli Lagrangian L** is the minimization problem

$$\min_{\sigma: \mathbb{R}^d \times [0,1] \rightarrow \mathbb{R}} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(\sigma(t, x), \dot{\sigma}(t, x)) \rho_0(x) dx dt : \sigma(1, \cdot) \# \rho_0 = \rho_1 \right\},$$

where $\sigma(\cdot, x) \in W^{1,1}$ for every x and $\sigma(0, \cdot) = id$.

With the results established in the first chapter, the admissible families of trajectories in the Lagrangian problem may be assumed to be of class C^2 with respect to time.

Let ρ_0 and ρ_1 be probability densities on \mathbb{R}^d . By proposition 3.2.11 and theorem 3.2.13, the Monge optimal transport problem from ρ_0 to ρ_1 for the cost c induced by the Tonelli Lagrangian L has a solution $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ which pushes forward ρ_0 to ρ_1 . In this case, the solution to the Monge optimal transport problem satisfies

$$\int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx = \int_{\mathbb{R}^d} \int_0^1 L(\sigma(t, x), \dot{\sigma}(t, x)) dt \rho_0(x) dx \quad (3.3)$$

where $\sigma(\cdot, x)$ is an optimal trajectory from x to $T(x)$.

More substantial relationships between the minimizers of the Monge, Kantorovich dual, and Lagrangian optimal transport problems will be established in the proceeding subsections.

3.3.1 Further Properties of Cost Functions Induced by Lagrangians

Let μ and ν be absolutely continuous probability measures on \mathbb{R}^d . Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex in velocity and let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty)$ be the induced cost function. Let T be the Monge map for the Monge optimal transport problem from μ to ν for the cost c . Let u and v be Kantorovich potentials for the corresponding Kantorovich dual problem.

This subsection will establish that for almost every x , there is a unique action minimizing curve from x to $T(x)$. This does not follow directly from proposition 2.2.23; it requires a more subtle argument. The first step will be to show that the Kantorovich potential u is differentiable almost everywhere.

The following theorem from Figalli-Gigli [19] establishes the differentiability of the Kantorovich potentials when the cost function is given by the square Riemannian distance on a non-compact Riemannian manifold. As noted by Figalli-Gigli, this theorem follows from a generalization of the arguments in Appendix C of Gangbo-McCann [21].

Theorem 3.3.2 (Theorem 1 from Figalli-Gigli). *Let M be a Riemannian manifold of dimension n with distance function $d : M \times M \rightarrow [0, +\infty)$. Let $c(x, y) = \frac{1}{2}d^2(x, y)$. Let μ and ν be probability measures on M . Let u and v be Kantorovich potentials for the Kantorovich dual problem from μ to ν for the cost c . Let*

$$\partial^c u(x) = \{y \in M : u(x) = v(y) - c(x, y)\}$$

and

$$\partial^c u = \cup_{x \in M} (\{x\} \times \partial^c u(x)).$$

Let $D = \{x \in M : u(x) < +\infty\}$. Then, the potential u is locally semiconvex in $\text{int}D$, the set $\partial^c u(x)$ is non-empty for all $x \in \text{int}D$, and $\partial^c u$ is locally bounded in $\text{int}D$. Moreover, the set $D \setminus \text{int}D$ is $(n - 1)$ -rectifiable.

Under the hypotheses of this theorem, there is an immediate corollary about the differentiability of the Kantorovich potential u .

Corollary 3.3.3. *With the hypotheses of the previous theorem (that is, on a Riemannian manifold with the cost given by $c = \frac{1}{2}d^2$), the Kantorovich potential u is differentiable almost everywhere.*

Proof. Locally semiconvex functions are differentiable almost everywhere. ■

As noted in Remark 3 of Figalli-Gigli [19], these results continue to hold when the cost function $\frac{1}{2}d^2$ is replaced with a cost function induced by a Tonelli Lagrangian. This is recorded in the following proposition.

Proposition 3.3.4. *Let μ and ν be probability measures on \mathbb{R}^d . Let c be a cost function induced by a Tonelli Lagrangian and let u and v be Kantorovich potentials for the Kantorovich dual problem from μ to ν for the cost c . Then, the function u is semiconcave, differentiable almost everywhere, and its derivative ∂u is locally bounded.*

The following proposition and corollary establish the differentiability of the cost function c at almost every point of the form $(x, T(x))$. This proposition and its proof are a modification of Lemma 7 from McCann [30].

Proposition 3.3.5. *Suppose that the Kantorovich potential u is differentiable at x . If equality holds in*

$$c(x, y) + u(x) - v(y) \geq 0, \quad (3.4)$$

then y is the endpoint after unit time of the flow of Lagrangian system

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v) \\ (x_0, v_0) = (x, \nabla_p H(x, -\partial u(x))). \end{cases}$$

Proof. Suppose that u is differentiable at the point x and suppose that equality holds in (3.4). Then, for all $z \in \mathbb{R}^d$,

$$\begin{aligned} c(z, y) + u(z) - v(y) &\geq 0 \\ &= c(x, y) + u(x) - v(y). \end{aligned}$$

Rearranging and eliminating $v(y)$ yields

$$c(z, y) \geq c(x, y) + u(x) - u(z) \quad (3.5)$$

$$\geq c(x, y) + u(x) - u(x) - \partial u(x) \cdot (z - x) + o(z - x) \quad (3.6)$$

$$= c(x, y) - \partial u(x) \cdot (z - x) + o(z - x). \quad (3.7)$$

Equation (3.6) follows from (3.5) since $\partial u(x)$ is a subdifferential of u at x . This proves $-\partial u(x)$ is a subdifferential of $c(\cdot, y)$ at x . And, by corollary ??, the function $c(\cdot, y)$ is always superdifferentiable. Therefore, $c(\cdot, y)$ is differentiable at x .

By a simple modification of the proof of proposition 2.2.23, whenever $c(\cdot, y)$ is differentiable, there is a unique optimal curve from x to y . From proposition 3.2.18, when $c(x, y) + u(x) - v(y) = 0$ and both $c(\cdot, y)$ and u are differentiable at the point x , then

$$\nabla_x c(x, y) = -\partial u(x)$$

and from corollary 2.2.22,

$$\nabla_x c(x, y) = -\nabla_v L(x, \dot{\sigma}(0)),$$

where σ is an optimal curve from x to y . (In particular, since $\nabla_x c(x, y)$ exists, a unique optimal curve is guaranteed to exist by proposition 2.2.23.) Since $\nabla_p H(x, \nabla_v L(x, v)) = v$,

$$\nabla_p H(x, \partial u(x)) = \nabla_p H(x, \nabla_v L(x, \dot{\sigma}(0))) = \dot{\sigma}(0).$$

Therefore, the solution to

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v) \\ (x_0, v_0) = (x, \nabla_p H(x, -\partial u(x))) \end{cases}$$

is the unique optimal curve σ going from x to y . ■

Corollary 3.3.6. *For almost every x , the partial derivative*

$$\nabla_z c(z, T(x)) \Big|_{z=x} \tag{3.8}$$

exists and there is a unique optimal curve from x to $T(x)$.

Proof. By proposition 3.3.4, the Kantorovich potential u is differentiable at almost every $x \in \mathbb{R}^d$. By proposition 3.3.5 and its proof, when u is differentiable at x and $c(x, y) + u(x) - v(y) = 0$, then $c(\cdot, y)$ is differentiable at x . And then, by proposition 2.2.23 there exists a unique optimal curve from x to y .

By theorem 3.2.13, every optimal transport plan is concentrated on the graph of the Monge map, and by theorem 3.2.17, $c(x, y) + u(x) + v(y) = 0$ for almost every point (x, y) in the support of any optimal transport plan. So, for almost every x ,

$$c(x, T(x)) + u(x) - v(T(x)) = 0.$$

Therefore, for almost every x , the hypotheses of theorem 3.3.5 are satisfied at the point $(x, T(x))$, which proves that

$$\nabla_z c(z, T(x)) \Big|_{z=x}$$

exists and there is a unique optimal curve from x to $T(x)$. ■

3.3.2 Regularity of Optimal Trajectories

In section 2.2.1, optimal trajectories were found to be of class C^2 in time. This section establishes the regularity of optimal trajectories with respect to position. This is more subtle, and involves considering optimal trajectories with respect to their position after a short time, rather than the true initial position. In this framework, optimal trajectories and their time derivatives will turn out to both be locally Lipschitz with respect to their (after a short time) “initial” positions. First, however, the injectivity of optimal trajectories will be established.

Proposition 3.3.7. *Let $\sigma^{x_0} : [0, 1] \rightarrow \mathbb{R}^d$ be the solution of the Lagrangian system*

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v) \\ (x_0, v_0) = (x_0, \nabla_p H(x_0, -\partial u(x_0))) \end{cases}$$

Let $\sigma_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be given by

$$\sigma_t(x) = \sigma(t, x) = \sigma^x(t).$$

Then, $\sigma_1 = T$ and, for every $t \in [0, 1)$, the map $\sigma_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is injective.

Proof. By corollary 3.3.4, the Kantorovich potential u is differentiable almost everywhere, so the map σ is well defined in $L^1([0, 1] \times \mathbb{R}^d; \mathbb{R}^d)$.

By corollary 3.3.6, for almost every x , the partial derivative $\nabla_z c(z, T(x))|_{z=x}$ exists and there is a unique action minimizing curve from x to $T(x)$. From the proofs of

proposition 3.3.5 and corollary 3.3.6, this curve is the solution to the Lagrangian system

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v) \\ (x_0, v_0) = (x, \nabla_p H(x, -\partial u(x))). \end{cases}$$

By the uniqueness of solutions to Lagrangian systems, it follows that for almost every x ,

$$\sigma^x(1) = T(x) \implies \sigma^1 = T.$$

And again by the uniqueness of solutions to Lagrangian systems, if $t \in [0, 1)$ and

$$\sigma^{x_1}(t) = \sigma^{x_2}(t),$$

then $x_1 = x_2$, which proves that σ_t is injective for all $t \in [0, 1)$. ■

An important tool for understanding the regularity of optimal trajectories with respect to initial conditions is Mather's shortening lemma.

Theorem 3.3.8 (Mather's Shortening Lemma; Cor. 8.2 from [39]). *Let M be a smooth Riemannian manifold. Let d denote Riemannian distance on M . Let $L : TM \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex with respect to velocity. Let c be the induced cost function. Then, for any compact set $K \subset M$, there exists a constant C_K such that, whenever x_1, y_1, x_2, y_2 are four points in K which satisfy*

$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1)$$

and γ_i is the action minimizing curve from x_i to y_i , then, for any $t_0 \in (0, 1)$,

$$\sup_{t \in [0, 1]} d(\gamma_1(t), \gamma_2(t)) \leq \frac{C_K}{\min t_0, 1 - t_0} d(\gamma_1(t_0), \gamma_2(t_0)).$$

Regularity for optimal trajectories will proceed in four steps. First, a local boundedness lemma will be proven. Then, the family of trajectories $\sigma_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ will be shown to be Lipschitz continuous with respect to its time s initial conditions when $s \in (0, 1)$ —regularity fails to hold with respect to initial conditions at times 0. The time derivative map $\dot{\sigma}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ will be shown to be Hölder-1/2 continuous. Finally, $\dot{\sigma}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ will be shown to be locally Lipschitz.

Lemma 3.3.9. *Let $\rho_0, \rho_1 \in \mathcal{P}^{ac}(\mathbb{R}^d)$. Let L be a Tonelli Lagrangian and let $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the family of trajectories solving the optimal transport problem from ρ_0 to ρ_1 for the cost induced by L . Let $t \in (0, 1)$. For each $y \in \mathbb{R}^d$ such that $y = \sigma_{t_0}(x)$, there exists a neighbourhood $N \ni y$ and a compact set K such that whenever $\sigma_{t_0}(z) \in N$, then the endpoints of the curve $\sigma^z : [0, 1] \rightarrow \mathbb{R}^d$ are contained in K .*

Proof. Let $\gamma : [0, t_0] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\eta : [0, 1 - t_0] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be given by

$$\begin{cases} \gamma(t, y) = \sigma(t_0 - t, x) & \text{where } \sigma(t_0, x) = y \\ \eta(t, y) = \sigma(t_0 + t, x) & \text{where } \sigma(t_0, x) = y. \end{cases}$$

Then, γ solves the time t_0 optimal transport problem from $\rho_{t_0} = (\sigma_{t_0})_{\#}\rho_0$ to ρ_0 and η solves the optimal transport problem from ρ_{t_0} to ρ_1 .

By proposition 3.3.7, the trajectories $\gamma(\cdot, y)$ and $\eta(\cdot, y)$ are solutions to the Lagrangian system where the initial velocities are C^1 functions of Kantorovich potentials for the respective optimal transport problems. By proposition 3.3.4, the potentials are locally bounded, so there are neighbourhoods $N_\gamma, N_\eta \ni y$ on which the families of curves γ and η are bounded.

Let N be the intersection of N_γ and N_η . The proof concludes by noting that the image of a curve $\sigma^x : [0, 1] \rightarrow \mathbb{R}^d$ with $\sigma^x(t_0) = y$ is the union of the images of $\gamma(\cdot, y)$ and $\eta(\cdot, y)$. ■

Proposition 3.3.10. *Let $\rho_0, \rho_1 \in \mathcal{P}^{ac}(\mathbb{R}^d)$. Let L be a Tonelli Lagrangian which is strongly convex in velocity. Let c be the induced cost function and let $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the family of trajectories solving the optimal transport problem from ρ_0 to ρ_1 . Let $\epsilon > 0$. Let $\tilde{\sigma} : [\epsilon, 1 - \epsilon] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be given by*

$$\tilde{\sigma}(t, y) = \sigma(t, \sigma_t^{-1}(y)).$$

Then, the map $\tilde{\sigma}_t = \tilde{\sigma}(t, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is locally Lipschitz for all $t \in [\epsilon, 1 - \epsilon]$.

Proof. Fix $t \in [\epsilon, 1 - \epsilon]$. By lemma 3.3.9, given $\sigma_t x = y$, there is a neighbourhood $N \ni y$ such that all trajectories $\tilde{\sigma}^z$ which intersect N at time t are contained in a compact set K .

Since σ solves the optimal transport problem, for almost every pair $x, y \in \mathbb{R}^d$,

$$c(\sigma_0(x), \sigma_1(x)) + c(\sigma_0(y), \sigma_1(y)) \leq c(\sigma_0(x), \sigma_1(z)) + c(\sigma_0(y), \sigma_1(x)).$$

After possibly redefining σ on a set of measure zero, it satisfies the hypotheses of Mather's lemma (3.3.8). So, by Mather's lemma,

$$\sup_{s \in [0,1]} \|\sigma_s(x) - \sigma_s(y)\| \leq \begin{cases} \frac{C_K}{\min t, 1-t} \|\sigma_t(x) - \sigma_t(y)\| \\ \frac{C_K}{\min \epsilon, 1-\epsilon} \|\sigma_\epsilon(x) - \sigma_\epsilon(y)\|. \end{cases}$$

In particular, there are constants C_1 and C_2 depending only on K and ϵ such that

$$C_1 \|\sigma_\epsilon(x) - \sigma_\epsilon(y)\| \leq \|\sigma_t(x) - \sigma_t(y)\| \leq C_2 \|\sigma_\epsilon(x) - \sigma_\epsilon(y)\|.$$

Rewriting this in terms of $\tilde{\sigma}$ instead of σ , and setting $w = \sigma_\epsilon(x)$ and $z = \sigma_\epsilon(y)$, it follows that

$$C_1 \|z - w\| \leq \|\tilde{\sigma}_t(z) - \tilde{\sigma}_t(w)\| \leq C_2 \|z - w\|.$$

Therefore, $\tilde{\sigma}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is locally Lipschitz. \blacksquare

Proposition 3.3.11. *With the same set-up as proposition 3.3.10, the map $\tilde{\sigma}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is locally Hölder-1/2 continuous.*

Proof. Suppose that $x_n \rightarrow x_0$. Then, Taylor expanding $\tilde{\sigma}$ in time yields

$$\tilde{\sigma}_s(x_n) - \tilde{\sigma}_s(x_0) = \tilde{\sigma}_t(x_n) - \tilde{\sigma}_t(x_0) + (\dot{\tilde{\sigma}}_t(x_n) - \dot{\tilde{\sigma}}_t(x_0))(t - s) + O((s - t)^2).$$

By the triangle inequality, it follows that

$$\|\tilde{\sigma}_s(x_n) - \tilde{\sigma}_s(x_0)\| \geq \|\dot{\tilde{\sigma}}_t(x_n) - \dot{\tilde{\sigma}}_t(x_0)\| (s - t) - \|\tilde{\sigma}_t(x_n) - \tilde{\sigma}_t(x_0)\| - K(s - t)^2.$$

By the previous proposition (3.3.10), both $\|\tilde{\sigma}_s(x_n) - \tilde{\sigma}_s(x_0)\|$ and $\|\tilde{\sigma}_t(x_n) - \tilde{\sigma}_t(x_0)\|$ are bounded by a scalar multiple of $\|x_n - x_0\|$. Hence, there is a constant M such that

$$M \|x_n - x_0\| \geq \|\dot{\tilde{\sigma}}_t(x_n) - \dot{\tilde{\sigma}}_t(x_0)\| (s - t) - K(s - t)^2. \quad (3.9)$$

Set

$$s - t = \sqrt{\frac{2M + 1}{K} \|x_n - x_0\|}.$$

Substituting this value of $s - t$ into equation (3.9) and rearranging yields

$$\|\dot{\tilde{\sigma}}_t(x_n) - \dot{\tilde{\sigma}}_t(x_0)\| \leq C \|x_n - x_0\|$$

for positive constant C . Therefore, $\dot{\tilde{\sigma}}_t$ is locally Lipschitz. Since this argument was independent of t , this is true for every $t \in [\epsilon, 1 - \epsilon]$. ■

Proposition 3.3.12. *With the same set-up as proposition 3.3.10, the map $\dot{\tilde{\sigma}}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is locally Lipschitz for all $t \in [\epsilon, 1 - \epsilon]$.*

Proof. Fix $t \in [\epsilon, 1 - \epsilon]$. Let $y = \sigma_t(x)$ and let $N \ni y$ be an open set around y where all trajectories passing through N are bounded, as in lemma 3.3.9. Let $\Omega = \text{sig}\tilde{m}a_t^{-1}(N)$ and let $K \subset \Omega$ be a compact set containing an open set containing $\tilde{\sigma}_t^{-1}(y)$. Let $(x_n)_{n=1}^\infty$ be a sequence in K . For notational convenience, let

$$y_n = \tilde{\sigma}_t(x_n) \quad \text{and} \quad v_n = \dot{\tilde{\sigma}}_t(x_n).$$

The result will follow upon demonstrating that

$$\sup_{n,m} \frac{\|v_n - v_m\|}{\|x_n - x_m\|} \leq C$$

for some constant C .

The proof proceeds by contradiction. Suppose that this supremum is infinite. By the previous proposition, $\dot{\tilde{\sigma}}_t$ is continuous, so the sequence (v_n) is bounded on compact sets. Hence, the supremum can only diverge along a subsequence (x_n) where $\|x_m - x_n\| \rightarrow 0$ as $m, n \rightarrow \infty$. After passing to a convergent subsequence, suppose that $x_n \rightarrow x_0$ as $n \rightarrow \infty$.

Define $(q_n, p_n) : [\epsilon, 1 - \epsilon] \rightarrow T^*\mathbb{R}^d$ by taking the global Legendre transform of the sequence (y_n, v_n) :

$$(q_n, p_n)(t) = (y_n, \nabla_v L(y_n, v_n)).$$

Since the global Legendre transform is a C^1 -diffeomorphism, this sequence is also convergent. (As well, since $y_n = \tilde{\sigma}_t(x_n)$ and $v_n = \dot{\tilde{\sigma}}_t(x_n)$, both may be viewed, for fixed n , as functions from $[\epsilon, 1 - \epsilon]$ to \mathbb{R}^d .)

Let $\Psi_{t,s} : T^*\mathbb{R}^d \rightarrow T^*\mathbb{R}^d$ denote the Hamiltonian flow from time t to time s (see definition 2.4.4). Let π_q denote projection onto position and π_p denote projection onto

momentum. The map $\Psi_{t,s}$ is of class C^1 , so can be Taylor expanded around (q_0, p_0) :

$$\Psi_{t,s}(q_n, p_n) = \Psi_{t,s}(q_0, p_0) + \begin{pmatrix} \pi_q D_q \Psi_{s,t} & \pi_q D_p \Psi_{s,t} \\ \pi_p D_q \Psi_{s,t} & \pi_p D_p \Psi_{s,t} \end{pmatrix} \begin{pmatrix} q_n - q_0 \\ p_n - p_0 \end{pmatrix} + o(\|q_n - q_0, p_n - p_0\|). \quad (3.10)$$

Here, the derivative of the Hamiltonian flow is evaluated at (q_0, p_0) . Rearranging equation (3.10) and considering only the position component yields

$$q_n(s) - q_0(s) = \begin{pmatrix} \pi_q D_q \Psi_{s,t} & \pi_q D_p \Psi_{s,t} \end{pmatrix} \begin{pmatrix} q_n(t) - q_0(t) \\ p_n(t) - p_0(t) \end{pmatrix} + o(\|q_n - q_0, p_n - p_0\|). \quad (3.11)$$

Recall that $q_n(s) = \tilde{\sigma}_s(x_n)$ and $q_0(s) = \tilde{\sigma}_s(x_0)$. From applying Mather's lemma, it follows that, for some constants C_1 and C_2 ,

$$\begin{aligned} & C_1 \|x_n - x_0\| + o(\|q_n - q_0, p_n - p_0\|) \\ & \geq \|\pi_q D_q \Psi_{s,t} \cdot (q_n(t) - q_0(t)) + \pi_q D_p \Psi_{s,t} \cdot (p_n(t) - p_0(t))\| \end{aligned} \quad (3.12)$$

$$\geq C_2 \|x_n - x_0\| - o(\|q_n - q_0, p_n - p_0\|). \quad (3.13)$$

Since $\pi_q D_q \Psi_{s,t}$ is bounded (as $\Psi_{s,t}$ is of class C_1) and $q_i(t) = \tilde{\sigma}_t(x_i)$ is locally Lipschitz, it follows that, with different constants than above,

$$\begin{aligned} C_1 \|x_n - x_0\| + o(\|q_n - q_0, p_n - p_0\|) & \geq \|\pi_q D_p \Psi_{s,t} \cdot (p_n(t) - p_0(t))\| \\ & \geq C_2 \|x_n - x_0\| - o(\|q_n - q_0, p_n - p_0\|). \end{aligned}$$

If $\pi_q D_p \Psi_{s,t}$ is positive definite, then this will imply that $p_i(t) = \nabla_v L(\tilde{\sigma}_t(x_i), \dot{\tilde{\sigma}}_t(x_i))$ is locally Lipschitz with respect to x , and since the Legendre transform is a C^1 -diffeomorphism, it will follow that $\dot{\tilde{\sigma}}_t(x)$ is locally Lipschitz with respect to x , which will complete the proof.

Finally, from proposition 2.4.9, $\pi_q D_p \Psi_{s,t}(p, q) = (t - s) \nabla_{pp} H + o(t - s)$. The matrix $\nabla_{pp} H(\Psi_{s,t}(p, q))$ is positive definite and bounded on each compact set. Since (p, q) is contained in a compact set (which depends on the set K from the beginning of the proof), so, the remainder term in the derivative is of the form $C(t - s)^2$. Therefore, for $t - s$ sufficiently small, $\pi_q D_p \Psi_{s,t}(p, q)$ is positive definite, which concludes the proof. \blacksquare

3.4 The Eulerian Optimal Transport Problem

In [3], Benamou and Brenier reformulate the classical quadratic cost optimal transport problem in an Eulerian framework, motivated by fluid mechanics. The equivalent formulation in the setting of costs induced by Tonelli Lagrangians is the following:

Definition 3.4.1. Let ρ_0 and ρ_1 be probability densities on \mathbb{R}^d . The **Eulerian optimal transport problem from ρ_0 to ρ_1 for the cost c induced by the Tonelli Lagrangian L** is the minimization problem

$$\min_{\rho, V} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(x, V(t, x)) \rho(t, x) dx dt : \rho|_{t=0} = \rho_0, \rho|_{t=1} = \rho_1, \text{ and } \frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho V) = 0 \right\},$$

where $[0, 1] \ni t \mapsto \rho_t \in \mathcal{P}^{ac}(\mathbb{R}^d)$ and V is a time-varying vector field with the property that V is locally Lipschitz on $(0, 1) \times \mathbb{R}^d$ and locally bounded on $[0, 1] \times \mathbb{R}^d$.

The local Lipschitz condition is imposed so that only vector fields which define flows are considered. A theory of flows of non-smooth vector fields has been developed (see [1] and [13] for example), following from the work of DiPerna-Lions [15], but will not be considered here.

In the above definition, the differential equation $\partial_t \rho + \nabla_x \cdot (\rho V) = 0$ is understood in the sense of distributions. That is, a pair (ρ, V) satisfies $\partial_t \rho + \nabla_x \cdot (\rho V) = 0$ if for all $\varphi \in C_c^\infty([0, 1] \times \mathbb{R}^d)$,

$$0 = \int_{\mathbb{R}^d} \int_0^1 \varphi [\partial_t \rho + \nabla_x \cdot (\rho V)] dt dx \tag{3.14}$$

$$\begin{aligned} &:= \int_{\mathbb{R}^d} \left(\varphi(t, x) \rho(t, x) \Big|_{t=0}^{t=1} - \int_0^1 \rho(t, x) \partial_t \varphi(t, x) dt \right) dx \\ &\quad - \int_{\mathbb{R}^d} \int_0^1 \nabla_x \varphi(t, x) \cdot \rho(t, x) V(t, x) dt dx. \end{aligned} \tag{3.15}$$

Definition 3.4.2. The differential equation $\partial_t \rho + \nabla_x \cdot (\rho V) = 0$ is called the **continuity equation**.

Proposition 3.4.3. Let $V : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a locally bounded locally Lipschitz time-varying vector field in the sense of definition 3.4.1. Let $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the flow of V , defined by

$$\dot{\sigma}(t, x) = V(t, \sigma(t, x)),$$

where $\sigma_0 = id$.

Let ρ_0 be a probability density on \mathbb{R}^d and let $\rho_t = (\sigma_t)_\# \rho_0$. Then, the pair (ρ_t, V) solves the continuity equation in the sense of distributions.

Proof. It follows from the definition of the pushforward (for a measurable set A , that $(\sigma_t)_\# \rho_0(A) = \rho_0(\sigma_t^{-1}(A))$) that for f locally integrable,

$$\int_{\mathbb{R}^d} f(t, x) \rho_t(x) dx = \int_{\mathbb{R}^d} f(t, x) (\sigma_t)_\# \rho_0(x) dx = \int_{\mathbb{R}^d} f(t, \sigma(t, x)) \rho_0(x) dx. \quad (3.16)$$

Let $\varphi \in C_c^\infty([0, 1] \times \mathbb{R}^d)$. Then, by equation (3.14),

$$\begin{aligned} & \int_{\mathbb{R}^d} \int_0^1 \varphi [\partial_t \rho + \nabla_x \cdot (\rho V)] dt dx \\ & := \int_{\mathbb{R}^d} \left(\varphi(t, x) \rho(t, x) \Big|_{t=0}^{t=1} - \int_0^1 \rho(t, x) \partial_t \varphi(t, x) dt \right) dx \\ & \quad - \int_{\mathbb{R}^d} \int_0^1 \nabla_x \varphi(t, x) \cdot \rho(t, x) V(t, x) dt dx \end{aligned} \quad (3.17)$$

$$\begin{aligned} & = \int_{\mathbb{R}^d} \varphi(1, x) \rho_1(x) - \varphi(0, x) \rho_0(x) dx \\ & \quad - \int_0^1 \int_{\mathbb{R}^d} [\partial_t \varphi(t, x) + \nabla_x \varphi(t, x) \cdot V(t, x)] \rho(t, x) dx dt. \end{aligned} \quad (3.18)$$

Using the identity from eq. (3.16), the double integral in eq. (3.18) simplifies to

$$\int_0^1 \int_{\mathbb{R}^d} [\partial_t \varphi(t, x) + \nabla_x \varphi(t, x) \cdot V(t, x)] \rho(t, x) dx dt \quad (3.19)$$

$$= \int_0^1 \int_{\mathbb{R}^d} [\partial_t \varphi(t, \sigma(t, x)) + \nabla_x \varphi(t, \sigma(t, x)) \cdot V(t, \sigma(t, x))] \rho_0(x) dx dt \quad (3.20)$$

$$= \int_0^1 \int_{\mathbb{R}^d} \underbrace{[\partial_t \varphi(t, \sigma(t, x)) + \nabla_x \varphi(t, \sigma(t, x)) \cdot \dot{\sigma}(t, x)]}_{= \frac{d}{dt} \varphi(t, \sigma(t, x))} \rho_0(x) dx dt \quad (3.21)$$

Applying the fundamental theorem of calculus to equation (3.21) yields

$$\int_0^1 \int_{\mathbb{R}^d} \frac{d}{dt} \varphi(t, \sigma(t, x)) \rho_0(x) dx dt \quad (3.22)$$

$$= \int_{\mathbb{R}^d} \varphi(t, \sigma(t, x)) \Big|_{t=0}^{t=1} \rho_0(x) dx \quad (3.23)$$

$$= \int_{\mathbb{R}^d} \varphi(1, \sigma(1, x)) \rho_0(x) - \varphi(0, \sigma(0, x)) \rho_0(x) dx \quad (3.24)$$

$$= \int_{\mathbb{R}^d} \varphi(1, x) \rho_1(x) - \varphi(0, x) \rho_0(x) dx. \quad (3.25)$$

The identity (3.16) is applied again to transform (3.24) to (3.25).

Substituting (3.25) back into equation (3.18) yields

$$\begin{aligned} \int_{\mathbb{R}^d} \int_0^1 \varphi [\partial_t \rho + \nabla_x \cdot (\rho V)] dt dx &= \int_{\mathbb{R}^d} \varphi(1, x) \rho_1(x) - \varphi(0, x) \rho_0(x) dx \\ &\quad - \int_{\mathbb{R}^d} \varphi(1, x) \rho_1(x) - \varphi(0, x) \rho_0(x) dx \\ &= 0, \end{aligned}$$

which proves the proposition. \blacksquare

With the regularity results for optimal trajectories in the Lagrangian optimal transport problem in sections 3.3.2, 3.3.1, and 2.2.1, there is a distinguished candidate minimizer for the Eulerian optimal transport problem.

Proposition 3.4.4. *Let ρ_0 and ρ_1 be probability densities on \mathbb{R}^d . Let L be a Tonelli Lagrangian which is strongly convex in velocity and let c be the induced cost function. Let $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the family of optimal trajectories solving the Lagrangian optimal transport problem from ρ_0 to ρ_1 . Then, there is a vector field $V : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ which satisfies*

$$\dot{\sigma}(t, x) = V(t, \sigma(t, x)) \text{ for all } t, x,$$

and V is locally Lipschitz on $(0, 1) \times \mathbb{R}^d$ and locally bounded on $[0, 1] \times \mathbb{R}^d$.

Proof. The vector field V is constructed directly by $\dot{\sigma}(t, x) = V(t, \sigma(t, x))$ and then extended to all of \mathbb{R}^d .

Such a vector field is well defined: from proposition 3.3.7, the trajectories $\sigma_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are injective for all $t \in [0, 1)$. Therefore, $V(t, \cdot)$ is well-defined on $\{\sigma_t(x) : x \in \mathbb{R}^d\}$.

At $t = 0$, $\sigma_t = id$, so $\dot{\sigma}_0(x) = V(0, x)$ and from proposition 3.3.7 and 3.3.4, the initial velocities of optimal trajectories are locally bounded. Solving the optimal transport problem backwards, from ρ_1 to ρ_0 yields the same result for $t = 1$.

Let $t \in (0, 1)$. Let $0 < \epsilon < t$. Using the notation from 3.3.10, denote $z = \sigma_\epsilon(x)$.

Then,

$$\begin{cases} \sigma(t, x) = \tilde{\sigma}(t, z) \\ \dot{\sigma}(t, x) = \dot{\tilde{\sigma}}(t, z) \end{cases}$$

and

$$V(t, \tilde{\sigma}(t, z)) = \dot{\tilde{\sigma}}(t, z).$$

From proposition 3.3.12, the map $\dot{\tilde{\sigma}}(t, \cdot)$ is locally Lipschitz, and in the proof of proposition 3.3.10, the map $\tilde{\sigma}(t, \cdot)$ was seen to be bi-Lipschitz. Together, this implies that V is Lipschitz for $t \in (0, 1)$.

This vector field can be extended globally to be locally Lipschitz, which completes the proof. ■

3.5 Five Equivalent Formulations of the Kantorovich Cost

Let L be a Tonelli Lagrangian which is strongly convex in the second argument. Let c be the induced cost function.

The goal of this section is to prove the equivalence of the different forms of the optimal transport problem and understand the relationships between the minimizers in each formulation of the problem.

Theorem 3.5.1. *Let L be a Tonelli Lagrangian which is strongly convex in the second argument. Let c be the induced cost function. Let ρ_0 and ρ_1 be probability densities on*

\mathbb{R}^d . Then,

$$W_c(\rho_0, \rho_1) := \inf_{\gamma} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\rho_0, \rho_1) \right\} \quad (3.26)$$

$$= \sup_{u, v} \left\{ \int_{\mathbb{R}^d} v(y) \rho_1(y) dy - \int_{\mathbb{R}^d} u(x) \rho_0(x) dx : v(y) - u(x) \leq c(x, y) \right. \\ \left. \text{for } \mu\text{-a.e. } x, \nu\text{-a.e. } y \right\} \quad (3.27)$$

$$= \inf_T \left\{ \int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx : T_{\#} \rho_0 = \rho_1 \right\} \quad (3.28)$$

$$= \inf_{\phi} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), \dot{\phi}(t, x)) \rho_0(x) dx dt : \phi(1, \cdot)_{\#} \rho_0 = \rho_1 \right\} \quad (3.29)$$

$$= \inf_{\rho, V} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(x, V(t, x)) \rho(t, x) dx dt : \rho|_{t=0} = \rho_0, \rho|_{t=1} = \rho_1, \right. \\ \left. \frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho V) = 0, V \text{ loc. Lipschitz for } t \in (0, 1), \right. \\ \left. V \text{ loc. bounded for } t \in [0, 1] \right\} \quad (3.30)$$

Proof. If $W_c(\rho_0, \rho_1)$ is infinite, there is nothing to show, so suppose that $W_c(\rho_0, \rho_1) < +\infty$. Since ρ_0 and ρ_1 are probability distributions, they are automatically absolutely continuous with respect to Lebesgue measure.

(3.26)=(3.27): From section 4.2, Kantorovich duality holds (so lines (3.26) and (3.27) are equal) whenever the cost function is lower semi-continuous (see 3.2.16). From section 2.2.2, cost function induced by Tonelli Lagrangians are continuous (see ??). Hence, (3.26) and (3.27) are equal.

(3.26)=(3.28): From section 4.1, if the hypothesis of theorem 3.2.13 are satisfied, then an optimal transport map from ρ_0 to ρ_1 exists and the Monge and Kantorovich costs coincide, so (3.26) and (3.28) are equal. Cost functions induced by Tonelli Lagrangians are continuous and bounded below. Since ρ_0 and ρ_1 are probability densities, they are absolutely continuous with respect to Lebesgue measure by assumption. The Kantorovich cost is assumed finite. Thus, all but two of the hypotheses of 3.2.13 are satisfied.

The remaining hypotheses of 3.2.13 are that the family of maps $x \mapsto c(x, y) = c_y(x)$ is locally semi-concave in x locally uniformly in y and the cost c satisfies the twist condition. By proposition 3.2.11, costs induced from Tonelli Lagrangians satisfy these properties.

Hence, the Monge and Kantorovich costs from ρ_0 to ρ_1 for the cost c are equal.

(3.28)=(3.29): Suppose that T is a Borel measurable map pushing forward ρ_0 to ρ_1 . For each x , let $\phi(\cdot, x)$ be the action minimizing curve from x to $T(x)$. Then,

$\phi(1, x) = T(x)$, so $\phi(1, \cdot)$ pushes forward ρ_0 to ρ_1 . Further,

$$c(x, T(x)) = \int_0^1 L(\phi(t, x), \dot{\phi}(t, x)) dt,$$

from which it follows that

$$\begin{aligned} \int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx &= \int_{\mathbb{R}^d} \int_0^1 L(\phi(t, x), \dot{\phi}(t, x)) dt \rho_0(x) dx \\ &= \int_{\mathbb{R}^d} \int_0^1 L(\phi(t, x), \dot{\phi}(t, x)) \rho_0(x) dt dx. \end{aligned}$$

Let $(T_n)_{n=1}^\infty$ be a minimizing sequence for (3.29). Let $\phi_n(\cdot, x)$ be the correspond minimizing curve from x to $T_n(x)$. Then,

$$\begin{aligned} &\inf_{\phi(1, \cdot) \# \rho_0 = \rho_1} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), \dot{\phi}(t, x)) \rho_0(x) dx dt \right\} \\ &\leq \lim_{n \rightarrow \infty} \int_0^1 \int_{\mathbb{R}^d} L(\phi_n(t, x), \dot{\phi}_n(t, x)) \rho_0(x) dx dt \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} c(x, T_n(x)) \rho_0(x) dx \\ &= \inf_{T \# \rho_0 = \rho_1} \left\{ \int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx \right\}. \end{aligned}$$

Similarly, if $\phi : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfies $\phi(1, \cdot)$ pushes forward ρ_0 to ρ_1 , then the map T_ϕ given by $T_\phi(x) = \phi(1, x)$ pushes forward ρ_0 to ρ_1 , and an analogous argument as above shows that

$$\inf_{T \# \rho_0 = \rho_1} \left\{ \int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx \right\} \leq \inf_{\phi(1, \cdot) \# \rho_0 = \rho_1} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), \dot{\phi}(t, x)) \rho_0(x) dx dt \right\}.$$

Hence,

$$\inf_{T \# \rho_0 = \rho_1} \left\{ \int_{\mathbb{R}^d} c(x, T(x)) \rho_0(x) dx \right\} = \inf_{\phi(1, \cdot) \# \rho_0 = \rho_1} \left\{ \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), \dot{\phi}(t, x)) \rho_0(x) dx dt \right\}.$$

(3.29)=(3.30): Let (ρ, V) be an admissible pair in the Eulerian optimal transport problem. Let $\phi : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the flow of V . Then, ϕ is in the admissible class for

the Lagrangian optimal transport problem, and using equation (3.16),

$$\begin{aligned} & \int_0^1 \int_{\mathbb{R}^d} L(x, V(t, x)) \rho(t, x) dx dt \\ &= \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), V(t, \phi(t, x))) \rho_0(x) dx dt \\ &= \int_0^1 \int_{\mathbb{R}^d} L(\phi(t, x), V(t, \phi(t, x))) \rho_0(x), \end{aligned}$$

so every admissible pair in the Eulerian optimal transport problem yields an admissible family of trajectories in the Lagrangian problem, and both have equal cost. Hence, the Lagrangian cost is no greater than the Eulerian cost. That is, (3.29) \leq (3.30).

Let $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the family of optimal trajectories for the Lagrangian optimal transport problem, given in proposition 3.3.7. From proposition 3.4.4, this family of trajectories is the flow of an admissible vector field V in the Eulerian optimal transport problem, together with $\rho_t := (\sigma_t)_\# \rho_0$, which are of equal cost in the optimal transport problem. Therefore, (3.30) \leq (3.29), which concludes the proof. \blacksquare

The following theorem describes the relationships between the optimizers in the various optimal transport problems.

Theorem 3.5.2. *Let ρ_0 and ρ_1 be probability densities on \mathbb{R}^d . Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be a Tonelli Lagrangian which is strongly convex in the second argument with corresponding Hamiltonian $H : T^*\mathbb{R}^d \rightarrow \mathbb{R}$. Let c be the induced cost function. Let $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the Monge map for the Monge problem from ρ_0 to ρ_1 for the cost c . Let (u, v) be the pair of Kantorovich potentials for the Kantorovich dual problem. Let σ be the family of optimal trajectories for the Lagrangian problem. Let (ρ, V) be the optimal pair for the Eulerian problem. Then*

(a) *The optimal trajectories satisfy $\sigma(0, x) = x$, $\sigma(1, x) = T(x)$, and σ solves*

$$\begin{cases} \dot{x} = v \\ \frac{d}{dt} \nabla_v L(x, v) = \nabla_x L(x, v) \\ (x_0, v_0) = (x_0, \nabla_p H(x_0, -\partial u(x_0))). \end{cases}$$

(b) *The interpolant probability densities satisfy $\rho_t = (\sigma_t)_\# \rho_0$*

(c) *The vector field satisfies $\dot{\sigma}(t, x) = V(t, \sigma(t, x))$, with $\dot{\sigma}(t, x) = \nabla_p H(x, -\partial u_t(x))$ where (u_t, v_t) are the Kantorovich potentials for the time $1 - t$ optimal transport problem from ρ_t to ρ_1 .*

Chapter 4

The Otto Calculus for Higher Order Functionals

Around 2000, Benamou and Brenier introduced an Eulerian formulation of the optimal transport problem [3]. With this formulation of the problem, it was realized that Wasserstein space had a structure in many ways analogous to a Riemannian manifold. At the same time, Otto introduced a Riemannian calculus that allowed for computations of derivatives and Hessians in Wasserstein space ([34] [35]). This calculus has come to be called the *Otto calculus*; it has proven to be a useful tool for understanding the geometry of Wasserstein space. In [35], using this calculus, Villani and Otto predicted that in the optimal transportation problem on Riemannian manifolds with the cost given by squared Riemannian distance, there would be a non-negative Ricci curvature condition associated with displacement convex functionals. This Ricci curvature bound was independently discovered and proven, at essentially the same time, by Cordero-Erausquin, McCann, and Schmuckenschläger via distinct methods in [12].

In *Optimal Transport: Old and New*, Villani poses two problems regarding the Otto calculus. He first asks whether it is possible to extend the Otto Calculus to the setting of the optimal transportation problem with costs induced by Tonelli Lagrangians, rather than only the classical cost of distance squared. This problem is resolved; when working in a purely Eulerian framework, rather than in a Riemannian setting, this problem becomes straightforward. Paul Lee has worked in a similar direction, from a Hamiltonian perspective, in [26].

Villani then asks whether it is possible to extend the Otto calculus to allow formal computations of displacement Hessians of functionals involving derivatives of densities of probability measures. This problem is more subtle; its resolution is the main contribution of this chapter.

In particular, the displacement Hessian of a functional involving any number of derivatives of a probability measure on an n -dimensional manifold will be computed. In one dimension, this will yield a new class of displacement convex functionals heretofore unknown. This class will include, as a special case, the functionals considered by Carrillo and Slepčev [8], which thus far were the only known examples of displacement convex functionals involving derivatives of densities.

In order that this chapter be readable independent of the rest of this thesis, several definitions and basic facts will be restated.

4.1 Setup

For this chapter, in order to avoid subtle issues of regularity, all objects will be assumed smooth in both time and position. At the conclusion of the chapter there will be a brief discussion of some of these regularity issues.

Let L be a Tonelli Lagrangian and let $\rho_0, \rho_1 \in \mathcal{P}^{ac}(\mathbb{R}^d)$ or in $\mathcal{P}^{ac}(\mathbb{T}^d)$ where $\mathbb{T}^d \cong \mathbb{R}^d/\mathbb{Z}^d$ be smooth probability densities (with the assumption of smoothness, probability measures and their densities will be treated interchangeably). For the rest of this section, definitions will be stated for \mathbb{R}^d , but apply equally well on \mathbb{T}^d ; the 1-torus will be denoted by S^1 . A smooth map

$$[0, 1] \ni t \mapsto \rho_t \in \mathcal{P}^{ac}(\mathbb{R}^d)$$

is a **smooth curve** in $\mathcal{P}^{ac}(\mathbb{R}^d)$. A curve in $\mathcal{P}^{ac}(\mathbb{R}^d)$ will be denoted (ρ_t) , while the probability measure at time t will be denoted ρ_t . When there is no ambiguity, the subscripts will be omitted.

The **Eulerian formulation of the optimal transportation problem from ρ_0 to ρ_1 with cost induced by L** is the infimal problem

$$\inf_{\rho, V} \left\{ \int_{\mathbb{R}^d} \int_0^1 L(x, V(t, x)) \rho(t, x) dt dx \mid \rho|_{t=0} = \rho_0, \rho|_{t=1} = \rho_1, \dot{\rho} + \nabla \cdot (\rho V) = 0 \right\},$$

where (ρ_t) is a smooth curve in $\mathcal{P}^{ac}(\mathbb{R}^d)$, with $\rho(t, x) = \rho_t(x)$, and V is a time-varying vector field. The equation $\dot{\rho} + \nabla \cdot (\rho V) = 0$ is called the **continuity equation**. The derivative with respect to time is given by $\dot{X} = \frac{\partial}{\partial t} X$, while ∇ and $\nabla \cdot$ refer to the spatial gradient and spatial divergence, respectively. Let σ denote the characteristic curves of

V . That is,

$$\dot{\sigma}(t, x) = V(t, \sigma(t, x)).$$

If (ρ, V) is an optimal pair in the Eulerian optimal transport problem, then the pair (ρ, V) also solves the following transport equations:

$$\dot{\rho} = -\nabla \cdot (\rho V) \quad \dot{V} = -V'V + W$$

where V' refers to the Jacobian matrix of spatial derivatives of V and

$$W(t, x) = \ddot{\sigma}(t, z) \Big|_{z=\sigma_t^{-1}(x)} = \frac{d^2}{dt^2} \sigma(t, z) \Big|_{z=\sigma_t^{-1}(x)}$$

(More generally, these transport equations are satisfied by all pairs of smooth densities and time-varying vector fields which solve the continuity equation) The term W corresponds to the acceleration of the characteristics of V and will often be assumed to be 0. (for instance in the classical optimal transportation problem with quadratic cost on \mathbb{R}^d , it is zero).

If (ρ, V) is a smooth optimizer in the Eulerian optimal transport problem, then (ρ_t) is called a **smooth geodesic in Wasserstein space**.

Let $\mathcal{F} : [0, 1] \rightarrow \mathbb{R}$ be given by

$$\mathcal{F}(t) = \int F((D^\alpha \rho_t(x))_{|\alpha| \leq n}) dx, \quad (4.1)$$

where n is an integer, α is a multi-index of length d , N is the number of multi-indices of length d and order at most n , and $F : \mathbb{R}^N \rightarrow \mathbb{R}$ is a smooth function.

In the smooth setting, the second derivative of \mathcal{F} is given by

$$\frac{d^2}{dt^2} \mathcal{F}(t) = \int_{\mathbb{R}^d} \frac{d^2}{dt^2} F((D^\alpha \rho_t(x))_{|\alpha| \leq n}) dx, \quad (4.2)$$

Definition 4.1.1. A functional $\mathcal{F} : [0, 1] \rightarrow \mathbb{R}$ is **displacement convex** if it is convex along every geodesic (ρ) in Wasserstein space.

The **displacement Hessian of \mathcal{F}** is the second derivative of \mathcal{F} computed along an arbitrary geodesic in Wasserstein space.

There are three key identities needed to compute displacement Hessians (two of them

are the transport equations from above).

$$\text{Key identities: } \begin{cases} \dot{\rho} = -\nabla \cdot (\rho V) & (4.3) \\ \dot{V} = -V'V + W & (4.4) \\ \ddot{\rho} = \nabla \cdot ((\nabla \cdot (\rho V)) V + \rho V'V - \rho W). & (4.5) \end{cases}$$

4.2 One derivative and one dimension

Let (ρ) be a geodesic in $\mathcal{P}^{ac}(\mathbb{R})$ or $\mathcal{P}^{ac}(S^1)$. Let $\mathcal{F} : [0, 1] \rightarrow \mathbb{R}$ be the functional given by

$$\mathcal{F}(t) = \int_{\mathbb{R}} F(\rho, \rho') dx,$$

where $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a smooth function. Denote by F_0 the partial derivative of F with respect to the first argument and by F_1 the partial derivative with respect to the second argument.

The computation of the displacement Hessian for functionals of the above form, involving only one derivative of the probability density, which itself has a one dimensional domain, is simpler than the general computation and yields interesting examples. It will be computed first.

Let X be a time-varying vector field on \mathbb{R} . Let $\Psi_0(\rho, X) = (\rho X)' - \rho' X$ and $\Psi_1(\rho, X) = (\rho X)'' - \rho'' X$. That is, $\Psi_i(\rho, X)$ the remainder of $(\rho X)^{(i+1)}$ after subtracting off all terms where X appears undifferentiated.

Theorem 4.2.1. *Let $\mathcal{F} : [0, 1] \rightarrow \mathbb{R}$ be as above. The displacement Hessian of \mathcal{F} is*

$$\begin{aligned} \mathcal{F}''(t) = & \int F_{00} (\Psi_0(\rho, V))^2 + 2F_{01} (\Psi_0(\rho, V)) (\Psi_1(\rho, V)) + F_{11} (\Psi_1(\rho, V))^2 \\ & + F_1 \left(2\rho' (V')^2 + 4\rho V'' V' \right) + F(W') - F_0(\Psi_0(\rho, W)) - F_1(\Psi_1(\rho, W)) dx. \end{aligned} \quad (4.6)$$

Proof. As all functions are assumed smooth, \mathcal{F} is differentiated twice:

$$\mathcal{F}''(t) = \int F_{00} (\dot{\rho})^2 + 2F_{01} (\dot{\rho}) (\dot{\rho}') + F_{11} (\dot{\rho}')^2 \quad (4.7)$$

$$+ F_0(\ddot{\rho}) + F_1(\ddot{\rho}') dx \quad (4.8)$$

Substituting the first key identity (4.3) into line (4.7) of the displacement Hessian yields

$$F_{00}(\dot{\rho})^2 + 2F_{01}(\dot{\rho})(\dot{\rho}') + F_{11}(\dot{\rho}')^2 \quad (4.9)$$

$$= F_{00}((\rho V)')^2 + 2F_{01}((\rho V)')((\rho V)'') + F_{11}((\rho V)'')^2 \quad (4.10)$$

$$= F_{00}(\Psi_0 + \rho'V)^2 + 2F_{01}(\Psi_0 + \rho'V)(\Psi_1 + \rho''V) + F_{11}(\Psi_1 + \rho''V)^2 \quad (4.11)$$

$$\begin{aligned} &= F_{00}(\Psi_0)^2 + 2F_{01}(\Psi_0)(\Psi_1) + F_{11}(\Psi_1)^2 \\ &+ \left(\underbrace{F_{00}(\rho') + F_{01}(\rho'')}_{=(F_0)'} \right) (V(\rho'V + 2\Psi_0)) + \left(\underbrace{F_{01}(\rho') + F_{11}(\rho'')}_{=(F_1)'} \right) (V(\rho''V + 2\Psi_1)). \end{aligned} \quad (4.12)$$

Substituting this back into the displacement Hessian and integrating by parts then yields

$$\begin{aligned} \mathcal{F}''(t) &= \int F_{00}(\Psi_0)^2 + 2F_{01}(\Psi_0)(\Psi_1) + F_{11}(\Psi_1)^2 \\ &+ F_0(\ddot{\rho} - (V(\rho'V + 2\Psi_0))') + F_1(\ddot{\rho}' - (V(\rho''V + 2\Psi_1))') dx. \end{aligned} \quad (4.13)$$

Substituting the third key identity (4.5) into the coefficient on F_0 (and remembering that $\Psi_0(\rho, V) = \rho V'$) results in

$$\ddot{\rho} - (V(\rho'V + 2\Psi_0))' = ((\rho V)'V + \rho V'V - \rho W)' - (\rho'V^2 + 2\rho V'V)' \quad (4.14)$$

$$= -(\rho W)' \quad (4.15)$$

$$= -(\rho'W + \Psi_0(\rho, W)). \quad (4.16)$$

Similarly, the coefficient on F_1 simplifies to

$$(\ddot{\rho}' - (V(\rho''V + 2\Psi_1))') = ((\rho V)'V + \rho V'V - \rho W)'' - (\rho''V^2 + 2V(\rho V'' + 2\rho'V'))' \quad (4.17)$$

$$= (2\rho(V')^2)' - (\rho W)'' \quad (4.18)$$

$$= (2\rho(V')^2)' - (\rho''W + \Psi_1(\rho, W)). \quad (4.19)$$

Substituting both (4.16) and (4.19) into (4.13) results in the following expression for the

displacement Hessian

$$\begin{aligned} \mathcal{F}''(t) = & \int F_{00} (\Psi_0(\rho, V))^2 + 2F_{01} (\Psi_0(\rho, V)) (\Psi_1(\rho, V)) + F_{11} (\Psi_1(\rho, V))^2 \\ & + F_1 (2\rho(V')^2) - F_0(\Psi_0(\rho, w)) - F_1(\Psi_1(\rho, w)) - \underbrace{(F_0(\rho') + F_1(\rho''))}_{=(F)'} W dx. \end{aligned} \quad (4.20)$$

Integrating by parts completes the computation. ■

In [8], Carrillo and Slepčev prove that the functional

$$E(\rho) = \int \left(\frac{d}{dx} \rho^\beta(t, x) \right)^2 dx$$

is displacement convex on $S^1 \cong \mathbb{R}/\mathbb{Z}$ with cost given by Euclidean distance squared whenever $\beta \in [-\frac{3}{2}, -1]$.

Example 4.2.1 (Carrillo-Slepčev functionals). Let $F(x, y) = \beta^2 x^{2\beta-2} y^2$. Then,

$$E(\rho) = \int F(\rho, \rho') dx.$$

Let $\theta = 2\beta - 2$. The derivatives of F are

$$\begin{cases} F_0 = \beta^2 \theta x^{\theta-1} y^2 \\ F_{00} = \beta^2 \theta (\theta - 1) x^{\theta-2} y^2 \\ F_{01} = 2\beta^2 \theta x^{\theta-1} y \\ F_1 = 2\beta x^\theta y \\ F_{11} = 2x^\theta. \end{cases}$$

Note that in the classical optimal transportation problem with Euclidean distance squared, the acceleration term W is identically 0. Then, using the formula from theorem 4.2.1 (eq. (4.6)), the displacement Hessian of E is

$$\begin{aligned} \frac{d^2}{dt^2} E(\rho) = & \int \beta^2 \left\{ \theta(\theta - 1) \rho^{\theta-2} (\rho')^2 (\rho V')^2 + 2(2\theta \rho^{\theta-1} \rho') (\rho V') (\rho V'' + 2\rho' V') \right. \\ & \left. + 2\rho^\theta (\rho V'' + 2\rho' V')^2 + 2\rho^\theta \rho' (2\rho' (V')^2 + 4\rho V'' V') \right\} dx. \end{aligned} \quad (4.21)$$

This is, in fact, a quadratic form in $(\rho' V')$ and $(\rho V'')$. After collecting like terms, it can

be rewritten as

$$\frac{d^2}{dt^2}E(\rho) = \int \beta^2 \rho^\theta \left\{ (\theta + 3)(\theta + 4)(\rho'V')^2 + 4(\theta + 4)(\rho'V')(\rho V'') + 2(\rho V'')^2 \right\} dx \quad (4.22)$$

$$= \int \beta^2 \rho^\theta \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix}^T \begin{pmatrix} (\theta + 3)(\theta + 4) & 2(\theta + 4) \\ 2(\theta + 4) & 2 \end{pmatrix} \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix} dx. \quad (4.23)$$

It is straightforward to check that

$$\begin{pmatrix} (\theta + 3)(\theta + 4) & 2(\theta + 4) \\ 2(\theta + 4) & 2 \end{pmatrix}$$

is a positive definite matrix whenever $\theta \in (-5, -4)$ and is positive semi-definite for $\theta = -5, -4$. Hence, the functional E is displacement convex whenever $\theta \in [-5, -4] \iff \beta \in [-3/2, -1]$. \blacklozenge

Following this example, it is straightforward to construct a new class of displacement convex functionals involving derivatives of densities. A simple convexity lemma will be needed.

Lemma 4.2.2. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be non-negative and of class C^2 . Let $\gamma \in \mathbb{R} \setminus \{0\}$. Then,*

$$f''(x)f(x) + (\gamma - 1)(f'(x))^2 \geq 0$$

if $g(x) = \frac{1}{\gamma}f^\gamma(x)$ is convex.

Proof. The second derivative of g is

$$\begin{aligned} g''(x) &= \frac{1}{\gamma} (\gamma f^{\gamma-1}(x) f'(x))' \\ &= (\gamma - 1) f^{\gamma-2}(x) (f'(x))^2 + f^{\gamma-1}(x) f''(x) \\ &= f^{\gamma-2} (f''(x) f(x) + (\gamma - 1)(f'(x))^2). \end{aligned}$$

Since f is non-negative, the sign of g is the same as that of $f''(x)f(x) + (\gamma - 1)(f'(x))^2$. \blacksquare

A new class of displacement convex functionals is now presented.

Theorem 4.2.3. *Let $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ be given by $F(x, y) = G(x)|y|^\alpha$ where G is of class C^2 and $\alpha \geq 2$. Let $G(x) = x^{-2\alpha}H(x)$. If $H(x) > 0$ for all $x > 0$ and*

$$H(x)^{-\frac{1}{\alpha-1}}$$

is concave on $[0, \infty)$, then the functional

$$\mathcal{F}(t) = \int G(\rho)|\rho'|^\alpha dx \quad (4.24)$$

is displacement convex for the classical optimal transport problem with cost given by Euclidean distance squared.

Note that, for instance, every functional of the form

$$\mathcal{F}(t) = \int \left(\frac{d}{dx} \Xi(\rho) \right)^2 dx$$

can be written in the form of eq. (4.24) with $\alpha = 2$.

Proof. The derivatives of F are

$$\begin{cases} F_0 = G'(x)|y|^\alpha \\ F_{00} = G''(x)|y|^\alpha \\ F_{01} = \alpha G'(x)y^{\alpha-1}\text{sign}(y) \\ F_1 = \alpha G(x)y^{\alpha-1}\text{sign}(y) \\ F_{11} = \alpha(\alpha-1)G(x)|y|^{\alpha-2}. \end{cases}$$

Then, using theorem 4.2.1, the displacement Hessian of \mathcal{F} is

$$\begin{aligned} \mathcal{F}''(t) = \int |\rho'|^{\alpha-2} \left\{ (G''(\rho)\rho^2 + 4\alpha G'(\rho)\rho + 2\alpha(2\alpha-1)G(\rho)) (\rho'V')^2 \right. \\ \left. + (2\alpha G'(\rho)\rho + 4\alpha^2 G(\rho)) (\rho'V')(\rho V'') \right. \\ \left. + (\alpha(\alpha-1)G(\rho)) (\rho V'')^2 \right\} dx \end{aligned} \quad (4.25)$$

$$= \int |\rho'|^{\alpha-2} \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix}^T \mathcal{A} \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix} dx. \quad (4.26)$$

Where \mathcal{A} is the 2×2 matrix

$$\mathcal{A} = \begin{pmatrix} G''(\rho)\rho^2 + 4\alpha G'(\rho)\rho + 2\alpha(2\alpha-1)G(\rho) & \alpha G'(\rho)\rho + 2\alpha^2 G(\rho) \\ \alpha G'(\rho)\rho + 2\alpha^2 G(\rho) & \alpha(\alpha-1)G(\rho) \end{pmatrix}.$$

Upon substituting $x^{-2\alpha}H(x) = G(x)$, the matrix simplifies to

$$\mathcal{A} = \rho^{-2\alpha} \begin{pmatrix} H''(\rho)\rho^2 & \alpha H'(\rho)\rho \\ \alpha H'(\rho)\rho & \alpha(\alpha - 1)H(\rho) \end{pmatrix}.$$

Recalling that $\alpha \geq 2$, the functional \mathcal{F} is displacement convex whenever H is chosen such that for all $x > 0$, the following two conditions hold

$$\begin{cases} x^{-2\alpha}H(x) > 0 & \iff H(x) > 0 \\ \frac{\alpha(\alpha-1)H''(x)H(x) - \alpha^2(H'(x))^2}{x^{2\alpha-2}} > 0 & \iff H''(x)H(x) - \frac{\alpha}{\alpha-1}(H'(x))^2 > 0. \end{cases}$$

Applying lemma 4.2.2 to the second expression completes the proof. \blacksquare

In the above proof, the positive definiteness of the matrix \mathcal{A} was found to be a sufficient condition for displacement convexity. There is a partial converse to this theorem. In particular, if the $(1, 1)$ -entry of \mathcal{A} ,

$$\xi(x) = G'''(x)x^2 + 4\alpha G''(x)x + 2\alpha(2\alpha - 1)G'(x), \tag{4.27}$$

is negative at some point $c > 0$, then the functional \mathcal{F} considered in the above theorem fails to be displacement convex, as sketched in the following example.

Example 4.2.2. Let \mathcal{F} be as in theorem 4.2.3, where G is assumed to be of class C^2 . If $\xi(c) < 0$ for $c > 0$ (from eq. (4.27)), then the displacement Hessian of \mathcal{F} can be made negative in the classical setting.

Let $\tilde{\rho}$ be the function given by

$$\tilde{\rho}(x) = c\chi_{[0, \frac{1}{c}]}(x) + \begin{cases} \delta \sin\left(\frac{x}{\epsilon}\right) & \text{if } x \in [0, \frac{1}{c}] \\ 0 & \text{otherwise,} \end{cases}$$

where χ_A denotes the characteristic function of A . The numbers δ and ϵ will be specified later (they'll be very small). Let ρ be a smoothed version of $\tilde{\rho}$, but maintaining the oscillatory behaviour of $\tilde{\rho}$ on $[0, \frac{1}{c}]$ (except possibly right near the endpoints), and ensuring that the ρ has total mass one.

In the classical setting, by Brenier's theorem, given any convex function ϕ , the map $T = \nabla\phi$ is the optimal transport map from ρ to $T_{\#}\rho$, and the corresponding optimal vector field is $V = T - I$, where I denotes the identity map.

Let $\phi(x) = x^2$. Then, $V'(x) = 1$ and $V''(x) = 0$. Therefore, along the geodesic from

ρ to $T_{\#}\rho$, the displacement Hessian of \mathcal{F} is

$$\begin{aligned}\mathcal{F}''(t) &= \int |\rho'|^{\alpha-2} \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix}^T \mathcal{A} \begin{pmatrix} \rho'V' \\ \rho V'' \end{pmatrix} dx. \\ &= \int |\rho'|^{\alpha-2} (\rho')^2 \xi(\rho) dx.\end{aligned}$$

Since G is of class C^2 , the function ξ is continuous. When δ is taken to be sufficiently small, $\rho(x) \approx c$ on $[0, 1/c]$ (except possibly very near the endpoints). Then, by choosing $\epsilon > 0$ sufficiently small, $|\rho'|$ can be made as large as desired, and

$$\begin{aligned}\mathcal{F}''(t) &= \int |\rho'|^{\alpha-2} (\rho')^2 \xi(\rho) dx \\ &\approx \int_0^{1/c} |\rho'|^\alpha \xi(c) dx,\end{aligned}$$

which is negative.

This example can be extended to S^1 with measures that are bounded away from zero by carrying out the same construction on an interval I of length less than $\min\{\frac{1}{2c}, \frac{1}{8}\}$, and adding a constant function on the rest of S^1 . \blacklozenge

4.3 A General Formula for the Displacement Hessian

As in the beginning of the chapter, let $\mathcal{F} : [0, 1] \rightarrow \mathbb{R}$ be given by

$$\mathcal{F}(t) = \int_{\mathbb{R}^d} F((D^\alpha \rho_t(x))_{|\alpha| \leq n}) dx,$$

where n is an integer, α is a multi-index of length d , N is the number of multi-indices of length d and order at most n , and $F : \mathbb{R}^N \rightarrow \mathbb{R}$ is a smooth function.

Let $X : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Let

$$\Psi_\alpha(\rho, X) = D^\alpha \nabla \cdot (\rho X) - (\nabla D^\alpha \rho) X. \quad (4.28)$$

That is, $\Psi_\alpha(\rho, X)$ is the remainder of $D^\alpha \nabla \cdot (\rho X)$ after subtracting off all terms where X appears undifferentiated. Similarly, define

$$\Phi_\alpha(\rho, X) = D^\alpha \nabla \cdot ((\nabla \cdot (\rho X) X) + \rho X' X) - 2X \cdot \nabla D^\alpha (\nabla \cdot (\rho X)) + x \cdot (D^\alpha \rho'') X. \quad (4.29)$$

Less obviously, $\Phi_\alpha(\rho, X)$ is the remainder of $D^\alpha \nabla \cdot ((\nabla \cdot (\rho X)X) + \rho X'X)$ after subtracting off all terms where V appears undifferentiated.

Proposition 4.3.1. *Let $X : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be smooth. Then,*

$$\Phi_\alpha(\rho, X) = D^\alpha \nabla \cdot ((\nabla \cdot (\rho X)X) + \rho X'X) - 2X \cdot \nabla D^\alpha (\nabla \cdot (\rho X)) + x \cdot (D^\alpha \rho'')X$$

and

$$\Psi_\alpha(\rho, X) = D^\alpha \nabla \cdot (\rho X) - (\nabla D^\alpha \rho)X.$$

contains no terms where X is undifferentiated.

Proof. Let $[a, b] = ab - ba$ denote the commutator of a and b . Writing out Ψ and Φ in coordinates reveals them to be sums of commutator:

$$\begin{aligned} \Psi_\alpha(\rho, V) &= \sum_{i=1}^d \{D^\alpha D_i V_i - V_i D^\alpha D_i\}(\rho) \\ &= \sum_{i=1}^d [D^\alpha D_i, V_i](\rho), \end{aligned}$$

and

$$\begin{aligned} \Phi_\alpha(\rho, V) &= \sum_{i,j=1}^d \{D^\alpha D_i D_j V_i V_j - 2V_i D^\alpha D_i D_j V_j + V_i V_j D^\alpha D_i D_j\}(\rho) \\ &= \sum_{i,j=1}^d [[D^\alpha D_i D_j, V_i], V_j](\rho). \end{aligned}$$

Given any differential operator L of order k and any smooth function f , the commutator $[L, f]$ is a differential operator of order $k-1$. (This can be seen, for instance, by induction on the degree of multi-indices, with the base case $[\frac{\partial}{\partial x_i}, f](g) = g \frac{\partial f}{\partial x_i}$).

Each summand $[D^\alpha D_i, V_i](\rho)$ of Ψ therefore contains no terms where V_i appears undifferentiated, which can also be seen by simply expanding out the commutator.

Since $D^\alpha D_i$ is a differential operator of order at least 2, $[D^\alpha D_i D_j, V_i]$ is a differential operator of order at least 1, and therefore $[[D^\alpha D_i D_j, V_i], V_j]$ contains no terms where V_i or V_j appear undifferentiated. \blacksquare

Theorem 4.3.2 (The Hessian of \mathcal{F}). *Let (ρ_t) be a smooth curve in $\mathcal{P}^{ac}(\mathbb{R}^d)$, let V_t be a time-dependent smooth vector field on \mathbb{R}^d , and let \mathcal{F} be a functional of the form as*

described above. If (ρ_t, V_t) solve the continuity equation, eq. (4.3), then the displacement Hessian of \mathcal{F} is

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{F} &= \int_{\mathbb{R}^d} \sum_{|\alpha|, |\beta| \leq n} F_{\alpha, \beta} \Psi_\alpha(\rho, V) \Psi_\beta(\rho, V) dx \\ &+ \int_{\mathbb{R}^d} \sum_{|\alpha| \leq n} F_\alpha \{ \Phi_\alpha(\rho, V) - 2(\nabla \cdot V) \Psi_\alpha(\rho, V) - \Psi_\alpha(\rho, W) \} dx \\ &+ \int_{\mathbb{R}^d} F \{ (\nabla \cdot V)^2 - \text{tr}(V')^2 + \nabla \cdot W \} dx. \end{aligned} \quad (4.30)$$

Remark. In the Riemannian setting, with the transport cost given by squared Riemannian distance, the function $\nabla \cdot W$ corresponds to the Bakry-Emery tensor (see appendix A).

Hence, in the Riemannian setting, the displacement Hessian of \mathcal{F} is a quadratic form in the derivatives of V of order $0, \dots, n+1$ whose coefficients depend on $(D^\alpha \rho)_{|\alpha| \leq n}$, the function F , and its first and second partial derivatives, where the 0th order terms of V only appear in $\nabla \cdot W$. (This is where the non-negative Ricci curvature condition comes from).

Proof of theorem 4.3.2. The integral formula for the displacement Hessian of \mathcal{F} is given in eq. (4.2).

By the chain rule, the integrand is can be rewritten as

$$\frac{d^2}{dt^2}\mathcal{F}((D^\alpha \rho)_{|\alpha| \leq n}) = \sum_{|\alpha|, |\beta| \leq n} F_{\alpha, \beta} D^\alpha(\dot{\rho}) D^\beta(\dot{\rho}) + \sum_{|\alpha| \leq n} F_\alpha D^\alpha(\ddot{\rho}). \quad (4.31)$$

From the key identities eq. (4.3) and (4.5), the time derivatives $\dot{\rho}$ and $\ddot{\rho}$ can be rewritten as expressions involving spatial gradients and Hessians of ρ . Upon substituting these identities into eq. (4.31), the double sum becomes an expression involving spatial derivatives of ρ up to order $n+1$, and up to order $n+2$ in the other sum.

The displacement Hessian will be rewritten to eliminate all derivatives of ρ of order greater than n .

By Eqs. (4.3), (4.4), and (4.28), the factors in the double sum are given by

$$\begin{aligned} D^\alpha \dot{\rho} &= -D^\alpha \nabla \cdot (\rho V) \\ &= -\Psi_\alpha(\rho, V) - V \cdot \nabla D^\alpha \rho. \end{aligned}$$

By the equality of mixed partials, $F_{\alpha,\beta} = F_{\beta,\alpha}$,

$$\begin{aligned} \sum_{|\alpha|,|\beta|\leq n} F_{\alpha,\beta} D^\alpha(\dot{\rho})D^\beta(\dot{\rho}) &= \sum_{|\alpha|,|\beta|\leq n} F_{\alpha,\beta} \Psi_\alpha(\rho, V)\Psi_\beta(\rho, V) \\ &+ \sum_{|\alpha|\leq n} \left(\sum_{|\beta|\leq n} F_{\alpha,\beta} V \cdot \nabla D^\beta \rho \right) (2D^\alpha \nabla \cdot (\rho V) - V \cdot \nabla D^\alpha \rho). \end{aligned} \quad (4.32)$$

The sum in the large parentheses can be simplified with the help of the chain rule,

$$\sum_{|\beta|\leq n} F_{\alpha,\beta} (V \cdot \nabla D^\beta \rho) = \nabla F_\alpha \cdot V. \quad (4.33)$$

where ∇F denotes the spatial gradient of the composition $F((D^\alpha \rho)_{|\alpha|\leq n})$. Integrating eq. (4.32) yields, with an integration by parts on the last term,

$$\begin{aligned} \int_{\mathbb{R}^d} \sum_{|\alpha|,|\beta|\leq n} F_{\alpha,\beta} D^\alpha(\dot{\rho})D^\beta(\dot{\rho}) dx &= \int_{\mathbb{R}^d} \sum_{|\alpha|,|\beta|\leq n} F_{\alpha,\beta} \Psi_\alpha(\rho, V)\Psi_\beta(\rho, V) dx \\ &- \int_{\mathbb{R}^d} \sum_{|\alpha|\leq n} F_\alpha \nabla \cdot \{(2D^\alpha \nabla \cdot (\rho V) - V \cdot \nabla D^\alpha \rho)V\} dx. \end{aligned} \quad (4.34)$$

The contribution of $\ddot{\rho}$ to the integrand in Eq. (4.31) is given by eq. (4.5):

$$D^\alpha \ddot{\rho} = D^\alpha \nabla \cdot (\nabla \cdot (\rho V)V + \rho V'V - \rho W).$$

We add this term to the summands in last integral of Eq. (4.34). Expanding in components we obtain

$$\begin{aligned} D^\alpha \ddot{\rho} - 2\nabla \cdot ((D^\alpha \nabla \cdot (\rho V)V) + \nabla \cdot (V \cdot \nabla D^\alpha \rho)V) \\ = \sum_{i,j=1}^d D_i \{D^\alpha D_j V_i V_j - 2V_i D^\alpha D_j V_j + V_i V_j D^\alpha D_j\}(\rho) - \sum_{i=1}^d D^\alpha D_i W_i(\rho) \end{aligned} \quad (4.35)$$

$$\begin{aligned} &= \Phi_\alpha(\rho, V) - \Psi_\alpha(\rho, W) - W \cdot \nabla D^\alpha \rho \\ &+ \sum_{i,j=1}^d \left\{ -2[D_i, V_i]D^\alpha D_j V_j + [D_i, V_i]V_j D^\alpha D_j + V_i [D_i, V_j]D^\alpha D_j \right\}(\rho). \end{aligned} \quad (4.36)$$

Singling out the terms where V appear without derivatives,

$$\sum_{i,j=1}^d \{ -[D_i, V_i] V_j D^\alpha D_j + V_i [D_i, V_j] D^\alpha D_j \}(\rho) = ((\nabla \cdot V)V - V'V) \cdot \nabla D^\alpha \rho,$$

we see that Eq. (4.35) becomes

$$\begin{aligned} D^\alpha \ddot{\rho} - 2 \overline{\nabla} \cdot ((D^\alpha \nabla \cdot (\rho V)V) + \nabla \cdot (V \cdot \nabla D^\alpha \rho)V) \\ = \Phi_\alpha(\rho, V) - 2(\nabla \cdot V)\Psi_\alpha(\rho, V) - \Psi_\alpha(\rho, W) - ((\nabla \cdot V)V - V'V + W) \cdot \nabla D^\alpha \rho. \end{aligned} \quad (4.37)$$

Inserting Eqs. (4.34) and (4.37) into Eq. (4.31) yields

$$\begin{aligned} \frac{d^2}{dt^2} \mathcal{F} &= \int_{\mathbb{R}^d} \sum_{|\alpha|, |\beta| \leq n} F_{\alpha, \beta} D^\alpha(\dot{\rho}) D^\beta(\dot{\rho}) + \sum_{|\alpha| \leq n} F_\alpha D^\alpha(\ddot{\rho}) dx \\ &= \int_{\mathbb{R}^d} \sum_{|\alpha|, |\beta| \leq n} F_{\alpha, \beta} \Psi_\alpha(\rho, V) \Psi_\beta(\rho, V) dx \\ &\quad + \int_{\mathbb{R}^d} \sum_{|\alpha| \leq n} F_\alpha \{ \Phi_\alpha(\rho, V) - 2(\nabla \cdot V)\Psi_\alpha(\rho, V) - \Psi_\alpha(\rho, W) \} dx \\ &\quad - \int_{\mathbb{R}^d} \sum_{|\alpha| \leq n} F_\alpha ((\nabla \cdot V)V - V'V + W) \cdot \nabla D^\alpha \rho dx. \end{aligned}$$

Since

$$\sum_{|\alpha| \leq n} F_\alpha \nabla D^\alpha \rho = \nabla F$$

by the chain rule, the last line equals

$$- \int_{\mathbb{R}^d} \nabla F \cdot ((\nabla \cdot V)V - V'V + W) dx = \int_{\mathbb{R}^d} F \nabla \cdot ((\nabla \cdot V)V - V'V + W) dx.$$

The identity $\nabla \cdot ((\nabla \cdot V)V - V'V) = (\nabla \cdot V)^2 - \text{tr}((V')^2)$ completes the proof of Eq. (4.30). \blacksquare

Example 4.3.1 (Carrillo-Slepčev functionals on \mathbb{R}^d). Let (ρ) be a geodesic in $\mathcal{P}^{ac}(\mathbb{R}^d)$. Let \mathcal{F} be given by

$$\mathcal{F}(t) = \int \rho^\beta \|\nabla \rho\|^{2\alpha} dx. \quad (4.38)$$

Using the displacement Hessian formula (4.30), it can be (very tediously) checked that

the displacement Hessian of \mathcal{F} is, assuming that the W term is identically zero,

$$\begin{aligned}
\mathcal{F}''(t) = & \int \rho^\beta \|\nabla \rho\|^{2\alpha} (1 - 2\alpha - \beta) ((\nabla \cdot V)^2 - \text{tr}((V')^2)) \\
& + \rho^\beta \|\nabla \rho\|^{2\alpha} \left\{ (\beta(\beta - 1) + 2\alpha\beta + 2\alpha + 8\alpha(\alpha - 1)) \|\nabla \rho\|^2 (\nabla \cdot V)^2 \right. \\
& \quad + 2\alpha \nabla \rho^T ((V'^T)^2) \nabla \rho + (2\alpha(1 + \beta) + 16) (\nabla \cdot V) \nabla \rho^T V'^T \nabla \rho \\
& \quad + 2\alpha \|V'^T \nabla \rho\|^2 + 2\alpha \rho \nabla \rho \cdot \nabla \text{tr}((V')^2) + 4\alpha \rho \nabla \rho^T V'^T \nabla (\nabla \cdot V) \\
& \quad + 4\alpha \rho \nabla (\nabla \cdot V)^T V'^T \nabla \rho + [2\alpha(2 + \beta) + 8\alpha(\alpha + 1)] \rho (\nabla \cdot V) \nabla \rho^T \nabla (\nabla \cdot V) \\
& \quad \left. + 2\alpha \rho^2 \|\nabla (\nabla \cdot V)\|^2 \right\} \\
& + \rho^\beta \|\nabla \rho\|^{2\alpha-4} 8\alpha(\alpha - 1) \left\{ (\nabla \rho^T V'^T \nabla \rho) (\nabla \rho^T V'^T \nabla \rho) \right. \\
& \quad \left. + (\rho \nabla \rho^T \nabla (\nabla \cdot V)) (\rho \nabla \rho^T \nabla (\nabla \cdot V)) + 2\rho \nabla \rho^T V' \nabla \rho \nabla \rho^T \nabla (\nabla \cdot V) \right\} dx.
\end{aligned}$$

It is not clear whether there are any values of α and β such that this functional is displacement convex. (Perhaps in the classical case on \mathbb{R}^d , where $V = T - I = \nabla \phi$ for some function ϕ , there is a simplification, as V' is a symmetric matrix, but such a simplification is not apparent to me.)

4.4 Epilogue: Issues of regularity

This chapter demonstrates how the Otto calculus can be extended to compute displacement Hessians for functionals depending on derivatives of densities, and shows how such a concept is fruitful, by constructing a new class of displacement convex functionals. But, the question remains as to whether this is merely a formal exercise.

For the functionals found in theorem 4.2.3, the answer is, thankfully, no: these functionals really are displacement convex. An identical approximation scheme as used in [8] carries through for these functionals. If (ρ) is a geodesic on $\mathcal{P}^{ac}(S^1)$ for which the functional $\int G(\rho)|\rho'|^\alpha dx$ is finite, then ρ can be checked to be bounded away from 0 and ∞ , and can be approximated by densities in H^1 , which can then be approximated by smooth densities.

In the more general setting of \mathbb{R}^d with the transport cost induced by a Tonelli Lagrangian, the optimal vector field V corresponding to a geodesic (ρ) will only be locally Lipschitz in the spatial variable. If ρ_0 and ρ_1 are absolutely continuous, the corresponding optimal vector field V will, flowing ρ_0 to ρ_1 , preserve absolute continuity, but will

not, in general, have sufficient regularity to preserve stronger regularity properties which ρ_0 and ρ_1 have. If ρ_0 and ρ_1 are both of class C^∞ , it does not appear to be the case that ρ_t will be, for instance, of class C^1 , at any intermediate time along a geodesic.

When considering functionals of 0th order, however, there is a natural approach to approximate minimizers (ρ, V) in the Eulerian optimal transport problem with smooth solutions to the continuity equation, which will be explored in the next chapter. This will (N.B. Hopefully! This isn't fully sorted out yet...) yield a rigorous Otto calculus for functionals of 0th order.

Finally, although not an issue of regularity, there is one question I would like to highlight.

Question 4.4.1. Let (ρ, V) be a smooth minimizer in the Eulerian optimal transport problem. In computing displacement Hessians, the expression

$$\nabla \cdot ((\nabla \cdot V)V - V'V) = (\nabla \cdot V)^2 - \text{tr}((V')^2)$$

seems to inevitably arise (this is, of course, equal to zero when working on \mathbb{R}). Is there a geometric interpretation of $(\nabla \cdot V)^2 - \text{tr}((V')^2)$ on \mathbb{R}^d when $d > 1$?

Appendix A

Otto Calculus Computations - Riemannian Setting

A.1 Preliminaries

Let g be a Riemannian metric on \mathbb{R}^d . That is, g is a smooth map from \mathbb{R}^d to the space of positive definite $d \times d$ real matrices. The entries of $g(x)$ will be denoted $g_{ij}(x)$. When there is no ambiguity, the argument x will be omitted. In this appendix, Einstein summation notation will be used. The entries of the inverse matrix of g will be denoted g^{ij} .

Let $L : T\mathbb{R}^d \rightarrow \mathbb{R}$ be given by

$$L(x, v) = \frac{1}{2}v^T g(x)v = \frac{1}{2}v^i v^j g_{ij}. \quad (\text{A.1})$$

Let (ρ, V) be a geodesic in Wasserstein space. Let φ be the corresponding potential, defined via

$$-D\varphi_t(x) = D_v L(x, V_t(x)). \quad (\text{A.2})$$

Consider the entropy functional

$$\mathcal{F}(t) = \int \rho \log(\rho) dx. \quad (\text{A.3})$$

If $F(x) = x \log(x)$, then, for $x \geq 0$,

$$\begin{cases} p(x) = xF'(x) - F(x) = x \\ p_2(x) = x^2F''(x) - xF'(x) + F(x) = 0. \end{cases} \quad (\text{A.4})$$

My formal displacement Hessian of \mathcal{F} is

$$\begin{aligned} \mathcal{F}''(t) &= \int \rho^2 F''(\rho) (\nabla \cdot V)^2 \\ &\quad + [F(\rho) - \rho F'(\rho)] [(\nabla \cdot V)^2 - \text{tr}(DV^2) + \nabla \cdot W] dx \end{aligned} \quad (\text{A.5})$$

$$= \int p_2(\rho) (\nabla \cdot V)^2 + p(\rho) [\text{tr}(DV^2) - \nabla \cdot W] dx \quad (\text{A.6})$$

$$= \int \rho [\text{tr}(DV^2) - \nabla \cdot W] dx. \quad (\text{A.7})$$

where various derivatives are the standard derivatives are \mathbb{R}^d . The function W is defined as follows:

Let σ be a solution to the Euler-Lagrange equation

$$\begin{cases} \frac{d}{dt} \nabla_v L(\sigma, \dot{\sigma}) = \nabla_x L(\sigma, \dot{\sigma}) \\ (\sigma, \dot{\sigma})(x, 0) = (x, V_0(x)) \end{cases}$$

Expanding and rearranging the Euler-Lagrange equation yields,

$$\begin{aligned} \ddot{\sigma}(x, t) &= (D_v \nabla_v L(\sigma(x, t), \dot{\sigma}(x, t)))^{-1} \\ &\quad \times [\nabla_x L(\sigma(x, t), \dot{\sigma}(x, t)) - D_x \nabla_v L(\sigma(x, t), \dot{\sigma}(x, t)) \dot{\sigma}(x, t)]. \end{aligned} \quad (\text{A.8})$$

Finally, noting that

$$\begin{cases} \sigma(\sigma_t^{-1}(x), t) = x \\ \dot{\sigma}(x, t) = V(\sigma(x, t), t) \iff \dot{\sigma}(\sigma_t^{-1}(x), t) = V(x, t), \end{cases}$$

the map W is defined

$$W(x, t) = \ddot{\sigma}(z, t) \Big|_{z=\sigma_t^{-1}(x)} \quad (\text{A.9})$$

$$= (D_v \nabla_v L(x, V_t(x)))^{-1} [\nabla_x L(x, V_t(x)) - D_x \nabla_v L(x, V_t(x)) V_t(x)]. \quad (\text{A.10})$$

When there is no ambiguity, the time subscripts will be omitted.

Villani's formal displacement Hessian of \mathcal{F} is

$$\begin{aligned} \mathcal{F}''(t) &= \int (-\Delta\varphi + \nabla A \cdot \nabla\varphi)^2 p_2(\rho) dx \\ &\quad + \int \left[\|\nabla^2\varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A)(\nabla\varphi) \right] p(\rho) dx \end{aligned} \quad (\text{A.11})$$

$$= \int \rho \left[\|\nabla^2\varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A)(\nabla\varphi) \right] dx \quad (\text{A.12})$$

where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm of a matrix and the function A is defined via

$$dx = e^{-A} dvol. \quad (\text{A.13})$$

Since $dvol = \sqrt{\det g} dx$, it follows that $A = \frac{1}{2} \log \det g$.

The derivatives in Villani's formal displacement Hessian are derivatives with respect to the Riemannian metric.

A.2 Coordinate computations

A.2.1 My displacement Hessian

In this subsection, the derivatives are standard derivatives on \mathbb{R}^d .

Derivatives of L in coordinates

Various derivatives of L are computed in coordinates. The gradient with respect to velocity is

$$\nabla_v L(x, v) = \nabla_v \left(\frac{1}{2} v^i v^j g_{ij} \right) = \begin{pmatrix} v^j g_{1j} \\ \vdots \\ v^j g_{dj} \end{pmatrix}. \quad (\text{A.14})$$

Then,

$$D_v \nabla_v L(x, v) = g. \quad (\text{A.15})$$

The gradient in position is

$$\nabla_x L(x, v) = \nabla_x \left(\frac{1}{2} v^i v^j g_{ij} \right) = \frac{1}{2} \begin{pmatrix} v^i v^j \partial_1 g_{ij} \\ \vdots \\ v^i v^j \partial_d g_{ij} \end{pmatrix}. \quad (\text{A.16})$$

and

$$D_x \nabla_v L(x, v) = D_x \frac{1}{2} \begin{pmatrix} v^j g_{1j} \\ \vdots \\ v^j g_{dj} \end{pmatrix} \quad (\text{A.17})$$

$$= \begin{pmatrix} v^j \partial_1 g_{1j} & \cdots & v^j \partial_d g_{1j} \\ \vdots & & \vdots \\ v^j \partial_1 g_{dj} & \cdots & v^j \partial_d g_{dj} \end{pmatrix}. \quad (\text{A.18})$$

Finally,

$$D_x \nabla_v L(x, v) v = \begin{pmatrix} v^j \partial_1 g_{1j} & \cdots & v^j \partial_d g_{1j} \\ \vdots & & \vdots \\ v^j \partial_1 g_{dj} & \cdots & v^j \partial_d g_{dj} \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^d \end{pmatrix} \quad (\text{A.19})$$

$$= \begin{pmatrix} v^i v^j \partial_i g_{1j} \\ \vdots \\ v^i v^j \partial_i g_{dj} \end{pmatrix}. \quad (\text{A.20})$$

W in coordinates

In coordinates, the map W is given by

$$W(x, t) = (D_v \nabla_v L(x, V(x)))^{-1} [\nabla_x L(x, V(x)) - D_v \nabla_x L(x, V(x)) V(x)] \quad (\text{A.21})$$

$$= \begin{pmatrix} g^{11} & \cdots & g^{1d} \\ \vdots & & \vdots \\ g^{d1} & \cdots & g^{dd} \end{pmatrix} \left[\frac{1}{2} \begin{pmatrix} V^i V^j \partial_1 g_{ij} \\ \vdots \\ V^i V^j \partial_d g_{ij} \end{pmatrix} - \begin{pmatrix} V^i V^j \partial_i g_{1j} \\ \vdots \\ V^i V^j \partial_i g_{dj} \end{pmatrix} \right] \quad (\text{A.22})$$

$$= \frac{1}{2} \begin{pmatrix} V^i V^j g^{1k} \partial_k g_{ij} \\ \vdots \\ V^i V^j g^{dk} \partial_k g_{ij} \end{pmatrix} - \begin{pmatrix} V^i V^j g^{1k} \partial_i g_{kj} \\ \vdots \\ V^i V^j g^{dk} \partial_i g_{kj} \end{pmatrix}. \quad (\text{A.23})$$

Noting that $2v^i v^j \partial_i g_{kj} = 2v^i v^j \partial_j g_{ki} = v^i v^j (\partial_i g_{kj} + \partial_j g_{ki})$, equation (A.23) can be rewritten as

$$W(x, t) = \frac{1}{2} V^i V^j \begin{pmatrix} g^{1k} (\partial_k g_{ij} - \partial_i g_{kj} - \partial_j g_{ki}) \\ \vdots \\ g^{dk} (\partial_k g_{ij} - \partial_i g_{kj} - \partial_j g_{ki}) \end{pmatrix} \quad (\text{A.24})$$

$$= V^i V^j \begin{pmatrix} -\Gamma_{ij}^1 \\ \vdots \\ -\Gamma_{ij}^d \end{pmatrix}, \quad (\text{A.25})$$

where Γ_{ij}^l is the Christoffel symbol

$$\Gamma_{ij}^l = \frac{1}{2} g^{lk} (\partial_i g_{kj} + \partial_j g_{ki} - \partial_k g_{ij}). \quad (\text{A.26})$$

My Displacement Hessian in coordinates

From equation (A.7), for $\mathcal{F} = \int \rho \log \rho dx$, my displacement Hessian is

$$\mathcal{F}'' = \int \rho [\text{tr}(DV^2) - \nabla \cdot W] dx. \quad (\text{A.27})$$

In coordinates, $\text{tr}(DV^2) = (\partial_i V^j)(\partial_j V^i)$. From equation (A.25), the formula for W is

$$W(x, t) = V^i V^j \begin{pmatrix} -\Gamma_{ij}^1 \\ \vdots \\ -\Gamma_{ij}^d \end{pmatrix}. \quad (\text{A.28})$$

Therefore, $\nabla \cdot W$ is

$$\nabla \cdot W = -\partial_k (V^i V^j \Gamma_{ij}^k) \quad (\text{A.29})$$

$$= -2(\partial_k V^i) V^j \Gamma_{ij}^k - V^i V^j \partial_k \Gamma_{ij}^k \dots \quad (\text{A.30})$$

Substituting equation (A.30) into (A.7) yields

$$[\text{tr}(DV^2) - \nabla \cdot W] = (\partial_i V^j)(\partial_j V^i) + 2(\partial_k V^i) V^j \Gamma_{ij}^k + V^i V^j \partial_k \Gamma_{ij}^k \quad (\text{A.31})$$

A.2.2 Villani's displacement Hessian

In this section, the derivatives are with respect to the Riemannian metric. Several formulas are recalled. The **Christoffel symbol of the second kind** is

$$\Gamma_{ij}^k = \frac{1}{2}g^{kl}(\partial_i g_{lj} + \partial_j g_{li} - \partial_l g_{ij}) \quad (\text{A.32})$$

and, when $j = k$, this simplifies to

$$\Gamma_{ik}^k = \frac{1}{2}g^{jk}\partial_i g_{kj}. \quad (\text{A.33})$$

The **Christoffel symbol of the first kind** is

$$\Gamma_{lij} = g_{lk}\Gamma_{ij}^k. \quad (\text{A.34})$$

Christoffel symbols satisfy the symmetry relations

$$\Gamma_{ij}^k = \Gamma_{ji}^k \quad \text{and} \quad \Gamma_{lij} = \Gamma_{lji}. \quad (\text{A.35})$$

The vanishing of the covariant derivative of the Riemannian metric yields the identity

$$0 = \partial_l g_{ik} - g_{km}\Gamma_{il}^m - g_{im}\Gamma_{kl}^m \iff \partial_l g_{ik} = g_{km}\Gamma_{il}^m + g_{im}\Gamma_{kl}^m = \Gamma_{kil} + \Gamma_{ikl}. \quad (\text{A.36})$$

The **Ricci tensor** is

$$R_{ji} = \partial_k \Gamma_{ji}^k - \partial_i \Gamma_{jk}^k + \Gamma_{kl}^k \Gamma_{ji}^l - \Gamma_{il}^k \Gamma_{jk}^l. \quad (\text{A.37})$$

Derivatives of A in coordinates

The volume distortion map A is given by $A = \frac{1}{2} \log \det g = \log \sqrt{\det g}$, so its i^{th} partial derivative is

$$\partial_i A = \frac{1}{\sqrt{\det g}} \frac{1}{2\sqrt{\det g}} \partial_i \det g. \quad (\text{A.38})$$

By Jacobi's formula,

$$\partial_i (\det g) = \det g (\text{tr}(g^{-1} \partial_i g)), \quad (\text{A.39})$$

where $\partial_i g_{kl}$ is the (k, l) -entry of $\partial_i g$. Substituting (A.39) into (A.38) yields

$$\partial_i A = \frac{1}{2} \text{tr}(g^{-1} \partial_i g) \quad (\text{A.40})$$

$$= \frac{1}{2} g^{jk} \partial_i g_{kj} \quad (\text{A.41})$$

Equation (A.41) follows from equation (A.40) because

$$\text{tr}(g^{-1} \partial_i g) = g^{jk} \partial_i g_{kj}. \quad (\text{A.42})$$

Equation (A.41) can be rewritten as

$$\partial_i A = \Gamma_{ik}^k. \quad (\text{A.43})$$

Then, using the formula for the Hessian of a scalar function with respect to a Riemannian metric, the Hessian of A can be written in three forms

$$\nabla^2 A = \begin{cases} \partial_i \partial_j A - \Gamma_{ij}^k \partial_k A = \partial_i \Gamma_{jl}^l - \Gamma_{ij}^k \Gamma_{kl}^l & (\text{A.44}) \\ \frac{1}{2} (\partial_i \partial_j A + \partial_j \partial_i A) - \Gamma_{ij}^k \partial_k A = \frac{1}{2} (\partial_i \Gamma_{jl}^l + \partial_j \Gamma_{il}^l) - \Gamma_{ij}^k \Gamma_{kl}^l & (\text{A.45}) \\ \partial_j \partial_i A - \Gamma_{ij}^k \partial_k A = \partial_j \Gamma_{il}^l - \Gamma_{ij}^k \Gamma_{kl}^l & (\text{A.46}) \end{cases}$$

The term $Ric + \nabla^2 A$ in (A.12) is called the **Bakry-Emery tensor**. Using equation (A.44) to write $\nabla^2 A$ in coordinates, the Bakry-Emery tensor can be written in the form

$$Ric + \nabla^2 A = R_{ij} + (\partial_i \Gamma_{jl}^l - \Gamma_{ij}^k \Gamma_{kl}^l) \quad (\text{A.47})$$

$$= (\partial_k \Gamma_{ij}^k - \partial_j \Gamma_{ik}^k + \Gamma_{kl}^k \Gamma_{ij}^l - \Gamma_{jl}^k \Gamma_{ik}^l) + (\partial_i \Gamma_{jl}^l - \Gamma_{ij}^k \Gamma_{kl}^l) \quad (\text{A.48})$$

$$= \partial_k \Gamma_{ij}^k - \Gamma_{jl}^k \Gamma_{ik}^l \quad (\text{A.49})$$

Derivatives of φ in coordinates

The potential φ is defined via

$$V_t(x) = \nabla_p H(x, -D\varphi_t(x)) \iff -D\varphi = D_v L(x, V(x)). \quad (\text{A.50})$$

Then, the i^{th} coordinate of $\nabla \varphi$ is given by raising indices:

$$(\nabla \varphi)^i = g^{ij} \partial_j \varphi. \quad (\text{A.51})$$

Since $L(x, v) = \frac{1}{2}v^i v^j g_{ij}(x)$, the j^{th} component of $D\varphi$ is

$$-\partial_j \varphi = (-D\varphi)_j = (D_v L(x, V(x)))_j = g_{ji} V^i. \quad (\text{A.52})$$

Therefore,

$$(\nabla \varphi)^i = -V^i \quad (\text{A.53})$$

Using equation (A.52), the Hessian of φ can be written in three forms:

$$\nabla^2 \varphi = \begin{cases} \partial_i \partial_j \varphi - \Gamma_{ij}^k \partial_k \varphi = -\partial_i (g_{jl} V^l) + \Gamma_{ij}^k g_{kl} V^l & (\text{A.54}) \\ \frac{1}{2}(\partial_i \partial_j \varphi + \partial_j \partial_i \varphi) - \Gamma_{ij}^k \partial_k \varphi = -\frac{1}{2} [\partial_i (g_{jl} V^l) + \partial_j (g_{li} V^l)] + \Gamma_{ij}^k g_{kl} V^l & (\text{A.55}) \\ \partial_i \partial_j \varphi - \Gamma_{ij}^k \partial_k \varphi = -\partial_j (g_{li} V^l) + \Gamma_{ij}^k g_{kl} V^l. & (\text{A.56}) \end{cases}$$

Lemma A.2.1. *Let g and V be as above. Then,*

$$(\partial_i g_{jl}) V^l - (\partial_j g_{il}) V^l = g_{il} (\partial_j V^l) - g_{jl} (\partial_i V^l) \quad (\text{A.57})$$

Proof. By the symmetry of mixed partial derivatives,

$$\partial_i \partial_j \varphi = \partial_j \partial_i \varphi. \quad (\text{A.58})$$

Substituting equation (A.52) into equation (A.58) yields

$$\partial_i (g_{jl} V^l) = \partial_j (g_{il} V^l).$$

Applying the product rule and rearranging completes the proof. \blacksquare

The adjoint of $\nabla^2 \varphi$ can similarly be written in three forms, corresponding to equations (A.54), (A.55), and (A.56):

$$\nabla^2 \varphi^* = \begin{cases} g^{is} [-\partial_s (g_{rl} V^l) + \Gamma_{sr}^k g_{kl} V^l] g^{rj} & (\text{A.59}) \\ g^{is} \left[-\frac{1}{2} (\partial_s (g_{rl} V^l) + \partial_j (g_{ls} V^l)) + \Gamma_{sr}^k g_{kl} V^l \right] g^{rj} & (\text{A.60}) \\ g^{is} [-\partial_r (g_{ls} V^l) + \Gamma_{sr}^k g_{kl} V^l] g^{rj} & (\text{A.61}) \end{cases}$$

The Hilbert-Schmidt norm of $\nabla^2\varphi$ is given by

$$\|\nabla^2\varphi\|_{HS}^2 := \text{tr}(\nabla^2\varphi^*\nabla^2\varphi) \quad (\text{A.62})$$

$$= (\nabla^2\varphi^*)^{ij} (\nabla^2\varphi)_{ji} \quad (\text{A.63})$$

Using the forms of $\nabla^2\varphi$ and $\nabla^2\varphi^*$ from equations (A.54) and (A.59),

$$(\nabla^2\varphi^*)^{ij} = g^{is} [-\partial_s (g_{rl}V^l) + \Gamma_{sr}^k g_{kl}V^l] g^{rj} \quad (\text{A.64})$$

$$(\nabla^2\varphi)_{ji} = -\partial_j (g_{il}V^l) + \Gamma_{ji}^k g_{kl}V^l. \quad (\text{A.65})$$

Substituting equation (A.65) into (A.63) yields

$$\|\nabla^2\varphi\|_{HS}^2 = g^{is} [-\partial_s (g_{rl}V^l) + \Gamma_{sr}^k g_{kl}V^l] g^{rj} [-\partial_j (g_{in}V^n) + \Gamma_{ji}^k g_{kn}V^n] \quad (\text{A.66})$$

$$= g^{is} [-\partial_s g_{rl} V^l - g_{rl} (\partial_s V^l) + g_{kl} \Gamma_{sr}^k V^l] g^{rj} \quad (\text{A.67})$$

$$\times [-\partial_j g_{in} V^n - g_{in} (\partial_j V^n) + g_{kn} \Gamma_{ji}^k V^n]$$

The summand in (A.67) consisting of only derivatives of V is

$$g^{is} (-g_{rl} (\partial_s V^l)) g^{rj} (-g_{in} (\partial_j V^n)) = g^{is} g_{il} (\partial_j V^l) g^{rj} g_{rn} (\partial_s V^n) \quad (\text{A.68})$$

$$= (\partial_j V^s) (\partial_s V^j). \quad (\text{A.69})$$

Recalling the definition of the Christoffel symbol of the first kind from equation (A.34), the summand in (A.67) containing only undifferentiated V terms is

$$g^{is} [-\partial_s g_{rl} V^l + g_{kl} \Gamma_{sr}^k V^l] g^{rj} [-\partial_j g_{in} V^n + g_{kn} \Gamma_{ji}^k V^n] \quad (\text{A.70})$$

$$= g^{is} [-\partial_s g_{rl} V^l + \Gamma_{lsr} V^l] g^{rj} [-\partial_j g_{in} V^n + \Gamma_{nji} V^n] \quad (\text{A.71})$$

From the identity given in (A.36),

$$\begin{cases} -\partial_s g_{rl} = -\Gamma_{lrs} - \Gamma_{rls} \\ -\partial_j g_{in} = -\Gamma_{nij} - \Gamma_{inj}. \end{cases} \quad (\text{A.72})$$

Substituting (A.72) into equation (A.71) yields

$$g^{is} [- (\partial_s g_{rl}) V^l + \Gamma_{lsr} V^l] g^{rj} [- (\partial_j g_{in}) V^n + \Gamma_{nji} V^n] \quad (\text{A.73})$$

$$= g^{is} [-\Gamma_{lrs} - \Gamma_{rls} + \Gamma_{lsr}] V^l g^{rj} [-\Gamma_{nij} - \Gamma_{inj} + \Gamma_{nji}] V^n \quad (\text{A.74})$$

$$= [-g^{rj} \Gamma_{rls}] V^l [-g^{is} \Gamma_{inj}] V^n \quad (\text{A.75})$$

$$= \Gamma_{ls}^j \Gamma_{nj}^s V^l V^n. \quad (\text{A.76})$$

The summand in equation (A.67) containing cross terms (products of V with derivatives of V) is

$$\begin{aligned} & - g^{is} [- (\partial_s g_{rl}) V^l + g_{kl} \Gamma_{sr}^k V^l] g^{rj} g_{in} (\partial_j V^n) \\ & + g^{is} [-g_{rl} (\partial_s V^l)] g^{rj} [- (\partial_j g_{in}) V^n + g_{kn} \Gamma_{ji}^k V^n]. \end{aligned} \quad (\text{A.77})$$

In the first summand, there is a product of the form $g^{is} g_{in}$ and in the second summand, there is a product of the form $g^{rj} g_{rl}$. Contracting indices and substituting in Christoffel symbols of the first kind allows (A.77) to be rewritten

$$\begin{aligned} & - g^{is} [- (\partial_s g_{rl}) V^l + g_{kl} \Gamma_{sr}^k V^l] g^{rj} g_{in} (\partial_j V^n) \\ & + g^{is} [-g_{rl} (\partial_s V^l)] g^{rj} [- (\partial_j g_{in}) V^n + g_{kn} \Gamma_{ji}^k V^n] \end{aligned} \quad (\text{A.78})$$

$$\begin{aligned} & = - [- (\partial_n g_{rl}) V^l + g_{kl} \Gamma_{nr}^k V^l] g^{rj} (\partial_j V^n) \\ & + g^{is} [- (\partial_s V^j)] [- (\partial_j g_{in}) V^n + g_{kn} \Gamma_{ji}^k V^n] \end{aligned} \quad (\text{A.79})$$

$$= -g^{rj} (\partial_j V^n) [- (\partial_n g_{rl}) + \Gamma_{lnr}] V^l - g^{is} [(\partial_s V^j)] [- (\partial_j g_{in}) + \Gamma_{nji}] V^n. \quad (\text{A.80})$$

Using the identities in (A.72) and the symmetry relation (A.35), it follows that

$$\begin{cases} - (\partial_n g_{rl}) + \Gamma_{lnr} = -\Gamma_{lrn} - \Gamma_{rln} + \Gamma_{lnr} = -\Gamma_{rln} \\ - (\partial_j g_{in}) + \Gamma_{nji} = -\Gamma_{nij} - \Gamma_{inj} + \Gamma_{nji} = -\Gamma_{inj}. \end{cases} \quad (\text{A.81})$$

Substituting the identities in (A.81) into equation (A.80) yields

$$- g^{rj} (\partial_j V^n) [- (\partial_n g_{rl}) + \Gamma_{lnr}] V^l - g^{is} [(\partial_s V^j)] [- (\partial_j g_{in}) + \Gamma_{nji}] V^n \quad (\text{A.82})$$

$$= g^{rj} \Gamma_{rln} (\partial_j V^n) V^l + g^{is} \Gamma_{inj} (\partial_s V^j) V^n \quad (\text{A.83})$$

$$= \Gamma_{ln}^j (\partial_j V^n) V^l + \Gamma_{nj}^s (\partial_s V^j) V^n. \quad (\text{A.84})$$

Relabelling indices and using the symmetry of the Christoffel symbol in equation (A.84)

allows for a simplification:

$$\Gamma_{ln}^j (\partial_j V^n) V^l + \Gamma_{nj}^s (\partial_s V^j) V^n = 2\Gamma_{ij}^k (\partial_k V^j) V^i. \quad (\text{A.85})$$

The Hilbert Schmidt norm of $\nabla^2\varphi$ is the sum of the terms (A.76), (A.76), and (A.85). That is,

$$\|\nabla^2\varphi\|_{HS}^2 = \underbrace{(\partial_j V^s) (\partial_s V^j)}_{\text{derivative terms}} + \underbrace{\Gamma_{jl}^k \Gamma_{ik}^l V^j V^i}_{\text{no derivative terms}} + \underbrace{2\Gamma_{ij}^k (\partial_k V^j) V^i}_{\text{cross terms}}$$

Villani's displacement Hessian in coordinates

Recall from (A.12) that, for $\mathcal{F} = \int \rho \log \rho dx$, Villani's displacement Hessian is

$$\mathcal{F}''(t) = \int \rho \left[\|\nabla^2\varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A) (\nabla\varphi) \right] dx. \quad (\text{A.86})$$

From equation (A.49), in coordinates,

$$\text{Ric} + \nabla^2 A = \partial_k \Gamma_{ij}^k - \Gamma_{jl}^k \Gamma_{ik}^l.$$

Since $(\text{Ric} + \nabla^2 A) (\nabla\varphi) := (\text{Ric} + \nabla^2 A)_{ij} (\nabla\varphi)^i (\nabla\varphi)^j$ and $(\nabla\varphi)^i = -V^i$,

$$(\text{Ric} + \nabla^2 A) (\nabla\varphi) = (\partial_k \Gamma_{ij}^k - \Gamma_{jl}^k \Gamma_{ik}^l) V^i V^j.$$

From equation (A.86),

$$\|\nabla^2\varphi\|_{HS}^2 = (\partial_j V^s) (\partial_s V^j) + \Gamma_{jl}^k \Gamma_{ik}^l V^j V^i + 2\Gamma_{ij}^k (\partial_k V^j) V^i.$$

Therefore, the term $\|\nabla^2\varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A) (\nabla\varphi)$ in the integrand in Villani's displacement Hessian (A.12) is

$$\begin{aligned} \|\nabla^2\varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A) (\nabla\varphi) &= [(\partial_j V^s) (\partial_s V^j) + \Gamma_{jl}^k \Gamma_{ik}^l V^j V^i + 2\Gamma_{ij}^k (\partial_k V^j) V^i] \\ &\quad + (\partial_k \Gamma_{ij}^k - \Gamma_{jl}^k \Gamma_{ik}^l) V^i V^j \end{aligned} \quad (\text{A.87})$$

$$= (\partial_j V^s) (\partial_s V^j) + 2\Gamma_{ij}^k (\partial_k V^j) V^i + \partial_k \Gamma_{ij}^k V^i V^j \quad (\text{A.88})$$

A.3 Comparison of Displacement Hessians

From equations (A.12) and (A.88), Villani's displacement Hessian takes the form

$$\mathcal{F}''(t) = \int \rho \left[\|\nabla^2 \varphi\|_{HS}^2 + (\text{Ric} + \nabla^2 A) (\nabla \varphi) \right] dx \quad (\text{A.89})$$

$$= \int \rho \left[(\partial_j V^s) (\partial_s V^j) + 2\Gamma_{ij}^k (\partial_k V^j) V^i + \partial_k \Gamma_{ij}^k V^i V^j \right] dx \quad (\text{A.90})$$

And from equations (A.7) and (A.31), my displacement Hessian takes the form

$$\mathcal{F}'' = \int \rho \left[\text{tr}(DV^2) - \nabla \cdot W \right] dx \quad (\text{A.91})$$

$$= \int \rho \left[(\partial_i V^j) (\partial_j V^i) + 2\Gamma_{ij}^k (\partial_k V^i) V^j + (\partial_k \Gamma_{ij}^k) V^i V^j \right] dx. \quad (\text{A.92})$$

This result is recorded in a theorem.

Theorem A.3.1. *In the Riemannian setting, my displacement Hessian for the entropy functional agrees with Villani's formula for the displacement Hessian in Formula 15.7 in [39].*

Bibliography

- [1] Luigi Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Inventiones Mathematicae **158** (2004), no. 2, 227–260.
- [2] Dominique Bakry and Michel Émery, *Diffusions Hypercontractives*, Séminaire de Probabilités XIX 1983/84, Springer, 1985, pp. 177–206.
- [3] Jean-David Benamou and Yann Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik **84** (2000), no. 3, 375–393.
- [4] Patrick Bernard and Boris Buffoni, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society **9** (2007), 85–121.
- [5] Yann Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on pure and applied mathematics **44** (1991), no. 4, 375–417.
- [6] Haim Brezis, *Functional analysis, sobolev spaces and partial differential equations*, Springer, 2010.
- [7] Giuseppe Buttazzo, Mariano Giaquinta, and Stefan Hildebrandt, *One-dimensional Variational Problems: An Introduction*, Clarendon Press Oxford, 1998.
- [8] José A Carrillo and Dejan Slepčev, *Example of a displacement convex functional of first order*, Calculus of Variations and Partial Differential Equations **36** (2009), no. 4, 547–564.
- [9] Francis Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, vol. 264, Springer, 2013.
- [10] Maria Colombo, Gianluca Crippa, and Stefano Spirito, *Renormalized solutions to the continuity equation with an integrable damping term*, Calculus of variations and partial differential equations **54** (2015), no. 2, 1831–1845.

- [11] Dario Cordero-Erausquin, Wilfrid Gangbo, and Christian Houdré, *Inequalities for Generalized Entropy and Optimal Transportation*, Contemporary Mathematics **353** (2004), 73–94.
- [12] Dario Cordero-Erausquin, Robert J McCann, and Michael Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Inventiones mathematicae **146** (2001), no. 2, 219–257.
- [13] Gianluca Crippa and Camillo De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, Journal für die reine und angewandte Mathematik (Crelles Journal) **2008** (2008), no. 616, 15–46.
- [14] Sara Daneri and Giuseppe Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM Journal on Mathematical Analysis **40** (2008), no. 3, 1104–1122.
- [15] Ronald J DiPerna and Pierre-Louis Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Inventiones Mathematicae **98** (1989), no. 3, 511–547.
- [16] Albert Fathi and Alessio Figalli, *Optimal Transportation on Non-compact Manifolds*, Israel Journal of Mathematics **175** (2010), no. 1, 1–59.
- [17] Alessio Figalli, Luigi Ambrosio, and Cédric Villani, *Optimal Transportation and Action-minimizing Measures*, Edizioni della Normale, 2008.
- [18] Alessio Figalli, Wilfrid Gangbo, and Türkay Yolcu, *A Variational Method for a Class of Parabolic PDEs*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie V **10** (2011), no. 1, 207.
- [19] Alessio Figalli and Nicola Gigli, *Local semiconvexity of Kantorovich potentials on non-compact manifolds*, ESAIM: Control, Optimisation and Calculus of Variations **17** (2011), no. 03, 648–653.
- [20] Gerald B Folland, *Real analysis: modern techniques and their applications*, John Wiley & Sons, 2013.
- [21] Wilfrid Gangbo and Robert J McCann, *The geometry of optimal transportation*, Acta Mathematica **177** (1996), no. 2, 113–161.
- [22] J.K. Hale, *Ordinary differential equations*, Dover Books on Mathematics Series, Dover Publications, 2009.

- [23] Leonid Vitalevich Kantorovich, *On a problem of Monge (English translation)*, Journal of Mathematical Sciences **133** (2006), no. 4, 1383–1383.
- [24] Leonid Vitalievich Kantorovich, *On the translocation of masses*, Dokl. Akad. Nauk SSSR, vol. 37, 1942, pp. 199–201.
- [25] John M Lee, *Introduction to Smooth Manifolds*, 1 ed., 2012.
- [26] Paul WY Lee, *Displacement interpolations from a Hamiltonian point of view*, Journal of Functional Analysis **265** (2013), no. 12, 3163–3203.
- [27] Robert J McCann, Personal Communication.
- [28] ———, *A convexity theory for interacting gases and equilibrium crystals*, Ph.D. thesis, Princeton University, 1994.
- [29] ———, *A Convexity Principle for Interacting Gases*, Advances in mathematics **128** (1997), no. 1, 153–179.
- [30] ———, *Polar factorization of maps on Riemannian manifolds*, Geometric & Functional Analysis GAFA **11** (2001), no. 3, 589–608.
- [31] Robert J McCann and Nestor Guillen, *Five Lectures on Optimal Transportation: Geometry, Regularity and Applications*, Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the Séminaire de Mathématiques Supérieure (SMS) Montréal (2011), 145–180.
- [32] Gaspard Monge, *Mémoire sur la théorie des déblais et des remblais*, De l’Imprimerie Royale, 1781.
- [33] James R Munkres, *Analysis on Manifolds*, Westview Press, 1997.
- [34] Felix Otto, *The geometry of dissipative evolution equations: the porous medium equation*, (2001).
- [35] Felix Otto and Cédric Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, Journal of Functional Analysis **173** (2000), no. 2, 361–400.
- [36] Felix Otto and Michael Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM Journal on Mathematical Analysis **37** (2005), no. 4, 1227–1255.

- [37] Walter Schachermayer and Josef Teichmann, *Characterization of optimal transport plans for the Monge-Kantorovich problem*, Proceedings of the American Mathematical Society **137** (2009), no. 2, 519–529.
- [38] Cédric Villani, *Topics in Optimal Transportation*, no. 58, American Mathematical Society, 2003.
- [39] ———, *Optimal Transport: Old and New*, vol. 338, Springer, 2008.
- [40] Robert M Wald, *General relativity*, University of Chicago press, 2010.