June 2022

AN OPTIMAL TRANSPORTATION THEORY FOR INTERACTING PATHS

Rene Cabrera
University of Massachusetts Amherst

AN OPTIMAL TRANSPORTATION THEORY FOR INTERACTING PATHS

A Dissertation Presented

by

RENÉ CABRERA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2022

Department of Mathematics and Statistics

AN OPTIMAL TRANSPORTATION THEORY FOR INTERACTING PATHS

A Dissertation Presented

by

RENÉ CABRERA

Approved as to style and content by:

Nestor Guillen, Chair

Alejandro Morales, Member

Enrique Suárez, Member

Jun Kitagawa, Member

Nathaniel Whitaker, Department Head
Mathematics and Statistics

Dedication

To my love and support, Cynthia

ACKNOWLEDGMENTS

My debts are substantial: Foremost and first I want to thank my advisor, Nestor Guillen, for believing in me. I am indebted to him for his patience, guidance, and support. From helping me pass the quals to writing a thesis. Working with Nestor has been enjoyable and mathematically stimulating. I have certainly learned a lot of math from him (and that math is not pretentious). This thesis would have not been completed if it wasn't for Nestor's encouragement, knowledge, pedagogy, and helpful and encouraging discussions, and for being responsible for shaping me to become a mathematician. Thanks, Nestor, for your support, guidance, helpful and insightful discussions ranging from mathematical research, academic career goals advice and gentle politics. Although Nestor moved to Texas, we continued to work together for the past three years virtually and through workshop visits and a long research visit to Austin. Nestor's guidance and advice and assistance have surely left their mark.

I owe much thanks to my committee members for agreeing to be part of my thesis defense committee: I would like to express my deepest gratitude to Jun Kitagawa, Alejandro Morales, and Enrique Suárez. Thanks to Jun for making me aware of a standard construction in optimal transport known as dynamical couplings, and for inviting me to the Canadian Mathematical Society Meeting in *applications to optimal transport* to give a talk on December 2020, and for writing me a letter of recommendation for postdoc position to which I am sure helped me get the postdoc.

Also, thanks for supporting me in the NSF's MPS-ASCEND fellowship application and agreeing to be one of my sponsors. Thanks to Alejandro for being such a kind friend to me and Cynthia and making us feel included. Also for writing me a letter of recommendation for postdoc positions and the NSF-MPS-ASCEND fellowship which, I believe, helped me get both. I really appreciated Enrique's positive email congratulating me on my first preprint, and a big thanks to him for the chat we had on March ten on zoom and giving me sage advice on my postdoc position. I would also like to thank Russell Schwab (from the math dept. at Michigan State University), my advisor's co-author for writing a meticulous project summary and agreeing to be my sponsor for an NSF funding. Albeit I personally never met them, I corresponded with them via email and they were pleased and eager to sponsor me—and displayed enthusiam to work with me—for the NSF-MPS-ASCEND fellowship, to which I was awarded but declined due to an RTG postdoctoral position I had already accepted earlier.

The University of Massachusetts at Amherst (UMass) has been my academic home for the past six years and nine months. On August, 2015, I came to UMass with a very nice funding package for my first year from the Graduate School's Northeast Alliance for Graduate Education and the Professoriate (NEAGEP) program. The program helped me achieve my sub-goal of doing well my first semester. The director at the time, Dr Heyda Martinez, provided me, and my cohort, with monthly dinners to decompress and share our first year experiences. The next academic year Dr Sandra L. Petersen, from the Molecular Neuroendocrinology of Reproduction department at UMass, came to help direct NEAGEP. She provided me with a support system I needed when I couldn't pass my quals. She went above and beyond to help me cope with stress, and was prepared to aid me if I couldn't move forward passing quals. I did, however! Much thanks go to both of them for their help and support.

From time to time administrations go under bureaucratic changes and NEAGEP was not immune to this change. Subsequently, NEAGEP morphed into two, but united, vital programs: The Reel Scholars and The Spaulding-Smith Fellows from the Graduate Schools' Office of Inclusion and Engagement program (OIE). Thanks goes to Dr. Wilmore Webley, from the department of microbiology, who is the associate dean of the program. He always gives sound advice and it's really nice to know someone is there to listen—and roots for you. Thanks also goes to Danielle Litwin-Moskowitz, assistant director of the OIE. I was granted the Spaulding-Smith Fellowship for my last academic year. It awarded me $25 000 for the Fall of 2021 to Spring 2022. This funding helped write my first research article, expand my research topic, apply to academic jobs, and finish my dissertation. OIE continued the tradition to have monthly luncheons. Thanks to these luncheons I have met incredible people, in particular non-black people of color and people of color who share my background.

The mathematics and statistics department has been supportive of my academic and teaching work. Special thanks to the department for employing Cynthia. Thanks go to Ilona Trousdale and Terry Mullen for making the department very efficient and giving me math classes I really wanted to teach such as differential equations. Thanks go to Jacob Lagerstrom for quickly helping me get Nestor Guillen's signature in the Graduate School's book record as my advisor. Kaitlyn O'Konis did a fantastic job in the bureaucratic aspect of my dissertation program and kept me updated on the due dates and the paper work, thanks very much! I appreciate professor Nathaniel Whitaker for sharing with Cynthia and I his story and being able to connect with him. Also thanks for the ride from the department tea to our apartment! Several colleagues of mine made grad school bearable and quite enjoyable; they're the following. A thank you to Brian Burrell for selecting me to be a math grad TA for the, at the time, South Central Scholars program summer 2018 and Thrive scholars in Summer 2019, and for writing me a teaching letter of recommendation. I would also like to thank Filip Dul, Dionyssios Mantzavinos, Tetsuya Nakamura, Sebastián Torres, Sanket Sabnis, and my roommates (for easing the task of finding an apt remotely from LA) Sean Hart and Hannah Choi. These folks were part of my support system and always made me feel like I was home.

In Summer 2006, I attended Pasadena City College in California; there is where a mathematical spark was born. MathJam, a program that helped students in basic math skills, was responsible for that spark. I met great faculty there: Warm gratitude goes to Professor of mathematics Lyman Chaffee for always being there—for me—in office hours to help me with my calculus and discrete homeworks, and for his patience every time I asked questions. And honestly for his sincerity in taking a genuine interest in my success even after so many years. Special thanks go to Dr. Brock Klein, director of MathJam; the staff in the Teaching and Learning Center for the support, and Mathematics Engineering Science Achievement program for providing a safe space until 10 PM to do work, and for gifting me a generous $10,000 scholarship to take to UCLA. Thanks go to Professor Jude Thaddeus Socrates for teaching me multivariable calculus and linear algebra. From UCLA, I wish to express my sincerest gratitude to Professor Robert F. Brown (Bob) for accepting my REU application and helping me solve a cool research problem in algebraic topology; and, for believing me. I really appreciated Bob's expert advice on grad school and what comes after. Special thanks to my Alhambra High School literature teacher Mr. Wayne Hooper for inspiring me academically, to read more books, and be a better student. Dr Tran—a substitute American history teacher my junior year—is responsible for getting me to think about a PhD in general, thank you!

I want to express my deepest gratitude to my parents Mario René Cabrera and Esperánza de Cabrera López who, although came from Guatemala and México with very little to no education, new of the importance of having one. Thanks go to my siblings: Silvia Cabrera, Rocío Cabrera, Elvis Cabrera, and Virginia-Ruby Cabrera. In particular, thanks to Elvis for sketching the picture of Lemma 3.14, well done! It was not easy at all to move across the country 3000 miles from Los Angeles, CA to Amherst, MA; leave family behind and study a PhD. I thank them for their remote support. Lastly but certainly not least I want to thank my wife Cynthia Cabrera for coming live with me in Amherst, MA shortly after our marriage; and seriously making my life way, way better. I really couldn't have finished the dissertation without her support and encouragement. Thank you Cynthia for both helping me submit my first preprint to the arxiv and figuring how to transfer the department's template to overleaf. I sincerely appreciate her listening to all my math talks before hand and helping me prepare for them in advance, and attending them. She's my singleton set groupie.

I am a nontraditional student. I didn't know what *college* meant during high school. As soon as I learnt, in my junior and senior year, what college was all about, I enrolled in Pasadena City College and after four years transferred to UCLA and within three years got my bachelors in math. Then I applied to PhD programs and got rejected from every college I applied to. Soon after, I enrolled to CalState LA where I earned a masters, and after two years I applied to UMass and got admitted. Having to overcome many hurdles in UMass and learning how to accept and cope with failure, after seven years, I am finishing my thesis (just now). Knowing this about myself and how long it took me to pursue my dream of studying a PhD in

math, I truly and sincerely believe that anyone, that wants to, can learn math and have the potential to go to graduate school. We just learn at different paces. Just keep your nose to the grindstone and eyes on the prize. The last thing I want to say is that I could have never been able to do this on my own; someone (a group of people) was (were) always there for me and help me pave the way.

ABSTRACT

AN OPTIMAL TRANSPORTATION THEORY FOR INTERACTING PATHS

MAY 2022

RENE CABRERA,

B.S., UNIVERSITY OF CALIFORNIA LOS ANGELES

M.S., CALIFORNIA STATE UNIVERSITY LOS ANGELES

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Nestor Guillen

In this work we study a modification of the Monge-Kantorovich problem taking into account path dependence and interaction effects between particles. We prove existence of solutions under mild conditions on the data, and after imposing stronger conditions, we characterize the minimizers by relating them to an auxiliary Monge-Kantorovich problem of the more standard kind. With this notion of how particles interact and travel along paths, we produce a dual problem. The main novelty here is to incorporate an interaction effect to the optimal path transport problem. This covers for instance, N-body dynamics when the underlying measures are discrete.

Lastly, our results include an extension of Brenier's theorem on optimal transport maps.

TABLE OF CONTENTS

ACKNOWLEDGMENTS v

ABSTRACT xi

LIST OF FIGURES xv

CHAPTER

1. INTRODUCTION 1

   1.1 Background and classical optimal transportation 1
   1.2 Summary of main results 7
   1.3 Outline of the thesis 16

2. MINIMAL PATHS 18

   2.1 Length of paths 18
   2.2 Lagrangian and minimal paths 23

3. MINIMIZERS AND THEIR PROPERTIES 31

   3.1 Existence of minimizers 31
   3.2 The dual problem and the endpoint cost function 35
   3.3 Potentials and cyclically monotone sets 41
   3.4 Optimal plans given by maps 53

4. THE OPTIMAL PATH PROBLEM WITH INTERACTION 59

   4.1 Existence of minimizers for the optimal path with interaction 59
   4.2 The dual problem 66
   4.3 Supergradient of a concave function 70
   4.4 The effective cost 73

5. DISCRETE OPTIMAL TRANSPORTATION WITH INTERACT-ING PATHS 79

   5.1 Cyclical monotonicity with interacting paths 79
   5.2 Cyclical monotonicity on quadratic measures 81
   5.3 N-body dynamics with interacting paths 84

6. A BENAMOU-BRENIER THEORY FOR INTERACTING PATHS 89

   6.1 Classical Benamou-Brenier theory reviewed 91
   6.2 Fluid mechanics with interacting paths 95

APPENDICES

A. BOCHNER'S THEOREM 101

B. BOCHNER-SCHWARTZ THEOREM 107

C. THE DIFFERENTIABILITY OF THE END POINT COST FUNCTION 110

BIBLIOGRAPHY 114

LIST OF FIGURES

Figure Page

1: This picture indicates how the pull-back mapping $T := (e_0, e_1)^{-1}(W) \cap \Omega_{min}$ (the above tubo) is obtained from $W := B_r(x_i) \times B_r(y_i)$ (the product of balls centered about x_i and y_i)........48

CHAPTER 1

INTRODUCTION

1.1 Background and classical optimal transportation

In 1781 Gaspard Monge initiated the problem of how to transfer mass from an initial location onto a final location in the most efficient way possible [24]. He interpreted this problem mathematically using Euclidean geometry. To wit, the cost of transferring a unit of mass from location x to location y was interpreted as the Euclidean distance $c(x, y) := |x-y|$. But this turned out to be quite difficult to solve. In the mid 20th century, Kantorovich studied a relaxation of this problem, that reduces it to a linear optimization problem [13]. In the early 90's, Brenier's work [5] gave a new impetus to the field which has considerably expanded and matured in the last three decades. The interested reader can find more about the history of the field, for example in the books by Villani [24], [25] and Santambrogio [19]. As a result, the optimal transportation in its modern formulation (see below) is known as the Monge-Kantorovich problem (MKP). Monge set the precedent to study the optimal path transport problem to incorporate particle trajectories [24]. Benamou and Brenier, however, [2] were the ones who intentionally reintroduced the time dependent variable to the optimal transport problem in the case of the quadratic cost function. The optimal transport problem may also be viewed as a distance problem between two probability measures, and the time-dependent minimization problem may be viewed as a minimal path problem [24, Ch 5].

In this thesis, we introduce a variant of the MKP that incorporates path dependence and interaction effects. We study a new minimization problem where we add to $\mathcal{E}_0(\pi)$ an interaction term: $\int_\Omega \mathcal{U}(\gamma, \pi) d\pi(\gamma)$. We prove solutions (using the Condition 1) of:

$$\inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} \mathcal{E}_0(\pi) + \int_\Omega \mathcal{U}(\gamma, \pi) d\pi(\gamma) \tag{1.1}$$

exist. A similar result to Theorem 1.6 applies to this new formulation incorporating interaction (Theorem 1.9). We also formulate a duality to characterize solutions of (1.1). Defining an effective cost (4.14) with interaction, we prove something very similar to Theorem 1.7; namely, Theorem 1.11 which says that the optimal plan is given by a general optimal map of paths characterized by a general result of Brenier, Gangbo and McCann [25], [5], [10].

Instead of looking at a cost function $c(x, y)$, that represents how much it costs to send a unit of mass at point x to point y, we consider continuous paths γ such that $\gamma(0) = x$ indicates the initial point along γ and $\gamma(1) = y$ the arrival point along γ, and associate to such a path

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

indicating *how much transportation along that path costs*. Here, L is a Lagrangian, e.g., $L(\gamma(t), \dot{\gamma}(t), t) = \|\dot{\gamma}(t)\|^2 - V(\gamma(t), t)$ (more on this in Chapter 2). More concretely, if Ω denotes the set of such continuous rectifiable paths, then the total cost of transporting along all paths in the plan is the functional:

$$\mathcal{E}_0(\pi) := \int_\Omega c(\gamma) d\pi(\gamma).$$

This can be thought of as an explicit path dependent version of the MKP. Monge set some precedent to this line of reasoning [16]. In this manuscript we explore this further and show it reduces to the traditional optimal transport problem. Motivated by the modeling of traffic networks, Carlier, Jimenez, and Santambrogio [6] studied a related transport problem involving paths, as well. The way [6] models congestion effects is different from the present paper: in [6] the objective functional considers intensity through paths while here the functional involves an interaction potential. The other novelty in our work comes from introducing an interaction term. Probability measures over the space of paths are frequently applied in both probability theory and mathematical physics. For example, Hynd recently considered such measures in the study of 1D sticky particle systems [12]. Whether this or related PDE models could benefit from the point of view in this paper is an interesting question. At any rate, our optimal path problem with interaction term reduces to Brenier's result and thus solves our optimal path problem with interaction effect (1.1).

Let us briefly review the modern description of Monge's problem for the quadratic cost function, $c(x, y) = |x-y|^2$. One considers two density functions $f(x), g(y) \geq 0$ on $\mathbb{R}^n$. Then if we have two probability measures μ on $X \subset \mathbb{R}^n$ and ν on $Y \subset \mathbb{R}^n$, then $\mu(x) = f(x)\ dx$ and $\nu(y) = g(y)\ dy$. Suppose T is any (Borel) measurable function of $X \subset \mathbb{R}^n$ to $Y \subset \mathbb{R}^n$ such that T *pushes* μ forward to ν. This is denoted by $T_\sharp\mu = \nu$ and it means that for any measurable (Borel) subset

$$B \subset Y,\ \nu(B) := T_\sharp\mu(B) = \mu(T^{-1}(B)) \text{ and where } T^{-1}(B) := \{x \in \mathbb{R}^n : T(x) \in B\}$$

Equivalently, $\int_{T^{-1}(B)} f(x)dx = \int_B g(y)dy$ for all Borel sets B. When T is measure preserving ($T_\sharp\mu = \nu$), using the change of variables formula for any continuous function $h \in C^0(\bar{Y})$,

$$\int_Y h(y)\, d\nu(y) = \int_X h(T(x))\, d\mu(x) \tag{1.2}$$

is another characterization.

Monge's problem is: Minimize the total *transportation cost*

$$\int_X c(x, T(x)) d\mu(x) = \int_X \|x - T(x)\|^2\, d\mu(x) \tag{1.3}$$

among all T pushing forward μ to ν [1].

In 1940 Leonid Kantorovich [14] in some sense "relaxed" Monge's problem [1] and [24] by introducing a linear program formulation for the problem. Concretely, again for the quadratic cost function, he considered probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with left and right marginals μ, ν, respectively; namely, $\mu[A] = \pi[A \times \mathbb{R}^n]$ and $\pi[\mathbb{R}^n \times B] = \nu[B]$ for any (Borel) measurable subsets A, B of $\mathbb{R}^n$. An equivalent criterion for π to have left and right marginals μ, ν is the following linearity of π [24]:

$$\forall(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu), \qquad \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \psi(y)]\, d\pi(x, y) = \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y).$$

Kantorovich's problem is: Minimize

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2\, d\pi(x, y) \tag{1.4}$$

among all π having marginals μ, ν. The set of such measures is denoted by $\Pi(\mu, \nu)$, and it is never empty, as it contains the product measure $\mu \otimes \nu$; also $\Pi(\mu, \nu)$ is convex. Let us show this in the next proposition.

Proposition 1.1 *The set*

$$\Pi(\mu, \nu) := \{\pi : \mu[A] = \pi[A \times \mathbb{R}^n],\ \pi[\mathbb{R}^n \times B] = \nu[B];\ \textit{for all measurable subsets } A, B \subset \mathbb{R}^n\}$$

is convex.

Proof. That $\Pi(\mu,\nu) \neq \emptyset$, follows from μ and ν being probability measures and considering the product measure $\pi := \mu \otimes \nu$,

$$\pi[A \times \mathbb{R}^n] := \mu \otimes \nu[A \times \mathbb{R}^n] = \mu[A] \otimes \nu[\mathbb{R}^n] = \mu[A]$$

(since $\nu\,[\mathbb{R}^n] = 1$). A similar calculation for ν works the same.

For convexity, let $\pi_0, \pi_1 \in \Pi(\mu,\nu)$. Then $\mu[A] = \pi_i[A \times \mathbb{R}^n]$ and $\pi_i[\mathbb{R}^n \times B] = \nu[B]$ for all $i = 0, 1$. Consider, for all $0 \leq t \leq 1$, the probability $\pi_t = (1-t)\pi_0 + t\pi_1$. Then we have, for all measurable subsets A, B

$$\begin{aligned}\pi_t[A \times \mathbb{R}^n] &= ((1-t)\pi_0 + t\pi_1)[A \times \mathbb{R}^n] = (1-t)\pi_0[A \times \mathbb{R}^n] + t\pi_1[A \times \mathbb{R}^n] \\ &= (1-t)\mu[A] + t\mu[A] = \mu[A],\end{aligned}$$

and

$$\begin{aligned}\pi_t[\mathbb{R}^n \times B] &= ((1-t)\pi_0 + t\pi_1)[\mathbb{R}^n \times B] = (1-t)\pi_0[\mathbb{R}^n \times B] + t\pi_1[\mathbb{R}^n \times B] \\ &= (1-t)\nu[B] + t\nu[B] = \nu[B].\end{aligned}$$

Consequently, π_t lies inside $\Pi(\mu,\nu)$. ◇

So this problem is actually a linear minimization problem with convex constraints. Whenever π satisfies the marginal condition we say π is admissible. A basic result in functional analysis applying continuity and compactness arguments is the existence of minimizers of functionals [1], [19], and [24].

Remark 1.2 *As mentioned earlier, Kantorovich's problem is a relaxation of Monge's problem [1]. To illustrate this, if* $(Id, T) : X \to X \times Y$ *is defined by* $(Id, T)(x) := (x, T(x))$ *and* $T_\sharp\mu = \nu$*, then* $(Id, T)_\sharp\mu \in \Pi(\mu,\nu)$*, i.e., in a sense Kantorovich's problem contains Monge's. More concretely, for all measurable subsets* $B \subset X$*, we* *have*

$$\begin{aligned}(Id,T)_\sharp\mu[B\times Y] &= \mu\left[(Id,T)^{-1}(B\times Y)\right]\\ &= \mu\left[\{x\in B:\ x\in B\}\right]\\ &= \mu[B],\end{aligned}$$

and for all measurable subsets $A\subset Y$ *such that* $T_\sharp\mu[A]=\nu[A]$*,*

$$\begin{aligned}(Id,T)_\sharp[X\times A] &= \mu\left[(Id,T)^{-1}(X\times A)\right]\\ &= \mu\left[\{x\in X:\ T(x)\in A\}\right]\\ &= \mu\left[T^{-1}(A)\right] = T_\sharp\mu[A]\\ &= \nu[A].\end{aligned}$$

An important property we require of costs on paths is that of *coercivity*, to prove existence of minimizers for (1.5) for energy $c(\gamma)=\int_0^1 L(\dot\gamma,\gamma,t)\,dt$ with Lagrangian $L(\dot\gamma,\gamma,t)=\frac{1}{2}|\dot\gamma|^2-V(\gamma(t),t)$. This is to, geometrically speaking, avoid paths that are very large or that oscillate a lot. In this case the interpretation is that the cost $c(\gamma)$ would be rather large and thus would be too costly, and we wish to eschew this in our theory (see Condition 1 for further details).

The optimal path (Kantorovich) problem is to minimize

$$\inf_{\pi\in\Pi_{\mathrm{path}}(\mu_0,\mu_1)}\int_\Omega c(\gamma)\,d\pi(\gamma), \tag{1.5}$$

this problem admits a solution under general conditions (Theorem 1.6). This is analogues to a result that extends Brenier [5] and Gangbo and McCann [10]. Indeed it turns out that these minimizers are given by maps (Theorem 1.7)

$$\Gamma: X\times[0,1]\longrightarrow\Omega,$$

with the properties

$$\Gamma(x,0)=x\quad\text{for all}\quad x\in X;\quad \Gamma(x,1)=T(x),\quad T_\sharp\mu_0=\mu_1.$$

For a map T that solves an auxiliary transport problem, the optimal measure will be given by $\pi_\Gamma := (\Gamma)_\sharp \mu_0$ and solves the Kantorovich's problem (1.5) uniquely. Here, μ_0 is absolutely continuous with respect to Lebesgue; and, in turn T solves the Monge's path dependent problem for the auxilary cost, $c_{\mathbf{e}}(x, y)$, (2.2). For all x in the support of μ_0, the minimal path is of the form $t \mapsto \gamma(t; x, y)$, just like in the content of Theorem 1.7. And thus for $T(x)$ the optimal map pushing μ_0 forward to μ_1 with respect to c_e, then the optimal map of paths is given by the composition $\Gamma(x, t) := \gamma(t; x, T(x))$.

We also establish the dual problem in view of these settings to characterize the minimizers of the optimal path problem.

1.2 Summary of main results

Let $X \subset \mathbb{R}^n$ be the closure of a bounded, open, and connected set. The space of probability measures will be denoted by $\mathcal{P}(X)$. Consider the space of continuous paths,

$$\Omega := \{\gamma : [0, 1] \to X \mid \gamma \text{ is continuous } \}.$$

If we endow Ω with a metric $\|\gamma_1 - \gamma_2\| := \max_{0 \leq t \leq 1}\{d(\gamma_1(t), \gamma_2(t))\}$, then Ω becomes a complete metric space. In addition as X is compact, Ω is a Polish. Let $B_R(0) := \{y \in \mathbb{R}^n : \ \|x - y\| < R\}$ be an open ball. For all intents and purposes, we may simply take $X := \overline{B_R(0)} \subset \mathbb{R}^n$ for a large $R > 0$. Let μ_0, $\mu_1 \in \mathcal{P}(X)$.

A well known result in functional analysis states that *if X is compact, then X and $C(X)$ are both separable.* We prove this result and then apply it to the space of continuous paths.

Let us suppose that X is compact. For each m consider the collection of finite balls $\mathcal{B}_m$ of radius $1/m$ covering X. Take the sequences of centers $\{c_k^m\}$ of balls in $\bigcup_{m=1}^{\infty} \mathcal{B}_m$. Thus, the sequence of balls of radius $1/m$ centered at c_k^m for all k, $\{B(c_k^m, 1/m)\}_{k=1}^{\infty}$ is countable. So is the sequence $\{c_k^m\}$ countable. We shall show that this set is dense in X. Fix $x \in X$ and $\varepsilon > 0$. Since for each m the sequence of these balls covers X, there is a finite number N_m so that $X = \bigcup_{k=1}^{N_m} B(c_k, 1/m)$. Then $x \in \{B(c_k^m, 1/m)\}$ so that if we Choose $N > N_m > 1/\varepsilon$ such that for all $m \geq N$, $d(c_k^m, x) < 1/m \leq 1/N < \varepsilon$. Since ε was arbitrary, X is dense.

To show $C(X)$ is dense, we argue as follows. Let X be a compact metric space with corresponding metric d. Let $\{\mathcal{B}_k^{(m)} : 1 \leq k \leq N_m\}$ be a collection of open balls of radius $1/m$ covering X. Let $\{\eta_k^{(m)} : 1 \leq k \leq N_m\}$ be a partition of unity subordinate to $\{\mathcal{B}_k^{(m)} : 1 \leq k \leq N_m\}$. Let Y be a collection of all possible linear combinations of $\{\eta_k^{(m)} : 1 \leq k \leq N_m\}$ with rational coefficients. Then this set is countable. We will prove that this set is dense in $C(X)$. Fix $u \in C(X)$ and $\varepsilon > 0$. As X is compact, u is uniformly continuous so that we can find $\delta > 0$ such that for any pair of points $x_1, x_2 \in X$, $d(x_1, x_2) < \delta/2$ implies $|u(x_2) - u(x_1)| < \varepsilon/2$. Choose m such that $1/m < \delta/2$. Consider the set $\{\mathcal{B}_k^{(m)} : 1 \leq k \leq N_m\}$. If $x_1, x_2 \in \mathcal{B}_k^{(m)}$, then $d(x_1, x_2) \leq 1/m < \delta/2$ implies $|u(x_2) - u(x_1)| < \varepsilon/2$. Pick $x_k \in \mathcal{B}_k^{(m)}$ and $r_k \in \mathbb{Q}$ for all k so that

$$|r_k - u(x_k)| \leq \varepsilon/2.$$

Let $v(x) = \sum_{k=1}^{N_m} r_k \eta_k^{(m)}(x)$, so $v \in Y$. Then for every $x \in X$,

$$|u(x) - v(x)| = \left| \sum_{k=1}^{N_m} u(x)\eta_k^{(m)}(x) - \sum_{k=1}^{N_m} r_k \eta_k^{(m)}(x) \right| \leq \sum_{k=1}^{N_m} \eta_k^{(m)}(x) \left| u(x) - r_k \right|. \quad (1.6)$$

Now, if $x \in \mathcal{B}_k^{(m)}$ we have

$$|u(x) - r_k| \leq |u(x) - u(x_k)| + |u(x_k) - r_k| < \varepsilon.$$

If $x \notin \mathcal{B}_k^{(m)}$, then $\eta_k^{(m)}(x) = 0$ and there is nothing to show. Hence, from (1.6), $|u(x) - v(x)| \leq \sum_{k=1}^{N_m} \eta_k^{(m)}(x)\varepsilon = \varepsilon$. Then $\|u - v\| < \varepsilon$, and since ε was arbitrary, Y is dense in $C(X)$ and so $C(X)$ is separable. In particular, we reiterate the statement that if X is compact, then Ω is separable under the topology of uniformly convergence.

Consider probability measures $\pi \in \mathcal{P}(\Omega)$ with the following admissibility condition.

Definition 1.3 *Define $e_t : \Omega \to X$, the evaluation map by*

$$e_t(\gamma) = \gamma(t) \quad \textit{for all} \quad 0 \leq t \leq 1$$

In particular we have $e_0(\gamma) = \gamma(0)$ and $e_1(\gamma) = \gamma(1)$; this merely indicates the initial and final end-points of the path γ, respectively. The admissibility condition on π is now given in the next definition.

Definition 1.4 *Given $\mu_0, \mu_1 \in \mathcal{P}(X)$, we say that $\pi \in \mathcal{P}(\Omega)$ is admissible if the following holds*

$$(e_0)_\sharp \pi = \mu_0, \ (e_1)_\sharp \pi = \mu_1$$

The set of all probability measures $\pi \in \mathcal{P}(\Omega)$ satisfying Definition 1.4 will be denoted by $\Pi_{\text{path}}(\mu_0, \mu_1)$. Such measures π are also known as dynamical couplings. This notion is well known in the classical optimal transport literature, see Villani's discussion in [25, Chapter 7]. Moreover, the set Π_{path} represents transport plans with associated paths and it is reminiscent of the standard admissible measures, in the classical Kantorovich measures $\Pi(\mu, \nu)$. We will revisit this description in Lemma 3.14, where we show a probability measure on the space of paths projects to a solution of the MKP in Euclidean space.

Lemma 1.5 *The set*

$$\Pi_{path}(\mu_0, \mu_1) = \{\pi \in \mathcal{P}(\Omega) : \; (e_0)_\sharp \pi = \mu_0 \quad and \quad (e_1)_\sharp \pi = \mu_1\}$$

is nonempty and convex.

Proof. To show $\Pi_{\text{path}}(\mu_0, \mu_1) \neq \emptyset$, consider $\mu_0 \otimes \mu_1 \in \mathcal{P}(X \times X)$. Let

$$g : X \times X \to \Omega$$

be defined by $g_t(x, y) = \gamma_{x,y}(t)$ for all $0 \leq t \leq 1$, the unique minimal path

$$(x, y) \mapsto \gamma_{x,y} \quad \text{from} \quad \gamma_{x,y}(0) = x \quad \text{to} \quad \gamma_{x,y}(1) = y.$$

Fashion a measure in $\mathcal{P}(\Omega)$,

$$\widehat{\pi} := g_\sharp(\mu_0 \otimes \mu_1).$$

Then the claim is that $\widehat{\pi}$ belongs to $\Pi_{\text{path}}(\mu_0, \mu_1)$. Indeed, for all $B \subset X$,

$$\widehat{\pi}[B] = g_\sharp(\mu_0 \otimes \mu_1)[B] = (\mu_0 \otimes \mu_1)\left[g^{-1}(B)\right].$$

Then

$$(e_0)_\sharp \widehat{\pi}[B] = \widehat{\pi}\left[e_0^{-1}(B)\right] = (\mu_0 \otimes \mu_1)\left[g^{-1}(e_0^{-1}(B))\right] = \mu_0[B].$$

Similarly applying the same argument one gets the second marginal. So $\widehat{\pi}$ is in $\Pi_{\text{path}}(\mu_0, \mu_1)$ and thus it's nonempty.

For the convexity, fix π_0 and π_1 in $\Pi_{\text{path}}(\mu_0, \mu_1)$. Take

$$\pi_t := (1 - t)\pi_0 + t\pi_1 \quad \forall 0 \leq t \leq 1.$$

We show $\pi_t \in \Pi_{\text{path}}(\mu_0, \mu_1)$. For every measurable subset $B \subset X$ and using the linearity of π_t: $\pi_t[B] = (1-t)\pi_0[B] + t\pi_1[B]$,

$$\begin{aligned}(e_0)_\sharp \pi_t[B] &= (1-t)(e_0)_\sharp \pi_0[B] + t(e_0)_\sharp \pi_1[B] \\ &= (1-t)\mu_0[B] + t\mu_0[B] = \mu_0[B].\end{aligned}$$

Similarly, for all measurable $A \subset X$,

$$\begin{aligned}(e_1)_\sharp \pi_t[A] &= (1-t)(e_1)_\sharp \pi_0[A] + t(e_1)_\sharp \pi_1[A] \\ &= (1-t)\mu_1[A] + t\mu_1[A] = \mu_1[A].\end{aligned}$$

$\diamond$

Let $c : \Omega \to \mathbb{R}$ be a cost function. Define the linear functional $\mathcal{E}_0 : \mathcal{P}(\Omega) \to \mathbb{R}$ by

$$\mathcal{E}_0(\pi) := \int_\Omega c(\gamma) d\pi(\gamma).$$

The *optimal path problem* that we wish to study is then:

Problem A. Given $\mu_0, \mu_1 \in \mathcal{P}(X)$, minimize $\mathcal{E}_0(\pi)$ among all $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$.

Costs satisfying the following two conditions will prove essential:

Condition 1 *The function $c : \Omega \to \mathbb{R}$ is bounded from below, lower semi-continuous, and it has the following coercivity property: given any two positive numbers M and N, the set*

$$\Omega_{M,N} := \{\omega \in \Omega \mid |\omega(0)| \leq M,\ c(\omega) \leq N\},$$

is a compact subset of Ω.

Condition 2 *The function $c : \Omega \to \mathbb{R}$ is of the form*

$$c(\gamma) = \int_0^1 \tfrac{1}{2}|\dot{\gamma}|^2 - V(\gamma(t), t)\, dt,$$

where $V : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ is bounded and has spatial first and second derivatives bounded uniformly in t, with $\nabla_x V(x,t)$ satisfying an L-Lipschitz condition with $L < 2/3$.

An example of a path (function) that oscillates a lot near the origin for which it does not satisfy Condition 1 is $\omega(t) = t^2 \sin\left(\frac{1}{t^2}\right)$ on $[0,1]$ with $V(\omega(t),t) = 0$. Its derivative exists on $[0,1]$ but ω' is not Lebesgue integrable on $[0,1]$. Thus, $\omega \notin \Omega_{M,N}$. In particular, c does not satisfy the bound $c(\omega) \leq N$. Essentially since $L^2([0,1]) \subset L^1([0,1])$ by the Cauchy-Schwarz inequality and the fact that the interval $[0,1]$ has finite measure, we will show $\int_0^1 \omega'$ blows up at infinity, and if ω' is not Lebesgue integrable, then certainly ω' will fail to be in $L^2([0,1])$. Indeed to prove this, we need to prove that the following is not integrable:

$$\omega'(0) = \begin{cases} 2t\sin\left(1/t^2\right) - 2\frac{1}{t}\cos\left(1/t^2\right) & \text{for} \quad t \in (0,1] \\ 0 & \text{for} \quad t = 0 \end{cases}$$

The sum of integrable functions will be integrable. Then ω' is not integrable is equivalent to $-\frac{1}{t}\cos(1/t^2)$ not being integrable. Let

$$f(t) := \left| \frac{1}{t} \cos\left(\frac{1}{t^2}\right) \right|.$$

Noticing that for $y \in [2\pi k - \pi/4, 2\pi k + \pi/4]$,

$$\cos(y) \geq \frac{1}{\sqrt{2}}.$$

For integers $k \geq 1$, define

$$a(k) := \frac{1}{\sqrt{2\pi k - \pi/4}}; \quad b(k) := \frac{1}{\sqrt{2\pi k + \pi/4}}.$$

Then, $f(t) \geq \frac{1}{\sqrt{2}}\frac{1}{a(k)}$ for $t \in [b(k), a(k)]$. If $I_k := [b(k), a(k)]$, then

$$f(t) \geq \sum_{k=1}^{K} \frac{1}{\sqrt{2}} \frac{1}{a(k)} \chi_{I_k}(t),$$

for arbitrary K. Therefore,

$$\int f \geq \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \left(1 - \frac{b(k)}{a(k)}\right)$$
$$= \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \left(1 - \sqrt{\frac{2\pi k - \pi/4}{2\pi k + \pi/4}}\right).$$

Applying the inequality $\frac{x}{2} \leq 1 - \sqrt{1-x}$, for $x \in [0,1]$, for $x = \frac{\pi/2}{2\pi k + \pi/4}$ we get

$$\int f \geq \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \frac{\pi/4}{2\pi k + \pi/4}$$
$$= \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \frac{1}{8k+1}$$
$$\geq \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \frac{1}{8(k+1)}.$$

The latter sum diverges as K tends to infinity; we are done.

Our first Theorem (Theorem 1.6) is on existence under general conditions.

Theorem 1.6 *Suppose μ_0 and μ_1 have compact support and that c satisfies Condition 1, then **Problem A** has at least one solution.*

Now under stronger conditions, the optimal plan ends up being unique, and it is characterized by a general map solving Monge's problem, this follows both Brenier's [25] and Gangbo's and McCann's results [10]:

Theorem 1.7 *Suppose μ_0 and μ_1 have compact support, $\mu_0 << dx$, and c satisfies Condition 2, then there is at most one solution to **Problem A** given by an optimal map, $\Gamma : spt(\mu_0) \to \Omega$ (see Section 3.4).*

For the next problem we add an interaction term to $\mathcal{E}_0(\pi)$. Given a continuous function $\mathcal{K} : \Omega \times \Omega \to \mathbb{R}$ define $\mathcal{U} : \Omega \times \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ by

$$\mathcal{U}(\gamma, \pi) := \int_{\Omega} \mathcal{K}(\gamma, \sigma)\, d\pi(\sigma). \tag{1.7}$$

Thus the new linear functional with interaction that we will study is the following:

$$\mathcal{E}(\pi) := \mathcal{E}_0(\pi) + \int_\Omega \mathcal{U}(\gamma, \pi)\ d\pi(\gamma) = \int_\Omega c(\gamma) d\pi(\gamma) + \int_\Omega \left(\int_\Omega \mathcal{K}(\gamma, \sigma) d\pi(\sigma) \right) d\pi(\gamma). \tag{1.8}$$

The function $\mathcal{K}(\gamma, \sigma)$ can be thought of as measuring interactions between γ and σ. Then $\mathcal{U}(\gamma, \pi)$ measures the total interaction between γ and a distribution of paths given by π. The integral term $\int_\Omega \mathcal{U}(\gamma, \pi)\ d\pi(\gamma)$ in (1.1) is the "total" cost or energy from these interactions.

The new optimal transportation problem with interacting paths is thus:

Problem B. Given $\mu_0, \mu_1 \in \mathcal{P}(X)$, minimize (1.1), given by (1.8), among all $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$.

This problem proved to be both interesting and subtle. Interesting because the objective functional (1.8) is not linear in π, and subtle because it might not even be convex in general.

As a particular example consider $\mathcal{K}(\gamma, \sigma)$ given by an integral over the time interval $[0, 1]$ of an exponential function:

$$\mathcal{K}(\gamma, \sigma) := \int_0^1 \theta \exp\{-\beta |\gamma(t) - \sigma(t)|^2\} dt, \tag{1.9}$$

for $\beta > 0$ and $\theta > 0$. Note that if γ and σ are very close to each other, $\mathcal{K}(\gamma, \sigma)$ is very close to 0. If, on the other hand, σ and γ are a large distance away from each other, then $\mathcal{K}(\gamma, \sigma)$ will be small. More generally, Bochner's theorem asserts $\mathcal{K}$ is convex if $\mathcal{K}(\gamma, \sigma) = \int_0^1 \kappa(\gamma(t) - \sigma(t))\ dt$, where $\kappa(\gamma(t) - \sigma(t))$ is the Fourier transform of a finite, positive measure, as given in Reed's and Simon's *Funtional Analysis I* book [17, Theorem IX.9]. Indeed, Bochner implies that the set of Fourier transforms of the finite, positive measures on $\mathbb{R}^n$ is exactly the cone of functions of positive type. Please refer to the Appendix A for more on this matter.

Let $d\pi_t(\cdot) := (e_t)_\sharp\pi$. As κ is the Fourier transform of a finite positive measure on $\mathbb{R}^n$, Bochner implies $H(\pi) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\kappa(x-y)d\pi_t(x)d\pi_t(y)$ is convex in π. As a result the quadratic functional in (1.8) is convex in π. Indeed, using Fubini to rearrange the integrals

$$\int_\Omega\int_\Omega\left(\int_0^1\kappa(\gamma(t)-\sigma(t))dt\right)d\pi(\sigma)d\pi(\gamma) = \int_0^1\int_\Omega\int_\Omega\kappa(\gamma(t)-\sigma(t))d\pi(\sigma)d\pi(\gamma)dt$$
$$(d\pi_t = (e_t)_\sharp\pi) = \int_0^1 H(\pi)dt$$

is convex is convex in π. More concretely, for all $0 \le \alpha \le 1$,

$$\int_0^1 H((1-\alpha)\pi_0 + \alpha\pi_1)\,dt \le \int_0^1((1-\alpha)H(\pi_0) + \alpha H(\pi_1))\,dt$$
$$= (1-\alpha)\int_0^1 H(\pi_0)dt + \alpha\int_0^1 H(\pi_1)dt.$$

Remark 1.8 *For concreteness we will focus on the Gaussian interaction but other kernels are covered by our methods as well. For example, $\kappa(x) = |x|^{2-n}$ $(n \ge 3)$. The kernel $\mathcal{K}(\gamma,\sigma) = \int_0^1|\gamma(t)-\sigma(t)|^{2-n}\,dt$ is convex, as the Bochner-Schwartz theorem [17, Theorem IX.10] implies $|\gamma(t)-\sigma(t)|^{2-n}$ is the Fourier transform of a positive measure of at most polynomial growth. The Coulomb repulsion force of interacting particles $\kappa(x-y) = |x-y|^{2-n}$ is an important class of examples that applies to our theory equally well for the interaction term. The Coulomb potential is characterized in a forthcoming paper in collaboration with Nestor Guillen. Appendix B contains Bochner-Schwartz theorem and gives an account on the Coulomb potential.*

In the same spirit to Theorems 1.6 and 1.7, we study a new optimal transportation with interacting paths problem (1.1), and show existence of minimizers of (1.8) and characterize the minimizers. For the remaining Theorems, we will only consider $\mathcal{K}(\gamma,\sigma)$ given by (1.9). We note that, in a sense, **Problem B** contains

Problem A as a special case. In our investigations we first analyzed **Problem A** and using this analysis as a foot-hold we approached **Problem B.**

Theorem 1.9 *Suppose μ_0 and μ_1 have compact support and that c satisfies Condition 1, then the Kantorovich problem with interaction,* ***Problem B****, has at least one solution.*

Theorem 1.10 *A solution to* ***Problem B*** *is a solution to* ***Problem A*** *with some effective cost.*

Theorem 1.11 *Suppose μ_0 and μ_1 have compact support, $\mu_0 << dx$, c satisfies Condition 2, and $\mathcal{K}$ is as in (1.9) for some θ and β. Then, there is $\theta_0 > 0$ depending on θ and β such that if $\theta \in (0, \theta_0)$ then* ***Problem B*** *has a unique solution given by a map.*

1.3 Outline of the thesis

Following the introduction and main results, Chapter 2 looks at minimal paths, Lagrangians, and costs of paths; using Lagrangians with a potential. We supply several results indicating that an energy functional with a potential achieves its minimum on Ω, and that for such a path minimizing an endpoint cost function c_e with endpoints fixed is differentiable. Chapter 2 deals with existence of minimizers and their properties. We prove Theorem 1.6, and produce the dual problem. Lastly, we prove Theorem 1.7. The fourth chapter, Chapter 4, gives an account of the optimal path with interaction effects. We prove existence of minimizers using a modulus of continuity argument and the coercivity property. We acquire the dual path dependence problem with interaction. We also prove Theorems 1.9

and 1.11 in this section. In Chapter 5 we look at a quadratic measure (Definition 5.3) and prove some essential properties that will let us describe cyclically monotonicity with interacting paths. We also formally show, in this chapter, that the minimizers of (5.1) are solutions to a differential equation (Theorem 5.7). The last chapter, Chapter 6, is an account of fluid mechanics' view point of optimal transportation, and an extension of the Benamou-Brenier minimization problem to interacting paths. Three appendices provide Bochner's statement (Appendix A) and Bochner's and Schwartz' (Appendix B) generalization on the convexity of the Coulomb potential and the other Appendix C establishes the differentiability of the end point cost function.

CHAPTER 2

MINIMAL PATHS

We start with a word of caution. The terms "curve" and "path", interchangeably, mean the same thing. In any rate, let $\gamma : [0,1] \to \mathbb{R}^n$ be a Lipschitz continuous path in $[0,1]$ that is twice differentiable. It is intuitively clear, so to speak, that the shortest path on a curve with endpoints x and y, is the straight line emanating from x to y. In order to shed more light on this matter, we will make this a bit more rigorous. Let $\langle \cdot, \cdot \rangle$ denote the Euclidean inner product in $\mathbb{R}^n$ and "$\dot{}$" denote a time derivative.

2.1 Length of paths

Definition 2.1 *Let X be a metric space. If $\gamma : [0,1] \to X$ is a piecewise smooth path segment, the length of a path γ is defined by*

$$L(\gamma) := \int_0^1 |\dot{\gamma}(t)| \, dt \tag{2.1}$$

A curve $\gamma : [0,1] \longrightarrow \mathbb{R}^n$ is said to be a *polygonal curve* if it is piecewise linear, that is: there is a partition $t_0 = 0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = 1$ of $[0,1]$ and points $p_0, p_1, \ldots, p_n$ such that

$$\gamma(t) = p_k + t(p_{k+1} - p_k) \qquad \text{for} \qquad t \in [t_{k-1}, t_k] \quad \text{for} \quad i \in \{1, \ldots, n\}.$$

Note that n is the "number of sides" of the polygonal line (assuming no three successive points are colinear). Among all polygonal curves from x_0 to x_1, the one with the shortest length is always the straight line ($n = 1$).

If we proceed by induction on n, keeping k fixed, we can prove this affirmatively. For the base case $n = 2$ suppose we are given a polygonal curve γ which can be partitioned piecewise linearly by three non colinear points p_0, p_1, p_2 on γ with partition $\mathbf{P} := \{0 = t_0 < t_1 < t_2 =: 1\}$. Then by the triangle inequality we have

$$|p_0 - p_1| = |p_0 - p_2 + p_2 - p_1| \leq |p_0 - p_2| + |p_2 - p_1|,$$

and the line from p_0 to p_1 given by $\gamma(t) := p_0 + t(p_1 - p_0)$ for $t \in [t_1, t_2]$ is the shortest length among the other two polygonal lines on γ going from p_0 to p_2 and then from p_2 to p_1.

Assume inductively for all $n - 1$ with partition $\mathbf{P}' := \{0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} = 1\}$ that

$$|p_0 - p_1| \leq |p_0 - p_2| + \cdots + |p_{n-1} - p_1|,$$

invoking the triangle inequality $n - 1$ times. So that among all polygonal curves from p_0 to p_1, the straight line has the shortest length for all $n - 1$ polygonal lines. Now adjoin a point p_n on γ to acquire the partition $\mathbf{P}'' := \{0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = 1\}$. Note that $\mathbf{P}' \subset \mathbf{P}''$. Applying the triangle inequality to the three non colinear points p_{n-1}, p_n, p_1 on γ, we have

$$|p_{n-1} - p_1| \leq |p_{n-1} - p_n| + |p_n - p_1|.$$

Applying this to the induction hypothesis yields

$$|p_0 - p_1| \leq |p_0 - p_2| + \cdots + |p_{n-1} - p_1| \leq |p_0 - p_2| + \cdots + |p_{n-1} - p_n| + |p_n - p_1|.$$

Thus, among all polygonal curves from p_0 to p_1, the one with shortest length is always the straight line from p_0 to p_1.

To make this thesis as self contained as possible, we go over the details of minimal paths. Let us talk about their length, taken from Carrillo et al in [7]. Namely the following definition.

Definition 2.2 *For a given continuous path $\gamma : [0,1] \to \mathbb{R}^n$ equipped with the supnorm topology, $d(\gamma_1, \gamma_2) := \max_{0 \le t \le 1} \|\gamma_1(t) - \gamma_2(t)\|$, its length $\mathcal{L}(\gamma)$ is defined as a supremum over finite partitions $\mathfrak{P} = \{t_i \ : \ 0 = t_0 < t_1 < \cdots < t_n = 1\}$ by*

$$\mathcal{L}(\gamma) := \sup_{\mathfrak{P} \subset [0,1]} \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})).$$

Notice that the triangle inequality shows $d(\gamma(0), \gamma(1)) \le \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1}))$, and upon taking the supremum over finite partitions on the right-hand side we acquire $d(\gamma(0), \gamma(1)) \le \mathcal{L}(\gamma)$. This is another slick way of showing that among all polygonal piecewise linear curves—with endpoints $\gamma(0) = x$ and $\gamma(1) = y$, the one with “shortest path” is the straight line!

Lemma 2.3 *Given a curve $\gamma : [0,1] \to \mathbb{R}^n$ which is Lipschitz continuous in $[0,1]$ and twice differentiable in $(0,1)$, define*

$$c(\gamma) = \frac{1}{2} \int_0^1 \|\dot{\gamma}(t)\|^2 dt.$$

Fix two points x_0 and x_1 in $\mathbb{R}^n$, and denote by $\Omega(x_0, x_1)$ the set of all γ's such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. Let $\gamma_0 \in \Omega(x_0, x_1)$ be a global minimizer of $c(\gamma)$.
Then, if $\phi : [0,1] \to \mathbb{R}^n$ is Lipschitz, twice differentiable in $(0,1)$, and $\phi(0) = \phi(1) = 0$, then

$$\int_0^1 \langle \dot{\gamma}_0(t), \dot{\phi}(t) \rangle dt = 0$$

Consequently, $\ddot{\gamma}_0 \equiv 0$*, and therefore,* γ_0 *must be equal to* $\gamma_0(t) = x_0 + t(x_1 - x_0)$*, and that* $c(\gamma_0) = \frac{1}{2}|x_1 - x_0|^2$

Proof. The proof stems from the methods of the calculus of variations, most notably the Euler-Lagrange equation. Define $\gamma_\varepsilon = \gamma_0 + \varepsilon\phi$ for $\phi \in C_c^2([0,1])$, namely; ϕ is Lipschitz and twice differentiable on $(0,1)$ and vanishes on the boundary points of the closed interval $[0,1]$. Suppose γ_0 minimizes the "energy" c, that is, γ_0 satisfies $c(\gamma_0) \leq c(\eta)$ for any other $\eta \in \Omega(x_0, x_1)$.

Let $f(\varepsilon) = \gamma_\varepsilon$. So since γ_0 is a minimizer of c, then f is optimal at $\varepsilon = 0$, and $c(\gamma_0)$ achieves its minimum. In other words f has a minimizer at $\varepsilon = 0$. This means

$$\begin{aligned}
0 &= f'(0) \\
&= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} c(\gamma_\varepsilon) \\
&= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \frac{1}{2}\int_0^1 \left\|\frac{d}{dt}\gamma_0(t) + \varepsilon\frac{d}{dt}\phi(t)\right\|^2 dt \\
&= \frac{1}{2}\int_0^1 2\cdot\left(\frac{d}{dt}\gamma_0(t) + \varepsilon\frac{d}{dt}\phi(t)\right)\cdot\dot{\phi}(t)\Big|_{\varepsilon=0} dt \\
&= \int_0^1 \dot{\gamma}_0(t)\cdot\dot{\phi}(t)dt := \int_0^1 \langle\dot{\gamma}_0(t), \dot{\phi}(t)\rangle\, dt
\end{aligned}$$

where we used the fact that ϕ vanishes in the boundary points of $[0,1]$.

As a corollary, we will get $\ddot{\gamma}_0 = 0$. To this end, an application of integration by parts and the divergence theorem give us such result:

$$\begin{aligned} 0 = \int_0^1 \dot{\gamma}_0(t) \cdot \dot{\phi}(t)\; dt &= \int_0^1 \Big(\frac{d}{dt}(\dot{\gamma}_0 \cdot \phi(t)) - \ddot{\gamma}_0(t)\phi(t)\Big) dt \\ &= -\int_0^1 \ddot{\gamma}_0(t) \cdot \phi(t)\; dt. \\ &:= -\int_0^1 \langle \ddot{\gamma}_0(t), \phi(t) \rangle dt. \end{aligned}$$

Since this is true for all $t \in [0,1]$ and all ϕ under the inner product, then it follows at once that $\ddot{\gamma}_0$ is identically zero.

Conclusion: $\ddot{\gamma}_0 = 0$, and therefore γ_0 must equal to

$$\gamma_0(t) = x_0 + t(x_1 - x_0),$$

and that $c(\gamma_0) = \frac{1}{2}|x_0 - x_1|^2$. ◇

Lemma 2.4 *Let γ be another curve such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$ with $\gamma_0(t) = x_0 + t(x_1 - x_0)$. Then*

$$c(\gamma_0) \leq c(\gamma)$$

with equality holding if and only if $\gamma(t) = \gamma_0(t)$ for all t. This shows explicitly that c achieves its minimum in the set $\Omega(x_0, x_1)$.

Proof. Since $\gamma_0 = \gamma$ on the boundary of $[0,1]$, then $\phi := \gamma - \gamma_0$ vanishes on the boundary points of $[0,1]$. Then $\gamma = \phi + \gamma_0$ and

$$\begin{aligned} \frac{1}{2}\int_0^1 \|\dot{\gamma}(t)\|^2 dt &= \frac{1}{2}\int_0^1 \|\dot{\phi}(t) + \dot{\gamma}_0(t)\|^2 dt \\ &= \frac{1}{2}\int_0^1 \|\dot{\phi}(t)\|^2 dt + \int_0^1 \dot{\gamma}_0(t) \cdot \dot{\phi}(t) dt + \frac{1}{2}\int_0^1 \|\dot{\gamma}_0(t)\|^2 dt \\ &= \frac{1}{2}\int_0^1 \|\dot{\phi}(t)\|^2 dt + \frac{1}{2}\int_0^1 \|\dot{\gamma}_0(t)\|^2 dt \\ &\geq \frac{1}{2}\int_0^1 \|\dot{\gamma}_0(t)\|^2 dt. \end{aligned}$$

Therefore, $c(\gamma_0) \leq c(\gamma)$. ◇

A solution of $\ddot{\gamma}_0(t) = 0$ is a minimizing "geodesic" curve from endpoints $\gamma_0(0) = x$ to $\gamma_0(1) = y$, and as we saw in Lemma 2.3, the solution to $\ddot{\gamma}_0$ that minimizes the energy c is given by $\gamma_0(t) = (1-t)x + ty$.

2.2 Lagrangian and minimal paths

In [25], Villani explains a construction in optimal transport that an *action*, $\mathcal{A}$, which measures the cost of displacing along a continuous path γ, defined on a time interval, is used to consider a cost, c, by minimizing the action among paths that go from the initial point of the path, x, to the final point of the path, y,

$$c(x, y) = \inf\{\mathcal{A}(\gamma) : \gamma(0) = x,\ \gamma(1) = y,\ \gamma \in C([0, 1]; \mathbb{R}^n)\}.$$

We drew from this in defining the endpoint cost function (2.2). A typical and classical example of such an action is the kinetic energy, $\mathcal{A}(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2/2dt$. More generally, an action is given by Lagrangians as Villani details in [25]. And following this line of thinking heavily, is where we begin our study of our cost action functional. We cover some results pertaining a Lagrangian $L(\dot{\gamma}, \gamma, t)$ and its action functional, which is given by the time integral of $L(\dot{\gamma}, \gamma, t)$ along the path, which defines a cost function $c(\gamma) := \int_0^1 L(\dot{\gamma}, \gamma, t)\, dt$, just like in Villani's general example of $\mathcal{A}(\gamma)$. The action functional will be interpreted as the cost of that path. We next consider a Lagrangian with a given potential $V(x, t)$. All these preliminary propositions will be used in Chapter 3 and Chapter 4.

The following elementary proposition will be useful in what follows. It indicates how close a path is to a linear path.

Proposition 2.5 *Suppose $\gamma : [0,1] \to \mathbb{R}^n$ is C^2 and $\delta > 0$ is such that*

$$|\ddot{\gamma}(t)| \leq \delta \quad \forall t \in [0,1].$$

Then for every $t \in [0,1]$ we have

$$|\gamma(t) - (\gamma(0) + t[\gamma(1) - \gamma(0)])| \leq \sqrt{n}\delta \quad \textit{and} \quad |\dot{\gamma}(t) - (\gamma(1) - \gamma(0))| \leq \sqrt{n}\delta.$$

Proof. Let $\mathbf{e}(t) := \gamma(t) - (\gamma(0) + t(\gamma(1) - \gamma(0)))$. Then $\mathbf{e}(0) = \mathbf{e}(1) = \mathbf{0}$, and $\mathbf{e}(t)$ is continuously differentiable on $[0,1]$. Furthermore, $\ddot{\mathbf{e}}(t) = \ddot{\gamma}(t)$ and $|\ddot{\mathbf{e}}(t)| \leq \delta$. Therefore for any $0 \leq t_1 < t_2 \leq 1$, we have

$$\dot{\mathbf{e}}(t_2) - \dot{\mathbf{e}}(t_1) = \int_{t_1}^{t_2} \ddot{\mathbf{e}}(t)dt.$$

So $|\dot{\mathbf{e}}(t_2) - \dot{\mathbf{e}}(t_1)| \leq \delta$ for all $0 \leq t_1 < t_2 \leq 1$.

In coordinate components of $\mathbf{e}(t)$, we will show that for each $i = 1, \ldots, n$ there is some $t_i \in [0,1]$ such that $\dot{\mathbf{e}}_i(t_i) = 0$. Thus, in this case, $|\dot{\mathbf{e}}_i(t)| \leq \delta$ for all $t \in [0,1]$. Then

$$|\dot{\mathbf{e}}(t)| \leq \sqrt{\delta^2 + \cdots + \delta^2} = \sqrt{n}\delta.$$

According to the mean value theorem, applied to $\mathbf{e}_i(t_i)$, there exists $\bar{t}_i \in [0,1]$ such that

$$0 = \mathbf{e}_i(1) - \mathbf{e}_i(0) = \dot{\mathbf{e}}_i(\bar{t}_i).$$

Therefore for all $t \in [0,1]$, $\mathbf{e}(t) = \mathbf{e}(0) + \int_0^t \dot{\mathbf{e}}(s)ds = \int_0^t \dot{\mathbf{e}}(s)ds$. Then $|\mathbf{e}(t)| \leq \sqrt{n}\delta$ and the proposition follows at once.

◇

We will focus on Lagrangians of the form

$$L(\dot{\gamma}(t), \gamma(t), t) := \frac{1}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t), t),$$

where $V : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ is a continuously differentiable potential which is bounded from below. Then we consider the cost function

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

We see that this corresponds to the classical Lagrangian $L(x, v) = \frac{1}{2}|v|^2 - V(x,t)$, where v is the velocity (or time-derivative) of the path γ at t, with endpoints x and y of γ fixed and $V(x,t)$ some potential [25]. It is well known that if $V(x,t) \in C^1$, the minimizers of $L(x,v)$ with endpoints fixed satisfy Newton's dynamical equation

$$\frac{d^2x}{dt^2} = -\nabla_x V(x,t).$$

Everything that follows can be done for more general Lagrangians, but we will focus on the Lagrangian above for the sake of concreteness.

Proposition 2.6 *Let $\gamma : [0,1] \to \mathbb{R}^n$ be a Lipschitz path which is twice differentiable in $(0,1)$. Then the minimizers of $c(\gamma)$, with endpoints x and y of γ fixed, satisfy the equation*

$$\ddot{\gamma}(t) = -\nabla_x V(\gamma(t), t)$$

Proof. This is just the Euler-Lagrange equation applied to the cost functional, see [8, Ch 8]. ◇

Notice that when $V \equiv 0$, we get $\ddot{\gamma} = 0$. If V is small enough, then we can expect that minimal γ's are close to straight lines. We now quantify this intuition.

Corollary 2.7 *Suppose the Lagrangian is given by $L(\dot{\gamma}, \gamma, t) = \frac{1}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t), t)$ and $V : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ is continuously differentiable and bounded from below. If $\delta > 0$ is such that $\|\nabla V\|_\infty \leq \delta$, then if γ is minimal,*

$$|\gamma(t) - (\gamma(0) + t[\gamma(1) - \gamma(0)])| \leq \sqrt{n}\delta \quad \textit{and} \quad |\dot{\gamma}(t) - (\gamma(1) - \gamma(0))| \leq \sqrt{n}\delta.$$

Proof. Choose a minimizer γ of $c(\gamma)$. Since $\delta > 0$ is such that $|\nabla V(\gamma(t), t)| \leq \delta$ for all $t \in [0, 1]$, Proposition 2.6 applies to show that $|\ddot{\gamma}(t)| \leq \delta$ for all $t \in [0, 1]$. Then applying Proposition 2.5 establishes the corollary. $\diamond$

Let us show now that for any pair of points x, y, there is a unique minimal γ between them, provided V satisfies a smallness condition.

Proposition 2.8 *Let V be bounded and of class C^2 and such that $\nabla V(x, t)$ satisfies the spatial L-Lipschitz condition for $L < 1$. Then for all $x, y \in \mathbb{R}^n$ there is a unique path $\gamma(\cdot, x, y)$ which minimizes $c(\gamma)$ among all paths from x to y, and this function is Lipschitz continuous in x, y.*

Proof. Let $\omega(t)$ be a solution for the boundary value problem

$$\begin{aligned}
&\ddot{\gamma}(t) = -\nabla_x V(\gamma(t), t), \\
&\gamma(0) = x, \\
&\gamma(1) = y.
\end{aligned}$$

Then $\omega(t)$ will satisfy the integral equation

$$\omega_{x,y}(t) = (1-t)x + ty + t\int_0^1 (1-s)\nabla V(\omega(s), s)\, ds - \int_0^t (t-s)\nabla V(\omega(s), s)\, ds.$$

To establish the uniqueness condition and Lipschitz continuity, fix two pair of points, x_1, y_1 and x_2, y_2. Suppose two such solutions exist, call them $\gamma_{x_1,y_1}(t)$ and $\gamma_{x_2,y_2}(t)$. Then both satisfy the integral equation. Denote $d(\gamma_{x_1,y_1}(t), \gamma_{x_2,y_2}(t)) := \sup_{[0,\tau]} |\gamma_{x_1,y_1}(t) - \gamma_{x_2,y_2}(t)|$. Then we have

$$\begin{aligned}
|\gamma_{x_1,y_1}(t) - \gamma_{x_2,y_2}(t)| \leq\ & |(1-t)(x_1 - x_2) + t(y_1 - y_2)| \\
& + t\int_0^1 (1-s)|\nabla V(\gamma_{x_1,y_1}(s), s) - \nabla V(\gamma_{x_2,y_2}(s), s)|\, ds \\
& + \int_0^t (t-s)|\nabla V(\gamma_{x_1,y_1}(s), s) - \nabla V(\gamma_{x_2,y_2}(s), s)|\, ds
\end{aligned}$$

$$\leq |x_1 - x_2| + |y_1 - y_2| + L \int_0^1 (1-s)|\gamma_{x_1,y_1}(s) - \gamma_{x_2,y_2}(s)|\, ds$$
$$+ L \int_0^t (t-s)|\gamma_{x_1,y_1}(s) - \gamma_{x_2,y_2}(s)|\, ds$$

$$\leq |x_1 - x_2| + |y_1 - y_2| + L \int_0^1 (1-s)\, ds \sup_{[0,t]} |\gamma_{x_1,y_1}(s) - \gamma_{x_2,y_2}(s)|$$
$$+ L \sup_{[0,t]} \int_0^t (t-s)\, ds \ \sup_{[0,t]} |\gamma_{x_1,y_1}(s) - \gamma_{x_2,y_2}(s)|.$$

Noticing that the term with the common factor equals

$$\int_0^1 (1-s)\, ds + \sup_{0\leq t\leq 1} \int_0^t (t-s)\, ds = 1,$$

and after rearranging, the above estimate equals the estimate, since $L < 1$ by assumption

$$d(\gamma_{x_1,y_1}, \gamma_{x_2,y_2}) \leq |x_1 - x_2| + |y_1 - y_2| + L\, d(\gamma_{x_1,y_1}, \gamma_{x_2,y_2})$$
$$\iff d(\gamma_{x_1,y_1}, \gamma_{x_2,y_2}) \leq \frac{1}{1-L}(|x_1 - x_2| + |y_1 - y_2|),$$

we are done with the proof of the proposition. ◇

For a given cost $c : \Omega \to \mathbb{R}$, we introduce the *endpoint* function between an initial point x and a final point y, which is obtained by minimizing the linear functional among paths γ that go from x to y. So for any pair of points $x, y \in \mathbb{R}^n$, let γ

$$c_{\mathbf{e}}(x,y) := \inf_{\gamma(0)=x,\ \gamma(1)=y} c(\gamma). \tag{2.2}$$

Lemma 2.9 *Let $c_e(x,y)$ be as in (2.2). Then c_e is differentiable with respect to x and y. Moreover, $c_e(x,y) = c_e(y,x)$, and $\nabla_y c_e(x,y) = y - x - \int_0^1 t\nabla_x V(\gamma_{x,y}(t), t)\, dt$.*

Proof. Let us first show the symmetric condition. The endpoint cost $c_{\mathbf{e}}(y,x)$ between y and x is acquired by minimizing the linear functional among paths that go from y to x. Let us denote such path as $\overline{\gamma}(t)$. More concretely, let $s : [0,1] \to [0,1]$ be given by $s(t) = 1-t$ for all $t \in [0,1]$. Then $\overline{\gamma}(t) = \gamma(s(t))$ is a path from y to x. Let $1 - t_0 = s(r_0)$ where $r_0 \in [0,1]$ is in the domain of s. Then

$$\begin{aligned}\dot{\overline{\gamma}}(r) = \frac{d}{dr}\overline{\gamma} &= \frac{d}{dr}\gamma(s(r)) \\ &= \frac{d\gamma}{dt}\frac{ds}{dr} \\ &= -\dot{\gamma}(t).\end{aligned}$$

So that then using the change of variables $s(t) = 1-t$, one can see $c(\gamma) = c(\overline{\gamma})$, and then

$$c_{\mathbf{e}}(x,y) = c_{\mathbf{e}}(y,x).$$

To show the differentiability of $c_{\mathbf{e}}(x,y)$ at (x,y) as a function of x and y, by the symmetry condition, it suffices to only show it is differentiable with respect to y. Indeed, pick a minimal path γ_* from $\gamma_*(0) = x$ to $\gamma_*(1) = y$. In this case we have $c_{\mathbf{e}}(x,y) = c(\gamma_*)$. Consider another path, not necessarily minimal, $\gamma_h(t) := \gamma_*(t) + th \cdot \hat{e}$ from x to $y + h\hat{e}$, where $h \neq 0$ and $\hat{e} := (0, \ldots, 1, \ldots, 0)$. From this, we know $c_{\mathbf{e}}(x, y + h\hat{e}) \leq c(\gamma_h(t))$. Putting this together, we get

$$\lim_{h\to 0} \frac{c_{\mathbf{e}}(x, y+h\hat{e}) - c_{\mathbf{e}}(x,y)}{h} \leq \lim_{h\to 0} \frac{c(\gamma_h(t)) - c(\gamma_*(t))}{h}. \tag{2.3}$$

Now, since the cost of the path $\gamma_h(t)$ can be differentiated ($c(\gamma_h)$ is explicit) we know that the right-hand side of (2.3) is well-defined. That the left-hand side limit exists[1] follows from a standard argument (using compactness and uniqueness of solutions to linear ODE's). Then it follows $\langle \nabla_y c_{\mathbf{e}}(x,y), \hat{e}\rangle \leq \frac{d}{dh}|_{h=0} c(\gamma_h(t))$. To establish the reverse inequality, consider another path $\gamma_h(t) = \gamma_*(t) + th(-\hat{e})$ from x to $y - h\hat{e}$. From this, (2.3) becomes

[1]We prove this fact in Appendix C

$$\lim_{h\to 0}\frac{c_{\mathbf{e}}(x,y-h\hat{e})-c_{\mathbf{e}}(x,y)}{h}\leq \lim_{h\to 0}\frac{c(\gamma_h(t))-c(\gamma_*(t))}{h}. \tag{2.4}$$

A change of variables with $z := y - h\hat{e}$; noticing $z \to y$ as $h \to 0$, the left hand side of (2.4) yields the reverse inequality, so that

$$-\lim_{h\to 0}\frac{c_{\mathbf{e}}(x,y+h\hat{e})-c_{\mathbf{e}}(x,y)}{h}=\langle -\nabla_y c_{\mathbf{e}}(x,y),\hat{e}\rangle,$$

while the right hand side of (2.4) with $\gamma_h(t) = \gamma_*(t) - th\cdot\hat{e}$ gives

$$\frac{d}{dh}\Big|_{h=0}c(\gamma_h(t))=\Big\langle -\int_0^1\Big(\dot{\gamma}_*(t)-\nabla_x V(\gamma_*(t),t)t\Big)\,dt,\hat{e}\Big\rangle.$$

Putting this together in (2.3) gives the reverse inequality and thus

$$\langle\nabla_y c_{\mathbf{e}}(x,y),\hat{e}\rangle=\frac{d}{dh}|_{h=0}c(\gamma_h)=\Big\langle\int_0^1\Big(\dot{\gamma}_*(t)-\nabla_x V(\gamma_*(t),t)t\,dt\Big),\hat{e}\Big\rangle.$$

Therefore the differentiability of $c_{\mathbf{e}}(x,y)$ with respect to y follows and therefore $\nabla_y c_{\mathbf{e}}(x,y) = y - x - \int_0^1 t\ \nabla_x V(\gamma_*(t),t)\ dt.$ ◇

Finally, we show $\nabla c_{\mathbf{e}}(x,y)$ satisfies the twist condition from [25].

Lemma 2.10 *Let V be bounded and of class C^2 and let ∇V be L-Lipschitz continuous with $L < 2/3$. For all $y \in \mathbb{R}^n$, and $x_1 \neq x_2$, we have*

$$\nabla_y c_e(x_1,y)\neq\nabla_y c_e(x_2,y)$$

Proof. Lemma 2.9 tells us

$$\nabla_y c_{\mathbf{e}}(x,y)=y-x-\int_0^1 t\ \nabla_x V(\gamma_{x,y}(t),t)\ dt.$$

Now for all y,

$$\nabla_y c_{\mathbf{e}}(x_1,y)-\nabla_y c_{\mathbf{e}}(x_2,y)=x_2-x_1-\int_0^1 t\Big(\nabla_{x_1}V(\gamma_{x_1,y}(t),t)-\nabla_{x_2}V(\gamma_{x_2,y}(t),t)\Big)\ dt.$$

The reverse triangle inequality and invoking Proposition 2.8, most notably the bound at the end of its proof and the Lipschitz condition on ∇V, we have

$$
\begin{aligned}
|\nabla_y c_{\mathbf{e}}(x_1, y) - \nabla_y c_{\mathbf{e}}(x_2, y)| &\geq |x_1 - x_2| - L \int_0^1 t|\gamma_{x_1,y}(t) - \gamma_{x_2,y}(t)| \, dt \\
&\geq |x_1 - x_2| - L\Big(\frac{1}{1-L}\Big)|x_1 - x_2| \int_0^1 t \, dt \\
&= |x_1 - x_2|\Big(1 - \frac{1}{2}\frac{L}{1-L}\Big) \\
&> 0.
\end{aligned}
$$

Since $L < 2/3$. ◇

CHAPTER 3

MINIMIZERS AND THEIR PROPERTIES

3.1 Existence of minimizers

Before showing that minimizers of a linear functional on paths *do* exist, a few helpful results are in order. A common theme in measure theory is how well a measure is stable. For instance in a compact metric space X, if $\{\mu_k\}$ is a sequence of probability measures bounded above by a constant $C > 0$,

$$\mu_k(X) \leq C \quad \text{for all} \quad k,$$

then one can find a subsequence of measures $\{\mu_{k_j}\}$ and another nonnegative measure μ such that $\mu_{k_j} \rightharpoonup \mu$ as $j \to \infty$, i.e. $\int \phi \, d\mu_{k_j} \to \int \phi \, d\mu$ as $j \to \infty$ for all $\phi \in C(X)$. A more general result due to Prokhorov, Theorem 3.2, includes this line of reasoning. First, however, we need the following definition.

Definition 3.1 *Let X be a metric space and let $\{\mu_k\}$ be a sequence of nonnegative Borel measures in X. The sequence is said to be tight if for each $\varepsilon > 0$, there exists a compact set $X_\varepsilon \subset X$ such that*

$$\mu_k(X \setminus X_\varepsilon) \leq \varepsilon \quad \textit{for all} \quad k.$$

Prokhorov's theorem, which can be found in either [24] or [19], is the following.

Theorem 3.2 (Prokhorov) *Let X be a separable metric space, and let $\{\mu_k\}$ be a sequence of nonnegative Borel measures. If the sequence $\{\mu_k\}$ is tight, then there exists a subsequence $\{\mu_{k_j}\}$ and another nonnegative Borel measure μ such that*

$$\mu_{k_j} \rightharpoonup \mu \quad as \quad j \to \infty.$$

We need two additional propositions before we prove existence of minimizers. And Condition 1 will help us with the proof of Theorem 1.6 as we will momentarily see.

Proposition 3.3 *Condition 2 implies Condition 1.*

Proof.

We have $c(\gamma) := \int_0^1 \frac{1}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t),t)\ dt$ is bounded from below and lower semi-continuous. To be more concrete, Proposition 3.4 establishes the lower semi-continuity part. That $c(\gamma)$ is bounded from below follows from

$$\int_0^1 \frac{1}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t),t)\ dt \geq -K,$$

as V is bounded by $K > 0$. Recall the set $\Omega_{M,N}$ from Condition 1. We will show this set is compact for all $N > 0$. To this end, we will show $\Omega_{M,N}$ is uniformly bounded and uniformly continuous, and then invoke Arzela-Ascoli. We first show the latter. Given $\varepsilon > 0$, fix $0 \leq t_2 < t_1 \leq 1$. We have to show $|\gamma(t_1) - \gamma(t_2)| \leq \varepsilon$. Indeed, up to a factor of $\frac{1}{2}$, we have

$$\begin{aligned}
|\gamma(t_1) - \gamma(t_2)| \leq \int_{t_2}^{t_1} |\dot{\gamma}(t)|\ dt &\leq \left(\int_0^1 |\dot{\gamma}(t)|^2\ dt\right)^{1/2} \left(\int_{t_2}^{t_1} 1\ dt\right)^{1/2} \\
&= \left(\int_0^1 |\dot{\gamma}(t)|^2 - V(\gamma(t),t) + V(\gamma(t),t) dt\right)^{1/2} |t_1 - t_2|^{1/2} \\
&\leq \left(\int_0^1 |\dot{\gamma}(t)|^2 - V(\gamma(t),t) dt + K\right)^{1/2} |t_1 - t_2|^{1/2} \\
&\leq \sqrt{N+K}\ |t_1 - t_2|^{1/2}.
\end{aligned}$$

This shows γ has a bounded Hölder seminorm. Now we check $\Omega_{M,N}$ is uniformly bounded. This follows from the fact that $|\gamma(0)| \leq M$ for every $\gamma \in \Omega_{M,N}$ and the modulus of continuity given above. $\diamond$

Proposition 3.4 *The cost function $c(\gamma)$ given by Condition 2 is lower semi-continuous.*

Proof. Assume $\gamma_n \to \gamma$ in the sup-norm in $[0, 1]$, we are going to show $\liminf_{n\to\infty} c(\gamma_n) \geq c(\gamma)$. The cost function $c(\gamma)$ has two terms. The first term satisfies

$$\liminf_{n\to\infty} \int_0^1 |\dot{\gamma_n}(t)|^2\, dt \geq \int_0^1 |\dot{\gamma}(t)|^2\, dt$$

from [8, Ch8.2.2, Theorem 1], as γ is Lipschitz continuous and $|\dot{\gamma}|^2$ smooth, convex and bounded below. While the second term has the potential V which is uniformly continuous; this means that given $\varepsilon > 0$, there is $\delta > 0$ such that

$$\|\gamma_n - \gamma\| \leq \delta \implies |V(\gamma_n, t) - V(\gamma, t)| \leq \varepsilon$$

in $[0, 1]$. In turn this then means that the second term satisfies $\lim_{n\to\infty} \int_0^1 V(\gamma_n, t)\, dt = \int_0^1 V(\gamma, t)\, dt$. Putting this together with the above liminf establishes the result. $\diamond$

We are now ready to prove Theorem 1.6.

Proof. [**Proof of Theorem 1.6**]

Since $c(\gamma) \geq -K$, then $\mathcal{E}_0(\pi) \geq -K$ for all $\pi \in \mathcal{P}(\Omega)$, thus the infimum of $\mathcal{E}_0$ in $\Pi_{\text{path}}(\mu_0, \mu_1)$ is finite. Let $\{\pi_k\}_k$ be a minimizing sequence $\mathcal{E}_0(\pi_k) \longrightarrow \inf_{\pi\in\Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{E}_0(\pi)$. Note that since $\{\pi_k\}_k$ is minimizing, we can find a constant $C > 0$ for which $\mathcal{E}_0(\pi_k) \leq C$ for all k.

Suppose $M > 0$ is such that spt $(\mu_0) \subset B_M(0)$. Define for $N > 0$, $X_N := \{\gamma \in \Omega \mid c(\gamma) \leq N, N > 0\}$ then, for every admissible π we have thus

$$\pi(X_N) = \pi(\Omega_{M,N}).$$

One can see this from the fact that $(e_0)_\sharp \pi = \mu_0$ implies that spt $(\pi) \subset e_0^{-1}(\text{spt } (\mu_0))$. So the equality follows from π being supported in the set $\{\gamma \in \Omega : \ |\gamma(0)| \leq M\}$. In particular,

$$\pi(\Omega_{M,N}^c) = 1 - \pi(\Omega_{M,N}) = 1 - \pi(X_N) = \pi(X_N^c).$$

We have $c(\gamma) > N$ in X_N^c, so

$$\pi(X_N^c) = \frac{1}{N} \int_{X_N^c} N \, d\pi(\gamma) \leq \frac{1}{N} \int_{X_N^c} c(\gamma) \, d\pi(\gamma),$$

which implies $\mathcal{E}_0(\pi)/N \geq \pi(X_N^c)$. Applying this to each π_k,

$$\pi_k(\Omega_{M,N}^c) \leq \frac{1}{N} \mathcal{E}_0(\pi_k) \leq \frac{C}{N} \quad \text{for all} \quad k \tag{3.1}$$

Returning to inequality (3.1), for each $N > 0$ given $\varepsilon := \frac{C}{N} > 0$,

$$\pi_k\left(\Omega_{M,N}^c\right) \leq \frac{C}{N} = \varepsilon \quad \text{for all} \quad k.$$

By Condition 1, $\Omega_{M,N}$ is compact, then $\{\pi_k\}_k$ is tight. Hence Prokhorov (Theorem 3.2) says there exists a subsequence $\{\pi_{k_j}\}_j \subset \Pi_{\text{path}}(\mu_0, \mu_1)$ and another Borel probability measure ρ such that

$$\pi_{k_j} \rightharpoonup \rho \quad \text{as} \quad j \to \infty.$$

The claim is that then $\rho \in \Pi_{\text{path}}(\mu_0, \mu_1)$. Recall the change of variables formula (1.2). We will use an extension of this with the evaluation map to show ρ lies in $\Pi_{\text{path}}(\mu_0, \mu_1)$. Given any test function $\phi \in C_b(\Omega)$,

$$\begin{aligned} \int_\Omega \phi(x) \, d((e_0)_\sharp \rho(x)) = \int_\Omega \phi(\gamma(0)) \, d\rho(\gamma) &= \lim_{j \to \infty} \int_\Omega \phi(\gamma(0)) d\pi_{k_j}(\gamma) \\ &= \lim_{j \to \infty} \int_\Omega \phi(\gamma) d((e_0)_\sharp \pi_{k_j}) \\ &= \int_\Omega \phi(\gamma) \, d\mu_0. \end{aligned}$$

From the arbitrariness of $\phi \in C_b(\Omega)$ we get $(e_0)_\sharp \rho = \mu_0$. Similarly, by the same argument we can show $(e_1)_\sharp \rho = \mu_1$. Thus, $\rho \in \Pi_{\text{path}}(\mu_0, \mu_1)$.

To conclude, we must show $\mathcal{E}_0(\rho) \le \liminf_{j\to\infty} \mathcal{E}_0(\pi_{k_j})$. By Proposition 3.4, $c(\gamma)$ is nonnegative and lower semi-continuous. Then it is well known [25, Ch 4] that c can be written as the limit of a nondecreasing, sequence of bounded continuous functions $\{c_n\}_{n\ge 0}$ for $c_n \le c$. The monotone convergence theorem shows

$$\begin{aligned}\mathcal{E}_0(\rho) = \int_\Omega c(\gamma)\ d\rho(\gamma) &= \lim_{n\to\infty} \int_\Omega c_n(\gamma)\ d\rho(\gamma) \\ &= \lim_{n\to\infty} \lim_{j\to\infty} \int_\Omega c_n(\gamma)\ d\pi_{k_j}(\gamma) \\ &\le \liminf_{j\to\infty} \int_\Omega c(\gamma)\ d\pi_{k_j}(\gamma).\end{aligned}$$

We are done since $\mathcal{E}_0(\rho) \le \liminf_{j\to\infty} \int_\Omega c(\gamma)\ d\pi_{k_j}(\gamma) = \inf_{\pi\in\Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{E}_0(\pi)$ but the definition of infimum $\inf_{\pi\in\Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{E}_0(\pi) \le \mathcal{E}_0(\rho)$ yields,

$$\mathcal{E}_0(\rho) = \inf_{\pi\in\Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{E}_0(\pi),$$

and so $\mathcal{E}_0(\pi)$ attains its infimum at ρ.

◇

3.2 The dual problem and the endpoint cost function

In this section we are going to study the *endpoint cost function* introduced in Section 2, equation (2.2). Namely,

$$c_{\mathbf{e}}(x, y) := \inf\{c(\gamma) \mid \gamma(0) = x \quad \text{and} \quad \gamma(1) = y\}.$$

Examples.

1. If $c(\gamma) := \int_0^1 \frac{1}{2}|\dot\gamma(t)|^2\, dt$, then $c_{\mathbf{e}}(x, y) = \frac{1}{2}|x - y|^2$.

2. If $c(\gamma) := \int_0^1 \frac{1}{p}|\dot{\gamma}(t)|^p \, dt, \; p \geq 1$, in $\mathbb{R}^n$, then $c_{\mathbf{e}}(x,y) = \frac{1}{p}|x-y|^p$.

It is not too difficult to show this. In fact we can find in [24], [25]:

Claim. If h is a convex function defined on $\mathbb{R}^n$, then $\inf\{\int_0^1 h(\dot{\gamma}(t)) \, dt \; : \; \gamma(0) = x, \gamma(1) = y\} = h(y-x)$.

We have $\int_0^1 h(\dot{\gamma}(t)) \, dt \geq h\Big(\int_0^1 \dot{\gamma}(t) dt\Big) = h(y-x)$, by Jensen's inequality [24], [25].

A very important toy problem that we will use extensively in the "dual problem" stems from Lagrangian multipliers. *The toy problem is to minimize a function $f(x)$ subject to a linear constraint $Ax = b$, where A is a real $m \times n$ matrix, x and b are column $n \times 1$ vectors, respectively. The formal problem is*

$$\min_x f(x)$$

$$\textit{subject to } Ax = b,$$

where the only variable is $x \in \mathbb{R}^n$, and the rest are constants: $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$.

The solution can be acquired by applying the method of Lagranage multipliers with Lagrangian multiplier λ as follows:

$$\mathfrak{L}(x, \lambda) := f(x) + \lambda \cdot (Ax - b).$$

Then seek "saddle points" of this via solving

$$\min_x \left\{ f(x) + \sup_\lambda \lambda \cdot (Ax - b) \right\}.$$

We will illustrate how to solve an actual *linear programming* problem from optimal transportation. Linear programming can be thought of as the study of the minimization or maximization of linear problems subject to linear inequalities. The problem that we will solve to illustrate this method comes from a version of a problem in Solomon's book [22, Chapter 10.5].

Problem C.

$$\text{minimize}_{x_{ij}\geq 0} \sum_{i,j=1}^{n,m} c_{ij}x_{ij}$$

$$\text{subject to } Ax = b$$

Solution. Let $b = (\mu_1, \dots, \mu_n, \nu_1, \dots, \nu_m)$, let $x = (x_{11}, x_{12}, \dots, x_{1n}; x_{21}, x_{22}, \dots, x_{nm})$, and let $\lambda = (\phi_1, \dots, \phi_n; \psi_1, \dots, \psi_m)$ be the Lagrangian multipliers. The constraints that the matrix A ought to satisfy are $A := (x_{ij})_{i.j}$ and

$$\sum_{i=1}^{n} x_{ij} = \nu_j \ \forall j \quad \text{and} \quad \sum_{j=1}^{m} x_{ij} = \mu_i \ \forall i$$

Then the Lagrange function is given by

$$\begin{aligned}\Lambda(x_{ij}, \phi_i, \psi_j) &:= \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} - \sum_{i=1}^{n} \phi_i \left[\sum_{j=1}^{m} x_{ij} - \mu_i\right] - \sum_{j=1}^{m} \psi_j \left[\sum_{i=1}^{n} x_{ij} - \nu_j\right] \\ &= \sum_{i=1}^{n} \phi_i\mu_i + \sum_{j=1}^{m} \psi_j\nu_j + \sum_{i,j=1}^{n,m} (c_{ij} - \phi_i - \psi_j)x_{ij}\end{aligned}$$

The dual functional can be defined by the minimum of Λ over the $x_{i,j}$

$$\mathfrak{D}(\phi_i, \psi_j) := \min_{x_{ij}} \Lambda(x_{ij}, \phi_i, \psi_j) = \min_{x_{ij}\geq 0} \left\{\sum_{i=1}^{n} \phi_i\mu_i + \sum_{j=1}^{m} \psi_j\nu_j + \sum_{i,j=1}^{n,m} (c_{ij} - \phi_i - \psi_j)x_{ij}\right\}.$$

In particular,

$$\mathfrak{D}(\phi_i, \psi_j) = \begin{cases} \sum_{i=1}^{n} \phi_i\mu_i + \sum_{j=1}^{m} \psi_j\nu_j & \text{if } c_{ij} - \phi_i - \psi_j \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

The Lagrange multiplier function $\Lambda(x;\phi,\psi)$ can be written

$$\Lambda(x;\phi,\psi) = \langle c, x\rangle - \phi^T(x\mathbb{I} - \mu) - \psi^T(x^T\mathbb{I} - \nu),$$

where $\langle\cdot,\cdot\rangle$ indicates the Euclidean inner product on $\mathbb{R}^n$ and $\mathbb{I}$ the vector consisting of all ones. Taking the gradient of the Lagrangian with respect to x, shifting to vector notation, yields

$$0 = \nabla_x\Lambda(x;\phi,\psi) = c - \phi\mathbb{I}^T - \mathbb{I}\psi^T.$$

The Karush-Kuhn-Tucker (KKT) conditions from Solomon's book [22, Chapter 10] infers that x is a critical point for minimizing $\Lambda(x;\phi,\psi)$ whenever there exist multipliers ϕ and ψ such that $0 = \nabla_x\Lambda(x;\phi,\psi)$. To this end, say x^* is a critical point of $\Lambda(x;\phi,\psi)$ for which $c - \phi\mathbb{I}^T - \mathbb{I}^T\psi = 0$. Substituting this into the Lagrange multiplier function Λ, we get

$$\Lambda(x_{ij};\phi_i,\psi_j) = \sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j.$$

The constraint $A^T\lambda \leq c$ turns out to be $\phi_i + \psi_j \leq c_{ij}$ for which we have let

$$b = (\mu_1,\ldots,\mu_n,\nu_1,\ldots,\nu_m) \in \mathbb{R}^n \times \mathbb{R}^m,$$

$$A^T = \sum_{j,i=1}^{m,n} x_{ji},$$

and

$$\lambda := (\phi_1,\ldots,\phi_n;\psi_1,\ldots,\psi_m) \in \mathbb{R}^n \times \mathbb{R}^m,$$

the Lagrange multipliers. □

Remark 3.5 *The solution above shows that one can recover the relation*

$$\max\left\{\sum_{i,j=1}^{n,m} b_{i,j}y_{i,j} : \; Ay \leq c\right\} = \min\left\{\sum_{i,j=1}^{n,m} c_{i,j}x_{i,j} : Ax = b; \; x \geq 0\right\}.$$

The KKT conditions of the minimization problem above were used to solve for the multipliers of the maximization one. Let us now see how we can apply this to the Kantorovich's problem (1.5).

The Kantorovich problem associated to paths is the problem of minimizing the linear functional $\pi \mapsto \int c(\gamma)\ d\pi$ subject to the linear constraints $(e_0)_\sharp \pi = \mu_0$, $(e_1)_\sharp \pi = \mu_1$, and $\pi \geq 0$. Linear minimization problems with convex constraints of this type admit a natural dual problem [1]. We now define this dual problem using the method of Lagrange multipliers that was done in the solution to **Problem C**.

For any pair $(\phi, \psi) \in L^1(d\mu_0) \times L^1(d\mu_1)$ and $\pi \in \mathcal{P}(\Omega)$, we define

$$\begin{aligned} L(\pi; \phi, \psi) := \int_\Omega c(\gamma)\ d\pi(\gamma) + \int_X \phi(x) d\mu_0(x) - \int_\Omega \phi(\gamma(0)) d\pi(\gamma) \\ + \int_X \psi(x) d\mu_1(y) - \int_\Omega \psi(\gamma(1)) d\pi(\gamma) \end{aligned}$$

Then after some rearrangement

$$\begin{aligned} L(\pi; \phi, \psi) = \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) \\ + \int_\Omega (c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1))))\ d\pi(\gamma). \end{aligned}$$

Minimizing over $\pi \in \mathcal{M}(\Omega)$, we get:

Proposition 3.6 *Given any pair of functions $\phi \in L^1(\mu_0)$, $\psi \in L^1(\mu_1)$, we define*

$$D(\phi, \psi) := \inf_{\pi \geq 0} L(\pi; \phi, \psi).$$

Then we have

$$D(\phi, \psi) = \begin{cases} \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) & \text{if } \phi(x) + \psi(y) \leq c_e(x, y) \\ -\infty & \text{otherwise} \end{cases}$$

Proof. Since the first two terms in the latter $L(\pi;\phi,\psi)$ are independent of π,

$$D(\phi,\psi)=\int_X \phi(x)d\mu_0(x)+\int_X \psi(y)d\mu_1(y)+\inf_\pi \int_\Omega (c(\gamma)-(\phi(\gamma(0))+\psi(\gamma(1))))\,d\pi(\gamma).$$

If there exists some γ_0 for which $c(\gamma_0)-\phi(\gamma_0(0))-\psi(\gamma_0(1))<0$, then we can make the minimum $-\infty$. Just take a Dirac mass at γ_0 with very large mass

$$\pi=\lambda\delta_{\gamma_0} \quad \text{for} \quad \lambda>0;$$

letting $\lambda\to+\infty$, $D(\phi,\psi)=-\infty$. Else, for all $\gamma\in\Omega$, $c(\gamma)-\phi(\gamma(0))-\psi(\gamma(1))\geq 0$, therefore the third term is ≥ 0 for all π, and taking $\pi=0$ we get the inf equal to zero, $D(\phi,\psi)=\int_X \phi\,d\mu_1+\int_X \psi\,d\mu_1$. ◇

Thus we arrive at the dual problem:

Definition 3.7 *(Dual problem.) Let $(\phi,\psi)\in L^1(\mu_0)\times L^1(\mu_1)$. Let*

$$\Pi^*:=\{\phi,\psi:\phi(x)+\psi(y)\leq c_{\mathbf{e}}(x,y)\}$$

The dual problem consists of finding the following supremum

$$\sup\left\{\int_X \phi(x)d\mu_0(x)+\int_X \psi(y)d\mu_1(y):(\phi,\psi)\in\Pi^*\right\}.$$

The latter is reminiscent of a dual problem coming from the standard theory of optimal transportation [1, Section 1.3]. The following is the main fact about the dual problem, we recall X is the closure of bounded, open, and connected set and we recall the function $c_{\mathbf{e}}(x,y)$ defined by (2.2).

Theorem 3.8 *Let $c:\Omega\to\mathbb{R}$ be lower semi-continuous. Let $\mu_0\in\mathcal{P}(X)$ and $\mu_1\in\mathcal{P}(X)$. Assume that $c_{\mathbf{e}}(x,y)\leq f(x)+g(y)$ for some $f\in L^1(d\mu_0), g\in L^1(d\mu_1)$, then*

$$\inf_{\pi\in\Pi_{path}}\left\{\int_\Omega c(\gamma)\,d\pi(\gamma)\right\}=\sup_{(\phi,\psi)\in\Pi^*}\left\{\int_X \phi(x)\,d\mu_0(x)+\int_X \psi(y)d\mu_1(y)\right\} \qquad (3.2)$$

The proof of this theorem will be given in the end of Section 3.3. For now we go over some important consequences of Theorem 3.8.

Corollary 3.9 *Under the assumptions of Theorem 3.8, if* $\pi_0 \in \Pi_{path}(\mu_0, \mu_1)$ *achieves the infimum, then for* π_0*-a.e.* $\gamma \in \Omega$ *we have* $c(\gamma) = c_{\mathbf{e}}(\gamma(0), \gamma(1))$*.*

Proof. Let $\pi_0 \in \Pi_{path}$ and $(\phi_0, \psi_0) \in \Pi^*$ be optimal. Then by Theorem 3.8 we have

$$\begin{aligned}\int_\Omega c(\gamma)\, d\pi_0(\gamma) &= \int_X \phi_0(x)\, d\mu_0(x) + \int_X \psi_0(y)\, d\mu_1(y)\\ &= \int_\Omega (\phi_0(\gamma(0)) + \psi_0(\gamma(1)))\, d\pi_0(\gamma).\end{aligned}$$

Therefore, $\int_\Omega (c(\gamma) - (\phi_0(\gamma(0)) - \psi_0(\gamma(1))))\ d\pi_0(\gamma) = 0$. Since the integrand is ≥ 0, not just π_0-a.e. it follows that $c(\gamma) = \phi_0(\gamma(0)) + \psi_0(\gamma(1))$ π_0-a.e. γ. Since $\phi_0(x) + \psi_0(y) \leq c_{\mathbf{e}}(x, y)$, $c(\gamma) = \inf\{c(\gamma) : \gamma(0) = x, \gamma(1) = y\}$ for π_0-a.e. γ. ⋄

Corollary 3.9 indicates that whenever π is optimal, the support of π lies in the set of all minimal paths, $\Omega_{\min}$. An explicit proof of this statement, that does not rely on Theorem 3.8, is given in Lemma 3.13.

3.3 Potentials and cyclically monotone sets

In this section we further study the properties from the inequality of Definition 3.7. In some sense, we follow the standard theory of using concave potentials ϕ, ψ [1] associated to Theorem 3.8 coming from the theory of superdifferentiability. Namely, a function ϕ is called *c-concave* if $\phi = \phi^*$ as given in (3.3).

Let us write and study such properties. For $\phi \in L^1(d\mu_0)$ and $\psi \in L^1(d\mu_1)$, and for μ_0-a.e. x and μ_1-a.e y for which $\gamma \in \Omega$ is a path from $\gamma(0) = x$ to $\gamma(1) = y$, we have the inequality from Definition 3.7. Then the "concavity" *transforms* stemming from [24] are the following. Given $(\phi, \psi) \in \Pi^*$, for μ_1-a.e. y

$$\psi(y) \leq c_{\mathbf{e}}(x,y) - \phi(x).$$

Taking the infimum with respect x,

$$\psi(y) \leq \inf_{x \in X} [c_{\mathbf{e}}(x,y) - \phi(x)] := \phi^*(y) \tag{3.3}$$

Similarly, for μ_0-a.e. x,

$$\phi(x) \leq c_{\mathbf{e}}(x,y) - \psi(y),$$

implementing the infimum with respect to y,

$$\phi(x) \leq \inf_{y \in X} [c_{\mathbf{e}}(x,y) - \psi(y)] := \psi^*(x). \tag{3.4}$$

Calling $\mathcal{J}(\phi, \psi) := \int_X \phi(x)\, d\mu_0(x) + \int_X \psi(y)\, d\mu_1(y)$, the linear functional from the supremum in Theorem 3.8, we can witness from (3.3) that

$$\mathcal{J}(\phi, \phi^*) \geq \mathcal{J}(\phi, \psi).$$

And from (3.3) and (3.4) for μ_0-a.e. x

$$\phi^{**}(x) =: \inf_{y \in X} [c_{\mathbf{e}}(x,y) - \phi^*(y)] \geq \phi(x),$$

and

$$\mathcal{J}(\phi^{**}, \phi^*) \geq \mathcal{J}(\phi, \phi^*) \geq \mathcal{J}(\phi, \psi).$$

This heuristic shows that the pair (ϕ, ϕ^*) maximizes the dual problem of Definition 3.7.

The *superdifferential* set defined for a c-concave function ϕ is:

$$\partial\phi := \Big\{\gamma \in \Omega : \ \phi(x) + \phi^*(y) = c_{\mathbf{e}}(x,y)\Big\}.$$

We now give an account on the theory of *cyclical monotone* sets that includes paths. For the classical definition please see Villani's, Santambrogio's, or Ambrosio's and Gigli's account on the theory in [24], [19], [1].

Let $\{(x_i, y_i)\}_{i=1}^N$ be a set of pair of points in X such that each x_i is contained in the support of μ_0 and each y_i is contained in the support of μ_1. Then each path $\gamma_i \in \Omega$ in the support of π is such that

$$\gamma_i(0) = x_i \quad \forall i \quad \text{and} \quad \gamma_i(1) = y_i \quad \forall i.$$

The *shift*, as we will call it for the moment, of the final points of paths give rise to new shifted paths $\widetilde{\gamma}$ defined by

$$\widetilde{\gamma}_i(t) = \gamma_i(t) + th_i(t), \quad 0 \le t \le 1, \ h_i \ll 1 \ \forall i, \quad \text{and} \quad \gamma_i(1) + h_i(1) := \gamma_{i+1}(1), \tag{3.5}$$

with the convention $\gamma_{N+1}(1) = \gamma_1(1)$. The shifted paths $\{\gamma_i\}_{i=1}^N$ with endpoints set $\{(x_i, y_i)\}_{i=1}^N$ are such that (3.5) holds and

$$\widetilde{\gamma}_i(0) = x_i \quad \forall i \quad \text{and} \quad \widetilde{\gamma}_i(1) = y_{i+1 \bmod N} \quad \forall i$$

with the convention $\gamma_{N+1}(1) = \gamma_1(1)$.

On the endpoints of γ_i, $\{(x_i, y_i)\}$ is $c_{\mathbf{e}}$-cyclically monotone if it satisfies the following definition.

Definition 3.10 *We say that the set* $\{(x_i, y_i)\}$ *is* $c_{\mathbf{e}}$-cyclically monotone *if for all* $i = 1, \ldots, n$,

$$\sum_{i=1}^{n} c_{\boldsymbol{e}}(x_i, y_i) \le \sum_{i=1}^{n} c_{\boldsymbol{e}}(x_i, y_{\tau(i)}).$$

for any permutations τ *on* n *letters.*

Remark 3.11 *We can, for simplicity, take the permutation* $\tau(i)$ *in Definitions 5.1 and 5.2 to mean the following:* $y_{\tau(i)} = y_{i+1}$ *with the convention* $y_{n+1} = y_1$.

Definition 3.12 *A continuous path* $\gamma : [0,1] \to X$ *from its initial point* x *to its final point* y *is* minimal *with respect to the cost function* c *if for all other paths* $\gamma' : [0,1] \to X$ *having the same initial and final points of* γ *satisfy*

$$c(\gamma) \leq c(\gamma').$$

A vital result is to show that if π is optimal in $\Pi_{\text{path}}(\mu_0, \mu_1)$ then the support of π will be contained in the set of all minimal paths, Ω_{min}. The proof of the discrete case of Lemma 3.13 appears to be more descriptive. To that end, we first show the discrete case, then the continuous one.

For a Borel set B, $\pi \llcorner B$ is the restriction of π to B, namely the measure defined by

$$[\pi \llcorner B](A) = \pi(B \cap A), \quad \text{for every Borel set } A.$$

Lemma 3.13 *Let* $c : \Omega \to \mathbb{R}$ *be a lower semicontinuous cost function. Suppose* π *is optimal. Then if* $\gamma_0 \in spt(\pi)$, γ_0 *is minimal.*

Proof. [Proof Discrete case.] We merely show this in the discrete case. Let $\gamma \in \Omega$ be a path with initial and final endpoints fixed. Consider a competitor path $\gamma' := \gamma_c$ having the same endpoints as γ associated to π_c in the following way. For $0 \leq \lambda_i \leq 1$, $i \in \{1, 2, \dots, k\}$ and Dirac delta δ_γ,

$$\pi_c = \lambda_1 \delta_{\gamma_c} + \sum_{k=2}^{m} \lambda_k \delta_{\gamma_k}.$$

Then define an optimal π by

$$\pi = \sum_{k=1}^{m} \lambda_k \delta_{\gamma_k}.$$

By hypothesis $\mathcal{E}_0(\pi) \leq \mathcal{E}_0(\pi_c)$ so that

$$\sum_{k=1}^{m} \lambda_k c(\gamma_k) \leq \lambda_1 c(\gamma_c) + \sum_{k=2}^{m} \lambda_k c(\gamma_k),$$

and in particular for all $k \geq 1$, cancelling out common terms yields

$$c(\gamma_1) \leq c(\gamma_c).$$

Then in this case γ_1 is minimal

By induction on k, keeping m fixed, on all paths γ_k applying the same argument on γ_k we get $c(\gamma_k) \leq c(\gamma_c)$ for all $k \geq 1$. Therefore, $c(\gamma) \leq c(\gamma_c)$. $\diamond$

Proof. [Proof Continuous case.] Suppose $\gamma_0 \in$ spt (π) and $\pi \in \Pi_{path}(\mu_0, \mu_1)$ is optimal. Let γ' be the minimal path from $x_0 := \gamma'(0)$ to $y_0 := \gamma'(1)$, arguing by contradiction, if γ_0 is not minimal, there is $\varepsilon > 0$ such that

$$c(\gamma') < c(\gamma_0) - \varepsilon.$$

Let $W := B_r(x_0) \times B_r(y_0)$ for $r > 0$ (to be determined later) and define what we will call a *tubo*

$$T := (e_0, e_1)^{-1} (W) \, .$$

This is an open set in Ω containing the path γ_0. Consider the measure, $\bar{\pi} := \pi \llcorner T / \pi [T]$; $\pi [T]$ will be positive as $\gamma_0 \in$ spt (π). Take $0 < \epsilon_0 < \pi [T]$. Define the measures

$$\nu_x := (e_0)_\sharp \bar{\pi} \quad \text{and} \quad \nu_y := (e_1)_\sharp \bar{\pi}.$$

Now build a measure $\widetilde{\pi} \in \Pi(\nu_x, \nu_y)$ as follows. Let $g : X \times X \to \Omega$ be defined by

$$g_t(x, y) = \gamma_{x,y}(t)$$

such that $\gamma_{x,y}$ is the minimal path from x to y. Then set $\widetilde{\pi} := (g)_\sharp (\nu_x \otimes \nu_y)$ and define

$$\pi' := \pi - \epsilon_0 \bar{\pi} + \epsilon_0 \widetilde{\pi}.$$

That π' is positive follows from $\pi - \epsilon_0 \bar{\pi}$ being positive, which is thanks to $\epsilon_0 < \pi[T]$. The marginals of π' share the same marginals of π. Concretely,

$$(e_0)_\sharp \pi' = \mu_0 - \epsilon_0 \nu_x + \epsilon_0 (e_0)_\sharp \widetilde{\pi},$$

and for every Borel subset $B \subset X$, $(\nu_x \otimes \nu_y)\left[g^{-1} \circ e_0^{-1}(B)\right] = \nu_x[B]$. Apply the same argument to get second marginal.

Finally, we will show $\int c(\gamma) d\pi(\gamma) - \int c(\gamma) d\pi'(\gamma) > 0$, thereby contradicting the optimality of π. Let $m : \Omega \to \Omega_{\min},\ \gamma \mapsto \gamma_{\min}$ be such that $\gamma_{\min}(0) = \gamma(0)$ and $\gamma_{\min}(1) = \gamma(1)$. So

$$c \circ m(\gamma) = c(\gamma_{\min}) \leq c(\gamma).$$

Fix $\delta > 0$ such that if $\gamma \in T$ ($r > 0$ small), then

$$\|\gamma_{\min} - \gamma'\|_\infty < \delta \ \text{ (if } r < r(\delta) := \delta/2).$$

The lower semi-continuity of c says $c(\gamma_{\min}) \geq c(\gamma') + \varepsilon/2$. On the other hand, for all $\gamma \in \operatorname{spt}(\widetilde{\pi})$, the continuity of g and lower semi continuity of c provides,

$$c(\gamma) \leq c(\gamma') + \varepsilon/4.$$

Putting this all together yields,

$$\begin{aligned}
\int c(\gamma) d\pi(\gamma) - \int c(\gamma) d\pi'(\gamma) &= \epsilon_0 \int c(\gamma)\, d\bar{\pi}(\gamma) - \epsilon_0 \int c(\gamma)\, d\widetilde{\pi}(\gamma) \\
&\geq \epsilon_0 \int c(\gamma_{\min})\, d\bar{\pi}(\gamma) - \epsilon_0 \int c(\gamma)\, d\widetilde{\pi}(\gamma) \\
&\geq \epsilon_0 \int \left(c(\gamma') + \frac{\varepsilon}{2}\right)\, d\bar{\pi}(\gamma) - \epsilon_0 \int \left(c(\gamma') + \frac{\varepsilon}{4}\right)\, d\widetilde{\pi}(\gamma) \\
&= \epsilon_0 \, \frac{\varepsilon}{2} > 0.
\end{aligned}$$

By definition $\overline{\pi}[\Omega] = (\pi \llcorner T)[\Omega]/\pi[T] = 1$ and when we used the fact that both $\widetilde{\pi}$ and $\bar{\pi}$ are supported on T and have unit mass due to the rescaling. $\diamond$

Next we show that optimal plans in Ω have cyclical monotone support. The picture complementing the proof of Lemma 3.14 is given below:

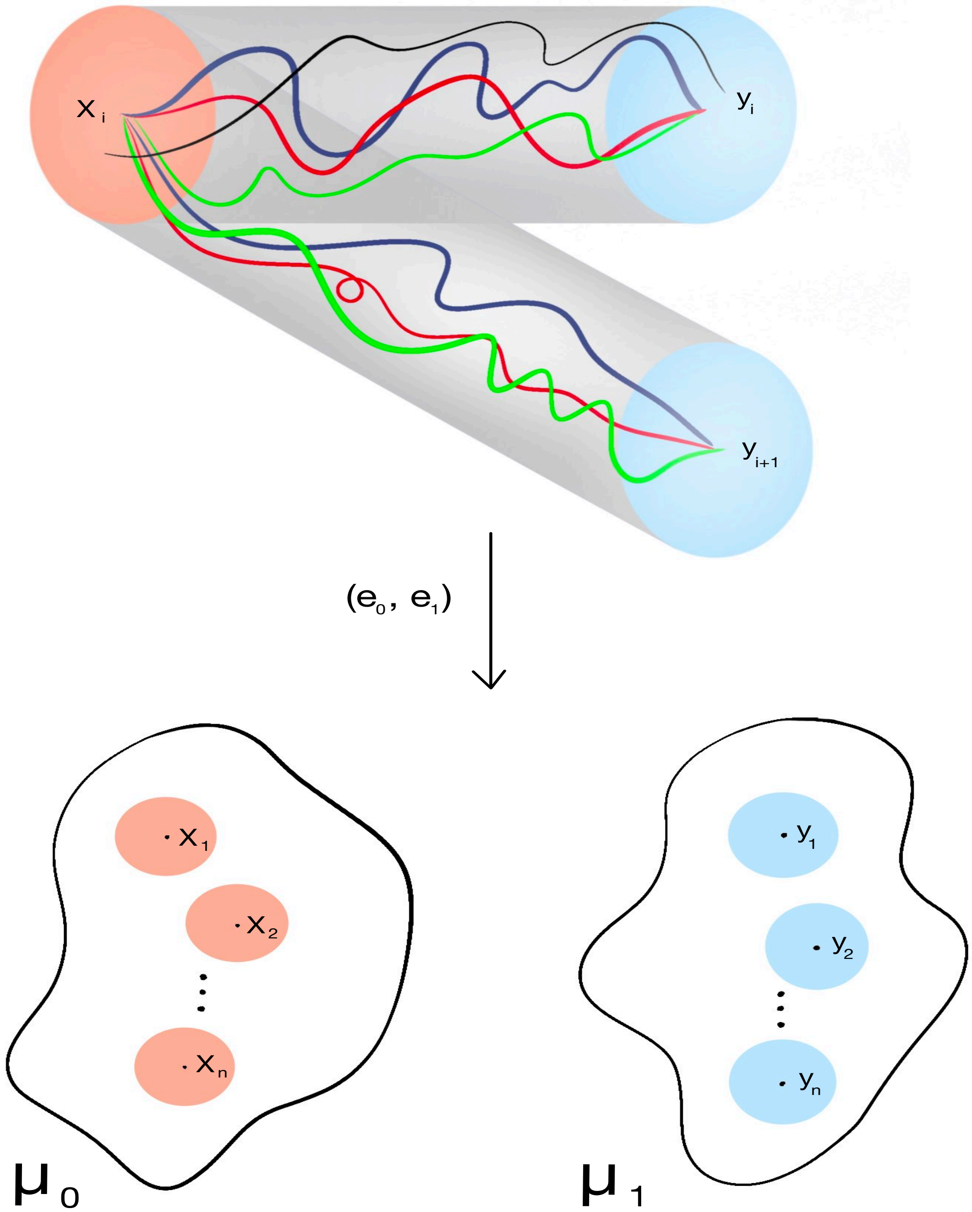

Figure 1: This picture indicates how the pull-back mapping $T := (e_0, e_1)^{-1}(W) \cap \Omega_{\min}$ (the above *tubo*) is obtained from $W := B_r(x_i) \times B_r(y_i)$ (the product of balls centered about x_i and y_i).

Lemma 3.14 *Suppose* $c : \Omega \to \mathbb{R}$ *is a continuous cost function and* $\pi \in \Pi_{path}(\mu_0, \mu_1)$ *optimal with respect to* c*. Let* $\breve{\pi} := (e_0, e_1)_\sharp \pi \in \Pi(\mu_0, \mu_1)$*. Then the support of* $\breve{\pi}$*, spt* $(\breve{\pi})$*, is* c_e*-cyclically monotone. Moreover,* $\breve{\pi}$ *is optimal with respect to* $c_e(x, y)$ *defined by (2.2).*

Proof. We follow a classical proof from standard optimal transportation in Santambrogio's book [19]; the new concept is to incorporate paths.

Let $\breve{\pi} := (e_0, e_1)_\sharp \pi$ be a transport plan from μ_0 onto μ_1 given by pushing forward π through the coupled evaluation map $(e_0, e_1) : \Omega \to X \times X$. Suppose $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$ is optimal such that $\text{spt}\,(\pi) \subset \Omega_{\min}$, the set of all minimal paths by Lemma 3.13. Suppose by way of contradiction that spt $(\breve{\pi})$ is not c_e-cyclically monotone. Then there exist n, τ, and minimal paths γ_i from x_i to y_i, and $\widetilde{\gamma}_i$ from x_i to $y_{\tau(i)}$ in spt π, respectively $\{(x_i, y_i)\} \subset$ spt $\breve{\pi}$, such that

$$\sum_{i=1}^{n} c(\gamma_i) > \sum_{i=1}^{n} c(\widetilde{\gamma}_i),$$

where the path $\widetilde{\gamma}_i$ satisfies $\widetilde{\gamma}_i(0) = \gamma_i(0)$ and $\widetilde{\gamma}_i(1) = \gamma_{\tau(i)}(1)$ for all $i = 1, \ldots, n$. Then since γ_i and $\widetilde{\gamma}_i$ are minimal, $c(\gamma_i) = c_e(x_i, y_i)$ and $c(\widetilde{\gamma}_i) = c_e(x_i, y_{\tau(i)})$, and thus the above inequality equals

$$\sum_{i=1}^{n} c_e(x_i, y_i) > \sum_{i=1}^{n} c_e(x_i, y_{\tau(i)}).$$

Given $\varepsilon > 0$, take

$$\varepsilon < \frac{1}{2n} \left(\sum_{i=1}^{n} c_e(x_i, y_i) - c_e(x_i, y_{\tau(i)}) \right). \tag{3.6}$$

By continuity of c, there exists an open neighborhood called a *tubo*

$$T_i := (e_0, e_1)^{-1} \left(B_r(x_i) \times B_r(y_i) \right) \cap \Omega_{\min}$$

such that for all $i = 1, \dots, n$ and all $\gamma \in T_i$, $c(\gamma) > c(\gamma_i) - \varepsilon$ and for all γ in

$$\widetilde{T}_i := (e_0, e_1)^{-1} \left(B_r(x_i) \times B_r(y_{\tau(i)}) \right) \cap \Omega_{\min},$$

we have $c(\gamma) < c(\widetilde{\gamma}_i) + \varepsilon$.

Now define the measures

$$\pi_i := \pi \llcorner T_i / \pi \left[T_i \right], \quad \nu_{x,i} := (e_0)_\sharp \pi_i, \quad \text{and} \quad \nu_{y,i} := (e_1)_\sharp \pi_i,$$

and note that $\pi \left[T_i \right]$ will be positive for each i since γ_i is contained in the support of π. Equivalently $(x_i, y_i) \in \operatorname{spt}(\check{\pi})$. Take $0 < \epsilon_0 < \min_i \pi \left[T_i \right]$.

Construct a measure $\widetilde{\pi}_i \in \Pi \left(\nu_{x,i}, \nu_{y,\tau(i)} \right)$, for every i, in the following way. Let g be a map

$$g : X \times X \to \Omega \quad \text{defined by} \quad (x, y) \mapsto \gamma_{x,y},$$

that is, $g(x, y) = \gamma_{x,y}(t)$, the minimal path between x and y such that $\gamma_{x,y}(0) = x$ and $\gamma_{x,y}(1) = y$. Then the estimates $c(\gamma) > c(\gamma_i) - \varepsilon$ for all $\gamma \in T_i$ and $c(\gamma) < c(\tilde{\gamma}_i) + \varepsilon$ for all $\gamma \in \widetilde{T}_i$ coincide with the estimates $c_e(x, y) > c_e(x_i, y_i) - \varepsilon$ and $c_e(x, y) < c_e(x_i, y_{\tau(i)}) + \varepsilon$ for all $(x, y) \in B_r(x_i) \times B_r(y_i)$ and all $(x, y) \in B_r(x_i) \times B_r(y_{\tau(i)})$, respectively. Take $\widetilde{\pi}_i := (g)_\sharp \left(\nu_{x,i} \otimes \nu_{y,\tau(i)} \right)$.

Now define

$$\widetilde{\pi} := \pi - \epsilon_0 \sum_{i=1}^{n} \pi_i + \epsilon_0 \sum_{i=1}^{n} \widetilde{\pi}_i.$$

That $\widetilde{\pi}$ is positive follows from $\pi - \epsilon_0 \sum_{i=1}^n \pi_i$ being positive as $\epsilon_0 < \min_i \pi \left[T_i \right]$. More concretely, since it suffices to check $\pi - \epsilon_0 \sum_{i=1}^n \pi_i > 0$, the condition $\epsilon_0 \pi_i < \pi / n$ is enough. Indeed, as $\epsilon_0 \pi_i = \frac{\epsilon_0}{\pi[T_i]} \pi \llcorner T_i$ and $\epsilon_0 / \pi \left[T_i \right] \leq 1/n$.

The marginals of $\widetilde{\pi}$ share the marginals of π:

$$(e_0)_\sharp \widetilde{\pi} = \mu_0 - \epsilon_0 \sum_{i=1}^{n} \nu_{x,i} + \epsilon_0 \sum_{i=1}^{n} (e_0)_\sharp (g)_\sharp \left(\nu_{x,i} \otimes \nu_{y,\tau(i)} \right),$$

and for all Borel subsets $B \subset X$, $\left(\nu_{x,i} \otimes \nu_{y,\tau(i)}\right)\left[g^{-1}\left(e_0^{-1}(B)\right)\right] = \nu_{x,i}[B]$ is the measure $\nu_{x,i}$ containing the points over the entire first copy of X. The second marginal follows the same story:

$$(e_1)_\sharp \widetilde{\pi} = \mu_1 - \epsilon_0 \sum_{i=1}^{n} \nu_{y,i} + \epsilon_0 \sum_{i=1}^{n} (e_1)_\sharp (g)_\sharp \left(\nu_{x,i} \otimes \nu_{y,\tau(i)}\right);$$

for all Borel $A \subset X$, $\left(\nu_{x,i} \otimes \nu_{y,\tau(i)}\right)\left[g^{-1}\left(e_1^{-1}(A)\right)\right] = \nu_{y,\tau(i)}[A]$ is the measure $\nu_{y,\tau(i)}$ containing the points over the entire second copy of X.

Finally, the estimate $\int c \; d\pi - \int c \; d\widetilde{\pi}$ is positive, thereby contradicting the optimality of π:

$$\begin{aligned} \int c(\gamma)\; d\pi(\gamma) - \int c(\gamma)\; d\widetilde{\pi}(\gamma) &= \epsilon_0 \sum_{i=1}^{n} \int c(\gamma)\; d\pi_i(\gamma) - \epsilon_0 \sum_{i=1}^{n} \int c(\gamma)\; d\widetilde{\pi}_i(\gamma) \\ &\geq \epsilon_0 \sum_{i=1}^{n} \left(c(\gamma_i) - \varepsilon\right) - \epsilon_0 \sum_{i=1}^{n} \left(c(\widetilde{\gamma}_i) + \varepsilon\right) \\ &= \epsilon_0 \left(\sum_{i=1}^{n} c_e(x_i, y_i) - c_e(x_i, y_{\tau(i)}) - 2n\varepsilon \right) > 0, \end{aligned}$$

where we used that π_i is supported on T_i, $\widetilde{\pi}_i$ supported on $\widetilde{T}_i$ and have unit mass by rescaling the measures by $\pi\left[T_i\right]$.

To end the proof, we apply the standard theory of optimal transportation, from [1, Theorem 1.13], to the endpoints (x_i, y_i) contained in the support of $\breve{\pi}$ to find that $\breve{\pi}$ is optimal with respect to $c_e(x, y)$, since spt $\breve{\pi}$ is c_e-cyclically monotone. ⋄

Remark 3.15 *The previous Lemma 3.14 says that any optimal plan* π *in* $\Pi_{path}(\mu_0, \mu_1)$ *"projects" to a solution of the Monge-Kantorovich problem with cost* $c_e(x, y)$ *and coupling* $\breve{\pi} := (e_0, e_1)_\sharp \pi$. *This is the same construction as the standard theory in optimal transportation that one may find in Villani's book [25, Chapter 7]. There it is called the* dynamical optimal coupling*, first mentioned in the Introduction Section 1. In Villani's book [25, Ch 7] he shows the standard demonstration in the* *classical optimal transport the existence of a unique transport plan $\pi_{0,1}$ between μ_0 and μ_1. In both cases we get the same dynamical optimal coupling, but our proofs are different.*

Armed with these results we may, and we actually do, prove (3.8).

Proof. [**Proof of Theorem 3.8**] With all the assumptions of Theorems 3.8 and Lemma 3.14, let $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$; notice that for any pair $(\phi, \psi) \in L^1(d\mu_0) \times L^1(d\mu_1)$ satisfying inequality of Definition 3.7,

$$\int_\Omega c_{\mathbf{e}}(x,y)d\pi(\gamma) \geq \int_\Omega (\phi(x) + \psi(y))d\pi(\gamma) = \int_X \phi(x)d\mu_0(x) + \int_X \psi(y)d\mu_1(y).$$

Next, take infimum over admissible π on the left-hand side and take the supremum over $(\phi, \psi) \in \Pi^*$ on the right-hand side to get the "$\geq$" part.

To prove the reverse inequality, choose an optimal $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$. Since spt $\check{\pi}$ is c_e-cyclically monotone by Lemma 3.14, the classical theory of optimal transport applies to show there is a lower semi-continuous concave function ϕ such that spt $\check{\pi} \subset \partial\phi$ for which $\phi \in L^1(d\mu_0)$ and $\phi^* \in L^1(d\mu_1)$. Then

$$\int_\Omega c_{\mathbf{e}}(x,y)d\pi(\gamma) = \int_\Omega (\phi(x) + \phi^*(y))d\pi(\gamma) = \int_X \phi(x)d\mu_0(x) + \int_X \phi^*(y)d\mu_1(y) \tag{3.7}$$

Now, the claim is then that (ϕ, ϕ^*) solves the maximization problem of Definition 3.7. More concretely, to prove that (ϕ, ϕ^*) solves the maximization problem, we note that

1. $\tilde{\phi}(x) + \tilde{\phi}^*(y) = c_{\mathbf{e}}(x,y)$ on the support of the optimal π, sptπ.
2. $\phi(x) + \phi^*(y) \leq c_{\mathbf{e}}(x,y)$ on $X \times X$.

Then

$$\begin{aligned}\int_X \tilde{\phi}(x)d\mu_0(x) + \int_X \tilde{\phi}^*(y)d\mu_1(y) &= \int_\Omega (\tilde{\phi}(x) + \tilde{\phi}^*(y))d\pi(\gamma) = \int_\Omega c_{\mathbf{e}}(x,y)\, d\pi(\gamma) \\ &\geq \int_\Omega (\phi(x) + \phi^*(y))d\pi(\gamma) = \int_X \phi(x)d\mu_0(x) + \int_X \phi^*(y)d\mu_1(y),\end{aligned}$$

and so $\tilde{\phi}$ solves the maximization problem. This with (3.7) establishes the proof. $\diamond$

3.4 Optimal plans given by maps

The goal of this section is to prove uniqueness of the minimizer of the optimal path Kantorovich problem (1.5) and also show it is given by a map $\Gamma(x)$, provided μ_0 is absolutely continuous with respect to the Lebesgue measure: $\mu_0(x) << dx$. This uses, and extends, results of Brenier [5] and Gangbo-McCann [10] in the classical optimal transport theory.

Let Ω_{min} denote the set of all minimal paths. If $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$ is optimal Corollary 3.9 indicated spt π is contained in Ω_{min}. In particular, the calculations and discussions of Section 2 and Section 3.2 amount to the existence of a minimal path $\gamma_* \in \Omega$, minimizing for $c(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 - V(\gamma(t), t)\, dt$ in Ω, in the support of π contained in Ω_{min}.

The uniqueness will follow at once. And what's more, the mapping Γ will be given by (3.9) below and such that the optimal transport plan $\pi_\Gamma := (\Gamma)_\sharp \mu_0$ will lie in $\Pi_{\text{path}}(\mu_0, \mu_1)$.

We are going to look for mappings

$$\Gamma : X \times [0,1] \to X \tag{3.8}$$

of the form $\Gamma(x,t)$ for every x with the following properties $\Gamma(x,t)$ for every x such that

$$\begin{cases} \Gamma(x,0) & = x \text{ for all } \quad x \in X \\ \Gamma(x,1) & = T(x) \text{ for } \quad T_\sharp \mu_0 = \mu_1 \end{cases} \tag{3.9}$$

where $T : X \to X$ is a measurable map pushing $\mu_0 \mapsto \mu_1$.

Once and for all we shall consider functions from $X \times [0,1]$ to X to be in one-to-one correspondence with continuous functions from X to Ω; we shall denote them with the same letter Γ. Let Γ be such that $\Gamma(x)(0) = x$ and $\Gamma(x)$ be a minimal path for every x, and let $T(x) = \Gamma(x)(1)$. To this end, since $\Gamma(x,t)$ defines a mapping from $X \times [0,1]$ to X and $\Gamma(x)(t)$ a mapping from X to Ω to which there is a one-to-one correspondence between them, we shall think of the equivalence $\Gamma(x,t) \Leftrightarrow \Gamma(x)(t)$. The maps Γ we will work with have the property: $\Gamma(x,0) = x$ for all $x \in X$.

The next result can be intuitively thought of as saying if T maps μ_0 to μ_1, then Γ corresponds to an admissible measure in the path space.

Lemma 3.16 *Define* $\pi_\Gamma := (\Gamma)_\sharp \mu_0$. *Then* $\pi_\Gamma \in \Pi_{path}(\mu_0, \mu_1)$.

Proof. In a moment we will unpack some definitions regarding push-forwards thru evaluation maps taken from Bernot et al in [4]. First recall the evaluation maps $e_0, e_1 : \Omega \to X$ are such that $e_0(\gamma) = \gamma(0)$ and $e_1(\gamma) = \gamma(1)$. For measurable subsets $A, B \subset X$ and $\pi \in \mathcal{P}(\Omega)$ we have

$$(e_0)_\sharp \pi_\Gamma[A] := \pi_\Gamma[e_0^{-1}(A)] = \pi_\Gamma[\{\gamma \in \Omega : \gamma(0) \in A\}]$$

$$(e_1)_\sharp \pi_\Gamma[B] := \pi_\Gamma[e_1^{-1}(B)] = \pi_\Gamma[\{\gamma \in \Omega : \gamma(1) \in B\}]$$

To show $(\Gamma)_\sharp(\mu_0)$ lies in $\Pi_\gamma(\mu_0, \mu_1)$, we must show the following two things. The first $(e_0)_\sharp(\Gamma)_\sharp(\mu_0) = \mu_0$ and the second $(e_1)_\sharp(\Gamma)_\sharp(\mu_0) = \mu_1$. Notice that $\Gamma : X \to \Omega$ pushes forward μ_0 to μ_1, by construction. Indeed, observe that the composition $e_1 \circ \Gamma$ coincides with T, and since $T_\sharp \mu_0 = \mu_1$, $(e_1)_\sharp(\Gamma_\sharp)\mu_0 = \mu_1$.

Let $A, B \subset X$ be measurable subsets. Then $\pi_\Gamma[A] := (\Gamma)_\sharp(\mu_0)[A] = \mu_0[\Gamma^{-1}(A)]$

so that

$$\begin{aligned}(e_0)_\sharp \pi_\Gamma[A] &= \pi_\Gamma[e_0^{-1}(A)] \\ &= \mu_0[\Gamma^{-1}(e_0^{-1}(A))] \\ &= \mu_0[A].\end{aligned}$$

Similarly, we have $\pi_\Gamma[B] := (\Gamma)_\sharp(\mu_0)[B] = \mu_0[\Gamma^{-1}(B)]$. Then

$$\begin{aligned}(e_1)_\sharp \pi_\Gamma[B] &= \pi_\Gamma[e_1^{-1}(B)] \\ &= \mu_0[\Gamma^{-1}(e_1^{-1}(B))] \\ &= \mu_1[B].\end{aligned}$$

This completes the proof. ◇

The important result we use from the standard optimal theory is that whenever μ_0 is absolutely continuous with respect to Lebesgue [1], [24], the optimal plan will be concentrated on the graph of T. Furthermore, the classical theory of optimal transportation already contains a theorem for a Lagrangian cost function, see Villani's book [24, Chapter 5]: the following has a unique solution,

$$\inf\left\{\int_{\mathbb{R}^n} c_L(x, T(x))\, d\mu_0(x) : \gamma_0 = \text{id},\ \gamma_1 = T \text{ such that } T_\sharp\mu_0 = \mu_1\right\}, \tag{3.10}$$

where $c_L(x,y) := \min\left\{\int_0^1 L(\dot\gamma(t))\, dt : \gamma(0) = x,\ \gamma(1) = y\right\}$ and L is a strictly convex Lagrangian cost function satisfying $L(0) = 0$. The solution enjoys Brenier's characterization [5] and it is a consequence of Theorem 10.28 in Villani's book [25]; so it is given by

$$\gamma_t(x) = x - t\nabla^* L(\nabla\psi(x)), \quad 0 \le t \le 1,$$

where ψ is a c_L-concave function (see Section 3.3) for which $[\text{id} - \nabla^* L(\nabla\psi)]_\sharp\mu_0 = \mu_1$, and ∇^* denotes the Legendre transform or in this case can be thought of as the inverse, $\nabla^{-1}L = \nabla^* L$. If such solution exists, then the optimizer should interpolate between $\gamma_0(x) = x$ and $\gamma_1(x) = x - \nabla^* L(\nabla\psi(x))$, according to Villani [24, Chapter 5]. Monge's classical minimization Problem (1.3) and Problem (3.10) are compatible provided $c_L(x, y) = \min\left\{\int_0^1 L(\dot\gamma_t)dt : \gamma_0 = x,\ \gamma_1 = y\right\}$. In this case solutions of the time-dependent minimization optimal transport problem have to satisfy for μ_0-a.e. x, $c(x, T(x)) = \int_0^1 L(\dot\gamma_t)dt$. To solve this problem, we took a different approach.

Returning to our main point of view of optimal path optimal transport theory and possessing this knowledge, we are ready to derive an optimal path map stemming from the set of minimal paths which will include the mapping (3.9): $\Gamma : X \to \Omega$. Let $e_0 : \Omega \to X$ be the evaluation map and recall $X := \overline{B_R(0)} \subset \mathbb{R}^n$.

Proof. [**Proof of Theorem 1.7**] Theorem 1.6 says there is an optimal plan π_* in the space of paths solving the optimal path problem (1.5), **Problem A**; while Lemma 3.14 tells us that π_* projects to the classical optimal transport solution with respect to $c_{\mathbf{e}}(x, y)$, namely $\check\pi_* := (e_0, e_1)_\sharp \pi_*$.

Lemma 2.9 says $c_{\mathbf{e}}(x, y)$ is differentiable and Lemma 2.10 says it is injective in its domain. Since π_* is optimal, Lemma 3.14 implies the support of $\check\pi_*$ is c_e-cyclically monotone. The classical theory of optimal transportation [1, Theorem 1.13] says that spt $\check\pi_*$ is contained in $\partial\phi$, and as ϕ is locally Lipschitz apply Rademacher's theorem and that $\mu_0 \ll dx$, ϕ is differentiable μ_0-a.e. All of the above show $c_{\mathbf{e}}(x, y)$ satisfies the assumptions of Theorem 10.28 in Villani's book [25] and so it applies to give a unique transport map T pushing μ_0 forward to μ_1, solving the optimal transport problem with respect to $c_{\mathbf{e}}(x, y)$.

Define a map Γ as in (3.9) containing the above data on T. Let $\gamma_{x,y}(t)$ be the minimal path between x and y. Then $c(\gamma_{x,y}(t)) = c_{\mathbf{e}}(x, y)$, and using this information, let $\Gamma(x,t) := \gamma_{x,T(x)}(t)$. More precisely, Γ is given as the composition,

$$X \xrightarrow{(Id\times T)} X \times X \xrightarrow{\gamma_{x,y}} \Omega_{\min} \quad .$$
$$\underbrace{\qquad\qquad\qquad\qquad}_{\Gamma}$$

So since spt π_* lies in $\Omega_{\min}$ by Lemma 3.13, spt π_* is concentrated on the graph of the mapping Γ. Therefore, Γ uniquely solves the Monge optimal path problem.

Next, we show π_* coincides with $\Gamma_\sharp\mu_0$ on a subset of minimal paths $\Omega_{\min} \subset \Omega$. To this aim, on the endpoints we have

$$\begin{aligned}
\pi_*[\Omega_{\min}] &= \pi_*[(\Gamma \circ e_0)^{-1}(\Omega_{\min})] \\
&= (\Gamma \circ e_0)_\sharp \pi_*[\Omega_{\min}] \\
&= (e_0)_\sharp \pi_*[\Gamma^{-1}(\Omega_{\min})] \\
&= \mu_0[\Gamma^{-1}(\Omega_{\min})] \\
&= (\Gamma)_\sharp \mu_0[\Omega_{\min}].
\end{aligned}$$

The identity map $(\Gamma \circ e_0)^{-1}$ on $\Omega_{\min}$ on its endpoints implies the first equation, the definition of the push forward imply the second and third equations, $\pi_* \in \Pi_{\text{path}}(\mu_0, \mu_1)$ provided by Lemma 3.16 implies the fourth equation, and the definition of push forward of Γ implies the last.

Calling $\pi_* := \pi_\Gamma$, the transport plan π_Γ is unique. Suppose π_1 and π_2 solve the optimal path problem (1.5). Then Γ_1 and Γ_2 are optimal maps given by (3.9) with corresponding maps $T_1 : X \to X$ and $T_2 : X \to X$, respectively. Then $\pi_1 := (\Gamma_1)_\sharp\mu_0$ and $\pi_2 := (\Gamma_2)_\sharp\mu_0$ are optimal transport admissible plans by Lemma 3.16; and, consequently Γ_1 and Γ_2 are Monge solutions by Lemma 3.14. Then, by the linearity of π_Γ, $\overline{\pi} := \frac{1}{2}[\pi_1 + \pi_2]$ is optimal as well. And Theorem 1.7 says $\overline{\pi}$ is also a solution of (1.5). Then this means $\overline{\pi} := (\overline{\Gamma})_\sharp\mu_0$, for $\overline{\Gamma}$ as in (3.9), solving Monge as well. Thus, $\overline{\pi}$ is concentrated on the graph of $\overline{\Gamma}$. But this is impossible unless $\Gamma_1 = \Gamma_2 (= \overline{\Gamma})$ μ_0-a.e. In which case, $\pi_1 = \pi_2$. $\diamond$

Remark 3.17 *Minimizers are thus given by maps. A nontrivial question to think about then is: are these maps continuous? The regularity of optimal transport maps is an important and active area of research. One should note that when there is no interaction term, one could apply the standard optimal transport regularity theory [11], [9], [15] to understand regularity in the path dependent case. However, in the case one has interaction terms, then regularity becomes much more difficult and poses a natural and interesting problem.*

CHAPTER 4

THE OPTIMAL PATH PROBLEM WITH INTERACTION

In this chapter we consider an added interaction term to the functional $\mathcal{E}_0(\pi)$. We recall the linear functional in (1.8):

$$\mathcal{E}(\pi) := \int_\Omega c(\gamma)\, d\pi(\gamma) + \int_\Omega \int_\Omega \mathcal{K}(\gamma, \sigma)\, d\pi(\sigma)\, d\pi(\gamma),$$

where the Kernel $\mathcal{K}$ is given by (1.9). The plan is to determine properties of minimizers of (1.8) over Π_{path} and to determine whether the optimal plans are given by maps.

4.1 Existence of minimizers for the optimal path with interaction

The next result is proving existence of minimizers for the cost $c(\gamma)$ with added interaction term, namely proving Theorem 1.9. That is, we prove existence of solutions of (1.1), where $\mathcal{U}$ is given by (1.8) with (1.9).

Proof. [**Proof of Theorem 1.9**] What will help us achieve the existence of a minimizer is two-fold: Condition 1 to help us get enough compactness, just like in the proof of Theorem 1.6 and the next Lemma 4.2 which will allow us to use continuity to pass to the limit in the iterated integrals.

Since $\mathcal{K} \geq 0$ we have $\mathcal{E}_0(\pi) \leq \mathcal{E}(\pi)$ for all π in $\Pi_{\text{path}}(\mu_0, \mu_1)$. Thus $\inf_\pi \mathcal{E}(\pi) \geq \inf_\pi \mathcal{E}_0(\pi) \geq 0$; since the infimum is finite, there is a minimizing sequence $\{\pi_k\}_k$.

On the other hand using the same compact set, X_N, from Theorem 1.6 and coercivity condition, Condition 1, we deduce

$$\begin{aligned}\pi(X_N^c) &= \int_{X_N^c} d\pi(\gamma) = \frac{1}{N}\int_{X_N^c} N\, d\pi(\gamma) \\ &\leq \frac{1}{N}\int_\Omega c(\gamma)\, d\pi(\gamma) \\ &= \frac{1}{N}\mathcal{E}_0(\pi) \leq \frac{1}{N}\mathcal{E}(\pi).\end{aligned}$$

Then

$$\mathcal{E}(\pi_k) \longrightarrow \inf_{\pi \in \Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{E}(\pi) \quad \text{as} \quad k \to \infty.$$

Since $\{\pi_k\}_k$ is a minimizing sequence, then $\mathcal{E}(\pi_k) \leq C$ for all k. Since $\pi(X_N) = \pi(\Omega_{M,N})$, we then have

$$\pi_k(\Omega_{M,N}^c) \leq \frac{\mathcal{E}(\pi_k)}{N} \leq \frac{C}{N} \quad \text{for all} \quad k.$$

So given $\varepsilon := \frac{C}{N} > 0$, $\pi_k(\Omega_{M,N}^c) \leq \frac{C}{N} = \varepsilon$ for all k. This says that the sequence $\{\pi_k\}_k$ is tight, as $\Omega_{M,N}$ is compact (Condition 1); then Prokhorov's theorem (Theorem 3.2) tells us that there exists a subsequence $\{\pi_{k_j}\}_j$ in $\Pi_{\text{path}}(\mu_0, \mu_1)$ and a Borel probabilty measure φ such that $\pi_{k_j} \rightharpoonup \varphi$ as $k \to \infty$. From the proof of Theorem 1.6, we see φ is an element of $\Pi_{\text{path}}(\mu_0, \mu_1)$.

Next we will show

$$\mathcal{E}(\varphi) \leq \liminf_{j\to\infty} \mathcal{E}(\pi_{k_j}).$$

Here the proof differs from the non-interaction one. Lemma 4.2 below; the dominated convergence theorem give

$$\begin{aligned}\int_\Omega c(\gamma)\, d\varphi(\gamma) + \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\varphi(\sigma)\, d\varphi(\gamma) &= \lim_{n\to\infty}\Big(\int_\Omega c_n(\gamma)d\varphi(\gamma)\\ &+ \int_\Omega\int_\Omega \mathcal{K}_n(\gamma,\sigma)d\varphi(\sigma)d\varphi(\gamma)\Big)\\ &= \lim_{n\to\infty}\lim_{j\to\infty}\int_\Omega c_n(\gamma)d\pi_{k_j}(\gamma)\\ &+ \int_\Omega\int_\Omega \mathcal{K}_n(\gamma,\sigma)d\pi_{k_j}(\sigma)d\pi_{k_j}(\gamma)\\ &\le \liminf_{j\to\infty}\Big(\int_\Omega c(\gamma)d\pi_{k_j}(\gamma)\\ &+ \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi_{k_j}(\sigma)d\pi_{k_j}(\gamma)\Big).\end{aligned}$$

We have therefore proved

$$\mathcal{E}(\varphi) \le \liminf_{j\to\infty} \mathcal{E}(\pi_{k_j}),$$

and therefore φ is indeed a minimizer of $\mathcal{E}(\pi)$, as wanted to be shown. ◇

An important property that will be used in the proof of Lemma 4.2 comes form the following definition.

Definition 4.1 *Let X be a metric space and let $f : X \to \mathbb{R}$ be a function and let $b : [0, +\infty] \to [0, +\infty]$ be a function. We say that f has modulus of continuity b if for all $x, y \in X$, $|f(x) - f(y)| \le b(d(x, y))$.*

Lemma 4.2 *Let $(\pi_k)_k$ be a sequence of probability measures on Ω. Suppose $\mathcal{U}(\gamma, \pi)$, defined in (1.7), is bounded and π_k, π have finite total mass. If $\pi_k \rightharpoonup \pi$ weakly as $k \to \infty$, then*

$$\int_\Omega \mathcal{U}(\gamma, \pi_k)\, d\pi_k(\gamma) \longrightarrow \int_\Omega \mathcal{U}(\gamma, \pi)\, d\pi(\gamma) \quad \textit{as} \quad k \longrightarrow \infty.$$

Proof. Recall the definitions of (1.8) and (1.9). In terms of this notation we will show

$$\left|\int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)\ d\pi_k(\sigma)\ d\pi_k(\gamma) - \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)\ d\pi(\sigma)\ d\pi(\gamma)\right| \longrightarrow 0 \quad \text{as} \quad k\to\infty.$$

For each compact set $\Omega_{M,N}$ arising from the coercivity property with given $\varepsilon := \frac{C}{N}$, we shall show

$$\left|\int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)\ d\pi_k(\sigma)\ d\pi_k(\gamma) - \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)\ d\pi(\sigma)\ d\pi(\gamma)\right| \to 0 \quad \text{as} \quad k\to\infty.$$

Let us both unpack this. Indeed we have this expression is equal to

$$\begin{aligned}\Big|\int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi_k(\sigma)d\pi_k(\gamma) - \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)d\pi_k(\gamma) + \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)d\pi_k(\gamma)\\ - \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)d\pi(\gamma)\Big|,\end{aligned} \tag{4.1}$$

applying the triangle inequality the above (4.1) is less than or equal to

$$\begin{aligned}\int_\Omega\left|\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi_k(\sigma) - \int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)\right| d\pi_k(\gamma) + \Big|\int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)d\pi_k(\gamma)\\ - \int_\Omega\int_\Omega \mathcal{K}(\gamma,\sigma)d\pi(\sigma)d\pi(\gamma)\Big|\\ := I_k + J_k.\end{aligned} \tag{4.2}$$

Let us first investigate J_k. Since $\pi_k \rightharpoonup \pi$ as k tends to infinity, we know that

$$\int_\Omega f(\sigma)\ d\pi_k(\sigma) \longrightarrow \int_\Omega f(\sigma)\ d\pi(\sigma) \quad \text{for all} \quad f\in C_b(\Omega).$$

Fix γ and apply this to $f(\sigma) := \mathcal{K}(\gamma,\sigma)$. So

$$\mathcal{U}(\gamma,\pi_k) \longrightarrow \mathcal{U}(\gamma,\pi) \quad \text{as} \quad k\to\infty.$$

Claim. *If $\mathcal{K}(\gamma,\sigma)$ has a b-modulus of continuity in the first coordinate, then $\mathcal{U}(\gamma,\pi)$ has a b-modulus of continuity in the first coordinate independent of σ.*

Indeed, using the hypothesis that $\mathcal{K}$ has a b modulus of continuity and then integrating in σ for any π probability measure on Ω, we have

$$\begin{aligned} |\mathcal{U}(\gamma_1,\pi)-\mathcal{U}(\gamma_2,\pi)| &= \Big| \int_\Omega \mathcal{K}(\gamma_1,\sigma)d\pi(\sigma) - \int_\Omega \mathcal{K}(\gamma_2,\sigma)d\pi(\sigma)\Big| \\ &\leq \int_\Omega \Big|\mathcal{K}(\gamma_1,\sigma)-\mathcal{K}(\gamma_2,\sigma)\Big| d\pi(\sigma) \\ &\leq b(\|\gamma_1-\gamma_2\|) \end{aligned} \tag{4.3}$$

Then $\mathcal{U}(\gamma,\pi)$ is continuous and bounded, so since $\pi_k \rightharpoonup \pi$, we conclude that (recall (4.2)) $J_k \to 0$ as $k \to \infty$.

Next we study I_k. This one is a bit more delicate. The idea is to break the integral on the compact set $\Omega_{M,N}$ and outside the compact set and use Arzela-Ascoli.

$$\begin{aligned} I_k = &\int_{\Omega_{M,N}} \Big|\mathcal{U}(\gamma,\pi_k)-\mathcal{U}(\gamma,\pi)\Big|\, d\pi_k(\gamma) + \int_{\Omega\setminus\Omega_{M,N}} \Big|\mathcal{U}(\gamma,\pi_k)-\mathcal{U}(\gamma,\pi)\Big|\, d\pi_k(\gamma) \\ &:= I_{0,k}+I_{1,k}. \end{aligned} \tag{4.4}$$

Let us look at $I_{0,k}$. We know that $\{\mathcal{U}(\gamma,\pi_k)\}$ is equicontinuous. From the last estimate in (4.3), it is clear that $\{\mathcal{U}(\gamma,\pi_k)\}_k$ is equibounded and equicontinuous. Therefore, by Arzela-Ascoli, there exists a subsequence $\{\mathcal{U}(\gamma,\pi_{k_j})\}_j$ of $\{\mathcal{U}(\gamma,\pi_k)\}_k$ that converges uniformly. Now we claim that

$$\mathcal{U}(\gamma,\pi_{k_j}) \longrightarrow \mathcal{U}(\gamma,\pi) \quad \text{uniformly as} \quad j \to \infty.$$

Ineed, since we saw the pointwise convergence of $\{\mathcal{U}(\gamma,\pi_k)\}$, then this sequence is Cauchy. Suppose now towards sake of a contradiction that $\{\mathcal{U}(\gamma,\pi_{k_j})\}_j$ does not converge uniformly to $\mathcal{U}(\gamma,\pi)$. Then we can find $\varepsilon > 0$, such that for each integer $N > 0$, there exists some $j_0 \geq N$ such that

$$|\mathcal{U}(\gamma,\pi_{k_{j_0}})-\mathcal{U}(\gamma,\pi)| \geq \varepsilon.$$

As $\{\mathcal{U}(\gamma, \pi_k)\}$ is Cauchy, then given any $\varepsilon > 0$, there exists a $K > 0$ such that

$$|\mathcal{U}(\gamma, \pi_m) - \mathcal{U}(\gamma, \pi_k)| < \varepsilon/2 \quad \text{for all} \quad m, k \geq K.$$

The pointwise convergence of $\{\mathcal{U}(\gamma, \pi_k)\}$ gives that for any $\varepsilon > 0$, we can find an integer $K' > 0$ such that

$$|\mathcal{U}(\gamma, \pi_k) - \mathcal{U}(\gamma, \pi)| < \varepsilon/2 \quad \text{for all} \quad k \geq K'.$$

Then taking $K'' := \max(K, K')$,

$$|\mathcal{U}(\gamma, \pi_{k_j}) - \mathcal{U}(\gamma, \pi)| \leq |\mathcal{U}(\gamma, \pi_{k_j}) - \mathcal{U}(\gamma, \pi_m)| + |\mathcal{U}(\gamma, \pi_m) - \mathcal{U}(\gamma, \pi)| < \varepsilon \quad \text{for all } j, m \geq K'',$$

the desired contradiction. Therefore $\mathcal{U}(\gamma, \pi_{k_j})$ converges uniformly to the continuous $\mathcal{U}(\gamma, \pi)$. Relabeling $\{\mathcal{U}(\gamma, \pi_{k_j})\}$ to $\{\mathcal{U}(\gamma, \pi_k)\}$, subsequently $\{\mathcal{U}(\gamma, \pi_k)\}$ converges uniformly to $\mathcal{U}(\gamma, \pi)$ for all sufficiently large k.

This means that for any given $\varepsilon > 0$, we can find an integer $N > 0$, not depending on γ, with $N > 3/\varepsilon$ such that

$$\sup_{\gamma \in X_n} |\mathcal{U}(\gamma, \pi_k) - \mathcal{U}(\gamma, \pi)| \leq \frac{1}{k} \leq \frac{1}{N} < \frac{\varepsilon}{3} \quad \text{for all} \quad k \geq N \quad \text{sufficiently large.}$$

So that then for each fixed n,

$$I_{0,k} \leq \sup_{\gamma \in X_n} |\mathcal{U}(\gamma, \pi_k) - \mathcal{U}(\gamma, \pi)| \leq \frac{\varepsilon}{3} \quad \text{for all} \quad k \quad \text{sufficiently large.}$$

Subsequently, the uniform convergence—hence strong convergence—allows us to conclude $I_{0,k} \to 0$ as $k \to \infty$.

Let us now turn to $I_{1,k}$. From each compact $\Omega_{M,N}$ arising from the Condition 1 with $C/N > 0$, let $\varepsilon > 0$ be given such that $C/N =: \varepsilon/3$. Then quite simply for each fixed $N > 0$,

$$I_{1,k} \le \pi(\Omega \setminus \Omega_{M,N}) \sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |\mathcal{U}(\gamma, \pi_k) - \mathcal{U}(\gamma, \pi)| \le \frac{C}{N}\Big(\sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |\mathcal{U}(\gamma, \pi_k)| + \sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |\mathcal{U}(\gamma, \pi)| \Big) \le 2\frac{\varepsilon}{3} \quad \text{for all} \quad k.$$

Putting all this together in (4.4) and hence in (4.2) yields

$$I_k \le \frac{\varepsilon}{3} + 2\frac{\varepsilon}{3} = \varepsilon, \quad \text{for all} \quad k \quad \text{sufficiently large.}$$

The lemma is now proved. ◇

Remark 4.3 *Note that the above claim held true for general $\mathcal{K}$ having a b modulus of continuity. But for the more specific example, if we define $\mathcal{K}$ as*

$$\mathcal{K}(\gamma, \sigma) := \theta \int_0^1 e^{-\beta|\gamma(t) - \sigma(t)|^2}\, dt,$$

where $\beta > 0$ and $\theta > 0$, the claim also holds true. In fact, since $e^{-|x|^2}$ is Lipschitz,

$$\Big| e^{-|x|^2} - e^{-|y|^2} \Big| \le C|x - y|.$$

Interchanging the values of x and y we achieve the Lipschitz estimate. Thus, $\exp\{-\beta|\sigma(t) - \gamma(t)|^2\}$ *is Lipschitz. Then $\mathcal{K}(\gamma, \sigma) := \int_0^1 e^{-\beta|\gamma(t)-\sigma(t)|^2} dt$ is uniformly Lipschitz in each coordinate, $|\mathcal{K}(\gamma_1, \sigma) - \mathcal{K}(\gamma_2, \sigma)| \le C\|\gamma_1 - \gamma_2\|$. Then*

$$\begin{aligned} |\mathcal{U}(\gamma_1, \pi) - \mathcal{U}(\gamma_2, \pi)| &= \Big| \int_\Omega \mathcal{K}(\gamma_1, \sigma) d\pi(\sigma) - \int_\Omega \mathcal{K}(\gamma_2, \sigma) d\pi(\sigma) \Big| \\ &\le \int_\Omega \Big| \mathcal{K}(\gamma_1, \sigma) - \mathcal{K}(\gamma_2, \sigma) \Big| d\pi(\sigma) \\ &\le C\|\gamma_1 - \gamma_2\|. \end{aligned}$$

4.2 The dual problem

For the rest of the paper we only consider the interactions given by $\mathcal{K}(\gamma, \sigma) = \theta \int_0^1 \exp\{-\beta|\gamma(t) - \sigma(t)|^2\}\, dt$, for some $\theta, \beta > 0$. In order for the results here to be as "smooth" as possible, we will make a small notational change to the functional $\mathcal{E}(\pi)$, (1.8). Namely, without losing generality,

$$\mathcal{E}(\pi) := \int_\Omega c(\gamma) d\pi(\gamma) + \int_\Omega 2\, \mathcal{U}(\gamma, \pi) d\pi(\gamma). \tag{4.5}$$

The difference is the factor of 2 in front of $\mathcal{U}(\gamma, \pi)$.

We start with a heuristic discussion of the dual problem and Lagrange multipliers (we make this more rigorous in the next section). Just like in Section 3.2 we wish to minimize a functional subject to linear constraints. The novelty here is we have a nonlinear functional. Let us elaborate on this, the interaction term $\int_\Omega \Big(2\ \int_\Omega \mathcal{K}(\gamma, \sigma)\ d\pi(\sigma)\Big)\ d\pi(\gamma)$ is actually *quadratic* with respect to π. We venture into what we did in the beginning of Section 3.2 to produce the required constraint of the optimal path Kantorovich duality with interaction. For the moment let us abandon rigor and see where this takes us—it will take us to the correct dual problem when we look at interaction terms.

For $(\phi, \psi) \in C_c^0(X) \times C_c^0(X)$ and $\pi \in \mathcal{M}^+(\Omega)$, define

$$\begin{aligned}\mathcal{L}(\pi; \phi, \psi) := \int_\Omega c(\gamma) d\pi(\gamma) + \int_\Omega 2\, \mathcal{U}(\gamma, \pi) d\pi(\gamma) - \int_\Omega \phi(\gamma(0)) d\pi(\gamma) + \int_X \phi(x) d\mu_0(x) \\ - \int_\Omega \psi(\gamma(1)) d\pi(\gamma) + \int_X \psi(y) d\mu_1(y).\end{aligned}$$

After rearranging

$$\begin{aligned}\mathcal{L}(\pi; \phi, \psi) &= \int_\Omega \Big(c(\gamma) + 2\, \mathcal{U}(\gamma, \pi) - (\phi(\gamma(0)) + \psi(\gamma(1)))\Big)\, d\pi(\gamma) \\ &\qquad + \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) \\ &= \mathcal{Q}(\pi; \phi, \psi) + \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y),\end{aligned}$$

where

$$\mathcal{Q}(\pi;\phi,\psi) := \int_\Omega c(\gamma) + 2\,\mathcal{U}(\gamma,\pi) - (\phi(\gamma(0)) + \psi(\gamma(1)))\ d\pi(\gamma).$$

Notice that $\mathcal{Q}(\pi;\phi,\psi)$ contains the quadratic term that we talked about in the prequel. Then the dual function is

$$\mathcal{D}(\phi,\psi) := \inf_{\pi\geq 0}\left\{\mathcal{Q}(\pi;\phi,\psi) + \int_X \phi(x)\ d\mu_0(x) + \int_X \psi(y)\ d\mu_1(y)\right\}. \tag{4.6}$$

The business at hand is to minimize the quadratic term $\mathcal{Q}(\pi;\phi,\psi)$ with respect to π. But how can one minimize such quadratic term in π? This can be answered if we recognize this as minimizing a quadratic functional in infinite dimensions and compare to the more tangible problem in finite dimension.

The minimization of $\mathcal{Q}(\pi;\phi,\psi)$ has the form: $\min_{p\geq 0} q(p)$ for

$$q(p) := \langle b,p\rangle + \langle Ap,p\rangle, \tag{4.7}$$

where $\langle\ ,\ \rangle$ is the Euclidean inner product, $b\in\mathbb{R}^n$ is positive, and $A\in\mathbb{R}^{n\times n}$ a positive semi-definite matrix. The term $\langle Ap,p\rangle$ can be interpreted as the quadratic term in (4.6), that is, the term $\int_\Omega \mathcal{U}(\gamma,\pi)d\pi(\gamma)$. Futher, p represents π, while Ap represents $2\,\mathcal{U}(\gamma,\pi)$, and b represents $c(\gamma)-\phi(\gamma(0))-\psi(\gamma(1))$. Then

$$\langle b,p\rangle := \Big\langle c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1))), \pi\Big\rangle = \int_\Omega \big(c(\gamma) - (\ \phi(\gamma(0)) + \psi(\gamma(1))\ )\big)\ d\pi(\gamma);$$

$$\langle Ap,p\rangle := \Big\langle 2\,\mathcal{U}(\gamma,\pi), \pi\Big\rangle = \int_\Omega 2\Big(\mathcal{U}(\gamma,\pi)\Big)\ d\pi(\gamma).$$

Thanks to Fubini and Tonelli the inner products above make (kind of) sense and their sum equal $\mathcal{Q}(\pi;\phi,\psi)$. Then to minimize $\mathcal{Q}(\pi;\phi,\psi)$ with respect to $\pi\geq 0$ in $\mathcal{P}(\Omega)$, it suffices to minimize (4.7) with respect to $p\geq 0$. Take a derivative of $q(p)$ with respect to p, set it equal to zero, and solve for minimum. To be cautious we will use vector notation. Take a gradient with respect to p

$$\nabla_p q(p) = b + 2Ap.$$

We have the constraint $p \geq 0$, so at the minimizer $p_{\min}$ we have, by the Karush-Kuhn-Tucker (KKT) conditions [22, Ch 10.2.2],

$$\nabla_p q(p) \geq 0 \text{ (coordinate wise)},$$
$$\nabla_{p_i} q(p) = 0 \text{ if } p_i > 0.$$

Let us interpret this in our infinite dimensional problem, here $\pi_{\min}$ corresponds to the $p_{\min}$, and we have

$$\mathcal{U}(\gamma, \pi_{\min}) \geq -\frac{1}{2}(c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1))), \ \forall\, \gamma \in \Omega$$
$$\mathcal{U}(\gamma, \pi_{\min}) = -\frac{1}{2}(c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1))), \ \forall\, \gamma \in \operatorname{spt}(\pi_{\min}).$$

Rearranging, this can be written as

$$\phi(\gamma(0)) + \psi(\gamma(1)) \leq c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_{\min}), \ \forall\, \gamma \in \Omega, \tag{4.8}$$

$$\phi(\gamma(0)) + \psi(\gamma(1)) = c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_{\min}), \ \forall\, \gamma \in \operatorname{spt}(\pi_{\min}). \tag{4.9}$$

Thus, we have a good guess about the condition that characterizes the optimal π in (4.6). More concretely, and mimicking what we have already done in Section 3.2, if there exists a path γ for which $c(\gamma) + 2\,\mathcal{U}(\gamma, \pi) - (\phi(\gamma(0)) + \psi(\gamma(1))) < 0$, then we can make $\min_{\pi \geq 0} \mathcal{Q}(\pi; \phi, \psi) = -\infty$ and in turn make $\mathcal{D}(\phi, \psi) = -\infty$. Otherwise, for all $\sigma \in \Omega$, and keeping γ fixed (remembering the proof of Proposition 3.6), $\phi(\gamma(0)) + \psi(\gamma(1)) \leq c(\gamma) + 2\,\mathcal{U}(\gamma, \pi)$ π-a.e.

Note that for any $\pi \in \mathcal{P}(\Omega)$, we may consider the auxiliary dual problem,

$$D_\pi(\phi, \psi) = \sup \left\{ \int_X \phi(x)\; d\mu_0(x) + \int_X \psi(y)\; d\mu_1(y) : \phi(\gamma(0)) + \psi(\gamma(1)) \leq c(\gamma) + 2\,\mathcal{U}(\gamma; \pi) \right\} \tag{4.10}$$

For the moment let us assume that Theorem 3.8's equality holds for (4.5) having constraint (4.8). Suppose the following holds:

$$\inf \{\mathcal{E}(\pi) :\ \pi \in \Pi_{\text{path}}(\mu_0, \mu_1)\} = \sup \{\mathcal{D}(\phi, \psi)\} \tag{4.11}$$

Suppose (ϕ_0, ψ_0) maximizes (4.10) and π_0 is optimal for (4.5). Then, thanks to (4.8-30) $c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0) = \phi_0(\gamma(0)) + \psi_0(\gamma(1))$. Furthermore, strong duality (4.11) ensures

$$\begin{aligned}\int_\Omega [c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0)]\, d\pi_0(\gamma) &= \int_X \phi_0(x) d\mu_0(x) + \int_X \psi_0(y)\, d\mu_1(y)\\ &= \int_\Omega [\phi_0(\gamma(0)) + \psi_0(\gamma(1))] d\pi_0(\gamma).\end{aligned}$$

Equivalently

$$\int_\Omega [c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0) - (\phi_0(\gamma(0)) + \psi_0(\gamma(1)))] d\pi_0(\gamma) = 0.$$

The integrand is nonnegative by (4.8), so it has to vanish π_0-a.e. Moreover, if we further stipulate something very similar like the endpoint cost function (2.2) to $c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0)$, say (4.14) for which the constraint $\phi_0(x) + \psi_0(y) \leq \inf_{\gamma(0)=x, \gamma(1)=y}\ c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0)$ holds π_0-a.e., then we would have

$$\begin{aligned}c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0) &= \phi_0(\gamma(0)) + \psi_0(\gamma(1))\\ &= \phi_0(x) + \psi_0(y)\\ &\leq \inf_{\gamma(0)=x,\ \gamma(1)=y} [c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0)],\end{aligned}$$

which in turn would imply $c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0) = \inf_{\gamma(0)=x,\ \gamma(1)=x} [c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_0)]$.

With this in mind, we state the interaction analogue of Theorem 3.8, i.e., we establish (4.11), but prove it in the end of Section 4.3 after some preliminary discussion and a bit of mathematical machinery coming from Definition 4.5.

Theorem 4.4 *Strong duality holds. Namely,*

$$\inf_{\pi \in \Pi_{path}(\mu_0, \mu_1)} \int_\Omega \Big(c(\gamma) + 2\,\mathcal{U}(\gamma; \pi) \Big) d\pi(\gamma) = \sup_{\phi, \psi} \mathcal{D}(\phi, \psi).$$

4.3 Supergradient of a concave function

We give a rigorous account to our discussion in Section 4.2. The background material needed in this section comes from Rockafellar [18, §23]. Namely, we use the fact that x is in the maximization set of f—a concave function—if and only if 0 is in the supergradient of f.

From (4.6), let $q(\phi, \psi) := \inf_{\pi \geq 0} \mathcal{Q}(\pi; \phi, \psi)$. So that for all ϕ, ψ there is a certain $\pi_* \geq 0$ such that,

$$q(\phi, \psi) = \mathcal{Q}(\pi_*; \phi, \psi) \leq \mathcal{Q}(\pi; \phi, \psi), \quad \forall \pi \geq 0. \tag{4.12}$$

In this case, from equation (4.6),

$$\mathcal{D}(\phi, \psi) := \inf_{\pi \geq 0} \mathcal{L}(\pi; \phi, \psi) = \inf_{\pi \geq 0} \mathcal{Q}(\pi; \phi, \psi) + \mathcal{J}(\phi, \psi) = q(\phi, \psi) + \mathcal{J}(\phi, \psi).$$

It is immediate to see that $\mathcal{J}(\phi, \psi)$ is linear in ϕ, ψ. For each fixed $\pi \geq 0$, $\mathcal{Q}(\pi; \phi, \psi)$ is linear in ϕ, ψ and so $q(\phi, \psi)$ is a minimum of a family of linear functionals. In other words, for all (ϕ_0, ψ_0) we can find π_0 such that $q(\phi_0, \psi_0) = \mathcal{Q}(\pi_0; \phi_0, \psi_0)$ and $q(\phi_0, \psi_0) \leq \mathcal{Q}(\pi; \phi, \psi)$. Thus, q is concave in ϕ, ψ. The superdifferentiability analogue of $\mathcal{D}(\phi, \psi)$ is given—just as in [18, §23].

Remark 4.5 *The* superdifferential *of $\mathcal{D}$ is made of functions of the form*

$$\mathfrak{l}_\pi(\phi, \psi) := -\int_\Omega [\phi(\gamma(0)) + \psi(\gamma(1))]\, d\pi(\gamma) + \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y),$$

for some $\pi \geq 0$. In particular, $\mathfrak{l}_\pi$ is in $\partial\mathcal{D}(\phi_0, \psi_0)$ if

$$\mathcal{D}(\phi, \psi) \leq \mathcal{D}(\phi_0, \psi_0) + \mathfrak{l}_\pi(\phi - \phi_0, \psi - \psi_0) \quad \forall (\phi, \psi).$$

Notice the pair (ϕ_0, ψ_0) is a maximum of $\mathcal{D}$ if and only if the 0 functional is in $\partial\mathcal{D}(\phi_0, \psi_0)$, i.e., if there is a π such that $\mathfrak{l}_\pi(\phi, \psi) = 0$ for all (ϕ, ψ).

This aligns perfectly well with Rockafellar's statement [18]: π—having the form $\mathfrak{l}_\pi$—is in the maximization set of $\mathcal{D}(\phi, \psi)$ for all (ϕ, ψ) if and only if 0 is in the supergradient of $\mathcal{D}$. That is, if and only if $\mathfrak{l}_\pi$ is identically zero.

A consequence of this indicates that if there exists $\pi \geq 0$ such that for all ϕ, ψ

$$0 = \mathfrak{l}_\pi = -\int_\Omega [\phi(\gamma(0)) + \psi(\gamma(1))] d\pi(\gamma) + \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y),$$

then, $(e_0)_\sharp \pi = \mu_0$ and $(e_1)_\sharp \pi = \mu_1$ and thusly π is admissible.

Armed with this discussion, we are ready to prove Theorem 4.4.

Proof. [**Proof of Theorem 4.4**] Let (ϕ_*, ψ_*) be the maximum of $\mathcal{D}(\phi, \psi)$. In view of the above discussion (and Remark 4.5), there exists $\pi_* \geq 0$ such that $\mathfrak{l}_{\pi_*} \in \partial\mathcal{D}(\phi_*, \psi_*)$ and $\mathfrak{l}_{\pi_*} = 0$. Equivalently, $q(\phi_*, \psi_*) = \mathcal{Q}(\pi_*; \phi_*, \psi_*)$. In particular, $\pi_* \in \Pi_{\text{path}}(\mu_0, \mu_1)$. Then, for all (ϕ, ψ) and all admissible π, and applying the inequality (4.12) we have

$$\begin{aligned}
\int_\Omega c(\gamma) + 2\,\mathcal{U}(\gamma, \pi_*)\; d\pi_*(\gamma) &= \mathcal{L}(\pi_*, \phi_*, \psi_*) \\
&= \mathcal{Q}(\pi_*; \phi_*, \psi_*) + \int_X \phi_*(x) d\mu_0(x) + \int_X \psi_*(y) d\mu_1(y) \\
&\leq \mathcal{Q}(\pi; \phi_*, \psi_*) + \int_X \phi_*(x) d\mu_0(x) + \int_X \psi_*(y) d\mu_1(y) \\
&= \int_\Omega c(\gamma) + 2\,\mathcal{U}(\gamma, \pi)\; d\pi(\gamma) \quad \forall\, \pi \quad \text{admissible}
\end{aligned}$$

The latter inequality indicates π_* minimizes $\int (c(\gamma) + 2\,\mathcal{U}(\gamma; \pi)) d\pi$. So since π_* is optimal, then

$$\phi_*(\gamma(0)) + \psi_*(\gamma(1)) = c(\gamma) + 2\,\mathcal{U}(\gamma; \pi_*), \quad \text{in the support of } \pi_*.$$

Therefore,

$$\int_X \phi_*(x)\, d\mu_0(x) + \int_X \psi_*(y)\, d\mu_1(y) = \int_\Omega (\phi_*(\gamma(0)) + \psi_*(\gamma(1)))\, d\pi_*(\gamma)$$
$$= \int_\Omega (c(\gamma) + 2\,\mathcal{U}(\gamma;\pi_*))\, d\pi_*(\gamma)$$

π_*-a.e. But (ϕ_*, ψ_*) solves the maximization (4.6); thus

$$\sup \mathcal{D}(\phi,\psi) = \inf_{\pi\in\Pi_{\text{path}}} \mathcal{E}(\pi).$$

$\diamond$

Remark 4.6 *The above result says that any solution $(\phi_*.\psi_*)$ to the dual problem produces a solution π_* to the primal problem.*

A consequence of strong duality is a characterization of optimizers using the auxiliary dual problem. This is the content of the next result.

Theorem 4.7 *If π_0 solves the primal problem, then (ϕ_0, ψ_0) solves the auxiliary dual problem.*

Proof. First of all using the auxiliary dual problem (4.10), $D_\pi(\phi,\psi) \leq \mathcal{E}(\pi)$ for all (ϕ,ψ); for all admissible π. For the reverse inequality, we proceed as follows. Let π_0 be some minimizer for the primal problem. Define (ϕ_0, ψ_0) maximizing (4.10). From the previous Section 4.2, any pair (ϕ_0, ψ_0) that achieves the supremum are such that

$$\phi_0(\gamma(0)) + \psi_0(\gamma(1)) = c(\gamma) + 2\,\mathcal{U}(\gamma;\pi_0) \quad \forall\gamma \in \text{spt}(\pi_0). \tag{4.13}$$

This implies, as $\pi_0 \in \Pi_{\text{path}}(\mu_0, \mu_1)$,

$$\begin{aligned} D_\pi(\phi_0, \psi_0) &= \int_X \phi_0(x)\, d\mu_0(x) + \int_X \psi_0(y)\, d\mu_1(y) \\ &= \int_\Omega (\phi_0(\gamma(0)) + \psi_0(\gamma(1)))\, d\pi_0(\gamma) \\ &= \int_\Omega (c(\gamma) + 2\,\mathcal{U}(\gamma; \pi_0))\, d\pi_0(\gamma) \\ &= \mathcal{E}(\pi_0). \end{aligned}$$

The second to last equation follows from (4.13).

Remembering the Lagrangian multipliers functional from Section 4.2,

$$\mathcal{L}(\pi_0; \phi_0, \psi_0) = \int_\Omega (c(\gamma) + 2\,\mathcal{U}(\gamma; \pi_0) - (\phi_0(\gamma(0)) + \psi_0(\gamma(1)))\, d\pi_0(\gamma) + \int_X \phi_0\, d\mu_0 + \int_X \psi_0\, d\mu_1,$$

and thanks to (4.13) the first integral vanishes (and so $\mathcal{Q}(\pi_0; \phi_0, \psi_0)$ vanishes as well), and thus

$$\begin{aligned} \mathcal{L}(\pi_0; \phi_0, \psi_0) &= \int_X \phi_0(x)\, d\mu_0(x) + \int_X \psi_0(y)\, d\mu_1(y) \\ &= D_\pi(\phi_0, \psi_0). \end{aligned}$$

Hence, from (4.6), we have

$$\mathcal{D}(\phi_0, \psi_0) \leq \mathcal{L}(\pi_0; \phi_0, \psi_0) = \mathcal{E}(\pi_0).$$

But by the KKT conditions, the inequality becomes an equality, and hence $\mathcal{D}(\phi_0, \psi_0) = \mathcal{E}(\pi_0)$. ◇

4.4 The effective cost

Toward the end of Section 2 we introduced the *endpoint function* (2.2) which represented the minimum cost function of a path going from x to y. Now for the problem with interaction we will introduce what we call an "effective" endpoint cost function that includes the interaction term.

Therefore given π, define $c_0(\gamma) := c(\gamma) + 2\,\mathcal{U}(\gamma;\pi)$ as the *effective cost*. The analogue of the endpoint cost function (2.2) is

$$c_{\text{eff}}(x,y) := \inf_{\gamma(0)=x,\gamma(1)=y} c_0(\gamma). \tag{4.14}$$

In what follows we deal explicitly with

$$c(\gamma) = \int_0^1 \frac{1}{2}|\dot\gamma(t)|^2 dt, \quad \text{and} \quad \mathcal{U}(\gamma;\pi) := \theta \int_\Omega \int_0^1 e^{-|\gamma(t)-\sigma(t)|^2} dt\, d\pi(\sigma),$$

for some $\theta << 1$ to be determined later. We start by showing that the optimal path of $c_0(\gamma)$ is sufficiently close to the optimal path of $c(\gamma)$, if θ is small.

Proposition 4.8 *Fix $\pi \in \mathcal{P}(\Omega)$ and $\theta > 0$. Then for any path $\gamma : [0,1] \to X$ which is Lipschitz in $[0,1]$ and twice differentiable in $(0,1)$ define,*

$$c_0(\gamma) = \int_0^1 \frac{1}{2}|\dot\gamma(t)|^2 dt + 2\theta \int_0^1 \int_\Omega e^{-|\gamma(t)-\sigma(t)|^2} d\pi(\sigma)\, dt$$

Given x, y, let γ_0 be optimal for $c_0(\gamma)$. Then γ_0 is of the form $\gamma_0(t) = \gamma_{euc}(t) + \theta E(\theta,t)$; where γ_{euc} is optimal for $c(\gamma)$ and where $E(\theta,t)$ satisfies the estimate

$$\sup_{0\le t\le 1} |E(\theta,t)| \le \sqrt{n}.$$

Proof. Notice that $d\pi_t(z) := (e_t)_\sharp \pi$, therefore

$$\int_\Omega e^{-|\gamma(t)-\sigma(t)|^2} d\pi(\sigma) = \int_{\mathbb{R}^n} e^{-|\gamma(t)-z|^2} d\pi_t(z), \text{ for all } \gamma(t) \text{ and all } t \in [0,1];$$

so for every π, the interaction term produces a potential function,

$$V(x,t) = -2\,\theta \int_{\mathbb{R}^n} e^{-|x-z|^2}\, d\pi_t(z).$$

We quickly review the Euler-Lagrange equation solutions of critical points of $c_0(\gamma)$. Fix $\varepsilon > 0$ and consider $\varphi : [0,1] \to X$, a Lipschitz, twice differentiable function in

$(0,1)$ such that $\varphi(0) = \varphi(1) = 0$. Perturbing the minimizer γ_0 by $\varphi \in C^2([0,1])$, we get $\gamma_\varepsilon := \gamma_0 + \varepsilon\varphi$. Then looking at the Euler-Lagrange equation we have

$$0 = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} c_0(\gamma_\varepsilon) = \int_0^1 \Big(-\ddot{\gamma}_0(t) - 4\theta \int_{\mathbb{R}^n} e^{-|\gamma_0(t)-z|^2}(\gamma_0(t) - z)\, d\pi_t(z)\Big) \cdot \varphi(t)\, dt$$

for all $\varphi \in C^2([0,1])$ for all $0 \le t \le 1$. Hence, γ_0 solves the equation

$$\ddot{\gamma}_0(t) = -4\theta \int_{\mathbb{R}^n} e^{-|\gamma_0(t)-z|^2}(\gamma_0(t) - z) d\pi_t(z).$$

We estimate the exponential expression inside the integral. For radial $r = |x - z|$,

$$\Big| e^{-|x-z|^2}(x - z) \Big| \le e^{-r^2} r \le 1.$$

Then

$$|\ddot{\gamma}_0(t)| \le 4\theta \int_{\mathbb{R}^n} \Big| e^{-|\gamma_0(t)-z|^2}(\gamma_0(t) - z) \Big| d\pi_t(z) \le 4\theta.$$

So thanks to Proposition 2.5, we have that γ_0 is of the form $\gamma_0(t) = \gamma_{euc}(t) + \theta\, E(\theta, t)$. ◇

For our last result, we need the following derivative calculation:

Proposition 4.9 *Let γ_0 be a minimizer of $c_0(\gamma)$ in Ω. For all $0 \le t \le 1$, let $\gamma_h(t) := \gamma_0(t) + th\hat{e}$ be a path from $\gamma_h(0) = \gamma_0(0)$ to $\gamma_h(1) = \gamma_0(1) + h\hat{e}$, where $h \neq 0$ and $\hat{e} := \langle 0, \ldots, 1, \ldots, 0\rangle$ is a unit vector in $\mathbb{R}^n$. Then for all y*

$$\nabla_y c_{\mathit{eff}}(x, y) = x - y - 4\theta \int_0^1 \int_{\mathbb{R}^n} t \exp\{-|\gamma_0(t) - z|^2\}(\gamma_0(t) - z)\, d\pi_t(z)\, dt. \quad (4.15)$$

Proof. The proof of Lemma 2.9 can be incorporated here with $c_{\mathbf{e}}$ replaced by (4.14). ◇

We now provide a Lipschitz bound on $T_{\mathbf{z}}(\mathbf{w}) := e^{-|\mathbf{w}-\mathbf{z}|^2}(\mathbf{w} - \mathbf{z})$ that will help us prove Lemma 4.11.

Proposition 4.10 *Given* $\mathbf{w}, \mathbf{u}, \mathbf{z} \in \mathbb{R}^n$, *we have*

$$|T_{\mathbf{z}}(\mathbf{w}) - T_{\mathbf{z}}(\mathbf{u})| \leq n^2 |\mathbf{w} - \mathbf{u}|.$$

Proof. We firstly bound a partial derivative of $T_{\mathbf{z}}(\mathbf{w})$. We have

$$\partial_{w_j} T^i_{\mathbf{z}}(\mathbf{w}) = e^{-|\mathbf{w}-\mathbf{z}|^2}(\delta_{ij} - 2(w_i - z_i)(w_j - z_j)).$$

For radial $r = |\mathbf{w} - \mathbf{z}|$, we get an upper bound,

$$\left|\partial_{w_j} T^i_{\mathbf{z}}(\mathbf{w})\right| \leq e^{-|\mathbf{w}-\mathbf{z}|^2}(1 + |\mathbf{w} - \mathbf{z}|^2) \leq e^{-r^2}(1 + r^2) \leq 1.$$

Secondly, borrowing notation and calculations from Spivak's *Calculus on Manifolds* [23], we observe that

$$T^i_{\mathbf{z}}(\mathbf{w}) - T^i_{\mathbf{z}}(\mathbf{u}) = \sum_{j=1}^{n} [T^i_{\mathbf{z}}(w_1, \ldots, w_j; u_{j+1}, \ldots, u_n) - T^i_{\mathbf{z}}(w_1, \ldots, w_{j-1}; u_j, \ldots, u_n)].$$

As $T_{\mathbf{z}}(\mathbf{w})$ is continuously differentiable, the mean value theorem gives us for every i and j

$$T^i_{\mathbf{z}}(w_1, \ldots, w_j; u_{j+1}, \ldots, u_n) - T^i_{\mathbf{z}}(w_1, \ldots, w_{j-1}; u_j, \ldots, u_n) = (w_j - u_j)\partial_{w_j} T^i_{\mathbf{z}}(v_{ij}),$$

for some v_{ij}. The absolute value of the right-hand side of this is $|w_j - u_j| \; |\partial_{w_j} T^i_{\mathbf{z}}(v_{ij})| \leq |w_j - u_j|$. Then $|T^i_{\mathbf{z}}(\mathbf{w}) - T^i_{\mathbf{z}}(\mathbf{u})| \leq n|w_j - u_j| \leq n|\mathbf{w} - \mathbf{u}|$, since each $|w_j - u_j| \leq |\mathbf{w} - \mathbf{u}|$. Hence, $|T_{\mathbf{z}}(\mathbf{w}) - T_{\mathbf{z}}(\mathbf{u})| \leq \sum_{i=1}^{n} |T^i_{\mathbf{z}}(\mathbf{w}) - T^i_{\mathbf{z}}(\mathbf{u})| \leq \sum_{i=1}^{n} n|\mathbf{w} - \mathbf{u}| = n^2|\mathbf{w} - \mathbf{u}|$, and the proposition follows. ◇

Lemma 4.11 *Let* $\theta_0 := \frac{1}{\sqrt{2}} n^{-5/4}$. *If* $\theta < \theta_0$, *then for all* $y \in \mathbb{R}^n$, *and* $x_1 \neq x_2$,

$$\nabla_y \Big(c_{\mathit{eff}}(x_2, y) - c_{\mathit{eff}}(x_1, y) \Big) \neq 0.$$

Remark 4.12 *The above holds independent of* π.

Proof. For all $\theta > 0$, let $\gamma_{\theta,i}(t) = \gamma_{0,i}(t) + \theta E(\theta, t)$ be an optimal path from x_i to y with respect to $c_0(\gamma_h)$; where $\gamma_{0,i}(t) = x_i + t(y - x_i)$—geodesic path for $c(\gamma)$—for $i = 1, 2$ and $\theta \cdot E(\theta, t)$ an error term. From Proposition 4.9, we get for all y

$$\nabla_y\Big(c_{\text{eff}}(x_2, y) - c_{\text{eff}}(x_1, y)\Big) = x_1 - x_2 - 4\theta \int_0^1 \int_{\mathbb{R}^n} t\Big[T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))\Big] d\pi_t(z)dt.$$

Thanks to Proposition 4.10, we can estimate the expression inside the brackets. Namely, since we saw $|\ddot{\gamma_0}(t)| \leq 4\theta$, Proposition 2.5 applies to give

$$\begin{aligned}\Big|T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))\Big| &\leq n^2|\gamma_{\theta,2}(t) - \gamma_{\theta,1}(t)| \\ &\leq n^2 \max_{0\leq t\leq 1} |\gamma_{\theta,2}(t) - \gamma_{\theta,1}(t)| \\ &\leq n^{\frac{5}{2}}\theta|x_2 - x_1|.\end{aligned}$$

In particular,

$$\begin{aligned}\Big|4\theta \int_0^1 \int_{\mathbb{R}^n} t[T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))]d\pi_t(z)\, dt\Big| &\leq 4\theta \int_0^1 t \int_{\mathbb{R}^n} n^{\frac{5}{2}}\theta|x_2 - x_1|d\pi_t(z)\, dt \\ &= 2n^{\frac{5}{2}}\theta^2|x_2 - x_1|.\end{aligned}$$

Let θ_0 be such that $\delta := 2n^{\frac{5}{2}}\theta_0^2 < 1$. The reverse triangle inequality applies to show

$$\begin{aligned}\Big|\nabla_y\Big(c_{\text{eff}}(x_2, y) - c_{\text{eff}}(x_1, y)\Big)\Big| &\geq |x_1 - x_2| - 4\theta\Big|\int_0^1 t \int_{\mathbb{R}^n} [T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))]d\pi_t(z)\, dt\Big| \\ &\geq |x_1 - x_2| - 2n^{\frac{5}{2}}\theta^2|x_2 - x_1| \\ &= (1 - 2n^{\frac{5}{2}}\theta^2)|x_1 - x_2| > (1 - \delta)|x_1 - x_2|.\end{aligned}$$

Therefore, since $\delta < 1$, $|\nabla_y(c_{\text{eff}}(x_2, y) - c_{\text{eff}}(x_1, y))| > 0$, and the proof is complete.

◇

Theorem 1.11 now follows as a corollary of Lemma 4.11 and Theorem 1.7.

Proof. [**Proof of Theorem 1.11**] To prove this theorem, we must show that c_{eff} satisfies the assumptions of Brenier and Gangbo and McCann's theorems or

Theorem 10.28 from [25]. Remark 4.12 tells that we can write the effective cost as $c_{\text{eff},\pi_{\min}}$. This $\pi_{\min}$ will help us attain $c_{\mathbf{eff},\pi_{\min}}$ implicitly depending on $\pi_{\min}$. Pick $\pi_{\min}$ by Theorem 1.9. Then $\pi_{\min}$ is optimal with respect to $c_{\mathbf{eff},\pi_{\min}}$. Since $\theta < \theta_0$, Lemma 4.11 tells us $\nabla_y c_{\text{eff},\pi_{\min}}$ is injective, and Lemma 2.2 tells us $c_{\mathbf{eff},\pi_{\min}}$ is differentiable, and $\mu_0 \ll \ dx$, then $c_{\mathbf{eff},\pi_{\min}}$ satisfies the assumptions of Brenier and Gangbo and McCann's results. Therefore, there is a unique transport map T solving **Problem B**.

$\diamond$

The proof of Theorem 1.10 now follows as a corollary from Theorems 1.6, 1.9 and 1.9, 1.11. Indeed,

Proof. [**Proof of Theorem 1.10**] From the previous results stated above, one has the following. If π_0 is a minimizer of $\mathcal{E}_0(\pi)$, then π_0 is a minimizer of $\mathcal{E}(\pi)$ without the interaction term containing $c_0(\gamma)$. $\diamond$

CHAPTER 5

DISCRETE OPTIMAL TRANSPORTATION WITH INTERACTING PATHS

In this section we prove some properties of the quadratic measure (Definition 5.3). The properties will help us define *cyclically monotonicity* with interacting paths. Moreover, when the underlying measures are discrete we produce the N-body dynamics of the optimal transport problem. Using this, we also establish that the critical points of $\mathcal{E}(\pi)$, (5.1), give rise to weak solutions of (5.4).

5.1 Cyclical monotonicity with interacting paths

We start this section by giving the definition of the classical cyclical monotonicity taken from Amrosio's and Gigli's guide in [1].

Definition 5.1 *A subset $S \subset \mathbb{R}^n \times \mathbb{R}^n$ is* cyclically monotone *if $(x_i, y_i) \in S$ for all $i = 1, \ldots, n$*

$$\sum_{i=1}^{n} \langle x_i, y_i \rangle \geq \sum_{i=1}^{n} \langle x_i, y_{\tau(i)} \rangle,$$

for any permutation τ of n letters $[n] := \{1, \ldots, n\}$.

In the classical optimal transportation, one of the fundamental results, due to Rockafellar [18], is that a set S in $\mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if and only if there exists a convex and lower semicontinuous function φ such that the set S is concentrated in the graph of the subdifferential of φ. This result was extended by Knott and Smith, given in Villani's book [24], to be general with the aid of the following definition.

Definition 5.2 *Let X and Y be metric spaces, and $c : X \times Y \to \mathbb{R}$ a lower semicontinuous cost function. A subset $\mathcal{S} \subset X \times Y$ is* c-cyclically monotone *if $(x_i, y_i) \in \mathcal{S}$ for all $i = 1, \ldots, n$,*

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{\tau(i)})$$

for all permutations τ on $[n]$.

The general extended version of Rockafellar's result, due to Knott and Smith, is that if $c(x, y)$ is lower semicontinuous and bounded from below, $\mu_0 \in \mathcal{P}(X)$ and $\mu_1 \in \mathcal{P}(Y)$, and π a probability measure in $X \times Y$ with marginals μ_0 and μ_1, then the following are equivalent: π is optimal, the support of π is c-cyclically monotone, and that there exists a c-concave function φ for which the support of π is contained in the c-superdifferential of φ (please resort to Ambrosio's and Gigli's *A user's guide to optimal transpor* [1] for more information).

Definition 5.1 heuristically means that sending the point x_i to y_i for $i = 1, \ldots, n$ is globally much less expensive than sending the point x_i to $y_{\tau(i)}$. This criterion is independent of path. To incorporate paths, consider the following.

5.2 Cyclical monotonicity on quadratic measures

In this section we acquire the analog of the discrete version of cyclical monotonicity which incorporates interacting paths. In order to gain an understanding of this, we will rely on the quadratic measure, with respect to π, we saw in Section 4.2. In this section we formally introduce it and prove a few properties that will lead to the definition of the cyclical monotonicity with interacting paths.

Definition 5.3 *A quadratic measure $\mathcal{Q}(\cdot,\cdot)$ is measure a functional*

$$\mathcal{Q} : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R},$$

defined by

$$\mathcal{Q}(\pi_1, \pi_2) := \int_\Omega \int_\Omega K(\gamma, \sigma) d\pi_1(\sigma) d\pi_2(\gamma).$$

Then arranging and applying the notations from (1.7) and (1.8), fixing π and letting $\mathcal{E}(\pi) = \mathcal{E}_0(\pi) + \mathcal{Q}(\pi, \pi)$, we obtain

$$\mathcal{E}(\pi) := \int_\Omega c(\gamma)\ d\pi(\gamma) + \int_\Omega \int_\Omega K(\gamma, \sigma) d\pi(\sigma)\ d\pi(\gamma). \tag{5.1}$$

Stipulate further that $K : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}$ is convex, and is no longer as in equation (1.9), and $c(\gamma) = \int_0^1 \frac{1}{2} \|\dot{\gamma}(t)\|^2 dt$.

Lemma 5.4 *To any pair $\gamma, \sigma \in \Omega$ consider*

$$\mathcal{E}(\pi) = \int_\Omega c(\gamma) d\pi(\gamma) + \int_\Omega \int_\Omega K(\gamma, \sigma) d\pi(\sigma) d\pi(\gamma).$$

Then

(1) For any pair of measures π and μ: $\mathcal{E}(\pi+\mu) = \mathcal{E}(\pi) + \mathcal{E}_0(\mu) + 2\mathcal{Q}(\pi, \mu) + \mathcal{Q}(\mu, \mu)$.

(2) If π is a minimizer of $\mathcal{E}(\pi)$, then $\mathcal{E}_0(\mu) + 2\mathcal{Q}(\pi, \mu) \geq 0$.

Proof. *(1)* This is merely the quadratic formula on measures using the linearity of the "energy and potential" functional $\mathcal{E}(\pi)$:

$$\begin{aligned}
\mathcal{E}(\pi+\mu) &= \int_\Omega c(\gamma)d(\pi+\mu)(\gamma) + \int_\Omega\int_\Omega K(\gamma,\sigma)d(\pi+\mu)(\gamma)d(\pi+\mu)(\sigma) \\
&= \int_\Omega c(\gamma)d\pi(\gamma) + \int_\Omega c(\gamma)d\mu(\gamma) + \int_\Omega\int_\Omega K(\gamma,\sigma)d\pi(\gamma)d\pi(\sigma) \\
&+ \int\int K(\gamma,\sigma)d\mu(\gamma)d\pi(\sigma) + \int\int K(\gamma,\sigma)d\pi(\gamma)d\mu(\sigma) + \int\int K(\gamma,\sigma)d\mu(\gamma)d\mu(\sigma) \\
&= \int c(\gamma)d\pi(\gamma) + \int c(\gamma)d\mu(\gamma) + \int\int K(\gamma,\sigma)d\pi(\gamma)d\pi(\sigma) \\
&+ 2\int\int K(\gamma,\sigma)d\pi(\gamma)\, d\mu(\sigma) + \int\int K(\gamma,\sigma)d\mu(\gamma)\, d\mu(\sigma) \\
&= \mathcal{E}(\pi) + \mathcal{E}_0(\mu) + 2\mathcal{Q}(\pi,\mu) + \mathcal{Q}(\mu,\mu).
\end{aligned}$$

(2) Let $\varepsilon > 0$ be given. For any Borel probability measure μ perturb π by μ, namely we get $\pi + \varepsilon\mu$ such that

$$\mathcal{E}(\pi+\varepsilon\mu) \geq \mathcal{E}(\pi) \quad \forall\varepsilon,$$

since π is optimal. By *(1)* using the linearity of functional and splitting of the measures in the integrals we have

$$\begin{aligned}
\mathcal{E}(\pi+\varepsilon\mu) &= \mathcal{E}_0(\pi+\varepsilon\mu) + \mathcal{Q}(\pi+\varepsilon\mu, \pi+\varepsilon\mu) \\
&= \mathcal{E}_0(\pi) + \varepsilon\mathcal{E}_0(\mu) + \mathcal{Q}(\pi,\pi) + 2\varepsilon\mathcal{Q}(\pi,\mu) + \varepsilon^2\mathcal{Q}(\mu,\mu) \\
&= \mathcal{E}(\pi) + \varepsilon\Big(\mathcal{E}_0(\mu) + 2\mathcal{Q}(\pi,\mu)\Big) + \varepsilon^2\mathcal{Q}(\mu,\mu) \\
&\geq \mathcal{E}(\pi) \quad \forall\varepsilon.
\end{aligned}$$

So the function of ε of the right in the last inequality is always ≥ 0 for all ε; and, since $\mathcal{E}(\pi)$ achieves its minimum at $\varepsilon = 0$, since π is a critical point, the function of ε: $f(\varepsilon) := \varepsilon(\mathcal{E}_0(\mu) + 2Q(\pi,\mu)) + \varepsilon^2 Q(\mu,\mu)$ equals 0 when $\varepsilon = 0$. Its derivative with respect to ε is

$$\frac{d}{d\varepsilon} f(\varepsilon)\Big|_{\varepsilon=0} = \mathcal{E}_0(\mu) + 2Q(\pi,\mu) + 2\varepsilon Q(\mu,\mu)\Big|_{\varepsilon=0}. \tag{5.2}$$

So since this must vanish for $\varepsilon = 0$, we conclude that for all $\mu \in \mathcal{P}(\Omega)$, (5.2) yields

$$\mathcal{E}_0(\mu) + 2Q(\pi,\mu) \geq 0$$

$\diamond$

The above inequality, however, is not very telling. In any case, we have an idea what the term a priori could mean heuristically: the discrete analog of the cyclical monotonicity with interacting paths. Recall the characterization of shifted paths in (3.5). To wit, the admissibility of the paths $\widetilde{\gamma}_k$ from x_k to y_{k+1} is thus:

$$\gamma_k(0) = x_k \quad \text{and} \quad \gamma_k(1) = y_k,$$

$$\widetilde{\gamma}_k(0) = \gamma_k(0) \quad \text{and} \quad \widetilde{\gamma}_k(1) = \gamma_{k+1 \bmod n}(1).$$

Theorem 5.5 *Suppose π is a minimizer for*

$$\mathcal{E}(\pi) \quad \textit{over} \quad \Pi_{path}(\mu_0,\mu_1).$$

Let $\{\gamma_i\}_{i=1}^n \subset \{\omega_i\}_{i=1}^N$ be contained in $\Omega(x_k,y_k)$ and $\{\widetilde{\gamma}_i\}_{i=1}^n \subset \Omega(x_k,y_k)$ be admissible, where for each integer $k \geq 1$,

$$\Omega(x_k,y_k) := \{\gamma_k : [0,1] \to X : \ \gamma_k(0) = x_k, \ \gamma_k(1) = y_k\}.$$

Then

$$\sum_{l=1}^n c(\gamma_l) + 2\sum_{l=1}^n \sum_{k=1}^N \lambda_k K(\omega_k,\gamma_l) \leq \sum_{l=1}^n c(\widetilde{\gamma}_l) + 2\sum_{l=1}^n \sum_{k=1}^N \lambda_k K(\omega_k,\widetilde{\gamma}_l)$$

Proof. Firstly let $\varepsilon > 0$ be given. Let δ_{γ_m} be the Dirac point measure assigning unit mass to spt(π) whenever $\gamma_m \in \text{spt}(\pi) \subset X$ for all m. Let $\pi = \sum_{k=1}^{N} \lambda_k \delta_{\omega_k}$ and $\mu = \sum_{l=1}^{n} \delta_{\widetilde{\gamma}_l} - \delta_{\gamma_l}$. Since π is a minimizer for $\mathcal{E}(\pi)$ by Lemma 5.4 (*2*),

$$\mathcal{E}_0(\mu) + 2Q(\pi, \mu) \geq 0.$$

Plugging our π, μ to the above we get

$$\sum_{l=1}^{n} c(\widetilde{\gamma}_l) - c(\gamma_l) + 2\sum_{l=1}^{n}\sum_{k=1}^{N} \lambda_k K(\omega_k, \widetilde{\gamma}_l) - 2\sum_{l=1}^{n}\sum_{k=1}^{N} \lambda_k K(\omega_k, \gamma_l) \geq 0.$$

This clearly establishes the conclusion of the theorem. $\diamond$

Theorem 5.5 tells us that the c-cyclically monotonicity property is somehow preserved in the discrete optimal transport with interacting paths. When the underlying measures are discrete, the N-body dynamics from Newton's theory, applying the Euler-Lagrane equation, which follows from Lemma 5.4, give us a variation of Newton's equation. This is the content of the next section.

5.3 N-body dynamics with interacting paths

This section is concerned with N-body dynamics when the underlying measures are discrete. We prove a variation of Newton's dynamical equation. Namely, the critical points of (5.1) give rise to a solution, in the weak sense, of the differential equation (5.4). Namely, since Theorem 1.9 says π is a minimizer of (1.8), the discussion in the end of Section 4.2, with the help of the discussion on Section 2.2 on Lagrangian minimal paths, imply that any path in the support of π must be minimal with respect to the effective cost (4.14), and the Euler-Lagrange equation implies that γ is a solution of

$$\ddot{\gamma}(t) = -\int_{\Omega} \nabla K(\gamma(t) - \sigma(t))\, d\pi(\sigma).$$

Recall the map $\Gamma : X \to \Omega$ defined by (3.9). We will perturb this map by a test function φ as follows. Given any $\varepsilon > 0$, let Γ_ε be defined as

$$\Gamma_\varepsilon(x, 0) = x \quad \text{for all } x \in X \quad \text{and} \quad \Gamma_\varepsilon(x, 1) = T_\varepsilon(x)\ \forall \varepsilon,$$

where $(T_\varepsilon)_\sharp \mu_0 = \mu_1$ for any pair of proability measures $\mu_0, \mu_1 \in \mathcal{P}(X)$, and consider a test function $\zeta \in C_c^2(X)$ such that

$$T_\varepsilon(x) := T(x) + \varepsilon\zeta(x).$$

Then

$$\Gamma_\varepsilon(x, t) := \Gamma(x, t) + \varepsilon\varphi(x, t) + o(\varepsilon),$$

where $\varphi \in C_c^2(X \times [0, 1])$. Suppose Γ_ε is smooth in ε.

Lemma 3.16 tells us $\pi_\varepsilon := (\Gamma_\varepsilon)_\sharp \mu_0$ is admissible. Moreover, Γ_ε has marginals μ_0 and μ_1. One can see this from the above, i.e., when $\varepsilon = 0$, we recover $T(x)$ such that T pushes forward μ_0 to μ_1. Furthermore for every $\varepsilon > 0$, since φ vanishes, in the time component, at the boundary points of $[0, 1]$ we have at $t = 0$

$$\Gamma_\varepsilon(x, 0) = \Gamma(x, 0) = x \quad \forall \varepsilon$$

and at $t = 1$

$$\Gamma_\varepsilon(x, 1) = \Gamma(x, 1) = T(x) \quad \forall \varepsilon.$$

Hence, Γ_ε recovers the properties of (3.9). Using this knowledge we can show that if π_Γ, as defined in Lemma 3.16, is a critical point of $\mathcal{E}_0(\pi)$, then π_Γ is concentrated on straight lines.

Lemma 5.6 *If* $\pi_\Gamma = (\Gamma)_\sharp\mu_0$ *is a minimizer of* $\mathcal{E}_0(\pi)$*, then* π_Γ *is concentrated on constant speed geodesics.*

Proof. We have, from Chapter 2, with $c(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|^2 dt$ and $\pi_\Gamma = (\Gamma)_\sharp\mu_0$, that

$$\mathcal{E}_0(\pi_\Gamma) = \int_X \left(\int_0^1 |\partial_t \Gamma(x,t)|^2 \right) d\mu_0(x). \tag{5.3}$$

Suppose that Γ is a critical point of (5.3). Then we have $\mathcal{E}_0(\Gamma_\varepsilon) \geq \mathcal{E}_0(\Gamma)$ for all $\varepsilon > 0$, and Γ_ε is given by perturbing Γ by φ (as is given above). Using the Euler-Lagrange equation, for all $\varepsilon > 0$,

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{E}_0(\pi_\varepsilon) = 2 \int_X \int_0^1 \partial_t \Gamma \cdot \partial_t \varphi \, dt \, d\mu_0(x) \geq 0.$$

Applying integration by parts to the above yields

$$\begin{aligned} \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{E}_0(\pi_\varepsilon) &= -2 \int_X \int_0^1 \partial_{tt} \Gamma \cdot \varphi dt \, d\mu_0(x) \quad \forall \varphi \\ &\geq 0. \end{aligned}$$

Since φ is compactly supported on $X \times [0,1]$, and by symmetry of φ, putting $-\varphi$ into the above inequality, we get equality.

Then the claim is that for μ_0-a.e. x, $\partial_{tt}\Gamma(x,t) \equiv 0$ for all $0 \leq t \leq 1$. Then this would mean that $\Gamma(x,t)$ has to be linear in t. Therefore, π_Γ would be supported on straight lines. The proof of this fact is as follows.

Let us prove the claim. Since $\int_0^1 \partial_{tt}\Gamma \cdot \varphi \, dt \geq 0$ is bounded and supported on $\text{spt}(\pi_\Gamma)$, it will suffice to show

$$\int_0^1 \partial_{tt} \Gamma \cdot \varphi \, dt = 0 \quad \forall \, \varphi.$$

To that end, for each $k \geq 1$ consider the measurable set

$$G_k := \left\{ x \in \text{spt}(\pi_\Gamma) : \int_0^1 \partial_{tt}\Gamma(x,t) \cdot \varphi(x,t) dt \geq 1/k, \ \forall \varphi \in C_c^2(X \times [0,1]) \right\}.$$

Then

$$\mu_0(G_k) = \int_{G_k} d\mu_0(x) = k \int_{G_k} \frac{1}{k}\, d\mu_0(x)$$
$$\leq k \int_X \left(\int_0^1 \partial_{tt}\Gamma(x,t) \cdot \varphi(x,t) dt \right)\, d\mu_0(x).$$

In particular,

$$\frac{1}{k}\mu_0(G_k) \leq \int_X \left(\int_0^1 \partial_{tt}\Gamma(x,t) \cdot \varphi(x,t) dt \right)\, d\mu_0(x) \quad \forall \varphi.$$

Since $\int_X \left(\int_0^1 \partial_{tt}\Gamma(x,t) \cdot \varphi(x,t) dt \right)\, d\mu_0(x) = 0$ for all φ, then $\mu_0(G_k) = 0$ for all k. As

$$G := \left\{ x \in \mathrm{spt}(\pi_\Gamma) : \int_0^1 \partial_{tt}(x,t) \cdot \varphi(x,t) dt > 0,\ \forall \varphi \in C_c^2(X \times [0,1]) \right\} = \bigcup_{k=1}^{\infty} G_k,$$

from this we extract

$$\mu_0(G) \leq \sum_{k=1}^{\infty} \mu_0(G_k) = 0 \quad \forall k,$$

and therefore conclude

$$\int_0^1 \partial_{tt}\Gamma(x,t) \cdot \varphi(x,t)\, dt = 0 \quad \forall \varphi.$$

Then we see that for μ_0-a.e x, $\partial_{tt}\Gamma(x,t) = 0$ for all $0 \leq t \leq 1$, and the claim now follows. $\diamond$

Now for the meat of this section.

Theorem 5.7 *The critical points of the functional (5.1) give rise to weak solutions of the differential equation*

$$-\partial_{tt}\Gamma(x,t) + \int_X \nabla K(\Gamma(x,t), \Gamma(y,t)) d\mu_0(y) \equiv 0. \tag{5.4}$$

Proof. Thanks to Lemma 3.16, π_Γ is admissible for the functional (5.1). The change of variables formula (1.2) gives

$$\mathcal{E}(\pi_\Gamma) = \int_X \int_0^1 |\partial_{tt}\Gamma(x,t)|^2\, dt\, d\mu_0(x) + \int_X \int_X \int_0^1 K(\Gamma(x,t),\Gamma(y,t))dt\, d\mu_0(x) d\mu_0(y).$$

By hypothesis π_Γ is a critical point of (5.1), and applying Lemma 5.4, (*2*),

$$\mathcal{E}_0(\mu) + 2\ \mathcal{Q}(\pi,\mu) \geq 0,$$

for any pair of measures $\pi, \mu \in \mathcal{P}(\Omega)$. Considering the perturbations, Γ_ε, of Γ given by φ satisfying the prescribed conditions of (3.9), the quadratic functional in Lemma 5.4's (*2*), with $\pi_\varepsilon = (\Gamma_\varepsilon)_\sharp \mu_0$, for all ε, takes the form

$$F(\varepsilon) := \int_X \int_0^1 |\partial_t\Gamma_\varepsilon(x,t)|^2\, dt\, d\mu_0(x) + \int_X \int_X \int_0^1 K(\Gamma_\varepsilon(x,t),\Gamma(y,t))dt \mu_0(x)\, d\mu_0(y),$$

and it is greater than or equal to zero. So since Γ is a minimizer of $F(\varepsilon)$, applying the Euler-Lagrange equation to $F(\varepsilon)$, we get at $\varepsilon = 0$, that $F(\varepsilon)$ achieves its minimum. That is

$$\begin{aligned}\frac{\partial F(\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0} &= 2\int_X \int_0^1 \left(-\partial_{tt}\Gamma(x,t) + \int_X \nabla K(\Gamma(x,t),\Gamma(y,t))d\mu_0(y)\right)\cdot \varphi(x,t) dt\, d\mu_0(x)\\ &\geq 0 \quad \forall \varphi.\end{aligned}$$

Again, by symmetry, putting $-\varphi$ in the above inequality yields an equality. Performing the same argument as in the proof of the claim in Lemma 5.6, but using the measurable set

$$E_k := \left\{ x \in \operatorname{spt}(\pi_\Gamma) : \int_0^1 \left(-\partial_{tt}\Gamma(x,t) + \int_X \nabla K(\Gamma(x,t),\Gamma(y,t))d\mu_0(y)\right) dt \geq 1/k, \forall\varphi \right\},$$

instead of G_k and $E := \left\{ x : \int_0^1 \left(-\partial_{tt}\Gamma(x,t) + \int_X \nabla K(\Gamma(x,t),\Gamma(y,t))d\mu_0(y)\right) dt > 0, \forall\varphi \right\}$ instead of G, we establish equation (5.4) for μ_0-a.e. x. ◇

CHAPTER 6

A BENAMOU-BRENIER THEORY FOR INTERACTING PATHS

In this Chapter we start by looking at the optimal transport problem in the context of fluid mechanics, following the work of Benamou and Brenier in [2]. We start by introducing two source densities and a time-dependent density flowing between them on the interval $(0, 1)$. Assume both the source and the target densities have bounded support such that at time zero we recover the source density and at time one it evolves to the target density. Such a flow travels through a vector field that preserves mass. This is done by the continuity equation which can be described as a mass preserving flow.

In Lagrangian coordinates, we can follow the flow of a particle evolving from the source density at time zero to the target density at time one for which the flow conserves mass. Using this information we can compute the total transport cost—the cost function being the Euclidean distance squared between x and $\nabla u(x)$, the gradient of a convex function is the optimal transport solution when the cost function is quadratic [5], equals the minimum over the density and velocity field of the linear functional known as the Benamou-Brenier functional

$$\int_0^1 \int_{\mathbb{R}^n} \frac{1}{2}\rho\|v\|^2 dxdt.$$

To actually solve the minimum Benamou and Brenier [2] incorporated an augmented Lagrangian to solve a saddle point problem. Instead of minimizing over the density and velocity, they minimized over the density and momentum (which is equal to the density times the velocity). After they wrote the new functional and incorporated the momentum term, they applied the Lagrangian multiplier, used integration by parts, and got the new linear functional with integrand momentum squared over twice the density. The integrand term then equals to a supremum term. In this case, they turned this saddle point problem into a minimum-maximum problem of an inner-product of three vector-valued terms, one of which involves the time and spatial gradient of the multiplier; which can be solved.

Following Benamou's and Brenier's theory, we consider the cost function as the kinetic energy of a path and add a total cost of congestion between paths γ and σ: $\int K(\gamma, \sigma) d\pi(\sigma)$. To study this effectively, we require a probability distribution of paths to be given by a map and a time-dependent density to be the projection of such probability distribution via an evaluation map to a closed ball, where the source marginals live. Then using this, the total kinetic energy of paths equals to the total kinetic energy of the velocity and a density dependent on time. We can see that the time derivative of the path stands for the velocity of the path at some time. In the same spirit, the total cost of transporting along all paths is equal to the total spatial term $K(x, y)$ times the time-dependent spatial source densities. Putting all of this together we obtain

$$\int_{\mathbb{R}^n} \int_0^1 \left(\frac{1}{2} \|v(x,t)\|^2 + \int_{\mathbb{R}^n} K(x,y) \rho(y,t) dy \right) \rho(x,t) dt dx \tag{6.1}$$

subject to the continuity equation. Before we elaborate more on this, let us talk about Benamou's and Brenier's fluid mechanics solution to the optimal transport problem.

6.1 Classical Benamou-Brenier theory reviewed

In [2], Benamou and Brenier used fluid mechanics to develop a differential formulation of optimal transport. Consider some probability densities $\mu_0, \mu_1 \in \mathcal{P}(X)$. Let $\rho(x,t) \geq 0$ be a density for $x \in \mathbb{R}^n$, $0 \leq t \leq 1$. We want

$$\rho(x,0) = \mu_0(x),\ \rho(x,1) = \mu_1(x). \tag{6.2}$$

We may think of $\mu_0(x) = f(x)$ and $\mu_1(x) = g(x)$ as density functions, for all intents and purposes. Introduce a velocity field $v(x,t) \in \mathbb{R}^n$ where the particles can move around which is mass preserving. So we describe the mass preserving flow through the continuity equation

$$\partial_t \rho(x,t) + \nabla_x(\rho(x,t)v(x,t)) = 0. \tag{6.3}$$

In Lagrangian coordinates, if $X(x,t)$ is the position of a particle, then $X(x,t)$ follows a particle trajectory through the flow

$$\begin{aligned} \tfrac{dX(x,t)}{dt} &= v(X(x,t),t) \\ X(x,0) &= x. \end{aligned} \tag{6.4}$$

In the (x,t)-plane, the density flow evolves from the source density $X(x,0)$ at time $t=0$ to the target density $X(x,1)$ at time $t=1$. Through the evolution of particles, this can be thought of as $X(x,0) = \mu_0$ and $X(x,1) = \mu_1$ or more precisely as $X(x,0) = \rho(x,0)$ and $X(x,1) = \rho(x,1)$. In fact, if the evolution of the particles' flow is set in $\mathbb{R}^n \times [0,1]$ and $E \subset \mathbb{R}^n$, then for every $x \in E$ the flow of the particle trajectory from $\rho(x,0)$ evolving to $\rho(x,1)$ has an intermediate density, $\rho(x,t)$ within $E \times [0,1]$. Moreover, the mass in E at $t=0$ must coincide with the mass in $X(E,t)$ at some certain time t, to preserve mass. Then the characterization of measure preserving mass from the Introduction Chapter 1 applies here to mean

$$\int_E f(x)dx = \int_{X(E,t)} \rho(x,t)dx.$$

In other symbols, $X_\sharp f(x) = \rho(x,t)$ and $X_\sharp f = g$.

Benamou and Brenier, then, in [2], reformulated the time-dependent optimal transportation problem in a fluid mechanics setting. Their first result was the following proposition.

Proposition 6.1 (Benamou-Brenier [2](2000))

$$\min\left\{\int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\rho(x,t)|v(x,t)|^2\, dtdx : (6.2) \textit{ and } (6.3) \textit{ hold}\right\}$$
$$= \min_{T_\sharp f = g}\left\{\int_{\mathbb{R}^n}\frac{1}{2}\left|T(x)-x\right|^2 f(x)\, dx\right\}.$$

Remark 6.2 *We will see that the minimum in the right hand side is achieved and that the $X(x,1)$ turns out to be the minimal map, $\nabla\Psi(x)$, where $\Psi(x)$ is convex by Brenier's result [5].*

Proof. We start by showing the minimum on the left is no smaller than the other one. We have

$$\begin{aligned}
\inf_{X_\sharp f = g}\int_{\mathbb{R}^n}\left|X(x,1)-x\right|^2 f(x)\, dx &\le \int_{\mathbb{R}^n}\left|X(x,1)-x\right|^2 f(x)\, dx \\
(6.4) &= \int_{\mathbb{R}^n}\left|X(x,1)-X(x,0)\right|^2 f(x)dx \\
&= \int_{\mathbb{R}^n}\left|\int_0^1 \frac{d}{dt}X(x,t)dt\right|^2 f(x)dx \\
(6.4) &\le \int_{\mathbb{R}^n}\int_0^1 \left|v(X(x,t),t)\right|^2 f(x)\, dtdx,
\end{aligned}$$

where we used the fact that $X(x,t)$ solves the particle trajectory differential equation flow (6.4). Since $(X(x,t))_\sharp f = \rho(x,t)$,

$$\int_{\mathbb{R}^n}|v(X(x,t),t)|^2 f(x)dx = \int_{\mathbb{R}^n}\left|v(x,t)\right|^2\rho(x,t)\, dx$$

Thus we conclude that

$$\inf_{X_\sharp f = g}\int_{\mathbb{R}^n}\left|X(x,1)-x\right|^2 f(x)dx \le \int_{\mathbb{R}^n}\int_0^1 \rho(x,t)\left|v(x,t)\right|^2 dt\, dx.$$

Since $X(x,1) := T(x)$ satisfies (6.2) such that $T : f \mapsto g$, and by Brenier's Theorem, the infimum on the left hand side is attained for the optimum map $T(x) := \nabla\Psi(x)$ such that $\nabla\Psi$ pushes forward $f \mapsto g$, assuming f and g are "nice" enough, we conclude the proof of the proposition. $\diamond$

Remark 6.3 *The above result also shows that*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 \, d\pi(x,y) \leq \int_{\mathbb{R}^n}\int_0^1 \rho(x,t)|v(x,t)|^2 dtdx,$$

just let $\pi := (id \times T)_\sharp\mu$*,* $\mu(x) = f(x)dx$*, and* $y = T(x) := X(x,1)$*; for which* $\frac{dT_t(x)}{dt} = v(T_t(x),t)$ *and* $T_0(x) = x$*. In fact,* $X(x,t) = x + t\nabla\Psi(x)$*;* $v(x,t) = \nabla\Psi(x)$*.*

The question now arises, *how can we actually solve this minimization problem of the left hand side of Proposition 6.1?* That is, do minimizers exist of $\inf_{\rho,v}\int\int_0^1 \rho|v|^2 dtd$? In order to answer this question, we first resort to the *toy problem* discussed in Section 3.2. Namely, in order to minimize the Benamou-Brenier functional we must write its corresponding Lagrange function with Lagrangian multiplier. To this end, we shall write the flow problem, the minimization of the left hand side of Proposition 6.1 subject to (6.2) and (6.3) with multiplier ϕ as a Lagrange function.

First we write the continuity equation in the distributional sense multiplied by the multiplier ϕ. Indeed, by integration by parts and $\phi \in C_c^2(\mathbb{R}^n \times [0,1])$ we have

$$\int_{\mathbb{R}^n}\int_0^1 (\partial_t\rho_t + \nabla_x(\rho_t v_t))\, \phi \, dtdx = 0$$

implies

$$\int_{\mathbb{R}^n} (\phi(x,1)\mu_1 - \phi(x,0)\mu_0)\, dx + \int_{\mathbb{R}^n}\int_0^1 (-\rho_t\partial_t\phi - \rho_t v_t \nabla_x\phi)\, dtdx = 0.$$

Setting $G(\phi) := \int_{\mathbb{R}^n} (\phi(x,1)\mu_1 - \phi(x,0)\mu_0)\, dx$ and using $E(x,t) = \rho(x,t)v(x,t)$, the *toy problem* indicates that we may write the Lagrangian as follows:

$$\mathfrak{L}(\phi, E, \rho) := \int_{\mathbb{R}^n} \int_0^1 \left(\frac{|E(x,t)|^2}{2\rho(x,t)} - \rho(x,t)\partial_t \phi - E(x,t)\nabla_x \phi \right) dtdx + G(\phi). \quad (6.5)$$

We seek to minimize the *Lagrangian* (6.5) by solving the "saddle point" problem:

$$\inf_{\rho,E} \sup_{\phi} \mathfrak{L}(\phi, \rho, E).$$

This is exactly what Benamou and Brenier did in [2]. Since the integrand of (6.5) is convex with respect to $|E|^2/2\rho$, we can write this as a supremum of several affine functions. Furthermore to that statement, if we let $h(\rho, E) := |E|^2/2\rho$, then the following claim holds true.

Proposition 6.4

$$h(\rho, E) = \sup_{(a,b)\in K} \left\{ a\rho + b \cdot E;\ \rho \geq 0 \right\},$$

where $K := \left\{ (a,b) \in \mathbb{R} \times \mathbb{R}^n :\ a + \frac{|b|^2}{2} \leq 0 \right\}$.

Proof. Let (a^*, b^*) achieve the supremum. Then $\sup\{a\rho + b \cdot E\} = a^*\rho + b^* \cdot E$. So since $\rho \geq 0$, we want a^* to be as large as possible. This is possible using $a^* = -|b^*|^2/2$. Then

$$\sup_{(a,b)\in K} \{a\rho + b \cdot E\} = \sup_{b\in K} \left\{ -\frac{|b|^2}{2}\rho + b \cdot E \right\}.$$

Since the inside of the right hand side supremum is concave, then we can take the gradient with respect to b; set it equal to zero, and find the critical point: $b^* = E/\rho$. In particular, $a^* = -|E|^2/2\rho$ and thus $a^*\rho + b^* \cdot E = |E|^2/2\rho$. ◇

The saddle point problem is thus

$$\inf_{\rho,E} \sup_{\phi} \left\{ \int_{\mathbb{R}^n} \int_0^1 (a - \partial_t \phi) \cdot \rho + (b - \nabla_x \phi) \cdot E\ dxdt + G(\phi) \right\}. \quad (6.6)$$

Let $\mathbf{r} = \begin{pmatrix} \rho \\ E \end{pmatrix}$, $\xi = \begin{pmatrix} a \\ b \end{pmatrix}$, and $\nabla_{t,x}\phi = \begin{pmatrix} \partial_t \phi \\ \nabla_x \phi \end{pmatrix}$. Then the saddle point problem (6.6) can be written as follows

$$\inf_{\mathbf{r}} \sup_{\phi,\xi} \langle \xi - \nabla_{t,x}\phi, \mathbf{r} \rangle + G(\phi),$$

where $\langle \cdot, \cdot \rangle$ indicates $\int_{\mathbb{R}^n} \int_0^1 \cdots \; dtdx$. Let $\epsilon > 0$ be sufficiently small. The *augmented Lagrange problem* is

$$\inf_{\mathbf{r}} \sup_{\phi,\, \xi} \langle \xi - \nabla_{t,x}\phi, \mathbf{r} \rangle + G(\phi) - \frac{\epsilon}{2} \| \xi - \nabla_{t,x}\phi \|^2.$$

Optimize these three unknowns independently.

6.2 Fluid mechanics with interacting paths

We want to apply the same mathematical reasoning to the case with interacting paths or to the nonlinear functional (6.1). That is, we will prove a version of Proposition 6.1 that incorporates a nonlinear functional (5.1). Before we do so we collect important results from Santambrogio's book [19, Ch 5].

The first thing to note is that the square root of the minimum value of the Kantorovich's problem (1.4) defines the *Wasserstein distance* of measures on compact sets $X, Y \subset \mathbb{R}^n$:

$$W_2(\mu, \nu) := \left(\min \left\{ \int_{X \times Y} \|x - y\|^2 d\pi(x, y) : \; \pi \in \Pi(\mu, \nu) \right\} \right)^{1/2}, \; \mu \in \mathcal{P}(X), \; \nu \in \mathcal{P}(Y).$$

One may find in Ambrosio's and Gigli's guide [1], that, on a compact set D, this quantity defines a distance over $\mathcal{P}(D)$. Moreover, if we endow $\mathcal{P}(D)$ with the distance $W_2(\mu, \nu)$ we get the *Wasserstein space* $\mathbb{W}_2(D) := (\mathcal{P}(D), W_2(\mu, \nu))$.

Theorem 6.5 (Theorem 5.14 [19]) *Let $(\rho_t)_{t\in[0,1]}$ be an absolutely continuous curve in $\mathbb{W}_2(D)$ and $D \subset \mathbb{R}^n$ compact. Then for a.e. $t \in [0,1]$, there exists a vector field $v_t \in L^2(\mathbb{R}^n)$ such that*

i. the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$ is satisfied in the weak sense,

ii. for a.e. t, we have $\int_D \rho_t |v_t|^2 dx \leq |\dot{\rho}|(t)$ (where $|\dot{\rho}|(t)$ denotes the metric derivative at time t of the curve $t \mapsto \rho_t$ w.r.t the distance W_2);

Conversely, if $(\rho_t)_{t\in[0,1]}$ is a family of measures in $\mathcal{P}(D)$ and for each t we have a vector field $v_t \in L^2(\mathbb{R}^n)$ with $\int \rho_t |v_t|^2 dx < +\infty$ solving the continuity equation (6.3), then $(\rho_t)_{t\in[0,1]}$ is absolutely continuous in $\mathbb{W}_2(D)$, and for a.e. t, we have $|\dot{\rho}|(t) \leq \int \rho_t |v_t|^2 \, dx$.

Remark 6.6 *Observe that the above theorem holds true for $L^p(\mathbb{R}^n)$ and $\mathbb{W}_p$ and $\mathcal{P}_p$ for $p > 1$.*

Proposition 6.7 (Proposition 5.31 [19]) *For every Lipschitz curve $(\rho_t)_t$ in $\mathbb{W}_2(D)$, there exists a measure $\pi \in \mathcal{P}(\Omega)$ such that $\rho_t = (e_t)_\sharp \pi$ and $\int_\Omega c(\omega) d\pi(\omega) \leq \int_0^1 |\dot{\rho}|(t)^2 \, dt$.*

Both the proofs of Theorem 6.5 and Proposition 6.7 can be found in Santambrogio's book [19].

Now we motivate the new problem. Define

$$c(\gamma) := \int_0^1 \tfrac{1}{2} \|\dot{\gamma}(t)\|^2 \, dt, \quad \text{and} \quad \mathcal{A}(\pi) := \int_\Omega c(\gamma) \, d\pi(\gamma) + \int_\Omega \int_\Omega K(\gamma, \sigma) \, d\pi(\sigma) d\pi(\gamma),$$

where

$$K(\gamma, \sigma) = \int_0^1 \kappa(\gamma(t) - \sigma(t)) \, dt$$

Stipulate further that $\kappa : \mathbb{R}^n \to \mathbb{R}$ is a function of positive type. Suppose that π is given by a map $\Gamma(x,t)$, and let $\rho_t = (e_t)_\# \pi$. Then, applying Fubini and using change of variables (1.2)

$$\begin{aligned}\int_\Omega \int_0^1 \tfrac{1}{2}\|\dot\gamma(t)\|^2 \, dt \, d\pi(\gamma) &= \int_0^1 \int_\Omega \tfrac{1}{2}\|\dot\gamma(t)\|^2 \, d\pi(\gamma) dt \\ &= \int_0^1 \int_\Omega \tfrac{1}{2}\|v(\gamma(x,t),t)\|^2 \, d\pi(\gamma) dt \\ &= \int_0^1 \int_{\mathbb{R}^d} \tfrac{1}{2}\|v(x,t)\|^2 \rho_t(x) \, dxdt.\end{aligned}$$

Here $v(x,t)$ is a vector field defined via the formula

$$\dot\gamma(t) = v(\gamma(t),t) \text{ for } \pi\text{-a.e.}\gamma.$$

We have, then, likewise,

$$\int_\Omega \int_\Omega K(\gamma,\sigma) \, d\pi(\sigma) d\pi(\gamma) = \int_0^1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \kappa(x-y)\rho_t(x)\rho_t(y) \, dxdydt. \tag{6.7}$$

This shows that

$$\mathcal{A}(\pi) = \int_0^1 \int_{\mathbb{R}^d} \left(\tfrac{1}{2}\|v\|^2 + \int_{\mathbb{R}^d} \kappa(x-y)\rho_t(y) \, dy \right) \rho_t(x) \, dxdt,$$

where v and ρ are such that (6.3) hold. This suggests that

$$\min_{\pi \in \Pi_{\text{path}}(\mu_0,\mu_1)} \mathcal{A}(\pi) = \inf_{v,\rho} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \left| v(x,t) \right|^2 \rho(x,t) dxdt + \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y)\rho(x,t)\rho(y,t) dxdydt.$$

A rigorous proof of this identity is the main result of this chapter (Theorem 6.9). Now define

$$\begin{aligned}\mathcal{Q}(\rho) &= \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y)\rho(x,t)\rho(y,t) \, dxdydt \\ &= \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y) d\rho(x) d\rho(y) \, dt.\end{aligned}$$

Let

$$\mathcal{B}(\rho,E) = \int_0^1 \int_{\mathbb{R}^d} \frac{|E(x,t)|^2}{2\rho(x,t)} \, dxdt + \mathcal{Q}(\rho).$$

Recall that π in $\Pi_{\text{path}}(\mu_0, \mu_1)$ satisfies the following constraints,

$$(e_0)_{\#}\pi = \mu_0,\ (e_1)_{\#}\pi = \mu_1, \tag{6.8}$$

while the continuity equation conserves mass with initial and final conditions,

$$\partial_t \rho + \nabla(E) = 0 \tag{6.9}$$

$$\rho(x,0) = \mu_0,\ \rho(x,1) = \mu_1 \tag{6.10}$$

Using Benamou's, Carlier's, and Santambrogio's paper on *Variational Mean Field Games* [3], and Santambrogio's *Lecture notes on variational mean field games* [20] provides a proof, combining both Theorem 6.5 and Proposition 6.7, of the next proposition.

Proposition 6.8 (Benamou et al Proposition 3.1 [3]) *Let $D \subset \mathbb{R}^n$ be a bounded domain. Suppose (ρ, v) satisfies the continuity equation (6.3) in the weak sense and $\int_0^1 \int_D \rho|v|^2 dxdt < +\infty$. Then there exists a representative of ρ such that $t \mapsto \rho_t \in \mathbb{W}_2(D)$ is absolutely continuous and $|\dot{\rho}|(t) \leq \int_D \rho|v|^2 dx$ a.e. Moreover, there exists a probability measure $\pi \in \mathcal{P}(\Omega)$ such that $\rho_t = (e_t)_\sharp \pi$ and*

$$\int_\Omega c(\gamma) d\pi(\gamma) \leq \frac{1}{2}\int_0^1 \int_D \rho|v|^2 dx\ dt.$$

Conversely, if $\rho_t = (e_t)_\sharp \pi$ for a probability measure $\pi \in \mathcal{P}(\Omega)$ with $\int_\Omega c(\gamma) d\pi(\pi) < +\infty$, then $t \mapsto \rho_t \in \mathbb{W}_2(D)$ is absolutely continuous and there exists a time-dependent family of vector fields $v_t \in L^2(D)$ such that (6.3) holds in the weak sense and

$$\frac{1}{2}\int_0^1 \int_D \rho|v|^2 dx\ dt \leq \int_\Omega c(\gamma)\ d\pi(\gamma).$$

Consequently, $\frac{1}{2}\int_0^1 \int_D \rho|v|^2 dx\ dt = \int_\Omega c(\gamma)\ d\pi(\gamma)$. Applying this proposition to the interaction term (6.7), then this proposition allows one to rewrite the minimum of $\mathcal{A}(\pi)$ subject to (6.8) as the minimum of $\mathcal{B}(\rho, E)$ subject to the constraints (6.9) and (6.10). Namely, the following result.

Theorem 6.9 *The minimum*

$$\min\{\ \mathcal{B}(\rho, E) : \rho, E \text{ for which (6.9) and (6.10) hold }\}$$

has the same value as

$$\min\{\ \mathcal{A}(\pi) : \pi \text{ for which (6.8) holds }\}.$$

Using the variational techniques in convex duality from Benamou et al [3] and Santambrogio [20], we managed to produce duality for the functional that incorporates interaction such as the functional (6.1).

Theorem 6.10 *Let* $H(\rho) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \kappa(x-y)d\rho(x)d\rho(y)$*, where* κ *is convex, and* $H^*(q) = \sup_\rho \langle q, \rho\rangle - H(\rho)$*, where* $q := \frac{1}{2}|\nabla\phi|^2 - \partial_t\phi$*. Then*

$$\min_{\rho,v}\{\mathcal{B}(\rho, v) : (6.3) \textit{ and } (6.10) \textit{ hold }\} = \sup_{\phi}\left\{\left(G(\phi) - \int_0^1 H^*(q)dt\right) : \phi \in C^1(\mathbb{R}^n \times [0,1])\right\}.$$

Proof. We apply the same reasoning that was done in [20] and [3]. Consider the functional

$$\mathcal{B}(\rho, v) = \int_0^1 \int \frac{1}{2}\rho(x,t)|v(x,t)|^2 dx\ dt + \int_0^1 \int\int \kappa(x-y)d\rho(x)d\rho(y)\ dt.$$

Applying the *toy problem* from Section 3.2 and writing a Lagrange equation with multiplier ϕ and using integration by parts, just like how we did in the end of the previous section, we have

$$\min_{\rho,v}\left\{\mathcal{B}(\rho, v) + \sup_{\phi}\left\{\int_0^1 (\rho\partial_t\phi + (\rho v)\nabla\phi)\, dxdt + G(\phi)\right\}\right\},$$

where $G(\phi) = \int(\phi(x,1)\mu_1 - \phi(x,0)\mu_0)dx$; and we may write this since the sup in ϕ takes a value 0 if the constraint is satisfied, else $+\infty$. Using a minmax principle (without justification) allows us to swap the inf and sup in the above to rewrite it as

$$\sup_{\phi}\left\{G(\phi)+\inf_{\rho,v}\int_0^1\int\left(\frac{1}{2}\rho|v|^2+\rho\partial\phi+\rho v\nabla\phi\right)dxdt+\int_0^1\int\int\kappa(x-y)d\rho(x)d\rho(y)dt\right\}.$$

Then minimizing inside the integrals with respect to v, to get $v=-\nabla\phi$. Now, we focus on the integrand of the infimum, and subsititute v for $-\nabla\phi$ to get

$$\begin{aligned}
&\inf_{\rho}\left\{-\frac{1}{2}\rho|\nabla\phi|^2+\rho\partial_t\phi+\int\int\kappa(x-y)d\rho(x)d\rho(y)\right\}\\
&\quad=\inf_{\rho}\left\{-\rho\left(\frac{1}{2}|\nabla\phi|^2-\partial_t\phi\right)+H(\rho)\right\}\\
&\quad=\inf_{\rho}\{\langle-\rho,q\rangle+H(\rho)\}=-\sup_{\rho}(\langle q,\rho\rangle-H(\rho))=-H^*(q),
\end{aligned}$$

where $q=\frac{1}{2}|\nabla\phi|^2-\partial_t\phi$ and $\langle\cdot,\cdot\rangle$ indicates $\int_{\mathbb{R}^n}\cdots dx$. The fact that the convex Fenchel-Rockafellar H^* holds is due to Bochner's theorem given in Reed's and Simon's book in [17, Theorem IX.9]. Namely, since $\kappa(x-y)$ is convex in $\mathbb{R}^n$, $H(\rho)$ is convex as well. In other words, if $H(\rho)=\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\kappa(x-y)\;dxdy$ where κ is the Fourier transform of a finite, positive measure on $\mathbb{R}^n$ then H is convex. (In appendix A, you will find Bochner's theorem: as any convex combinations or scalar multiples of functions of positive type form a cone of functions of positive type, the set of Fourier sets of finite positive measures on $\mathbb{R}^n$ is the cone of functions of positive type.) Returning to the business at hand, the minimum of $\mathcal{B}(\rho,v)$ is equal to $\sup_{\phi}\left\{G(\phi)-\int_0^1 H^*(q)dt\right\}$. ◇

APPENDIX A

BOCHNER'S THEOREM

In this appendix we gather a useful kit of results stemming from Reed's and Simon's book [17] to complement the Introduction 1 and Section 6.2. Most notably go over Bochner's result.

Theorem A.1 (Bochner's theorem) *[17, Theorem IX.9] The set of Fourier transforms of the finite, positive measures on $\mathbb{R}^n$ is exactly the cone of functions of positive type.*

The proof of this result needs some preliminary work. Let us recall the definition of the Fourier transform on the Schwartz spaces $\mathcal{S}(\mathbb{R}^n)$ and $\mathcal{S}'(\mathbb{R}^n)$ as presented in [17].

Definition A.2 *Suppose $f \in \mathcal{S}(\mathbb{R}^n)$. The* Fourier transform *of f is denoted by $\hat{f}$ and defined by*

$$\hat{f}(\lambda) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\lambda \cdot x} f(x)dx,$$

where $\lambda \cdot x = \sum_{i=1}^{n} \lambda_i x_i$. The inverse Fourier transform *of f is denoted $\check{f}$ and defined by*

$$\check{f}(\lambda) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{i\lambda \cdot x} f(x)dx.$$

We write $\hat{f} = \mathfrak{F}f$, for the Fourier transform of f.

The Fourier transform operators enjoy relevant properties such as continuity, boundedness, linear bicontinuous bijection from $\mathcal{S}(\mathbb{R}^n)$ to itself, and nice convolution properties. Please refer to the text by Reed and Simon [17] to learn more.

The most important property for us of the Fourier transform is associated to the finite positive measures on $\mathbb{R}^n$. Indeed, the Fourier transforms of the finite positive measures on $\mathbb{R}^n$ will be defined as follows.

Suppoe we have

$$\hat{\mu}(\lambda) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-i\lambda\cdot x}\, d\mu(x).$$

Then if $\varphi \in \mathcal{S}(\mathbb{R}^n)$, we have

$$\begin{aligned}\int_{\mathbb{R}^n} \hat{\mu}(\lambda)\varphi(\lambda)\, d\lambda &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-i\lambda\cdot x} d\mu(x) \right) \varphi(\lambda) d\lambda \\ \text{(Fubini)} &= \int_{\mathbb{R}^n} \left((2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot\lambda} \varphi(\lambda) d\lambda \right) d\mu(x) \\ &= \int_{\mathbb{R}^n} \hat{\varphi}(x) d\mu(x).\end{aligned}$$

This then shows that the above definition coincides with the Fourier transform of finite positive measures. More concretely, if $\lambda_1, \ldots, \lambda_n \in \mathbb{R}^n$ and $\xi = \langle \xi_1, \xi_2, \ldots, \xi_N \rangle \in \mathbb{C}^N$. Then

$$\begin{aligned}\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\mu}(\lambda_i - \lambda_j)\overline{\xi}_j \xi_i &= \sum_{i,j}^{N} (2\pi)^{-n/2} \int e^{-i\lambda_i\cdot x}\overline{e^{-i\lambda_j\cdot x}}\; \overline{\xi}_j \xi_i d\mu(x) \\ &= \int \left| \sum_{i=1}^{N} \xi_i e^{-i\lambda_i\cdot x} \right|^2 d\mu(x) \geq 0.\end{aligned}$$

Hence the function $\hat{\mu}(\lambda)$ has the property that for any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}^n$, $\{\hat{\mu}(\lambda_i - \lambda_j)\}_{i,j}$ is a positive operator on $\mathbb{C}^N$. That $\hat{\mu}$ is bounded follows from the following calculation:

$$|\hat{\mu}(\lambda)| \leq (2\pi)^{-n/2} \int_{\mathbb{R}^n} \left|e^{-i\lambda\cdot x}\right| d\mu(x)$$
$$= (2\pi)^{-n/2}\mu(\mathbb{R}^n).$$

Now is an appropriate time to state the definition of a function of positive type.

Definition A.3 *A complex-valued, bounded, continuous function f on $\mathbb{R}^n$ that has the property that $\{f(\lambda_i - \lambda_j)\}_{i,j}$ is a positive matrix on $\mathbb{C}^N$ for each N and all $\lambda_1, \ldots, \lambda_n \in \mathbb{R}^n$ is called a* function of positive type*.*

Three important properties follow. The first, if $N = 1$, $x \in \mathbb{R}^n$, then $f(0) \geq 0$, as $f(0)$ is a positive operator on $\mathbb{C}^1$. For $N = 2$ and $\lambda_1 = x$ and $\lambda_2 = 0$, we acquire the 2×2 matrix

$$\begin{pmatrix} f(0) & f(x) \\ f(-x) & f(0) \end{pmatrix}.$$

By the Definition A.3, this matrix has to be positive and so self-adjoint with positive determinant. This gives the second and third property, respectively: $f(x) = -f(-x)$ and $|f(x)| \leq f(0)$.

One can see, from the above, that any linear combination or scalar multiples of functions of positive type gives functions of positive type, so these functions make up a cone, and this statement amounts to Bochner's theorem, Theorem A.1. The proof of this relies heavily on a generalization of Stone's theorem. In order to state it, it requires a bit of mathematical machinery that, of course, can be found in [17]. For the moment, all one really needs to know to state it, are the following. Let P_Ψ be an operator defined by the characteristic function $\chi_\Psi(A)$ of the measurable set $\Psi \subset \mathbb{R}$, A an operator (not necessarily bounded). Then the family of operators $\{P_\Psi\}$ satisfy the properties:

(a) Each P_Ψ is an orthogonal projection,

(b) $P_\emptyset = 0$, $P_{(-\infty,\infty)} = \mathrm{Id}$,

(c) $P_{\Psi_1} P_{\Psi_2} = P_{\psi_1 \cap \Psi_2}$.

Given φ in a Hilbert space $\mathcal{H}$, $(\varphi, P_\Psi \varphi)$ is a well-defined Borel measure on $\mathbb{R}$ denoted by $d(\varphi, P_\lambda \varphi)$. If g is a bounded Borel function we can define $g(A)$ by

$$(\varphi, g(A)\varphi) = \int_{-\infty}^{\infty} g(\lambda) d(\varphi, P_\lambda \varphi).$$

For g an unbounded complex-valued function, define

$$D_g := \left\{ \varphi : \int_{-\infty}^{\infty} |g(\lambda)|^2 d(\varphi, P_{\lambda\varepsilon}) < \infty \right\}.$$

Then, assuming this is dense in $\mathcal{H}$ which it is and may be found in [17], an operator $g(A)$ is defined on D_g by

$$(\varphi, g(A)\varphi) = \int_{-\infty}^{\infty} g(\lambda) d(\varphi, P_\lambda \varphi)$$

In particular, for φ, ψ in some operator, $(\varphi, A\psi) = \int_{-\infty}^{\infty} \lambda d(\varphi, P_\lambda \psi)$.

Another important definition one needs to know before stating Stone's theorem is that of a unitary operator $U(t) = e^{it\,A}$, for a self-adjoint operator A. If for all $s, t \in \mathbb{R}$, $U(t)$ is unitay, $U(s+t) = U(s)U(t)$ and $U(0) = \mathrm{Id}$, and if $\varphi \in \mathcal{H}$ and $t \to t_0$, then $U(t)\varphi \to U(t_0)\varphi$, then U is said to be *strongly continuous*. In essence, Stone's theorem states that every strongly continuous unitary group arises as the exponential of a self-adjoint operator. The generalized version is:

Theorem A.4 (Stone's theorem) *[17, Theorem VIII.12] Let* $\boldsymbol{t} \mapsto U(\boldsymbol{t}) := U(t_1, \ldots, t_n)$ *be a strongly continuous map of* $\mathbb{R}^n$ *into the unitary operators on a separable Hilbert space* $\mathcal{H}$ *satisfying* $U(\boldsymbol{t} + \boldsymbol{s}) = U(\boldsymbol{t})U(\boldsymbol{s})$ *and* $U(0) = Id$. *Let* D *be the set of finite linear combinations of vectors of the form*

$$\varphi_f = \int_{\mathbb{R}^n} f(\boldsymbol{t}) U(\boldsymbol{t}) \varphi \, dt \quad \varphi \in \mathcal{H}, \quad f \in C_0^\infty(\mathbb{R}^n).$$

Then D is domain of essential self-adjointness for each of the generators A_j of the one-parameter subgroups $U(0, \ldots, t_j, \ldots, 0)$, each $A_j : D \to D$ and the A_j commute for $j = 1, \ldots, n$. Furthermore, there is a projection-valued measure P_Ψ on $\mathbb{R}^n$ so that

$$(\varphi, P_\lambda \psi) = \int_{\mathbb{R}^n} e^{i\lambda \cdot \boldsymbol{t}} d(\varphi, P_\lambda \psi)$$

for all $\varphi, \psi \in \mathcal{H}$.

Proof. The proof can be found in *Methods of modern mathematical physics: Functional analysis Vol I* in [17, Ch VIII]. $\diamond$

With this mathematical machinery we can finally prove Theorem A.1. **Proof.** [**Proof of Theorem A.1**] The proof is a reprise of Reed's and Simon's proof in [17, Ch IX]. Remember that we already showed, when we defined the finite, positive measures on $\mathbb{R}^n$, that the Fourier transforms of a finite positive measure are functions of positive type. All that remains is to show the converse: that a function of positive type is given by the Fourier transform of a finite positive measure. So suppose f is of positive type. Let $\mathcal{K}$ be the set of complex-valued functions on $\mathbb{R}^n$ which vanish except on a set of measure zero. Then

$$(\psi, \varphi)_f = \sum_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) \overline{\psi}(\mathbf{x}) \varphi(\mathbf{y})$$

has all the properties of a well-defined inner product, but except that we may have $(\psi, \varphi)_f = 0$ for some $\varphi \neq 0$. To remedy this, let $\mathcal{N}$ be the set of such φ's. Moding out by this set, $\mathcal{K}/\mathcal{N}$ defines a well-defined pre-Hilbert space under $(\cdot, \cdot)_f$. For $\mathbf{t} \in \mathbb{R}^n$, define $U_\mathbf{t}$ on $\mathcal{K}$ by $(U_\mathbf{t}\varphi)(\mathbf{x}) = \varphi(\mathbf{x} - \mathbf{t})$. Now, since $U_\mathbf{t}$ preserves $(\cdot, \cdot)_f$, equivalence classes are also preserved and it restricts to an isometry on $\mathcal{K}/\mathcal{N}$. We can show the same for $U_{-\mathbf{t}}$ and may extend to a unitary operator $\widetilde{U}_\mathbf{t}$ on $\mathcal{H} := \overline{\mathcal{K}/\mathcal{N}}$.

Since $\widetilde{U}_{\mathbf{t+s}} = \widetilde{U}_{\mathbf{t}}\widetilde{U}_{\mathbf{s}}$ and $\widetilde{U}_0 = \mathrm{Id}$, and as f is continuous, $\widetilde{U}$ is strongly continuous, and thus the mapping $\mathbf{t} \mapsto \widetilde{U}(\mathbf{t})$ satisfies the hypothesis of Theorem A.1. Applying such result, we can find a projection-valued measure P_λ on $\mathbb{R}^n$ such that

$$(\varphi, \widetilde{U}_{\mathbf{t}}\psi)_f = \int_{\mathbb{R}^n} e^{i\mathbf{t}\cdot\lambda} d(\varphi, P_\lambda \psi).$$

Let $\widetilde{\varphi}_0$ be the class of the characteristic function $\varphi_0(\mathbf{x}) = 1$ if $x = 0$ and $\varphi_0(\mathbf{x}) = 0$ otherwise. Then

$$f(\mathbf{t}) = (\widetilde{U}_{\mathbf{t}}\widetilde{\varphi}_0, \widetilde{\varphi}_0)_f = (\widetilde{\varphi}_0, \widetilde{U}_{\mathbf{-t}}\widetilde{\varphi}_0)_f = \int_{\mathbb{R}^n} e^{-i\mathbf{t}\cdot\lambda} d(\widetilde{\varphi}_0, P_\lambda \widetilde{\varphi}_0).$$

And the above expression is that of the Fourier transform of a finite positive measure. $\diamond$

APPENDIX B

BOCHNER-SCHWARTZ THEOREM

The generalized version of Bochner's theorem A.1, due to Schwartz [21], which includes distributions is the following.

Theorem B.1 *[17, Theorem IX.10] A distribution $T \in \mathcal{D}'(\mathbb{R}^n)$ is a distribution of positive type if and only if $T \in \mathcal{S}'(\mathbb{R}^n)$ and T is the Fourier transform of a positive measure of at most polynomial growth.*

Some discussion and definitions are in order to fully understand this result. To generalize Theorem A.1 considering functions of positive type to distributions, we need the following. Suppose $f(x)$ is a bounded, continuous function. Then $f(x)$ will be of positive type if and only if

$$\int\int f(x-y)\overline{\varphi(y)}\varphi(x)\,dx\,dy \geq 0$$

for all $\varphi \in C_0^\infty(\mathbb{R}^n)$, see Reed's and Simon's book [17]. The condition above can be rewritten in the following way if we consider $(\overline{\tilde{\varphi}} * \varphi)(\tau) := \int \overline{\tilde{\varphi}(x-\tau)}\varphi(x)\,dx$, namely,

$$\int\int f(\tau)\overline{\varphi(x-\tau)}\varphi(x)\,d\tau\,dx = \int f(\tau)(\overline{\tilde{\varphi}} * \varphi)(\tau)\,d\tau \geq 0,$$

where $\tilde{\varphi}(x) = \varphi(-x)$. Then we have the definition:

Definition B.2 *A distribution $T \in \mathcal{D}'(\mathbb{R}^n)$ is said to be of* positive type *if $T(\overline{\tilde{\varphi}} * \varphi) \geq 0$ for all $\varphi \in \mathcal{D}(\mathbb{R}^n)$.*

A goal of this appendix is to demonstrate that the functional (1.8) with the Coulomb potential is convex, which boils down in showing that the quadratic functional $\int_\Omega \int_\Omega \int_0^1 |\gamma - \sigma|^{2-n}\, dt\, d\pi(\sigma)\, d\pi(\gamma)$ $(n \geq 3)$ is convex in π.

To show that the quadratic term with the Coulomb potential $\kappa(x-y) = |x-y|^{2-n}$ is convex in π, given

$$Q(\pi_s) := \int_\Omega \int_\Omega \int_0^1 |\gamma(t) - \sigma(t)|^{2-n} dt\; d\pi_s(\sigma)\; d\pi_s(\gamma)$$

it suffices to show, for π_0 and $\pi_1 \in \mathcal{P}(\Omega)$ with $\pi_s = (1-s)\pi_0 + s\pi_1$, that $\frac{d^2}{ds^2} Q(\pi_s) \geq 0$. Indeed,

$$\frac{d^2}{ds^2} Q(\pi_s) = 2 \int_0^1 \int_\Omega \int_\Omega |\gamma(t) - \sigma(t)|^{2-n} d(\pi_0(\sigma) - \pi_1(\sigma)) d(\pi_0(\gamma) - \pi_1(\gamma)) dt.$$

Let $\omega(\sigma) = \pi_0(\sigma) - \pi_1(\sigma)$ and $\omega(\gamma) = \pi_0(\gamma) - \pi_1(\gamma)$, for a signed measure ω. Let $d\omega_t = (e_t)_\sharp \omega$ Then this is equivalent to

$$\int_0^1 \int_\Omega \int_\Omega |\gamma(t) - \sigma(t)|^{2-n} d\omega(\sigma) d\omega(\gamma) dt = \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y)|^{2-n} d\omega_t(x) d\omega_t(y) dt.$$

Since the measures ω can be approximated under the weak-* topology by finite measures against $C_c^\infty(\mathbb{R}^n)$ test functions, given finite measures $\rho \in \mathcal{M}(\mathbb{R}^n)$ with $\rho = \varphi(x)dx$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$, the above integral can be approximated by

$$\int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^{2-n} d\rho(x) d\rho(y)\; dt = \int_0^1 \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^{2-n} \varphi(x)\varphi(y)\; dx\; dy\; dt.$$

As the above discussion will indicate that the Coulomb potential κ will be of positive type if and only if the above integral is ≥ 0, and Theorem B.1 applies to takes care of the rest.

Let us consider the fundamental solution of the Laplace equation for $n \geq 3$:

$$\Phi(x) = C_n |x|^{2-n}, \quad C_n \text{ a dimensional constant,}$$

given in Evans [8, Ch 2]. Considering convolutions, for $f \in C_c^2(\mathbb{R}^n)$, we can write

$$\begin{aligned} u(x) &= \int_{\mathbb{R}^n} \Phi(x-y) f(y)\, dy \\ &= C_n \int_{\mathbb{R}^n} |x-y|^{2-n} f(y)\, dy. \end{aligned}$$

In Evans' book [8, Ch 2] one may find $u \in C_c^2(\mathbb{R}^n)$ and solves the Poisson equation $-\Delta u = f$.

Armed with this knowledge we readily show what we set to prove for the Coulomb potential, that the quadratic term in (1.8) is convex in π, that is $\frac{d^2}{ds^2} Q(\pi_s) \geq 0$. More concretely,

$$\begin{aligned} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x-y|^{2-n} \varphi(x)\varphi(y)\, dx\, dy &= -\int_{\mathbb{R}^n} \Delta u_\varphi(x) \int_{\mathbb{R}^n} |x-y|^{2-n} \varphi(y)\, dy\, dx \\ &= -C_n \int \Delta u_\varphi(x) u_\varphi(x)\, dx \\ &= C_n \int |\nabla u|^2\, dx \geq 0, \end{aligned}$$

where we applied integration by parts to get the last equation.

APPENDIX C

THE DIFFERENTIABILITY OF THE END POINT COST FUNCTION

In this appendix we prove the following limit exists:

$$\lim_{h\to 0} \frac{c_{\mathbf{e}}(x, y + h\hat{e}) - c_{\mathbf{e}}(x, y)}{h}.$$

This stems from observing if $\gamma_{x,y}(t)$ is the minimal path from $x := \gamma_{x,y}(0)$ to $y := \gamma_{x,y}(1)$, $c_{\mathbf{e}}(x, y) = c(\gamma_{x,y}(t))$ and, if we momentarily suppose that the above limit exists, then said limit will equal the left-hand side of the following equation:

$$\nabla_y c_{\mathbf{e}}(x, y) = \frac{\partial c}{\partial y} \nabla_y \gamma_{x,y},$$

by the chain rule. Indeed,

$$\nabla_y \left(c_{\mathbf{e}}(x, y)\right) = \int_0^1 \dot{\gamma}_{x,y}(t) \cdot \nabla_y \left(\dot{\gamma}_{x,y}(t)\right) \nabla_y \gamma_{x,y}(t) - \nabla V(\gamma_{x,y}(t), t) \nabla_y \gamma_{x,y}(t)\, dt,$$

and since

$$\nabla_y \left(\dot{\gamma}_{x,y}(t)\right) = \nabla_y \left(\frac{d}{dt}\gamma_{x,y}(t)\right) = \frac{d}{dt}\, \nabla_y \left(\gamma_{x,y}(t)\right),$$

to show the above limit exists it suffices to show the following limit exists:

$$\lim_{h\to 0} \frac{\gamma_{x,y+h\,\hat{e}}(t) - \gamma_{x,y}(t)}{h}.$$

As an example of the above phenomenon of commutative of derivatives, consider the straight line minimal path in $\mathbb{R}^n$ given by $\gamma_{x,y}(t) := (1-t)x + ty$. Then:

1) $\nabla_y \left(\frac{d}{dt}\gamma_{x,y}(t)\right) = 1$

2) $\frac{d}{dt}\left(\nabla_y \gamma_{x,y}(t)\right) = 1.$

We get the same conclusion for ∇_x :

1) $\nabla_x \left(\frac{d}{dt}\gamma_{x,y}(t)\right) = -1$

2) $\frac{d}{dt}\left(\nabla_x \gamma_{x,y}(t)\right) = -1.$

Lemma C.1 *The end point cost function $c_{\boldsymbol{e}}(x,y) := \inf_{\gamma(0)=x,\, \gamma(1)=y} c(\gamma)$ is differentiable with respect to y. Moreover,*

$$\nabla_y c_{\boldsymbol{e}}(x,y) = \lim_{h\to 0} \frac{c_{\boldsymbol{e}}(x, y+h\hat{e}) - c_{\boldsymbol{e}}(x,y)}{h}.$$

Proof. For all $t \in [0,1]$, let $\theta_h : [0,1] \to X$ be defined by

$$\theta_h(t) := \frac{\gamma_{x,y+h\,\hat{e}}(t) - \gamma_{x,y}(t)}{h}, \quad (h \neq 0).$$

From the aforementioned, to prove the lemma it suffices to prove $\lim_{h\to 0}\theta_h(t)$ exists. This will be accomplished by applying a "standard"[1] argument of compactness plus uniqueness of linear ODEs.

Thanks to Proposition 2.6, $\gamma_{x,y+h\hat{e}}(t)$ solves

$$\ddot{\gamma}_{x,y+h\hat{e}}(t) = -\nabla V(\gamma_{x,y+h\hat{e}}(t), t).$$

Then $\theta_h(t)$ solves the boundary value problem:

$$\ddot{\theta}_h(t) = -\frac{1}{h}\left(\nabla V(\gamma_{x,y+h\hat{e}}(t), t) - \nabla V(\gamma_{x,y}(t), t)\right),$$

$$\theta_h(0) = 0,$$

$$\theta_h(1) = \hat{e}.$$

[1]Well known method for establishing existence of limits in the *fully-nonlinear PDE community.*

Assuming $V \in C^2(X)$, Taylor expanding about h, we get

$$\begin{aligned} \ddot{\theta}_h(t) &= -D^2V(\gamma_{x,y}(t),t)\theta_h(t) + o_h(1)(t) \\ \theta_h(0) &= 0 \\ \theta_h(1) &= \hat{e}. \end{aligned} \tag{C.1}$$

Proposition 2.8 tells us that $\theta_h(t)$ is equibounded:

$$\|\theta_h(t)\|_\infty \leq \frac{1}{1-L} =: C_0 \quad \forall\, h, t.$$

Since ∇V is L-Lipschitz, and another application of Proposition 2.8, we have

$$\left\|\ddot{\theta}_h(t)\right\|_\infty \leq \frac{L}{1-L} =: C_1 \quad \forall\, h, t,$$

and so $\ddot{\theta}_h$ is equibounded as well. This implies $\theta_h(t)$ is equicontinuous. For the reader's benefit we show how this follows. More concretely, for any interval $[p,q]$ such that $0 \leq p < q \leq 1$ with $q - p = 1$, the Mean Value Theorem applies to show there is some $\tau \in [p,q]$ such that

$$\theta_h(q) - \theta_h(p) = \dot{\theta}_h(\tau).$$

Then $\left\|\dot{\theta}_h(\tau)\right\|_\infty \leq 2\ C_0$. Another application of the Mean Value Theorem shows there is some $\xi \in [r,s]$ such that

$$\frac{\dot{\theta}_h(s) - \dot{\theta}_h(r)}{s-r} = \ddot{\theta}_h(\xi) \quad (0 \leq p \leq r < s \leq q \leq 1).$$

Using this, pick any $\eta \in [p,q]$ so that

$$\left\|\dot{\theta}_h(\eta) - \dot{\theta}_h(\tau)\right\| \leq C_1,$$

while the reverse triangle inequality gives us

$$C_1 \geq \left\|\dot{\theta}_h(\eta) - \dot{\theta}_h(\tau)\right\| \geq \left\|\dot{\theta}_h(\eta)\right\| - \left\|\dot{\theta}_h(\tau)\right\|$$

which implies

$$\left\|\dot{\theta}_h(\eta)\right\| \leq 2\,C_0 + C_1 := C_2 \quad \forall\,\eta \in [p,q].$$

For any $0 \leq p \leq r < x < y < s \leq q \leq 1$, once again the Mean Value Theorem applies to show

$$\|\theta_h(y) - \theta_h(x)\| \leq C_2|y-x|,$$

and hence θ_h is equicontinuous. We are ready to prove the existence of the limit in γ. For any sequence h_k, such that $h_k \to 0$ as $k \to \infty$, the family of functions $\{\theta_{h_k}(t)\}$ from $[0,1]$ to X are both equibounded and equicontinuous. Arzelá-Ascoli applies to show that the sequence $\{\theta_{h_k}\}$ admits a subsequence $\{\theta_{\widetilde{h}_k}\}$ uniformly converging, as $k \to \infty$, to a continuous function $\theta : [0,1] \to X$. From here, using (C.1), one can show θ is $C^2([0,1])$ and solves uniquely

$$\ddot{\theta}(t) = -D^2V(\gamma_{x,y}(t),t)\theta(t)$$

$$\theta(0) = 0,$$

$$\dot{\theta}(0) = \hat{e}.$$

◇

BIBLIOGRAPHY

[1] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[3] Jean-David Benamou, Guillaume Carlier, and Filippo Santambrogio. Variational mean field games. In *Active Particles, Volume 1*, pages 141–171. Springer, 2017.

[4] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. *Optimal transportation networks: models and theory.* Springer, 2008.

[5] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[6] Guillaume Carlier, Chloé Jimenez, and Filippo Santambrogio. Optimal transportation with traffic congestion and wardrop equilibria. *SIAM Journal on Control and Optimization*, 47(3):1330–1350, 2008.

[7] José A Carrillo, Robert J McCann, and Cédric Villani. Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179(2):217–263, 2006.

[8] Lawrence C Evans. Partial differential equations. *Graduate studies in mathematics*, 19(2), 1998.

[9] Alessio Figalli, Young-Heon Kim, and Robert J McCann. Hölder continuity and injectivity of optimal maps. *Archive for Rational Mechanics and Analysis*, 209(3):747–795, 2013.

[10] Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[11] Nestor Guillen and Jun Kitagawa. On the local geometry of maps with c-convex potentials. *Calculus of Variations and Partial Differential Equations*, 52(1):345–387, 2015.

[12] Ryan Hynd. Sticky particles and the pressureless euler equations in one spatial dimension. *arXiv preprint arXiv:1804.09340*, 2018.

[13] Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management science*, 6(4):366–422, 1960.

[14] Leonid V Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4):1381–1382, 2006.

[15] Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Archive for rational mechanics and analysis*, 177(2):151–183, 2005.

[16] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[17] Michael Reed, Barry Simon, Barry Simon, and Barry Simon. *Methods of modern mathematical physics*, volume 1. Elsevier, 1972.

[18] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[19] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[20] Filippo Santambrogio. Lecture notes on variational mean field games. In *Mean Field Games*, pages 159–201. Springer, 2020.

[21] Laurent Schwartz. Théorie des distributions à valeurs vectorielles. i. In *Annales de l'institut Fourier*, volume 7, pages 1–141, 1957.

[22] Justin Solomon. *Numerical algorithms: methods for computer vision, machine learning, and graphics*. CRC press, 2015.

[23] Michael Spivak. *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. CRC press, 2018.

[24] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[25] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.