AN OPTIMAL TRANSPORTATION PRINCIPLE FOR INTERACTING PATHS AND CONGESTION

RENE CABRERA

ABSTRACT. 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.

Contents

1. Introduction 2
1.1. Related work 3
1.2. MKP 3
1.3. Summary of main results 4
1.4. Outline of the paper 7
2. Minimal paths 7
3. Minimizers and their properties 12
3.1. Existence of minimizers 12
3.2. The dual problem and the endpoint cost function 14
3.3. Potentials and cyclically monotone sets 16
3.4. Optimal plans given by maps 22
4. The optimal path problem with interaction 24
4.1. Existence of minimizers for the optimal path with interaction 24
4.2. The dual problem 28
4.3. Supergradient of a concave function 31
4.4. The effective cost 33
5. Acknowledgments 36
Appendix A. Bochner’s Theorem 37
Appendix B. The differentiability of the end point cost function 39
References 41
1. Introduction

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 [21]. 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 [11]. In the early 90’s, Brenier’s work [4] 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 [21, 22] and Santambrogio [17]. 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 [21]. 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 [21, Ch 5].

In this paper, we introduce a new variant of the MKP that incorporates interacting paths and congestion. We prove solutions (using the Condition 1) of:

$$\inf_{\pi \in \Pi_{path}(\mu_0, \mu_1)} E_0(\pi) + \int_{\Omega} U(\gamma, \pi) d\pi(\gamma)$$

exist. The functional $E_0(\pi)$ is the total cost of transporting along all paths in the plan, and this is reminiscent of the MKP but using an explicit path dependence approach. The interacting (latter) term in $\int_{\Omega} U(\gamma, \pi) d\pi(\gamma)$ measures the total congestion between paths $\gamma$ and a distribution of paths given by $\pi$. The optimal transport solution, $\pi$, to (1) is unique (Theorem 1.10) and is given by a map, $\Gamma$. We also formulate a duality (Theorem 4.4) to characterize solutions of (1). Defining an effective cost (35) with interaction, we prove an extension and a general version of Theorem 1.5; namely, Theorem 1.10 which says that the optimal plan is given by a general optimal map of paths2 characterized by a general result of Brenier, Gangbo and McCann [22, 4, 8]. The central lemmas of the paper that make this possible comes from $c_\infty$-cyclical monotonicity (Definition 3.9) of optimal transport that incorporates path dependence and interaction effects. Namely Lemmas 3.11 and 3.12.

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 $\gamma$ such that $\gamma(0) = x$ indicates the initial point along $\gamma$ and $\gamma(1) = y$ the arrival point along $\gamma$, 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 Section 2). More concretely, if $\Omega$ denotes the set of such continuous rectifiable paths, then the total cost of transporting along all paths in the plan is the functional:

$$E_0(\pi) := \int_{\Omega} c(\gamma) d\pi(\gamma).$$

This can be thought of as an explicit path dependent version of the MKP. Monge set some precedent to this line of reasoning [14]. Villani [22] explored this line of reasoning by describing
a time-dependent version of optimal transport. In this manuscript we explore this further and show it reduces to the traditional optimal transport problem.

### 1.1. Related work.
Motivated by the modeling of traffic networks, Carlier, Jimenez, and Santambrogio [5] studied a related transport problem involving paths, as well. The way [5] models congestion effects is different from the present paper: in [5] the objective functional considers intensity through paths while here the functional involves an interaction potential. The 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 [10]. 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].

### 1.2. MKP.
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 \( f(x) \) and \( g(y) \). Then if we have two probability measures \( \mu \) on \( X \subset \mathbb{R}^n \) and \( \nu \) 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 \( \mu \) forward to \( \nu \). This is denoted by \( T_\sharp \mu = \nu \) and it means that for any measurable (Borel) subset \( B \subset Y \), \( \nu(B) := T_\sharp \mu(B) = \mu(T^{-1}(B)) \) 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 \( h \in C^0(Y) \),

\[
\int_Y h(y) \, d\nu(y) = \int_X h(T(x)) \, d\mu(x)
\]

is another characterization.

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

\[
\int_X c(x, T(x)) \, d\mu(x) = \int_X |x - T(x)|^2 \, d\mu(x)
\]

among all \( T \) pushing forward \( \mu \) to \( \nu \) [1].

In 1940 Leonid Kantorovich [12] in some sense “relaxed” Monge’s problem [1] and [21] by introducing a linear program formulation for the problem. Concretely, again for the quadratic cost function, Kantorovich considered probability measures \( \pi \) on \( \mathbb{R}^n \times \mathbb{R}^n \) with left and right marginals \( \mu, \nu \), respectively; namely, \( \mu[A] = \pi[A \times \mathbb{R}^n] \) and \( \nu[\mathbb{R}^n \times B] = \pi[\mathbb{R}^n \times B] \) for any (Borel) measurable subsets \( A, B \) of \( \mathbb{R}^n \). An equivalent criterion for \( \pi \) to have left and right marginals \( \mu, \nu \) is the following linearity of \( \pi \) [21]:

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

**Kantorovich’s problem** is: Minimize

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

among all \( \pi \) having marginals \( \mu, \nu \). 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. So this problem is actually a linear minimization problem with convex constraints. Whenever \( \pi \) satisfies
the marginal condition we say \( \pi \) is admissible. A basic result in functional analysis applying continuity and compactness arguments is the existence of minimizers of functionals [1], [7], and [21].

Remark 1.1. As mentioned earlier, Kantorovich’s problem is a relaxation of Monge’s problem [11]. To illustrate this, if \((Id, T) : X \rightarrow X \times Y\) is defined by \((Id, T)(x) := (x, T(x))\) and \(T_\mu = \nu\), then \((Id, T)_t \mu \in \Pi(\mu, \nu)\), i.e., in a sense Kantorovich’s problem contains Monge’s.

An important property we require of costs on paths is that of coercivity, to prove existence of minimizers for (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 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

\[
\inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} \int_\Omega c(\gamma) \, d\pi(\gamma), \tag{5}
\]

this problem admits a solution under general conditions (Theorem 1.4). A similar result to Theorem 1.4 applies to a new formulation in which we incorporate interacting paths (Theorem 1.8). The former is analogous to a result that extends Brenier [4] and Gangbo and McCann [8]. Indeed it turns out that these minimizers are given by maps (Theorem 1.5)

\[ \Gamma : X \times [0, 1] \rightarrow \Omega, \]

with the properties

\[ \Gamma(x, 0) = x \quad \text{for all} \quad x \in X; \quad \Gamma(x, 1) = T(x); \quad T_1 \mu_0 = \mu_1. \]

For a map \( T \) that solves an auxiliary transport problem, the optimal measure will be given by \( \pi_T := (\Gamma)^*_t \mu_0 \) and solves the Kantorovich’s problem (5) uniquely. Here, \( \mu_0 \) is absolutely continuous with respect to Lebesgue; and, in turn \( T \) solves the Monge’s path dependent problem for the auxiliary cost, \( c_e(x, y) \), (9). For all \( x \) in the support of \( \mu_0 \), the minimal path is of the form \( t \mapsto \gamma_{x, y}(t) \), just like in the content of Theorem 1.5. And thus for \( T(x) \) the optimal map pushing \( \mu_0 \) forward to \( \mu_1 \) with respect to \( c_e \), then the optimal map of paths is given by the composition \( \Gamma(x, t) := \gamma_{x, T(x)}(t) \).

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

1.3. Summary of main results. Let \( \mathbb{R}^n \) be the \( n \)-dimensional Euclidean space and let \( X \subset \mathbb{R}^n \) be a bounded domain. The space of probability measures will be denoted by \( \mathcal{P}(X) \). Consider the space of continuous paths,

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

If we endow \( \Omega \) with a metric \( ||\gamma_1 - \gamma_2|| := \max_{0 \leq t \leq 1} \{d(\gamma_1(t), \gamma_2(t))\} \), then \( \Omega \) becomes a complete metric space. In addition as \( X \) is compact, \( \Omega \) 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 \( X := B_R(0) \subset \mathbb{R}^n \) for a large \( R > 0 \). Let \( \mu_0, \mu_1 \in \mathcal{P}(X) \).

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

Definition 1.2. Define \( e_t : \Omega \rightarrow X \), the evaluation map by

\[ e_t(\gamma) = \gamma(t) \quad \text{for all} \quad 0 \leq t \leq 1. \]
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 $\gamma$, respectively. The admissibility condition on $\pi$ is now given in the next definition.

**Definition 1.3.** Given $\mu_0, \mu_1 \in \mathcal{P}(X)$, we say that $\pi \in \mathcal{P}(\Omega)$ is *admissible* if the following holds

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

The set of all probability measures $\pi \in \mathcal{P}(\Omega)$ satisfying Definition 1.3 will be denoted by $\Pi_{\text{path}}(\mu_0, \mu_1)$. Such measures $\pi$ are also known as dynamical couplings. This notion is well known in the classical optimal transport literature, see Villani’s discussion in [22, Chapter 7]. Moreover, the set $\Pi_{\text{path}}$ 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 3.12, where we show a probability measure on the space of paths projects to a solution of the MKP in Euclidean space.

Let $c : \Omega \to \mathbb{R}$ be a cost function. Define the linear functional $E_0 : \mathcal{P}(\Omega) \to \mathbb{R}$ by

$$E_0(\pi) := \int_{\Omega} c(\gamma) d\pi(\gamma).$$

The *optimal path problem* is then:

**Problem A.** Given $\mu_0, \mu_1 \in \mathcal{P}(X)$, minimize $E_0(\pi)$ 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 $\Omega$.

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

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

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 $L < 2/3$.

Our first Theorem (Theorem 1.4) is on existence under general conditions.

**Theorem 1.4.** Suppose $\mu_0$ and $\mu_1$ 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 [22] and Gangbo’s and McCann’s results [8]:

**Theorem 1.5.** Suppose $\mu_0$ and $\mu_1$ 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, $\Gamma : \text{spl}(\mu_0) \to \Omega$ (see Section 3.4).

For the next problem we add an interaction term to $E_0(\pi)$. Given a continuous function $K : \Omega \times \Omega \to \mathbb{R}$ define $U : \Omega \times \mathcal{P}(\Omega) \to \mathbb{R}$ by

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

(6)
Thus the new functional with interaction that we will study is the following:

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

(7)

The function \( \mathcal{K}(\gamma, \sigma) \) can be thought of as measuring interactions between \( \gamma \) and \( \sigma \). Then \( \mathcal{U}(\gamma, \pi) \) measures the total interaction between \( \gamma \) and a distribution of paths given by \( \pi \). The integral term \( \int_{\Omega} \mathcal{U}(\gamma, \pi) \, d\pi(\gamma) \) in (7) 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 (7), given by (7), 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 (7) is not linear in \( \pi \), and subtle because it might not even be convex in general.

As a particular example of interest, and for concreteness, consider \( \mathcal{K}(\gamma, \sigma) \) given by an integral over the time interval \([0, 1]\) of an exponential function: in general \( \mathcal{K}(\gamma, \sigma) = \int_0^1 \kappa(\gamma(t) - \sigma(t)) \, dt \), where \( \kappa : \mathbb{R}^n \to \mathbb{R} \) is of positive type,

\[ \mathcal{K}(\gamma, \sigma) := \int_0^1 \theta \exp \left\{ -\beta |\gamma(t) - \sigma(t)|^2 \right\} \, dt, \]  

(8)

for \( \beta > 0 \) and \( \theta > 0 \). Note that if \( \gamma \) and \( \sigma \) are very close to each other, \( \mathcal{K}(\gamma, \sigma) \) is very close to 0. If, on the other hand, \( \sigma \) and \( \gamma \) are a large distance away from each other, then \( \mathcal{K}(\gamma, \sigma) \) will be small. More generally, Bochner’s theorem implies \( \mathcal{K} \) is convex if \( \mathcal{K}(\gamma, \sigma) = \int_0^1 \kappa(\gamma(t) - \sigma(t)) \, dt \), where \( \kappa \) is the Fourier transform of a finite, positive measure, as given in Reed’s and Simon’s *Functional Analysis I* book [15, Theorem IX.9]. (See Appendix A).

**Remark 1.6.** Let \( d\pi_t(\cdot) := (e_t)_y \pi \). As \( \kappa \) 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 \( \pi \). As a result the quadratic functional term in (7) is convex in \( \pi \). Indeed, the following computation shows it:

\[
\int_{\Omega} \int_{\Omega} \left( \int_0^1 \kappa(\gamma(t) - \sigma(t)) \, dt \right) d\pi(\sigma) d\pi(\gamma) = \int_0^1 \int_{\Omega} \int_{\Omega} \kappa(\gamma(t) - \sigma(t)) d\pi(\sigma) d\pi(\gamma) dt
\]

\[
(d\pi_t = (e_t)_y \pi) = \int_0^1 H(\pi) dt
\]

Consequently, the functional (7) is convex.

**Remark 1.7.** For concreteness we will focus on the Gaussian interaction but other kernels are covered by our methods as well. For example, \( \kappa(x) = |x|^{2-n} \) \((n \geq 3)\). The kernel \( \mathcal{K}(\gamma, \sigma) = \int_0^1 |\gamma(t) - \sigma(t)|^{2-n} \, dt \) is convex, as the Bochner-Schwartz theorem [15, 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 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. For Bochner’s theorem, Bochner-Schwartz theorem, and the Coulomb kernel see Appendix A.

In the same spirit to Theorems 1.4 and 1.5, we study a new optimal transportation with interacting paths problem (1), and show existence of minimizers of (7) and characterize the minimizers. For the remaining Theorems, we will only consider \( \mathcal{K}(\gamma, \sigma) \) given by (8). 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 1.8. Suppose $\mu_0$ and $\mu_1$ have compact support and that $c$ satisfies Condition 1 then the Kantorovich problem with interaction, Problem B, has at least one solution.

Theorem 1.9. A solution to Problem B is a solution to Problem A with some effective cost.

Theorem 1.10. Suppose $\mu_0$ and $\mu_1$ have compact support, $\mu_0 << dx$, $c$ satisfies Condition 3 and $K$ is as in [9] for some $\theta$ and $\beta$. Then, there is $\theta_0 > 0$ depending on $\theta$ and $\beta$ such that if $\theta \in (0, \theta_0)$ then Problem B has a unique solution given by a map.

1.4. Outline of the paper. Following the introduction and main results, Section 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 $\Omega$, and that for such a path minimizing an endpoint cost function $c_e$ with endpoints fixed is differentiable. Section 3 deals with existence of minimizers and their properties. We prove Theorem 1.4 and produce the dual problem. Lastly, we prove Theorem 1.5. The fourth section 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 the interaction effects. We also prove Theorems 1.8 and 1.10 in this section. Two appendices provide Bochner’s statement on the convexity of the Coulomb potential and the other establishes the differentiability of the end point cost function [9].

2. Minimal paths

In [22], Villani explains a construction in optimal transport that an action, $A$, which measures the cost of displacing along a continuous path $\gamma$, 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, $y$, $c(x, y) = \inf\{A(\gamma) : \gamma(0) = x, \gamma(1) = y, \gamma \in C([0, 1]; \mathbb{R}^n)\}$.

Similarly, we use this classical construction to define the endpoint cost function [9]. A typical and classical example of such an action is the kinetic energy, $A(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2/2dt$. More generally, an action is given by Lagrangians as Villani details in [22]. And following this line of thinking heavily, 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 $A(\gamma)$. 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 Section 3 and Section 4.

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

Proposition 2.1. Suppose $\gamma : [0, 1] \to \mathbb{R}^n$ is $C^2$ and $\delta > 0$ is such that $|\dot{\gamma}(t)| \leq \delta$ \ \forall t \in [0, 1]$. Then for every $t \in [0, 1]$ we have $|\gamma(t) - (\gamma(0) + t[\gamma(1) - \gamma(0)])| \leq \sqrt{\delta}$ and $|\dot{\gamma}(t) - (\gamma(1) - \gamma(0))| \leq \sqrt{\delta}$.

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

In coordinate components of $e(t)$, we will show that for each $i = 1, \ldots, n$ there is some $t_i \in [0, 1]$ such that $\dot{e}_i(t_i) = 0$. Thus, in this case, $|e_i(t)| \leq \delta$ for all $t \in [0, 1]$. Then
\[
|\dot{e}(t)| \leq \sqrt{\delta^2 + \cdots + \delta^2} = \sqrt{n}\delta.
\]
According to the mean value theorem, applied to $e_i(t_i)$, there exists $\tilde{t}_i \in [0, 1]$ such that
\[
0 = e_i(1) - e_i(0) = \dot{e}_i(\tilde{t}_i).
\]
Therefore for all $t \in [0, 1]$, $e(t) = e(0) + \int_0^t \dot{e}(s) \, ds = \int_0^t \dot{e}(s) \, ds$. Then $|e(t)| \leq \sqrt{n}\delta$ and the proposition follows at once.

We will focus on Lagrangians of the form
\[
L(\gamma(t), \gamma(t), t) := \frac{1}{2} |\dot{\gamma}(t)|^2 - V(\gamma(t), t),
\]
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
\[
c(\gamma) = \int_0^1 L(\gamma(t), \gamma(t), t) \, dt.
\]
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 $\gamma$ at $t$, with endpoints $x$ and $y$ of $\gamma$ fixed and $V(x, t)$ some potential, just like in Villani’s Lagrangian example in [22]. 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
\[
\frac{d^2 x}{dt^2} = -\nabla_x V(x, t).
\]

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

**Proposition 2.2.** 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 $\gamma$ 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 Evans’ [6] Ch 8].

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

**Corollary 2.3.** Suppose the Lagrangian is given by $L(\gamma, \gamma, t) = \frac{1}{2} |\dot{\gamma}(t)|^2 - V(\gamma(t), t)$ 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 $\gamma$ is minimal,
\[
|\gamma(t) - (\gamma(0) + t[\gamma(1) - \gamma(0)])| \leq \sqrt{n}\delta
\]
and
\[
|\gamma(t) - (\gamma(1) - \gamma(0))| \leq \sqrt{n}\delta.
\]
Proposition 2.2 applies to show that $|\gamma(t)| \leq \delta$ for all $t \in [0,1]$. Then applying Proposition 2.1 establishes the corollary.

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

**Proposition 2.4.** 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 $\gamma_{x,y}(\cdot)$ 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

$$
\gamma(t) = -\nabla_2 V(\gamma(t), t),
\gamma(0) = x,
\gamma(1) = y.
$$

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 $d(\gamma_{x_1,y_1}(t), \gamma_{x_2,y_2}(t)) := \sup_{[0,t]} |\gamma_{x_1,y_1}(t) - \gamma_{x_2,y_2}(t)|$. Then we have

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

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

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

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

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

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

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. \( \square \)

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 \( \gamma \) that go from \( x \) to \( y \). So for any pair of points \( x, y \in \mathbb{R}^n \), let

\[
c_e(x, y) := \inf_{\gamma(0)=x, \gamma(1)=y} c(\gamma).
\]

**Lemma 2.5.** Let \( c_e(x, y) \) be as in \([\ref{eq:cost}]\). 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 \( c_e(y, x) \) 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 \( \tilde{\gamma}(t) \). More concretely, let \( s : [0, 1] \to [0, 1] \) be given by \( s(t) = 1 - t \) for all \( t \in [0, 1] \). Then \( \tilde{\gamma}(t) = \gamma(s(t)) \) 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

\[
\tilde{\gamma}(r) = \frac{d}{dr} \gamma = \frac{d}{dr} \gamma(s(r)) = \frac{d\gamma}{dt} \frac{ds}{dr} = -\dot{\gamma}(t).
\]

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

\[
c_e(x, y) = c_e(y, x).
\]

To show the differentiability of \( c_e(x, y) \) 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 \( \gamma_0 \) from \( \gamma_0(0) = x \) to \( \gamma_0(1) = y \). In this case we have \( c_e(x, y) = c(\gamma_0) \). Consider another path, not necessarily minimal, \( \gamma_h(t) := \gamma_0(t) + th \cdot \hat{e} \) from \( x \) to \( y + h\hat{e} \), where \( h \neq 0 \) and \( \hat{e} := (0, \ldots, 1, \ldots, 0) \). From this, we know \( c_e(x, y + h\hat{e}) \leq c(\gamma_h(t)) \). Putting this together, we get

\[
\lim_{h \to 0} \frac{c_e(x, y + h\hat{e}) - c_e(x, y)}{h} \leq \lim_{h \to 0} \frac{c(\gamma_h(t)) - c(\gamma_0(t))}{h}.
\]

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 \([\ref{eq:diff}]\) is well-defined. That the left-hand side limit exists\(^2\) follows from a standard argument (using compactness and uniqueness of solutions to linear ODE’s). Then it follows \( \langle \nabla_y c_e(x, y), \hat{e} \rangle \leq \frac{d}{dh} \big|_{h=0} c(\gamma_h(t)) \). To establish the reverse inequality, consider another path \( \gamma_h(t) = \gamma_0(t) + th(-\hat{e}) \) from \( x \) to \( y - h\hat{e} \). From this, \([\ref{eq:diff}]\) becomes

\[
\lim_{h \to 0} \frac{c_e(x, y - h\hat{e}) - c_e(x, y)}{h} \leq \lim_{h \to 0} \frac{c(\gamma_h(t)) - c(\gamma_0(t))}{h}.
\]

\(^2\)We prove this fact in Appendix B
A change of variables with \( z = y - h\hat{e} \); noticing \( z \to y \) as \( h \to 0 \), the left hand side of (11) yields the reverse inequality, so that
\[
- \lim_{h \to 0} \frac{c_{e}(x, y + h\hat{e}) - c_{e}(x, y)}{h} = \langle -\nabla_{y}c_{e}(x, y), \hat{e} \rangle,
\]
while the right hand side of (11) with \( \gamma_{h}(t) = \gamma_{0}(t) - th \cdot \hat{e} \) gives
\[
\frac{d}{dh} \bigg|_{h=0} c(\gamma_{h}(t)) = \left\langle -\int_{0}^{1} \left( \gamma_{0}(t) - \nabla_{x}V(\gamma_{0}(t), t)t \right) dt, \hat{e} \right\rangle.
\]
Putting this together in (10) gives the reverse inequality and thus
\[
\langle \nabla_{y}c_{e}(x, y), \hat{e} \rangle = \frac{d}{dh} \bigg|_{h=0} c(\gamma_{h}) = \left\langle \int_{0}^{1} \left( \dot{\gamma}_{0}(t) - \nabla_{x}V(\gamma_{0}(t), t)t \right) dt, \hat{e} \right\rangle.
\]

Therefore the differentiability of \( c_{e}(x, y) \) with respect to \( y \) follows and therefore \( \nabla_{y}c_{e}(x, y) = y - x - \int_{0}^{1} t \nabla_{x}V(\gamma_{0}(t), t) dt. \)

Finally, we show \( \nabla c_{e}(x, y) \) satisfies the twist condition from [22].

**Lemma 2.6.** Let \( V \) be bounded and of class \( C^{2} \) and let \( \nabla V \) 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 2.5 tells us
\[
\nabla_{y}c_{e}(x, y) = y - x - \int_{0}^{1} t \nabla_{x}V(\gamma_{x,y}(t), t) dt.
\]
Now for all \( y \),
\[
\nabla_{y}c_{e}(x_{1}, y) - \nabla_{y}c_{e}(x_{2}, y) = x_{2} - x_{1} - \int_{0}^{1} t \left( \nabla_{x_{1}}V(\gamma_{x_{1},y}(t), t) - \nabla_{x_{2}}V(\gamma_{x_{2},y}(t), t) \right) dt.
\]

The reverse triangle inequality and invoking Proposition 2.4, most notably the bound at the end of its proof and the Lipschitz condition on \( \nabla V \), we have
\[
|\nabla_{y}c_{e}(x_{1}, y) - \nabla_{y}c_{e}(x_{2}, y)| \geq |x_{1} - x_{2}| - L \int_{0}^{1} t|\gamma_{x_{1},y}(t) - \gamma_{x_{2},y}(t)| dt
\]
\[
\geq |x_{1} - x_{2}| - L \left( \frac{1}{1-L} \right) |x_{1} - x_{2}| \int_{0}^{1} t dt
\]
\[
= |x_{1} - x_{2}| \left( 1 - \frac{1}{2} \frac{L}{1-L} \right)
\]
\[
> 0.
\]
Since \( L < 2/3 \).
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. An important tool in measure theory is compactness of measures. 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 $\mu$ 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

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

Prokhorov’s theorem, which can be found in either [21] or [17], is the following.

**Theorem 3.2** (Prokhorov’s Theorem). 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 $\mu$ such that

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

We need two additional propositions before we prove existence of minimizers. And Condition 3.1 will help us with the proof of Theorem 1.4 as we will momentarily see.

**Proposition 3.3.** Condition 3 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

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

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

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

$$\lim \inf_{n \rightarrow \infty} \int_0^1 |\gamma_n(t)|^2 \, dt \geq \int_0^1 |\gamma(t)|^2 \, dt$$

from [6, Ch. 8.2.2, Theorem 1], as $\gamma$ 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 $\lim_{n \rightarrow \infty} \int_0^1 V(\gamma_n, t) \, dt = \int_0^1 V(\gamma, t) \, dt$. Putting this together with the above liminf establishes the result.

We are now ready to prove Theorem 1.4

Proof of Theorem 1.4. Since $c(\gamma) \geq -K$, then $E_0(\pi) \geq -K$ for all $\pi \in \mathcal{P}(\Omega)$, thus the infimum of $E_0$ in $\Pi_{\text{path}}(\mu_0, \mu_1)$ is finite. Let $\{\pi_k\}_k$ be a minimizing sequence $E_0(\pi_k) \rightarrow \inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} E_0(\pi)$. Note that since $\{\pi_k\}_k$ is minimizing, we can find a constant $C > 0$ for which $E_0(\pi_k) \leq C$ for all $k$.

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

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

One can see this from the fact that $(\varepsilon_0)_{\pi} = \mu_0$ implies that $\text{spt} (\pi) \subset \varepsilon_0^{-1}(\text{spt} (\mu_0))$. So the equality follows from $\pi$ 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 $E_0(\pi)/N \geq \pi(X_N^c)$. Applying this to each $\pi_k$,

$$\pi_k(\Omega_{M,N}^c) \leq \frac{1}{N} E_0(\pi_k) \leq \frac{C}{N} \quad \text{for all} \quad k \quad \text{(12)}$$

Returning to inequality (12), for each $N > 0$ given $\varepsilon := \frac{C}{N} > 0$,

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

By Condition 4, $\Omega_{M,N}$ 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 $\rho$ such that

$$\pi_{k_j} \rightharpoonup \rho \quad \text{as} \quad j \rightarrow \infty.$$
The claim is that then \( \rho \in \Pi_{\text{path}}(\mu_0, \mu_1) \). Recall the change of variables formula \([2]\). We will use an extension of this with the evaluation map to show \( \rho \) lies in \( \Pi_{\text{path}}(\mu_0, \mu_1) \). Given any test function \( \phi \in C_b(\Omega) \),

\[
\int_{\Omega} \phi(x) \, d((e_0)_x^2 \rho(x)) = \int_{\Omega} \phi(\gamma(0)) \, d\rho(\gamma) = \lim_{j \to \infty} \int_{\Omega} \phi(\gamma(0)) \, d\pi_{k_j}(\gamma)
\]

\[
= \lim_{j \to \infty} \int_{\Omega} \phi(\gamma) \, d((e_0)_x^2 \pi_{k_j})
\]

\[
= \int_{\Omega} \phi(\gamma) \, d\mu_0.
\]

From the arbitrariness of \( \phi \in C_b(\Omega) \) we get \( (e_0)_x^2 \rho = \mu_0 \). Similarly, by the same argument we can show \( (e_1)_x^2 \rho = \mu_1 \). Thus, \( \rho \in \Pi_{\text{path}}(\mu_0, \mu_1) \).

To conclude, we must show \( E_0(\rho) \leq \liminf_{j \to \infty} E_0(\pi_{k_j}) \). By Proposition 3.3, \( c(\gamma) \) is nonnegative and lower semi-continuous. Then it is well known \([22, \text{Ch 4}]\) that \( c \) can be written as the limit of a nondecreasing, sequence of bounded continuous functions \( \{c_n\}_{n \geq 0} \) for \( c_n \leq c \). The monotone convergence theorem shows

\[
E_0(\rho) = \int_{\Omega} c(\gamma) \, d\rho(\gamma) = \lim_{n \to \infty} \int_{\Omega} c_n(\gamma) \, d\rho(\gamma)
\]

\[
= \lim_{n \to \infty} \lim_{j \to \infty} \int_{\Omega} c_n(\gamma) \, d\pi_{k_j}(\gamma)
\]

\[
\leq \liminf_{j \to \infty} \int_{\Omega} c(\gamma) \, d\pi_{k_j}(\gamma).
\]

We are done since \( E_0(\rho) \leq \liminf_{j \to \infty} \int_{\Omega} c(\gamma) \, d\pi_{k_j}(\gamma) = \inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} E_0(\pi) \) but the definition of infimum \( \inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} E_0(\pi) \leq E_0(\rho) \) yields,

\[
E_0(\rho) = \inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} E_0(\pi),
\]

and so \( E_0(\pi) \) attains its infimum at \( \rho \). \( \square \)

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\([3]\) equation (9). Namely,

\[
c_e(x, y) := \inf\{c(\gamma) \mid \gamma(0) = x \quad \text{and} \quad \gamma(1) = y\}.
\]

Examples.

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

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

It is not too difficult to show this. In fact we can find in Villani’s books \([21, 22]\):  

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 \( \int_0^1 h(\dot{\gamma}(t)) \, dt \geq h\left( \int_0^1 \dot{\gamma}(t) \, dt \right) = h(y - x) \), by Jensen’s inequality \([21, 22]\).

The Kantorovich problem associated to paths is the problem of minimizing the linear functional \( \pi \to \int c(\gamma) \, d\pi \) subject to the linear constraints \( (e_0)_x^2 \pi = \mu_0 \), \( (e_1)_x^2 \pi = \mu_1 \), and \( \pi \geq 0 \). Linear minimization problems with convex constraints of this type admit a natural dual problem, see Solomon’s book \([19, \text{Ch 10}]\). We now define this dual problem, as is usually done, using Lagrange multipliers.
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
\[
L(\pi; \phi, \psi) := \int_{\Omega} c(\gamma) \, d\pi(\gamma) + \int_X \phi(x) d\mu_0(x) - \int_{\Omega} \phi(\gamma(0)) d\pi(\gamma) + \int_X \psi(y) d\mu_1(y) - \int_{\Omega} \psi(\gamma(1)) d\pi(\gamma).
\]

Then after some rearrangement
\[
L(\pi; \phi, \psi) = \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) + \int_{\Omega} (c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1)))) \, d\pi(\gamma).
\]

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

**Proposition 3.5.** 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
\[
D(\phi, \psi) = \begin{cases} \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) & \text{if } \phi(x) + \psi(y) \leq c_{e}(x, y) \quad \forall x, y \\ -\infty & \text{otherwise} \end{cases}
\]

**Proof.** Since the first two terms in $L(\pi; \phi, \psi)$ are independent of $\pi$,
\[
D(\phi, \psi) = \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) + \inf_{\pi} \int_{\Omega} (c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1)))) \, d\pi(\gamma).
\]

If there exists some $\gamma_0$ 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 $\gamma_0$ 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 $\geq 0$ for all $\pi$, and taking $\pi = 0$ we get the inf equal to zero, $D(\phi, \psi) = \int_X \phi \, d\mu + \int_X \psi \, d\mu_1$. \[\square\]

Thus we arrive at the dual problem:

**Definition 3.6.** (Dual problem.) Let $(\phi, \psi) \in L^1(\mu_0) \times L^1(\mu_1)$. Let
\[
\Pi^\ast := \{ \phi, \psi : \phi(x) + \psi(y) \leq c_{e}(x, y) \}.
\]

The dual problem consists of finding the following supremum
\[
\sup \left\{ \int_X \phi(x) d\mu_0(x) + \int_X \psi(y) d\mu_1(y) : (\phi, \psi) \in \Pi^\ast \right\}.
\]

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 $c_{e}(x, y)$ defined by [1].

**Theorem 3.7.** 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 $c_{e}(x, y) \leq f(x) + g(y)$ for some $f \in L^1(d\mu_0), g \in L^1(d\mu_1)$, then
\[
\inf_{\pi \in \Pi_{\text{path}}} \left\{ \int_{\Omega} c(\gamma) \, d\pi(\gamma) \right\} = \sup_{(\phi, \psi) \in \Pi^\ast} \left\{ \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y) \right\} \quad (13)
\]

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 3.7.

**Corollary 3.8.** Under the assumptions of Theorem 3.7, if \( \pi_0 \in \Pi_{\text{path}}(\mu_0, \mu_1) \) achieves the infimum, then for \( \pi_0 \)-a.e. \( \gamma \in \Omega \) we have \( c(\gamma) = c_\epsilon(\gamma(0), \gamma(1)) \).

**Proof.** Let \((\phi_0, \psi_0) \in \Pi^*\) be optimal. Then by Theorem 3.7 we have
\[
\int_\Omega c(\gamma) \, d\pi_0(\gamma) = \int_X \phi_0(x) \, d\mu_0(x) + \int_X \psi_0(y) \, d\mu_1(y)
= \int_\Omega (\phi_0(\gamma(0)) + \psi_0(\gamma(1))) \, d\pi_0(\gamma).
\]
Therefore,
\[
\int_\Omega (c(\gamma) - (\phi_0(\gamma(0)) + \psi_0(\gamma(1)))) \, d\pi_0(\gamma) = 0.
\]
Since the integrand is \( \geq 0 \), not just \( \pi_0 \)-a.e. it follows that \( c(\gamma) = \phi_0(\gamma(0)) + \psi_0(\gamma(1)) \) \( \pi_0 \)-a.e. \( \gamma \).
Since \( \phi_0(x) + \psi_0(y) \leq c_\epsilon(x, y) \), \( c(\gamma) = \inf \{ c(\gamma) : \gamma(0) = x, \gamma(1) = y \} \) for \( \pi_0 \)-a.e. \( \gamma \).

Corollary 3.8 indicates that whenever \( \pi \) is optimal, the support of \( \pi \) lies in the set of all minimal paths, \( \Omega_{\min} \). An explicit proof of this statement, that does not rely on Theorem 3.7, is given in Lemma 3.11.

3.3. **Potentials and cyclically monotone sets.** In this section we further study the properties from the inequality of Definition 3.6. In some sense, we follow the standard theory of using concave potentials \( \phi, \psi \) as in Ambrosio’s and Gigli’s guide to optimal transport in [1] associated to Theorem 3.7 coming from the theory of superdifferentiability. Namely, a function \( \phi \) is called \( c \)-concave if \( \phi = \phi^* \) as given in (14).

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 \( \mu_0 \)-a.e. \( x \) and \( \mu_1 \)-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 3.6. Then the “concavity” transforms stemming from [21] are the following. Given \((\phi, \psi) \in \Pi^*\), for \( \mu_1 \)-a.e. \( y \)
\[
\psi(y) \leq c_\epsilon(x, y) - \phi(x).
\]
Taking the infimum with respect to \( x \),
\[
\psi(y) \leq \inf_{x \in X} [c_\epsilon(x, y) - \phi(x)] := \phi^*(y)
\] (14)

Similarly, for \( \mu_0 \)-a.e. \( x \),
\[
\phi(x) \leq c_\epsilon(x, y) - \psi(y),
\]
implementing the infimum with respect to \( y \),
\[
\phi(x) \leq \inf_{y \in X} [c_\epsilon(x, y) - \psi(y)] := \psi^*(x).
\] (15)

Calling \( J(\phi, \psi) := \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y) \), the linear functional from the supremum in Theorem 3.7, we can witness from (14) that
\[
J(\phi, \phi^*) \geq J(\phi, \psi).
\]
And from (14) and (15) for \( \mu_0 \)-a.e. \( x \).
\[ \phi^{**}(x) =: \inf_{y \in X} [c_e(x, y) - \phi^*(y)] \geq \phi(x), \]

and

\[ J(\phi^{**}, \phi^*) \geq J(\phi, \phi^*) \geq J(\phi, \psi). \]

This heuristic shows that the pair \((\phi, \phi^*)\) maximizes the dual problem of Definition 3.6.

The superdifferential set defined for a \(c\)-concave function \(\phi\) is:

\[ \partial \phi := \{ \gamma \in \Omega : \phi(x) + \phi^*(y) = c_e(x, y) \}. \]

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 [21], [17], [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 \(\mu_0\) and each \(y_i\) is contained in the support of \(\mu_1\). Then each path \(\gamma_i \in \Omega\) in the support of \(\pi\) 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 \(\tilde{\gamma}_i\) defined by

\[ \tilde{\gamma}_i(t) = \gamma_i(t) + th_i(t), \quad 0 \leq t \leq 1, \quad h_i(1) := \gamma_{i+1}(1), \]

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 (16) holds and

\[ \tilde{\gamma}_i(0) = x_i \quad \forall i \quad \text{and} \quad \tilde{\gamma}_i(1) = y_{i+1 \mod N} \quad \forall i \]

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

**Definition 3.9.** We say the set \(\{(x_i, y_i)\}\) is \(c_e\)-cyclically monotone if for all \(i = 1, \ldots, n\),

\[ \sum_{i=1}^n c_e(x_i, y_i) \leq \sum_{i=1}^n c_e(x_i, y_{\tau(i)}), \]

for any permutation \(\tau\) on \(n\) letters.

One last definition of equal importance is that of minimal paths.

**Definition 3.10.** A continuous path \(\gamma : [0, 1] \to X\) is a minimal path from its initial point \(x\) to its final point \(y\) if for all other paths \(\gamma' : [0, 1] \to X\) having the same initial and final points of \(\gamma\) satisfy

\[ c(\gamma) \leq c(\gamma'). \]

For a Borel set \(B\), \(\pi \downarrow B\) is the restriction of \(\pi\) to \(B\), namely the measure defined by

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

**Lemma 3.11.** Let \(c : \Omega \to \mathbb{R}\) be a lower semicontinuous cost function. Suppose \(\pi\) is optimal. Then if \(\gamma_0 \in \text{spt}(\pi)\), \(\gamma_0\) is minimal.
Proof. Suppose $\gamma_0 \in \text{spt} (\pi)$ and $\pi \in \Pi_{\text{path}} (\mu_0, \mu_1)$ is optimal. Let $\gamma'$ be the minimal path from $x_0 := \gamma'(0)$ to $y_0 := \gamma'(1)$, arguing by contradiction, if $\gamma_0$ is not minimal, there is $\varepsilon > 0$ such that

$$c(\gamma') < c(\gamma_0) - \varepsilon/2.$$  

Let $W := B_r(x_0) \times B_r(y_0)$ for $r > 0$ (to be specified later), $(e_0, e_1) : \Omega \to X \times X$ by $\gamma \mapsto (\gamma(0), \gamma(1))$, and define what we will call a tube

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

This is an open set in $\Omega$ containing the path $\gamma_0$. Consider the measure, $\tilde{\pi} := \pi \cdot T/\pi [T]; \pi [T]$ will be positive as $\gamma_0 \in \text{spt} (\pi)$. Take $0 < e_0 < \pi [T]$. Define the measures

$$\nu_x := (e_0)_x \tilde{\pi} \quad \text{and} \quad \nu_y := (e_1)_y \tilde{\pi}.$$ 

Now build a measure $\bar{\pi} \in \Pi (\nu_x, \nu_y)$ 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 $\bar{\pi} := (g)_{*} (\nu_x \otimes \nu_y)$ and define

$$\pi' := \bar{\pi} - e_0 \nu_x + e_0 \nu_y.$$ 

That $\pi'$ is positive follows from $\pi - e_0 \nu_x$ being positive, which is thanks to $e_0 < \pi [T]$. The marginals of $\pi'$ share the same marginals of $\pi$. Concretely,

$$(e_0)_{x} \pi' = \mu_0 - e_0 \nu_x + e_0 (e_0)_{x} \tilde{\pi},$$

and for every Borel subset $B \subset X$, $(\nu_x \otimes \nu_y) [g^{-1} \circ e_0^{-1} (B)] = \nu_x [B]$. 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 $\pi$. Let $m : \Omega \to \Omega_{\text{min}}, \gamma \mapsto \gamma_{\text{min}}$ be defined as the minimal $\gamma$ between $x := \gamma_{\text{min}}(0) = \gamma(0)$ and $y := \gamma_{\text{min}}(1) = \gamma(1)$. That the map $m$ is well-defined follows from Proposition 2.4: it tells us that there exists a unique minimal path $\gamma$ between $x$ and $y$. So

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

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

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

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

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

Putting this all together yields,

$$\int c(\gamma) d\pi (\gamma) - \int c(\gamma) d\pi' (\gamma) = c_0 \int c(\gamma) d\tilde{\pi} (\gamma) - \varepsilon_0 \int c(\gamma) d\tilde{\pi} (\gamma)$$

$$\geq c_0 \int c(\gamma_{\text{min}}) d\tilde{\pi} (\gamma) - \varepsilon_0 \int c(\gamma) d\tilde{\pi} (\gamma)$$

$$\geq c_0 \int \left( c(\gamma') + \frac{\varepsilon}{2} \right) d\tilde{\pi} (\gamma) - \varepsilon_0 \int \left( c(\gamma') + \frac{\varepsilon}{4} \right) d\tilde{\pi} (\gamma)$$

$$= c_0 \frac{\varepsilon}{4} > 0.$$
By definition $\pi[\Omega] = (\pi^\top T)[\Omega]/\pi[T] = 1$. Furthermore, we used the fact that both $\tilde{\pi}$ and $\tilde{\pi}$ are supported on $T$ and have unit mass due to the rescaling.

Next we show that optimal plans in $\Omega$ have cyclical monotone support. The picture complementing the proof of Lemma 3.12 is given below:

$$T_i := (e_0, e_1)^{-1}(W_i) \cap \Omega_{\text{min}}$$

$W_i := B_r(x_i) \times B_r(y_i)$

**Figure 1**: This picture indicates how the pull-back mapping $T_i$ (the above tubo) is obtained from $W_i$. 
Lemma 3.12. Suppose \( c : \Omega \to \mathbb{R} \) is a continuous cost function and \( \pi \in \Pi_{\text{path}}(\mu_0, \mu_1) \) optimal with respect to \( c \). Let \( \bar{\pi} := (e_0, e_1)\pi \in \Pi(\mu_0, \mu_1) \). Then the support of \( \bar{\pi} \), \( \text{spt}(\bar{\pi}) \), is \( c_\varepsilon \)-cyclically monotone. Moreover, \( \bar{\pi} \) is optimal with respect to \( c_\varepsilon(x, y) \) defined by \( [9] \).

Proof. We follow a classical proof from standard optimal transportation, in particular see the presentation in Santambrogio’s book [17, Theorem 1.38]. The new concept in the proof is to incorporate path dependence.

Let \( \tilde{\pi} := (e_0, e_1)^e\pi \) be a transport plan from \( \mu_0 \) onto \( \mu_1 \) obtained by pushing forward \( \pi \) 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 \( \text{spt}(\pi) \subset \Omega_{\text{min}} \), the set of all minimal paths by Lemma 3.11. Suppose by way of contradiction that \( \text{spt}(\tilde{\pi}) \) is not \( c_\varepsilon \)-cyclically monotone. Then there exist \( n, \tau \), and minimal paths \( \gamma_i \) from \( x_i \) to \( y_i \), and \( \tilde{\gamma}_i \) from \( x_i \) to \( y_{\tau(i)} \) in \( \pi \), respectively \( \{(x_i, y_i)\} \subset \text{spt} \pi \), such that

\[
\sum_{i=1}^n c(\gamma_i) > \sum_{i=1}^n c(\tilde{\gamma}_i),
\]

where the path \( \tilde{\gamma}_i \) satisfies \([16]\) or \( \gamma_i(0) = \gamma_i(0) \) and \( \gamma_i(1) = \gamma_{\tau(i)}(1) \) for all \( i = 1, \ldots, n \). Then since \( \gamma_i \) and \( \tilde{\gamma}_i \) are minimal, \( c(\gamma_i) = c_e(x_i, y_i) \) and \( c(\tilde{\gamma}_i) = c_e(x_i, y_{\tau(i)}) \), 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

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

By continuity of \( c \), there exists an open neighborhood called a tubo

\[
T_i := (e_0, e_1)^{-1}(B_{r_i}(x_i) \times B_{r_i}(y_i)) \cap \Omega_{\text{min}}
\]

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

\[
\gamma_i := (e_0, e_1)^{-1}(B_{r_i}(x_i) \times B_{r_i}(y_{\tau(i)})) \cap \Omega_{\text{min}},
\]

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

Now define the measures

\[
\pi_i := \pi_{\upharpoonright T_i}/\pi [T_i], \quad \nu_{x,i} := (e_0)^\pi_i, \quad \text{and} \quad \nu_{y,i} := (e_1)^\pi_i,
\]

and note that \( \pi [T_i] \) will be positive for each \( i \) since \( \gamma_i \) is contained in the support of \( \pi \). Equivalently \( (x_i, y_i) \in \text{spt} (\tilde{\pi}) \). Take \( 0 < \varepsilon_0 < \min_i \pi [T_i] \).

Construct a measure \( \tilde{\pi}_i \in \Pi(\nu_{x,i}, \nu_{y,\tau(i)}) \), 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 \( \gamma_{x,y}(0) = x \) 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 \( \gamma \in \tilde{T}_i \) 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 \( (x, y) \in B_{r_i}(x_i) \times B_{r_i}(y_i) \) and all \( (x, y) \in B_{r_i}(x_i) \times B_{r_i}(y_{\tau(i)}) \), respectively. Take \( \tilde{\pi}_i := (g)_\pi (\nu_{x,i} \otimes \nu_{y,\tau(i)}) \).

Now define

\[
\tilde{\pi} := \pi - \varepsilon_0 \sum_{i=1}^n \pi_i + \varepsilon_0 \sum_{i=1}^n \tilde{\pi}_i.
\]
That $\tilde{\pi}$ is positive follows from $\pi - \epsilon_0 \sum_{i=1}^n \pi_i$ being positive as $\epsilon_0 < \min_i \pi [T_i]$. 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 $\epsilon_0 \pi_i = \frac{\epsilon_0}{\pi [T_i]} \pi_i [T_i]$ and $\epsilon_0 / \pi [T_i] \leq 1 / n$.

The marginals of $\tilde{\pi}$ share the marginals of $\pi$:

$$(e_0)_2 \pi = \mu_0 - \epsilon_0 \sum_{i=1}^n \nu_{x,i} + \epsilon_0 \sum_{i=1}^n (e_0)_2 (g_2) (\nu_{x,i} \otimes \nu_{y,\tau(i)}) ,$$

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

$$(e_1)_2 \pi = \mu_1 - \epsilon_0 \sum_{i=1}^n \nu_{y,i} + \epsilon_0 \sum_{i=1}^n (e_1)_2 (g_2) (\nu_{x,i} \otimes \nu_{y,\tau(i)}) ;$$

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

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

$$\int c(\gamma) \, d\pi(\gamma) - \int c(\gamma) \, d\tilde{\pi}(\gamma) = \epsilon_0 \sum_{i=1}^n \int c(\gamma_i) \, d\pi_i(\gamma_i) - \epsilon_0 \sum_{i=1}^n \int c(\gamma_i) \, d\tilde{\pi}_i(\gamma_i)$$

$$\geq \epsilon_0 \sum_{i=1}^n (c(\gamma_i) - \epsilon) - \epsilon_0 \sum_{i=1}^n (c(\gamma_i) + \epsilon)$$

$$= \epsilon_0 \sum_{i=1}^n c_e(x_i, y_i) - c_e(x_i, y_{\tau(i)}) - 2n\epsilon > 0 ,$$

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

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 $\tilde{\pi}$ to find that $\tilde{\pi}$ is optimal with respect to $c_e(x, y)$, since spt $\tilde{\pi}$ is $c_e$-cyclically monotone. 

**Remark 3.13.** The previous Lemma 3.12 says that any optimal plan $\pi$ in $\Pi_{\text{path}}(\mu_0, \mu_1)$ “projects” to a solution of the Monge-Kantorovich problem with cost $c_e(x, y)$ and coupling $\tilde{\pi} := (e_0, e_1)_2 \pi$. Recall the notion of dynamical couplings from Section 1.3. 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 Theorem 3.7.

**Proof of Theorem 3.7.** With all the assumptions of Theorems 3.7 and Lemma 3.12 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 3.6

$$\int \Omega c_e(x, y) \, d\pi(\gamma) \geq \int \Omega (\phi(x) + \psi(y)) \, d\pi(\gamma) = \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y) .$$

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

To prove the reverse inequality, choose an optimal $\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)$. Since spt $\tilde{\pi}$ is $c_e$-cyclically monotone by Lemma 3.12 the classical theory of optimal transport applies to show
there is a lower semi-continuous concave function \( \phi \) such that \( \text{spt} \hat{\pi} \subset \partial \phi \) for which \( \phi \in L^1(d\mu_0) \) and \( \phi^* \in L^1(d\mu_1) \). Then
\[
\int_{\Omega} c_e(x, y) d\pi(\gamma) = \int_{\Omega} (\phi(x) + \phi^*(y)) d\pi(\gamma) = \int_{X} \phi(x) d\mu_0(x) + \int_{X} \phi^*(y) d\mu_1(y) \tag{18}
\]
Now, the claim is then that \((\phi, \phi^*)\) solves the maximization problem of Definition 3.6. More concretely, to prove that \((\phi, \phi^*)\) solves the maximization problem, we note that
1. \( \hat{\phi}(x) + \hat{\phi}^*(y) = c_e(x, y) \) on the support of the optimal \( \hat{\pi} \), \text{spt} \( \hat{\pi} \).
2. \( \phi(x) + \phi^*(y) \leq c_e(x, y) \) on \( X \times X \). Then
\[
\int_{X} \hat{\phi}(x) d\mu_0(x) + \int_{X} \hat{\phi}^*(y) d\mu_1(y) = \int_{\Omega} (\hat{\phi}(x) + \hat{\phi}^*(y)) d\pi(\gamma)
\]
\[
= \int_{\Omega} c_e(x, y) d\pi(\gamma)
\]
\[
\geq \int_{\Omega} (\phi(x) + \phi^*(y)) d\pi(\gamma)
\]
\[
= \int_{X} \phi(x) d\mu_0(x) + \int_{X} \phi^*(y) d\mu_1(y),
\]
and so \( \hat{\phi} \) solves the maximization problem. This with \ref{18} establishes the proof.

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 \ref{5} and also show it is given by a map \( \Gamma(x) \), provided \( \mu_0 \) is absolutely continuous with respect to the Lebesgue measure: \( \mu_0(x) << dx \). This uses, and extends, results of Brenier \cite{4} and Gangbo-McCann \cite{8} in the classical optimal transport theory.

Let \( \Omega_{\text{min}} \) denote the set of all minimal paths. If \( \pi \in \Pi_{\text{path}}(\mu_0, \mu_1) \) is optimal Corollary 3.8 indicated \( \text{spt} \pi \) is contained in \( \Omega_{\text{min}} \). In particular, the calculations and discussions of Section 2 and Section 3.2 amount to the existence of a minimal path \( \gamma_e \in \Omega \), minimizing for \( c(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 - V(\gamma(t), t) \, dt \) in \( \Omega \), in the support of \( \pi \) contained in \( \Omega_{\text{min}} \).

The uniqueness will follow. Moreover, the mapping \( \Gamma \) will be given by \ref{20} below and such that the optimal transport plan will be given by \( \pi_{\Gamma} := \pi_{\sharp} \mu_0 \)

We are going to look for mappings
\[
\Gamma : X \times [0, 1] \to X \tag{19}
\]
of the form \( \Gamma(x, t) \) for every \( x \) with the following properties
\[
\left\{\begin{array}{ll}
\Gamma(x, 0) = x & \text{for all } x \in X \\
\Gamma(x, 1) = T(x) & \text{for } T \sharp \mu_0 = \mu_1
\end{array}\right. \tag{20}
\]
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 functions from \( X \) to \( \Omega \); we shall denote them with the same letter \( \Gamma \).

The next result says that if \( T \) above maps \( \mu_0 \) to \( \mu_1 \), then \( \Gamma \) maps \( \mu_0 \) to an admissible measure in the path space.

Lemma 3.14. Define \( \pi_{\Gamma} := (\Gamma)_{\sharp} \mu_0 \). Then \( \pi_{\Gamma} \in \Pi_{\text{path}}(\mu_0, \mu_1) \).

Proof. In a moment we will unpack some definitions regarding push-forwards thru evaluation maps (see Bernot et al in \cite{3} Section 3 for an elementary introduction). 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 \( \Pi_{\text{path}}(\mu_0, \mu_1) \), 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 \( T : X \to X \) pushes forward \( \mu_0 \) to \( \mu_1 \). Then, observe that the composition \( e_1 \circ \Gamma \) coincides with \( T \), and since \( T \sharp \mu_0 = \mu_1 \), \( (e_1 \circ \Gamma)_\sharp \mu_0 = \mu_1 \).

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

\[
(e_0)_\sharp \pi \Gamma[A] = \pi \Gamma[e_0^{-1}(A)] = \mu_0[\Gamma^{-1}(e_0^{-1}(A))]
\]

\[
= \mu_0[A].
\]

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

\[
(e_1)_\sharp \pi \Gamma[B] = \pi \Gamma[e_1^{-1}(B)] = \mu_0[\Gamma^{-1}(e_1^{-1}(B))]
\]

\[
= \mu_1[B],
\]

and this completes the proof. \( \square \)

The important result we use from the standard optimal theory is that whenever \( \mu_0 \) is absolutely continuous with respect to Lebesgue [1], [21], 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 [21] Chapter 5: the following has a unique solution,

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

(21)

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 [1] and it is a consequence of Theorem 10.28 in Villani’s book [22]; so it is given by

\[
\gamma_t(x) = x - t \nabla^* L(\nabla \psi(x)), \quad 0 \leq t \leq 1,
\]

where \( \psi \) is a \( c_L \)-concave function (see Section 3.3) for which \( [\text{id} - \nabla^* L(\nabla \psi)] \sharp \mu_0 = \mu_1 \), and \( \nabla^* \) 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 [21] Chapter 5. Monge’s classical minimization Problem [3] and Problem [21] 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 \( \mu_0 \)-a.e. \( x, c_L(x, T(x)) = \int_0^1 L(\dot{\gamma}(t)) \, dt \). 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 (20): \( \Gamma : X \to \Omega \). Let \( e_0 : \Omega \to X \) be the evaluation map and recall \( X := B_R(0) \subset \mathbb{R}^n \).

**Proof of Theorem 1.5** Theorem 1.4 says there is an optimal plan \( \pi_* \) in the space of paths solving the optimal path problem (5), Problem A; while Lemma 3.12 tells us that \( \pi_* \) projects to the classical optimal transport solution with respect to \( c_e(x,y) \), namely \( \tilde{\pi}_* := (e_0, e_1) \# \pi_* \).

Lemma 2.5 says \( c_e(x,y) \) is differentiable and Lemma 2.6 says it is injective in its domain. Since \( \pi_* \) is optimal, Lemma 3.12 implies the support of \( \tilde{\pi}_* \) is \( c_e \)-cyclically monotone. The classical theory of optimal transportation [1, Theorem 1.13] says that \( \text{spt} \, \tilde{\pi}_* \) is contained in \( \partial \phi \), and as \( \phi \) is locally Lipschitz apply Rademacher’s theorem and that \( \mu_0 \ll dx \), \( \phi \) is differentiable \( \mu_0 \)-a.e. All of the above show \( c_e(x,y) \) satisfies the assumptions of Theorem 10.28 in Villani’s book [22] and so it applies to give a unique transport map \( T \) pushing \( \mu_0 \) forward to \( \mu_1 \), solving the optimal transport problem with respect to \( c_e(x,y) \).

Define a map \( \Gamma \) as in (20) containing the above data on \( T \). Let \( \gamma_{x,y}(t) \) be the minimal path between \( x \) and \( y \). Then \( c(\gamma_{x,y}(t)) = c_e(x,y) \), and using this information, let \( \Gamma(x,t) := \gamma_{x,T(x)(t)} \). More precisely, \( \Gamma \) is given as the composition,

\[
\begin{align*}
X & \xrightarrow{(Id \times T)} X \times X \xrightarrow{\gamma_{x,y}} \Omega_{\min}.
\end{align*}
\]

So since \( \text{spt} \, \pi_* \) lies in \( \Omega_{\min} \), \( \text{spt} \, \pi_* \) is concentrated on the graph of the mapping \( \Gamma \). Therefore, \( \Gamma \) uniquely solves the Monge optimal path problem. \( \square \)

**Remark 3.15.** 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 of Guillen-Kitagawa [9], Figalli-Kim-McCann [7], and Ma-Trudinger-Wang [13] 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.

4. The optimal path problem with interaction

In this section we consider a added interaction term to the functional \( E_0(\pi) \), from Section 3. We recall the linear functional in (7):

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

where the Kernel \( K \) is given by (8). The plan is to determine properties of minimizers of (7) over \( \Pi_{\text{path}} \) 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 1.8. That is, we prove existence of solutions of (1), where \( \mathcal{U} \) is given by (7) with (8).

**Proof of Theorem 1.8.** 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 1.4 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 $\pi$ 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 1.4 and coercivity condition, Condition 1, we deduce

$$\pi(X^c_N) = \int_{X^c_N} d\pi(\gamma) = \frac{1}{N} \int_{X^c_N} N d\pi(\gamma) \leq \frac{1}{N} \int_{\Omega} c(\gamma) d\pi(\gamma) = \frac{1}{N} \mathcal{E}_0(\pi) \leq \frac{1}{N} \mathcal{E}(\pi).$$

Then

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

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

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

So given $\varepsilon := \frac{C}{N} > 0$, $\pi_k(\Omega_{M,N}) \leq \frac{C}{N} = \varepsilon$ 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 probability measure $\varphi$ such that $\pi_{k_j} \rightharpoonup \varphi$ as $k \to \infty$. From the proof of Theorem 1.4, we see $\varphi$ is an element of $\Pi_{\text{path}}(\mu_0, \mu_1)$.

Next we will show

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

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

$$\int_\Omega c(\gamma) d\varphi(\gamma) + \int_\Omega \int_\Omega K(\gamma, \sigma) d\varphi(\sigma) d\varphi(\gamma) = \lim_{n \to \infty} \left( \int_\Omega c_n(\gamma) d\varphi(\gamma) + \int_\Omega \int_\Omega K_n(\gamma, \sigma) d\varphi(\sigma) d\varphi(\gamma) \right)$$

$$= \lim_{n \to \infty} \lim_{j \to \infty} \int_\Omega c_n(\gamma) d\pi_{k_j}(\gamma) + \int_\Omega \int_\Omega K_n(\gamma, \sigma) d\pi_{k_j}(\sigma) d\pi_{k_j}(\gamma)$$

$$\leq \lim_{j \to \infty} \left( \int_\Omega c(\gamma) d\pi_{k_j}(\gamma) + \int_\Omega \int_\Omega K(\gamma, \sigma) d\pi_{k_j}(\sigma) d\pi_{k_j}(\gamma) \right).$$

We have therefore proved

$$\mathcal{E}(\varphi) \leq \lim_{j \to \infty} \inf_{j \to \infty} \mathcal{E}(\pi_{k_j}),$$

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

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 $x, y \in X$, $|f(x) - f(y)| \leq b(d(x, y))$. 
Lemma 4.2. Let \((\pi_k)_k\) be a sequence of probability measures on \(\Omega\). Suppose \(U(\gamma, \pi)\), defined in (7), is bounded and \(\pi_k, \pi\) have finite total mass. If \(\pi_k \rightarrow \pi\) weakly as \(k \rightarrow \infty\), then
\[
\int_{\Omega} U(\gamma, \pi_k) \, d\pi_k(\gamma) \longrightarrow \int_{\Omega} U(\gamma, \pi) \, d\pi(\gamma) \quad \text{as} \quad k \rightarrow \infty.
\]

Proof. Recall the definitions of (7) and (8). In terms of this notation we will show \(\pi\) in (6), is bounded and
\[
\left| \int_{\Omega} U(\gamma, \pi) \, d\pi(\gamma) \right| \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty.
\]

For each compact set \(\Omega_{M,N}\) arising from the coercivity property with given \(\varepsilon := \frac{C}{\pi}\), we shall show
\[
\left| \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi_k(\sigma) \, d\pi_k(\gamma) - \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \, d\pi(\gamma) \right| \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty.
\]

Let us both unpack this. Indeed we have this expression is equal to
\[
\left| \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi_k(\sigma) \, d\pi_k(\gamma) - \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \, d\pi(\gamma) \right| + \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \, d\pi_k(\gamma)
\]
\[
- \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi_k(\sigma) \, d\pi(\gamma),
\]

applying the triangle inequality the above (22) is less than or equal to
\[
\int_{\Omega} \left| \int_{\Omega} K(\gamma, \sigma) \, d\pi_k(\sigma) - \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \right| \, d\pi_k(\gamma) + \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \, d\pi_k(\gamma)
\]
\[
- \int_{\Omega} \int_{\Omega} K(\gamma, \sigma) \, d\pi(\sigma) \, d\pi(\gamma),
\]

(23)

Let us first investigate \(J_k\). Since \(\pi_k \rightarrow \pi\) 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 \(\gamma\) and apply this to \(f(\sigma) := K(\gamma, \sigma)\). So
\[
U(\gamma, \pi_k) \longrightarrow U(\gamma, \pi) \quad \text{as} \quad k \rightarrow \infty.
\]

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

Indeed, using the hypothesis that \(K\) has a \(b\) modulus of continuity and then integrating in \(\sigma\) for any \(\pi\) probability measure on \(\Omega\), we have
\[
|U(\gamma_1, \pi) - U(\gamma_2, \pi)| = \left| \int_{\Omega} K(\gamma_1, \sigma) \, d\pi(\sigma) - \int_{\Omega} K(\gamma_2, \sigma) \, d\pi(\sigma) \right|
\]
\[
\leq \int_{\Omega} |K(\gamma_1, \sigma) - K(\gamma_2, \sigma)| \, d\pi(\sigma)
\]
\[
\leq b(\|\gamma_1 - \gamma_2\|)
\]

(24)

Then \(U(\gamma, \pi)\) is continuous and bounded, so since \(\pi_k \rightarrow \pi\), we conclude that (recall (23)) \(J_k \rightarrow 0\) as \(k \rightarrow \infty\).
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.

$$I_k = \int_{\Omega_{M,N}} \left| U(\gamma, \pi_k) - U(\gamma, \pi) \right| \, d\pi_k(\gamma) + \int_{\Omega \setminus \Omega_{M,N}} \left| U(\gamma, \pi_k) - U(\gamma, \pi) \right| \, d\pi_k(\gamma)$$

$$:= I_{0,k} + I_{1,k}.$$  

Let us look at $I_{0,k}$. We know that $\{U(\gamma, \pi_k)\}$ is equicontinuous. From the last estimate in (24), it is clear that $\{U(\gamma, \pi_k)\}_k$ is equibounded and equicontinuous. Therefore, by Arzela-Ascoli, there exists a subsequence $\{U(\gamma, \pi_{k_j})\}_j$ of $\{U(\gamma, \pi_k)\}_k$ that converges uniformly. Now we claim that

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

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

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

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

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

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

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

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

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

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

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

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

So that then for each fixed $n$,  

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

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

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 $N > 0$,

$$I_{1,k} \leq \pi(\Omega \setminus \Omega_{M,N}) \sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |U(\gamma, \pi_k) - U(\gamma, \pi)| \leq \frac{C}{N} \left( \sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |U(\gamma, \pi_k)| + \sup_{\gamma \in \Omega \setminus \Omega_{M,N}} |U(\gamma, \pi)| \right)$$

$$\leq \frac{2\varepsilon}{3} \quad \text{for all} \quad k.$$  

Putting all this together in (25) and hence in (23) yields
\[ I_k \leq \frac{\varepsilon}{3} + 2 \frac{\varepsilon}{3} = \varepsilon, \quad \text{for all} \quad k \quad \text{sufficiently large}. \]

The lemma is now proved. \hfill \Box

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

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

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

\[ |e^{-|x|^2} - e^{-|y|^2}| \leq C|x - y|. \]

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

4.2. The dual problem. For the rest of the paper we only consider the interactions given by

\[ K(\gamma, \sigma) := \theta \int_0^1 e^{-\beta |\gamma(t) - \sigma(t)|^2} dt, \quad \text{for some} \quad \theta, \beta > 0. \]

In order for the results here to be as “smooth” as possible, we will make a small notational change to the functional \( E(\pi), (7) \).

Namely, without losing generality,

\[ E(\pi) := \int_{\Omega} c(\gamma) d\pi(\gamma) + \int_{\Omega} 2U(\gamma, \pi) d\pi(\gamma). \]

(26)

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

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 \( \int_{\Omega} \left( 2 \int_{\Omega} K(\gamma, \sigma) d\pi(\sigma) \right) d\pi(\gamma) \) is actually quadratic with respect to \( \pi \). 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 \( (\phi, \psi) \in C^0_c(X) \times C^0_c(X) \) and \( \pi \in M^+(\Omega) \), define

\[
L(\pi; \phi, \psi) := \int_{\Omega} c(\gamma) d\pi(\gamma) + \int_{\Omega} 2U(\gamma, \pi) d\pi(\gamma) - \int_{\Omega} \phi(\gamma(0)) d\pi(\gamma) + \int_{X} \phi(x) d\mu_0(x) \\
- \int_{\Omega} \psi(\gamma(1)) d\pi(\gamma) + \int_{X} \psi(y) d\mu_1(y).
\]
After rearranging
\[ L(\pi; \phi, \psi) = \int_\Omega \left( c(\gamma) + 2U(\gamma, \pi) - (\phi(\gamma(0)) + \psi(\gamma(1))) \right) d\pi(\gamma) \]
\[ + \int_X \phi(x)d\mu_0(x) + \int_X \psi(y)d\mu_1(y) \]
\[ = Q(\pi; \phi, \psi) + \int_X \phi(x)d\mu_0(x) + \int_X \psi(y)d\mu_1(y), \]
where
\[ Q(\pi; \phi, \psi) := \int_\Omega c(\gamma) + 2U(\gamma, \pi) - (\phi(\gamma(0)) + \psi(\gamma(1))) \ d\pi(\gamma). \]

Notice that \( Q(\pi; \phi, \psi) \) contains the quadratic term that we talked about in the prequel. Then the dual function is
\[ D(\phi, \psi) := \inf_{\pi \geq 0} \left\{ Q(\pi; \phi, \psi) + \int_X \phi(x)d\mu_0(x) + \int_X \psi(y)d\mu_1(y) \right\}. \tag{27} \]

The business at hand is to minimize the quadratic term \( Q(\pi; \phi, \psi) \) with respect to \( \pi \). But how can one minimize such quadratic term in \( \pi \)? 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 \( Q(\pi; \phi, \psi) \) has the form: \( \min_{p \geq 0} q(p) \) for
\[ q(p) := \langle b, p \rangle + \langle Ap, p \rangle, \tag{28} \]
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 \eqref{27}, that is, the term \( \int_\Omega U(\gamma, \pi)d\pi(\gamma) \). Further, \( p \) represents \( \pi \), while \( Ap \) represents \( 2U(\gamma, \pi) \), and \( b \) represents \( c(\gamma) - \phi(\gamma(0)) - \psi(\gamma(1)) \). Then
\[ \langle b, p \rangle := \langle c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1))), \pi \rangle = \int_\Omega \left( c(\gamma) - (\phi(\gamma(0)) + \psi(\gamma(1)) \right) d\pi(\gamma); \]
\[ \langle Ap, p \rangle := \left\langle 2U(\gamma, \pi), \pi \right\rangle = \int_\Omega 2\left(U(\gamma, \pi) \right) d\pi(\gamma). \]

Thanks to Fubini and Tonelli the inner products above make (kind of) sense and their sum equal \( Q(\pi; \phi, \psi) \). Then to minimize \( Q(\pi; \phi, \psi) \) with respect to \( \pi \geq 0 \) in \( P(\Omega) \), it suffices to minimize \eqref{28} 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 \cite[Ch 10.2.2]{19},
\[ \nabla_p q(p) \geq 0 \text{ (coordinate wise),} \]
\[ \nabla_p q(p) = 0 \text{ if } p_i > 0. \]
Let us interpret this in our infinite dimensional problem, here $\pi_{\text{min}}$ corresponds to the $p_{\text{min}}$, and we have

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

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

Rearranging, this can be written as

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

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

Thus, we have a good guess about the condition that characterizes the optimal $\pi$ in (27). More concretely, and mimicking what we have already done in Section 3.2, if there exists a path $\gamma$ for which $c(\gamma) + 2U(\gamma, \pi) - (\phi(\gamma(0)) + \psi(\gamma(1))) < 0$, then we can make $\min_{\pi \geq 0} Q(\pi; \phi, \psi) = -\infty$ and in turn make $D(\phi, \psi) = -\infty$. Otherwise, for all $\sigma \in \Omega$, and keeping $\gamma$ fixed (remembering the proof of Proposition 3.5), $\phi(\gamma(0)) + \psi(\gamma(1)) \leq c(\gamma) + 2U(\gamma, \pi) \text{ a.e.}$

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

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

(31)

For the moment let us assume that Theorem 3.7's equality holds for (26) having constraint (29). Suppose the following holds:

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

(32)

Suppose $(\phi_0, \psi_0)$ maximizes (31) and $\pi_0$ is optimal for (26). Then, thanks to (29-30) $c(\gamma) + 2U(\gamma, \pi_0) = \phi_0(\gamma(0)) + \psi_0(\gamma(1))$. Furthermore, strong duality (32) ensures

$$\int_\Omega [c(\gamma) + 2U(\gamma, \pi_0)] \, d\pi_0(\gamma) = \int_X \phi_0(x) \, d\mu_0(x) + \int_X \psi_0(y) \, d\mu_1(y)$$

$$= \int_\Omega [\phi_0(\gamma(0)) + \psi_0(\gamma(1))] \, d\pi_0(\gamma).$$

Equivalently

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

The integrand is nonnegative by (29), so it has to vanish $\pi_0$-a.e. Moreover, if we further stipulate something very similar like the endpoint cost function (9) to $c(\gamma) + 2U(\gamma, \pi_0)$, say (35) for which the constraint $\phi_0(x) + \psi_0(y) \leq \inf_{\gamma(0) = x, \gamma(1) = y} c(\gamma) + 2U(\gamma, \pi_0)$ holds $\pi_0$-a.e., then we would have

$$c(\gamma) + 2U(\gamma, \pi_0) = \phi_0(\gamma(0)) + \psi_0(\gamma(1))$$

$$= \phi_0(x) + \psi_0(y)$$

$$\leq \inf_{\gamma(0) = x, \gamma(1) = y} [c(\gamma) + 2U(\gamma, \pi_0)],$$

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

With this in mind, we state the interaction analogue of Theorem 3.7, i.e., we establish (32), 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,
\[
\inf_{\pi \in \Pi_{\text{path}}(\mu_0, \mu_1)} \int_{\Omega} \left( c(\gamma) + 2\, U(\gamma; \pi) \right) d\pi = \sup_{\phi, \psi} D(\phi, \psi),
\]

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 [16, §23]. Namely, we use the fact that \( x \in \text{supergradient of } f \) if and only if 0 is in the supergradient of \( f \).

From [27], let \( q(\phi, \psi) := \inf_{\pi \geq 0} Q(\pi; \phi, \psi) \). So that for all \( \phi, \psi \), there is a certain \( \pi_* \geq 0 \) such that,
\[
q(\phi, \psi) = Q(\pi_*; \phi, \psi) \leq Q(\pi; \phi, \psi), \quad \forall \pi \geq 0.
\]

In this case, from equation [27],
\[
D(\phi, \psi) := \inf_{\pi \geq 0} L(\pi; \phi, \psi) = \inf_{\pi \geq 0} Q(\pi; \phi, \psi) + J(\phi, \psi) = q(\phi, \psi) + J(\phi, \psi),
\]
where \( J(\phi, \psi) = \int \phi \, d\mu_0 + \int \psi \, d\mu_1 \). It is immediate to see that \( J(\phi, \psi) \) is linear in \( \phi, \psi \). For each fixed \( \pi \geq 0 \), \( Q(\pi; \phi, \psi) \) is linear in \( \phi, \psi \) and \( q(\phi, \psi) \) is a minimum of a family of linear functionals. In other words, for all \((\phi_0, \psi_0)\) we can find \( \pi_0 \) such that \( q(\phi_0, \psi_0) = Q(\pi_0; \phi_0, \psi_0) \) and \( q(\phi_0, \psi) \leq Q(\pi; \phi_0, \psi_0) \). Thus, \( q \) is concave in \( \phi, \psi \). The superdifferentiability analogue of \( D(\phi, \psi) \) is given—just as in [16, §23].

Remark 4.5. The superdifferential of \( D \) is made of functions of the form
\[
I_{\pi}(\phi, \psi) := - \int_{X} [\phi(\gamma(0)) + \psi(\gamma(1))] \, d\pi(\gamma) + \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y),
\]
for some \( \pi \geq 0 \). In particular, \( I_{\pi} \) is in \( \partial D(\phi_0, \psi_0) \) if \( D(\phi, \psi) \leq D(\phi_0, \psi_0) + I_{\pi}(\phi - \phi_0, \psi - \psi_0) \) \( \forall (\phi, \psi) \).

Notice the pair \((\phi_0, \psi_0)\) is a maximum of \( D \) if and only if the 0 functional is in \( \partial D(\phi_0, \psi_0) \), i.e., if there is a \( \pi \) such that \( I_{\pi}(\phi, \psi) = 0 \) for all \((\phi, \psi)\).

This aligns perfectly well with Rockafellar’s statement [16]: \( \pi \)—having the form \( I_{\pi} \)—is in the maximization set of \( D(\phi, \psi) \) for all \((\phi, \psi)\) if and only if 0 is in the supergradient of \( D \). That is, if and only if \( I_{\pi} \) is identically zero.

A consequence of this indicates that if there exists \( \pi \geq 0 \) such that for all \( \phi, \psi \)
\[
0 = I_{\pi} = - \int_{\Omega} [\phi(\gamma(0)) + \psi(\gamma(1))] \, d\pi(\gamma) + \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y),
\]
then, \((e_0)_{\pi} = \mu_0 \) and \((e_1)_{\pi} = \mu_1 \) and thusly \( \pi \) is admissible.

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

Proof of Theorem 4.4. Let \((\phi_*, \psi_*)\) be the maximum of \( D(\phi, \psi) \). In view of the above discussion (and Remark 4.5), there exists \( \pi_* \geq 0 \) such that \( I_{\pi_*} \in \partial D(\phi_*, \psi_*) \) and \( I_{\pi_*} = 0 \). Equivalently, \( q(\phi_*, \psi_*) = Q(\pi_*; \phi_*, \psi_*) \). In particular, \( \pi_* \in \Pi_\text{path}(\mu_0, \mu_1) \). Then, for all \((\phi, \psi)\) and all admissible \( \pi \), and applying the inequality (33) we have
\[
\int_{\Omega} c(\gamma) + 2 \mathcal{U}(\gamma, \pi_*) \, d\pi_*(\gamma) = \mathcal{L}(\pi_*, \phi_*, \psi_*)
\]
\[
= \mathcal{Q}(\pi_*; \phi_*, \psi_*) + \int_X \phi_*(x) \, d\mu_0(x) + \int_X \psi_*(y) \, d\mu_1(y)
\]
\[
\leq \mathcal{Q}(\pi; \phi_*, \psi_*) + \int_X \phi_*(x) \, d\mu_0(x) + \int_X \psi_*(y) \, d\mu_1(y)
\]
\[
= \int_{\Omega} c(\gamma) + 2 \mathcal{U}(\gamma; \pi_*) \, d\pi(\gamma) \quad \forall \pi \text{ admitable}
\]

The latter inequality indicates \(\pi_*\) minimizes \(\int (c(\gamma) + 2 \mathcal{U}(\gamma; \pi)) \, d\pi\). So since \(\pi_*\) is optimal, then
\[
\phi_*(\gamma(0)) + \psi_*(\gamma(1)) = c(\gamma) + 2 \mathcal{U}(\gamma; \pi_*), \quad \text{in the support of } \pi_*.
\]

Therefore,
\[
\int_X \phi_*(x) \, d\mu_0(x) + \int_X \psi_*(y) \, d\mu_1(y) = \int_{\Omega} (\phi_*(\gamma(0)) + \psi_*(\gamma(1))) \, d\pi_*(\gamma)
\]
\[
= \int_{\Omega} (c(\gamma) + 2 \mathcal{U}(\gamma; \pi_*)) \, d\pi_*(\gamma)
\]
\(\pi_*\)-a.e. But \((\phi_*, \psi_*)\) solves the maximization \([27]\): thus
\[
\sup \mathcal{D}(\phi, \psi) = \inf_{\pi \in \Pi_{\text{path}}} \mathcal{E}(\pi).
\]

\begin{remark}
The above result says that any solution \((\phi_*, \psi_*)\) to the dual problem produces a solution \(\pi_*\) to the primal problem.
\end{remark}

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

\begin{theorem}
If \(\pi_0\) solves the primal problem, then \((\phi_0, \psi_0)\) solves the auxiliary dual problem.
\end{theorem}

\begin{proof}
First of all using the auxiliary dual problem \([31]\), \(D_\pi(\phi, \psi) \leq \mathcal{E}(\pi)\) for all \((\phi, \psi)\); for all admissible \(\pi\). For the reverse inequality, we proceed as follows. Let \(\pi_0\) be some minimizer for the primal problem. Define \((\phi_0, \psi_0)\) maximizing \([31]\). From the previous Section 4.2, any pair \((\phi_0, \psi_0)\) that achieves the supremum are such that
\[
\phi_0(\gamma(0)) + \psi_0(\gamma(1)) = c(\gamma) + 2 \mathcal{U}(\gamma; \pi_0) \quad \forall \gamma \in \text{spt}(\pi_0).
\]
This implies, as \(\pi_0 \in \Pi_{\text{path}}(\mu_0, \mu_1)\),
\[
D_\pi(\phi_0, \psi_0) = \int_X \phi_0(x) \, d\mu_0(x) + \int_X \psi_0(y) \, d\mu_1(y)
\]
\[
= \int_{\Omega} (\phi_0(\gamma(0)) + \psi_0(\gamma(1))) \, d\pi_0(\gamma)
\]
\[
= \int_{\Omega} (c(\gamma) + 2 \mathcal{U}(\gamma; \pi_0)) \, d\pi_0(\gamma)
\]
\[
= \mathcal{E}(\pi_0).
\]

The second to last equation follows from \([34]\).
Proposition 4.8. Fix $s$ sufficiently close to the optimal path of $c_\theta$ for some $\theta > 0$ and thanks to (34) the first integral vanishes (and so $Q(\pi_0; \phi, \psi)$ vanishes as well), and thus

$$L(\pi_0; \phi, \psi) = \int_X \phi(x) \, d\mu_0(x) + \int_X \psi(y) \, d\mu_1(y)$$

$$= D_\pi(\phi_0, \psi_0).$$

Hence, from (27), we have

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

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

4.4. The effective cost. Toward the end of Section 2 we introduced the endpoint function (9) 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 $\pi$, define $c_0(\gamma) := c(\gamma) + 2U(\gamma; \pi)$ as the effective cost. The analogue of the endpoint cost function (9) is

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

In what follows we deal explicitly with

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

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

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 $\theta$ 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,

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

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

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

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

$$\int_\Omega e^{-|\gamma(t) - \sigma(t)|^2} \, d\pi(\sigma) = \int_{\mathbb{R}^n} e^{-|\gamma(t) - z|^2} \, d\pi_t(z),$$

for all $\gamma(t)$ and all $t \in [0, 1]$; so for every $\pi$, 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 $\varphi(0) = \varphi(1)$. Then we have

$$\varphi(0) = \varphi(1) = x, \quad \varphi(t) = x + \int_0^t \dot{\varphi}(s) \, ds$$

for all $t \in [0, 1]$. Therefore, the solution to the Euler-Lagrange equation satisfies

$$\varphi(0) = \varphi(1) = x, \quad \varphi(t) = x + \int_0^t \dot{\varphi}(s) \, ds$$

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

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

for all \( \varphi \in C^2([0,1]) \) for all \( 0 \leq t \leq 1 \). Hence, \( \gamma_0 \) 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 \( r := |x - z| \),

\[
\left| e^{-|x-z|^2}(x - z) \right| \leq e^{-r^2} r \leq 1.
\]

Then

\[
|\dot{\gamma}_0(t)| \leq 4\theta \int_{\mathbb{R}^n} \left| e^{-|\gamma_0(t)-z|^2}(\gamma_0(t) - z) \right| d\pi_t(z) \leq 4\theta.
\]

So thanks to Proposition 2.1 we have that \( \gamma_0 \) is of the form \( \gamma_0(t) = \gamma_{\text{euc}}(t) + \theta E(\theta, t) \). \( \square \)

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

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

\[
\nabla_y c_{\text{eff}}(x, y) = x - y - 4\theta \int_0^1 \int_{\mathbb{R}^n} t \exp\left\{ -|\gamma_0(t) - z|^2 \right\} (\gamma_0(t) - z) d\pi_t(z) dt.
\]

**Proof.** The proof of Lemma 2.5 can be incorporated here with \( c_e \) replaced by \( c_\varepsilon \). \( \square \)

We now provide a Lipschitz bound on \( T_z(w) := e^{-|w-z|^2}(w - z) \) that will help us prove Lemma 4.1.

**Proposition 4.10.** Given \( w, u, z \in \mathbb{R}^n \), we have

\[
|T_z(w) - T_z(u)| \leq n^2|w - u|.
\]

**Proof.** We firstly bound a partial derivative of \( T_z(w) \). We have

\[
\partial_{w_j} T_z^i(w) = e^{-|w-z|^2}(\delta_{ij} - 2(w_i - z_i)(w_j - z_j)).
\]

For radial \( r := |w - z| \), we get an upper bound,

\[
\left| \partial_{w_j} T_z^i(w) \right| \leq e^{-|w-z|^2}(1 + |w - z|^2) \leq e^{-r^2}(1 + r^2) \leq 1.
\]

Secondly, borrowing notation and calculations from Spivak’s *Calculus on Manifolds* [20], we observe that

\[
T_z^i(w) - T_z^i(u) = \sum_{j=1}^n \left[ T_z^i(w_1, \ldots, w_j; u_{j+1}, \ldots, u_n) - T_z^i(w_1, \ldots, w_{j-1}; u_j, \ldots, u_n) \right].
\]

As \( T_z(w) \) is continuously differentiable, the mean value theorem gives us for every \( i \) and \( j \)

\[
T_z^i(w_1, \ldots, w_j; u_{j+1}, \ldots, u_n) - T_z^i(w_1, \ldots, w_{j-1}; u_j, \ldots, u_n) = (w_j - u_j) \partial_{w_j} T_z^i(v_{ij}),
\]

for some \( v_{ij} \). The absolute value of the right-hand side of this is \( |w_j - u_j| |\partial_{w_j} T_z^i(v_{ij})| \leq |w_j - u_j| \). Then \( |T_z^i(w) - T_z^i(u)| \leq n|w_j - u_j| \leq n|w - u| \), since each \( |w_j - u_j| \leq |w - u| \). Hence,
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 $x_1 \neq x_2$,

$$\nabla_y \left( c_{\text{eff}}(x_2, y) - c_{\text{eff}}(x_1, y) \right) \neq 0.$$

Remark 4.12. The above holds independent of $\pi$.

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 $c_0(\gamma_i)$; 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 $y$

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

Thanks to Proposition [4.10] we can estimate the expression inside the brackets. Namely, since we saw $|\gamma_0(t)| \leq 4\theta$, Proposition [2.1] applies to give

$$|T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))| \leq n^2 |\gamma_{\theta,2}(t) - \gamma_{\theta,1}(t)|$$

$$\leq n^2 \max_{0 \leq t \leq 1} |\gamma_{\theta,2}(t) - \gamma_{\theta,1}(t)|$$

$$\leq n^\frac{5}{2} \theta |x_2 - x_1|.$$

In particular,

$$\left| 4\theta \int_0^1 \int_{\mathbb{R}^n} t |T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))| d\pi_t(z) dt \right| \leq 4\theta \int_0^1 \int_{\mathbb{R}^n} n^\frac{5}{2} \theta |x_2 - x_1| d\pi_t(z) dt$$

$$= 2n^\frac{5}{2} \theta^2 |x_2 - x_1|.$$

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

$$|\nabla_y \left( c_{\text{eff}}(x_2, y) - c_{\text{eff}}(x_1, y) \right)| \geq |x_1 - x_2| - 4\theta \int_0^1 \int_{\mathbb{R}^n} |T_z(\gamma_{\theta,2}(t)) - T_z(\gamma_{\theta,1}(t))| d\pi_t(z) dt$$

$$\geq |x_1 - x_2| - 2n^\frac{5}{2} \theta^2 |x_2 - x_1|$$

$$= (1 - 2n^{\frac{5}{2}} \theta^2) |x_1 - x_2| > (1 - \delta) |x_1 - x_2|.$$ 

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

Theorem 1.10 now follows as a corollary of Lemma 4.11 and Theorem 1.5.

Proof of Theorem 1.10. To prove this theorem, we must show that $c_{\text{eff}}$ satisfies the assumptions of Brenier and Gangbo and McCann’s theorems or Theorem 10.28 from [22]. Remark 4.12 tells that we can write the effective cost as $c_{\text{eff}, \pi_{\text{min}}}$. This $\pi_{\text{min}}$ will help us attain $c_{\text{eff}, \pi_{\text{min}}}$ implicitly depending on $\pi_{\text{min}}$. Pick $\pi_{\text{min}}$ by Theorem 1.8. Then $\pi_{\text{min}}$ is optimal with respect to $c_{\text{eff}, \pi_{\text{min}}}$. Since $\theta < \theta_0$, Lemma 4.11 tells us $\nabla_y c_{\text{eff}, \pi_{\text{min}}}$ is injective, and Lemma 9 tells us $c_{\text{eff}, \pi_{\text{min}}}$ is differentiable, and since $\mu_0 \ll dx$, then $c_{\text{eff}, \pi_{\text{min}}}$ satisfies the assumptions of Brenier and Gangbo and McCann’s results. Therefore, there is a unique transport map $T$ solving Problem B.

The proof of Theorem 1.9 now follows as a corollary from Theorems 4.4, 1.8 and, 1.10. Indeed,
Proof of Theorem 1.9. From the previous results stated above, one has the following. If $\pi_0$ is a minimizer of $\mathcal{E}_0(\pi)$, then $\pi_0$ is a minimizer of $\mathcal{E}(\pi)$ without the interaction term containing $c_0(\gamma)$.

\[\square\]

5. Acknowledgments

My deepest gratitude goes to Nestor Guillen for bringing optimal transport to light! And for introducing this problem to me and for all his support and guidance throughout this research project. I would also like to thank Jun Kitagawa for reading a preliminary version of this note and making me aware of a standard construction in optimal transport known as dynamical couplings.
APPENDIX A. BOCHNER’S THEOREM

A goal of this appendix is to demonstrate that the functional (7) with the Coulomb potential, $\kappa(x-y) := |x-y|^{2-n}$, is convex in $\pi$, 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 $\pi$. In order to establish this we rely on Bochner’s theorem as presented in Reed’s and Simon’s book [15]. (see Theorem A.1 below).

The generalized version of Bochner’s theorem [15 Theorem IX.9], due to Schwartz [18], which includes distributions, is an extension of functions of positive type to distributions (see definition A.2).

**Theorem A.1.** [15 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. Suppose $f(x)$ is a bounded, continuous function. Then $f(x)$ is said to be of positive type if

$$\int \int f(x-y)\varphi(y)\varphi(x) \, dx \, dy \geq 0$$

for all $\varphi \in C_0^\infty(\mathbb{R}^n)$, [15]. The condition above can be rewritten in the following way if we consider the convolution $(\tilde{\varphi} * \varphi)(\tau) := \int \tilde{\varphi}(x-\tau)\varphi(x) \, dx$, namely,

$$\int \int f(\tau)\varphi(x-\tau)\varphi(x) \, d\tau \, dx = \int f(\tau)(\tilde{\varphi} * \varphi)(\tau) \, d\tau \geq 0,$$

where $\tilde{\varphi}(x) = \varphi(-x)$. Then we have the definition:

**Definition A.2.** A distribution $T \in \mathcal{D}'(\mathbb{R}^n)$ is said to be of positive type if $T(\tilde{\varphi} * \varphi) \geq 0$ for all $\varphi \in \mathcal{D}(\mathbb{R}^n)$.

We shall apply this to show the quadratic function given by a $\kappa$ of positive type,

$$Q(\pi) := \int_{\Omega} \int_{\Omega} \int_{0}^{1} \kappa(\gamma(t) - \sigma(t)) \, dt \, d\pi(\sigma) \, d\pi(\gamma)$$

is convex in $\pi$. It suffices to show, for $\pi_0$ 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,

$$\frac{d^2}{ds^2}Q(\pi_s) = 2 \int_{0}^{1} \int_{\Omega} \int_{\Omega} \kappa(\gamma(t) - \sigma(t)) d(\pi_0(\sigma) - \pi_1(\sigma)) \, d\pi(\gamma) \, d\pi(\sigma) \, dt.$$ 

Let $\omega = \pi_0 - \pi_1$. Let $d\omega_t = (e_t) \omega$. Then, up to a factor of 2, this latter integral is equal to

$$\int_{0}^{1} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y) d\omega_t(x) d\omega_t(y) \, dt.$$ 

Since the measures $\omega$ can be approximated under the weak-* topology by finite measures against $C_0^\infty(\mathbb{R}^n)$ test functions, given finite measures $\rho \in \mathcal{M}(\mathbb{R}^n)$ with $\rho = \varphi(x) dx$ for all $\varphi \in C_0^\infty(\mathbb{R}^n)$, the above integral can be approximated by

$$\int_{0}^{1} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y) \rho(x) \rho(y) \, dx \, dy \, dt = \int_{0}^{1} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa(x-y) \varphi(x) \varphi(y) \, dx \, dy \, dt.$$
Observe \( \kappa(x) = C_n |x|^{2-n} \) has Fourier transform equal to \(|\xi|^{-2} \geq 0\). From here one can see \( \kappa \) is a distribution of the Fourier transform of a positive measure with polynomial growth, so by Bochner the above is \( \geq 0 \). This proves the statement at the start of this appendix.

It is worthwhile to mention an alternative proof using integration by parts. Let us consider the fundamental solution of the Laplace equation for \( n \geq 3 \):

\[
\Phi(x) = C_n|x|^{2-n}, \quad C_n \geq 0 \text{ a dimensional constant,}
\]

(see Evans [6, Ch 2]). Considering convolutions, for \( f \in C^2_c(\mathbb{R}^n) \), we can write

\[
u(x) = \int_{\mathbb{R}^n} \Phi(x - y)f(y) \, dy = C_n \int_{\mathbb{R}^n} |x - y|^{2-n}f(y) \, dy.
\]

Thus [6] Ch 2], \( u \in C^2_c(\mathbb{R}^n) \) and it 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 (7) is convex in \( \pi \), that is \( \frac{d^2}{ds^2} Q(\pi_s) \geq 0 \). More concretely,

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^{2-n} \phi(x) \phi(y) \, dx \, dy = -\int_{\mathbb{R}^n} \Delta \phi(x) \int_{\mathbb{R}^n} |x - y|^{2-n} \phi(y) \, dy \, dx
\]

\[
= -C_n \int \Delta \phi(x) \phi(x) \, dx
\]

\[
= C_n \int |\nabla \phi|^2 \, dx \geq 0,
\]

where we applied integration by parts to get the last equation.
Appendix B. The differentiability of the end point cost function

In this appendix we prove the following limit exists:

$$\lim_{h \to 0} \frac{c_e(x, y + h \hat{e}) - c_e(x, y)}{h}. \quad (37)$$

This stems from observing if $\gamma_{x,y}(t)$ is the minimal path from $x := \gamma_{x,y}(0)$ to $y := \gamma_{x,y}(1)$, $c_e(x, y) = c(\gamma_{x,y}(t))$ and, if we momentarily suppose that the above limit exists, then said limit will equal the left-hand side of the following equation:

$$\nabla_y c_e(x, y) = \frac{\partial c}{\partial y} \nabla_y \gamma_{x,y},$$

by the chain rule. Indeed,

$$\nabla_y (c_e(x, y)) = \int_0^1 \dot{\gamma}_{x,y}(t) \cdot \nabla_y (\dot{\gamma}_{x,y}(t)) \, dt,$$

and since

$$\nabla_y (\dot{\gamma}_{x,y}(t)) = \nabla_y \left( \frac{d}{dt} \gamma_{x,y}(t) \right) = \frac{d}{dt} \nabla_y (\gamma_{x,y}(t)),$$

to show the limit (37) exists it suffices to show the following limit exists:

$$\lim_{h \to 0} \frac{\gamma_{x,y+h \hat{e}}(t) - \gamma_{x,y}(t)}{h}.$$  

Lemma B.1. The end point cost function $c_e(x, y) := \inf_{\gamma(0)=x, \gamma(1)=y} c(\gamma)$ is differentiable with respect to $y$. Moreover,

$$\nabla_y c_e(x, y) = \lim_{h \to 0} \frac{c_e(x, y + h \hat{e}) - c_e(x, y)}{h}.$$  

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” argument of compactness plus uniqueness of linear ODEs.

Thanks to Proposition 2.2, $\gamma_{x,y+h \hat{e}}(t)$ solves

$$\dot{\gamma}_{x,y+h \hat{e}}(t) = -\nabla V(\gamma_{x,y+h \hat{e}}(t), t).$$

Then $\theta_h(t)$ solves the boundary value problem:

$$\ddot{\theta}_h(t) = -\frac{1}{h} \left( \nabla V(\gamma_{x,y+h \hat{e}}(t), t) - \nabla V(\gamma_{x,y}(t), t) \right),$$

$$\theta_h(0) = 0,$$

$$\theta_h(1) = \hat{e}.$$

$^3$This is commonly used, for instance, in the context of viscosity solutions.
Assuming $V \in C^2(X)$, Taylor expanding about $h$, we get
\[
\ddot{\theta}_h(t) = -D^2V(\gamma_{x,y}(t), t)\theta_h(t) + o_h(1)(t), \tag{38}
\]
\[
\theta_h(0) = 0, \tag{39}
\]
\[
\theta_h(1) = \hat{e}. \tag{40}
\]

Proposition 2.4 tells us that $\theta_h(t)$ is equibounded:
\[
\|\theta_h(t)\|_{\infty} \leq \frac{1}{1-L} =: C_0 \quad \forall \ h, t.
\]

Since $\nabla V$ is $L$-Lipschitz, and another application of Proposition 2.4, 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 $\left\|\ddot{\theta}_h(\tau)\right\|_{\infty} \leq 2 C_0$. 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 $\theta_h$ is equicontinuous. We are ready to prove the existence of the limit in $\gamma$. 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 $\{\theta_{h_{k_j}}\}$ uniformly converging, as $k \to \infty$, to a continuous function $\theta : [0, 1] \to X$. From here, using (38), one can show $\theta$ is $C^2([0, 1])$ and uniquely solves
\[
\ddot{\theta}(t) = -D^2V(\gamma_{x,y}(t), t)\theta(t)
\]
\[
\theta(0) = 0,
\]
\[
\dot{\theta}(0) = \hat{e}.
\]
References

[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] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. Optimal transportation networks: models and theory. Springer, 2008.
[4] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Math., 44(4):375–417, 1991.
[5] 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.
[6] Lawrence C Evans. Partial differential equations. Graduate studies in mathematics, 19(2), 1998.
[7] 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.
[8] Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. Acta Mathematica, 177(2):113–161, 1996.
[9] Nestor Guillen and Jun Kitagawa. On the local geometry of maps with e-convex potentials. Calculus of Variations and Partial Differential Equations, 52(1):345–387, 2015.
[10] Ryan Hynd. Sticky particles and the pressureless euler equations in one spatial dimension. arXiv preprint arXiv:1804.09340, 2018.
[11] Leonid V Kantorovich. Mathematical methods of organizing and planning production. Management science, 6(4):366–422, 1960.
[12] Leonid V Kantorovich. On the translocation of masses. Journal of mathematical sciences, 133(4):1381–1382, 2006.
[13] 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.
[14] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. Histoire de l’Académie Royale des Sciences de Paris, 1781.
[15] Michael Reed, Barry Simon, Barry Simon, and Barry Simon. Methods of modern mathematical physics, volume 1. Elsevier, 1972.
[16] Ralph Tyrell Rockafellar. Convex analysis. Princeton university press, 2015.
[17] Filippo Santambrogio. Optimal transport for applied mathematicians. Birkhäuser, NY, 55(58-63):94, 2015.
[18] Laurent Schwartz. Théorie des distributions à valeurs vectorielles. i. In Annales de l’institut Fourier, volume 7, pages 1–141, 1957.
[19] Justin Solomon. Numerical algorithms: methods for computer vision, machine learning, and graphics. CRC press, 2015.
[20] Michael Spivak. Calculus on manifolds: a modern approach to classical theorems of advanced calculus. CRC press, 2018.
[21] Cédric Villani. Topics in optimal transportation. Number 58. American Mathematical Soc., 2003.
[22] Cédric Villani. Optimal transport: old and new, volume 338. Springer, 2009.

(Rene Cabrera) Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305

Email address: cabrera@math.umass.edu