university of groningen

faculty of science and engineering

mathematics and applied mathematics

Aubry-Mather Theory through Optimal Transportation

Bachelor's Project Mathematics

July 2023

Student: F. Voronine

First supervisor: Dr. M. Seri

Second assessor: Dr. A. Sterk

Contents

Chapter 0

Introduction

Optimal transport theory, as the name suggests, is the study of efficient movement and allocation of resources. The field can be traced back to the XVIII$^{\text{th}}$ century where the French mathematician Gaspard Monge aimed to understand how to move soil from the ground with a given density distribution to construct fortifications described by another density distribution so as to minimize the total effort [13]. It turns out that these early ideas play a role in the foundations of the field of linear programming, the study of optimizing linear functions subject to linear constraints. Moreover, those same ideas were rediscovered in the early XX$^{\text{th}}$ by Leonid Kantorovich and Leonid Wasserstein while formalizing the notion of distance between density distributions using tools from functional analysis [23]. Nowadays, the product of these investigations can be applied in a wide range of areas including economics [7], meteorology [14] and machine learning [22] to name a few. While optimal transport theory is rooted in application, it has proven to be equally interesting from a purely mathematical point of view, connecting seemingly disparate fields such as partial differential equations, information theory and gradient flows [9].

Another important field that is closely related to optimal transport theory is the study of dynamical systems. It is known that for integrable Hamiltonian systems, the phase space is foliated by invariant submanifolds that are diffeomorphic to tori whose dynamics are conjugate to rigid rotation [2]. For the near-integrable case, the celebrated KAM theorem (named after Kolmogorov, Arnold and Moser) states that some invariant tori with quasi-periodic dynamics survive when the perturbation is small [16]. A generalization of this notion is captured by Aubry-Mather theory which describes these invariant orbits for any perturbation using the principle of least action. In the context of Tonelli Lagrangians, these invariant orbits are described by the Mather and Aubry sets, which can be seen as solutions to variational problems [6]. These invariant sets provide valuable insight on the resulting dynamics of the perturbed system and are intricately linked to the weak solutions of the Hamilton-Jacobi equation [20].

In Chapter 1, we present the foundations of optimal transport theory following *Introduction to Optimal Transport* by Matthew Thorpe [21]. In particular, we begin by giving a clear formulation of the optimal transport problems considered by Monge and Kantorovich respectively. Both problems can be described as minimization problems where the objective function corresponds to the total cost of transportation. Moreover, we consider the dual Kantorovich problem, a corresponding maximization problem which provides valuable information about the Kantorovich problem. Lastly, we provide sufficient conditions for the existence of a solution to the Monge problem in a Euclidean setting and characterize solutions to the Kantorovich problem in a general setting.

In Chapter 2, we present, without proof, a summary of the results from my work [26] which is based on *Action-Minimizing Methods in Hamiltonian Dynamics* by Alfonso Sorrentino [20]. A statement is presented if it relevant in the explanation of the relationship between optimal transportation and Aubry-Mather theory provided in the bibliographical notes of Chapter 5 in *Optimal Transport: Old and New* by Cédric Villani [24]. We begin by defining Tonelli Lagrangians and Hamiltonians on compact manifolds which provide a robust framework describing many known dynamical systems. Moreover, we consider the corresponding action-minimizing measures and curves which provide information on the Mather and Aubry sets respectively. Lastly, we describe a connection between optimal transport and Aubry-Mather theory using the properties of weak KAM solutions to the Hamilton-Jacobi equation.

I would like to express my deepest gratitude to my supervisor, Marcello Seri, for being an invaluable source of guidance, encouragement, inspiration and patience throughout this bachelor project. I am immensely grateful for the opportunity to engage in thought-provoking conversations in the goal of understanding optimal transportation and Aubry-Mather theory. I warmly extend my gratitude to Alef Sterk for being an inspiring teacher interested in becoming my second supervisor. Lastly, I would like to thank my friends and family for their love and support without which this enterprise would not be possible.

Chapter 1

Optimal Transportation

This chapter is based on *Introduction to Optimal Transport* by Matthew Thorpe [21] and *Optimal Transport: Old and New* by Cédric Villani [24].

1.1 Formulating the Optimal Transport Problem

1.1.1 The Monge Problem

The field of optimal transportation can be traced back to the XVIII$^{\text{th}}$ century where the French mathematician Gaspard Monge set out to determine the most efficient manner of transporting mass from the ground to construct fortifications during the Napoleonic wars. In a modern mathematical framework, the initial configuration of mass in the ground can be represented using a probability measure μ on a measure space X, whereas the final desired configuration can be seen as a probability measure ν on a measure space Y. Given measurable sets $A \subseteq X$ and $B \subseteq Y$, the quantity $\mu(A)$ represents the proportion of the mass contained in the set A in the initial configuration whereas $\nu(B)$ represents the proportion of mass that needs to be transported to the set B in the final configuration. Using induced measures, we can combine a map $T : X \to Y$ with a probability measure μ on X to construct a new probability measure $T_{\#}\mu$ on Y. We say that T transports μ to ν if the induced measure $T_{\#}\mu$ coincides with the measure ν.

Definition 1.1. A map $T : X \to Y$ is a transport map from μ to ν if

$$\nu(A) = T_{\#}\mu(A) = \mu(T^{-1}(A)) \quad \text{for all measurable sets } A \subseteq Y.$$

Inherent to the problem is a measurable cost function $c : X \times Y \to [0, +\infty]$ of moving a unit of mass from a point in X to a point in Y.

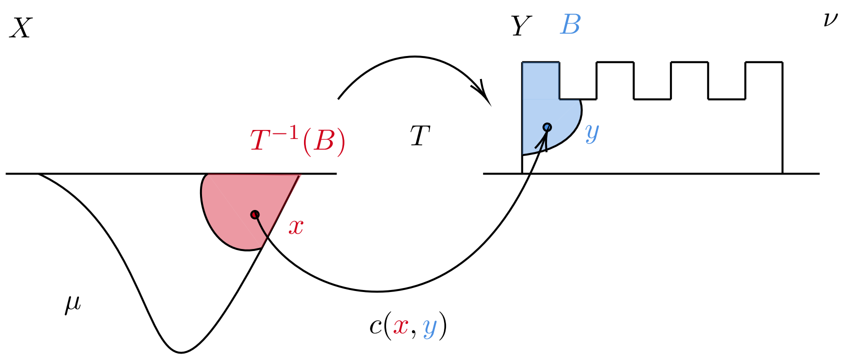


FIGURE 1.1: The setup of the Monge formulation of the Optimal Transport Problem.

Given a transport map $T : X \to Y$, the total cost of transporting μ to ν can be computed by integrating the cost function with respect to μ over the space X, where the second argument in the cost function is replaced by the endpoint $T(x)$.

Definition 1.2. Let $T : X \to Y$ be a transport map from μ to ν. The total cost of transporting μ to ν by T is defined by

$$\mathbb{M}(T) = \int_X c(x, T(x)) \, d\mu(x).$$

To determine the most efficient way of transporting μ to ν, we minimize the total cost of transportation over the set of all transport maps from μ to ν. This minimization problem is known as the Monge problem.

Problem 1.3. Let μ and ν be probability measures on X and Y respectively. The Monge problem asks to find a transport map from μ to ν minimizing the total cost of transportation. That is, find a map $T^\dagger : X \to Y$ satisfying

$$\mathbb{M}(T^\dagger) = \min_{T_\# \mu = \nu} \mathbb{M}(T).$$

Remark 1.4. It may happen that the set of transport maps from μ to ν is empty. To illustrate, consider the Dirac measures

$$\mu = \delta_{x_1} \quad \text{and} \quad \nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2},$$

where $x_1 \in X$ and $y_1, y_2 \in Y$ are distinct. This implies that

$$\nu(\{y_1\}) = \frac{1}{2} \quad \text{but} \quad \mu(T^{-1}(y_1)) \in \{0, 1\}.$$

This means that no transport map can exist from μ to ν. From a different perspective, this can be seen from the fact that transport maps send the entirety of the mass located at x_1 to either y_1 or y_2. As a result, it is impossible to achieve the desired configuration using transport maps since mass is not allowed to split.

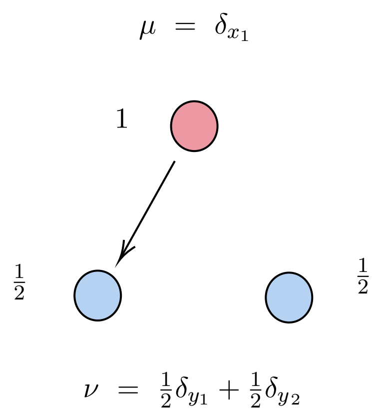


FIGURE 1.2: Non-existence of a transport map from μ to ν.

1.1.2 The Kantorovich Problem

The main drawback of the Monge problem is that mass is not allowed to split. Given a transport map $T : X \to Y$, a point mass located at $x \in X$ will be entirely transported to $T(x) \in Y$. Recall from Remark 1.4 that the Monge problem can lead to the non-existence of a transport map between suitably chosen discrete measures. As a result, we relax the optimal transport problem by allowing mass to split.

Instead of considering transport maps, we consider probability measures π on the product space $X \times Y$ where $\pi(A \times B)$ represents the proportion of the mass that is transported from a measurable set $A \subseteq X$ to a measurable set $B \subseteq Y$. For consistency, the proportion of the mass that leaves a measurable set $A \subseteq X$, which is given by $\pi(A \times Y)$, must coincide with $\mu(A)$. Similarly, the proportion of the mass that gets transported to a measurable set $B \subseteq Y$, which is given by $\pi(X \times B)$, must coincide with $\nu(B)$. We call such measures transport plans from μ to ν.

Definition 1.5. A probability measure π on $X \times Y$ is a transport plan from μ to ν if

$$\pi(A \times Y) = \mu(A) \quad \text{and} \quad \pi(X \times B) = \nu(B) \quad \text{for all measurable sets } A \subseteq X, B \subseteq Y.$$

Notation 1.6. We write $\Pi(\mu, \nu)$ for the set of all transport plans from μ to ν. In addition, we write $P^X : X \times Y \to X$ and $P^Y : X \times Y \to Y$ for the projection onto the space X and Y respectively. By definition, it follows that

$$P^X_\# \pi = \mu \quad \text{and} \quad P^Y_\# \pi = \nu.$$

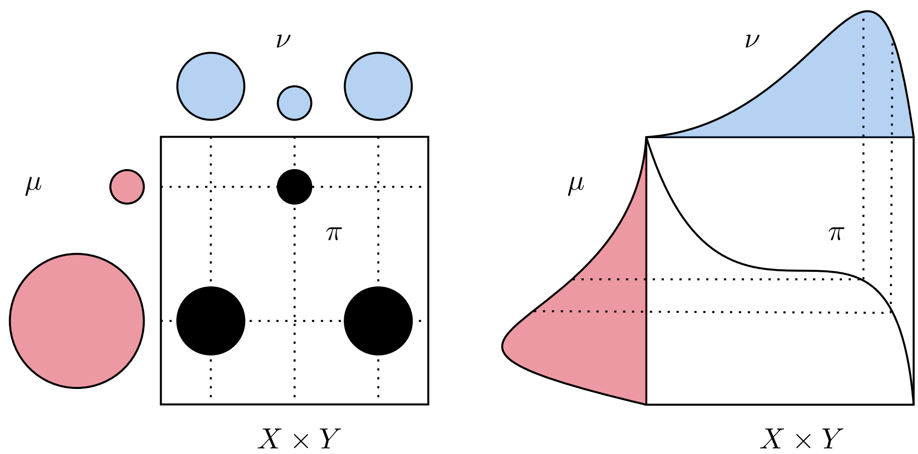


FIGURE 1.3: Transport plans from μ to ν in the case of discrete measures and absolutely continuous measures respectively.

Remark 1.7. In the case of discrete measures of the form

$$\mu = \sum_{i=1}^{m} \alpha_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{j=1}^{n} \beta_j \delta_{y_j},$$

where α_i represents the proportion of the mass initially at x_i and β_j represents the proportion of the mass to be transported to y_j, we have that

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j = 1,$$

and all transport plans from μ to ν satisfy the conditions

$$\mu(x_i) = \sum_{j=1}^{n} \pi(x_i, y_j) \quad \text{and} \quad \nu(y_j) = \sum_{i=1}^{m} \pi(x_i, y_j).$$

In the case where μ and ν are absolutely continuous with respect to the Lebesgue measure, we have that transport plans satisfy

$$\mu(A) = \int_{X\times Y} \mathbb{1}_{A\times Y}\, d\pi \quad \text{and} \quad \nu(B) = \int_{X\times Y} \mathbb{1}_{X\times B}\, d\pi.$$

For an illustration of the conditions satisfied by transport plans, see Figure 1.3.

Remark 1.8. The set of all transport plans $\Pi(\mu, \nu)$ is non-empty since the product measure $\mu \otimes \nu \in \Pi(\mu, \nu)$. This follows from the fact that $\mu \otimes \nu$ is the unique measure on $X \times Y$ satisfying

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B) \quad \text{for all measurable sets } A \subseteq X, B \subseteq Y.$$

Moreover, the set $\Pi(\mu, \nu)$ is convex since, for arbitrary $\pi, \eta \in \Pi(\mu, \nu)$ and $t \in [0, 1]$, we have

$$\begin{aligned} (t\pi + (1-t)\eta)\,(A \times Y) &= t\pi(A \times Y) + (1-t)\eta(A \times Y) \\ &= t\mu(A) + (1-t)\mu(A) \\ &= \mu(A). \end{aligned}$$

Similarly, we obtain our second condition

$$(t\pi + (1-t)\eta)\,(X \times B) = \nu(B),$$

from which we can deduce that $t\pi + (1-t)\eta \in \Pi(\mu, \nu)$.

Given a transport plan $\pi \in \Pi(\mu, \nu)$, the total cost of transporting μ to ν is obtained by integrating the cost function with respect to π over the product space $X \times Y$.

Definition 1.9. Let $\pi \in \Pi(\mu, \nu)$. The total cost of transporting μ to ν is defined by

$$\mathbb{K}(\pi) = \iint_{X\times Y} c(x, y)\; d\pi(x, y).$$

To obtain the most efficient way of transporting μ to ν, we minimize the total cost of transportation over the set of all transport plans from μ to ν. This minimization problem is known as the Kantorovich problem.

Problem 1.10. Let μ and ν be probability measures on X and Y respectively. The Kantorovich problem asks to find a transport plan from μ to ν minimizing the total cost of transportation. That is, find $\pi^\dagger \in \Pi(\mu, \nu)$ satisfying

$$\mathbb{K}(\pi^\dagger) = \min_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi).$$

Remark 1.11. Given a transport map $T : X \to Y$ from μ to ν, we can explicitly construct a corresponding transport plan $\pi \in \Pi(\mu, \nu)$. By the Radon-Nikodym theorem, we can construct a measure π on the product space $X \times Y$ satisfying

$$d\pi(x, y) = \delta_{y=T(x)}\; d\mu(x).$$

This constructed probability measure π is a transport plan from μ to ν since

$$\pi(A \times Y) = \int_A \delta_{T(x)\in Y}\, d\mu(x) = \mu(A),$$
$$\pi(X \times B) = \int_X \delta_{T(x)\in B}\, d\mu(x) = \mu(T^{-1}(B)) = T_{\#}\mu(B) = \nu(B).$$

Moreover, by the Fubini-Tonelli theorem, we have

$$\begin{aligned}\mathbb{K}(\pi) &= \iint_{X\times Y} c(x,y)\, d\pi(x,y)\\ &= \int_X \int_Y c(x,y)\delta_{y=T(x)}\, dy\, d\mu(x)\\ &= \int_X c(x,T(x))\, d\mu(x)\\ &= \mathbb{M}(T).\end{aligned} \tag{1.1}$$

Since the transport maps from μ to ν correspond to a subset of the transport plans from μ to ν, we obtain

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) \le \inf_{T_{\#}\mu=\nu} \mathbb{M}(T). \tag{1.2}$$

It turns out that (1.2) still holds even in the absence of minimizers for the Monge problem. In particular, let $\epsilon > 0$ be arbitrary and suppose that the inequality

$$\mathbb{M}(T^{\dagger}) \le \min_{T_{\#}\mu=\nu} \mathbb{M}(T) + \epsilon,$$

holds for some transport map $T^{\dagger} : X \to Y$. This implies that

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) \le \inf_{T_{\#}\mu=\nu} \mathbb{M}(T) + \epsilon.$$

Since $\epsilon > 0$ was chosen arbitrarily, it follows that (1.2) holds. In the case that the optimal transport plan $\pi^{\dagger}$ satisfies

$$d\pi^{\dagger}(x,y) = \delta_{y=T^{\dagger}(x)}\, d\mu(x),$$

we have that $T^{\dagger}$ is an optimal transport map and equality is achieved in (1.2) using (1.1).

In addition, the Kantorovich problem is more versatile in practice compared to the Monge problem. This can be seen from the fact that Kantorovich problem is a convex optimization problem since the constraints are convex by Remark 1.8 and the cost function is typically convex. Moreover, many practical transportation problems can be modelled using discrete measures.

Example 1.12. Suppose that m factories produce bread which need to be transported to n bakeries so as to minimize the total cost of transportation. We write α_i for the proportion of the total bread produced by factory x_i and β_j for the proportion of the total bread that needs to be sent to bakery y_j. Recall from Remark 1.7 that this situation can be modelled using discrete measures μ and ν of the form

$$\mu = \sum_{i=1}^{m} \alpha_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{i=1}^{n} \beta_j \delta_{y_j},$$

where the coefficients $\alpha_i, \beta_j \geq 0$ satisfy

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j = 1.$$

We write $c_{ij} = c(x_i, y_j)$ for the cost of moving a single loaf of bread from factory x_i to bakery y_j and $\pi_{ij} = \pi(x_i, y_j)$ for the proportion of the total bread that is moved from factory x_i to bakery y_j by a transport plan π. By Remark 1.7, the total proportion of bread that is moved from factory x_i needs to be equal to α_i and the total proportion of bread that is received by bakery y_j needs to be equal to β_j. Thus, we have

$$\pi \in \Pi(\mu, \nu) \quad \text{if and only if} \quad \pi_{ij} \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} \pi_{ij} = \alpha_i \quad \text{and} \quad \sum_{i=1}^{m} \pi_{ij} = \beta_j.$$

In the context of discrete measures, the Kantorovich problem asks to find a transport plan $\pi^\dagger \in \Pi(\mu, \nu)$ such that

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij}^\dagger = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij}.$$

Note that this is a linear programming problem which can be solved using various linear programming algorithms including the simplex method and entropic regularization methods [15, Sections 3.1 and 4.1]. For a concrete example, suppose that we have 3 factories producing bread which need to be transported to 6 bakeries according to the following scheme.

(A) Factories and Bakeries throughout Groningen

Factory	Production
1	300
2	400
3	300

Bakery	Inventory
1	100
2	100
3	100
4	100
5	500
6	100

(B) Production and Inventory

Furthermore, we assume that the cost function is given by

$$(c_{ij}) = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 3 \\ 3 & 3 & 2 & 1 & 1 & 3 \\ 1 & 2 & 3 & 2 & 1 & 1 \end{pmatrix}.$$

By the above discussion, we obtain the probability measures

$$\mu = \sum_{i=1}^{3} \alpha_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{i=1}^{6} \beta_j \delta_{y_j},$$

where

$$\alpha_1 = \alpha_3 = 0.3,\ \alpha_2 = 0.4, \quad \text{and} \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_6 = 0.1,\ \beta_5 = 0.5.$$

The above coefficients are obtained by dividing the production and inventory capacities by 1000, which corresponds to the total bread in circulation in the economy. Moreover, transport plans must satisfy

$$\pi_{ij} \geq 0 \quad \text{and} \quad \sum_{j=1}^{6} \pi_{ij} = \alpha_i \quad \text{and} \quad \sum_{i=1}^{3} \pi_{ij} = \beta_j.$$

The Kantorovich problem asks to find a transport plan $\pi^\dagger \in \Pi(\mu, \nu)$ such that

$$\sum_{i=1}^{3} \sum_{j=1}^{6} c_{ij} \pi^\dagger_{ij} = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i=1}^{3} \sum_{j=1}^{6} c_{ij} \pi_{ij}. \tag{1.3}$$

By applying any linear programming algorithm to (1.3), we obtain a optimal transport plan

$$(\pi^\dagger_{ij}) = \begin{pmatrix} 0.1 & 0.1 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0.1 \end{pmatrix},$$

which means that that the total amount of bread that needs to be sent from each factory to each bakery can be obtained by multiplying the entries of $\pi^\dagger$ by 1000, namely

$$1000 \cdot (\pi^\dagger_{ij}) = \begin{pmatrix} 100 & 100 & 100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 300 & 0 \\ 0 & 0 & 0 & 0 & 200 & 100 \end{pmatrix}.$$

1.1.3 Existence of Minimizers to the Kantorovich Problem

We now prove the existence of an optimal transport plan in the case that the cost function is lower semi-continuous and the spaces X and Y are completely separable metric spaces. Such spaces are called Polish spaces, since they were extensively studied by Polish mathematicians including Sierpiński, Kuratowski and Tarski among others. Before proving the claim, we state without proof some classical results from Functional Analysis and Measure Theory [12, Chapter 2].

Notation 1.13. We write $\mathcal{P}(X)$ for the set of Borel probability measures on X and $C_b^0(X)$ for the set of all bounded continuous functions.

Definition 1.14. A finite Borel measure μ on X is tight if for every $\epsilon > 0$ there exists a compact set $K \subset X$ such that $\mu(X \backslash K) < \epsilon$. A finite tight measure is also called a Radon measure. A set Γ of Borel probability measures is tight if all measures in Γ are tight.

In other words, tight measures are well-approximated from within by compact sets. Moreover, the notion of tightness guarantees that the measures are "compatible" with the topology of the space X. For instance, a measure $\mu \in \mathcal{P}(X)$ may not have a well-defined support [10]. In the context of a completely separable metric space, such pathologies do not occur for finite Borel measures.

Theorem 1.15. If X is a completely separable metric space, then every finite Borel measure on X is tight.

Lastly, we equip the space $\mathcal{P}(X)$ with the weak* topology. In the context of probability measures, the canonical pairing between a probability measure μ and a bounded continuous function f on X is given by

$$\langle \mu, f \rangle = \int_X f \, d\mu.$$

Definition 1.16. A sequence of probability measures $\mu_n \in \mathcal{P}(X)$ converges weak*ly to $\mu \in \mathcal{P}(X)$, written $\mu_n \overset{*}{\rightharpoonup} \mu$, if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu \quad \text{for all } f \in C_b^0.$$

The notion of tightness of measures is closely related to compactness in the weak* topology. The relationship between the two concepts is established by Prokhorov's theorem.

Theorem 1.17. Let (X, d) be a completely separable metric space and let $\Gamma \subset \mathcal{P}(X)$. Then, the following statements are equivalent.

1. The set $\bar{\Gamma}$ is sequentially compact in $\mathcal{P}(X)$ in the weak* topology.
2. The set Γ is tight.

Now that we have introduced all the necessary tools, we are ready to prove the existence of minimizers to the Kantorovich problem.

Theorem 1.18. Let μ and ν be probability measures on Polish spaces X and Y respectively. Suppose that $c : X \times Y \to [0, \infty]$ is lower semi-continuous. Then, there exists a transport plan $\pi^\dagger \in \Pi(\mu, \nu)$ solving the Kantorovich problem.

Proof. We prove this statement using a standard argument from calculus of variations. We first prove that the feasible set $\Pi(\mu, \nu)$ is compact with respect to the weak* topology. Therefore, if π_n is a minimizing sequence satisfying

$$\mathbb{K}(\pi_n) \to \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) \quad \text{as } k \to \infty,$$

we can extract a converging subsequence converging to $\pi^\dagger \in \Pi(\mu, \nu)$. Note that such a minimizing sequence exists by [8, Section 39, Remark 1]. Since the limit of the subsequence agrees with the limit of the sequence π_n, we have that $\pi_n \overset{*}{\rightharpoonup} \pi^\dagger$. Since the cost function c is lower semi-continuous, we can use [4, Theorem A.3.12] to deduce that $\mathbb{K}$ is also lower semi-continuous which implies that

$$\lim_{n \to \infty} \mathbb{K}(\pi_n) \geq \mathbb{K}(\pi^\dagger).$$

Hence, the measure $\pi^\dagger$ is a minimizer to the Kantorovich problem.

Thus, it remains to be shown that the feasible set $\Pi(\mu,\nu)$ is compact in the weak* topology. From Remark 1.8, we know that $\Pi(\mu,\nu)$ is non-empty. Let $\delta > 0$ be arbitrary. By Theorem 1.15, there exist compact sets $K \subset X$ and $L \subset Y$ such that

$$\mu(X\backslash K) \leq \frac{\delta}{2} \quad \text{and} \quad \nu(Y\backslash L) \leq \frac{\delta}{2}.$$

If $(x,y) \in (X \times Y)\backslash(K \times L)$, then either $x \notin K$ or $y \notin L$. Hence, it follows that $(x,y) \in X \times (Y\backslash L)$ or $(x,y) \in (X\backslash K) \times Y$. This implies that for any transport plan $\pi \in \Pi(\mu,\nu)$, we have

$$\begin{aligned} \pi((X \times Y)\backslash(K \times L)) &\leq \pi(X \times (Y\backslash L)) + \pi((X\backslash K) \times Y) \\ &= \nu(Y\backslash L) + \mu(X\backslash K) \\ &\leq \delta. \end{aligned}$$

This shows that the set of transport plans $\Pi(\mu,\nu)$ is tight. By Theorem 1.17, we can deduce that the closure of $\Pi(\mu,\nu)$ is compact with respect to the weak* topology. Thus, it suffices to show that $\Pi(\mu,\nu)$ is weak*ly closed. Suppose that π_n converges weak*ly to $\pi \in \mathcal{P}(X \times Y)$. That is,

$$\int_{X\times Y} f(x,y)\ d\pi_n(x,y) \to \int_{X\times Y} f(x,y)\ d\pi(x,y) \quad \text{for all } f \in C_b^0(X \times Y). \tag{1.4}$$

Pick a bounded continuous function $f \in C_b^0(X\times Y)$ depending only on x. This means that $f(x,y) = \tilde{f}(x)$ for some $\tilde{f} \in C_b^0(X)$. Since $\pi_n \in \Pi(\mu,\nu)$, the left-hand side of Equation (1.4) yields

$$\begin{aligned} \int_{X\times Y} f(x,y)\ d\pi_n(x,y) &= \int_{X\times Y} \tilde{f}(x)\ d\pi_n(x,y) \\ &= \int_X \tilde{f}(x)\ d\mu(x). \end{aligned}$$

On the other hand, the right-hand side of Equation (1.4) yields

$$\begin{aligned} \int_X f(x,y)\ d\pi(x,y) &= \int_X \tilde{f}(x)\ d\pi(x,y) \\ &= \int_X \tilde{f}(x)\ dP^X_\# \pi(x), \end{aligned}$$

where P^X denotes the projection onto X. Thus, we have that

$$\int_X \tilde{f}(x)\ d\mu(x) \to \int_X \tilde{f}(x)\ dP^X_\# \pi(x).$$

Since this holds for all $\tilde{f} \in C_b^0(X)$, we can deduce that $P^X_\# \pi = \mu$. The same reasoning can be used to prove that $P^Y_\# \pi = \nu$. This shows that $\pi \in \Pi(\mu,\nu)$ and the set of transport plans is weak*ly closed.

□

Conclusion

In essence, the Monge problem can be viewed as an optimization problem where we minimize the total cost of transportation over the set of all transport maps. However, this problem is not always defined for arbitrary initial and final configurations since mass is not allowed to split. To remedy this problem, we allow for mass splitting by considering the Kantorovich problem where transport plans are used instead of transport maps. We have shown that we can guarantee the existence of a minimizer to the Kantorovich problem under general conditions.

1.2 Kantorovich Duality

1.2.1 Informal Proof of Kantorovich Duality

It turns out that the Kantorovich problem admits a dual problem. That is, there exists a corresponding maximization problem whose solution provides information on the primal minimization problem. In fact, we show that the duality gap is zero which means that both problems yield the same optimal value. This is known as strong duality. We start by providing a statement of the dual problem.

Theorem 1.19. Let μ and ν be probability measures on Polish spaces X and Y respectively. Let $c : X \times Y \to [0, +\infty]$ be a lower semi-continuous cost function. We define the function

$$\begin{aligned} \mathbb{J} : L^1(\mu) \times L^1(\nu) &\to \mathbb{R}, \\ (\varphi, \psi) &\mapsto \int_X \varphi \, d\mu + \int_Y \psi \, d\nu. \end{aligned}$$

Furthermore, we define the set Φ_c by

$$\Phi_c = \{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu) : \varphi(x) + \psi(y) \leq c(x, y)\},$$

where the inequality holds for μ-almost all $x \in X$ and ν-almost all $y \in Y$. Then, we have

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) = \sup_{(\varphi,\psi) \in \Phi_c} \mathbb{J}(\varphi, \psi). \tag{1.5}$$

Remark 1.20. A maximizer of the Kantorovich dual problem may not exist if the cost function is sufficiently pathological. Thus, a supremum over Φ_c is employed in the dual problem. In Section 1.2.3, we provide sufficient conditions for the existence of such a maximizer. In practice, most cost functions satisfy these conditions.

Remark 1.21. This Kantorovich dual has an intuitive interpretation which is attributed to L. Caffarelli [25, Chapter 1, 1.1.3]. Suppose that we own a bread manufacturing company and we need to transport the bread from the factories to bakeries spread throughout the city. Recall that the cost of transporting a bread from factory x to bakery y is $c(x, y)$. Now, a clever shipping company offers to ship the bread from the factories to the bakeries according to the following price scheme.

- We pay $\varphi(x)$ for loading the bread into the trucks at factory x.
- We pay $\psi(y)$ for unloading the trucks at bakery y.

To make the offer attractive, the shipping company ensures that we have to pay at most the cost of transporting the bread ourselves. In other words, we have $\varphi(x) + \psi(y) \leq c(x, y)$. However, the Kantorovich duality theorem states that the shipping company can find a price scheme (φ, ψ) such that the money gained by the shipping company from the transaction is exactly the cost of transporting the bread to the bakeries ourselves. In other words, there exist φ, ψ such that $\varphi(x) + \psi(y) = c(x, y)$.

Proof of Theorem 1.19. The proof relies on a minimax principle which allows us to interchange an infimum and a supremum. This step will be subsequently made rigorous in Section 1.2.2. Let $\mathcal{M}_+(X \times Y)$ denote the space of non-negative Radon measures on the product space $X \times Y$. We start by writing the Kantorovich problem as

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) = \inf_{\pi \in \mathcal{M}_+(X\times Y)} \left(\mathbb{K}(\pi) + \begin{cases} 0 & \text{if } \pi \in \Pi(\mu, \nu) \\ +\infty & \text{else} \end{cases} \right). \tag{1.6}$$

Observe that

$$\begin{Bmatrix} 0 & \text{if } \pi \in \Pi(\mu,\nu) \\ +\infty & \text{else} \end{Bmatrix} = \sup_{(\varphi,\psi)} \left(\int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} [\varphi(x)+\psi(y)] \, d\pi(x,y) \right),$$

where the supremum is taken over all $(\varphi,\psi) \in C_b^0(X) \times C_b^0(Y)$. This means that the left-hand side of (1.6) can be expressed as

$$\inf_{\pi \in \mathcal{M}_+(X\times Y)} \sup_{(\varphi,\psi)} \left(\int_{X\times Y} c(x,y) \, d\pi(x,y) + \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} [\varphi(x)+\psi(y)] \, d\pi(x,y) \right).$$

Assuming that the order of the supremum and infimum can be interchanged, we can rewrite the above equation as

$$\begin{aligned} &\sup_{(\varphi,\psi)} \inf_{\pi \in \mathcal{M}_+(X\times Y)} \left(\int_{X\times Y} c(x,y) \, d\pi(x,y) + \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} [\varphi(x)+\psi(y)] \, d\pi(x,y) \right) \\ &= \sup_{(\varphi,\psi)} \left(\int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \sup_{\pi \in \mathcal{M}_+(X\times Y)} \int_{X\times Y} [\varphi(x)+\psi(y)-c(x,y)] \, d\pi(x,y) \right) \end{aligned} \tag{1.7}$$

We now compute the value of the supremum inside the brackets. We distinguish two cases.

- Suppose that $\varphi(x_0)+\psi(y_0)-c(x_0,y_0) > 0$ for some $(x_0,y_0) \in X \times Y$. In this case, we can consider the Borel measure $\pi = \lambda\delta_{(x_0,y_0)}$. By letting $\lambda \to +\infty$, we can deduce that the supremum is infinite.
- Suppose that $\varphi(x)+\psi(y)-c(x,y) \leq 0$ for $d\mu \otimes d\nu$-almost every $(x,y) \in X \times Y$, we see that the supremum is attained for the zero measure $\pi = 0$. Hence, we can conclude that

$$\sup_{\pi \in \mathcal{M}_+(X\times Y)} \int_{X\times Y} [\varphi(x)+\psi(y)-c(x,y)] \, d\pi(x,y) = \begin{cases} 0 & \text{if } (\varphi,\psi) \in \Phi_c \\ +\infty & \text{else.} \end{cases}$$

Thus, we substitute the above in (1.7) to conclude that

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) = \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi).$$

□

1.2.2 Rigorous Proof of Kantorovich Duality

Note that in the proof of Theorem 1.19, we interchange an infimum and a supremum without proof. In this section, we make this step rigorous by proving Theorem 1.19 in two steps. We start with the simpler step.

Lemma 1.22. Under the same assumptions as Theorem 1.19, we have

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) \geq \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi).$$

Proof. Let $\pi \in \Pi(\mu,\nu)$ be arbitrary. We start by showing that the inequality

$$\varphi(x)+\psi(y) \leq c(x,y). \tag{1.8}$$

holds for π-almost every $(x, y) \in X \times Y$. Let $(\varphi, \psi) \in \Phi_c$ be arbitrary. Then, there exist subsets $A \subset X$ and $B \subset Y$ such that $\mu(A) = 1$ and $\nu(B) = 1$ satisfying

$$\varphi(x) + \psi(y) \le c(x, y)$$

for all $(x, y) \in A \times B$. Note that

$$\pi(A^c \times B^c) \le \pi(A^c \times Y) + \pi(X \times B^c) = \mu(A^c) + \nu(B^c) = 0.$$

This implies that

$$\begin{aligned} \pi(A \times B) &= \pi(X \times B) - \pi(A^c \times B) \\ &= \nu(B) - \pi(A^c \times Y) + \pi(A^c \times B^c) \\ &= 1 - \mu(A^c) + \pi(A^c \times B^c) \\ &= 1. \end{aligned}$$

This means that the inequality $\varphi(x)+\psi(y) \le c(x, y)$ holds for π-almost every $(x, y) \in X \times Y$. This allows us to conclude that

$$\begin{aligned} \mathbb{K}(\pi) = \int_{X\times Y} c(x, y)\ d\pi(x, y) &\ge \int_{X\times Y} [\varphi(x) + \psi(y)]\ d\pi(x, y) \\ &= \int_X \varphi\ d\mu + \int_Y \psi\ d\nu = \mathbb{J}(\varphi, \psi). \end{aligned} \tag{1.9}$$

By taking the infimum over all $\pi \in \Pi(\mu, \nu)$ on the left-hand side of (1.9) and the supremum over all $(\varphi, \psi) \in \Phi_c$ on the right-hand side of (1.9), we obtain the desired result. □

The proof of the reverse inequality uses a powerful tool from convex analysis called the Fenchel-Rockafellar duality theorem [25, Theorem 1.9] which relates a primal minimization problem to a corresponding dual maximization problem using the Legendre-Fenchel transform.

Definition 1.23. Let $\varphi : X \to \bar{\mathbb{R}}$ be a real-valued function on X. Then, the Legendre-Fenchel transforms φ^* and φ^{**} are defined by

$$\begin{aligned} \varphi^* : X^* &\to \bar{\mathbb{R}} \\ x^* &\mapsto \sup_{x\in X}(\langle x^*, x\rangle - \varphi(x)), \\ \varphi^{**} : X^{**} &\to \bar{\mathbb{R}} \\ x^{**} &\mapsto \sup_{x^*\in X^*}(\langle x^{**}, x^*\rangle - \varphi^*(x^*)), \end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the canonical pairing between covectors and vectors.

Remark 1.24. The Legendre-Fenchel transform gives rise to the Legendre-Fenchel inequality

$$\langle x^*, x\rangle \le \varphi(x) + \varphi^*(x^*),$$

where $x \in X$ and $x^* \in X^*$.

Theorem 1.25. Let E be a normed vector space and $\Theta, \Xi : E \to \mathbb{R} \cup \{+\infty\}$ be convex functions. Suppose that there exists $z_0 \in E$ such that Θ and Ξ are both finite at z_0 and Θ is continuous at z_0. Then,

$$\inf_{z\in E}(\Theta(z) + \Xi(z)) = \max_{z^*\in E^*}(-\Theta^*(-z^*) - \Xi^*(z^*)). \tag{1.10}$$

Remark 1.26. Note that the supremum on the right-hand side of (1.10) is attained for some $z^* \in E^*$.

We now consider the reverse inequality in Lemma 1.27. The proof of this step is much longer and significantly more involved. We only present the case where X and Y are compact and the cost function is continuous. The reader may safely ignore the details of the proof on the first reading. The proof of the statement in complete generality can be found in [25, pp. 28-32].

Lemma 1.27. Under the same assumptions as Theorem 1.19, we have

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) \leq \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi).$$

The statement is proved in three successive parts, with assumptions of increasing generality.

1. The spaces X and Y are compact and the cost function is continuous.
2. The spaces X and Y are no longer compact, but the cost function c is still continuous.
3. The cost function c is only assumed to be lower semi-continuous.

Proof of Lemma 1.27.1. Suppose that X and Y are compact and that the cost function c is continous. We start by showing that all the conditions in Theorem 1.25 are satisfied. Let $E = C_b^0(X\times Y)$ be equipped with the supremum norm. Then, the Riesz-Markov-Kakutani representation theorem [18, Theorem 6.19] states that the dual space E^* is given by the space of Radon measures $\mathcal{M}(X\times Y)$. We define the functions

$$\begin{aligned}
\Theta : C_b^0(X\times Y) &\to \mathbb{R}\cup\{+\infty\} \\
u &\mapsto \begin{cases} 0 & \text{if } u(x,y) \geq -c(x,y) \\ +\infty & \text{otherwise,} \end{cases} \\
\Xi : C_b^0(X\times Y) &\to \mathbb{R}\cup\{+\infty\} \\
u &\mapsto \begin{cases} \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) & \text{if } u(x,y) = \varphi(x)+\psi(y) \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Despite the fact that the representation of $u(x,y) = \varphi(x)+\psi(y)$ is not unique (for instance, take $\tilde{\varphi}(x) = \varphi(x)+s$ and $\tilde{\psi}(y) = \psi(y)-s$ where $s\in\mathbb{R}$) the map Ξ is still well-defined.

We first show that Θ is convex. Let $u, v \in C_b^0(X\times Y)$ and $\lambda\in[0,1]$ be arbitrary. We distinguish two cases.

- If both $\Theta(u), \Theta(v) < \infty$, we have $u(x,y), v(x,y) \geq -c(x,y)$ which implies that

  $$\lambda u(x,y) + (1-\lambda)v(x,y) \geq -c(x,y).$$

  Hence, we can deduce that

  $$\Theta(\lambda u + (1-\lambda)v) = 0 = \lambda\Theta(u) + (1-\lambda)\Theta(v).$$

- If $\Theta(u) = +\infty$ or $\Theta(v) = +\infty$, we immediately obtain

$$\Theta(\lambda u + (1-\lambda)v) \leq \lambda\Theta(u) + (1-\lambda)\Theta(v).$$

We now show that the map Ξ is convex. Let $u, v \in C_b^0(X \times Y)$ and $\lambda \in [0,1]$ be arbitrary. Again, we distinguish two cases.

- If both $\Xi(u), \Xi(v) < \infty$, we have that

$$u(x,y) = \varphi(x) + \psi(y) \quad \text{and} \quad v(x,y) = \tilde{\varphi}(x) + \tilde{\psi}(y).$$

  By grouping terms that depend on x and y respectively, we obtain

$$\lambda u(x,y) + (1-\lambda)v(x,y) = [\lambda\varphi(x) + (1-\lambda)\tilde{\varphi}(x)] + [\lambda\psi(y) + (1-\lambda)\tilde{\psi}(y)].$$

  By applying the definition of the map Ξ, we get

$$\Xi(\lambda u+(1-\lambda)v) = \int_X [\lambda\varphi+(1-\lambda)\tilde{\varphi}]\, d\mu + \int_Y [\lambda\psi+(1-\lambda)\tilde{\psi}]\, d\nu = \lambda\Xi(u)+(1-\lambda)\Xi(v).$$

- If $\Xi(u) = +\infty$ or $\Xi(v) = +\infty$, we immediately obtain

$$\Xi(\lambda u + (1-\lambda)v) \leq \lambda\Xi(u) + (1-\lambda)\Xi(v).$$

We now check the remaining conditions of Theorem 1.25 with $z_0 \equiv 1$. We see that both $\Theta(z_0), \Xi(z_0) < \infty$ and the fact that Θ is continous at $z_0 \equiv 1$ follows from an $\epsilon - \delta$ argument.

Let $\epsilon > 0$ be arbitrary. Pick $\delta = \frac{1}{2}$. Then, we have

$$\|u - 1\|_\infty < \delta \quad \Rightarrow \quad \sup_{(x,y)\in X\times Y} |u(x,y) - 1| < \frac{1}{2}.$$

Hence, it follows that $u(x,y) > 0$ for all $(x,y) \in X \times Y$. Thus, we obtain

$$|\Theta(u) - \Theta(1)| = 0 < \epsilon.$$

Thus, Theorem 1.25 yields

$$\inf_{u\in E}(\Theta(u) + \Xi(u)) = \max_{\pi\in E^*}(-\Theta^*(-\pi) - \Xi^*(\pi)). \tag{1.11}$$

By considering the left-hand side of (1.11), we obtain

$$\inf_{u\in E}(\Theta(u)+\Xi(u)) \geq \inf_{\substack{\varphi(x)+\psi(y)\geq -c(x,y)\\ \varphi\in L^1(\mu),\psi\in L^1(\nu)}} \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) = -\sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi).$$

By considering the right-hand side of (1.11), we compute the Legendre-Fenchel transform of the functions Θ and Ξ. Since $E^* = \mathcal{M}(X \times Y)$, we note that the canonical pairing between $u \in E$ and $\pi \in E^*$ is given by

$$\langle \pi, u \rangle = \int_{X\times Y} u\, d\pi.$$

We first compute the Legendre-Fenchel transform of Θ.

$$\begin{aligned}\Theta^*(-\pi) &= \sup_{u\in E}\left(\langle \pi, u\rangle - \Theta(u)\right)\\ &= \sup_{u\in E}\left(-\int_{X\times Y}[u-\Theta(u)]\ d\pi\right)\\ &= \sup_{u\geq -c} -\int_{X\times Y} u\ d\pi\\ &= \sup_{u\leq c}\int_{X\times Y} u\ d\pi.\end{aligned}$$

We distinguish two cases.

- If $\pi \in \mathcal{M}_+(X\times Y)$ is a non-negative Radon measure on $X\times Y$, it follows that

$$\sup_{u\leq c}\int_{X\times Y} u\ d\pi = \int_{X\times Y} c\ d\pi.$$

- If $\pi \in \mathcal{M}(X\times Y)\backslash\mathcal{M}_+(X\times Y)$ then taking $u\to-\infty$ yields

$$\sup_{u\leq c}\int_{X\times Y} u\ d\pi = \infty.$$

To summarize, we have

$$\Theta^*(-\pi) = \begin{cases}\int_{X\times Y} c\ d\pi & \text{if } \pi\in\mathcal{M}_+(X\times Y)\\ +\infty & \text{otherwise.}\end{cases}$$

We now compute the Legendre-Fenchel transform of Ξ.

$$\begin{aligned}\Xi^*(\pi) &= \sup_{u\in E}\left(\langle\pi,u\rangle - \Xi(u)\right)\\ &= \sup_{u\in E}\left(\int_{X\times Y} u\ d\pi - \Xi(u)\right)\\ &= \sup_{u(x,y)=\varphi(x)+\psi(y)}\left(\int_{X\times Y} u\ d\pi - \int_X \varphi(x)\ d\mu(x) - \int_Y \psi(y)\ d\nu(y)\right)\\ &= \sup_{u(x,y)=\varphi(x)+\psi(y)}\left(\int_X \varphi(x)\ d(P^X_\#\pi-\mu)(x) + \int_Y \psi(y)\ d(P^Y_\# - \nu)(y)\right).\end{aligned}$$

Hence, we obtain

$$\Xi^*(\pi) = \begin{cases}0 & \text{if } \pi\in\Pi(\mu,\nu)\\ +\infty & \text{otherwise.}\end{cases}$$

Thus, from (1.11), we obtain

$$-\min_{\pi\in\Pi(\mu,\nu)}\mathbb{K}(\pi) = \max_{\pi\in E^*}(-\Theta^*(-\pi)-\Xi^*(\pi)) = \inf_{u\in E}(\Theta(u)+\Xi(u)) \geq -\sup_{(\varphi,\psi)\in\Phi_c}\mathbb{J}(\varphi,\psi),$$

which implies that

$$\min_{\pi\in\Pi(\mu,\nu)}\mathbb{K}(\pi) \leq \sup_{(\varphi,\psi)\in\Phi_c}\mathbb{J}(\varphi,\psi).$$

□

Proof of Lemma 1.27.2 and 1.27.3. The proof of these parts is longer and significantly more involved. For the sake of clarity, these proofs are omitted and can be found in [25, pp. 28-32]. □

1.2.3 Existence of Maximizers to the Dual Kantorovich Problem

In this section, we provide sufficient conditions for the existence of a maximizer to the Kantorovich dual problem. To do so, we introduce a tool pioneered by Rüschendorf [19] which is similar in structure to the Legendre-Fenchel transform. This tool transforms real-valued functions defined on the initial space X into real-valued functions defined on the target space Y using the cost function c that inherent to the optimal transport problem.

Definition 1.28. Let $\varphi : X \to \bar{\mathbb{R}}$ be a real-valued function on X. Then the c-transforms φ^c and φ^{cc} are defined by

$$\begin{aligned} \varphi^c : Y &\to \bar{\mathbb{R}} \\ y &\mapsto \inf_{x \in X} (c(x,y) - \varphi(x)), \\ \varphi^{cc} : X &\to \bar{\mathbb{R}} \\ x &\mapsto \inf_{y \in Y} (c(x,y) - \varphi^c(y)). \end{aligned}$$

Example 1.29. In the context of factories and bakeries, the c-transform $\varphi^c(y)$ represents the best possible price such that we can unload the bread at bakery y since the cost of unloading the bread satisfies

$$\psi(y) \leq c(x,y) - \varphi(x),$$

and we are taking the greatest lower bound. Similarly, $\psi^c(x)$ represents the best possible price such that we can load the bread at factory x.

For probability measures on Polish spaces, the existence of a minimizer to the Kantorovich dual problem follows whenever the cost function can be bounded above by the sum of two integrable functions, one defined on the initial space and the other on the target space.

Theorem 1.30. Let μ and ν be probability measures on Polish spaces X and Y respectively and $c : X \times Y \to [0, \infty]$ be a cost function. Suppose that there exist integrable functions $c_X \in L^1(\mu)$ and $c_Y \in L^1(\nu)$ such that the inequality $c(x,y) \leq c_X(x) + c_Y(y)$ holds for μ-almost every $x \in X$ and ν-almost every $y \in Y$. Define

$$M = \int_X c_X \, d\mu + \int_Y c_Y \, d\nu < \infty.$$

Then, there exists $(\varphi^\dagger, \psi^\dagger) \in \Phi_c$ such that

$$\sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi) = \mathbb{J}(\varphi^\dagger, \psi^\dagger),$$

where the pair $(\varphi^\dagger, \psi^\dagger)$ can be chosen such that $(\varphi^\dagger, \psi^\dagger) = (\eta^{cc}, \eta^c)$ for some $\eta \in L^1(\mu)$.

To prove this statement, we start by restricting the pairs $(\varphi, \psi) \in \Phi_c$ that we need to consider as candidates for the maximizer of the Kantorovich dual problem. Namely, only c-transform pairs need to be considered.

Lemma 1.31. Let μ and ν be probability measures on Polish spaces X and Y respectively and $c : X \times Y \to [0, +\infty]$ be a cost function. Then, for all $a \in \mathbb{R}$ and $(\tilde{\varphi}, \tilde{\psi}) \in \Phi_c$, we have that $(\varphi, \psi) = (\tilde{\varphi}^{cc} - a, \tilde{\varphi}^c + a)$ satisfies the inequalities

$$\mathbb{J}(\varphi, \psi) \geq \mathbb{J}(\tilde{\varphi}, \tilde{\psi}) \quad \text{and} \quad \varphi(x) + \psi(y) \leq c(x, y),$$

for μ-almost every $x \in X$ and ν-almost every $y \in Y$. Furthermore, if the inequalities

$$\mathbb{J}(\tilde{\varphi}, \tilde{\psi}) > -\infty \quad \text{and} \quad M < +\infty$$

hold, where M is defined as in Theorem 1.30 and there exist $c_X \in L^1(\mu), c_Y \in L^1(\nu)$ such that $\varphi \leq c_X$ and $\psi \leq c_Y$, then the pair $(\varphi, \psi) \in \Phi_c$.

Proof. It is clear that for all $a \in \mathbb{R}$, we have $\mathbb{J}(\varphi - a, \psi + a) = \mathbb{J}(\varphi, \psi)$ for all $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. To show that $\mathbb{J}(\varphi, \psi) \geq \mathbb{J}(\tilde{\varphi}, \tilde{\psi})$, it suffices to show that $\varphi = \tilde{\varphi}^{cc} \geq \tilde{\varphi}$ and that $\psi = \tilde{\varphi}^c \geq \tilde{\psi}$.

Using the fact that $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq c(x, y)$, we can deduce that

$$\begin{aligned}
\psi(y) &= \inf_{x \in X} (c(x, y) - \tilde{\varphi}(x)) \geq \tilde{\psi}(y), \\
\varphi(x) &= \inf_{y \in Y} \sup_{z \in X} (c(x, y) - c(z, y) + \tilde{\varphi}(z)) \geq \tilde{\varphi}(x).
\end{aligned}$$

where the last inequality is obtained by setting $z = x$. Moreover, by the definition of the c-transform, we obtain

$$\varphi(x) + \psi(y) = \inf_{z \in Y} (c(x, z) - \tilde{\varphi}(z) + \tilde{\varphi}^c(y)) \leq c(x, y)$$

where the last inequality is obtained by setting $z = y$. Since $\varphi(x) + \psi(y) \leq c(x, y)$, to prove that $(\varphi, \psi) \in \Phi_c$, it remains to be shown that $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. Using the fact that $\mathbb{J}(\varphi, \psi) \geq \mathbb{J}(\tilde{\varphi}, \tilde{\psi})$, we obtain

$$\int_X [\varphi - c_X] \, d\mu + \int_Y [\varphi(y) - c_Y] \, d\nu = \mathbb{J}(\varphi, \psi) - M \geq \mathbb{J}(\tilde{\varphi}, \tilde{\psi}) - M.$$

By assumption, we know that $\varphi - c_X \leq 0$ and $\psi - c_Y \leq 0$. Hence, we can deduce that both integrals on the left-hand side of the inequality are negative. Thus, it follows that

$$\begin{aligned}
\|\varphi - c_X\|_{L^1(\mu)} + \|\psi - c_Y\|_{L^1(\nu)} &= -\int_X [\varphi - c_X] \, d\mu - \int_Y [\psi - c_Y] \, d\nu \\
&\leq M - \mathbb{J}(\tilde{\varphi}, \tilde{\psi}).
\end{aligned}$$

This means that both norms are finite and so $\varphi - c_X \in L^1(\mu)$ and $\psi - c_Y \in L^1(\nu)$. Hence, we have shown that $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. □

Under the same assumptions as Theorem 1.30, we can guarantee the existence of a maximizing sequence in Φ_c for the maximization problem in the right-hand side of (1.5). Moreover, this sequence can be chosen to satisfy an upper bound condition.

Lemma 1.32. Under the same assumptions as Theorem 1.30, there exists a sequence $(\varphi_k, \psi_k) \in \Phi_c$ such that

$$\mathbb{J}(\varphi_k, \psi_k) \to \sup_{(\varphi, \psi) \in \Phi_c} \mathbb{J}(\varphi, \psi),$$

where the sequences φ_k, ψ_k satisfy the pointwise upper bounds

$$\varphi_k(x) \le c_X(x) \quad \text{and} \quad \psi_k(y) \le c_Y(y),$$

for all $k \in \mathbb{N}$, $x \in X$ and $y \in Y$.

Proof. Let $(\tilde{\varphi}_k, \tilde{\psi}_k) \in \Phi_c$ be a maximizing sequence satisfying

$$\mathbb{J}(\tilde{\varphi}_k, \tilde{\psi}_k) \to \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) \quad \text{as } k \to \infty.$$

Such a maximizing sequence exists by [8, Section 41, Remark 1]. Since we have the inequalities

$$0 \le \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) \le \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) \le M < \infty,$$

we can deduce that the sequences $\tilde{\varphi}_k, \tilde{\psi}_k$ can never attain the values $\pm\infty$. We define a new sequence $(\varphi_k, \psi_k) = (\tilde{\varphi}_k^{cc} - a_k, \tilde{\psi}_k^c + a_k)$ where

$$a_k = \inf_{y\in Y}(c_Y(y) - \tilde{\varphi}_k^c(y)).$$

By Lemma 1.31, we can deduce that the sequence (φ_k, ψ_k) is a maximizing sequence with $(\varphi_k, \psi_k) \in \Phi_c$ if we can show that $\varphi_k \le c_X$ and $\psi_k \le c_Y$.

We first show that the sequence (φ_k, ψ_k) is well-defined. We do this by showing that $a_k \in \mathbb{R}$. Since $(\tilde{\varphi}_k, \tilde{\psi}_k) \in \Phi_c$, we have that for all $y \in Y$,

$$\tilde{\varphi}_k(x) \le c(x, y) - \tilde{\psi}_k(y).$$

Since both the cost function c and the sequence $\tilde{\psi}_k(y)$ are bounded, there exists some $y_0 \in Y$ and $b_0 \in \mathbb{R}$ which may depend on k such that

$$\tilde{\varphi}_k(x) \le c(x, y_0) + b_0.$$

From this, we can conclude that

$$\tilde{\varphi}_k^c(y_0) = \inf_{x\in X}(c(x, y_0) - \tilde{\varphi}_k(x)) \ge -b_0.$$

This means that we can construct an have an upper bound for a_k since

$$a_k \le c_Y(y_0) - \varphi_k^c(y_0) \le c_Y(y_0) + b_0 < \infty.$$

To show that a_k is bounded below, by assumption, we have

$$c_Y(y) - \tilde{\varphi}_k^c(y) = \sup_{x\in X}(c_Y(y) - c(x, y) + \tilde{\varphi}_k(x)) \ge \sup_{x\in X}(\tilde{\varphi}_k(x) - c_X(x)) \ge \tilde{\varphi}_k(x_0) - c_X(x_0),$$

for any $x_0 \in X$. This means that

$$a_k = \inf_{y\in Y}(c_Y(y) - \tilde{\varphi}_k^c(y)) \ge \tilde{\varphi}_k(x_0) - c_X(x_0).$$

We now show that the pair (φ_k, ψ_k) satisfies the pointwise upper bounds. Note that

$$\psi_k = \tilde{\varphi}_k^c(y) + a_k \le c_Y(y).$$

For the second bound, note that

$$\begin{aligned}\varphi_k(x) - c_X(x) &= \tilde{\varphi}_k^{cc}(x) - a_k - c_X(x)\\ &= \inf_{y\in Y}(c(x,y) - \tilde{\varphi}_k^c(y) - a_k - c_X(x))\\ &\le \inf_{y\in Y}(c_Y(y) - \tilde{\varphi}_k^c(y) - a_k)\\ &= 0.\end{aligned}$$

Thus, we can conclude that

$$\varphi_k(x) \le c_X(x).$$

□

We use Lemma 1.31 and Lemma 1.32 to prove the existence of a maximizer to the Kantorovich dual problem under the assumptions of Theorem 1.30.

Proof of Theorem 1.30. Let $(\varphi_k, \psi_k) \in \Phi_c$ be a maximizing sequence obtained from Lemma 1.32. We define the sequences

$$\begin{aligned}\varphi_k^{(l)}(x) &= \max\{\varphi_k(x) - c_X(x), -l\} + c_X(x),\\ \psi_k^{(l)}(y) &= \max\{\psi_k(y) - c_Y(y), -l\} + c_Y(y).\end{aligned}$$

Observe that we have

$$\varphi_k \le \varphi_k^{(l)} \quad \text{and} \quad \psi_k \le \psi_k^{(l)}. \tag{1.12}$$

Moreover, for a fixed $k \in \mathbb{N}$, we have that $\varphi_k^{(l)}$ and $\psi_k^{(l)}$ are monotonically decreasing sequences of functions. Furthermore, for all $k, l \in \mathbb{N}$, we have the inequalities

$$\begin{aligned}-l \le \varphi_k^{(l)} - c_X \le 0,\\ -l \le \psi_k^{(l)} - c_Y \le 0.\end{aligned} \tag{1.13}$$

Lastly, we obtain

$$\begin{aligned}\varphi_k^{(l)}(x) + \psi_k^{(l)}(y) &\le \max\{\varphi_k(x) - c_X(x) + \psi_k(y) - c_Y(y), -l\} + c_X(x) + c_Y(y)\\ &\le \max\{c(x,y) - c_X(x) - c_Y(y), -l\} + c_X(x) + c_Y(y),\end{aligned} \tag{1.14}$$

where the last inequality follows from the fact that $(\varphi_k, \psi_k) \in \Phi_c$. Since $L^p(\mu)$ is a reflexive Banach space for all $p \in (1, \infty)$ and the sequence $\varphi_k^{(l)}$ is a bounded sequence for a fixed $l \in \mathbb{N}$, we have that the set

$$\overline{\left\{\varphi_k^{(l)}\right\}_{k\in\mathbb{N}}}$$

is weakly compact subset of $L^p(\mu)$ for all $p \in (1, \infty)$. By choosing $p = 2$, we can deduce that, there exists a subsequence of $\varphi_k^{(l)}$ converging weakly to $\varphi^{(l)} \in L^1(\mu)$. Let I_1 denote the indices of the corresponding weakly converging subsequence of $\varphi_k^{(1)}$. By repeating the above argument with $\{\varphi_k^{(2)}\}_{k\in I_1}$, we have that there exists a subsequence of $\{\varphi_k^{(2)}\}_{k\in I_1}$ converging weakly to $\varphi^{(2)} \in L^1(\mu)$. Proceeding inductively, we obtain

$$\begin{aligned}
\varphi_k^{(1)} &\rightharpoonup \varphi^{(1)} \quad \text{along } I_1 \subseteq \mathbb{N}, \\
\varphi_k^{(2)} &\rightharpoonup \varphi^{(2)} \quad \text{along } I_2 \subseteq I_1, \\
&\vdots \\
\varphi_k^{(l)} &\rightharpoonup \varphi^{(l)} \quad \text{along } I_l \subseteq I_{l-1}, \\
&\vdots
\end{aligned}$$

Consider the indices

$$I = \{k_n \in \mathbb{N} : k_n \text{ is the } n^{\text{th}} \text{ element of } I_n\}.$$

This implies that

$$\varphi_k^{(l)} \rightharpoonup \varphi^{(l)} \in L^1(\mu) \quad \text{along } I \subseteq \mathbb{N} \text{ for all } l \in \mathbb{N}.$$

Using the same diagonalization argument, we also deduce that for some subset $J \subseteq \mathbb{N}$ of indices, we have

$$\psi_k^{(l)} \rightharpoonup \psi^{(l)} \in L^1(\nu) \quad \text{along } J \subseteq \mathbb{N} \text{ for all } l \in \mathbb{N}.$$

Since weak limits preserve order, for all $l \in \mathbb{N}$, we obtain

$$\begin{aligned}
c_X \geq \varphi^{(1)} \quad &\text{and} \quad \varphi^{(l)} \geq \varphi^{(l+1)}, \\
c_Y \geq \psi^{(1)} \quad &\text{and} \quad \psi^{(l)} \geq \psi^{(l+1)}.
\end{aligned}$$

By applying the monotone convergence theorem, we get

$$\begin{aligned}
\lim_{l\to\infty} \int_X \varphi^{(l)} \, d\mu &= \int_X \varphi^\dagger \, d\mu, \\
\lim_{l\to\infty} \int_Y \psi^{(l)} \, d\nu &= \int_Y \psi^\dagger \, d\nu,
\end{aligned}$$

where $\varphi^\dagger$ and $\psi^\dagger$ are the pointwise limits of $\varphi^{(l)}$ and $\psi^{(l)}$ respectively. We claim that $(\varphi^\dagger, \psi^\dagger)$ is a maximizer to the Kantorovich dual problem. To do so, we show that $(\varphi^\dagger, \psi^\dagger) \in \Phi_c$ and $\mathbb{J}(\varphi^\dagger, \psi^\dagger) \geq \mathbb{J}(\varphi, \psi)$ for all $(\varphi, \psi) \in \Phi_c$. To prove the latter, note that for all $l \in \mathbb{N}$, we have

$$\sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) = \lim_{k\to\infty} \mathbb{J}(\varphi_k, \psi_k) \leq \lim_{k\to\infty} \mathbb{J}(\varphi_k^{(l)}, \psi_k^{(l)}) = \mathbb{J}(\varphi^{(l)}, \psi^{(l)}),$$

where the inequality follows from (1.12). Thus, we have

$$\sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) \leq \lim_{l\to\infty} \mathbb{J}(\varphi^{(l)}, \psi^{(l)}) = \mathbb{J}(\varphi^\dagger, \psi^\dagger).$$

To prove that $(\varphi^\dagger, \psi^\dagger) \in \Phi_c$, we let $l \to \infty$ in (1.14) to deduce

$$\varphi^\dagger(x) + \psi^\dagger(y) \leq c(x, y).$$

It remains to show integrability of φ and ψ. Note that

$$\begin{aligned}\sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi,\psi) - M &\leq \int_X [\varphi^\dagger - c_X]\, d\mu + \int_Y [\psi^\dagger - c_Y]\, d\nu \\ &\leq 0.\end{aligned} \tag{1.15}$$

By letting $l \to \infty$ in (1.13), we have that $\varphi^\dagger - c_X \leq 0$ and $\psi^\dagger - c_Y \leq 0$. Thus, we can deduce that both integrals in (1.15) are finite. Hence, it follows that $\varphi^\dagger - c_X \in L^1(\mu)$ and $\psi^\dagger - c_Y \in L^1(\nu)$. Thus, we conclude $\varphi^\dagger \in L^1(\mu)$ and $\psi^\dagger \in L^1(\nu)$.

To prove the last part of Theorem 1.30, we use Lemma 1.31 to deduce that for all $a \in \mathbb{R}$, we have

$$\mathbb{J}(\varphi^\dagger,\psi^\dagger) \leq \mathbb{J}((\varphi^\dagger)^{cc} - a, (\varphi^\dagger)^c + a) = \mathbb{J}((\varphi^\dagger)^{cc}, (\varphi^\dagger)^c).$$

Thus, it suffices to show that the pair $((\varphi^\dagger)^{cc}, (\varphi^\dagger)^c) \in L^1(\mu) \times L^1(\nu)$. We define

$$a = \inf_{y\in Y}(c_Y(y) - (\varphi^\dagger)^c(y)).$$

Using the same argument showing that $a_k \in \mathbb{R}$ in the proof of Lemma 1.32, we have that $a \in \mathbb{R}$. Since a is a lower bound, we have

$$(\varphi^\dagger)^c(y) + a \leq c_Y(y).$$

Moreover, by the definition of c-transform, we have

$$(\varphi^\dagger)^{cc}(x) - a = \inf_{y\in Y}(c(x,y) - (\varphi^\dagger)^c(y) - a) \leq \inf_{y\in Y}(c_X(x) + c_Y(y) - (\varphi^\dagger)^c(y) - a) \leq c_X(x).$$

Using Lemma 1.31, we have $((\varphi^\dagger)^{cc} - a, (\varphi^\dagger)^c + a) \in L^1(\mu) \times L^1(\nu)$. Hence, we have $((\varphi^\dagger)^{cc}, (\varphi^\dagger)^c) \in L^1(\mu) \times L^1(\nu)$. □

Conclusion

Since the Kantorovich problem is a convex optimization problem, it is not surprising that it admits a corresponding dual problem. This maximization problem provides valuable information about the original minimization problem since the duality gap is equal to zero. To guarantee the existence of a maximizer, some additional assumptions on the cost function must be imposed. In this case, the maximizers take the form of c-transform pairs.

1.3 Existence and Characterization of Transport

1.3.1 Euclidean Setting

After proving the existence of minimizers to the Kantorovich problem in Section 1.1.3 and the existence of maximizers to the dual Kantorovich problem in Section 1.2.3, we now characterize the optimality of transport plans in the Euclidean setting. In particular, we consider probability measures μ and ν on subsets of $\mathbb{R}^n$ where the cost function is given by $c(x,y) = \frac{1}{2}\|x-y\|^2$. To guarantee that the cost function is well-behaved, we make some assumptions on the integrability of $\|x\|^2$ and $\|y\|^2$.

Definition 1.33. Let μ and ν be probability measures on $X, Y \subseteq \mathbb{R}^n$ respectively. We say that μ and ν have finite second moments if

$$\int_X \|x\|^2 \, d\mu(x) < \infty \quad \text{and} \quad \int_Y \|y\|^2 \, d\nu(y) < \infty.$$

Before characterizing the optimality of transport plans, a brief reminder of the necessary tools from convex analysis is provided. More details can be found in [17].

Definition 1.34. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, the subdifferential of φ at x is defined by

$$\partial\varphi(x) = \{y \in \mathbb{R}^n : \ \varphi(z) \geq \varphi(x) + \langle y, z - x\rangle \quad \text{for all } z \in \mathbb{R}^n\}.$$

Moreover, we write

$$\partial\varphi = \{(x,y) \in X \times Y : y \in \partial\varphi(x)\}.$$

Theorem 1.35. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a proper, lower semi-continuous convex function. For all $x, y \in \mathbb{R}^n$, we have the characterization

$$\varphi(x) + \varphi^*(y) = x \cdot y \quad \text{if and only if} \quad y \in \partial\varphi(x).$$

Theorem 1.36. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, the following statements hold.

1. The map φ is differentiable almost everywhere with respect to the Lebesgue measure.
2. If the map φ is differentiable at x, then its subgradient at x is given by

$$\partial\varphi(x) = \{\nabla\varphi(x)\}.$$

Theorem 1.37. Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function. Then, the following are equivalent.

1. The map φ is convex and lower semi-continuous.
2. We have $\varphi = \psi^*$ for some proper function $\psi : \mathbb{R}^n \to \mathbb{R}$.
3. We have $\varphi^{**} = \varphi$.

Equipped with the tools from convex analysis, we state a characterization for the optimality of transport plans known as the Knott-Smith criterion. A generalization to the case where X and Y are Polish spaces equipped with a general cost function is provided in Section 1.3.2.

Theorem 1.38. Let μ and ν be probability measures on $X, Y \subseteq \mathbb{R}^n$ respectively with finite second moments. Then, the transport plan $\pi^\dagger \in \Pi(\mu,\nu)$ is a minimizer to the Kantorovich problem with cost $c(x,y) = \frac{1}{2}\|x-y\|^2$ if and only if there exists a convex, lower semi-continuous function $\tilde{\varphi} \in L^1(\mu)$ such that

$$\text{supp } \pi^\dagger \subseteq \partial\tilde{\varphi},$$

which is equivalent to $y \in \partial\tilde{\varphi}(x)$ for $\pi^\dagger$-almost every $(x,y) \in X \times Y$. Furthermore, the pair $(\tilde{\varphi}, \tilde{\varphi}^*)$ satisfies

$$\mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) = \min_{(\overline{\varphi},\overline{\psi})\in\tilde{\Phi}} \mathbb{J}(\overline{\varphi}, \overline{\psi}),$$

where the set $\tilde{\Phi}$ is defined by

$$\tilde{\Phi} = \left\{(\overline{\varphi},\overline{\psi}) \in L^1(\mu) \times L^1(\nu) : \overline{\varphi}(x) + \overline{\psi}(y) \geq x \cdot y\right\}.$$

Proof. We start by expressing the Kantorovich problem in the Euclidean context. Let $(\varphi, \psi) \in \Phi_c$. Define

$$\tilde{\varphi}(x) = \frac{1}{2}\|x\|^2 - \varphi(x) \quad \text{and} \quad \tilde{\psi}(y) = \frac{1}{2}\|y\|^2 - \psi(y).$$

Since both μ and ν have finite second moments, we have that $\tilde{\varphi} \in L^1(\mu)$ and $\tilde{\psi}(y) \in L^1(\nu)$. Moreover, we have

$$\tilde{\varphi}(x) + \tilde{\psi}(y) = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \varphi(x) - \psi(y) \geq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x-y\|^2 = x \cdot y$$

Furthermore, it holds that

$$\begin{aligned}\varphi(x) + \psi(y) &= \frac{1}{2}\|x\|^2 - \tilde{\varphi}(x) + \frac{1}{2}\|y\|^2 - \tilde{\psi} \\ &\leq \frac{1}{2}\left(\|x\|^2 + \|y\|^2 - x \cdot y\right) \\ &= c(x,y).\end{aligned}$$

Hence, we have shown that $(\varphi, \psi) \in \Phi_c$ if and only if $(\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi}$. Moreover, if

$$M = \frac{1}{2}\int_X \|x\|^2 \, d\mu(x) + \frac{1}{2}\int_Y \|y\|^2 \, d\nu(y),$$

it also follows that $\mathbb{J}(\tilde{\varphi}, \tilde{\psi}) = M - \mathbb{J}(\varphi, \psi)$. Thus, any transport plan $\pi \in \Pi(\mu,\nu)$ satisfies

$$\mathbb{K}(\pi) = \frac{1}{2}\int_{X\times Y} \|x-y\|^2 \, d\pi(x,y) = M - \int_{X\times Y} x \cdot y \, d\pi(x,y),$$

which implies that

$$M - \mathbb{J}(\tilde{\varphi}, \tilde{\psi}) = \mathbb{J}(\varphi, \psi) \leq \mathbb{K}(\pi) = M - \int_{X\times Y} x \cdot y \, d\pi(x,y).$$

Applying Theorem 1.19, we obtain

$$\max_{(\overline{\varphi},\overline{\psi})\in\tilde{\Phi}} M - \mathbb{J}(\overline{\varphi}, \overline{\psi}) = \max_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) = \min_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) = \min_{\pi\in\Pi(\mu,\nu)} M - \int_{X\times Y} x \cdot y \, d\pi(x,y),$$

which yields

$$\min_{(\overline{\varphi},\overline{\psi})\in\tilde{\Phi}} \mathbb{J}(\overline{\varphi},\overline{\psi}) = \max_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} x\cdot y \; d\pi(x,y). \tag{1.16}$$

Note that if $\pi^\dagger$ minimizes $\mathbb{K}$, then $\pi^\dagger$ also maximizes the integral in the right-hand side of (1.16) and vice versa. Similarly, if $(\varphi,\psi)\in\Phi_c$ maximizes $\mathbb{J}$, then $(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}$ minimizes $\mathbb{J}$ and vice versa. Using Theorem 1.30, we can further characterize the pair $(\tilde{\varphi},\tilde{\psi})$ as

$$\tilde{\varphi}(x) = \frac{1}{2}\|x\|^2 - \varphi^{cc}(x) \quad \text{and} \quad \tilde{\psi}(y) = \frac{1}{2}\|y\|^2 - \varphi^{c}(y).$$

This means that

$$\begin{aligned}
\tilde{\psi}(y) &= \frac{1}{2}\|y\|^2 - \varphi^{c}(y)\\
&= \sup_{x\in X}\left(\frac{1}{2}\|y\|^2 + \varphi(x) - \frac{1}{2}\|x-y\|^2\right)\\
&= \sup_{x\in X}(x\cdot y - \tilde{\varphi}(x))\\
&= \tilde{\varphi}^*(y),
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\varphi}(x) &= \frac{1}{2}\|x\|^2 - \varphi^{cc}(x)\\
&= \sup_{y\in Y}\left(\frac{1}{2}\|x\|^2 - \frac{1}{2}\|x-y\|^2 + \varphi^{c}(y)\right)\\
&= \sup_{y\in Y}\left(\frac{1}{2}\|x\|^2 - \frac{1}{2}\|x-y\|^2 + \varphi^{c}(y)\right)\\
&= \sup_{y\in Y}\left(\frac{1}{2}\|x\|^2 - \frac{1}{2}\|x-y\|^2 + \frac{1}{2}\|y\|^2 - \tilde{\psi}(y)\right)\\
&= \sup_{y\in Y}\left(\frac{1}{2}\|x\|^2 - \frac{1}{2}\|x-y\|^2 + \frac{1}{2}\|y\|^2 - \tilde{\varphi}^*(y)\right)\\
&= \sup_{y\in Y}(x\cdot y - \tilde{\varphi}^*(y))\\
&= \tilde{\varphi}^{**}(x).
\end{aligned}$$

In other words, pairs minimizing $\mathbb{J}$ are of the expected form $(\tilde{\varphi}^{**},\tilde{\varphi}^*)$. Using Theorem 1.37, we have that $\tilde{\varphi}$ is convex and lower semi-continuous with $\tilde{\varphi}^{***} = \tilde{\varphi}^*$. Thus, we have shown that

$$\mathbb{J}(\tilde{\varphi},\tilde{\varphi}^*) = \min_{(\overline{\varphi},\overline{\psi})\in\tilde{\Phi}} \mathbb{J}(\overline{\varphi},\overline{\psi}), \tag{1.17}$$

where $\tilde{\varphi}$ is a proper, convex and lower semi-continuous function.

Now that we have expressed the Kantorovich problem in the Euclidean context, we prove the Knott-Smith criterion. Let $\pi^\dagger\in\Pi(\mu,\nu)$ be a minimizer and $\tilde{\varphi}$ be a proper, convex, lower semi-continuous function as given in (1.17). By (1.16), we have

$$\int_X \tilde{\varphi}(x)\; d\mu(x) + \int_Y \tilde{\varphi}^*(y)\; d\nu(y) = \int_{X\times Y} x\cdot y\; d\pi^\dagger(x,y),$$

which can be expressed as

$$\int_{X\times Y} [\tilde{\varphi}(x) + \tilde{\varphi}^*(y) - x\cdot y]\; d\pi^\dagger(x,y) = 0. \tag{1.18}$$

Using Remark 1.24, we obtain

$$x \cdot y \leq \tilde{\varphi}(x) + \tilde{\varphi}^*(y),$$

from which we can deduce that the integral in (1.18) is non-negative. This implies that

$$\tilde{\varphi}(x) + \tilde{\varphi}^*(y) = x \cdot y \text{ for } \pi^\dagger\text{-almost every } (x, y) \in X \times Y.$$

Hence, we can deduce that $y \in \partial\tilde{\varphi}(x)$ for $\pi^\dagger$-almost every $(x, y) \in X \times Y$ by Theorem 1.35.

Conversely, suppose that $y \in \partial\tilde{\varphi}(x)$ for $\pi^\dagger$-almost every $(x, y) \in X \times Y$ where $\varphi \in L^1(\mu)$ is a proper, convex and lower semi-continuous function. Using Theorem 1.35, we have that $\tilde{\varphi}(x) + \tilde{\varphi}^*(y) = x \cdot y$ for $\pi^\dagger$-almost every $(x, y) \in X \times Y$. This implies that

$$\int_X \tilde{\varphi}(x)\ d\mu(x) + \int_Y \tilde{\varphi}^*(y)\ d\nu(y) = \int_{X\times Y} x \cdot y\ d\pi^\dagger(x, y).$$

To show that $(\tilde{\varphi}, \tilde{\varphi^*}) \in \tilde{\Phi}$, it suffices to show that $\varphi^* \in L^1(\nu)$. By Remark 1.24, note that

$$\tilde{\varphi}^*(y) \geq x_0 \cdot y - \tilde{\varphi}(x_0),$$

for some $x_0 \in X$. Define $f(y) = x_0 \cdot y - \tilde{\varphi}(x_0)$. This means that

$$\begin{aligned}
\|\tilde{\varphi}^* - f\|_{L^1(\nu)} &= \int_Y [\tilde{\varphi}^*(y) - f(y)]\ d\nu(y) \\
&= \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) - \int_X \tilde{\varphi}(x)\ d\mu(x) - \int_Y f(y)\ d\nu(y) \\
&\leq \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) + \int_X |\tilde{\varphi}(x)|\ d\mu(x) + \int_Y |f(y)|\ d\nu(y) \\
&= \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) + \|\tilde{\varphi}\|_{L^1(\mu)} + \int_Y |x_0 \cdot y|\ d\nu(y) + |\tilde{\varphi}(x_0)| \\
&\leq \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) + \|\tilde{\varphi}\|_{L^1(\mu)} + \frac{1}{2}\|x_0\|^2 + \frac{1}{2}\int_Y \|y\|^2\ d\nu(y) + \tilde{\varphi}(x_0) \\
&< \infty,
\end{aligned}$$

since ν has finite second moments. This means that $\tilde{\varphi}^* - f \in L^1(\nu)$ and so $\tilde{\varphi}^* \in L^1(\nu)$ since $f \in L^1(\nu)$. This shows that the pair $(\tilde{\varphi}, \tilde{\varphi}^*) \in \tilde{\Phi}$. Hence, using (1.16), we obtain the chain of equalities

$$\min_{(\overline{\varphi}, \overline{\psi}) \in \tilde{\Phi}} \mathbb{J}(\overline{\varphi}, \overline{\psi}) = \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) = \int_{X\times Y} x \cdot y\ d\pi^\dagger(x, y) = \max_{\pi \in \Pi(\mu,\nu)} \int_{X\times Y} x \cdot y\ d\pi(x, y).$$

Thus, we have that $\pi^\dagger$ is a minimizer to the Kantorovich problem. □

The Knott-Smith criterion is useful in providing sufficient conditions for the existence and uniqueness of an optimal transport plan under a slight strengthening of the assumptions on the probability measure μ. This is known as Brenier's theorem.

Definition 1.39. A probability measure μ does not give mass to small sets if $\mu(A) = 0$ for all measurable sets $A \subseteq \mathbb{R}^n$ with Hausdorff dimension at most $n - 1$ [5].

Theorem 1.40. Let μ and ν be probability measures on $X, Y \subseteq \mathbb{R}^n$ respectively with finite second moments with the added assumption that μ does not give mass to small sets. Then, there exists a unique solution $\pi^\dagger \in \Pi(\mu,\nu)$ to the Kantorovich problem with cost $c(x,y) = \frac{1}{2}\|x-y\|^2$ of the form

$$\pi^\dagger = (\mathrm{Id} \times \nabla\varphi)_\# \mu \quad \text{which is equivalent to} \quad d\pi^\dagger(x,y) = \delta_{y=\nabla\varphi(x)}\, d\mu(x),$$

where $\nabla\varphi$ is the gradient of a convex function defined μ-almost everywhere satisfying $(\nabla\varphi)_\#\mu = \nu$. In other words, the map $\nabla\varphi$ is a transport map from μ to ν.

Proof. Suppose that $\pi^\dagger$ is a minimizer to the Kantorovich problem obtained from Theorem 1.18. By [1, Theorem 5.3.1], we can write

$$\pi^\dagger(A \times B) = \int_A \pi^\dagger(B|x)\, d\mu(x),$$

where $\{\pi^\dagger(\cdot|x)\}_{x\in X}$ is a family of probability measures on Y. By Theorem 1.38, we have

$$\operatorname{supp} \pi^\dagger(\cdot|x) \subseteq \partial\varphi(x) \quad \text{for } \mu\text{-almost every } x \in X$$

for some $\varphi \in L^1(\mu)$ proper, convex and lower semi-continuous. By Theorem 1.36, we have that $\partial\varphi(x) = \{\nabla\varphi(x)\}$ for almost every $x \in X$ with respect to the Lebesgue measure. Since μ does not give mass to small sets, the previous statement also holds for μ-almost every $x \in X$. This means that

$$\operatorname{supp} \pi^\dagger(\cdot|x) \subseteq \{\nabla\varphi(x)\} \quad \text{for } \mu\text{-almost every } x \in X.$$

Hence, it follows that the measures $\pi^\dagger(\cdot|x) = \delta_{\nabla\varphi(x)}$ for μ-almost every $x \in X$. Hence, we can express $\pi^\dagger$ as

$$\pi^\dagger = (\mathrm{Id} \times \nabla\varphi)_\# \mu.$$

Moreover, for a measurable subset $B \subseteq Y$, we have

$$\begin{aligned}
\nu(B) &= \pi^\dagger(\mathbb{R}^n \times B) \\
&= (\mathrm{Id} \times \nabla\varphi)_\#\mu(\mathbb{R}^n) \\
&= \mu((\mathrm{Id} \times \nabla\varphi)^{-1}(\mathbb{R}^n \times B)) \\
&= \mu(\{x \in X : (\mathrm{Id} \times \nabla\varphi)(x) \in \mathbb{R}^n \times B\}) \\
&= \mu(\{x \in X : \nabla\varphi(x) \in B\}) \\
&= \mu((\nabla\varphi)^{-1}(B)) \\
&= (\nabla\varphi)_\#\mu(B),
\end{aligned}$$

from which we can deduce that $(\nabla\varphi)_\#\mu = \nu$. It remains to be shown that the solution is unique. Suppose that $\bar{\varphi}$ is another convex function with $(\nabla\bar{\varphi})_\#\mu = \nu$. By Theorem 1.38, we have that $\operatorname{supp}(\mathrm{Id} \times \nabla\bar{\varphi})_\#\mu \subseteq \partial\bar{\varphi}$ which implies that $(\mathrm{Id} \times \nabla\bar{\varphi})_\#\mu$ is an optimal transport plan. Moreover, the pair $(\bar{\varphi}, \bar{\varphi}^*)$ minimizes $\mathbb{J}$ over the set $\tilde{\Phi}$. Thus, we have the identity

$$\int_X \bar{\varphi}\, d\mu + \int_Y \bar{\varphi}^*\, d\nu = \int_X \varphi\, d\mu + \int_Y \varphi^*\, d\nu.$$

We can use the above in combination with the fact that $y \in \partial\varphi(x)$ for $\pi^\dagger$-almost every $(x, y) \in X \times Y$ to obtain

$$\begin{aligned}\int_{X\times Y} [\bar{\varphi}(x) + \bar{\varphi}^*(y)]\, d\pi^\dagger(x, y) &= \int_{X\times Y} [\varphi(x) + \varphi^*(y)]\, d\pi^\dagger(x, y)\\ &= \int_{X\times Y} x \cdot y\, d\pi^\dagger(x, y)\\ &= \int_{X\times Y} x \cdot y\, d(\mathrm{Id} \times \nabla\varphi)_{\#}\mu(x, y)\\ &= \int_X x \cdot \nabla\varphi(x)\, d\mu(x).\end{aligned}$$

In addition, we also have

$$\begin{aligned}\int_{X\times Y} [\bar{\varphi}(x) + \bar{\varphi}^*(y)]\, d\pi^\dagger(x, y) &= \int_{X\times Y} [\bar{\varphi}(x) + \bar{\varphi}^*(y)]\, d(\mathrm{Id} \times \nabla\varphi)_{\#}\mu(x, y)\\ &= \int_X [\bar{\varphi}(x) + \bar{\varphi}(\nabla\varphi(x))]\, d\mu(x).\end{aligned}$$

Thus, we can deduce that

$$\int_X [\bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x)) - x \cdot \nabla\varphi(x)]\, d\mu(x) = 0.$$

This means that

$$\bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x)) - x \cdot \nabla\varphi(x) = 0 \quad \text{for } \mu\text{-almost every } x \in X.$$

By Theorem 1.35, we can deduce that

$$\nabla\varphi(x) \in \partial\bar{\varphi}(x) \quad \text{for } \mu\text{-almost every } x \in X,$$

and so we conclude

$$\nabla\varphi(x) = \nabla\bar{\varphi}(x) \quad \text{for } \mu\text{-almost every } x \in X.$$

□

Note that Brenier's theorem can be used to prove the existence and uniqueness of an optimal transport map from μ to ν in the Euclidean context.

Corollary 1.41. Under the same assumptions as Theorem 1.40, we have that $\nabla\varphi$ is the unique solution to the Monge problem

$$\frac{1}{2}\int_X \|x - \nabla\varphi(x)\|^2\, d\mu(x) = \frac{1}{2}\inf_{T_{\#}\mu=\nu} \int_X \|x - T(x)\|^2\, d\mu(x).$$

Proof. Let $\pi^\dagger$ be the unique minimizer to the Kantorovich problem and $T^\dagger = \nabla\varphi$ be the corresponding transport map from μ to ν obtained from Theorem 1.40. Recall from Remark 1.11 that

$$\min_{\pi\in\Pi(\mu,\mu)} \mathbb{K}(\pi) \leq \inf_{T_{\#}\mu=\nu} \mathbb{M}(T).$$

Using the fact that $T^\dagger(x) = y$ for $\pi^\dagger$-almost every $(x, y) \in X \times Y$, we have

$$\begin{aligned}
\mathbb{M}(T^\dagger) &= \frac{1}{2} \int_X \|x - T^\dagger(x)\|^2 \, d\mu(x) \\
&= \frac{1}{2} \int_{X \times Y} \|x - T^\dagger(x)\|^2 \, d\pi^\dagger(x, y) \\
&= \frac{1}{2} \int_{X \times Y} \|x - y\|^2 \, d\pi^\dagger(x, y) \\
&= \min_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) \\
&\leq \inf_{T_\# \mu = \nu} \mathbb{M}(T).
\end{aligned}$$

Since $T^\dagger$ is a transport map from μ to ν, we have

$$\mathbb{M}(T^\dagger) = \min_{T_\# \mu = \nu} \mathbb{M}(T),$$

and so $T^\dagger$ is an optimal transport map from μ to ν. Furthermore, this map is unique by the uniqueness of $\pi^\dagger$. □

Lastly, explicitly determining the optimal transport map in a Euclidean setting is equivalent to solving a second order, nonlinear partial differential equation.

Corollary 1.42. Under the same assumptions as Theorem 1.40, suppose that μ and ν are absolutely continuous with respect to the Lebesgue measure, with

$$\frac{d\mu}{dx} = f \quad \text{and} \quad \frac{d\nu}{dy} = g,$$

where $f : X \to [0, +\infty]$ and $g : Y \to [0, +\infty]$ are non-negative measurable functions. Moreover, assume that $T = \nabla\varphi$ is a C^1-diffeomorphism. Then, solving the Monge problem is equivalent to solving the Monge-Ampère equation

$$g \circ \nabla\varphi \cdot \det\left(\nabla^2 \varphi\right) = f.$$

Proof. Since T is a transport map from μ to ν, it follows that

$$\int_X \psi(T(x)) \, d\mu(x) = \int_Y \psi(y) \, d\nu(y),$$

for all non-negative functions $\psi : Y \to \mathbb{R}$. By absolute continuity of the measures, we have

$$\int_X \psi(T(x)) f(x) \, dx = \int_Y \psi(y) g(y) \, dy.$$

Since T is a C^1-diffeomorphism, we make a change of variables $y = T(x)$ to obtain

$$\int_X \psi(T(x)) f(x) \, dx = \int_X \psi(T(x)) g(T(x)) |\det(\nabla T(x))| \, dx.$$

Since ψ is an arbitrary non-negative function, these two integrals are equal if and only if

$$f(x) = g(T(x)) |\det(\nabla T(x))|.$$

By Theorem 1.40, we have that $T = \nabla\varphi$ where φ is a convex function. Hence, we have

$$(g \circ \nabla\varphi)(x)\det(\nabla^2\varphi(x)) = f(x).$$

The absolute value signs can be omitted by convexity of φ since $\nabla^2\varphi(x) \geq 0$ for all $x \in X$. □

1.3.2 General Setting

In this section, we generalize tools from convex analysis using the cost function inherent to the optimal transport problem to prove a statement similar to the Knott-Smith criterion in the general setting using an approach developed by Rüschendorf [19]. We start by recalling the definition of c-transform, which is used as a stepping stone to proving more sophisticated results.

Definition 1.43. Let $\varphi : X \to \bar{\mathbb{R}}$ be a real-valued function on X. Then, the c-transforms φ^c and φ^{cc} are defined by

$$\begin{aligned}\varphi^c : Y &\to \bar{\mathbb{R}}\\ y &\mapsto \inf_{x\in X}(c(x,y) - \varphi(x)),\\ \varphi^{cc} : X &\to \bar{\mathbb{R}}\\ x &\mapsto \inf_{y\in Y}(c(x,y) - \varphi^c(y)).\end{aligned}$$

Definition 1.44. The map $\varphi : X \to \bar{\mathbb{R}}$ is c-concave if there exists $\zeta : Y \to \bar{\mathbb{R}}$ such that $\varphi = \zeta^c$. Equivalently, we have $\varphi^{cc} = \varphi$.

Definition 1.45. A subset $\Gamma \subseteq X \times Y$ is c-cyclically monotone if for all $N \in \mathbb{N}$ and any family $(x_1, y_1), \ldots, (x_N, y_N) \in \Gamma$, it follows that

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

where we use the convention $y_{N+1} = y_1$.

The notion of cyclical monotonicity has an intuitive interpretation in the context of factories and bakeries. Suppose we pick N arbitrary elements in Γ and any element (x_i, y_i) has a factory at x_i and a bakery at y_i. Given a transport plan, we attempt to decrease the total cost of transportation by rerouting a unit of bread from y_1 to y_2 that is closer to x_1. This procedure leads to a gain of $c(x_1, y_2) - c(x_1, y_1)$. However, this means that bakery y_2 will have an excess of bread in their inventory.

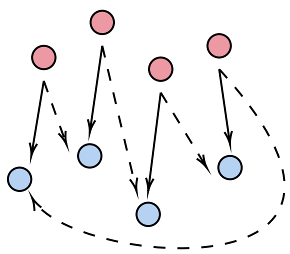


FIGURE 1.5: The initial transport plan is represented by the solid arrows. An attempt in improving the total cost of transportation by rerouting is represented by the dashed arrows.

Thus, another unit of bread needs to be rerouted to another bakery y_3. We proceed inductively until we reach factory x_N rerounting bread to y_1, which yields a new transport plan. This new transport plan decreases the total cost of transportation if

$$\sum_{i=1}^{N} c(x_i, y_{i+1}) < \sum_{i=1}^{N} c(x_i, y_i).$$

In essence, c-cyclically monotone sets are optimal in the sense that the total cost of transportation can never be reduced by rerouting bread to other bakeries.

In the Euclidean setting, recall that the subdifferential of a convex represents the set of gradients such that the linearization remains below the graph of the function. A similar notion exists for concave functions where we consider the set of gradients such that the linearization remains above the graph of the function. This is known as the superdifferential of a concave function, which can be generalized to a general setting as follows.

Definition 1.46. Let $\varphi : X \to \bar{\mathbb{R}}$ be a c-concave function. Then, the c-superdifferential of φ at x is defined by

$$\partial^c \varphi(x) = \{y \in Y : \ \varphi(z) \leq \varphi(x) + c(z, y) - c(x, y) \quad \text{for all } z \in X\}.$$

Moreover, we write

$$\partial^c \varphi = \{(x, y) \in X \times Y : y \in \partial^c \varphi(x)\}.$$

It turns out that a similar version of Theorem 1.35 exists for c-superdifferentials. In particular, the c-superdifferential can be viewed as the set of points in $X \times Y$ which achieve equality in the constraints of the dual Kantorovich problem when we consider c-transform pairs as in Lemma 1.31. In addition, we have that c-superdifferentials are c-cyclically monotone sets.

Theorem 1.47. Let $\varphi : X \to \bar{\mathbb{R}}$ be a c-concave function. Then, the c-superdifferential can be expressed as

$$\partial^c \varphi = \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Moreover, the c-superdifferential $\partial^c \varphi$ is c-cyclically monotone.

Proof. Since $\varphi^c(y) \leq c(x, y) - \varphi(x)$, we have that

$$\begin{aligned}
c(x, y) = \varphi(x) + \varphi^c(y) \quad &\Leftrightarrow \quad c(x, y) \leq \varphi(x) + \varphi^c(y) \\
&\Leftrightarrow \quad c(x, y) \leq \varphi(x) - \varphi(z) + c(z, y) \quad \text{for all } z \in X \\
&\Leftrightarrow \quad \varphi(z) \leq \varphi(x) + c(z, y) - c(x, y) \quad \text{for all } z \in X \\
&\Leftrightarrow \quad y \in \partial^c \varphi(x),
\end{aligned}$$

which gives the first result. To show that $\partial^c\varphi$ is c-cyclically monotone, consider N arbitrary elements $(x_1, y_1), \dots, (x_N, y_N) \in \partial^c\varphi$. Then, we obtain

$$\begin{aligned}\sum_{i=1}^N c(x_i, y_i) &= \sum_{i=1}^N [\varphi(x_i) + \varphi^c(y_i)] \\ &= \sum_{i=1}^N [\varphi(x_i) + \varphi^c(y_{i+1})] \\ &\leq \sum_{i=1}^N c(x_i, y_{i+1}).\end{aligned}$$

This shows that $\partial^c\varphi$ is c-cyclically monotone. □

At last, we consider a characterization of the optimality of transport plans in a general setting. The statement is a special case of [24, Theorem 5.10] where the cost function is assumed to be continuous and positive. Moreover, stronger conditions on the functions bounding the cost function from above are imposed.

Theorem 1.48. Let μ and ν be probability measures on Polish spaces X and Y respectively and $c : X \times Y \to [0, +\infty]$ be a continuous cost function. Suppose that there exist $c_X \in C^0 \cap L^1(\mu)$, $c_Y \in C^0 \cap L^1(\nu)$ satisfying $c(x, y) \leq c_X(x) + c_Y(y)$. Then, we have

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) = \max_{\varphi \in L^1(\mu)} \mathbb{J}(\varphi, \varphi^c), \tag{1.19}$$

and there exists a c-cyclically monotone set $\Gamma \subseteq X \times Y$ such that for all $\pi \in \Pi(\mu, \nu)$ and any c-concave $\varphi \in L^1(\mu)$,

1. The measure π is optimal in the Kantorovich problem if and only if $\pi(\Gamma) = 1$.

2. The map φ is optimal in the dual Kantorovich problem if and only if $\Gamma \subseteq \partial^c\varphi$.

Proof. Note that (1.19) follows by applying Theorems 1.18, 1.19 and 1.30. By Theorem 1.47, we have that the c-superdifferential of a c-concave function is a c-cyclically monotone set. We define

$$\Gamma_{\max} = \bigcap_{\varphi \text{ optimal}} \partial^c\varphi,$$

where the intersection is taken over all optimal maps in the Kantorovich dual problem. We show that $\Gamma_{\max}$ characterizes optimality in both Kantorovich problems.

Let $\pi^\dagger \in \Pi(\mu, \nu)$ be a minimizer to the Kantorovich problem and $(\varphi, \varphi^c) \in \Phi_c$ a maximizer to the dual Kantorovich problem. It is clear that $\Gamma_{\max} \subseteq \partial^c\varphi$. On the other hand, we use (1.19) to obtain

$$\int_{X\times Y} c \, d\pi^\dagger = \int_X \varphi \, d\mu + \int_Y \varphi^c \, d\nu.$$

Since $\pi^\dagger$ is a transport plan from μ to ν, we deduce that

$$\int_{X\times Y} [c(x, y) - \varphi(x) - \varphi^c(y)] \, d\pi^\dagger(x, y) = 0.$$

Using the constraint given by Φ_c, we can deduce that the integrand is non-negative. Thus, it follows that the equality

$$\varphi(x) + \varphi^c(y) = c(x, y),$$

holds $\pi^\dagger$-almost everywhere. By Theorem 1.47, we can deduce that supp $\pi^\dagger \subseteq \partial^c\varphi$ for all optimal maps φ. Hence, it also follows that supp $\pi^\dagger \subseteq \Gamma_{\max}$, which implies that $\pi^\dagger(\Gamma_{\max}) = 1$.

Conversely, suppose that $\pi \in \Pi(\mu, \nu)$ satisfies $\pi(\Gamma_{\max}) = 1$. Then, it follows that supp $\pi \subseteq \partial^c\varphi$ for all optimal map φ. Thus, it follows that

$$\int_{X\times Y} c\ d\pi = \int_{X\times Y} [\varphi(x) + \varphi^c(y)]\ d\pi(x, y) = \int_X \varphi\ d\mu + \int_Y \varphi^c\ d\nu,$$

from which we can conclude that the transport plan π is a minimizer by Theorem 1.19. On the other hand, let $\varphi \in L^1(\mu)$ be a c-concave function such that $\Gamma_{\max} \subseteq \partial^c\varphi$. For the moment, we assume that $\varphi^c \in L^1(\nu)$ to show that φ leads to a maximizing pair. Using Theorem 1.47, we can deduce that $\varphi(x) + \varphi^c(y) = c(x, y)$ for π-almost every $(x, y) \in X \times Y$. This implies that

$$\int_{X\times Y} c(x, y)\ d\pi(x, y) = \int_X \varphi\ d\mu + \int_Y \varphi^c\ d\nu.$$

Thus, we can conclude that (φ, φ^c) is a maximizing pair. To show that $(\varphi, \varphi^c) \in \Phi_c$, it suffices to show that $\varphi^c \in L^1(\nu)$. Using the constraint given by Φ_c, we have that

$$\varphi^c(y) \leq c(x_0, y) - \varphi(x_0),$$

for some $x_0 \in X$. Define $f(y) = c(x_0, y) - \varphi(x_0)$. This means that

$$\begin{aligned}
\|\varphi^c - f\|_{L^1(\nu)} &= \int_Y f(y) - \varphi^c(y)\ d\nu(y) \\
&= \int_Y f(y)\ d\nu(y) + \int_X \varphi(x)\ d\mu(x) - \mathbb{J}(\varphi, \varphi^c) \\
&\leq \int_Y |f(y)|\ d\nu(y) + \int_X |\varphi(x)|\ d\mu(x) - \mathbb{J}(\varphi, \varphi^c) \\
&= \int_Y |c(x_0, y)|\ d\nu(y) + |\varphi(x_0)| + \|\varphi\|_{L^1(\mu)} - \mathbb{J}(\varphi, \varphi^c) \\
&\leq |c_X(x_0)| + \int_Y |c_Y(y)|\ d\nu(y) + |\varphi(x_0)| + |\varphi|_{L^1(\mu)} - \mathbb{J}(\varphi, \varphi^c) \\
&< \infty,
\end{aligned}$$

since c_Y is integrable. This means that $\varphi^c - f \in L^1(\nu)$ and so $\varphi^c \in L^1(\nu)$ since $f \in L^1(\nu)$. This shows that the pair $(\varphi, \varphi^c) \in \Phi_c$.

□

Remark 1.49. The same proof can be used to obtain a more general version of Theorem 1.48 exists where the cost function can take negative values as long as it can be bounded below.

Remark 1.50. From Theorem 1.48, we see that the set $\Gamma_{\max}$ is a maximal set characterizing optimality. A minimal set characterizing optimality can also be obtained by taking

$$\Gamma_{\min} = \bigcup_{\pi \text{ optimal}} \operatorname{supp} \pi.$$

We first consider the Kantorovich problem. Suppose that π is an optimal transport plan. Then, we have that $\operatorname{supp} \pi \subseteq \Gamma_{\min}$, which implies that $\pi(\Gamma_{\min}) = 1$. Conversely, in the case that $\pi(\Gamma_{\min}) = 1$, we can deduce that $\pi(\Gamma_{\max}) = 1$ since $\Gamma_{\min} \subseteq \Gamma_{\max}$. Hence, we have that π is an optimal transport plan by the proof of Theorem 1.48.

The fact that $\Gamma_{\min}$ characterizes optimality in the dual Kantorovich problem follows from the proof of Theorem 1.48 and the inclusions $\Gamma_{\min} \subseteq \Gamma_{\max} \subseteq \partial^c\varphi$ for any optimal map φ.

Conclusion

In the Euclidean setting, we use tools from convex analysis to obtain a characterization of optimal transport plans known as the Knott-Smith criterion. Moreover, Brenier's theorem provides sufficient conditions for the existence of a transport map in the Monge problem. Lastly, we use Rüschendorf's theoretical framework to generalize the tools from convex analysis to deduce a general statement of the Knott-Smith criterion. This optimality criterion yields two c-cyclically monotone sets which characterize optimality in both Kantorovich problems. The first minimal set consists of the union of the supports of all optimal transport plans whereas the second maximal set is defined by the intersection of the c-superdifferentials of all optimal maps in the dual Kantorovich problem.

Chapter 2

Aubry-Mather Theory

In this chapter, we present, without proof, a summary of the results from my work [26] which is based on *Action-Minimizing Methods in Hamiltonian Dynamics* by Alfonso Sorrentino [20].

2.1 Tonelli Lagrangians and Hamiltonians on Compact Manifolds

2.1.1 Lagrangian Setting

Many problems in classical mechanics come down to solving minimization problems. In fact, most laws of nature can be stated as variational principles. Arguably, the most famous such principle is the principle of least action, which states that the trajectories of a system correspond to the stationary points of the system's action functional. This action functional is defined in terms of a Lagrangian, a function encoding the dynamics of the system from which the equations of motion can be obtained. However, we can modify the Lagrangian in different ways without changing the equations of motion. This recurring theme of investigating mathematical objects that remain preserved under certain transformations is central to Aubry-Mather theory which describes invariant sets on the tangent bundle providing valuable information about the dynamics of the system.

Definition 2.1. A Lagrangian on a smooth, compact and connected manifold M is a function $L : \mathrm{T}M \to \mathbb{R}$ of smoothness class C^2. The action functional along a continuous piecewise C^1 curve $\gamma : [a, b] \to M$ is defined by

$$A_L(\gamma) = \int_a^b L(\gamma(t), \dot{\gamma}(t))\, dt.$$

Theorem 2.2. The curves $\gamma : [a, b] \to M$ extremizing the action functional A_L correspond correspond to the solutions to the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t)) = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t)). \tag{2.1}$$

Observe that by applying the time derivative in the left-hand side of (2.1), we obtain

$$\frac{\partial^2 L}{\partial v^2}(\gamma(t), \dot{\gamma}(t))\ddot{\gamma}(t) = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t)) - \frac{\partial^2 L}{\partial v \partial x}(\gamma(t), \dot{\gamma}(t))\dot{\gamma}(t).$$

Hence, it follows that the Legendre condition

$$\det \frac{\partial^2 L}{\partial v^2} \neq 0,$$

guarantees the invertibility of the Hessian of the function L in v, which allows us to define a vector field X_L on the tangent bundle such that the solutions of

$$\ddot{\gamma}(t) = X_L(\gamma(t), \dot{\gamma}(t))$$

are precisely the curves satisfying the Euler-Lagrange equation.

Definition 2.3. Suppose that L is a Lagrangian satisfying the Legendre condition. Then, the vector field X_L is called the Euler-Lagrange vector field and its corresponding flow Φ_t^L is the Euler-Lagrange flow associated with L.

However, note that our current definition for the Lagrangian is too broad to carry out any meaningful analysis of physical systems. To remain in a general setting, we impose some conditions which correspond to the properties satisfied by the Lagrangians of many known systems.

Definition 2.4. A function $L : \mathrm{T}M \to \mathbb{R}$ is a Tonelli Lagrangian if

1. The function L is of smoothness class C^2.
2. The Hessian of the function L in v, given by
$$\frac{\partial^2 L}{\partial v^2}(x, v)$$
is positive definite as a quadratic form for all $(x, v) \in \mathrm{T}M$. We say that L is strictly convex in each fiber.
3. The function L satisfies
$$\lim_{\|v\|_x \to \infty} \frac{L(x, v)}{\|v\|_x} = +\infty$$
for all $x \in M$. We say that L is superlinear in each fiber.

Remark 2.5. If L is a Tonelli Lagrangian, then strict convexity in each fiber implies that the Legendre condition is satisfied. Hence, the Euler-Lagrange vector field X_L and its corresponding Euler-Lagrange flow Φ_t^L are well-defined.

Remark 2.6. Observe that L is superlinear in each fiber if and only if for all $x \in M$, it follows that for all $A \in \mathbb{R}$, there exists $B \in \mathbb{R}$ such that

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

In a physical setting, the principle of conservation of energy states that the total energy of a system is conserved over time. Thus, to give a meaningful definition of energy in the Lagrangian framework, its time derivative must vanish. In fact, the energy of a system can be expressed in terms of the Lagrangian as follows.

Definition 2.7. The energy of the system described by a Tonelli Lagrangian L is defined by

$$\begin{aligned} E : \mathrm{T}M &\to \mathbb{R} \\ (x, v) &\mapsto \left\langle \frac{\partial L}{\partial v}(x, v), v \right\rangle_x - L(x, v). \end{aligned}$$

Moreover, the level sets of the energy function E are called energy levels.

Theorem 2.8. The energy of the system described by a Tonelli Lagrangian L is conserved under the Euler-Lagrange flow. In other words, for all $(x, v) \in \mathrm{T}M$ and all $t \in \mathbb{R}$, we have

$$\frac{d}{dt} E(\Phi_t^L(x, v)) = 0.$$

Corollary 2.9. The energy levels are compact and invariant under the Euler-Lagrange flow.

2.1.2 Hamiltonian Setting

For a physical system described by a Lagrangian L, we can shift perspective and apply the Legendre-Fenchel transform to obtain the Hamiltonian, a function defined on the cotangent bundle that is intricately linked to the total energy of the system.

Definition 2.10. Let $L : \mathrm{T}M \to \mathbb{R}$ be a Lagrangian. Then, the Hamiltonian corresponding to L is defined by

$$\begin{aligned} H : \mathrm{T}^*M &\to \mathbb{R} \\ (x, p) &\mapsto \sup_{v \in \mathrm{T}_x M} \left(\langle p, v \rangle_x - L(x, v) \right), \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical pairing between covector fields and vector fields.

We can mirror the properties of Tonelli Lagrangians to define Tonelli Hamiltonians.

Definition 2.11. A function $H : \mathrm{T}^*M \to \mathbb{R}$ is a Tonelli Hamiltonian if

1. The function H is of smoothness class C^2.
2. The Hessian of the function H in p, given by

   $$\frac{\partial^2 H}{\partial p^2}(x, p),$$

   is positive definite as a quadratic form for all $(x, p) \in \mathrm{T}^*M$. We say that H is strictly convex in each fiber.
3. The function H satisfies

   $$\lim_{\|p\|_x \to \infty} \frac{H(x, p)}{\|p\|_x} = +\infty,$$

   for all $x \in M$. We say that H is superlinear in each fiber.

Remark 2.12. Observe that H is superlinear in each fiber if and only if for all $x \in M$, it follows that for all $A \in \mathbb{R}$, there exists $B \in \mathbb{R}$ such that

$$H(x, p) \geq A\|p\|_x - B.$$

However, note that the definition of the Hamiltonian is similar in structure to the definition of the energy function. In light of this fact, we define a function which canonically identifies the tangent and cotangent bundles.

Definition 2.13. Let L be a Tonelli Lagrangian. Then, the Legendre transform is a C^1-diffeomorphism defined by

$$\begin{aligned} \mathcal{L} : \mathrm{T}M &\to \mathrm{T}^*M \\ (x, v) &\mapsto \frac{\partial L}{\partial v}(x, v). \end{aligned}$$

In addition, we can restate the Legendre-Fenchel inequality in the context of Lagrangians and Hamiltonians on smooth compact manifolds.

Theorem 2.14. Let $L : \mathrm{T}M \to \mathbb{R}$ be a Lagrangian and $H : \mathrm{T}^*M \to \mathbb{R}$ be the corresponding Hamiltonian. Then, for all $(x, v) \in \mathrm{T}M$ and all $(x, p) \in \mathrm{T}^*M$, we have

$$\langle p, v\rangle_x \leq L(x, v) + H(x, p),$$

with equality whenever

$$p = \frac{\partial L}{\partial v}(x, v).$$

Remark 2.15. We can use the Legendre-Fenchel equality to describe the energy in terms of the energy and the Legendre transform as follows.

$$E(x, v) = H \circ \mathcal{L}(x, v) = \left\langle \frac{\partial L}{\partial v}(x, v), v \right\rangle_x - L(x, v).$$

In other words, the Hamiltonian measures the total energy of the system where the initial conditions are given by elements in the cotangent bundle. However, something more can be said about the relationship between Tonelli Lagrangians and Tonelli Hamiltonians. It turns out that the properties of Tonelli Lagrangians are preserved under the Legendre-Fenchel transform.

Theorem 2.16. Let $L : \mathrm{T}M \to \mathbb{R}$ be a Tonelli Lagrangian. Then, the corresponding Hamiltonian $H : \mathrm{T}^*M \to \mathbb{R}$ is a Tonelli Hamiltonian.

Remark 2.17. Given a Tonelli Hamiltonian, we can construct a vector field on the cotangent bundle described by Hamilton's equations

$$X_H(x(t), p(t)) = \left(\frac{\partial H}{\partial p}(x(t), p(t)), -\frac{\partial H}{\partial x}(x(t), p(t))\right).$$

Definition 2.18. Suppose that H is a Tonelli Hamiltonian. Then, the vector field X_H is called the Hamiltonian vector field and its corresponding flow Φ_t^H is the Hamiltonian flow associated with H.

Remark 2.19. Legendre transform is a conjugacy between the Euler-Lagrange flow on the tangent bundle and the Hamiltonian flow on the cotangent bundle. Thus, the following diagram commutes.

$$\begin{array}{ccc} \mathrm{T}M & \xrightarrow{\Phi_t^L} & \mathrm{T}M \\ {\scriptstyle \mathcal{L}}\downarrow & & \downarrow{\scriptstyle \mathcal{L}} \\ \mathrm{T}^*M & \xrightarrow{\Phi_t^H} & \mathrm{T}^*M \end{array}$$

Since the Legendre transform $\mathcal{L}$ and the Hamiltonian flow Φ_t^H are of smoothness class C^1, we can deduce by commutativity that the Euler-Lagrange flow is of smoothness class C^1.

Conclusion

Given a Lagrangian, the equations of motion of the corresponding physical system can be obtained by considering the solutions to the Euler-Lagrange equation. These solutions exist whenever the Legendre condition is satisfied. To work in a general setting, we assume that the Lagrangians are strictly convex and superlinear in each fiber. Similar properties can be imposed on the corresponding Hamiltonian. These assumptions provide the right foundation to study of rich invariants describing the dynamics of the system.

2.2 Action-Minimizing Measures and Curves for Tonelli Lagrangians

2.2.1 Action-Minimizing Measures and Mather Sets

We now shift our attention to the minimization of the action functional over different spaces. For the moment, our objective is to minimize the action over the space of probability measures on the tangent bundle that are invariant under the Euler-Lagrange flow.

Definition 2.20. Let L be a Tonelli Lagrangian. Then, the average action of a probability measure $\mu \in \mathcal{P}(\mathrm{T}M)$ is defined by

$$A_L(\mu) = \int_{\mathrm{T}M} L \, d\mu.$$

We aim to minimize this quantity over the set of all probability measures that are invariant under the Euler-Lagrange flow Φ_t^L.

Definition 2.21. A probability measure $\mu \in \mathcal{P}(\mathrm{T}M)$ is invariant if

$$\Phi_t^L{}_{\#}\mu = \mu \quad \text{for all } A \subseteq \mathrm{T}M \text{ measurable.}$$

Notation 2.22. We write $\mathfrak{M}(L)$ for the set of all probability measures $\mu \in \mathcal{P}(\mathrm{T}M)$ that are invariant under Φ_t^L with finite average action.

Moreover, the set $\mathfrak{M}(L)$ is non-empty by a classical result of Kryloff and Bogoliouboff [11] stating that non-empty energy levels contain the support of an invariant measure.

Theorem 2.23. Suppose that $k \in \mathbb{R}$ and the energy level

$$E^{-1}(k) = \{(x, v) \in \mathrm{T}M : E(x, v) = k\},$$

is non-empty. Then, there exists a probability measure $\mu \in \mathfrak{M}(L)$ supported on the energy level $E^{-1}(k)$.

We can give a name to invariant measures minimizing the average action over the space $\mathfrak{M}(L)$. These measures encode valuable information about the dynamics of the system.

Definition 2.24. A probability measure $\mu \in \mathfrak{M}(L)$ is an action-minimizing measure of L if it satisfies

$$A_L(\mu) = \min_{\tilde{\mu} \in \mathfrak{M}(L)} A_L(\tilde{\mu}).$$

To perform analysis on the average action, we need to specify a topology on $\mathfrak{M}(L)$. Let $C_l^0(\mathrm{T}M)$ denote the space of continuous functions $f : \mathrm{T}M \to \mathbb{R}$ having at most linear growth. In other words, functions satisfying

$$\sup_{(x,v) \in \mathrm{T}M} \frac{|f(x, v)|}{1 + \|v\|_x} < \infty.$$

We endow $\mathfrak{M}(L)$ with the vague topology, the weak* topology induced by $C_l^0(\mathrm{T}M)$. Namely, this corresponds to the topology where

$$\mu_n \overset{*}{\rightharpoonup} \mu \quad \text{if and only if} \quad \int_{\mathrm{T}M} f \, d\mu_n \to \int_{\mathrm{T}M} f \, d\mu \quad \text{for all } f \in C_l^0(\mathrm{T}M).$$

Theorem 2.25. The set $\mathfrak{M}(L)$ is compact and the average action functional is lower semi-continuous with respect to the vague topology.

Corollary 2.26. If L is a Tonelli Lagrangian, then there exists an action-minimizing measure μ of L.

What is interesting is that modifying the Lagrangian by a closed 1-form does not have any effect on the dynamics of the system. In other words, the Euler-Lagrange flow on the tangent bundle remains unchanged. We formalize this as follows.

Definition 2.27. Let η be a closed 1-form on M. Then, we can represent η as a function on the tangent bundle as follows.

$$\begin{aligned} \hat{\eta} &: \mathrm{T}M \to \mathbb{R} \\ (x, v) &\mapsto \langle \eta(x), v \rangle_x. \end{aligned}$$

Definition 2.28. Let L be a Tonelli Lagrangian and η be a closed 1-form on M. Then, the Lagrangian shift of L by η is the function defined by

$$\begin{aligned} L_\eta &: \mathrm{T}M \to \mathbb{R} \\ (x, v) &\mapsto L(x, v) - \hat{\eta}(x, v). \end{aligned}$$

Theorem 2.29. Let η be a closed 1-form on M. Then, both L and L_η have the same Euler-Lagrange flow on the tangent bundle $\mathrm{T}M$.

We can also consider the Hamiltonian corresponding to the Lagrangian shift by a closed 1-form.

Theorem 2.30. Let L be a Tonelli Lagrangian and η a closed 1-form. Then the Hamiltonian corresponding to the Lagrangian shift of L by η is given by

$$\begin{aligned} H_\eta &: \mathrm{T}^*M \to \mathbb{R} \\ (x, p) &\mapsto H(x, \eta(x) + p). \end{aligned}$$

However, it turns out that exact 1-forms do not contribute to the average action.

Theorem 2.31. Let L be a Tonelli Lagrangian and $\mu \in \mathfrak{M}(L)$. If η is an exact 1-form on M, then

$$\int_{\mathrm{T}M} \hat{\eta} \; d\mu = 0.$$

Thus, shifting the Lagrangian by a closed 1-form has no effect on the dynamics of the system. Moreover, shifting the Lagrangian by an exact 1-form does not contribute to the average action. This means that adding exact 1-forms does not change the action-minimizing measures of the system, but shifts the values attained by the average action. In light of this fact, we consider families of shifted Lagrangians parametrized over de Rham cohomology classes.

Notation 2.32. We write η_c for a closed 1-form of cohomology class $c \in \mathrm{H}^1(M; \mathbb{R})$.

Thus, the need to distinguish action-minimizing measures obtained from different Lagrangian shifts motivates the following definition.

Definition 2.33. Let η_c be a closed 1-form of cohomology class c. A measure $\mu \in \mathfrak{M}(L)$ is a c-action-minimizing measure for L if it satisfies

$$A_{L_{\eta_c}}(\mu) = \min_{\tilde{\mu} \in \mathfrak{M}(L)} A_{L_{\eta_c}}(\tilde{\mu}).$$

Remark 2.34. Note that the cohomology class of an action-minimizing measure is independent not only of the dynamics of the system but also of the measure. Rather, the cohomology of an action-minimizing measure is intrinsic to the choice of Lagrangian describing the system. To illustrate, suppose that μ is a 0-action-minimizing measure for L. Then, the quantity

$$A_L(\mu) = \int_{\mathrm{T}M} L \, d\mu = \int_{\mathrm{T}M} [L_{\eta_c} - (-\hat{\eta}_c)] \, d\mu,$$

is minimized. So μ is also a $(-c)$-action-minimizer for L_{η_c}.

We can thus define a function which maps each cohomology class c to the negative value of the minimal average action corresponding to the Lagrangian shift by a closed 1-form of cohomology class c. The presence of a negative sign is a matter of convention.

Definition 2.35. Mather's α-function is a well-defined convex function given by

$$\begin{aligned} \alpha : \mathrm{H}^1(M;\mathbb{R}) &\to \mathbb{R} \\ c &\mapsto -\min_{\mu \in \mathfrak{M}(L)} A_{L_{\eta_c}}(\mu). \end{aligned}$$

We can thus relate Mather's α-function with c-action-minimizing measures as follows.

Notation 2.36. We write $\mathfrak{M}_c(L)$ for the subset of c-action-minimizing measures for L and

$$\mathfrak{M}_c = \mathfrak{M}_c(L) = \{\mu \in \mathfrak{M}(L) \colon A_{L_{\eta_c}}(\mu) = -\alpha(c)\}.$$

We are now ready to define the Mather set of cohomology class c.

Definition 2.37. The Mather set of cohomology class c is defined by

$$\tilde{\mathcal{M}}_c = \bigcup_{\mu \in \mathfrak{M}_c} \operatorname{supp} \mu \subseteq \mathrm{T}M.$$

The projection on the base manifold $\mathcal{M}_c = \pi\left(\tilde{\mathcal{M}}_c\right) \subseteq M$ is the projected Mather set of cohomology class c.

Theorem 2.38. The Mather set of cohomology class c is non-empty, closed and invariant under the Euler-Lagrange flow.

2.2.2 Action-Minimizing Curves and Aubry Sets

In this section, we move away from action-minimizing measures and focus on action-minimization problems over spaces of absolutely continuous curves. Using a similar idea as in the previous section, we consider Lagrangian shifts by closed 1-forms since these do not change the Euler-Lagrange flow. To set the scene, we start by fixing two endpoints $x, y \in M$ and real numbers $a < b$. We ask whether there exists an absolutely continuous curve $\gamma : [a, b] \to M$ with $\gamma(a) = x$ and $\gamma(b) = y$ minimizing the action obtained from the shifted Lagrangian. For notational convenience, we introduce the following notation.

Notation 2.39. We write $\mathrm{AC}([a, b], M)$ for the set of absolutely continuous curves $\gamma : [a, b] \to M$. Moreover, for $x, y \in M$, we write

$$C_{[a,b]}(x, y) = \{\gamma \in \mathrm{AC}([a, b], M) : \gamma(a) = x, \gamma(b) = y\}.$$

Lastly, for a fixed $T > 0$, we write $C_T(x, y)$ for the set of curves $C_{[0,T]}(x, y)$.

Theorem 2.40. Let L be a Tonelli Lagrangian and η_c a closed 1-form of cohomology class c. For all $x, y \in M$ and real numbers $a < b$, there exists a curve $\gamma \in C_{[a,b]}(x, y)$ which minimizes the action

$$A_{L_{\eta_c}}(\gamma) = \int_a^b L_{\eta_c}(\gamma(t), \dot{\gamma}(t))\ dt.$$

We can give a name to the minimizing curves obtained from the above theorem.

Definition 2.41. A c-Tonelli minimizer is a curve $\gamma \in C_{[a,b]}(x, y)$ minimizing the action

$$A_{L_{\eta_c}}(\gamma) = \int_a^b L_{\eta_c}(\gamma(t), \dot{\gamma}(t))\ dt.$$

Remark 2.42. Note that c-Tonelli minimizers only depend on the cohomology class of η_c and not on the choice of representative since adding an exact 1-form df to η_c results in a shift in the values attained by the action. This can be seen from

$$A_{L_{\eta_c+df}}(\gamma) = A_{L_{\eta_c}}(\gamma) + f(y) - f(x).$$

In other words, adding an exact 1-form does not have any effect on the selection of the minimizing curves. Moreover, Tonelli minimizers exhibit a certain smoothing property.

Theorem 2.43. Let $\gamma \in C_{[a,b]}(x, y)$ be a c-Tonelli minimizer of smoothness class C^1. Then, the curve γ is of smoothness class C^2 and satisfies the Euler-Lagrange equation.

Until now, we have considered absolutely continuous curves defined on an interval. However, we can also consider action-minimizing curves defined on the whole real line. We can do this in two ways. The first approach considers curves minimizing the action over any given time length.

Definition 2.44. An absolutely continuous curve $\gamma\colon \mathbb{R} \to M$ is a c-minimizer for L if for all real numbers $a < b$, we have

$$A_{L_\eta}(\gamma|_{[a,b]}) = \min A_{L_\eta}(\sigma),$$

where the minimum is taken over all curves $\sigma \in C_{[a,b]}(\gamma(a), \gamma(b))$.

The second formulation can be viewed as a reinforcement of the first formulation by adding the condition that the minimum of the action be realized over all absolutely continuous curves connecting the endpoints, regardless of the time taken.

Definition 2.45. An absolutely continuous curve $\gamma : \mathbb{R} \to M$ is a c-time-free minimizer for L if, for all real numbers $a < b$, we have

$$A_{L_\eta}(\gamma|_{[a,b]}) = \min A_{L_\eta}(\sigma),$$

where the minimum is taken over all $\sigma \in C_{[a',b']}(\gamma(a), \gamma(b))$.

Remark 2.46. Clearly, we have that c-time-free minimizers are also c-minimizers. What is less obvious is that c-time-free minimizers are sensitive to the vertical shifts in the Lagrangian. Suppose that $\gamma : [0, T] \to M$ and $\sigma : [0, T'] \to M$ are curves satisfying

$$\int_0^T L_\eta(\gamma(t), \dot{\gamma}(t))\ dt < \int_0^{T'} L_\eta(\sigma(t), \dot{\sigma}(t))\ dt.$$

Define $k \in \mathbb{R}$ by the inequality

$$\frac{1}{T'-T}\left(\int_0^{T'} L_\eta(\sigma(t),\dot{\sigma}(t))\ dt - \int_0^{T} L_\eta(\gamma(t),\dot{\gamma}(t))\ dt\right) < k.$$

Then, adding this constant k to the Lagrangian reverses the inequality

$$A_{L_\eta+k}(\gamma) \geq A_{L_\eta+k}(\sigma).$$

Some critical notions in the analysis of absolutely continuous curves defined on the whole real line include the Mañé potential and critical value.

Definition 2.47. Let η be a closed 1-form and $k \in \mathbb{R}$. Then, the Mañé potential is defined by

$$\begin{aligned}\phi_{\eta,k}&\colon M\times M \to \mathbb{R}\cup\{-\infty\}\\ \phi_{\eta,k}(x,y) &= \inf_{T>0}\min_{\gamma\in C_T(x,y)} A_{L_\eta+k}(\gamma).\end{aligned}$$

Definition 2.48. The Mañé critical value is defined by

$$\begin{aligned}c(L_\eta) &= \sup\{k\in\mathbb{R}\colon \text{there exists a closed curve } \gamma \text{ such that } A_{L_\eta+k}(\gamma)<0\}\\ &= \inf\{k\in\mathbb{R}\colon \text{for all closed curves } \gamma, \text{ we have } A_{L_\eta+k}(\gamma)\geq 0\}\\ &< \infty.\end{aligned}$$

We now focus on the properties of the Mañé potential and critical value.

Theorem 2.49. Let $k \in \mathbb{R}$ and $x, y, z \in M$. Then, the Mañé potential satisfies the triangle inequality

$$\phi_{\eta,k}(x,y) \leq \phi_{\eta,k}(x,z) + \phi_{\eta,k}(z,y).$$

Theorem 2.50. Let $x, y \in M$ and $k < c(L_\eta)$. Then, we have

$$\phi_{\eta,k}(x,y) = -\infty.$$

Theorem 2.51. Let $x, y \in M$ and $k \geq c(L_\eta)$. Then, the following properties hold.

1. The Mañé potential is finite.
2. The Mañé potential is Lipschitz.
3. The Mañé potential satisfies $\phi_{\eta,k}(x,x) = 0$.
4. The Mañé potential satisfies $\phi_{\eta,k}(x,y) + \phi_{\eta,k}(y,x) \geq 0$.

Theorem 2.52. Let $x, y \in M$ be distinct and $k > c(L_\eta)$. Then, the Mañé potential satisfies

$$\phi_{\eta,k}(x,y) + \phi_{\eta,k}(y,x) > 0.$$

In light of these properties, the Mañé critical value can be equivalently expressed in terms of the Mañé potential.

Corollary 2.53. The Mañé critical value can be equivalently expressed as

$$\begin{aligned}c(L_\eta) &= \inf\{k\in\mathbb{R}:\ \text{there exist } x,y\in M \text{ with } \phi_{\eta,k}(x,y) > \infty\}\\ &= \sup\{k\in\mathbb{R}:\ \text{there exist } x,y\in M \text{ with } \phi_{\eta,k}(x,y) = -\infty\}.\end{aligned}$$

We can give a complete characterization of c-time-free minimizers for $L+k$ in terms of the Mañé potential.

Corollary 2.54. An absolutely continuous curve $\gamma : \mathbb{R} \to M$ is a c-time-free minimizer for $L + k$ if, for all real numbers $a < b$, we have

$$\int_a^b [L_\eta(\gamma(t), \dot{\gamma}(t)) + k] \, dt = \phi_{\eta,k}(\gamma(a), \gamma(b)).$$

We have seen that the case $k < c(L_\eta)$ is uninteresting since we have that $\phi_{\eta,k}(x, y) = -\infty$ for all $x, y \in M$. It turns out that the case $k > c(L_\eta)$ is also uninteresting since there always exists a c-time-free minimizer for $L + k$ connecting any two points in finite time. This notion is captured in the following theorem.

Theorem 2.55. Let $x, y \in M$ be distinct and $k > c(L_\eta)$. Then, there exists $T > 0$ and $\gamma \in C_T(x, y)$ such that

$$A_{L_\eta+k}(\gamma) = \phi_{\eta,k}(x, y).$$

Remark 2.56. The most interesting case to consider is the case $k = c(L_\eta)$. This corresponds to the least possible value of $k \in \mathbb{R}$ such that c-time-free minimizers can exist.

Definition 2.57. An absolutely continuous curve $\gamma \colon \mathbb{R} \to M$ is called c-semi-static for L if, for all real numbers $a < b$, we have

$$\int_a^b [L_\eta(\gamma(t), \dot{\gamma}(t)) + c(L_\eta)] \, dt = \phi_{\eta,c(L_\eta)}(\gamma(a), \gamma(b)).$$

In other words, c-semi-static curves for L are c-time-free minimizers for $L + c(L_\eta)$. Since adding a constant to the Lagrangian has no effect on the selection of minimizers, it follows that c-semi-static curves are also c-minimizers for L and are thus integral curves to the Euler-Lagrange flow. Recall from Theorem 2.51.4 , that for all $(x, y) \in M$ the Mañé potential satisfies

$$\phi_{\eta,c(L_\eta)}(x, y) \geq -\phi_{\eta,c(L_\eta)}(y, x).$$

This means that for c-semi-static curves, the minimal action required to join x to y is not necessarily the same as the minimal action required to join y back to x in absolute value. Thus, we can reinforce the notion of c-semi-static curves to incorporate the above insight.

Definition 2.58. An absolutely continuous curve $\gamma \colon \mathbb{R} \to M$ is c-static for L if, for all real numbers $a < b$, we have

$$\int_a^b [L_\eta(\gamma(t), \dot{\gamma}(t)) + c(L_\eta)] \, dt = -\phi_{\eta,c(L_\eta)}(\gamma(b), \gamma(a)).$$

Remark 2.59. Let $\gamma : \mathbb{R} \to M$ be a c-static curve. By Theorem 2.51.4, we have that for all $a < b$,

$$\begin{aligned}\int_a^b [L_\eta(\gamma(t), \dot{\gamma}(t)) + c(L_\eta)] \, dt &= -\phi_{\eta,c(L_\eta)}(\gamma(b), \gamma(a)) \\ &\leq \phi_{\eta,c(L_\eta)}(\gamma(a), \gamma(b)).\end{aligned}$$

The reverse inequality is obtained by the definition of the Mañé potential. Thus, we can conclude that for all c-static curves γ, we have

$$\int_a^b [L_\eta(\gamma(t), \dot{\gamma}(t)) + c(L_\eta)]\, dt = \phi_{\eta, c(L_\eta)}(\gamma(a), \gamma(b)).$$

In other words, any c-static curve is also a c-semi-static curve and thus also an integral curve to the Euler-Lagrange flow.

We are now ready to define the Aubry set of cohomology class c.

Definition 2.60. The Aubry set of cohomology class c is defined by

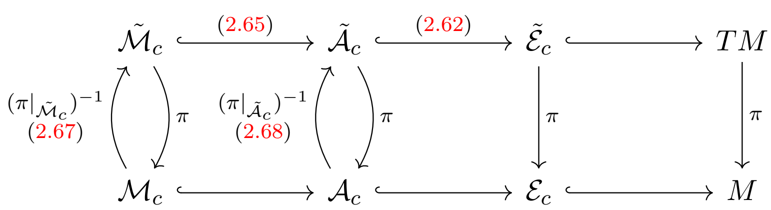
$$\tilde{\mathcal{A}}_c = \bigcup \{(\gamma(t), \dot{\gamma}(t)) \colon \gamma \text{ is a } c\text{-static curve and } t \in \mathbb{R}\} \subseteq \mathrm{T}M.$$

The projection on the base manifold $\mathcal{A}_c = \pi\left(\tilde{\mathcal{A}}_c\right) \subseteq M$ is the projected Aubry set of cohomology class c.

Theorem 2.61. The Aubry set of cohomology class c is non-empty, closed and invariant under the Euler-Lagrange flow.

2.2.3 Mather's Graph Theorems

Until now, we have defined two important sets which describe the dynamics of the system. On one hand, we investigated Mather sets which consist of supports of action-minimizing measures. On the other hand, we considered Aubry sets which consist of orbits of static curves. Recall that energy levels also give information about the dynamics of the system. We present an overview of the relationships between these sets summarized by the commutative diagram below,



where

$$\tilde{\mathcal{E}}_c = \{(x, v) \in \mathrm{T}M : E(x, v) = c(L_\eta) \overset{(2.64)}{=} \alpha(c)\}$$

denotes the energy level corresponding to the energy $\alpha(c)$ and $\mathcal{E}_c = \pi(\tilde{\mathcal{E}}_c)$ denotes its projection onto the base manifold. Theorems 2.67 and 2.68 are known as Mather's graph theorems which state that the Mather and Aubry sets can be viewed as graphs over the manifold M. In other words, the projection onto the base manifold is bijective over these sets.

Theorem 2.62. All c-static curves have energy equal to $c(L_\eta)$. In other words,

$$\tilde{\mathcal{A}}_c \subseteq \tilde{\mathcal{E}}_c = \{(x, v) \in \mathrm{T}M : E(x, v) = c(L_\eta)\}.$$

Remark 2.63. Recall from Theorem 2.61 that the Aubry set $\tilde{\mathcal{A}}_c$ of cohomology class c is closed and from Theorem 2.23 that the energy level $\tilde{\mathcal{E}}_c$ is compact. Thus, it follows from Theorem 2.62 that $\tilde{\mathcal{A}}_c$ is compact.

Theorem 2.64. The Mañé critical value coincides with Mather's α-function. In other words, if η is a closed 1-form of cohomology class c, we have

$$c(L_{\eta_c}) = \alpha(c).$$

Theorem 2.65. A probability measure $\mu \in \mathfrak{M}(L)$ is a c-action-minimizing measure of L if and only if supp $\mu \subseteq \tilde{\mathcal{A}}_c$. In particular, we have $\tilde{\mathcal{M}}_c \subseteq \tilde{\mathcal{A}}_c$.

Remark 2.66. Note the striking similarity between Theorem 1.38 and Theorem 2.65. This relationship is investigated further in Section 2.3.

Theorem 2.67. The projection map of the Mather set of cohomology class c onto the manifold M

$$\pi|_{\tilde{\mathcal{M}}_c} : \tilde{\mathcal{M}}_c \to \mathcal{M}_c$$

is injective. Moreover, its inverse

$$(\pi|_{\tilde{\mathcal{M}}_c})^{-1} : \mathcal{M}_c \to \tilde{\mathcal{M}}_c$$

is Lipschitz.

Theorem 2.68. The projection map of the Aubry set of cohomology class c onto the manifold M

$$\pi|_{\tilde{\mathcal{A}}_c} : \tilde{\mathcal{A}}_c \to \mathcal{A}_c$$

is injective. Moreover, its inverse

$$(\pi|_{\tilde{\mathcal{A}}_c})^{-1} : \mathcal{A}_c \to \tilde{\mathcal{A}}_c$$

is Lipschitz.

Conclusion

In the case of a Tonelli Lagrangian, we can define two sets which encode valuable information about the dynamics of the system. The first consists of the supports of all action-minimizing measures. The second consists of the orbits of all static curves which are defined in terms of the Mañé potential and critical value. Moreover, Mather's celebrated graph theorems state that the projection onto the base manifold restricted to these sets forms a bijection, which has important dynamical consequences.

2.3 Optimal Transportation on Compact Manifolds

2.3.1 Weak KAM Theory and the Hamilton-Jacobi Equation

Weak KAM theory can be viewed as a functional analytic perspective of Aubry-Mather theory. Historically, the field emerged through the study of existence and properties of solutions to the Hamilton-Jacobi equation

$$H(x, \eta(x) + d_x u) = k, \tag{2.2}$$

where $H : \mathrm{T}^*M \to \mathbb{R}$ is a Tonelli Hamiltonian, η is a closed 1-form of cohomology class c and $k \in \mathbb{R}$ are given. The Hamilton-Jacobi equation can be viewed as a special case of the Hamilton-Jacobi-Bellman equation which has numerous applications in dynamic programming and optimal control [3]. Recall from Theorem 2.30 that considering 1-forms of different cohomologies corresponds to considering different Lagrangian shifts. We say that $u : M \to \mathbb{R}$ is a classical solution to the Hamilton-Jacobi equation if it is of smoothness class C^1 and satisfies (2.2) for all $x \in M$. It turns out that such solutions can exist for at most one value of k.

Theorem 2.69. There exists a unique $k \in \mathbb{R}$ which admits a classical solution to the Hamilton-Jacobi equation.

Moreover, we can consider classical subsolutions to the Hamilton-Jacobi equation. That is, functions $u : M \to \mathbb{R}$ of smoothness class C^1 satisfying

$$H(x, \eta(x) + d_x u) \le k,$$

for all $x \in M$. We aim to characterize classical subsolutions to the Hamilton-Jacobi equation in a manner that is independent of the regularity of the function.

Theorem 2.70. Let $u : M \to \mathbb{R}$ be a function of smoothness class C^1. Then, the function u is a classical subsolution to the Hamilton-Jacobi equation if and only if for all $a < b$ and all $\gamma \in \mathrm{AC}([a, b], M)$, it holds that

$$u(\gamma(b)) - u(\gamma(a)) \le \int_a^b L_\eta(\gamma(t), \dot{\gamma}(t))\, dt + k(b - a).$$

We can use the inequality in Theorem 2.70 to extend the notion of subsolution to the Hamilton-Jacobi equation to functions that are continuous rather than of smoothness class C^1. This allows us to bypass the differential present in the Hamilton-Jacobi equation.

Definition 2.71. Let $u : M \to \mathbb{R}$ be a continuous function. Then, the function u is dominated by $L_\eta + k$, written $u \prec L_\eta + k$, if for all $a < b$ and all $\gamma \in \mathrm{AC}([a, b], M)$, it holds that

$$u(\gamma(b)) - u(\gamma(a)) \le \int_a^b L_\eta(\gamma(t), \dot{\gamma}(t))\, dt + k(b - a).$$

Similarly as in the case of subsolutions to the Hamilton-Jacobi equation, we aim to characterize classical solutions to the Hamilton-Jacobi equation in a way that is independent of the regularity of the solution.

Theorem 2.72. Let $u : M \to \mathbb{R}$ be a C^1 function and $k \in \mathbb{R}$. The following are equivalent.

1. The function u is a classical solution to the Hamilton-Jacobi equation

2. We have $u \prec L_\eta + k$ and for all $x \in M$, there exists an absolutely continuous curve $\gamma_x : \mathbb{R} \to M$ such that $\gamma_x(0) = x$ and for all $[a,b] \subseteq \mathbb{R}$, we have

$$u(\gamma_x(b)) - u(\gamma_x(a)) = \int_a^b L_\eta(\gamma_x(t), \dot{\gamma}_x(t))\, dt + k(b-a).$$

3. We have $u \prec L_\eta + k$ and for all $x \in M$, there exists an absolutely continuous curve $\gamma_x : (-\infty, 0] \to M$ such that $\gamma_x(0) = x$ and for all $a < b \leq 0$, we have

$$u(\gamma_x(b)) - u(\gamma_x(a)) = \int_a^b L_\eta(\gamma_x(t), \dot{\gamma}_x(t))\, dt + k(b-a).$$

4. We have $u \prec L_\eta + k$ and for all $x \in M$, there exists an absolutely continuous curve $\gamma_x : [0, \infty) \to M$ such that $\gamma_x(0) = x$ and for all $0 \geq a < b$, we have

$$u(\gamma_x(b)) - u(\gamma_x(a)) = \int_a^b L_\eta(\gamma_x(t), \dot{\gamma}_x(t))\, dt + k(b-a).$$

Again, we can use Theorem 2.72 to extend the notion of solution to the Hamilton-Jacobi equation to functions that are continous rather than of smoothness class C^1. This allows us to bypass the differential present in the Hamilton-Jacobi equation.

Definition 2.73. Let $u : M \to \mathbb{R}$ be a continuous function such that $u \prec L_\eta + k$. Then, an absolutely continuous curve $\gamma : I \to M$ is (u, L_η, k)-calibrated on I if for any $[a,b] \subseteq I$, we have

$$u(\gamma(b)) - u(\gamma(a)) = \int_a^b L_\eta(\gamma(t), \dot{\gamma}(t))\, dt + k(b-a).$$

Moreover, it turns out that the unique value of $k \in \mathbb{R}$ admitting classical solutions to the Hamilton-Jacobi equation in Theorem 2.69 corresponds to Mather's α-function $\alpha(c)$, which coincides with the Mañé critical value $c(L_\eta)$ by Theorem 2.64. In this case, the calibrated curves satisfy a stronger property.

Theorem 2.74. Let $u : M \to \mathbb{R}$ be a continuous function such that $u \prec L_\eta + k$ and $\gamma : [a,b] \to M$ be a (u, L_η, k)-calibrated curve on $[a,b]$. Then, the following statements hold.

1. The curve γ is a c-Tonelli minimizer.

2. The curve γ is a c-time-free minimizer if $k = \alpha(c)$.

Remark 2.75. By Theorem 2.43, we can deduce that any $(u, L_\eta, \alpha(c))$-calibrated curve γ is a solution to the Euler-Lagrange equation and is of smoothness class C^2.

As in Remark 2.56, it is interesting to study the solutions to the Hamilton-Jacobi equation in the case where $k = \alpha(c)$. Such functions play an important role in weak KAM theory since they carry valuable information about the dynamics of the system. We can give a name to dominated functions in the case $k = \alpha(c)$.

Definition 2.76. Let $u : M \to \mathbb{R}$ be a continuous function. Then, the function u is said to be critically dominated if $u \prec L_\eta + \alpha(c)$.

Now that we have extended the properties of classical solutions of the Hamilton-Jacobi equation to continuous functions, we formalize the notion of weak solutions to the Hamilton-Jacobi equation as follows.

Definition 2.77. Let $u : M \to \mathbb{R}$ be a continuous function satisfying $u \prec L_\eta + k$.

1. The function u is a weak KAM solution of the negative type if for each $x \in M$, there exists $\gamma_x : (-\infty, 0] \to M$ such that $\gamma_x(0) = x$ and γ_x is (u, L_η, k)-calibrated.

2. The function u is a weak KAM solution of the positive type if for each $x \in M$, there exists $\gamma_x : [0, +\infty) \to M$ such that $\gamma_x(0) = x$ and γ_x is (u, L_η, k)-calibrated.

The main result of weak KAM theory states that we can guarantee the existence of weak KAM solutions to the Hamilton-Jacobi equation whenever $k = \alpha(c)$. Moreover, these weak KAM solutions are closely related to the Aubry set defined in Section 2.2.2.

Theorem 2.78. There is a unique value of $k \in \mathbb{R}$ for which weak KAM solutions of a positive or negative type may exist. This value coincides with Mather's α-function $\alpha(c)$. Moreover, for any $u \prec L_\eta+\alpha(c)$, there exist a weak KAM solution of the negative type u_- and a weak KAM solution of the positive type u_+ such that $u_- = u = u_+$ on the projected Aubry set $\mathcal{A}_c$.

In light of this, the Aubry set can be equivalently defined from the perspective of weak KAM theory. Since weak KAM solutions to the Hamilton-Jacobi equation give to calibrated curves, we can consider the set of initial conditions in the tangent bundle such that the projected Euler-Lagrange flow yields a calibrated curve.

Definition 2.79. Let $u : M \to \mathbb{R}$ be a continuous function. The Aubry set of the function u is defined by

$$\tilde{\mathfrak{I}}(u) = \{(x, v) \in \mathrm{T}M : \gamma_{(x,v)}(s) = \pi\left(\Phi_s^L(x, v)\right) \text{ is } (u, L_\eta, \alpha(c))\text{-calibrated on } \mathbb{R}\}.$$

Theorem 2.80. The Aubry set defined in Section 2.2.2 can be equivalently expressed as

$$\tilde{\mathcal{A}}_c = \bigcap_{u \prec L_\eta+\alpha(c)} \tilde{\mathfrak{I}}(u).$$

Remark 2.81. In the case where $M = \mathbb{T}^d$ is a d-dimensional torus, we can relate the Aubry set $\tilde{\mathcal{A}}_c$ of cohomology class c defined on the tangent bundle to a KAM torus $\mathcal{T}_c$ of rotation vector ρ defined on the cotangent bundle using the Legendre transform

$$\tilde{\mathcal{A}}_c = \mathcal{L}^{-1}(\mathcal{T}_c).$$

To be precise, a KAM torus $\mathcal{T}_c$ of rotation vector ρ is a set of the form

$$\mathcal{T}_c = \{(x, c + du) : x \in \mathbb{T}^d\},$$

where $c \in \mathbb{R}^d$ and $u : \mathbb{T}^d \to \mathbb{R}$ such that the following properties hold.

1. The set $\mathcal{T}_c$ is invariant under the Hamiltonian flow Φ_t^H.

2. The Hamiltonian flow Φ_t^H on $\mathcal{T}_c$ is conjugate to uniform rotation on $\mathbb{T}^d$. This means that there exists a diffeomorphism $\varphi : \mathbb{T}^d \to \mathcal{T}_c$ such that

$$\varphi^{-1} \circ \Phi_t^H \circ \varphi = R_\rho^t$$

holds for all $t \in \mathbb{R}$, where the uniform rotation R_ρ^t is defined by

$$\begin{aligned} R_\rho^t : \mathbb{T}^d &\to \mathbb{T}^d \\ x &\mapsto x + \rho t \mod \mathbb{Z}^d. \end{aligned}$$

In other words, the following diagram commutes.

$$\begin{array}{ccc} \mathbb{T}^d & \xrightarrow{R_\rho^t} & \mathbb{T}^d \\ \mathcal{L}\big\downarrow & & \big\downarrow\mathcal{L} \\ \mathcal{T}_c & \xrightarrow{\Phi_t^H} & \mathcal{T}_c \end{array}$$

Thus, we see that Aubry sets form a generalization KAM tori and thus can be seen as a weak form of KAM theory.

2.3.2 Weak KAM Theory as an Optimal Transport Problem

Finally, we illustrate how aspects of optimal transport theory play a background role in the theory of dynamical systems. Due to a clash in notation, we keep the symbol c for the cohomology class of a closed 1-form and write $\bar{c} : X \times Y \to \mathbb{R}$ for the cost function inherent to the optimal transport problem. We attempt to make precise the ideas related to Aubry-Mather theory described in the bibliographical notes of Chapter 5 in [24]. In particular, we investigate how the $\bar{c}$-cyclically monotone sets described in Remark 1.50 generalize the Aubry and Mather sets from the theory of dynamical systems. Throughout this section, we assume that X and Y correspond to a smooth compact and connected manifold M, η is a closed 1-form of cohomology class c and the cost function is given by the Mañé potential with $k = \alpha(c)$. Namely

$$\begin{aligned} \bar{c} : M \times M &\to \mathbb{R} \\ (x, y) &\mapsto \phi_{\eta,\alpha(c)}(x, y). \end{aligned}$$

This choice of cost function is motivated by Theorem 2.74 from which it follows that the right-hand side of the inequality in Definition 2.71 is effectively the Mañé potential. Using Theorem 2.51.2, we can deduce that the cost function is continuous and Theorem 2.49 implies that we can bound the cost function above by the sum of two continuous and integrable functions on M. Note that integrability follows by continuity of the Mañé potential and by compactness of M. Moreover, we can deduce that the cost function can be bounded below since the Mañé potential will attain its minimum value on $M \times M$. By Remark 1.49, we can use Theorem 1.48 to obtain two $\bar{c}$-cyclically monotone sets which characterize the optimality in the Kantorovich problems, namely

$$\Gamma_{\min} = \bigcup_{\pi \text{ optimal}} \operatorname{supp} \pi \quad \text{and} \quad \Gamma_{\max} = \bigcap_{\varphi \text{ optimal}} \partial^{\bar{c}} \varphi.$$

To determine the initial and final measures required to model the above situation accurately, we need the first set to coincide with the Mather set of cohomology class c. In other words, the optimal transport plans must correspond to c-action-minimizing measures. Since Theorem 1.48 is devoid of a smooth structure, we can use Theorem 2.68 to uniquely identify

$$\begin{aligned} \iota : \mathcal{A}_c \times \mathcal{A}_c &\to \tilde{\mathcal{A}}_c \\ (x, y) &\mapsto (x, v), \end{aligned} \tag{2.3}$$

where $v \in \mathrm{T}_x M$ denotes the unique element in the tangent space used to reach y along a c-static curve. This tangent vector exists by Theorem 2.68. Note that we can restrict ourselves to the projected Aubry set since points outside the projected Aubry set do not contribute to the total cost of transportation since there does not exist a c-static curve joining them. Instead, such points remain stationary by Theorem 2.51.3.

As a result, we must consider an optimal transport problem from μ to μ which minimizes the total cost of transportation over all possible invariant measures $\mu \in \mathfrak{M}(L)$. Recall from Theorem 2.25 that the set of invariant measures $\mathfrak{M}(L)$ is compact and the map

$$\mathfrak{M}(L) \to \mathbb{R}$$
$$\mu \mapsto \max_{(\varphi,\psi)\in\Phi_{\bar{c}}} \int_{M\times M} [\varphi + \psi]\, d\mu,$$

is lower semi-continous with respect to the vague topology since it is the pointwise supremum of a collection of lower semi-continous functions. Using the proof of Theorem 1.18, we can deduce that there exists a probability measure $\bar{\mu} \in \mathfrak{M}(L)$ which minimizes the total cost of transportation. By Theorem 1.19, the corresponding optimal transport plan $\pi^\dagger$ is a c-action-minimizing measure.

2.3.3 Optimality of the Aubry Set

Now that we have defined the relevant optimal transport problem, we explicitly show that the Aubry set defined in Section 2.2.2 corresponds to the set $\Gamma_{\max}$ described in Remark 1.50. As a stepping stone, it is clear that critically dominated functions correspond to elements in the constraint in the corresponding dual Kantorovich problem in Theorem 1.19. Moreover, it turns out that the weak KAM solutions to the Hamilton-Jacobi equation correspond to optimal maps φ in the Kantorovich dual problem.

Theorem 2.82. Let $u : M \to \mathbb{R}$ be a weak KAM solution to the Hamilton-Jacobi equation with $k = \alpha(c)$ and define $\varphi(x) = -u(x)$. Then, we have that $\varphi^{\bar{c}}(y) = u(y)$ and the $\bar{c}$-transform pair $(\varphi, \varphi^{\bar{c}})$ is a maximizer to the dual Kantorovich problem.

Proof. Recall that the c-transform of φ is given by

$$\begin{aligned}\varphi^{\bar{c}}(y) &= \inf_{x\in X} (\bar{c}(x,y) - \varphi(x)) \\ &= \inf_{x\in X} (\bar{c}(x,y) + u(x)).\end{aligned}$$

Since u is critically dominated, we have that $u(y) \le u(x) + \bar{c}(x,y)$, from which we can deduce that

$$u(y) \le \inf_{x\in X} (\bar{c}(x,y) + u(x)) = \varphi^{\bar{c}}(y).$$

To prove the reverse inequality, suppose that

$$u(y_0) < \inf_{x\in X} (u(x) + \bar{c}(x,y_0)),$$

for some $y_0 \in Y$. Then, we have $u(y_0) - u(x) < \bar{c}(x,y_0)$ for all $x \in X$. This contradicts Theorem 2.78 which guarantees the existence of a $(u, L_\eta, \alpha(c))$-calibrated curve passing through y_0. Thus, we have shown that $u(y) = \varphi^{\bar{c}}(y)$, which implies that the equality

$$\varphi(x) + \varphi^{\bar{c}}(y) = \bar{c}(x,y),$$

holds for all $(x, y) \in \mathcal{A}_c \times \mathcal{A}_c$ by Theorem 2.78. Since points outside the projected Aubry set do not contribute to the total cost of transportation since there does not exist a c-static curve joining those points, we have that the $\bar{c}$-transform pair is a maximizer to the dual Kantorovich problem. □

Moreover, we have that calibrated curves correspond to the $\bar{c}$-superdifferentials of the weak KAM solutions. In particular, after identifying the product projected Aubry set with the tangent bundle using (2.3), we obtain the Aubry set of a weak KAM solution.

Theorem 2.83. Let $u : M \to \mathbb{R}$ be a weak KAM solution to the Hamilton-Jacobi equation with $k = \alpha(c)$ and define $\varphi(x) = -u(x)$. Then, we have

$$\tilde{\mathfrak{I}}(u) = \iota(\partial^{\bar{c}}\varphi).$$

Proof. Let $(x, v) \in \tilde{\mathfrak{I}}(u)$ be arbitrary. Then, it follows that

$$\begin{aligned} \gamma_{(x,v)} &: \mathbb{R} \to M \\ s &\mapsto \pi(\Phi_s^L(x, v)). \end{aligned}$$

is a $(u, L_\eta, \alpha(c))$-calibrated curve on $\mathbb{R}$. Let $t > 0$ and define $y = \gamma_{(x,v)}(t)$. By Theorem 2.68, we can deduce that $(x, v) = \iota(x, y)$. Since the curve $\gamma_{(x,v)}$ is $(u, L_\eta, \alpha(c))$-calibrated, we have

$$\varphi(x) + \varphi^{\bar{c}}(y) = \bar{c}(x, y),$$

from which we can deduce that $(x, y) \in \partial^{\bar{c}}\varphi$. Hence, we have shown that $(x, v) \in \iota(\partial^{\bar{c}}\varphi)$. For the reverse inclusion, let $(x, v) \in \iota(\partial^{\bar{c}}\varphi)$. Again, consider the curve

$$\begin{aligned} \gamma_{(x,v)} &: \mathbb{R} \to M \\ s &\mapsto \pi(\Phi_s^L(x, v)). \end{aligned}$$

By Theorem 2.61, we have Aubry set is invariant under the Euler-Lagrange flow from which we can deduce that $\gamma_{(x,v)}(s) \in \partial^{\bar{c}}\varphi(x)$ for all $s \in \mathbb{R}$. Thus, we can deduce that for all $a < b$, we have

$$u(\gamma_{(x,v)}(b)) - u(\gamma_{(x,v)}(a)) = \int_a^b L_\eta(\gamma_{(x,v)}(t), \dot{\gamma}_{(x,v)}(t))\, dt + \alpha(c)(b - a).$$

Thus, we can conclude that $\gamma_{(x,v)}$ is $(u, L_\eta, \alpha(c))$-calibrated on $\mathbb{R}$. Hence, we have shown that $(x, v) \in \tilde{\mathfrak{I}}(u)$. □

Corollary 2.84. The set $\Gamma_{\max}$ defined in the proof of Theorem 1.48 corresponds to the Aubry set after the identification given in (2.3).

$$\tilde{\mathcal{A}}_c = \bigcap_{u \prec L_\eta + \alpha(c)} \tilde{\mathfrak{I}}(u) = \iota\left(\bigcap_{\varphi \text{ optimal}} \partial^{\bar{c}}\varphi\right).$$

Proof. The first equality coincides with Theorem 2.80. We now show the second equality. By Theorem 2.78, any critically dominated function admits a weak KAM solution which yields a maximizer to the dual Kantorovich problem by Theorem 2.82. On the other hand, it is clear that an optimal map φ to the dual Kantorovich problem yields a critically dominated function. By applying Theorem 2.83, we obtain

$$\bigcap_{u \prec L_\eta + \alpha(c)} \tilde{\mathfrak{I}}(u) = \bigcap_{\varphi \text{ optimal}} \iota(\partial^{\bar{c}} \varphi).$$

Since ι is injective, we obtain

$$\bigcap_{u \prec L_\eta + \alpha(c)} \tilde{\mathfrak{I}}(u) = \iota \left(\bigcap_{\varphi \text{ optimal}} \partial^{\bar{c}} \varphi \right).$$

□

Conclusion

In a sense, classical solutions to the Hamilton-Jacobi equation are rare since they can only exist whenever the right-hand side coincides with Mather's α-function. As a result, weaker solutions are defined in a manner that is consistent with the properties satisfied by the classical solutions. We have made an attempt at establishing a relationship between these weak KAM solutions and the optimal maps in a suitably chosen dual Kantorovich problem. The minimal set characterizing optimality coincides with the Mather set whereas the maximal set characterizing optimality corresponds to the Aubry set from the theory of dynamical systems.

Chapter 3

Conclusion

In this thesis, we investigate the meeting point between optimal transportation and the theory of dynamical systems.

When it comes to optimal transportation, we have seen that the Monge and Kantorovich problems represent cost-minimization problems through different perspectives. The former can be viewed as an optimization problem over the set of transport maps whereas the latter can be viewed an an optimization problem over the set of transport plans. However, the Monge problem is not always defined for any given initial and final configurations since mass splitting is prohibited. On the other hand, it is known that the Kantorovich problem admits a minimizer under fairly general conditions. Moreover, since the Kantorovich problem is a convex optimization problem, it is not surprising that there exists a corresponding dual problem. Solving this dual maximization problem is equivalent to solving the primal minimization problem since there is no duality gap. To guarantee the existence of a maximizer, the cost function must be sufficiently well-behaved. In the Euclidean setting, we employ tools from convex analysis to obtain a characterization of optimal transport plans known as the Knott-Smith criterion. Moreover, Brenier's theorem provides sufficient conditions for the existence of a transport map in the corresponding Monge problem. Lastly, we employ Rüschendorf's theoretical framework to generalize tools from convex analysis to deduce a general version of the Knott-Smith criterion. This general optimality criterion yields two c-cyclically monotone sets which characterize optimality in both Kantorovich problems. The first minimal set consists of the union of the supports of all optimal transport plans whereas the second maximal set is defined by the intersection of the c-superdifferentials of all optimal maps in the dual Kantorovich problem.

When it comes to Aubry-Mather theory, we have seen that for a given Lagrangian, the equations of motion of the corresponding physical system can be obtained by considering the solutions to the Euler-Lagrange equation. These solutions exist whenever the Legendre condition is satisfied. To guarantee this condition, we work with strictly convex, superlinear Lagrangians. Similar properties can be imposed on the corresponding Hamiltonian to obtain rich invariant sets describing the dynamics of the system. The first consists of the supports of all action-minimizing measures known as the Mather set. The second consists of the orbits of all static curves known as the Aubry set. Mather's celebrated graph theorems state that the projection onto the base manifold restricted to these sets forms a bijection, which has important dynamical consequences. We have made an attempt at establishing a relationship between weak KAM solutions to the Hamilton-Jacobi equation and the optimal maps in a suitably chosen dual Kantorovich problem. The minimal set characterizing optimality coincides with the Mather set whereas the maximal set characterizing optimality corresponds to the Aubry set.

Bibliography

[1] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Springer, Jan. 2005.

[2] V. Arnold. *Mathematical methods of classical mechanics*. Springer, Nov. 1978.

[3] M. Caputo. *Dynamic Programming and the Hamilton-Jacobi-Bellman Equation.* Cambridge University Press, June 2012.

[4] P. Dupuis and R.S. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley Series in Probability and Statistics, Feb. 1997.

[5] Á. Farkas and J. Fraser. *On the equality of Hausdorff measure and Hausdorff content*. 2015.

[6] A. Fathi. *Weak KAM Theorem in Lagrangian Dynamics*. Jan. 2014.

[7] A. Galichon. *The unreasonable effectiveness of optimal transport in economics*. 2021.

[8] I. Gelfand and S. Fomin. *Calculus of Variations*. Prentice-Hall, 1964.

[9] R. Jordan, D. Kinderlehrer, and F. Otto. "The Variational Formulation of the Fokker–Planck Equation". In: *SIAM Journal on Mathematical Analysis* 29.1 (1998), pp. 1–17.

[10] P. Kozarzewski. "On existence of the support of a Borel measure". In: *Demonstratio Mathematica* 51 (2018), pp. 76–84.

[11] N. Kryloff and N. Bogoliouboff. "La Theorie Generale De La Mesure Dans Son Application A L'Etude Des Systemes Dynamiques De la Mecanique Non Lineaire". In: *Annals of Mathematics* 38.1 (1937), pp. 65–113.

[12] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.

[13] G. Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704.

[14] G. Papayiannis, G. Galanis, and A. Yannacopoulos. "Model aggregation using optimal transport and applications in wind speed forecasting". In: *Environmetrics* 29.8 (2018), pp. 35–41.

[15] G. Peyré. *Course notes on Computational Optimal Transport*. Nov. 2021.

[16] J. Pöschel. *A lecture on the classical KAM theorem*. 2009.

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

[18] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.

[19] L. Rüschendorf. "On c-optimal random variables". In: *Statistics Probability Letters* 27.3 (1996), pp. 267–270.

[20] A. Sorrentino. *Action-minimizing methods in Hamiltonian dynamics : an introduction to Aubry-Mather theory*. May 2015.

[21] M. Thorpe. *Introduction to Optimal Transport*. Mar. 2018.

[22] L. Torres, L. Pereira, and M. Amini. *A Survey on Optimal Transport for Machine Learning: Theory and Applications*. 2021.

[23] A. Vershik. *Kantorovich Metric: Initial History and Little Known Applications*. 2005.

[24] C. Villani. *Optimal Transport: Old and New*. Springer, Oct. 2008.

[25] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[26] F. Voronine. *An Invitation to Aubry-Mather and Weak KAM Theory*. 2023.