

OPTIMAL TRANSPORT AND APPLICATIONS

AIDAN COPINGA

ABSTRACT. In the 18th century, a problem was created about the best way to transport munitions to each barracks in France. This problem, now called optimal transportation, has shown useful in many fields of mathematics, from PDEs to image processing and machine learning. This paper introduces the Monge-Kantorovich Problem and its dual problem. Using this, we are able to discuss existence of transport plans and maps of the Monge-Kantorovich problem. Finally, we cover the Wasserstein Space and applications with respect to gradient flows.

CONTENTS

Acknowledgments	1
1. Formulation of Optimal Transport	1
1.1. Problem	1
1.2. Monge Formulation	2
1.3. Kantorovich Formulation	3
2. Kantorovich Duality	4
2.1. Kantorovich Duality	4
2.2. Fenchel-Rockafellar and Convex Analysis Results	5
2.3. Proof of Kantorovich Duality	6
3. Existence and Characterization	6
3.1. Existence of Transport Plans	6
3.2. Existence of Maximizers of the Dual Problem	6
3.3. Existence of Transport Maps	6
4. Wasserstein	6
4.1. Wasserstein Distances	6
4.2. Wasserstein Topology	6
5. Gradient Flows and PDEs	6
5.1. Gradient Flow Motivation	6
5.2. Benamou-Brenier Theorem	6
6. Bibliography	6
References	6

ACKNOWLEDGMENTS

1. FORMULATION OF OPTIMAL TRANSPORT

1.1. Problem.

Start with a simple example where we have a pile of dirt, and a hole to fill completely with the dirt. To transport the dirt from this pile to the hole, clearly, the volume of the hole and the pile must be the same. As is suggested in [1], we normalize this mass to 1.

To model this rigorously, allow the pile to be space X and the hole to be space Y . It's natural to define probability measures μ, ν on X, Y respectively. In other words, for measurable $A \subset X$ and $B \subset Y$, $\mu(A)$ is the volume in the pile X and $\nu(B)$ is the volume to be filled in hole Y .

Now, introduce cost function $c : X \times Y \rightarrow \mathbb{R}$ such that for points $x \in X, y \in Y$, $c(x, y)$ represents how much energy is exhausted transporting point x in the pile to point y in the hole.

We can now introduce the transportation problem.

Definition 1.1 (Transportation Problem). Given measure spaces X, Y and μ, ν respectively on X, Y and $c : X \times Y \rightarrow \mathbb{R}$. The Transportation Problem is realizing the transport plan from Figure 1 while minimizing c .

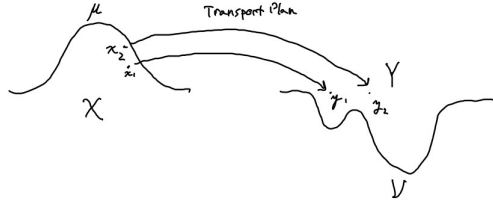


FIGURE 1. Transportation Problem

There are two primary formulation for the transportation problem, the Monge and the Kantorovich formulation. The Monge formulation came historically before Kantorovich's, so I will be introducing it first.

1.2. Monge Formulation.

Figure 1 uses the Lagrangian framework in order to describe Definition 1.1. What this means is that particles $x \in X$ are observed individually as they are transported to $y \in Y$, and importantly, mass cannot be split, or the flows $T : X \rightarrow Y$ must be bijective.

From here, we introduce the Monge formulation, which observes a transport map T that transports μ to ν . Before we define it, we must define what is meant by transporting one measure to another.

Definition 1.2 (Transport Maps). We say that $T : X \rightarrow Y$ transports $\mu \in \mathcal{P}(X)$ to $\nu \in \mathcal{P}(Y)$ if

$$\mu(T^{-1}(B)) = \nu(B) \quad \text{for all } \nu\text{-measurable } B \in Y.$$

This is typically called the *pushforward* of μ and denoted as $T_{\#}\mu$.

In order to describe the Monge Formulation, we denote the set of all T such that $T_{\#}\mu = \nu$ with

$$(1.1) \quad \mathcal{T}(\mu, \nu) = \{T : X \rightarrow Y \mid T_{\#}\mu = \nu, T \text{ measurable}\}$$

Definition 1.3 (Monge Formulation of the transport problem).

$$(1.2) \quad \mathbb{M} = \underset{\mathcal{T}(\mu, \nu)}{\text{minimize}} \left\{ \int_X c(x, T(x)) d\mu(x) \right\}$$

In other words, Definition 1.3 is minimizing the energy required from transporting $X \xrightarrow{T(x)} Y$.

Example 1.4 (Discrete Monge Problem). Consider $\mu = \frac{2}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2}$ and $\nu = \frac{2}{3}\delta_{y_1} + \frac{1}{3}\delta_{y_2}$. We see that the only transport map that exists is the map where $T(x_1) = y_1$ and $T(x_2) = y_2$ as such mass is not split. However, in order to make the optimal transport map less trivial, we can allow $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$ to create the discrete Monge optimal transportation problem.

Now, we can allow $j = \sigma(i)$ where $\sigma \in S_n$ with S_n be the permutations of $\{1, 2, \dots, n\}$. In other words, σ is a transport map where we can define the discrete monge problem as

$$\mathbb{M}_{\text{disc}} = \inf_{\sigma \in S_n} \left\{ \frac{1}{n} \sum_{i=1}^n c(x_i, y_{\sigma(i)}) \right\}.$$

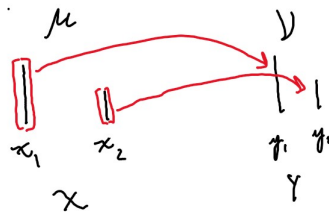


FIGURE 2. We see that $x_1 \rightarrow y_1$ and $x_2 \rightarrow y_2$ is the only valid transport map T such that mass is not split. These are solutions to what are discrete optimal transport problems.

1.3. Kantorovich Formulation.

It's natural to ask what happens on the product space $X \times Y$. The Kantorovich does exactly this by defining *transportation plans* as a product measure on $X \times Y$.

Definition 1.5 (Transport Plans). A transport plan, $\pi \in \mathcal{P}(X \times Y)$, is valid if all mass taken from a point $x \in X$ must correspond to $d\mu(x)$ and all mass taken from a point $y \in Y$ must correspond to $d\nu(y)$, or in other terms,

$$(1.3) \quad \pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B).$$

To define Kantorovich's Formulation, we define the set of all probability measures that follow (1.3) with

$$(1.4) \quad \Pi(\mu, \nu) = \{ \pi \in \mathcal{P}(X \times Y) : (1.3) \text{ holds for all measurable } A \subset X, B \subset Y \}.$$

Definition 1.6 (Kantorovich Formulation of the transport problem).

$$(1.5) \quad \mathbb{K} = \underset{\Pi(\mu, \nu)}{\text{minimize}} \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) \right\}.$$

This formulation is typically referred to as a relaxation of Definition 1.3, or that Monge's problem is stronger than Kantorovich's.

Proposition 1.7. *The Monge Formulation can be written as a Kantorovich Formulation.*

Proof. Because mass cannot be split in the Monge Formulation, this means that we can rewrite $d\pi(x, y)$ in Definition 1.6 with

$$d\pi(x, y) = d\pi_T(x, y) \equiv d\mu(x) \delta[y = T(x)]$$

where δ is the dirac measure and $T(x)$ is as defined in Definition 1.2. As $\int_Y \xi(x, y) \delta[y = T(x)] = \xi(x, T(x))$ for any nonnegative function $\xi : X \times Y \rightarrow \mathbb{R}$, we have that

$$\int_{X \times Y} c(x, y) d\pi(x, y) = \int_X c(x, T(x)) d\mu(x).$$

In order for π_T to be in $\Pi(\mu, \nu)$, we consider (1.3)

$$\pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B)$$

so we see that this is the case where we have measurable $B \in Y$ with

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

Of course, we allowed $T(x)$ to follow Definition 1.2, so this condition is satisfied. Furthermore, this is equivalent to the Kantorovich formulation of the transportation problem. \square

We see that the Monge Formulation is the stronger case, where we restrict mass in X from being able to be split. However, Proposition 1.7 can also be framed as stating that a Kantorovich Formulation with mass that cannot be split is a Monge Problem.

Example 1.8 (Discrete Kantorovich Problem). Consider $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$ and $\nu = \sum_{i=1}^n \beta_i \delta_{y_i}$ where $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i = 1$ and $\alpha_i \geq 0$ and $\beta_i \geq 0$. Unlike Example 1.4, $\alpha_i = \beta_i = \frac{1}{n}$ need not be true. Transport plans can be represented as $\pi = \pi_{ij}$ [2] where π is a $n \times n$ bistochastic matrix in the sense that

$$B = \left\{ \pi \mid \pi_{ij} \geq 0 \text{ for all } 1 \leq i, j \leq n, \quad \sum_{i=1}^n \pi_{ij} = 1 \text{ for all } 1 \leq j \leq n, \quad \sum_{j=1}^n \pi_{ij} = 1 \text{ for all } 1 \leq i \leq n. \right\}$$

So now the discrete Kantorovich problem becomes

$$(1.6) \quad \mathbb{K}_{\text{disc}} = \inf_{\pi \in B} \sum_{i,j} \pi_{ij} c(x_i, y_j).$$

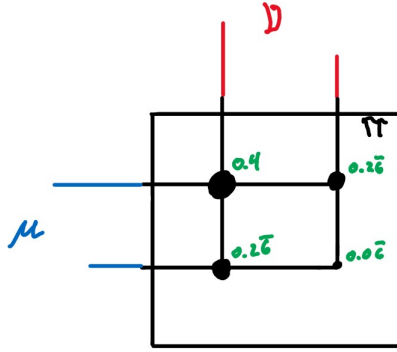


FIGURE 3. Letting $\mu = \frac{2}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2}$ and $\nu = \frac{2}{3}\delta_{y_1} + \frac{1}{3}\delta_{y_2}$, we see there is no concentration along the diagonal for π , hence mass is split, but this transport plan is still valid.

2. KANTOROVICH DUALITY

For linear optimization problems, it is well known that they admit a dual problem, and Kantorovich in 1942 showed duality for [Definition 1.6](#) for $c(x, y) = d(x, y)$ where d is the distance metric in \mathbb{R}^n , but it still holds in general Polish spaces. To begin to understand Kantorovich's dual problem, we can think of Caffarelli's Shipper's Problem described in [\[1\]](#).

Example 2.1 (Shipper's Problem). It costs $c(x_1, y_1)$ dollars for a factory to transport coal from mine x_1 to factory y_1 . I tell the mine owner that I can charge them $\phi(x_1)$ dollars to pick up at mine x_1 and charge $\psi(y_1)$ to deliver to factory y_1 , then, in order for the factory owner to agree to my terms,

$$\phi(x) + \psi(y) \leq c(x, y)$$

for every location x and destination y . However, I can make this sum as close as I want to $c(x, y)$ because I have the liberty to choose ϕ, ψ . In other words, the most money I can make is equal to the least effort it would take the factory owner (just simply transporting it with cost $c(x, y)$) which is exactly what a dual problem would suggest.

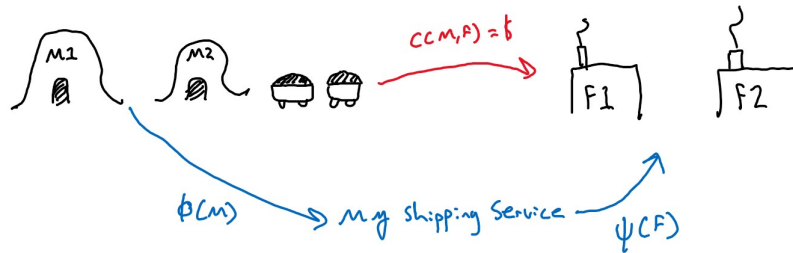


FIGURE 4. The easiest thing for the owners to do is pay $c(M, F)$, but they can pay less by going through my shipping service which costs $\phi(M) + \psi(F)$.

2.1. Kantorovich Duality.

Theorem 2.2 (Kantorovich Duality). *Let X and Y be Polish spaces, let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and let $c : X \otimes Y \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ be a lower semicontinuous cost function.*

Whenever $\pi \in \mathcal{P}(X \otimes Y)$ and $(\phi, \psi) \in L^1(d\mu) \otimes L^1(d\nu)$, define

$$I(\pi) = \int_{X \otimes Y} c(x, y) d\pi(x, y), \quad J(\phi, \psi) = \int_X \phi d\mu + \int_Y \psi d\nu.$$

Define $\Pi(\mu, \nu)$ to be the set of all Borel probability measures π on $X \otimes Y$ such that for all measurable subsets $A \subset X$ and $B \subset Y$,

$$\pi(A \otimes Y) = \mu(A), \quad \pi(X \otimes B) = \nu(B),$$

and define Φ_c to be the set of all measurable functions (ϕ, ψ) satisfying

$$(2.1) \quad \phi(x) + \psi(y) \leq c(x, y)$$

for $d\mu$ -almost every $x \in X$, $d\nu$ -almost every $y \in Y$. Then,

$$(2.2) \quad \inf_{\Pi(\mu, \nu)} I(\pi) = \sup_{\Phi_c} J(\phi, \psi).$$

We can construct a minimax principle in order to give an informal proof of [Theorem 2.2](#). We see

$$(2.3) \quad \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi = \inf_{\pi \in M_+(X \times Y)} \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) + \begin{cases} 0 & \pi \in \Pi(\mu, \nu) \\ +\infty & \text{otherwise.} \end{cases} \right\}$$

where M_+ is the set of nonnegative Borel measures on $X \times Y$. We can rewrite the righthand side of (2.3) with

$$\begin{cases} 0 & \pi \in \Pi(\mu, \nu) \\ +\infty & \text{otherwise.} \end{cases} = \sup_{(\phi, \psi) \in L^1(d\mu) \otimes L^1(d\nu)} \left\{ \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu - \int_{X \times Y} [\phi(x) + \psi(y)] d\pi(x, y) \right\}.$$

as $\pi \in \Pi$ follows the constraints defined in (1.3). Now, the Monge-Kantorovich problem can be written as

$$(2.4) \quad \inf_{\pi \in M_+(X \times Y)} \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) + \sup_{(\phi, \psi)} \left\{ \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu - \int_{X \times Y} [\phi(x) + \psi(y)] d\pi(x, y) \right\} \right\}.$$

Informally exchanging the inf and sup, we get that the Monge-Kantorovich problem is

$$(2.5) \quad \sup_{(\phi, \psi)} \left\{ \inf_{\pi \in M_+(X \times Y)} \left\{ \int_{X \times Y} c(x, y) - [\phi(x) - \psi(y)] d\pi(x, y) \right\} + \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu \right\}.$$

Now, note that

$$\inf_{\pi \in M_+(X \times Y)} \left\{ \int_{X \times Y} c(x, y) - [\phi(x) - \psi(y)] d\pi(x, y) \right\} = \begin{cases} 0 & \phi(x) + \psi(y) \leq c(x, y) \\ +\infty & \text{otherwise.} \end{cases}$$

so the Kantorovich-Monge problem becomes the maximization (dual) problem

$$(2.6) \quad \mathbb{K}(\mu, \nu) = \sup_{\Phi_c} \left\{ \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu \right\}.$$

We've assumed that $(\phi, \psi) \in C_b(X) \times C_b(Y)$ as $\phi(x) + \psi(y) \leq c(x, y)$ holding everywhere is different to holding almost everywhere. Moreover, we need to make this minimax principle rigorous.

2.2. Fenchel-Rockafellar and Convex Analysis Results. In order to prove [Theorem 2.2](#), we need a rigorous minimax principle. We start by presenting some results and definitions from convex and functional analysis.

Definition 2.3 (Legendre-Fenchel Transform). Let $\phi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex. Then, the Legendre-Fenchel Transform for $f \in E^*$ is

$$(2.7) \quad \phi^*(f) = \sup_{x \in E} [\langle f, x \rangle - \phi(x)].$$

Definition 2.4 (Polish Spaces). A Polish space, X , satisfies the following

- (1) A Borel probability measure μ on X is regular
- (2) A probability measure μ on X is concentrated on a σ -compact set.
- (3) Any tight family of probability measures in $P(X)$ is relatively sequentially compact in $P(X)$.
- (4) Let X be a metric space. If F a nonnegative lower semi-continuous function on X then it can be written as the supremum of an increasing sequence of uniformly continuous nonnegative functions.
- (5) If K is a compact metric space then $C(K)$ is separable.

Details of these are presented in [1] and [4].

Now, we introduce the following from [3]

Theorem 2.5 (Fenchel-Rockafellar). Let E be a normed vector space and E^* be its topological dual space. Let $\phi, \psi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex. If there is some $x_0 \in E$ such that $\phi(x_0), \psi(x_0) < +\infty$ and ϕ is continuous at x_0 , then

$$\inf_{x \in E} \{\phi(x) + \psi(x)\} = \sup_{f \in E^*} \{-\phi^*(-f) - \psi^*(f)\} = \max_{f \in E^*} \{-\phi^*(-f) - \psi^*(f)\} = -\min_{f \in E^*} \{-\phi^*(-f) - \psi^*(f)\}$$

where $\phi^*(f)$ for $f \in E^*$ is the Legendre-Fenchel transform.

Proof. To be proven. I can base this off of Brezis. □

Moreover, we can check that [Theorem 2.5](#) is a minimax principle.

2.3. Proof of Kantorovich Duality. Finally, we can begin to prove Kantorovich Duality as stated in [Theorem 2.2](#). The following lemmas come from [\[1\]](#).

Lemma 2.6. *Under the same conditions as [Theorem 2.2](#),*

$$(2.8) \quad \sup_{\Phi_c} J(\phi, \psi) \leq \inf_{\Pi(\mu, \nu)} I(\pi).$$

Proof. Let $(\phi, \psi) \in \Phi_c$ and $\pi \in \Pi(\mu, \nu)$. Then,

$$J(\phi, \psi) = \int_X \phi d\mu + \int_Y \psi d\nu = \int_{X \times Y} [\phi(x) + \psi(y)] d\pi(x, y)$$

as π has marginals μ, ν .

Let N_x, N_y be null sets of μ and ν respectively. Because [\(2.1\)](#) holds almost everywhere, [\(2.1\)](#) holds for $(x, y) \in N_x^c \times N_y^c$. Moreover, $\pi(N_x \times Y) = \mu(N_x) = 0$ and $\pi(X \times N_y) = \nu(N_y) = 0$ so $\pi((N_x^c \times N_y^c)^c) = 0$ so [\(2.1\)](#) holds for π . Thus,

$$(2.9) \quad \int_{X \times Y} [\phi(x) + \psi(y)] d\pi(x, y) \leq \int_{X \times Y} c(x, y) d\pi(x, y).$$

[\(2.8\)](#) is recovered upon taking the supremum on the left-hand side and the infimum on the right-hand side. \square

Lemma 2.7. *Under the same conditions as [Theorem 2.2](#), we have*

$$\sup_{\Phi_c} J(\phi, \psi) \geq \inf_{\Pi(\mu, \nu)} I(\pi)$$

[\[1\]](#) presents the proof in 3 steps of increasing generality:

1. Assuming X, Y are compact and $c(x, y)$ are continuous.
2. Keeping the assumption that $c(x, y)$ is continuous, but relaxing compactness of X, Y .
3. Only assuming that $c(x, y)$ is lower semi-continuous.

Only 1. will use the minimax principle discussed with [Theorem 2.5](#) so the proof will be provided below.

Proof of 1. \square

3. EXISTENCE AND CHARACTERIZATION

3.1. Existence of Transport Plans.

3.2. Existence of Maximizers of the Dual Problem.

3.3. Existence of Transport Maps.

4. WASSERSTEIN

4.1. Wasserstein Distances.

4.2. Wasserstein Topology.

5. GRADIENT FLOWS AND PDES

5.1. Gradient Flow Motivation.

5.2. Benamou-Brenier Theorem.

6. BIBLIOGRAPHY

REFERENCES

- [1] Villani C. Topics in Optimal Transportation. American Mathematical Society, 2003.
- [2] Thorpe M. Introduction to Optimal Transport. University of Cambridge, 2018.
- [3] Brezis H. Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer Press, 2011.
- [4] Billingsley P. Convergence of Probability Measures. John Wiley & Sons Inc., 1999.