# Optimal Transport by Brittany Hamfeldt
# Lecture Note - 1

Nafis Faisal

February 2025

## 1 Introduction

### 1.1 What is Optimal Transport?

In Optimal Transport or OT, the core problem studied is the different ways or "transport plans" to move masses (or probability distributions) from a source mass distribution to a target mass distribution while minimizing some cost function, that is, an optimal transport plan. This can be thought of as moving a pile of dirt from one place (a source set $X$) to another (a target set $Y$) having some source and target mass distributions or densities. One can also consider the problem of finding the optimal route of moving things (suppose coal) from its source (e.g. mines) to various places where it is needed (e.g. factories) as an OT problem. This is how the problem was first posed by Gaspard Monge in 1781.

### 1.2 Why study Optimal Transport?

On the applied side of things, OT is used in MRI scans (considering the scan density as mass density), image processing (considering pixel brightness as density), optimizing positioning of mirrors and reflectors to allow light to be reflected in a specific intensity distribution, studying the early structure of the universe, seismology, mesh generation for 3-D modelling and also in Machine Learning.

OT is also interesting on a more theoretical level as it offers insights into various other branches of mathematics especially differential and Riemannian geometry, probability and statistics, partial differential equations, numerical analysis and optimization among others.

## 2 The Monge Formulation of Optimal Transport

### 2.1 A first try

Monge posed the OT problem as finding a transport map $T(x)$ which maps the mass at each point $x$ in the source set $X$ having mass density $f(x)$ to a point $y$ in the target set $Y$ having mass density $g(y)$ such that it minimizes a specific cost function $C(x, T(x))$. The source and target are supported on the sets $X$ and $Y$.

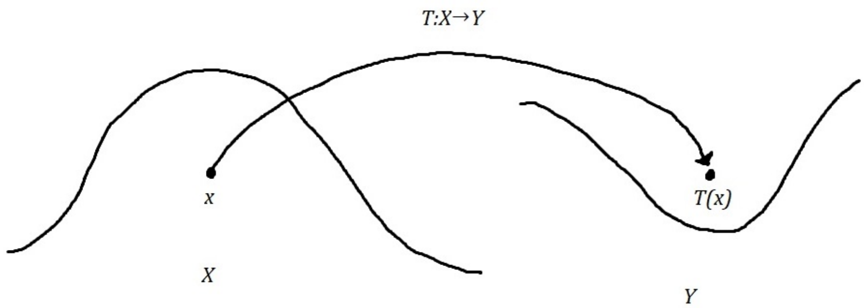


Figure 1: Transportation function

Mathematically we can formulate it as,

$$\min_{T(x)} \int_X C(x,y) f(x) dx \tag{1}$$

As of now, we consider the source set $X$ being integrated over along with the target set as subsets of $\mathbb{R}^n$, although it can be any other set satisfying the properties required for the problem and the optimal transport formulation. The cost function can be the $l_1$ norm-induced metric $|x - T(x)|$ or the $l_2$ norm-induced metric $\frac{1}{2}|x - T(x)|^2$ (called the quadratic cost) or any other cost function as required by the specific case being solved for. Monge formulated the optimal transport problem first using $|x - T(x)|$ as cost. We are assuming that the source and target sets to be subsets of the same space, so as to allow the metric to be defined between them.

## 2.2 Generalizing the formulation

We need to generalize this problem using measures on the source and target set. It is also required that there is mass balance between the source and the target both globally and locally, that is,

i The total mass in the source set $X$ equal the total mass in the target set $Y$

ii The mass in a subset $A$ in $Y$ equal the mass coming from its pre-image under $T$, $T^{-1}(A) \subset X$

Denoting the source measure as $\mu$ defined on $X$ and target measure as $\nu$ on $Y$, these two conditions can be stated as,

i $\mu(X) = \nu(Y)$, and

ii $\mu(T^{-1}(A)) = \nu(A)$, where $A \subset Y$

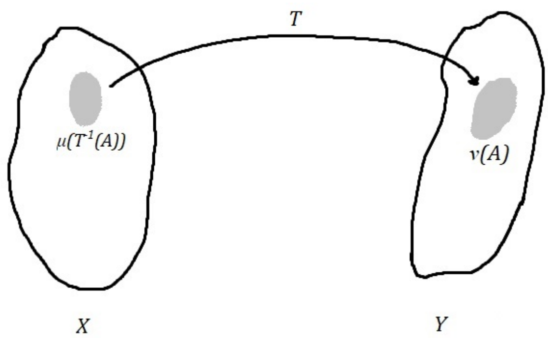


Figure 2: Pushforward of the measure $\mu$ through $T$

This $\mu(T^{-1}(A))$ is called the pushforward of $\mu$ through $T$, denoted by $T_{\#}\mu$, a terminology borrowed from the study of differential manifolds. Mass conservation can now be written more succinctly as,

$$T_{\#}\mu = \nu, \forall A \subset Y \tag{2}$$

The Monge formulation can now be written more generally as,

$$\min_{T(x)}\{\int_X C(x, T(x))d\mu(x) \mid T_{\#}\mu = \nu\} \tag{3}$$

Some questions pertaining to this generalized Monge formulation need to be pondered over before getting our hands dirty by solving for specific cases.

- Does a solution to this (a transport map $T$ in this case) exist?
- If it exists, is it unique?
- Is it stable?
- Is this feasible or is the problem well-posed?
- If all of the above is true, then how can one compute the solution efficiently?

Such questions should be considered while delving into any mathematical problem.

## 2.3 Examples

**Example 1.** ***Book Moving problem*** *Consider two books kept side by side on a shelf (or a number line). We want to move the books one unit to the right. Assuming both books to be same, the order of the books does not matter. Two quite obvious transport methods are-*

1. *to move the first book over the other to the right of the second book, or,*
2. *to slide both the books one unit to the right*

*But which is optimal? To determine this, we will calculate the cost for both methods using both the Monge and the quadratic cost.*

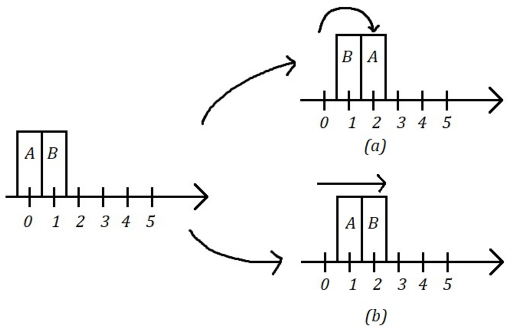


Figure 3: (a)Moving one book over two unit (b) Sliding both books one unit

*For the Monge cost function, the cost is same in both cases (2+0 = 1+1), thus both the methods are optimal, making the solution non-unique. On the other hand, for the quadratic cost function, sliding leads to a cost of 1 while moving the first book over leads to a cost of 4. This time, the optimal transport plan is to slide the books and this solution is unique.* ■

This example shows how the OT solutions depends on the particular cost function being used. Generally, it is easier to work with the quadratic cost and it has been studied extensively, mostly due to its convex nature.

**Example 2.** ***Mines and Factories*** *Let's consider two mines at points $x_1$ and $x_2$ (marked with dots) having same point masses (think of a truckload of coal) that need to be transported to two factories (marked with crosses) at points $y_1$ and $y_2$ in $\mathbb{R}^2$. Assume that both factories are equidistant from both mines. Then what is the optimal transport plan?*

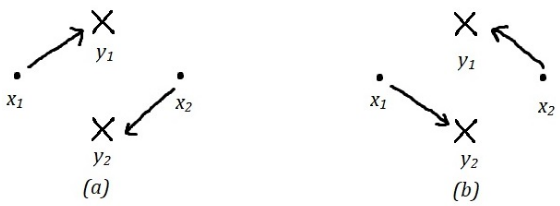


Figure 4: Two transport methods having same cost

*We can transport the mass at $x_1$ to $y_1$ and that at $x_2$ to $y_2$ in straight lines, leading to a cost of $2d$. Or, we can transport the mass at $x_1$ to $y_2$ and that at $x_2$ to $y_1$, also costing $2d$. This time, the optimal transport plan is non-unique. We will find this to be a recurring problem while working with point masses.* ■

**Example 3.** ***Point mass splitting*** *Now let's assume there is one source: a point mass of 1 unit at point $x$ and two target point masses at $y_1$ and $y_2$ that can contain $\frac{1}{2}$ mass each, in $\mathbb{R}^2$. Then what is the optimal transport plan?*

*It should be observed that the Monge formulation does not allow for the mass to split. As such, the point mass at $x$ can either be transported to $y_1$ or to $y_2$ but not both, missing one of the targets completely in both cases. In other words, a solution is not feasible using this formulation.* ■

Such issues with the Monge formulation led to the development of a different formulation of the Optimal Transport problem called the Kantorovich formulation. The Monge formulation is still useful in many cases but not all. Yet some insight can be found from the Monge formulation as we will see in the next example.

**Example 4.** ***Quadratic cost in one dimension*** *Throughout this problem we assume that the properties of the sets and functions declared throughout are "nice enough" to allow for the operations done to be well-defined.*

*Let's consider two mass density functions $f$ and $g$ defined on the source set $X = \mathbb{R}$ and the target set $Y = \mathbb{R}$ respectively. Let's assume an optimal transport plan exists between $X$ and $Y$, given by $T : X \to Y$ and the cost function be the quadratic cost. So the OT problem is,*

$$min\{\frac{1}{2}\int_{\mathbb{R}}(x - T(x)^2 f(x)dx \mid \int_{T^{-1}(A)} f(x)dx = \int_A g(y)dy, \forall A \subset \mathbb{R}\}$$

*Choosing $\epsilon > 0$ pick two points in $X$- $x_1$, $x_2$ and make two open intervals around them, denoted by $I_1$ and $I_2$ respectively such that,*

$$\int_{I_1} f(x)dx = \epsilon = \int_{I_2} g(y)dy$$

*Also define,*

$$y_i = T(x_i),\ J_i = T(I_i),\ where\, i = 1, 2$$

*Let's permute part of the map and define a new measure preserving map $\tilde{T}(x)$ as,*

$$\begin{gathered}\tilde{T}(x_1) = y_2,\ \tilde{T}(x_2) = y_1,\\ \tilde{T}(I_1) = J_2,\ \tilde{T}(I_2) = J_1,\\ \tilde{T}(x) = T(x),\ \forall x \notin I_1 \cup I_2\end{gathered}$$

*As $T$ is assumed to be optimal,*

$$\frac{1}{2}\int_{\mathbb{R}}(x - T(x))^2 f(x)dx \le \frac{1}{2}\int_{\mathbb{R}}(x - \tilde{T}(x))^2 f(x)dx$$

*Everything cancels on both sides of this inequality besides the cross terms leaving,*

$$-\int_{I_1} xT(x)f(x)dx - \int_{I_2} xT(x)f(x)dx \le -\int_{I_1} x\tilde{T}(x)f(x)dx - \int_{I_2} x\tilde{T}(x)f(x)dx$$

*Dividing both sides $\epsilon$ while collecting terms, we get,*

$$\frac{1}{\epsilon}\int_{I_1} x(\tilde{T}(x) - T(x))f(x)dx + \frac{1}{\epsilon}\int_{I_2} x(\tilde{T}(x) - T(x))f(x)dx \le 0$$

*Taking limit $\epsilon \to 0$,*

$$\begin{aligned}\frac{1}{\epsilon}\int_{I_1} x_1(y_2 - y_1)f(x)dx + \frac{1}{\epsilon}\int_{I_2} x_2(y_1 - y_2)f(x)dx \le 0\\ \implies \frac{1}{\epsilon}x_1(y_2 - y_1)\epsilon + \frac{1}{\epsilon}x_2(y_1 - y_2)\epsilon \le 0\\ \implies (x_2 - x_1)(y_2 - y_1) \ge 0\end{aligned}$$

*Thus, if $x_1 < x_2$ then $y_1 < y_2$ too, that is, the optimal mapping $T(x)$ is monotone, which means that the optimal transport map slides things along for the quadratic cost function in one dimension.* ■

Inspite of not being feasible for some cases, the Monge formulation led to some interesting insights.

# 3 The Kantorovich Formulation of Optimal Transport

## 3.1 Introducing mass splitting

We need to generalize to allow the mass from a point to split. As before, there is a source measure $\mu$ supported on a set $X$ and a target measure $\nu$ supported on a set $Y$. As mass splitting is allowed, we need to keep track of the amount of mass transported from each point $x \in X$ to each point $y \in Y$, making the transport map a function of two variables. This information is stored in a new measure $\pi(x, y)$ defined on the product space $X \times Y$. $\pi(x, y)$ measures how much mass is moved from point $x$ to $y$ and, more generally, $\pi(A, B)$ measures how much mass is moved from set $A \subset X$ to $B \subset Y$. Unlike before, it is not required that the whole of the mass at $x$ is moved to $y$, or from $A$ to $B$. Also observe that there is no restrictions on the sets $A$ and $B$, so nothing needs to be assumed about their connectedness, compactness or boundedness.

**Example 5.** *Consider a source of unit point mass at $x_0$ and two target point masses of $\frac{1}{3}$ units at $y_1$ and $\frac{2}{3}$ units at $y_2$. Then, whatever the transport plan might be, it must ensure,*

- $\pi(x_0, y_1) = \frac{1}{3}$
- $\pi(x_0, y_2) = \frac{2}{3}$ ■

Notice that this time the formulation allows for the mass to split allowing the solution to be feasible, unlike the Monge formulation.

## 3.2 Mass conservation

Ensuring mass conservation leads to restrictions being imposed on the measure $\pi(x, y)$. First, consider the measure $\pi(x, Y)$. This measures the amount of mass that goes from the point $x$ in the source set $X$ to the entire target set $Y$. This must equal to the amount of mass at point $x$, that is, it must go somewhere in $Y$. Similarly, $\pi(X, y)$ denotes the amount of mass coming from the whole of the source set $X$ to the point $y \in Y$. This also must equal the mass at $y$, in other words, it must come from somewhere in $X$. These two conditions can be written as,

$$\begin{aligned} \pi(x, Y) = \mu(x) \, or, \, \pi(A, Y) = \mu(A) \\ \pi(X, y) = \nu(y) \, or, \, \pi(X, B) = \nu(B) \\ \forall A \subset X \, and \, B \subset Y \end{aligned} \tag{4}$$

where, $\pi(x, Y)$ is called the marginal of $\pi$ on $X$ (with the $Y$ dependence integrated out) and $\pi(X, y)$ is called the marginal of $\pi$ on $Y$ (with the $X$ dependence integrated out).

## 3.3 Putting it all together

Weighing mass going from $x$ to $y$ ($\pi(x, y)$ by the cost of moving between these two points, $C(x, y)$, we get the pointwise cost. Since the cost and the transport measure both are functions of both source and target positions, we integrate over the product space of sets $X$ and $Y$.

$$inf\{\int_{X \times Y} C(x, y) d\pi(x, y) \mid \pi \in \Pi(\mu, \nu)\} \tag{5}$$

where $\Pi(\mu, \nu)$ is the set of measures on $X \times Y$ whose marginals on $X$ and $Y$ are $\mu$ and $\nu$ respectively. This is the Kantorovich formulation of OT.

The Kantorovich formulation is feasible in most cases. It is also noteworthy that this is linear in the transport measure $\pi$ unlike the Monge formulation which was clearly non-linear in terms of the transport map $T$.

**Some special cases:**

- Discrete OT: both source and target set consists of Dirac (or point) masses
- Continuous OT: $\mu$ and $\nu$ are absolutely continuous with mass densities $f$ and $g$ respectively
- Semi discrete OT: $\mu$ is absolutely continuous and $\nu$ consists of Dirac masses