Optimal Transport - Introduction

📅Date: 2023-01-10
🕐Time: 14:12

Bipin Koirala

Table of Contents

Introduction 🔍

Monge Formulation

Back in the 18^{th} century, Gaspard #Monge wanted to find an optimal way to transport/rearrange a pile of dirt into castle walls or other desired shapes. Later this task was coined as Monge's problem. Mathematically,

$$\begin{aligned}\text{optimize} &= \min_T \int_{\mathbb{R}^N} \mathrm{c}(x, T(x)) f(x) dx \\ &= \min_T \int_{\mathbb{R}^N} \big|T(x) - x\big| f(x) dx\end{aligned}$$

where, $T(x)$ is some optimal transformation of x , and sometimes the cost $c(x, T(x))$ is replaced by | . | --a distance metric.

- In general, instead of working directly with probability densities we want to setup the problem up using measures. Let μ be a source measure and ν be a target measure. For example, $\mu(X)$ tells us how much mass is present in the set X.

- Since we are simply transporting the mass from one measure to another, the total mass should be constant.

$$\mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$$

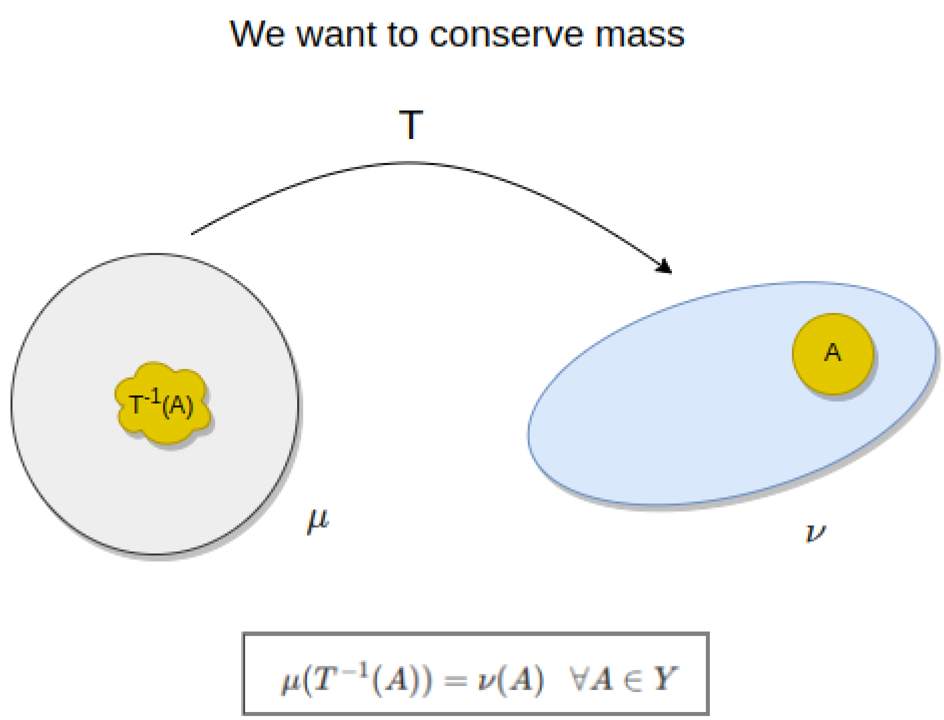
  Lot of the times the total mass is 1 and also often interpreted as probability measures.
- Now we seek the transport map $T(x)$ where source is supported on X and target on Y.

$$T : X \rightarrow Y$$

- Furthermore, we want to conserve mass not only globally but also locally.
  Let A be a subset in $\nu(Y)$ and if we want to find where it came from in the set $\mu(X)$; we can take $T^{-1}(A)$. Since, we the local mass is conserved; the following needs to be true:

$$\mu(T^{-1}(A)) = \nu(A) \quad \forall A \in Y$$

  Here; $\mu(T^{-1}(A))$ is called the #push-forward of μ through T. This is denoted by $T_{\#}\mu$!



$$\therefore T_{\#}\mu = \nu \quad \text{(aka mass conservation)}$$

- Therefore the #Monge formulation of #Optimal-Transport is the following:

$$\min\left\{ \int_{\mathbb{R}^N} c(x, T(x)) \ d\mu(x) \ \middle| T_{\#}\mu = \nu \right\}$$

Here, $d\mu(x)$ weights how much mass we're removing from X at a time.
Some issues that we may face when solving the above formulation are; feasibility and uniqueness of solution, stability, and figuring out a suitable cost function. Most of the time quadratic cost is often sought after. For example, in Book-moving problem, the cost $c(x, y) = |x - y|$ does not provide a unique solution but $c(x, y) = \frac{1}{2}|x - y|^2$ provides a unique solution.

Limitation of Monge's Formulation

In mines and factories setting, the number of mines does not necessarily have to equal to the number of factories. If we want to split a single mass(1) source into two targets each with mass $\frac{1}{2}$; the transformation $T(X)$ does not allow for splitting of the original mass. Clearly, #Monge formulation of the optimal transport doesn't work in this setting.

#Kantorovich formulation allows us to generalize the #Monge formulation. Kantorovich problem aims to seek a transport plan rather than a transport map which allows for the mass to go to different places i.e. it allows the mass to be split.

We have a source measure μ supported on X and a target measure ν supported on Y. We now want to learn how much mass gets moved from x to y. We store this information in another measure called π and is defined on product space $X \times Y$.

For example, suppose there is a mine at $x = 0$ with 1 unit of resource and also a factory at $y = 0, 1$ with $\frac{1}{3}$ and $\frac{2}{3}$ units of resources respectively. Here;

$$\pi(0,0) = \frac{1}{3}$$
$$\pi(0,1) = \frac{2}{3}$$

As a side note here, $\pi(0, \mathbb{R}) = 1$. In general, let $A \subset X$ and $B \subset Y$, then $\pi(A, B)$ tells us how much mass is transported from A to B.

Consider $x \in X$, and $\pi(x, Y)$ (quantifies how much mass is transported from a point x to all the potential targets in Y). By the conservation of mass, we can write.

$$\pi(x, Y) = \mu(x)$$

More generally, $\pi(A, Y) = \mu(A)$ and we say that μ is the marginal of π on X. Similarly, when $\pi(X, B) = \nu(B)$, we say ν is the marginal of π on Y.

Kantorovich Formulation

We know the cost $c(x, y)$ is weighted by the amount of mass we're moving from x to y.

$$\inf \int_{X\times Y} c(x,y)\, d\pi(x,y) \big| \pi \in \Pi(\mu,\nu)$$

Here, $\Pi(\mu,\nu)$ consists of measures whose marginals on X and Y are μ and ν respectively.

Special Cases:

- Discrete Optimal Transport: Dirac masses $\rightarrow$ Dirac masses
- Continuous Optimal Transport: μ and ν are continuous functions with densities f and g respectively
- Semi-discrete Optimal Transport: μ is absolutely continuous and ν consists of Dirac mass

Monge's Formulation in 1-D

Goal: Find $T(x)$ for

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

In simple terms, the constraint part tells us that the mass A in the source must be equal to the mass in the target region.
Alternatively,

$$\int_X h(T(x)) \, f(x) \, dx = \int_Y h(y) \, g(y) \, dy \ \ \forall \, h \in C^0(X)$$

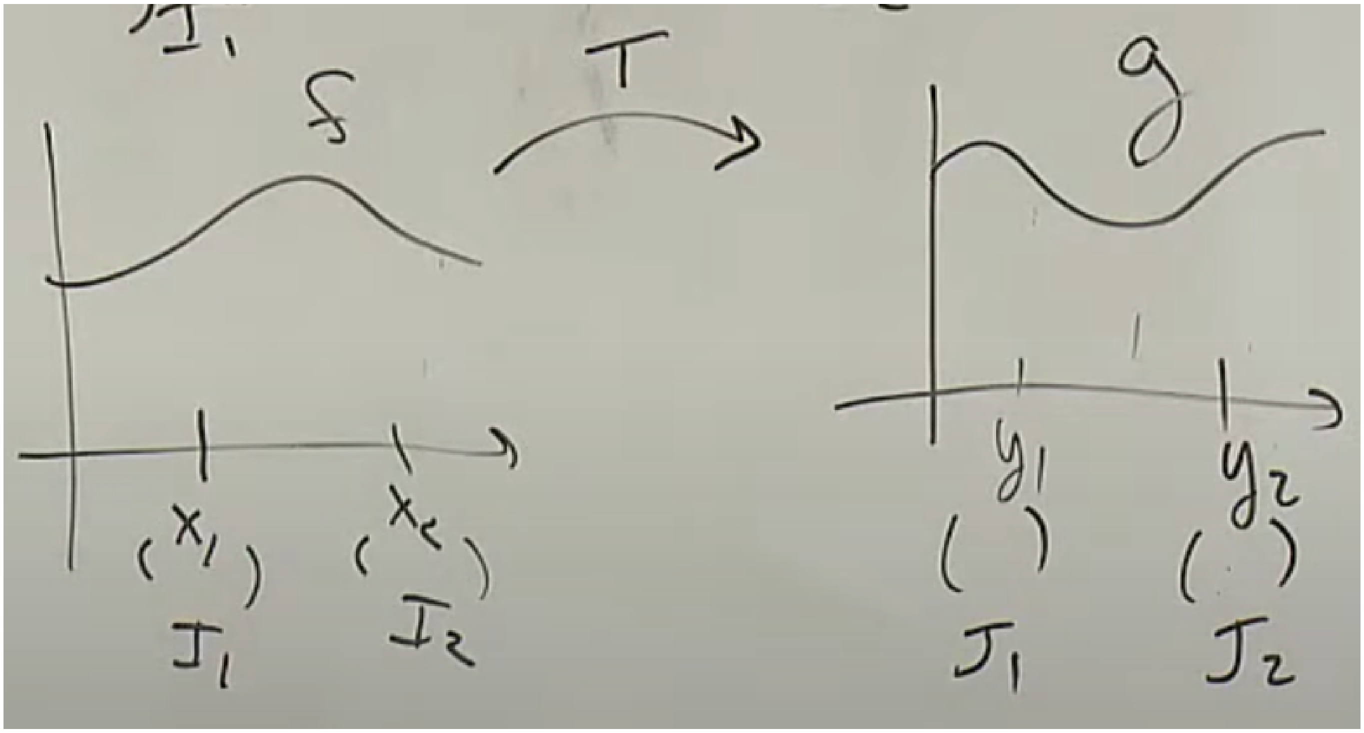
The above expression tells us that given any function h, the map $T(x)$ should preserve measure and also preserves what happens when we integrate over y.

Properties of optimal map

Pick two points, x_1 and x_2 such that $x_1 < x_2$ and $\epsilon > 0$. Make two little open intervals $x_1 \in I_1$, $x_2 \in I_2$ s.t.

$$\int_{I_1} f(x) \, dx = \int_{I_2} f(x) \, dx = \epsilon$$

i.e. total mass on I_1 is the same as total mass on I_2.



Here, $y_i = T(x_i)$ and $J_i = T(I_i)$.

Let's 'permute' part of the map and create a new measure-preserving map s.t.

$$\begin{aligned} \widetilde{T}(x_1) = y_2, \quad \widetilde{T}(x_2) = y_1 \\ \widetilde{T}(I_1) = J_2, \quad \widetilde{T}(I_2) = J_1 \\ \widetilde{T}(x) = T(x) \ \text{ if } x \notin I_1 \cup I_2 \end{aligned}$$

Now, under 'nice assumptions', if T was optimal.

$$\begin{aligned} \frac{1}{2}\int_{\mathbb{R}} (x - T(x))^2 \ f(x)dx \leq \frac{1}{2}\int_{\mathbb{R}} (x - \widetilde{T}(x))^2 \ f(x)dx \\ \Rightarrow -\int_{I_1} xT(x)f(x)dx - \int_{I_2} xT(x)f(x)dx \leq -\int_{I_1} x\widetilde{T}(x)f(x)dx - \int_{I_2} x\widetilde{T}(x)f(x)dx \\ \Rightarrow \frac{1}{\epsilon}\int_{I_1} x\big(\widetilde{T}(x) - T(x)\big) \ f(x) \ dx + \frac{1}{\epsilon}\int_{I_2} x\big(\widetilde{T}(x) - T(x)\big) \ f(x) \ dx \leq 0 \end{aligned}$$

As $\epsilon \to 0$:

$$\begin{aligned} x_1(y_2 - y_1) + x_2(y_1 - y_2) \leq 0 \\ \Rightarrow (y_2 - y_1)(x_2 - x_1) \geq 0 \end{aligned}$$

Since, $x_1 \leq x_2$ the above expression tells that the quantity $(y_2 - y_1)$ is positive. In other words, in the case of quadratic cost, the optimal transport map is a #monotone function in $\mathbb{R}$.

Can we construct a monotone map?

We use Cumulative Distribution Function (CDF) to construct the monotone map $T(x)$. Here,

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{and} \quad G(y) = \int_{-\infty}^{y} g(t)dt$$

We expect, $F(x) = G\big[T(x)\big]$ so we get an exact solution via $T(x) = G^{-1}F(x)$.