Optimal Transport

Preliminaries and Applications

Kirby

Department of Mathematics
University of Utah[1]

Feburary 2022

[1]This talk is neither property of nor endorsed by the University of Utah.

Optimal Transportation

History

Originally a question of optimally moving ammunition from factory to battlefield by Gaspard Monge in 1781 during Napoleonic France.

Optimal Transportation

History

Originally a question of optimally moving ammunition from factory to battlefield by Gaspard Monge in 1781 during Napoleonic France.

Formulation

Let X and Y be measure spaces with probability measures[a] μ, ν respectively. In order to transport $x \in X$ to $y \in Y$, we let $\mu(X) = \nu(Y)$. The optimal transport problem is minimizing the total effort (through nonnegative $c : X \times Y \to \mathbb{R}$) over μ and ν.

[a] The talk will simply label these as measures, but they're all assumed to be probability measures, or measures with $\mu(X) = 1$.

Monge Formulation

Transport Map

We call a function $T : X \to Y$ a *transport map* if $T_{\#}\mu = \nu$ (that is, $T_{\#}\mu$ is the pushforward of ν). Furthermore, we say T transports μ to ν.

Monge Formulation

Transport Map

We call a function $T : X \to Y$ a *transport map* if $T_{\#}\mu = \nu$ (that is, $T_{\#}\mu$ is the pushforward of ν). Furthermore, we say T transports μ to ν.

Monge Formulation

Let

$$\mathcal{T}(\mu, \nu) = \{T : X \to Y | T_{\#}\mu = \nu, \ \ T \text{ measurable}\} \tag{1}$$

Then we say that $\mathbb{M}(\mu, \nu)$ is the minimizer

$$\mathbb{M}(\mu, \nu) = \inf_{T \in \mathcal{T}(\mu, \nu)} \int_X c(x, T(x)) d\mu(x). \tag{2}$$

Monge Formulation Restrictions

Splitting Mass

Because we have $T : X \to Y$ with $T_{\#}\mu = \nu$, T is bijective, so it cannot *split mass*. This means for $y_1, y_2 \in Y$, a point $x \in X$ cannot have $T(x) = \{y_1, y_2\}$.

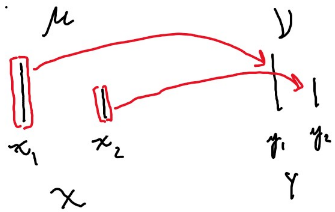

Figure: If we let $\mu = \frac{2}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2}$ and $\nu = \frac{2}{3}\delta_{y_1} + \frac{1}{3}\delta_{y_2}$ the only valid transport map is in red.

Kantorovich Formulation

Transport Plan

We call a measure $\pi \in \mathcal{P}(X \times Y)$ a *transport plan* if for all measurable $A \subset X$ and $B \subset Y$,

$$\pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B). \tag{3}$$

Kantorovich Formulation

Transport Plan

We call a measure $\pi \in \mathcal{P}(X \times Y)$ a *transport plan* if for all measurable $A \subset X$ and $B \subset Y$,

$$\pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B). \tag{3}$$

Kantorovich Formulation

Let

$$\Pi(\mu, \nu) = \{\pi | \pi \in \mathcal{P}(X \times Y) \text{ constrained by } (3)\} \tag{4}$$

Then we say that $\mathbb{K}(\mu, \nu)$ is the minimizer

$$\mathbb{K}(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y) d\pi(x, y). \tag{5}$$

Kantorovich Formulation as a Relaxation

We no longer have to be restricted to not splitting mass!

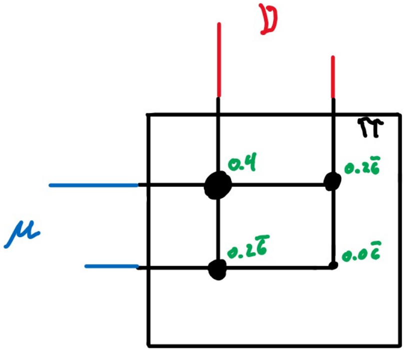

Figure: Using the same $\mu = \frac{2}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2}$ and $\nu = \frac{2}{3}\delta_{y_1} + \frac{1}{3}\delta_{y_2}$, we can define π to no longer be diagonal (not surjective)

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 mine x and factory y. I can make this sum as close to $c(x, y)$ as I want, so I can maximize my profit while minimizing the factory owner's effort.

Shipper's Problem cont.

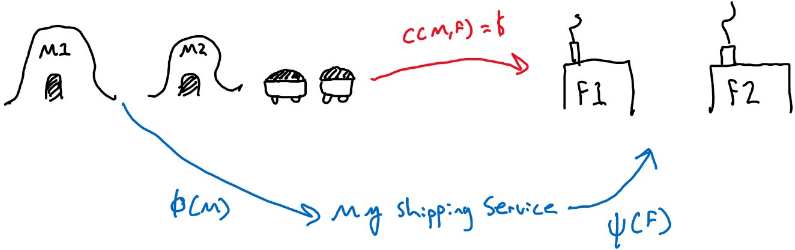

Figure: The owner has an easy option of just paying $c(M, F)$ for each mine and factory, or they can choose my strategy and pay me $\phi(M) + \psi(F)$.

Deriving Duality

For a measure π, we see that we have $I[\pi, \phi, \psi]$ defined as

$$\sup_{\phi \in C(X), \psi \in C(Y)} \left\{ \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu - \int_{X \times Y} (\phi(x) + \psi(y) d\pi) \right\}$$

is 0 when $\pi \in \Pi(\mu, \nu)$ and $+\infty$ otherwise. Now, writing this as a Kantorovich Problem for any positive radon measure π,

$$\mathbb{K}(\mu, \nu) = \inf_{\pi \geq 0} \left\{ \int_{X \times Y} c(x,y) d\pi(x,y) + \sup_{\phi \in C(X), \psi \in C(Y)} I[\pi, \phi, \psi] \right\} \quad (6)$$

We can informally$\star$ exchange the sup and inf to yield

$$\sup_{\phi, \psi} \left\{ \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu + \inf_{\pi \in \Pi} \int_{X \times Y} c(x,y) - (\phi(x) + \psi(y)) d\pi \right\} \quad (7)$$

Kantorovich Duality

Following from (7), we see that

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

This gives us the dual problem

Dual Problem

For $\mu\in\mathcal{P}(X),\nu\in\mathcal{P}(Y)$ and a nonnegative cost function $c: X\times Y\to\mathbb{R}$, the Kantorovich dual problem is

$$\sup_{\phi\in C(X),\psi\in C(Y)}\left\{\int_X \phi(x)d\mu(x)+\int_Y \psi(Y)d\nu(y)\right\} \tag{8}$$

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

Discussing Informal Exchange & Applications of Duality

Informal$\star$

There's more to the proof of Kantorovich Duality in general, but they all rely on a rigorous minimax principle.
In Villani, this is done using Fenchel-Rockafeller Duality and Legendre Transforms in section 1.1.
Message me at kirby#4923 with specific questions.

Discussing Informal Exchange & Applications of Duality

Informal$\star$

There's more to the proof of Kantorovich Duality in general, but they all rely on a rigorous minimax principle.
In Villani, this is done using Fenchel-Rockafeller Duality and Legendre Transforms in section 1.1.
Message me at kirby#4923 with specific questions.

Applications

Duality is important in the existence of transport maps (Brenier Polar Factorization).
The Monge-Kantorovich problem also can be shown to be a "continuous" linear programming problem. The proof of which is in "Partial Differential Equations and Monge-Kantorovich Mass Transfer" by Evans.

Wasserstein Distance

Let's start by considering Kantorovich-Monge solutions to the transport problem with cost functions $c(x, y) = d(x, y)^p$ where d is the distance on X and Y. Let $\Omega \subset X$ and set $x_0 \in X$.

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(X) : \int_\Omega d(x, x_0)^p < +\infty \right\}$$

is the admissable class of measures μ, even if Ω is unbounded.

Lemma (Kantorovich-Monge forms a metric.)

Let $W_p(\mu, \nu) = \left[\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)^p d\pi(x, y) \right]^{1/p}$ *defined for* $\mu, \nu \in \mathcal{P}_P(X)$ *be the Wasserstein Distance. This forms a metric on* X*.*

Wasserstein Distance cont.

We can informally interpret the Wasserstein Distance as the horizontal distance between measures.

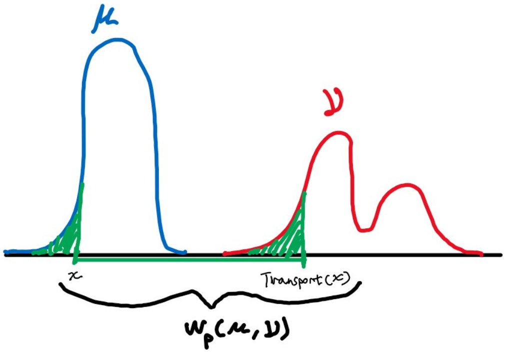

Figure: The optimal transport (in green) between μ and ν describes the distance between points in μ and ν informally. Hence, this is the Wasserstein Distance.

Wasserstein Distance Triangle Inequality

Assume $\mu, \nu, \eta \in \mathcal{P}_p(X)$ are absolutely continuous. Let T be the optimal map from μ to η and S be the optimal map from η to ν. Then $S \circ T \in \mathcal{T}(\mu, \nu)$ because $(S \circ T)_{\#}\mu = S_{\#}(T_{\#}\mu) = S_{\#}\eta = \nu$. Hence,

$$
\begin{aligned}
W_p(\mu, \nu)) &\leq \left(\int_X d(x, (S \circ T)(x))^p d\mu(x) \right)^{1/p} = \|d(\mathrm{Id}, (S \circ T))\|_{L^p(\mu)} \\
&\leq \|d(\mathrm{Id}, T)\|_{L^p(\mu)} + \|d(T, S \circ T)\|_{L^p(\mu)} \\
&= W_p(\mu, \eta) + \|d(\mathrm{Id}, S)\|_{L^p(\eta)} \\
&= W_p(\mu, \eta) + \int_X d(x, S(x))^p d\eta(x) = W_p(\mu, \eta) + W_p(\eta, \nu).
\end{aligned}
$$

Wasserstein Topology

We see that if $x_n \to x$ in X then $W_p(\delta_{x_n}, \delta_x) = d(x_n, x)^{\inf(1,p)} \to 0$. In general, this looks like

Theorem (Wasserstein distances metrize weak convergence)

Let $p \in (0, \infty)$, let $(\mu_k)_{k=1}^{\infty}$ be a sequence of measures in $\mathcal{P}_p(X)$ and $\mu \in \mathcal{P}(X)$, then the following are equivalent[a]:

1. $W_p(\mu_k, \mu) \to 0$
2. $\mu_k \to \mu$ *in the weak sense, meaning for $h \in C_b(X)$,*

$$\lim_{k\to\infty} \int h d\mu_k = \int h d\mu.$$

[a] Theorem 7.12 in Villani is much stronger and gives more equivalences, but due to time, I just want to make this statement.

Furthermore, we usually look at $p = 2$, and say $\mathbb{W}_2$ is the space $\mathcal{P}_2(X)$ endowed with the metric W_2.

Continuity Equation

The continuity equation is defined with

$$p_t + \nabla \cdot (p\vec{v}) = 0 \tag{9}$$

where p is the fluid density and $\vec{v}$ is the flow vector field.
Now, consider the smooth vector field $\vec{v} : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$. Let $\phi_t : \mathbb{R}^n \to \mathbb{R}^n$ be the flow map such that for each $x \in \mathbb{R}^n$,

$$\partial_t \phi_t(x) = \vec{v}(t, \phi_t(x)), \quad \phi_0(x) = x.$$

Given some measure p_0 on $\mathbb{R}^n$, we can look at the family of measures

$$p_t := (\phi_t)_{\#} p_0.$$

Continuity Equation

Theorem

The family of measures p_t is the unique distributional solution of the IVP of (9) with initial data $p_t|_{t=0} = p_0$.

Proof.

Let $\phi \in C_c^\infty(\mathbb{R}^n)$.

$$\int_0^T \int_{\mathbb{R}^n} \partial_t \phi(t,x) + \vec{v}(t,x) \cdot \nabla \phi(t,x) dp_t(x) dt$$

$$= \int_0^T \int_{\mathbb{R}^n} \partial_t \phi(t,\Phi_t(x)) + \vec{v}(t,\Phi_t(x)) \cdot \nabla \phi(t,\Phi_t(x)) dp_0(x) dt$$

$$= \int_{\mathbb{R}^n} \int_0^T \frac{d}{dt}[\phi(t,\Phi_t(x))] dt dp_0(x) = -\int_{\mathbb{R}^n} \phi(0,x) dp_0(x).$$

Hence p_t is a distributional solution. Uniqueness follows considering $\mu_t = \phi_t - \nu_t$ where ν_t is another solution. □

Bernamou-Brenier

Restrict $0 \leq t \leq 1$ and fix source (p_0) and destination (p_1) densities.

Theorem (Bernamou-Brenier)

The minimum effort required to transport p_0 to p_1 by the flow of the vector field $\vec{v}$ with p_t,

$$\inf \left\{ A[p_t, v] \middle| \partial_t p_t + \nabla \cdot (p_t \vec{v}) = 0, p_t|_{t=0} = p_0, p_t|_{t=1} = p_1 \right\} \tag{10}$$

is equivalent to the squared 2−Wasserstein Distance $W_2(p_0, p_1)^2$ where

$$A[p_t, v] = \int_0^1 \int_{\mathbb{R}^n} |\vec{v}(t,x)|^2 p_t(x) dxdt$$

A is also called the action where the inner integral is the kinetic energy E at time t. The action can also be interpreted as the average kinetic energy over $0 \leq t \leq 1$.

Bernamou-Brenier

Proving one side of the equality.

Choose admissable $\vec{v}, p_t$. We know tht $p_t = (\Phi_t)_{\#} p_0$ where Φ is the flow map of V.

$$\int_{\mathbb{R}^n} |\vec{v}(t,x)|^2 p_t(x)dx = \int_{\mathbb{R}^n} |\vec{v}(t,x)|^2 d(\Phi_t)_{\#} p_0$$
$$= \int_{\mathbb{R}^n} |\vec{v}(t,\Phi_t(x))|^2 p_0(x)dx = \int_{\mathbb{R}^n} |\partial_t \Phi_t(x)|^2 p_0 dx$$

Using Fubini's theorem and Jensen's inequality,

$$A[p_t, v] = \int_0^1 \int_{\mathbb{R}^n} |\partial_t \Phi_t(x)|^2 p_0 dxdt = \int_{\mathbb{R}^n} \left(\int_0^1 |\vec{v}(t,\Phi_t(x))|^2 dt \right) p_0(x)dx$$
$$\geq \int_{\mathbb{R}^n} \left(\int_0^1 \partial_t \Phi_t(x) dt \right)^2 p_0(x) = \int_{\mathbb{R}^n} |\Phi_1(x) - x|^2 p_0(x)dx$$
$$\geq W_2(p_0, p_1)^2.$$

$\square$

Why does this matter?

Gradient Flows in $\mathbb{R}^n$ have structure

$$\dot{x}(t) = -\nabla E(x(t)) \tag{11}$$

which may remind one of ODE flows. Usually E is convex and arise from problems

$$\min_{x \in \mathbb{R}^n} \{E(x)\}.$$

Wasserstein-Distances define geodesics in $\mathbb{W}_p$, or rather, $\mathbb{W}_p$ is a Geodesic Space.

Questions? Concerns?

This was my first talk, so it might have been rough around the edges, but I hoped you enjoyed it.

Recommended resources:

1. Cedric Villani - Topics in Optimal Transportation (good overview)
2. Fillipo Santambrogio - Optimal Transport for Applied Mathematicians (covers applications nicely)
3. Luigi Ambrosio - Gradient Flows (this is dry and very technical)
4. Matthew Thorpe - Introduction to Optimal Transport (nice summary of optimal transport, and covers geodesics in Wasserstein)

This talk was meant to spur questions or ideas, and applications were covered without proof, so ask me about them, or look up at the resources here :)