MATH 8803: Optimal Transport: Theory and Applications

Rui Gong

September 11, 2025

Acknowledgements

These notes are based on the MATH 8803 lectures given by Professor *Tobias Ried* in Fall 2025 at Georgia Institute of Technology.

Contents

1 Introduction

1.1 Monge's Original Formulation of Optimal Transport

Consider a measure μ, another measure ν, and x, y in the supports of μ and ν respectively, *what is the optimal way of moving x to y?*

1. Pile and hole
   - Pile and hole should have the same volume $\rightarrow$ normalize to 1.
   - modern way of thinking about pile and hole: probability measure on some metric space X and Y respectively, $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$. It could be point cloud: $\mu = \sum_i \alpha_i \delta_{x_i}$; or continuous densities: $\mu(dx) = f(x)dx$.
2. Transport pile to hole region: Transport described by a map $T : X \rightarrow Y$. Notice that T may be discontinuous. We need T to be measurable.
3. Transport Cost: $c : X \times Y \rightarrow [0, \infty) \cup \{\infty\}$, where $c(x, y)$ represents the cost of moving one unit of mass from x to y (how it is transported does not matter). Implicit Assumption: cost only depends on initial and final. Typical cost: $c(x, y) = |x - y|$; $c(x, y) = |x - y|^2$; $c(x, y) = \frac{1}{|x-y|}$.
4. Filling the hole completely: $\mu(T^{-1}(B)) = \nu(B)$ for every measurable $B \subseteq Y$.

Definition 1.1: Push-Forward Measure

Let $\mu \in \mathcal{X}$ be a probability measure on X, $T : X \rightarrow Y$ be a measurable map between metric space X, Y. Then push-forward (or image) measure *of μ under T is the measure $T_{\#}\mu$ on Y defined by $\mu(T^{-1}(B)) = T_{\#}\mu(B)$ for every $B \subseteq Y$ measurable.*

A little bit of functional analysis

- $C_b(X)$: the Banach space of bounded continuous function on X endowed with the norm $\|f\|_\infty := \sup_{x \in X} |f(x)|$.
- $C_0(X) \subseteq C_b(X)$: closed subspace (w.r.t. $\|\cdot\|_\infty$), which is the space of continuous functions vanishing at ∞: $f \in C_0(X)$ if $f \in C_b(X)$ and for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq X$, such that $|f| < \epsilon$ on $X \setminus K_\epsilon$.
- $\mathcal{M}(X)$: space of finite signed measures on X. $\lambda \in \mathcal{M}(X)$ if
  (a) $\lambda(A) \in \mathbb{R}$ for any (Borel) measurable $A \subseteq X$.
  (b) for every countable disjoint union $A = \cup_{i \in \mathbb{N}} A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$, these holds
    - $\sum_{i \in \mathbb{N}} |\lambda(A_i)| < \infty$
    - $\sum_{i \in \mathbb{N}} \lambda(A_i) = \lambda(A)$.

  To every $\lambda \in \mathcal{M}(X)$, we can associate a unique non-negatable measure $|\lambda| \in \mathcal{M}_+(X)$ via $|\lambda|(A) := \sup\{\sum_{i \in \mathbb{N}} |\lambda(A_i)| : A = \cup_{i \in \mathbb{N}} A_i, A_i \cap A_j = \emptyset \text{ for } i \neq j\}$, the total variation measure of λ. $\|\lambda\| := |\lambda|(X)$ is a norm on $\mathcal{M}(X)$.

Theorem 1.2: Riesz Representation Theorem

Suppose X is separable, and locally compact. Then $\mathcal{M}(X) \cong [C_0(X)]^*$ (the dual space of $C_0(X)$). That is, every continuous linear functional $L : C_0(X) \rightarrow \mathbb{R}$ is represented in a unique way by an element of $\mathcal{M}(X)$, i.e., there exists a unique measure $\mu_L \in \mathcal{M}(X)$ s.t. $L(\varphi) = \int_X \varphi d\mu_L$.

Remark. Consider a special case of $T_{\#}\mu = \nu$; assume that T is a C^1-diffeomorphism between X, Y and $X, Y \subseteq \mathbb{R}^d$ open, and that $\mu(dx) = f(x)dx$, $\nu(dy) = g(y)dy$. Then for any $B \subseteq \mathbb{R}^d$ measurable:

$$(T_{\#}\mu)(B) = \mu(T^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{T^{-1}(B)}(x) f(x) dx = \int_{T^{-1}(B)} f(x) dx.$$

Write $y = T(x), dy = |\det DT(x)|dx$, we can write

$$(T_{\#}\mu)(B) = \nu(B) = \int_B g(y)dy = \int_{T^{-1}(B)} g(T(x))|\det DT(x)|dx,$$

which implies $f(x) = g(T(x))|\det DT(x)|$ for almost every $x \in X$ (technically remark: $f \geq \alpha$ for some $\alpha > 0$ on X).

Reminder: $T_{\#}\mu = \nu$ means:

1. $(T_{\#}\mu)(B) = \mu(T^{-1}(B)) = \nu(B)$ for any measurable subset $B \subseteq Y$.
2. $\int_y \varphi d(T_{\#}\mu) = \int_X \varphi \circ T d\mu = \int_y \varphi d\nu$, $\forall \varphi \in C_0(y)$.

A quick remark on the change of variables formula:

$$\int_y \varphi d(T_{\#}\mu) = \int_X \varphi(T(x))\mu(dx).$$

Definition 1.3: Monge's Optimal Tranposrt Problem

Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$*,*

$$\min \ I[T] = \int_X c(x, T(x))\mu(dx) \qquad (\mathcal{M})$$

over all transport maps $T : X \to Y$ *(i.e., all measurable maps from* X *to* Y *such that* $T_{\#}\mu = \nu$*).*

Remark.

- I is a highly nonlinear functional of T subject to the nonlinear constraint $T_{\#}\mu = \nu$.
- Functional relatively simple: depends only locally on T (or its pointwise values)
  - no coupling between different values of T.
  - without constraint could just minimize pointwise, i.e. find minimum $y_{\min}(x)$ of $y \mapsto c(x, y)$ for each x and get $T(x) = y_{\min}(x)$.
- constraint complicated: nonlocal, couples values of T. If we could restrict to smooth diffeomorphisms, problem requires solving highly nonlinear PDE. Also, it is not even clear whether T such that $T_{\#}\mu = \nu$ exists for given μ, ν.

1.2 The Kantorovich Optimal Transport Problem

Comparing to Monge's OT problem, Kantorovich OT problem allows measure splitting, so we are looking for probability measure on $X \times Y$.

- Pile and hole: $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$.
- Transport: probability measure $\gamma \in \mathcal{P}(X \times Y)$ (transport plan).

  $$\gamma(A \times B) = \int_{A \times B} \gamma(dxdy)$$

  is the amount of mass moved from measurable $A \subseteq X$ to measurable $B \subseteq Y$. All the mass of μ has to be transported somewhere, hence, $\gamma(A \times Y) = \mu(A)$ for all $A \subseteq X$ measurable.
- Transport cost: Let $c(x, y)$ be the cost of moving one unit of mass from x to y, then the total cost is

  $$\int_{X \times Y} c(x, y)\gamma(dxdy) = c[\gamma].$$

- Filling hole completely: $\gamma(X \times B) = \nu(B)$ for all $B \subseteq Y$ measurable. That is, the amount of mass transported to B has to be the volume of the hole in region B.

Remark. Note that, from above, $\gamma(X \times Y) = \mu(X) = \nu(Y) = 1$, so $\gamma \in \mathcal{P}(X \times Y)$, and $\gamma(A \times Y), \gamma(X \times B)$ defined marginals.

Definition 1.4: Marginals

Let $\gamma \in \mathcal{P}(X \times Y)$.

- *Marginal w.r.t. X: $M_X\gamma \in \mathcal{P}(X)$ defined via*

$$(M_X\gamma)(A) = \gamma(A \times Y) = \int_{A \times Y} \gamma(dxdy), \; \forall \text{ measurable } A \subseteq X,$$

- *Marginal w.r.t. Y: $M_Y\gamma \in \mathcal{P}(Y)$ defined via*

$$(M_Y\gamma)(B) = \gamma(X \times B) = \int_{X \times B} \gamma(dxdy), \; \forall \text{ measurable } B \subseteq Y.$$

Remark. Transport plans are probability measures on $X \times Y$ with marginals $M_X\gamma = \mu$, $M_Y\gamma = \nu$, γ is a *coupling* of the probability measure μ and ν.

Let $\Pi(\mu, \nu)$ be the set of all couplings between μ and ν.

Lemma 1.5

Let $\varphi \in L^1(X, \mu)$ and $\psi \in L^1(Y, \nu)$. Then for any coupling $\gamma \in \Pi(\mu, \nu)$. These hold

(M1) $\int_{X \times Y} \varphi(x)\gamma(dxdy) = \int_X \varphi(x)(M_X\gamma)(dx) = \int_X \varphi(x)\mu(dx)$.

(M2) $\int_{X \times Y} \psi(y)\gamma(dxdy) = \int_Y \psi(y)(M_Y\gamma)(dy) = \int_Y \psi(y)\nu(dy)$.

Proof sketch. Any function $\varphi \in L^1(X, \mu)$ can be approximated by simple functions: $\varphi = \lim_{n\to\infty} \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}$, for $A_j \subseteq X$ measurable.

$$\begin{aligned}\int_{X \times Y} \varphi(x)\gamma(dxdy) &= \lim_{n\to\infty} \sum_{j=1}^n \alpha_j \int_{A_j \times Y} \gamma(dxdy) = \lim_{n\to\infty} \sum_{j=1}^n \alpha_j \mu(A_j) \\ &= \lim_{n\to\infty} \int_X \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}(x)\mu(dx) = \int_X \varphi(x)\mu(dx).\end{aligned}$$

□

Definition 1.6: Couplings

Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. A probability measure $\gamma \in \mathcal{P}(X \times Y)$ is called coupling of μ and ν if $M_X\gamma = \mu$, $M_Y\gamma = \nu$. The set of all couplings between μ and ν is called $\Pi(\mu, \nu)$.

Definition 1.7: Kantorovich Optimal Transport Problem

Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$,

$$\min \; C[\gamma] = \int_{X \times Y} c(x, y)\gamma(dxdy) \tag{$\mathcal{K}$}$$

over all couplings $\gamma \in \Pi(\mu, \nu)$.

Structure of ($\mathcal{K}$):

(1) $\gamma \mapsto C[\gamma]$ linear function.

(2) $M_X\gamma = \mu$, $M_Y\gamma = \nu$ linear constraints.

(3) $\Pi(\mu,\nu)$ is a convex set: if $\gamma_1, \gamma_2 \in \Pi(\mu,\nu)$, $\lambda \in (0,1)$, then

 (a) $\lambda\gamma_1 + (1-\lambda)\gamma_2 \in \mathcal{P}(X\times Y)$, since $\lambda\gamma_1(X\times Y) + (1-\lambda)\gamma_2(X\times Y) = 1$, and $\lambda\gamma_1(Z) + (1-\lambda)\gamma_2(Z) \geq 0$ for any $Z \subseteq X \times Y$ measurable.

 (b) $(\lambda\gamma_1 + (1-\lambda)\gamma_2)(A\times Y) = \lambda\gamma_1(A\times Y) + (1-\lambda)\gamma_2(A\times Y) = \lambda\mu(A) + (1-\lambda)\mu(A) = \mu(A)$; analogously, $(\lambda\gamma_1 + (1-\lambda)\gamma_2)(X\times B) = \nu(B)$, for any $A \subseteq X, B \subseteq Y$ measurable.

 $\Longrightarrow$ ($\mathcal{K}$) is a linear programming problem. *But*: it is in inifinte dimensions.

(4) Existence of couplings is trivial: independent coupling (product measure) $\gamma = \mu \otimes \nu$ defined via $\gamma(A\times B) = \mu(A)\nu(B)$ for every measurable $A \subseteq X, B \subseteq Y$, is a coupling of μ and ν.

(5) ($\mathcal{K}$) is higher-dimensional than ($\mathcal{M}$) in the following sense: consider transport plan given by a density $\tilde{\gamma} : \mathbb{R}^{2d} \to \mathbb{R}$. Discretize $\mathbb{R}^d$ by ℓ gridpoints then $\tilde{\gamma}$ corresponds to ℓ^2 real numbers. Transport $T : \mathbb{R}^d \to \mathbb{R}^d$ however only corresponds to ℓd real numbers. e.g. $\ell =$ number of pixels in a 2D picture, say $\ell = 500 \times 500$, then $\ell d = 500000$, but $\ell^2 = 62.5^9$.

1.3 Monge VS Kantorovich

Kantorovish problem is a relaxation of Monge Problem in the following sense: Monge = restriction of Kantorovish to sparse plan.

$$\gamma_T(dxdy) = \delta_{T(x)}(dy)\mu(dx)$$

for $T : X \to Y$ measurable such that $T_\#\mu = \nu$. For any $\varphi \in C_0(X\times Y)$,

$$\begin{aligned}\int_{X\times Y} \varphi d\gamma_T &= \int_{X\times Y} \varphi(x,y)\gamma_T(dx\times dy) = \int_{X\times Y} \varphi(x,y)\delta_{T(x)}(dy)\mu(dx) = \int_X \varphi(x,T(x))\mu(dx)\\ &= \int_X \varphi((id,T)(x))\mu(dx) = \int_X \varphi\circ(id,t)d\mu = \int_{X\times Y}\varphi(x,y)[(id,T)_\#\mu](dx\times dy).\end{aligned}$$

In other words, Monge measures that

$$\begin{aligned}\operatorname{supp}\gamma &= \{(x,y)\in X\times Y : \gamma(B_\epsilon(x,y)) > 0 \text{ for every } \epsilon > 0\}\\ &= \cap\{Z \subseteq X\times Y \text{ closed } : \gamma(Z) = 1\}\\ &\subseteq \operatorname{graph} T = \{(x,T(x)) \in X\times Y : x \in X\}\end{aligned}$$

Lemma 1.8

Let $T : X \to Y$ be measurable such that $T_\#\mu = \nu$. Then

(i) $C[\gamma_T] = I[T]$.

(ii) $\gamma_T \in \Pi(\mu,\nu)$

(iii) ($\mathcal{K}$) $\leq$ ($\mathcal{M}$).

Proof. (i) $C[\gamma_T] = \int_{X\times Y} c(x,y)\gamma_T(dx\times dy) = \int_{X\times Y} c(x,y)((id,T)_\#\mu)(dxdy) = \int_X c(x,T(x))\mu(dx) = I[T]$.

(ii) Let $A \subseteq X$, $B \subseteq Y$ measurable. Then

$$
\begin{aligned}
(M_X\gamma_T)(A) &= \gamma_T(A\times Y) = \int_{A\times Y} d\gamma_T = \int_X \mathbb{1}_{A\times Y}(x,y)\ \underbrace{\gamma_T(dx\times dy)}_{((id,T)_{\#}\mu)(dxdy)} \\
&= \int_X \underbrace{\mathbb{1}_{A\times Y}(x,T(x))}_{\mathbb{1}_A(x)}\mu(dx) \\
&= \int_A d\mu = \mu(A) \\
(M_Y\gamma_T)(B) &= \int_X \underbrace{\mathbb{1}_{X\times B}(x,T(x))}_{\mathbb{1}_B(x)}\mu(dx) = \int_Y \mathbb{1}_B(y)(T_{\#}\mu)(dy) \\
&= \int_A d\nu = \mu(A)
\end{aligned}
$$

(iii) By (i) and (ii),

$$
(\mathcal{M}) = \inf_{T:X\to Y \text{ measurable}, T_{\#}\mu=\nu} I[T] = \inf_{T:X\to Y \text{ measurable}, T_{\#}\mu=\nu} C[\gamma_T] \geq \inf_{\gamma\in\Pi(\mu,\nu)} C[\gamma] = (\mathcal{K}),
$$

by $\gamma_T \in \Pi(\mu,\nu)$.

$\square$