

Day 2

Polar Factorization Theorem (Brenier)

In same setting as Brenier's Theorem. Any other valid transport map T s.t. $T = T_0 \circ S$ where S is a measure preserving map $\mu \rightarrow \mu$.

Eg. If $S = Id$ $T = T_0$

(*) This is generalization of Polar Factorization of matrix
Remarks any matrix $M = PO \rightarrow$ orthonormal

$\downarrow$
PSD

If $X \sim N(0, I_d)$ $MX \sim N(0, MM^T)$ $M = P \cdot O$ $\left\{ \begin{array}{l} \phi(x) = \frac{x^T P x}{2} \\ \nabla \phi(x) = P x = T_0(x) \end{array} \right.$
 $\downarrow$
OT Map

Duality

discrete case $\rightarrow$

$$\max_{\alpha, \beta} \frac{1}{n} \sum_{i=1}^n \alpha_i + \frac{1}{n} \sum_j \beta_j$$

s.t. $\alpha_i + \beta_j \leq c_{ij}$

General case $\rightarrow$
 Dual form)

$$\max_{\substack{f \in L^1(\mu) \\ g \in L^1(\nu)}} \int f d\mu + \int g d\nu$$

s.t. $f(x) + g(y) \leq c(x, y)$

(*) Strong duality If c is lower semi-continuous & bounded below the strong duality holds.

Structural Remarks when $p=1$

1. Fix f , optimize over g s.t. $g(y) \leq \inf_x [c(x, y) - f(x)] = f^c(y)$

(*) optimal g is f^c when f is fixed.

$\downarrow$
 c -transform / conjugate of

2. $p=1$ $c(x, y) = \|x - y\| \rightarrow$ euclidean cost

(*) c -transform of f is 1-Lipschitz

(*) c -transform of 1-Lipschitz f is $-f$

3. Assume they are true — (iterative argument of 1)

$$\max_{f \in L^1(\mu)} \int (f^c)^c d\mu + \int (f^c)^c d\nu$$

* f^c is 1-Lipschitz & $(f^c)^c = -f^c$

* when $c(x, y) = \|x - y\|$ the optimization

$$\max_{f \in 1\text{-Lipschitz}} \int f d\mu - \int f d\nu$$

↗ Kantorovich-Rubinstein duality

$$W_1(\mu, \nu) = \sup_{f \in 1\text{-Lipschitz}} \left[\int f d\mu - \int f d\nu \right]$$

IPM

* W_1 is an Integral Probability Metric (IPM) which is defined

$$\text{as } \rho(\mu, \nu) = \sup_{f \in \mathcal{F}} \left| \int f d\mu - \int f d\nu \right|$$

* TV distance is an IPM.

$$TV(\mu, \nu) = \sup_{f \in \{ \|f\|_\infty \leq 1 \}} \left| \int f d\mu - \int f d\nu \right| \rightarrow \text{big class of function}$$

* Maximum Mean discrepancy (MMD) is also an IPM

$$MMD(\mu, \nu) = \sup_{f, \|f\|_k \leq 1} \left| \int f d\mu - \int f d\nu \right|$$

↓

↗ RKHS Norm

RKHS unit ball {not very big}

Proof of property in ②

$$g(y) = \inf_x \left[\|x - y\| - f(x) \right]$$

$$\tilde{g}(y) = \|x_0 - y\| - f(x_0) \leftarrow 1\text{-Lipsh}$$

$$\begin{aligned} \tilde{g}(y) - \tilde{g}(y') &= \|x_0 - y\| - \|x_0 - y'\| \\ &\leq \|y - y'\| \end{aligned}$$

Conclusion. $g = f^c$ is 1-Lipschitz as pointwise infimum of 1-Lip chitz $\tilde{g}(y)$ function

Property-2

If f is 1-Lipschitz $(-f)^c = f$

$$g(y) = \inf_x [\|x-y\| + f(x)] \leftarrow c \text{ transform of } -f$$

$$g(y) \leq f(y) \quad \left| \quad \begin{aligned} f(y) &\leq \|x-y\| + f(x) \\ f(y) &\leq \inf_x [\|x-y\| + f(x)] = g(y) \end{aligned} \right.$$

$$\Rightarrow f = g \quad \text{i.e. } (-f)^c = f$$

Duality for squared Euclidean Cost

Starting point: $\sup_{\substack{f \in L^1(\mu) \\ g \in L^1(\nu)}} \int f d\mu + \int g d\nu$ (General Kantorovich Dual form)

$$\text{s.t. } f(x) + g(y) \leq \|x-y\|_2^2$$

⊛ Optimal f & g - Kantorovich potential

$$\begin{aligned} \text{⊛ } f(x) &= \|x\|^2 - 2\phi(x) \\ g(y) &= \|y\|^2 - 2\psi(y) \end{aligned} \quad \left. \vphantom{\begin{aligned} f(x) &= \|x\|^2 - 2\phi(x) \\ g(y) &= \|y\|^2 - 2\psi(y) \end{aligned}} \right\} \text{ define } \phi \text{ \& } \psi$$

Rewrite
dual $\rightarrow$
Form

$$\sup_{\phi, \psi} \int \|x\|^2 d\mu + \int \|y\|^2 d\nu - 2 \int \phi(x) d\mu - 2 \int \psi(y) d\nu$$

$$\text{s.t. } \phi(x) + \psi(y) \geq x^T y$$

$$\Rightarrow \inf_{\phi, \psi} \int \phi d\mu + \int \psi d\nu$$

$$\text{s.t. } \phi(x) + \psi(y) \geq x^T y$$

$$\psi(y) \geq \sup_x [x^T y - \phi(x)] := \phi^*(y) \quad \left(\begin{array}{l} \text{Frenchel} \\ \text{Conjugate} \end{array} \right)$$

$$\boxed{\inf_{\phi \in L^1(\mu)} \int \phi d\mu + \int \phi^* d\nu} \leftarrow \text{semi-dual}$$

$\nearrow$ Always convex

Remark

⊛ If you satisfy conditions of Brenier's Theorem

$$T_0 := \nabla \phi_0 \text{ where } \phi_0 = \arg \inf_{\phi \in L^1(\mu)} \int \phi d\mu + \int \phi^* d\nu$$

$$\otimes \phi(x) + \phi^*(\nabla \phi(x)) = \langle x, \nabla \phi(x) \rangle$$

Statistical Optimal Transport

↓
convex.

⊛ μ, ν are not known

→ $x_1, \dots, x_n \sim \mu$ & $y_1, \dots, y_n \sim \nu$ μ is absolute continuous.

Questions

⊛ Target of inference

1. Estimate T_0 : $\frac{1}{n} \sum \nabla \phi(x_i) \approx T_0$? → how many samples do I need?

Applications

⇒ Generative modeling. $x_1, \dots, x_n \sim \mathcal{N}(\mu, \Sigma)$
 $y_1, \dots, y_n \sim \text{real images}$

⇒ Waddington OT: broken correspondence samples

2. Estimate $W_2(\mu, \nu)$ or $W_1(\mu, \nu)$ → Hypothesis testing

Hypothesis Testing

Two sample $H_0: \mu = \nu$

$H_1: W_p(\mu, \nu) \geq \varepsilon$

3. Density estimation $x_1, \dots, x_n \sim \mu$

construct $\hat{\mu}$ so that $W_p(\hat{\mu}, \mu)$ is small → What is sample complexity?

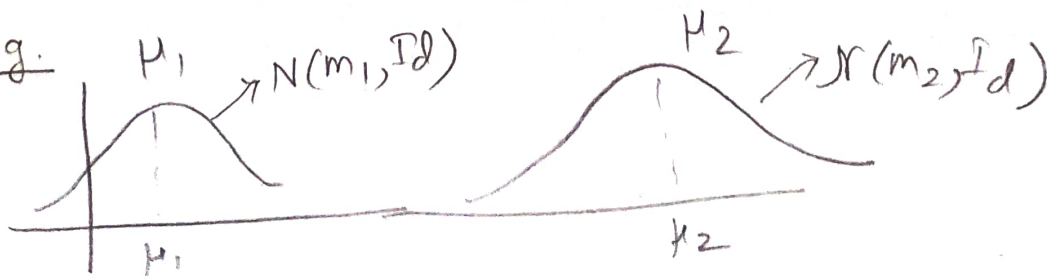
4. Barycenter $\mu_1, \dots, \mu_k \leftarrow k \text{ distribution}$ Average of dist.

$$\bar{\mu} = \arg \min_{\mu} \sum_i W_p(\mu, \mu_i)$$



"shape preserving property"

E.g.



⇒ One defn of mid-point = $\frac{1}{2} \mathcal{N}(m_1, \Sigma_d) + \frac{1}{2} \mathcal{N}(m_2, \Sigma_d)$

⇒ Wasserstein mid-point will be $\mathcal{N}(\bar{m}, \Sigma_d)$

Estimating the OT map

Dual Estimators

Primal Estimators

$$\phi_0 = \underset{\phi \in L(\mu)}{\operatorname{argmin}} \int \phi d\mu + \int \phi^* d\nu$$

Brenier
potential

↳ Looks like ERM

ERM principle

$$\hat{\phi} = \underset{\phi \in \Phi}{\operatorname{argmin}} \int \phi d\mu_n + \int \phi^* d\nu_n$$

where

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\nu_n = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}$$

$$\hat{T} = \nabla \hat{\phi}$$

Reference Makkuva et al. (2020.)

Input: Convex NN (Amos et al.)