Recap:

Monge: $\displaystyle \min_{T_\# \mu = \nu} \int c(x, T(x))\, d\mu \quad \leftarrow$

Kantorovich: $\displaystyle \min_{\gamma \in \Gamma_{\mu,\nu}} \int c(x,y)\, d\gamma \quad \leftarrow$

Brenier: If μ is absolutely continuous, μ, ν have 2 bdd moments, & $c(x,y) = \|x-y\|^2$

$T_0 = \nabla \varphi_0 \quad \varphi_0$ is convex

& $\quad \gamma_0 \sim (x, \nabla\varphi_0(x))$.

Polar Factorization: Any transport map T can be decomposed as $T = T_0 \circ S$ where S is measure-preserving.

Wasserstein distance: $\displaystyle W_p(\mu,\nu) = \left[\min_{\gamma \in \Gamma_{\mu,\nu}} \int \|x-y\|^p\, d\gamma \right]^{1/p}$

Duality:

General dual to Kantorovich:

$$\sup_{\substack{f \in L^1(\mu) \\ g \in L^1(\nu)}} \int f\, d\mu + \int g\, d\nu \qquad f(x) + g(y) \le c(x,y)$$

(a) For $p = 1$, (ie) $c(x,y) = \underbrace{\|x-y\|}$

Equivalent dual & strong duality show that

$$\underline{\underline{W_1(\mu,\nu)}} = \sup_{\substack{f \in L^1(\mu) \\ \underline{f:\ 1\text{-Lipschitz}}}} \underline{\int f\, d\mu} - \underline{\int f\, d\nu}.$$

(ie) W_1 is an integral probability metric over class of 1-Lipschitz functions.

(b) For $p = 2$: Semi-dual $\quad c(x,y) = \|x-y\|^2$

$$\left[\min_{\varphi \in L^1(\mu)} \underbrace{\int \varphi\, d\mu} + \underbrace{\int \varphi^* d\nu} \right] \leftarrow$$

$$\varphi^*(y) = \sup_x \left[x^T y - \varphi(x) \right] \leftarrow$$

& if conditions of Brenier's theorem hold then $\quad T_0 = \nabla \varphi_0$ where φ_0 optimizes semi-dual.

Statistical OT:

<u>Setting</u>: $\quad X_1, \dots X_n \sim \mu$ ✓
$\quad Y_1, \dots Y_n \sim \nu$ ✓

Want to estimate:

(1) T_0 (if μ is absolutely continuous)

(2) $W_p(\mu, \nu)$

(3) Want to construct $\hat{\mu}$ so that $W_p(\hat{\mu}, \mu)$ is small.

(4) Estimate Barycenter

$$\bar{\mu} = \arg\min_{\mu} \sum_j W_2(\mu, \mu_j)$$

Dual estimators:

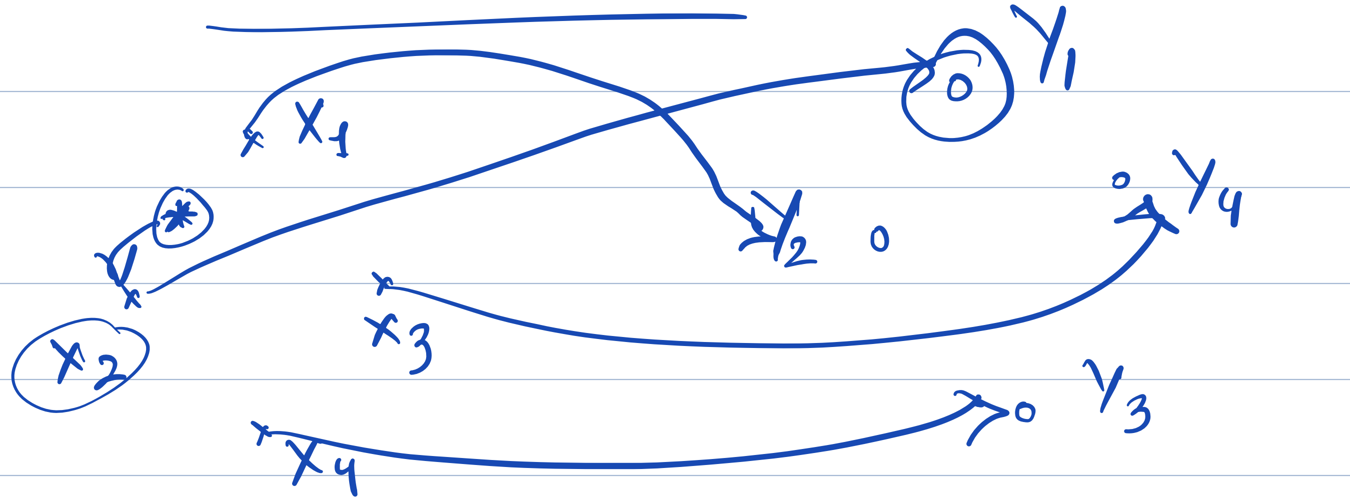
$$\hat{\varphi} = \arg\min_{\varphi \in \Phi} \int \varphi \, d\mu_n + \int \varphi^* \, d\nu_n$$

$$\boxed{\hat{T} := \nabla \hat{\varphi}}$$

* What statistical properties does it have?

φ_0 — Brenier potential.

Primal Estimators:



"compute an in-sample matching".

Hungarian alg / Auction alg. (solve LP)

we want an extension of in-sample map.

→ use nearest neighbour principle

$$\hat{T}(x_{new}) := \hat{T}(X_{nn}(x_{new}))$$

→ what stat. prop does it have?

Plugin estimators:

$X_1, \dots X_n \sim \mu$. , $Y_1, \dots Y_n \sim \nu$

Construct density estimates $\hat{\mu}$, $\hat{\nu}$ ←

$$\hat{T} = \underbrace{\text{OT Map}(\hat{\mu}, \hat{\nu})}$$

→ what prop. does it have?

$$\left\{ \begin{array}{l} \hat{\mu} \to N \text{ samples.} \\ \hat{\nu} \to \text{" "} \end{array} \right\}$$

Risk:

$$R(\hat{T}, T_0) = \mathbb{E}\left[\underbrace{\int \|\hat{T}(x) - T_0(x)\|_2^2 \, d\mu}\right]$$

$\sim \mathcal{L}^2$ risk measure.

Brenier regularity: $\rightarrow \nu = (\nabla \varphi_0)_{\#}\mu.$

φ_0 - strongly convex & smooth.

$$\rightarrow \boxed{\frac{I_d}{\lambda} \lesssim \nabla^2 \varphi_0 \lesssim \lambda I_d} \; *$$

(1) μ, ν are supported on Ω

$\mu, \nu > 0$.

Ω - ~~connected~~, convex set with smooth boundaries.

then "Caffarelli regularity theory" will certify Brenier regularity.

Monge - Ampere: $\rightarrow \mu = \nu(\nabla \varphi_0(x)) \, |\det \nabla^2 \varphi_0(x)|$

(2) $\underline{\mu,\nu}$ ~~&~~ are smooth & strongly $\underbrace{\text{log-concave}}$.

$$|\det \nabla^2 \varphi_0| = \underset{\sim}{f}.$$

Dual stability:

$$S_{\mu,\nu}(\varphi) = \int \varphi \, d\mu + \int \varphi^* \, d\nu \leftarrow$$

Lemma: Functional has quadratic growth, for φ is smooth & strongly convex,

$$\underbrace{\|\nabla\varphi - \nabla\varphi_0\|^2_{L^2(\mu)}} \lesssim \underbrace{S_{\mu,\nu}(\varphi)} - \underbrace{S_{\mu,\nu}(\varphi_0)} \lesssim \|\nabla\varphi - \nabla\varphi_0\|^2_{L^2(\mu)}$$

Statistical analysis:

$$\hat{\varphi} = \underset{\varphi \in \Phi}{\operatorname{argmin}} \, S_{\mu_n,\nu_n}(\varphi).$$

transport map/potential selection.

$\to \Phi \in \{\varphi_1, \dots \varphi_M\}$.

→ $\varphi_0 \in \Phi$

→ bdd domain, $\|\varphi_j\|_\infty \leq B$, $\|\varphi_j^*\|_\infty \leq B$.

→ All potentials are smooth & strongly convex.

Basic inequality:

→ $S_{\mu_n,\nu_n}(\hat{\varphi}) \leq S_{\mu_n,\nu_n}(\varphi_0)$

→ $S_{\mu,\nu}(\hat{\varphi}) - S_{\mu,\nu}(\varphi_0)$

$$= S_{\mu,\nu}(\hat{\varphi}) - S_{\mu_n,\nu_n}(\hat{\varphi}) + S_{\mu_n,\nu_n}(\hat{\varphi}) - S_{\mu_n,\nu_n}(\varphi_0) + S_{\mu_n,\nu_n}(\varphi_0) - S_{\mu,\nu}(\varphi_0)$$

$$\leq S_{\mu,\nu}(\hat{\varphi}) - S_{\mu_n,\nu_n}(\hat{\varphi}) + S_{\mu_n,\nu_n}(\varphi_0) - S_{\mu,\nu}(\varphi_0)$$

$$\leq 2 \max_{j \in \{1,\ldots,M\}} \left| \underbrace{S_{\mu,\nu}(\varphi_j) - S_{\mu_n,\nu_n}(\varphi_j)} \right|$$

$\|\nabla\hat{\varphi} - \nabla\varphi_0\|^2_{L^2(\mu)} \leq \boxed{2(\max)}$ which is small.

$$\| \hat{T} - T_0 \|^2_{L^2(\mu)} \lesssim \epsilon \lesssim \sqrt{\frac{\log M}{n}}$$

Fix a φ

$$\underbrace{\int \varphi \, d\mu + \int \varphi^* d\nu}_{} - \underbrace{\int \varphi \, d\mu_n - \int \varphi^* d\nu_n}_{}.$$

Hoeffding's ineq + union bound.

with prob at least $1-\delta$

$$\max_{j \in \{1, \dots, M\}} \left| S_{\mu,\nu}(\varphi_j) - S_{\mu_n, \nu_n}(\varphi_j) \right| \lesssim 2B \sqrt{\frac{\log(M/\delta)}{2n}}$$

Proof of quadratic growth:

φ – smooth & strongly convex.

$$S_{\mu,\nu}(\varphi) - S_{\mu,\nu}(\varphi_0)$$

$$= \int \varphi \, d\mu + \int \varphi^* d\nu - \int \varphi_0 \, d\mu - \int \varphi_0^* d\nu.$$

Facts about Fenchel conjugates:

$$\varphi^*(y) = \sup_x \left[x^T y - \varphi(x) \right]$$

$$\varphi(x) + \varphi^*(y) \geq x^T y \qquad \text{for any } (x,y) \leftarrow$$

$y = \nabla\varphi(x)$ then equality holds

$$\varphi(x) + \varphi^*(\nabla\varphi(x)) = x^T \nabla\varphi(x). \quad \star$$

$$(\nabla\varphi_0)_\# \mu = \nu.$$

$$\boxed{\int \varphi_0^* \, d\nu = \int \varphi_0^*(\nabla\varphi_0(x)) \, d\mu(x)}$$

$$S_{\mu,\nu}(\varphi_0) = \int \varphi_0 \, d\mu + \int \varphi_0^* d\nu.$$

$$= \int \left[\varphi_0(x) + \varphi_0^*(\nabla\varphi_0(x)) \right] d\mu$$

$$= \int \langle x, \nabla\varphi_0(x)\rangle \, d\mu .$$

$S_{\mu,\nu}(\varphi) - S_{\mu,\nu}(\varphi_0)$

$$= \int \varphi \, d\mu + \underbrace{\int \varphi^* \, d\nu}_{\varphi} - \underbrace{\int \langle x, \nabla\varphi_0(x)\rangle \, d\mu}$$

$\rightarrow \varphi^*(\nabla\varphi(x)) + \varphi(x) = \langle x, \nabla\varphi(x)\rangle.$

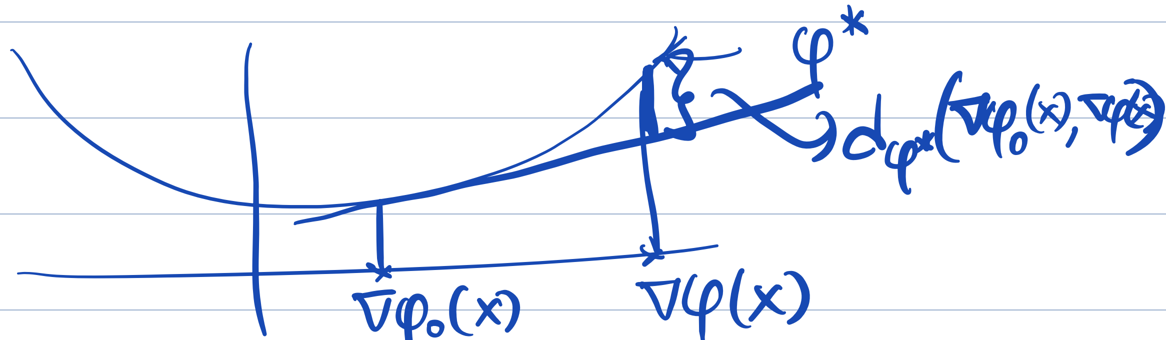
Fenchel equality.

$$= \int \Big[\varphi^*(\nabla\varphi_0(x)) - \varphi^*(\nabla\varphi(x)) - \langle x, \nabla\varphi_0(x) - \nabla\varphi(x)\rangle\Big] d\mu .$$

$x = \nabla\varphi^*(\nabla\varphi(x))$

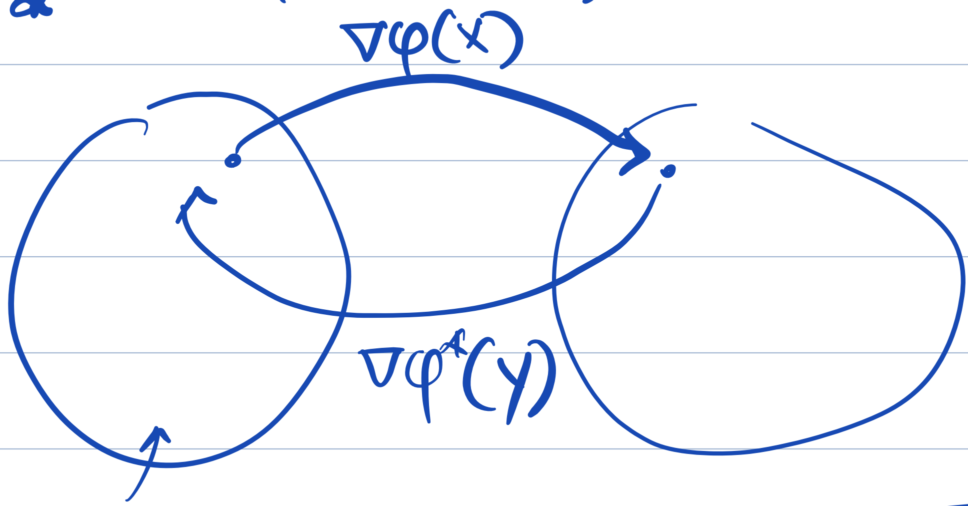
$$= \int \Big[\varphi^*(\nabla\varphi_0(x)) - \varphi^*(\nabla\varphi(x)) - \langle \nabla\varphi^*(\nabla\varphi(x)), \nabla\varphi_0(x) - \nabla\varphi(x)\rangle\Big] d\mu .$$

$$= \int d_{\varphi^*}(\nabla\varphi_0, \nabla\varphi)\, d\mu .$$



Strong-convexity & smoothness:

$$\|\nabla\varphi_0 - \nabla\varphi\|_2^2 \lesssim d_{\varphi^*}(\nabla\varphi_0, \nabla\varphi) \lesssim \|\nabla\varphi_0 - \nabla\varphi\|_2^2$$



Primal Stability (plugin estimators)

$$\hat{T} = OT(\hat{\mu}, \hat{\nu}).$$

$$T_0 = OT(\mu, \nu)$$

→ when is $\hat{T} \approx T_0$.

→ maybe $\hat{\mu} \approx \mu$, $\hat{\nu} \approx \nu$

in what distance?

Lemma: μ, ν, $\nu = (\nabla \varphi_0)_\# \mu$.

φ_0 is smooth & strongly convex.

then for any $\underline{\hat{\mu}, \hat{\nu}}$, we can construct $\hat{T}$ such that

$$\left[\int \| \underline{\hat{T}(x)} - \underbrace{T_0(x)} \|^2 d\underline{\hat{\mu}}(x) \lesssim W_2^2(\hat{\mu}, \mu) + W_2^2(\hat{\nu}, \nu) \right]$$

OT coupling: $\hat{\gamma}$, Barycentric projection of $\hat{\gamma}$.

$$\hat{T}(x) = \mathbb{E}_{(x,y)\sim \underline{\hat{\gamma}}} [Y \mid X = x] \leftarrow$$

Wasserstein density estimation

$W_2^2(\hat{\mu}, \mu)$ / $W_1(\hat{\mu}, \mu) \leftarrow$

φ smooth and strongly convex

$$d_\varphi(x,y) = \underbrace{\varphi(x) - \varphi(y) - \nabla\varphi(y)^T(x-y)}$$

$\frac{1}{\lambda}\|x-y\|^2 \leq \quad\quad \leq \lambda\|x-y\|^2$

$\gamma(X, Y) \quad\quad \underbrace{\gamma(Y \mid X = x)}$.