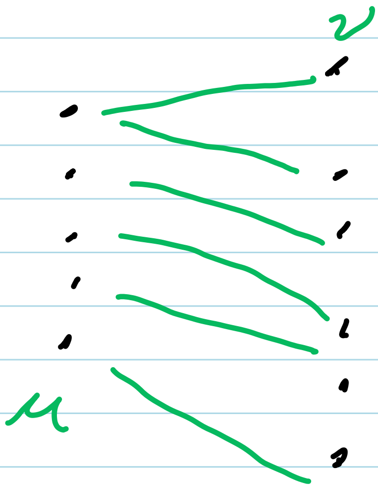
Wasserstein-2 Monge Problem

$$W_2^2(\mu,\nu) = \inf_{T \,:\, T_\# \mu = \nu} \int \|x - T(x)\|^2 \, d\mu$$

$\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$

Problem:



$\exists!\, T$

Kantorovich Problem

$$W_2^2(\mu,\nu) = \inf_{\gamma \in \Pi_\mu^\nu} \int \|x-y\|^2 \, d\gamma(x,y)$$

Definition of Π_μ^ν :

$\gamma \in \Pi_\mu^\nu$ if

(*) $\gamma \in P(X \times Y)$

$\gamma(A \times \mathbb{R}^d) = \mu(A)$

$\gamma(\mathbb{R}^d \times B) = \nu(B)$

Dual Problem

$$\sup_{f,g \in L^1(\mathbb{R}^d)} \int f(x)\, d\mu + \int g(y)\, d\nu$$

$$f(x) + g(y) \leq \|x-y\|^2$$

$$(*) \quad \sup_{f,g \in C_b} \int f\, d\mu + \int g\, d\nu - \int f(x) + g(y)\, d\gamma(x,y)$$

$$= \begin{cases} 0 & \text{if } \gamma \in \Pi_\mu^\nu \\ \infty & \text{else} \end{cases}$$

$$(KP) \quad \inf_{\gamma \in M_+} \left\{ \int \|x-y\|^2\, d\gamma + (*) \right\}$$

$$(KP) \geq \sup_{f,g} \left\{ \int f \, d\mu + \int g \, d\nu + \underbrace{\inf_{\gamma} \left\{ \int \|x-y\|^2 - f(x) - g(y) \, d\gamma \right\}}_{} \right\}$$

$$(**) = \begin{cases} 0 & \text{if } f(x) + g(y) \leq \|x-y\|^2 \\ -\infty & \text{else} \end{cases}$$

$$\Rightarrow \quad DKP$$

Brenier Thm (1992)

$\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, & μ has density.

$X \sim \mu$. The following are equiv.

i) $\bar{\gamma}$ is opt coupling for KP

ii) $\exists$ convex fctn φ s.t.

$$(X, \nabla \varphi(X)) \sim \bar{\gamma}$$

ie $T(x) = \nabla \varphi(x)$, then $T_{\#}\mu = \nu$.

iii) $KP = DKP$ & $\bar{f}(x) = \|x\|^2 - 2\varphi(x)$

$\bar{g}(x) = \|x\|^2 - 2\varphi^*(x)$

Cyclic Monotonicity

On $\mathbb{R}$: φ convex $\Rightarrow$ $(x-y)(\varphi'(x) - \varphi'(y)) \geq 0$

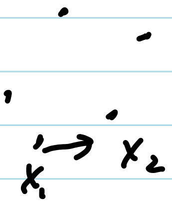
On $\mathbb{R}^d$: " $\Rightarrow$ $\langle x-y, \tilde{\nabla}\varphi(x) - \tilde{\nabla}\varphi(y) \rangle \geq 0$

$\tilde{\nabla}\varphi \in \partial\varphi$

For $\{x_i\}_{i=1}^{K} \subseteq \mathbb{R}^d$, φ convex $\Rightarrow$

$$\sum_{i=1}^{K} \langle x_i - x_{i+1}, \nabla\varphi(x_i) \rangle \geq 0$$

w/ $x_{K+1} = x_1$



Def: (x_i, y_i) is cyc. monotone if

$$\sum_{i=1}^{K} \langle x_i - x_{i+1}, y_i \rangle \geq 0$$

Rockafellar Thm:

A is cyc monotone iff $\exists \varphi$ convex s.t

$$A \in \partial \varphi$$

$$\shortparallel$$

$$\{(x, g) : x \in \mathbb{R}^d, g \in \partial f(x)\}$$

Opt $\bar{\gamma} \in \Pi_\mu^\nu \Rightarrow \text{supp}(\bar{\gamma})$ is C.M.

Prop: Let $\bar{\gamma} \in \Pi_\mu^\nu$ be opt coupling

Then $\text{supp}(\bar{\gamma})$ is cyc. monotone.

Intuition: CM $\Rightarrow$

$$\sum_{i=1}^{K} \|x_i - y_i\|^2 \le \sum_{i=1}^{K} \|x_{i+1} - y_i\|^2$$

Discrete OT: σ is opt perm.

$$\sum_{i=1}^{K} \|a_i - b_{\sigma(i)}\|^2 \le \sum_{i=1}^{K} \|a_i - b_{\tau(i)}\|^2$$

$i \Rightarrow ii$

$\bar{\gamma}$ be opt coupling

$\Rightarrow$ supp $(\bar{\gamma})$ is cyc monotone

$\Rightarrow$ $\exists$ convex φ s.t.

$\bar{\gamma}\big(Y \in \partial \varphi(X)\big) = 1$

$\Rightarrow$ If assume domain is bdd & compact

then φ is diff a.e.

$\Rightarrow$ $\bar{\gamma}\big(Y = \nabla \varphi(X)\big) = 1 \Rightarrow \big(X, \nabla \varphi(X)\big) \sim \bar{\gamma}$

ii $\Rightarrow$ iii

$$\|x-\nabla\varphi(x)\|^2 = \|x\|^2 + \|\nabla\varphi(x)\|^2 - 2\langle x, \nabla\varphi(x)\rangle$$

Convex conj of φ: $\varphi^*(y) = \sup_x \{\langle x, y\rangle - \varphi(x)\}$

φ^* satisfies $\varphi(x) + \varphi^*(\nabla\varphi(x)) = \langle \nabla\varphi(x), x\rangle$

$$\|x-\nabla\varphi\|^2 = \overbrace{\|x\|^2 - 2\varphi(x)}^{\bar{f}} + \overbrace{\|\nabla\varphi(x)\|^2 - \varphi^*(\nabla\varphi(x))}^{\bar{g}}$$

$$\int \|x-y\|^2 d\gamma = \int \bar{f}(x)\, d\mu + \underbrace{\int \bar{g}(\nabla\varphi(x))\, d\mu}_{\int \bar{g}(y)\, d\nu}$$

(continued)

$(\bar{f}, \bar{g})$ satisfies const.

$$\bar{f}(x) + \bar{g}(y) = \|x\|^2 + \|y\|^2 - 2(\varphi(x) + \varphi^*(y))$$

$$\uparrow$$

$$\langle x, y \rangle$$

$$= \|x - y\|^2$$

iii $\Rightarrow$ i

$$\int \|x-y\|^2 \, d\bar{\gamma} = \int \bar{f}(x) \, d\mu + \int \bar{g} \, d\nu$$

$$= \int \left(\bar{f}(x) + \bar{g}(y)\right) d\gamma$$

any coupling

$$\leq \int \|x-y\|^2 \, d\gamma$$

$\bar{\gamma}$ is opt

$\square$

Uniqueness

Thm: Same asspt.

$(x, \nabla \varphi(x)) \sim \bar{\gamma}$ is opt.

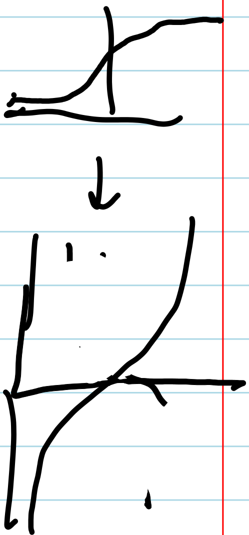
$\nabla \varphi$ is unique in sense that

$\exists$ convex ψ s.t $\nabla \psi_{\#} \mu = \nu$

then $\nabla \varphi = \nabla \psi$ a.e.

Application I : 1D OT

CDF



Thm: $\mu, \nu \in P(\mathbb{R})$, μ has density

$$W_2^2(\mu, \nu) = \int_0^1 |F_\mu^\dagger(u) - F_\nu^\dagger(u)|^2 \, du$$

where F_μ is CDF of μ

$F_\mu^\dagger$ is inverse CDF

Pf: μ density $\Rightarrow$ $F_\mu \circ F_\mu^\dagger = Id$

$U \sim Unif([0,1])$, $X \sim F_\mu^\dagger(U)$ $\quad Y \sim F_\nu^\dagger(U)$

(continued)

$X \sim \mu \quad , \quad Y \sim \nu$

$Y = F_\nu^+ \circ F_\mu(X)$ & $F_\nu^+ \circ F_\mu$ (*) is monotone inc.

$\Rightarrow$ (*) is grad of convex fctn

$\Rightarrow \Big(X, F_\nu^+ \circ F_\mu(X)\Big) \sim \gamma$ opt $\square$

Application II: OT for Gaussians

$$\mu = \mathcal{N}(m_1, \Sigma_1)$$

$$\nu = \mathcal{N}(m_2, \Sigma_2)$$

$$T(x) = Ax + b$$

$$T^*(x) = \Sigma_1^{-1/2}\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\Sigma_1^{-1/2}(x - m_1) + m_2$$

$$W_2^2(\mu, \nu) = \|m_1 - m_2\|^2 + \operatorname{tr}\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\right)$$

(continued)

Lower bound $W_2(\nu_1, \nu_2)$:

Application III: ICNNs

$\exists$ NN w/ skip connection s.t.

$f_\theta : \mathbb{R}^d \to \mathbb{R}$ is convex

$\mu = \mathcal{N}(0, I)$, $\nu = p_{data}$

Train NN to minimize

$W_2(\nu, (\nabla f_\theta)_\# \mu)$

then ∇f_θ is approximate OT map