Generalizing beyond TV distance

TV perturbations + bounded covariance + mean estimation

TV perturbations → any Wasserstein metric

bounded covariance → resilience

mean estimation → any loss (regression)

Plan for today

- define Wasserstein ①
- generalize resilience to Wasserstein
    - deletion → "friendly" perturbations
    - prove analog of midpoint lemma
- prove modulus of continuity bound
- concrete: 2^{nd} moment estimation under W_1

① Wasserstein distance

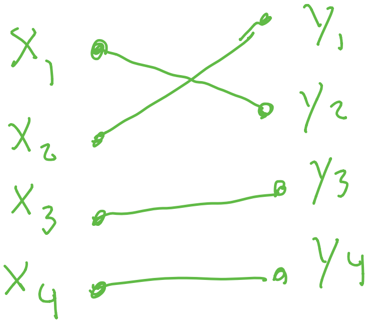
Definition (Coupling).

Given p on $\mathcal{X}$, q on $\mathcal{Y}$, a coupling $\pi_{XY} \in \Pi(p,q)$ ↙ space of all couplings is a distribution on $\mathcal{X} \times \mathcal{Y}$ s.t. $\pi_X = p$ and $\pi_Y = q$.

Intuition: matching

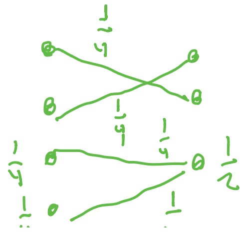
$X = \{x_1, x_2, x_3, x_4\}$ $\quad$ $Y = \{y_1, y_2, y_3, y_4\}$

p, q : uniform



$$\begin{array}{c|cccc|c} & y_1 & y_2 & y_3 & y_4 & \\ \hline x_1 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ x_2 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} \\ x_3 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ x_4 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ \hline & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \end{array}$$

$X = \{x_1, x_2, x_3, x_4\}$ with $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$ $\quad$ $Y = \{y_1, y_2, y_3\}$ with $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$



$$\begin{bmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$$

4 4

Definition (W$_c$ distance)

Let $c(x,y)$ be non-negative cost.

For $\pi \in \Pi(p,q)$, let cost of π be $\mathbb{E}_{(x,y)\sim\pi}[c(x,y)]$

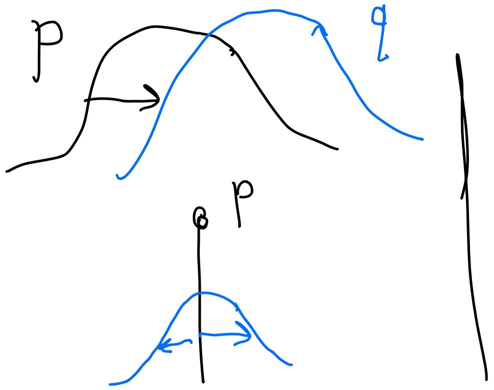
$$W_c(p,q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{\pi}[c(x,y)].$$

Examples. $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$

$c(x,y) = \mathbb{I}(x \neq y) \Rightarrow$ TV distance

$c(x,y) = \|x-y\|_2 \Rightarrow$ "earthmover" distance

W_1 distance



Primal: $\inf_{\pi \in \Pi(p,q)} \mathbb{E}_{\pi}[c(x,y)]$

Assume c is a metric.

Dual: f is Lipschitz w.r.t. c is $|f(x)-f(y)| \leq c(x,y)$

$\mathcal{L}(\mathcal{X}, c)$: all Lipschitz functions w.r.t. c

$\sup_{f \in \mathcal{L}(\mathcal{X},c)} \mathbb{E}_p[f(x)] - \mathbb{E}_q[f(x)]$

$\pi \in \Pi(p,q)$

Thm (Kantesavich- Rubenstein theorem) Primal = Dual.

Primal ≥ Dual

any function f lowerbounds cost of any coupling π

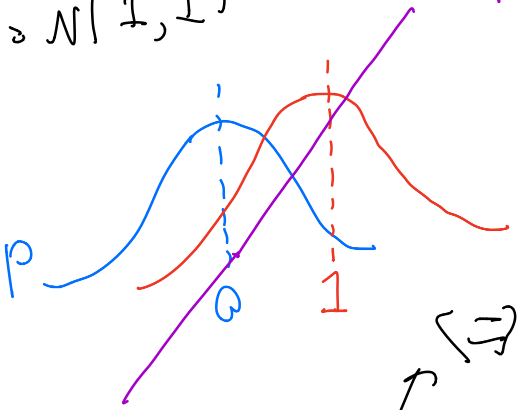
$$\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{y\sim q}[f(y)]$$

$$= \mathbb{E}_{(x,y)\sim\pi_{XY}}[f(x) - f(y)] \le \mathbb{E}_{(x,y)\sim\pi_{XY}}[c(x,y)]$$

$p = N(0,1)$

$q = N(1,1)$

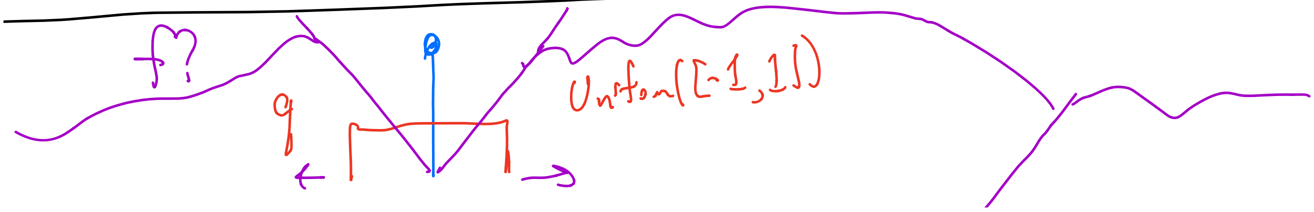
f? $f(x) = x$

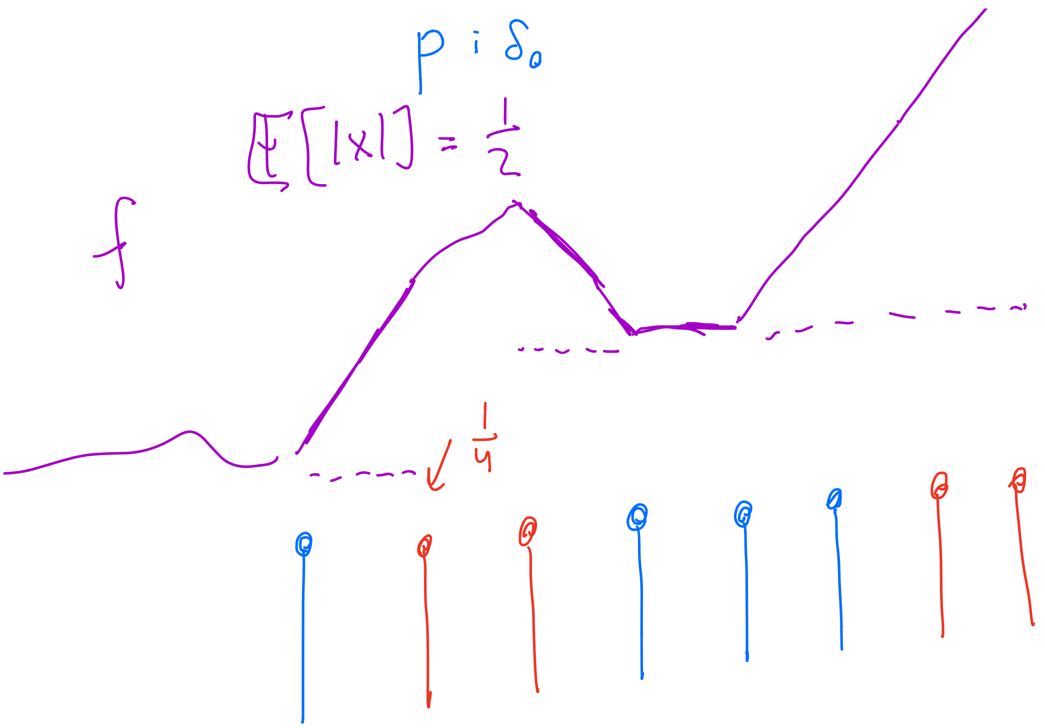


Complementary slackness:
Under optimal π and f,
$f(x) - f(y) = c(x,y)$
almost surely.

$\Leftarrow$ If x matched to y under coupling, then $f(x) - f(y) = c(x,y)$.

$f(x+1) = f(x) + 1 \quad \forall x$





$\widetilde{TV}$

$$TV = \sup_{E} |p(E) - q(E)|$$

$$= \sup_{\substack{f \text{ w/} \\ a \le f(x) \le a+1}} |\mathbb{E}_p[f] - \mathbb{E}_q[f]|$$

$$f(x) = \begin{cases} 1; & x \in E \\ 0; & x \notin E \end{cases}$$

exactly w.r.t. Lipschitz $\mathbb{1}[x \neq y]$

instead of all f, only take $\tilde{\ }$

$f = \mathbb{I}[\langle x, v \rangle \leq c]$

$c(x,y) = \mathbb{I}[x \neq y]$; TV

$c(x,y) = \|x-y\|_2$; W_1

$c(x,y)$; $\|x-y\|_2^\alpha$: W_α $(\alpha \in (0,1])$

↳ interpolating b/t TV and W_1

As $\alpha \to 0$, $W_\alpha \to TV$

$c(x,y) = \|x-y\|_0$

Note, All above c are metrics.

Proposition, If c is a metric, then W_c is also a metric.

Pf. $\underset{\pi_1}{W(p,q)} + \underset{\pi_2}{W(q,r)} \geq \underset{\pi_3}{W(p,r)}$.

$$\pi_3(p,r) = \int \pi_1(p,q)\,\pi_2(q,r)\,dq.$$

W_α : $\alpha \geq 1$

$$W_2 : \left(\inf_{\pi} \mathbb{E}_{\pi}\left[\|x-y\|_2^{\overset{\alpha}{\downarrow}2}\right]\right)^{1/2 \;\leftarrow\; \frac{1}{\alpha}}$$

Resilience for W_c

TV picture:

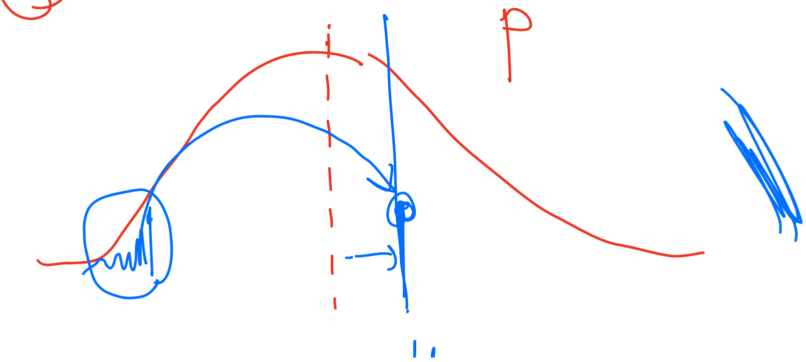
$$\begin{tikzcd}
p_1 \arrow[rr, dash, "TV \leq \varepsilon"] \arrow[dr, "\varepsilon\text{-deletion}"'] & & p_2 \arrow[dl, "\varepsilon\text{-deletion}"] \\
& r &
\end{tikzcd}$$

"midpoint"

Need to generalize this picture

---

① Generalize deletion



μ_p μ_r

Claim: μ_r attainable via deletion

$\iff$ attainable by moving mass <u>towards</u> μ_r.

Definition (Friendly perturbations)

Given dist^n p on $\mathcal{X}$, $f : \mathcal{X} \to \mathbb{R}$, cost $c(x,y)$, say r is an ε-friendly perturbation of p (w.r.t. f) if $\exists$ coupling $\pi \in \Pi(p,r)$ s.t.

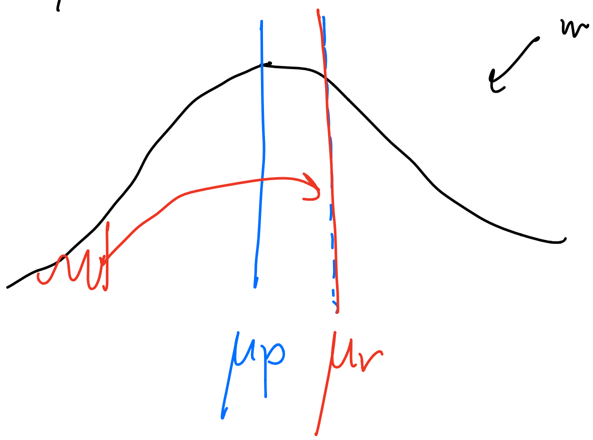
(low cost) ① $\mathbb{E}_{\pi}[c(x,y)] \le \varepsilon$

(moves towards mean) ② $\underbrace{f(y)}_{\text{end } (r)}$ is between $\underbrace{f(x)}_{\text{start } (p)}$ and $\underbrace{\mathbb{E}_{z \sim r}[f(z)]}_{\text{mean}}$ <u>almost surely</u>.

$c(x,y) = \mathbb{I}[x \neq y]$

ε-friendly ⟺ ε-deletions

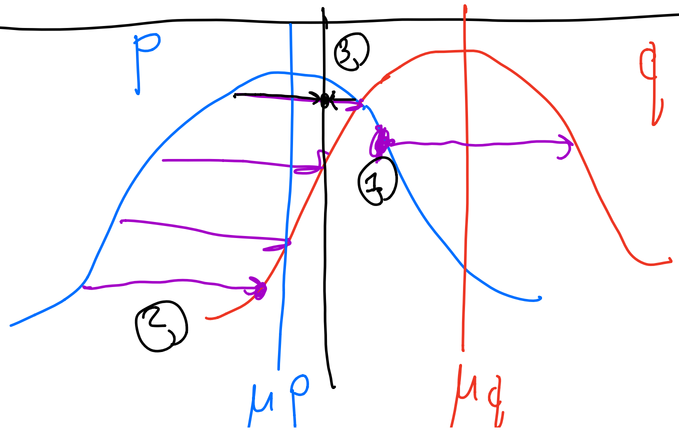
$c(x,y) = \|x-y\|_2$



when $W_c \neq TV$, might not move all the way to the mean.

Goal: Prove midpoint lemma.

If $W_c(p,q) \leq \varepsilon$, then for any f there is an r that is ε-friendly perturbation of both p and q.



μ_r

Given "guess" μ, can define r as follows:

$x \longrightarrow y$ under coupling π

$f(x) \quad \mu \quad f(y)$

① $\mu < f(x)$: stay at x for r

② $\mu > f(y)$: go to y for r

③ $\mu \in [f(x), f(y)]$: go to μ

$\Rightarrow$ can check: distance $\leq \varepsilon$

Intermediate Value Property.

For any $x, y \in \mathcal{X}$ and μ s.t. $f(x) < \mu < f(y)$, $\exists$ z s.t. $f(z) = \mu$ and $\max(c(x,z), c(z,y)) \leq c(x,y)$.

Holds for TV, W_1 $\rightarrow$

Midpoint Lemma. If IVP holds, and $W_c(p,q) \leq \varepsilon$, then $\exists$ r that is ε-friendly for both p and q:

Pf. Given μ, [illegible] $: \mu \leq \min(\circ)$

random variable $\rightarrow$ $S_{xy}(\mu) = \begin{cases} \min(f(x), f(y)) & \\ \max(f(x), f(y)) : & \mu \geq \max(\cdot) \\ \mu & ; \text{ else} \end{cases}$

$z_{xy}(\mu)$: z s.t. $f(z) = S_{xy}(\mu)$

π_{xy} : $\mathbb{E}_{\pi}[c(x,y)] \leq \varepsilon$

marginal: r

$(x, y) \mapsto (x, z_{xy}(\mu))$ $\leftarrow$ ε-friendly w.r.t. μ

$(y, z_{xy}(\mu))$ $\leftarrow$ ε-friendly w.r.t. μ

Claim $\mu = \underbrace{\mathbb{E}_{r(\mu)}[f(x)]}_{F(\mu)}$ for some μ.

$\mu = F(\mu)$ for some μ

$\rightarrow$ F is monotonically increasing and bounded