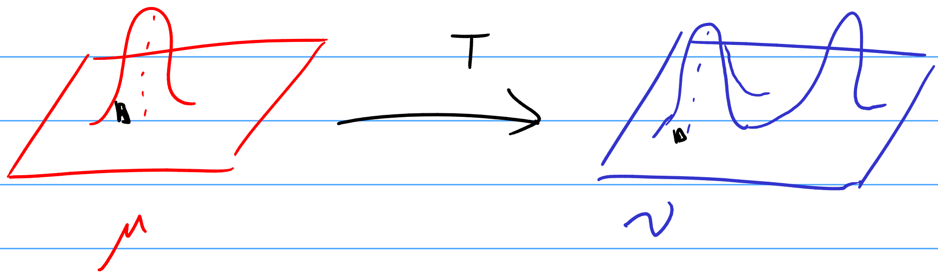
A (rough) introduction to OT

- $\Omega \subset \mathbb{R}^d$, compact
- $\mu, \nu \in P(\Omega)$ are prob. measures (usually w/ 2+ moments)



T is a map that turns one pile of sand (μ) into another (ν).

$T : \Omega \to \Omega$, $T_\# \mu = \nu$; $\forall A \subseteq \Omega$ msble, $\nu(A) = \mu(T^{-1}(A))$

pushforward

$$\min_T \int c(x, T(x))\, d\mu(x) \quad \text{s.t.} \quad T_\# \mu = \nu \quad (M)$$

w/ $c(x,y) = h(x-y)$ (e.g. $h(\cdot) = \|\cdot\|^2$)

- Hard to solve numerically
- Existence of maps is not guaranteed
- Modulo details on cost (strictly cvx), if $\mu, \nu \ll$ Leb, then T_* exists (as does T_*^{-1})

"optimal plans"

$$\pi \in \Gamma(\mu,\nu) \iff \int \pi(x,y)\,dy = \mu(x), \quad \int \pi(x,y)\,dx = \nu(y)$$

$$\inf_{\pi} \iint c(x,y)\, d\pi(x,y) \quad \text{s.t. } \pi \in \Pi(\mu,\nu) \qquad (K)$$

$\to$ much more reasonable problem & significantly more tractable.

$\hookrightarrow$ now, solving for a "distribution"

Discrete form of (K)

$\mu, \nu \in \mathbb{R}^n$, w/ $\sum_{i=1}^{n} \mu_i = 1$

(K) $\quad \min \ \langle \pi, C \rangle \qquad \pi^T \mathbb{1}_n = \mu_n, \ \ \pi \mathbb{1}_n^T = \nu_n$

is a linear program

$\pi \in \mathbb{R}^{n\times n}$

C: cost matrix $c(x,y) = \|x-y\|_2^2 \in \mathbb{R}^{n\times n}$

$$\Rightarrow \min_{\pi} \sum_{i,j} \pi_{ij} C_{ij} \quad \text{s.t. } \pi^T \mathbb{1}_n = \mu_n, \ \ \pi \mathbb{1}_n^T = \nu_n$$

$\hookrightarrow$ $\tilde{O}(n^3)$ time to solve, where $\tilde{O}$ hides polylog factors of n.

Duality in OT

Recall

$$OT(\mu,\nu) = \min_{T} \int c(x, T(x))\, d\mu(x) \quad \text{s.t. } T_{\#}\mu = \nu \qquad (M)$$

$$T_\# \mu = \nu \iff \forall \varphi \in C(\Omega), \quad \int \varphi(y)\, d\nu(y) = \int \varphi(T(x))\, d\mu(x)$$

$$OT(\mu,\nu) = \min_T \sup_\varphi \int c(x, T(x))\, d\mu(x) + \int \varphi\, d\nu - \int \varphi(T(x))\, d\mu(x)$$

$$= \min_T \sup_\varphi \int c(x,T(x)) - \varphi(T(x))\, d\mu(x) + \int \varphi\, d\nu$$

$\mu \ll Leb$

$$\text{"="} \sup_\varphi \min_T \int c(x,T(x)) - \varphi(T(x))\, d\mu(x) + \int \varphi\, d\nu$$

$$= \sup_\varphi \int \underbrace{\min_T c(x,T(x)) - \varphi(T(x))}_{=:\ \varphi^c}\, d\mu + \int \varphi\, d\nu$$

$$= \sup_\varphi \int \varphi^c\, d\mu + \int \varphi\, d\nu$$

↳ "Kantorovich potentials"

-> Possible papers / topics

- Brenier's Theorem w/ applications
- "Fast OT w/ back & forth method" ~ $O(n \log n)$

↓

What's the deal w/ Entropy?

For $\mu, \nu \in P(\Omega)$,

$$\underset{\text{Entropy}}{H(\mu|\nu)} = \begin{cases} \int \log\left(\frac{d\mu}{d\nu}\right) d\mu & \mu \ll \nu \\ +\infty & \text{o.w.} \end{cases}$$

$$OT(\mu,\nu) = \min_{\pi \in \Gamma(\mu,\nu)} \int c \cdot d\pi \qquad (K) \quad \rightsquigarrow \tilde{O}(n^3)$$

$$OT_\varepsilon(\mu,\nu) = \min_{\pi \in \Gamma(\mu,\nu)} \int c\, d\pi + \varepsilon H(\pi | \mu \otimes \nu)$$

↳ 1-strongly cvx, as $\varepsilon \to 0$, we recover π^*

- 2013, Cuturi published "Sinkhorn" $\rightsquigarrow \tilde{O}(n^2)$
- Since then, this has been sped up to $\tilde{O}(n)$ w/ some extra assumption