Study Group Lect: O.T.

Total Variation Dist

$(\Omega, \mathcal{F})$ $\quad p, q \quad P, Q$

$$TV(p,q) := \underbrace{\sup_{A\in\mathcal{F}} P(A) - Q(A)} \leftarrow \ell_\infty \text{ notion}$$

$$= \frac{1}{2} \sum_{x\in\Omega} |p(x) - q(x)| \leftarrow \ell_1 \text{ notion}$$

$$\sup_{A} P(A) - q(A) = \sup_{A} \int \underbrace{1_A}_{f} dP - \int \underbrace{1_A}_{f} dQ$$

This motivates us:

$$d(p,q) = \sup_{f\in \underline{\underline{D}}} \left| \int f\,dP - \int f\,dQ \right|$$

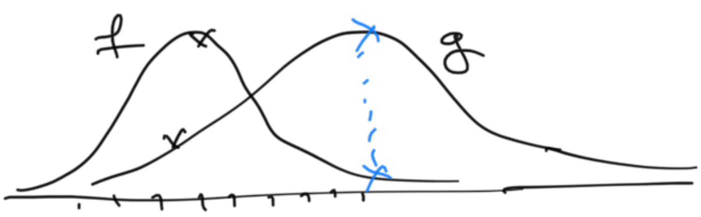
class of Borel sets

$$D = \{ 1_A : A \in \mathcal{B} \} \to TV$$

$$D = \{ 1_{(-\infty, x]} : x \in \mathbb{R} \} \to \text{Kolmogorov distance}$$

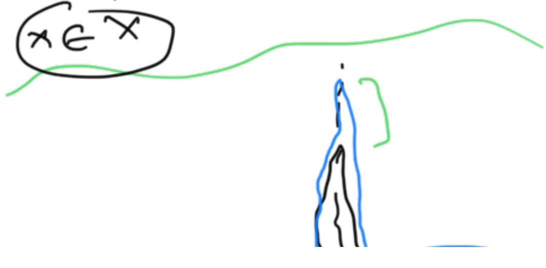
$$\text{Kolm}(p,q) \leq TV(p,q)$$

Take a step back:

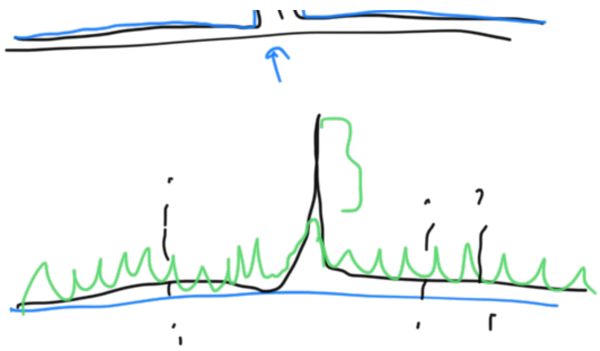


Trial #1:

$$\sup_{x\in\mathcal{X}} |f(x) - g(x)|$$



Trial #2

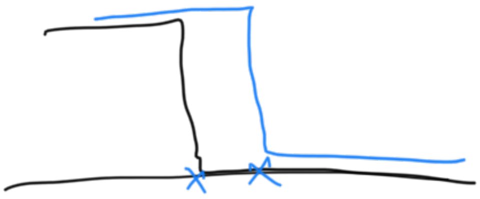


$$\underbrace{l_1(f,g) = \sum_{x \in X} |f(x) - g(x)|}_{\text{pointwise close}}$$

↪ think of funs topologically

Trial #3

shift +ε



+ε

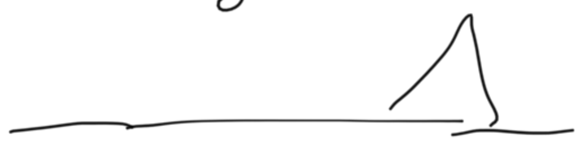
How to capture this instance?

⇓

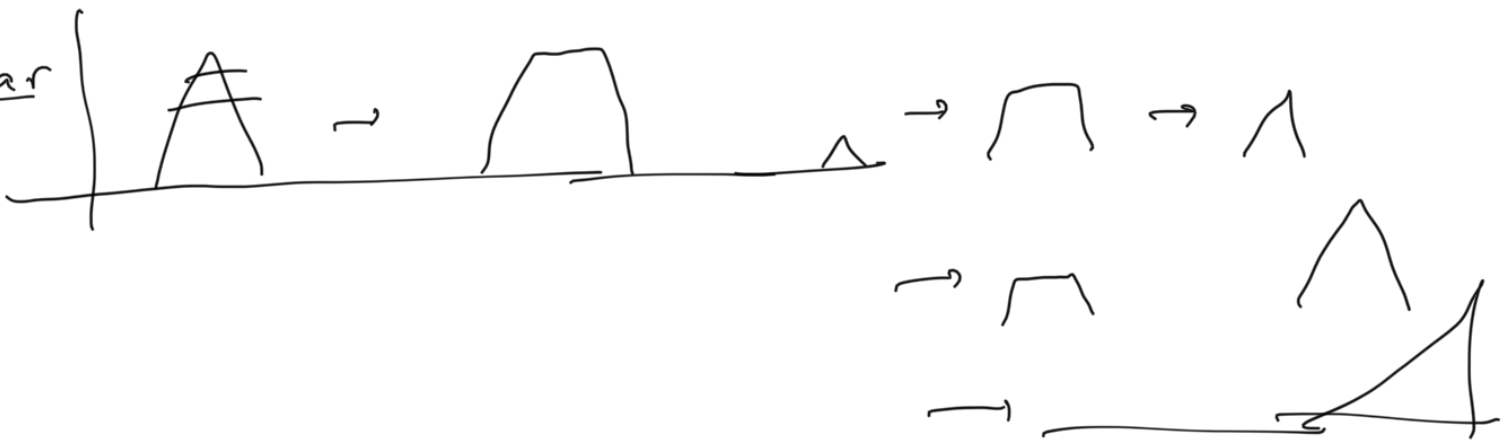
Mass Transportation

f, g → $\boxed{\text{Interpolation}}$

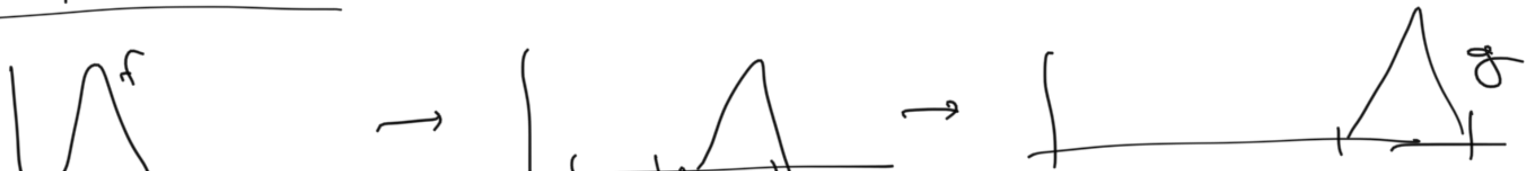
Initial | Final
f | g



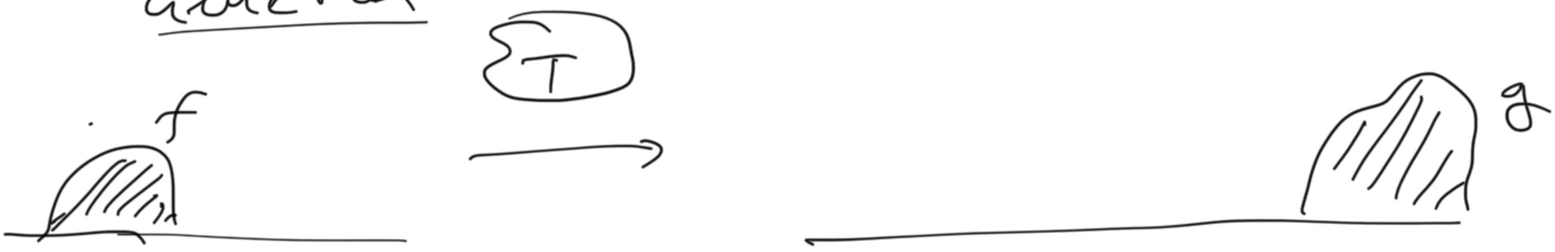
Linear



Displacement



General

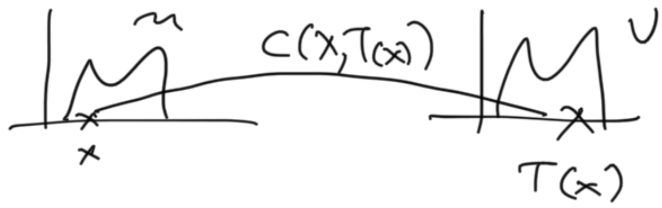


(+) mass preserved (later).

Optimal Transport

It is possible to define a geometry of a space of measures μ, ν & define distances between μ & ν.

Monge (1784) μ, ν on X



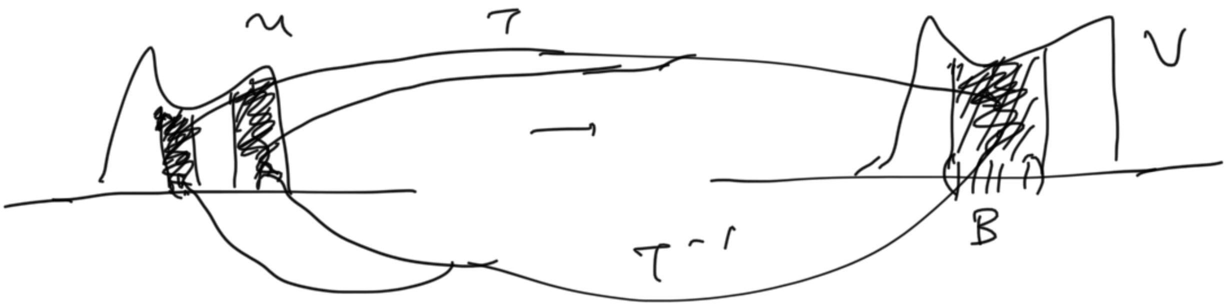
$$\inf_{T:X\to X} \int_X \underbrace{c(\overset{\text{initial}}{x}, \overset{\text{final}}{T(x)})}_{\text{cost}}\, \mu(x)\,dx$$

$T : x \mapsto T(x)$

μ is the "initial weight of the points"

s.t.

(I) $\forall\, B \subset X$: $\underbrace{\int_B \nu(x)\,dx}_{\text{final space}} = \underbrace{\int_{T^{-1}(B)} \mu(x)\,dx}_{\text{initial space}}$



mass preservation : global : $\int_X \mu = \int_X \nu$

(I) is $\boxed{\text{local}}$: it tells you that T cannot "create" "destroy" mass or desire

The feasible sols are called transportation maps

Obj : Monge for measures

$$\mu, \nu \qquad \mu(X) = \nu(Y)$$
$$X \quad Y$$

$$\inf_{T : X \to Y} \int_X c(x, T(x))\, d\mu$$

s.t. $T_{\#}\mu = \nu$ → push forward measure of μ to ν

$\to B \in \mathcal{F}$: $T_{\#}\mu(B) = \mu\left(T^{-1}(B)\right)$

The condition tells you that:

$$\mu\left(T^{-1}(B)\right) = \nu(B) \quad \forall B \in \mathcal{F}_Y$$

$$(M) \quad \inf \int \|x - T(x)\|^p \, d\mu$$

s.t.

$$T_{\#}\mu = \nu$$

$p \geq 1$

$\|\cdot\|_p$

$\|x+y\|_p \leq \|x\|_p + \|y\|_p$

$0 < p < 1$

A minimizer T^*, if one exists, is the optimal transport map

Q1: When it exists?

Q2: Is it unique?

Examples :

(#1) Discrete Monge

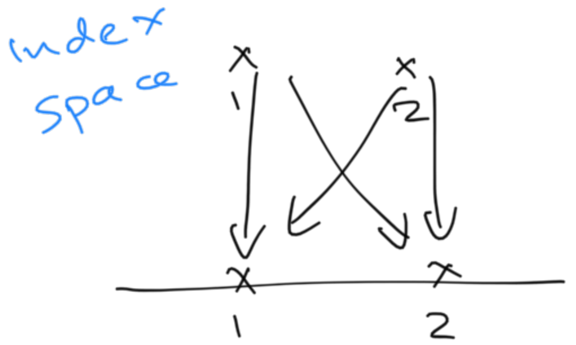
$$\mu = \frac{1}{n} \sum \delta_{x_i} \qquad \min_{\sigma \in S_n} \frac{1}{n} \sum c(x_i, y_{\sigma(i)})$$

$$\nu = \frac{1}{n} \sum \delta_{y_j}$$

$c(x,y)$ arbitrary

(online k-server)

(dist. of routing)



Each transport is a permutation.

#2) Optimality tester for $\underline{c(x,y) = |x-y|}$

Let's say that T is a trans. from μ to ν.

How do I know if T is opt?

Let $u : X \to \mathbb{R}$ $\boxed{1\text{-Lip}}$ on $c(x,y) = |x-y|$

$$\int_X \boxed{u(x)} \, (d\mu - d\nu) \overset{T \text{ is}}{\underset{\text{tr.}}{=}} \int_X \big(u(T(x)) - u(x)\big) \, d\mu$$

$$\overset{\text{Lip}}{\leq} \underbrace{\int_X \underbrace{|T(x) - x|}\, d\mu}, \quad u \text{ arbitrary}$$

If equality hold for some u

$\Rightarrow$ T must be opt

(#3) Minimizers are not unique

$$\mu = \mathbb{1}_{[0,n]}\, dx \qquad u(x) = x$$

$$\nu = \mathbb{1}_{[1,n+1]}\, dx$$

$$T_1(x) = x+1$$

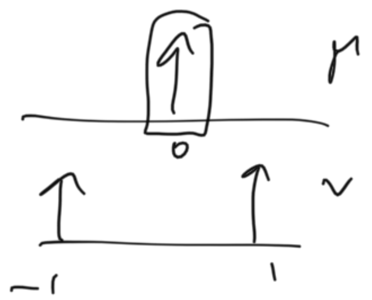
$$T_2(x) = \begin{cases} x+n & \text{on } [0,1) \\ x & \text{on } [1,n] \end{cases}$$

(#4) $T_{\#}\mu = \nu \to$ non linear

$\downarrow$

Sol may not exist

$X = [-1,1] \qquad \mu = \delta_0$



$$\nu = \underline{\underline{\tfrac{1}{2}}}\, \delta_{-1} + \underline{\underline{\tfrac{1}{2}}}\, \delta_1$$

Cannot Split mass in (M)

Kantorovich relaxation

deal with

$$\min_{\boxed{T}} \int_X c(x, T(x))\, d\mu \ : \ T_{\#}\mu = \nu \quad \text{is nonconvex}$$

Trick: Relax the nature of the transportation mass can be split.

Work in the product space (coupling appears)

$$\underbrace{X}\times\underbrace{Y}$$

Opt & look for a measure γ on the $\underbrace{X\times Y}$

$$(K)\quad \inf_{\gamma}\int_{X\times Y} c(x,y)\,d\gamma(x,y)\,,\quad \gamma\geq 0$$

s.t. "coupled space"

$$\gamma\in M(\mu,\nu)$$

$\pi_1(x,y)=x$

marginals $\begin{cases}\pi_{1\#}\gamma=\mu\\ \pi_{2\#}\gamma=\nu\end{cases}$

γ are the transp. <u>plans</u>

Discrete example

Any trans. plan from μ to ν is a DS matrix

(K):

$$\inf_{B\in BN(n)}\frac{1}{n}\sum_{i,j}B(i,j)\,c(x_i,y_j)$$

$\sigma\in S_n$ α^γ

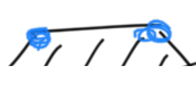
$$A=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\alpha^\gamma$$

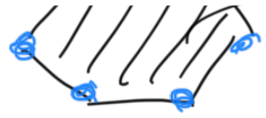
rook

Relaxed version of S_n

Doubly Stoch Matrices

Birkhoff–VN Poly

 space of



DS

BN

$$\boxed{\mathcal{E}(BN) = S_n}$$

Relation (M) – (K)

Prop: Every trans. map $T : X \to X$ determines a plan γ_T by:

$$\gamma_T := (Id \times T)_{\#}\, \mu$$

$$\underbrace{\min_{\gamma \in M(\mu,\nu)} \int_{X\times Y} c(x,y)\, d\gamma}_{} \leq \inf_{\substack{T: \\ T_{\#}\mu = \nu}} \int_X c(x, T(x))\, d\mu$$

p=2 Brenier

(X, d) $\quad \mu, \nu$ over X $\quad$ Kantorovich Wasserstein distance

$$W_d^{(p)}(\mu,\nu) = \left(\inf_{\gamma \in M(\mu,\nu)} \int_{X\times X} d(x,y)^p\, d\gamma(x,y) \right)^{\frac{1}{p}}$$

$p \geq 1$

$p=1$: Earth-mover distance.

Duality $\quad W_d(\mu,\nu) = \sup_{\substack{f: X \to \mathbb{R} \\ 1\text{-Lip}}} \left| \int f\, d\mu - \int f\, d\nu \right|$

$\|f(x) - f(y)\| \leq d(x,y)$

Obs #1 $\quad$ Existence of opt. plans $\quad$ open set

Obs. II: Existence

Assume c is lower semicontinuous in $X \times X$

Then there exists $\gamma \in \mathcal{M}(X \times X)$ solving (MK)

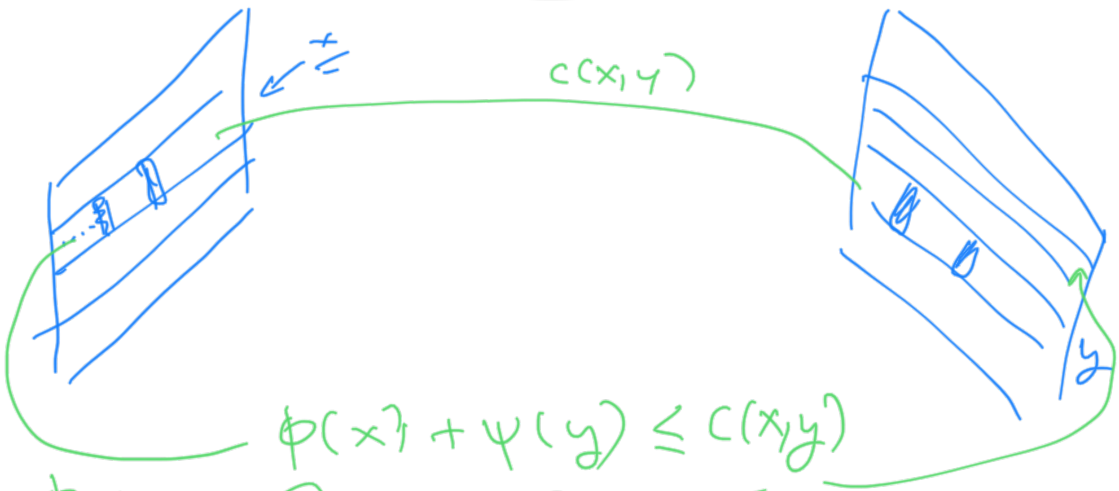
Kantorovich duality

Trick: whenever you can apply the min-max principle → duality

$$\underbrace{\inf_{\gamma} \int_{X \times X} c(x,y)\, d\gamma}_{} = \sup_{\Phi_c} \left[\int_X \phi(x)\, d\mu + \int_X \psi\, d\nu \right]$$

$$\Phi_c = \{ (\phi, \psi) \in L^1(d\mu) \times L^1(d\nu) : \underbrace{\phi(x) + \psi(y) \le c(x,y)}_{} \}$$

~~crazy~~ constraint → $\phi(x)+\psi(y) \le c(x,y)$



Company $c(x,y)$

ϕ, ψ may be negative prices

Lipschitzness.

Duality proof (Find it!)

(X, d) metric

Wass → TV

$d_{...}(\mu, \nu) = \sup \left| \int f d\mu - \int f d\nu \right|$

$$W_d(\mu,\nu) = \sup_{\|f\|_{Lip}\le 1} \left| \int f\,d\mu - \int f\,d\nu \right|$$

$|f(x) - f(x')| \le 1$

$d = \text{Hamming} = \mathbb{1}\{x \ne x'\}$

assume $f : X \to [0,1]$

$$W_{Ham}(\mu,\nu) = \sup_{f : X \to [0,1]} \left| \int f\,d\mu - \int f\,d\nu \right|$$

$$\overset{?}{=} \inf_M M[X \ne Y] = \|\mu - \nu\|_{TV}$$

$$W_d(\mu,\nu) = \inf_\gamma \mathbb{E}_\gamma[d(X,Y)]$$

$d = \mathbb{1}_{x \ne x'}$ $\qquad TV = \inf_{(X,Y)} \Pr[X \ne Y].$

Intuitive!

$$\inf_M \mathbb{E}_M[d(x,y)] = \inf_M \int_{X \times X} d(x,y)\,dM$$

$$= \sup_{\boxed{\|f\|_{Lip} \le 1}} \int f(d\mu - d\nu)$$

TV

$$\inf_{(X,Y)} \Pr[X \ne Y] = \sup_{\boxed{A \subseteq X}} \mu(A) - \nu(A)$$

Intuition

Perms

"" "" polytope " "... of Lip. funs

DS

"$\Sigma(\mathring{p}ol_y) = \oint A$"

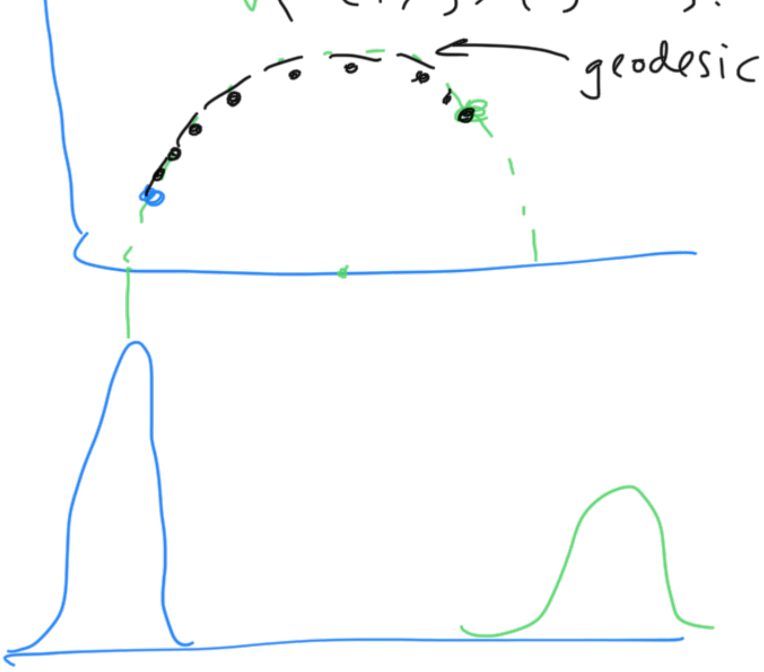
Divergence (KL)	Geometry

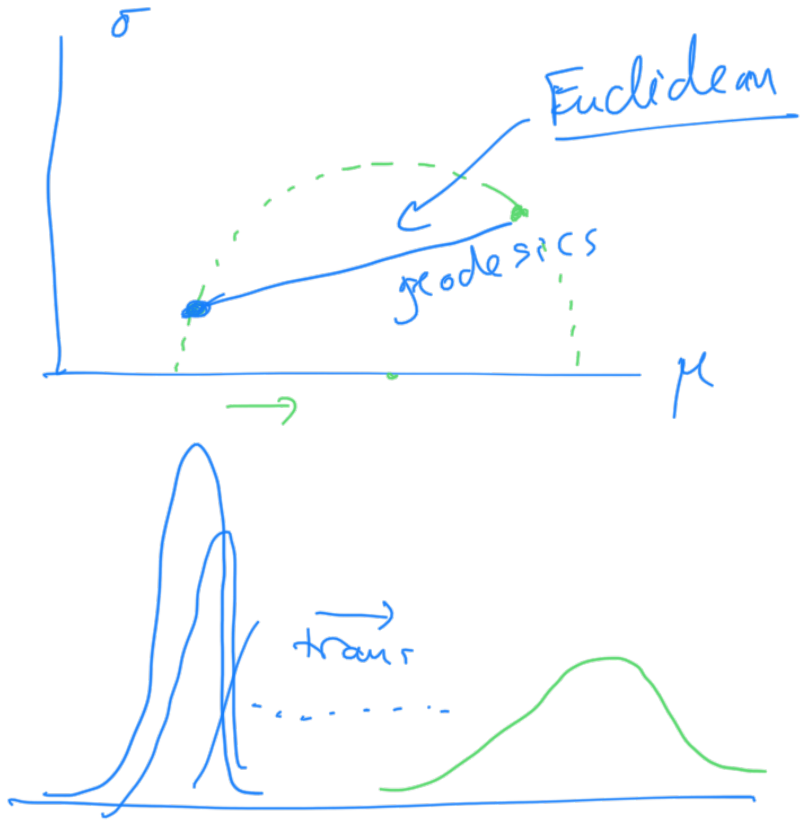
KL | Wass

Gaussians | Hyperbolic | Euclidean

Hyperbolic Poincaré half plane

$\{ (x,y) \mid y>0 \}$.

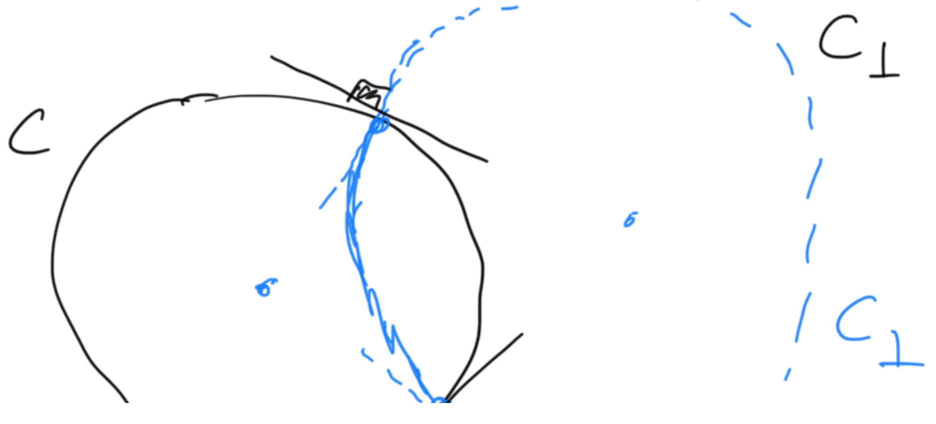




Intuition

"log-ratio" ⇒ Hyperbolicity.

Hyperbolic geometry



$C_\perp$ perp to C

Poincaré dist

model.