Lecture 15

Reminders
- Solutions for 2-3 exercises
- Revise article by March 7th
- Makeup Lecture

Recall:

goal: characterize topology of W_p

One last ingredient...

Def: Given X Polish, $\mu \in P(X)$,
$\operatorname{supp} \mu := \{x \in X : \mu(U) > 0 \ \forall \ U \ni x \text{ open}\}$

Fact: $\operatorname{supp} \mu$ is the smallest closed set C s.t. $\mu(X \setminus C) = 0$.

Thm: (c-monotonicity) Given $\mathcal{X}, \mathcal{Y}$ Polish
$c: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ cts, bdd below,
$\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$, $\mathbb{K}_c(\mu \otimes \nu) < +\infty$

Then, for any $\gamma \in \Gamma(\mu, \nu)$,
γ is optimal
$\Updownarrow$
for any $(x_1, y_1), \ldots, (x_n, y_n) \in \operatorname{supp} \gamma$, $\sigma \in S_n$
$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^n c(x_i, y_i).$$
"supp γ is c-cylically monotone"

Pf: Exercise 31 proof for $(\mathbb{R}^d, |\cdot|)$ and $c(x,y) = |x-y|^2$.

General case: AGS Prop 7.1.3.

As a consequence of this theorem...

<u>Cor</u>: Suppose X is a Polish space and $\mu_n \to \mu$, $\nu_n \to \nu$ narrowly. Then if $\gamma_n \in \Gamma(\mu_n, \nu_n)$ are optimal for W_p, $p \geq 1$, there exists a subseq γ_{n_k} s.t. $\gamma_{n_k} \to \gamma \in \Gamma(\mu, \nu)$ narrowly and γ is optimal for W_p.

Pf: Exercise 32.

Now, we can characterize the topology of W_p.

Thm: If $\mathcal{X}$ is a Polish space, for any $\mu_n, \mu \in \mathcal{P}_p(\mathcal{X})$,

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{array}{l} \mu_n \to \mu \text{ narrowly} \\ m_p(\mu_n) \to m_p(\mu) \end{array}$$

Fact: $d^p(x,y) \le (d(x,x_0) + d(x_0,y))^p$
$\le 2^{p-1}(d^p(x,x_0) + d^p(x_0,y))$

Now, we show "$\Leftarrow$".

Suppose $\mu_n \to \mu$ narrowly
$m_p(\mu_n) \to m_p(\mu)$

Last time: For all $\varepsilon > 0$, $\exists$ $R > 0$ s.t.

$$\int_{B_R^c(x_0)} d^p(x_0,y)\,d\mu < \varepsilon, \quad \limsup_{n\to\infty} \int_{B_R^c(x_0)} d^p(x_0,y)\,d\mu_n < \varepsilon$$

Let γ_n be OT plans from μ_n to μ.
It suffices to show

$$\limsup_{n\to\infty} \underbrace{\int d^p(x,y)\, d\gamma_n}_{=(*)} = 0.$$

Choose a subsequence γ_{n_k} s.t.

$$\lim_{k\to\infty} \int d^p(x,y)\, d\gamma_{n_k} = (*).$$

By Cor, there exists a further subseq, also denoted γ_{n_k} with the property $\gamma_{n_k} \to \gamma$ narrowly and γ is the OT map from μ to μ, so $0 = W_p^p(\mu,\mu) = \mathbb{K}_p(\gamma) \Rightarrow x=y$ γ a.e.

Thus, for all $R>0$,

$$\left.\begin{aligned}\lim_{k\to\infty} &\int d^p(x,y)\wedge R^p\, d\gamma_{n_k} \\ &= \int d^p(x,y)\wedge R^p\, d\gamma = 0\end{aligned}\right\} (**)$$

Likewise,

$$\limsup_{k\to\infty} \int_{d(x,y)\geq R} d^p(x,y) - R^p \, d\gamma_{n_k}(x,y)$$

$$\leq \int_{(B^c_{R/2}(x_0)\times \mathcal{X}) \cup (\mathcal{X}\times B^c_{R/2}(x_0))} 2\left(d^p(x,x_0) \vee d^p(x_0,y)\right) d\gamma_{n_k}(x,y)$$

$$\leq \underbrace{\int_{B^c_{R/2}(x_0)\times \mathcal{X}} 2\left(d^p(x,x_0)\vee d^p(x_0,y)\right) d\gamma_{n_k}(x,y)}_{\mathrm{I}}$$

$$+ \underbrace{\int_{\mathcal{X}\times B^c_{R/2}(x_0))} 2\left(d^p(x,x_0)\vee d^p(x_0,y)\right)}_{\mathrm{II}}$$

$$\frac{I}{2} = \int_{B^c_{R/2}(x_0)\times B^c_{R/2}(x_0)} \left(d^p(x,x_0) \vee d^p(x_0,y)\right) d\gamma_{n_k}(x,y)$$

$$+ \int_{B^c_{R/2}(x_0)\times B_{R/2}(x_0)} d^p(x,x_0)\, d\gamma_{n_k}(x,y)$$

$$\leq \int_{B^c_{R/2}(x_0)\times B^c_{R/2}(x_0)} d^p(x,x_0) + d^p(x_0,y)\, d\gamma_{n_k}(x,y)$$

$$+ \int_{B^c_{R/2}(x_0)\times X} d^p(x,x_0)\, d\gamma_{n_k}(x,y)$$

$$\leq \int_{B^c_{R/2}(x_0)} d^p(x,x_0)\, d\mu_n(x) + \int_{B^c_{R/2}(x_0)} d^p(x_0,y)\, d\mu(y)$$

$$+ \int_{B^c_{R/2}(x_0)} d^p(x,x_0)\, d\mu_n(x)$$

Arguing similarly for II,

for all $\varepsilon > 0$, $\exists R > 0$ s.t.

$$\limsup_{k\to\infty} \int_{d(x,y)\geq R} d^p(x,y) - R^p \, d\gamma_{n_k}(x,y) < \varepsilon.$$

Combining with $(**)$, $\forall \varepsilon > 0$

$$\lim_{k\to\infty} \int d^p(x,y) \, d\gamma_{n_k} < \varepsilon.$$

This completes the proof. $\square$

Ex: $\mu_n = c_n \delta_n + (1-c_n)\delta_0$

CLAIM: $\exists\, c_n \to 0$ (so that $\mu_n \to \delta_0$ narrowly) with $M_2(\mu_n) \not\to 0$, so $\mu_n \not\to \delta_0$ in W_2.

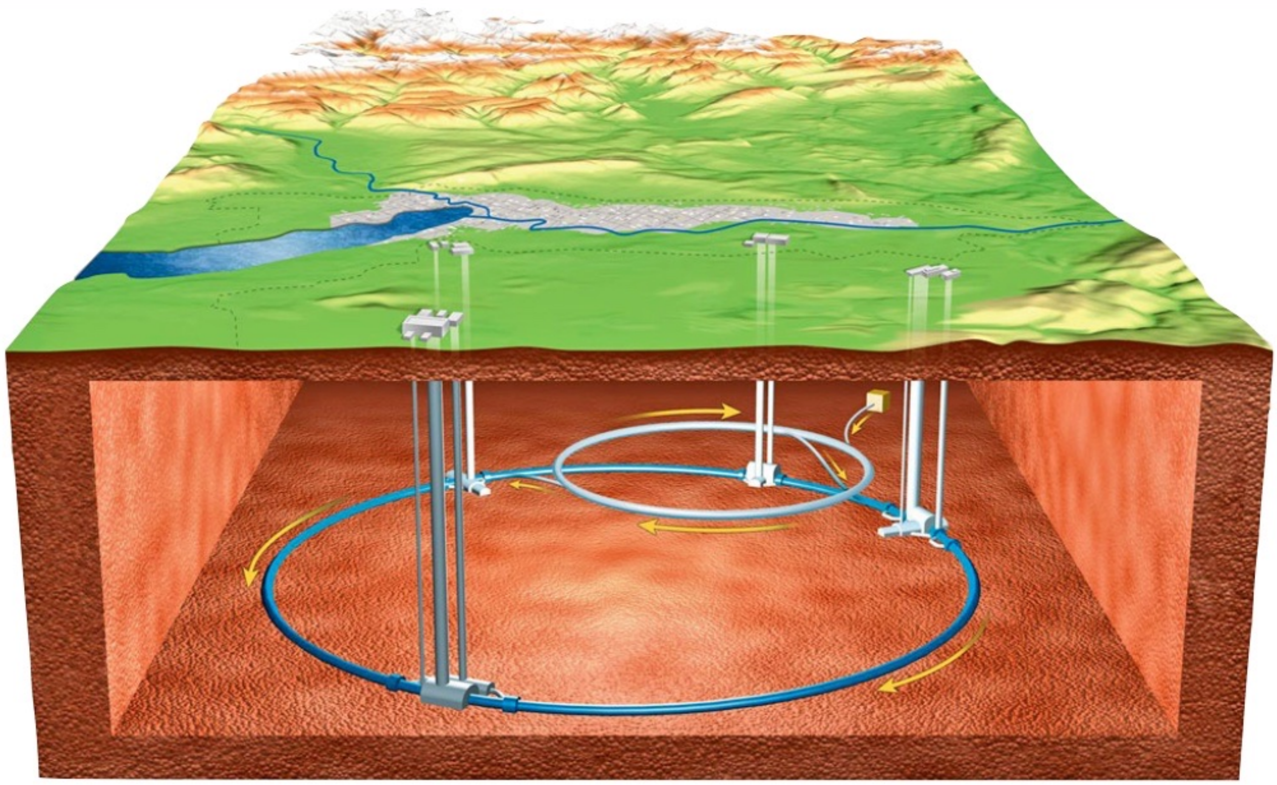
<u>Fact</u>: $W_p(\mu_n,\mu)\to 0$ iff
$\int f\,d\mu_n \to \int f\,d\mu$
for all $f\in C(\mathcal{X})$ with
$$|f(x)|\le C_1 + C_2 d^p(x,x_0)$$
for some $C_1, C_2\ge 0$, $x_0\in\mathcal{X}$

"continuous functions with at most p-growth"

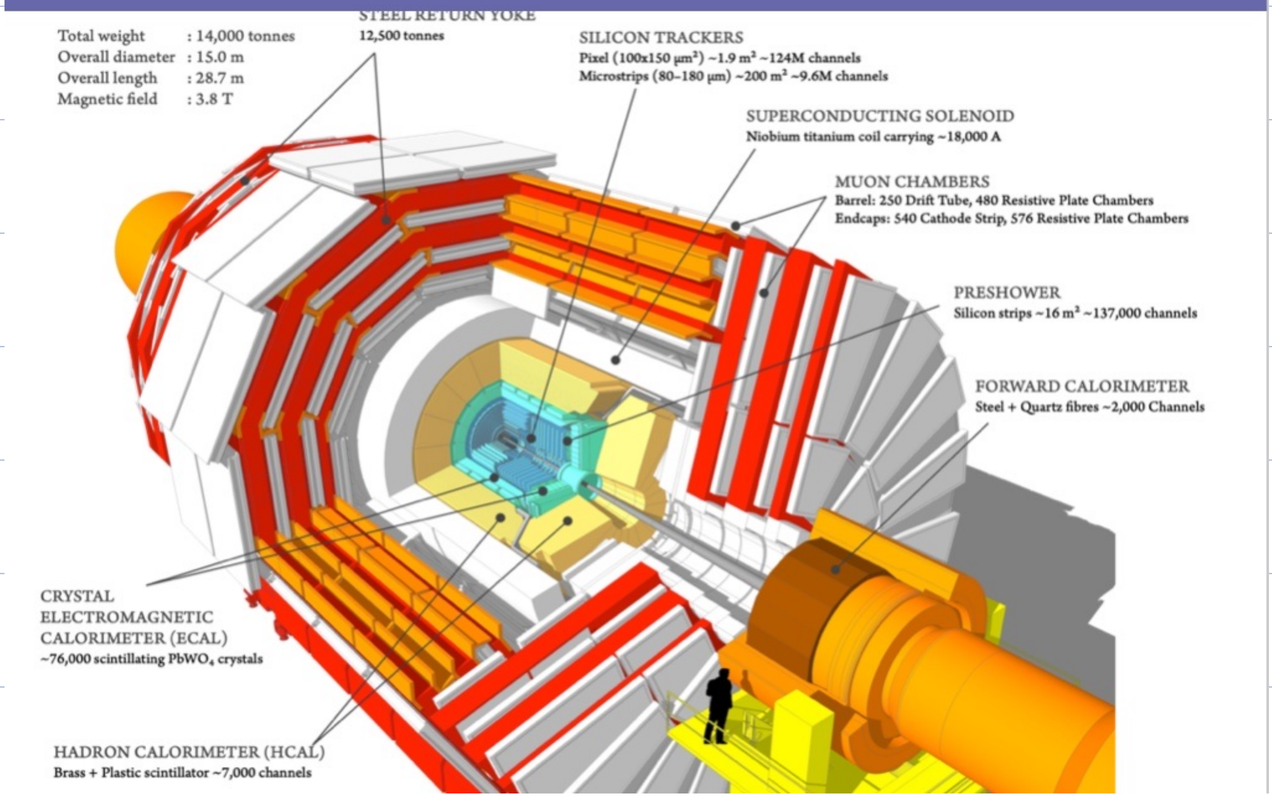
<u>Rmk</u>: Another consequence of the preceeding theorem is
$(\mathcal{X},d)$ Polish $\Rightarrow (\mathcal{P}_p(\mathcal{X}), W_p)$ is Polish
for any $p\ge 1$.

Application of Metric Properties of OT:

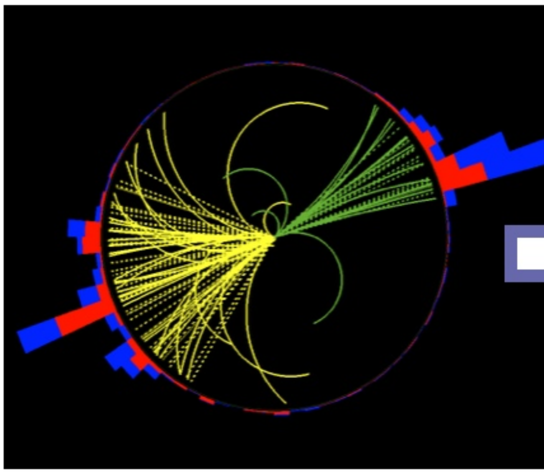
Large Hadron Collider



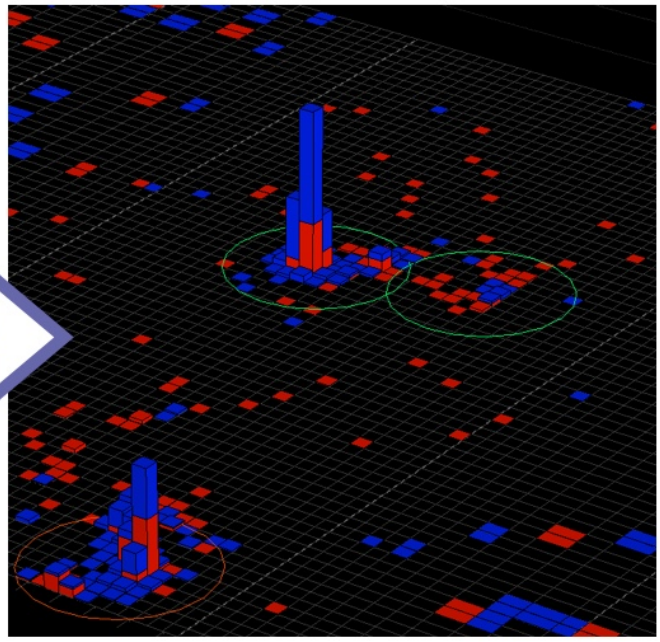
CMS Detector



Jet events on the calorimeter

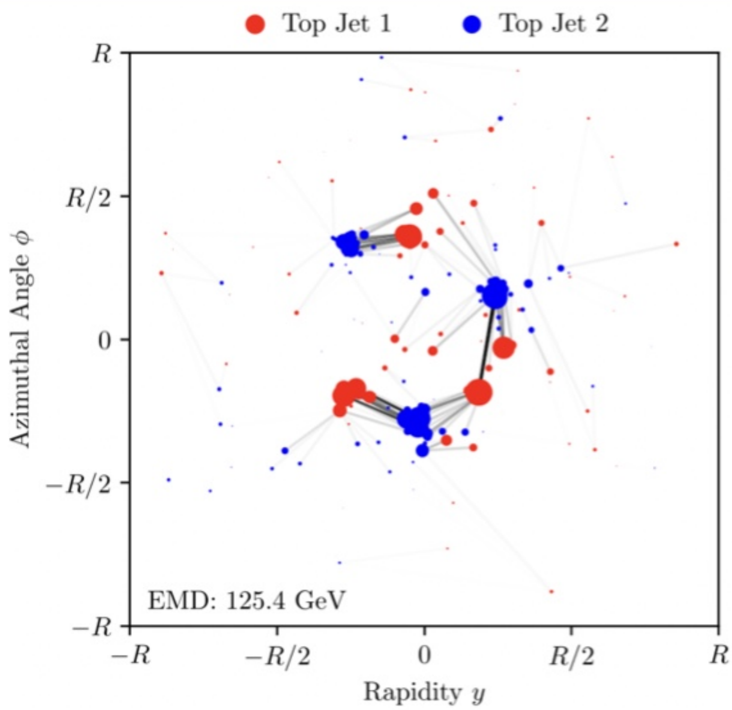


Cross section of cylindrical detector



Unroll cylinder and cluster into jets

Jet tagging



[Komiske, Metodiev, Thaler, 2019]

Goal of jet tagging: use hCal measurements to classify what type of event occurred at the parton level.

Key features of hCal data:

- spatial location is meaningful
- minimal overlapping support
- low resolution
  (pT measured ~200 locations; Fashion MNIST 784 pixels)

normalization…

Previous work

[Komiske, Metodiev, Thaler 2019], [Komiske, et. al. 2020]

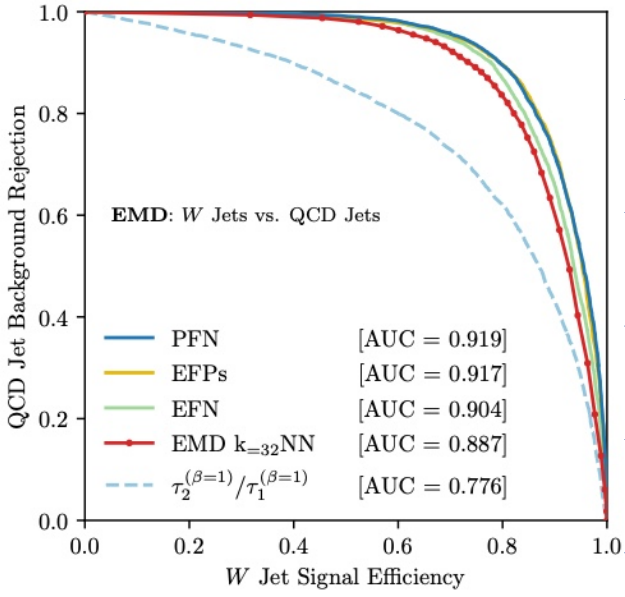
1) Compute W_1 distance between images
2) Apply KNN (balanced 100K training sample, 20K test sample)

Benefits:

- outperforms classical collider observables
- approaches accuracy of NN, superior interpretability

Challenges:

- requires $\mathcal{O}(N^2)$ evaluations of OT distance: ~16 years on a laptop using POT library
- large storage burden



Idea: Leverage good geometry of W_2 to improve computational efficiency

Next Topic: connect OT to PDE

- characterize "smooth" (in time) curves in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ as solutions of the continuity equation
- "dynamic" characterization of W_2
- geometric implications

The Continuity Equation (CE)

$$(CE)\begin{cases} \partial_t \mu_t + \nabla\cdot(\mu_t v_t) = 0 \\ \mu_t|_{t=0} = \mu_0 \end{cases}$$

Def: Given $T>0$, $\mu_0 \in \mathcal{M}(\mathbb{R}^d)$, $v : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ measurable, $\mu : [0,T] \to \mathcal{M}(\mathbb{R}^d)$ is a weak solution of (CE) if...

(i) $\forall \varphi \in C_c^\infty(\mathbb{R}^d)$,

(a) $t \mapsto \int_{\mathbb{R}^d} \varphi \, d\mu_t$ is abs cts

(b) $\frac{d}{dt} \int_{\mathbb{R}^d} \varphi \, d\mu_t = \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot v(x,t) \, d\mu_t(x)$ for a.e. $t \in [0,T]$

(ii) $\int_0^T \int_{\mathbb{R}^d} |v(x,t)| \, d\mu_t(x) \, dt < +\infty$

Rmk: (i)(a) ensures $\{\mu_t\}_{t \in [0,T]}$ is a Borel family, so that the integral in (ii) is well defined. (See Exercise 33.)

Rmk: If μ is a soln of (CE) in the above sense, then we also have

$\varphi(x,t) = \alpha(x)\beta(t)$

$$\int_0^T \int_{\mathbb{R}^d} \left(\partial_t \varphi(x,t) + \nabla\varphi(x,t)\cdot v(x,t)\right) d\mu_t(x)\,dt = 0$$

$$\forall \varphi \in C_c^\infty(\mathbb{R}^d \times (0,T))$$

See Ambrosio, Brue, Semola, Prop 16.3.

The preceding defn of solns to (CE) relies on Eulerian perspective. There is also a Lagrangian perspective.

Def: (solution of ODE)

Given $v : \mathbb{R}^d \times (0,T) \to \mathbb{R}^d$ meas, $x_0 \in \mathbb{R}^d$,

we say $x : [0,T] \to \mathbb{R}^d$ is a solution of

$$(\text{ODE}) \begin{cases} \dot{x}(t) = v(x(t),t) \\ x(0) = x_0 \end{cases}$$

if

(i) x is abs cts on $[0,T]$

(ii) $x(t) = x_0 + \int_0^t v(x(s),s)\,ds \quad \forall t \in [0,T]$.

A classical result in ODE is...

Thm (Cauchy-Lipschitz)

Suppose

- $v : \mathbb{R}^d \times (0,T) \to \mathbb{R}^d$ meas
- $\|v(\cdot,t)\|_{Lip} < +\infty \quad \forall t \in (0,T)$
  $\|v(\cdot,t)\|_{Lip} \in L^1_{loc}([0,T])$
- $\exists C > 0$ s.t. $|v(x,t)| \leq C(1+|x|) \; \forall x,t$

Then, $\forall x_0 \in \mathbb{R}^d$, $\exists!$ soln of ODE and

(i) $|x(t)| \leq f(t)|1+|x||$ for $f : (0,T) \to \mathbb{R}$ depending on C.

(ii) $|x(t) - y(t)| \leq \exp\left(\int_0^t \|v(\cdot,s)\|_{Lip} \, ds\right) |x_0 - y_0|$

solns of ODE

c.f. Ambrosio, Brue, Semola Thm 16.2

Fix $v(x,t)$. Suppose that a soln of (ODE) exists for all $x_0 \in \mathbb{R}^d$. In this case, we may consider the flow map induced by v:

$$X_t(y) := x(t), \text{ where } x(t) \text{ solves (ODE) w/ i.c. } x_0 = y.$$

Then, for all $t \in [0,T]$, $X_t : \mathbb{R}^d \to \mathbb{R}^d$ is well defined.

Next time, we will use this flow map to characterize solns of (CE). :)