Lecture 14

Reminders
- Solutions for 2-3 exercises
- Select article to revise by Fri, Feb 21

Recall:

Def: Given $(\mathcal{X}, d)$ Polish, $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $p \geq 1$,

$$W_p(\mu,\nu) := \min_{\gamma \in \Gamma(\mu,\nu)} \left(\mathbb{K}_p(\gamma)\right)^{1/p}.$$

Rmk: for $p \leq q$
- $W_p(\mu,\nu) \leq W_q(\mu,\nu)$
- $W_q(\mu,\nu) \leq \operatorname{diam}(\mathcal{X})^{1-p/q} W_p(\mu,\nu)^{p/q}$

$$\mathcal{P}_p(\mathcal{X}) := \Big\{\mu \in \mathcal{P}(\mathcal{X}) : \underbrace{\int d^p(x,x_0)\,d\mu(x)}_{=: m_p(\mu)} < +\infty \text{ for some } x_0 \in \mathcal{X}\Big\}$$

<u>Thm</u> (Disintegration): Given Polish spaces $\mathcal{X}, \mathcal{Y}$ and $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, let $\mu := \pi_{\mathcal{X}} \# \gamma$. Then there exists unique (μ-a.e.)

$$\{\gamma_x\}_{x \in \mathcal{X}} \subseteq \mathcal{P}(\mathcal{Y})$$

s.t., $\forall$ meas $f : \mathcal{X} \times \mathcal{Y} \to [0, +\infty]$,

- $x \mapsto \int_{\mathcal{Y}} f(x,y) \, d\gamma_x(y)$ is meas
- $\int_{\mathcal{X} \times \mathcal{Y}} f(x,y) \, d\gamma(x,y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x,y) \, d\gamma_x(y) \, d\mu(x)$

<u>Cor</u>: Given Polish space $\mathcal{X}$, $\gamma^{12} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, $\gamma^{23} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$ s.t.

$$\pi_2 \# \gamma^{12} = \pi_1 \# \gamma^{23} =: \mu.$$

Then $\exists$ $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{X})$ s.t.

$$\pi_{1,2} \# \gamma = \gamma^{12}, \qquad \pi_{2,3} \# \gamma = \gamma^{23}.$$

Thm: Suppose (X,d) is a Polish space. Then $\forall p \geq 1$, $(W_p, \mathcal{P}_p(X))$ is a metric space.

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

Next goal: characterize topology of W_p

Exercise 29: Suppose $\mu_n \to \mu$ and $\nu_n \to \nu$ narrowly and $\gamma_n \in \Gamma(\mu_n, \nu_n)$. Then $\exists\ \gamma_{n_k}$ s.t. $\gamma_{n_k} \to \gamma \in \Gamma(\mu, \nu)$ narrowly.

Prop: (W_p jointly lsc wrt narrow conv.)
Suppose $\mathcal{X}$ is a Polish space and $\mu_n, \nu_n \in \mathcal{P}(\mathcal{X})$ satisfy $\mu_n \to \mu$, $\nu_n \to \nu$ narrowly. Then

$$\liminf_{n\to\infty} W_p(\mu_n, \nu_n) \geq W_p(\mu, \nu).$$

Pf: Exercise 30

Now, we can characterize the topology of W_p.

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

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

Pf: First show "$\Rightarrow$"

Suppose $\lim_{n\to\infty} W_p(\mu_n,\mu)=0$.

Note that, $\forall \nu \in P_p(X)$

$$W_p^p(\nu, \delta_{x_0}) = \int d^p(x,y)\, d\gamma_* = m_p(\nu)$$

$\pi^{2\#}\gamma_* = \delta_{x_0} \Leftrightarrow y = x_0 \;\; \gamma_*\text{-a.e.}$

$$\underbrace{\lim_{n\to\infty} W_p^p(\mu_n, \delta_{x_0})}_{\lim_{n\to\infty} m_p(\mu_n)} = \underbrace{W_p^p(\mu, \delta_{x_0})}_{m_2(\mu)}$$

Now, we show $\mu_n \to \mu$ narrowly.

$$W_p(\mu_n,\mu) \geq W_1(\mu_n,\mu)$$
$$= \dots$$

$$\ldots = \sup_{\psi \in C(X),\ \|\psi\|_{Lip} \leq 1,\ \psi \in L^1(|\mu_n - \mu|)} \int \psi \, d(\mu_n - \mu)$$

$$\geq \sup_{\psi \in C_b(X),\ \|\psi\|_{Lip} \leq 1} \int \psi \, d(\mu_n - \mu)$$

Thus, for any $\psi \in C_b(X)$, $\|\psi\|_{Lip} \leq 1$,

$$\lim_{n \to \infty} \int \psi \, d\mu_n = \int \psi \, d\mu$$

More generally, for any $\psi \in C_b(X)$ with $\|\psi\|_{Lip} < +\infty$,

$$\lim_{n \to \infty} \int \psi \, d\mu_n = \lim_{n \to \infty} \frac{\|\psi\|_{Lip}}{\|\psi\|_{Lip}} \int \psi \, d\mu_n = \int \psi \, d\mu$$

Finally, for any $f \in C_b(X)$, Exercise 9 on the Moreau-Yosida Regularization shows that $\exists \{g_k\}_{k\in\mathbb{N}} \subseteq C_b(X)$, $\|g_k\|_{Lip} < +\infty$ s.t. $g_k \nearrow f$ pointwise and $g_k \geq \inf f$.

Thus,

$$\liminf_{n\to\infty} \int f \, d\mu_n \geq \liminf_{k\to\infty} \liminf_{n\to\infty} \int g_k \, d\mu_n$$
$$= \liminf_{k\to\infty} \int g_k \, d\mu$$
$$\overset{MCT}{\geq} \int f \, d\mu$$

Similarly, since $-f \in C_b(X)$

$$\liminf_{n\to\infty} -\int f \, d\mu_n \geq -\int f \, d\mu$$
$$\Leftrightarrow \int f \, d\mu \geq \limsup_{n\to\infty} \int f \, d\mu_n$$

Thus $\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$ and $\mu_n \to \mu$ narrowly.

Now, we show "$\le$."

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

Choose γ_n OT plans from μ_n to μ.
We must show $\lim_{n\to\infty} \int d^p(x,y)\, d\gamma_n = 0$

Step 1: γ_n doesn't put "too much" mass at infinity

- Since $m_p(\mu) < +\infty$, by MCT
$$\lim_{R\to+\infty} \int_{B_{R/2}(x_0)} d^p(x_0,y)\, d\mu(y) = m_p(\mu) < +\infty$$
Thus, $\forall\, \varepsilon > 0$, $\exists\, R$ s.t.
$$\int_{B^c_{R/2}(x_0)} d^p(x_0,y)\, d\mu = m_p(\mu) - \int_{B_{R/2}(x_0)} d^p(x_0,y)\, d\mu(y) < \varepsilon$$

- Thus, for all $\varepsilon > 0$, $\exists$ R suff large s.t.

$$\limsup_{n\to\infty} \int_{B^c_{R/2}(x_0)} d^p(x_0,y)\,d\mu_n(y)$$

$$= \limsup_{n\to\infty} m_p(\mu_n) + - \int_{B_{R/2}(x_0)} d^p(x_0,y)\,d\mu_n$$

$$\leq m_p(\mu) - \liminf_{n\to\infty} \int \underbrace{d^p(x_0,y)\,\mathbb{1}_{B_{R/2}(x_0)}(y)}_{\text{lsc, bdd below}}\,d\mu_n$$

Portmanteau

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

$$< \varepsilon$$

- Thus μ_n doesn't give too much mass at infinity

Fact: $d(x,y) \leq 2\max(d(x,x_0), d(x_0,y))$

$d(x,y) \geq c \implies$ either $d(x,x_0) \geq c/2$
or $d(x_0,y) \geq c/2$

• $$\limsup_{n\to\infty} \int_{d(x,y)\geq R} d^p(x,y)\, d\gamma_n(x,y)$$

$$\leq \limsup_{n\to\infty} 2^p \int_{(B^c_{R/2}(x_0)\times \mathcal{X})\cup(\mathcal{X}\times B^c_{R/2}(x_0))} \overbrace{\max(d^p(x,x_0), d^p(x_0,y))}^{f(x,y)}\, d\gamma_n$$

$$\leq \limsup_{n\to\infty} 2^p \int_{B^c_{R/2}(x_0)\times \mathcal{X}} f(x,y)\, d\gamma_n + 2^p \int_{\mathcal{X}\times B^c_{R/2}(x_0)} f(x,y)\, d\gamma_n$$

$\leftarrow B^c_{R/2}(x_0) \rightarrow$

doesn't change value of integrand

Will fix this step next time.

For now, take the conclusion of this estimate: $\forall \varepsilon>0$, $\exists R>0$ suff large so that

$$\limsup_{n\to\infty} \int_{d(x,y)\geq R} d^p(x,y)\, d\gamma_n < \varepsilon.$$

Consider

$$\limsup_{n\to\infty} \int d^p(x,y)\,d\gamma_n$$

choose a subsequence γ_{n_k} so that

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

Since $\mu^{n_k}\to\mu$, $\nu^{n_k}\to\nu$ narrowly, $\exists$ further subsequence $\gamma_{n_{k_\ell}}$ s.t. $\gamma_{n_{k_\ell}} \to \gamma \in \Gamma(\mu,\nu)$.

abbreviate γ_{n_k} for simplicity

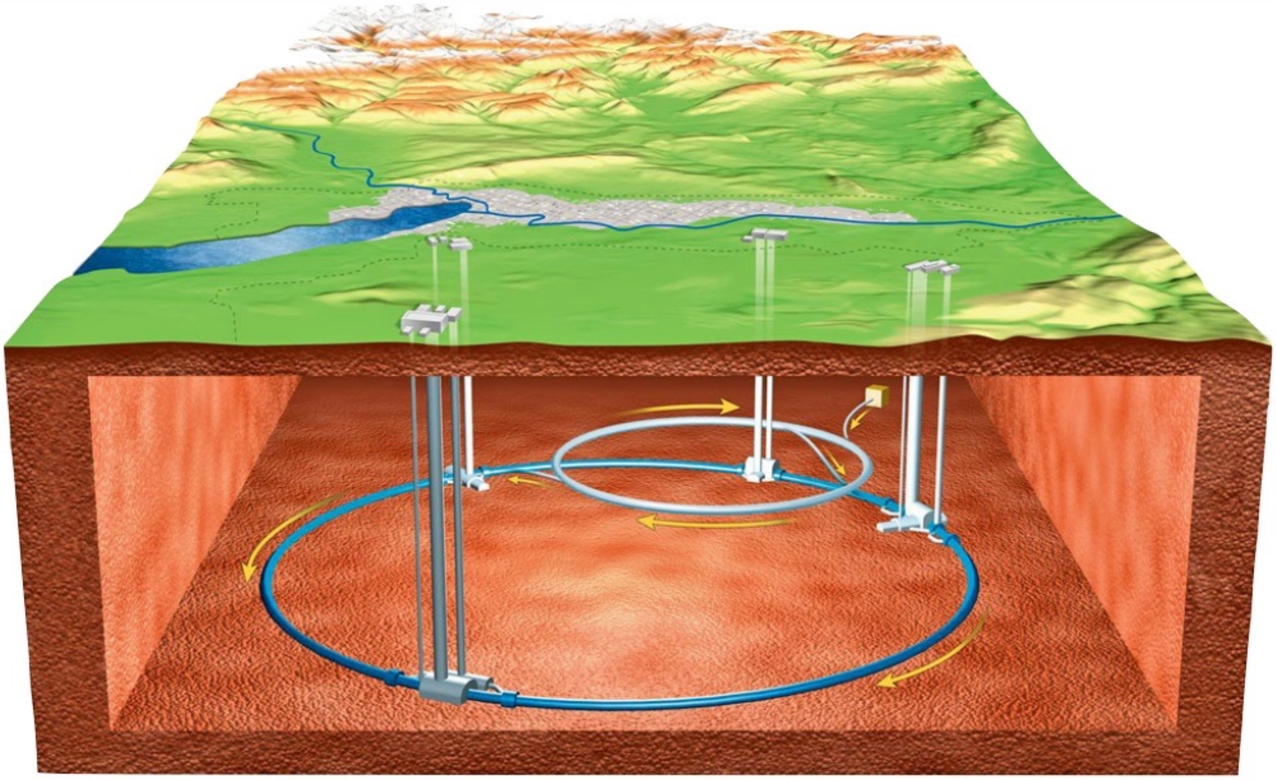
Furthermore, Portmanteau, since $d^p(x,y)$ is lsc and bdd below

$$\mathbb{K}_p(\gamma) \leq \liminf_{k\to\infty} \mathbb{K}_p(\gamma_{n_k})$$

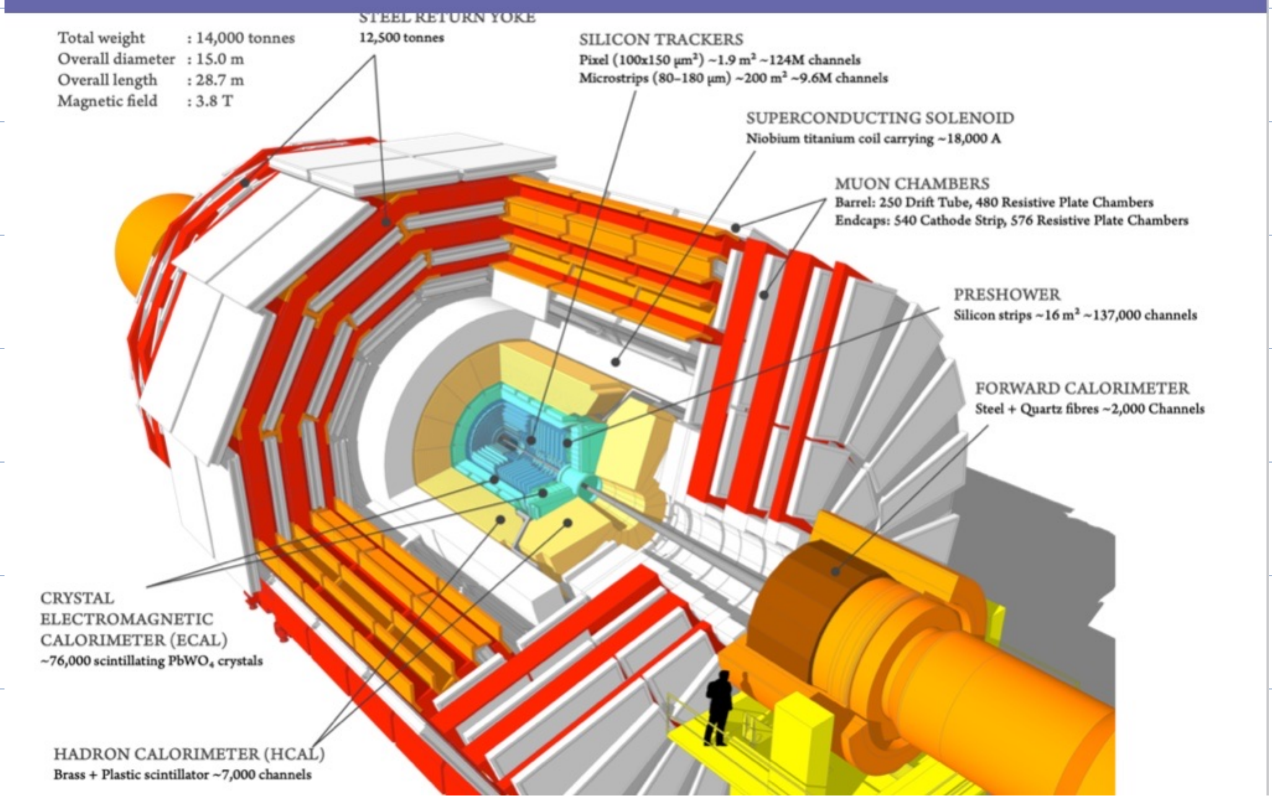
Finish next time...

Application of Metric Properties of OT:

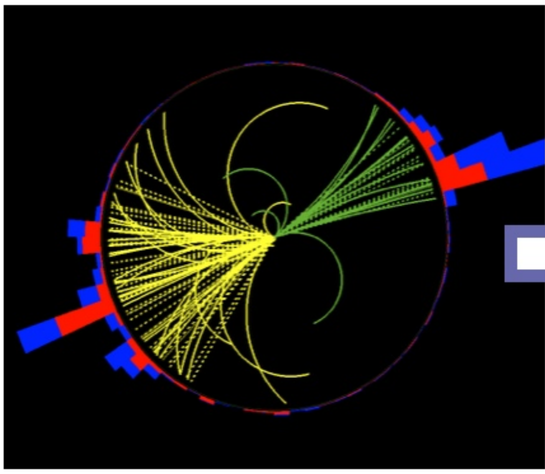
Large Hadron Collider



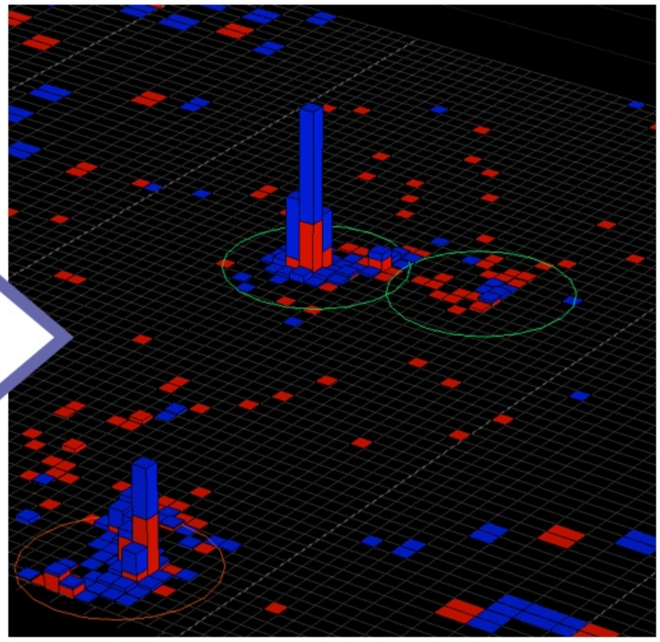
CMS Detector



Jet events on the calorimeter

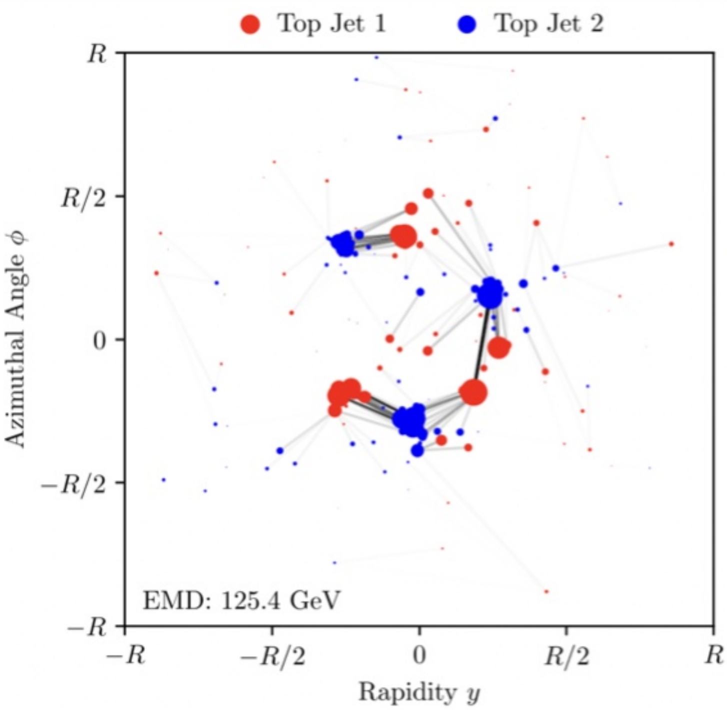


Cross section of cylindrical detector



Unroll cylinder and cluster into jets

Jet tagging



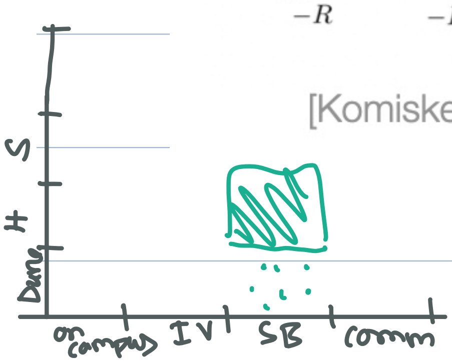
[Komiske, Metodiev, Thaler, 2019]

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]

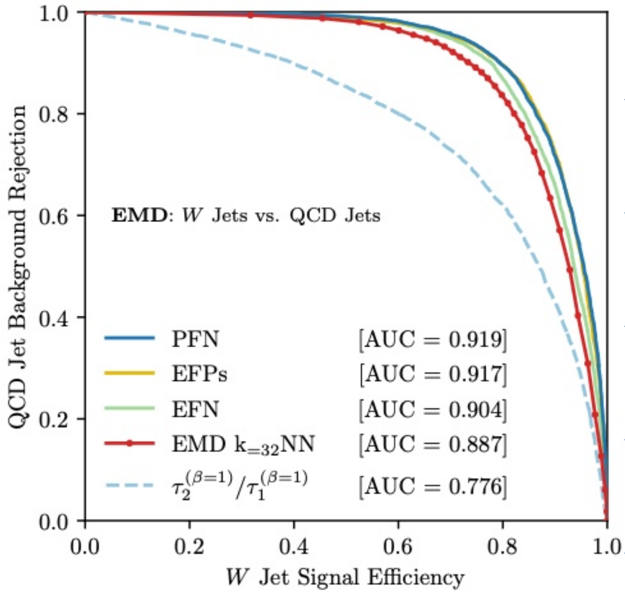
1) Compute W_1 distance between images
2) Apply KNN (balanced 100K training sample, 20K test sample)

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