SYSU-PKU Collider Physics Forum

For Young Scientists

3/15/2023

Which Metric on the Space of Collider Events?

Speaker: Tianji Cai

Collaborators: Junyi Cheng, **Nathaniel Craig**, Katy Craig, Bernhard Schmitzer, Matthew Thorpe.

Based on 2008.08604, 2102.08807, 2111.03670, and work in progress.

Department of Physics

UC SANTA BARBARA

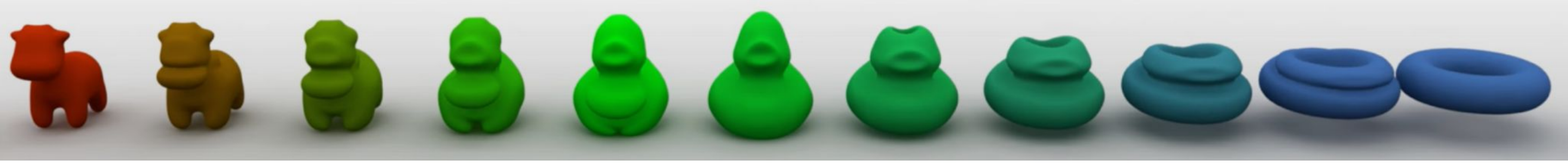


Image Credit: Solomon, et al. SIGGRAPH 2015

Contents

- **Introduction** — *What is it & Why should we care?*
- **Theory of Optimal transport (OT)** — *Small doses of not-too-scary math…*

  History; Balanced OT; Unbalanced OT; Linearized OT.
- **Optimal Transport for Collider Physics** — *Successful debut!*

  Physics Background; A Unified Framework; Jet Tagging.
- **Summary & Outlook** — *Let’s have more fun.* 👻

1. Introduction

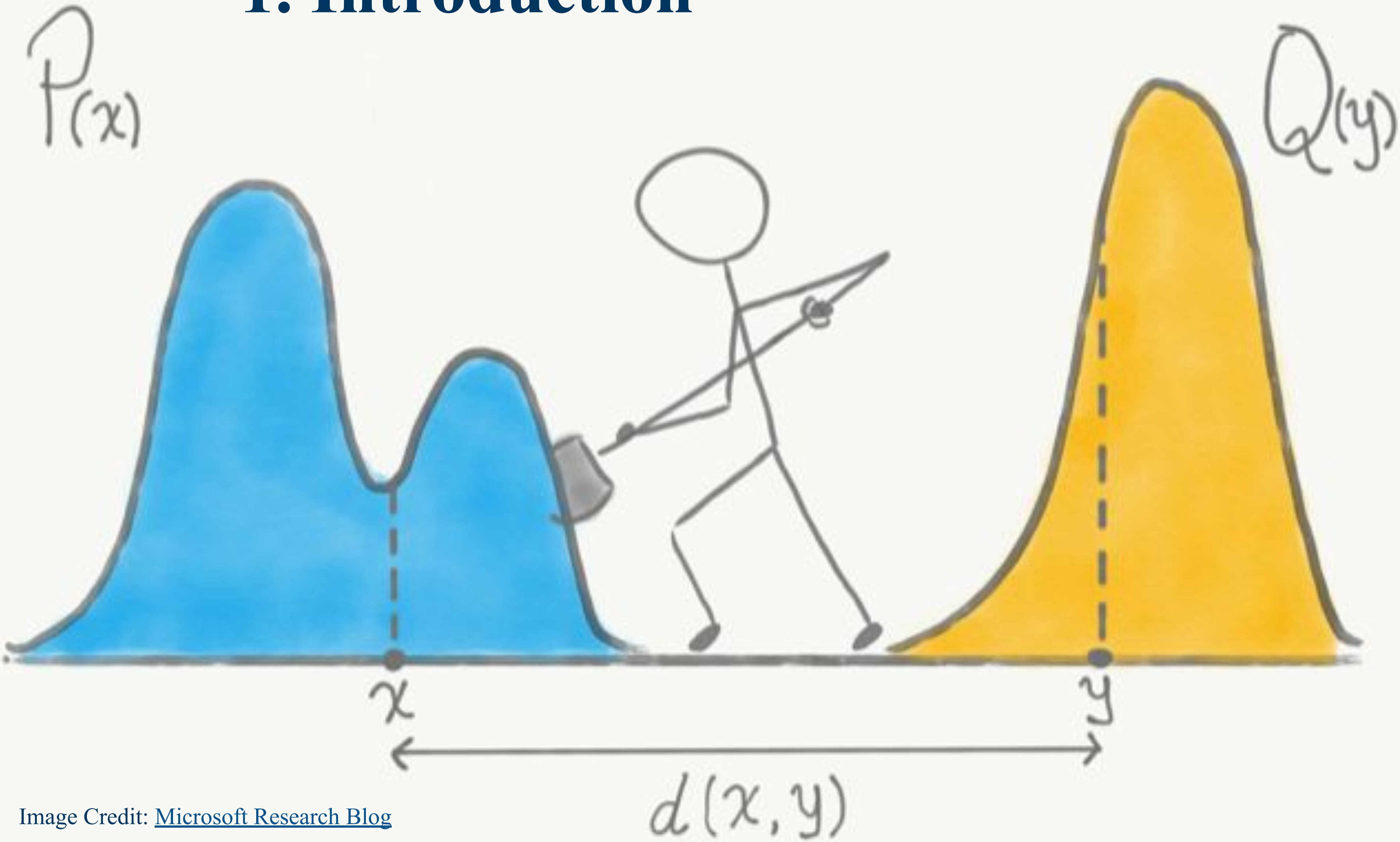
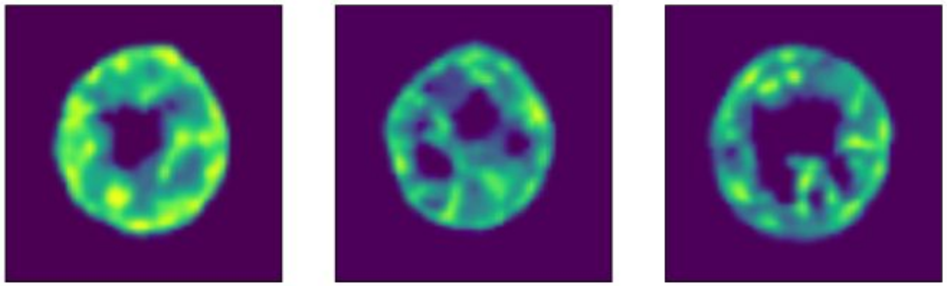


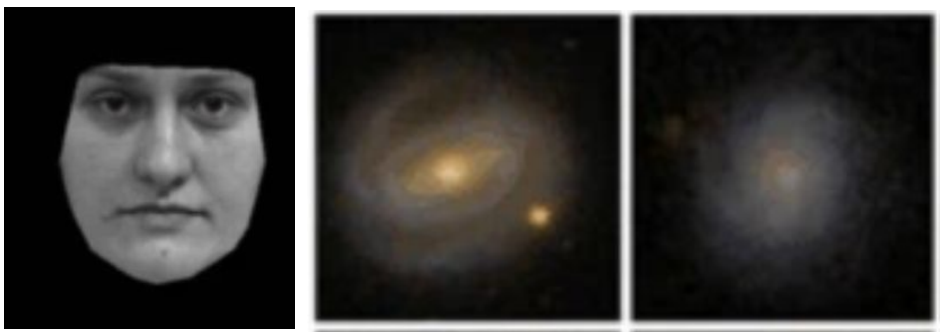
Image Credit: Microsoft Research Blog

Typical Scenario

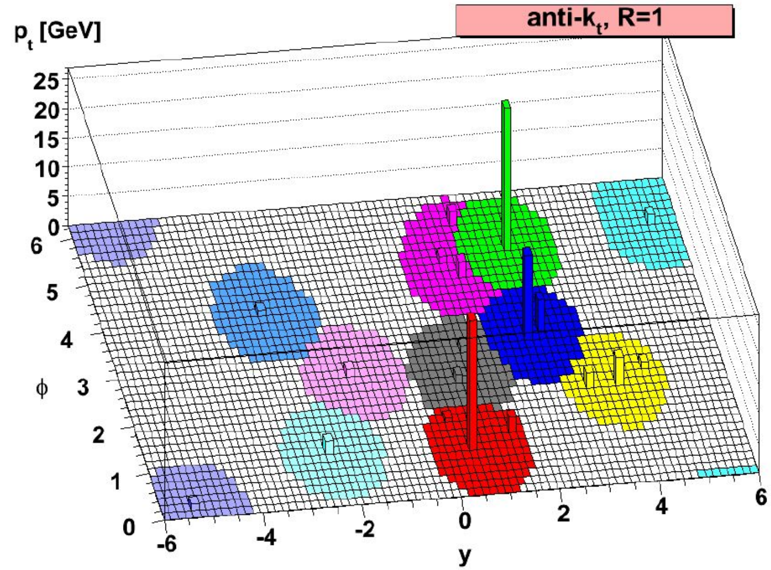
Given two distributions, discrete or continuous…

Example images of liver cells, Fig 5.4.





Example facial images and galaxy images, Fig 6.



Example jet image, Fig 2.2 in *Hadronic Jets*.

Q1: How to define a distance between them?

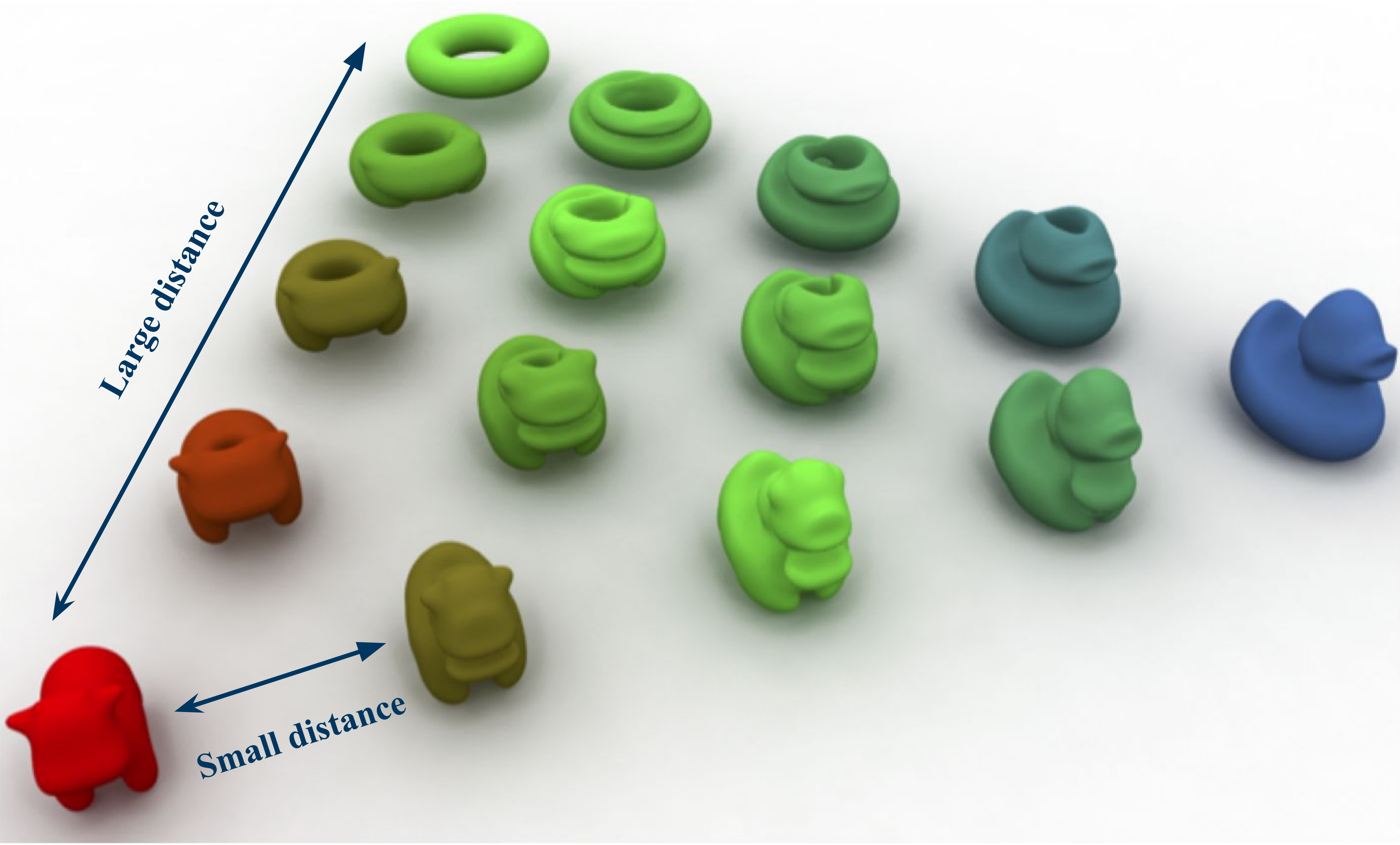
+

Q2: How to rearrange one to look like the other?

Our Physics Goal here is…

To find a way to quantify the distance between collider events/jets.

How to transform a cow into a duck and into a donut…



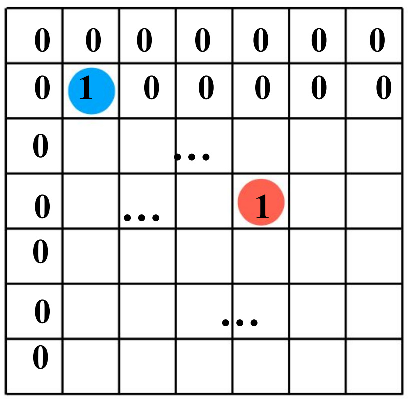
Credit: Fig 4 in paper.

Why is the normal Euclidean distance not enough?

Consider two particles with unit energy.

Image-based approach

Bin on N-bin grid, represent energy distribution by vectors in $\mathbf{R}^N$, compute Euclidean distance between vectors

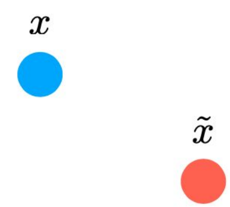


v = [0, …, 1, 0, …, 0, 0, …, 0]
v'= [0, …, 0, 0, …, 1, 0, …, 0]

$$d_{\ell^2(\mathbb{R}^N)}(\mathcal{E}, \tilde{\mathcal{E}}) = \left(\sum_{i=1}^{N} |v_i - \tilde{v}_i|^2 \right)^{1/2} = \sqrt{2}$$

regardless of positions

OT-based approach



$$W_p(\mathcal{E}, \tilde{\mathcal{E}}) = \|x - \tilde{x}\|$$

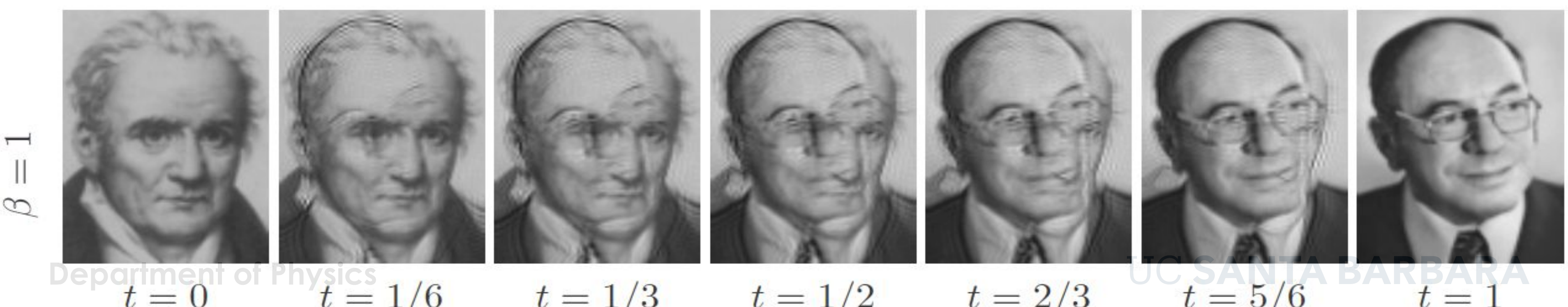
Invaluable if the relative distribution of pixels carries meaning

OT preserves the underlying geometry of the ground space!

Monge ***Kantorovich***

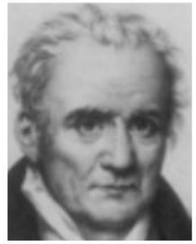
2. Theory of Optimal Transport

Image Credit: N. Papadakis, et al. [10.1137/130920058]



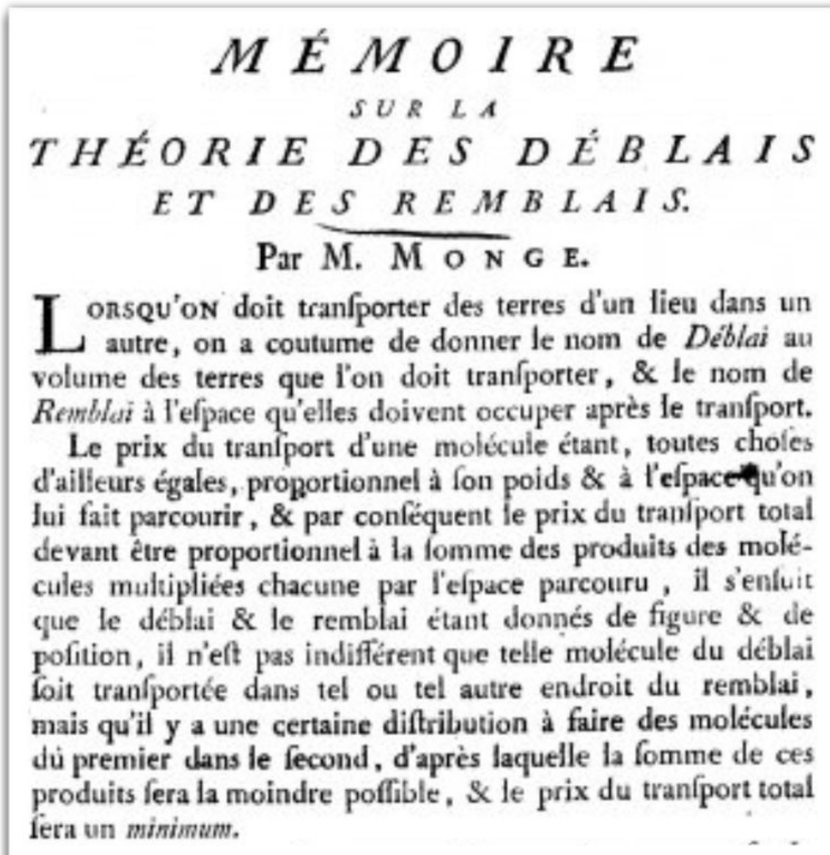
2.1 A Brief History of Optimal Transport

Gaspard Monge
1781



Fundamental problem of **optimal transport:**

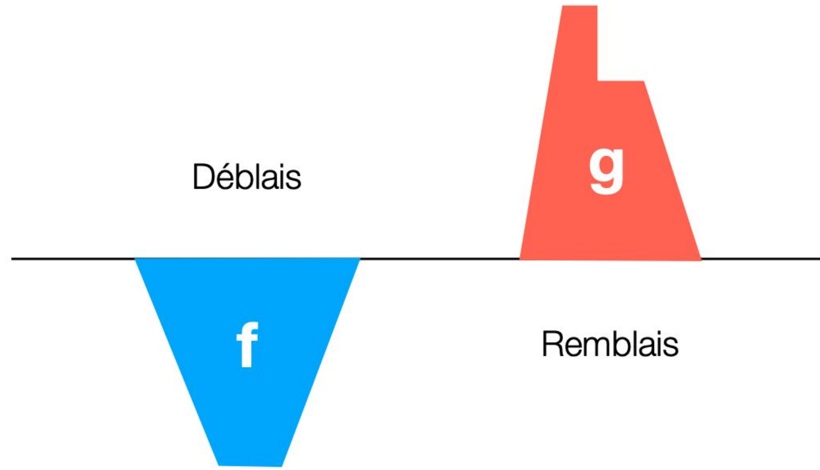
How to rearrange **f** to look like **g**
with the **least amount of "work"?**



MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.

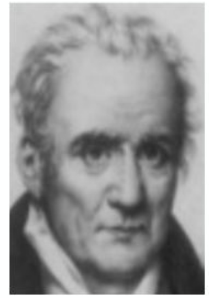


In other words, how can we optimally transport *f* to *g*?

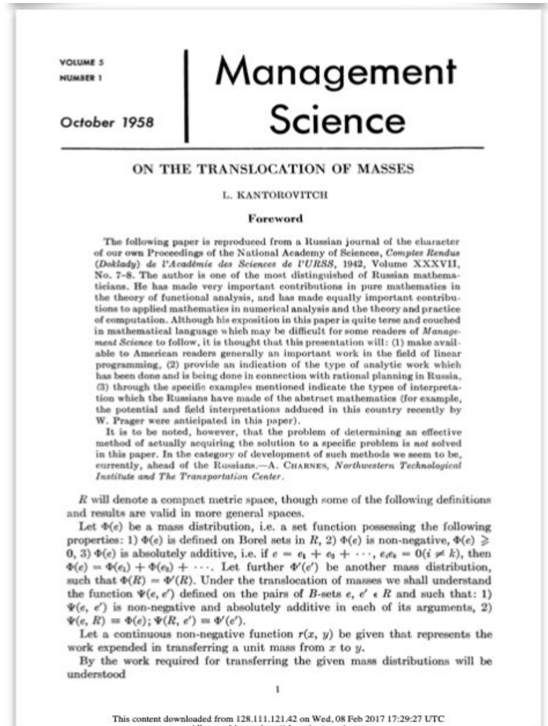
Sounds pretty simple?
Took over 100 years to formulate the problem mathematically!

OT in the Modern Time

1781: Gaspard Monge,
Mémoire sur la théorie des deblais et des remblais
(On cuttings and embankments)



1942: Leonid Kantorovich,
On the translocation of masses



VOLUME 5
NUMBER 1
October 1958
Management Science

ON THE TRANSLOCATION OF MASSES

L. KANTOROVITCH

Foreword

The following paper is reproduced from a Russian journal of the character of our own Proceedings of the National Academy of Sciences, *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, 1942, Volume XXXVII, No. 7–8. The author is one of the most distinguished of Russian mathematicians. He has made very important contributions in pure mathematics in the theory of functional analysis, and has made equally important contributions to applied mathematics in numerical analysis and the theory and practice of computation. Although his exposition in this paper is quite terse and couched in mathematical language which may be difficult for some readers of *Management Science* to follow, it is thought that this presentation will: (1) make available to American readers generally an important work in the field of linear programming, (2) provide an indication of the type of analytic work which has been done and is being done in connection with rational planning in Russia, (3) through the specific examples mentioned indicate the types of interpretation which the Russians have made of the abstract mathematics (for example, the potential and field interpretations adduced in this country recently by W. Prager were anticipated in this paper).

It is to be noted, however, that the problem of determining an effective method of actually acquiring the solution to a specific problem is *not* solved in this paper. In the category of development of such methods we seem to be, currently, ahead of the Russians.—A. Charnes, *Northwestern Technological Institute and The Transportation Center.*

R will denote a compact metric space, though some of the following definitions and results are valid in more general spaces.

Let $\Phi(e)$ be a mass distribution, i.e. a set function possessing the following properties: 1) $\Phi(e)$ is defined on Borel sets in R, 2) $\Phi(e)$ is non-negative, $\Phi(e) \geqslant 0$, 3) $\Phi(e)$ is absolutely additive, i.e. if $e = e_1 + e_2 + \cdots$, $e_i e_k = 0 (i \neq k)$, then $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \cdots$. Let further $\Phi'(e')$ be another mass distribution, such that $\Phi(R) = \Phi'(R)$. Under the translocation of masses we shall understand the function $\Psi(e, e')$ defined on the pairs of B-sets $e, e' \epsilon R$ and such that: 1) $\Psi(e, e')$ is non-negative and absolutely additive in each of its arguments, 2) $\Psi(e, R) = \Phi(e); \Psi(R, e') = \Phi'(e')$.

Let a continuous non-negative function $r(x, y)$ be given that represents the work expended in transferring a unit mass from x to y.

By the work required for transferring the given mass distributions will be understood

1

1999: Felix Otto,
The geometry of dissipative evolution equations: the porous medium equation

2000: Felix Otto, Cedric Villani,
Generalization of an inequality by Talagrand, as a consequence of the logarithmic Sobolev inequality

2010: Cedric Villani wins Fields medal

2018: Alessio Figalli wins Fields medal

OT gives a family of well-defined metrics between distributions.

2.2 Balanced Optimal Transport

Let's first look at the Earth Mover's Distance, one simple example of OT distance.

$$EMD(E, E') = \min_{f_{ij} \in \Gamma_{E,E'}} \sum_{ij} f_{ij}\theta_{ij}$$

theta$_{ij}$: ground metric between particles i and j

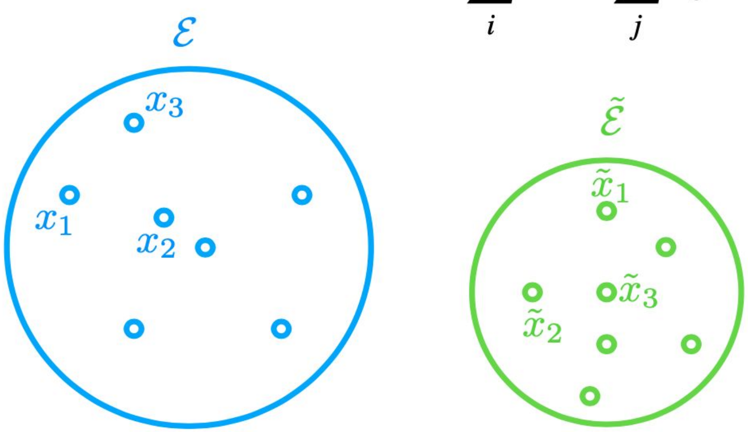
$$\Gamma_{E,E'} = \left\{ f_{ij} : f_{ij} \geq 0, \sum_j f_{ij} = E_i, \sum_i f_{ij} = E'_j \right\}$$

f$_{ij}$: the amount of mass moved from particle i to particle j.

Consider two "events" $\mathscr{E}$, $\tilde{\mathscr{E}}$

Collections of particles at locations $x_i, \tilde{x}_j$ in a rectangular domain with masses $E_i, \tilde{E}_j \geq 0$

Assume same total mass $\sum_i E_i = \sum_j \tilde{E}_j$

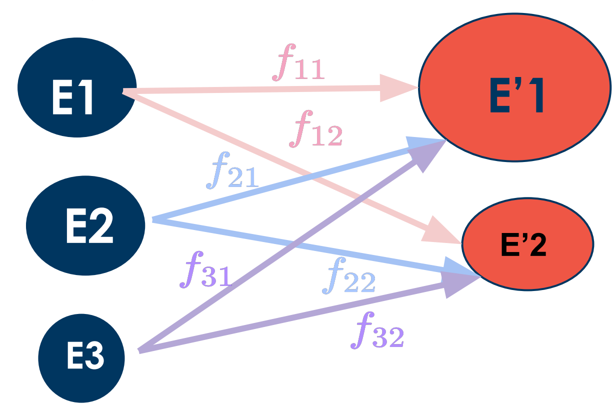


Example:

$$E_1 + E_2 + E_3 = E'_1 + E'_2$$

$$f_{11} + f_{12} = E_1$$
$$f_{21} + f_{22} = E_2$$
$$f_{31} + f_{32} = E_3$$

$$f_{11} + f_{21} + f_{31} = E'_1$$
$$f_{12} + f_{22} + f_{32} = E'_2$$



Generalize to the p-Wasserstein distances

$$W_p(\mathcal{E},\tilde{\mathcal{E}}) = \min_{g_{ij}\in\Gamma(\mathcal{E},\tilde{\mathcal{E}})} \left(\sum_{ij} g_{ij}\|x_i - \tilde{x}_j\|^p\right)^{1/p}$$

$$\Gamma(\mathcal{E},\tilde{\mathcal{E}}) = \left\{ g_{ij} : g_{ij} \geq 0, \sum_j g_{ij} = E_i, \sum_i g_{ij} = \tilde{E}_j \right\}$$

p=1: Earth Mover's Distance (EMD)
p=2: Monge-Kantorovich Distance / 2-Wasserstein (W_2) Distance

We focus on the W_2 Distance.

Answer: W_2 has a (weak) Riemannian structure => can linearize!

Two Equivalent Formulations of W_2 Distance:

Kantorovich formulation (static):

$$W_2(\mathcal{E}, \tilde{\mathcal{E}}) = \min_{r_{ij} \in \Gamma(\mathcal{E}, \tilde{\mathcal{E}})} \left(\sum_{ij} r_{ij} \|x_i - \tilde{x}_j\|^2 \right)^{1/2}$$

Benamou-Brenier formulation (dynamic):

$$\partial_t \rho + \operatorname{div} \omega = 0$$

Continuity Equation

⇔ Charge conservation

ρ: charge density

ω: current density

No source/sink

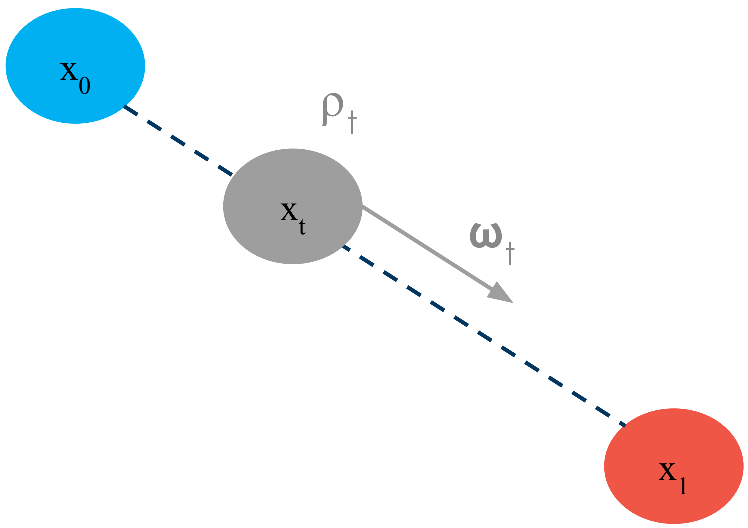
$$J_{\mathrm{W}}(\rho, \omega) := \begin{cases} \int_{[0,1]\times\Omega} \|\frac{\mathrm{d}\omega}{\mathrm{d}\rho}\|^2 \mathrm{d}\rho & \text{if } \rho \geq 0, \omega \ll \rho \\ +\infty & \text{else.} \end{cases}$$

Cost Functional

$$\mathrm{W}_2(\mu_0, \mu_1)^2 := \inf \left\{ J_{\mathrm{W}}(\rho, \omega) | (\rho, \omega) \in \mathcal{CE}(\mu_0, \mu_1) \right\}.$$

Minimizer

Example: Two dirac masses at position x_0 and x_1.



Let $X(x_0, x_1; t) = (1-t)\, x_0 + t\, x_1\,.$

$$W_2(\delta_{x_0}, \delta_{x_1})^2 = \|x_0 - x_1\|^2\,.$$

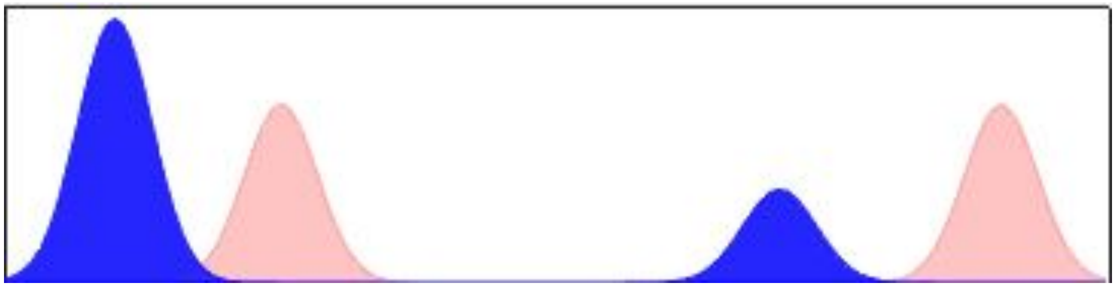
$$\rho_t = \delta_{X(x_0,x_1;t)}$$

$$\omega_t = \delta_{X(x_0,x_1;t)} \cdot \partial_t X(x_0, x_1; t)\,.$$

A Dirac-to-Dirac geodesic in W_2 consists of a single Dirac traveling along the straight line between them at constant speed.

2.3 Unbalanced Optimal Transport

Balanced Transport: Mass can only be transported, not created or destroyed. Therefore, the total mass of the two distributions must be exactly equal.



Unbalanced Transport: Mass can be transported, created and destroyed. The total mass of the two distributions can be different (but also can be the same too).

Credit: L. Chizat, G. Peyré, B. Schmitzer, F-X. Vialard on G. Peyré's Github.

Balanced OT $\in$ Unbalanced OT

Before introducing unbalanced OT, we were doing … for EMD.

Same as Standard EMD

Extra piece to account for the unequal total mass

$$\mathrm{EMD}_R(\mathcal{E},\mathcal{E}') := \min_{f_{ij}\in\tilde{\Gamma}_{\leqslant(\mathcal{E},\mathcal{E}')}} \frac{1}{R}\sum_{ij} f_{ij}\theta_{ij} + \left|\sum_i E_i - \sum_j E'_j\right| \quad (1.2)$$

$$\tilde{\Gamma}_{\leqslant(\mathcal{E},\mathcal{E}')} := \left\{ f_{ij} : f_{ij} \geqslant 0,\ \sum_j f_{ij} \leqslant E_i,\ \sum_i f_{ij} \leqslant E'_j,\ \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)\right\}. \quad (1.3)$$

Different conditions for the unbalanced case

Ad-hoc!

Now we can incorporate the total mass difference in a natural way by allowing mass to be created and destroyed in addition to being transported.

The Unbalanced Twin of W_2: Hellinger-Kantorovich (HK) Distance

Kantorovich type formulation (static):

$$c(x_0, x_1) := \begin{cases} -2\log(\cos(\|x_0 - x_1\|)) & \text{if } \|x_0 - x_1\| < \frac{\pi}{2} \\ +\infty & \text{else.} \end{cases}$$

$$J_{\mathrm{SM}}(\pi) := \int_{\Omega^2} c \,\mathrm{d}\pi + \sum_{i\in\{0,1\}} \mathrm{KL}(\mathrm{P}_{i\sharp}\pi|\mu_i).$$

$$\mathrm{HK}(\mu_0, \mu_1)^2 = \inf\left\{ J_{\mathrm{SM}}(\pi) \,\middle|\, \pi \in \mathcal{M}_+(\Omega^2) \right\}$$

Bonus: HK also enjoys a (weak) Riemannian structure => can linearize!

Benamou-Brenier-type formulation (dynamic):

$$\partial_t \rho + \operatorname{div}\omega = \zeta$$

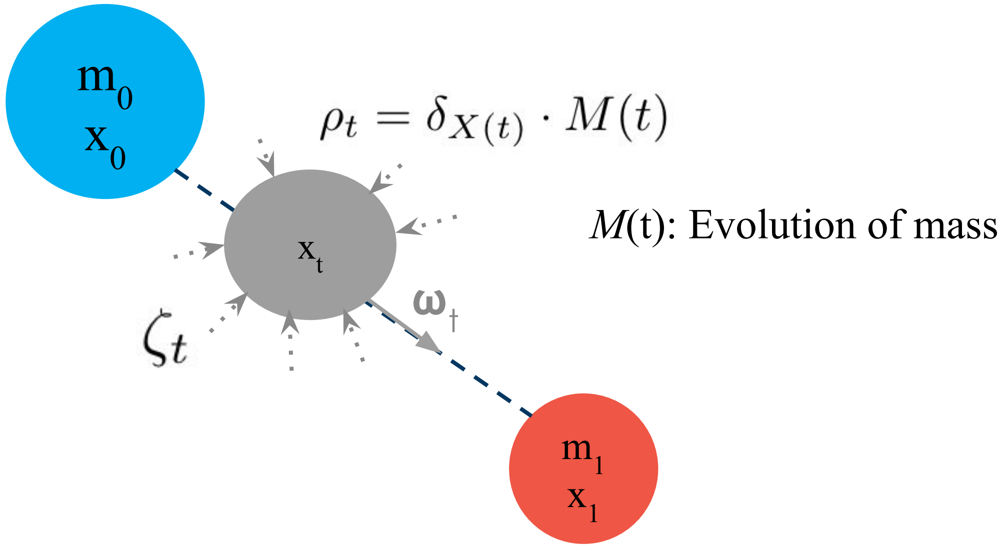
Continuity Equation with Source

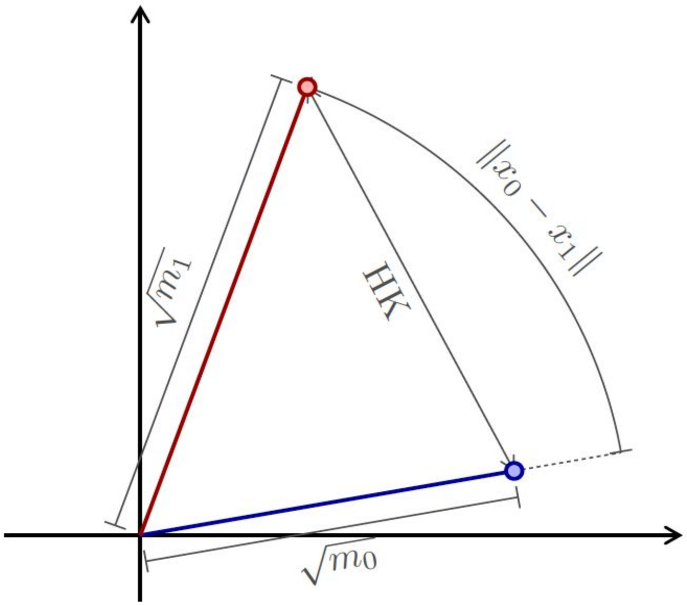
$$J_{\mathrm{HK}}(\rho, \omega, \zeta) := \begin{cases} \int_{[0,1]\times\Omega} \left(\left\| \frac{\mathrm{d}\omega}{\mathrm{d}\rho} \right\|^2 + \frac{1}{4} \left(\frac{\mathrm{d}\zeta}{\mathrm{d}\rho} \right)^2 \right) \mathrm{d}\rho & \text{if } \rho \geq 0, \omega, \zeta \ll \rho, \\ +\infty & \text{else.} \end{cases}$$

Additional term takes care of mass creation/destruction

$$\mathrm{HK}(\mu_0, \mu_1)^2 := \inf\left\{ J_{\mathrm{HK}}(\rho, \omega, \zeta) | (\rho, \omega, \zeta) \in \mathcal{CES}(\mu_0, \mu_1) \right\}$$

Example: Again two dirac masses at position x_0 and x_1 with mass m_0 and m_1, respectively.





$$\mathrm{HK}(\delta_{x_0} \cdot m_0, \delta_{x_1} \cdot m_1)^2 = m_0 + m_1 - 2\sqrt{m_0\, m_1}\,\overline{\cos}(\|x_0 - x_1\|)$$

where $\overline{\cos}(s) = \cos\left(\min\{s, \frac{\pi}{2}\}\right)$.

For $\|x_0 - x_1\| \le \frac{\pi}{2}$ $\quad$ $\mathrm{HK}(\delta_{x_0} \cdot m_0, \delta_{x_1} \cdot m_1) = \|\sqrt{m_0} - \sqrt{m_1}\exp(i\|x_0 - x_1\|)\|$

When $\|x_0 - x_1\| > \frac{\pi}{2}$ $\quad$ Mass at x_0 is destroyed, mass at x_1 created. No transport!

HK Distance has an intrinsic length scale κ!

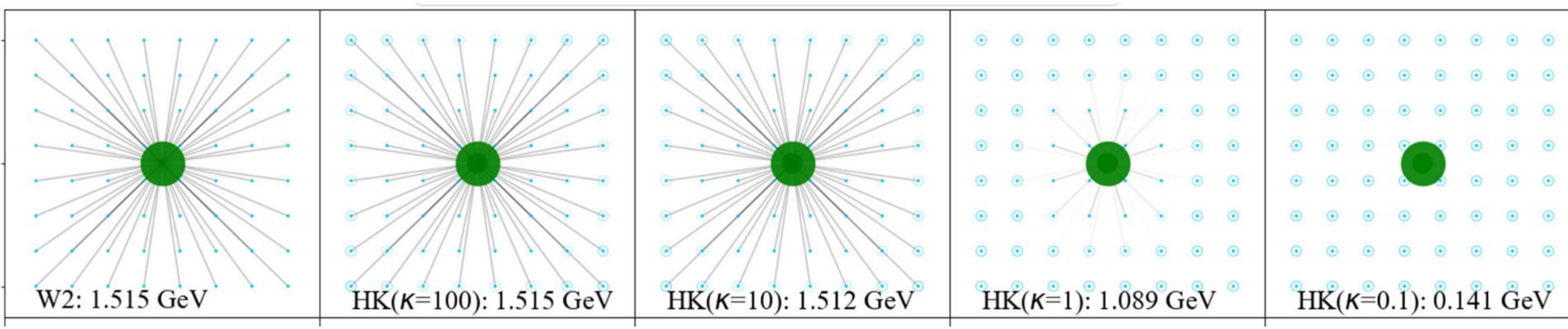
$$J_{\mathrm{HK},\kappa}(\rho,\omega,\zeta) := \begin{cases} \int_{[0,1]\times\Omega} \left(\left\| \frac{\mathrm{d}\omega}{\mathrm{d}\rho} \right\|^2 + \frac{\kappa^2}{4} \left(\frac{\mathrm{d}\zeta}{\mathrm{d}\rho} \right)^2 \right) \mathrm{d}\rho & \textit{if } \rho \geq 0, \omega, \zeta \ll \rho, \\ +\infty & \textit{else.} \end{cases}$$

Intrinsic length scale κ>0 controls the relative importance of the transport part of the cost and the creation/destruction part.

$\mathrm{HK}_\kappa(\mu_0,\mu_1) \to \mathrm{W}_2(\mu_0,\mu_1)$

$\mathrm{HK}_\kappa(\mu_0,\mu_1)/\kappa \to$ Hellinger distance (~ Euclidean)

κ: ∞ → 0



HK Distance: Normalized vs. Unnormalized Distributions

$$\mathrm{HK}(m_0 \cdot \mu_0, m_1 \cdot \mu_1)^2 = \sqrt{m_0 \cdot m_1} \cdot \mathrm{HK}(\mu_0, \mu_1)^2 + (\sqrt{m_0} - \sqrt{m_1})^2$$

Unnormalized Distributions
Transport plan: $\sqrt{(m_0 m_1)}\pi$

Normalized Distributions
Transport plan: π

Unbalanced HK on ***un*****normalized** measures can be obtained from HK from **normalized** measures.

Local mass discrepancies more **important than** the differences in the **total** mass of the measures.

In analysis, **first normalize** all samples before computing HK, then recover the total mass difference either via the above equation, or keeping the total masses to be separate features.

Practical Limitations of OT Distances:

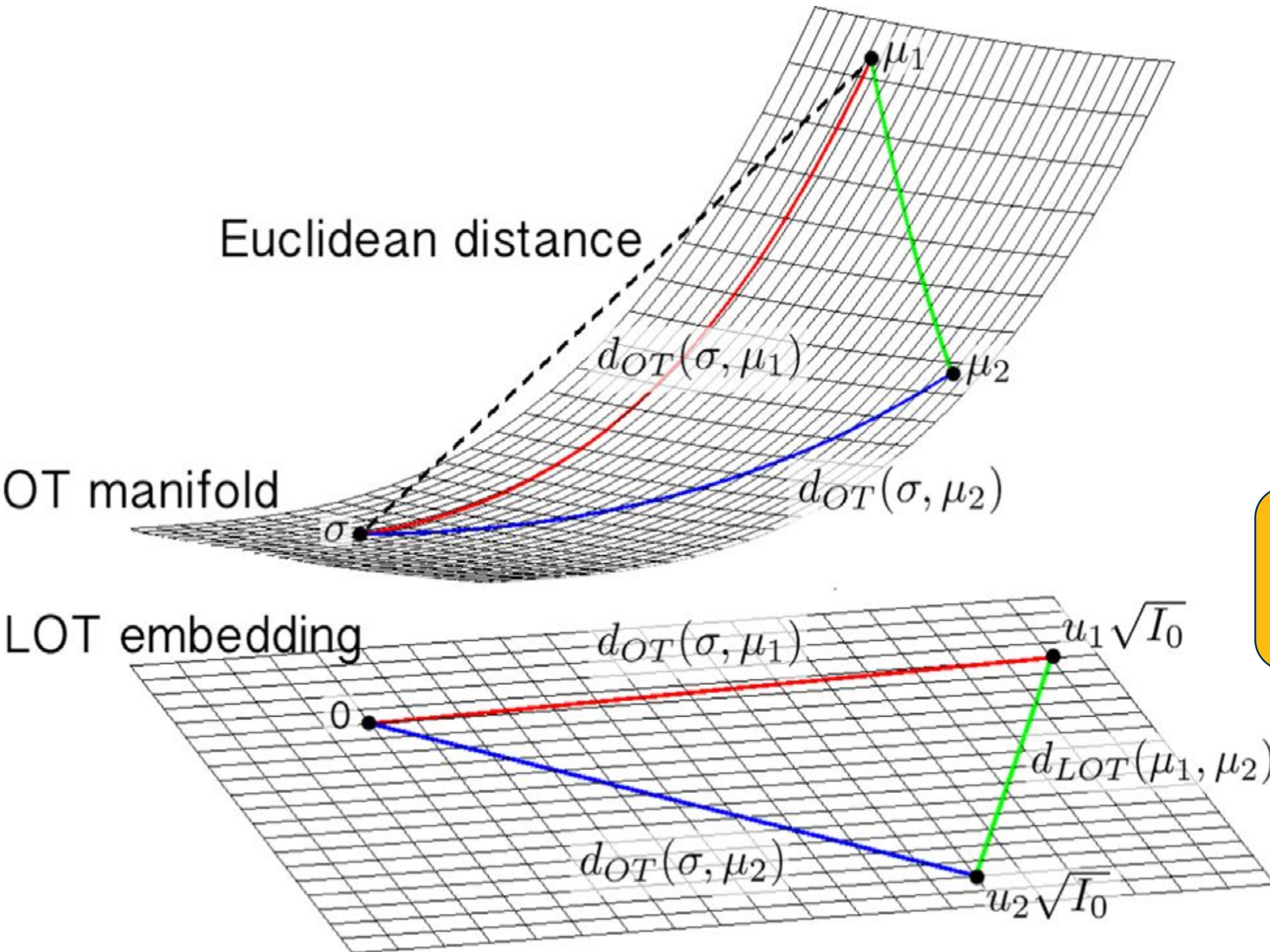
- A dataset with N samples
- T_{OT}: Time to compute one pair of OT distance (~ 0.1 secs)
- T_2: Time to compute one pair of Euclidean distance (~ 10^{-3} secs)

OT for the whole dataset takes time on the order of $N(N-1)/2 \times T_{OT}$.
=> Too long for large datasets!

Compute the OT distances between **100k** events takes **~16 years** on a desktop.

2.4 Linearized Optimal Transport

Goal: Reduce computation to N × T_{OT} + N(N-1)/2 × T_2 by using **Linearized Optimal Transport** (LOT). [Wang et al.]

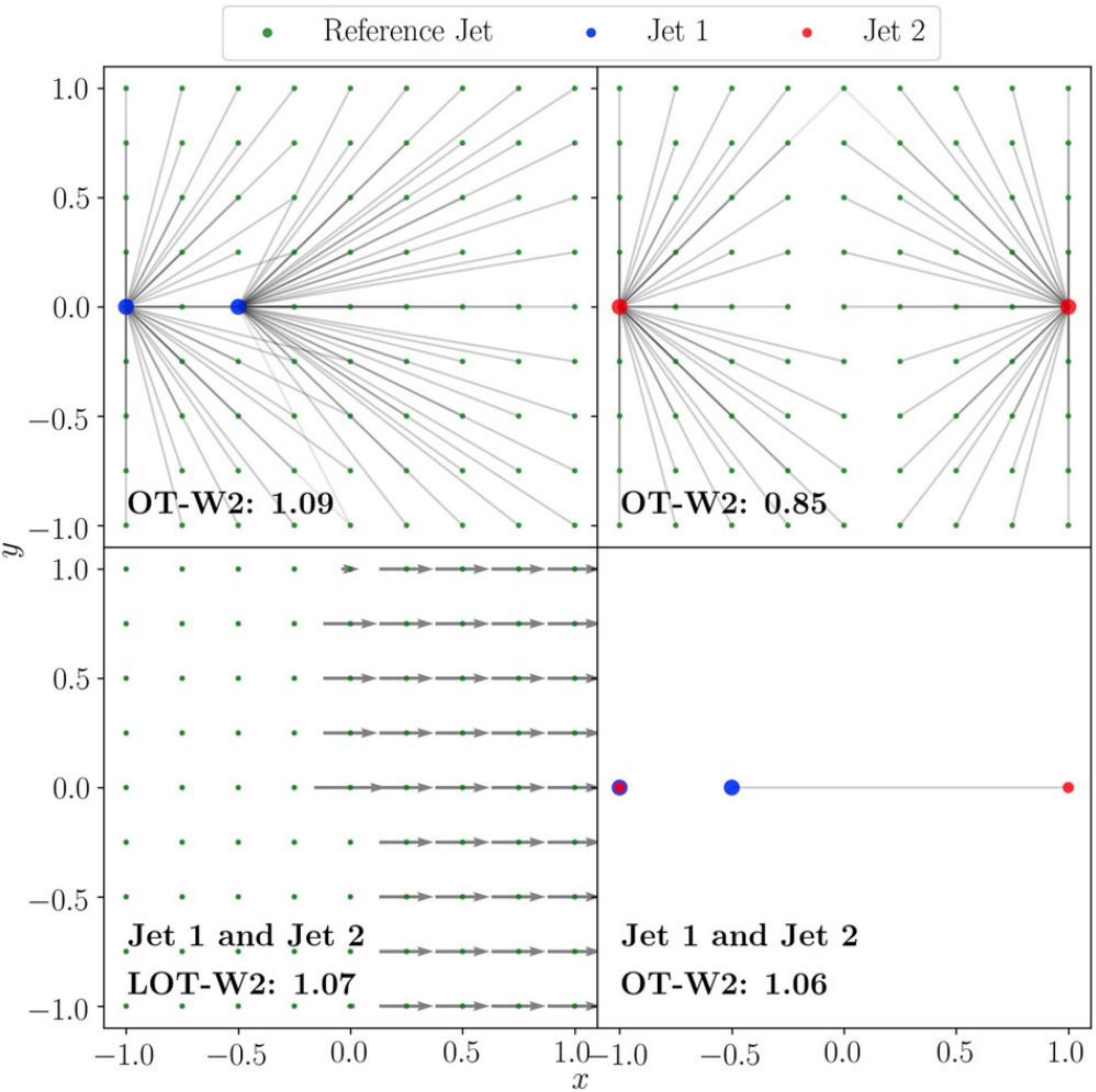


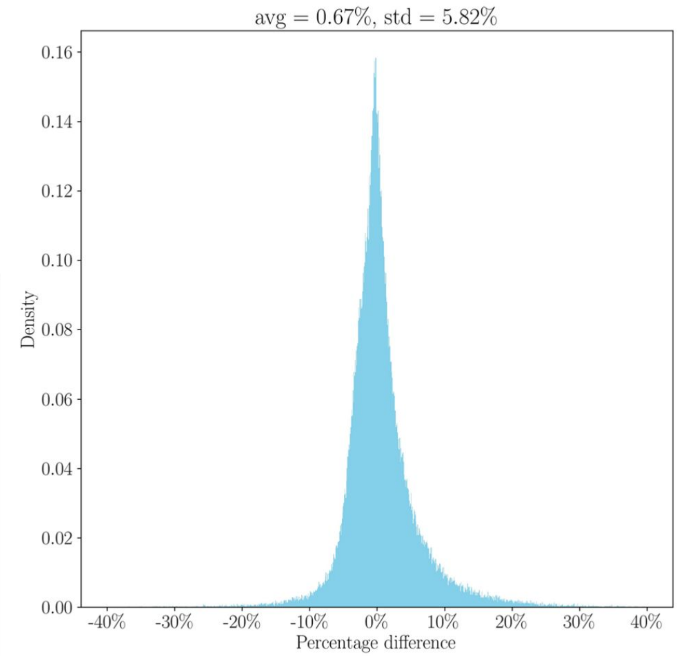
Basic Idea: Project onto the tangent plane at a chosen reference event, then compute Euclidean distances.

Now only minutes on your desktop!

Source: Fig 1, S. Kolouri et al. / Pattern Recognition 51 (2016) 453–462

A Simple Example

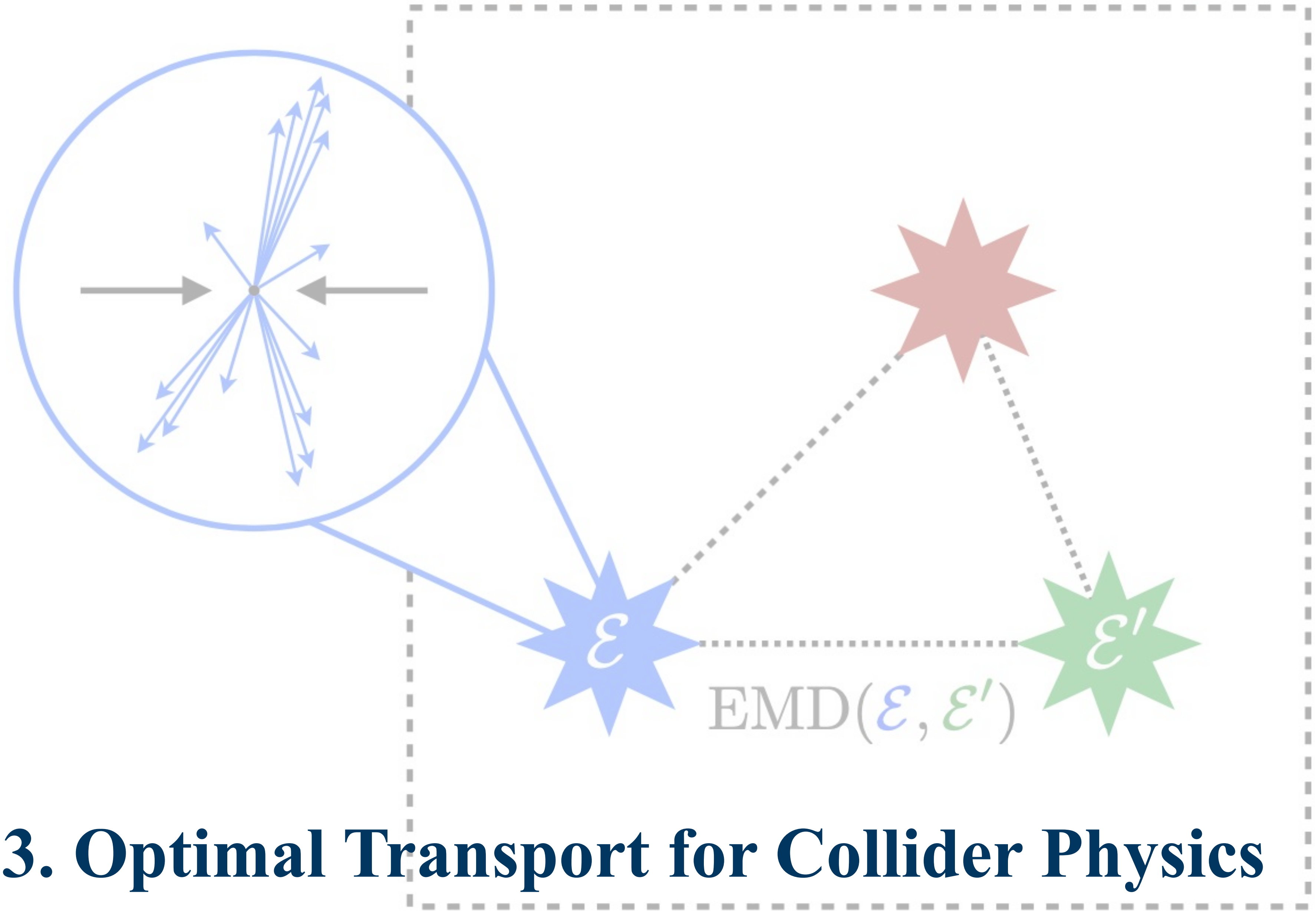




Percentage difference between linearized W_2 and W_2 distances for 500 W and QCD jets.

The linear approx for both W_2 and HK is pretty good! 👻

[Komiske, Metodiev, Thaler 2004.04159]



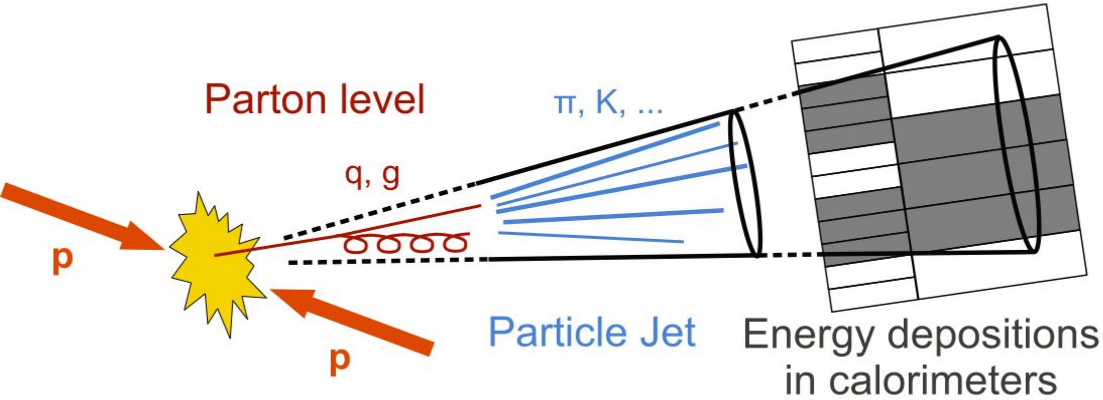
3. Optimal Transport for Collider Physics

3.1 Physics Background

Simplified Perspective: Events as energy flow on calorimeters.

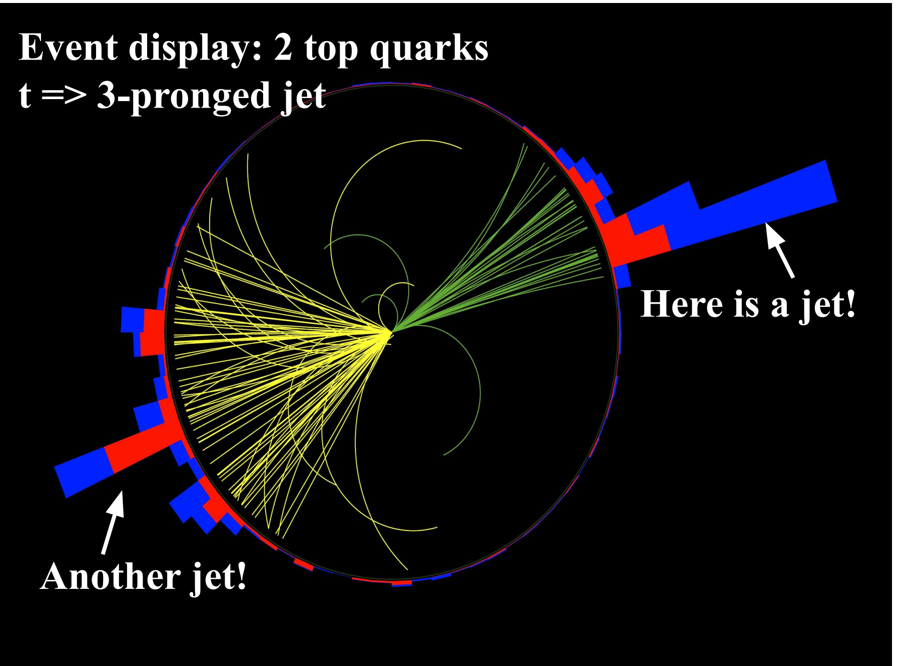
Phenomenological Tool: Jets and Jet substructure.

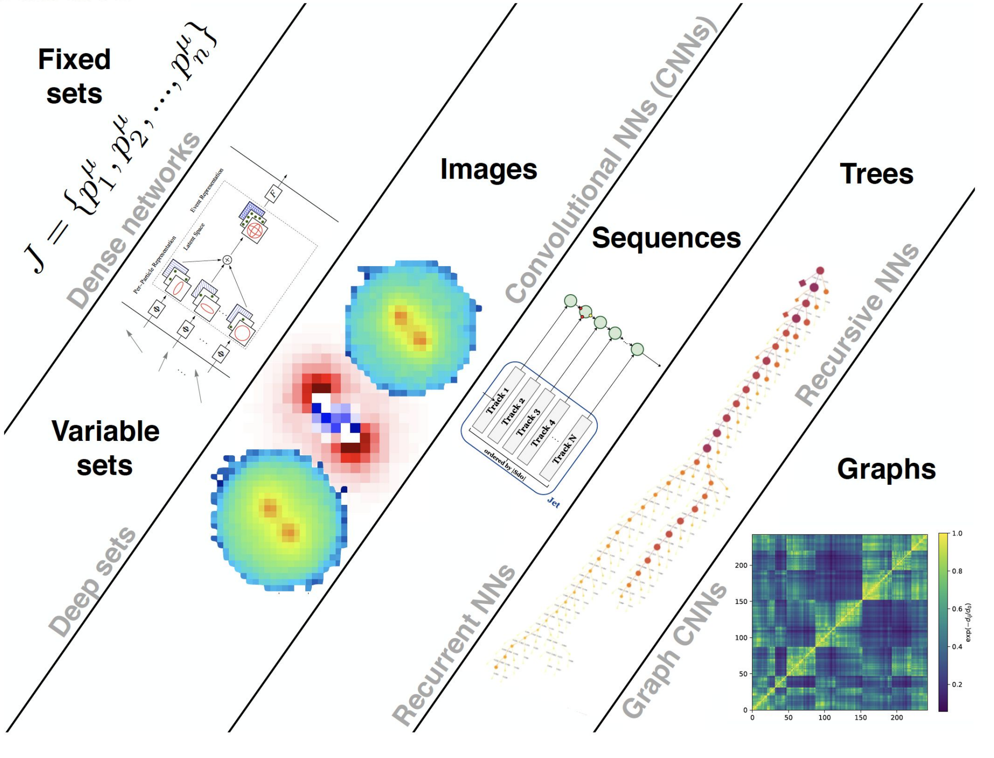
$$\mathcal{E}(\hat{n}) = \sum_{i \in J} E_i \, \delta(\hat{n} - \hat{n}_i)$$



Source: CMS website, 1709.04464v2.

Common ML Methods





Why optimal transport as the metrics on jets?

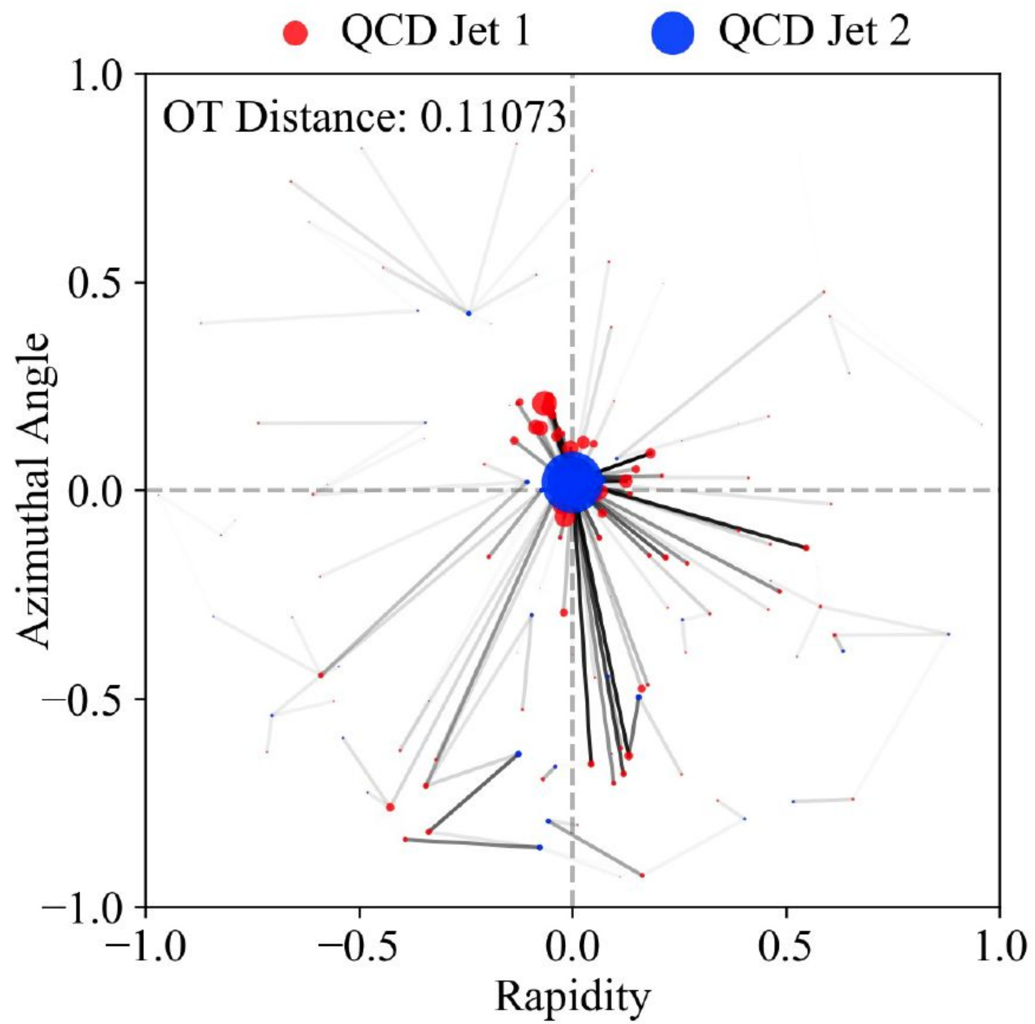
Goal: Give the space of collider events a physically meaningful metric.

Jets are discrete distributions on the y-ϕ plane. We want a distance between them. => **Of course OT!**

Motive 1: Jet tagging with fast, simple-to-use, easy-to-interpret machine learning models.

Motive 2: Unify the concepts and techniques in QFT and jet physics through the geometric language of a metricized collider space.

Related Works: Jesse Thaler's group at MIT.



3.2 A Unified Framework

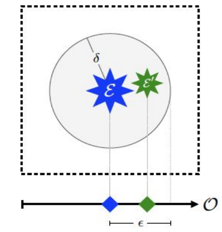
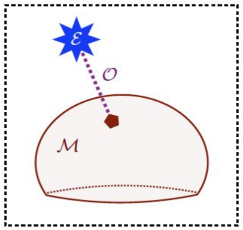
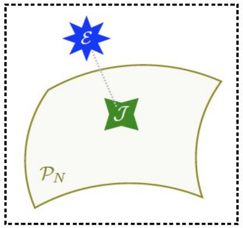
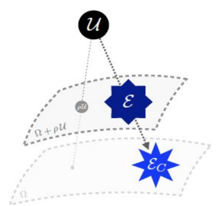
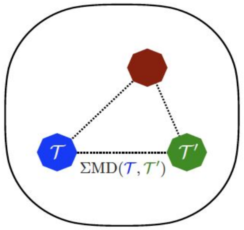
IRC safety: EMD continuity everywhere except a negligible set of events.

Collider observables: the distance of closest approach between an event and a manifold of events.

Jet algorithm: approximates an M-particle event with N < M objects called jet.

Pileup mitigation: finds the event that, when combined with an amount rho of uniform radiation U, is closest to the given event.

A distance between theories: EMD as the ground metric and cross sections as weights.

Concept	Equation	Illustration
Infrared and Collinear Safety [27–39]	$\text{EMD}(\mathcal{E},\mathcal{E}') < \delta \implies |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon$	
Observables Event Shapes [40–45] Jet Shapes [46–48]	$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}'\in\mathcal{M}} \text{EMD}(\mathcal{E},\mathcal{E}')$	
Jets Cone Finding [49, 50] Seq. Rec. [51–54]	$\mathcal{J}(\mathcal{E}) = \arg\min_{\mathcal{J}\in\mathcal{P}_N} \text{EMD}(\mathcal{E},\mathcal{J})$	
Pileup Subtraction [55–61]	$\mathcal{E}_C(\mathcal{E},\rho) = \arg\min_{\mathcal{E}'\in\Omega} \text{EMD}(\mathcal{E},\mathcal{E}'+\rho\,\mathcal{U})$	
Theory Space	$\mathcal{T}(\mathcal{E}) = \sum_{i=1}^{N} \sigma_i\,\delta(\mathcal{E}-\mathcal{E}_i)$	

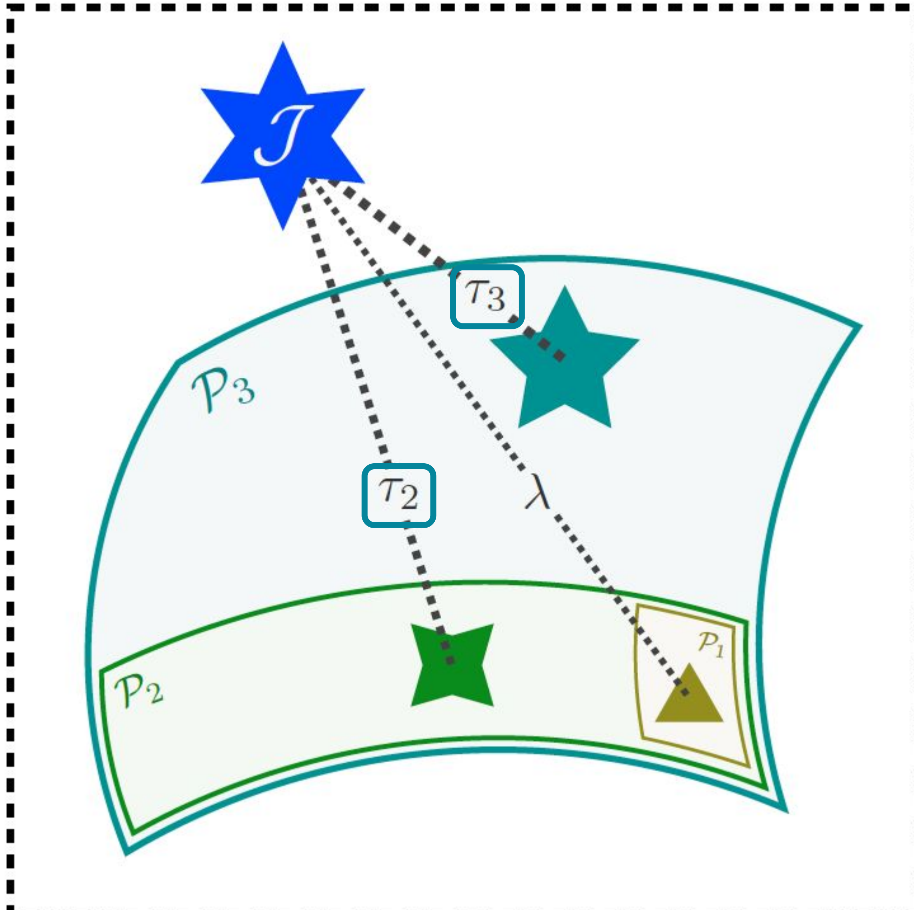
Source: Table 1, 2004.04159.

For example, *N*-subjettiness can be defined using the EMD distance.

$$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\hat{n}_1,\cdots,\hat{n}_N} \sum_{i=1}^{M} E_i \min\left(\theta_{i1}^{\beta}, \theta_{i2}^{\beta}, \cdots, \theta_{iN}^{\beta}\right)$$

$$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{J}, \mathcal{J}').$$

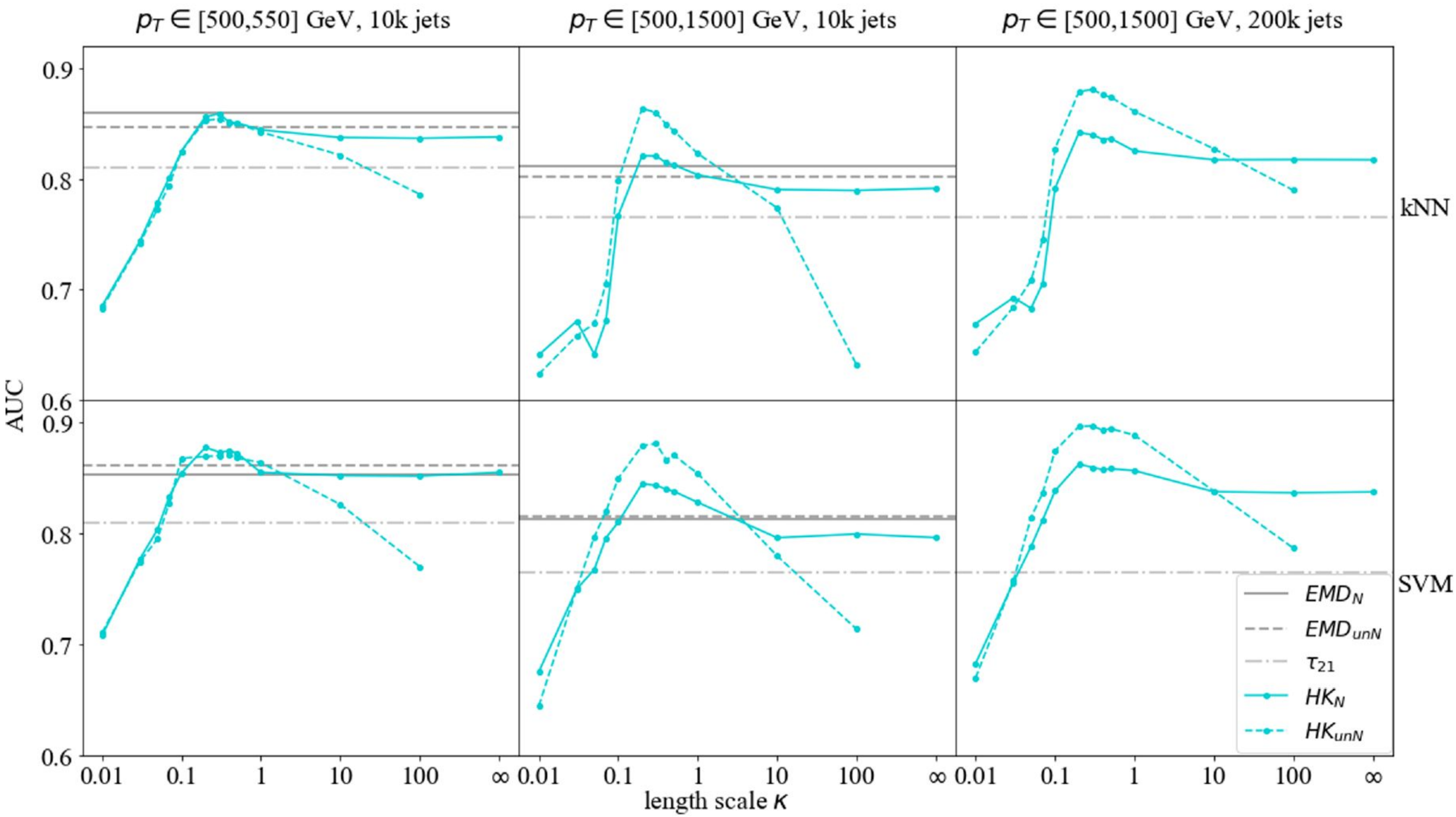
$$\tau_{N,N-1}^{(\beta)} = \frac{\tau_N^{(\beta)}}{\tau_{N-1}^{(\beta)}}$$



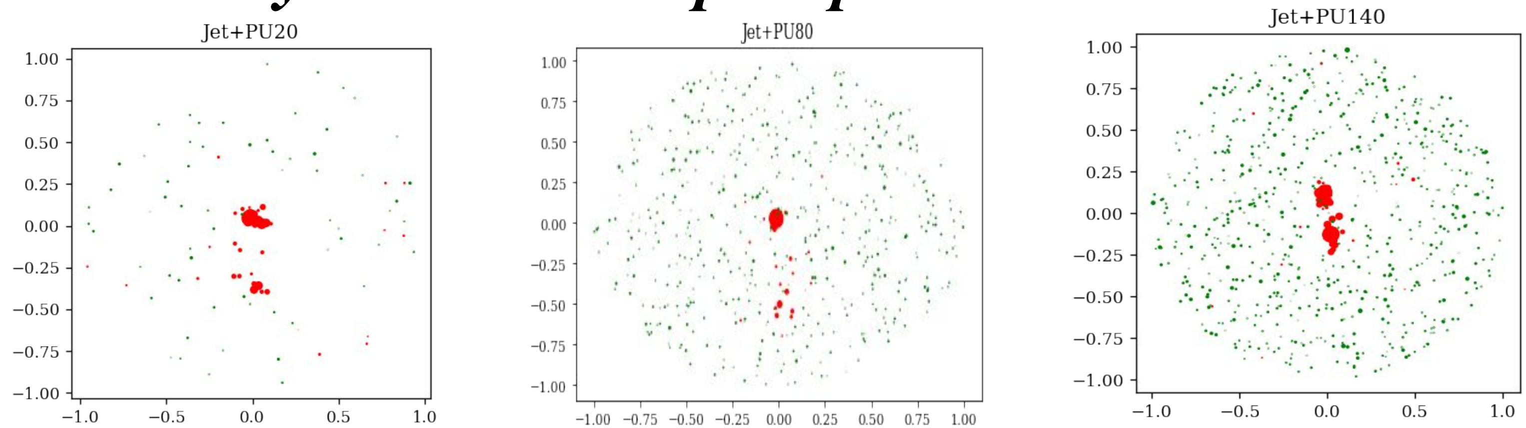
3.3 Jet Tagging

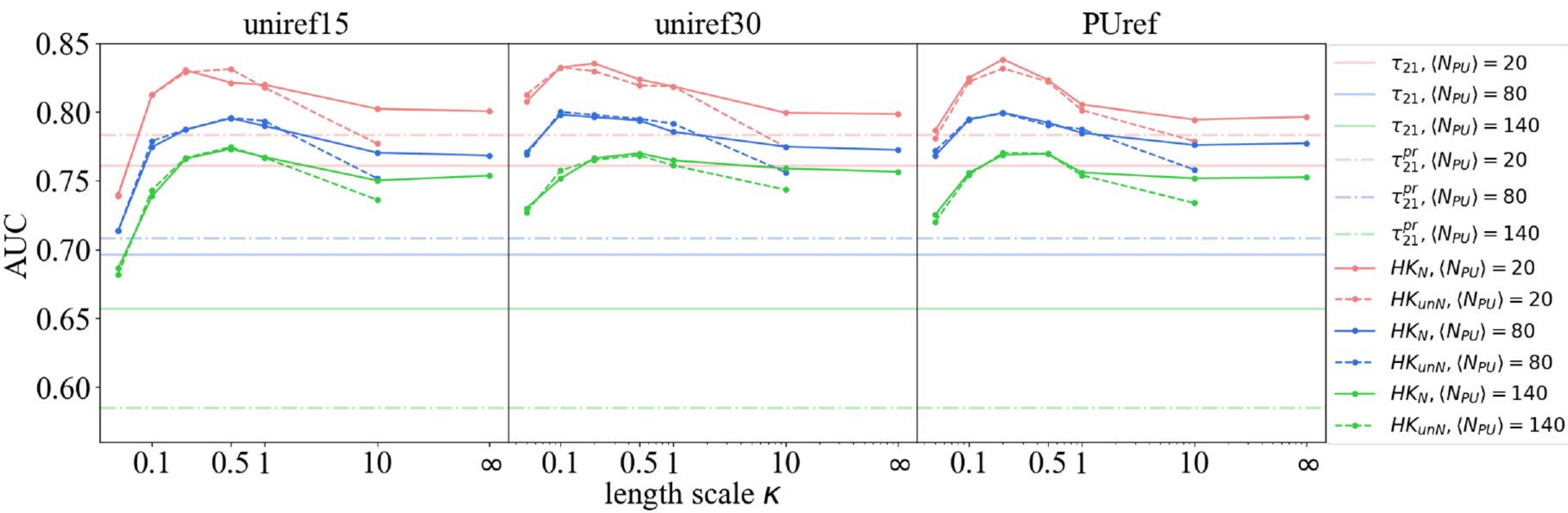
Please refer to our paper for more details.

- Task is to classify simulated QCD vs. W boson jets.
- Linearized W_2 and HK distances with various **κ**
- Apply simple ML models such as kNN and SVM to classify the jets.

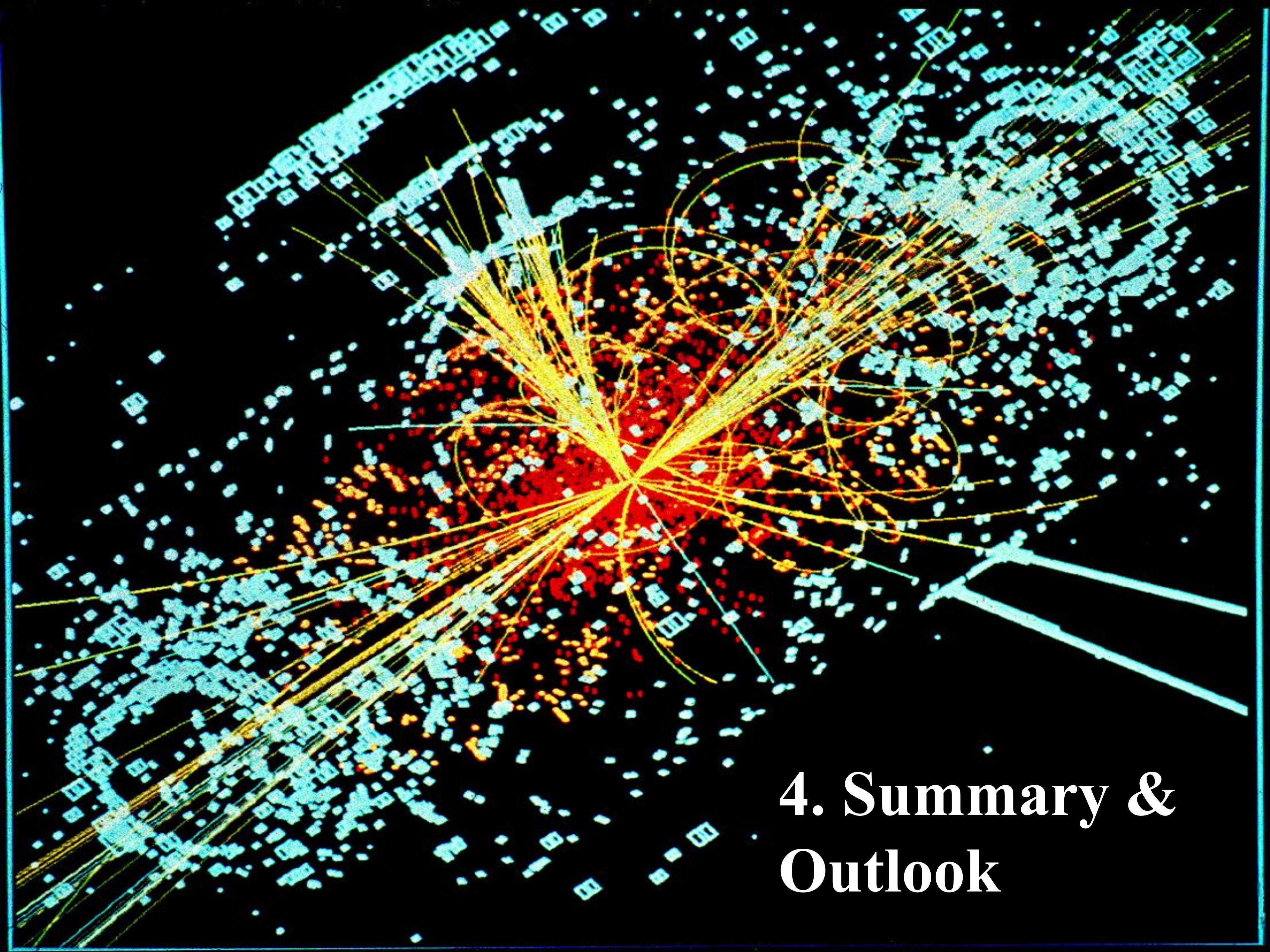


Is OT relatively insensitive to pileup?





Linearized W_2 and HK distances on 10k WQCD jets with different pileups (Poisson distributions with mean N_{PU} = ***20, 80, 140****) compared with τ_{21} on pruned jets in the same dataset.*



4. Summary & Outlook

Take-home Messages

Optimal Transport is GREAT!

Especially suitable whenever you want to define a distance between two distributions with rich substructures.

One example of such distributions is LHC jets, immensely important to collider physics.

Potentially many other use cases in physics and beyond.

OT also has many theoretical implications, e.g., Renormalization Group Flow as OT.

It's now your turn to play with OT!

Future Directions Along the Current Lines of Research

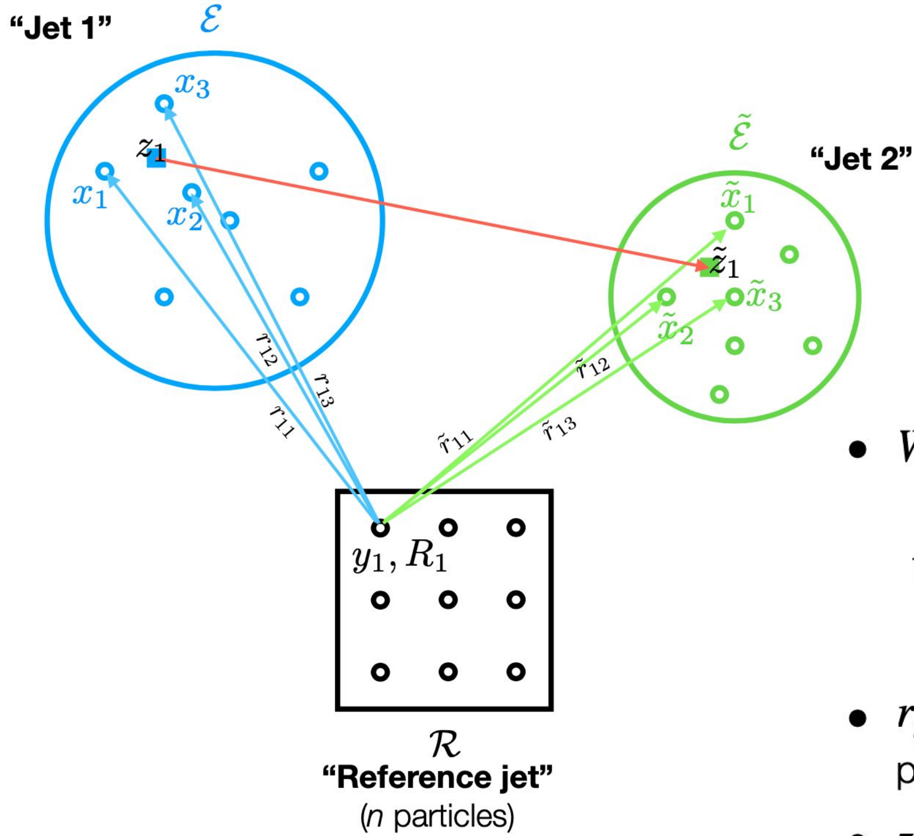
- ***More Developments of the OT framework***
  - *Multi-species OT*
  - *OT with Invariances*
  - *OT on Multi-scaled Distributions*
  - …
- ***OT for Collider Physics***
  - *OT for Anomaly Detection*
  - *Study the OT Manifold*
  - *OT for a Theory Space*
  - …
- ***OT for Other HEP Fields***
  - *OT for Dark Matter*
  - *Your turn now!*
  - …

THANKS!

Presented by Tianji Cai

tianji_cai@ucsb.edu

Backup



Procedure to Linearize the W_2 Distance

- $W_2(\mathscr{E}, \tilde{\mathscr{E}})$: **2-Wasserstein distance**

$$W_2(\mathcal{E}, \tilde{\mathcal{E}}) = \min_{r_{ij} \in \Gamma(\mathcal{E}, \tilde{\mathcal{E}})} \left(\sum_{ij} r_{ij} \|x_i - \tilde{x}_j\|^2 \right)^{1/2}$$

- r_{ij} : **transport plan** (minimizer of W_2) from *i*th particle in reference to *j*th particle in event.
- z_i : **barycenter** (avg. of locations to which *i*th particle is sent, weighted by transport plan)

$$z_i = \frac{1}{R_i} \sum_j r_{ij} x_j$$

Map from $\mathscr{E}$ to a vector in $\mathbb{R}^{2n}$

- $LOT_{r,\tilde{r}}(\mathscr{E}, \tilde{\mathscr{E}})$: LOT approximation of the 2-Wasserstein distance between events $\mathscr{E}$, $\tilde{\mathscr{E}}$

$$LOT_{r,\tilde{r}}(\mathcal{E}, \tilde{\mathcal{E}}) = \left(\sum_i R_i \|z_i - \tilde{z}_i\|^2 \right)^{1/2}$$

Can think of LOT as an approximation to W_2 distance, but also converges to a true metric $W_{2,\mathscr{R}}(\mathscr{E}, \tilde{\mathscr{E}})$ in its own right in continuum limit of reference $\mathscr{R}$; this metric bounded as

$$W_2(\mathcal{E}, \tilde{\mathcal{E}}) \leq W_{2,\mathcal{R}}(\mathcal{E}, \tilde{\mathcal{E}}) \leq C W_2(\mathcal{E}, \tilde{\mathcal{E}})^{2/15}$$

Linearizing the HK distances is similar, but more complicated.

Source Distribution: $\mu_0 = u_0 \cdot \sigma + \mu_0^\perp$

Target Distribution: $\mu_1 = u_1 \cdot T_\sharp \sigma + \mu_1^\perp$

Two Distributions have different total mass.

$\mu_1^\perp$: Mass of particles that are created from nothing.

Spatial movement of mass particles: $v_t := \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$

Change of mass of moving particles and of those that disappear entirely at t=1: $\alpha_t := \frac{\mathrm{d}\tilde{\zeta}_t}{\mathrm{d}\rho_t} - 2(1-t)\frac{\mathrm{d}\mu_0^\perp}{\mathrm{d}\rho_t}$

$$v_0(x) = \begin{cases} \frac{T(x)-x}{\|T(x)-x\|} \cdot \sqrt{\frac{u_1(T(x))}{u_0(x)}} \cdot \sin(\|T(x)-x\|) & \sigma\text{-a.e.}, \\ 0 & \mu_0^\perp\text{-a.e.}, \end{cases}$$

$$\alpha_0(x) = \begin{cases} 2\left(\sqrt{\frac{u_1(T(x))}{u_0(x)}} \cdot \cos(\|T(x)-x\|) - 1\right) & \sigma\text{-a.e.}, \\ -2 & \mu_0^\perp\text{-a.e.}. \end{cases}$$

$$\mathrm{HK}(\mu_0, \mu_1)^2 = \int_\Omega \left[\|v_0\|^2 + \tfrac{1}{4}(\alpha_0)^2\right] \mathrm{d}\mu_0 + \|\mu_1^\perp\|.$$

The tangent space now consists of a velocity field and a mass growth field.

For detailed descriptions, please refer to our math paper.