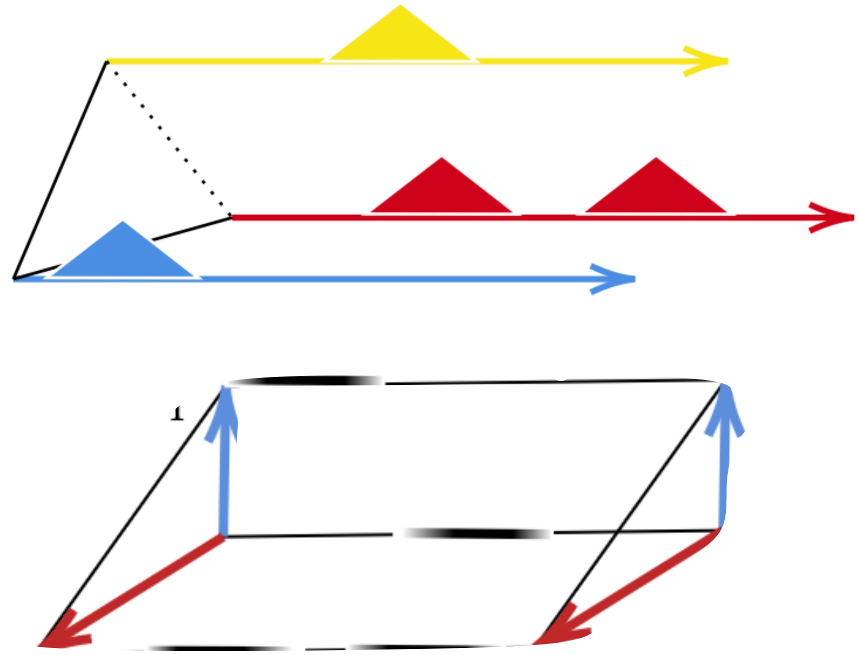
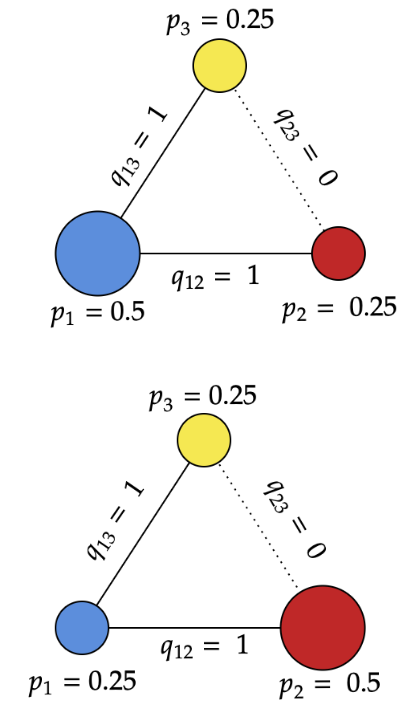




Vector Valued Optimal Transport:

From Dynamic to Static Formulations

Katy Craig
University of California, Santa Barbara

joint with Nicolas García Trillos (Wisconsin) and Đorđe Nicolić (UCSB)

Dynamics of Density Operators
IPAM
April 28, 2025

Plan

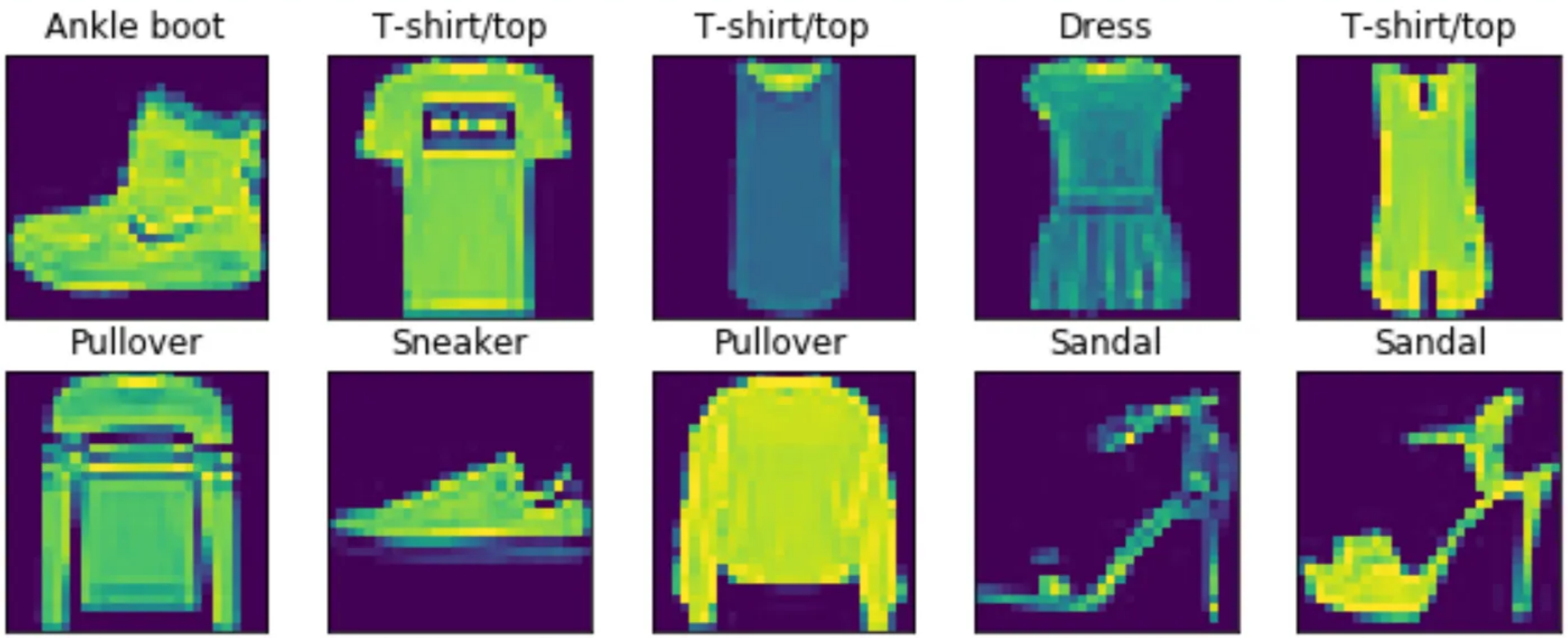
1. Motivation for vector valued OT

2. Previous work on vector valued and graph OT

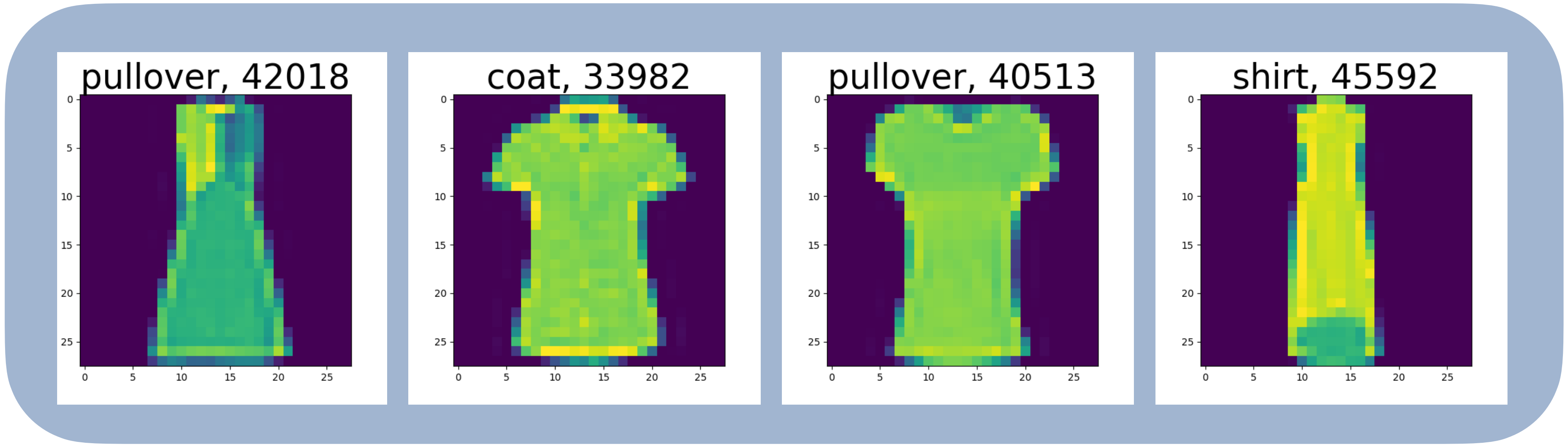
3. New dynamic and static (semi)-metrics

4. Metric comparison

Motivation: image classification



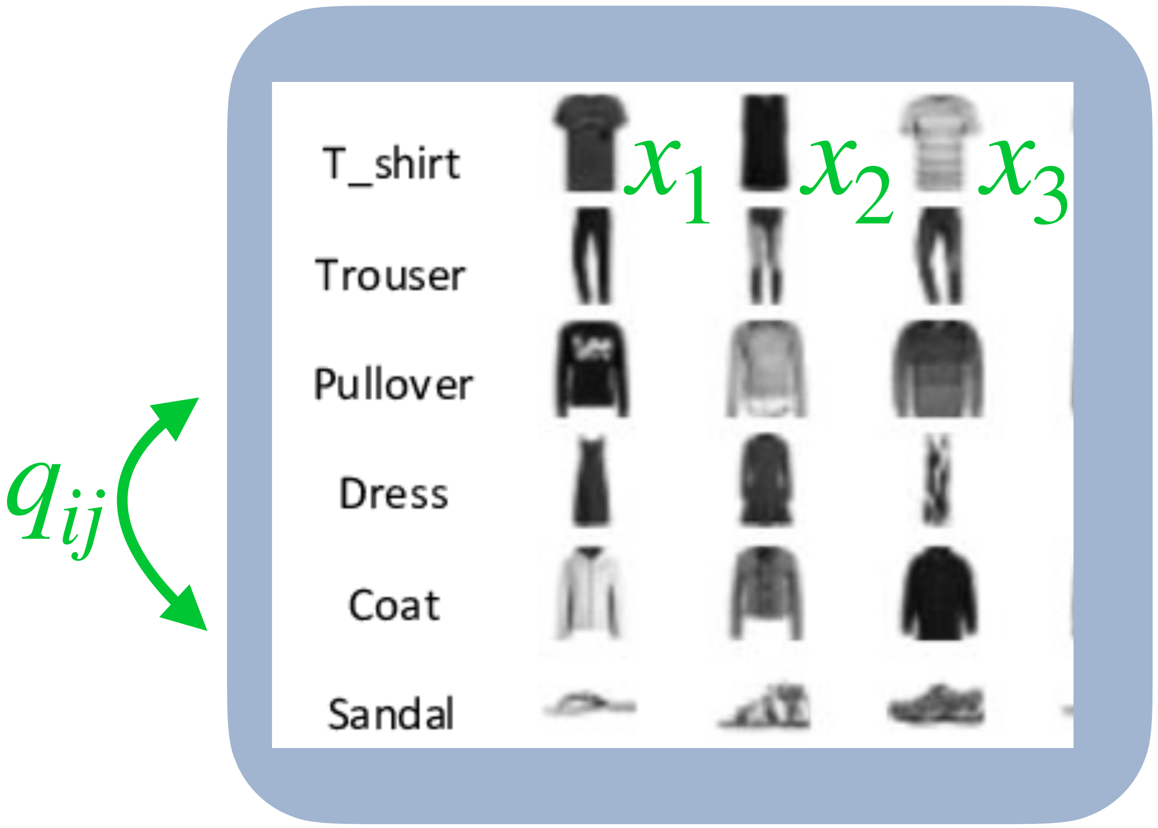
Mild misclassifications in Fashion MNIST— for example…



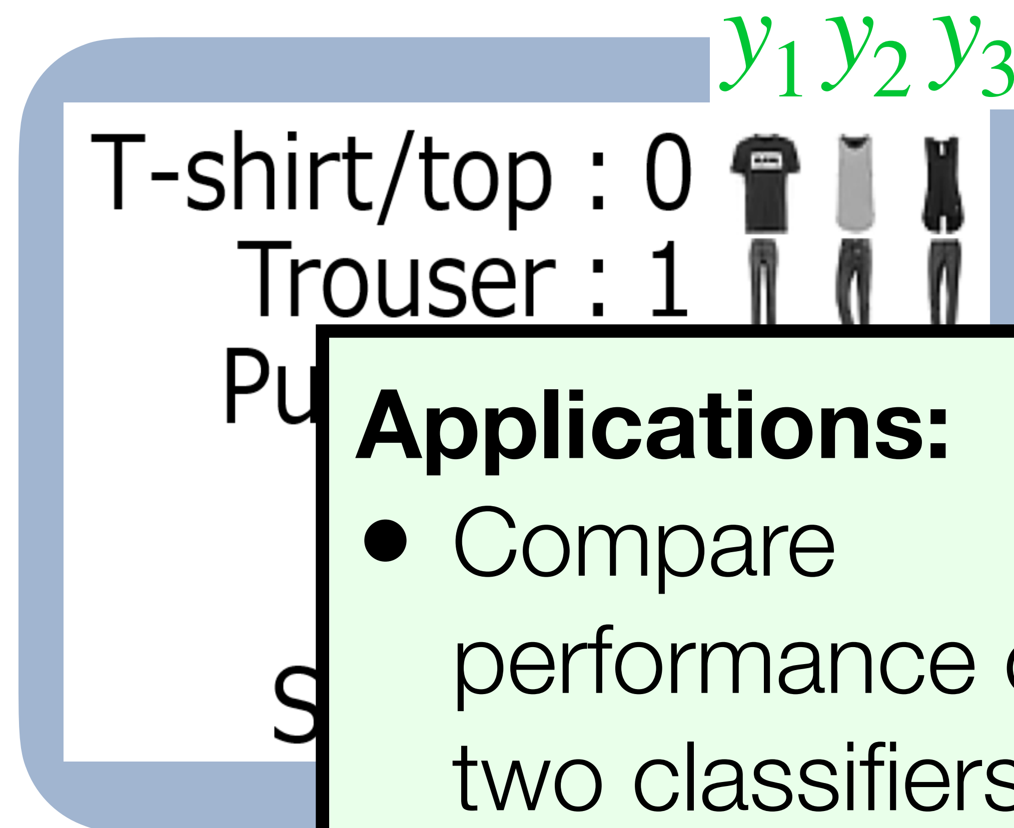
[Müller, Markert 2019]

Motivation: image classification

Labeled Dataset #1



Labeled Datas



Applications:

- Compare performance of two classifiers
- Compare structure of two labeled datasets (ATLAS/CMS)

How different are these datasets?

Represent images as points $x_k, y_l \in \mathbb{R}^d$

Similarity between images $= ||x_k - y_l||$

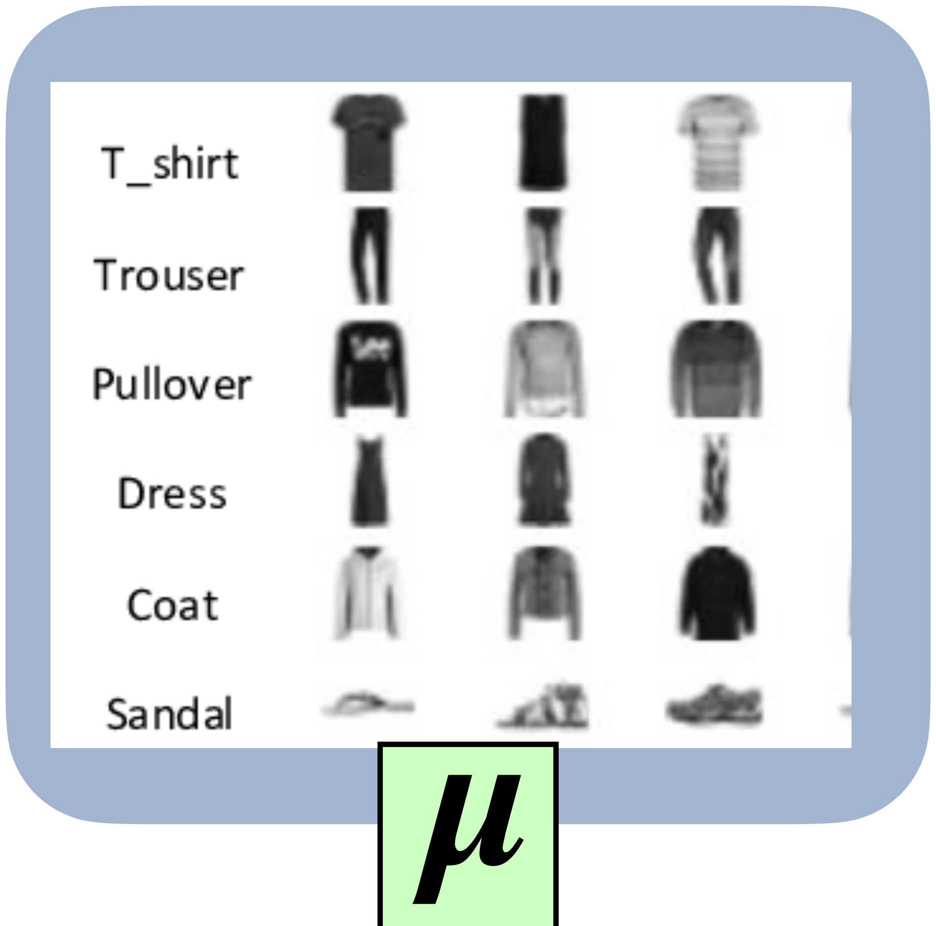
Dissimilarity between label i and label $j = q_{ij}$

Datasets are similar if their images & labels are similar.

Motivation: image classification

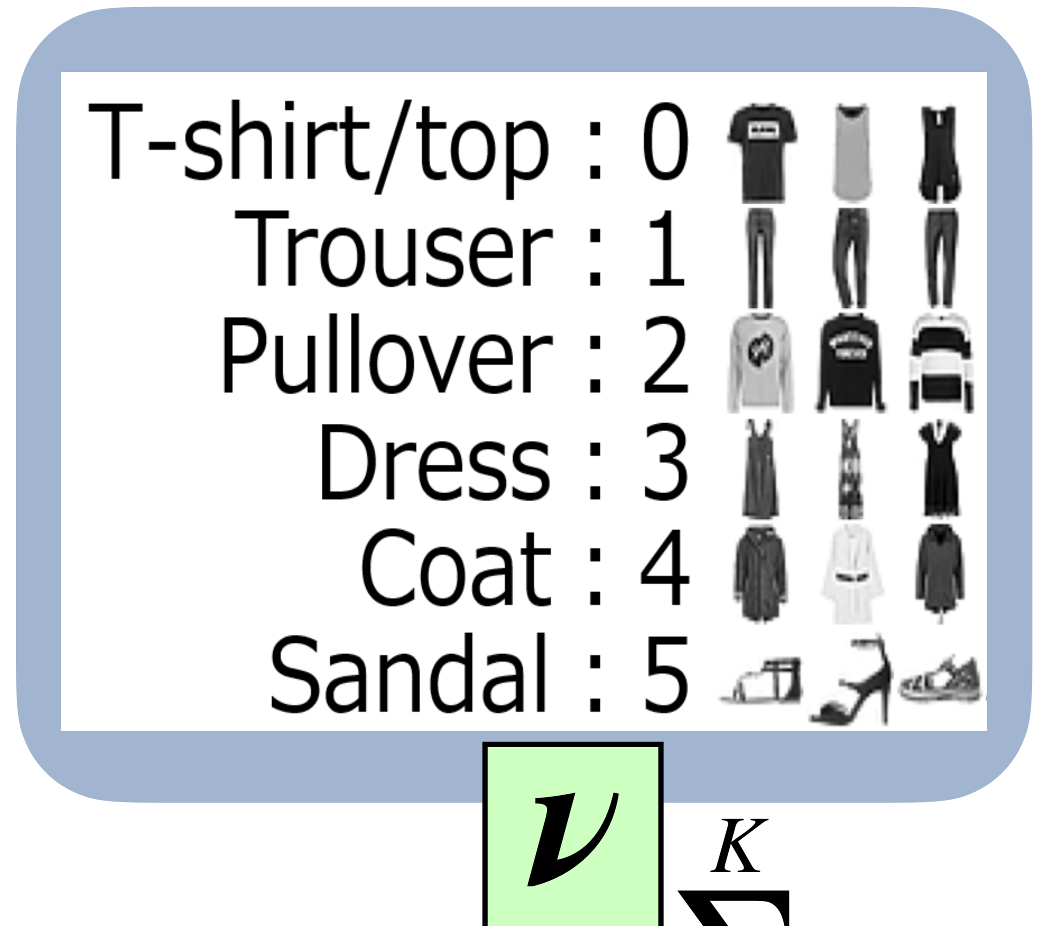
Labeled Dataset #1

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{bmatrix}$$



Labeled Dataset #2

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix}$$



Represent collection of images with label i as $\mu_i = \sum_{k=1}^{K} \delta_{x_k} m_k$

Represent datasets as vector valued measures $\boldsymbol{\mu} = [\mu_i]_{i=1}^{n}$

Datasets are similar if $\boldsymbol{d}(\boldsymbol{\mu}, \boldsymbol{\nu}) \ll 1$ for some distance **d**.

Motivation: gradient flows

Given a target vector valued measure $\boldsymbol{\mu}$, a loss functional $\mathscr{L}$, and a constraint set C, find a minimizer of

$$\min_{\rho \in C} \mathscr{L}(\boldsymbol{\rho}, \boldsymbol{\mu}) \,.$$

Given a distance **d** between vector valued measures, could flow to critical points via **gradient descent**. In discrete time,

$$\boldsymbol{\rho}_\tau^{n+1} = \operatorname{argmin}_{\rho \in C} \mathscr{L}(\boldsymbol{\rho}, \boldsymbol{\mu}) + \frac{1}{2\tau} \boldsymbol{d}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_\tau^n),$$

which, for nice $\mathscr{L}$, is stable wrt perturbations in **d**.

For classification or gradient flows, what is an appropriate notion of distance **d** between vector valued measures?

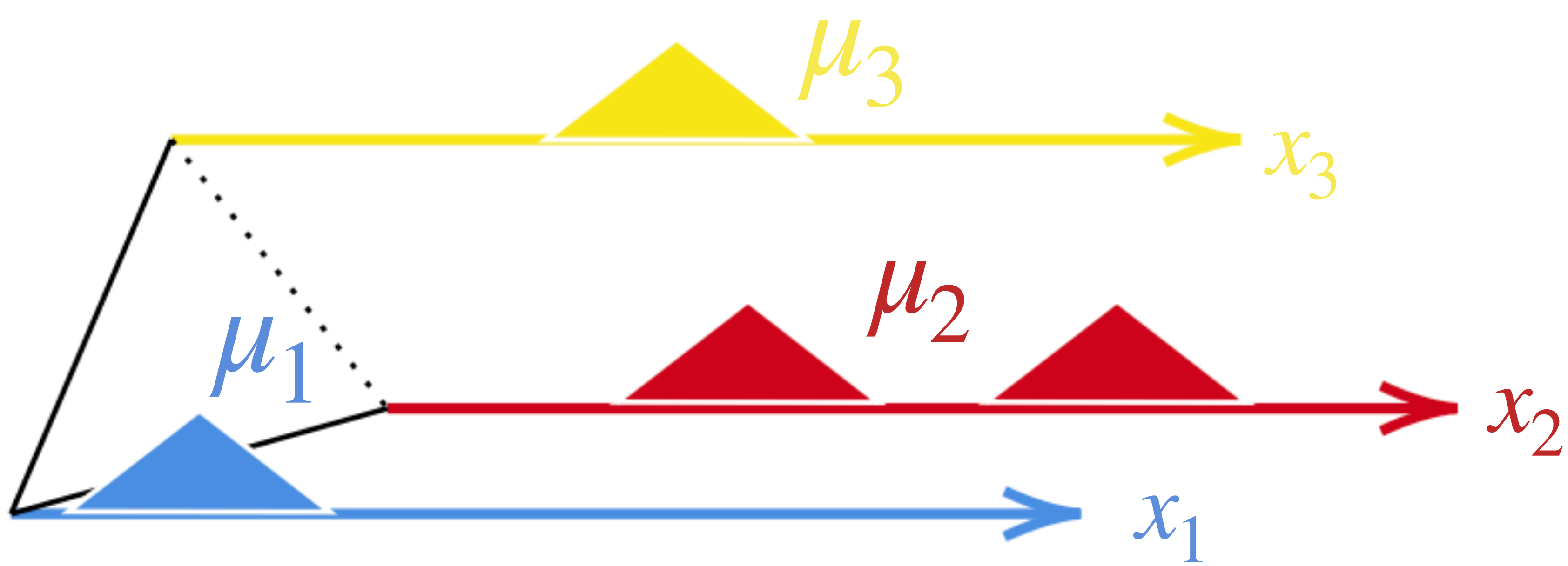
Vector valued measures

Fix $\boldsymbol{\Omega} \subseteq \mathbb{R}^d$ closed. Let $\boldsymbol{G}$ denote an $\boldsymbol{n}$ node graph.

Consider vector valued measures with total mass one:

$$\mathscr{P}(\Omega \times G) := \left\{ \boldsymbol{\mu} = [\mu_i]_{i=1}^n : \mu_i \in \mathscr{M}(\Omega), \underbrace{\sum_{i=1}^n \mu_i}_{=: \bar{\mu}} \in \mathscr{P}(\Omega) \right\}$$

For example, when $\boldsymbol{\Omega} = \mathbb{R}$ and $\boldsymbol{G}$ has three nodes:



Plan

1. Motivation for vector valued OT

2. Previous work on vector valued and graph OT

3. New dynamic and static (semi)-metrics

4. Metric comparison

Previous work on vector valued OT

There are several existing notions of distance on $\mathscr{P}(\Omega \times G)$.

[Chen, Georgiou, Tannenbaum '18]: define dynamic distance via a product space structure on $(W_G, \mathscr{P}(G))$ & $(W_2, \mathscr{P}(\mathbb{R}^d))$

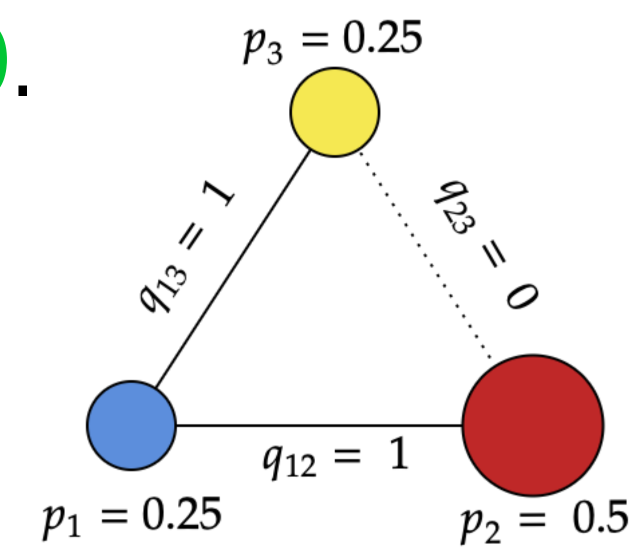
[Bacon '20]: define static distance via Kantorovich formulation; transport plans γ_{ij} move mass from μ_i to ν_j at cost $c_{ij}(x, y)$; prove strong duality; compatibility of c_{ij} to ensure metric.

Also W_1-type metrics [Todeschi, Metvier, Mirebeau '25], [Ryu, Chen, Li, Osher '18].

We seek a W_2-type distance on $\mathscr{P}(\Omega \times G)$ that unites dynamic and static perspectives, with a (formal) Riemannian structure for gradient flows and linearization.

Graph Wasserstein metric

Let G denote an n node weighted graph, where the weight of the edge connecting node i to node j is $q_{ij} \geq 0$.



$$\mathscr{P}(G) = \left\{ p = \sum_{i=1}^{n} p_i \delta_i : p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

$$W_G^2(p, \tilde{p}) := \min \frac{1}{2} \sum_{i,j=1}^{n} \int_0^1 |v_{ij,t}|^2 \theta(p_{i,t}, p_{j,t}) q_{ij} dt$$

$$\text{such that} \quad \partial_t p_t + \nabla_G \cdot (\check{p}_t v_t) = 0,\ p_0 = p,\ p_1 = \tilde{p}$$

$$\check{p}_{ij} v_{ij} := \theta(p_i, p_j) v_{ij}$$

Examples: $\theta(p_i, p_j) = (p_i + p_j)/2,$ $\sqrt{p_i p_j},$ or $\displaystyle\int_0^1 p_i^{1-s} p_j^s ds$

Graph Wasserstein metric

Thm [Maas '11]: For $\mathscr{G}$ connected, $(\mathscr{P}_{>0}(G), W_G)$ is a smooth Riemannian manifold.

Rmk: The manifold is not complete, and the geodesic between two points in the interior can touch the boundary [Gangbo, Li, Mou '19], so branching is possible.

To see role of q_{ij}, absorb weights into velocity: for $q_{ij} > 0$,

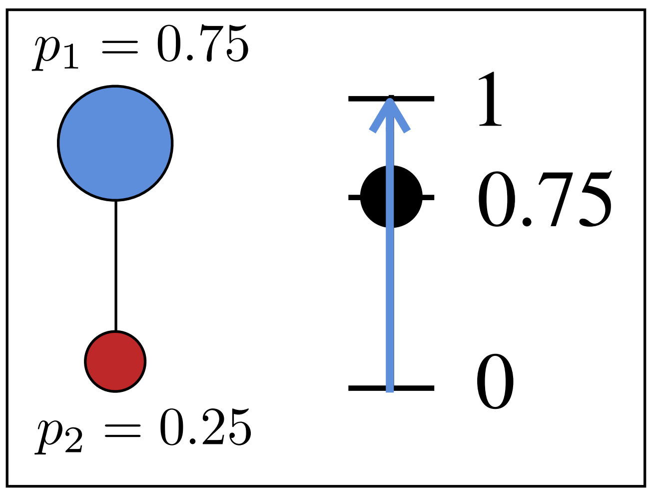
$$W_G^2(p, \tilde{p}) = \min \frac{1}{2} \sum_{i,j=1}^{n} \int_0^1 |v_{ij,t}|^2 \theta(p_{i,t}, p_{j,t}) q_{ij}^{-1} dt$$

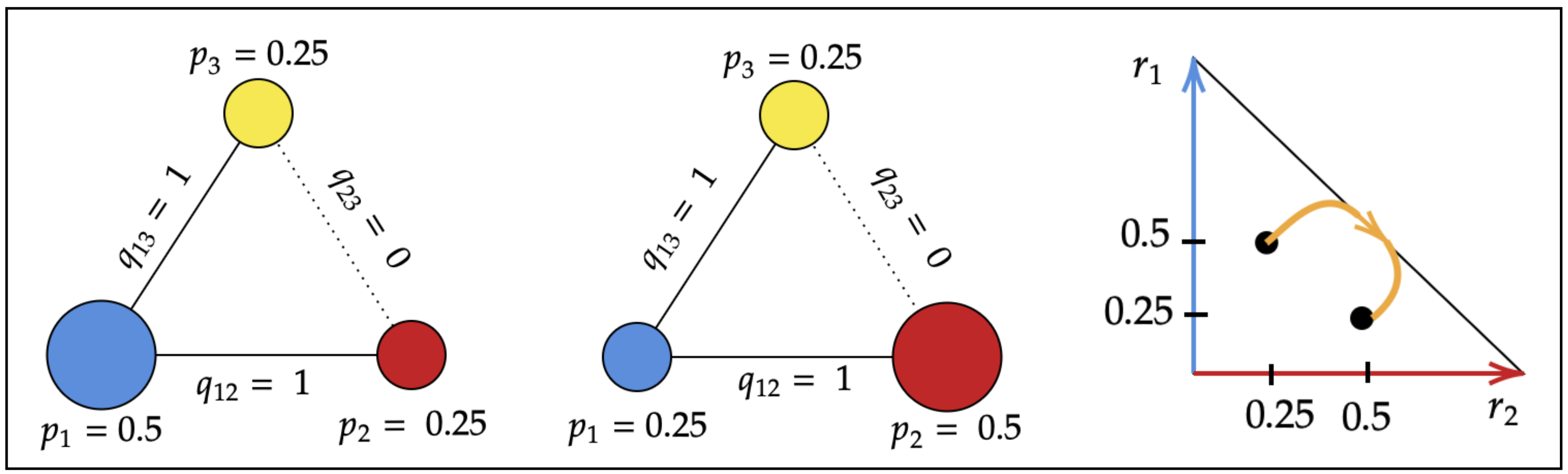
$$\text{such that} \quad \partial_t p_t + \nabla_{[n]} \cdot (\check{p}_t v_t) = 0, \; p_0 = p, \; p_1 = \tilde{p}$$

Induced geometry on simplex

Via the isometry $\boldsymbol{p} : \Delta^{n-1} \to \mathscr{P}(G) : r \mapsto \left[r_1, r_2, \ldots, 1 - \Sigma_i r_i\right]$,

$$\Delta^{n-1} = \left\{ r \in \mathbb{R}_+^{n-1} : \sum_{i=1}^{n} r_i \leq 1 \right\} \cong \mathscr{P}(G)$$





Induced distance: $d_{\Delta^{n-1}}(r, \tilde{r}) = W_G(\boldsymbol{p}(r), \boldsymbol{p}(\tilde{r}))$.

Prop: [c.f. Maas, Erbar '12]
$(\Delta^{n-1}, d_{\Delta^{n-1}})$ is topologically equivalent to $(\Delta^{n-1}, \|\cdot\|_{\mathbb{R}^{n-1}})$;
on $(\Delta^{n-1})^{\circ}$, it is a smooth Riemannian manifold.

Plan

1. Motivation for vector valued OT

2. Previous work on vector valued and graph OT

3. New dynamic and static (semi)-metrics

4. Metric comparison

Continuity equation

$$\theta(\rho_i, \rho_j) := \theta\left(\frac{d\rho_i}{d\bar{\rho}}, \frac{d\rho_j}{d\bar{\rho}}\right)\bar{\rho}$$

Given $\boldsymbol{\Omega} \subseteq \mathbb{R}^d$ closed, convex … graph, we consider the following continuity equation on $\boldsymbol{\Omega} \times G$:

For velocities $\boldsymbol{u}_t(x) = [u_{i,t}(x)]_{i=1}^n$ and $\boldsymbol{v}_t(x) = [v_{ij,t}(x)]_{i,j=1}^n$,

$$\partial_t \boldsymbol{\rho} + \nabla \odot (\boldsymbol{\rho} \odot \boldsymbol{u}) + \nabla_G \cdot (\check{\boldsymbol{\rho}} \boldsymbol{v}) = 0$$

$$(\boldsymbol{\rho} \odot \boldsymbol{u})_i(x) = \rho_i(x) u_i(x) \qquad (\check{\boldsymbol{\rho}} \boldsymbol{v})_{ij}(x) = \theta(\rho_i(x), \rho_j(x)) v_{ij}(x)$$

For $\boldsymbol{v}$ symmetric, ith component of $\boldsymbol{\rho}_t(x) = [\rho_{i,t}(x)]_{i=1}^n$ satisfies

$$\partial_t \rho_{i,t} + \nabla \cdot (u_{i,t} \rho_{i,t}) = \sum_{j=1}^n \theta(\rho_{i,t}, \rho_{j,t}) v_{ij,t} q_{ij}$$

Let $\mathscr{C}(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}})$ denote the solutions satisfying $\boldsymbol{\rho}_0 = \boldsymbol{\rho}$, $\boldsymbol{\rho}_1 = \tilde{\boldsymbol{\rho}}$.

Dynamic metric

Given a solution of the continuity eqn, we consider the action

$$\|(\boldsymbol{u},\boldsymbol{v})\|_{\boldsymbol{\rho}}^2 := \sum_{i=1}^{n}\int_\Omega |u_i|^2\rho_i + \sum_{i,j=1}^{n}\int_\Omega |v_{ij}|^2\theta(\rho_i,\rho_j)q_{ij}$$

This leads to the dynamic distance:

$$W_{\Omega\times G}^2(\boldsymbol{\mu},\boldsymbol{\nu}) = \inf\int_0^1 \|(\boldsymbol{u}_t,\boldsymbol{v}_t)\|_{\boldsymbol{\rho}_t}^2 dt$$

$$\text{such that } (\boldsymbol{\rho},\boldsymbol{u},\boldsymbol{v}) \in \mathscr{C}(\boldsymbol{\mu},\boldsymbol{\nu})$$

Thm: [C., García Trillos, Nikolic '25]: $W_{\Omega\times G}$ is a metric on $\mathscr{P}_2(\Omega\times G)$ and a minimizer exists.

Dynamic metric comparison

Continuity equation: $\partial_t \boldsymbol{\rho} + \nabla \cdot (\boldsymbol{\rho} \odot \boldsymbol{u}) + \nabla_G \cdot (\check{\boldsymbol{\rho}} \boldsymbol{v}) = 0$

<u>C., García Trillos, Nikolic</u>

$$(\check{\boldsymbol{\rho}} \boldsymbol{v})_{ij} = \theta(\rho_i, \rho_j) v_{ij}$$

<u>Chen, Georgiou, Tannenbaum</u>

$$(\check{\boldsymbol{\rho}} \boldsymbol{v})_{ij} = \rho_i (v_{i,j})_+ - \rho_j (v_{ij})_-$$

Action: $\|(\boldsymbol{u}, \boldsymbol{v})\|_{\boldsymbol{\rho}}^2 := \sum_i \int |u_i|^2 \rho_i + (\mathrm{II})$

$$(\mathrm{II}) = \sum_{i,j=1}^{n} \int_\Omega |v_{ij}|^2 \theta(\rho_i, \rho_j) q_{ij}$$

$$(\mathrm{II}) = \sum_{i,j=1}^{n} \int_\Omega (v_{ij})_+^2 \rho_i q_{ij}$$

Thm [Chen, Georgiou, Tannenbaum '18], [Esposito, Patacchini, Schlichting, Slepčev '21]: The CGT formulation is not symmetric and leads to a **Finslerian structure**.

From dynamic to static

Potential analogy with Hellinger-Kantorovich?

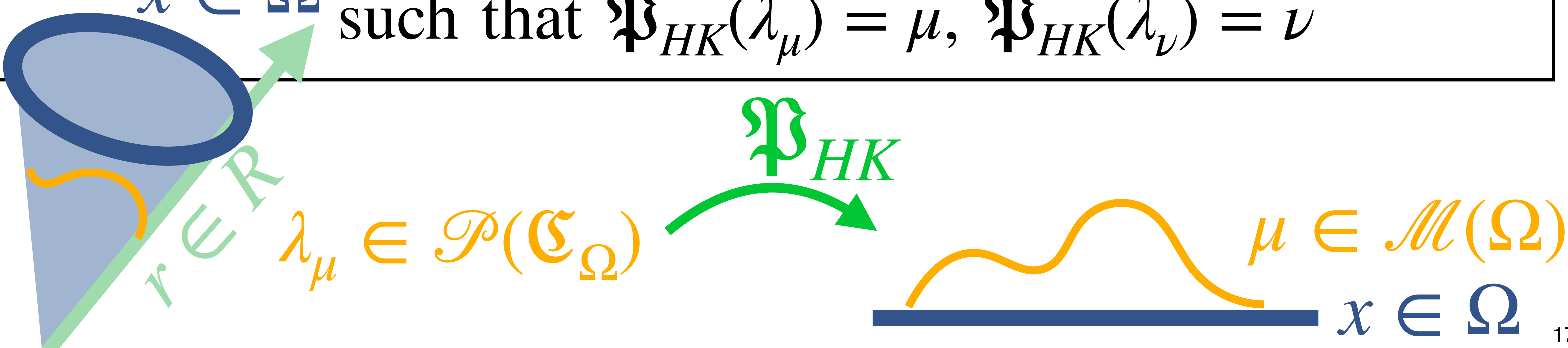
$$HK^2(\mu,\nu) = \inf \int_0^1 \int_\Omega (|v_t|^2 + 4\xi_t^2)\rho_t dt$$

$$\text{such that } \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 4\rho_t \xi_t, \quad \rho_0 = \mu, \rho_1 = \nu$$

Thm [Liero, Mielke, Savaré '16]: Given the projection operator, $\mathfrak{P}_{HK} : \mathscr{P}(\mathfrak{C}_\Omega) \to \mathscr{M}(\Omega) : \lambda \mapsto \pi_\Omega \#(r^2 d\lambda(x,r))$, we have

$$HK(\mu,\nu) = \inf W_{\mathfrak{C}_\Omega}(\lambda_\mu, \lambda_\nu)$$

$$\text{such that } \mathfrak{P}_{HK}(\lambda_\mu) = \mu,\ \mathfrak{P}_{HK}(\lambda_\nu) = \nu$$



Static

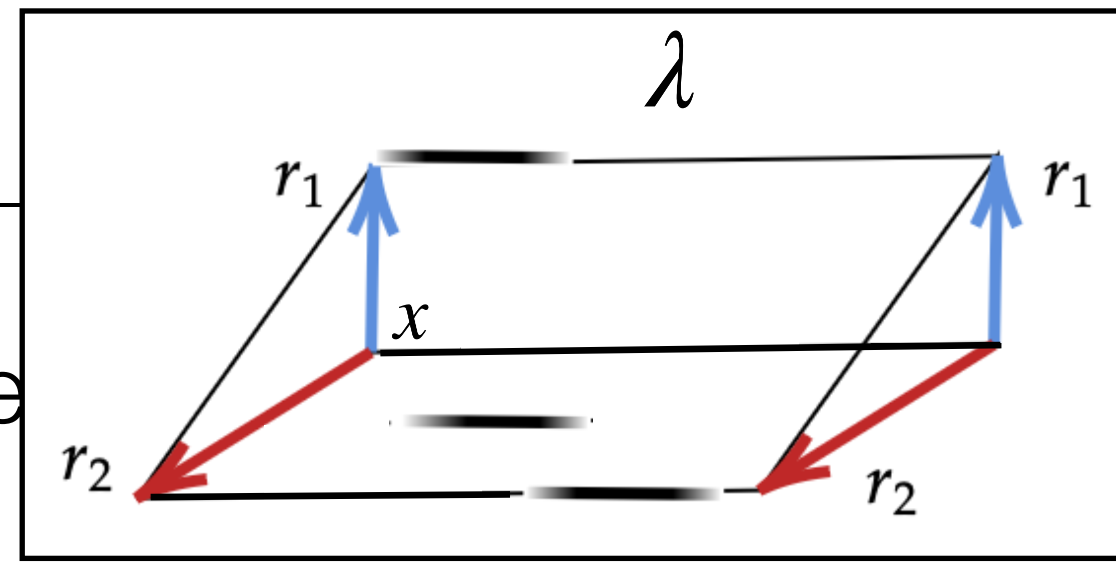
$$\int_{\mathbb{R}^d} \eta(x)d\mu_n(x) = \int_{\mathbb{R}^d\times\Delta^{n-1}} \left(1-\sum_{i=1}^{n} r_i\right)\eta(x)d\lambda(x,r)$$

Motivated

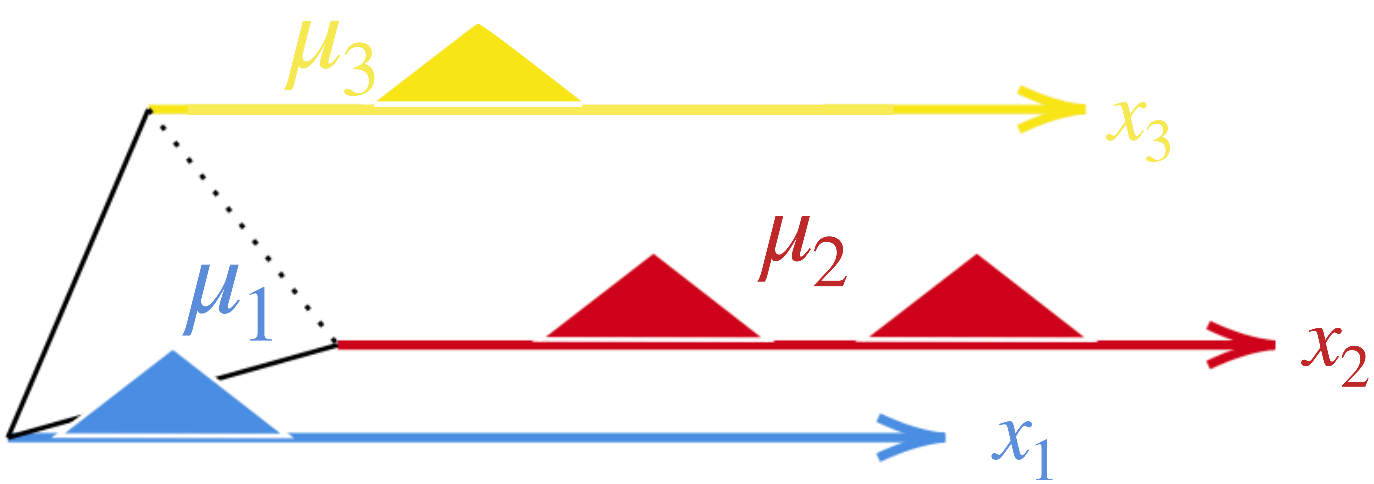
$$\mathfrak{P} : \mathscr{P}(\mathbb{R}^d \times \Delta^{n-1}) \to \mathscr{P}(\mathbb{R}^d \times G) : \lambda \mapsto \pi_{\mathbb{R}^d}\#(\boldsymbol{p}(r)\lambda(x,r))$$

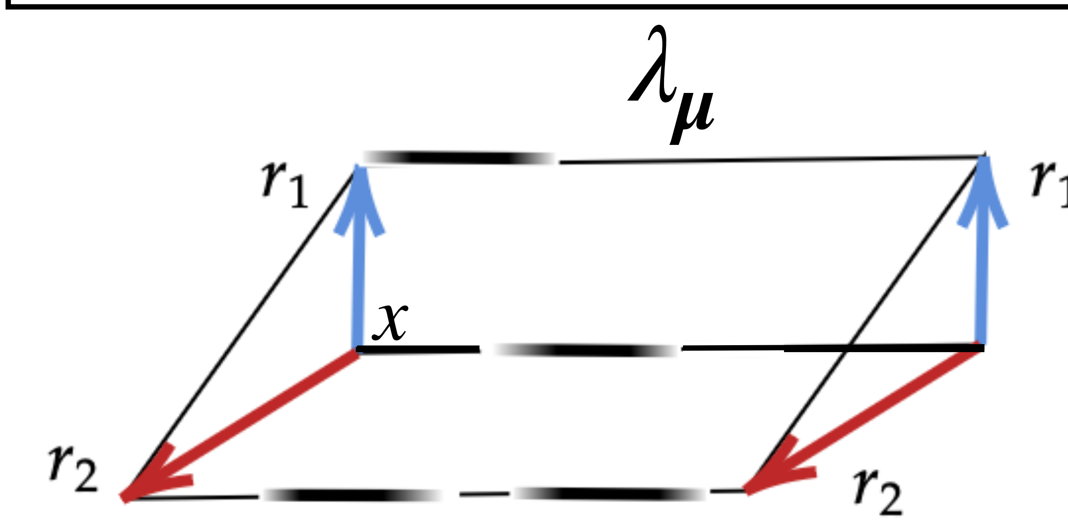
In other words, $\boldsymbol{\mu} = \mathfrak{P}\lambda$ iff for $i = 1,\ldots,n-1$, $\eta \in C_b(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \eta(x)d\mu_i(x) = \int_{\mathbb{R}^d\times\Delta^{n-1}}$$



Call $\lambda_{\boldsymbol{\mu}}(x,r) := \sum_{i=1}^{n} \mu_i(x) \otimes \delta_{e_i}(r)$ the





Static (semi) met

$D_{\mathbb{R}^d \times G}$ fails the triangle inequality

$$D_{\mathbb{R}^d\times G}(\boldsymbol{\mu},\boldsymbol{\nu}) := \inf\left\{W_{\mathbb{R}^d\times\Delta^{n-1}}(\lambda_1,\lambda_2) : \mathfrak{P}\lambda_1 = \boldsymbol{\mu},\ \mathfrak{P}\lambda_2 = \boldsymbol{\nu}\right\}$$

$$W_{2,\mathscr{W}}(\boldsymbol{\mu},\boldsymbol{\nu}) := W_{\mathbb{R}^d\times\Delta^{n-1}}(\lambda_{\boldsymbol{\mu}},\lambda_{\boldsymbol{\nu}})$$

Prop [C., García Trillos, Nikolic '25]: On $\mathscr{P}_2(\mathbb{R}^d \times G)$, $D_{\mathbb{R}^d\times G}$ is a semi-metric, $W_{2,\mathscr{W}}$ is a metric, and $D_{\mathbb{R}^d\times G} \le W_{2,\mathscr{W}}$.

Prop [C., García Trillos, Nikolic '25]:

$$W^2_{2,\mathscr{W}}(\boldsymbol{\mu},\boldsymbol{\nu}) = \min \sum_{i,j=1}^{n} \iint \|x-\tilde{x}\|^2 + W_G^2(\delta_j,\delta_i)\,d\gamma_{ij}(x,\tilde{x})$$

$$\text{s.t. } \gamma_{ij} \in \mathscr{P}(\mathbb{R}^{2d}),\ \mu_i = \sum_{j=1}^{n} \pi_1 \# \gamma_{ij},\ \nu_j = \sum_{i=1}^{n} \pi_2 \# \gamma_{ij}$$

c.f. [Erbar, Maas '12] on graph, [Bacon '20] for cost $c_{ij}(x,\tilde{x})$

Plan

1. Motivation for vector valued OT
2. Previous work on vector valued and graph OT
3. New dynamic and static (semi)-metrics
4. Metric comparison

Comparison of metrics

How are these various (semi) metrics related?

Thm [C., García Trillos, Nikolic '25]: For $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathscr{P}(\Omega \times G)$,

$$W_{\Omega \times G}(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq D_{\mathbb{R}^d \times G}(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq W_{2,\mathscr{W}}(\boldsymbol{\mu}, \boldsymbol{\nu})$$

Cor [C., García Trillos, Nikolic '25]: Suppose $\Omega \subseteq \mathbb{R}^d$ is bounded. For $C_q \sim \max_i \sum_j q_{ij}$ and $C_d \sim \operatorname{diam}(\Omega \times \Delta^{n-1})$,

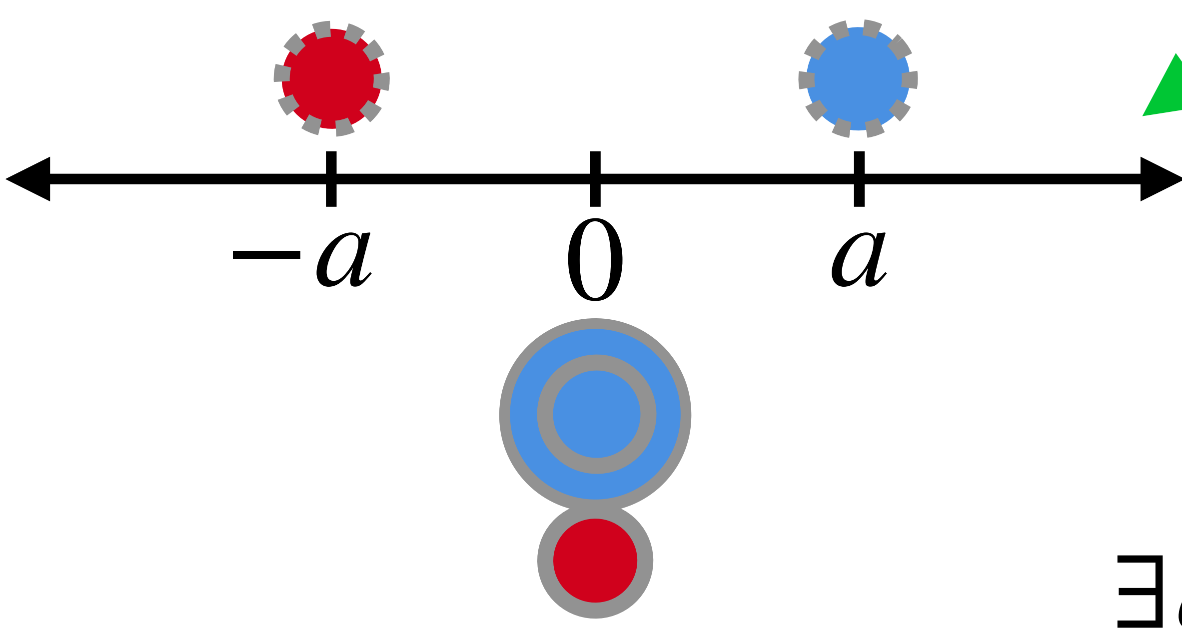
$$\min\{1, C_q^{-1/2}\} d_{BL}(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \dots \leq \max\{1, C_d^{3/2} n^{1/4}\} d_{BL}^{1/2}(\boldsymbol{\mu}, \boldsymbol{\nu})$$

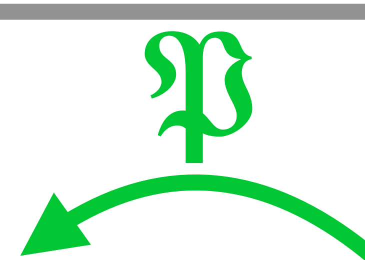
In particular, on a bounded domain, all are topologically equiv.

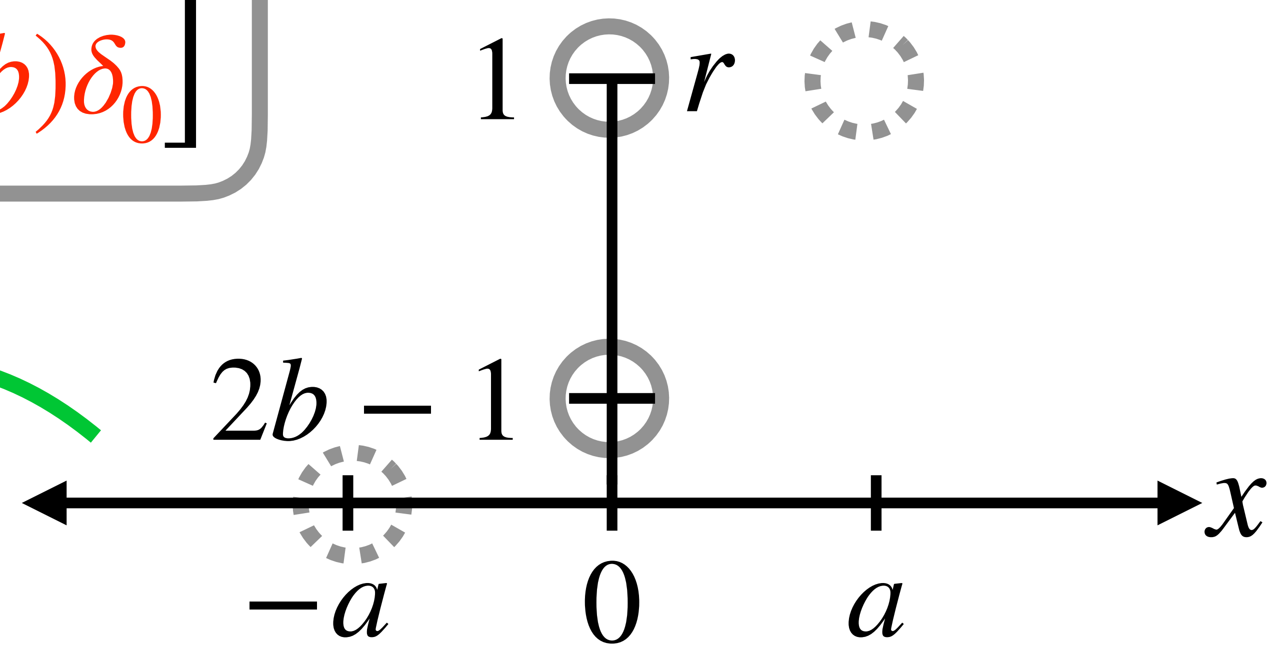
Example

$$\boldsymbol{\mu} = \frac{1}{2}\begin{bmatrix} \delta_a \\ \delta_{-a} \end{bmatrix}$$

$$\boldsymbol{\nu} = \begin{bmatrix} b\delta_0 \\ (1-b)\delta_0 \end{bmatrix}$$







$$W_{\mathbb{R}\times G}(\boldsymbol{\mu},\boldsymbol{\nu}) \leq a + d_{[0,1]}(1/2, b)$$

$\exists a, b$ s.t.

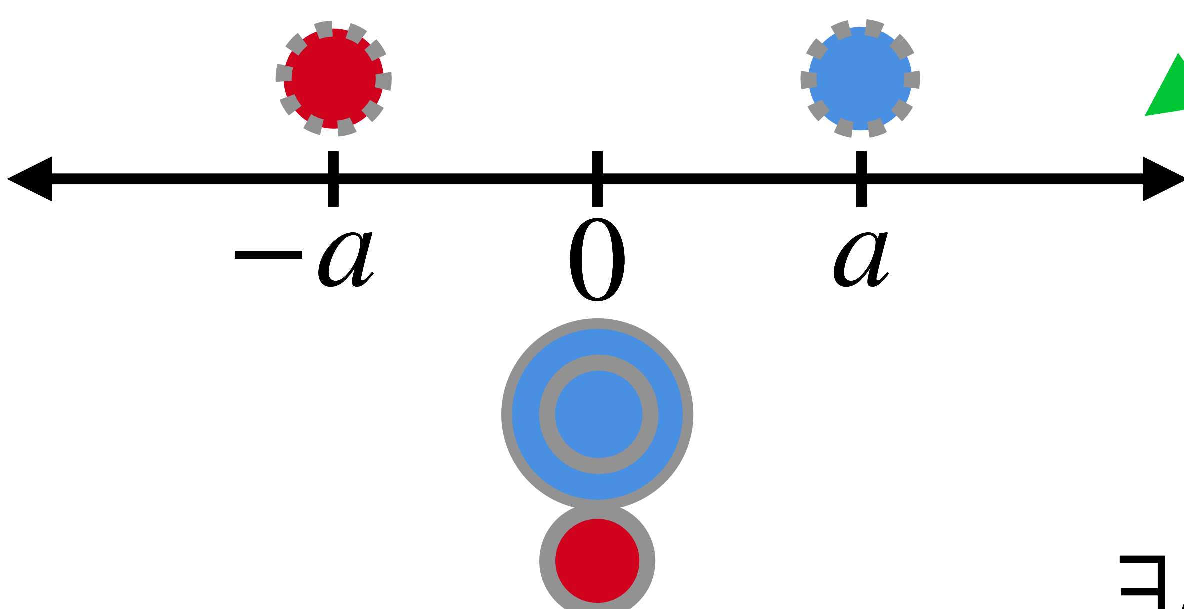
$$<$$

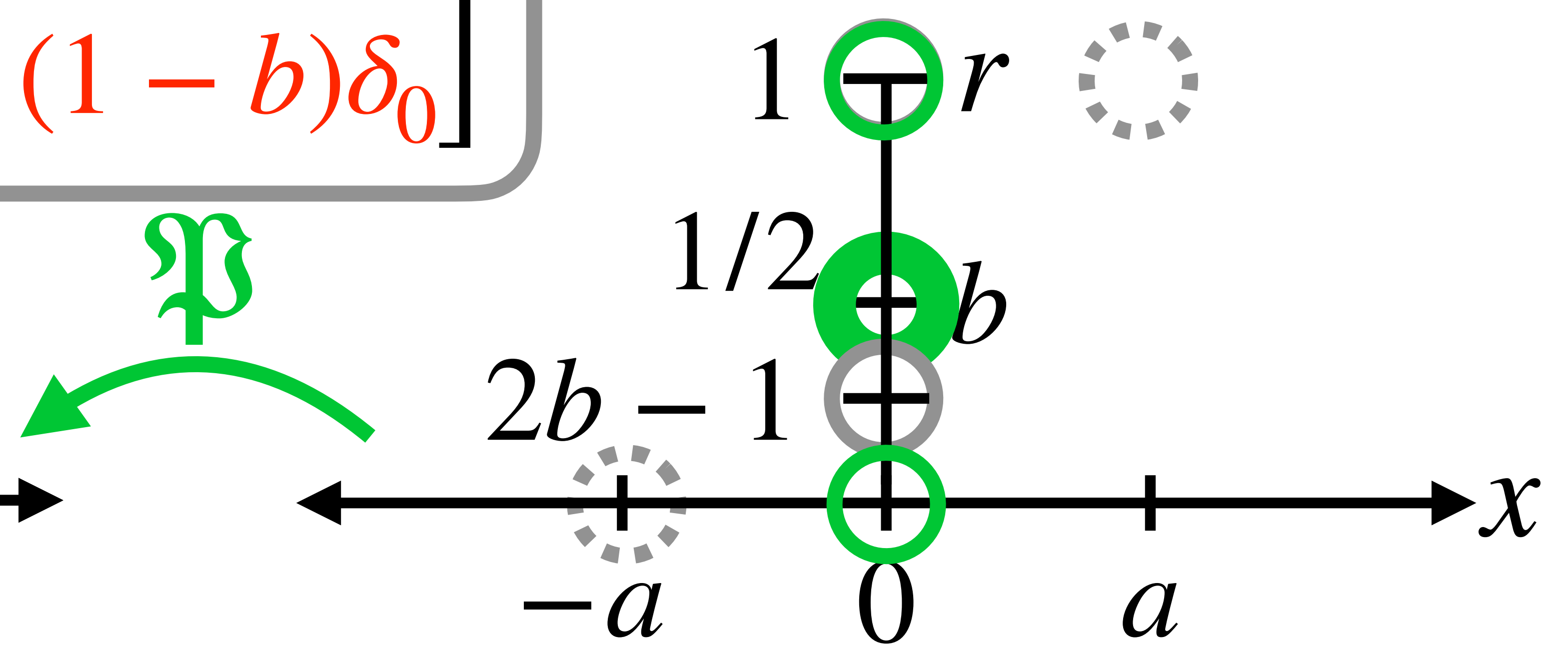
$$D_{\mathbb{R}\times G}(\boldsymbol{\mu},\boldsymbol{\nu}) = \sqrt{a^2 + \frac{1}{2} d^2_{[0,1]}(2b-1, 0)}$$

Example

$$\boldsymbol{\mu} = \frac{1}{2}\begin{bmatrix} \delta_a \\ \delta_{-a} \end{bmatrix}$$

$$\boldsymbol{\nu} = \begin{bmatrix} b\delta_0 \\ (1-b)\delta_0 \end{bmatrix}$$





$$W_{\mathbb{R}\times G}(\boldsymbol{\mu},\boldsymbol{\nu}) \leq a + d_{[0,1]}(1/2, b)$$

$\exists a, b$ s.t. $<$

$$D_{\mathbb{R}\times G}(\boldsymbol{\mu},\boldsymbol{\nu}) = \sqrt{a^2 + \frac{1}{2} d^2_{[0,1]}(2b-1, 0)}$$

Which metric to use?

Dynamic metric: **gradient flows**; for $\mathscr{L}(\boldsymbol{\rho}, \boldsymbol{\mu}) := \sum_{i=1}^{n} KL(\rho_i \,|\, \mu_i)$,

$$\partial_t \rho_{i,t} = \Delta \rho_{i,t} - \nabla \cdot (\nabla \log(\mu_i) \rho_{i,t}) + \sum_{j=1}^{n} \theta(\rho_{i,t}, \rho_{j,t}) \log\left(\frac{\rho_{i,t}/\mu_i}{\rho_{j,t}/\mu_j}\right) q_{ij}$$

Static metric: **classification and linearization**

$$d^2_{LOT}(\boldsymbol{\mu}, \boldsymbol{\nu}) := \int_{\mathbb{R}^d \times \Delta^{n-1}} \|T_{\boldsymbol{\mu}}(x, r) - T_{\boldsymbol{\nu}}(x, r)\|^2 d\lambda_{\text{ref}}(x, r)$$

Caveat: existence of $T_{\boldsymbol{\mu}}$ satisfying $T_{\boldsymbol{\mu}} \# \lambda_{\text{ref}} = \lambda_{\boldsymbol{\mu}}$ open, since $\mathbb{R}^d \times \Delta^{n-1}$ may be a branching space [Cavallett, Mondino '17]

Thank you!

Gradient flows

Energy:

$$E(\boldsymbol{\mu}) = \int_{\mathbb{R}^d} f(\boldsymbol{\mu}(x))dx + \sum_{i=1}^{n} \int_{\mathbb{R}^d} V_i(x)\mu_i(x)dx$$

$$+ \frac{1}{2} \sum_{i,j=1}^{n} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mu_i(x) W_{ij}(x-y) \mu_j(y) dx dy,$$

Gradient flows:

$$\partial_t \mu_i(x) = \nabla \cdot \left(\mu_i(x) \left(\nabla \partial_i f(\boldsymbol{\mu}(x)) + \nabla V_i(x) + \sum_{k=1}^{n} \nabla W_{ik} * \mu_k(x) \right) \right)$$

$$- \sum_{j=1}^{n} \left(\partial_i f(\boldsymbol{\mu}(x)) - \partial_j f(\boldsymbol{\mu}(x)) + V_i(x) - V_j(x) + \sum_{k=1}^{n} (W_{ik} - W_{jk}) * \mu_k(x) \right)$$

$$\cdot \theta\left(\mu_i(x), \mu_j(x)\right) q_{ij},$$

Graph operators

$$\nabla_{\mathcal{G}} : \mathbb{R}^n \to \mathbb{R}^{n \times n}(\mathbb{R}) : \phi \mapsto [\phi_j - \phi_i]_{i,j=1}^n$$

$$\mathrm{div}_{\mathcal{G}} : \mathbb{R}^{n \times n} \to \mathbb{R}^n : v \mapsto \Big[-\frac{1}{2} \sum_j (v_{ij} - v_{ji}) q_{ij} \Big]_{i=1}^n$$