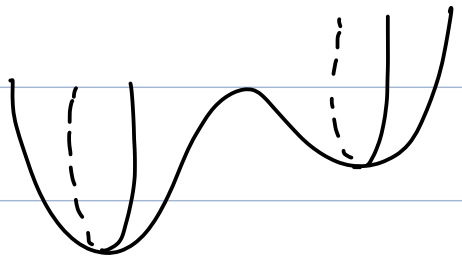
Dynamic Optimal Transport

Plan

Dynamic 2-Wasserstein Metric

⇝ geometry

⇝ gradient flows



Generalizations of W_2

① Graph Wasserstein Metric

② Vector-valued Optimal Transport

References

AGS Ambrosio, Gigli, Savaré, Gradient Flows, 2005

M Mass, "Gradient flows of the entropy", 2011

Em Erbar, Maas, "Ricci Curvature...," 2012

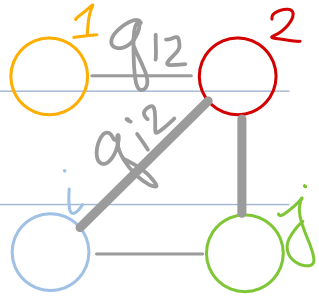
GLM Gangbo, Li, Mou, "Geodesic of Minimal...", 2017

CGN C., García Trillos, Nikolić, "Vector Valued", 2025

Recall: Graph Wasserstein Metric

Let G denote an n node, connected, symmetric, weighted graph.

$q_{ij} = q_{ji}$



Let $\rho, \tilde{\rho} \in \mathcal{P}(G) = \{\rho \in \mathbb{R}_+^n : \sum_{i=1}^n \rho_i = 1\}$

$$W_G^2(\rho, \tilde{\rho}) = \inf_{(\rho_t, v_t)} \int_0^1 \frac{1}{2} \sum_{i,j=0}^n |v_{ij,t}|^2 \check{\rho}_{ij,t} q_{ij} \, dt$$

$\check{\rho}_t$: $\langle v_t, v_t \rangle \rho_t$

$$\partial_t \rho_t + \operatorname{div}_G(v_t \check{\rho}_t) = 0$$

$$\rho_0 = \rho, \rho_1 = \tilde{\rho}$$

where $\check{\rho} = [\check{\rho}_{ij}]_{i,j=1}^n = [\theta(\rho_i, \rho_j)]_{i,j=1}^n$

Examples of θ:

① $\theta(\rho_i, \rho_j) \in \frac{\rho_i + \rho_j}{2}$, ② $\theta(\rho_i, \rho_j) = \sqrt{\rho_i \rho_j}$, ...

Geometry:

Thm (M) $(\mathcal{P}_{>0}(G), W_G)$ is a smooth Riemannian manifold.

Bad News:

- $(\mathcal{P}_{>0}(G), W_G)$ is not complete; geodesics between points in interior can touch bdy. GLM
- the minimizer of dynamic problem does not necessarily exist but if it exists and is strictly positive, it is a geodesic. (as a curve in the space of probability measures)

$\{\{$

For those who want a deep dive into this last statement, see pdf of previous lecture notes for more details and an open problem...

α

Kantorovich formulation

Thm (M) There is no Kantorovich formulation

Pf: Suppose G is a two node graph and we had a Kantorovich form.

$$\tilde{W}_G^2(\rho,\tilde{\rho}) = \inf_{\gamma \in \Gamma(\rho,\tilde{\rho})} \sum_{i,j=1}^{2} d(x_i,x_j)\gamma_{ij}$$

$$= \underbrace{d(x_1,x_2)}_{c} |\rho_1 - \tilde{\rho}_1|$$

Suppose $\rho : [0,1] \to \mathcal{P}(G)$ is a geodesic. Then

$$c|\rho_{1,t} - \rho_{1,s}| = \tilde{W}_G^2(\rho_t,\rho_s) = |t-s|^2 \tilde{W}_G^2(\rho_0,\rho_1)$$

$$= |t-s|^2 c\, |\rho_{1,0} - \rho_{1,1}|$$

Thus $t \mapsto \rho_{1,t}$ is $\frac{1}{2}$-Hölder cts, hence constant; the only geodesics in $W_{\tilde{G}}$ are constant curves.

OTOH W_G has a rich class of nonconstant geos → (minimizers of dynamic form.) □

Gradient Flows.

Given $E : \mathcal{P}(G) \to \mathbb{R}$ smooth, $\rho_0 \in \mathcal{P}(G)$ what smooth $\rho : [0,1] \to \mathcal{P}(G)$ makes E decrease as quickly as possible?

$\in \mathbb{R}^n_+$

$$\begin{aligned}
\frac{d}{dt} E(\rho_t) &= \langle \nabla_{\mathbb{R}^n} E(\rho_t), \partial_t \rho_t \rangle_{\mathbb{R}^n} \\
&= \langle \nabla_{\mathbb{R}^n} E(\rho_t), -\operatorname{div}_G(v_t \tilde{\rho}_t) \rangle_{\mathbb{R}^n} \\
&= \langle \nabla_G \nabla_{\mathbb{R}^n} E(\rho_t), v_t \tilde{\rho}_t \rangle_G \\
&= \langle \nabla_G \nabla_{\mathbb{R}^n} E(\rho_t), v_t \rangle_{\rho_t}
\end{aligned}$$

$$\text{WGGF eqn: } \partial_t \rho_t - \operatorname{div}_G(-\nabla_G \nabla_{\mathbb{R}^n} E(\rho_t)\rho_t) = 0$$

Example:

$$\theta(\rho_i, \rho_j) = \frac{\rho_i - \rho_j}{\log\rho_i - \log\rho_j} = \int_0^1 \rho_i^s \rho_j^{1-s} ds$$

$$E(\rho) = \sum_i \rho_i \log \rho_i \Rightarrow \nabla_{\mathbb{R}^n} E(\rho) = [\log\rho_i + 1]_{i=1}^n$$

Given an initial condition $\rho_0 \in \mathcal{P}(G)$, the WG gradient flow of E is

$$\partial_t \rho_t - \operatorname{div}_G(\nabla_G[\nabla_{\mathbb{R}^n} E(\rho_t)]\check{\rho}_t) = 0$$

$$\Updownarrow$$

$$\partial_t \rho_t - \operatorname{div}_G([\log\rho_{i,t} - \log\rho_{j,t}]_{i,j=1}^n \check{\rho}_t) = 0$$

$$\Updownarrow \quad \nabla_G \rho_t$$

$$\partial_t \rho_t - \operatorname{div}_G([\rho_{i,t} - \rho_{j,t}]_{i,j=1}^n) = 0$$

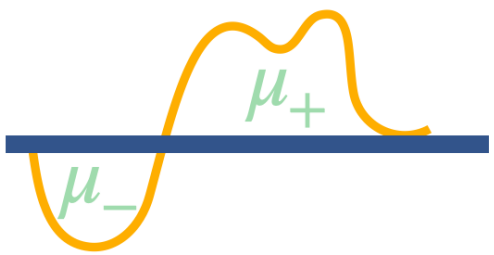
$$\Updownarrow$$

$$\partial_t \rho_t = \Delta_G \rho_t$$

Vector-Valued Optimal Transport [C., García-Trillos, Nikolić, '25]

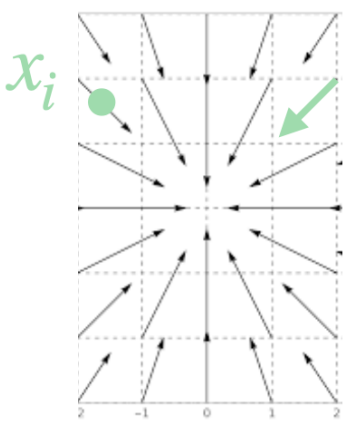
Motivation:

Signed measures



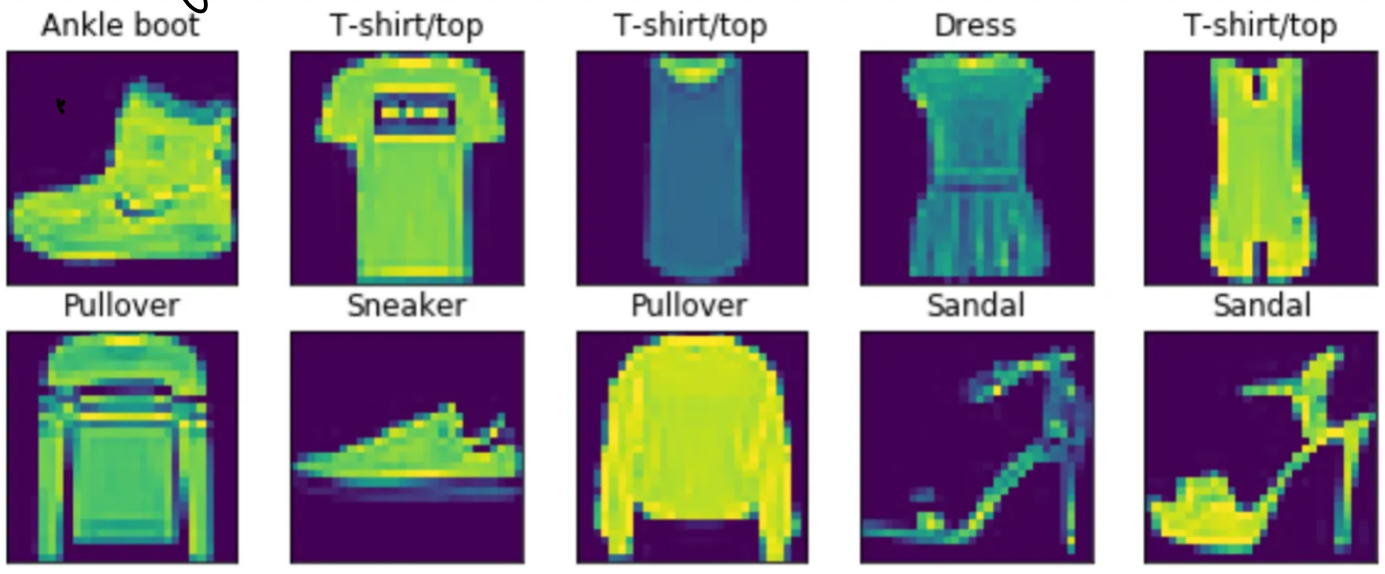
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_+ \\ \mu_- \end{bmatrix}$$

Vector fields

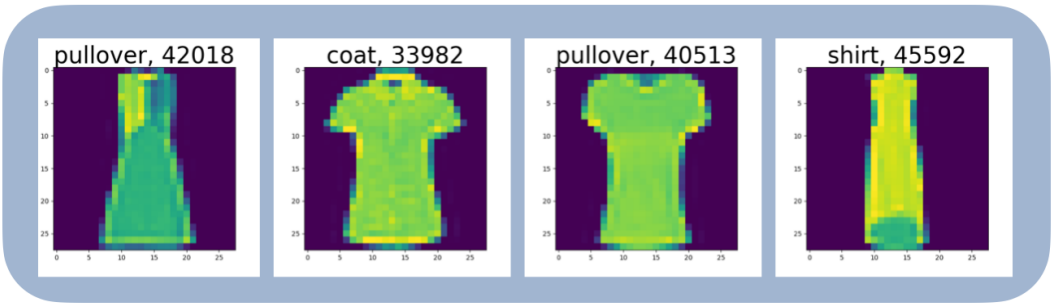


$$\boldsymbol{\mu} = \sum_{i=1}^{N} \begin{bmatrix} m_i^1 \\ m_i^2 \end{bmatrix} \delta_{x_i}$$

Image Classification

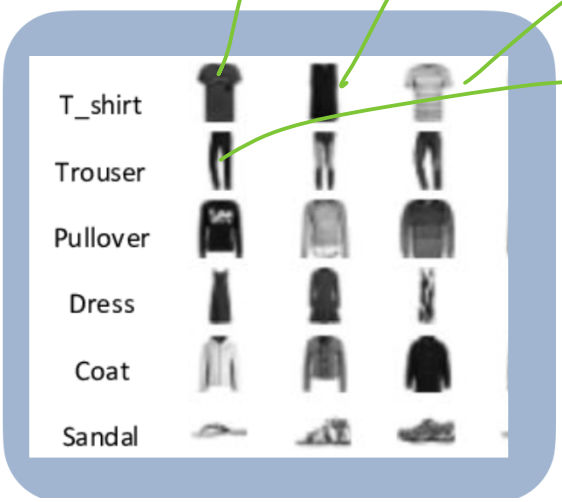


Mild misclassifications in Fashion MNIST— for example…



[Müller, Markert 2019]

Labeled Dataset #1



Labeled Dataset #2

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.

VV OT Formulation:

$$\mu_i = \sum_{k \in K_i} \delta_{x_k} m_k, \quad K_i = \{k : x_k \text{ has label } i\}$$

labeled dataset $= [\mu_i]_{i=1}^n = \underline{\mu}$

Goal: Find a distance d on the space of vector valued measures s.t. $d([\mu_i]_{i=1}^n, [\nu_i]_{i=1}^n) << 1$ if the measures are similar,

e.g., the datasets have similar images and labels.

Applications:

- compare the performance of two classifiers
- compare the structure of two labeled datasets
- optimization over the space of vv measures via GF

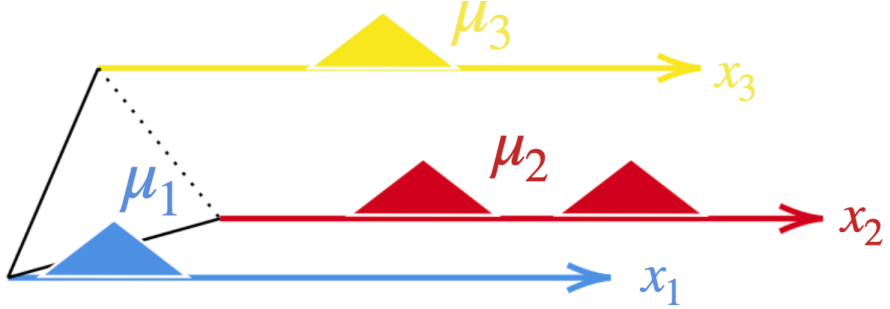
Given $\Omega \subseteq \mathbb{R}^d$ closed, convex

G n node sym, conn graph

we consider vv measures in

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

Example:



Continuity eqn:

Given $[\mu_{i,t}(x)]_{i=1}^n$, $[v_{ij,t}(x)]_{i,j=1}^n$

$$\partial_t \rho + \nabla \cdot (\rho \circ \underline{u}) + \operatorname{div}_G(\check{\rho}\, \underline{v}) = 0$$

For $v_{ij} = -v_{ji}$, each component satisfies

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

Action:

$$\|(\underline{u},\underline{v})\|_{\underline{\rho}}^2 := \sum_i \int_\Omega |u_i|^2 \rho_i + \sum_{i,j} \int_\Omega |v_{ij}|^2 \check{\rho}_{ij} q_{ij}$$

This leads to the dynamic vector valued distance...

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

$$(\underline{\rho},\underline{u},\underline{v}) \in \mathcal{C}(\underline{\mu},\underline{\nu})$$

Thm $W_{\Omega\times G}$ is a metric $\mathcal{P}_2(\Omega\times G)$ and a minimizer exists.

Kantorovich Formulation

- On one hand, we don't expect a Kantorovich formulation to exist, since when $\Omega = \{x_0\}$, $W_{\Omega \times G}$ reduces to W_G.
- On the other hand, via analogy with Hellinger-Kantorovich, maybe there exists a static metric via liftings?

nice subset of Eucl. space

Fact: $\forall \underline{\mu} \in \mathcal{P}(\Omega \times G)$, $\exists \lambda_{\underline{\mu}} \in \mathcal{P}(\Omega \times \mathcal{P}(G))$ s.t. $\mathcal{P} \lambda_{\underline{\mu}} = \underline{\mu}$, where $\mathcal{P}$ is a linear projection operator.

Thm For all $\underline{\mu}, \underline{\nu} \in \mathcal{P}(\Omega \times G)$,

$$W_{\Omega \times G}(\underline{\mu}, \underline{\nu}) \leq \inf_{\substack{\mathcal{P} \lambda_{\underline{\mu}} = \underline{\mu} \\ \mathcal{P} \lambda_{\underline{\nu}} = \underline{\nu}}} W_{\Omega \times \mathcal{P}(G)}(\lambda_{\underline{\mu}}, \lambda_{\underline{\nu}})$$

Good News: If Ω is bounded,

- LHS is bdd below d_{BL}
- RHS is bdd above $(d_{BL})^{1/2}$

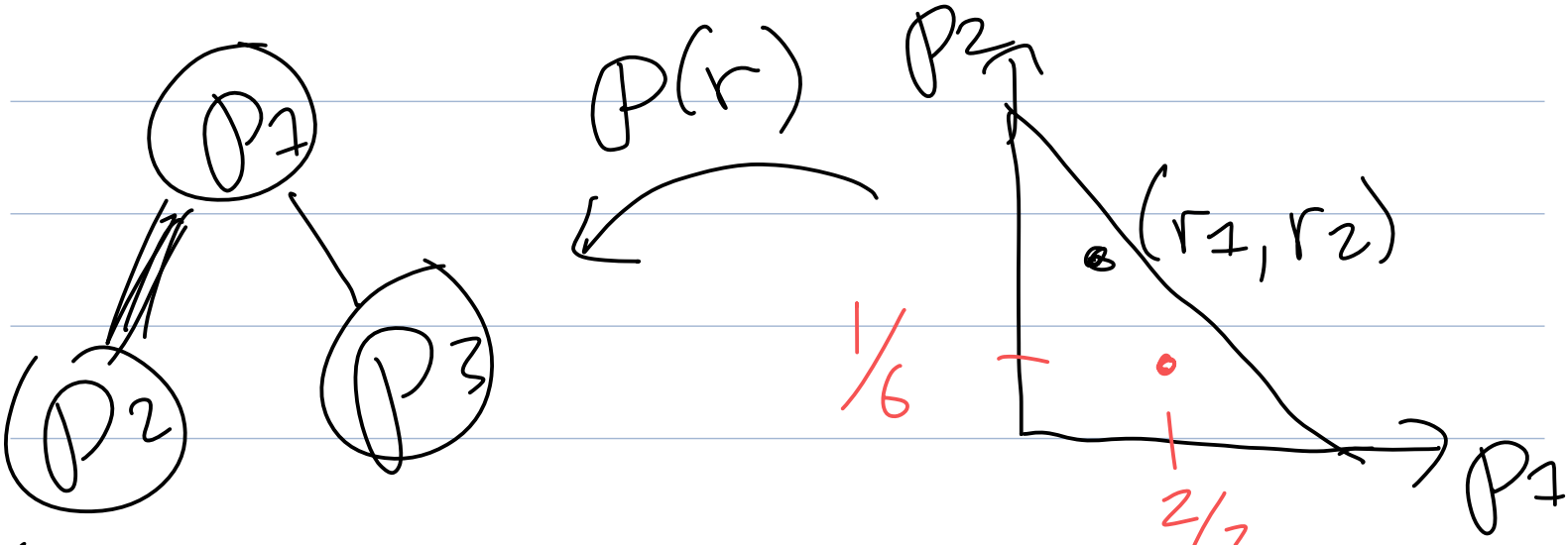
$\Rightarrow$ Topological equivalence.

...very reminiscent of relationship between W_2 and OLOT

$$W_2(\mu,\nu) \leq LOT_\sigma(\mu,\nu) \leq C\left(W_2(\mu,\nu)\right)^{1/6}$$

Future Work

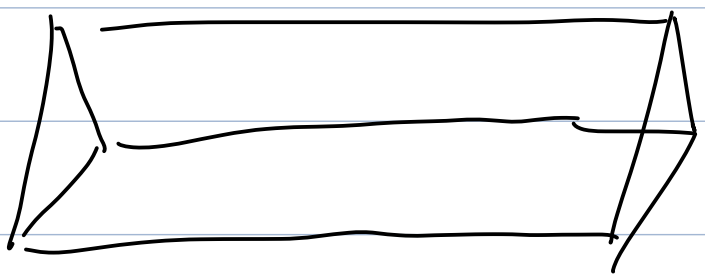
- Linearization
- Gradient flows
- Applications



$$(P(G), W_G) \cong (\Delta^{n-1}, d_{\Delta^{n-1}})$$

$$P(\Omega \times P(G)) = P(\Omega \times \Delta^{n-1})$$

$$\cap \lambda$$



$$(T\lambda)_i = d\mu_i(x) = \int_{\Delta^{n-1}} \rho_i(r)\, d\lambda(x, r)$$