Linearization of balanced and unbalanced optimal transport

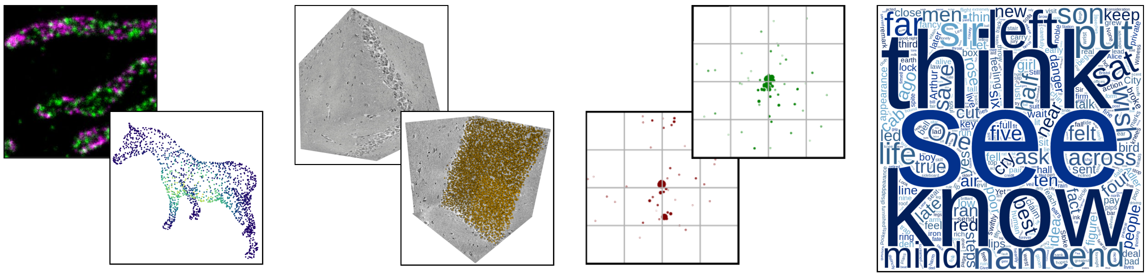
Bernhard Schmitzer

IFIP TC7 online lecture, April 2023

1 Introduction to optimal transport

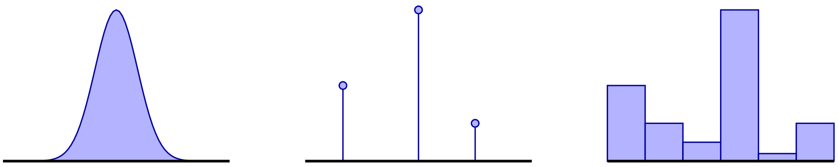
1.1 Measures for data modelling

Comparing and understanding data



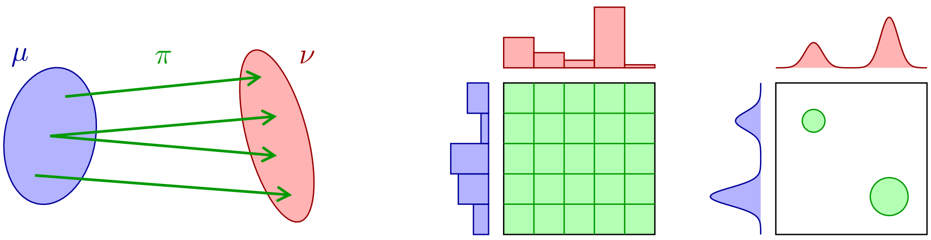
- 'Are two samples similar?'

Language: probability measures $\mathcal{P}(X)$ on metric space (X, d)



- similarity of samples $\leftrightarrow$ metric on $\mathcal{P}(X)$

1.2 Kantorovich formulation of optimal transport



Couplings

- $\Pi(\mu, \nu) := \{\pi \in \mathcal{M}_+(X \times X) : \mathrm{P}_{1\sharp}\pi = \mu,\ \mathrm{P}_{2\sharp}\pi = \nu\}$
- **marginals**: $\mathrm{P}_{1\sharp}\pi(A) := \pi(A \times X)$, $\mathrm{P}_{2\sharp}\pi(B) := \pi(X \times B)$
- **rearrangement** of mass, generalization of **map**

Optimal transport [**Kantorovich, 1942**]

$$C(\mu, \nu) := \inf \left\{ \int_{X \times X} c(x, y)\,\mathrm{d}\pi(x, y) \middle| \pi \in \Pi(\mu, \nu) \right\}$$

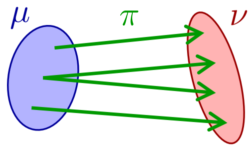
- **cost function** $c : X \times X \to \mathbb{R}$ for moving unit mass from x to y
- **convex problem**: linear program

Wasserstein distance on probability measures $\mathcal{P}(X)$

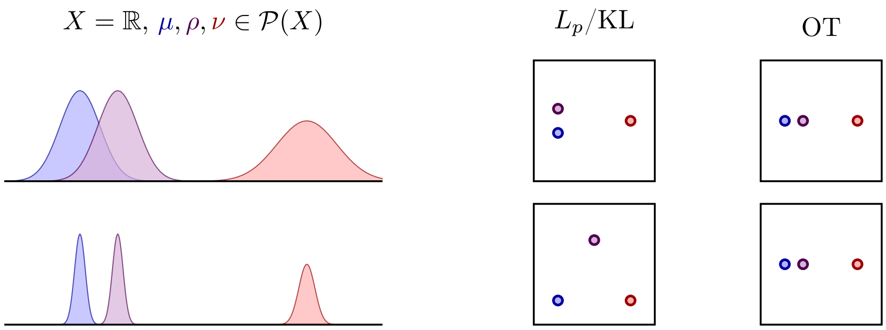
$$W_p(\mu,\nu) := (C(\mu,\nu))^{1/p} \quad \text{for } c(x,y) := d(x,y)^p, \quad p \in [1,\infty)$$

1.3 Some important properties of Wasserstein distances

$$W_2(\mu,\nu) := \inf \left\{ \int_{X\times X} d(x,y)^2 \, \mathrm{d}\pi(x,y) \middle| \pi \in \Pi(\mu,\nu) \right\}^{1/2}$$



- **intuitive**, **robust** to positional noise

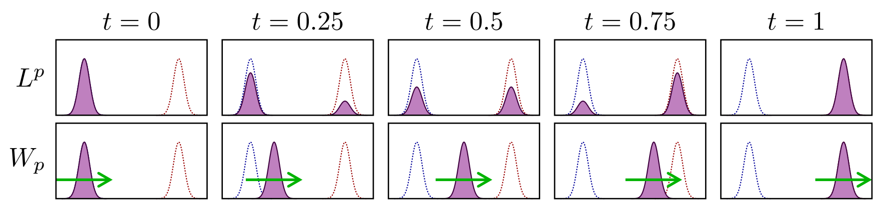


Transport maps [Brenier, 1991]

- $[X = \mathbb{R}^d,\ \mu \ll \mathcal{L},\ c = d^p] \Rightarrow [\pi = (\mathrm{id}, T)_\sharp \mu]$
- $W_2(\mu,\nu)^2 = \int_X \|T(x) - x\|^2 \, \mathrm{d}\mu(x)$

Displacement interpolation [McCann, 1997]

- $[(X,d)$ length space$] \Rightarrow [(\mathcal{P}(X), W_p)$ length space$]$
- $X = \mathbb{R}^d$: $\rho_t := [(1-t)\cdot \mathrm{id} + t \cdot T]_\sharp \mu$
- **velocity field** v_t: mass particle starting at x travels with constant speed along straight line to $T(x)$



Dynamic formulation: Benamou–Brenier formula (on $X = \mathbb{R}^d$)
[Benamou and Brenier, 2000]

- (weak) **continuity equation**: mass ρ, velocity field v

$$\mathcal{CE}(\mu,\nu) := \{(\rho, v) \colon \partial_t \rho + \nabla(v \cdot \rho) = 0,\ \rho_0 = \mu,\ \rho_1 = \nu\}$$

- **least action principle**: minimize Lagrangian / kinetic energy

$$W_2(\mu,\nu)^2 = \inf_{(\rho,v)\in\mathcal{CE}(\mu,\nu)} \int_0^1 \int_X \|v_t\|^2 \, \mathrm{d}\rho_t \mathrm{d}t$$

- $(\mathcal{P}(X), W_2)$ has weak Riemannian structure [Otto, 2001]

1.4 Wasserstein distances: what now?

Attractive properties

- ✓ **intuitive, robust, flexible** metric for probability measures
- ✗ **numerically involved**, ✓but good solvers exist
- ✓ **rich geometric structure** (barycenter, interpolation, Riemannian flavour…)

Challenge #1

- ✗ analyzing point clouds in **non-linear metric space** is tricky
  - ✓ approximate Euclidean embeddings
  - ✗ interpretation not obvious
- ✗ requires computation of **all pairwise distances**
- ✓ remedy through **local linearization** [Wang et al., 2012]

Challenge #2

- ✗ W_2 susceptible to small non-local **mass fluctuations**
- ✓ remedy through **unbalanced transport**, in particular **Hellinger–Kantorovich distance**

In this talk: combine both ingredients

2 Interlude: a little bit of Riemannian geometry

2.1 Basic concepts

Riemannian manifold $\mathbb{M}$

- locally homeomorphic to $\mathbb{R}^d$, **tangent space** $T_z\mathbb{M} \simeq \mathbb{R}^d$ at z
- at each point: inner product $\langle\cdot,\cdot\rangle_z$ and norm $\|\cdot\|_z$: **angles and speed**
- **examples:** $\mathbb{R}^d$, torus, sphere

Length and distance

- $\text{length}(\gamma) := \int_0^1 \|\dot\gamma(t)\|_{\gamma(t)}\,\mathrm{d}t$ for $\gamma \in C^1([0,1],\mathbb{M})$
- distance $d(x,y) := \inf\{\text{length}(\gamma)|\gamma(0)=x,\ \gamma(1)=y\}$
- $d(x,y)^2 = \inf\left\{\int_0^1 \|\dot\gamma(t)\|^2_{\gamma(t)}\,\mathrm{d}t\,\middle|\,\gamma(0)=x,\ \gamma(1)=y\right\}$
- minimal γ called **geodesics**, generalization of straight line

2.2 Local linearization of Riemannian manifold

Exponential map $\text{Exp}_z : T_z\mathbb{M} \to \mathbb{M}$

- $\text{Exp}_z(v)$ = start walking at z with velocity v until time 1
- 'follow curvature' of $\mathbb{M}$

Inverse: logarithmic map $\text{Log}_z : \mathbb{M} \to T_z\mathbb{M}$

- may not be defined on full $\mathbb{M}$ → cut-locus
- **Thm:** $\|\text{Log}_z(y)\|_z = d(z,y)$

Local linearization of d

- $\text{Lind}_z(x,y) := \|\text{Log}_z(x) - \text{Log}_z(y)\|_z$

✗ $\text{Lind}_z \neq d$ on curved manifolds, ✓error small when x, y, z close and bounded curvature R

$$d(x,y)^2 = \text{Lind}_z(x,y)^2 + O(R\cdot\varepsilon^4) \quad \text{if} \quad d(z,x) = d(z,y) = O(\varepsilon)$$

✓ $(T_z\mathbb{M}, \text{Lind}_z)$ is linear ⇒ many data analysis tools available

- **interpretation:** approximate curved surface locally by tangent plane

3 Linearization of Wasserstein-2

3.1 Riemannian structure of Wasserstein-2

Recall Benamou–Brenier formula (on $X = \mathbb{R}^d$)

- (weak) **continuity equation**: mass ρ, velocity field v

$$\mathcal{CE}(\mu,\nu) := \{(\rho,v)\colon \partial_t\rho + \nabla(v\cdot\rho) = 0,\ \rho_0 = \mu,\ \rho_1 = \nu\}$$

- **least action principle**: minimize Lagrangian / kinetic energy

$$W_2(\mu,\nu)^2 = \inf_{(\rho,v)\in\mathcal{CE}(\mu,\nu)} \int_0^1 \int_X \|v_t\|^2\,\mathrm{d}\rho_t\mathrm{d}t = \inf_{(\rho,v)\in\mathcal{CE}(\mu,\nu)} \int_0^1 \|v_t\|^2_{\rho_t}\mathrm{d}t$$

Formal comparison with Riemannian geometry

$$d(x,y)^2 = \inf\left\{\int_0^1 \|\dot\gamma(t)\|^2_{\gamma(t)}\,\mathrm{d}t \,\middle|\, \gamma(0)=x,\ \gamma(1)=y\right\}$$

$\Rightarrow$ v_t represents tangent vector

Logarithmic and exponential map for W_2

- let $\pi = (\mathrm{id}, T)_\sharp \mu$ optimal for $W_2^2(\mu, \nu)$

$$\mathrm{Log}_\mu(\nu) = v_0 = T - \mathrm{id}, \qquad \mathrm{Exp}_\mu(v_0) = (\mathrm{id} + v_0)_\sharp \mu$$

3.2 Local linearization of Wasserstein-2

Proposed for data analysis in [Wang et al., 2012]

- set of samples $\{\nu_i\}_{i=1}^N$, **'reference'** measure μ
- represent ν_i by optimal T_i for $W_2(\mu, \nu_i)$, **Lagrangian** representation

$$\mathrm{Log}_\mu(\nu_i) = T_i - \mathrm{id}$$

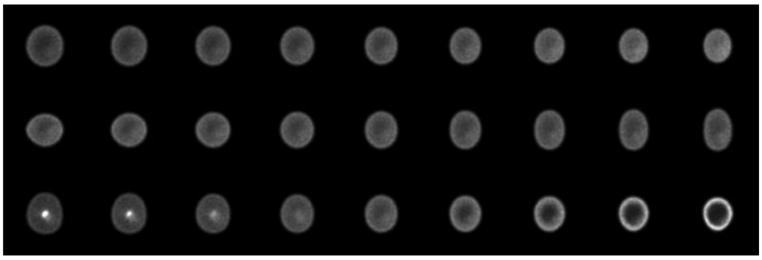
✓ approximate distance

$$\mathrm{Lin}W_2(\nu_i, \nu_j) := \|\mathrm{Log}_\mu(\nu_i) - \mathrm{Log}_\mu(\nu_j)\|_{L^2(\mu,\mathbb{R}^d)} = \|T_i - T_j\|_{L^2(\mu,\mathbb{R}^d)}$$

- $\{T_i - \mathrm{id}\}_{i=1}^N$ lie in $L^2(\mu, \mathbb{R}^d)$ $\Rightarrow$ vector space

✓ only OT problems $W_2(\mu, \nu_i)$ need to be solved, not all $W_2(\nu_i, \nu_j)$

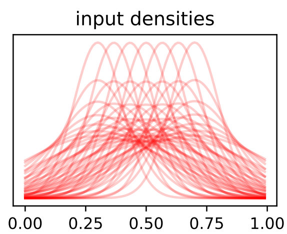
✓ simple post-processing (dimensionality reduction, classifiers, ...)



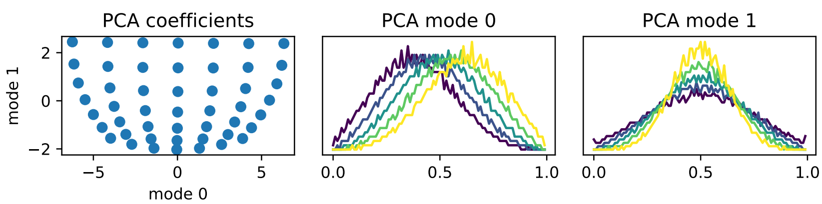
3.3 A simple numerical example

Input data:

- (truncated) Gaussians with different means and variances on $[0, 1]$

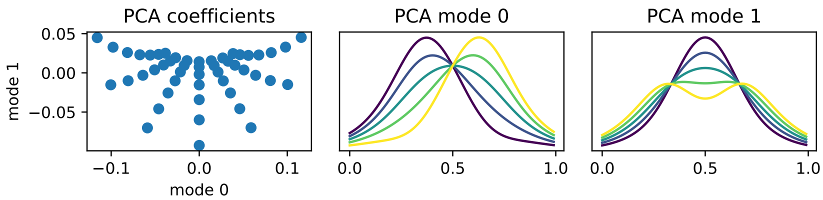


LinW_2**-analysis and PCA embedding**



- captured variance by two modes: $> 99\%$

L^2-analysis and PCA embedding



- captured variance by two modes: $\approx 90\%$

3.4 Basic properties and some references

Approximation quality $\text{Lin}W_2$ **vs** W_2

- upper bound: $\text{Lin}W_2(\nu_i, \nu_j) \geq W_2(\nu_i, \nu_j)$, proof via gluing lemma,
  $\Rightarrow$ non-negative curvature of $(\mathcal{P}(\mathbb{R}^d), W_2)$
- $\text{Lin}W_2(\nu_i, \nu_j) = W_2(\nu_i, \nu_j)$ on $(\mathcal{P}_2(\mathbb{R}), W_2)$, isometric embedding into $L^2([0,1])$
- scale and translation are 'simple flat submanifolds' of $\mathcal{P}(\mathbb{R}^d)$:
  $$\{T_\# \mu | T : x \mapsto s \cdot x + t, s \in \mathbb{R}_{++}, t \in \mathbb{R}^d\}$$
  can be embedded isometrically into $L^2(\mu)$
- map $\nu \mapsto \text{Log}_\mu(\nu)$ is continuous in $(W_2, L^2(\mu))$, but not Lipschitz or even Hölder, Hölder regularity only under additional regularity assumptions [Gigli, 2011; Delalande and Merigot, 2021]
- there are always tangent vectors along which we can only move in one direction

Approximation by discretization

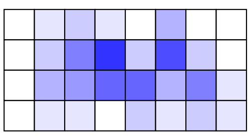
- approximate Monge map and logarithm by **barycentric projection**, convergence as $(\mu_n, \nu_n) \stackrel{*}{\rightharpoonup} (\mu, \nu)$ [Sarrazin and Schmitzer, 2023]

Other interesting directions

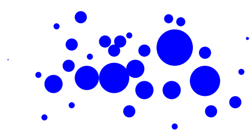
- **multiple support points** for classification [Khurana et al., 2022]
- Linearized **Gromov-Wasserstein** distance [Beier et al., 2021]
- Many nice applications to real data, 'sliced linearized OT', by Kolouri, Rohne et al.

3.5 Comparing Eulerian and Lagrangian representation

Eulerian



Lagrangian

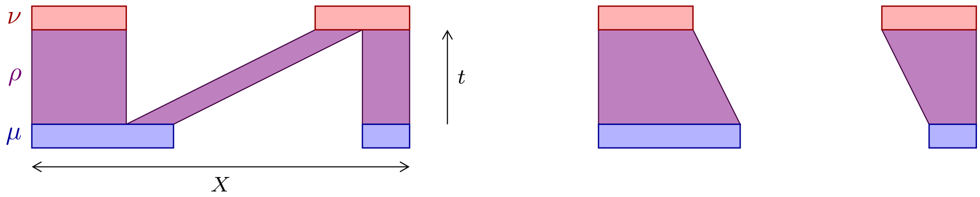


x	y	m
1.1	0.2	0.1
1.9	-0.1	0.2
⋮	⋮	⋮

- better choice depends on problem / context
- Eulerian representation sensitive to 'horizontal perturbations'
- Lagrangian representation order invariant, but consistent order makes comparison easier
- LinOT provides canonical order, 'know which list items to compare'

4 Hellinger–Kantorovich distance

[Kondratyev et al., 2016; Chizat et al., 2018b; Liero et al., 2018]



- **unbalanced continuity equation**: mass ρ, velocity v, **source** α

$$\mathcal{CE}(\mu,\nu) := \{(\rho, v, \alpha) \colon \partial_t \rho + \nabla(v \cdot \rho) = \alpha \cdot \rho,\ \rho_0 = \mu,\ \rho_1 = \nu\}$$

- **unbalanced Benamou–Brenier formula:**

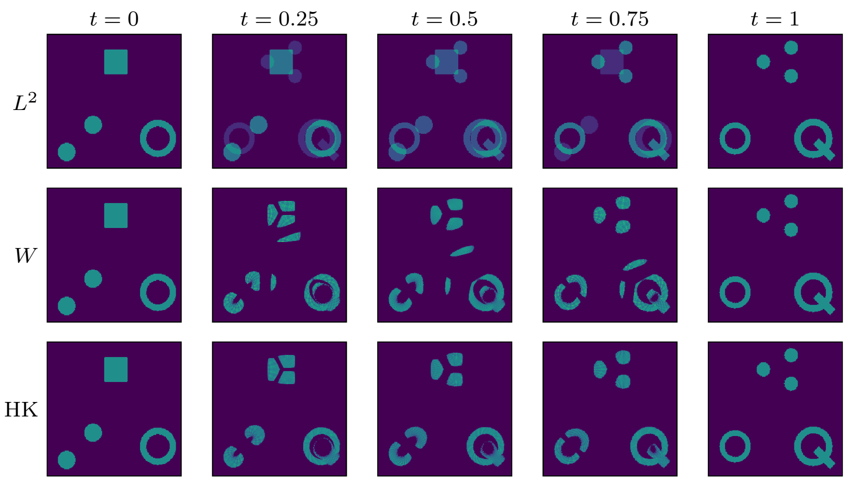
$$\mathrm{HK}(\mu,\nu)^2 := \inf_{(\rho,v,\alpha)\in\mathcal{CE}(\mu,\nu)} \int_{[0,1]\times X} \left[\|v_t\|^2 + \frac{\kappa^2}{4}\alpha_t^2\right] \mathrm{d}\rho_t \mathrm{d}t$$

- other **unbalanced models:** [Dolbeault et al., 2009; Caffarelli and McCann, 2010; Piccoli and Rossi, 2016]...
- **Thm:** HK is geodesic distance on **non-negative measures**
  - geodesics well understood, weak Riemannian structure
  - transport up to $\frac{\kappa\pi}{2}$, pure Hellinger after that,
    choose κ by physical intuition and cross-validation,
    equiv. to spatial scaling
- **Thm:** Kantorovich-type soft-marginal formulation

$$\mathrm{HK}(\mu,\nu)^2 = \kappa^2 \min_{\pi\in\mathcal{M}_+(X\times X)} \int_{X\times X} c\,\mathrm{d}\pi + \mathrm{KL}(\mathrm{P}_1\pi|\mu) + \mathrm{KL}(\mathrm{P}_2\pi|\nu)$$

for $c(x,y) = \begin{cases} -2\log\cos(\|x-y\|/\kappa) & \text{if } \|x-y\| < \frac{\kappa\pi}{2} \\ +\infty & \text{else} \end{cases}$

- simple **numerical approximation** via **entropic regularization**
  and **Sinkhorn**-type algorithm [Chizat et al., 2018a]
- **barycenters** [Chung and Phung, 2020; Friesecke et al., 2021; Bonafini et al., 2023]

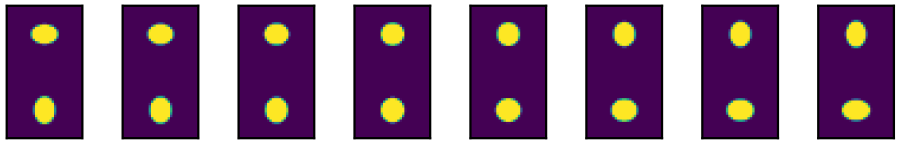


4.1 Hellinger–Kantorovich distance: local linearization

[Cai et al., 2022]

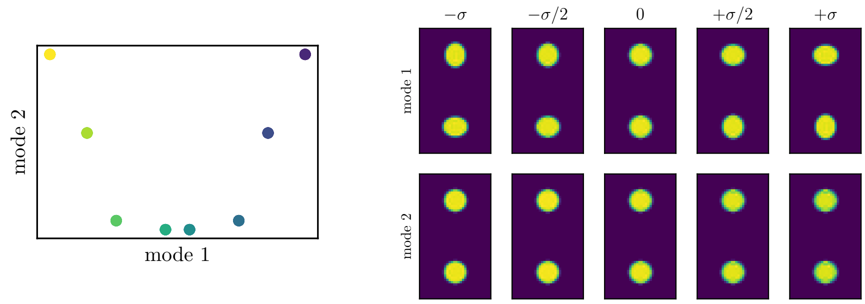
$$\mathrm{HK}(\mu,\nu)^2 := \inf_{(\rho,v,\alpha)\in\mathcal{CE}(\mu,\nu)} \int_{[0,1]\times X} \left[\|v_t\|^2 + \frac{1}{4}\alpha_t^2\right] \mathrm{d}\rho_t \mathrm{d}t$$

Example: (varying ellipticities and radii)

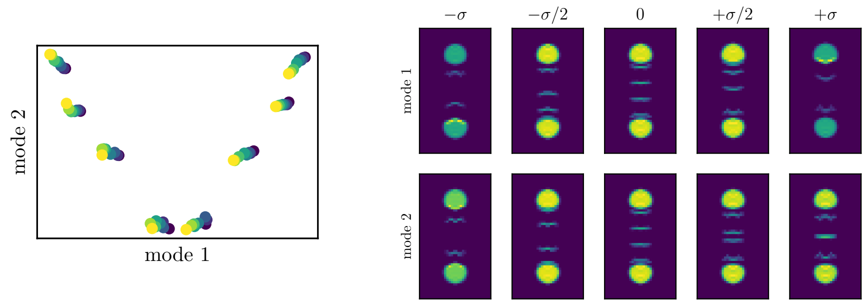


$$\mathrm{Log}_{\mu}^{W_2}(\nu) = v_0 \qquad \mathrm{Log}_{\mu}^{\mathrm{HK}}(\nu) = (v_0, \alpha_0, \sqrt{\nu^{\perp}})$$

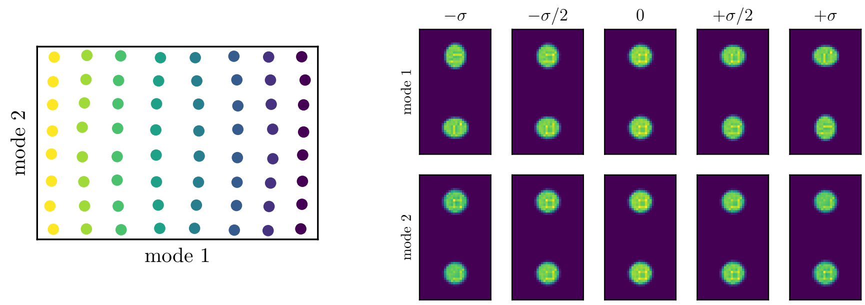
PCA in tangent space: W_2, constant radii



PCA in tangent space: W_2, small radii variations



PCA in tangent space: HK, small radii variations



4.2 Hellinger–Kantorovich distance: logarithmic map

Logarithmic map for W_2:

- $\pi = \underset{\pi \in \mathcal{M}_+(X \times X)}{\mathrm{argmin}} \int d^2 \, \mathrm{d}\pi + \iota_{\{\mu\}}(\mathrm{P}_X \pi) + \iota_{\{\nu\}}(\mathrm{P}_Y \pi) = (\mathrm{id}, T)_{\sharp}\mu$
- $\mathrm{Log}_{\mu}^{W_2}(\nu) = v_0 = T - \mathrm{id}$
- discrete approximation by barycentric projection: $T_i = \frac{1}{\mu_i} \sum_j \pi_{i,j} \, y_j$

Logarithmic map for HK: [**Cai et al., 2022; Sarrazin and Schmitzer, 2023**]

- $\pi = \underset{\pi \in \mathcal{M}_+(X \times X)}{\operatorname{argmin}} \int c^2 \, \mathrm{d}\pi + \mathrm{KL}(\mathrm{P}_X \pi | \mu) + \mathrm{KL}(\mathrm{P}_Y \pi | \nu) = (\mathrm{id}, T)_\sharp \sigma$
- $u = \frac{\mathrm{d}\sigma}{\mathrm{d}\mu}$, $\nu^\perp$: part that is singular w.r.t. $T_\sharp \sigma$

$$v_0 = \frac{T - \mathrm{id}}{\|T - \mathrm{id}\|} \tan(\|T - \mathrm{id}\|) u \qquad \alpha_0 = 2(u-1)$$

- $\mathrm{Log}_\mu^{\mathrm{HK}}(\nu) = (v_0, \alpha_0, \sqrt{\nu^\perp})$

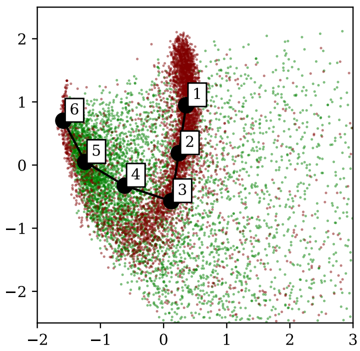
Additional results [**Sarrazin and Schmitzer, 2023**]

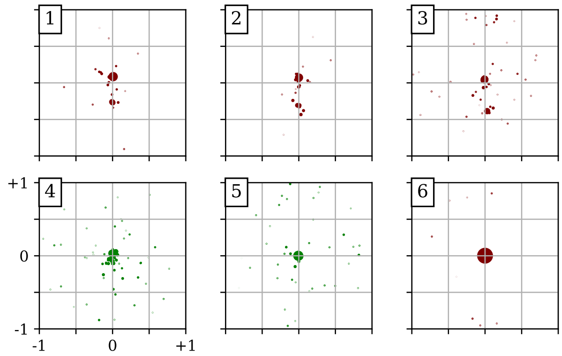
- dual perspective: W_2: $v_0 = -\frac{1}{2}\nabla\phi$, HK: $(v_0, \alpha_0) = (-\frac{1}{2}\nabla\phi, -2\phi)$
- convergence of barycentric projection approximation for W_2 and HK
- extension to OT on manifolds
- extension to **spherical** Hellinger–Kantorovich distance [Laschos and Mielke, 2019]

5 Example applications

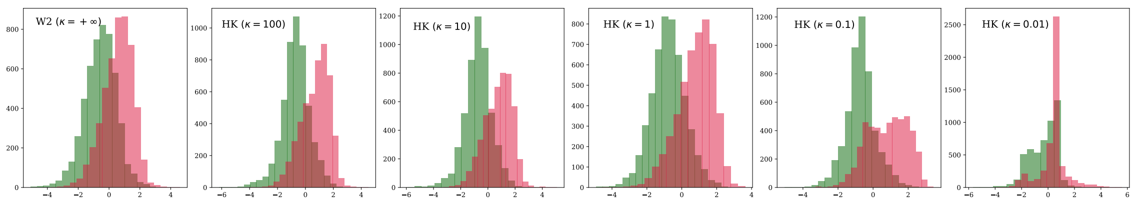
Classification of particle jets [Cai et al., 2022]

- mass represents energy absorbed in detector plane
- separate weak (red) vs strong (green) decay channels

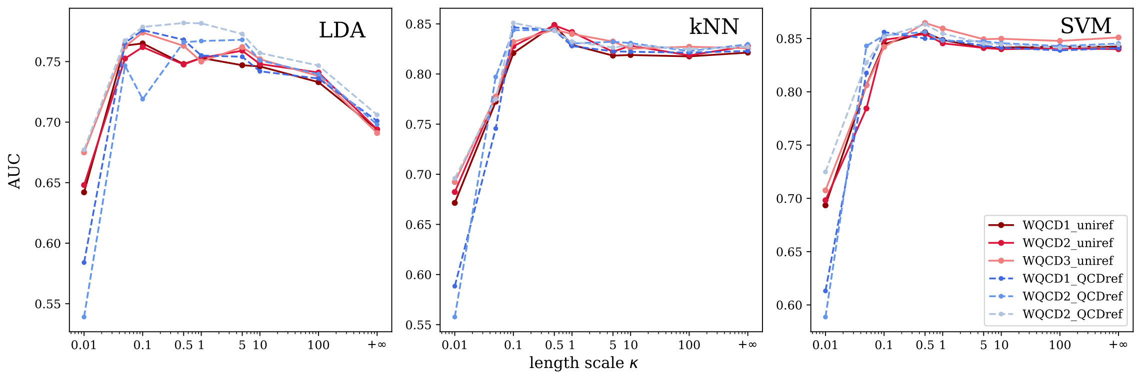




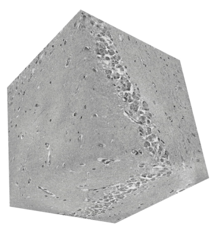
- LDA: better separation with unbalanced HK metric

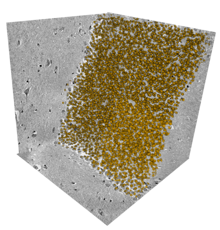


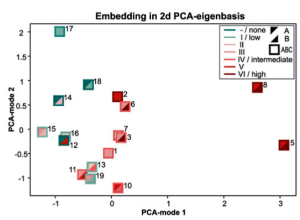
- AUC curves for various classifiers

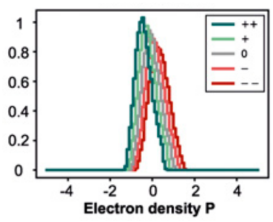


Linearized OT for cell nuclei statistics [Eckermann et al., 2021; Frost et al., 2023]







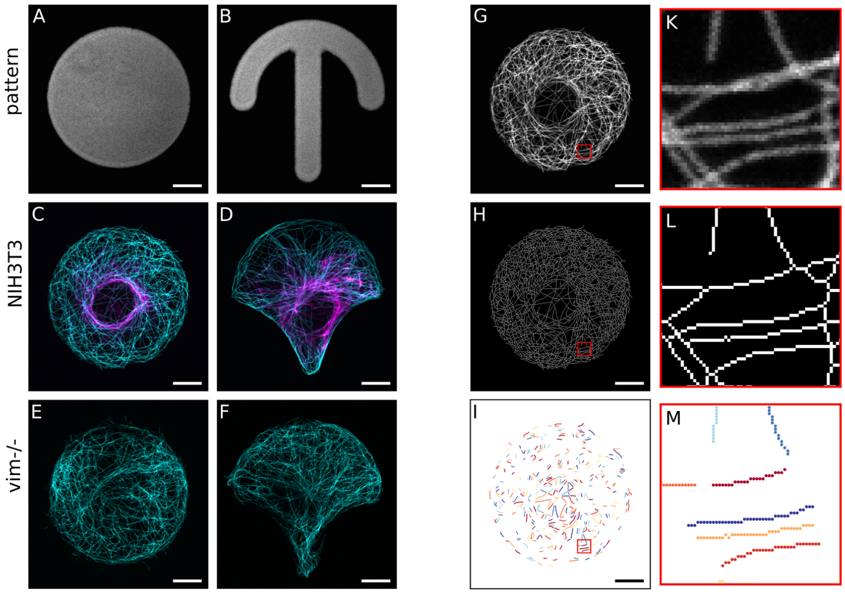


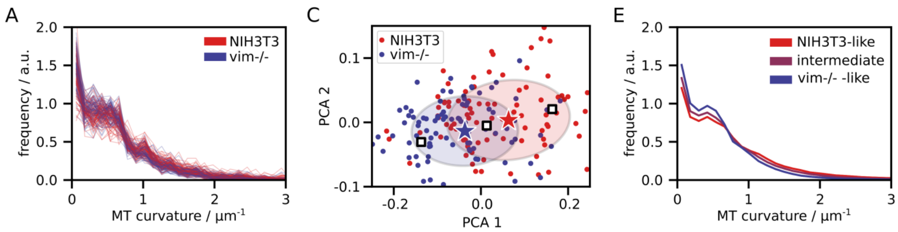
- collaboration with Salditt group, x-ray physics, Göttingen:
  phase contrast x-ray tomography, high resolution 3d images of tissue samples
- **segmentation** of cell nuclei, feature extraction,
  each sample → nuclei distribution on **feature space**
- application to Alzheimer and multiple sclerosis data, try to discover systematic shift in cell (nuclei) population

✗ interpretation of tangent vectors possibly in principle, but still tricky in practice

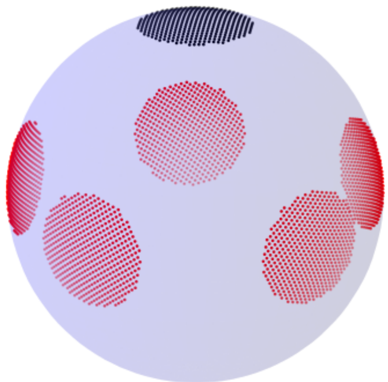
Linearized OT for microtubule curvature analysis in cell microscopy

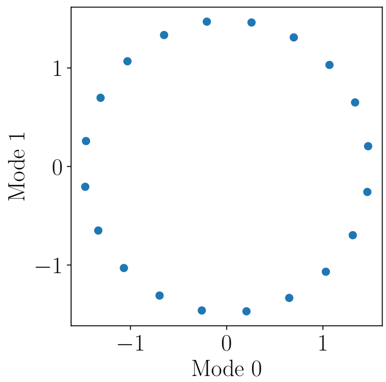
- collaboration with Koester group, x-ray physics, Göttingen:
  fluorescence microscopy images of cells; study impact of vimentin on microtubule curvature
- **segmentation** of microtubule network, extraction of curvature distribution,
  each cell (region) → distribution on **curvature**





Linearized OT on sphere [Sarrazin and Schmitzer, 2023]





6 Conclusion

6.1 Overview

Optimal transport

- ✓ **intuitive, robust, flexible** metric for probability measures
- ✓ **rich geometric structure** (Riemannian flavour. . .)
- ✓ accessible by **convex optimization**

Local linearization of OT [Wang et al., 2012]

- ✓ **Lagrangian representation**: combine OT metric with **linear structure**
- ✓ **intuitive interpretation** of tangent vectors
- ✓ **useful representation** for subsequent machine learning analysis

Unbalanced transport: Hellinger–Kantorovich distance

- ✓ more **robust to mass fluctuations**
- ✓ carries over to **linearization** [Cai et al., 2022; Sarrazin and Schmitzer, 2023]
- ✓ hyperparameter κ **easy to tune**
- ✓ formulas look scary, but **numerics** almost the same

Example code

- https://github.com/bernhard-schmitzer/UnbalancedLOT
- https://gitlab.gwdg.de/bernhard.schmitzer/linot

6.2 Open questions

How well does the linear approximation work?

- for general Wasserstein-2 case: [Gigli, 2011; Mérigot et al., 2020; Delalande and Merigot, 2021], expect: better on 'nice sub-manifolds'
- still open for HK

Riemannian structure of HK metric

- is {range of the logarithmic map} = {domain of the exponential map} convex?
- ✓ dual perspective of logarithmic map (W_2: $v_0 = -\frac{1}{2}\nabla\phi$. HK: $(v_0, \alpha_0) = (-\frac{1}{2}\nabla\phi, -2\phi)$)
- ✓ regularity of logarithmic map?

Beyond simple one-point-linearization

- multiple support points? 'local triangulation' of a sub-manifold?
- barycentric subspace analysis [Pennec, 2018; Bonneel et al., 2016]?

Statistical questions

- how robust is the analysis under sampling of the samples?
- what if samples are themselves only empirical measures?

Better interpretation of tangent vectors

- relevant for medical imaging

References

F. Beier, R. Beinert, and G. Steidl. On a linear Gromov–Wasserstein distance. arXiv:2112.11964, 2021.

J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

M. Bonafini, O. Minevich, and B. Schmitzer. Hellinger–Kantorovich barycenter between Dirac measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 29:19, 2023. doi: 10.1051/cocv/2022088.

N. Bonneel, G. Peyré, and M. Cuturi. Wasserstein barycentric coordinates: Histogram regression using optimal transport. *ACM Trans. Graph.*, 35(4), 2016. doi: 10.1145/2897824.2925918.

Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Annals of Math.*, 171(2):673–730, 2010.

T. Cai, J. Cheng, B. Schmitzer, and M. Thorpe. The linearized Hellinger–Kantorovich distance. *SIAM J. Imaging Sci.*, 15(1):45–83, 2022. doi: 10.1137/21M1400080.

L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced optimal transport problems. *Math. Comp.*, 87:2563–2609, 2018a. doi: 10.1090/mcom/3303.

L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and Fisher–Rao metrics. *Found. Comp. Math.*, 18(1):1–44, 2018b. doi: 10.1007/s10208-016-9331-y.

N.-P. Chung and M.-N. Phung. Barycenters in the Hellinger–Kantorovich space. *Appl Math Optim*, 2020. doi: 10.1007/s00245-020-09695-y.

A. Delalande and Q. Merigot. Quantitative stability of optimal transport maps under variations of the target measure. arXiv:2103.05934, 2021.

J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009. doi: 10.1007/s00526-008-0182-5.

M. Eckermann, B. Schmitzer, F. van der Meer, J. Franz, O. Hansen, C. Stadelmann, and T. Salditt. Three-dimensional virtual histology of the human hippocampus based on phase-contrast computed tomography. *PNAS*, 118(48):e2113835118, 2021. doi: 10.1073/pnas.2113835118.

G. Friesecke, D. Matthes, and B. Schmitzer. Barycenters for the Hellinger–Kantorovich distance over Rd. *SIAM J. Math. Anal.*, 53(1):62–110, 2021. doi: 10.1137/20M1315555.

J. Frost, B. Schmitzer, M. Toepperwien, M. Eckermann, J. Franz, C. Stadelmann, and T. Salditt. 3d virtual histology reveals pathological alterations of cerebellar granule cells in multiple sclerosis. *Neuroscience*, 2023. doi: 10.1016/j.neuroscience.2023.04.002.

N. Gigli. On Holder continuity-in-time of the optimal transport map towards measures along a curve. *Proceedings of the Edinburgh Mathematical Society*, 54(2):401–409, 2011.

L. V. Kantorovich. O peremeshchenii mass. *Doklady Akademii Nauk SSSR*, 37(7–8):227–230, 1942.

V. Khurana, H. Kannan, A. Cloninger, and C. Moosmüller. Supervised learning of sheared distributions using linearized optimal transport. arXiv:2201.10590, 2022.

S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.

V. Laschos and A. Mielke. Geometric properties of cones with applications on the Hellinger–Kantorovich space, and a new distance on the space of probability measures. *J. Funct. Anal.*, 276(11):3529–3576, 2019. doi: 10.1016/j.jfa.2018.12.013.

M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018. doi: 10.1007/s00222-017-0759-8.

R. J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1): 153–179, 1997.

Q. Mérigot, A. Delalande, and F. Chazal. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3186–3196, 2020.

F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001. doi: 10.1081/PDE-100002243.

X. Pennec. Barycentric subspace analysis on manifolds. *Ann. Statist.*, 46(6A):2711–2746, 2018. doi: 10.1214/17-AOS1636.

B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Arch. Rat. Mech. Analysis*, 222(3):1339–1365, 2016. doi: 10.1007/s00205-016-1026-7.

C. Sarrazin and B. Schmitzer. Linearized optimal transport on manifolds. arXiv:2303.13901, 2023.

W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *Int. J. Comp. Vision*, 101:254–269, 2012. doi: 10.1007/s11263-012-0566-z.