A Barycentric Coding Model for Measures on Graphs

Synthesis and analysis of measures via discrete transport geometry

David Gentile · FSU Topology & Geometry Seminar · 29 September 2026

Structure of the talk

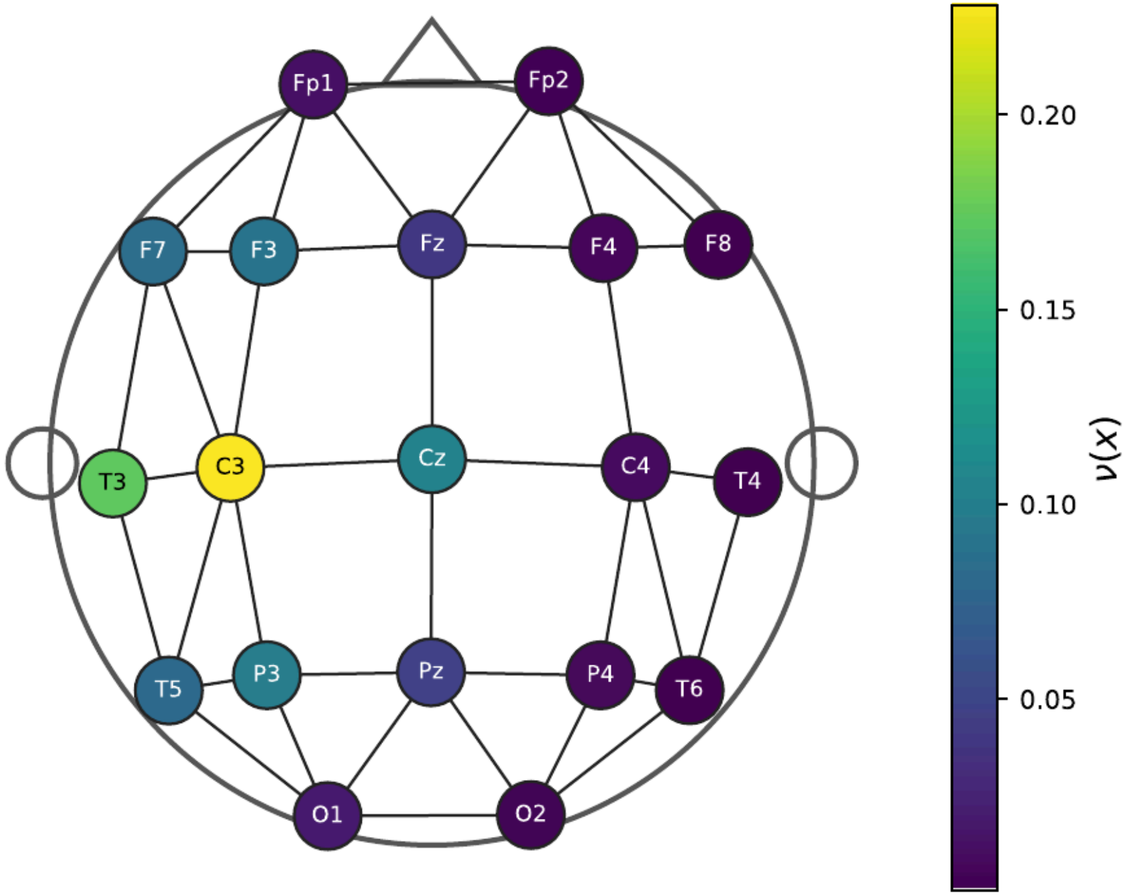
> **New in this talk:** exponential and logarithmic maps for the discrete transport geometry computed by Hamiltonian shooting, with no time discretization of the variational problem, put to work in barycentric synthesis and analysis for measures on graphs.

1. Background: the Kantorovich problem, the Wasserstein metric W_2, and the barycentric coding model (BCM) it supports
2. Why the Kantorovich picture breaks on a graph
3. Solution 1: keep Kantorovich and regularize (entropic transport on a graph metric)
4. Solution 2: start from Benamou–Brenier instead, and build a Riemannian structure on measures that sees the edges; implement the BCM there
5. New: computing the exponential and logarithmic maps of that structure by Hamiltonian shooting
6. Numerical experiments, and where this might go next

Spatial Signal Processing

Signals that live on graphs

- Social, transportation, sensor, and neuronal networks all carry data on the vertices of a weighted graph; today's data are probability measures on the vertices
- There is no grid and no coordinates, only the edges: the geometry of the domain *is* the graph
- **The question:** how do we average such measures, and decompose one into others, in a way that respects that geometry?

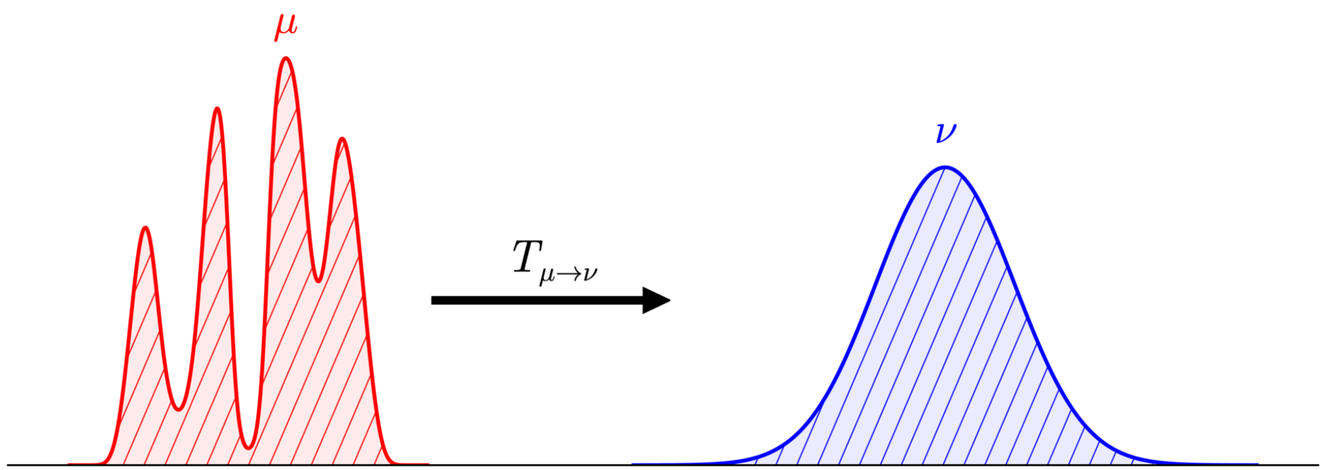


A synthetic measure on the standard 10-20 system graph

Shuman, Narang, Frossard, Ortega, Vandergheynst. "The emerging field of signal processing on graphs." *IEEE Signal Processing Magazine* 30.3 (2013): 83–98.

Optimal Transport:

A Tool for Comparing Probability Distributions



Optimal Transport Defines a Metric on Probability Measures

Kantorovich Problem, Wasserstein Geometry

- Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be finite variance probability measures on $\mathbb{R}^d$. Then the optimal transport cost, denoted by W_2, between μ and ν is given by the following optimization problem:

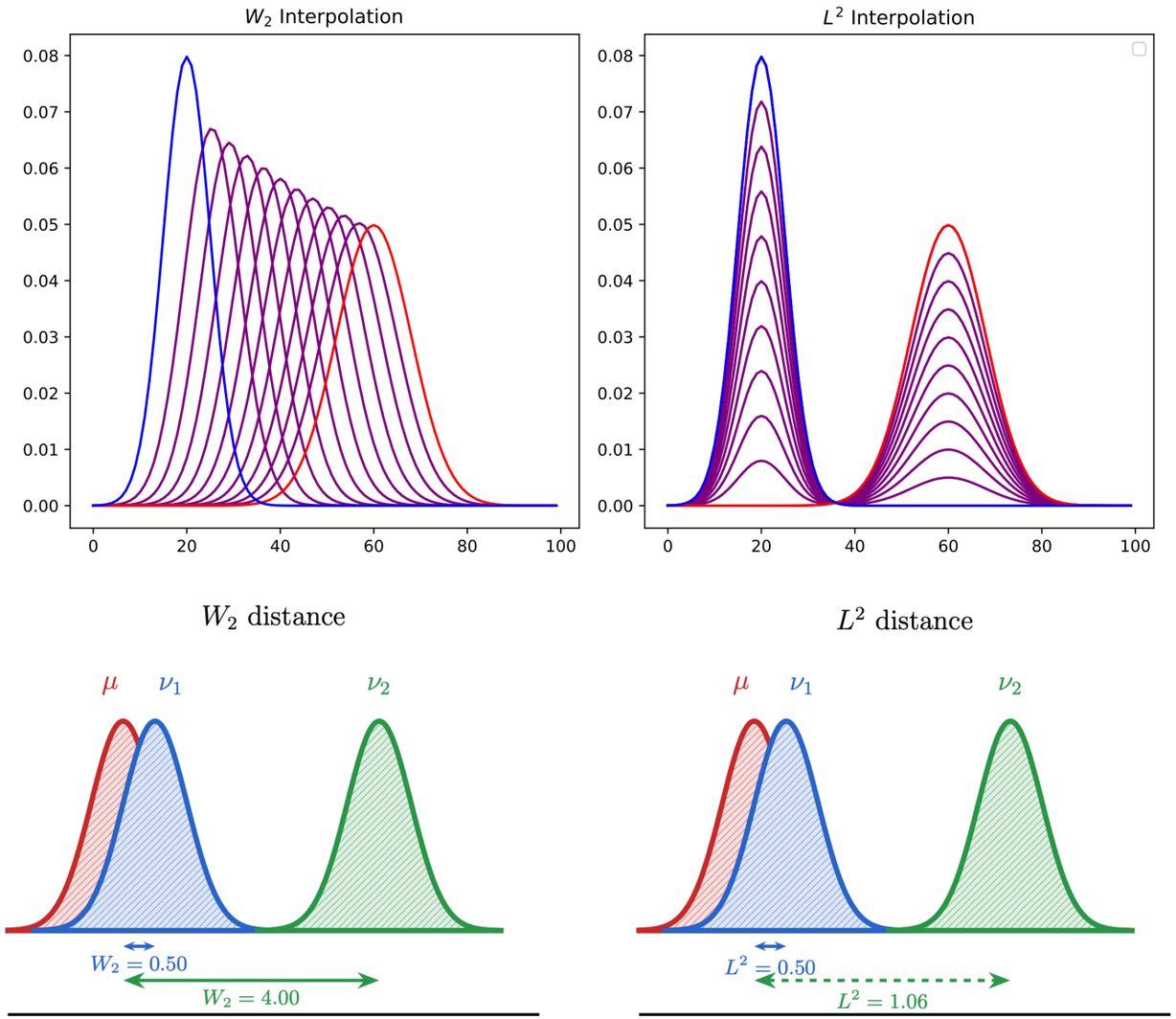
$$W_2^2(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}.$$

- This is a special case of the Kantorovich problem (KP), but for historical reasons the distance between measures defined by this problem is referred to as the $(L^2$-)Wasserstein distance

Background on Optimal Transport

How does transport encode information about the underlying domain?

- The W_2 distance between two distributions tells us how expensive it is to move one to the other “physically”
- This contrasts with, e.g., the L^2 difference between the densities of two distributions, where mass “teleports”



Barycentric Coding Models

Averaging without linear structure

- In $\mathbb{R}^d$, a barycenter (AKA a weighted average) $\bar{x} = \sum \lambda_i x_i$ of points $\{x_i\}_{i=1}^p$ is *also* a minimizer for the function

$$x \mapsto \sum_i \frac{\lambda_i}{2} \|x - x_i\|^2$$

- The latter is expressed through a squared distance, so it generalizes. With a transport distance it defines *transport barycenters* (Agueh–Carlier 2011 for W_2): for a dictionary $\mathcal{R} = \{\nu_i\}_{i=1}^p$ and weights $\lambda \in \Delta^{p-1}$, i.e. $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$,

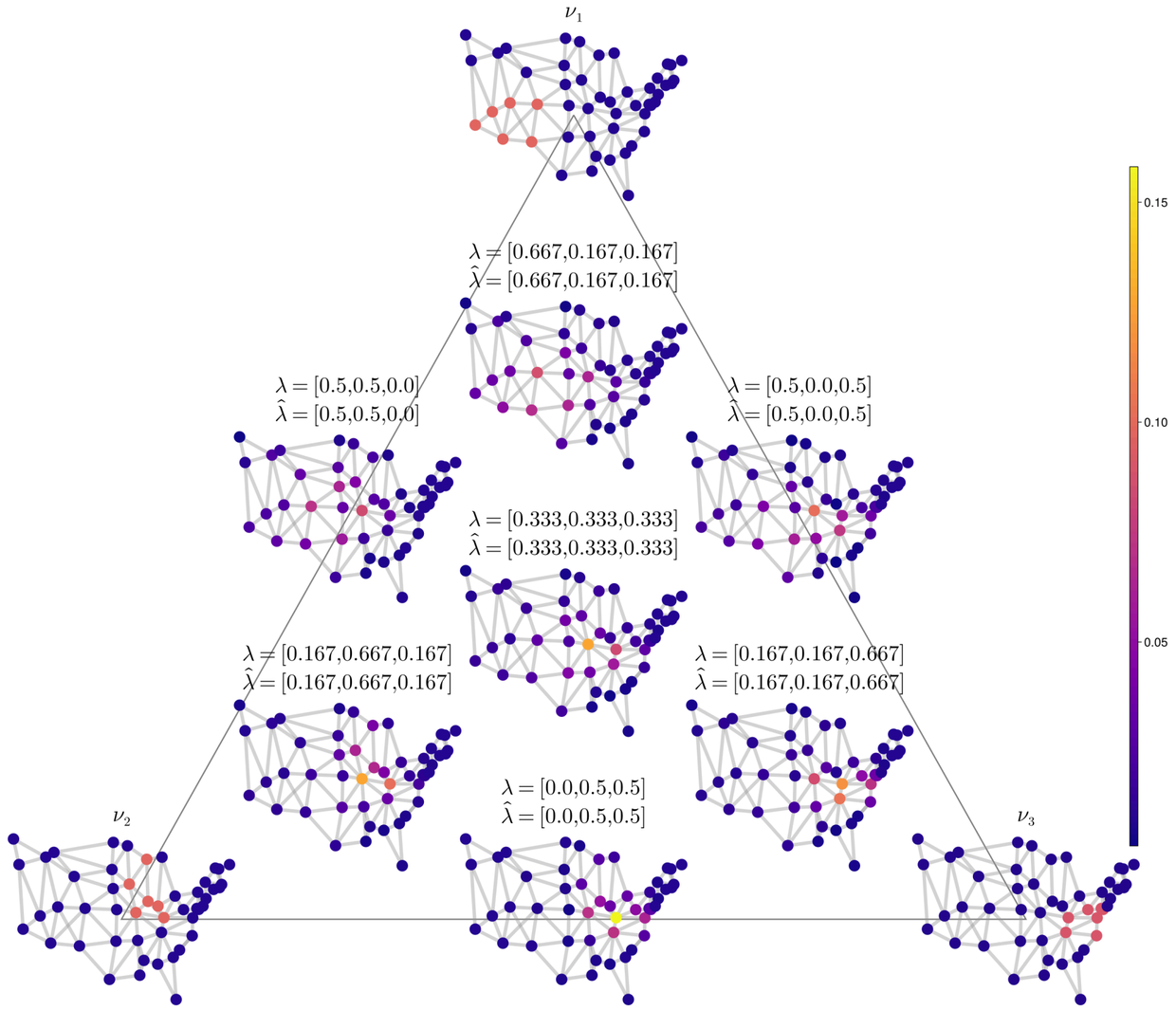
$$\nu_\lambda = \arg\min \left\{ \mathcal{J}_{\mathcal{R},\lambda}[\nu] = \sum_i \frac{\lambda_i}{2} W_2^2(\nu, \nu_i) \right\}$$

- Two jobs: ***synthesis***, given $\mathcal{R}, \lambda$ compute ν_λ; ***analysis***, given ν_0 and $\mathcal{R}$ recover λ

Agueh, Carlier (2011); Werenski et al. (2022); in GW geometry, Chowdhury, Needham (2020) and Díaz Martín, Medri, Murphy (2025).

Example BCM

- Vertices of the triangle: the dictionary atoms, three measures on the graph of the contiguous US states
- Points inside: syntheses of the atoms
- Each is annotated with its true λ and the $\hat{\lambda}$ recovered by analysis: to 4×10^{-8} with Hamiltonian shooting; the paper's Chambolle-Pock analysis of the same measures lands at 10^{-2}–10^{-3}
- Computed with the graph transport geometry of this talk; by the end you will know how



Definitions

Graphs, Markov chains, and measures on them

- A graph $\mathcal{X} = (V, E)$, simple and connected, with symmetric edge weights $\omega : E \to \mathbb{R}^+$ ($\omega \equiv 1$ if unspecified), represented by the transition kernel Q of its random walk and the steady state distribution π:

$$Q(x,y) = \frac{\omega(x,y)}{\deg_\omega(x)} \text{ for } (x,y) \in E,\ 0 \text{ otherwise}; \qquad \pi(x) = \frac{\deg_\omega(x)}{\sum_{y \in V} \deg_\omega(y)}; \qquad \pi(x)Q(x,y) = \pi(y)Q(y,x)$$

- A *function* φ on $\mathcal{X}$ lives on the vertices, $\varphi \in \mathbb{R}^{|V|}$; a *vector field* Φ lives on the edges, $\Phi \in \mathbb{R}^{|V| \times |V|}$
- Probability measures are densities with respect to π:

$$\mathcal{P}(\mathcal{X}) = \left\{ \nu \in \mathbb{R}_+^{|V|} : \sum_{x \in \mathcal{X}} \nu(x)\pi(x) = 1 \right\}$$

  a simplex; its *interior* is $\{\nu > 0\}$, its boundary the measures that vanish somewhere

Why is transport on graphs a problem?

Degeneracy of the Classical Transport Geometry when $|\Omega| < \infty$

- Exercise to the reader: Let $\mathcal{X}_2 = \{a, b\}$ be the two-point graph with a single edge. Characterize every coupling for a fixed pair of measures $\nu_\alpha, \nu_\beta \in \mathcal{P}(\mathcal{X}_2)$ and *solve KP explicitly*
- Conclude that $(\mathcal{P}(\mathcal{X}_2), W_2) \simeq \left([0,1], \sqrt{|x-y|}\right)$
- Further conclude that if $\nu(t) : [0,1] \to \mathcal{P}(\mathcal{X}_2)$ is a geodesic curve then it is constant! So you can't even define a shortest-path between two measures on a two point graph.
- **Moral of the story:** transport on a discrete space breaks if we come at it from the POV of coupled probability measures / KP

Two Ways Out

What do we do computationally about the degenerate transport geometry?

- The Kantorovich problem is degenerate on a graph, but it is a linear program on matrices, and the BCM only needs a distance and a way to minimize $\mathcal{J}$
- **Solution 1 (static):** keep the Kantorovich problem. Regularize it so it stops being degenerate, and feed it a cost matrix built from the graph. Cheap, and the graph is seen only through pairwise distances
- **Solution 2 (dynamic):** replace the Kantorovich problem by the fluid-mechanical picture of W_2 (Benamou–Brenier, coming up): W_2 as the least kinetic energy of a flow carrying one measure to the other. Kinetic energy and flows make sense on a graph, and they define a Riemannian metric on $\mathcal{P}(\mathcal{X})$ that lives on the edges; the price is solving differential equations on the graph instead of a linear program

Solution 1: Keep Kantorovich, Regularize

Entropic transport on a ground metric built from the graph

A “Static” Approach to Barycenters of Measures

Entropic regularization: four steps to a BCM on a graph

1. **Cuturi 2013.** Regularize the Kantorovich problem with an entropy term: strongly convex, and Sinkhorn’s algorithm computes transport costs at “lightspeed”

$$\mathrm{OT}_{reg}(\mu,\nu) = \inf\left\{\langle \gamma, C\rangle + \varepsilon_{reg} H(\gamma) : \gamma \in \Pi(\mu,\nu)\right\}, \quad H(\gamma) = -\sum_{x,y\in\mathcal{X}} \gamma(x,y)\left(\log \gamma(x,y) - 1\right)$$

2. **Benamou, Carlier, Cuturi, Nenna, Peyré 2015.** Regularized barycenters of empirical measures by iterative Bregman projections
3. **Bonneel, Peyré, Cuturi 2016.** On a fixed support, differentiate through Sinkhorn to fit the weights: a first BCM with both synthesis and analysis
4. **G., Murphy 2025.** Plug a cost matrix built from the graph into that scheme (next slide)

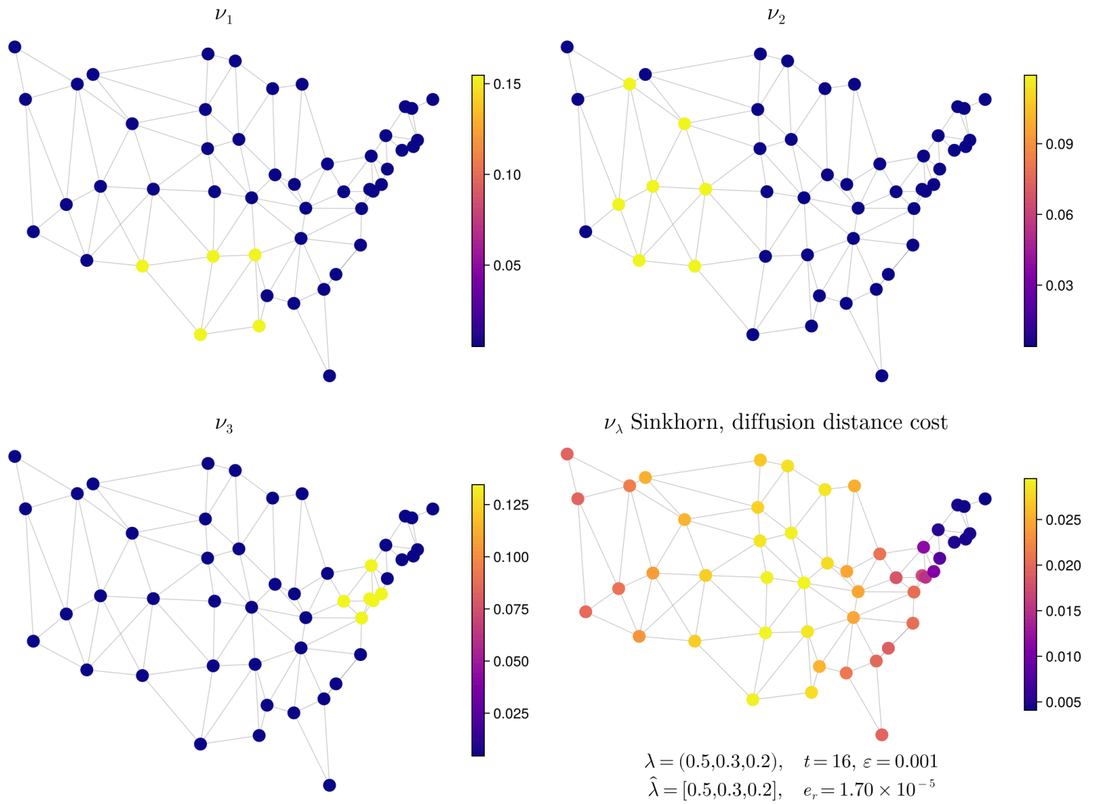
Cuturi. “Sinkhorn distances: lightspeed computation of optimal transport.” *NeurIPS* 26 (2013). Bonneel, Peyré, Cuturi. “Wasserstein barycentric coordinates: histogram regression using optimal transport.” *ACM Trans. Graph.* 35.4 (2016): 71.

Benamou, Carlier, Cuturi, Nenna, Peyré. “Iterative Bregman projections for regularized transportation problems.” *SIAM J. Sci. Comput.* 37.2 (2015): A1111–A1138.

Solution 1 in Practice

Regularized barycenters on the US states graph (G., Murphy, SampTA 2025)

- The cost matrix comes from the graph: squared shortest-path distance, squared *diffusion distance* of the random walk at time t, or a Fermat distance; Bonneel et al.'s fixed-support scheme does the rest
- It works: synthesis and analysis both run, and the coordinates are recovered to a relative error of 2×10^{-5}
- The drawback: a pairwise distance matrix does not encode edge connectivity, so the graph is seen only through distances; the entropy biases every output, and the geometry degenerates again as $\varepsilon_{reg} \to 0$. What would a barycenter that sees the edges *directly* look like?

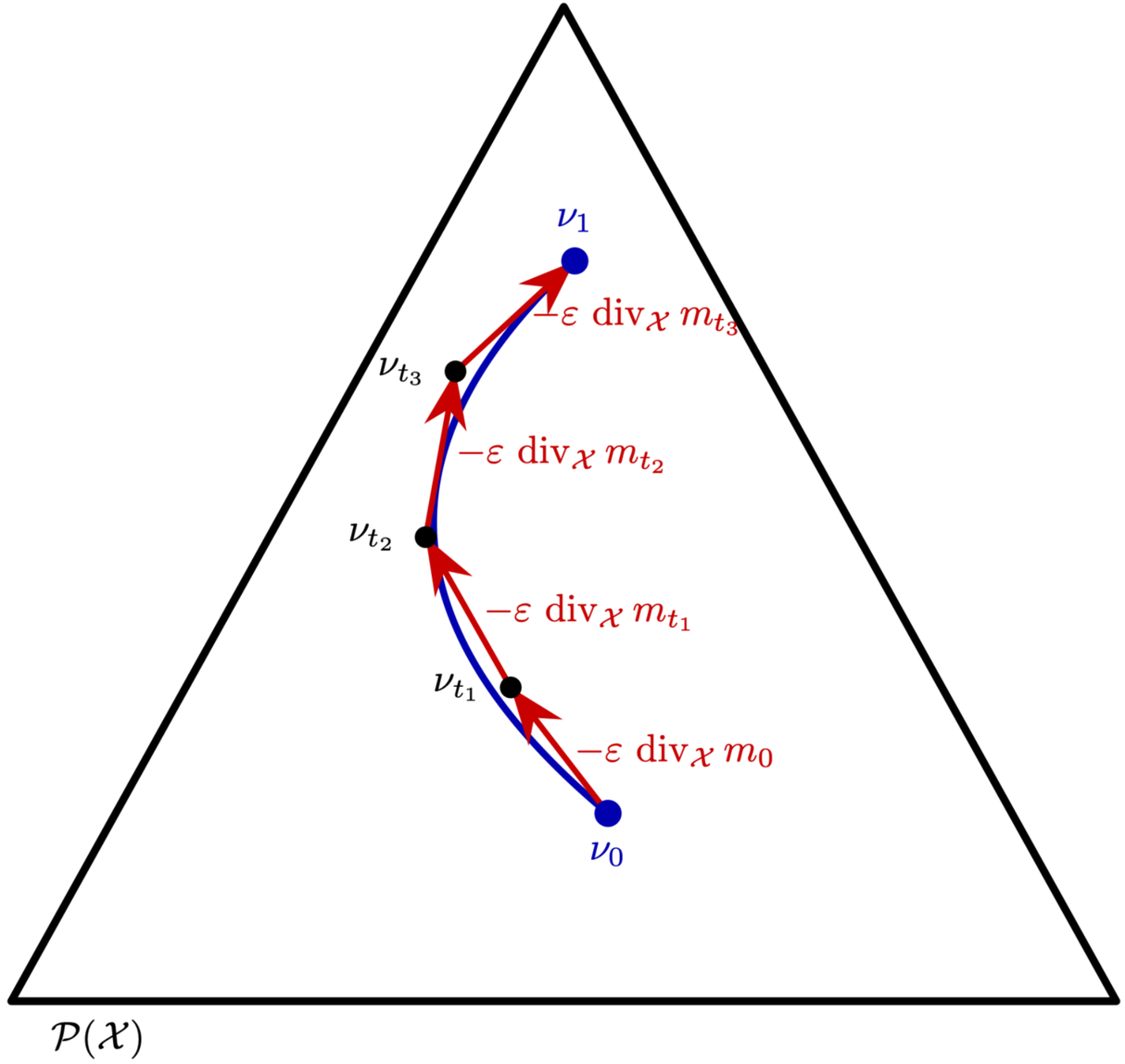


The SampTA experiment redone with the current code: squared diffusion-distance cost at $t = 16$, $\varepsilon_{reg} = 10^{-3}$, $\lambda = (0.5, 0.3, 0.2)$; the recovered $\hat{\lambda}$ and its relative error are under the panel

G., Murphy. "Regularized 2-Wasserstein barycenters on discrete metric spaces." *SampTA* 2025.

Solution 2: Start from Benamou–Brenier

A Riemannian structure for transport on graphs



The Fluid-Mechanical Picture of W_2

Benamou-Brenier Formula

- A second characterization of W_2: it arises from a Riemannian structure on an "infinite-dimensional manifold" of densities, via the *Benamou–Brenier formula*

$$W_2^2(\rho_0, \rho_1) = \inf \left\{ \mathcal{A}(\rho(x,t), V(x,t)) = \int_0^1 \|V(t,x)\|^2_{L^2(\rho(t,x))}\, dt \right\}$$

- The infimum is taken over pairs $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, $V : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ which satisfy the *continuity equation with temporal boundary conditions:*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0; \qquad \rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1$$

Benamou, Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem." *Numer. Math.* 84.3 (2000): 375–393. The Riemannian reading is Otto's: "The geometry of dissipative evolution equations: the porous medium equation." *Comm. PDE* 26 (2001).

Continuity Equations on Graphs

The admissible mean

- The construction that follows is due to Maas (2011), independently Mielke (2011) and Chow–Huang–Li–Zhou (2012); its Riemannian geometry is developed in Erbar–Maas (2012)
- Plan: take the right-hand side of Benamou–Brenier as the *definition* of a distance on $\mathcal{P}(\mathcal{X})$. On a graph, ∇ and $\nabla\cdot$ are easy (differences along edges, sums over neighbours); the trouble is the product ρV in the continuity equation, a function on *vertices* times a vector field on *edges*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0$$

- An *admissible mean* is a function $\theta : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$; it must be concave, lower semi-continuous, 1-homogeneous, symmetric, and C^2 on $(0,\infty)^2$.
- Examples:

$$\theta_{\log}(s,t) = \frac{s-t}{\log(s)-\log(t)}; \qquad \theta_{geo}(s,t) = \sqrt{st}$$

- For $\rho \in \mathcal{P}(\mathcal{X})$, define $\theta(\rho)(x,y) = \theta(\rho(x),\rho(y))$; now θ turns functions into vector fields

Continuity Equations on Graphs, Cont’d

Finally: we write down a continuity equation

- We now have the language to express a *discrete continuity equation*

$$\mathcal{CE}(\rho_A, \rho_B) := \left\{ (\rho, m) : \frac{\partial \rho}{\partial t} + \operatorname{div}_{\mathcal{X}}(\theta(\rho) \odot m) = 0,\ \rho(0,\cdot) = \rho_A,\ \rho(1,\cdot) = \rho_B \right\}$$

$$\mathcal{A}(\rho, m) := \frac{1}{2} \int_0^1 \sum_{i,j \in \mathcal{X}} \frac{m(t,i,j)^2}{\theta(\rho(t,i), \rho(t,j))}\, Q(i,j)\pi(i)\, dt.$$

$$\mathcal{W}(\rho_A, \rho_B) := \inf \left\{ \sqrt{\mathcal{A}(\rho, m)} : (\rho, m) \in \mathcal{CE}(\rho_A, \rho_B) \right\}.$$

Read: the momentum m on each edge is charged against the mean θ of the densities at its two ends, and $\mathcal{W}$ is the least such kinetic energy along a path from ρ_A to ρ_B. Benamou–Brenier, with ρV replaced by $\theta(\rho) \odot m$.

Maas, Jan. “Gradient flows of the entropy for finite Markov chains.” *Journal of Functional Analysis* 261.8 (2011): 2250–2292.

The $\mathcal{W}$ Geometry

Riemannian Geometry Objects: Tangent Space & Metric Tensor

- We'll need to understand two things about the Riemannian geometry of $(\mathcal{P}(\mathcal{X}), \mathcal{W})$: the metric tensor, and the exponential mapping
- First, we need to say what the tangent space is, and how we can compute the inner product of two vectors at a point.

$$T_\nu \mathcal{P}(\mathcal{X}) = \left\{ \nabla\varphi \in \mathbb{R}^{|V|\times|V|} : \varphi \in \mathbb{R}^{|V|} \right\},$$

$$\langle \nabla\varphi, \nabla\psi \rangle_\nu = \frac{1}{2} \sum_{x,y\in V} \theta(\nu(x), \nu(y))\, \nabla\varphi(x,y)\, \nabla\psi(x,y).$$

The $\mathcal{W}$ Geometry, Cont'd

The two maps the rest of the talk computes

- With $\gamma_{\nu,m}$ the geodesic from ν with initial velocity $m \in T_\nu \mathcal{P}(\mathcal{X})$:

$$\exp_\nu(m) = \gamma_{\nu,m}(1), \qquad \exp_\nu^{-1}(\gamma_{\nu,m}(1)) = m, \qquad \exp_\nu : U \subset T_\nu \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$$

- $\exp_\nu^{-1}$ is single-valued on the ball of the *injectivity radius*; the radius of $(\mathcal{P}(\mathcal{X}), \mathcal{W})$ is not known, and everything below assumes we stay inside it
- The geodesic equation of $(\mathcal{P}(\mathcal{X}), \mathcal{W})$ is explicit, as a Hamiltonian system in (ρ, φ); that is the shooting section

An Implementation of a Barycentric Coding Model for Graphs

What existed, and what we add

Barycentric coding on graphs, before and after

- **Before.** Barycentric synthesis and analysis for entropic W_2 on a fixed support (Bonneel, Peyré, Cuturi 2016): Solution 1. For $\mathcal{W}$, a convex program for one geodesic (Erbar et al. 2020), no barycenters
- **What we add.** A BCM on $(\mathcal{P}(\mathcal{X}), \mathcal{W})$ built from its Riemannian structure: intrinsic gradient descent for synthesis, a tangent-space Gram matrix for analysis, both driven by $\exp_\nu$ and $\exp_\nu^{-1}$
- **Two ways to compute those maps.** Erbar et al.'s space-time convex program (the paper): any pair of measures, sparsely supported ones included. Hamiltonian shooting on the geodesic equations (new): one to two orders of magnitude faster, interior measures only

Erbar, Rumpf, Schmitzer, Simon (2020); Chow, Li, Zhou (2020); Cui, Liu, Zhou (2021); Cui, Dieci, Zhou, *SIAM J. Sci. Comput.* (2022).

Synthesis: minimizing $\mathcal{J}$

Getting back towards barycenters

- Recall that the variance functional is expressed directly in terms of a metric!

$$\mathcal{J}[\nu] = \sum_{i=1}^{p} \frac{\lambda_i}{2} \mathcal{W}^2(\nu_i, \nu)$$

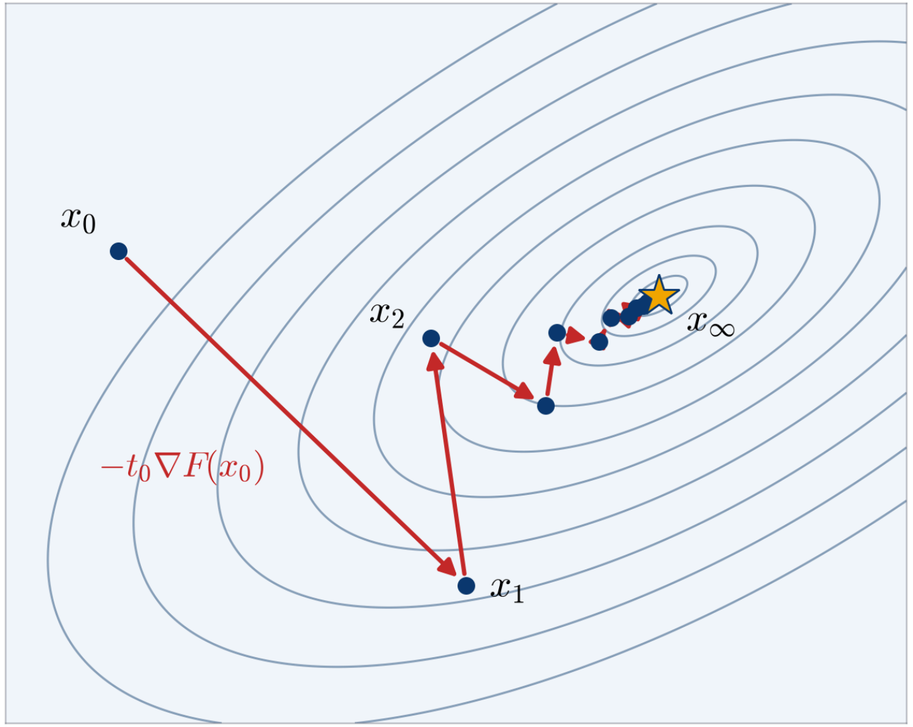
- Minimizers exist: $\nu \mapsto \mathcal{W}^2(\nu_i, \nu)$ is convex along *linear* interpolations (Erbar–Maas 2012), so $\mathcal{J}$ is convex and continuous on a compact convex set. Uniqueness, and convexity along *geodesics*, are open
- To find one, the oldest trick in the book: gradient descent. Find the direction in which $\mathcal{J}$ decreases fastest and take a small step that way
- Two problems: how do I know what direction to walk in, and how do I walk in it?

Erbar, Maas. “Ricci curvature of finite Markov chains via convexity of the entropy.” *Arch. Ration. Mech. Anal.* 206 (2012): 997–1038.

Synthesis Plan: Intrinsic Gradient Descent

When in Non-Euclidean spaces, do as the Non-Euclidean Optimizers Do

- Gradient descent, $x_{k+1} = x_k - t_k \nabla F(x_k)$: the same idea on $\mathcal{P}(\mathcal{X})$, with the *intrinsic gradient* of the Riemannian structure and steps along its geodesics



Absil, Mahony, Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton (2008). Huang, Absil, Gallivan. “ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds.” *ACM TOMS* (2016).

Synthesis Plan: Reading the Two Maps

What $\exp^{-1}$ and $\exp$ mean on a graph, and how the paper approximated the step

- **Direction, $\exp_\nu^{-1}(\nu_i)$:** the initial momentum $m_i(0)$ of the geodesic from ν to ν_i: an edge field, the tangent vector from the geometry section, the first output of any geodesic solver. The descent direction for $\mathcal{J}$ is $m = \sum_i \lambda_i m_i(0)$
- **Step, $\exp_\nu(\varepsilon m)$:** the point at time ε on the geodesic launched from ν with momentum m. Geodesics solve the continuity equation, so near the launch point the step is one Euler step of it:

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}_{\mathcal{X}}\left(\theta(\rho) \odot m\right) \quad \Longrightarrow \quad \exp_\nu(\varepsilon m) = \rho(\varepsilon) \approx \nu - \varepsilon \operatorname{div}_{\mathcal{X}}\left(\theta(\nu) \odot m\right)$$

- **The heuristic:** replace $\exp$ by this first-order update: momenta from the geodesic solver (next slides), one sparse product for the step, mass conserved because $\operatorname{div}_{\mathcal{X}}$ sums to zero
- **What it costs:** the error is $O(\varepsilon^2)$ per step, so ε must be small and the descent takes many steps, each needing p fresh geodesics

Synthesis Plan: Intrinsic Gradient Descent

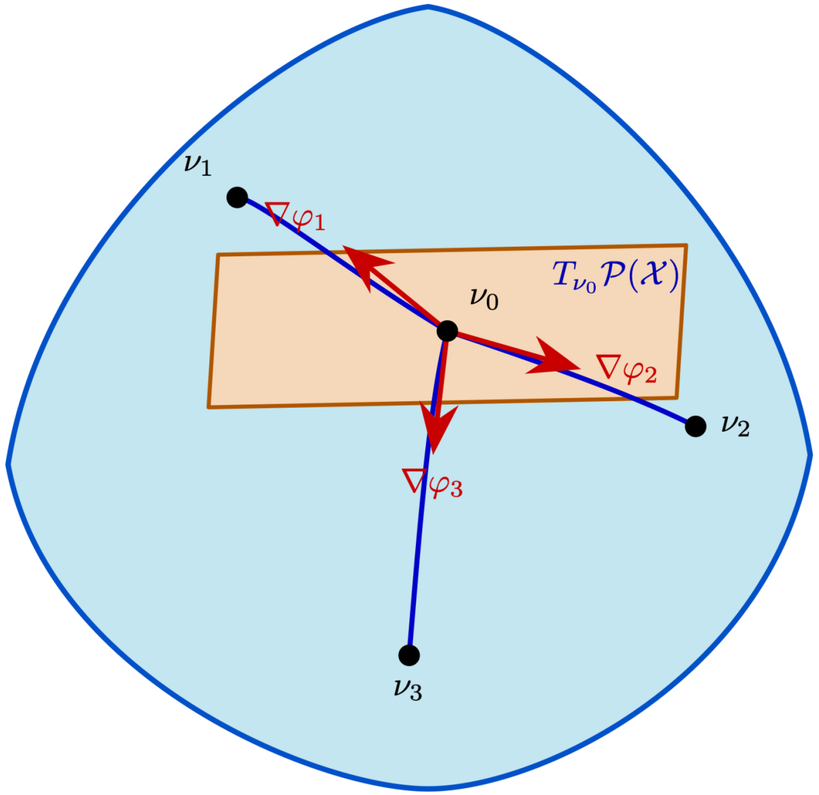
When in Non-Euclidean spaces, do as the Non-Euclidean Optimizers Do

- The update in Riemannian calculus: the direction is a sum of *logarithmic maps*, the step an *exponential map*. Those two maps are all the BCM needs. Last line: the paper's Euler step from the previous slide; shooting computes both maps directly

$$\nu_{k+1} = \exp_{\nu_k}\Big(-t_k \nabla_{\mathcal{W}}\Big\{\sum_{i=1}^{p}\frac{\lambda_i}{2}\mathcal{W}^2(\nu_i,\nu)\Big\}\Big)$$

$$= \exp_{\nu_k}\Big(t_k\Big(\sum_{i=1}^{p}\lambda_i \exp_{\nu_k}^{-1}(\nu_i)\Big)\Big)$$

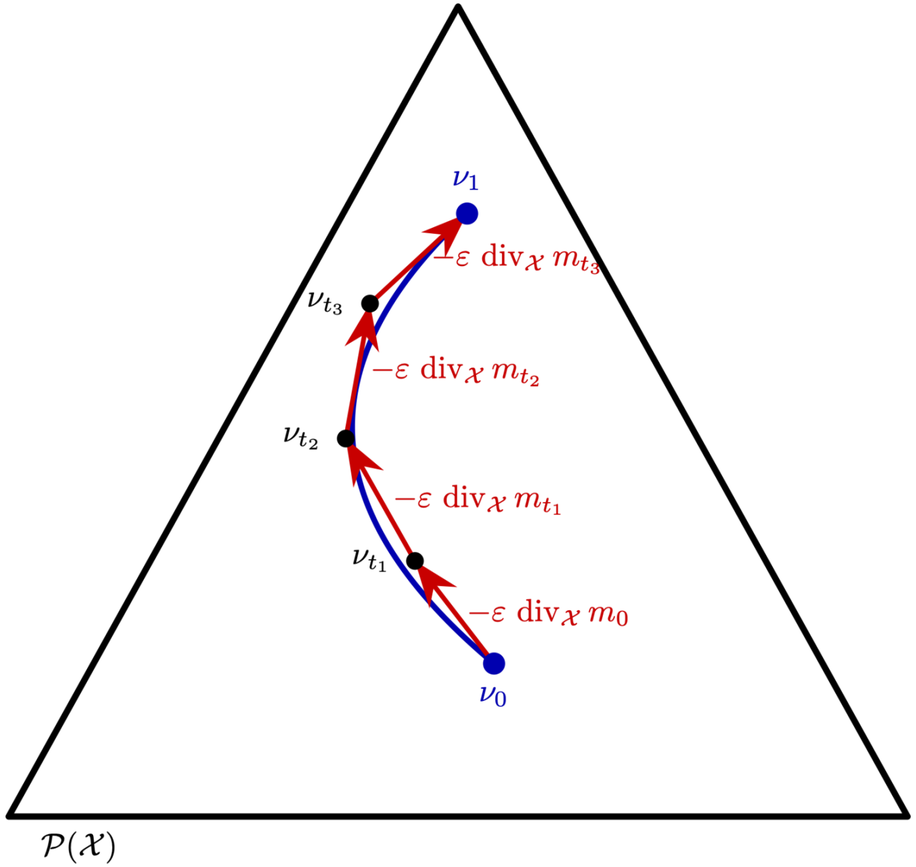
$$\approx \nu_k - t_k\ \mathrm{div}_{\mathcal{X}}\Big(\theta(\nu_k)\odot\sum_{i=1}^{p}\lambda_i m_i(0)\Big)$$



Approximating Geodesic Curves via Chambolle-Pock

Discretize space-time, then minimize the action as a convex program

- Erbar et al. (2020): approximate $\mathcal{W}$ by non-smooth convex optimization
- Discretize $[0, 1]$ into N steps, take a weak form of the CE, introduce slack variables to decouple the non-linear relations in the action, then minimize; the output is a discretized geodesic in the $\mathcal{W}$ geometry
- Works for *any* pair of measures, sparsely supported ones on the boundary included: the problem is convex and needs no interiority (remember this later)



Erbar, Matthias, et al. “Computation of optimal transport on discrete metric measure spaces.” *Numerische Mathematik* 144.1 (2020): 157–200.

Analysis Plan: Inverting The Exponential Map

Take it to the Tangent Space

- To analyze a measure ν_0 against references $\mathcal{R} = \{\nu_i\}_{i=1}^p$, we use a theorem of Werenski et al.
- **Theorem** (Werenski, Murphy, Aeron et al. 2022, for W_2 on $\mathbb{R}^d$). Let T_i be the optimal transport map from ν_0 to ν_i and

$$A_{ij} = \left\langle T_i - \mathrm{Id},\ T_j - \mathrm{Id} \right\rangle_{L^2(\nu_0)}.$$

  Then ν_0 is the barycenter of $\mathcal{R}$ with weights λ if and only if $\min_{\lambda \in \Delta^{p-1}} \lambda^\top A \lambda = 0$, and the minimizer is λ
- Why: the first-order condition for $\mathcal{J}$ is $\sum_i \lambda_i (T_i - \mathrm{Id}) = 0$, and $T_i - \mathrm{Id} = \exp_{\nu_0}^{-1}(\nu_i)$ is a tangent vector, so $\lambda^\top A \lambda = \| \sum_i \lambda_i \exp_{\nu_0}^{-1}(\nu_i) \|^2$
- On a graph: replace $T_i - \mathrm{Id}$ by the log map and $L^2(\nu_0)$ by the metric tensor at ν_0. Analysis is then p log maps, one $p \times p$ Gram matrix, and a small quadratic program over the simplex

Werenski, Matthew, et al. “Measure estimation in the barycentric coding model.” *International Conference on Machine Learning*. PMLR, 2022.

Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Basel: Birkhäuser Basel, 2005.

Hamiltonian Shooting

Exponential and logarithmic maps without the space-time program

Where the time goes

The BCM only ever uses the start of a geodesic

- Chambolle-Pock is a first-order method: reaching tolerances like $\delta_g = 10^{-10}$ takes 10^4–10^5 iterations, each with a global space-time elliptic solve, over $O(N \cdot |E|)$ variables
- But look at what synthesis and analysis actually consume:
  - Synthesis: $\exp_{\nu_k}$, and the *initial* momenta $m_i(0)$
  - Analysis: the *initial* tangent vectors $\exp_{\nu_0}^{-1}(\nu_i)$, to build the Gram matrix
- A geodesic is determined by its initial point and velocity, and satisfies an ODE of dimension $2n$, independent of any time grid N
- **Idea:** solve the geodesic ODE directly. $\exp$ becomes an initial value problem; $\exp^{-1}$ becomes a boundary value problem, solved by *shooting*

Geodesics as a Hamiltonian System

Position: the measure ρ. Momentum: the potential φ.

- Stationarity of the action in m puts optimal momenta in gradient form, $m_t = \theta(\rho_t) \odot \nabla\varphi_t$: velocities live in the tangent space from before
- Stationarity in ρ gives the second equation. Together:

$$\dot\rho(x) = \sum_y \theta\big(\rho(x),\rho(y)\big)\,\nabla\varphi(x,y)\,Q(x,y), \qquad \dot\varphi(x) = -\frac{1}{2}\sum_y \partial_1\theta\big(\rho(x),\rho(y)\big)\,\nabla\varphi(x,y)^2\,Q(x,y)$$

- These are Hamilton's equations (in the π-weighted pairing) for the kinetic energy

$$H(\rho,\varphi) = \frac{1}{4}\sum_{x,y} \theta\big(\rho(x),\rho(y)\big)\,\nabla\varphi(x,y)^2\,Q(x,y)\pi(x)$$

- First equation: the continuity equation. Second: a discrete Hamilton–Jacobi equation. On a graph this is the Wasserstein Hamiltonian flow of Chow–Li–Zhou, in the form written down by Cui–Liu–Zhou

Chow, Li, Zhou. "Wasserstein Hamiltonian flows." *J. Differential Equations* 268.3 (2020): 1205–1219. Cui, Liu, Zhou. "What is a stochastic Hamiltonian process on finite graph? An optimal transport answer." *J. Differential Equations* 305 (2021).

What the Hamiltonian buys us

Conservation laws, and the distance from initial data alone

- H is conserved along geodesics: constant speed, and a free accuracy check on any integrator
- Total mass is conserved: it is the conserved quantity of the gauge symmetry $\varphi \mapsto \varphi + c$
- The metric speed is $\|\dot{\rho}_t\|_{\mathcal{W}}^2 = 2H$, so

$$\mathcal{W}^2(\rho_A, \rho_B) = \int_0^1 2H \, dt = 2H(\rho_A, \varphi_0)$$

  The squared distance is read off the *initial* data

- The analysis Gram matrix is assembled exactly from the recovered potentials:

$$A_{ij} = \left\langle \nabla\varphi_i,\ \theta(\nu) \odot \nabla\varphi_j \right\rangle_Q$$

The Exponential Map

Shoot forward

- Given ν and a potential φ_0 (or a momentum m_0, converted by one weighted-Laplacian solve), integrate from $\rho(0) = \nu$, $\varphi(0) = \varphi_0$:

$$\exp_\nu(\nabla\varphi_0) = \rho(1)$$

- In our implementation: fixed-step RK4, 150 steps on $[0, 1]$; H and total mass drift by $\sim 10^{-12}$ relative over the unit horizon
- **Interior only.** The right-hand side of the ODE involves $\partial_1\theta(\rho(x), \rho(y))$, which blows up as $\rho(x) \to 0$: shooting needs $\rho > 0$ everywhere and does not apply to sparsely supported measures, where Chambolle-Pock still does. A step that would push some $\rho(x)$ below a positivity floor is bisected (up to 4 times) before giving up
- Replaces the first-order heuristic $\nu + t\,\mathrm{div}_{\mathcal{X}}\, m(0)$ in the synthesis update with an exponential map

The Logarithmic Map

Shooting: aim, fire, correct

- To compute $\exp_\nu^{-1}(\mu)$, solve for the initial potential:

$$F(\varphi_0) := \exp_\nu(\nabla\varphi_0) - \mu = 0$$

- Gauge $\langle \varphi_0, 1\rangle_\pi = 0$, and F conserves mass: $n-1$ equations in $n-1$ unknowns
- Damped Newton, with the Jacobian by forward-mode automatic differentiation through the integrator and a backtracking line search on $\|F\|_\pi$
- Initial guess: linearize the flow at $t = 0$. One sparse SPD solve:

$$L_\theta(\nu)\,\varphi_0 = \pi \odot (\mu - \nu), \qquad L_\theta(\nu) = \nabla^\top \operatorname{Diag}\big(\kappa \odot \theta(\nu)\big)\nabla$$

  with $\kappa(x,y) = Q(x,y)\pi(x)$ the symmetric edge weight: the graph Laplacian with edge weights $\kappa \odot \theta(\nu)$

- Warm starts: in an outer loop, the previous φ_0 is an excellent initial guess

---

Shooting for the geodesic BVP is classical in shape analysis: Klassen, Srivastava, Mio, Joshi, *IEEE PAMI* (2004). For Wasserstein–Hamiltonian systems: Cui, Dieci, Zhou, *SIAM J. Sci. Comput.* (2022).

Closing the Loop: the BCM by Shooting

The synthesis update from before, without the approximation

- **Synthesis.** With $\varphi_0^{(i)}$ the potential of $\exp_{\nu_k}^{-1}(\nu_i)$:

$$\nu_{k+1} = \exp_{\nu_k}\left(h\,\nabla \sum_{i=1}^{p} \lambda_i \varphi_0^{(i)}\right)$$

  Each iteration: p log maps, warm-started from the previous iterate, and one exp map. Halve h if the step hits the positivity floor or fails to decrease $\mathcal{J}$; stop when the descent direction's metric norm $\sqrt{2H}$ is small

- **Analysis.** Log-map the target to each reference, build $A_{ij} = \langle \nabla\varphi_i, \theta(\nu) \odot \nabla\varphi_j \rangle_Q$, solve the same quadratic program over the simplex

- Two different time grids. Shooting still integrates the geodesic ODE on a grid of K RK4 steps, but that grid discretizes an *ODE*: error $O(K^{-4})$, refined at will without touching Newton. Chambolle-Pock's N steps discretize the *action* being minimized: error $O(1/N)$, and every variable scales with N. So synthesis and analysis should use the same method

How fast?

Random interior endpoint pairs on grid graphs, 20 trials each

	25 nodes, 40 edges	100 nodes, 180 edges
Shooting $\exp^{-1}$: Newton iterations, cold / warm	3–4 / 2–3	3–4 / 3
Shooting $\exp^{-1}$: time	**92 ms**	**1.6 s**
Chambolle-Pock ($N = 32$, $\delta_g = 10^{-8}$): time	3.1 s (34×)	636.1 s (391×)
Round trip $\lVert \exp_\nu(\exp_\nu^{-1} \mu) - \mu \rVert_\pi$, worst	3×10^{-11}	9×10^{-10}
$\mathcal{W}^2$ vs. C-P, worst relative gap	3×10^{-2}	3×10^{-2}

Median wall-clock per geodesic, one CPU thread, after JIT warm-up. Chambolle-Pock is slow, so it was run on only 5 of the 25-node pairs and 1 of the 100-node pairs. The $\mathcal{W}^2$ gap is Chambolle-Pock's $O(1/N)$ discretization of the action; shooting's only discretization error is the integrator's, $O(K^{-4})$.

Shooting is one to two orders of magnitude faster than Chambolle-Pock, and the speedup grows with problem size: 34× at 25 nodes, 391× at 100 nodes.

Two Discretizations of the Same Geodesic

What each method discretizes, and what that costs

	Chambolle-Pock (the paper)	Hamiltonian shooting (new)
What is discretized	the action, on a space-time grid	the geodesic ODE, in time only
Unknowns	$O(N\,\lvert E\rvert)$: a path (ρ, m) on N time steps	$n-1$: the initial potential φ_0
Error	$O(1/N)$ in the time steps (Γ-convergence)	$O(K^{-4})$ in the RK4 steps, refined at will
Iteration	first order: 10^4–10^5 primal-dual steps to $\delta_g = 10^{-10}$	Newton: 3–4 steps cold, 2–3 warm, to 10^{-12}
Output	the whole path; the first-interval momentum	φ_0, hence $\exp^{-1}$ and $\mathcal{W}^2 = 2H$
Admissible data	any pair of measures, boundary included	interior of the simplex only
Guarantee	global optimum of the discretized problem, with a duality gap	a stationary path; minimality certified against the convex program

Shooting is an Interior Method

What happens near the boundary of the simplex

- $\partial_1\theta(s,t) \to \infty$ as $s \to 0$: like $\frac{1}{2}\sqrt{t/s}$ for the geometric mean, like $t/(s\log^2(s/t))$ for the logarithmic mean. Either way H is not C^1 up to the boundary
- This is geometry: a solution either exists for all time or, at some finite T, some $\rho(x) \to 0$ or $\varphi \to \infty$ (Cui–Liu–Zhou 2021, Remark 3.3); geodesics can touch the boundary even between interior endpoints, and uniqueness can fail there (Gangbo–Li–Mou 2019)
- In practice: mollify toward uniform, $\nu_\varepsilon = (1-\varepsilon)\nu + \varepsilon 1$, shoot for decreasing ε (warm-started), extrapolate; for boundary-supported data, Chambolle-Pock remains the reference

Gangbo, Li, Mou. “Geodesics of minimal length in the set of probability measures on graphs.” *ESAIM: COCV* 25 (2019): 78. Cui, Liu, Zhou. “What is a stochastic Hamiltonian process on finite graph? An optimal transport answer.” *J. Differential Equations* 305 (2021).

Validating the Code with Numerical Experiments

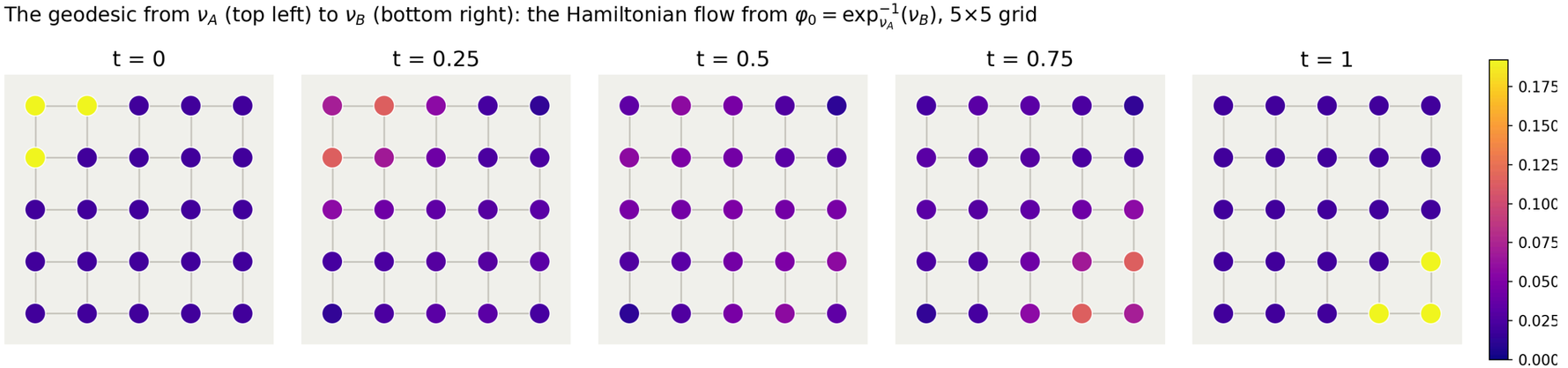
Validation

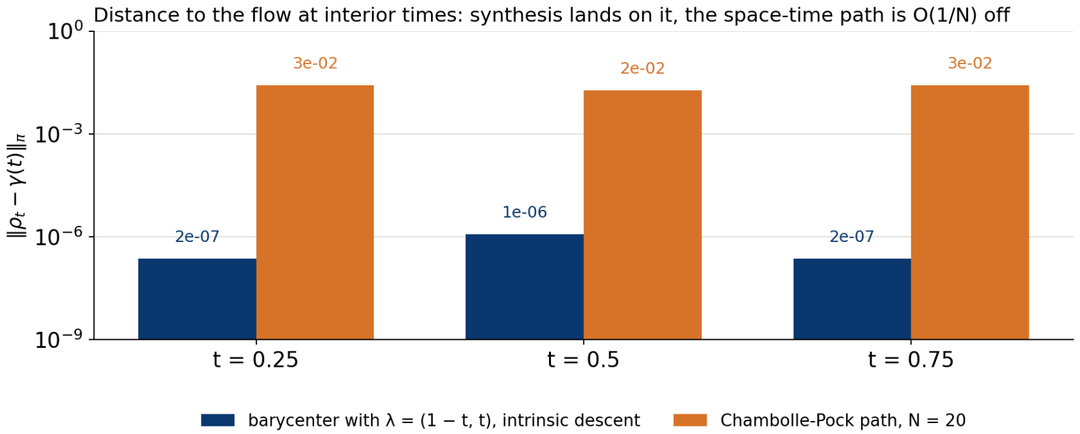
How do we know if the thing actually works?

- The code is validated by a suite of numerical experiments, looking at the following list of sanity checks:
  1. Invariants of the Hamiltonian flow along every integrated geodesic: H and total mass conserved (relative drift $\sim 10^{-12}$), round trip $\exp_\nu(\exp_\nu^{-1}\mu) = \mu$ to 10^{-10}, time-reversal symmetry, and $\sqrt{2H}$ against the closed-form distance on the two-point graph
  2. Agreement between independent solvers: do intrinsic descent with Chambolle-Pock geodesics and with Hamiltonian shooting land on the same barycenter?
  3. Round-trip consistency: synthesize with λ, analyze to recover $\hat{\lambda}$, check the relative error
  4. Comparison with the popular alternative: entropically regularized transport on a cost matrix built from the graph

A Geodesic Three Ways

Synthesis with $\lambda = (1-t, t)$, the geodesic equations, and Chambolle-Pock





Analysis of each synthesized measure recovers t

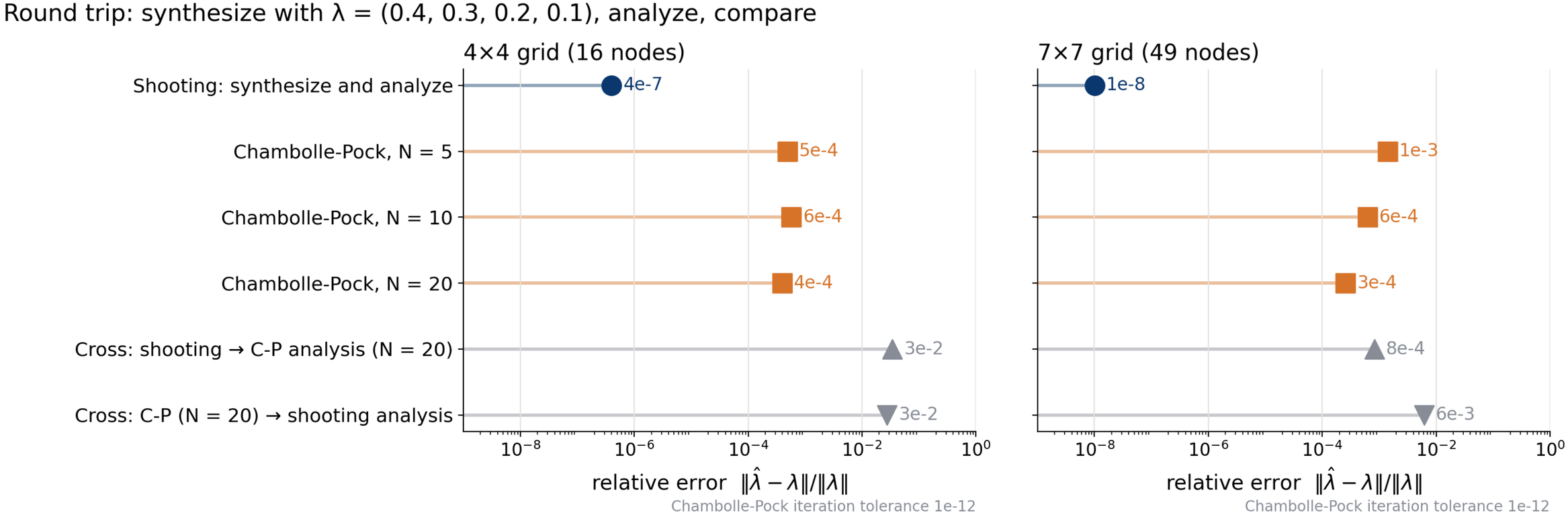
t (weight on ν_B)	t̂ recovered	|t̂ − t|
0	0.000000000000	0
0.25	0.249999999971	2.9e-11
0.5	0.500000000000	0
0.75	0.750000000029	2.9e-11
1	1.000000000000	0

W² of the path: shooting 30.2429; Chambolle-Pock (N = 20) 28.1597, below by its O(1/N) time discretization

Top: $\gamma(t)$, the Hamiltonian flow from $\varphi_0 = \exp_{\nu_A}^{-1}(\nu_B)$. Left: the barycenter with weights $(1-t, t)$ sits on γ to 10^{-6}; the Chambolle-Pock path at $N = 20$ is 10^{-2} off. Right: the coordinate recovered by analyzing the point at time t from each of the three sources

Coordinate Recovery

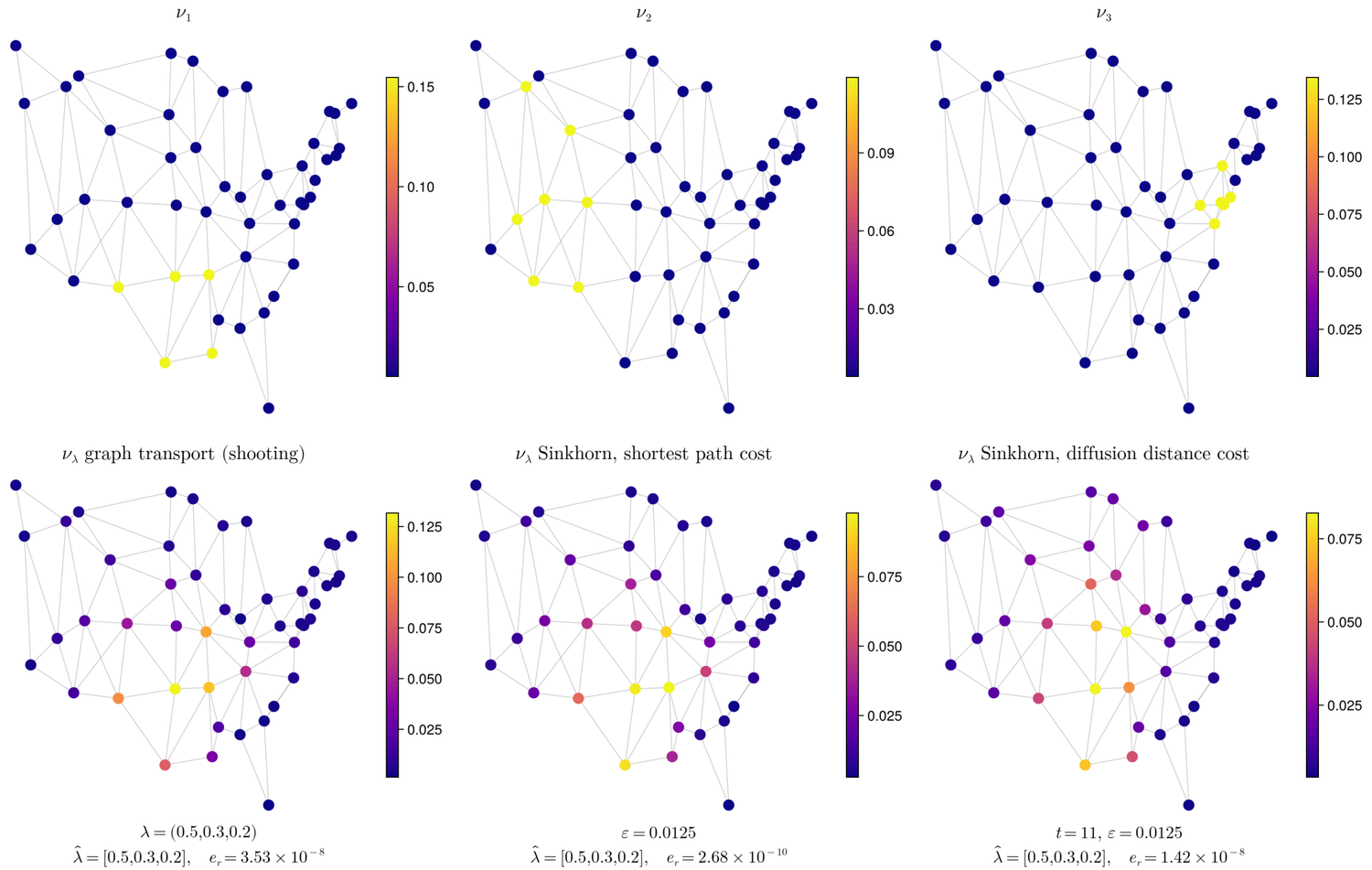
Shooting recovers λ to 10^{-7}; Chambolle-Pock to 10^{-3}, bounded by its iteration tolerance



Solution 1 vs. Solution 2

Entropic transport on a graph metric against graph transport

- Back to the fork: Sinkhorn on a graph metric (Solution 1) against graph transport by shooting (Solution 2), same three references
- Upside to entropic transport: very fast (about a second against a minute here); coordinate recovery is comparable, 10^{-8} to 10^{-10} for all three
- Downsides: distance matrices don't capture all connectivity information, entropy biases the output
- Takeaway: only the graph transport barycenter operates directly on the graph structure



Summing Up

Takeaways on what we did

- Barycenters need more than a metric: they need a geometry with geodesics, and the Kantorovich geometry on a graph has none (the two-point graph). The Benamou–Brenier picture survives the passage to a graph and gives the discrete transport metric $\mathcal{W}$ of Maas, a Riemannian structure in which barycenters can be computed
- Synthesis and analysis can both be handled by working within the Riemannian geometry of $(\mathcal{P}(\mathcal{X}), \mathcal{W})$: intrinsic gradient descent for synthesis, the tangent-space Gram matrix of Werenski et al. for analysis. Both consume only $\exp$ and $\exp^{-1}$
- Computing those maps was the obstacle: the space-time convex program of Erbar et al. gives them at $O(1/N)$ accuracy and $O(N|E|)$ cost
- New: Hamiltonian shooting computes $\exp$ and $\exp^{-1}$ directly from the geodesic ODE, accurate to fourth order in the integrator step (Runge–Kutta) rather than first order in a time grid, and one to two orders of magnitude faster on interior data

Looking Ahead

Future directions and ideas to explore

- **Convergence theory for the descent.** $\mathcal{J}$ is convex along linear, probably not along geodesic, interpolation; in $\mathbb{R}^d$ the same descent is saved by a Polyak–Łojasiewicz inequality. Does one hold on $(\mathcal{P}(\mathcal{X}), \mathcal{W})$?
- **Next: linearizing the BCM.** Near a base measure ν, replace $\exp_\nu^{-1}(\nu_i)$ by one weighted-Laplacian solve, $L_\theta(\nu)\,\varphi_i = \pi \odot (\nu_i - \nu)$; on graphs, the discrete transport distance is locally a Laplacian quadratic form (Robertson, Wan, Cloninger 2024); this is linear optimal transport (cf. Wilson, Needham, Srivastava 2024)
- **Toward geodesic PCA on graphs:** in the Wasserstein space over $\mathbb{R}$, geodesic PCA becomes convex PCA in a linearized, tangent-space picture (Bigot, Gouet, Klein, López 2017). The linearized BCM supplies that picture for measures on graphs

References

- My paper: G., Murphy. “Static and Dynamic Approaches to Computing Barycenters of Probability Measures on Graphs.” *arXiv preprint* *arXiv:2603.26940* (2026).
- My code: https://github.com/dcgentile/GraphTransportation.jl, with a Python implementation at https://github.com/dcgentile/graphtransport-py
- My site: https://pilesofsand.com
- Select references:
  - Benamou, Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.” *Numer. Math.* 84.3 (2000): 375–393.
  - Agueh, Carlier. “Barycenters in the Wasserstein space.” *SIAM J. Math. Anal.* 43.2 (2011): 904–924.
  - Cuturi. “Sinkhorn distances: lightspeed computation of optimal transport.” *NeurIPS* 26 (2013).
  - Bonneel, Peyré, Cuturi. “Wasserstein barycentric coordinates: histogram regression using optimal transport.” *ACM Trans. Graph.* 35.4 (2016): 71.
  - Werenski et al. “Measure estimation in the barycentric coding model.” *ICML*, PMLR, 2022.
  - Otto. “The geometry of dissipative evolution equations: the porous medium equation.” *Comm. PDE* 26 (2001): 101–174.
  - Absil, Mahony, Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton, 2008. Huang, Absil, Gallivan. “ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds.” *ACM TOMS* (2016).
  - Klassen, Srivastava, Mio, Joshi. “Analysis of planar shapes using geodesic paths on shape spaces.” *IEEE PAMI* 26.3 (2004): 372–383.
  - Chowdhury, Needham. “Gromov–Wasserstein averaging in a Riemannian framework.” *CVPR Workshops* 2020.
  - Díaz Martín, Medri, Murphy. “Gromov–Wasserstein barycenters: the analysis problem.” *arXiv:2507.09865* (2025).
  - Wilson, Needham, Srivastava. “Fused Gromov–Wasserstein variance decomposition with linear optimal transport.” *arXiv:2411.10204* (2024).
  - Bauer, Mémoli, Needham, Nishino. “Metric geometry of Lebesgue, Wasserstein, and Gromov–Wasserstein spaces: submetries, curvature, and geodesics.” *arXiv:2608.11680* (2026).
  - Maas. “Gradient flows of the entropy for finite Markov chains.” *J. Funct. Anal.* 261.8 (2011): 2250–2292.
  - Mielke. “A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems.” *Nonlinearity* 24.4 (2011): 1329–1346.
  - Chow, Huang, Li, Zhou. “Fokker–Planck equations for a free energy functional or Markov process on a graph.” *Arch. Ration. Mech. Anal.* 203 (2012): 969–1008.
  - Erbar, Maas. “Ricci curvature of finite Markov chains via convexity of the entropy.” *Arch. Ration. Mech. Anal.* 206 (2012): 997–1038.
  - Chewi, Maunu, Rigollet, Stromme. “Gradient descent algorithms for Bures–Wasserstein barycenters.” *COLT* 2020.
  - Erbar, Rumpf, Schmitzer, Simon. “Computation of optimal transport on discrete metric measure spaces.” *Numer. Math.* 144.1 (2020): 157–200.
  - Gangbo, Li, Mou. “Geodesics of minimal length in the set of probability measures on graphs.” *ESAIM: COCV* 25 (2019): 78. Erbar, Maas, Wirth. “On the geometry of geodesics in discrete optimal transport.” *Calc. Var. PDE* 58 (2019): 19.
  - Afsari, Tron, Vidal. “On the convergence of gradient descent for finding the Riemannian center of mass.” *SIAM J. Control Optim.* 51.3 (2013): 2230–2260.
  - Chow, Li, Zhou. “Wasserstein Hamiltonian flows.” *J. Differential Equations* 268.3 (2020): 1205–1219.
  - Cui, Liu, Zhou. “What is a stochastic Hamiltonian process on finite graph? An optimal transport answer.” *J. Differential Equations* 305 (2021): 428–457.
  - Cui, Liu, Zhou. “Wasserstein Hamiltonian flow with common noise on graph.” *SIAM J. Appl. Math.* 83.2 (2023): 484–509.
  - Liu, Wu, Ye, Zhou. “Parametrized Wasserstein Hamiltonian flow.” *SIAM J. Numer. Anal.* 63.1 (2025).
  - Zhou, Zuo, Liu, Li. “Accelerated Markov chain Monte Carlo algorithms on discrete states.” *arXiv:2505.12599* (2025).
  - Cui, Dieci, Zhou. “Time discretizations of Wasserstein–Hamiltonian flows.” *Math. Comp.* 91 (2022): 1019–1075.
  - Cui, Dieci, Zhou. “A continuation multiple shooting method for Wasserstein geodesic equation.” *SIAM J. Sci. Comput.* 44.5 (2022): A2918–A2943.
  - Chen, Cui, Dieci, Zhou. “Newton’s method for optimal transport problem on graphs.” *arXiv:2605.07408* (2026).
  - Bigot, Gouet, Klein, López. “Geodesic PCA in the Wasserstein space by convex PCA.” *Ann. Inst. H. Poincaré Probab. Statist.* 53.1 (2017): 1–26.
  - Robertson, Wan, Cloninger. “Resistance distance and linearized optimal transport on graphs.” *arXiv:2404.15261* (2024).

Thank you so much for your time and attention!

Please enjoy this picture of my cats.

