MULTI-ANCHOR LOT: FROM LOCAL TRANSPORT MAPS TO GLOBAL WASSERSTEIN GEOMETRY

Anonymous authors
Paper under double-blind review

ABSTRACT

Linearized optimal transport (LOT) provides Euclidean representations of distributions whose coordinate distances approximate the Wasserstein distance. However, linearization at a single reference cannot generally preserve the nonflat geometry of Wasserstein space. On smooth finite-dimensional submanifolds of Wasserstein space, we quantify this limitation through a local distortion bound governed by curvature and squared reference distance. This analysis motivates Multi-Anchor Linearized Optimal Transport (MALOT) for recovering global Wasserstein geometry from local transport maps. Starting from a multi-anchor Riemannian metric that gives greater weight to nearby references, we derive a symmetric two-point approximation with endpoint-dependent weights on squared LOT separations. Radial completion corrects each discrete coordinate block to preserve its reference radius. We prove that the resulting dissimilarities converge uniformly to the Wasserstein distance as the anchor covering radius and weight-localization error vanish, and establish coverage guarantees for farthest-point sampling. Simulations demonstrate improved recovery over LOT and decreasing error with finer anchor coverage; real-data studies evaluate distance recovery and downstream prediction.

1 INTRODUCTION

Learning from distribution-valued observations requires representations suitable for prediction (Szabó et al., 2016). Linearized optimal transport (LOT) represents probability distributions through optimal transport displacements from a shared reference (Wang et al., 2013; Moosmüller & Cloninger, 2023). After discretization, these displacements form Euclidean coordinates whose distances approximate the Wasserstein distance. These coordinates support visualization and learning on images and point clouds while retaining information about the arrangement of probability mass (Cloninger et al., 2025; Mahan et al., 2024). Reusable reference-to-target transports replace pairwise OT solves (Peyré & Cuturi, 2019).

The geometric fidelity of LOT, however, depends on where the reference is placed. In its ideal map formulation, a LOT coordinate is the initial velocity of the Wasserstein geodesic from the reference to a target. Its norm gives the exact reference-to-target distance, but subtracting two such velocities need not recover the distance between their endpoints. On nonflat smooth submanifolds of Wasserstein space, this distinction reflects the difference between geodesic separation and separation in one tangent space (Lott, 2008; Bhatia et al., 2019). We make this locality precise by bounding the absolute relative error in squared distance by $C\kappa R^2$, where κ bounds absolute sectional curvature and R bounds distance from the reference along the target-to-target geodesic within a compatible normal neighborhood. The bound identifies the reference radius as a controllable source of approximation error.

We address this limitation by using multiple anchors to localize the tangent approximation. Each anchor supplies the initial velocities of its own transport geodesics to the targets, providing a different local description of the same distribution family. We combine the quadratic forms that the anchor-specific LOT maps induce on a common target tangent space. Proximity weights favor nearby anchors, for which the local distortion bound is tighter. The resulting spatially varying Riemannian metric motivates *multi-anchor linearized optimal transport* (MALOT). Finite LOT increments and symmetrized endpoint weights give a weighted mean of squared anchor-specific separations, favor-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ing anchors near either endpoint. Taking its square root yields the MALOT dissimilarity without differentiating transport maps or optimizing paths. For discrete plans, a variance-based correction restores the exact reference radius in each LOT coordinate block, enabling the recovery guarantee.

MALOT addresses reference locality within LOT, complementing faster OT solvers (Cuturi, 2013; Scetbon et al., 2021; Gasteiger et al., 2021) and sliced or learned representations (Kolouri et al., 2019; Amir & Dym, 2025; Haviv et al., 2024; Tran et al., 2025). Fixed multi-reference concatenation (Khurana et al., 2023) retains a global Hilbert representation and remains subject to the fixed-embedding distortion obstruction in Section 4. In our rotating-shape simulation and cell-image study, concatenation has higher distance stress than LOT despite using multiple references. MALOT instead localizes each comparison through endpoint-dependent weights. Euclidean coordinates are obtained subsequently through multidimensional scaling (MDS).

Contributions.

1. **Geometric analysis and construction.** We quantify LOT's reference-dependent local error and construct a multi-anchor Riemannian metric whose intrinsic distance converges to Wasserstein distance under explicit geometric conditions. MALOT is a locally consistent two-point approximation of this metric.
2. **Discrete recovery and adaptive coverage.** We prove uniform recovery directly for discrete MALOT as anchor covering radius and weight-localization error vanish, without manifold assumptions. Farthest-point sampling provides label-free coverage guarantees; further results control triangle-inequality defects and MDS stability.
3. **Empirical evaluation.** Crescent and rotating-shape simulations test anchor coverage and distance recovery against LOT and concatenation. Real-data studies compare distance recovery and downstream prediction on cell images, object contours and viscous-fingering images, with additional datasets and component controls in the appendix.

Section 2 develops the local analysis, followed by the construction and guarantees in Sections 3 and 4. Section 5 reports experiments, and Section 6 concludes.

2 Optimal Transport and Its Linearization

2.1 Wasserstein Geometry

Let $\mathcal{P}_2(\mathbb{R}^d)$ denote probability measures on $\mathbb{R}^d$ with finite second moment. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the squared 2-Wasserstein distance is

$$W_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \,\mathrm{d}\pi(\boldsymbol{x},\boldsymbol{y}), \tag{1}$$

where $\Pi(\mu,\nu)$ contains the couplings with marginals μ and ν. For an absolutely continuous curve of smooth densities, a tangent perturbation is represented by its minimum-norm gradient velocity $v = \nabla\psi$, satisfying $\partial_t\mu_t + \nabla\cdot(\mu_t v_t) = 0$. The Wasserstein tangent inner product is

$$g_{W,\mu}(v,w) = \int_{\mathbb{R}^d} \langle v(\boldsymbol{x}), w(\boldsymbol{x})\rangle \,\mathrm{d}\mu(\boldsymbol{x}). \tag{2}$$

Its norm measures instantaneous transport speed (Lott, 2008; Sarrazin & Schmitzer, 2024). We use this description on a connected finite-dimensional smooth submanifold $\mathcal{M} \subset \mathcal{P}_2(\mathbb{R}^d)$ whose Riemannian speed agrees with ambient Wasserstein speed. This restriction makes the differential arguments precise without treating all of $\mathcal{P}_2(\mathbb{R}^d)$ as a smooth manifold.

2.2 LOT and Its Local Approximation Error

For an absolutely continuous reference ρ, Brenier's theorem supplies an optimal map $T_{\rho\to\mu}$, and the LOT coordinate is

$$\phi_\rho(\mu) = \mathrm{Log}_\rho(\mu) = T_{\rho\to\mu} - \mathrm{id} \in L^2(\rho;\mathbb{R}^d). \tag{3}$$

LOT compares two targets through

$$d_\rho^{\mathrm{LOT}}(\mu,\nu) = \|\phi_\rho(\mu) - \phi_\rho(\nu)\|_{L^2(\rho)}, \qquad \|v\|_{L^2(\rho)}^2 = \int \|v(\boldsymbol{x})\|_2^2 \,\mathrm{d}\rho(\boldsymbol{x}). \tag{4}$$

The norm averages squared displacement differences over the mass of the reference. It gives the radial identity $\|\phi_\rho(\mu)\|_{L^2(\rho)} = W_2(\rho,\mu)$. The distinction is between exact radial transport and approximate transverse comparison. To quantify it, define the reference-induced quadratic form

$$g_{\rho,\mu}(v,w) = \langle D\phi_\rho(\mu)[v], D\phi_\rho(\mu)[w]\rangle_{L^2(\rho)}.$$

It measures an infinitesimal motion at μ after expressing that motion in coordinates based at ρ.

Proposition 1 (Local reference distortion)**.** *Suppose ϕ_ρ agrees, under an isometric tangent identification, with the Riemannian logarithm on a normal neighborhood containing the closed normal ball of radius r_0 about ρ in $\mathcal{M}$. If the absolute sectional curvature along its radial geodesics is at most κ and $\kappa r_0^2 \le 1/4$, then, for μ in this neighborhood with $r = W_2(\rho,\mu) \le r_0$ and $v \in T_\mu\mathcal{M}$,*

$$\left|g_{\rho,\mu}(v,v) - g_{W,\mu}(v,v)\right| \le C_0\kappa r^2 g_{W,\mu}(v,v). \tag{5}$$

Here C_0 is a universal constant. If a Wasserstein minimizing geodesic from μ to ν stays in this normal neighborhood within radius $R \le r_0$ of ρ, then

$$\left|[d_\rho^{\mathrm{LOT}}(\mu,\nu)]^2 - W_2^2(\mu,\nu)\right| \le C_0\kappa R^2 W_2^2(\mu,\nu). \tag{6}$$

The proposition applies the classical Jacobi-field estimate to LOT coordinates (Lee, 2018). The absolute relative error bound scales quadratically with the reference radius and vanishes on flat compatible families. Bures–Wasserstein space, comprising nondegenerate Gaussian measures equipped with W_2, admits this logarithmic interpretation (Bhatia et al., 2019; Vesseron et al., 2026). The dependence on reference radius motivates replacing one expansion point by a spatially localized family.

For discrete measures, optimal plans need not be maps, and the preceding differential argument does not apply directly. Section 3 constructs a finite-plan counterpart, whose recovery is proved separately. Throughout, exact OT or EMD denotes W_2 under a quadratic ground cost, not the unrooted transport objective.

3 Multi-Anchor Linearized Optimal Transport

The reference-radius dependence of LOT suggests combining local linearizations according to where they are accurate. We first construct a spatially weighted metric and establish its coverage guarantee, then obtain MALOT by replacing infinitesimal displacements with finite reference-coordinate differences.

3.1 Local Weighting and Geometric Recovery

Let $\mathcal{R} = \{\rho_k\}_{k=1}^K$ be reference distributions, called anchors, and set $c_k(\mu) = W_2^2(\rho_k,\mu)$. On the smooth submanifold $\mathcal{M}$ of Section 2, let $\phi_k\colon \mathcal{M} \to \mathcal{H}_k = L^2(\rho_k;\mathbb{R}^d)$ be smooth exact LOT maps. Each map expresses the same tangent motion at a target in a different reference coordinate system. We give greater weight to nearby anchors, whose local distortion bound is smaller, by defining

$$\begin{aligned} \alpha_k(\mu) &= \frac{\exp\{-c_k(\mu)/\tau\}}{\sum_{s=1}^K \exp\{-c_s(\mu)/\tau\}}, \\ g_{\mathcal{R},\mu}(v,w) &= \sum_{k=1}^K \alpha_k(\mu)\langle D\phi_k(\mu)[v], D\phi_k(\mu)[w]\rangle_{\mathcal{H}_k}, \qquad \tau > 0. \end{aligned} \tag{7}$$

The summands are quadratic measurements on the common tangent space $T_\mu\mathcal{M}$, so their combination does not require aligning different anchor spaces. The positive weights sum to one and vary with the target. The temperature τ, measured in squared-distance units, determines how strongly a target favors its closest anchors. Unlike fixed concatenation, this construction changes the contribution of each reference across the distribution family.

The resulting tensor defines lengths by integrating local motion. On a connected open region $U \subset \mathcal{M}$, write

$$d_{\mathcal{R},U}(\mu,\nu) = \inf_{\gamma \subset U} \int_0^1 \sqrt{g_{\mathcal{R},\gamma(t)}(\dot\gamma(t),\dot\gamma(t))}\,\mathrm{d}t, \tag{8}$$

where γ ranges over piecewise smooth paths joining the endpoints. Let $d_{W,U}$ be the corresponding intrinsic distance for g_W and $h = \sup_{\mu\in U} \min_k W_2(\rho_k,\mu)$ the anchor covering radius. To apply the local distortion bound uniformly, assume each compatible normal neighborhood N_k in Proposition 1 contains $U \cap \{\mu : W_2(\rho_k,\mu) \le r_0\}$, with common r_0, κ and $\kappa r_0^2 \le 1/4$. For targets farther than r_0 from an anchor, assume $g_{\rho_k} \preceq (1+B)g_W$ with finite $B \ge 0$. These geometric constants are independent of K.

Theorem 1 (Multi-anchor geometric recovery)**.** *Under the preceding conditions, put* $A = \max\{C_0\kappa, (1+B)/r_0^2\}$ *and* $\delta = A(h^2 + \tau\log K)$*. If* $\delta < 1$*, the tensor in Equation* (7) *is a Riemannian metric, and for tangent vectors* v *and distinct endpoints,*

$$\begin{aligned} \left|g_{\mathcal{R}}(v,v) - g_W(v,v)\right| &\le \delta g_W(v,v), \\ \left|\frac{d_{\mathcal{R},U}}{d_{W,U}} - 1\right| &\le \frac{\delta}{1+\sqrt{1-\delta}}. \end{aligned} \tag{9}$$

When the evaluated pairs admit Wasserstein minimizing geodesics in U*,* $d_{W,U} = W_2$*. Thus* $h_K \to 0$ *and* $\tau_K \log K \to 0$ *give uniform relative distance recovery with fixed geometric constants.*

The theorem replaces the radius to one reference by the covering radius of the entire anchor bank. Its mechanism is the localization inequality $\sum_k \alpha_k(\mu)c_k(\mu) \le \min_k c_k(\mu) + \tau\log K$. The first term measures available coverage, while the second measures how much weight remains on more distant references. The condition $\delta < 1$ ensures positive definiteness and eventually holds in the recovery regime. Taking $\tau \log K = O(h^2)$ therefore gives $O(h^2)$ relative error for the intrinsic distance. The regularity conditions hold, for example, on suitable regions of Bures–Wasserstein space with covariance eigenvalues bounded above and away from zero. Appendix B gives the proof and specifies the mean and covariance bounds.

3.2 From the Local Metric to MALOT

Computing the intrinsic distance requires intermediate distributions, transport-map derivatives and path optimization. MALOT instead reuses the reference transports at the two targets. We replace $D\phi_k(\mu)[v]$ by the finite increment $\phi_k(\nu) - \phi_k(\mu)$ and average the two endpoint weights,

$$w_k(\mu,\nu) = \frac{\alpha_k(\mu) + \alpha_k(\nu)}{2}, \qquad D^2_{\mathcal{R}}(\mu,\nu) = \sum_{k=1}^K w_k(\mu,\nu)\|\phi_k(\mu) - \phi_k(\nu)\|^2_{\mathcal{H}_k}. \tag{10}$$

This symmetric two-point rule retains the local geometry of the constructed metric. For a smooth curve μ_t with initial velocity v, fixed K and $\tau > 0$, it satisfies $D^2_{\mathcal{R}}(\mu,\mu_t) = t^2 g_{\mathcal{R},\mu}(v,v) + o(t^2)$. Proposition 2 in Appendix B provides the formal statement.

Endpoint averaging also explains why local reference information can recover distances between widely separated targets. An anchor equal to either endpoint gives an exact LOT comparison with the other endpoint, regardless of their separation. An anchor close to either endpoint inherits a controlled error, as Section 4 establishes. Each target can therefore contribute its own nearby references, without requiring an anchor close to both targets or to their midpoint. The weighted comparison uses local accuracy where available while retaining LOT's directional displacement information.

For finite-support measures μ_i, let $\boldsymbol{\phi}_{ki} \in \mathbb{R}^{m_k d}$ be the barycentric LOT block from an m_k-atom anchor and let c_{ki} be the raw quadratic cost of its optimal plan. Write $\boldsymbol{\psi}_{ki}$ for the block defined in Remark 1 and $L_{k,ij} = \|\boldsymbol{\psi}_{ki} - \boldsymbol{\psi}_{kj}\|_2$. The implemented MALOT comparison is

$$\alpha_{ki} = \frac{\exp(-c_{ki}/\tau)}{\sum_s \exp(-c_{si}/\tau)}, \qquad w_{k,ij} = \frac{\alpha_{ki} + \alpha_{kj}}{2}, \qquad D_{ij} = \left(\sum_k w_{k,ij} L^2_{k,ij}\right)^{1/2}. \tag{11}$$

Thus the algorithm combines squared LOT separations, not target coordinates across unrelated reference spaces. All weights and blocks are obtained from the same reference transports, with no target-to-target OT solve.

Remark 1 (Preserving the reference distance for discrete plans)**.** *Barycentric projection can lose conditional variance when a discrete OT plan splits mass. Appending one scalar gives* $\boldsymbol{\psi}_{ki} = (\boldsymbol{\phi}_{ki}, \sqrt{c_{ki} - \|\boldsymbol{\phi}_{ki}\|^2}) \in \mathbb{R}^{m_k d+1}$ *and restores* $\|\boldsymbol{\psi}_{ki}\| = W_2(\rho_k, \mu_i)$*. The scalar is zero for deterministic plans and requires no additional transport solve. Appendix C.1 gives the variance identity, formulas and numerical conventions. Results omitting this scalar are explicitly labeled* MALOT without radial completion*.*

MALOT is symmetric, nonnegative and zero on the diagonal. Its pair-dependent weights need not satisfy the triangle inequality, so we use the term *dissimilarity*. The recovery theorem below controls its departure from a metric as well as its error relative to W_2.

3.3 Adaptive Anchors and Computation

Farthest-point sampling (FPS) builds the anchor bank where coverage is most needed. Starting from a stated training reference, it repeatedly selects the training measure farthest from its nearest selected anchor. Each new batch of reference transports both supplies the next LOT blocks and updates these nearest-anchor distances. FPS uses no labels and, with exact W_2, has covering radius at most twice the optimal discrete K-center radius (Gonzalez, 1985). We set

$$\tau = \eta h^2_{\text{train}} / \log(K+1), \qquad h^2_{\text{train}} = \max_{i \in \text{train}} \min_k c_{ki}. \tag{12}$$

The multiplier $\eta > 0$ controls localization relative to the attained coverage. It is fixed in simulations or selected on validation data and held fixed with the anchors. At zero coverage we use the zero-temperature limit, assigning equal weight to all nearest anchors. Algorithm 1 summarizes the complete procedure.

Algorithm 1 FPS-MALOT.

Require: Training measures $\{\mu_i\}_{i=1}^n$, initial training anchor ρ_1, budget $K \le n$, multiplier $\eta > 0$.
1: Initialize nearest costs $s_i = +\infty$ and the selected set with ρ_1.
2: **for** $k = 1, \dots, K$ **do**
3: Solve OT from ρ_k to each μ_i; store c_{ki} and $\boldsymbol{\psi}_{ki}$ as in Remark 1.
4: Update $s_i \leftarrow \min\{s_i, c_{ki}\}$ for every training measure.
5: **if** $k < K$ **then**
6: Select an unchosen measure maximizing s_i as ρ_{k+1} and add it to the selected set.
7: **end if**
8: **end for**
9: Set $h^2_{\text{train}} = \max_i s_i$ and τ by Equation (12).
10: Compute α_{ki} and every requested D_{ij} by Equation (11).
Ensure: Selected anchors, stored reference blocks and costs, and MALOT dissimilarities.

For n measures, maximum reference-block dimension q and per-pair transport cost T_{OT}, the method uses $O(nKT_{\text{OT}})$ transport work and $O(n^2Kq)$ arithmetic for a full matrix. Blocks require $O(nKq)$ storage, in addition to $O(n^2)$ for the matrix. The number of OT solves is reduced from quadratic to nK when $K \ll n$. Appendix C.3 compares LOT, MALOT, Sinkhorn and sliced Wasserstein under their respective computational assumptions.

When vectors are needed, classical MDS forms $\mathbf{B} = -\mathbf{H}\mathbf{D}^{\circ 2}\mathbf{H}/2$, with $\mathbf{H} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^{\mathsf{T}}/n$, and retains r positive eigenvalues to obtain $\mathbf{Y}_r = \mathbf{U}_r\mathbf{\Lambda}_r^{1/2}$. A new distribution requires only K reference transports and its MALOT distances to the training set before applying the fixed Gower projection (Gower, 1968). This separates reference-based distance estimation from the choice of Euclidean dimension. Appendix C.2 gives the extension formulas, cost and stability guarantee.

4 Wasserstein Recovery from Finite Anchors

Theorem 1 justifies the localized metric at the level of infinitesimal motion. We now analyze the finite MALOT comparison directly. The central observation is that preserving each reference distance makes one nearby endpoint sufficient for accurate pairwise recovery. This gives a coverage

guarantee for arbitrary finite-support distributions, without imposing the smoothness and geometric conditions of Section 3.1.

4.1 A Finite-Anchor Recovery Guarantee

Write $W_{ij} = W_2(\mu_i, \mu_j)$, $r_{ki} = W_2(\rho_k, \mu_i)$ and $h_i = \min_k r_{ki}$. For an evaluated family $\mathcal{D}$, let $h = \sup_{\mu\in\mathcal{D}} \min_k W_2(\rho_k, \mu)$. This is coverage of the target distributions themselves.

Theorem 2 (Discrete MALOT recovery)**.** *For arbitrary finite-support probability measures with exact optimal reference plans, MALOT in Equation* (11) *satisfies the following bounds, also valid for exact Brenier-map LOT in Equation* (10)*,*

$$|D_{ij} - W_{ij}| \le 2\sqrt{(h_i^2 + h_j^2)/2 + \tau \log K}, \qquad \sup_{\mu_i,\mu_j\in\mathcal{D}} |D_{ij} - W_{ij}| \le 2\sqrt{h^2 + \tau \log K}. \quad (13)$$

The guarantee holds at every finite anchor budget. At zero temperature it applies with $\tau \log K = 0$ under the nearest-anchor convention. Vanishing covering radius and $\tau_K \log K \to 0$ therefore imply uniform Wasserstein recovery. With $\tau = \eta h^2/\log(K+1)$, the maximum error is at most $2\sqrt{1+\eta}\, h$. For one exact radial reference, the corresponding bound is $2\sup_{\mu\in\mathcal{D}} W_2(\rho, \mu)$. Multiple anchors replace this global reference radius by the distance to the nearest available reference. If every evaluated distribution is itself an anchor, zero-temperature MALOT recovers all pairwise distances exactly.

The proof explains the role of endpoint weighting. For an anchor ρ_k, both its feature separation $L_{k,ij}$ and the target distance W_{ij} lie in the interval $[|r_{ki} - r_{kj}|, r_{ki} + r_{kj}]$. Their difference is therefore at most $2\min\{r_{ki}, r_{kj}\}$. Averaging squared anchor errors with the endpoint weights and applying the localization inequality from Section 3.1 yields Theorem 2. Thus nearby anchors control the finite comparison through radial exactness, while their LOT coordinates retain directional information. Appendix A gives the proof and a sharper, pair-specific certificate computable from the stored reference costs.

FPS makes this bound constructive. If a candidate pool approximates $\mathcal{D}$ within Wasserstein distance e, and $r^\star_{K,W}$ is its optimal discrete K-center radius, exact FPS gives $h \le e + 2r^\star_{K,W}$. Substitution in Theorem 2 connects the anchor budget to recovery through the attained coverage. Increasingly dense candidate pools on a totally bounded family permit e and $r^\star_{K,W}$ to vanish. For a fixed training pool, this guarantee concerns its covered region; query coverage is evaluated separately. Corollary 1 in Appendix A.3 additionally accounts for compressed anchors, inexact feasible plans and observation error.

The continuous and discrete results control different objects using the same localization mechanism. With coverage-scaled temperature, Theorem 1 gives $O(h^2)$ relative error for the smooth intrinsic metric, whereas Theorem 2 gives $O(h)$ absolute error for the computed dissimilarity. Where both sets of conditions hold, their common Wasserstein limit also controls the discrepancy between MALOT and the intrinsic metric, as quantified in Appendix B.6.

4.2 Metric Behavior and Euclidean Representation

Uniform recovery also bounds departures from metric behavior. If $\max_{ij} |D_{ij} - W_{ij}| \le \epsilon$, then

$$D_{i\ell} \le D_{ij} + D_{j\ell} + 3\epsilon, \qquad W_{ij} > \epsilon \implies D_{ij} > 0. \quad (14)$$

Hence the coverage bound controls both triangle defects and the possible collapse of separated distributions. These conclusions follow directly from the Wasserstein triangle inequality, with details in Appendix A.

Pairwise recovery differs structurally from finding a fixed Euclidean representation. Appendix A.5 exhibits four distributions for which every Hilbert representation, including LOT and fixed concatenation, has a sharp maximum relative distortion floor of 13.29%. MALOT with those four distributions as anchors recovers their distances as $\tau \downarrow 0$. The example identifies a fixed-representation obstruction that pair-adaptive comparisons can overcome.

MDS then provides coordinates at a chosen dimension, with a separate approximation target. For a fixed collection and retained rank, a positive retained spectrum and a separating eigengap ensure

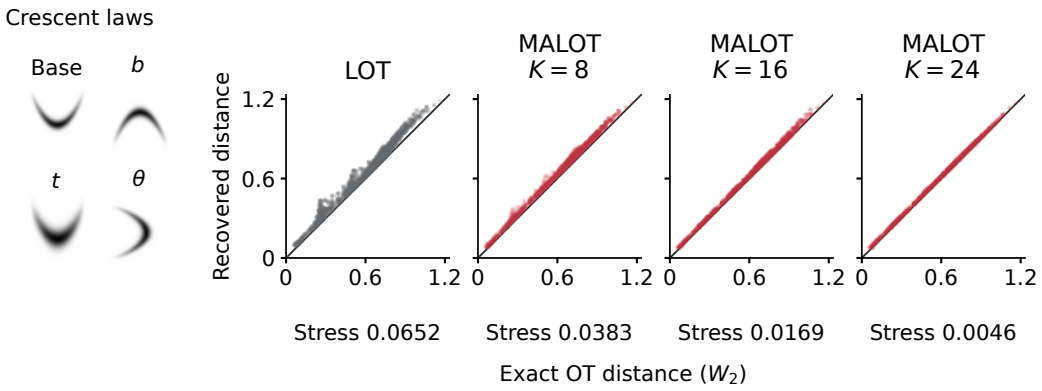

Figure 1: Crescent distributions and recovery with increasing anchor coverage. Left, a base density and changes in bend b, thickness t and rotation θ, using common coordinates and peak-normalized grayscale. Right, LOT and MALOT with $K = 8, 16, 24$ nested FPS anchors on the same 2,775 pairs. All recovery panels share axes and identity lines, without fitted rescaling. Stress decreases as the MALOT points approach the exact-OT diagonal.

that MALOT-based MDS converges, up to orthogonal alignment, to exact Wasserstein MDS as the distance error vanishes. The same conclusion holds for a fixed query when its reference-based distances converge. Proposition 3 in Appendix C.2 states the Gram perturbation and coordinate guarantees. Any Euclidean distortion already present in exact Wasserstein MDS remains at this limit.

5 Experiments

We evaluate the finite MALOT approximation. We examine whether finer anchor coverage improves recovery, whether adaptive localization contributes beyond fixed multi-reference averaging, and whether improved fidelity accompanies a downstream benefit. Exact W_2 for each experiment's stated distribution representation is the target, with recovery measured before MDS. For a set of unordered pairs, raw stress is $\{\sum_{i<j}(D_{ij} - W_{ij})^2 / \sum_{i<j} W_{ij}^2\}^{1/2}$. No scalar is fitted to the reported distances. MALOT uses endpoint weights and radial completion, with its removal explicitly labeled as an ablation. Appendices D and E provide the simulation and real-data protocols.

5.1 Crescent Laws Without Sampling Noise

To isolate geometric variation from sampling variability, we specify 25 centered crescent densities with varying bend and thickness and rotate them by $0°$, $45°$ and $90°$ to obtain 75 laws. Opposite bends can share their mean and covariance while differing in their transport geometry. Figure 1 places four representative densities beside the distance comparisons. Deterministic cell integration produces weighted measures on a 41×81 grid; no random samples are drawn.

At $K = 8$, FPS-MALOT reduces stress from 0.0652 for medoid LOT to 0.0383 and increases Spearman correlation from 0.9913 to 0.9969. Increasing the nested FPS bank to $K = 16$ and 24 reduces stress to 0.0169 and 0.0046, with the recovery plots approaching the identity line. The corresponding maximum absolute errors decrease from 0.0938 to 0.0509 and 0.0141, while the coverage radius decreases from 0.4910 to 0.2772 and 0.2011. With $\eta = 1$ fixed, the experiment varies the same coverage quantity that governs Theorem 2. Removing radial completion lowers stress to 0.0325 at $K = 8$, whereas completion improves stress at $K = 16, 24$. Appendix D.3 isolates this budget-dependent effect using matched component ablations.

5.2 Rotating Multi-Cluster Distributions

The rotating-shape experiment tests whether additional reference coordinates suffice without adaptive localization. It contains 120 distributions in twelve local families, with 80 uniform atoms per

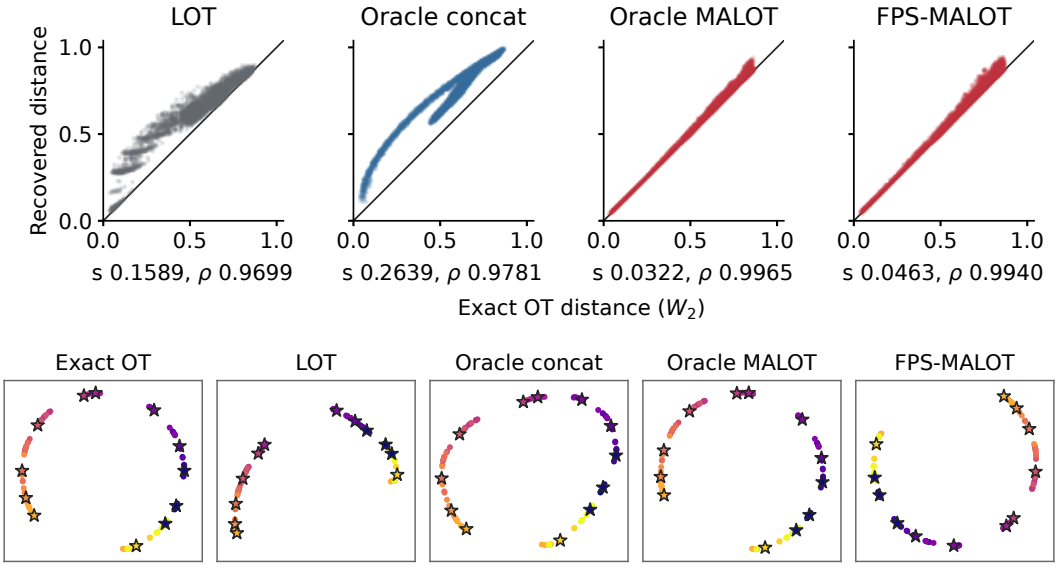

Figure 2: Rotating-shape simulation. Top, direct distance recovery, with raw stress s, Spearman correlation ρ and identity lines. Bottom, UMAP-style fuzzy-graph views from exact OT, LOT, oracle concat, oracle MALOT and FPS-MALOT, using seed 20 and the same embedding procedure. Colors index the twelve families, and stars identify FPS selections. Embedding axes are independently scaled and oriented.

measure and seed 7. Figure 3 shows four generating prototypes. LOT uses a free-support barycenter. Oracle concat and oracle MALOT share twelve generating prototypes, while FPS-MALOT chooses twelve observed distributions without labels. MALOT uses Equation (12) with $\eta = 1$.

Figure 2 combines distance recovery with five UMAP-style fuzzy-graph views. Across all 7,140 pairs, raw stress is 0.1589 for LOT and 0.2639 for normalized oracle concat, versus 0.0322 for oracle MALOT and 0.0463 for FPS-MALOT. Fixed oracle concatenation increases rather than reduces distortion in this setting. The matched oracle comparison isolates endpoint weighting, with radial completion inactive to numerical precision (Appendix D.3). These views illustrate the cyclic geometry and LOT's deformation. They use seed 20 and the same embedding procedure across methods, including the reweighted repulsion term specified in Appendix D. Appendix D documents the reference construction and visualization protocol.

5.3 Real-Data Distance Recovery and Prediction

We compare distance recovery and downstream prediction on three datasets in Table 1. Caulobacter contains 1,806 fluorescence frames from 301 tracked cells, represented by their native intensity-weighted pixels, for predicting division within 30 minutes (Lambert et al., 2018a). ETH-80 provides 3,280 views of 80 objects for category classification, using 64-atom weighted representations (Leibe & Schiele, 2003). Fingering uses 194 experimental images from nine viscous-displacement runs, aggregated on a 64×64 grid, for classifying stages before and after breakthrough (Keable et al., 2022). Train, validation and test sets are separated by acquisition date, physical object or experimental run, respectively.

The two MALOT variants on ETH-80 and Fingering use common reference banks and parameter values, specified in Appendix E.1. Caulobacter uses validation classification performance to select settings, with $n = 64$ fits for stress and full training sets for ACC. Its radial variant retains the no-radial parameters and refits MDS and the classifier. The table evaluates complete procedures, including their reference choices and downstream representations.

Without radial completion, MALOT reduces stress relative to Fixed LOT by 29.28%, 10.44% and 21.33% across the three datasets. Its downstream scores also improve over that row, although Con-

Table 1: Distance recovery and downstream prediction on three datasets. Entries are means over three Caulobacter date splits, five ETH-80 splits, and three Fingering run splits. Stress is raw distance-recovery error without fitted scaling. Caulobacter uses $n = 64$ for stress and all 576–630 training frames for ACC. Fixed LOT uses a fixed training sample for Caulobacter and a training barycenter for the other datasets. The Caulobacter radial variant keeps the no-radial parameters fixed. Both MALOT rows are shaded, and each column's best mean is bold, including ties. Appendices E.1 and E.3 give distance-recovery visualizations, dataset protocols and the cell-division task context.

Method	Caulobacter		ETH-80		Fingering	
	Stress ↓ $n = 64$	ACC ↑ Full (%)	Stress ↓	ACC ↑ (%)	Stress ↓	ACC ↑ (%)
Fixed LOT	0.13582	85.71	0.04121	80.55	0.00641	91.72
Concat LOT	0.15900	87.16	0.12091	**82.07**	0.00751	93.23
Sliced W_2	0.45997	86.59	0.54897	80.73	0.32782	92.47
MALOT w/o radial	**0.09605**	**87.48**	**0.03691**	81.98	0.00504	**93.99**
MALOT w/ radial	0.09668	**87.48**	0.04138	81.77	**0.00500**	**93.99**

cat LOT gives higher ETH-80 accuracy. Radial completion further lowers Fingering stress, while increasing stress on Caulobacter and ETH-80. Full-training Caulobacter ACC is unchanged, and both MALOT variants attain 93.99% ACC on Fingering. Appendix E.1 provides ETH-80 and Fingering protocols and distance-recovery plots. Appendix E.3 explains the Caulobacter prediction task and evaluation protocol using observed cell-division sequences. Appendices E and E.4 present the supplementary datasets and their protocols.

6 Conclusion

MALOT replaces a single LOT reference with local comparisons across multiple anchors. Endpoint-dependent weights adapt each comparison to nearby references, while radial completion preserves the reference-distance identity for finite transport plans. Under the stated geometric and coverage conditions, the intrinsic construction has $O(h^2)$ relative distance error and the finite two-point construction has $O(h)$ absolute distance error with coverage-scaled temperature. The simulations examine anchor coverage, endpoint weighting and radial completion, and the real-data studies evaluate distance recovery and downstream prediction using a common set of baselines.

AI Use Statement

Generative AI tools assisted with language editing and polishing and with generating synthetic datasets for the simulation experiments. The synthetic-data construction is described in Section 5 and Appendix D, while real-data observations come from the cited datasets. The authors are responsible for the final manuscript and for the validity of the data-generation procedures, analyses and conclusions, including AI-assisted content.

Ethics Statement

The studies use released datasets and involve no new human-subject data collection. The supplementary diffusion-MRI analysis evaluates diffusion-tensor geometry from released scans.

Reproducibility Statement

Appendix A proves the recovery claims, Appendix B treats local geometry, and Appendix C.2 specifies and analyzes MDS. Appendices D–E specify data representations, splits, reference construction, parameter selection and evaluation metrics for the reported experiments. The anonymous supplementary code includes the method implementation, experiment configurations, public-data preparation instructions and scripts for recomputing results or regenerating the reported figures and tables from supplied numerical records.

REFERENCES

Tal Amir and Nadav Dym. Fourier sliced-Wasserstein embedding for multisets and measures. In *International Conference on Learning Representations*, 2025.

R. S. Barlow, S. Meares, G. Magnotti, H. Cutcher, and A. R. Masri. Local extinction and near-field structure in piloted turbulent CH4/air jet flames with inhomogeneous inlets. *Combustion and Flame*, 162(10):3516–3540, 2015.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019.

Alexander Cloninger, Keaton Hamm, Varun Khurana, and Caroline Moosmüller. Linearized Wasserstein dimensionality reduction with approximation guarantees. *Applied and Computational Harmonic Analysis*, 74:101718, 2025. doi: 10.1016/j.acha.2024.101718. URL https://www.sciencedirect.com/science/article/abs/pii/S1063520324000952.

H. C. Cutcher, R. S. Barlow, G. Magnotti, and A. R. Masri. Turbulent flames with compositionally inhomogeneous inlets: Resolved measurements of scalar dissipation rates. *Proceedings of the Combustion Institute*, 36(2):1737–1745, 2017.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78): 1–8, 2021.

Johannes Gasteiger, Marten Lienen, and Stephan Günnemann. Scalable optimal transport in high dimensions for graph distances, embedding alignment, and more. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5616–5627, 2021.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

John C. Gower. Adding a point to vector diagrams in multivariate analysis. *Biometrika*, 55(3): 582–585, 1968.

Doron Haviv, Russell Zhang Kunes, Thomas Dougherty, Cassandra Burdziak, Tal Nawy, Anna Gilbert, and Dana Pe'er. Wasserstein wormhole: Scalable optimal transport distance with transformer. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17697–17718, 2024. URL https://proceedings.mlr.press/v235/haviv24a.html.

Daniel Keable, Alistair Jones, Samuel Krevor, Ann Muggeridge, and Samuel J. Jackson. The effect of viscosity ratio and Peclet number on miscible viscous fingering in a Hele-Shaw cell: A combined numerical and experimental study. *Transport in Porous Media*, 143:23–45, 2022.

Varun Khurana, Harish Kannan, Alexander Cloninger, and Caroline Moosmüller. Supervised learning of sheared distributions using linearized optimal transport. *Sampling Theory, Signal Processing, and Data Analysis*, 21(1):1, 2023.

Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo K. Rohde. Generalized sliced Wasserstein distances. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ambroise Lambert, Aster Vanhecke, Anna Archetti, Seamus Holden, Felix Schaber, Zachary Pincus, Michael T. Laub, Erin Goley, and Suliana Manley. Constriction rate modulation can drive cell size control and homeostasis in C. crescentus. *iScience*, 4:180–189, 2018a.

Ambroise Lambert, Aster Vanhecke, Anna Archetti, Seamus Holden, Felix Schaber, Zachary Pincus, Michael T. Laub, Erin Goley, and Suliana Manley. Constriction rate modulation can drive cell size control and homeostasis in C. crescentus. Data part 1. Zenodo dataset, 2018b. URL https://zenodo.org/records/1248441.

Ambroise Lambert, Aster Vanhecke, Anna Archetti, Seamus Holden, Felix Schaber, Zachary Pincus, Michael T. Laub, Erin Goley, and Suliana Manley. Constriction rate modulation can drive cell size control and homeostasis in C. crescentus. Data part 2. Zenodo dataset, 2018c. URL https://zenodo.org/records/1241005.

John M. Lee. *Introduction to Riemannian Manifolds*. Springer, 2018.

Bastian Leibe and Bernt Schiele. Analyzing appearance and contour based methods for object categorization. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 409–415, 2003.

C. A. Lindstrøm, J. Beinortaitė, J. Björklund Svensson, L. Boulton, J. Chappell, S. Diederichs, B. Foster, J. M. Garland, P. González Caminal, G. Loisch, F. Peña, S. Schröder, M. Thévenet, S. Wesch, M. Wing, J. C. Wood, R. D'Arcy, and J. Osterhoff. Emittance preservation in a plasma-wakefield accelerator. *Nature Communications*, 15(1):6097, 2024a.

Carl A. Lindstrøm, Judita Beinortaitė, Jonas Björklund Svensson, Lewis Boulton, James Chappell, Severin Diederichs, Brian Foster, Jimmy Garland, Pau González Caminal, Gregor Loisch, Felipe Peña, Sarah Schröder, Maxence Thévenet, Matthew Wing, Stephan Wesch, Jonathan C. Wood, Richard D'Arcy, and Jens Osterhoff. Experimental data, analysis scripts and simulations for "Emittance preservation in a plasma-wakefield accelerator". Zenodo dataset, 2024b. URL https://zenodo.org/records/11967839.

John Lott. Some geometric calculations on Wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, 2008.

Scott Mahan, Caroline Moosmüller, and Alexander Cloninger. Point cloud classification via deep set linearized optimal transport. *arXiv preprint arXiv:2401.01460*, 2024.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. URL https://arxiv.org/abs/1802.03426.

Caroline Moosmüller and Alexander Cloninger. Linear optimal transport embedding: Provable Wasserstein classification for certain rigid transformations and perturbations. *Information and Inference*, 12(1):363–389, 2023.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

Clément Sarrazin and Bernhard Schmitzer. Linearized optimal transport on manifolds. *SIAM Journal on Mathematical Analysis*, 56(4):4970–5016, 2024. doi: 10.1137/23M1564535. URL https://epubs.siam.org/doi/abs/10.1137/23M1564535.

Meyer Scetbon, Marco Cuturi, and Gabriel Peyré. Low-rank Sinkhorn factorization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9344–9354, 2021.

Zoltán Szabó, Bharath K. Sriperumbudur, Barnabás Póczos, and Arthur Gretton. Learning theory for distribution regression. *Journal of Machine Learning Research*, 17(152):1–40, 2016.

The University of Sydney, Clean Combustion Research Group. Database for turbulent piloted flames with compositionally inhomogeneous inlets. University of Sydney data repository, 2026. URL https://web.aeromech.usyd.edu.au/thermofluids/Inhomog_Meares/inhom_website.php. Webpage updated February 12, 2026; accessed September 26, 2026.

Viet-Hoang Tran, Minh-Khoi Nguyen-Nhat, Trang Pham, Thanh T. Chu, Tam Le, and Tan M. Nguyen. Distance-based tree-sliced Wasserstein distance. In *International Conference on Learning Representations*, 2025.

Nina Vesseron, Elsa Cazelles, Alice Le Brigant, and Thierry Klein. On the Wasserstein geodesic principal component analysis of probability measures. In *International Conference on Learning Representations*, 2026.

Wei Wang, Dejan Slepčev, Saurav Basu, John A. Ozolek, and Gustavo K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *International Journal of Computer Vision*, 101:254–269, 2013.

A Recovery Proofs

This appendix proves the statements for the implemented finite-support construction. The only transport-specific input to the principal bound is the exact reference radius. No differentiability, absolute continuity or common support is needed here.

A.1 Variance Identity and Radial Comparison

Lemma 1 (Radial completion)**.** *Let* $\mathbf{P}$ *be any feasible coupling from* $\rho = \sum_a a_a \delta_{\boldsymbol{z}_a}$ *to* $\mu = \sum_b b_b \delta_{\boldsymbol{x}_b}$*, with* $a_a > 0$*. Define* $\boldsymbol{t}_a = a_a^{-1} \sum_b P_{ab} \boldsymbol{x}_b$ *and* $\boldsymbol{\phi} = \mathrm{vec}[\sqrt{a_a}(\boldsymbol{t}_a - \boldsymbol{z}_a)]_a$*. Then*

$$c := \sum_{a,b} P_{ab} \|\boldsymbol{z}_a - \boldsymbol{x}_b\|^2 = \|\boldsymbol{\phi}\|^2 + V, \qquad V := \sum_{a,b} P_{ab} \|\boldsymbol{x}_b - \boldsymbol{t}_a\|^2 \geq 0.$$

Consequently $\boldsymbol{\psi} = (\boldsymbol{\phi}, \sqrt{V})$ *has norm* $\sqrt{c}$*. For an optimal plan this norm is* $W_2(\rho, \mu)$*.*

Proof. Expand $\boldsymbol{x}_b - \boldsymbol{z}_a = (\boldsymbol{t}_a - \boldsymbol{z}_a) + (\boldsymbol{x}_b - \boldsymbol{t}_a)$. For every anchor atom, $\sum_b P_{ab}(\boldsymbol{x}_b - \boldsymbol{t}_a) = \mathbf{0}$, so the cross term vanishes after summation. The two remaining terms are exactly $\|\boldsymbol{\phi}\|^2$ and V. Optimality is used only to identify the cost with W_2^2. □

Lemma 2 (One nearby endpoint suffices)**.** *Fix an anchor* ρ_k *and features in a common Hilbert space with norms* $\|\boldsymbol{\psi}_{ki}\| = r_{ki} = W_2(\rho_k, \mu_i)$*. Then*

$$\big|\|\boldsymbol{\psi}_{ki} - \boldsymbol{\psi}_{kj}\| - W_{ij}\big| \leq 2\min(r_{ki}, r_{kj}).$$

Proof. The Hilbert triangle and reverse triangle inequalities put the feature distance in $[|r_{ki} - r_{kj}|, r_{ki} + r_{kj}]$. The Wasserstein triangle and reverse triangle inequalities put W_{ij} in the same interval. Its width is $2\min(r_{ki}, r_{kj})$. □

Any radially exact representation is reliable for a pair when one endpoint is close to its reference. LOT additionally supplies directional displacement information, which can improve approximation before the covering radius is small. The lemma also explains why the same theorem applies to ideal LOT: its displacement norm is exactly its reference-to-target Wasserstein distance.

A.2 Proof of Theorem 2

We prove the sharper pair-specific bound first. Retaining the notation of Section 4, define

$$h_{\mathcal{R}}(\mathcal{D}) = \sup_{\mu \in \mathcal{D}} \min_k W_2(\rho_k, \mu), \qquad q_{ij} = \sum_k w_{k,ij} \min\{r_{ki}^2, r_{kj}^2\}. \tag{15}$$

Then

$$|D_{ij} - W_{ij}| \leq 2\sqrt{q_{ij}} \leq 2\sqrt{(h_i^2 + h_j^2)/2 + \tau \log K}. \tag{16}$$

The quantity q_{ij} is computable from exact stored reference costs and endpoint weights, without a target-pair OT solve. It allows a different endpoint to be close to each contributing anchor.

For each pair, write $w_k = w_{k,ij}$ and $L_k = L_{k,ij}$. The weights sum to one. The reverse triangle inequality in $\mathbb{R}^K$, followed by Lemma 2, gives

$$|D_{ij} - W_{ij}| = |\|(\sqrt{w_k}L_k)_k\|_2 - \|(\sqrt{w_k}W_{ij})_k\|_2| \\ \leq \left(\sum_k w_k(L_k - W_{ij})^2\right)^{1/2} \leq 2\left(\sum_k w_k \min(r_{ki}^2, r_{kj}^2)\right)^{1/2}.$$

For completeness, let $a_k = \exp(-c_k/\tau)/Z$, $c_* = \min_k c_k$ and $H(a) = -\sum_k a_k \log a_k$. The identity

$$\sum_k a_k c_k = -\tau \log Z + \tau H(a) \leq c_* + \tau \log K$$

follows from $Z \geq \exp(-c_*/\tau)$ and $H(a) \leq \log K$. Because the pair weights are the arithmetic mean of endpoint weights,

$$q_{ij} = \frac{1}{2}\sum_k (\alpha_{ki} + \alpha_{kj}) \min(r_{ki}^2, r_{kj}^2) \\ \leq \frac{1}{2}\left(\sum_k \alpha_{ki} r_{ki}^2 + \sum_k \alpha_{kj} r_{kj}^2\right) \leq \frac{h_i^2 + h_j^2}{2} + \tau \log K.$$

Taking a supremum proves Equation (13). At $\tau = 0$, each endpoint's weights are supported on its minimizers and the expectation is exactly h_i^2. This proves the zero-temperature statement as well.

If every evaluated measure is itself an anchor, each endpoint's selected anchors are identical to that measure in W_2. Every contributing L_k then equals W_{ij} by Lemma 2. Thus $D_{ij} = W_{ij}$ exactly at zero temperature, irrespective of optimal-plan nonuniqueness away from the endpoint.

A.3 FPS, Compression and Solver Error

Corollary 1 (FPS, approximate plans and observed measures)**.** *Let a finite candidate pool S approximate $\mathcal{D}$ within $e = \sup_{\mu\in\mathcal{D}} \min_{\sigma\in S} W_2(\mu, \sigma)$, and let $r^\star_{K,W}$ be its optimal discrete K-center radius. Exact-W_2 FPS gives $h_{\mathcal{R}}(\mathcal{D}) \leq e + 2r^\star_{K,W}$. If selected stored anchors are compressed within W_2 error ξ of their candidate measures, farthest-first selection using distances to these stored anchors gives*

$$h_{\mathcal{R}}(\mathcal{D}) \leq H_K := e + 2r^\star_{K,W} + \xi. \tag{17}$$

For feasible reference plans with primal suboptimality at most $\varepsilon_{\mathrm{OT}}$ and observed measures within s of their latent targets, the uniform recovery bound is

$$2\sqrt{H_K^2 + \varepsilon_{\mathrm{OT}} + \tau \log K} + \sqrt{\varepsilon_{\mathrm{OT}}} + 2s. \tag{18}$$

Here the FPS coverage statement uses exact selection distances; the solver term pertains to the plans used in the pairwise estimator.

Vanishing coverage, compression, solver, observation and temperature terms suffice for consistency through this bound. The following covering argument and plan-error calculation prove the corollary.

Lemma 3 (Farthest-first with compressed stored anchors)**.** *Let S be a finite subset of a metric space (X, d) and select $K \leq |S|$ distinct candidates $\sigma_1, \ldots, \sigma_K$ by farthest-first traversal. The distance used for selection is distance to stored anchors $\rho_1, \ldots, \rho_{k-1}$, with $d(\sigma_j, \rho_j) \leq \xi$. If $r^\star_K$ is the optimal K-center radius on S with centers in S, then*

$$\max_{\sigma\in S} \min_{j\leq K} d(\sigma, \rho_j) \leq 2r^\star_K + \xi.$$

Proof. Let the left side be H. The claim is immediate when $H \leq \xi$. Otherwise, a candidate already selected has distance at most $\xi < H$ to the stored set, so any farthest candidate at every preceding step with radius at least H is unselected. For $i < j$, the selection rule implies $d(\sigma_j, \rho_i) \geq H$ and hence $d(\sigma_j, \sigma_i) \geq H - \xi$. A final farthest point σ_{K+1} is unselected and satisfies the same separation from the K selected candidates. Of these $K + 1$ points, two belong to one ball in an optimal K-center cover. Their separation is at most $2r^\star_K$. Therefore $H - \xi \leq 2r^\star_K$. When $K = |S|$, $H \leq \xi$ directly. □

For a family $\mathcal{D}$ within distance e of the candidate pool, the triangle inequality adds e to this radius. Setting $\xi = 0$ recovers the standard FPS guarantee (Gonzalez, 1985). This FPS guarantee uses exact selection distances.

For a feasible, possibly suboptimal plan, write $c_{ki} = r_{ki}^2 + \varepsilon_{ki}$ with $0 \le \varepsilon_{ki} \le \varepsilon_{\mathrm{OT}}$ and $s_{ki} = \sqrt{c_{ki}}$. Lemma 1 gives $\|\boldsymbol{\psi}_{ki}\| = s_{ki}$ and $0 \le s_{ki} - r_{ki} \le \sqrt{\varepsilon_{\mathrm{OT}}}$. Using endpoint i, the triangle inequalities yield

$$|L_{k,ij} - W_{ij}| \le |L_{k,ij} - s_{kj}| + |s_{kj} - r_{kj}| + |r_{kj} - W_{ij}| \le 2s_{ki} + \sqrt{\varepsilon_{\mathrm{OT}}}.$$

Interchanging i and j gives the analogous inequality with s_{kj}. Weighted Minkowski and the same Gibbs expectation bound, now applied to the actual costs c_{ki} used in the weights, imply

$$|D_{ij} - W_{ij}| \le 2\sqrt{h_{\mathcal{R}}(\mathcal{D})^2 + \varepsilon_{\mathrm{OT}} + \tau \log K} + \sqrt{\varepsilon_{\mathrm{OT}}}.$$

Finally, if $W_2(\mu_i, \mu_i^\circ) \le s$ for latent targets μ_i°, then $|W_{ij} - W_2(\mu_i^\circ, \mu_j^\circ)| \le 2s$. Combining this with Lemma 3 proves Corollary 1. The covering radius in this application is that of the observed measures; the final $2s$ compares their pairwise target to the latent target.

A.4 Approximate Metric Properties

MALOT has a zero diagonal because the two stored feature blocks of an object are identical, and symmetry follows from symmetric endpoint weights. The square root of a nonnegative weighted sum proves nonnegativity. If $|D_{ij} - W_{ij}| \le \epsilon$ for all pairs, then

$$D_{i\ell} \le W_{i\ell} + \epsilon \le W_{ij} + W_{j\ell} + \epsilon \le D_{ij} + D_{j\ell} + 3\epsilon.$$

Also $D_{ij} \ge W_{ij} - \epsilon$, proving the separation statement. For a finite family of distinct measures, sufficiently small error enforces positive off-diagonal entries.

A.5 A Sharp Obstruction for Fixed Hilbert Representations

Consider $\mu_j = (\delta_{\boldsymbol{u}_j} + \delta_{-\boldsymbol{u}_j})/2$, where $\boldsymbol{u}_j = (\cos(j\pi/4), \sin(j\pi/4))$, $j = 0, 1, 2, 3$. An optimal coupling either preserves or swaps the two atoms, giving

$$W_2^2(\mu_i, \mu_j) = 2 - 2|\boldsymbol{u}_i^{\mathsf{T}} \boldsymbol{u}_j|.$$

The four cyclic sides have length $a = \sqrt{2 - \sqrt{2}}$; both diagonals have length $\sqrt{2}$. For any four Hilbert points $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_3$,

$$\|\boldsymbol{y}_0 - \boldsymbol{y}_2\|^2 + \|\boldsymbol{y}_1 - \boldsymbol{y}_3\|^2 \le \sum_{j=0}^{3} \|\boldsymbol{y}_j - \boldsymbol{y}_{j+1 \bmod 4}\|^2.$$

Indeed, the right side minus the left side is $\|\boldsymbol{y}_0 - \boldsymbol{y}_1 + \boldsymbol{y}_2 - \boldsymbol{y}_3\|^2$. If every distance has relative error at most $0 \le \varepsilon < 1$, this inequality forces $4(1-\varepsilon)^2 \le 4a^2(1+\varepsilon)^2$, hence $\varepsilon \ge (1-a)/(1+a)$. For $\varepsilon \ge 1$, this lower bound holds trivially. Equality is attained by a Euclidean square of side $2a/(1+a)$, proving sharpness.

Any fixed LOT feature map, fixed weighted concatenation or MDS embedding is a Hilbert representation and is subject to this bound. MALOT's pair-dependent weights are not one such representation. Using the four measures as anchors gives $h = 0$ and error at most $2\sqrt{\tau \log 4}$, which tends to zero. Since the minimum nonzero target distance is a, the corresponding maximum relative error is at most $2\sqrt{\tau \log 4}/a$. This exhibits a finite family on which MALOT escapes an unavoidable fixed-embedding distortion.

B Local Wasserstein Geometry and Its Finite Approximation

B.1 Regularity and the Objects Being Compared

The geometric analysis concerns a finite-dimensional smooth submanifold $\mathcal{M} \subset \mathcal{P}_2(\mathbb{R}^d)$ with a smooth positive definite metric g_W whose norm equals ambient Wasserstein speed. For absolutely continuous anchors, the exact LOT maps ϕ_ρ are assumed smooth on the region under consideration.

Logarithm compatibility means that, on a normal neighborhood of ρ, there is a linear isometry $I_\rho \colon T_\rho\mathcal{M} \to L^2(\rho;\mathbb{R}^d)$ such that

$$\phi_\rho(\mu) = I_\rho \operatorname{Log}_\rho^{\mathcal{M}}(\mu).$$

The isometry is fixed across all targets for a given anchor. It follows from LOT radial exactness that the norm of this Riemannian logarithm equals $W_2(\rho,\mu)$. Logarithm compatibility is an additional geometric assumption on $\mathcal{M}$.

We use a uniform normal radius $r_0 > 0$, a sectional-curvature bound $|\operatorname{Sec}| \le \kappa$ along the relevant radial geodesics, and $\kappa r_0^2 \le 1/4$. Each compatible neighborhood contains the exponential image of the closed tangent ball of radius r_0, with the curvature bound valid along every radial geodesic in that ball. Specifically, the compatible normal neighborhood N_ρ must contain every target $\mu \in U$ with $W_2(\rho,\mu) \le r_0$. An intrinsic normal radius alone does not imply this inclusion for a submanifold that can return ambient-close to a reference outside its local branch. References outside these neighborhoods require a separate finite derivative bound $g_\rho \preceq (1+B)g_W$. For example, such a B exists if anchors and targets range over compact subsets of a regular region and the reference maps depend jointly and smoothly on both, after isometric identifications of the reference Hilbert spaces. Compactness of the unit tangent bundle over the target set then bounds the largest eigenvalue of each pullback relative to g_W. All these constants must be independent of K in an asymptotic claim.

For connected open U, the intrinsic distance $d_{W,U}$ integrates g_W along paths constrained to U. Equality with ambient W_2 additionally requires appropriate Wasserstein geodesics to stay in U.

B.2 Proof of Proposition 1

Put $z = \operatorname{Log}_\rho^{\mathcal{M}}(\mu)$ with $\|z\| = r$ and let $\gamma(t) = \operatorname{Exp}_\rho(tz)$ for $0 \le t \le 1$. Parallel transport back to $T_\rho\mathcal{M}$ identifies the variation field of the exponential with an operator $F(t)$ satisfying

$$F''(t) + \mathcal{R}(t)F(t) = 0, \qquad F(0) = 0, \qquad F'(0) = I, \qquad \|\mathcal{R}(t)\| \le \kappa r^2.$$

The last inequality follows because the curvature operator $v \mapsto R(v,\dot\gamma)\dot\gamma$ is self-adjoint and its eigenvalues are controlled by sectional curvature. This is the usual Jacobi-field comparison for the exponential differential (Lee, 2018). Writing $a = \kappa r^2$ and integrating twice yields

$$F(t) = tI - \int_0^t (t-s)\mathcal{R}(s)F(s)\,\mathrm{d}s.$$

If $M_0 = \sup_{0<t\le 1} \|F(t)\|/t$, then $M_0 \le 1 + aM_0/6$. Thus $M_0 \le (1-a/6)^{-1}$ and

$$\|F(1) - I\| \le a/(6-a).$$

Set $e = a/(6-a)$. The singular values of $F(1)$ lie in $[1-e, 1+e]$, so the squared singular values of its inverse lie in $[(1+e)^{-2}, (1-e)^{-2}]$. For $a \le 1/4$, both endpoints differ from one by at most $4a$. Parallel transport and I_ρ are isometries, and the logarithm differential is the inverse exponential differential. This proves the absolute quadratic-form bound in Equation (5) with $C_0 = 4$. No curvature-derivative bound is needed for this estimate.

For the pairwise statement, take a Wasserstein minimizing geodesic γ contained in the indicated normal neighborhood and in $\mathcal{M}$. The chord of its LOT image is no longer than its image-curve length, whence

$$d_\rho^{\mathrm{LOT}}(\mu,\nu) \le \int_0^1 \|D\phi_\rho(\gamma(t))[\dot\gamma(t)]\|\,\mathrm{d}t \le \sqrt{1 + C_0\kappa R^2}\, W_2(\mu,\nu).$$

To control the other direction, put $z_\mu = \operatorname{Log}_\rho^{\mathcal{M}}(\mu)$ and $z_\nu = \operatorname{Log}_\rho^{\mathcal{M}}(\nu)$. Their joining line segment stays in the tangent ball of radius R. With $a_R = \kappa R^2$ and $e_R = a_R/(6-a_R)$, the same Jacobi estimate bounds the exponential differential on this segment by $1 + e_R$. Its exponential image joins μ to ν and has Wasserstein length at most $(1+e_R)\|z_\mu - z_\nu\|$. Logarithm compatibility identifies this last norm with $d_\rho^{\mathrm{LOT}}(\mu,\nu)$, giving

$$\frac{W_2(\mu,\nu)}{1+e_R} \le d_\rho^{\mathrm{LOT}}(\mu,\nu) \le \sqrt{1+4a_R}\, W_2(\mu,\nu).$$

Since $1 - (1+e_R)^{-2} \le 4a_R$, the two estimates prove Equation (6).

B.3 Proof of Theorem 1

At a target μ, write $r_k = W_2(\rho_k, \mu)$ and $g_k = g_{\rho_k,\mu}$. For $r_k \le r_0$, the explicit inclusion places μ in N_k, so Proposition 1 bounds $|g_k(v,v) - g_W(v,v)|$ by $C_0\kappa r_k^2 g_W(v,v)$. For $r_k > r_0$, the pullback is positive semidefinite and satisfies $g_k \preceq (1+B)g_W$. Hence $|g_k(v,v) - g_W(v,v)| \le (1+B)g_W(v,v)$. Therefore, with $A = \max\{C_0\kappa, (1+B)/r_0^2\}$,

$$\left|g_k(v,v) - g_W(v,v)\right| \le A r_k^2 g_W(v,v) \quad \text{for every } v \in T_\mu\mathcal{M}. \tag{19}$$

The constant A records both local curvature and regularity of distant references.

Let $Z = \sum_k \exp(-r_k^2/\tau)$ and $\mathcal{E}(\alpha) = -\sum_k \alpha_k \log \alpha_k$. The normalized exponential weights satisfy

$$\sum_k \alpha_k r_k^2 = -\tau \log Z + \tau \mathcal{E}(\alpha) \le \min_k r_k^2 + \tau \log K \le h^2 + \tau \log K.$$

Weighting Equation (19) and applying the triangle inequality gives the quadratic-form comparison in Equation (9). Smooth maps and $\tau > 0$ give smooth weights. For $\delta < 1$, this comparison gives $g_{\mathcal{R}} \succeq (1-\delta)g_W$, which ensures positive definiteness. Hence $g_{\mathcal{R}}$ is a Riemannian metric, and its intrinsic length distance is finite and is a metric on connected U. Completeness is unnecessary for this statement.

For every piecewise smooth connecting path in U, the tensor bound compares its two lengths by factors $\sqrt{1-\delta}$ and $\sqrt{1+\delta}$. Taking infima over the same path class bounds the ratio of the intrinsic distances by the same factors. If an ambient Wasserstein minimizing geodesic lies in U, equality of curve speeds gives $d_{W,U} = W_2$. For distinct endpoints, the relative error is bounded uniformly by

$$\left|\frac{d_{\mathcal{R},U}(\mu,\nu)}{d_{W,U}(\mu,\nu)} - 1\right| \le 1 - \sqrt{1-\delta} = \frac{\delta}{1+\sqrt{1-\delta}}.$$

This proves convergence with fixed geometric constants. On a family of finite $d_{W,U}$-diameter it also implies uniform absolute convergence.

For a finite candidate pool S, write $e_S = \sup_{\mu\in U} \min_{\sigma\in S} W_2(\mu,\sigma)$ and let $r_K^\star(S)$ be the optimal discrete K-center radius. Exact FPS implies $h \le e_S + 2r_K^\star(S)$ by the same covering argument used in Corollary 1. Thus a sequence with $e_S \to 0$, $r_K^\star(S) \to 0$ and $\tau_K \log K \to 0$ has geometric recovery. Total boundedness permits such sequences.

B.4 Bures–Wasserstein Space

For $\rho = \mathcal{N}(m_\rho, \Sigma_\rho)$ and $\mu = \mathcal{N}(m_\mu, \Sigma_\mu)$ with positive definite covariances, the optimal map is

$$T_{\rho\to\mu}(x) = m_\mu + A_{\rho\mu}(x - m_\rho), \qquad A_{\rho\mu} = \Sigma_\rho^{-1/2}(\Sigma_\rho^{1/2}\Sigma_\mu\Sigma_\rho^{1/2})^{1/2}\Sigma_\rho^{-1/2}.$$

Bures–Wasserstein space has a Euclidean mean component and a covariance component carrying the Bures–Wasserstein metric (Bhatia et al., 2019). At a fixed mean, the space is identified with the positive definite covariance matrices. Displacement interpolation preserves Gaussianity, so the optimal displacement is the Riemannian logarithm under the natural velocity-field identification. Choose compact anchor and target regions with uniformly bounded means and covariance eigenvalues bounded above and away from zero. Matrix square roots and the reference maps then have bounded derivatives on these regions, and curvature is bounded on the relevant compact regular neighborhoods. A sufficiently small uniform normal radius supplies the local conditions, including the ambient-ball inclusion because intrinsic distance equals ambient W_2 in the full Bures–Wasserstein space. Geodesic containment must still be checked for the chosen U or ensured by a geodesically convex region. The constants can deteriorate as covariances approach singularity. This example explains why the geometric result has content beyond flat location families while retaining explicit regularity boundaries.

B.5 Local Consistency of the Two-Point Rule

Proposition 2 (Consistency of the two-point rule)**.** *For smooth exact LOT maps on $\mathcal{M}$, fixed K, $\tau > 0$, and a smooth curve μ_t with $\mu_0 = \mu$ and $\dot\mu_0 = v$, Equation* (10) *satisfies*

$$D_{\mathcal{R}}^2(\mu, \mu_t) = t^2 g_{\mathcal{R},\mu}(v,v) + o(t^2). \tag{20}$$

The finite rule and the intrinsic metric therefore agree to leading order along each smooth target motion. This statement concerns fixed anchors and positive temperature; the separate coverage results control changing anchor banks.

Proof. For a fixed finite anchor bank, differentiability gives $\phi_k(\mu_t) - \phi_k(\mu) = tD\phi_k(\mu)[v] + o(t)$. Continuity of the weights gives $w_k(\mu, \mu_t) = \alpha_k(\mu) + o(1)$. Substitution into Equation (10) proves Equation (20). The derivatives of the weights affect only higher-order terms, because they multiply squared increments of order t^2. When $g_{\mathcal{R}}$ is positive definite, as ensured by Theorem 1 for $\delta < 1$, the squared intrinsic distance along the same curve also has leading term $t^2 g_{\mathcal{R},\mu}(v, v)$. □

B.6 Finite-Pair Approximation of the Intrinsic Metric

The local and finite-pair results can also be combined without interchanging their limits. Suppose the hypotheses of Theorem 1 hold on a region of Wasserstein diameter $\Delta < \infty$, with the geodesic-containment condition. For the exact-map MALOT rule, set $\zeta = h^2 + \tau \log K$ and assume $A\zeta < 1$. Theorem 2 and the geometric bound give

$$\sup_{\mu,\nu \in U} |D_{\mathcal{R}}(\mu, \nu) - d_{\mathcal{R},U}(\mu, \nu)| \le 2\sqrt{\zeta} + \frac{A\zeta}{1 + \sqrt{1 - A\zeta}}\,\Delta. \tag{21}$$

Indeed, insert $W_2(\mu, \nu)$ between the two terms and apply the triangle inequality for real numbers. Thus the symmetric rule approximates the constructed intrinsic metric uniformly as the anchor cover and temperature terms vanish. For empirical approximations to a smooth family, the observation, compression and feasible-plan errors in Corollary 1 must also vanish. The arbitrary finite-support recovery theorem itself requires none of the smoothness or geometric assumptions above.

C Numerical Implementation and Embedding

C.1 Discrete LOT Blocks and Radial Completion

For $\rho_k = \sum_{a=1}^{m_k} a_{ka}\delta_{\boldsymbol{z}_{ka}}$ and $\mu_i = \sum_{b=1}^{p_i} b_{ib}\delta_{\boldsymbol{x}_{ib}}$ in $\mathbb{R}^d$, let $\mathbf{P}_{ki}$ be an optimal coupling, with positive anchor masses a_{ka}. Its barycentric displacement and raw quadratic cost are

$$\begin{aligned} \boldsymbol{t}_{kia} &= a_{ka}^{-1} \sum_b P_{ki,ab}\boldsymbol{x}_{ib}, \qquad \boldsymbol{\phi}_{ki} = \operatorname{vec}\left[\sqrt{a_{ka}}(\boldsymbol{t}_{kia} - \boldsymbol{z}_{ka})\right]_a, \\ c_{ki} &= \sum_{a,b} P_{ki,ab}\|\boldsymbol{z}_{ka} - \boldsymbol{x}_{ib}\|^2, \qquad V_{ki} = \sum_{a,b} P_{ki,ab}\|\boldsymbol{x}_{ib} - \boldsymbol{t}_{kia}\|^2 = c_{ki} - \|\boldsymbol{\phi}_{ki}\|^2. \end{aligned} \tag{22}$$

Here $\boldsymbol{\phi}_{ki} \in \mathbb{R}^{m_k d}$ and $V_{ki} \ge 0$ is the conditional variance lost by barycentric projection. Lemma 1 proves this identity by decomposing each transport displacement into its conditional mean and residual. When mass splits, the LOT norm alone can underestimate the reference cost even for an optimal plan. The completed block and its pairwise separation are

$$\boldsymbol{\psi}_{ki} = (\boldsymbol{\phi}_{ki}, \sqrt{V_{ki}}) \in \mathbb{R}^{m_k d + 1}, \qquad L_{k,ij}^2 = \|\boldsymbol{\phi}_{ki} - \boldsymbol{\phi}_{kj}\|^2 + (\sqrt{V_{ki}} - \sqrt{V_{kj}})^2. \tag{23}$$

Thus $\|\boldsymbol{\psi}_{ki}\|^2 = c_{ki} = W_2^2(\rho_k, \mu_i)$, preserving the radial identity used in the discrete recovery proof. With deterministic plans, $V_{ki} = 0$ and the correction vanishes. The same plan supplies the block, variance and cost, so completion and weighting require no additional OT solves.

With one anchor, MALOT reduces to LOT for deterministic plans and to radially completed LOT for general split plans. The ablation without radial completion omits the final term in Equation (23) while retaining the endpoint weights. Multi-reference LOT concatenates blocks from fixed references (Khurana et al., 2023). We normalize this baseline by $1/\sqrt{K}$, giving squared distance $K^{-1}\sum_k \|\boldsymbol{\phi}_{ki} - \boldsymbol{\phi}_{kj}\|^2$. When radial completion vanishes, comparing MALOT and Concat LOT on the same reference bank isolates endpoint weighting.

For each reference and target, the feasible plan set is

$$\{\mathbf{P} \ge 0 : \ \mathbf{P}\mathbf{1} = \boldsymbol{a}, \quad \mathbf{P}^{\mathsf{T}}\mathbf{1} = \boldsymbol{b}\}.$$

We solve the linear transport problem with quadratic Euclidean ground costs using POT (Flamary et al., 2021). All weights are normalized to unit mass; zero-weight support points are removed. Target and reference supports can differ in size and location. The reference support remains fixed for every block bearing the same anchor index, so its coordinates can be subtracted across targets.

We evaluate the softmax after subtracting each endpoint's minimum cost. At zero temperature, its limiting convention is uniform weight over cost minimizers. For inexact feasible plans, c_{ki} is the primal quadratic cost, not an entropy-penalized objective. Corollary 1 gives the corresponding solver-error bound. In floating-point arithmetic, a tiny negative value of $c - \|\boldsymbol{\phi}\|^2$ can arise from cancellation. Only such numerical negatives are clipped to zero; solver feasibility and mass conservation are checked separately.

For deterministic uniform-to-uniform assignment plans, radial variance is zero. For compressed references or unequal atom masses, optimal plans can split mass and radial variance need not vanish. These cases explain why the same endpoint rule can need a radial correction in one dataset but not another.

C.2 MDS and the Fixed-Anchor Out-of-Sample Map

Let D_{ij} be training dissimilarities and define

$$R_i = \frac{1}{n}\sum_j D_{ij}^2, \qquad G = \frac{1}{n^2}\sum_{i,j} D_{ij}^2, \qquad \mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{D}^{\circ 2}\mathbf{H}.$$

Here $\mathbf{H} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^\mathsf{T}/n$ and $\mathbf{D}^{\circ 2}$ denotes entrywise squares. Write $\lambda_1 \geq \cdots \geq \lambda_n$ for the eigenvalues of $\mathbf{B}$. Retaining r positive eigenvalues gives $\mathbf{Y} = \mathbf{U}_r\mathbf{\Lambda}_r^{1/2}$. Negative eigenvalues are discarded.

For a new measure μ_*, keep the training anchors, preprocessing, temperature, row means and eigenvectors fixed. Its K reference transports give $\boldsymbol{\psi}_{k*}$ and α_{k*}. Compute d_{*i} by Equation (11), with one training and one new endpoint. Then the Gower extension (Gower, 1968) is

$$b_{*i} = -\frac{1}{2}\left(d_{*i}^2 - \frac{1}{n}\sum_j d_{*j}^2 - R_i + G\right), \qquad \boldsymbol{y}_* = \mathbf{\Lambda}_r^{-1/2}\mathbf{U}_r^\mathsf{T}\boldsymbol{b}_*.$$

This is an inductive projection through a fixed training eigensystem.

Proposition 3 (MDS recovery relative to exact Wasserstein MDS)**.** *For a fixed set of n measures with Wasserstein diameter Δ, suppose $\max_{ij}|D_{ij} - W_{ij}| \leq \epsilon$. Then, with $\|\cdot\|_2$ denoting the spectral norm,*

$$\|\mathbf{B} - \mathbf{B}_W\|_2 \leq n(2\Delta\epsilon + \epsilon^2)/2. \tag{24}$$

If the exact centered matrix $\mathbf{B}_W$ has $\lambda_r(\mathbf{B}_W) > 0$ and $\lambda_r(\mathbf{B}_W) > \lambda_{r+1}(\mathbf{B}_W)$, its rank-$r$ positive spectral truncation and that of $\mathbf{B}$ converge as $\epsilon \to 0$. Their MDS coordinates converge up to an orthogonal transformation, and all embedded pairwise distances converge.

For an additional fixed query, suppose its training distances also converge uniformly and are uniformly bounded. Under the same eigenvalue conditions, its Gower coordinates converge, under the same alignment, to the coordinates obtained using exact Wasserstein training and query distances.

Proof. The squared-distance perturbation obeys $|D_{ij}^2 - W_{ij}^2| \leq \epsilon(2\Delta + \epsilon)$. Since $\|\mathbf{H}\|_2 = 1$ and an n-by-n matrix with entries bounded by a has spectral norm at most na,

$$\|\mathbf{B} - \mathbf{B}_W\|_2 \leq \frac{n}{2}\epsilon(2\Delta + \epsilon).$$

Eigenvalues are continuous under this perturbation. The separating eigengap ensures convergence of the retained spectral subspace, by spectral-projector perturbation (Davis & Kahan, 1970). The positive lower eigenvalue bound makes the square-root factors continuous on that subspace. Choosing a common orthogonal alignment yields convergence of coordinates; convergence of finitely many Euclidean distances follows. Equivalently, each squared embedded distance is the continuous expression $B_{r,ii} + B_{r,jj} - 2B_{r,ij}$ in the truncated Gram matrix. Internal repeated retained eigenvalues do not prevent convergence up to rotation.

For the query, let ϵ_* bound query distance error and let $\Delta_* = \max_i W_2(\mu_*, \mu_i)$. The two query-square terms in $\boldsymbol{b}_*$ contribute at most $\epsilon_*(2\Delta_* + \epsilon_*)$ per coordinate after multiplication by $1/2$. The two training-square terms contribute at most $\epsilon(2\Delta + \epsilon)$. Thus

$$\|\widehat{\boldsymbol{b}}_* - \boldsymbol{b}_{*,W}\|_2 \le \sqrt{n}\{\epsilon_*(2\Delta_* + \epsilon_*) + \epsilon(2\Delta + \epsilon)\}.$$

For a fixed projector, multiplication by $\boldsymbol{\Lambda}_r^{-1/2}\mathbf{U}_r^\mathsf{T}$ has norm $\lambda_r^{-1/2}$. When the training eigensystem also varies, its aligned inverse-square-root factor converges by the same spectral-gap argument. The query vector is bounded, so the product converges. □

Proposition 3 holds for fixed n and fixed retained rank. Combining it with Theorem 2 gives convergence when the covering radius of the training and query distributions and the temperature term vanish.

C.3 Computational Comparison

Algorithm 1 in the main text implements reference selection and comparison.

Let each target have at most p atoms in $\mathbb{R}^d$ and each anchor at most m atoms. Write $T_{\mathrm{OT}}(m, p)$ for one exact transport solve including its cost matrix, and $q = md + 1$ for a completed block. For a dense n-object matrix, the principal workloads are

exact OT:	$O(n^2 T_{\mathrm{OT}}(p, p))$,
LOT:	$O(nT_{\mathrm{OT}}(m, p) + n^2 md)$,
MALOT:	$O(nKT_{\mathrm{OT}}(m, p) + n^2 Kq)$,
dense Sinkhorn:	$O(n^2 Ip^2)$ for I iterations per pair,
sliced Wasserstein:	$O(nLpd + nLp\log p + n^2 Lp)$

for L directions and equal-weight sorted projected samples in the last line. Weighted measures may instead use a quantile discretization; its resolution replaces p in the comparison stage. These expressions exclude construction of a barycenter reference, hyperparameter search and optional MDS. FPS additionally updates nearest-anchor costs; with exact costs already computed for its selected references, these updates require $O(nK)$ arithmetic. Support sizes additionally influence solver work and wall time.

Full MDS costs $O(n^3)$ and $O(n^2)$ memory. A query requires K reference solves, $O(nKq)$ distance arithmetic and $O(nr)$ fixed-projection work.

D Simulation and Sydney Protocols

D.1 Multi-Cluster Distributions

The simulation seed is 7. Twelve prototype shapes generate ten distributions each, with 80 uniform atoms per distribution. One shared base realization has four components with centers $(1.35, 0.05)$, $(-0.55, 0.95)$, $(-0.82, -0.55)$ and $(0.25, -0.20)$, and coordinatewise Gaussian standard deviations $(0.08, 0.22)$, $(0.18, 0.07)$, $(0.10, 0.12)$ and $(0.05, 0.05)$. Atom indices cycle through these components, allocating 20 atoms to each. After one Gaussian draw using NumPy's default generator, the 80 points are centered by their common empirical mean. All twelve prototypes rotate this same base by $2\pi k/12$, $k = 0, \ldots, 11$; all 120 measures also deform this shared realization, rather than independently resampling a mixture. Within-family perturbations use rotation noise 0.22, diagonal shear noise 0.025, off-diagonal shear noise 0.018 and point noise 0.025. Specifically, each measure rotates the base by $2\pi k/12 + N(0, 0.22^2)$, applies a matrix with independent diagonal entries $1 + N(0, 0.025^2)$ and off-diagonal entries $N(0, 0.018^2)$ about its mean, and adds independent coordinate noise $N(0, 0.025^2)$. All 7,140 unordered pairs are evaluated. The barycenter LOT baseline uses six free-support iterations. The oracle bank consists of twelve generating prototypes; FPS uses twelve observed measures, starting from the first endpoint of the largest entry of the full squared-OT matrix, with row-major tie breaking. This FPS initialization uses the full target distance matrix.

The temperature is fixed by $\eta = 1$ in Equation (12), giving $\tau = 0.0284383$ for the oracle bank and 0.0483831 for FPS. Table 2 also reports fixed-temperature sensitivities, specified in Appendix D.3. Radial completion changes this assignment-plan example only at machine precision. Its improvement therefore isolates endpoint weighting, not recovery of split mass.

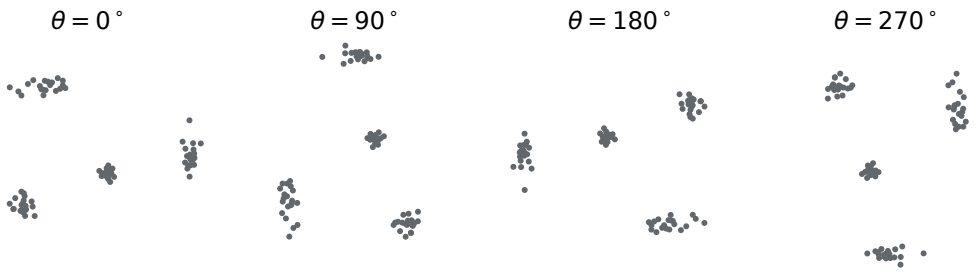

Figure 3: Four generating prototypes at equally spaced angles along the rotating-shape family, using 80-atom supports and common coordinate limits.

D.2 Analytic Crescent Distributions

Let U have density $3(1-u^2)/4$ on $[-1,1]$ and let Z be an independent standard Gaussian. The unrotated law is

$$(X,Y) = \{2U,\ b(U^2 - 1/5) + tZ\}.$$

We use $b \in \{-2.4, -1.2, 0, 1.2, 2.4\}$ and $t \in \{0.08, 0.14, 0.22, 0.32, 0.44\}$. Its mean is zero and its covariance is $\mathrm{diag}\{4/5, 8b^2/175 + t^2\}$. Consequently opposite nonzero bends have identical first and second moments but different laws. For $|x| < 2$, its density is

$$f_{b,t}(x,y) = \frac{3}{8}(1 - x^2/4)\frac{1}{t}\varphi\left(\frac{y - b(x^2/4 - 1/5)}{t}\right),$$

where φ is the standard Gaussian density. The rotated experiment applies physical rotations $0°, 45°, 90°$ to the 25 laws, producing 75 distributions. The four densities in Figure 1 use $(b,t,\theta) = (2.4, 0.22, 0°)$ as the base, changing only b to -2.4, t to 0.44, or θ to $90°$. The display is peak-normalized within each panel; transport uses normalized cell masses, not grayscale values. Some flat-law symmetries can yield equivalent shapes.

We integrate density over a 41×81 grid on $[-2,2] \times [-4.5, 4.5]$ using eight-point Gauss–Legendre quadrature and conditional Gaussian probabilities. Negligible mass is pruned under a 10^{-8} budget and the finite cell masses are normalized. The transport calculations use these weighted cell-center measures. Rotations move the centers and preserve their masses; raster interpolation is not used. The target is exact OT between these deterministically quantized measures.

The LOT reference is the squared-W_2 medoid, and FPS begins at it. The primary budget is $K = 8$ with $\eta = 1$. Budgets 16 and 24 are prespecified sensitivities. The three banks are prefixes of the same FPS sequence, so the target distributions and LOT baseline remain fixed while anchor coverage increases. The covering radius is 0.490989, 0.277185 and 0.201120 for $K = 8, 16, 24$, respectively, with temperatures 0.109716, 0.027118 and 0.012566 set by Equation (12). Figure 1 shows all three complete-method recovery plots, and Figure 4 summarizes the stress and rank-correlation trends alongside the ablation. Appendix D.3 separates the local LOT and residual radial contributions at each budget.

Visualization. The panels in Figure 2 use a UMAP-style fuzzy-graph construction (McInnes et al., 2018) with twelve neighbors, minimum distance 0.05, spread one and seed 20. For fuzzy-graph affinities p_{ij}, our NumPy implementation multiplies the repulsive part of the cross-entropy by $\beta = \sum_{i<j} p_{ij} / \sum_{i<j}(1 - p_{ij})$. It minimizes the resulting full-pair objective with L-BFGS-B, capped at 1,000 iterations, from classical MDS coordinates normalized to unit standard deviation plus Gaussian jitter of standard deviation 10^{-3}. Each panel has independently fitted coordinates, orientation and scale. The plots provide qualitative views of cyclic structure.

D.3 Radial Completion Ablation

We isolate radial completion by holding the target distributions, reference transports, anchor banks and temperatures fixed. Each setting compares complete MALOT with its no-radial ablation and a

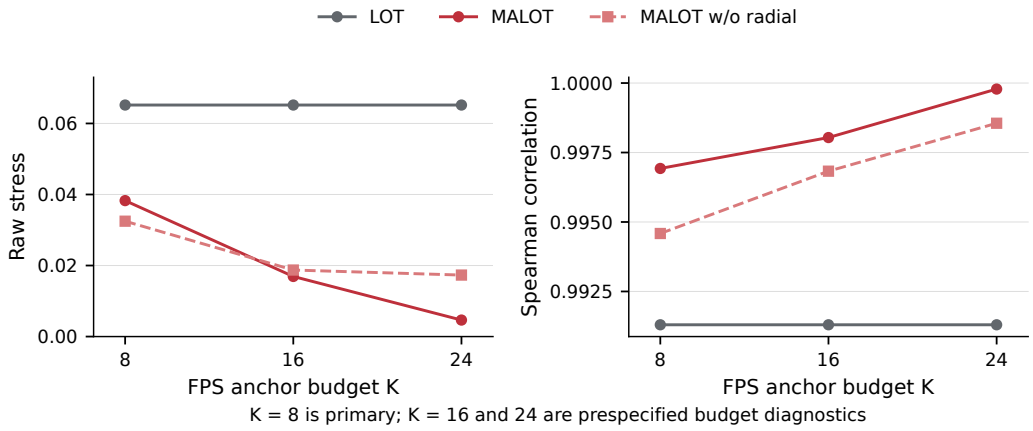

Figure 4: Prespecified crescent anchor-budget sensitivity. The complete method and its no-radial ablation share reference banks and temperatures at each budget.

residual-only control under identical endpoint weights $w_{k,ij}$ from Equation (11). Using the conditional variance V_{ki} in Equation (22), define

$$(D_{ij}^{\mathrm{loc}})^2 = \sum_k w_{k,ij} \|\boldsymbol{\phi}_{ki} - \boldsymbol{\phi}_{kj}\|^2,$$
$$(D_{ij}^{\mathrm{res}})^2 = \sum_k w_{k,ij} (\sqrt{V_{ki}} - \sqrt{V_{kj}})^2.$$

The three comparisons are $D_{ij} = \sqrt{(D_{ij}^{\mathrm{loc}})^2 + (D_{ij}^{\mathrm{res}})^2}$, D_{ij}^{loc} and D_{ij}^{res}, respectively. Thus completion adds a nonnegative squared residual-radius difference before taking the square root. For a separate anchor-radius control, we replace $\sqrt{V_{ki}}$ by $\sqrt{c_{ki}} = W_2(\rho_k, \mu_i)$ in the residual-only formula. This control retains distances to anchors but discards the LOT coordinates. The reverse triangle inequality bounds it above by $W_2(\mu_i, \mu_j)$.

Table 2 reports raw stress for ten settings against exact OT of the corresponding finite measures. The crescent LOT baseline uses the medoid, and the multi-cluster baseline uses a barycenter computed with six iterations. Concat divides the concatenated LOT coordinates by $\sqrt{K}$ and uses the same bank as MALOT. The unrotated crescent experiments use 25 laws on the 41×81 and 61×121 grids, including a fine-grid evaluation at the coarse-grid temperature. Their coverage temperatures are 0.0225249 and 0.0219667, respectively. The rotated experiments use 75 laws and nested banks with $K = 8, 16, 24$. The multi-cluster experiments use 120 measures with twelve oracle or FPS anchors, at the coverage temperatures and at fixed temperatures 0.0680219 and 0.0760779, respectively, for sensitivity analysis. These fixed temperatures are 1.6 times the 80th percentile of the nearest-anchor squared W_2 costs. The oracle bank uses generating prototypes, whereas FPS selects observed distributions.

For the rotated crescents at the primary budget $K = 8$, completion increases raw stress from 0.032481 to 0.038274, a 17.83% increase. At $K = 16$ and 24, it reduces stress by 9.65% and 73.16%. On the unrotated fine grid, stress decreases by 0.14%, while holding the coarse-grid temperature fixed slightly reverses this change.

In the multi-cluster experiment, targets and anchors each have 80 equal-mass atoms, and the optimal plans are assignments. The residual variance is zero up to floating-point precision. Consequently, the with-radial and no-radial stresses coincide to the reported precision, while residual-only distances are numerical noise and have stress approximately one. The matched comparison with concat therefore attributes the improvement to endpoint weighting rather than residual completion.

The crescent plans split unequal masses, so their residual variances are nonzero. Residual-only stress ranges from 0.863 to 0.920, and the anchor-radius control also has higher stress than both MALOT variants in every setting. Radial completion restores the reference-distance identity needed

Table 2: Simulation radial ablation measured by raw stress, with smaller values indicating better distance recovery. The two MALOT variants and both radius-only controls share the anchor bank and temperature within each row. Concat uses the same bank and LOT blocks, whereas LOT uses its stated single reference. Residual and anchor denote the two radius-only controls defined in Appendix D.3. Unrotated crescents use $K = 8$ and multi-cluster settings use $K = 12$. Coverage temperatures are used except in the three fixed-temperature sensitivity settings. The multi-cluster residual-only values round to one because their distances contain only floating-point residuals.

			MALOT		Radius only	
Setting	LOT	Concat	W/o radial	With radial	Residual	Anchor
Unrotated, coarse	0.037679	0.081733	0.015515	0.009672	0.888457	0.166743
Unrotated, fine	0.038437	0.087302	0.009975	0.009961	0.920209	0.156970
Unrotated, fine, fixed τ	0.038437	0.087302	0.010058	0.010147	0.920303	0.158141
Rotated, $K = 8$	0.065185	0.101268	0.032481	0.038274	0.862945	0.354869
Rotated, $K = 16$	0.065185	0.101912	0.018731	0.016923	0.879043	0.224211
Rotated, $K = 24$	0.065185	0.098516	0.017313	0.004647	0.901339	0.163153
Multi-cluster, oracle	0.158929	0.263940	0.032172	0.032172	1.000000	0.226496
Multi-cluster, FPS	0.158929	0.266105	0.046272	0.046272	1.000000	0.251292
Oracle, fixed τ	0.158929	0.263940	0.051195	0.051195	1.000000	0.262535
FPS, fixed τ	0.158929	0.266105	0.064997	0.064997	1.000000	0.288031

for discrete recovery, while the ablation shows how its contribution to distance accuracy varies with mass splitting, anchor coverage and temperature.

D.4 Sydney Flame Measurements

Sydney/Sandia measurements provide joint observations of mixture fraction, temperature, carbon dioxide and carbon monoxide (Barlow et al., 2015; Cutcher et al., 2017). Table 3 includes the distance comparison under acquisition shift. The no-radial MALOT variant has all-pair stress 0.002187, compared with 0.002509 for Fixed LOT. Adding radial completion at the same parameters gives 0.002527.

The Sydney/Sandia inhomogeneous-flame database supplies measured joint scalar records (The University of Sydney, Clean Combustion Research Group, 2026). We form 120 distributions from two acquisition years, five operating conditions, three axial stations $x/D \in \{10, 20, 30\}$ and four radial positions $r/D \in \{0, 0.4, 0.8, 1.2\}$. The operating-condition identifiers are `FA200-5GP-Lr75-45`, `FJ200-5GP-Lr300-59`, `FJ200-5GP-Lr75-57`, `FJ200-5GP-Lr75-80` and `FJ200-5GP-Lr75-103`. We use the centerline and positive radial side only; $D = 7.5$ mm, so the nonzero positions are 3, 6 and 9 mm. Each atom contains mixture fraction, temperature, carbon-dioxide mass fraction and carbon-monoxide mass fraction from one aligned measurement record. The common raw columns are `Fblgr`, `Tray`, `CO2` and `COLIF`. An acquisition line is reconstructed from file order: a coordinate reset or a forward gap greater than twice the median pixel spacing starts a new line. For each line and radial station, its nearest native record is retained within a grid-selection tolerance of 0.06 mm. Lines contributing to multiple radial objects never cross the train/validation/test partition. Finite negative species readings are retained; each object's retained records have uniform mass. The native measures contain 600–3,900 atoms, with $150{,}057$ atoms in total. Training-only feature standardization is common to all measures. Its center and coordinatewise variance weight the 60 training distributions equally and their atoms equally within each distribution; the variance includes both within- and between-distribution variation. The acquisition years also differ in spatial resolution, making the split a genuine acquisition shift.

The 60 distributions from 2013 train the references and preprocessing. The 20 distributions at $x/D = 20$ in 2015 select settings. The 40 distributions at $x/D = 10, 30$ in 2015 form the test set. Within these sets, stations and operating conditions are correlated; pair counts are not independent experimental replicates.

Mass-preserving K-means compression gives at most 256 weighted atoms per target. Weighted K-means uses one initialization and random seed 17; final cluster centers are recomputed as exact

weighted means and their masses are the assigned original masses. Empty clusters are removed. The barycenter and FPS references have 128 atoms. Barycenter optimization allows 1,000 iterations and stops at squared support displacement 10^{-7}. It uses equal training-distribution weights, uniform reference masses and initial support locations drawn from pooled training atoms with seed 17. FPS starts at the training distribution whose concatenated mean and coordinatewise variance is nearest, in squared Euclidean distance, to the average training summary; ties follow input order. It uses distances to the compressed anchors. This matches the compressed-anchor form of Lemma 3.

Validation raw stress selects $K \in \{2, 4, 8\}$ and $\eta \in \{1/16, 1/4, 1, 4\}$. The no-radial variant selects $K = 8, \eta = 4$, and radial completion uses the same reference bank and temperature without further selection. Concat selects $K = 8$ independently on the same validation data. The sliced baseline uses 128 fixed directions and 512 quantile evaluation points for the weighted projections. The sliced baseline approximates a different metric from W_2.

E Supplementary Real-Data Evidence

This appendix gives the main-dataset protocols and supplementary distance-recovery experiments. Training, validation and test sets separate reference construction, parameter selection and evaluation. All MALOT rows use the endpoint rule in Equation (11). The label *w/o radial* refers only to removal of Equation (23)'s variance term.

E.1 Distance-Recovery Visualizations and Main-Data Protocols

Figures 5 and 6 compare the recovered distances with exact W_2 for the ETH-80 and Fingering protocols in Table 1. The plots evaluate test-pair distances before downstream MDS, using the parameter values specified below and no fitted distance scaling. Caulobacter's prediction task, image representation and main-table protocols are described in Appendix E.3. Fixed LOT in the main table denotes one common reference for all targets in a split, obtained from a fixed training sample for Caulobacter and a training Wasserstein barycenter for the other datasets. Concat LOT denotes normalized equal-weight concatenation of FPS reference blocks. Its ETH-80 reference budget can differ from the MALOT budget within a split. Fingering uses the same selected bank and budget for concat and MALOT.

Parameter calibration. For ETH-80 and Fingering, the MALOT reference budgets and temperatures are calibrated by validation raw stress. The reported endpoint-weighted comparisons hold these budgets and temperatures fixed. For Fingering, the MDS dimension $q \in \{4, 8, 16, 32\}$ and SVC parameters $C, \gamma \in \{0.1, 1, 10\}$ are jointly selected on the validation split under this calibration distance, using training-fitted MDS. Both endpoint variants retain the resulting q, C, γ and separately refit MDS and the classifier. For Fingering, Fixed LOT, Concat LOT and Sliced W_2 select their SVC parameters independently in their respective native feature spaces. The supplementary programs recompute the reported fixed-configuration comparisons; they do not repeat the calibration search.

ETH-80 object contours. The eight categories each contain ten physical objects with 41 views per object (Leibe & Schiele, 2003). We center the released binary contour pixels, divide coordinates by their root-mean-square radius and form 64 mass-weighted centers, retaining orientation. At seeds $13, 29, 43, 71, 97$, each category contributes six training, two validation and two test objects, giving $1{,}968/656/656$ images. All views of an object remain in the same set. The stress diagnostic evaluates all $214{,}840$ unordered test-image pairs, and one-nearest-neighbor classification uses test-to-training distances directly. The barycenter has 64 atoms. MALOT uses $K = 16$ throughout, with temperatures $(0.00819418, 0.00755080, 0.00743475, 0.00444428, 0.00799098)$. Concat uses $(16, 16, 16, 4, 16)$ anchors from the corresponding banks. The reference banks use training objects only. Both endpoint variants use the listed temperatures. Distance recovery and classification use the same configuration within each split. Sliced W_2 uses 64 directions and 512 quantiles. Exact Wasserstein targets are computed on the 64-atom representations.

Figure 5 shows the larger positive deviations of Concat LOT from exact W_2. The no-radial MALOT variant has the lowest mean stress among the plotted methods, whereas radial completion raises its mean stress slightly above Fixed LOT.

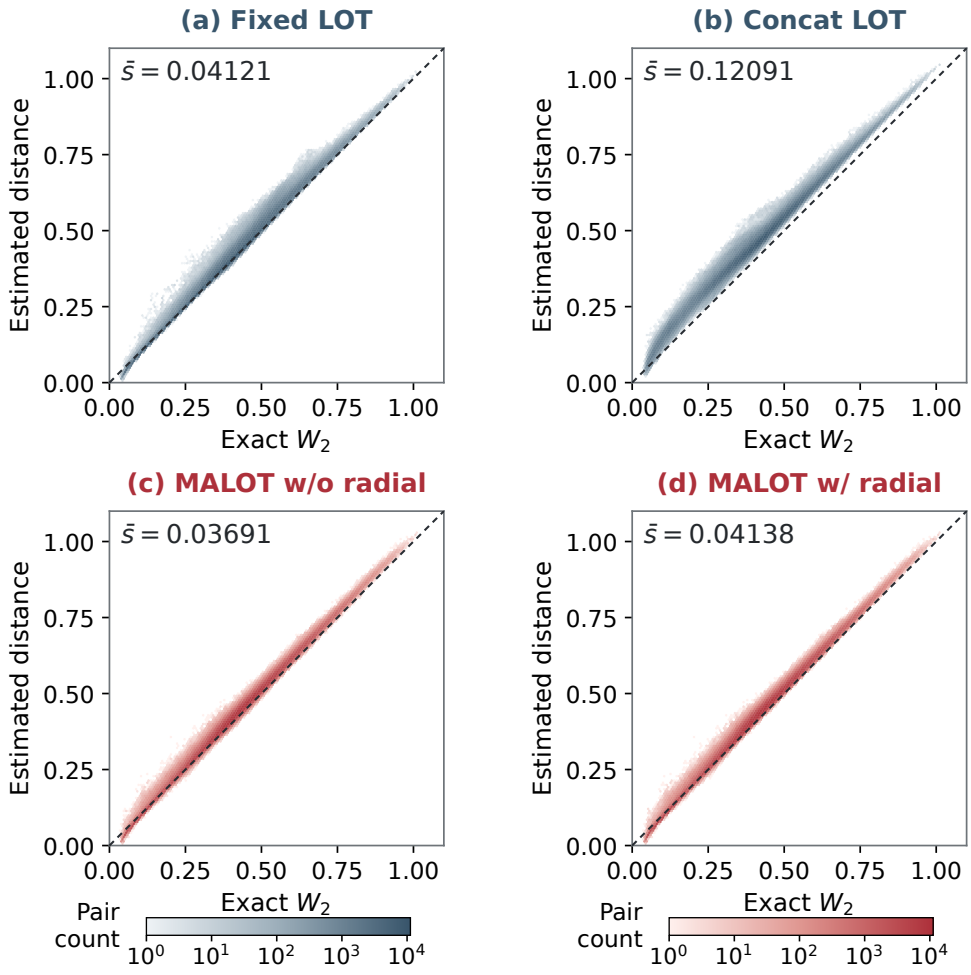

Figure 5: ETH-80 distance recovery over all five object-disjoint splits. Hexagonal bins include all 214,840 test pairs per split, totaling 1,074,200 pair observations, with shared axes, identity lines and logarithmic count normalization. The target is exact W_2 of the 64-atom contour representations. Displayed stress is the unweighted mean across splits, matching Table 1. Fixed LOT uses the training barycenter, and Concat LOT uses the reference prefixes specified above. The plots use no fitted distance scaling, and test pairs can recur across splits.

Viscous fingering. The released segmented experimental images record injected-fluid occupancy in a Hele-Shaw cell (Keable et al., 2022). We align the inlet direction once per run and aggregate pixel mass onto a fixed 64×64 unit-square grid, retaining 330–3,912 positive cells per frame. Each frame is normalized to unit mass, so it represents spatial occupancy rather than local concentration or total injected volume. The 194 retained frames belong to nine experimental runs. Labels indicate whether the independently measured volumetric breakthrough has occurred, excluding frames within the published measurement-error band. Seeds $13, 29, 43$ split the runs into five training, two validation and two test runs, giving $111/38/45$, $107/43/44$ and $106/44/44$ frames. Each split evaluates 32 within-run and 32 cross-run test pairs against exact grid-based OT. The barycenter and compressed FPS references use 128 atoms. Sliced W_2 uses 128 directions and 512 quantiles.

For the geometry comparison, we fix the reference budgets and temperatures at the values below. The budgets are $(4, 8, 2)$ and temperatures are $(0.0117804, 0.0152654, 0.0382923)$. Both MALOT variants use four-dimensional training-fitted MDS coordinates and a class-balanced RBF SVC. Fixed LOT, Concat LOT and Sliced W_2 use their native feature coordinates for classification. All classifier

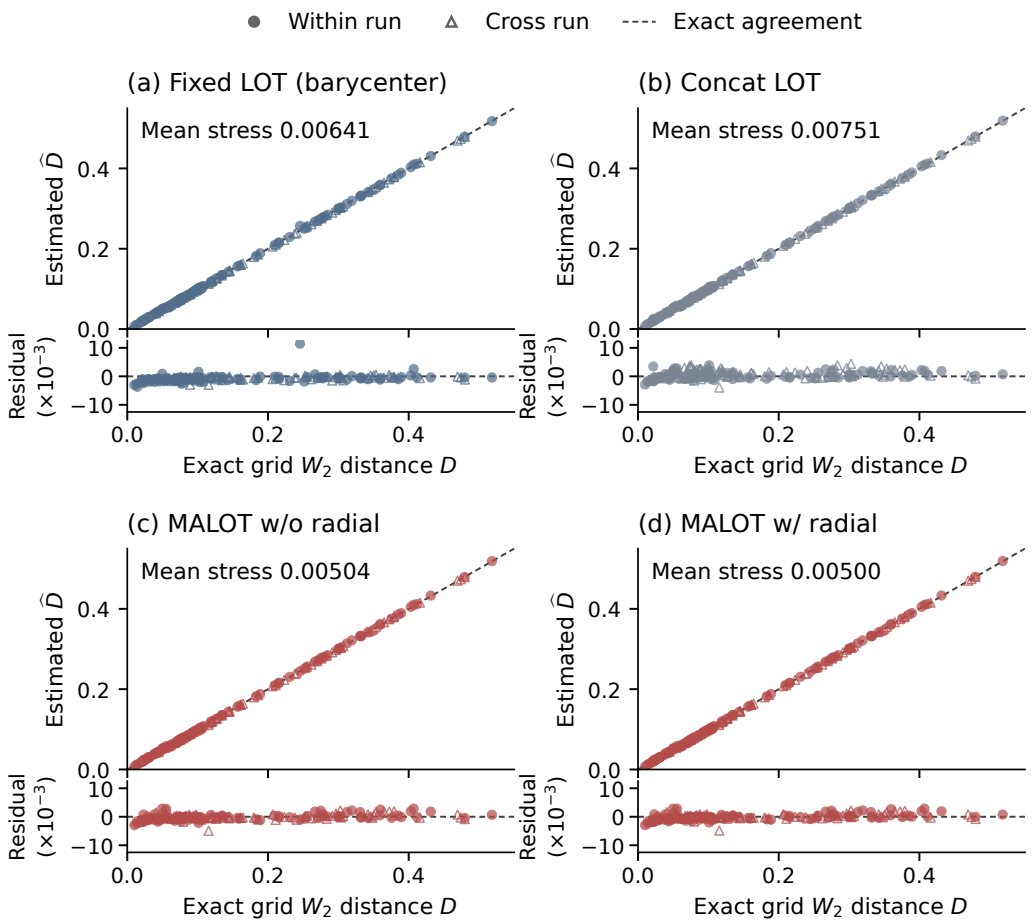

Figure 6: Fingering distance recovery on the 64×64 occupancy-grid measures. All 192 test-pair observations are shown across three splits, each containing 32 within-run and 32 cross-run pairs. The scatter panels share axes and dashed identity lines. Each lower strip shows the signed residual $\widehat{D} - D$ in units of 10^{-3} on a common vertical scale. Annotations give mean raw stress across splits, matching Table 1. Fixed LOT uses a 128-atom training barycenter. All methods use the same test pairs without fitted distance scaling.

features are centered by their training mean and divided by one training root-mean-square scale. For both MALOT variants, the fixed (C, γ) pairs are $(10, 0.1)$, $(1, 10)$ and $(10, 1)$. We report ordinary ACC on the test set.

Figure 6 shows the same held-out pairs for all four procedures. Because the estimated distances are close to exact grid W_2, common-scale residual panels expose differences that are difficult to see against the identity line. Both MALOT variants lower mean stress relative to Fixed LOT and Concat LOT, with a small further reduction from radial completion.

E.2 Beam-Profile Geometry

FLASHForward. Forty observed beam-profile shots (Lindstrøm et al., 2024a;b) are split into 20 training, ten validation and ten test shots using fixed temporal blocks from two diagnostic runs. Table 3 includes their distance-recovery results. The representation uses mass-preserving 4×4 pixel aggregation and 128-atom references. All 45 test pairs are evaluated. At the shared $K = 2$

reference bank, complete MALOT has stress 0.021190, compared with 0.022574 without radial completion, 0.033700 for Fixed LOT and 0.026987 for Concat LOT.

Reference and parameter selection. Reference selection and evaluation of the covering radius use training objects only. Validation raw stress selects the no-radial MALOT parameters from $K \in \{2, 4, 8\}$ and $\eta \in \{1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8\}$ in Equation (12), retaining the first minimizer in increasing grid order. The selected values are $(K, \eta) = (2, 1/4)$, and radial completion uses the same reference bank and parameters. Concat LOT selects $K = 2$ from the same budget grid by validation raw stress. Runs 20333 and 20334 each contribute 20 shots selected by integer-valued `linspace(0,49,20)`. Within each run, selected-shot positions 1–10, 11–15 and 16–20 form training, validation and test sets. Validation and test each evaluate all 45 pairs. The targets contain 1,990–2,373 atoms, while the training barycenter and FPS references contain 128 atoms. The sliced baseline uses 128 fixed random unit directions and 512 weighted quantile points per direction, without fitted scaling. Its random stream starts at seed $13 + 8103 = 8116$, draws the two 45-pair ordering permutations for validation and test, and then draws the projection directions.

Shared FPS reference construction. The FLASHForward and Caulobacter protocols initialize FPS at the training measure nearest the arithmetic mean of its weighted-mean/diagonal-variance summaries in unscaled Euclidean distance. Subsequent anchors maximize the minimum transport cost to the selected references, with the first index breaking ties. Each bank has eight references, compressed by mass-weighted K-means with one initialization and random state 13 before updating coverage. Caulobacter's task and classifier settings are given in Appendix E.3.

E.3 Caulobacter: Predicting Imminent Cell Division

Biological question. The Caulobacter experiment asks whether the spatial distribution of membrane fluorescence in a single image can distinguish a cell approaching division from one at an earlier stage. The released structured-illumination microscopy sequences follow individual *C. crescentus* cells through growth and constriction (Lambert et al., 2018a;b;c). We use the MTS-mCherry inner-membrane channel, which records the cell's spatial outline and developing constriction. Figure 7 illustrates how these distributions change along two observed trajectories. The prediction target is whether division will occur within the next 30 minutes. Both classes contain pre-division images; the classifier receives only the current frame, without the subsequent sequence, acquisition time or cell identifier.

From an image to a probability measure. The authors' processing reconstructs the microscopy images, corrects drift, subtracts background, clips negative intensities and crops the tracked cell. We retain every positive pixel of the released crop at its native $0.03\,\mu\mathrm{m}$ spacing, normalize intensities to unit mass and subtract the intensity-weighted spatial centroid. Thus, a frame is a two-dimensional distribution whose mass encodes cell shape and the spatial arrangement of membrane fluorescence. The representation preserves the native grid, physical size and orientation while removing translation and total brightness. Wasserstein distance quantifies changes in the normalized spatial distribution of membrane fluorescence.

Labels and sampling. The retained cohort contains 301 tracked cells and 1,806 pre-division frames from three acquisition dates, with six sampled frames per cell. The source's semi-automated tracking records a division flag at five-minute sampling intervals. Let $(t_L, t_R]$ bracket the transition from the last undivided frame to the first division flag. A frame at time t is positive if $t_R - t \leq 30$ minutes and negative if $t_L - t \geq 30$ minutes. Frames whose interval crosses the decision boundary are excluded. Only trajectories with an observable division endpoint are retained.

Evaluation used in the main table. Let A, B, C denote acquisition dates 2016-06-22, 07-13 and 09-26, with 630, 600 and 576 retained frames. The training/validation/test rotations are (A, B, C), (C, A, B) and (B, C, A), so no cell contributes to more than one subset within a split. Table 1 uses 64 class-stratified training frames per split for the geometry diagnostic, with sampling seeds 17, 18 and 19 and 32 fixed native test pairs per date. Its ACC column instead uses all 630, 576 and 600 available training frames in the respective rotations. Validation and test sets remain the complete held-out dates in both protocols.

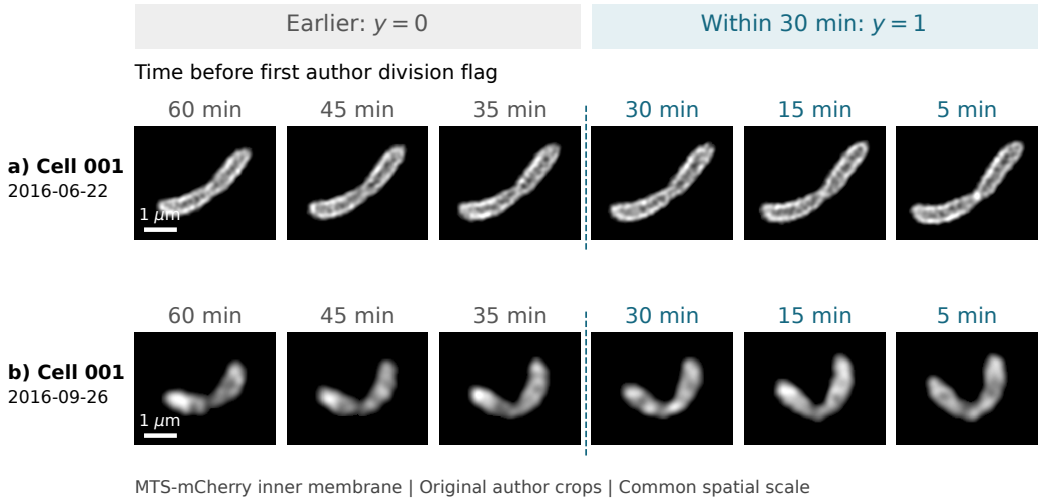

Figure 7: Membrane fluorescence along two Caulobacter trajectories from the evaluated cohort. Columns show time relative to the first author-recorded division flag; the dashed boundary separates the earlier and within-30-minute labels. The earliest and latest retained acquisition dates each contribute their lowest-numbered cell, selected without reference to classifier predictions. Images are author-processed native-grid membrane crops at a common physical scale, with fixed contrast within each row. Scale bars, $1\,\mu$m. The time sequence illustrates the retrospective labeling rule, whereas each prediction uses one image. The panels follow mother cells to the recorded division endpoint.

Reference and classifier settings. Targets retain 1,096–3,716 positive pixels, while references contain 128 atoms. Fixed LOT uses one fixed training sample, and Concat LOT uses equal-weight FPS reference blocks. The FPS bank contains eight training references, initialized and compressed as specified in Appendix E.2. Sliced W_2 uses 128 directions and 512 quantiles. MALOT searches $K \in \{2, 4, 8\}$, $\eta \in \{1/16, 1/4, 1, 4\}$ and requested positive MDS dimensions $q \in \{4, 8, 16, 32\}$. Training-fitted inductive MDS supplies its downstream coordinates, while the other methods use their native representations. The class-balanced RBF SVC search uses $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.1, 1, 10\}$. Settings are selected on the validation split, retaining the first configuration within 10^{-12} ties. The reported test metric is ordinary accuracy. At $n = 64$, the selected (K, η, q, C, γ) tuples are $(2, 1/4, 32, 0.1, 1)$, $(4, 1/16, 4, 100, 0.1)$ and $(2, 4, 4, 100, 0.1)$. For full training they are $(2, 4, 8, 10, 0.1)$, $(8, 1, 4, 100, 0.1)$ and $(2, 1/16, 8, 100, 0.1)$. The radial variant retains these parameters and the reference bank, then refits training MDS and the classifier after adding radial completion.

E.4 Additional Distance-Recovery Results

Table 3 compares the same five methods as Table 1 on four supplementary datasets. Fixed LOT uses a training Wasserstein barycenter, Concat LOT uses the equal-weight concatenation of FPS reference blocks, and sliced W_2 uses fixed projection directions without fitted scaling. The two MALOT variants share their reference bank, number of references and temperature, with radial completion added under the no-radial parameters. Sydney and FLASHForward use the protocols in Appendices D.4 and E.2. For SeaFlow and Stanford96, the MALOT settings are specified below, and Concat selects the number of references on validation data alone. Means weight splits or test subjects equally.

SeaFlow optical distributions. SeaFlow MGL1704 contains 296 hourly optical distributions from one cruise, represented by 128 weighted centers in three optical coordinates.[1] Five splits keep whole sampling days together and impose a six-hour buffer between subsets. The target is exact W_2 for

[1]Released cytograms and source code are available at `https://github.com/sangwon-hyun/flowmix`.

Table 3: Distance recovery with the same five methods as Table 1, measured by raw stress without fitted scaling. Sydney and FLASHForward each use one held-out split, SeaFlow averages five splits, and Stanford96 averages six test subjects. Fixed LOT uses one training barycenter per split. The two MALOT variants share their reference bank, K and temperature within each split. Targets are the stated finite representations for the first three datasets and analytic Gaussian W_2 for Stanford96. Both MALOT rows are shaded and each column's lowest stress is bold, including ties.

Method	Sydney	FLASHForward 4×4	SeaFlow 128 atoms	Stanford96 High FA
Fixed LOT	0.002509	0.033700	0.014902	0.010869
Concat LOT	0.014085	0.026987	0.015204	0.053707
Sliced W_2	0.505622	0.386205	0.480311	0.603845
MALOT w/o radial	**0.002187**	0.022574	**0.011271**	**0.008660**
MALOT w/ radial	0.002527	**0.021190**	0.011283	**0.008660**

these compressed distributions. For seeds $13, 29, 47, 71, 97$, MALOT uses $K = (16, 16, 8, 16, 16)$ and temperature multipliers $\eta = (1.6, 6.4, 0.05, 1.6, 1.6)$. Its temperature is η times the training 80th percentile of nearest-reference squared W_2 costs, with a 10^{-12} lower bound. For seeds $13, 29, 47, 71, 97$, Concat selects $K = (2, 16, 4, 2, 8)$ from $\{2, 4, 8, 16\}$ using validation raw stress on prefixes of the same training reference bank. The sliced baseline uses 64 fixed random directions with seed 20260907 and 512 weighted midpoint quantiles per direction. Both MALOT variants lower stress relative to Fixed LOT and Concat LOT in all five splits. The mean stress is slightly lower without radial completion.

Diffusion tensors. Stanford96 contains diffusion MRI scans from six subjects, with each voxel represented by a three-dimensional zero-mean Gaussian whose covariance is a fitted diffusion tensor.[2] Six rotations allocate three subjects to training, two to validation and one to testing. The high-anisotropy analysis fixes fractional anisotropy at least 0.5, uses $K = 64$ and temperature multiplier $\eta = 0.1$, and evaluates 10,000 sampled voxel pairs per test subject against analytic Gaussian W_2. Concat selects K from $\{2, 4, 8, 16, 32, 64\}$ using mean raw stress over the two validation subjects, with 10,000 fixed pairs per subject. In test-subject order SUB1–SUB6, its selected budgets are $(64, 2, 64, 64, 64, 64)$. Here, the temperature is η times the training 80th percentile of nearest-reference squared W_2 costs, with a 10^{-12} lower bound. Both endpoint variants lower stress relative to Fixed LOT and Concat LOT in all six subjects. The radial residual is zero for these deterministic Gaussian transport maps, so the two variants coincide. The sliced baseline uses the root mean squared one-dimensional Gaussian W_2 over 256 fixed random unit directions, with seed 2026 and no fitted scale.

[2]The six-subject 96-direction release is available at `https://purl.stanford.edu/rt034xr8593`.