CIRCULAR OPTIMAL PARTIAL TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

When two signals share only part of their content, optimal partial transport (OPT) can match that common content without assigning every observation a counterpart. For periodic signals, the samples lie on a circle, and their cyclic ordering requires more than a direct use of a line solver. PAWC computes the complete circular partial-transport profile for the W_1 cost on distinct support. Building on the PAWL and PAWC profile algorithms, we introduce COPT, an exact solver for circular OPT between empirical measures with a penalty on unmatched mass, an arbitrary convex radial cost, and possibly repeated support points. We extend PAWC's local augmentation and simultaneous-cut arguments to convex costs, using a rank refinement to handle repeated support and flat cost regions. We prove that COPT returns the complete profile of optimal costs over all transported cardinalities in $O(N \log N + K^2)$ time and $O(N)$ memory, where N is the number of atoms and K the largest transported cardinality, and in $O(pN \log^2 N)$ time for integer power costs. We further extend COPT to spherical measures through great-circle slicing and to repeated comparisons through a linear embedding, and we establish their metric properties. Through numerical experiments on correspondence recovery, partial shape matching, foreground retrieval, and directional data, we demonstrate the computational benefits of COPT and its robustness to missing signal and background clutter.

1 INTRODUCTION

Optimal transport (OT) provides a geometrically meaningful way of comparing probability measures and has become a standard tool in machine learning, computer vision, and signal processing (Peyré & Cuturi, 2019). Its balanced form must assign all the mass of one measure to the other. When observations are missing, occluded, or contaminated, this constraint forces points with no counterpart into the plan. Optimal partial transport (OPT) instead charges a penalty for the mass left unmatched (Caffarelli & McCann, 2010; Figalli, 2010; Figalli & Gigli, 2010), which has made it attractive for point-cloud registration (Bai et al., 2025), minibatch transport (Nguyen et al., 2022), and outlier-robust comparisons (Mukherjee et al., 2021; Nietert et al., 2022; Raghvendra et al., 2024). In this paper, we are interested in OPT between measures supported on the unit circle. Hue wraps around the color wheel, musical pitch repeats by octave, the normal directions of a planar contour form a circular measure (Horn, 1984), and phases and times of day are defined modulo one cycle. Moreover, sliced transport on the sphere projects spherical measures onto great circles (Bonet et al., 2023; Liu et al., 2025), so circular partial transport is also the building block of partial transport for spherical data (Kolouri, 2026). On the real line, OPT with a convex radial cost between empirical measures with unit atoms admits a monotone optimal plan after sorting (Appendix A.1). PAWL (Chapel & Tavenard, 2025) computes the whole profile of optimal partial W_1 costs, indexed by the transported mass, in $O(N \log N)$ time for N atoms. Kolouri (2026) formulated this profile problem on the circle and solved it for W_1 with PAWC, extending PAWL's profile construction. Its elegant free-gap argument identifies a single cut that is optimal for every transported cardinality, preserving the $O(N \log N)$ bound for distinct support and the geodesic W_1 cost.

Circular data have no distinguished origin. Balanced circular OT chooses a cutting point, unwraps the circle into an interval, and solves the line problem (Delon et al., 2010; Rabin et al., 2011). For partial transport with general convex costs, the cut also affects which atoms are transported. The direct exact method reruns a line solver at every support gap, which multiplies its cost by the number of atoms. Repeated support points add a second difficulty. A cut through a source copies

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.

and b target copies admits $(a+1)(b+1)$ ways of splitting the copies between the two ends of the interval, and cuts at support gaps alone can miss the optimum for costs such as $h(t) = t^2$.

Building directly on the profile approach of PAWL and its circular extension PAWC (Chapel & Tavenard, 2025; Kolouri, 2026), we develop Circular Optimal Partial Transport (COPT) for general convex radial costs h with $h(0) = 0$ and repeated support points. Starting from the empty plan, it activates one source and one target at a time. PAWC showed for W_1 on distinct support that an optimal plan at the next cardinality is obtained by activating the endpoints of a single circular cell between neighboring inactive atoms and rematching that cell in sorted order, so one run computes the whole profile; we show that this holds for every convex radial cost. A secondary rank cost resolves ties between coincident atoms and between arcs of equal cost without perturbing the locations, and a rank gap that survives every activation determines one cut, possibly through a repeated support point, at which sorting reads off the optimal plans of all cardinalities. For integer power costs, we adapt the balanced-interval data structure of Angrick et al. (2026) to the doubled circular order.

Contributions. Our contributions include 1) an extension of PAWC's profile algorithm to general convex radial costs and repeated support, with complexity guarantees (Theorem 1); 2) the rank-refined augmentation lemma and an extension of PAWC's simultaneous-cut theorem (Lemma 1 and Proposition 1); 3) applications of the profile to correspondence recovery, partial shape matching, and foreground-palette retrieval (Section 5); and 4) Spherical Sliced COPT and a Linear COPT embedding, following LOPT, LCOT, and LSSOT (Bai et al., 2023a; Díaz Martín et al., 2024; Liu et al., 2025), with metric guarantees and experiments on spherical and earthquake data (Section 6).

2 Related Work

Bai et al. (2023b) study empirical OPT on the line and give a sorted-line solver for a fixed penalty. One-dimensional partial transport also underlies sliced methods (Bonneel & Coeurjolly, 2019). PAWL (Chapel & Tavenard, 2025) computes the partial transport profile on the line in $O(N \log N)$ time for W_1 and quadratic time for squared distance. Angrick et al. (2026) reduced the $\Theta(n^2)$ cost of the same approach for $p > 1$ to $O(pn \log^2 n)$ for fixed integer p through an FFT-based interval oracle. Phatak et al. (2023) compute all optimal partial transports for general discrete measures, while Chapel et al. (2021) and Séjourné et al. (2022) solve unbalanced OT through regularization paths and Frank–Wolfe iterations, and Bonet et al. (2024) develop sliced unbalanced transport. We use partial linear programs (Chapel et al., 2020; Flamary et al., 2021) and the one-dimensional unbalanced solver of Gouvine (2023) as exact references.

On the circle, the balanced problem is classical (Werman et al., 1985; 1986; Cabrelli & Molter, 1995; 1998; Hundrieser et al., 2022), and Delon et al. (2010) and Rabin et al. (2011) give the cut and shift formulas for Monge costs. Thresholded EMD provides robust cyclic histogram comparisons (Pele & Werman, 2008; 2009). The PAWC method of Kolouri (2026) extends PAWL's profile construction to the circle. For $h(t) = t$ and distinct support, COPT recovers PAWC's partial-transport profile. We retain its circular cells and free-gap invariant and establish local augmentation for convex radial costs and repeated atoms through a rank-refined version of the endpoint exchange of Angrick et al. (2026) (Appendix C). Circular OPT also reduces to a cyclic alignment, which the separator algorithm of Maes (1990) solves in $O(nm \log(1+\min(n,m)))$ time (Appendix A.4), while Gregor & Thomason (1993) develop a guided search for cyclic alignment. Our algorithm instead grows the active sets by local augmentations, as PAWL and PAWC do, using the Monge structure of the unwrapped line problems (Karp & Li, 1975; Aggarwal et al., 1992). Our spherical and linear extensions build on LOPT (Bai et al., 2023a), LCOT (Díaz Martín et al., 2024), and LSSOT (Liu et al., 2025).

3 Background

We identify the unit circle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ with $[0, 1)$ under the identification $0 = 1$, and we equip it with the geodesic distance

$$\rho_{\mathbb{S}^1}(x, y) = \min_{k \in \mathbb{Z}} |x - y + k|, \qquad x, y \in [0, 1). \tag{1}$$

The transport cost is $c_{\mathbb{S}^1}(x, y) = h(\rho_{\mathbb{S}^1}(x, y))$, where $h : [0, 1] \to \mathbb{R}_+$ is finite and convex with $h(0) = 0$. These assumptions imply that h is nondecreasing; an upward jump at 1 is allowed. We

unwrap the circle at a cutting point z through

$$u_z(t) = (t - z) \bmod 1, \tag{2}$$

and use $h(|\tilde{x} - \tilde{y}|)$ on the resulting sorted lists.

For empirical measures whose atoms have unit mass, $\mu = \sum_{i=1}^{n} \delta_{x_i}$ and $\nu = \sum_{j=1}^{m} \delta_{y_j}$, the symmetric OPT problem with cost c and penalty $\lambda > 0$ is (Caffarelli & McCann, 2010; Figalli, 2010)

$$\min_{\gamma \in \mathbb{R}_+^{n \times m}} \sum_{i,j} c(x_i, y_j)\gamma_{ij} + \lambda \left(n + m - 2 \sum_{i,j} \gamma_{ij} \right) \quad \text{s.t.} \quad \gamma \mathbf{1}_m \le \mathbf{1}_n, \quad \gamma^T \mathbf{1}_n \le \mathbf{1}_m, \tag{3}$$

in which every unit of mass that is not transported is charged λ, and an optimizer can be chosen as a one-to-one partial plan (Bai et al., 2023b). On the line, Monge uncrossing produces a monotone optimum after sorting, and an alignment dynamic program computes its value $\mathrm{OPT}_\lambda^{\mathbb{R}}(X, Y)$ in $O(nm)$ time (Appendix A.1). Our algorithm instead evaluates balanced intervals, whose sorted matchings are optimal.

For balanced circular transport on distinct uniformly weighted supports, one suitable cut reduces the problem to sorted line matching (Rabin et al., 2011; Delon et al., 2010; Díaz Martín et al., 2024). Repeated atoms may require a cut that splits coincident copies between the two interval endpoints. Appendix A.2 recalls the quantile formula and its relation to such splits. The transported part of an optimal partial plan solves balanced circular transport on its active atoms; Proposition 1 extends the common-cut representation to every transported cardinality.

4 Method

We follow the profile construction of PAWL and PAWC (Chapel & Tavenard, 2025; Kolouri, 2026) and extend the augmentation and cut arguments to convex radial costs and repeated support. For two lists of unit atoms $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_m)$, write $N = n+m$ and $K = \min(n, m)$. Their circular OPT value is

$$\mathrm{COPT}_\lambda^{\mathbb{S}^1}(X, Y) = \min_M \sum_{(i,j) \in M} h(\rho_{\mathbb{S}^1}(x_i, y_j)) + \lambda(n + m - 2|M|), \tag{4}$$

where M ranges over partial matchings, with each source and target index used at most once. The one-to-one representation in Section 3 makes this equivalent to the empirical OPT linear program.

4.1 Cut Representation

Let us first define the minimum transport cost at each cardinality,

$$C_k^\circ = \min_{|M|=k} \sum_{(i,j) \in M} h(\rho_{\mathbb{S}^1}(x_i, y_j)), \qquad k = 0, \ldots, K. \tag{5}$$

Following Kolouri (2026), who introduced this circular profile for $h(t) = t$, we call $k \mapsto C_k^\circ$ the partial transport profile. Grouping the plans in equation 4 by cardinality, we obtain

$$\mathrm{COPT}_\lambda^{\mathbb{S}^1}(X, Y) = \min_{0 \le k \le K} \{C_k^\circ + \lambda(N - 2k)\}, \tag{6}$$

so one profile determines the solution for every penalty. We construct a common cut at which sorted line matchings attain every C_k°.

Let $\mathcal{G}$ be the support gaps, the connected components of $\mathbb{S}^1 \setminus \mathrm{supp}(X, Y)$, and choose one cut $z_G \in G$ from each. Moving a cut within a gap translates the unwrapped lists without changing their line cost. At a support point v with multiplicities a_v, b_v, a virtual cut (v, r, s) instead places r source copies and s target copies at endpoint 1, the remaining copies at 0, and all other atoms at $u_v(t)$. These cuts form the finite family

$$\mathcal{C}(X, Y) = \{z_G : G \in \mathcal{G}\} \cup \{(v, r, s) : 0 \le r \le a_v,\ 0 \le s \le b_v\}, \tag{7}$$

where the union over virtual cuts includes every support point v and integer split. Let X^σ, Y^σ be the sorted unwrapped lists.

Proposition 1 (Cut representation; extending Kolouri, 2026, Prop. 1 and Thm. 4)**.** *There is a cut $\sigma^\star \in \mathcal{C}(X,Y)$ at which, for every $k = 0, \dots, K$, the active set constructed by Lemma 1 has a sorted line matching of cost C_k°. Consequently,*

$$\mathrm{COPT}_\lambda^{\mathbb{S}^1}(X,Y) = \min_{\sigma \in \mathcal{C}(X,Y)} \mathrm{OPT}_\lambda^{\mathbb{R}}(X^\sigma, Y^\sigma) = \mathrm{OPT}_\lambda^{\mathbb{R}}(X^{\sigma^\star}, Y^{\sigma^\star}) \quad \textit{for every } \lambda > 0. \tag{8}$$

We prove this proposition in Appendix A.4.

4.2 Circular Augmentation

We start from the empty matching and increase the transported cardinality by one at each step. Previously matched atoms remain matched, but their partners may change. For instance, with $X = (0, \frac{1}{4})$, $Y = (\frac{1}{4}, \frac{1}{2})$, and $h(t) = t^2$, the optimal one-pair matching has zero cost, but keeping this pair at full mass costs $\frac{1}{4}$, whereas rematching the two pairs costs $\frac{1}{8}$.

We sort the union support in cyclic order, and we call a matched atom active and an unmatched atom inactive. Two consecutive inactive atoms u, v delimit a directed cell $I = [u, v]_{\circlearrowleft}$ whose open interior I° contains only active atoms. If a single inactive atom remains, its cell makes one full turn back to that atom. This is the cell representation of Kolouri (2026). When the endpoints of a cell have opposite types, activating them adds one source and one target. For an interval J with as many sources as targets, we let

$$W_h(J) = \sum_{j=1}^{|J \cap X|} h(|x_j^J - y_j^J|), \tag{9}$$

where x_j^J and y_j^J are the sorted source coordinates and the sorted target coordinates of J in its unwrapped order. By Monge uncrossing, this sorted matching is optimal on J. The cost of activating the endpoints of a candidate cell is

$$\Delta(I) = W_h(I) - W_h(I^\circ), \tag{10}$$

which includes the rematching of the interior of the cell and not merely a new edge between its endpoints (Figure 1).

To see why one cell suffices, consider first distinct locations and a strictly increasing h. No edge of a cardinality-constrained optimum passes over an inactive atom along its shortest arc, since replacing the endpoint of the same type by that inactive atom would shorten the edge. Consider now an augmenting path of minimum net cost. If one of its new edges passes over an inactive endpoint of the path, a shorter edge replaces the corresponding alternating prefix or suffix and lowers the net cost, so an optimal augmentation lies inside one current cell, including when only two inactive atoms remain.

For coincident copies, or on flat regions of h, shortening an arc need not strictly reduce its cost, and we refine the cost so that it does. We fix a total order of the copies at each location, with the source copies before the target copies. A directed arc then has an arc length ℓ and a rank length q, the number of consecutive gaps that it crosses, and we assign to each edge the refined cost

$$\widehat{c}(i,j) = \min_{\mathrm{lex}} \big\{ (h(\ell), q),\ (h(1-\ell), N-q) \big\}, \tag{11}$$

which we minimize lexicographically. The first component is the circular transport cost, and the second selects among its minimizers without perturbing the locations. With this refinement, shortening an arc that attains $\widehat{c}$ to a proper subarc strictly lowers the refined cost, and on an unwrapped interval both components satisfy the Monge inequality, so we compare candidate cells through the refined version of equation 10.

Lemma 1 (Circular augmentation; extending Kolouri, 2026, Thms. 1–3)**.** *Let $0 \le k < K$, and suppose that the current cardinality-k matching is optimal for the refined cost equation 11. Then its open current cells are balanced, and closing a candidate cell of minimum refined increment and matching its source and target atoms in sorted order produces an optimal cardinality-$(k+1)$ matching. In particular, the increment of the first component is $C_{k+1}^\circ - C_k^\circ$.*

We prove the lemma in Appendix A.3, where we carry out the endpoint exchange in detail and extend the line argument of Angrick et al. (2026) to the circle with ranked copies.

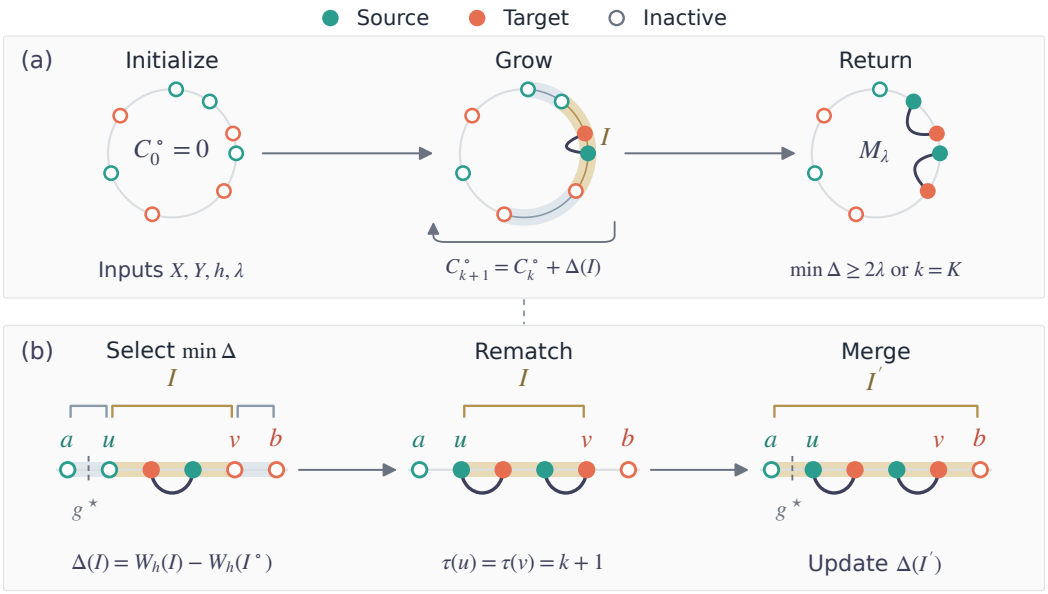


Figure 1: COPT at a glance. (a) Starting from $C_0^\circ = 0$, activate a cell of minimum refined increment until the next increment reaches 2λ or all K pairs are active. (b) Activate u, v, rematch I in sorted order, and record their activation rank. The three adjacent cells merge into I', which inherits a neighboring free gap and receives a new increment. Curves join matched atoms and hollow atoms are inactive.

4.3 Tied Support

Following PAWC's free-gap construction (Kolouri, 2026), each cell retains one original rank gap outside all selected cells. A merged cell inherits a gap from a neighbor; at the full-mass step, the complementary cell retains one. A positive-width gap defines an ordinary cut, while a gap between copies at v defines a virtual split. The source-before-target order only represents $(r, 0)$ and (a_v, s), but these suffice: Lemma 1 establishes optimality over all circular matchings, and the surviving gap represents every constructed optimum.

The virtual splits are necessary. For $X = (0, 0, \frac{1}{4}, \frac{3}{4})$ and $Y = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2}, \frac{3}{4})$ with $h(t) = t^2$, the full circular transport costs $\frac{1}{4}$, whereas every ordinary support-gap cut, which places all the copies at a support point at the same end of the interval, costs at least $\frac{3}{8}$. A surviving gap between the two coincident target copies recovers the smaller value. A sequence of ordinary cuts approaching a support point still places all of its copies at one end, so an intermediate split is not a limit of ordinary cuts, and Appendix A.4 represents it directly.

4.4 Algorithm and Complexity

If one of the two lists is empty, the empty plan with value $\lambda(n + m)$ is optimal and the profile reduces to $C_0^\circ = 0$. Otherwise, we start from the empty matching, keep the inactive atoms in a circular doubly linked list, and keep the candidate increments in a min-heap. After the cell $[u, v]_\circlearrowright$ has been selected, and unless u and v are the last two inactive atoms, the three regions (a, u), $[u, v]$, and (v, b) merge into one cell (Figure 1b). Each region is balanced, so their sorted matchings concatenate, and the cost of the new interior is

$$W_h((a, b)) = W_h((a, u)) + W_h([u, v]) + W_h((v, b)). \tag{12}$$

Only the newly formed cell needs a new key in the heap. During the growth, we record the activation rank τ of each atom, namely the cardinality at which it becomes active, and we keep one free rank gap. To recover a matching of any requested cardinality, we keep the atoms with $\tau \leq k$ and pair the sources and targets in sorted order from the common cut. Algorithm 1 and Figure 5 in Appendix A.5 list and illustrate these steps.

It remains to evaluate the cell costs. The direct implementation evaluates the new cell at step k on at most $k+1$ pairs, so all the queries take $O(K^2)$ time. For $h(t) = t$, the chain prefix sums of PAWL and PAWC (Chapel & Tavenard, 2025; Kolouri, 2026) evaluate every balanced interval, including its rank component, in constant time after a linear preprocessing. For higher integer powers, we use the balanced-interval data structure of Angrick et al. (2026) on the doubled sequence. At each node of a range tree, the binomial expansion of $(y-x)^p$ reduces the costs of all rank shifts to $p+1$ correlations computed by an FFT, and every balanced query combines $O(\log N)$ stored terms.

Theorem 1 (Profile computation; extending Kolouri, 2026, Thm. 5)**.** *Let $h : [0,1] \to \mathbb{R}_+$ be finite and convex with $h(0) = 0$, and assume that cost evaluations, arithmetic operations, and comparisons take constant time. Algorithm 1 computes every C_k° and a compact representation of corresponding optimal matchings in $O(N \log N + K^2)$ time and $O(N)$ space, for repeated locations and unequal cardinalities. For $h(t) = t$, prefix queries reduce the time to $O(N \log N)$ with $O(N)$ space. For $h(t) = t^p$ with a fixed integer $p \geq 2$, balanced-interval preprocessing reduces the time to $O(pN \log^2 N)$ with $O(N \log N)$ space. In all cases, recovering one requested matching takes $O(N)$ time.*

We prove the theorem in Appendix A.5, and Appendix B describes the floating-point comparison rule of the implementation. For a single penalty λ, the growth stops early. Since the increments $C_k^\circ - C_{k-1}^\circ$ are nondecreasing, we stop before activating the first cell whose increment is at least 2λ, and the direct implementation then takes $O(N \log N + k_\lambda^2)$ time for k_λ transported pairs.

5 Experiments

The main computational study uses the quadratic geodesic cost; Appendix B covers further convex costs, tied support, and comparisons with linear programming and direct cut enumeration.

5.1 Computing the Profile

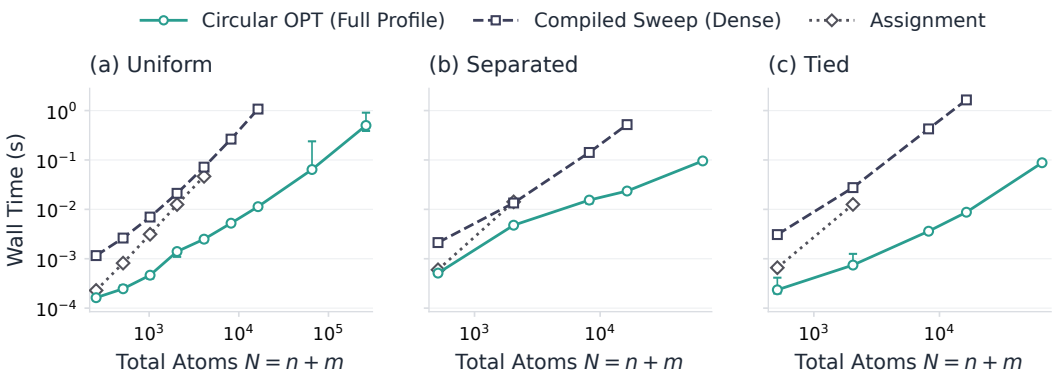


Figure 2: Quadratic-cost profile computation on uniform, separated, and tied inputs. COPT computes every cardinality and recovers one selected plan, whereas the cut sweep and the assignment solver return the solution for one penalty. Curves show median runtimes with interquartile ranges over three independent inputs and three repetitions per method, including sorting and preprocessing.

Figure 2 compares COPT with our compiled sliding-cut sweep, which repairs an optimal line matching as the cut moves, and with direct assignment. On uniform inputs with $n = m = 8192$, the full profile and a selected plan take 11.3 milliseconds, against 1.07 seconds for the sweep at one penalty. This speedup of $95\times$ rises to $188\times$ on repeated grid locations, and the profile remains practical at 131,072 atoms per measure, where it takes 0.50 seconds. On separated supports, which repeatedly rematch a growing interval, the interval preprocessing reduces the direct computation at 32,768 atoms per measure from 5.74 to 0.040 seconds. Appendix B reports query, memory, and single-penalty comparisons.

The profile also reduces the cost of selecting a penalty. On 324 pairs of Flowers validation palettes, evaluating 64 quadratic-cost penalties takes 0.032 seconds with one profile per pair, against 1.39

seconds for repeated early-stopped solves; the Huber-cost comparison shows the same benefit (Figure 9). In the shared W_1 setting, COPT has comparable runtimes to PAWC (Appendix C).

5.2 Matching Along the Profile

The profile describes the cost of transporting more and more observations, including those that have no counterpart. Figure 3 shows three plans recovered from one quadratic-cost profile. At low cardinality only a few nearby observations are matched. Increasing the mass recovers more of the visible observations, but eventually forces the clutter atoms into the matching. All three plans are read off the same stored profile without repeating the circular optimization.

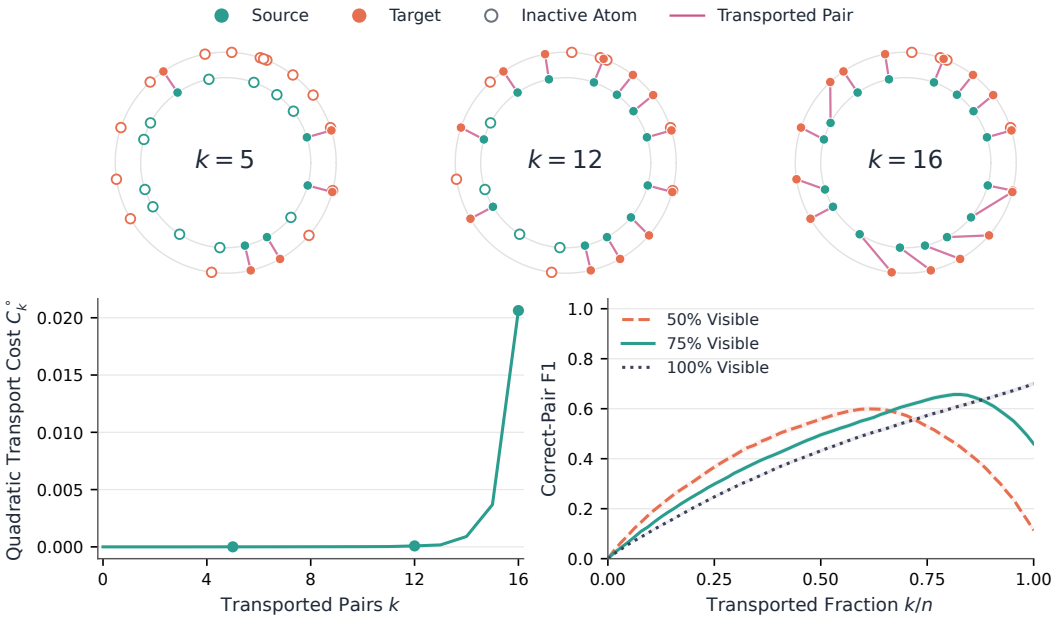


Figure 3: Correspondence recovery along a quadratic-cost profile. The example has sixteen source atoms, twelve noisy visible observations, and eight clutter atoms. Filled atoms are transported, and the three plans are marked on the cost curve. The recovery curves average 200 instances with 64 sources and 32 clutter atoms at each visible fraction, with pointwise 95% intervals.

To quantify this behavior, we retain 50%, 75%, or 100% of 64 source observations, perturb their circular locations, and add 32 uniform target samples. The correct-pair F1 first rises with the transported mass. When observations are missing, it peaks before all the source atoms are matched and decreases afterwards, so the useful mass depends on the observations and on their noise. The shape study below selects this fraction on validation data, and Appendices D and E examine correspondence at a fixed penalty and the sensitivity to the line origin.

5.3 Partial Shape Matching

The normal directions of a planar contour form a circular measure that rotates with the shape (Horn, 1984), so partial contours lead to partial matching on the circle, as in Kolouri (2026). We use the MPEG-7 silhouettes (Latecki et al., 2000) for rotation recovery and for the retrieval of other shapes of the same class. Each intact contour provides 128 normal samples. We generate randomly rotated queries from contiguous arcs covering 100%, 75%, or 50% of the contour. Independent uniform directions replace 0%, 20%, or 40% of the sampled normals. We compare partial W_1 (the PAWC objective) and quadratic-cost profiles, both computed by COPT, with full-mass quadratic matching and with balanced probability W_2, and every method searches 72 rotations. Fourteen validation classes select one transported fraction per cost from $\{0.4, 0.5, \ldots, 1\}$, namely 0.9 for W_1 and 0.9 for the quadratic cost. Rotation recovery then uses all 1,120 shapes of the other 56 classes, and

retrieval uses three queries per class. Appendix D.2 describes the descriptors, the selection rule, the database construction, and the complete results.

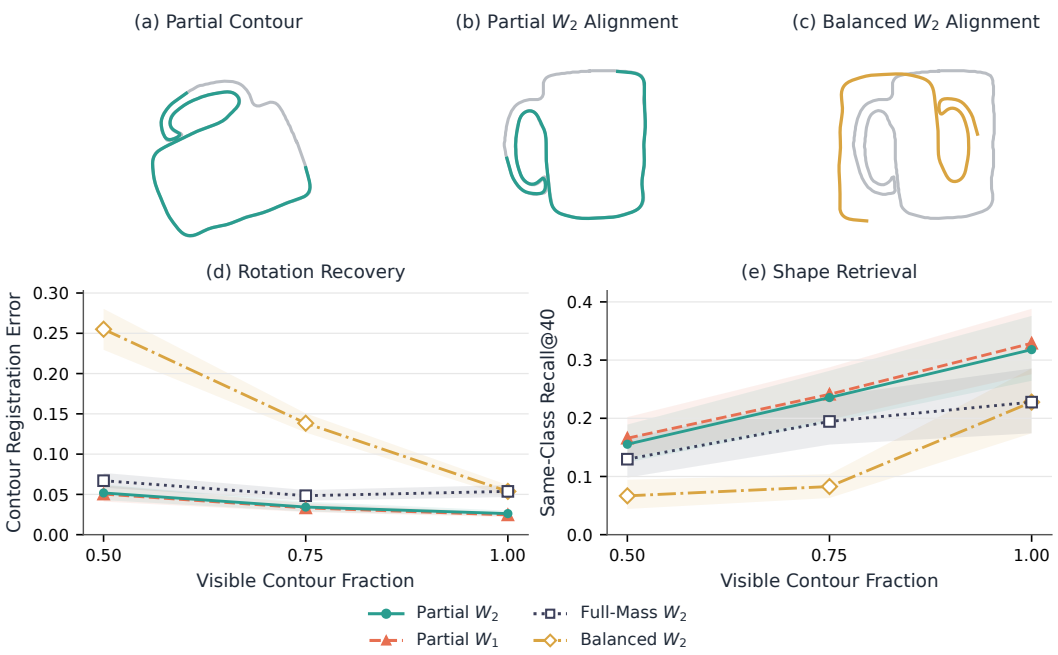


Figure 4: Partial shape matching through normal-angle measures. Gray shows the intact contour and color the observed arc before and after alignment in a validation example. The partial W_1 curve evaluates the PAWC objective (Kolouri, 2026) using COPT. The lower panels report test rotation recovery and class retrieval at 20% replaced normals, with 95% bootstrap intervals over 56 classes. The full visibility and corruption grid is shown in Figure 10.

With half of the contour visible and 20% replaced normals, quadratic-cost partial matching reduces the mean contour error from 0.255 for balanced transport to 0.052, against 0.067 at full mass and 0.050 for partial W_1. Full-mass matching can leave surplus database atoms unmatched, but partial matching can also exclude the corrupted query normals. At full visibility and 20% corruption, the quadratic-cost recall@40 increases from 22.8% at full mass to 31.8% with partial matching, and W_1 reaches 32.9%. The profile also reduces the cost of selecting the transported fraction. On the validation contours, searching seven fractions over 72 rotations takes 6.5 milliseconds with quadratic-cost profiles, against 217 milliseconds with repeated fixed-cardinality assignment, a 33-fold reduction of the complete rotation-and-mass search.

5.4 Foreground-Palette Retrieval

On Oxford Flowers, we retrieve foreground palettes while the foreground visibility and the real background clutter vary independently (Figure 18). We use the provided segmentations to draw disjoint pixel samples, select one parameter per method on twelve classes, and evaluate 216 images from 54 other classes. When the added background equals 60% of the original foreground count, COPT reaches 63.4% top-1 accuracy with the full foreground and 17.6% at 40% visibility, against 44.4% and 9.7% for fixed-mass circular transport and 4.2% and 1.4% for balanced OT. Its advantage over fixed mass with the full foreground is 19.0 percentage points (95% confidence interval 12.5–25.5). In the plans to the corresponding clean palettes, COPT retains 77.5% of the foreground atoms and rejects 91.2% of the added background. Appendix H reports the complete results, the comparison with the circular W_1 cost, and retrieval examples.

6 Spherical and Linear Extensions of COPT

6.1 Spherical Sliced COPT

We follow the great-circle slicing of Bonet et al. (2023), which Kolouri (2026) uses for fixed-mass partial transport. For $d \geq 2$ and an orthonormal two-frame $U \in \mathbb{V}_{d,2}$, we set $P^U(x) = U^\top x/\|U^\top x\|$ (with a fixed value when $U^\top x = 0$), read as an angle divided by 2π. For $p \geq 1$ and $\lambda > 0$, the measure-level OPT functional equation K.1 on the circle defines

$$\mathrm{SSCOPT}_{\lambda,p}(\mu,\nu) = \left(\int_{\mathbb{V}_{d,2}} \mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(P^U_\#\mu, P^U_\#\nu)\, d\sigma(U)\right)^{1/p}, \tag{13}$$

where σ is uniform on $\mathbb{V}_{d,2}$. In practice, we use a Monte Carlo average over slices shared by the measures being compared.

Proposition 2. *SSCOPT is a metric on finite positive Borel measures on $\mathbb{S}^{d-1}$. Every finite-slice estimator is a pseudo-metric, and a common orthogonal transformation of both measures preserves the σ-averaged value.*

This extends the continuous-density metric result of Liu et al. (2025) to finite measures with the penalized cost. Appendix K gives the proof and slice complexity. The fixed-mass SSPW discrepancy need not satisfy the triangle inequality (Kolouri, 2026, Remark 7).

With 45% uniform clutter, synthetic same-instance retrieval reaches 93.6% for SSCOPT_2, against 84.8% for fixed-mass SSCOT_2 and 90.4% for SSPW_1; SSCOPT_1 reaches 96.4%. Following Kolouri (2026), we use larger target batches for partial earthquake fitting, with $\gamma = 1.25$ throughout. At 30% contamination, the eight-run test geodesic energies are 6.55 ± 1.48 for SSCOPT_2 and 20.18 ± 2.44 for SSW_2, in 10^{-3} units (Figure 28). Appendix K reports the clean, fixed-mass, and chordal controls.

6.2 Linear COPT Embedding

Following LOPT and LCOT (Bai et al., 2023a; Díaz Martín et al., 2024), we fix $R = \sum_{i=1}^r \delta_{r_i}$ and choose a canonical optimal plan from R to each unit-atom target Y. Let $e_i(Y)$ be the endpoint matched to r_i, or the auxiliary point $\hat{\infty}$ if it is unmatched, and let η_Y be the created target mass. The Linear COPT embedding is $\Psi_R(Y) = (e_1(Y), \ldots, e_r(Y), \eta_Y)$, with distance

$$D^R_{\mathrm{geo}}(Y,Z)^p = \sum_{i=1}^r d_\lambda(e_i(Y), e_i(Z))^p + \mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(\eta_Y, \eta_Z), \tag{14}$$

where $d_\lambda(a,b)^p = \min\{\rho_{\mathbb{S}^1}(a,b)^p, 2\lambda\}$, $d_\lambda(a,\hat{\infty})^p = d_\lambda(\hat{\infty},a)^p = \lambda$, and $d_\lambda(\hat{\infty},\hat{\infty}) = 0$. Replacing the last term by $\lambda(|\eta_Y| + |\eta_Z|)$ defines the fixed-size discrepancy A^R. We prove in Appendix K that D^R_{geo} is a metric on unit-atom empirical measures and $\mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(Y,Z) \leq D^R_{\mathrm{geo}}(Y,Z)^p \leq A^R(Y,Z)^p$, with equality when one argument is R.

Concatenating the embeddings over eight slices on 320 ModelNet distributions with 48 atoms, the fixed-size form is $14.6\times$ faster than pairwise COPT at a Spearman correlation of 0.967, while refinement reaches 0.979 at a $1.54\times$ speedup. The separate retrieval study obtains 27.5% top-1 accuracy for refined LCOPT, 28.4% for COPT, and 19.4% for the fixed-size form. Appendix K gives the details.

7 Conclusion

Building on PAWL and PAWC (Chapel & Tavenard, 2025; Kolouri, 2026), we introduced COPT for the complete partial transport profile on the circle with convex radial costs and repeated support. Local augmentation and a surviving rank gap extend PAWC's common-cut construction to this setting, with the complexity guarantees of Theorem 1. In the quadratic-cost experiments, the full profile is 95–188 times faster than the single-penalty cut sweep. Our algorithm is limited to atoms of equal weight. Extending the local augmentations to unequal weights, which would carry Linear COPT to positive spherical measures (Liu et al., 2025), and finding faster interval queries for general convex costs are natural directions for future work.

Reproducibility Statement

Appendix A contains the proofs and the pseudocode, and Appendices B through K describe the experimental setup, the parameter selection, and the uncertainty conventions.

Statement on AI Use

We used generative AI tools to assist with mathematical reasoning and proofs, the design and analysis of experiments, synthetic data generation, code development and review, literature searches, and improvements to the writing and figures. We are responsible for the technical content, the experimental results, and the final text.

References

Alok Aggarwal, Amotz Bar-Noy, Samir Khuller, Dina Kravets, and Baruch Schieber. Efficient minimum cost matching using quadrangle inequality. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, pp. 583–592. IEEE, 1992. URL https://research.ibm.com/publications/efficient-minimum-cost-matching-using-quadrangle-inequality.

Sebastian Angrick, Jacobus Conradi, Mónika Csikós, Niko Hastrich, Danny Mittal, André Nusser, Krzysztof Onak, and Sharath Raghvendra. Computing all optimal partial p-Wasserstein matchings on the line. *arXiv preprint arXiv:2608.18875*, 2026. URL https://arxiv.org/abs/2608.18875.

Yikun Bai, Ivan Vladimir Medri, Rocío Díaz Martín, Rana Shahroz, and Soheil Kolouri. Linear optimal partial transport embedding. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 1492–1520. PMLR, 2023a. URL https://proceedings.mlr.press/v202/bai23c.html.

Yikun Bai, Bernhard Schmitzer, Matthew Thorpe, and Soheil Kolouri. Sliced optimal partial transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13681–13690, 2023b. URL https://openaccess.thecvf.com/content/CVPR2023/html/Bai_Sliced_Optimal_Partial_Transport_CVPR_2023_paper.html.

Yikun Bai, Huy Tran, Steven B. Damelin, and Soheil Kolouri. Partial transport for point-cloud registration. *Sampling Theory, Signal Processing, and Data Analysis*, 23(1):4, 2025. doi: 10.1007/s43670-025-00097-1.

Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. Spherical sliced-Wasserstein. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=jXQ0ipgMdU.

Clément Bonet, Kimia Nadjahi, Thibault Séjourné, Kilian Fatras, and Nicolas Courty. Slicing unbalanced optimal transport. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=AjJTg5M0r8.

Nicolas Bonneel and David Coeurjolly. SPOT: sliced partial optimal transport. *ACM Transactions on Graphics*, 38(4):89:1–89:13, 2019. doi: 10.1145/3306346.3323021.

Carlos A. Cabrelli and Ursula M. Molter. The Kantorovich metric for probability measures on the circle. *Journal of Computational and Applied Mathematics*, 57(3):345–361, 1995. doi: 10.1016/0377-0427(93)E0213-6.

Carlos A. Cabrelli and Ursula M. Molter. A linear time algorithm for a matching problem on the circle. *Information Processing Letters*, 66(3):161–164, 1998. doi: 10.1016/S0020-0190(98)00048-9.

Luis A. Caffarelli and Robert J. McCann. Free boundaries in optimal transport and Monge–Ampère obstacle problems. *Annals of Mathematics*, 171(2):673–730, 2010. doi: 10.4007/annals.2010.171.673.

Laetitia Chapel and Romain Tavenard. One for all and all for one: efficient computation of partial Wasserstein distances on the line. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=kzEPsHbJDv.

Laetitia Chapel, Mokhtar Z. Alaya, and Gilles Gasso. Partial optimal transport with applications on positive-unlabeled learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2903–2913, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/1e6e25d952a0d639b676ee20d0519ee2-Abstract.html.

Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. In *Advances in Neural Information Processing Systems*, volume 34, pp. 23270–23282, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/c3c617a9b80b3ae1ebd868b0017cc349-Abstract.html.

Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018. doi: 10.1090/mcom/3303.

Harald Cramér and Herman Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, s1-11(4):290–294, 1936. doi: 10.1112/jlms/s1-11.4.290.

Julie Delon, Julien Salomon, and Andrei Sobolevski. Fast transport optimization for Monge costs on the circle. *SIAM Journal on Applied Mathematics*, 70(7):2239–2258, 2010. doi: 10.1137/090772708.

Rocío Díaz Martín, Ivan Medri, Yikun Bai, Xinran Liu, Kangbai Yan, Gustavo K. Rohde, and Soheil Kolouri. LCOT: linear circular optimal transport. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/1baf53ac5c85ab5046bc7541dc3eb14a-Abstract-Conference.html.

Alessio Figalli. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2):533–560, 2010. doi: 10.1007/s00205-008-0212-7.

Alessio Figalli and Nicola Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 94(2):107–130, 2010. doi: 10.1016/j.matpur.2009.11.005. URL https://cvgmt.sns.it/paper/488/.

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. URL https://jmlr.org/papers/v22/20-451.html.

Gabriel Gouvine. An efficient algorithm for unbalanced 1D transportation. *arXiv preprint arXiv:2311.17704*, 2023. URL https://arxiv.org/abs/2311.17704.

Jens Gregor and Michael G. Thomason. Dynamic programming alignment of sequences representing cyclic patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):129–135, 1993. doi: 10.1109/34.192484.

Florian Heinemann, Marcel Klatt, and Axel Munk. Kantorovich–Rubinstein distance and barycenter for finitely supported measures: foundations and algorithms. *Applied Mathematics & Optimization*, 87:4, 2023. doi: 10.1007/s00245-022-09911-x.

Berthold K. P. Horn. Extended Gaussian images. *Proceedings of the IEEE*, 72(12):1671–1686, 1984. doi: 10.1109/PROC.1984.13073. URL https://people.csail.mit.edu/bkph/papers/EGI.pdf.

Shayan Hundrieser, Marcel Klatt, and Axel Munk. The statistics of circular optimal transport. In *Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale*, pp. 57–82. Springer, 2022. doi: 10.1007/978-981-19-1044-9_4.

Richard M. Karp and Shuo-Yen Robert Li. Two special cases of the assignment problem. *Discrete Mathematics*, 13(2):129–142, 1975. doi: 10.1016/0012-365X(75)90014-X.

Peter Knees, Ángel Faraldo, Perfecto Herrera, Richard Vogl, Sebastian Böck, Florian Hörschläger, and Mickael Le Goff. Two data sets for tempo estimation and key detection in electronic dance music annotated from user corrections. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pp. 364–370, 2015. URL https://www.cp.jku.at/research/papers/knees_etal_ismir_2015.pdf.

Soheil Kolouri. Partial optimal transport on the circle for all transported masses in $O(N \log N)$. *arXiv preprint arXiv:2608.23910*, 2026. URL https://arxiv.org/abs/2608.23910.

Carol L. Krumhansl and Edward J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4):334–368, 1982. doi: 10.1037/0033-295X.89.4.334.

Longin Jan Latecki, Rolf Lakämper, and Ulrich Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pp. 424–429, 2000. doi: 10.1109/CVPR.2000.855850. URL https://cis.temple.edu/~latecki/Papers/cvpr00.pdf.

Xinran Liu, Yikun Bai, Rocío Díaz Martín, Kaiwen Shi, Ashkan Shahbazi, Bennett A. Landman, Catie Chang, and Soheil Kolouri. Linear spherical sliced optimal transport: a fast metric for comparing spherical data. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/cc10fd7fe3e0d2af517cedf99fbc0651-Abstract-Conference.html.

Russell Lyons. Strong negative type in spheres. *Pacific Journal of Mathematics*, 307(2):383–390, 2020. doi: 10.2140/pjm.2020.307.383.

Maurice Maes. On a cyclic string-to-string correction problem. *Information Processing Letters*, 35 (2):73–78, 1990. doi: 10.1016/0020-0190(90)90109-B.

Brian McFee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference*, pp. 18–24, 2015. doi: 10.25080/Majora-7b98e3ed-003.

Debarghya Mukherjee, Aritra Guha, Justin M. Solomon, Yuekai Sun, and Mikhail Yurochkin. Outlier-robust optimal transport. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7850–7860. PMLR, 2021. URL https://proceedings.mlr.press/v139/mukherjee21a.html.

Raoul Müller, Dominic Schuhmacher, and Jorge Mateu. Metrics and barycenters for point pattern data. *Statistics and Computing*, 30:953–972, 2020. doi: 10.1007/s11222-020-09932-y.

Natural Earth. 1:110m physical vectors, 2026. URL https://github.com/nvkelso/natural-earth-vector/releases/tag/v5.1.2. Version 5.1.2 vector release.

New York City Taxi and Limousine Commission. TLC trip record data, 2026. URL https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

Khai Nguyen, Dang Nguyen, The-Anh Vu-Le, Tung Pham, and Nhat Ho. Improving mini-batch optimal transport via partial transportation. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16656–16690. PMLR, 2022. URL https://proceedings.mlr.press/v162/nguyen22e.html.

Sloan Nietert, Ziv Goldfeld, and Rachel Cummings. Outlier-robust optimal transport: duality, structure, and statistical analysis. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 11691–11719. PMLR, 2022. URL https://proceedings.mlr.press/v151/nietert22a.html.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 722–729, 2008. doi: 10.1109/ICVGIP.2008.47.

Ofir Pele and Michael Werman. A linear time histogram metric for improved SIFT matching. In *Computer Vision – ECCV 2008*, volume 5304 of *Lecture Notes in Computer Science*, pp. 495–508. Springer, 2008. doi: 10.1007/978-3-540-88690-7_37.

Ofir Pele and Michael Werman. Fast and robust Earth Mover's Distances. In *IEEE International Conference on Computer Vision*, pp. 460–467, 2009. doi: 10.1109/ICCV.2009.5459199.

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. doi: 10.1561/2200000073.

Abhijeet Phatak, Sharath Raghvendra, Chittaranjan Tripathy, and Kaiyi Zhang. Computing all optimal partial transports. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=gwcQajoXNF.

Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660, 2017. URL https://openaccess.thecvf.com/content_cvpr_2017/html/Qi_PointNet_Deep_Learning_CVPR_2017_paper.html.

Michael Quellmalz, Robert Beinert, and Gabriele Steidl. Sliced optimal transport on the sphere. *Inverse Problems*, 39(10):105005, 2023. doi: 10.1088/1361-6420/acf156.

Julien Rabin, Julie Delon, and Yann Gousseau. Transportation distances on the circle. *Journal of Mathematical Imaging and Vision*, 41(1–2):147–167, 2011. doi: 10.1007/s10851-011-0284-0.

Sharath Raghvendra, Pouyan Shirzadian, and Kaiyi Zhang. A new robust partial p-Wasserstein-based metric for comparing distributions. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 41867–41885. PMLR, 2024. URL https://proceedings.mlr.press/v235/raghvendra24a.html.

Abu Sajana Rahmathullah, Ángel F. García-Fernández, and Lennart Svensson. Generalized optimal sub-pattern assignment metric. In *Proceedings of the 20th International Conference on Information Fusion*, pp. 1–8, 2017. doi: 10.23919/ICIF.2017.8009645.

Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. Faster unbalanced optimal transport: translation invariant Sinkhorn and 1-D Frank–Wolfe. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 4995–5021. PMLR, 2022. URL https://proceedings.mlr.press/v151/sejourne22a.html.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: a class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013. doi: 10.1016/j.jspi.2013.03.018.

Huy Tran, Yikun Bai, Abihith Kothapalli, Ashkan Shahbazi, Xinran Liu, Rocío Díaz Martín, and Soheil Kolouri. Stereographic spherical sliced Wasserstein distances. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 48494–48564. PMLR, 2024. URL https://proceedings.mlr.press/v235/tran24a.html.

U.S. Geological Survey. ANSS comprehensive earthquake catalog (ComCat), 2026. URL https://earthquake.usgs.gov/fdsnws/event/1/. Magnitude at least 4.5, 2010–2024; retrieved September 2026.

Michael Werman, Shmuel Peleg, and Azriel Rosenfeld. A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing*, 32(3):328–336, 1985. doi: 10.1016/0734-189X(85)90055-6.

Michael Werman, Shmuel Peleg, Robert A. Melter, and T. Yung Kong. Bipartite graph matching for points on a line or a circle. *Journal of Algorithms*, 7(2):277–284, 1986. doi: 10.1016/0196-6774(86)90009-X.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: a deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, 2015. URL https://openaccess.thecvf.com/content_cvpr_2015/html/Wu_3D_ShapeNets_A_2015_CVPR_paper.html.

A PROOFS

NOTATION

We summarize the notation used in the profile sections and in the Linear COPT appendix. In the database comparisons of Section 6 and Appendix K, D denotes the number of distributions and n the number of atoms per distribution.

Symbol	Definition
Partial transport profiles (Section 4)	
n, m, N, K	$n = \lvert X\rvert$, $m = \lvert Y\rvert$, $N = n + m$, $K = \min(n, m)$.
h, λ	Finite convex $h : [0, 1] \to \mathbb{R}_+$ with $h(0) = 0$; $\lambda > 0$ is the penalty per unmatched atom.
$C^\circ_k, \widehat{C}_k$	$\min_{\lvert M\rvert=k} \sum_{(i,j)\in M} h(\rho_{\mathbb{S}^1}(x_i, y_j))$ and $\min_{\text{lex},\, \lvert M\rvert=k} \sum_{(i,j)\in M} \widehat{c}(i, j)$, respectively, over partial matchings M, $0 \le k \le K$.
ℓ, q	Arc length and number of rank gaps of the same directed arc.
$\widehat{c}(i, j)$	$\min_{\text{lex}}\{(h(\ell), q), (h(1 - \ell), N - q)\}$, the refined cost.
I, I°	$I = [u, v]_\circlearrowright$ is a cell between consecutive inactive atoms; I° is its open interior.
$W_h(J), \widehat{W}(J)$	$W_h(J) = \sum_{j=1}^{\lvert J\cap X\rvert} h(\lvert x^J_j - y^J_j\rvert)$ for balanced J. $\widehat{W}(J) = (W_h(J), \sum \lvert i - j\rvert)$, with the second sum over the same sorted pairs in unwrapped ranks. With $J = I^\circ$, this is the interior cost.
$\Delta(I)$	$W_h(I) - W_h(I^\circ)$; the refined increment is $\widehat{W}(I) - \widehat{W}(I^\circ)$.
δ_k, k_λ	$\delta_k = C^\circ_k - C^\circ_{k-1}$ for $1 \le k \le K$; $k_\lambda = \min \arg\min_{0\le k\le K}\{C^\circ_k + \lambda(N - 2k)\}$.
τ	Activation rank of an atom, the first cardinality at which it is matched.
$g^\star, \sigma^\star$	A rank gap outside every selected cell, and its common cut for all cardinalities, respectively.
$\sigma \in \mathcal{C}(X, Y)$	An ordinary support-gap or virtual cut; X^σ, Y^σ are its sorted unwrapped coordinates.
(v, r, s)	Virtual cut at v: r source copies and s target copies at v lie at endpoint 1, and the remaining copies at v lie at 0.
Linear COPT ($h(t) = t^p$, $p \ge 1$; Appendix K.3)	
R, r	Fixed reference $R = \sum_{i=1}^r \delta_{r_i}$, with $r = \lvert R\rvert$.
$e_i(Y)$	Target endpoint matched to r_i by the fixed optimal-plan selector, or $\hat\infty$ if r_i is unmatched.
d_λ	$d_\lambda(a, b)^p = \min\{\rho_{\mathbb{S}^1}(a, b)^p, 2\lambda\}$ for $a, b \in \mathbb{S}^1$; $d_\lambda(a, \hat\infty)^p = d_\lambda(\hat\infty, a)^p = \lambda$ and $d_\lambda(\hat\infty, \hat\infty) = 0$.
η_Y, q_Y	$\eta_Y = Y - \sum_{i:e_i(Y)\ne\hat\infty} \delta_{e_i(Y)}$, $q_Y = \lvert\eta_Y\rvert$.
$\Psi_R(Y), \Phi_R(Y)$	$(e_1(Y), \dots, e_r(Y), \eta_Y)$ and $(e_1(Y), \dots, e_r(Y), q_Y)$, respectively.
$D^R_{\text{geo}}(Y, Z)^p$	$\sum_{i=1}^r d_\lambda(e_i(Y), e_i(Z))^p + \text{COPT}^{\mathbb{S}^1}_{\lambda,p}(\eta_Y, \eta_Z)$.
$A^R(Y, Z)^p$	$\sum_{i=1}^r d_\lambda(e_i(Y), e_i(Z))^p + \lambda(q_Y + q_Z)$.
$T_{\text{COPT}}(r, n), L$	Time for one reference solve, including sorting and recovery; number of slices, respectively.

Throughout the proofs of the profile algorithm, we work with atoms of unit mass and with a fixed total order of the coincident copies. This order refines the circular coordinates, and an atom of rank i has rank $i + N$ after one turn. Refined costs are compared lexicographically. We write $\widehat{C}_k$ for the minimum sum of the refined costs equation 11 over cardinality-k matchings, whose first component is C°_k, and $\widehat{W}(J)$ for the sorted cost of an unwrapped interval J, whose two components sum $h(\lvert z_i - z_j\rvert)$ and $\lvert i - j\rvert$ over the pairs of the sorted matching.

A.1 MONGE STRUCTURE AND LINE OPT

We first recall why sorted supports lead to monotone optimal plans. Bai et al. (2023b) prove that empirical OPT admits an optimal one-to-one partial plan, and we extend the monotone representation to every convex $h : [0, 1] \to \mathbb{R}_+$ with $h(0) = 0$ through a Monge argument.

Our assumptions make h nondecreasing, since $h(s) \le (s/t)h(t) \le h(t)$ for $0 < s < t$, so that $f(t) = h(|t|)$ is convex on $[-1, 1]$. We claim that the cost matrix of $h(|x - y|)$ on sorted lists is Monge, that is, for $a \le b$ and $c \le d$ in a common unit interval,

$$h(|a - c|) + h(|b - d|) \le h(|a - d|) + h(|b - c|). \tag{A.1}$$

Indeed, fix $b - a \ge 0$. Since f is convex, the increment $f(t + b - a) - f(t)$ is nondecreasing in t, and we evaluate it at $t = a - d$ and $t = a - c$ and rearrange to obtain equation A.1. This lets us uncross every crossing pair of a plan without increasing its cost and obtain a monotone optimum. Tied coordinates or flat regions of h may leave several optima. In that case we order the coincident copies by their ranks in the cyclic order and uncross the inversions in that order. The same argument applies to $|i - j|$ on the unwrapped rank order, so both components of the refined cost admit the same sorted optimizer. The argument uses neither strict convexity nor distinctness, and it allows an upward jump of h at 1, because equation A.1 holds at the endpoints of the domain and every unwrapped cell has span at most one.

Finally, the monotone partial plans of a sorted instance are exactly the alignments of its two lists, and the dynamic program

$$D_{i,j} = \min\{D_{i-1,j} + \lambda,\ D_{i,j-1} + \lambda,\ D_{i-1,j-1} + h(|x_i - y_j|)\}, \qquad D_{i,0} = i\lambda, \quad D_{0,j} = j\lambda, \tag{A.2}$$

computes their optimum in $O(nm)$ time. By the Monge argument above, the unrestricted line problem has such a monotone optimum for every h in our class, and for a balanced interval J the sorted matching of its full source and target lists is $W_h(J)$.

A.2 Line OPT and Balanced Circular Transport

For the sorted line lists, the Monge inequality allows one to uncross any plan without increasing its cost, and the problem reduces to the search for a monotone partial map,

$$\min_L C(L), \qquad C(L) = \sum_{i \in \mathrm{dom}(L)} c(x_i, y_{L[i]}) + \lambda(n + m - 2|\,\mathrm{dom}(L)|). \tag{A.3}$$

Here $L : [1 : n] \to \{-1\} \cup [1 : m]$ sends each source index either to its target index or to -1 when it is not transported, $\mathrm{dom}(L) = \{i : L[i] \ne -1\}$, and $L|_{\mathrm{dom}(L)}$ is strictly increasing. The minimum is $\mathrm{OPT}^{\mathbb{R}}_{\lambda}(X, Y)$. Under strict convexity, the primal-dual algorithm of Bai et al. (2023b) has the same $O(nm)$ bound as the alignment dynamic program, with the shorter list as source.

For probability measures μ, ν on $\mathbb{S}^1$, the circular OT cost is

$$\mathrm{COT}_h(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{S}^1 \times \mathbb{S}^1} h(\rho_{\mathbb{S}^1}(x, y))\, d\gamma(x, y), \tag{A.4}$$

where $\Gamma(\mu, \nu)$ is the set of couplings with marginals μ and ν. Let $F_{\mu,z}(t) = \mu(u_z^{-1}([0, t)))$ be the cumulative distribution function after cutting at z, as in Definition 2.1 of Díaz Martín et al. (2024), and let $F^{-1}_{\mu,z}(r) = \inf\{t : F_{\mu,z}(t) > r\}$ be its quantile function. For two uniformly weighted sets of the same cardinality at pairwise-distinct locations, Rabin et al. (2011) prove

$$\mathrm{COT}_h(\mu, \nu) = \inf_{z \in [0,1)} \int_0^1 h\left(\left|F^{-1}_{\mu,z}(r) - F^{-1}_{\nu,z}(r)\right|\right)\, dr, \tag{A.5}$$

which Delon et al. (2010) and Díaz Martín et al. (2024) also discuss. With repeated atoms, its right-hand side can strictly exceed COT_h: a cut at z sends every copy at z to the same endpoint, whereas an optimal plan may require only some copies to cross. Section 4.3 gives an example. For general measures, Delon et al. (2010) allow a mass offset that may split atoms; virtual cuts recover this freedom for atoms of unit mass. The transported part of an optimal partial plan solves balanced circular transport on its active atoms, and Proposition 1 extends the cut representation to all cardinalities at once.

A.3 Proof of Lemma 1

The proof has three steps. We first show that one augmenting path leads from an optimal matching at cardinality k to an optimal matching at cardinality $k + 1$. We then show that this path lies inside

one current cell. Finally, we show that closing a cell of minimum increment produces an optimal matching and that the stored cell costs remain valid.

Step 1: successive augmentation. Let M minimize the refined cost among cardinality-k matchings and let M^+ minimize it among cardinality-$(k+1)$ matchings. Their symmetric difference decomposes into vertex-disjoint alternating paths and cycles, each of which changes the cardinality by -1, 0, or $+1$. Toggling a zero-change component in M produces another cardinality-k matching, so its net cost is nonnegative, and the same holds for any pair of components with changes $+1$ and -1. Since there is one more positive component than negative components, we may pair every negative component with a positive one and leave a single positive path P unpaired. Writing Φ for the added minus the removed cost, the preceding inequalities give

$$\Phi(P) \le \widehat{C}_{k+1} - \widehat{C}_k, \tag{A.6}$$

and $M \triangle P$ is a cardinality-$(k+1)$ matching, so optimality at cardinality $k+1$ provides the reverse inequality. It follows that an augmenting path of minimum net cost reaches an optimum at the next cardinality, and in particular the active source and target sets can be chosen nested. This is the usual successive-augmentation property of bipartite matching, also used by Angrick et al. (2026).

Step 2: locality. We now show that such an augmentation is local on the rank circle. For every matched edge, we choose an arc attaining its refined cost. If an inactive atom lies strictly inside that arc in the rank order, we replace the endpoint of the same type by the inactive atom. The new matching is feasible with the same cardinality, and the replacement arc has no greater arc length and strictly smaller rank length, so its refined cost is strictly smaller, which contradicts optimality. Consequently, no selected arc of an optimal matching passes over an inactive atom.

Let P be an augmenting path of minimum net cost from an inactive source u to an inactive target v, and set $M' = M \triangle P$. Every old inactive atom other than u and v remains inactive in M', and since both M and M' are optimal, none of their arcs passes over those atoms. An old edge cannot pass over u or v either, because both are inactive in M. Suppose now that a new edge (a, b) of P passes over v. Its alternating suffix Q from the target b to the target v starts with an edge of M and ends with an added edge. Toggling Q alone in M frees b, matches v, and keeps every internal vertex matched. This is a feasible exchange of equal cardinality, so the optimality of M forces $\Phi(Q) \ge_{\mathrm{lex}} 0$. Replacing (a, b) and this suffix by the single edge (a, v) produces a simple feasible augmenting path P', and since the arc from a to v is a proper subarc of the selected arc from a to b,

$$\Phi(P') - \Phi(P) = \widehat{c}(a, v) - \widehat{c}(a, b) - \Phi(Q) <_{\mathrm{lex}} 0, \tag{A.7}$$

which contradicts the minimality of P. Here we used that the refined cost of (a, v) is at most the refined cost of that proper subarc, whether or not the subarc attains it. If instead a new edge (a, b) passes over u, the prefix of P from u to the source a starts with an added edge and ends with an old edge. Toggling it in M matches u, frees a, and preserves the cardinality, so its net cost is nonnegative, and replacing the prefix and (a, b) by (u, b) produces the same strict improvement. This is the endpoint-shortcut argument of Angrick et al. (2026), applied to the two directed circular arcs and to the refined cost.

We have shown that no old inactive atom lies in the interior of an arc of P. Since consecutive arcs of P share an active atom, the whole path lies in one closed cell between consecutive old inactive atoms, and its endpoints have opposite types, so they delimit a candidate cell. The same reasoning covers the full-mass transition, since when only two inactive atoms remain their two complementary directed cells are the only two possible components.

Step 3: minimum increment. We finally show that a minimum cell increment reaches an optimum. Since no edge of M crosses an inactive atom, every open cell contains equally many active sources and targets. We apply Monge uncrossing inside its unwrapped interval, which preserves the active atoms without increasing either component. The old arcs attain their circular refined costs in this interval, and a strict decrease of their sorted cost, or a further decrease on returning to circular costs, would improve M. The sorted unwrapped cost $\widehat{W}(I^\circ)$ is then exactly the old interior cost. For any candidate cell I, replacing this matching by the sorted matching of I is feasible at cardinality $k+1$, and its circular refined cost is no greater than its unwrapped cost, so every candidate satisfies

$$\widehat{C}_{k+1} - \widehat{C}_k \le_{\mathrm{lex}} \widehat{W}(I) - \widehat{W}(I^\circ). \tag{A.8}$$

On the other hand, the optimal augmenting path of Step 2 lies inside some candidate $I^\star$, all of its arcs use the unwrapped interval of this cell, and uncrossing the matching that it produces preserves

the active atoms and cannot increase its cost, so $I^\star$ attains equality. We conclude that any candidate of minimum increment produces an optimal matching, and that its stored unwrapped cost equals its circular refined cost, since a strict inequality would again contradict optimality. This completes the induction from the empty matching.

It remains to check the stored cell costs. Before each activation, every open cell is balanced, its cached refined cost is its sorted unwrapped cost, and the union of these interior matchings attains $\widehat{C}_k$. Unless the last two inactive atoms have just left, the selected closed cell and its two neighboring interiors are balanced consecutive regions, so their sorted matchings concatenate and equation 12 holds for both components. All the other cells keep their costs and keys. Since inactive atoms are only removed, an old heap entry is valid exactly while both of its endpoints remain inactive and consecutive in its directed order, and such a surviving adjacency has not changed its interior, so only the merged cell needs a new key. This concludes the proof.

A.4 Proof of Proposition 1

We first prove one direction of the cut identity. Fix $\sigma \in \mathcal{C}(X, Y)$ and a feasible increasing partial plan of the unwrapped instance (X^σ, Y^σ). Mapping each transported pair back to the circle preserves the active atoms and the penalty, and since $\rho_{\mathbb{S}^1}(x, y) \le |\tilde{x} - \tilde{y}|$ and h is nondecreasing, it cannot increase the transport cost. Taking the infimum over the line plans and then the minimum over σ, every fixed-cut line value is at least the circular OPT value.

For the reverse inequality, we use the active sets constructed by Lemma 1. If $K = 0$, the empty matching is feasible at every cut with the same objective, and there is nothing to prove. Otherwise, we exhibit a gap outside all the selected cells. Call an original rank gap free until it belongs to a selected directed cell. Initially every cell consists of one such gap. If $[u, v]$ is selected and inactive atoms survive on both sides, the cells $[a, u]$, $[u, v]$, and $[v, b]$ merge, and since the two neighboring cells each retain a free gap, the merged cell inherits one. This remains true when $a = b$, which happens with three inactive atoms before the last activation, in which case the merged cell is the full turn based at the remaining atom. The other cells keep their gaps. When unequal sizes exhaust the smaller measure, we take a representative from any remaining cell, and with two inactive atoms the selected cell and its complement are distinct, so we retain a free gap of the latter. Hence a gap $g^\star$ outside every selected cell survives the whole sequence. This is the free-gap induction of Kolouri (2026), with rank gaps in place of distinct physical support gaps.

We now start the cyclic rank order immediately after $g^\star$. Every selected cell becomes an ordinary interval, and its sorted rematching agrees with the unwrapped interval used during its activation. Moreover, every selected cell either contains or is disjoint from each earlier selected cell. The maximal selected cells therefore form disjoint balanced intervals covering the active atoms, and their sorted matchings concatenate, as do the three balanced regions of equation 12. By induction over the activations, sorting the active sources and targets at this cut recovers at each cardinality the matching recorded by the algorithm, whose first component is C_k°.

If the surviving gap has positive arc length, any point of that support gap defines the required cut. Otherwise the gap lies between two copies at one coordinate v, and the numbers of source and target copies that precede it within the tied block determine the virtual split (v, r, s), because the rank order places those copies at the endpoint 1 and the remaining copies at the endpoint 0, without any limiting argument. The same cut $\sigma^\star$ then represents every C_k°. The fixed copy order need not represent every virtual split at a given support point. Its global sufficiency follows from Lemma 1, which establishes optimality over all circular matchings before we construct the common cut.

At this cut, we choose the active cardinality minimizing equation 6. Its unwrapped matching has exactly the circular objective, and the fixed-cut optimum cannot exceed the value of this feasible plan. Combined with the first direction, this proves the proposition with the same $\sigma^\star$ for every penalty.

Relation to cyclic alignment. For nonempty sorted circular lists $X^\uparrow$ and $Y^\uparrow$, let $\mathrm{Align}_{\lambda, c_{\mathbb{S}^1}}$ denote the alignment value with substitution cost $c_{\mathbb{S}^1}(x, y) = h(\rho_{\mathbb{S}^1}(x, y))$ and insertion and deletion cost λ. Then

$$\mathrm{COPT}_\lambda^{\mathbb{S}^1}(X, Y) = \min_{0 \le s < m} \mathrm{Align}_{\lambda, c_{\mathbb{S}^1}}(X^\uparrow, \mathrm{rot}_s\, Y^\uparrow). \tag{A.9}$$

Indeed, every alignment is a feasible circular partial matching, so its cost is at least the COPT value. Conversely, Proposition 1 provides an optimal cyclically ordered matching. Holding the source order fixed and rotating the target list so that the partner of the first matched source comes first realizes this matching as an alignment, the empty matching being immediate, and tied copies may be rotated individually with their labels chosen consistently, so one relative rotation attains the COPT value. The weighted alignment graph repeats its target columns, so the noncrossing shortest-path separator algorithm of Maes (1990) applies even though the substitution cost need not satisfy a triangle inequality, and rotating the shorter list leads to the stated $O(nm\log(1+\min(n,m)))$ bound.

A.5 Proof of Theorem 1

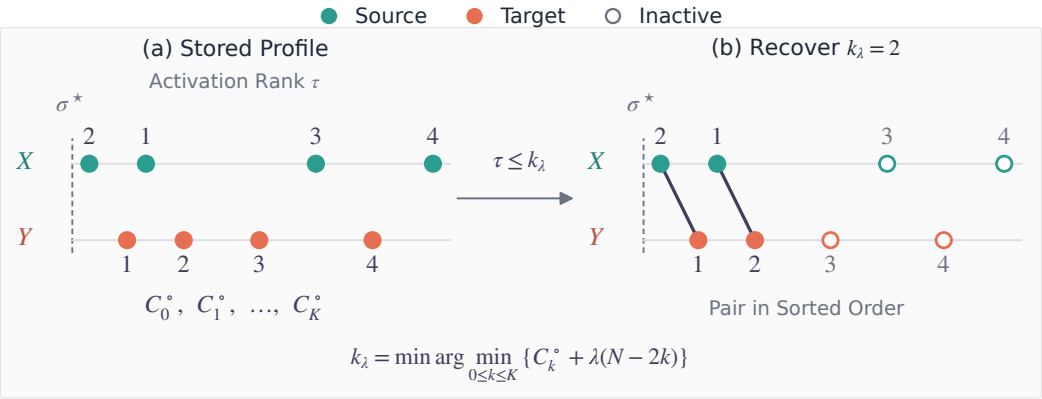


Figure 5: Recovering a plan from the stored profile. (a) The growth records the activation rank τ of every atom and one free gap, which defines the common cut $\sigma^\star$. (b) For a penalty λ, the cardinality k_λ minimizes $C_k^\circ + \lambda(N-2k)$, and the atoms with $\tau \le k_\lambda$ are kept and paired in sorted order from $\sigma^\star$.

Algorithm 1 Circular OPT by transported cardinality, adapting PAWC (Kolouri, 2026) with rank-refined increments

Require: Lists X, Y of atoms of unit mass and finite convex $h : [0,1] \to \mathbb{R}_+$ with $h(0) = 0$
if $\min(n,m) = 0$ **then**
  return $C_0^\circ = 0$, the empty matching, and any cut
end if
Sort the union support cyclically; initialize the circular list of inactive atoms
Set $C_0^\circ = 0$; store a free rank gap in each initial cell
Insert every candidate cell into a heap by its refined increment
for $k = 0, \ldots, \min(n,m) - 1$ **do**
  Extract a minimum candidate $I = [u,v]_\circlearrowleft$, discarding obsolete entries
  Set $C_{k+1}^\circ = C_k^\circ + \Delta(I)$ using the first component; record the activation of u, v
  if u, v are the last two inactive atoms **then**
    Retain a free gap from the complementary cell and stop
  end if
  Remove u, v; merge their cell with its two neighbors
  Inherit a free gap from a neighboring cell
  Insert the merged cell if its endpoints have opposite types
end for
if $n \neq m$ **then**
  Retain the free gap of the last merged cell
end if
return the profile, the activation ranks, and the surviving gap

We first prove correctness. We proceed by induction over the cardinalities and show that, after each activation, the maintained matching is optimal for the refined cost equation 11. The empty

matching is the base case, and for the induction step Lemma 1 shows that closing a cell of minimum increment reaches an optimum at the next cardinality. Thus every recorded first component is C_k°, and Proposition 1 provides a common cut for the whole sequence, including its final activation.

We next prove the running-time and memory bounds in a unit-cost arithmetic model, in which one evaluation of h and one arithmetic operation or comparison each take $O(1)$ time. Sorting the union support takes $O(N \log N)$ time and $O(N)$ storage. Initially every candidate has an empty interior and costs one edge evaluation, and there are at most $2K$ opposite-type adjacencies. Each activation creates at most one new candidate, whose interior contains at most k matched pairs at stage k, so that evaluating its sorted closed-cell cost takes $O(k+1)$ time, while its open-cell cost is already stored through equation 12. Summed over the stages, the query work is $O(N+K^2)$. There are $O(K)$ heap entries in total, including the obsolete entries that are later discarded, so the heap operations take $O(N \log N)$ time, and the linked neighbors, interior costs, gap representatives, activation ranks, and profile values all require linear storage.

For $h(t) = t$, we use the chain recurrence of Chapel & Tavenard (2025) and Kolouri (2026). On the doubled sequence $z_1, \dots, z_{2N}$, let $s_i = 1$ for a source and $s_i = -1$ for a target, let $B_t = \sum_{i \le t} s_i$ and $S_t = \sum_{i \le t} s_i z_i$ with zero initial values, and let p_t be the last index before t with $B_{p_t} = B_t$. Define

$$Q_0 = 0, \qquad Q_t = \begin{cases} Q_{p_t} + |S_t - S_{p_t}|, & p_t \text{ exists}, \\ 0, & \text{otherwise}. \end{cases} \tag{A.10}$$

Between two consecutive visits to the same differential rank, the relative rank keeps a constant sign, so every pair of the sorted matching of this block has the same orientation and the cost of the block is the absolute value of its signed coordinate sum. A balanced interval $[a, b]$ decomposes at these returns, and its cost is $Q_b - Q_{a-1}$. This argument only uses the fixed index order and not strict inequalities between coordinates, and replacing z_i by i computes the rank cost of the same sorted matching. A last-visit array indexed by the differential rank builds both tables in $O(N)$ time and space, and each query takes $O(1)$ time, and with the heap this proves the $O(N \log N)$ time and $O(N)$ space bounds for W_1. Only the queried intervals need to be balanced, so unequal input counts are allowed.

We now turn to higher integer powers. We double the sorted sequence, so that every candidate cell becomes a contiguous interval whose open and closed versions are both balanced, and we apply the data structure of Angrick et al. (2026) to this doubled order. At a node whose two children are spatially ordered, the identity

$$(y - x)^p = \sum_{j=0}^{p} \binom{p}{j} (-x)^j y^{p-j} \tag{A.11}$$

reduces the cross-child costs of all rank shifts to correlations. Each node stores $O(s)$ entries for s atoms, the FFT construction of a node costs $O(ps \log s)$, and summing over the levels of the tree, we obtain $O(pN \log^2 N)$ preprocessing and $O(N \log N)$ storage. The component $|i - j|$ has the same construction with $p = 1$ on the ranks, and integer prefix sums accumulate it.

We claim that balancedness lets each query use its full cross-child terms. Write the source indices of a query as $\alpha, \dots, \alpha + t - 1$ and its target indices as $\beta, \dots, \beta + t - 1$, so that the common matching offset is $d = \beta - \alpha$. A pair (x_j, y_{j+d}) with $j < \alpha$ has both endpoints before the query, and one with $j > \alpha + t - 1$ has both endpoints after it, so neither can cross a node split lying inside the query. Thus every offset-d cross-child pair at such a split belongs to the query, the precomputed cross-child term counts in full, and the recursion handles the two boundary parts while full nodes use their stored offset value. This visits $O(\log N)$ contributions, as proved in the balanced-interval lemma of Angrick et al. (2026). There are only $O(N)$ queries over the activation sequence, so they do not exceed the preprocessing bound, and the argument applies to unequal input sizes because only the queried intervals, not the whole doubled input, need to be balanced.

To recover a plan at cardinality k, we scan the cyclic order from the surviving gap, retain the atoms whose activation rank is at most k, and pair the source and target subsequences in order. The common-cut argument above establishes optimality, and the recovery takes $O(N)$ time. Explicitly outputting all the partner maps may take $\Theta(K^2)$ space or time. The theorem covers the profile, its compact history, and the recovery of one requested plan.

Finally, we consider a solve at a single penalty. The profile is a convex sequence, because the minimum-cost-flow relaxation with prescribed transported mass has a convex value function and unit capacities admit integral optima at every integer mass, so the increments $\delta_k = C^\circ_k - C^\circ_{k-1}$ are nondecreasing. Adding the k-th pair changes the penalized objective by $\delta_k - 2\lambda$, so rejecting the next activation as soon as its increment is at least 2λ selects the minimum optimal cardinality. Through k_λ activations the direct query work is $O(k_\lambda^2)$ in addition to sorting and initialization, and the one extra cell evaluated after the last accepted activation has at most $k_\lambda + 1$ pairs and does not change this bound. A gap in any current cell still avoids every accepted cell, so the recovery argument applies to a stopped profile as well. A completed profile also allows selecting the penalty afterwards, by searching its breakpoints without any further transport solve. This concludes the proof.

Floating-point implementation. Our implementation uses the same cell routine for direct and precomputed queries. Built-in costs use compiled loops, while a general callable uses direct interval evaluation. For W_1, compensated prefix sums evaluate the physical costs and integer prefix sums evaluate the rank lengths. For higher integer powers, we start with direct queries and switch to the range tree when the next query would exceed a construction-work budget, which we estimate by summing $s \log s$ over the mixed-color internal nodes and the pair work inside the constant-size leaves, with weights that reflect the power and the higher cost of modular integer convolutions. The budget has a fixed floor and is bounded by $O(pN \log^2 N)$. We skip the estimate when the upper bound on the direct query work does not exceed the floor, and otherwise computing it takes $O(N)$ time, so in exact arithmetic a single restart with the range tables preserves the stated worst-case order. We expand the cross-block costs of integer powers around the node split, so that their terms are nonnegative, and short intervals use direct summation. We also evaluate an interval directly when its FFT contribution is small relative to the estimated roundoff. On a binary grid, the oracle uses modular integer convolutions after bounding its integer tables and their conversion to binary floating point, which preserves the primary ties of the unscaled power cost within those bounds, and rank lengths always use integer arrays. The theorem itself is a statement about exact arithmetic. We describe the numerical validation and the comparison rule with the experiments.

B Validation and Runtime Details

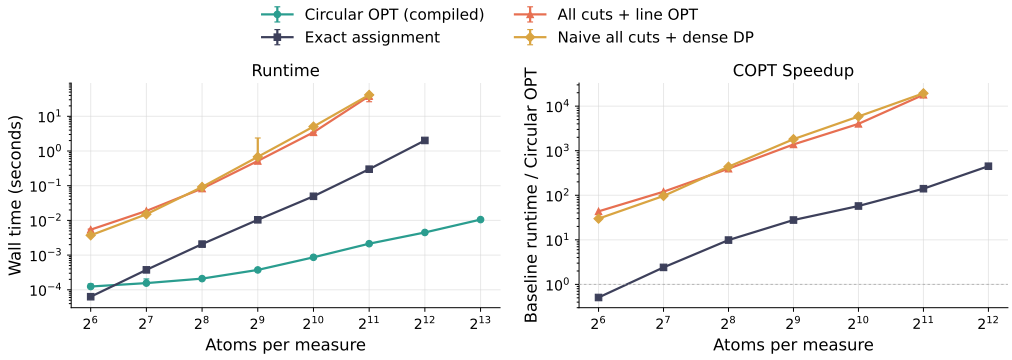


Figure 6: Single-penalty computation with local augmentations and penalty stopping. The all-cuts baselines solve either the alignment dynamic program or the line problem of Bai et al. (2023b) at every support gap. For balanced inputs with distinct support, the curves report median runtimes with interquartile ranges over five runs per method, including sorting, cost construction, and plan recovery.

We first verify exactness. At tractable sizes, we enumerate every support-gap cut and every tied-block split, and we compare the line optima with COPT and with a partial linear program. The implementations agree to floating-point precision for all the convex costs considered below. We then measure the single-penalty runtime with the linear-space implementation on one uniform random instance per size at $\lambda = 0.01$, on one Apple Silicon CPU thread with float64 arrays, after one untimed call of each method at each size. The timings include sorting, cost construction, optimization, and plan recovery, and the five repetitions measure the runtime variation on each input.

Figure 6 shows the comparison, in which the advantage over all-cuts line OPT reaches $18{,}000\times$ at $n = 2048$, where COPT is also $19{,}400\times$ faster than all-cuts dense dynamic programming.

The numerical implementation uses float64 for the physical costs and integers for the rank lengths. For built-in costs with an explicit multiplicative scale, we factor that scale out of both the cost and the penalty before comparing candidates, and the profile algorithm uses no absolute improvement threshold. The cross-block power expansions use nonnegative distances from the node split, short intervals use compensated direct summation, and when a binary grid and its coefficient bounds allow it, integer convolutions preserve the primary ties, while arbitrary floating inputs use FFT arithmetic. The numerical checks evaluate both the returned value and the cost of the recovered plan.

We then extend the measurements to larger and rectangular inputs, beyond the reach of direct cut enumeration. To compare the interval evaluators, we compute the full profile on uniform supports, on separated source and target intervals, on repeated grid locations, and on unequal sizes. Each setting uses three independent inputs and three interleaved repetitions, with one numerical-library thread and the same timing scope as above. The penalty only selects one plan from the computed profile and is not part of the profile computation. For the balanced inputs of Figure 2, the selection penalty is $10/n^2$ on uniform supports, 0.05 on separated supports, and 0.001 on the repeated grid. The adaptive evaluator keeps the direct queries when the preprocessing is unlikely to repay its cost. Its work budget estimates the mixed-color nodes of the range tree, since monochromatic nodes need no matching-cost tables, and includes the cost of the integer transforms on quantized inputs. The budget is bounded by the preprocessing order of Theorem 1, and at most one restart is needed after it is exceeded.

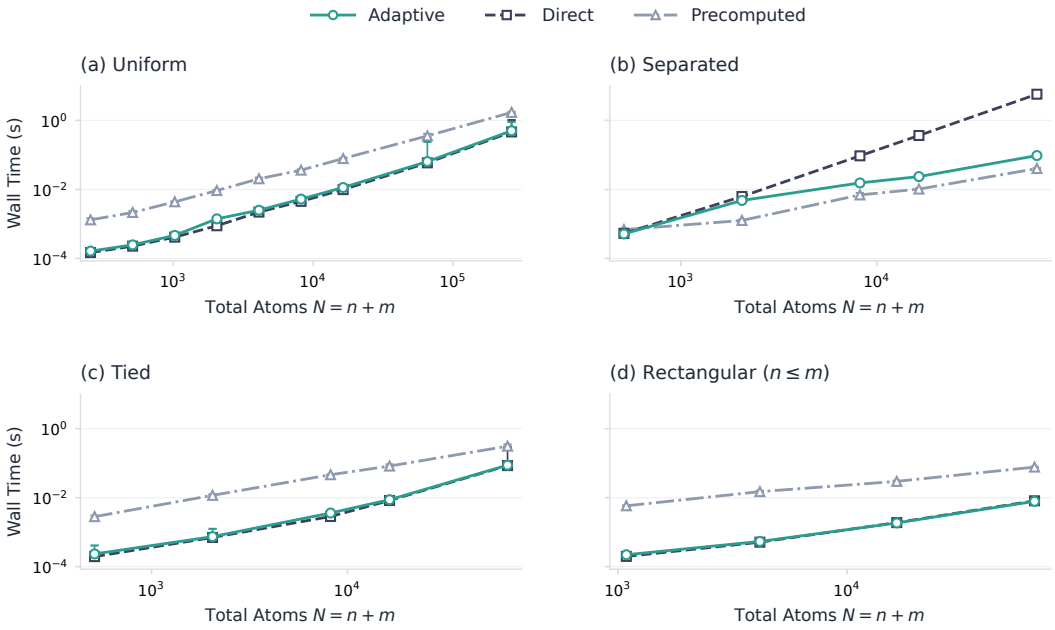


Figure 7: Direct, precomputed, and adaptive interval evaluation within the same cardinality-growth algorithm. Short cells favor direct evaluation, whereas separated supports repeatedly rematch a growing interval and benefit from preprocessing. The unequal-size panel fixes the source count at 64 while the target count increases. Medians and interquartile ranges over three inputs and three repetitions per setting.

At $n = m = 16{,}384$, the direct implementation uses 146 MiB of peak resident memory, against 196 MiB for the range oracle, 154 MiB for the linear-space sweep, and 407 MiB for the dense traceback, and at 65,536 atoms per measure the direct and precomputed evaluations use 177 and 417 MiB (Figure 8). These measurements include a common small-input compilation warmup and are medians over three fresh processes per setting. The range oracle trades additional storage for faster worst-case queries, while direct evaluation keeps linear space.

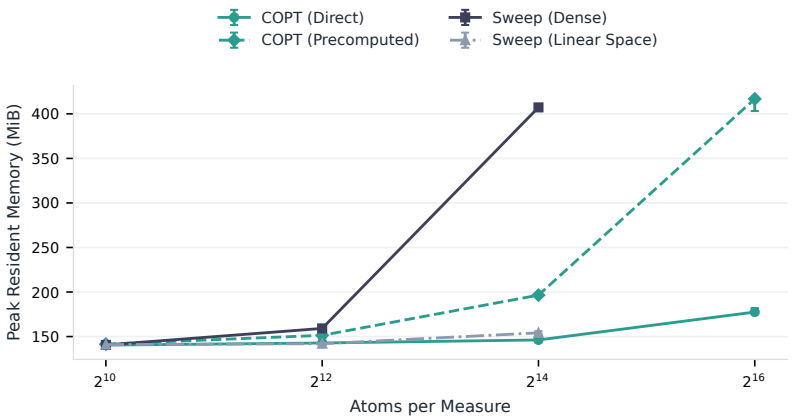


Figure 8: Peak resident memory of direct and precomputed profile evaluation, with the dense and linear-space cut sweeps as references. Each setting uses three fresh processes with the same compilation warmup.

We also evaluate penalty grids on the 324 query–database pairs formed by the 18 validation palettes of the fixed-composition Flowers study. Since this experiment only needs cost values, the repeated-profile reference stops at each requested penalty without recovering a plan, and the assignment reference builds one cost matrix per pair and reuses it across the grid. The quadratic-cost grids range from 2^{-15} to 2^{-4} on a logarithmic scale, the one-penalty case uses their geometric midpoint, and for the Huber cost with threshold 0.05 we scale this range by $h(1/2)/(1/2)^2$. For 64 quadratic-cost penalties, one profile per pair evaluates the grid in 0.032 seconds, against 1.39 seconds for repeated stopped solves and 0.619 seconds for cached-matrix assignment, and with the Huber cost the corresponding times are 0.030, 1.12, and 0.502 seconds (Figure 9). Both experiments use direct interval evaluation at these sizes, so the benefit does not come from the polynomial oracle. A single small-penalty solve can remain cheaper than the full profile, while denser grids share nearly all of their transport work.

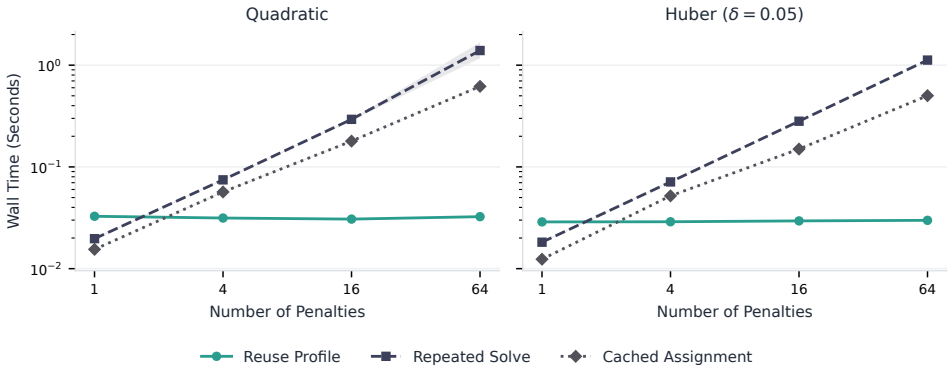


Figure 9: Evaluating penalty grids on the Flowers validation palettes with quadratic and Huber costs. The profile is computed once per pair, the repeated solves use penalty stopping, and assignment reuses the untruncated cost matrix. Medians over five repetitions, including the construction of the values needed by each method.

Table 1 distinguishes the effect of the cost from the effect of the geometry by crossing the quadratic and W_1 costs with the circular and fixed-cut geometries on the same empirical samples. We use one penalty grid for t^2 in both geometries and set $\lambda_1 = \sqrt{2\lambda_2}/2$ for W_1, which aligns the distance thresholds of individual edges, and where a tuning split is available we apply a common selection rule to all four variants. The line and circular objectives are evaluated on the original coordinates, with assignment checks for both costs. Across these tasks neither cost dominates. The quadratic distance leads on the origin-sensitive density and real-crop hue cases, whereas W_1 leads on Flowers

and on synthetic hue retrieval. The two circular costs recover every boundary-wrapped hue query and the four chroma scores coincide, so the circular geometry is what improves the wrap-around tasks, and COPT evaluates the circular quadratic cost without restarting a line solve at each cut.

Table 1: Top-1 retrieval accuracy with circular or fixed-cut geometry and quadratic or W_1 cost, evaluated on the same samples with aligned transport-radius grids.

Task (queries)	Circular t^2	Circular W_1	Fixed-cut t^2	Fixed-cut W_1
Flowers, fixed composition (54)	0.685	0.815	0.722	0.796
Flowers, boundary hue (14 of 54)	0.571	0.857	0.714	0.786
Chroma, KK templates (237)	0.363	0.363	0.363	0.363
Density, origin-sensitive (12)	1.000	0.917	0.667	0.667
Synthetic hue retrieval (90)	0.789	0.944	0.811	0.944
Boundary-wrapped hue (6)	1.000	1.000	0.167	0.167
Real-crop hue retrieval (4)	4/4	1/4	1/4	1/4

For the real-crop row, we use the per-case penalties of Figure 17. The Flowers entries use one validation-selected penalty per cost and geometry, and the origin-sensitive density row combines the two six-query variants of Figure 16 at their stated penalties. The chroma row uses the single-key cohort and the common numerical tie rule of Appendix I, with $\lambda_2 = 0.004$ and $\lambda_1 = \sqrt{0.008}/2$.

Using the same seeded instances as the cut-sweep validation, we evaluate the general-cost profile implementation against a scaled linear-programming oracle over the unit-capacity matching polytope. With $h(t) = t^2$, it reproduces the reference objective on 140 random instances including tied supports, and the returned plans attain the same value. We then test four power costs, a tolerance cost, a Huber cost, and an exponential cost. For each cost, the comparison comprises 60 random distinct, 60 tied-grid, 15 rectangular, and 8 adversarial instances with $n, m \leq 100$ and penalties from 10^{-5} to 3. All 1,001 LP comparisons agree, as do 40 brute-force tied instances for each cost, and additional integer-grid checks verify both components of the refined objective and the recovery at a common cut for every cardinality. As a control, the algorithm fails on $17/80$ instances for the concave cost $h(t) = \sqrt{t}$ and on $3/80$ for the capped cost $h(t) = \min\{t^2, 0.05\}$, with a relative excess as large as 10%, illustrating failures of cell rematching outside the convex class.

Two further checks address tied support. For the instance $X = (0.10, 0.25, 0.50, 0.50)$, $Y = (0.10, 0.25, 0.90, 0.90)$, and $\lambda = 1$, restricting the search to open support gaps returns 0.320, whereas the full split grid and COPT return the optimum 0.285, so the common virtual cut recovers the full split value without enumerating the splits. Virtual cuts affect the empirical comparisons only when large penalties allow transported arcs to cover most of the circle. On chroma samples with the quadratic cost, this first occurs at $\lambda \geq 0.032$, where the admissible radius spans about three pitch classes. There, ordinary cuts are suboptimal on 3.3% of the pairs, with a relative excess of up to 21%, and change one to four of the 30 top-1 retrievals. Over 720 chroma pairs, 231 circadian-activity pairs at three penalties, and 576 Oxford Flowers pairs per setting after quantizing the hue to 36 or 256 bins, the best ordinary cut is optimal for $|t|$ throughout and for t^2 at the validation-selected penalties. Small tied instances with $n, m \leq 8$ require a virtual split in 0.3–0.5% of the cases, again for the quadratic cost at large penalties, while the random tied grids with $n, m \leq 40$ always have an optimal ordinary cut.

C Relationship to PAWC

PAWC (Kolouri, 2026) computes the circular partial W_1 cost for every transported mass in $O(N \log N)$ time for distinct, uniformly weighted support. Under these assumptions and $h(t) = t$, our penalized value is the lower envelope of the PAWC profile,

$$\mathrm{COPT}^{\mathbb{S}^1}_{\lambda}(X, Y) = \min_k C^{\circ}_k + \lambda(n + m - 2k), \tag{C.1}$$

so that one PAWC run recovers our W_1 value for every λ. On 120 random distinct-support instances with $2 \leq n, m \leq 5$, the complete profiles of the public implementation agree with exhaustive partial matching to 10^{-9}, and our algorithm computes the same profile in this shared case.

We also compare the public full-profile calls on distinct uniform, separated, wrapped, alternating, and unequal-size inputs, including sorting and the compact recovery data. Across 231 paired inputs with five warmed repetitions on one CPU thread, the geometric mean of the PAWC-to-COPT runtime ratio is 1.08. The two implementations have comparable runtimes: for uniform inputs with $n = m =$ 131,072, the median times are 0.109 seconds for COPT and 0.110 seconds for PAWC.

Beyond this shared setting, the rank refinement of Section 4.3 handles repeated locations and flat regions of h, the interval oracle of Angrick et al. (2026) is applied on the balanced cells of the doubled circular order for integer powers, and general convex costs such as the Huber and tolerance costs use direct interval queries. For $p > 1$ with tied support, the common virtual cut can attain a smaller value than any ordinary cut, as the example of Section 4.3 and the chroma experiments at larger penalties show.

D Correspondence and Shape Matching

D.1 Matching Along the Profile

For each visible fraction $v \in \{0.5, 0.75, 1\}$, we draw 200 source measures with 64 independent uniform coordinates. A target contains a uniformly selected subset of $64v$ source atoms, perturbed by wrapped Gaussian noise of standard deviation 0.003, plus 32 additional uniform atoms, and the original source indices identify the genuine pairs. At cardinality k, the correct-pair F1 is $2r_k/(k + 64v)$, where r_k is the number of genuine pairs recovered by the quadratic-cost optimum. The curves of Figure 3 average the 200 instances, with pointwise intervals of 1.96 standard errors. The smaller illustration uses sixteen stratified source samples, twelve visible observations, and eight clutter atoms with the same noise model, and each of its plans is recovered from the profile at the displayed cardinality. The activation endpoints alone do not determine the matching partners.

D.2 Partial Shape Matching

MPEG-7 CE-Shape-1 contains twenty silhouettes in each of seventy classes (Latecki et al., 2000). We trace the largest outer contour at level 0.5, orient it counterclockwise, and resample it at 512 equally spaced arclength positions. A Fourier representation with 32 harmonics and a Gaussian spectral taper of width $1/256$ of the contour period smooths the pixel-scale corners. We evaluate its tangent analytically and obtain the outward normal by a quarter turn. Each database descriptor contains 128 independently stratified arclength samples. The query samples are drawn independently on a contiguous visible arc with a uniform starting position and rotation. The visibility $v \in \{1, 0.75, 0.5\}$ leaves $\mathrm{round}(128v)$ samples, and a fraction $c \in \{0, 0.2, 0.4\}$ of their normal directions is replaced by independent uniform directions. The corruption changes the normals used for matching, not the contour used to measure the registration error.

Fourteen randomly selected classes provide 280 validation shapes, and the remaining 56 classes provide 1,120 evaluation shapes. For each cost, the validation selects one transported fraction from $\{0.4, 0.5, \dots, 1\}$ by the mean contour-registration error over all nine conditions, keeping the larger fraction in case of a tie. All methods search $\Theta = \{j/72 : 0 \le j < 72\}$ in unit-circle coordinates. For a query X and a database shape Y, the partial score is

$$S_{p,\rho}(X, Y)^p = \frac{1}{k} \min_{\theta \in \Theta} C_k^{(p)}(X - \theta, Y), \qquad k = \mathrm{round}(\rho \min(n, m)), \tag{D.1}$$

where $C_k^{(p)}$ is the cardinality cost equation 5 with $h(t) = t^p$. COPT computes both the $p = 1$ and the $p = 2$ profiles. We minimize each mass over the rotations on its own, keeping the first grid index in case of equal computed costs, and since the minimum over rotations need not be a convex sequence in k, we do not apply monotone-marginal stopping to this outer search.

Full-mass matching takes $k = \min(n, m)$ at $p = 2$. Balanced W_2 instead assigns the probability weights $1/n$ and $1/m$ to the two descriptors. We evaluate its squared value by replicating the atoms to the common denominator $L = \mathrm{lcm}(n, m)$, which is at most 384, computing the full balanced circular cost with the periodic-shift solver of Delon et al. (2010), and dividing by L, and we check these values against assignment. The balanced and full-mass scores coincide when the input counts are equal, but not when the visibility reduces the query size. Repeated coordinates in the partial profiles are handled by the rank refinement of COPT, without any location perturbation.

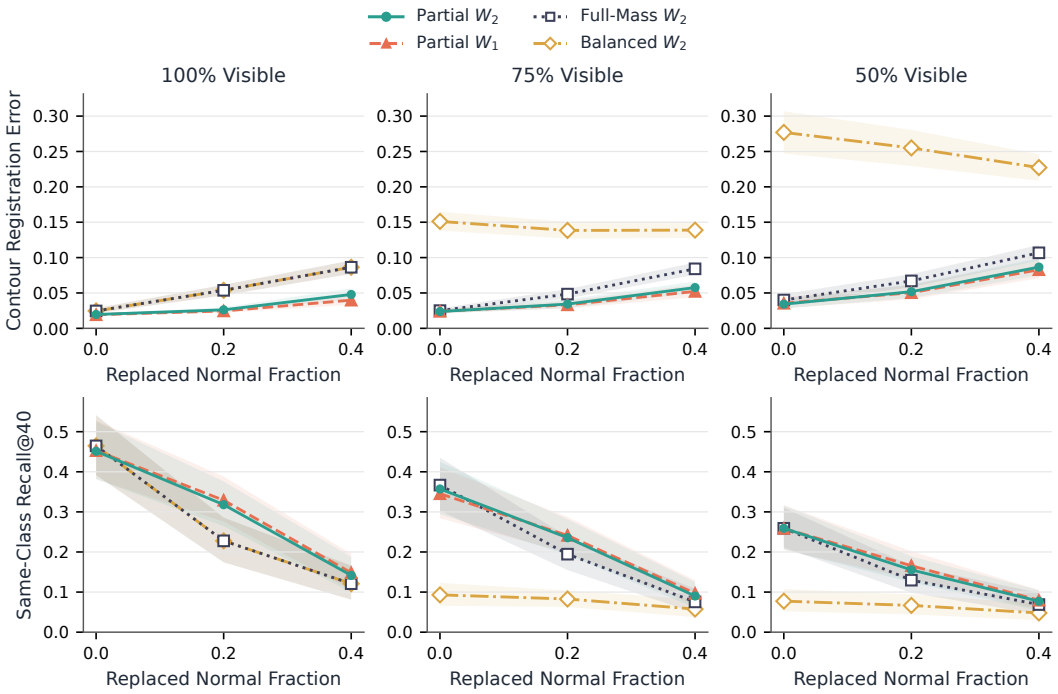


Figure 10: Rotation recovery and class retrieval over the visibility and normal-corruption grid. Each partial cost uses one transported fraction selected on validation. Shading marks 95% bootstrap intervals over the 56 evaluation classes. At full visibility, full-mass and balanced W_2 coincide. Open markers and distinct line styles distinguish their traces.

Registration compares a query with its own intact database contour. Since a symmetric shape can have several correct orientations, we measure the residual rotation geometrically. For a 256-point reference contour T, a true rotation α, and an estimate $\widehat{\alpha}$, let $Q = R_{\alpha-\widehat{\alpha}}T$ and let r_T be its root-mean-square radius about the rotation center. The normalized contour error is

$$E(T,Q) = \frac{1}{r_T}\left[\frac{1}{2|T|}\left(\sum_{x\in T}\min_{y\in Q}\|x-y\|^2 + \sum_{y\in Q}\min_{x\in T}\|x-y\|^2\right)\right]^{1/2}. \tag{D.2}$$

We also record the wrapped angular error and recover the selected partial plans to measure which genuine and replaced normals are retained. Retrieval uses three predetermined queries per evaluation class, excluding the original silhouette of each query from its database of 1,119 shapes. Recall@40 is the fraction of the nineteen remaining same-class shapes among the first forty results, and top-1 records whether the nearest shape has the class of the query. Scores tied within $10^{-12}\max(1,|s|)$ receive their expected credit under a uniform ordering of the tied group. Confidence intervals resample the 56 classes 10,000 times and keep the shapes, queries, methods, and corruption conditions of each class in one block.

The application timings use eight validation silhouettes, each compared with itself and with another validation shape at three visibility levels and 20% replaced normals. Each of these 48 comparisons returns the optimal rotation and score for seven transported fractions. The profile method computes all cardinalities at each of the 72 rotations, whereas assignment reuses the rotated cost matrix and solves one fixed-cardinality problem at each requested mass. Five interleaved repetitions follow compilation and a first execution, and we report the median of the per-comparison median times. Contour extraction and the registration error are outside this scoring task.

D.3 Fixed-Penalty Correspondence and Rotation Recovery

We also examine correspondence and rotation recovery when a signal crosses the line origin. The configurations are selected from a sweep over corruption rates, decoy offsets, noise scales, and penalties, using a score that combines the accuracy of COPT (Circular OPT in the figure legends) with its margin over the baselines. Figure 11 reports the recovered pairs and the rejection of clutter atoms as the corruption increases, and shows representative circular matchings. PAWL uses the squared line distance, and both PAWL and the fixed-mass partial LP transport the known visible count, which is $k = 80$ in the origin-sensitivity example. The retrieval studies select their parameters on their own validation splits.

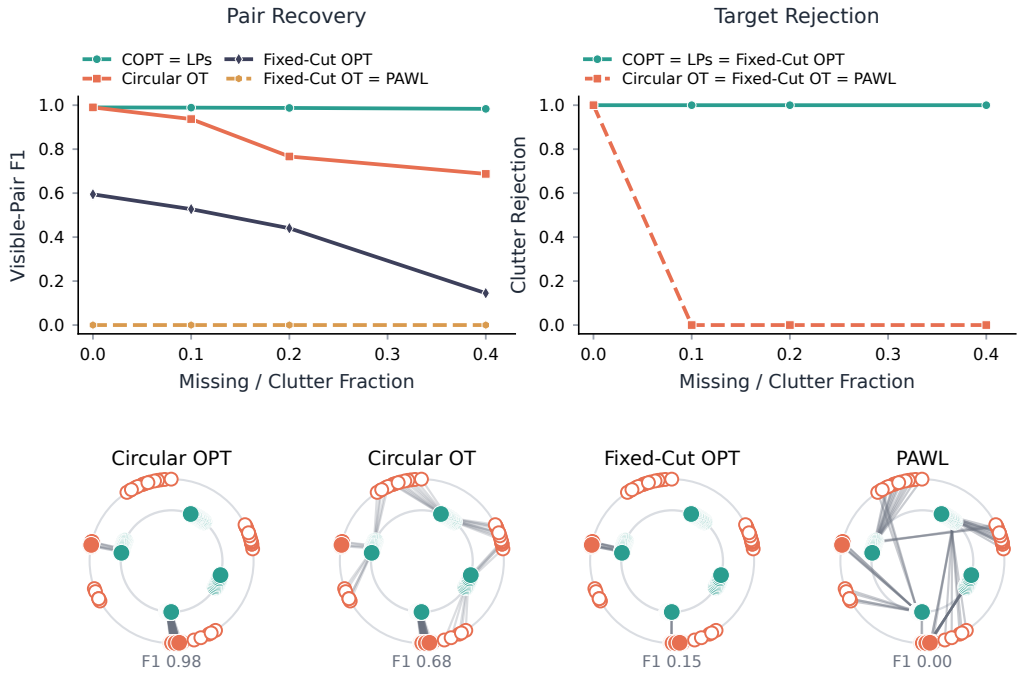


Figure 11: Active correspondence recovery. COPT keeps a high visible-pair F1 by leaving the clutter atoms unmatched, whereas balanced circular OT transports all targets and the line-based methods depend on the chosen origin. Methods with coincident curves share a legend entry, with LPs denoting the Lagrange and fixed-mass partial references.

For registration, we evaluate each method over the same rotation grid. We use asymmetric signals with one mode spanning the boundary, and we form the target by removing samples and adding nearby extra mass. COPT improves over the balanced circular methods whenever the extra mass changes the rotation score, while the fixed-cut partial methods solve the corresponding line problem after one chosen unwrapping (Figure 12).

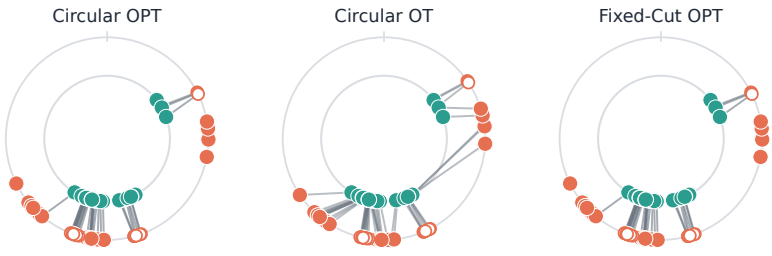


Figure 12: Rotation recovery with missing target samples and nearby extra target mass. Circular OPT keeps the estimate stable by leaving extra mass unmatched, while balanced baselines transport it.

E LINE-ORIGIN SENSITIVITY

Applying a line partial-transport solver to circular data requires a choice of unwrapping. To measure the dependence on this choice, we repeat the correspondence task while varying only the line origin. COPT is invariant to the origin by construction. Fixed-cut OPT and PAWL recover the active correspondence when the origin avoids the wrapped signal, but their F1 decreases when the cut passes through it. At its least favorable origin, fixed-cut OPT drops to an F1 of 0.47, and PAWL drops to 0.01 while matching every extra target sample, whereas COPT keeps an F1 of 1.00 for every origin (Figure 13).

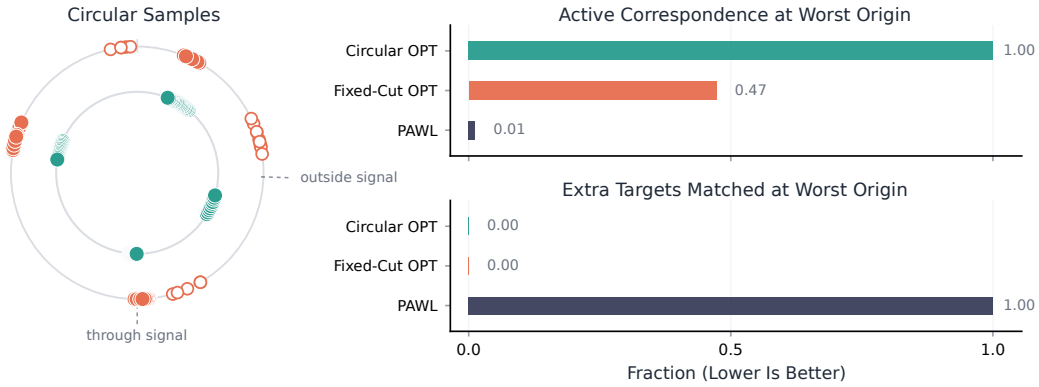


Figure 13: Sensitivity to the cut location. Left: circular samples and two cut locations. Right: worst-case active-pair F1 and fraction of clutter atoms matched at the worst origin over the full origin sweep.

F SYNTHETIC HUE RETRIEVAL

We also test finite hue samples with missing modes and with extra modes that wrap around the boundary. The task is to retrieve a clean target from a database whose other entries share only part of the wrapped hue support. We compute the circular and fixed-cut OPT and OT objectives exactly, as well as PAWL and the Lagrange and fixed-mass partial LP references, and we use LCOT as the balanced circular embedding baseline. On the constructed configuration, COPT retrieves the true clean target for all six queries with boundary-wrapped extra modes, while Circular OT, LCOT, fixed-cut OT, fixed-cut OPT, and PAWL retrieve it for at most one of the six (Figure 14). On the additional set of 90 queries with missing or extra hue modes (Figure 15, middle), fixed-cut OPT reaches a top-1 accuracy of 0.811 against 0.789 for COPT (Table 1), so that the boundary-wrapped examples show the clearest benefit of optimizing the cut.

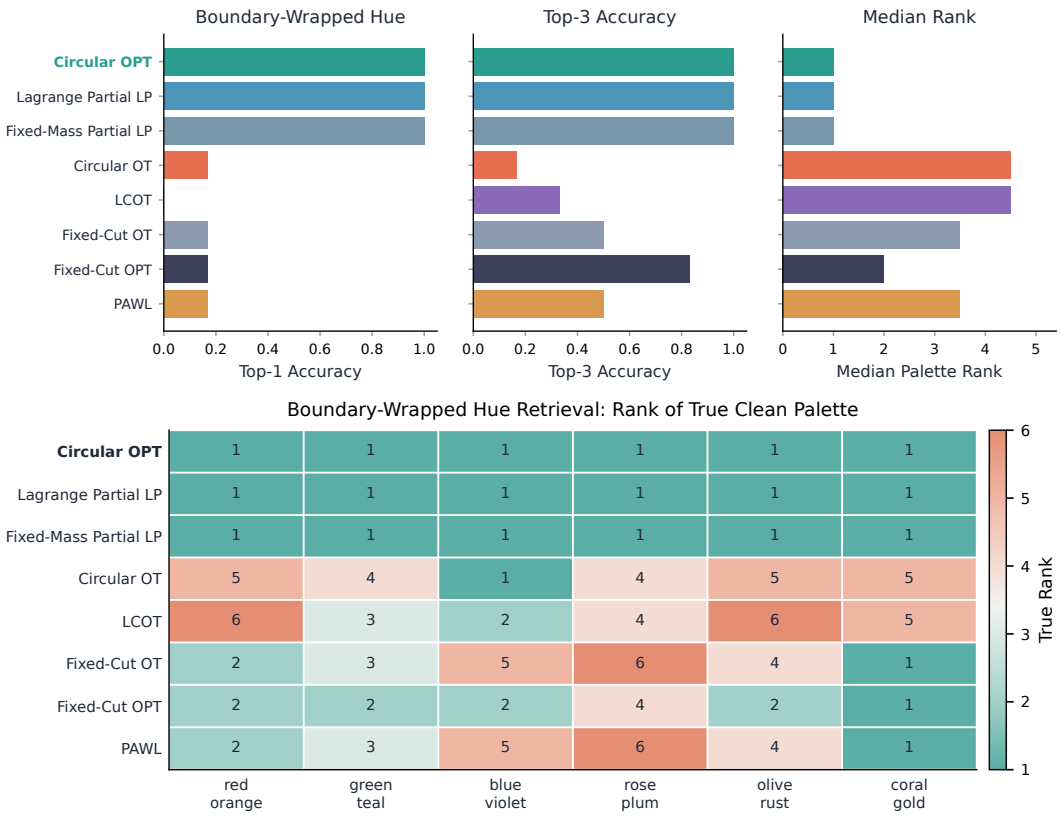


Figure 14: COPT and the Lagrange partial LP recover the clean palette in all six queries with boundary-wrapped extra modes, while each balanced, fixed-cut, LCOT, or PAWL baseline retrieves it for at most one query.

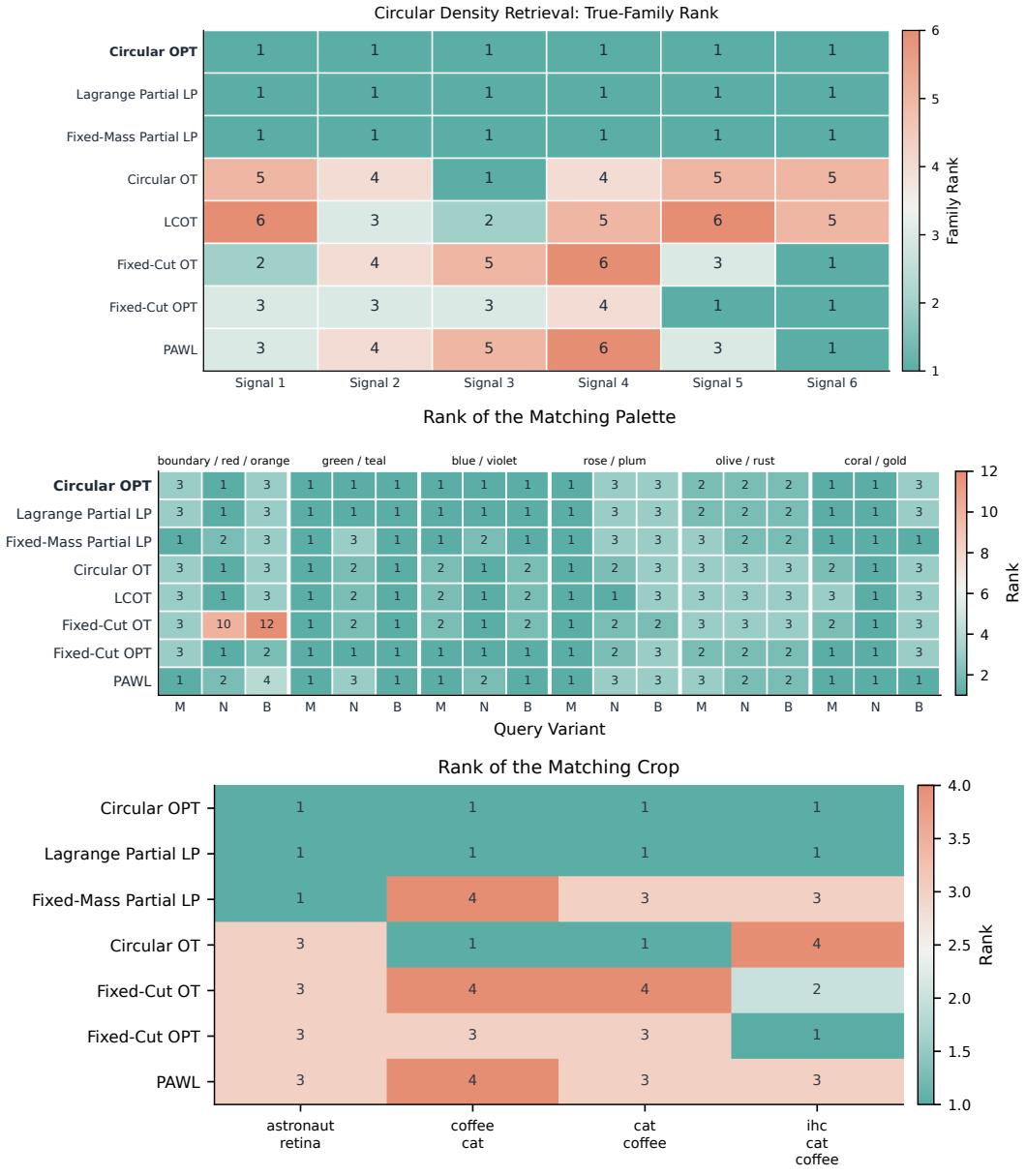


Figure 15: Retrieval rank heatmaps for the density and hue experiments. Top: rank of the true clean family in circular density retrieval. Middle: one replicate with missing modes (M), nearby added modes (N), and boundary distractors (B) for each clean palette. The aggregate synthetic hue results use all 90 queries. Bottom: rank of the matching crop in the real-crop hue retrieval cases, with per-case penalties between 0.0007 and 0.006.

G Circular Density and Image-Crop Retrieval

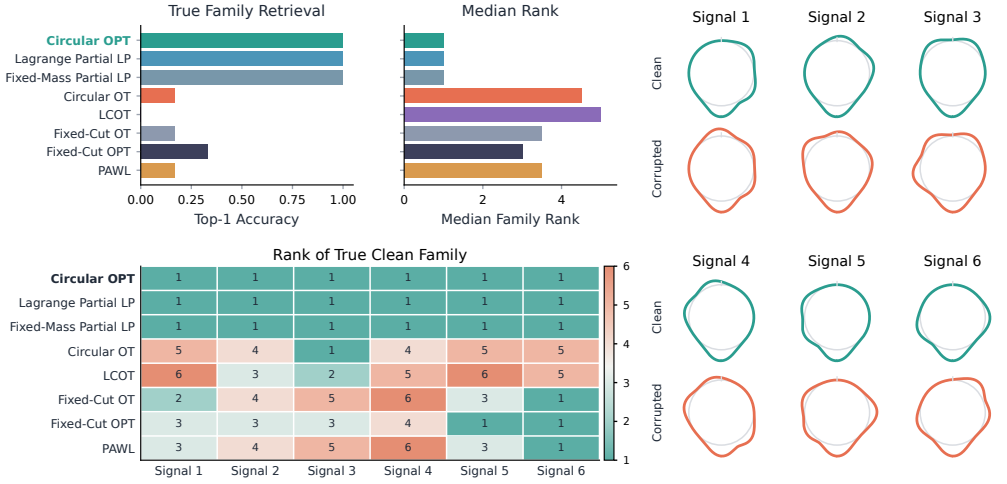


Figure 16: Circular density retrieval on a constructed configuration. Circular OPT and both dense partial LP references recover all six clean families, compared with two for the best fixed-cut partial baseline, one for balanced circular OT, and none for LCOT.

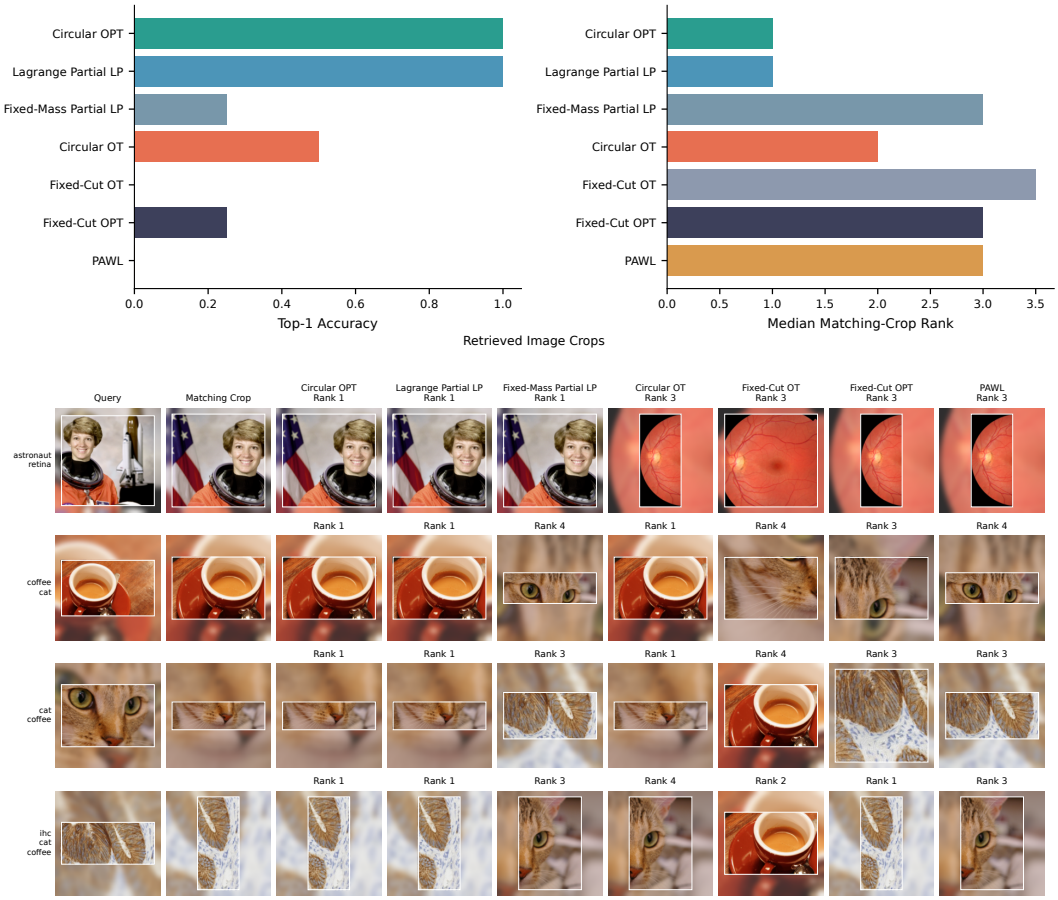


Figure 17: Hue retrieval for four image crops with distractor crops whose hue distributions overlap, using a separate penalty for each case. The summary compares retrieval accuracy and the rank of the matching crop, while the grid shows the top result from each method.

H Oxford Flowers Foreground Palette Retrieval

Oxford 102 Flowers is a fine-grained benchmark with 102 flower categories, 40–258 images per class, labels, splits, and segmentations, and with broad variation in scale, pose, illumination, and appearance (Nilsback & Zisserman, 2008). We decode the blue background marker in the provided segmentation images with a fixed per-channel tolerance of 20 levels for JPEG noise, and we define each database item as the hue palette of the segmented foreground. Saturation and value thresholds of 0.25 and 0.18 are fixed before any method selection. For each image, we deterministically draw disjoint 64-atom database and foreground-query samples, and the clutter sample consists of real background, leaf, and shadow pixels of the same image.

With a fixed seed, we select 12 classes for tuning and 54 disjoint classes for evaluation, with four images per class. The foreground visibility varies over 100, 70, and 40%, while the background clutter independently adds 0, 30, or 60% of the 64-atom base count. For each penalized method we select one penalty across all nine validation conditions and keep it for evaluation. For the fixed-mass method we select one transported fraction relative to the smaller atom count in the same way, while balanced circular OT assigns uniform probability weights within each measure and has no parameter. For the quadratic cost, COPT and fixed-cut OPT select $\lambda_2 = 10^{-5}$ and fixed-mass circular OT selects 50%. For circular W_1, we transform the quadratic-cost penalty grid through $\lambda_1 = \sqrt{2\lambda_2}/2$ and select $\lambda_1 \approx 0.007071$ on the same nine conditions. Our implementation evaluates COPT with the profile solver and the fixed-cut and fixed-mass baselines by assignment reductions, and balanced circular OT uses the circular one-dimensional solver. Tied distances receive fractional top-k credit, and confidence intervals resample the 54 evaluation classes in 4,000 bootstrap replicates.

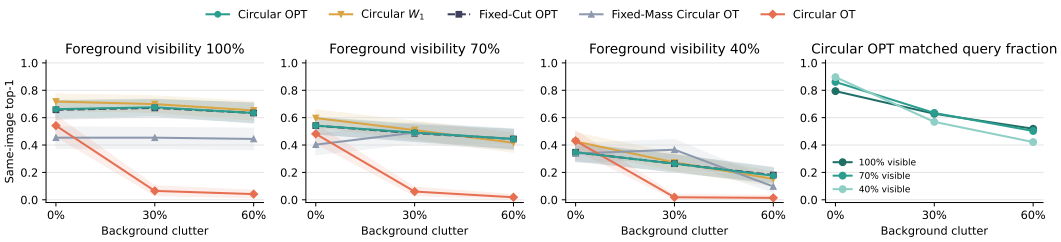


Figure 18: Foreground-palette retrieval on 216 Oxford Flowers images as foreground visibility and added background vary independently, both expressed relative to the original 64-atom foreground count. The first three panels compare retrieval accuracy with 95% bootstrap confidence intervals, and the last panel shows the fraction of query atoms transported by quadratic-cost COPT. Circular W_1 uses the linear geodesic cost, and each partial method uses one validation-selected parameter throughout.

With the full foreground, COPT improves on the globally tuned fixed-mass transport by 20.8 percentage points without clutter and by 19.0 points with 60% added background (95% confidence intervals 14.4–27.3 and 12.5–25.5). Its advantage is still 7.9 points when the foreground visibility falls to 40% at the same background level (95% confidence interval 2.8–13.0). Balanced circular OT is competitive on clean queries but degrades sharply as background is added. At full foreground visibility, COPT retains 79.4% of foreground atoms without clutter and 77.5% with 60% added background, while accepting 8.8% of the background atoms in the latter case. These rates are computed from the plans to the true clean palette. Fixed-cut OPT stays within one query of COPT in every condition, in agreement with the origin study, where a favorable cut recovers the circular plan.

Without added background, the linear cost improves retrieval by 5.6–7.9 percentage points at the three foreground levels. The two costs become closer as background is added. At the 60% background level, W_1 reaches 65.3% with the full foreground against 63.4% for the quadratic cost, while the quadratic cost leads by 2.8 and 2.3 points at 70- and 40% visibility, and the paired 95% intervals include zero in every condition with added background.

We also evaluate 54 images from 18 disjoint evaluation classes with queries that contain 70% foreground and 30% real background hue mass (Figure 19). We select all parameters on six disjoint tuning classes. The selected preprocessing uses 48 atoms, a saturation of at least 0.32, and a value of at least 0.24.

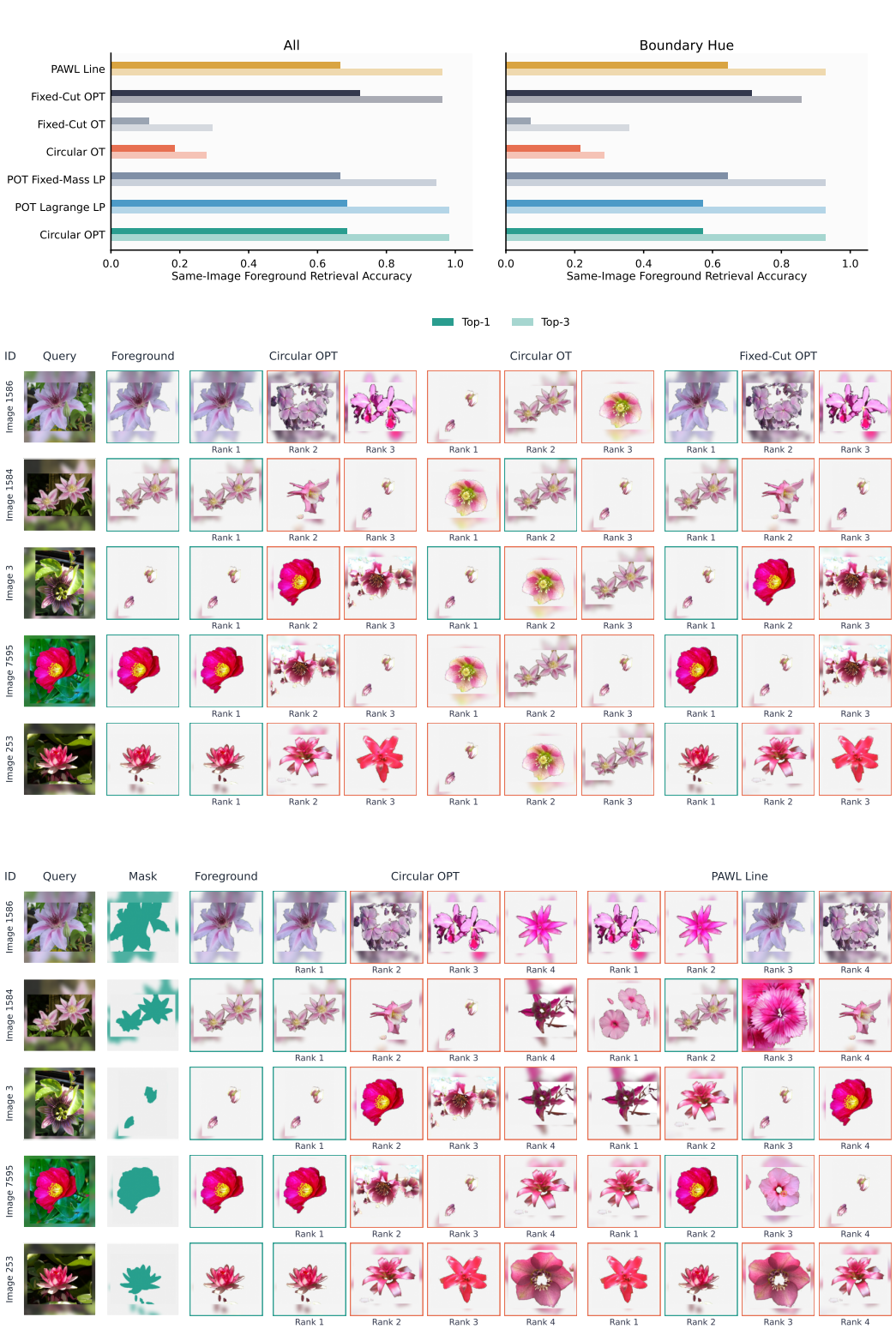


Figure 19: Oxford Flowers retrieval with queries containing 70% foreground and 30% background hues. COPT attains 68.5% top-1 accuracy over the 54 queries and 57.1% on the boundary-hue subset, against 18.5% and 11.1% overall for balanced circular OT and fixed-cut OT. The grids show the retrieved palettes, with teal borders marking the matching foreground image.

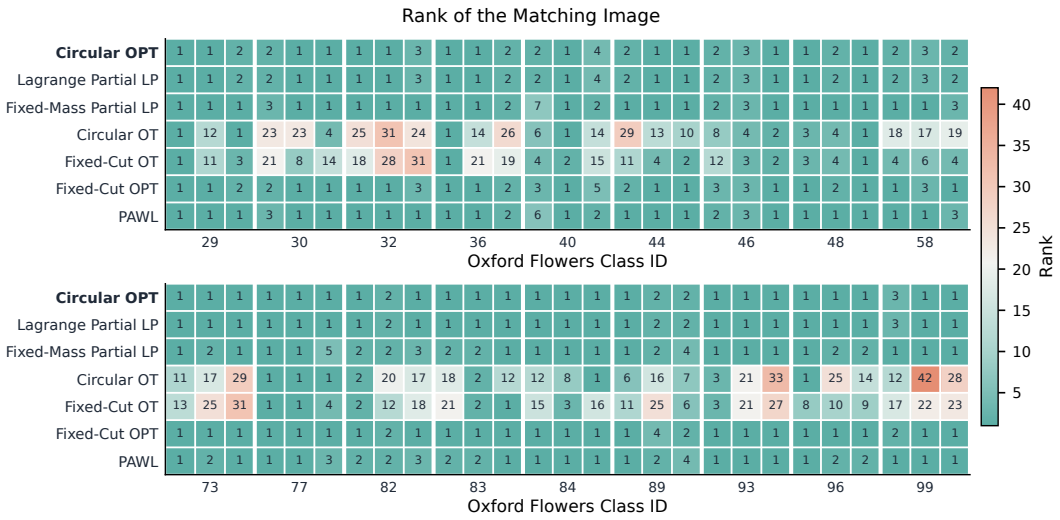


Figure 20: Rank of the matching image for every fixed-composition Oxford Flowers evaluation query. Columns are grouped as three images from each evaluation class, with the class IDs beneath each group.

I GiantSteps Chroma Key Retrieval

Pitch classes lie on a 12-point circle with B adjacent to C. We extract track-level chroma from the GiantSteps key dataset (Knees et al., 2015) with `librosa` (McFee et al., 2015) and convert each normalized profile into 72 atoms of unit mass by largest-remainder allocation. We keep the tracks with a single explicit major or minor annotation, excluding multikey, other-tonality, and unspecified labels. After requiring at least four tracks per key, the cohort contains 466 tracks, split into 229 for tuning and 237 for evaluation. The database contains the 24 major and minor Krumhansl–Kessler profiles (Krumhansl & Kessler, 1982), and the boundary subset contains the 60 evaluation tracks in the top quartile of raw chroma mass on C and B.

Validation selects $\lambda = 0.004$ for COPT, the Lagrange LP, and fixed-cut OPT, and 68 transported atoms for the fixed-mass circular and line methods. Sinkhorn UOT (Chizat et al., 2018) uses entropy and marginal penalty weights of 0.001 and 0.01. For every method, we group the distances within $10^{-12}(1 + |c_{\min}|)$ of the group minimum and order tied keys by their index.

COPT, the Lagrange LP, and fixed-cut OPT reach 36.3% top-1 accuracy overall and 41.7% on the boundary subset (Figure 21), and COPT and fixed-cut OPT identify the same correct queries. Fixed-mass circular and line transport reach 31.6% and 30.8%, against 20.3% and 15.6% for balanced circular and fixed-cut OT, while numerical Sinkhorn UOT reaches 57.0%. The selected transport radius barely exceeds one pitch class, so that the circular geometry acts mainly through the adjacency of C and B.

J Taxi Event-Time Localization

For the time-of-day study, we use the official NYC TLC yellow taxi trip records (New York City Taxi and Limousine Commission, 2026), restrict them to Manhattan pickup zones, and map the pickup time to the 24-hour circle. These records are natural circular event-time samples, in which late-night activity crosses midnight while holidays and city events add localized pickup mass that is absent from an ordinary weekday. We build weekday zone templates from non-event training days and evaluate test zone-days, using the COPT objective to score event days and the unmatched query atoms to localize the excess activity in time. COPT reaches an event-day AUROC of 0.890, close to the 0.889 of fixed-cut OPT, and the Lagrange LP agrees with COPT on the circular objective. The unmatched atoms follow the observed hourly excess of the query over the template (Figure 24).

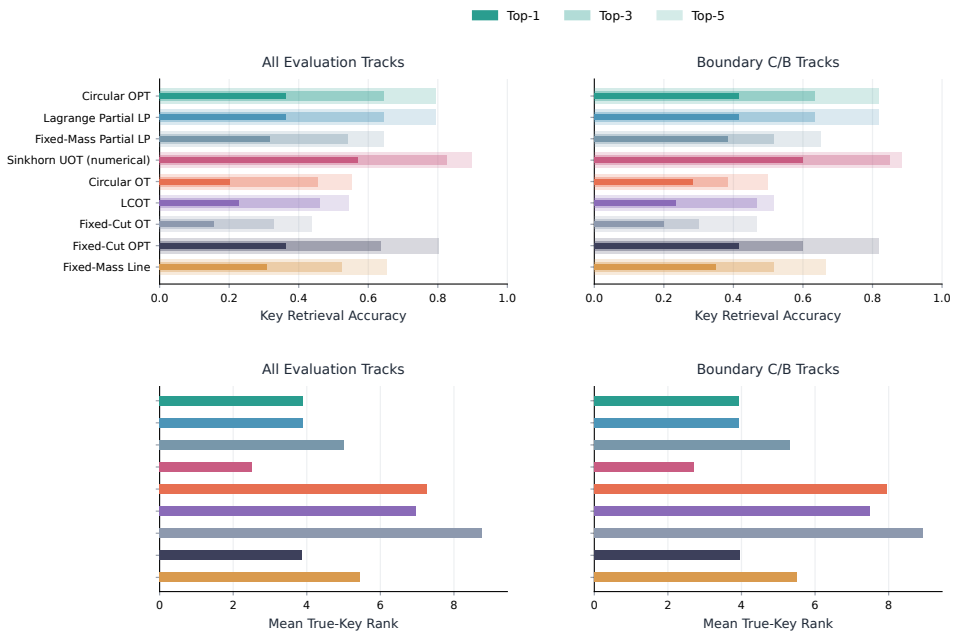


Figure 21: Key retrieval on GiantSteps with Krumhansl–Kessler templates. The plots report top-1, top-3, and top-5 accuracy and the mean true-key rank on the 237 evaluation tracks and on the 60-track boundary subset.

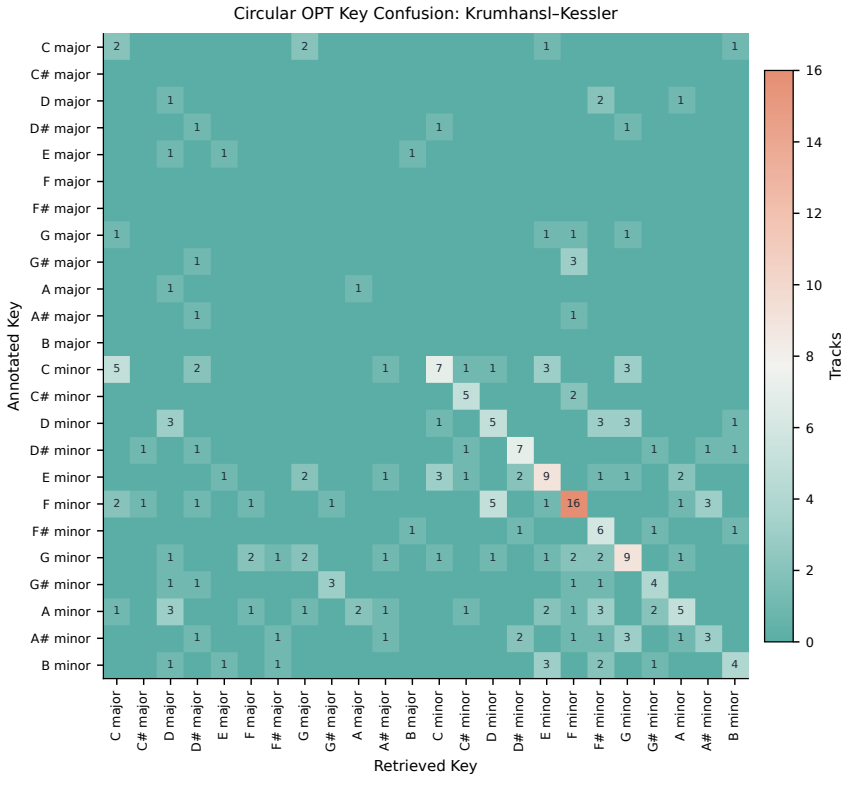


Figure 22: Key-confusion matrix of COPT on the GiantSteps evaluation split. Rows are the annotated keys and columns the retrieved keys, with the diagonal counting correct predictions. The largest counts occur in the frequent minor keys.

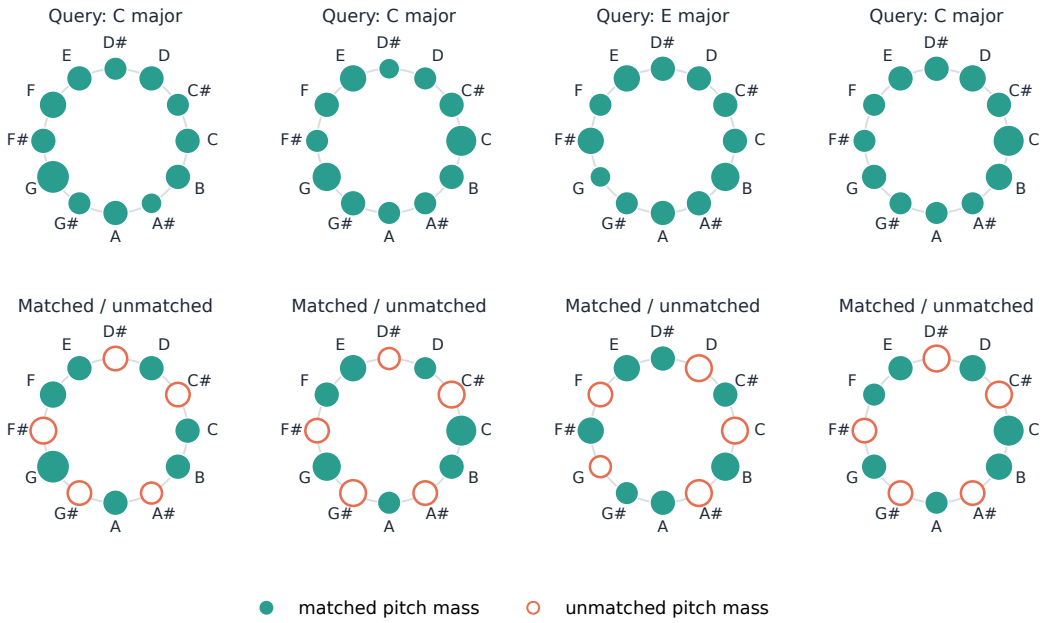


Figure 23: Unmatched pitch classes under sparse diatonic templates for four boundary-subset tracks, selected by the rank of the annotated key and then by objective value. Top: chroma samples. Bottom: matched and unmatched pitch mass. The displayed plans are one optimal matching. Other minimizers may leave other pitch classes unmatched.

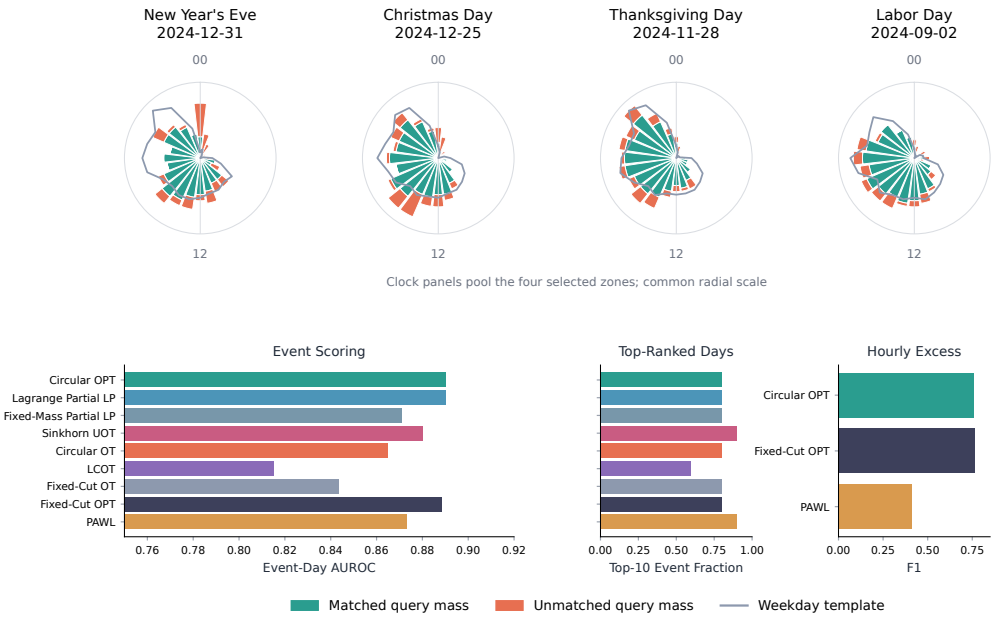


Figure 24: NYC taxi event-time localization and event scoring. Top: four zones pooled on a common radial scale, with the test event-day query mass split into matched and unmatched mass and the weekday template in gray. Bottom: event-day AUROC and top-10 event enrichment for all methods, and the hourly-excess localization F1 of the methods that return partial matchings, measured against the empirical hourly excess over the weekday template.

K Spherical and Linear COPT

K.1 Metric Properties of OPT

We first establish the base metric used in Proposition 2. For finite positive Borel measures on a compact metric space (Ω, d), $p \geq 1$, and $\lambda > 0$, we define

$$\mathrm{OPT}_{\lambda,p}(\mu,\nu) = \inf_{\substack{\gamma \in \mathcal{M}_+(\Omega^2) \\ \pi_{1\#}\gamma \leq \mu,\ \pi_{2\#}\gamma \leq \nu}} \left[\int_{\Omega^2} d(x,y)^p \, d\gamma(x,y) + \lambda\big(|\mu| + |\nu| - 2|\gamma|\big) \right], \tag{K.1}$$

where $|\cdot|$ denotes the total mass. On counting measures, equation K.1 is the pth power of the transport–transform metric with penalty $C = \lambda^{1/p}$ (Müller et al., 2020), on finite sets it equals the pth power of GOSPA with $\alpha = 2$ and cutoff $c = (2\lambda)^{1/p}$ (Rahmathullah et al., 2017), and for finitely supported measures $\mathrm{OPT}_{\lambda,p}^{1/p}$ coincides with the Kantorovich–Rubinstein distance of Heinemann et al. (2023) at $C = (2\lambda)^{1/p}$. Following the extended-space construction of Caffarelli & McCann (2010), which Bai et al. (2023b) use for discrete OPT, we adjoin an auxiliary point $\hat{\infty} \notin \Omega$ and define

$$\bar{d}(x,y) = \min\{d(x,y), (2\lambda)^{1/p}\}, \qquad \bar{d}(x,\hat{\infty}) = \bar{d}(\hat{\infty},x) = \lambda^{1/p}, \qquad \bar{d}(\hat{\infty},\hat{\infty}) = 0. \tag{K.2}$$

The function $\bar{d}$ is a metric on $\Omega \cup \{\hat{\infty}\}$, because truncating d at a constant preserves every triangle inside Ω, and a triangle through $\hat{\infty}$ follows from $(2\lambda)^{1/p} \leq 2\lambda^{1/p}$.

For finite positive measures μ and ν, we choose $M \geq |\mu| + |\nu|$ and augment both measures to the common mass M by

$$\bar{\mu} = \mu + (M - |\mu|)\delta_{\hat{\infty}}, \qquad \bar{\nu} = \nu + (M - |\nu|)\delta_{\hat{\infty}}. \tag{K.3}$$

The lower bound on M leaves enough auxiliary mass on both sides for every partial plan, including the empty one. Let $E = \{(x,y) : d(x,y)^p < 2\lambda\}$. Given a coupling of $\bar{\mu}$ and $\bar{\nu}$, we keep its real-to-real part on E and regard the complementary real mass as destroyed source mass and created target mass. Since a unit of capped real-to-real mass costs 2λ, which is exactly the sum of its deletion and creation penalties, this conversion preserves the objective. Conversely, given a partial plan, we first restrict it to E, which cannot increase its objective, and call the restricted plan γ. We then send $\mu - \pi_{1\#}\gamma$ to $\hat{\infty}$ and $\nu - \pi_{2\#}\gamma$ from $\hat{\infty}$ to the real space, whose masses are $|\mu| - |\gamma|$ and $|\nu| - |\gamma|$, and the remaining auxiliary-to-auxiliary mass is

$$M - |\mu| - |\nu| + |\gamma| \geq 0, \tag{K.4}$$

so that γ and these three auxiliary pieces form a coupling of $\bar{\mu}$ and $\bar{\nu}$. Writing $W_{p,\bar{d}}^p$ for the unnormalized transport cost between equal-mass finite measures, the two conversions prove

$$W_{p,\bar{d}}(\bar{\mu},\bar{\nu})^p = \mathrm{OPT}_{\lambda,p}(\mu,\nu). \tag{K.5}$$

Increasing M extends any coupling by zero-cost mass at $(\hat{\infty},\hat{\infty})$, while the conversion above maps every augmented coupling back to a partial plan with the same objective, so the value does not depend on M once the bound holds. For three finite measures we choose one M at least as large as the sum of their masses and apply the Wasserstein triangle inequality to the three augmented measures, and the definiteness of the OPT root follows from $\lambda > 0$ and the definiteness of d. We conclude that $\mathrm{OPT}_{\lambda,p}^{1/p}$ is a metric on finite positive measures.

K.2 Proof of Proposition 2

Great-circle slicing is one of several spherical constructions (Quellmalz et al., 2023; Tran et al., 2024; Liu et al., 2025). For continuous spherical densities, injectivity of the great-circle transform and metricity of spherical sliced Wasserstein were established by Liu et al. (2025, Corollary B.7 and Theorem D.3). We use the penalized circular metric and prove definiteness for finite positive measures.

We first verify that the integral in equation 13 is well defined. Our fixed value on $U^\top x = 0$ makes $(U,x) \mapsto P^U(x)$ Borel measurable, and integrating against a fixed finite measure makes $U \mapsto P_\#^U \mu$

measurable in the weak topology. By equation K.5, each circular COPT term is a Wasserstein pth power on a compact augmented circle and is continuous in both measure arguments. Therefore the integrand is measurable, and Fubini's theorem shows that the exceptional pairs have zero $\sigma \otimes \mu$ measure, so our convention does not change the integral against σ.

We now prove the pseudo-metric properties. The pushforwards are defined for every U, and equation K.5 makes the rooted circular term a metric, whose nonnegativity and symmetry pass through the integral. For three spherical measures μ, ν, and ξ, we apply the pointwise triangle inequality followed by Minkowski's integral inequality and obtain

$$\mathrm{SSCOPT}_{\lambda,p}(\mu,\xi) \le \mathrm{SSCOPT}_{\lambda,p}(\mu,\nu) + \mathrm{SSCOPT}_{\lambda,p}(\nu,\xi), \tag{K.6}$$

which proves that SSCOPT is a pseudo-metric. To prove the identity of indiscernibles, suppose that the integral vanishes. Since each circular term is a metric, $P^U_{\#}\mu = P^U_{\#}\nu$ for σ-almost every U. When $d = 2$, every frame is orthogonal and P^U is a circle isometry, so equality for any such frame implies $\mu = \nu$. We now treat $d \ge 3$.

For each fixed $x \in \mathbb{S}^{d-1}$, a uniformly chosen unit vector v satisfies $x \cdot v \neq 0$ almost surely. Fubini's theorem applied to $\mu+\nu$ therefore provides a unit vector v for which the equator $E_v = \{x : x \cdot v = 0\}$ has zero mass under both measures. Fix any unit $u \in v^\perp$ and let $U = [u, v]$. The pole set of P^U lies in E_v, so if $U_j \to U$, then $P^{U_j}(x) \to P^U(x)$ for $(\mu + \nu)$-almost every x. Testing against a continuous function on the circle and applying dominated convergence shows that the two projected measures depend weakly continuously on the frame at U. Since σ has full support, the full-measure set of frames on which the projections agree is dense. Approximating U by frames in this set gives $P^{[u,v]}_{\#}\mu = P^{[u,v]}_{\#}\nu$ for every unit $u \in v^\perp$.

On each open hemisphere $H_\pm = \{x : \pm x \cdot v > 0\}$, consider the gnomonic projection

$$G_v(x) = \frac{x}{x \cdot v} - v \in v^\perp. \tag{K.7}$$

Its restrictions to H_+ and H_- are homeomorphisms onto $v^\perp$, with respective inverses

$$z \longmapsto \frac{v + z}{\sqrt{1 + \|z\|^2}}, \qquad z \longmapsto -\frac{v + z}{\sqrt{1 + \|z\|^2}}. \tag{K.8}$$

Writing $P^{[u,v]}(x) = (q_1, q_2)$ in Cartesian circle coordinates, the sign of q_2 identifies the hemisphere and

$$\frac{q_1}{q_2} = \frac{x \cdot u}{x \cdot v} = u \cdot G_v(x). \tag{K.9}$$

Thus equality of the circular projections implies equality of every one-dimensional linear projection of $G_{v\#}(\mu|_{H_\pm})$ and $G_{v\#}(\nu|_{H_\pm})$. These finite measures have equal masses. After normalizing when this mass is nonzero, the Cramér–Wold theorem (Cramér & Wold, 1936) identifies the two measures on $v^\perp$, while the zero-mass case is immediate. Applying the inverse maps gives $\mu|_{H_\pm} = \nu|_{H_\pm}$. Since both measures assign zero mass to E_v, we conclude that $\mu = \nu$. This proves metricity on all finite positive Borel spherical measures. For $d = 3$, injectivity of the normalized semicircle transform on measures was established by Quellmalz et al. (2023, Theorem 4.10).

For the balanced reduction, suppose that $|\mu| = |\nu| = 1$ and $2\lambda \ge 2^{-p}$. The destroyed and created parts of any partial plan have equal mass, and we may couple them to complete the plan. Since the diameter of the unit circle is $1/2$, each added unit costs at most 2^{-p}, which is no more than its destruction and creation penalties, so some optimum transports all the mass on every slice and SSCOPT equals SSW_p. This is an equality of values, since at the cutoff the smallest-cardinality optimal selector need not return a full plan.

For an orthogonal matrix Q, direct substitution shows that $P^U(Qx) = P^{Q^\top U}(x)$, and the invariance of σ under $U \mapsto Q^\top U$ preserves the integral, which proves the common-rotation invariance. As for the runtime, one slice projection costs $O(dN)$, and the direct profile solver costs $O(N \log N + \min(n, m)^2)$ including sorting and recovery, so summing over L slices gives $O(L[dN + N \log N + \min(n, m)^2])$ time. For a fixed positive integer p, the interval oracle replaces the solve term by $O(pN \log^2 N)$. For a single penalty, early stopping reduces the direct bound to $O(L[dN + N \log N] + \sum_{\ell=1}^{L} k^2_{\lambda,\ell})$, where $k_{\lambda,\ell}$ is the cardinality retained on slice ℓ. These bounds

hold for atoms of unit mass, whereas the metric statements above are properties of the functional and also hold for unequal positive atom weights.

Finally, for $L \geq 1$ we consider the empirical estimate

$$\widehat{\mathrm{SSCOPT}}^{p}_{\lambda,p} = \frac{1}{L}\sum_{\ell=1}^{L} \mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(P^{U_\ell}_{\#}\mu, P^{U_\ell}_{\#}\nu). \tag{K.10}$$

Our projection convention defines this estimator for every deterministic family of slices, and samples from σ avoid the exceptional set almost surely. Since the empty plan has value $\lambda(|\mu| + |\nu|)$, each summand lies in $[0, \lambda(|\mu| + |\nu|)]$, and Hoeffding's inequality controls the error of the estimated pth power. The rooted finite-slice quantity is a pseudo-metric and is invariant under a joint rotation of the data and the slices. If the slices stay fixed, exact common-rotation invariance need not hold, but the Monte Carlo error vanishes as the estimate approaches the integral against σ.

K.3 Linear COPT Embedding

For $p \geq 1$ and $\lambda > 0$, we fix a reference $R = \sum_{i=1}^{r} \delta_{r_i}$ and list its representatives in $[0, 1)$ in nondecreasing order. We order each target multiset in the same way and apply Algorithm 1, placing the reference copies before the target copies at coincident locations, resolving equal refined increments by the ordered endpoint ranks, and letting a merged cell inherit the free gap of its left neighbor. For a penalty λ, we select the smallest optimal cardinality and recover the plan at the stored common cut. Identical copies occupy consecutive entries, so this selector depends only on the two empirical measures and not on the enumeration of their atoms. For an empirical target Y whose atoms have unit mass, we record an endpoint $e_i(Y) \in \mathbb{S}^1$ in each transported reference entry and the auxiliary point $\hat{\infty}$ in the others, we collect the target atoms omitted from the plan in η_Y, and we write $q_Y = |\eta_Y|$. We call $\Psi_R(Y) = (e_1(Y), \ldots, e_r(Y), \eta_Y)$ the Linear Circular Optimal Partial Transport embedding of Y, and $\Phi_R(Y) = (e_1(Y), \ldots, e_r(Y), q_Y)$ its compressed form.

We equip each entry with the metric $d_\lambda(a, b)^p = \min\{\rho_{\mathbb{S}^1}(a, b)^p, 2\lambda\}$, $d_\lambda(a, \hat{\infty})^p = d_\lambda(\hat{\infty}, a)^p = \lambda$, and $d_\lambda(\hat{\infty}, \hat{\infty}) = 0$, and we define

$$A^R(Y, Z)^p = \sum_{i=1}^{r} d_\lambda(e_i(Y), e_i(Z))^p + \lambda(q_Y + q_Z), \tag{K.11}$$

$$D^R_{\mathrm{geo}}(Y, Z)^p = \sum_{i=1}^{r} d_\lambda(e_i(Y), e_i(Z))^p + \mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(\eta_Y, \eta_Z). \tag{K.12}$$

The fixed-size form A^R uses $O(r)$ storage and comparison time and is exact against R, but distinct created masses may share one compressed embedding and $A^R(Y, Y)^p = 2\lambda q_Y$, so A^R is not a metric when $q_Y > 0$. Retaining the created mass requires $O(r + q_Y)$ storage.

Proposition 3. *D^R_{geo} is a metric on empirical circular measures whose atoms have unit mass, and*

$$\mathrm{COPT}^{\mathbb{S}^1}_{\lambda,p}(Y, Z) \leq D^R_{\mathrm{geo}}(Y, Z)^p \leq A^R(Y, Z)^p. \tag{K.13}$$

Both inequalities are equalities when one of the two targets is R.

K.4 Proof of Proposition 3

As in LCOT, the word "linear" refers to the linearization at R. The entries remain circle-valued and do not form a Euclidean vector space (Díaz Martín et al., 2024). The endpoints, the transported reference mass, and the created mass are the three components of the LOPT embedding (Bai et al., 2023a), of which Φ_R retains only the amount of created mass.

The metric d_λ is the restriction of $\bar{d}$ above to the circle. Taking the pth root in equation K.12 produces the ℓ_p product of the entry metrics and of the COPT metric between created masses of equation K.5. Our canonical plan rule makes the embedding single-valued, and every target satisfies the reconstruction identity

$$Y = \sum_{i : e_i(Y) \neq \hat{\infty}} \delta_{e_i(Y)} + \eta_Y. \tag{K.14}$$

Suppose that $D^R_{\mathrm{geo}}(Y,Z) = 0$. Then every entry distance vanishes, so $e_i(Y) = e_i(Z)$ for all i, including the entries equal to $\hat{\infty}$, and the created-mass term vanishes as well, so $\eta_Y = \eta_Z$. By equation K.14, $Y = Z$. With the product metric properties, this proves that D^R_{geo} is a metric.

To prove the first upper bound, we construct a target-to-target partial plan. We pair the active endpoints that occupy the same reference entry whenever their direct cost is below 2λ, we leave both endpoints unmatched when the cap is attained, we leave one active endpoint unmatched when the other entry contains $\hat{\infty}$, and we use an optimal plan between the created masses η_Y and η_Z. This plan is feasible and its objective is exactly $D^R_{\mathrm{geo}}(Y,Z)^p$, so the optimal COPT value cannot exceed it.

We next prove the exactness at the reference. The embedding of R uses a zero-cost full plan, so $\eta_R = 0$ and every entry contains its reference location, up to exchangeable ties. An optimal reference-to-target edge has cost at most 2λ, since removing an edge of larger cost would strictly lower the objective. The occupied entries retain these edge costs, the entries containing $\hat{\infty}$ charge λ for the unmatched reference atoms, and $\mathrm{COPT}_{\lambda,p}(0,\eta_Y) = \lambda|\eta_Y|$ charges the omitted target atoms, so the sum of these terms is exactly the reference-to-target objective. Hence the first inequality is an equality when one target is R.

Finally, the empty plan between the two created masses has cost $\lambda(q_Y + q_Z)$, which proves $D^R_{\mathrm{geo}}(Y,Z)^p \le A^R(Y,Z)^p$, and at the reference the same computation is exact because $\eta_R = 0$. This concludes the proof.

K.5 Stability of the Reference Embedding

The metric and upper-bound guarantees do not imply continuity with respect to COPT. To see this, take $R = \delta_0$, $p = 2$, $\lambda = 0.1$, $a = 0.2$, and $0 < \varepsilon < 0.1$, and set

$$Y_\varepsilon = \delta_{a-\varepsilon} + \delta_{1-a}, \qquad Z_\varepsilon = \delta_a + \delta_{1-a+\varepsilon}. \tag{K.15}$$

The unique reference match uses $a - \varepsilon$ for Y_ε and $1 - a + \varepsilon$ for Z_ε, so the occupied entry and the created mass exchange sides, and a direct computation shows

$$\mathrm{COPT}^{\mathbb{S}^1}_{\lambda,2}(Y_\varepsilon, Z_\varepsilon) = 2\varepsilon^2, \tag{K.16}$$

$$D^R_{\mathrm{geo}}(Y_\varepsilon, Z_\varepsilon)^2 = (2a - 2\varepsilon)^2 + (2a)^2 \longrightarrow 0.32. \tag{K.17}$$

Neither continuity nor a uniform multiplicative approximation bound holds for arbitrary targets, then, and no deterministic tie rule can remove this example, because both reference matches are unique for positive ε. The separated regime below avoids such changes of assignment and allows exact interpolation and barycentric synthesis.

K.6 Barycenters and Interpolation

We set $p = 2$. Let $Y_1, \dots, Y_M$ have nonnegative weights $w_1, \dots, w_M$ summing to one, and assume that no LCOPT embedding has created mass. For each reference entry, choose

$$\bar{e}_i \in \arg\min_{z \in \mathbb{S}^1 \cup \{\hat{\infty}\}} \sum_{j=1}^{M} w_j d_\lambda(z, e_i(Y_j))^2. \tag{K.18}$$

Proposition 4 (Realizable LCOPT barycenter)**.** *If a target $\bar{Y}$ has embedding $(\bar{e}_1, \dots, \bar{e}_r, 0)$ under the same deterministic reference encoder, then $\bar{Y}$ is a Fréchet barycenter of $Y_1, \dots, Y_M$ under D^R_{geo}.*

Proof. Let Y be an arbitrary target. Since no input has created mass,

$$\sum_{j=1}^{M} w_j D^R_{\mathrm{geo}}(Y, Y_j)^2 = \sum_{i=1}^{r}\sum_{j=1}^{M} w_j d_\lambda(e_i(Y), e_i(Y_j))^2 + \mathrm{COPT}^{\mathbb{S}^1}_{\lambda,2}(\eta_Y, 0) \tag{K.19}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{M} w_j d_\lambda(e_i(Y), e_i(Y_j))^2 + \lambda|\eta_Y|. \tag{K.20}$$

By equation K.18, each entry term is bounded from below by its minimum, and the created-mass term is nonnegative, whereas the stated embedding attains every entry minimum with no created mass. Therefore $\bar{Y}$ globally minimizes the Fréchet objective, which proves the proposition. □

An arbitrary collection of entry means need not lie in the range of the encoder, and we now identify a regime in which it does. Let $s = \min_{i \neq k} \rho_{\mathbb{S}^1}(r_i, r_k)$, with $s = \infty$ when $r \leq 1$, and let $R_\lambda = \sqrt{2\lambda}$. Suppose that every input contains at most one atom in each circular ball $B(r_i, \delta)$ and no atom outside these balls, where

$$0 < \delta < \tfrac{1}{4}, \qquad 2\delta < R_\lambda, \qquad 2R_\lambda + 2\delta < s. \tag{K.21}$$

Under these conditions, every within-entry cost is below 2λ and every cross-entry cost exceeds 2λ, so the reference plan assigns every atom to its unique nearby entry and creates no mass.

For entry i, let a_i be the total weight of the inputs containing that entry. If $a_i = 0$, we set $v_i = 0$ and omit the entry. Otherwise, let m_i be the weighted intrinsic mean of the present endpoints inside $B(r_i, \delta)$, and set

$$v_i = \sum_{j : e_i(Y_j) \neq \hat{\infty}} w_j \rho_{\mathbb{S}^1}(m_i, e_i(Y_j))^2. \tag{K.22}$$

To see that the cap does not change this mean, consider a nonempty subset of the present endpoints with positive weight. Their intrinsic mean lies on the same arc, since $2\delta < 1/2$, and if the omitted endpoints have total weight $b > 0$, adding them at that mean increases the quadratic cost by at most $(2\delta)^2 b$, which is less than their capped cost $2\lambda b$, so retaining every endpoint is optimal, and if the subset is empty the same conclusion follows by placing the atom at m_i. The best finite state has value $\lambda(1 - a_i) + v_i$, while the auxiliary state has value λa_i, and we retain entry i when

$$v_i < \lambda(2a_i - 1), \tag{K.23}$$

either state being optimal at equality. For coincident present endpoints this rule reduces to weighted majority retention.

Proposition 5 (Agreement with COPT)**.** *Under equation K.21, the target formed from the retained means m_i is a barycenter under D^R_{geo} and a global Fréchet barycenter under the metric $(\mathrm{COPT}^{\mathbb{S}^1}_{\lambda,2})^{1/2}$ among finite unit-atom measures.*

Proof. We first decompose the COPT Fréchet objective by entry. Under equation K.21, the transport neighborhoods of distinct entries are disjoint, so a candidate atom can have a cost below 2λ to the endpoints of at most one entry, and an atom outside every neighborhood only adds a destruction penalty and may be removed. The objective then decomposes over the protected entries.

We now fix one entry and let a candidate place q atoms in its neighborhood. With no candidate atom, the entry costs λa_i, and with one atom its minimum cost is $\lambda(1-a_i)+v_i$. For $q \geq 2$, the unmatched candidate mass contributes at least $\lambda(q - a_i)$, whereas the present endpoints have diameter below R_λ, so their weighted variance satisfies $v_i \leq (2\delta)^2/4 < \lambda/2$, and one atom is strictly cheaper than two or more. Comparing the zero- and one-atom states recovers equation K.23, and the minimizing state is the LCOPT coordinate construction of Proposition 4. Applying this argument to every entry proves both barycenter claims. □

If two targets satisfying equation K.21 occupy the same entries, we move each endpoint along its unique shortest circular arc. The intermediate targets remain in their respective balls and satisfy

$$D^R_{\mathrm{geo}}(Y_s, Y_t) = |t - s| D^R_{\mathrm{geo}}(Y_0, Y_1), \tag{K.24}$$

and the same identity holds for the rooted metric $(\mathrm{COPT}^{\mathbb{S}^1}_{\lambda,2})^{1/2}$ in this regime. If the active set changes, the path passes through the auxiliary state and does not define a continuous geodesic for the unit-atom metric.

When the inputs have created mass, the coordinatewise minimization also requires a COPT barycenter of the created masses, and the embedding obtained in this way must still be realizable. Likewise, independently averaged slice-wise spherical embeddings need not be realizable as the projections of one spherical measure.

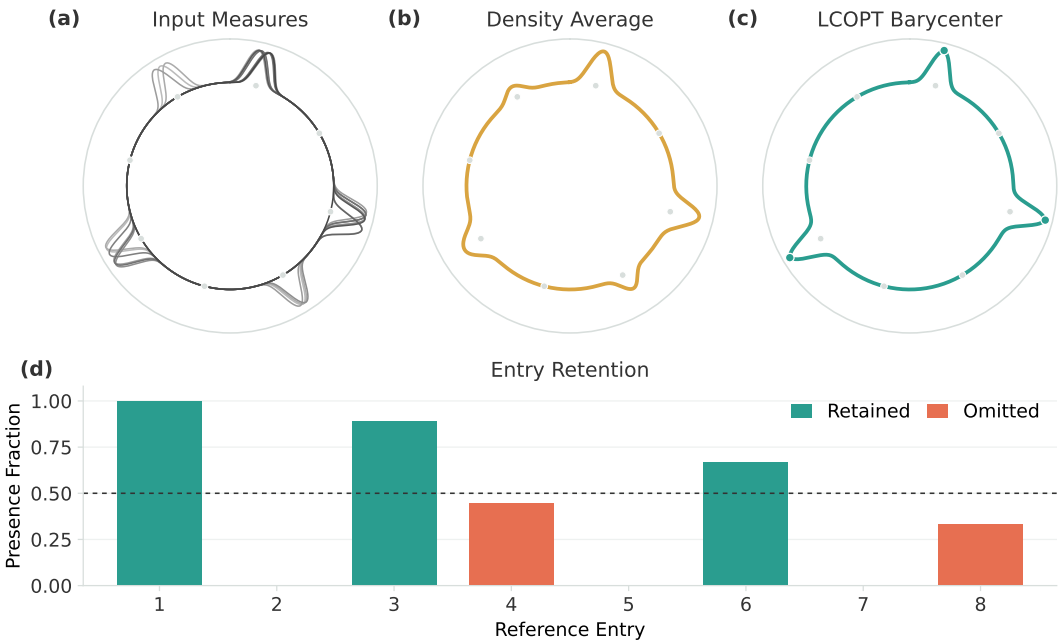


Figure 25: Barycenter of nine input measures that share three recurrent modes and contain two minority modes, with gray points marking the reference entries. The curves are kernel-smoothed displays of the atomic measures. The density average retains every observed mode, while LCOPT keeps the three entries satisfying equation K.23. Bar colors indicate whether each entry is retained under this rule, and the dashed line marks the threshold for zero within-entry variance. The parameters satisfy equation K.21, so the LCOPT output is also a COPT barycenter.

K.7 Linear COPT on the Sphere

We fix slices $U_1, \dots, U_L$ and one circular reference R_ℓ for each slice. Writing $Y_\ell = P_\#^{U_\ell} Y$, we define

$$\widehat{A}_{\mathcal{R},L}(Y,Z)^p = \frac{1}{L}\sum_{\ell=1}^{L} A^{R_\ell}(Y_\ell, Z_\ell)^p, \tag{K.25}$$

$$\widehat{D}_{\mathcal{R},L}(Y,Z)^p = \frac{1}{L}\sum_{\ell=1}^{L} D_{\mathrm{geo}}^{R_\ell}(Y_\ell, Z_\ell)^p. \tag{K.26}$$

The full slice-wise LCOPT embedding collects $\Psi_{R_\ell}(Y_\ell)$ over the L slices, and its compressed fixed-size embedding concatenates $\Phi_{R_\ell}(Y_\ell)$ and uses $\widehat{A}_{\mathcal{R},L}$ as a discrepancy. The refined form is a finite-slice pseudo-metric on empirical spherical measures and becomes a metric on any class for which the measure map $Y \mapsto (P_\#^{U_\ell} Y)_{\ell=1}^L$ is injective.

For a database of D distributions with n atoms each, $L \geq 1$ slices, and references with $r \geq 1$ atoms, we compute the fixed-size LCOPT in

$$O\big(DL[dn + T_{\mathrm{COPT}}(r,n)] + D^2 Lr\big) \tag{K.27}$$

time, where $T_{\mathrm{COPT}}(r,n)$ is the time of one reference-to-target solve, including sorting and plan recovery. The direct bound is $O((r+n)\log(r+n) + \min(r,n)^2)$, the accelerated bound for a fixed positive integer p is $O(p(r+n)\log^2(r+n))$, and early stopping replaces the squared minimum by the squared retained cardinality in each direct solve. Using the same solver for the created-mass refinement adds $\sum_{a<b}\sum_{\ell=1}^{L} T_{\mathrm{COPT}}(q_{a,\ell}, q_{b,\ell})$, where $q_{a,\ell}$ is the number of created atoms of distribution a on slice ℓ, while the entry-comparison term remains $O(D^2 Lr)$.

K.8 Spherical Retrieval

For the synthetic spherical mixtures, we form six validation classes and twelve disjoint evaluation classes on $\mathbb{S}^2$. Each class is a three-component von Mises–Fisher mixture with one shared mode of mass 0.50 and two class-specific modes of masses near 0.32 and 0.18, and each measure contains 32 atoms. We cross uniform and wrong-class structured clutter with the contamination levels 0, 0.15, 0.30, and 0.45, and we use 24 common slices, eight queries per class and condition, and eight evaluation seeds.

For ModelNet40, we use the official training and test partition (Wu et al., 2015; Qi et al., 2017). After centering each 2,048-point cloud, we normalize its points to unit radial directions, which retains the direction but not the radial distance or the surface connectivity. We use the training shapes for validation and two test shapes from each of the 40 classes for evaluation, the clean database and every query drawing disjoint point indices. Each empirical measure has 48 atoms, and every method uses 24 common slices whenever its construction admits them.

We evaluate a clean full view, a 70% view with 20% uniform clutter, a 50% view with 40% uniform clutter, and a 45% view with 50% clutter drawn from an object of another class. The clean database remains fixed across eight query and slice seeds. We report the same-shape top-1 accuracy, the reciprocal rank, and the same-class top-1 accuracy with 95% confidence intervals over the eight seeds, where same-shape retrieval asks for the database sample drawn from the same source mesh and same-class retrieval also accepts the other test mesh of the class. The intervals describe the variation in point sampling, corruption, and projections for the fixed set of 80 test shapes.

We select one λ for SSCOPT_2 and one for SSCOPT_1 on the validation data, and similarly one transported fraction for fixed-mass SSCOT_2 and for SSPW, while SSW_2 and LSSOT are balanced. For LSSOT, we use the implementation of the authors with a 1,024-point uniform reference, following their ModelNet configuration. Doubling the reference to 2,048 points reproduces the ModelNet results and changes one of the eight synthetic conditions of the first run by a single query. We also select the vMF-MMD concentration on the validation data. Each method uses one selected parameter across the evaluation conditions, and we compare this choice with the known visible fraction and with one fraction tuned for each condition.

The selection pools all validation conditions and maximizes the same-instance top-1 accuracy, resolving ties by the reciprocal rank, then by the same-class top-1 accuracy, and then by the larger parameter. For the synthetic and ModelNet40 studies, respectively, it selects $\lambda = 0.002$ and 0.000125 for SSCOPT_2, and 0.02 and 0.01 for SSCOPT_1. The selected transported fractions are 0.75 and 0.55 for both fixed-mass methods, and the vMF-MMD concentrations are 10 and 50. In both studies, the penalty grids double from $3.125 \cdot 10^{-5}$ to 0.064 for SSCOPT_2 and from 0.005 to 0.16 for SSCOPT_1, the candidate transported fractions are $\{0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 1\}$, and the vMF-MMD concentrations are $\{2, 5, 10, 20, 50, 100, 200\}$.

We compute SSCOPT_2 with the profile solver. The reported SSCOPT_1 values and the SSPW_1 controls use the public PAWC implementation (Kolouri, 2026): SSCOPT_1 selects the penalized minimum of each slice's PAWC W_1 profile, which equals the COPT value on distinct support (Appendix C). When a pooled projected gap is at most 10^{-13}, the PAWC wrapper separates the coordinates by replacing z_i with $(z_i + 10^{-11} i) \bmod 1$ in input order. SSCOPT_2 uses the original coordinates.

As an exact reference at equal cardinalities, we use an independent assignment formulation, which replaces every pair cost c_{ij} by $\min\{c_{ij}, 2\lambda\}$ and solves one complete assignment, using a thresholded ground cost as in Pele & Werman (2009), and an augmented version that enforces a fixed transported cardinality. Across 120 random trials, we compare the quadratic-cost penalized and fixed-mass assignments with brute-force partial matching, and the profile solver with the sweep on equal-size inputs, and we check the PAWC W_1 profiles against their brute-force values. All comparisons agree to floating-point precision. Further tests verify the auxiliary-point augmentation and the balanced reduction, the invariance under joint rotations of data and slices, the LCOPT identities and upper bounds, and the triangle inequalities of COPT and of refined LCOPT.

Figure 27 shows the expected Monte Carlo behavior. Increasing the number of shared slices reduces the estimation error, while rotating the data against a fixed finite set of slices leaves a residual rotation

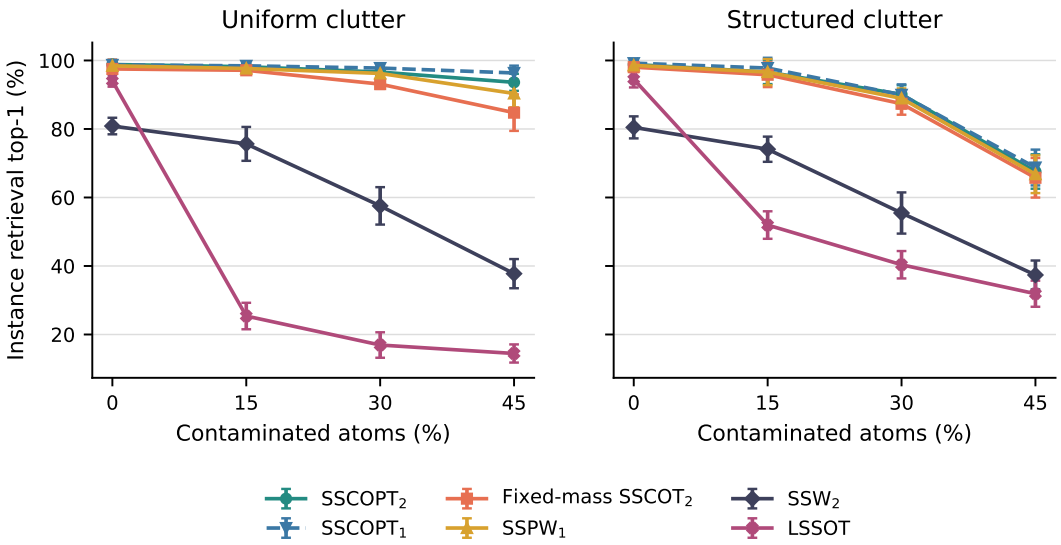


Figure 26: Spherical retrieval as uniform or structured clutter replaces an increasing fraction of each sample, showing the mean accuracy with 95% confidence intervals over eight runs.

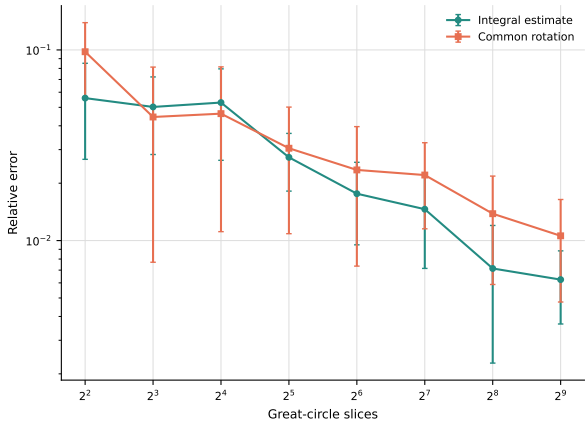


Figure 27: Across eight repetitions, increasing the number of projections reduces the SSCOPT$_2$ estimation error relative to a 2,048-slice reference and the error under a common rotation with fixed slices. Curves show the mean errors with 95% confidence intervals.

error that decreases as the estimate approaches the integral against σ. All retrieval comparisons use common slices, so that this sampling variation is shared across methods and pairs.

Table 2: Same-instance retrieval accuracy on synthetic spherical mixtures, in % with 95% confidence intervals over eight runs.

Method	Uniform 0%	Uniform 45%	Structured 0%	Structured 45%
SSCOPT_2	98.8 ± 1.4	93.6 ± 2.5	98.8 ± 1.0	67.6 ± 5.0
SSCOPT_1	98.7 ± 1.4	96.4 ± 2.1	99.2 ± 0.8	68.8 ± 5.2
Fixed-mass SSCOT_2	97.5 ± 2.7	84.8 ± 5.3	98.0 ± 1.4	65.8 ± 5.8
SSPW_1	98.4 ± 1.5	90.4 ± 4.1	98.6 ± 1.0	66.8 ± 5.5
SSW_2	80.9 ± 2.4	37.8 ± 4.3	80.5 ± 3.2	37.4 ± 4.2
LSSOT	94.0 ± 1.7	14.5 ± 2.7	94.5 ± 2.4	31.9 ± 3.8

Table 2 reports the two clutter mechanisms at the endpoints of the contamination range. Under 45% uniform clutter, the paired gains of SSCOPT_2 over fixed-mass SSCOT_2 and over SSPW_1 are 8.85 and 3.26 percentage points, with 95% confidence intervals of 5.63–12.08 and 1.43–5.08. Under structured clutter, the partial methods are close at 45%, while both balanced methods trail by a wide margin.

Table 3: ModelNet same-instance top-1 accuracy of the main spherical methods. The conditions are clean full views, 70% views with 20% uniform clutter, 50% views with 40% uniform clutter, and 45% views with 50% wrong-object clutter.

Method	Clean	20% Uniform	40% Uniform	50% Object
SSCOPT_2	50.2 ± 4.7	30.8 ± 4.9	21.4 ± 3.4	13.3 ± 2.7
Fixed-mass SSCOT_2	48.0 ± 4.5	32.7 ± 6.3	24.4 ± 2.8	12.7 ± 2.3
SSPW_1	45.6 ± 4.9	32.5 ± 5.0	23.6 ± 1.8	13.6 ± 3.0
SSW_2	23.6 ± 2.5	7.7 ± 2.0	4.7 ± 1.9	2.8 ± 1.3
LSSOT	23.9 ± 1.7	6.7 ± 1.8	3.8 ± 1.6	3.1 ± 1.4

Table 4: ModelNet retrieval with alternative transported fractions and with vMF-MMD. The known-overlap variants use the visible signal fraction, and the condition-tuned variants select a fraction on training shapes with the corresponding corruption.

Method	Clean	20% Uniform	40% Uniform	50% Object
SSCOPT_1	49.2 ± 5.0	31.2 ± 4.4	21.9 ± 3.0	13.9 ± 3.4
Fixed-mass SSCOT_2 (known overlap)	23.6 ± 2.5	24.2 ± 4.1	20.9 ± 3.6	14.7 ± 3.1
SSPW_1 (known overlap)	30.2 ± 3.7	27.2 ± 4.9	21.9 ± 3.4	14.7 ± 1.5
Fixed-mass SSCOT_2 (condition-tuned)	52.2 ± 4.7	32.7 ± 6.3	22.5 ± 1.8	12.7 ± 2.3
SSPW_1 (condition-tuned)	50.5 ± 5.0	32.5 ± 5.0	21.4 ± 1.9	13.6 ± 3.0
vMF-MMD	54.8 ± 3.8	30.3 ± 4.3	19.2 ± 3.5	12.7 ± 2.0

Among the spherical transport methods of Table 3, SSCOPT_2 has the highest clean mean, fixed-mass SSCOT_2 leads the mild and uniform-clutter conditions, and SSPW_1 leads the wrong-object condition by 0.3 points. All three partial methods remain well above SSW_2 and LSSOT as soon as the observations are incomplete or contaminated. Using the known overlap improves retrieval mainly under wrong-object contamination (Table 4), and the relative performance of the partial methods depends on the corruption.

K.9 Earthquake Fitting

For earthquake fitting, we follow Kolouri (2026) in using larger target batches for the partial methods and use a catalogue of USGS ComCat earthquakes of magnitude at least 4.5 from 2010 through 2024 (U.S. Geological Survey, 2026). The catalogue contains 112,484 cleaned coordinates, which we divide into training and test partitions with an 80/20 split and seed 20260818.

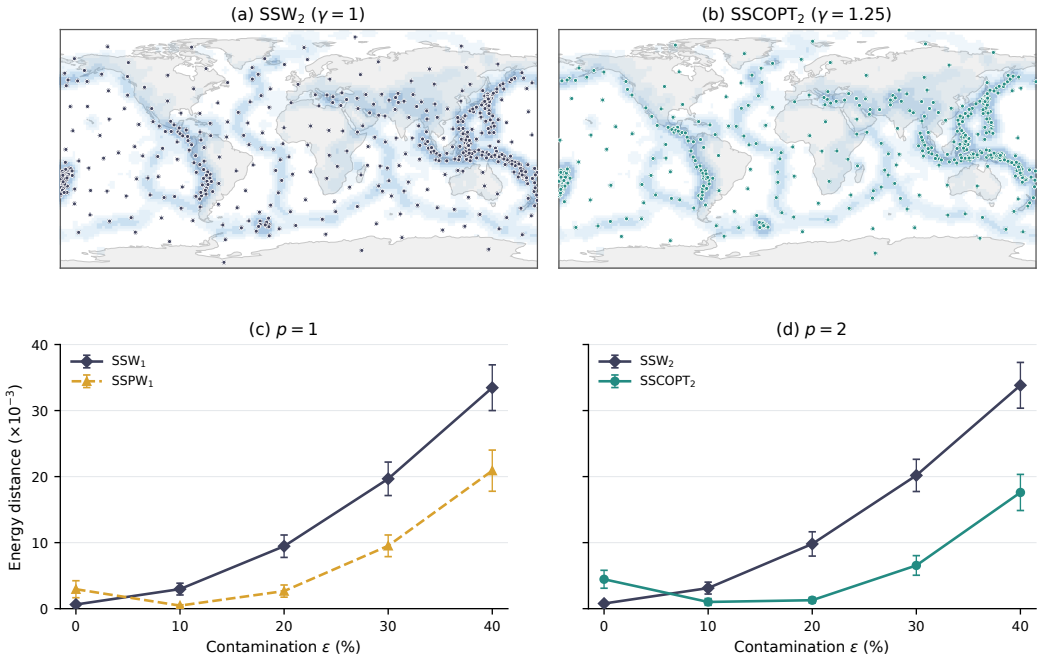


Figure 28: Fitting the USGS earthquake distribution under uniform contamination, following the experimental design of Kolouri (2026). The maps show the catalogue density in blue and the fitted particles at 30% contamination. The curves report mean test geodesic energy with 95% confidence intervals over eight runs. SSW uses 500 targets ($\gamma = 1$) and the partial methods use 625 ($\gamma = 1.25$), with one parameter setting per method across contamination levels.

We use the evaluation seeds $\{1103, 2207, 3313, 4421, 5527, 6637, 7753, 8861\}$. We fit the spatial density of the training catalogue and evaluate on test events. SSW$_1$ and SSPW$_1$ use the public PAWC solver; projected angles receive 10^{-9}-scale jitter in radians before each circular solve, with roundoff-scale separation if coincident coordinates remain. Within each run, all methods share the initial 500 uniform particles and the 4,000-event evaluation sample. Each fit draws 32 new slices per step and takes 2,000 Adam steps under a cosine learning-rate schedule from $8 \cdot 10^{-2}$ to $4 \cdot 10^{-3}$. We replace 0, 10, 20, 30, or 40% of each target batch by uniform spherical points. SSW uses 500 targets, and the partial methods use $\gamma = 1.25$ and 625 targets at every level. SSCOPT$_2$ uses $\lambda = 10^{-3}$ in the $[0, 1)$ circle parametrization at every contamination level, the value that an archived 300-step validation run with seed 20260902 selected from $\{2.5 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 4 \cdot 10^{-3}\}$. We did not keep the full configuration of that run, and a repeat with the current defaults selects $2 \cdot 10^{-3}$, so we report this penalty as a fixed setting. For the fixed-mass SSCOT$_2$ control, we transport all 500 model particles to the 625-target batch. We evaluate every fit by the geodesic energy distance to the test sample, and we draw the coastline from the Natural Earth 110m land polygons (Natural Earth, 2026).

Figure 29 summarizes the eight evaluation seeds. On clean batches, SSW$_1$ and SSW$_2$ reach the lowest mean errors, at 0.63 ± 0.21 and 0.78 ± 0.19. SSCOPT$_2$ improves on fixed-mass SSCOT$_2$ in seven of the eight seeds, with a mean paired advantage of 0.73×10^{-3} and a 95% interval of $[0.13, 1.33] \times 10^{-3}$. Their means are 4.45 ± 1.36 and 5.18 ± 1.58, and both remain above balanced SSW$_2$ in every clean run.

At each contaminated level, SSPW$_1$ and SSCOPT$_2$ improve on their balanced counterparts in all eight paired runs. For SSCOPT$_2$, the mean errors at 10, 20, 30, and 40% are 1.02, 1.28, 6.55, and 17.60, against 3.12, 9.80, 20.18, and 33.83 for SSW$_2$, in 10^{-3} units, and fixed-mass SSCOT$_2$ obtains 1.12, 1.30, 6.82, and 17.48 at the same levels. The two quadratic partial methods reach similar fitting errors, with paired confidence intervals containing zero at each contaminated level, while both keep at most about half of the SSW$_2$ error at every level.

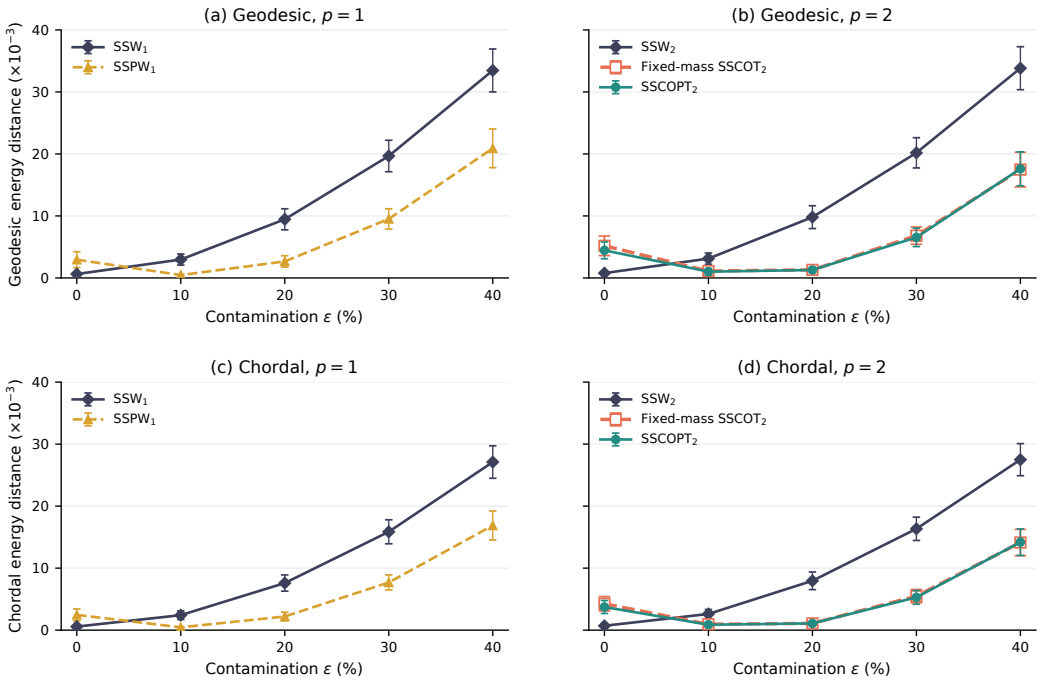


Figure 29: Fixed-mass SSCOT$_2$ and SSCOPT$_2$ reach similar earthquake fitting errors on the contaminated batches, and both improve on balanced SSW$_2$. The top row extends Figure 28 with the fixed-mass comparison, and the bottom row reports the chordal energy.

We also measure the chordal energy, which distinguishes probability distributions on the full sphere and complements the geodesic energy (Lyons, 2020; Székely & Rizzo, 2013). For all 200 fits, we use the test samples to compute

$$\mathcal{E}_{\mathrm{ch}}(\mu, \nu) = 2\mathbb{E}\|X - Y\|_2 - \mathbb{E}\|X - X'\|_2 - \mathbb{E}\|Y - Y'\|_2, \tag{K.28}$$

where X, X' and Y, Y' are independent draws from μ and ν. The empirical means include the zero diagonal terms, and the confidence intervals come from the eight paired runs. At 30% contamination, SSCOPT$_2$, fixed-mass SSCOT$_2$, and SSW$_2$ reach 5.29 ± 1.11, 5.49 ± 1.07, and 16.34 ± 1.88 in 10^{-3} units. SSCOPT$_2$ improves on SSW$_2$ in all eight seeds at every contaminated level, while its paired intervals against fixed-mass SSCOT$_2$ contain zero, and on clean batches SSW$_2$ reaches a lower chordal energy of 0.70 ± 0.17 against 3.73 ± 1.04 for SSCOPT$_2$, in agreement with the geodesic comparison.

For the gradients, we hold each discrete circular plan fixed while differentiating its slice cost, project the derivative onto the tangent spaces of the particles, and retract by normalization. On a neighborhood where the minimizing plan $\pi^\star$ is fixed and its cost branches are smooth, the finite minimum obeys the envelope identity

$$\nabla_\theta \min_\pi L(\theta, \pi) = \nabla_\theta L(\theta, \pi^\star), \tag{K.29}$$

whereas the objective may be nondifferentiable where the plan changes. We keep only the directions whose plans agree at the center and at both finite-difference endpoints. With a step of 10^{-6}, we record 32 accepted directions for each of $p = 1$ and $p = 2$. For $p = 1$, 32 of 55 attempts pass this check, whereas all 32 pass for $p = 2$, and the analytic and finite-difference derivatives agree in these comparisons.

The active fraction helps explain the similarity between the partial fits. SSPW transports all 500 model particles at every step, whereas SSCOPT selects the active particles through its penalty. In the displayed run (seed 1103), the mean slice-wise active source fraction of SSCOPT rises from 0.899 at step 1 to 1.000 at step 51, which brings it to the fixed-mass cardinality early in the optimization.

K.10 Linear COPT Comparisons

For the LCOPT study, we compare a same-size uniform reference, a double-size uniform reference, per-slice pooled quantiles, and a per-slice medoid on the validation set. We select the reference with the highest fixed-size Spearman correlation, resolving ties by the top-five neighbor overlap and then by the smaller storage, and we keep this choice for test retrieval and timing. The retrieval study compares 320 queries with 80 clean-gallery distributions using 24 slices. In the separate all-pairs timing study, we cross $n \in \{24, 48, 96\}$ atoms with $D \in \{40, 80, 160, 320\}$ distributions and use eight slices, with new frame and sampling seeds. Each timing is the median of five repetitions on one CPU worker with warmed compiled dependencies, and the time includes the LCOPT construction and every pairwise comparison after the shared spherical projection. The reference selection uses the training data, and we report its time outside the test timing. The encoding uses the cardinality-growth solver with direct interval evaluation at these atom counts. Pairwise COPT and the scalar costs between created masses use the compiled SciPy assignment reduction described above.

Table 5: LCOPT retrieval on test ModelNet shapes, reporting the retrieval accuracy, the mean reciprocal rank (MRR), and the distance correlation with COPT. The relative excess is the median percentage by which the rooted LCOPT distance exceeds the rooted COPT distance.

Method	Top-1 (%)	MRR (%)	Spearman	Relative Excess
COPT	28.4	40.6	1.000	0
Refined LCOPT	27.5	40.1	0.988	4.8%
Fixed-size LCOPT	19.4	31.4	0.981	5.2%

The created-mass refinement reaches 27.5% top-1 accuracy, against 28.4% for COPT, with a Spearman correlation of 0.988. The fixed-size embedding reaches a similar global correlation of 0.981 but a lower top-1 accuracy of 19.4%, which indicates that the compression changes the ordering of nearby candidates more than the overall ranking of the distances. The created-mass refinement also preserves more of the local neighbor structure. In the eight-slice timing study with 48 atoms and 320 distributions, its top-five neighbor overlap with COPT is 74.8%, against 57.2% for the fixed-size embedding and 29.8% for LSSOT.

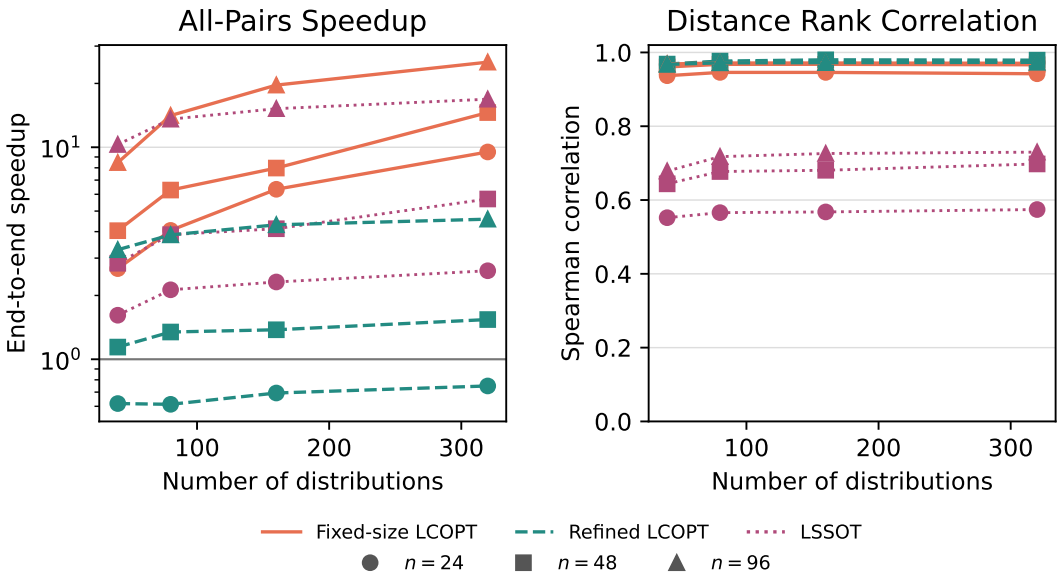


Figure 30: All-pairs comparison speed and distance correlation for LCOPT and LSSOT as the number of distributions grows. Each curve uses eight slices with a fixed atom count, and timings include reference encoding and all pairwise comparisons after the shared spherical projection.

For 48 atoms and 320 distributions (Figure 30), the fixed-size embedding takes 0.94 seconds, against 13.7 seconds for pairwise COPT, that is, a $14.6\times$ speedup at a correlation of 0.967, while the created-

mass refinement takes 8.9 seconds and reaches a correlation of 0.979, a $1.54\times$ speedup over pairwise COPT.