PARTIAL OPTIMAL TRANSPORT ON THE CIRCLE FOR ALL TRANSPORTED MASSES IN $O(N \log N)$

Anonymous authors
Paper under double-blind review

ABSTRACT

Partial optimal transport permits mass to remain unmatched, making it useful for comparing incomplete or contaminated distributions. On the real line, PAWL computes the entire transport profile—the optimal cost at every transported mass—in log-linear time. Extending this result to the circle requires resolving the cut optimization that already appears in balanced circular transport. In this paper, we introduce PAWC, an exact algorithm for circular partial 1-Wasserstein transport between discrete measures with distinct support points and a common atom weight. For N atoms in total and geodesic transport cost, PAWC computes the full partial-transport profile and a compact encoding of optimal plans in $O(N \log N)$ time and $O(N)$ memory. We prove that a single cut is simultaneously optimal for every transported mass and construct it through a *free-gap invariant* that makes local matching updates consistent with one unwrapping of the circle. This avoids solving a separate line problem at every support gap. Applying the solver to great-circle projections yields a partial extension of spherical sliced Wasserstein. The numerical benchmarks show agreement with the exact reference solvers and the predicted scaling. Applications demonstrate improved robustness in matching shape descriptors under occlusion and clutter and in spherical distribution fitting with contaminated targets.

1 INTRODUCTION

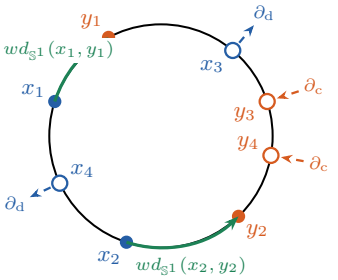
Optimal transport (OT) compares measures while respecting the geometry of their underlying domain (Villani, 2009; Peyré & Cuturi, 2019), but classical OT requires the measures to have equal total mass. Partial transport relaxes this constraint by matching only a portion of the available mass and allowing the remainder to stay unmatched. This flexibility isolates shared structure without forcing outliers, spurious detections, or occluded features into implausible correspondences (Nietert et al., 2022). The appropriate amount of mass to match, however, is rarely known beforehand. The natural object is therefore the entire partial-transport profile, $k \longmapsto C_k^\circ$, where C_k° is the minimum cost of matching exactly k pairs. This profile quantifies the tradeoff between coverage and geometric fidelity and allows the transported mass to be selected after examining the marginal costs of additional matches (Chapel & Tavenard, 2025).

On the real line, the underlying order makes partial transport particularly tractable. For a fixed penalty on unmatched mass, Bai et al. (2023) developed an exact primal-dual solver with worst-case complexity $O(n \max\{m, n\})$, or $O(N^2)$ for distributions of comparable size. More recently, Chapel & Tavenard (2025) introduced the PAWL algorithm, which computes exact partial transport plans for every transported mass in $O(N \log N)$ time using nested active sets, neighboring candidate pairs, balanced chains, and a heap of chain marginals. These efficiencies, however, rely fundamentally on the linear order and fixed boundary of the real line.

Extending these ideas to periodic domains is nontrivial. Periodic geometry arises naturally in angles, phases, orientations, time of day, hue, and great-circle slices of spherical data, yet the circle has no canonical first point, last point, or zero-flow boundary. For balanced W_1, this missing boundary condition appears as a free additive circulation in the flow formulation, or equivalently as a cut over which a lifted line problem must be optimized (Delon et al., 2010; Rabin et al., 2011). In partial transport, this topological freedom is coupled to the transported mass: as k varies, the active atoms, optimal circulation, and hence the optimal cut may all change. Consequently, no single predetermined

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.

(a) Fixed-mass partial OT on $\mathbb{S}^1_L$



solid: transported; hollow: unmatched

(b) Exact mass and cost accounting

$$\mathcal{J}_k(M) = \underbrace{w \sum_{(i,j)\in M} d_{\mathbb{S}^1}(x_i, y_j)}_{\text{transport cost}} + \lambda_d \underbrace{(n-k)w}_{\text{destroyed source mass}} + \lambda_c \underbrace{(m-k)w}_{\text{created target mass}} .$$

transport active source → active target
destruction unmatched source → ∂_d
creation ∂_c → unmatched target

Figure 1: Partial transport on the circle. (a) solid arcs are geodesic transports between selected atoms; hollow atoms are left unmatched, their mass destroyed at ∂_d or created at ∂_c. (b) at a fixed transported cardinality $|M| = k$ those two amounts are constants, so any fixed penalties λ_d, λ_c contribute a constant and the problem is exactly the cardinality-constrained matching of Equation (2).

cut is guaranteed to recover the full profile $k \mapsto C^\circ_k$. A direct exact approach cuts the circle at every support gap, runs PAWL on each resulting line, and takes the lower envelope of the resulting profiles. Although correct, this approach repeats an $O(N \log N)$ computation over $O(N)$ cuts, resulting in $O(N^2 \log N)$ complexity. The central algorithmic challenge is therefore to recover the entire circular partial-transport profile without paying separately for every possible cut.

Contributions. We show that the apparent quadratic cost of searching over circular cuts is unnecessary: the entire partial-transport profile on the circle can be recovered exactly in the same $O(N \log N)$ time required on the line. The key is a hidden cut-free linear structure (Section 3). Building on the standard nesting of optimal active sets, we show that each greedy step activates two cyclically neighboring inactive atoms and that every evolving cell contains an original gap untouched by all previously selected cells. This *free-gap invariant* reveals a single, algorithmically constructed cut under which all preceding circular updates are simultaneously valid line updates. As a result, our method constructs—rather than searches for—a gap that is optimal for every cardinality at once, yielding an exact greedy algorithm (Theorems 3 and 4) that returns the complete cost profile together with a compact representation of all nested active sets and transport plans in $O(N \log N)$ time and $O(N)$ memory (Section 4). We further lift the construction to spherical data through slicing over great circles (Section 5). Finally, Section 6 confirms exactness against an LP oracle and demonstrates scaling for the complete profile. Applications to shape-descriptor matching on MPEG-7 and robust fitting on $\mathbb{S}^2$, including a real earthquake catalogue, illustrate partial transport at prescribed masses.

2 Background

2.1 Partial optimal transport on the circle

Let $\mathbb{S}^1_L = \mathbb{R}/L\mathbb{Z}$, $L > 0$, with coordinates in $[0, L)$ and geodesic distance $d_{\mathbb{S}^1}(x, y) = \min\{|x - y|, L - |x - y|\}$, and let $\mu = w\sum_{i=1}^n \delta_{x_i}$ and $\nu = w\sum_{j=1}^m \delta_{y_j}$ be uniformly weighted empirical measures with $w > 0$; write $N = n + m$ and $K = \min(n, m)$. For a transported mass $s \in [0, Kw]$, define the circular partial 1-Wasserstein cost by

$$\mathrm{PW}_\circ(s) := \min_{\pi \in \mathbb{R}_+^{n\times m}} \Big\{ \sum_{i=1}^n \sum_{j=1}^m d_{\mathbb{S}^1}(x_i, y_j)\pi_{ij} \ : \ \sum_{j=1}^m \pi_{ij} \le w, \ \sum_{i=1}^n \pi_{ij} \le w, \ \sum_{i,j} \pi_{ij} = s \Big\}. \quad (1)$$

Thus $\mathrm{PW}_\circ(s)$ is the minimum cost of transporting exactly s units of mass, with at most the full mass w of each atom participating. Relaxing mass conservation to an inequality and fixing the transported total gives the partial problem studied by Caffarelli & McCann (2010); Figalli (2010) for quadratic costs between densities, including dependence on transported mass.

Assumption 1 (Standing assumptions)**.** *All atoms carry the same mass w; the points $\{x_i\} \cup \{y_j\}$ are pairwise distinct; and the ground cost is $d_{\mathbb{S}^1}$.*

At an integer mass $s = kw$, scaling the feasible set by w gives the cardinality-k bipartite matching polytope, so an optimal plan can be chosen without splitting atoms: it matches exactly k source–target pairs, transports mass w along each pair, and uses each atom at most once. Hence,

$$C_k^\circ := \mathrm{PW}_\circ(kw) = \min_{M:\,|M|=k} w \sum_{(x_i,y_j)\in M} d_{\mathbb{S}^1}(x_i, y_j), \qquad k = 0, \dots, K, \tag{2}$$

where M ranges over matchings, so no atom appears in more than one pair. We call $k \mapsto C_k^\circ$ the *profile* and denote by $A(M)$ the set of atoms incident to M, called its *active set*. For $0 \le k < K$ and $s \in [kw, (k+1)w]$, write $\lambda = (s - kw)/w$; the value interpolates linearly,

$$\mathrm{PW}_\circ(s) = (1-\lambda)\, C_k^\circ + \lambda\, C_{k+1}^\circ, \tag{3}$$

so the discrete profile $\{C_k^\circ\}_{k=0}^K$ determines $\mathrm{PW}_\circ(s)$ for every $s \in [0, Kw]$; see Appendix I.

2.2 Computing partial transport

General-purpose exact OT solvers are costly; standard alternatives include entropic regularisation (Cuturi, 2013), hierarchical solvers (Halmos et al., 2025), and the one-dimensional quantile formula underlying sliced transport (Rabin et al., 2012; Peyré & Cuturi, 2019; Kolouri et al., 2019). For partial transport, Phatak et al. (2023) compute the full mass profile in $O(N^3)$; Bai et al. (2023) give a quadratic one-dimensional solver; Bonneel & Coeurjolly (2019) solve injective partial assignment with quadratic worst-case cost; Séjourné et al. (2022) give linear-time iterations for one-dimensional unbalanced OT with mass controlled by penalties; and Chapel et al. (2021) obtain a regularisation path. PAWL (Chapel & Tavenard, 2025) is the first to return the entire partial profile on the line in $O(N \log N)$, and is what we extend.

On the circle, prior work is balanced: Delon et al. (2010) and Rabin et al. (2011) solve W_1 by optimising a cut or global circulation; Díaz Martín et al. (2024) linearise circular OT; and Bonet et al. (2023); Liu et al. (2025) use circular transport within spherical slicing. None addresses prescribed-mass partial transport on the circle.

Why the extension to the circle is non-trivial. For a *fixed* active set the circular cost is the minimum of the unwrapped line costs over the N support gaps (Proposition C.1 in Appendix C), so the envelope over gaps and cardinalities is an exact $O(N^2 \log N)$ reference solver (Corollary C.1, also there) — which we use as a baseline in Section 6 and as a trusted second opinion in testing. Example F.1 in Appendix F shows that reusing one fixed cut can be off by an unbounded factor. In the reported 5600-instance comparison of Appendix K.5, cutting at the widest support gap is suboptimal at some mass on 22.6% of instances, with a worst cost ratio of 3.2 to the exact optimum.

3 Structure of circular partial transport

3.1 Nested active sets and cyclic neighbours: $O(N)$ candidates per step

Theorem 1 (Nested optimal extension; proved in Appendix D)**.** *Let M_k be any optimal cardinality-k matching, $k < K$. Then there is an optimal cardinality-$(k+1)$ matching M_{k+1} with $A(M_{k+1}) = A(M_k) \cup \{x, y\}$ for a source atom x and a target atom y, both previously inactive.*

This is a standard consequence of successive shortest augmenting paths in min-cost flow (Ahuja et al., 1993), also used in PAWL (Chapel & Tavenard, 2025): active atoms are preserved while edges may be rematched. The circular geometry restricts the extension.

Theorem 2 (Cyclic-neighbour property; proved in Appendix D)**.** *In such an extension, x and y are consecutive in the cyclic ordering of the inactive set A_k^c.*

To identify candidate extensions, consider the inactive atoms in cyclic order. Only neighbouring source–target pairs need to be considered, giving at most $N - 2k \le N$ candidate cells at step k. We call the arc between two consecutive inactive atoms a *current cell*, a cell whose endpoints have opposite types a *candidate*, and write $\mathfrak{m}(\mathcal{C})$ for its *local marginal*: the increase in the cell's internal line-transport cost caused by activating its two endpoints (Definition E.2 in Appendix E). Under a cut avoiding this cell and all prior selections, activation rematches only its cell (Figure 2); and the increment is computable from that cell alone.

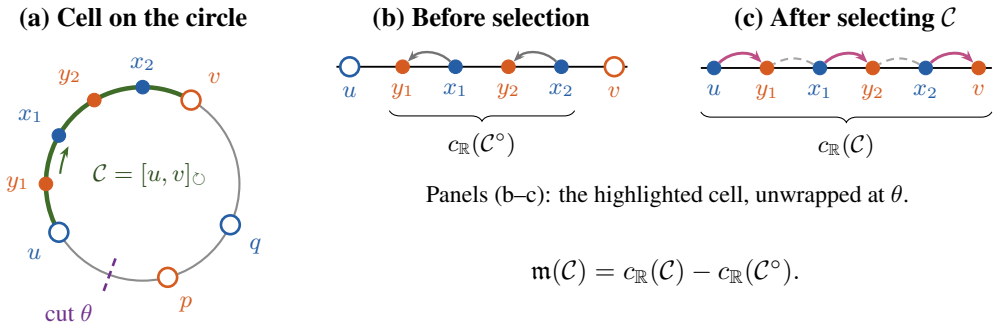


Figure 2: Selecting a cell activates its endpoints and rematches its interior. (a) Hollow atoms are inactive; solid atoms are active. The highlighted cell $\mathcal{C}$ joins cyclic neighbours u, v and has a balanced active interior. Other inactive atoms p, q lie on the complementary arc, so that arc is not a second u, v cell. Cutting at θ, outside $\mathcal{C}$ and all previous selections, gives the line order shown in (b–c). (b) Before selection, only the interior atoms are matched. (c) Selecting $\mathcal{C}$ activates u, v and replaces the two old matches (grey dashed) by the three sorted matches (magenta); it does not simply add a match from u to v. Matches outside the cell are unchanged. The marginal $\mathfrak{m}(\mathcal{C})$ subtracts the old interior cost from the new cell cost, as in PAWL's chain update (Chapel & Tavenard, 2025).

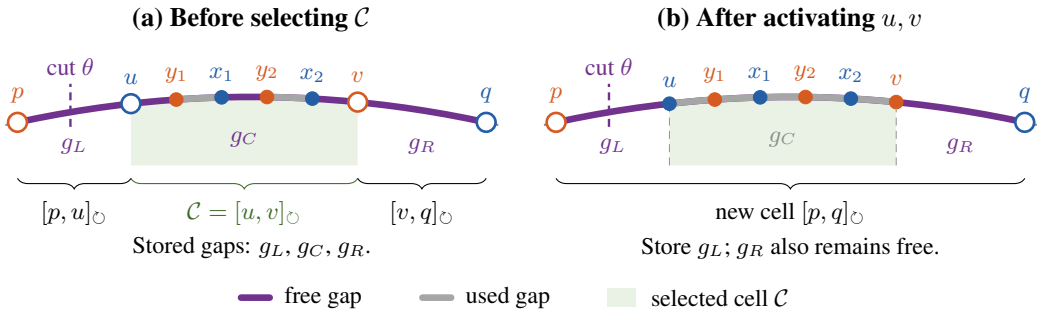


Figure 3: Free-gap inheritance for the update in Figure 2. The same atoms appear along the clockwise arc from p to q; the complementary arc is omitted. (a) Each cell stores a free gap: g_L, g_C, g_R. Grey gaps were covered by earlier selections. (b) Activating u, v (hollow to solid) merges the three cells into $[p, q]_{\circlearrowleft}$. Only gaps inside the shaded cell $\mathcal{C}$ become used. Both g_L and g_R remain free; the merged cell stores g_L in $O(1)$ time. The fixed cut $\theta \in g_L$ continues to avoid all selections.

3.2 The free-gap invariant: one cut for all $K + 1$ problems

The difficulty specific to the circle is that a local update is only meaningful relative to a cut, and cells are updated at different times. The following invariant is what makes one cut serve all of them. Call an original support gap *free* if it lies inside no cell that has been selected so far.

Lemma 1 (Free-gap invariant; proved in Appendix F)**.** *Before termination, every current cell contains at least one free original gap.*

After a selection, the merged cell inherits a free gap from an unselected neighbour (Figure 3), maintaining the invariant in constant time. For any candidate $\mathcal{C}$, a free gap in another cell avoids both $\mathcal{C}$ and all prior selections. This cut realizes the current optimum and makes the update feasible, so $C^\circ_{k+1} - C^\circ_k \leq \mathfrak{m}(\mathcal{C})$ (Lemmas F.2 and F.3 in Appendix F). Conversely, a nested optimal extension has a common optimal cut for its two active sets (Lemma D.1 in Appendix D). Under this cut, the line-neighbour and locality properties place the new pair in a candidate cell avoiding the cut and identify its marginal with the optimal increment. This second cut need not avoid past selections: line optimality supplies the required locality. Together, these two comparisons establish the greedy rule.

Theorem 3 (Exact greedy choice; proved in Appendix F)**.** *For $0 \leq k < K$, if A_k is circularly optimal and was produced from the empty set by valid cell selections, then $C^\circ_{k+1} - C^\circ_k = \min_{\mathcal{C} \in \mathfrak{C}_k} \mathfrak{m}(\mathcal{C})$, where $\mathfrak{C}_k$ is the set of current candidate cells, and activating the endpoints of any minimiser yields a circularly optimal A_{k+1}.*

Theorem 4 (Simultaneous optimal cut; proved in Appendix F)**.** *There is an original support gap $\theta^\star$ with $C^{\mathrm{line}}_{k,\theta^\star} = C^\circ_k$ for every $k = 0, \dots, K$, and the greedy construction produces one.*

Theorem 4 removes the outer loop over cuts: free-gap bookkeeping constructs a simultaneous optimal cut (Figure F.1 in Appendix F). Every plan is recovered by pairing the active sources and targets in unwrapped order (Appendix I). Full-mass-optimal cuts need not be unique, and full-mass optimality alone does not identify a cut that works for the entire profile (see Appendix K.5).

4 THE PAWC ALGORITHM: $O(1)$ MARGINALS, $O(N \log N)$ OVERALL

Algorithm 1 combines the greedy rule of Theorem 3, free-gap bookkeeping, and the following constant-time marginal queries. Following PAWL's chain preprocessing, let $z_1 < \cdots < z_N$ be the union support in $[0, L)$, with $\sigma_i = +1$ for a source and -1 for a target. Double it as

$$\widetilde{z}_i = z_i, \quad \widetilde{z}_{i+N} = z_i + L, \quad \widetilde{\sigma}_i = \widetilde{\sigma}_{i+N} = \sigma_i, \qquad i = 1, \dots, N. \tag{4}$$

A candidate cell from z_u to z_v becomes $[a, b]$ with $a = u$ and $b = v$ if $v > u$, or $b = v + N$ otherwise. Starting from $R_0 = S_0 = Q_0 = 0$, define for $t = 1, \dots, 2N$ the differential ranks, signed coordinate sums, and predecessors

$$R_t = \sum_{i=1}^{t} \widetilde{\sigma}_i, \qquad S_t = \sum_{i=1}^{t} \widetilde{\sigma}_i \widetilde{z}_i, \qquad p_t = \max\{0 \le s < t : R_s = R_t\}, \tag{5}$$

where p_t is undefined if the set is empty. Accumulate costs along each rank level:

$$Q_t = \begin{cases} Q_{p_t} + w|S_t - S_{p_t}|, & p_t \text{ exists}, \\ 0, & \text{otherwise}. \end{cases} \tag{6}$$

When p_t exists, $[p_t + 1, t]$ is a minimal balanced chain of line cost $w|S_t - S_{p_t}|$. An array storing the last index of each rank, initialized with index 0 for rank 0, computes all tables in $O(N)$ time and memory after sorting. For a balanced interval $[a, b]$, meaning $R_b = R_{a-1}$, its increasing-order line-matching cost is $c_{\mathbb{R}}([a, b]) = Q_b - Q_{a-1}$: the difference sums the intervening minimal-chain costs (proof in Appendix G). Candidate cells have balanced interiors and opposite-type endpoints, so this identity applies to both their full and interior costs.

Corollary 1 (Constant-time candidate marginal; proved in Appendix G)**.** *For a candidate cell represented by the balanced doubled interval $[a, b]$, $\mathfrak{m}([a, b]) = (Q_b - Q_{a-1}) - (Q_{b-1} - Q_a)$, with the second term zero when the interior is empty, leading to marginal costs of $O(1)$ after preprocessing.*

Theorem 5 (Correctness and complexity; proved in Appendix H)**.** *Under Assumption 1, Algorithm 1 returns the exact C°_k and activation ranks encoding optimal active sets, together with a gap $\theta^\star$ valid for all k, in $O(N \log N)$ time and $O(N)$ memory.*

In Algorithm 1, at most N initial candidates and one new candidate per activation give $O(N)$ heap entries. Lazy deletion pops each entry once at $O(\log N)$ cost. Marginal evaluation, list splicing, and free-gap inheritance each take $O(1)$, giving $O(N \log N)$ overall, including the initial sort.

In Section 6.1, we show that PAWC (Algorithm 1) agrees with the LP and cut-envelope cost profiles to numerical precision on diverse tested instances.

5 SLICING TO THE SPHERE

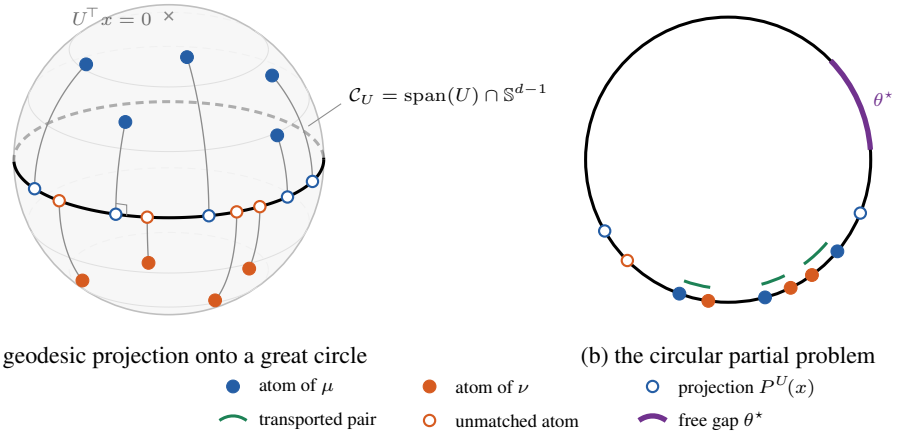
Directional data in d dimensions lives on $\mathbb{S}^{d-1}$, where exact OT is expensive. Slicing replaces it by an average over one-dimensional projections; on the sphere the natural projections are onto great circles (Bonet et al., 2023). For an orthonormal frame $U \in V_{d,2} := \{U \in \mathbb{R}^{d \times 2} : U^\top U = I_2\}$, the circle coordinate of the geodesic projection is $P^U(x) = U^\top x / \|U^\top x\|$. Its angle gives an instance of (2) with $L = 2\pi$. Write ω for the uniform probability measure on $V_{d,2}$; projections are defined for almost every U.

Definition 1 (Sliced spherical partial Wasserstein; Appendix J)**.** $\mathrm{SSPW}(s; \mu, \nu) = \int_{V_{d,2}} \mathrm{PW}_\circ(s; P^U_\# \mu, P^U_\# \nu)\, d\omega(U)$, estimated by the average over M i.i.d. slices.

Algorithm 1 PAWC: all exact partial W_1 costs on the circle

Input: Source points $\{x_i\}_{i=1}^n$, target points $\{y_j\}_{j=1}^m$, atom weight w, and circumference L
Output: Optimal costs $\{C_k^\circ\}_{k=0}^K$, activation ranks, and a simultaneous optimal cut $\theta^\star$

Sort the union support cyclically and form labels $\sigma_1, \ldots, \sigma_N$
Build the doubled sequence and precompute R, S, p, Q using (4)–(6)
Initialize a circular doubly linked list containing all support indices
For each initial directed cell, store its unique original gap as its free-gap representative
Insert every opposite-endpoint initial cell into a min-heap keyed by $\mathfrak{m}$ (Corollary 1)
$C_0^\circ \leftarrow 0$
for $k = 0, 1, \ldots, K-1$ **do**
    repeat
        Pop the minimum heap entry $(u, v, \mathfrak{m})$
    until u, v are inactive, $v = \mathrm{succ}(u)$, and u, v have opposite types
    Record activation rank $k+1$ for u and v
    $C_{k+1}^\circ \leftarrow C_k^\circ + \mathfrak{m}$
    Let $p \leftarrow \mathrm{pred}(u)$ and $q \leftarrow \mathrm{succ}(v)$ before deletion
    if only u, v are inactive **then**
        $\theta^\star \leftarrow$ free-gap representative of the complementary cell $[v, u]_\circlearrowleft$
        Remove u, v and **continue**
    end if
    Remove u, v and link p directly to q
    Let the new cell $[p, q]_\circlearrowleft$ inherit a free-gap representative from $[p, u]_\circlearrowleft$ or $[v, q]_\circlearrowleft$
    if p and q have opposite types **then**
        Compute $\mathfrak{m}([p, q]_\circlearrowleft)$ in $O(1)$ and insert the new candidate into the heap
    end if
end for
if $\theta^\star$ has not been set **then**
    Choose it as the free-gap representative of any remaining current cell
end if
return $\{C_k^\circ\}_{k=0}^K$, activation ranks, $\theta^\star$



(a) geodesic projection onto a great circle (b) the circular partial problem

Figure 4: Spherical slicing, after Bonet et al. (2023, Figure 1). (a) atoms of μ (blue) and ν (orange) on $\mathbb{S}^{d-1}$, the great circle $\mathcal{C}_U$ (dashed behind the sphere), and the geodesic projection of each onto it. (b) the circular problem that slice induces, with a cardinality-3 matching and a free gap $\theta^\star$. PAWC returns the optimal matching for *every* k in one $O(N \log N)$ sweep per slice.

Figure 4 shows one slice. Because PAWC returns the whole profile per slice, one pass over M slices gives the estimate at *every* transported mass, at cost $O(M(N \log N + dN))$. The estimator concentrates uniformly over the profile: $M \geq (\pi^2 K^2 w^2 / 2\varepsilon^2) \log(2(K+1)/\delta)$ slices suffice for ε-accuracy on the entire curve with probability $1 - \delta$ (Proposition J.1 in Appendix J). For $n = m$, its

endpoint equals nw times spherical sliced Wasserstein between the normalized measures, recovering the probability-measure version when $nw = 1$.

6 EXPERIMENTS

We evaluate numerical correctness and scaling, then study partial transport for shape matching and spherical fitting under controlled contamination. The applications compare transport costs; they do not evaluate automatic selection of transported mass. Protocols are described in Appendices K and L.

6.1 EXACTNESS AND SCALING

Across 504 instances from seven test families — uniform, clustered, near-antipodal, tiny-arc, unequal-size, perturbed-grid, and interleaved (Figure K.1) — PAWC agrees with the LP oracle to 4.7×10^{-15} and with cut enumeration to numerical precision. Checks of nestedness, the free-gap invariant, cell balance, feasibility, and the validity of $\theta^\star$ pass at every step (Appendix K).

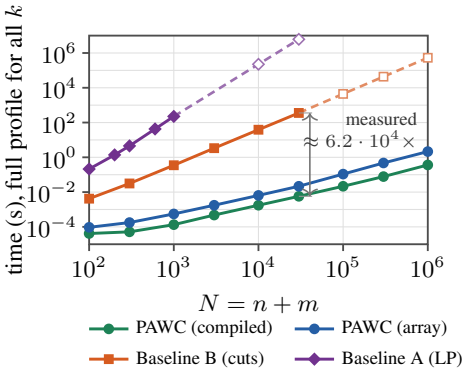


Figure 5: Time for the complete profile. The measured speedup over cut enumeration is 6.2×10^4 at $N = 3 \cdot 10^4$. Solid curves are measured; dashed segments extrapolate cut enumeration ($N^2 \log N$) and the LP oracle (N^3).

Figure 5 measures the complete profile up to $N = 10^6$ for PAWC and $N = 3 \cdot 10^4$ for cut enumeration. At the latter size, compiled PAWC takes 5.74 ms versus 357 s, about 6.2×10^4 times faster. The measured log–log slopes are 1.03–1.29 and 2.00, respectively; heap operation counts grow linearly (Appendix K). Timings use one core of an AMD Ryzen Threadripper PRO 9965WX, Python 3.11 and NumPy 2.4, excluding JIT compilation. For $n = m = 256$ and 2048, the complete profile takes 0.09 ms and 0.56 ms, versus 4.4 ms and 1.54 s for one mass, or 30 ms and 11.4 s for seven masses, using POT's general partial solver (Flamary et al., 2021). Reading additional masses from the computed profile requires no further transport solves.

6.2 PARTIAL MATCHING OF SHAPE DESCRIPTORS

The outward normal directions of a closed planar curve, sampled uniformly in arclength, form a measure on $\mathbb{S}^1_L$ — the two-dimensional extended Gaussian image (Horn, 1984). It is invariant to translation, scale and starting point and rotates rigidly with the shape, so comparing shapes up to rotation is alignment on the circle. Partial transport addresses two important failure modes: *occlusion* deletes part of the measure, *clutter* adds mass belonging to nothing. This descriptor is a controlled testbed for transport costs: we fix it and the rotation grid across methods. Its lack of spatial part structure limits retrieval accuracy relative to specialized shape descriptors (Ling & Jacobs, 2007).

Protocol. The database contains all 1400 silhouettes of MPEG-7 CE-SHAPE-1 (Latecki et al., 2000). Contours are traced at the 0.5 level, resampled to 128 points, and converted to normal angles. A query keeps a contiguous fraction v of the arclength, replaces a fraction c of the survivors by uniform normals, and is rotated by a uniformly random angle. Every method scores the same 360 rotations. Table 1 uses the known clutter level for orientation and retrieval; Figure 6 instead fixes $\rho = 0.8$. The "full OT" baseline uses $k = K$, which is balanced OT only when $n = m$. Three tasks: recover the rotation (success $< 5°$); rank the other 1399 shapes and report the MPEG-7 bullseye score; and score the pairs of the recovered plan against the arclength indices the query was built from (F_1; clutter samples have no correct partner).

Result. At full visibility and $c = 0.2$, retrieval improves from 0.071 for maximum-cardinality transport to 0.302 for PAWC, against 0.457 on clean data (Table 1). The paired gain is 0.231 ± 0.053 (reported 95% interval), with PAWC ahead on 95 of 138 queries. At $c = 0.4$, orientation success is 1.00 versus 0.45. Under occlusion alone, unequal support sizes already permit maximum-cardinality transport to leave template atoms unmatched. Correlation degrades more strongly: its retrieval score

Table 1: Shape matching on MPEG-7: visible fraction v and clutter fraction c. Orientation and correspondence use 212 queries per cell; retrieval uses 138 queries against a 1400-shape database. Orientation and retrieval use $\rho = 1 - c$, assuming known clutter. "full OT" uses $k = K$; "corr." selects the best of three bandwidths per cell; "NN" is mutual nearest neighbours. Correspondence alignment is described in Appendix L.

		orientation ($< 5°$)			retrieval (bullseye)			correspondence (F_1)		
v	c	PAWC	full OT	corr.	PAWC	full OT	corr.	PAWC	full OT	NN
1.00	0.0	**1.00**	**1.00**	**1.00**	**0.457**	**0.457**	0.384	**1.00**	**1.00**	0.77
1.00	0.2	**1.00**	0.62	0.98	**0.302**	0.071	0.198	**0.84**	0.21	0.69
1.00	0.4	**1.00**	0.45	0.92	**0.118**	0.038	0.077	**0.66**	0.12	0.53
0.70	0.0	**1.00**	**1.00**	0.96	**0.337**	**0.337**	0.126	**0.94**	**0.94**	0.77
0.70	0.2	**1.00**	0.83	0.90	**0.201**	0.106	0.066	**0.81**	0.45	0.69
0.70	0.4	**0.99**	0.62	0.73	**0.104**	0.039	0.050	**0.68**	0.29	0.56
0.50	0.0	**1.00**	**1.00**	0.85	**0.228**	**0.228**	0.052	**0.91**	**0.91**	0.77
0.50	0.2	**1.00**	0.83	0.73	**0.141**	0.089	0.044	**0.79**	0.58	0.68
0.50	0.4	**0.94**	0.67	0.57	**0.081**	0.048	0.037	**0.69**	0.41	0.57

Table 2: Spherical fitting: mean clean-target energy distance ($\times 10^{-3}$), eight seeds per cell. SSPW uses $\gamma = 1/(1-\varepsilon)$ rounded to the grid, assuming known contamination; it equals SSW at $\varepsilon = 0$. "gain" is the reduction from SSW; "wins" counts seeds ahead; "best" is the largest gain over six tested ratios, without held-out selection.

	synthetic vMF mixture					earthquake catalogue				
ε	SSW	SSPW	gain	wins	best	SSW	SSPW	gain	wins	best
0.0	0.50	**0.50**	0%	0/8	14%	0.79	**0.79**	0%	0/8	0%
0.1	1.97	**0.79**	60%	8/8	70%	3.23	**1.15**	64%	8/8	77%
0.2	7.13	**2.30**	68%	8/8	68%	10.25	**3.05**	70%	8/8	70%
0.3	15.92	**6.09**	62%	8/8	63%	21.86	**10.35**	53%	8/8	53%
0.4	26.40	**14.46**	45%	8/8	50%	36.40	**31.42**	14%	7/8	40%

falls from 0.384 to 0.052 as v decreases from 1 to 0.5 at $c = 0$. The fixed-ρ comparisons in Figure 6 also favor PAWC under clutter, without supplying c at evaluation. They do not establish how to select ρ on unseen data. A supplementary comparison with KL-relaxed Sinkhorn, selecting the best of three relaxation weights per cell, reports orientation success 0.71 versus 0.85 at $v = 0.35, c = 0.4$ and about 20 times the runtime (Appendix L).

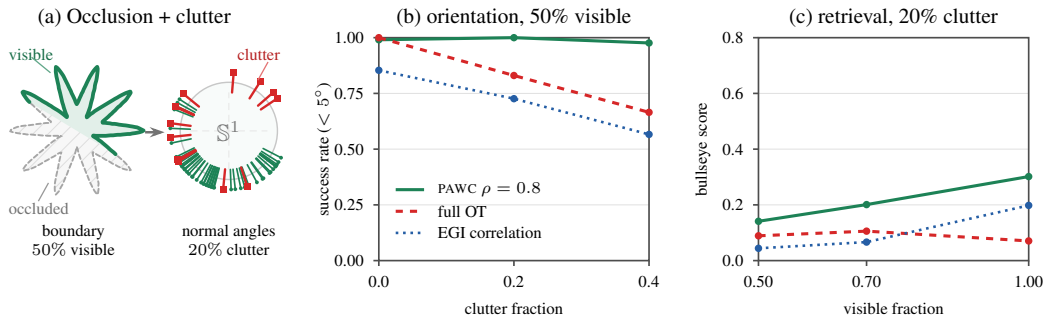


Figure 6: Partial matching of shape descriptors on MPEG-7. (a) From a partially visible boundary (green; occluded arc dashed) to its circular normal-angle measure; red squares mark clutter replacing 20% of the observed normals. (b) orientation success against clutter at half the boundary visible and (c) retrieval bullseye against the visible fraction at 20% clutter, using a fixed $\rho = 0.8$. Baselines are maximum-cardinality transport ("full OT") and extended-Gaussian-image correlation.

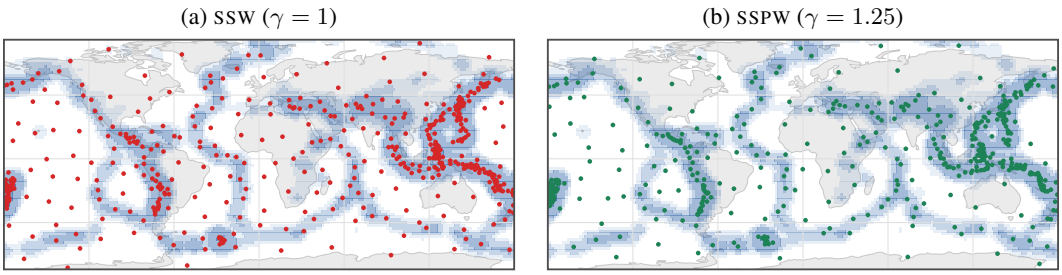


Figure 7: Fitting 500 particles to USGS target batches with $\varepsilon = 0.3$ contamination. Blue shows the catalogue's kernel density over Natural Earth coastlines; dots show fitted particles. SSW uses $\gamma = 1$ and the illustrative SSPW fit $\gamma = 1.25$, concentrating more particles near ridges and subduction zones.

6.3 Robust fitting on $\mathbb{S}^2$

We use SSPW as a training loss and compare against the method it generalises, spherical sliced Wasserstein (SSW), recovered here when $\gamma = 1$. We put partiality on the *target* side: trimming a model particle removes its direct transport gradient on that slice. We draw target batches of size $m = \gamma n$, $\gamma \geq 1$, and transport $k = K = n$, so every particle participates while a fraction $1 - 1/\gamma$ of the target batch remains unmatched. Thus $\rho = k/K = 1$ throughout, although target-side trimming increases with γ.

We fit 500 particles, initialised uniformly, for 2000 Adam steps with 32 slices redrawn per step; gradients are taken with the matching held fixed (envelope theorem) and retracted to the sphere. A fraction ε of each target batch is replaced by uniform points. We report the geodesic energy distance to 4000 *clean* target points, a metric that shares no implementation with the training loss. The two targets are a five-component von Mises–Fisher mixture with unequal masses and concentrations, and 94 953 earthquake epicentres of magnitude ≥ 4.5 from the USGS catalogue, 2010–2024 (U.S. Geological Survey, 2026), split 80/20 for training and evaluation.

Figure 8 shows broad minima at larger γ under contamination, while the clean-data curve favors modest oversampling. In Table 2, for $\varepsilon \geq 0.1$ the rule reduces energy distance by 45–68% on the synthetic mixture and 14–70% on the catalogue. It wins every seed except at catalogue $\varepsilon = 0.4$, where it wins seven of eight and trimming less performs better. At $\varepsilon = 0$, the rule equals SSW; the best tested γ improves the synthetic mean by 14%. These comparisons combine larger target batches with trimming and therefore do not isolate their effects. The rule also assumes known contamination; selecting γ without it remains to be evaluated.

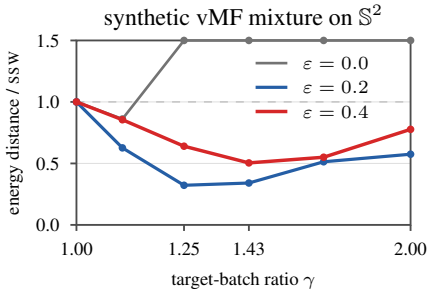


Figure 8: Energy distance against γ on the synthetic mixture, each curve divided by its own $\gamma = 1$ value. Increasing γ changes both batch size and trimming. The clean curve exceeds the vertical range at larger ratios.

7 Discussion

PAWC computes the full profile, compactly encodes optimal plans, and constructs a simultaneous optimal cut in $O(N \log N)$ time and $O(N)$ memory via the free-gap invariant. The result assumes equal atom weights and geodesic W_1 cost. Unequal weights require capacity-saturation events; coincident supports fall outside Theorem 5, and finite perturbations are exact only for the perturbed instance. Deterministic tie-breaking handles antipodal pairs and equal marginals. Selecting mass from the profile's marginals could extend our applications beyond prescribed masses.

For integer $p > 1$, Angrick et al. (2026) recently extend the PAWL framework on the line with FFT-based balanced-interval queries, computing the full profile in $O(pN \log^2 N)$ time. Their results could extend PAWC to geodesic p-Wasserstein costs by replacing its marginal oracle on the doubled support, subject to a circular correctness analysis, which we postpone to future work.

Reproducibility statement

Appendices K and L describe the numerical and application protocols. The supplied code includes reference and optimised PAWC implementations, LP and cut-envelope baselines, invariant tests, and walkthrough and scaling demos.

AI use statement

Generative AI tools assisted with editing, figure code, implementation scaffolding, validation scripts, proof checking and revision, and discussion of experimental design and interpretation. They also helped audit consistency between the manuscript, code, and reported results. Proposed additional experiments are not reported as completed experiments. The authors are responsible for the final content, including mathematical claims, empirical results, references, and AI-assisted artifacts.

References

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ, 1993.

Sebastian Angrick, Jacobus Conradi, Mónika Csikós, Niko Hastrich, Danny Mittal, André Nusser, Krzystof Onak, and Sharath Raghvendra. Computing all optimal partial p-Wasserstein matchings on the line. *arXiv preprint arXiv:2608.18875*, 2026. doi: 10.48550/arXiv.2608.18875. URL https://arxiv.org/abs/2608.18875.

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, 2023.

Thomas Bendokat, Ralf Zimmermann, and P.-A. Absil. A Grassmann manifold handbook: Basic geometry and computational aspects. *arXiv preprint arXiv:2011.13699*, 2020.

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.

Nicolas Bonneel and David Coeurjolly. SPOT: Sliced partial optimal transport. *ACM Transactions on Graphics*, 38(4):1–13, 2019.

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, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. *Advances in Neural Information Processing Systems*, 34:23270–23282, 2021.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26, 2013.

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 Vladimir Medri, Yikun Bai, Xinran Liu, Kangbai Yan, Gustavo Rohde, and Soheil Kolouri. LCOT: Linear circular optimal transport. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=49z97Y9lMq.

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.

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 http://jmlr.org/papers/v22/20-451.html.

Peter Halmos, Julian Gold, Xinhao Liu, and Benjamin Raphael. Hierarchical refinement: Optimal transport to infinity and beyond. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 21629–21661. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/halmos25a.html.

Berthold K. P. Horn. Extended Gaussian images. *Proceedings of the IEEE*, 72(12):1671–1686, 1984.

Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced wasserstein distances. *Advances in neural information processing systems*, 32, 2019.

Longin Jan Latecki, Rolf Lakämper, and Ulrich Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 424–429, 2000.

Haibin Ling and David W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.

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://openreview.net/forum?id=fgUFZAxywx.

Sloan Nietert, Ziv Goldfeld, and Rachel Cummings. Outlier-robust optimal transport: Duality, structure, and statistical analysis. In *International Conference on Artificial Intelligence and Statistics*, pp. 11691–11719. PMLR, 2022.

Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport: With Applications to Data Science*, volume 11 of *Foundations and Trends in Machine Learning*. Now Publishers, 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.

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.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision (SSVM 2011)*, pp. 435–446. Springer, 2012.

Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. Faster unbalanced optimal transport: Translation invariant Sinkhorn and 1-D Frank-Wolfe. In *International Conference on Artificial Intelligence and Statistics*, pp. 4995–5021. PMLR, 2022.

U.S. Geological Survey. ANSS comprehensive earthquake catalog (ComCat). https://earthquake.usgs.gov/fdsnws/event/1/, 2026. events of magnitude ≥ 4.5, 2010–2024; retrieved 18 August 2026.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2009. ISBN 9783540710509.

APPENDIX

The appendices contain the worked example, full proofs, numerical validation and application protocols. Their sections are lettered, and every statement, equation, figure and table carries the letter of its section (*Lemma D.3*, *Figure A.1*); a reference with a bare number (*Theorem 3*, *Figure 5*, *Section 6.2*) points to the main text.

Appendix A comes first and runs Algorithm 1 by hand on one instance, showing the heap, the linked list and the free-gap bookkeeping at every activation: it is the concrete account of what the algorithm does, and the mechanisms it exercises are the ones the rest of the appendix proves correct. Appendices B to I then contain the full development and every proof, in the order in which the main text uses them: notation and the cardinality-k formulation; the line structure inherited from PAWL and the cut envelope; nested active sets, the common-cut lemma and the cyclic-neighbour theorem; cells and their marginals; the free-gap invariant, the exact greedy theorem and the constructive simultaneous cut; the doubled-sequence preprocessing that makes candidate marginals $O(1)$; correctness and complexity; and plan recovery with the interpolation for non-integer masses. Appendix J develops the spherical construction and its concentration bound.

Appendix K reports the numerical validation — solvers and trust hierarchy, exactness against an LP oracle, the structural invariants, the complexity study, the one-cut experiment of Appendix K.5 and the spherical estimator's behaviour. Appendix L gives the protocols for the two applications in the main text, the unbalanced-transport comparison, and the in-situ exactness check.

A A WORKED EXAMPLE, STEP BY STEP

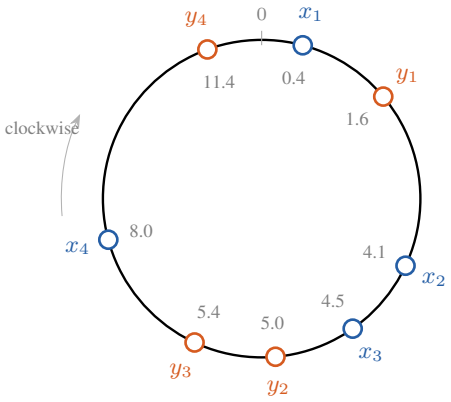
This appendix runs Algorithm 1 by hand on a single instance, so that every mechanism the paper introduces can be seen firing on concrete numbers before any of it is proved. We follow the notation in the main text; the statements the trace exercises — Lemma F.1, Theorem F.2, and Proposition G.1 — are proved in the sections that follow it. The mechanisms are: a wrap-around selection, a merged-cell selection with matching revocation, lazy heap skips, free-gap inheritance including the consumption of a selected cell's own representative, and the two-directed-cell endgame. We run PAWC on the circle of circumference $L = 12$ with unit weights $w = 1$ and geodesic cost $d_{\mathbb{S}^1}(u, v) = \min(|u - v|,\, L - |u - v|)$:

$$x = \{0.4,\ 4.1,\ 4.5,\ 8.0\} \quad \text{(blue circles, sources)},$$
$$y = \{1.6,\ 5.0,\ 5.4,\ 11.4\} \quad \text{(orange circles, targets)}.$$

The cyclic type pattern is $x\, y\, x\, x\, y\, y\, x\, y$. The instance is chosen so that every mechanism fires: a wrap-around selection, a merged-cell selection with matching revocation, four lazy heap skips, free-gap inheritance (including consumption of a selected cell's own representative), and the two-directed-cell endgame.

STEP 0 — INPUT ON S^1, AND THE NAIVE BASELINE

Cutting the circle open at any of the $N = 8$ inter-atom gaps and unrolling gives a line instance; the circular optimum is the lower envelope over cuts, $C^\circ_k = \min_r C^{\text{line}}_{k,r}$ (the cut envelope). Running PAWL once per cut costs $O(N^2 \log N)$. PAWC produces the same profile — and a single cut $\theta^\star$ valid for *all* k at once (Theorem F.2) — in one $O(N \log N)$ sweep. (Atom spacing in the circle diagrams is schematic, spread for legibility; the gray coordinates are the truth.)



Atoms are read clockwise from 0; hollow markers are *inactive* atoms (none transported yet). Note y_4 and x_1 sit on opposite sides of 0, only 1.0 apart along the wrap.

Step 1 — Precomputation on the doubled sequence (Appendix G)

Let $z_1 < \cdots < z_8$ be the sorted union support, with label $\sigma_i = +1$ for a source and $\sigma_i = -1$ for a target. To represent every clockwise cell as an ordinary interval, double the coordinates and labels:

$$\tilde{z}_i = z_i, \qquad \tilde{z}_{i+8} = z_i + L, \qquad \tilde{\sigma}_i = \tilde{\sigma}_{i+8} = \sigma_i, \qquad i = 1, \ldots, 8.$$

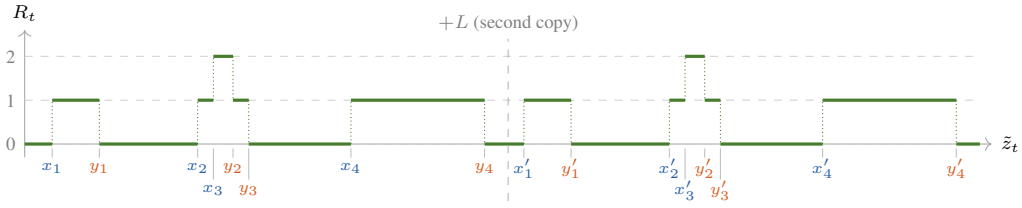
The prefix differential ranks R_t and signed coordinate sums S_t are

$$R_0 = S_0 = 0, \qquad R_t = \sum_{i=1}^{t} \tilde{\sigma}_i, \qquad S_t = \sum_{i=1}^{t} \tilde{\sigma}_i \tilde{z}_i, \qquad t = 1, \ldots, 16.$$

Let $p_t = \max\{0 \le s < t : R_s = R_t\}$ when this set is nonempty: it is the last previous index with the same rank. Define the maximal-chain prefix costs by

$$Q_0 = 0, \qquad Q_t = \begin{cases} Q_{p_t} + w|S_t - S_{p_t}|, & p_t \text{ exists}, \\ 0, & p_t \text{ does not exist}. \end{cases}$$

Thus Q_t accumulates minimal-chain costs along equal-rank predecessors. These are the preprocessing tables of PAWL on the doubled line; one sweep computes them by storing the latest index of each rank. In this balanced example ($n = m$), the rank walk repeats with period $N = 8$:



t	0	1	2	3	4	5	6	7	8
atom	—	x_1	y_1	x_2	x_3	y_2	y_3	x_4	y_4
$\tilde{z}_t$	—	0.4	1.6	4.1	4.5	5.0	5.4	8.0	11.4
R_t	0	1	0	1	2	1	0	1	0
S_t	0	0.4	−1.2	2.9	7.4	2.4	−3.0	5.0	−6.4
p_t	—	—	0	1	—	3	2	5	6
Q_t	0	0	1.2	2.5	0	3.0	3.0	5.6	6.4

t	9	10	11	12	13	14	15	16
atom	x'_1	y'_1	x'_2	x'_3	y'_2	y'_3	x'_4	y'_4
$\tilde{z}_t$	12.4	13.6	16.1	16.5	17.0	17.4	20.0	23.4
R_t	1	0	1	2	1	0	1	0
S_t	6.0	-7.6	8.5	25.0	8.0	-9.4	10.6	-12.8
p_t	7	8	9	4	11	10	13	14
Q_t	6.6	7.6	9.1	17.6	9.6	9.4	12.2	12.8

The tables make the recurrence concrete: $Q_6 = Q_2 + |S_6 - S_2| = 1.2 + |-3.0 - (-1.2)| = 3.0$. When p_t exists, the minimal chain ending at t starts at $p_t + 1$, one position after its predecessor; for example, $p_2 = 0$ gives $[1, 2] = \{x_1, y_1\}$. The $t = 0$ column is the empty prefix, needed whenever $a = 1$.

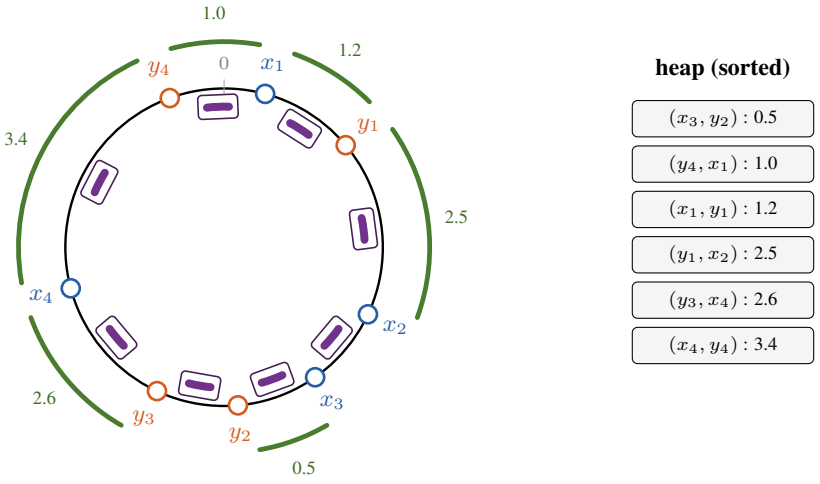
For a balanced interval $[a, b]$, meaning $R_b = R_{a-1}$, the cost is $c_{\mathbb{R}}([a, b]) = Q_b - Q_{a-1}$ (Proposition G.1); this difference is used only for balanced intervals. A candidate cell with endpoints a, b and $b - a < N$ has both a balanced outer interval and a balanced interior, so its marginal is (G.10):

$$\mathfrak{m}([a, b]) = \big(Q_b - Q_{a-1}\big) - \big(Q_{b-1} - Q_a\big).$$

Worked example: the cell $[x_2, y_3] = [3, 6]$ has $\mathfrak{m} = (Q_6 - Q_2) - (Q_5 - Q_3) = (3.0 - 1.2) - (3.0 - 2.5) = 1.8 - 0.5 = \mathbf{1.3}$. This cell does not exist yet — it is created when the three cells around $[x_3, y_2]$ merge at k=1, and it is selected at k=3 (Step 5).

Step 2 — Cells, free gaps, and the initial heap (Appendices E and F)

The inactive atoms partition S^1 into *cells*. Initially every cell is a single gap between cyclic neighbors; the six with opposite-type endpoints are *candidate cells* (green arcs) with marginal equal to their clockwise arc length; (x_2, x_3) and (y_2, y_3) are not candidates. Every cell contains a *free gap* — an original gap never absorbed by a selected cell (Lemma F.1) — and stores one boxed *representative*; initially each cell's representative is its own gap.

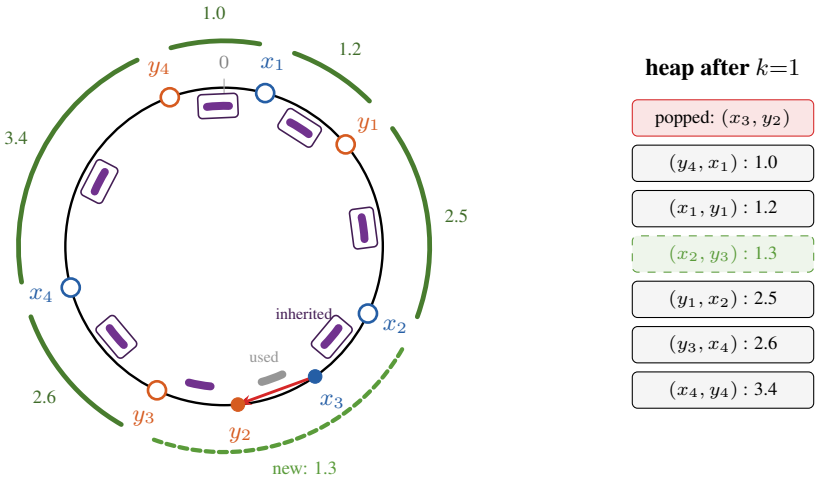


Note the wrap-around candidate (y_4, x_1) across the origin: on the circle it is a perfectly ordinary cell. The two same-type gaps carry free gaps and representatives too — the invariant is about *cells*, not candidates.

Step 3 — Iteration $k = 1$: pop (x_3, y_2); three cells merge into one

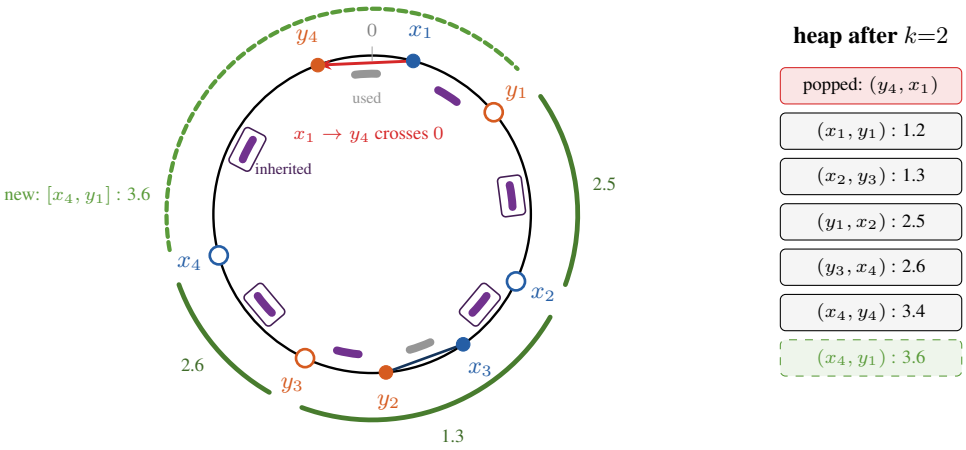
The cheapest cell (x_3, y_2), marginal 0.5, is selected: $x_3 \to y_2$, $C_1^\circ = 0.5$. Cells $[x_2, x_3]$, $[x_3, y_2]$, $[y_2, y_3]$ merge into $[x_2, y_3]$, whose endpoints have opposite types — a new candidate with marginal 1.3 (computed in Step 1). The selected cell's gap turns *used*; the merged cell inherits the boxed

representative of its *left* unselected neighbor $[x_2, x_3]$; the right neighbor's gap (y_2, y_3) stays free but unstored.



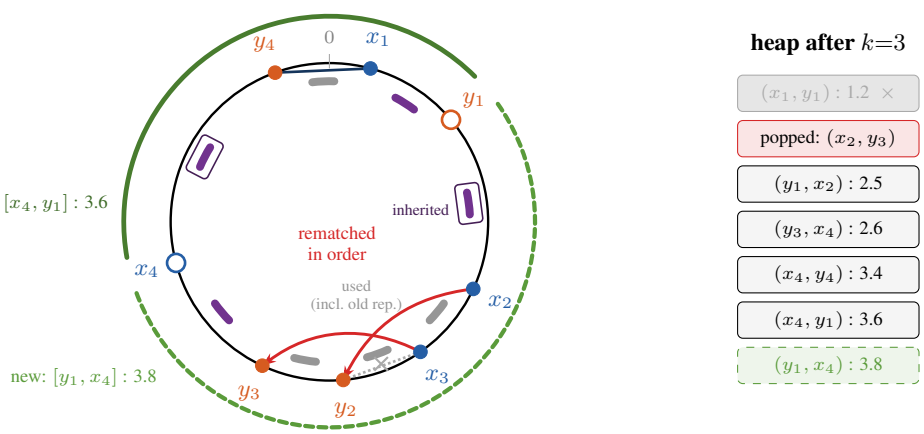
STEP 4 — ITERATION $k = 2$: THE WRAP-AROUND CELL IS SELECTED

Next cheapest is the wrap cell (y_4, x_1), marginal 1.0: the match $x_1 \to y_4$ crosses 0. $C_2^\circ = 1.5$. No cut placed at the top could ever find this plan — this is the step where the circle genuinely departs from every line relaxation cut near 0. Cells $[x_4, y_4]$, $[y_4, x_1]$, $[x_1, y_1]$ merge into $[x_4, y_1]$ (opposite types): new candidate, marginal $\mathfrak{m} = (Q_{10} - Q_6) - (Q_9 - Q_7) = 4.6 - 1.0 = \mathbf{3.6}$, inheriting the representative of $[x_4, y_4]$.



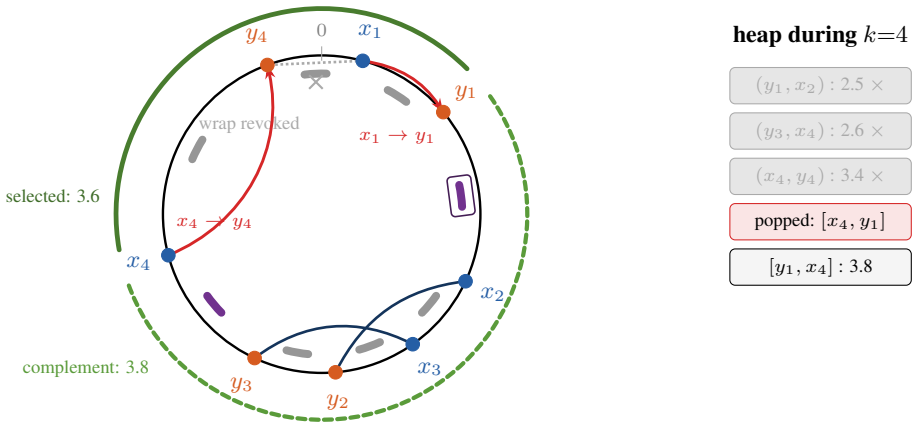
STEP 5 — ITERATION $k = 3$: A LAZY SKIP, THEN THE MERGED CELL FIRES

The heap's top, (x_1, y_1) : 1.2, is *stale*: x_1 was activated by the wrap selection. Lazy deletion discards it on pop (validity check: both endpoints inactive, cyclic successors, opposite types). Next is the merged cell $[x_2, y_3]$: 1.3 — valid. Its endpoints activate and the cell's interior is rematched in unwrapped order: $x_2 \to y_2$, $x_3 \to y_3$, *revoking* $x_3 \to y_2$. $C_3^\circ = 2.8$. Merging $[y_1, x_2] + [x_2, y_3] + [y_3, x_4] \to [y_1, x_4]$ gives a new candidate, marginal $(Q_7 - Q_1) - (Q_6 - Q_2) = 5.6 - 1.8 = \mathbf{3.8}$. Crucially, the selected cell's own boxed representative (x_2, x_3) is *consumed* along with all its gaps — this is exactly why Lemma F.1 makes the merged cell inherit from an *unselected* neighbor: here the left one, $[y_1, x_2]$, donating gap (y_1, x_2).



STEP 6 — ITERATION $k = 4$: THREE SKIPS, THE ENDGAME, AND A CIRCULATION SHIFT

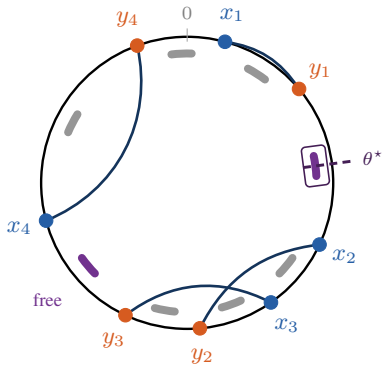
Three stale entries are popped and discarded in a row: (y_1, x_2), (y_3, x_4), (x_4, y_4) — each has an active endpoint. Only two inactive atoms remain, y_1 and x_4, of opposite types: the endgame. They bound *two* directed cells, $[x_4, y_1]$ through the top (marginal 3.6) and its complement $[y_1, x_4]$ through the bottom (3.8); both sit in the heap, and the pop order selects the cheaper automatically. Selecting $[x_4, y_1]$ activates x_4, y_1 and rematches its interior $\{y_4, x_1\}$ in unwrapped order: $x_4 \to y_4$, $x_1 \to y_1$ — *revoking the wrap match* $x_1 \to y_4$. The circulation that was optimal at $k = 2, 3$ is abandoned at $k = 4$: $C_4^\circ = 2.8 + 3.6 = 6.4$.



STEP 7 — THEOREM F.2: THE SIMULTANEOUS CUT $\theta^\star$

Two gaps survive as free: (y_1, x_2) and (y_3, x_4). The representative maintained through every merge — last held by the complementary cell $[y_1, x_4]$ — is $\theta^\star = (y_1, x_2)$: cutting there and running plain PAWL on the unrolled line reproduces C_k° for *every* k at once. The table below (each row an exhaustively computed line profile) shows why this is not automatic: a gap can be free *at some* k yet fail at another. In particular (x_4, y_4) is only consumed at the last step and matches $k \leq 3$ but fails at $k = 4$; and the wrap gap (y_4, x_1), used at $k = 2$, happens to carry no flow again at $k = 4$. A gap that remains free certifies compatibility with every selected update; zero flow at one cardinality is insufficient.

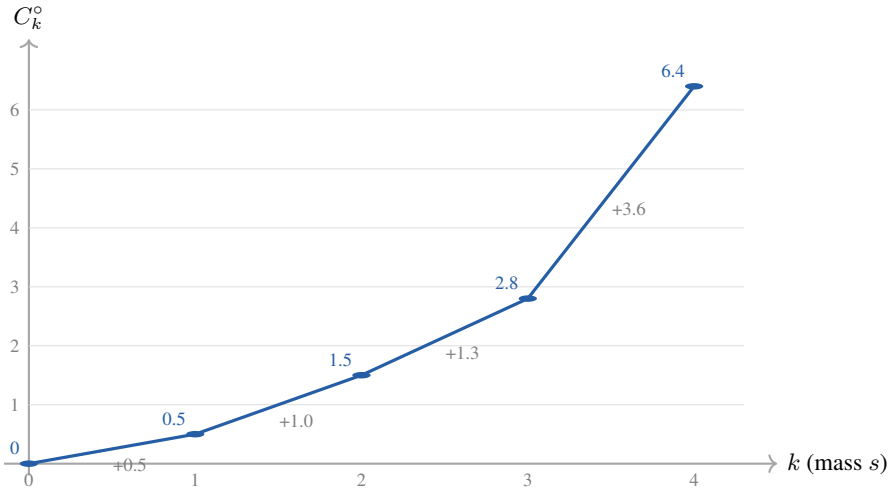
cut in gap	status	k=0	1	2	3	4	verdict
$(y_1, x_2) = \theta^\star$	free (representative)	0	0.5	1.5	2.8	6.4	matches all k
(y_3, x_4)	free	0	0.5	1.5	2.8	6.4	matches all k
(x_4, y_4)	used at k=4	0	0.5	1.5	2.8	**6.6**	fails at k=4
(y_4, x_1)	used at k=2	0	0.5	**1.7**	**3.0**	6.4	fails at k=2, 3
(x_3, y_2)	used at k=1	0	**1.0**	**3.5**	**6.1**	**17.6**	fails from k=1



cut at $\theta^\star$, unroll from x_2, keep atoms with activation rank $\leq k$, match in order $\Rightarrow$ optimal π^k for every k (plan recovery, Appendix I)

STEP 8 — THE OUTPUT: THE FULL CIRCULAR PROFILE

One sweep has produced C_k° for every k; fractional mass follows by linear interpolation by Proposition I.1. The increments 0.5, 1.0, 1.3, 3.6 are nondecreasing. The jump after $k = 3$ illustrates how the marginal profile can inform a later choice of transported mass; it does not itself prescribe a selection rule. The applications in the main paper use prescribed masses.



Summary of the run.

k	popped (skipped)	action	new candidate	rep. inherited	C_k°
1	(x_3, y_2):0.5	$x_3 \to y_2$	$[x_2, y_3]$:1.3	(x_2, x_3)	0.5
2	(y_4, x_1):1.0	$x_1 \to y_4$ *(wrap)*	$[x_4, y_1]$:3.6	(x_4, y_4)	1.5
3	((x_1, y_1)), $[x_2, y_3]$:1.3	rematch $x_2 \to y_2$, $x_3 \to y_3$	$[y_1, x_4]$:3.8	(y_1, x_2)	2.8
4	(3 stale), $[x_4, y_1]$:3.6	rematch; wrap revoked	—	—	6.4

Every marginal was a constant-time lookup in the Q-table of Step 1; every pop, skip, and insert a heap operation on at most $N+k$ entries; the free-gap representatives cost $O(1)$ bookkeeping per merge and delivered $\theta^\star$ for free. Total: $O(N \log N)$, compared with $O(N^2 \log N)$ for cut enumeration. The sweep reproduces the lower envelope of the profiles in Step 7; it does not compute the separate profile for every cut.

B NOTATION AND PROBLEM FORMULATION

Let the circle of circumference $L > 0$ be

$$\mathbb{S}^1_L = \mathbb{R}/L\mathbb{Z},$$

and represent its points by coordinates in $[0, L)$. Its geodesic distance is

$$d_{\mathbb{S}^1}(x, y) = \min\{|x - y|,\ L - |x - y|\}. \tag{B.1}$$

Consider two empirical measures

$$\mu = w \sum_{i=1}^{n} \delta_{x_i}, \qquad \nu = w \sum_{j=1}^{m} \delta_{y_j}, \qquad w > 0. \tag{B.2}$$

Assumption B.1 (Standing assumptions)**.** *All atoms have the same mass w; the points in $\{x_1, \dots, x_n, y_1, \dots, y_m\}$ are pairwise distinct; and the ground cost is the geodesic distance* (B.1)*. Let $N = n + m$ and $K = \min(n, m)$.*

For a transported mass $s \in [0, Kw]$, define

$$\mathrm{PW}_\circ(s) = \min_{\pi \in \mathbb{R}_+^{n \times m}} \sum_{i=1}^{n} \sum_{j=1}^{m} d_{\mathbb{S}^1}(x_i, y_j)\pi_{ij} \tag{B.3}$$

subject to

$$\sum_j \pi_{ij} \le w, \qquad \sum_i \pi_{ij} \le w, \qquad \sum_{i,j} \pi_{ij} = s. \tag{B.4}$$

At an integer transported mass $s = kw$, define the rescaled transport plan

$$\gamma_{ij} \coloneqq \frac{\pi_{ij}}{w}.$$

The feasible set then becomes

$$\mathcal{P}_k = \left\{ \gamma \in \mathbb{R}_+^{n \times m} \;\middle|\; \sum_{j=1}^{m} \gamma_{ij} \le 1,\ \forall i, \quad \sum_{i=1}^{n} \gamma_{ij} \le 1,\ \forall j, \quad \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} = k \right\}.$$

This is the cardinality-k bipartite matching polytope. Its extreme points are integral: at each extreme point,

$$\gamma_{ij} \in \{0, 1\}.$$

Since the objective is linear, an optimum is attained at an extreme point of $\mathcal{P}_k$. Therefore, there exists an optimal transport plan satisfying

$$\pi_{ij} \in \{0, w\}.$$

Moreover, the constraint

$$\sum_{i,j} \gamma_{ij} = k$$

implies that exactly k entries of γ are equal to one, while the row and column constraints ensure that no two selected entries share the same source or target. Hence, an optimal solution can be represented by a cardinality-k matching between k distinct source atoms and k distinct target atoms. We therefore write

$$C_k^\circ = \min_{M:\,|M|=k} w \sum_{(x_i, y_j) \in M} d_{\mathbb{S}^1}(x_i, y_j), \qquad k = 0, \dots, K, \tag{B.5}$$

where no source or target atom occurs in more than one edge of M.

The *active set* of a matching is

$$A(M) = \{x_i : \exists j, (x_i, y_j) \in M\} \cup \{y_j : \exists i, (x_i, y_j) \in M\}. \tag{B.6}$$

When only the active set matters, we write A_k for an optimal set containing k source and k target atoms.

Mass accounting. Figure 1 provides an accounting interpretation of the fixed-cardinality partial transport problem. At cardinality k, exactly kw units of mass are transported, while $(n-k)w$ units of source mass remain unmatched and $(m-k)w$ units of target capacity remain unfilled. These latter quantities may be interpreted as destroyed and created mass, respectively, and represented by arcs to dummy reservoirs. However, because k is fixed in (B.5), the total amounts of destruction and creation are prescribed. Consequently, constant per-unit destruction and creation penalties contribute only an additive constant, and PAWC needs to optimize only the geodesic transportation cost.

C Background: the line structure and the cut envelope

C.1 The line structure inherited from PAWL

On the line, uniformly weighted partial W_1 admits a particularly rigid combinatorial structure that makes it possible to compute the entire transported-mass profile efficiently. In particular, optimal active sets can be chosen to grow monotonically with the transported cardinality, and the transition from a cardinality-k solution to a cardinality-$(k+1)$ solution activates exactly one source atom and one target atom. Moreover, these newly activated atoms must be adjacent after removing the current active set, so each admissible update is represented by a balanced contiguous chain. For the absolute-value ground cost, the marginal cost of activating such a chain can be recovered from precomputed chain costs in constant time. These structural and computational properties form the basis of PAWL (Chapel & Tavenard, 2025) and are summarized below, as they will serve as the starting point for our extension from the line to the circle.

Throughout, all line costs are written with a single symbol. Let P be a finite set of support points carrying equally many source and target atoms, all of common weight w, and let $x_{(1)} < \cdots < x_{(j)}$ and $y_{(1)} < \cdots < y_{(j)}$ be its source and target atoms in increasing order. Such a P is called *balanced*, and we write

$$c_{\mathbb{R}}(P) = w \sum_{i=1}^{j} \left| x_{(i)} - y_{(i)} \right| \tag{C.1}$$

for its one-dimensional W_1 transport cost, the increasing-order matching being optimal on the line. We apply $c_{\mathbb{R}}$ to whatever set of atoms is under discussion and abbreviate accordingly: $c_{\mathbb{R}}(I)$ for the atoms lying in an interval I, $c_{\mathbb{R}}(A)$ for those of an active set A, $c_{\mathbb{R}}(\mathcal{C})$ for those of a circular cell $\mathcal{C}$ once a cut outside it has been fixed, and $c_{\mathbb{R}}([a,b])$ for those at positions a through b of the doubled sequence of Appendix G.1. In every case the argument is balanced, so (C.1) applies verbatim; only the description of which atoms are meant changes.

Lemma C.1 (Universal line-neighbor and locality principle)**.** *Consider two optimal active sets for uniformly weighted partial W_1 on the line satisfying*

$$A_{k+1} = A_k \cup \{u, v\},$$

where $u, v \notin A_k$ belong to opposite measures. Then u and v are consecutive in the ordered inactive set A_k^c. Consequently, the open interval between u and v contains only active atoms and is balanced. Moreover, passing from A_k to A_{k+1} changes the increasing-order optimal matching only inside the closed interval $[u, v]$; every matching edge outside this interval is preserved.

Proof. We give a direct proof of the line structure underlying Chapel & Tavenard (2025, Proposition 2 and Section 3.3). First, no matched segment in a globally optimal cardinality-k line matching can contain an inactive support atom z in its open interior. Replacing the matched endpoint of the same type as z by z preserves feasibility and cardinality, but strictly shortens that edge. Distinct support locations and $w > 0$ make the cost decrease strict, contradicting optimality.

A *maximal active block* is a maximal consecutive run of active atoms in the sorted union support. No edge of an optimal matching can join two such blocks, since it would cross an intervening inactive atom. Thus every maximal active block of an optimal active set is matched internally and is balanced. This holds for every optimal active set, without a uniqueness assumption.

Relabel the new atoms so that $u < v$. If an inactive atom remains in (u, v), the maximal active block of A_{k+1} containing u cannot contain v. Adding atoms never splits an old active block, and the inactive atom in (u, v) stays inactive in A_{k+1}, so the run of active atoms to the right of u stops before

v; this new block is therefore a union of whole maximal active blocks of A_k together with exactly one new atom, u. The old blocks are balanced; adding u gives imbalance $+1$ or -1. This contradicts the balance of every maximal active block of A_{k+1}. Hence u and v are consecutive in A_k^c.

Their open interval is empty or one maximal active block of A_k, so it is balanced. The active portions strictly before u and strictly after v are also unions of balanced blocks. Increasing-order matching treats these three consecutive balanced portions independently. After adding u, v, the middle portion becomes the balanced interval $[u, v]$ and the exterior portions are unchanged. Thus the only changed pairs lie in $[u, v]$, and the cost increment is $c_{\mathbb{R}}([u, v]) - c_{\mathbb{R}}((u, v))$. The same argument covers empty interior or exterior portions, including $k = 0$. □

Remark C.1 (Why the universal form matters)**.** The strict exchange argument above rules out every non-neighbor optimal nested extension; the lemma is therefore stronger than the existence of one suitably chosen nested sequence. This universal form is used in Theorem D.2, where the line-neighbor property must be applied to the particular nested pair obtained after choosing a common cut.

For a balanced line interval I and its interior I°, define

$$\mathfrak{m}(I) = c_{\mathbb{R}}(I) - c_{\mathbb{R}}(I^\circ), \tag{C.2}$$

with $c_{\mathbb{R}}$ as in (C.1). Under the hypotheses of Lemma C.1, this is the exact global line-cost increment for the interval between the newly activated endpoints. Balanced interior alone does not imply this global identity. This is the chain marginal used by PAWL (Chapel & Tavenard, 2025).

C.2 Circular flow, circulation, and cuts

This subsection concerns the balanced transport problem induced by a fixed active set. Specifically, let A contain k source atoms and k target atoms. The restricted measures

$$\mu_A = w \sum_{x_i \in A \cap \operatorname{supp} \mu} \delta_{x_i}, \qquad \nu_A = w \sum_{y_j \in A \cap \operatorname{supp} \nu} \delta_{y_j}$$

have equal total mass kw, so their transport cost is an ordinary balanced 1-Wasserstein problem on the circle. The partial nature of the overall problem enters through the subsequent optimization over the choice of the active set A.

To characterize one-dimensional transport on the circle, we first introduce the elementary arcs between consecutive support points and the cumulative mass imbalance associated with them. Let

$$z_1, \dots, z_N, \qquad N = n + m,$$

denote the cyclically ordered union of the source and target support points. In this circular representation, support indices are understood cyclically, so that $z_{N+1} = z_1$. Indices in the doubled sequence of Appendix G.1 instead refer to its ordinary linear order. The *original support gaps* are the open circular arcs

$$G_i \coloneqq (z_i, z_{i+1})_\circ, \qquad i = 1, \dots, N,$$

and their lengths are denoted by

$$\ell_i \coloneqq |G_i| = d_{\mathrm{cw}}(z_i, z_{i+1}),$$

where d_{cw} is the clockwise arc length. By construction, no source or target atom lies in the interior of a support gap. Under our distinct-support assumption, every gap has positive length and

$$\sum_{i=1}^{N} \ell_i = L.$$

These original support gaps should not be confused with the *current cells* introduced later: a current cell is an arc between consecutive inactive atoms and may contain several original support gaps as well as active atoms.

Now fix a balanced active set A, containing the same number of source and target atoms. Choose an arbitrary origin θ in one of the support gaps and unwrap the circle at θ, thereby identifying $\mathbb{S}^1_L$ with the interval $[0, L)$. Define the cumulative imbalance

$$H_A(t) = \#\big\{x \in A \cap \operatorname{supp} \mu : \ x \in [0, t)\big\} - \#\big\{y \in A \cap \operatorname{supp} \nu : \ y \in [0, t)\big\}.$$

Thus, $H_A(t)$ records the net number of active source atoms minus active target atoms encountered while moving clockwise from the cut θ to t. It increases by one when an active source atom is crossed, decreases by one when an active target atom is crossed, and is unchanged when an inactive atom is crossed. Consequently, H_A is integer-valued and constant on the interior of every original support gap G_i. We denote this constant value by

$$H_A(G_i).$$

Unlike on the line, the cumulative imbalance on the circle depends on where the circle is cut. Equivalently, a feasible circular flow is determined only up to an arbitrary constant circulation around the cycle. Accounting for this degree of freedom gives the standard one-dimensional circular W_1 representation

$$W_{1,\circ}(A) = w \min_{a\in\mathbb{R}} \int_0^L |H_A(t) - a| \, dt = w \min_{a\in\mathbb{R}} \sum_{i=1}^N \ell_i \, |H_A(G_i) - a| \, . \tag{C.3}$$

The scalar a represents the circulation level. The second equality follows because H_A is constant on each support gap. Hence, a minimizer $a^\star$ is a weighted median of the values $\{H_A(G_i)\}_{i=1}^N$, with each value weighted by the corresponding gap length ℓ_i (Delon et al., 2010; Rabin et al., 2011). Since all values $H_A(G_i)$ are integers, an integer-valued minimizing median may be chosen.

The same representation admits a direct interpretation as a minimum-cost flow problem on the cycle. Orient each original support gap G_i clockwise, and let f_i denote the signed amount of mass flowing through that gap, with $f_i > 0$ corresponding to clockwise flow. Let

$$b_i = \begin{cases} +1, & z_i \in A \cap \operatorname{supp}\mu, \\ -1, & z_i \in A \cap \operatorname{supp}\nu, \\ 0, & z_i \notin A, \end{cases}$$

denote the signed activity at the ith support point. Conservation of mass imposes the discrete divergence constraints

$$f_i - f_{i-1} = w b_i,$$

with cyclic indexing. These constraints determine the differences between successive edge flows, but they do not determine their common additive offset: adding the same constant circulation to every f_i preserves all divergence constraints. Consequently, every feasible circular flow can be written as

$$f_i(a) = w\big(H_A(G_i) - a\big)$$

for some scalar circulation level a. Since transporting an amount $|f_i(a)|$ across a gap of length ℓ_i incurs cost $\ell_i|f_i(a)|$, the minimum flow cost is

$$\min_{a\in\mathbb{R}} \sum_{i=1}^N \ell_i |f_i(a)| = w \min_{a\in\mathbb{R}} \sum_{i=1}^N \ell_i \, |H_A(G_i) - a| \, ,$$

which recovers (C.3).

This flow formulation makes the freedom in the choice of cut explicit: the circulation level a is a free primal variable, and every value of it gives a feasible flow. Because the minimizing circulation $a^\star$ may be chosen as a weighted median of the finitely many values $\{H_A(G_i)\}_{i=1}^N$, it can be chosen equal to $H_A(G_r)$ for at least one support gap G_r. The corresponding optimal flow then satisfies

$$f_r = w\big(H_A(G_r) - a^\star\big) = 0.$$

No mass crosses G_r, so cutting the circle anywhere inside this gap does not interrupt the optimal transport flow and reduces the circular problem to an equivalent transport problem on the line (Delon et al., 2010; Rabin et al., 2011).

Proposition C.1 (Cut representation for a fixed active set)**.** *For each original support gap G_r, choose any $\theta_r \in G_r$, cut the circle at θ_r, and unwrap it onto an interval of length L. Let $W_{1,r}^{\mathrm{line}}(A)$ denote the line W_1 cost of the active set A in this unwrapped representation. Then, for every balanced active set A,*

$$W_{1,\circ}(A) = \min_{r=1,\dots,N} W_{1,r}^{\mathrm{line}}(A). \tag{C.4}$$

The value $W_{1,r}^{\mathrm{line}}(A)$ does not depend on the precise location of θ_r inside G_r, because moving the cut within a gap changes all unwrapped coordinates by the same translation and does not alter their linear order.

Proof. A cut in gap G_r fixes the flow through that gap to zero, which corresponds in (C.3) to choosing $a = H_A(G_r)$. Conversely, an attained minimizing median $a^\star$ equals H_A on at least one positive-length gap; cutting there realizes the circular optimum as a line optimum. □

Corollary C.1 (Exact cut envelope)**.** *For each support gap G_r, let $C^{\mathrm{line}}_{k,r}$ denote the minimum transportation cost, with the absolute-value ground cost on the unwrapped line, among all matchings that pair exactly k source atoms with k target atoms. Then*

$$C^\circ_k = \min_{r=1,\dots,N} C^{\mathrm{line}}_{k,r}. \tag{C.5}$$

Consequently, running PAWL *for all N possible cuts yields an exact $O(N^2 \log N)$ reference solver for the complete circular partial-transport profile.*

Proof. Using Proposition C.1 and exchanging two finite minima,

$$C^\circ_k = \min_{A:|A\cap \mathrm{supp}\,\mu|=|A\cap \mathrm{supp}\,\nu|=k} \min_r W^{\mathrm{line}}_{1,r}(A) = \min_r C^{\mathrm{line}}_{k,r}.$$

Each line run returns all k in $O(N \log N)$ time (Chapel & Tavenard, 2025). □

D NESTED OPTIMAL ACTIVE SETS AND CYCLIC NEIGHBOURS

The purpose of this section is to establish the two structural properties that make an efficient incremental algorithm possible on the circle. First, we show that optimal solutions can be selected consistently across transported cardinalities: starting from an optimal cardinality-k solution, one can construct an optimal cardinality-$(k+1)$ solution by activating exactly one additional source atom and one additional target atom. Thus, the associated active sets may be chosen to form a nested sequence. Second, we exploit the one-dimensional geometry of the circle to sharply restrict which inactive source–target pair can be activated next. We show that the two newly activated atoms may be chosen to be consecutive in the cyclic ordering of the inactive support points.

These results replace the quadratic search over all inactive source–target pairs by a linear-size family of local candidates. The nestedness property is derived first from the successive-augmentation structure of min-cost flow, while the cyclic-neighbor property is obtained by cutting the circle at a suitable gap and reducing the transition to the corresponding line problem studied by PAWL.

D.1 NESTED OPTIMAL ACTIVE SETS

A priori, optimal matchings at two consecutive cardinalities need not appear compatible: an optimal cardinality-k matching and an independently computed optimal cardinality-$(k+1)$ matching may involve substantially different source and target atoms. For an incremental algorithm, however, we need a stronger statement. Namely, we need to know that the optima can be *chosen* so that increasing the transported cardinality preserves every previously active atom and activates exactly one additional source atom and one additional target atom. This nestedness property is the analogue of the monotone active-set structure exploited by PAWL on the line (Chapel & Tavenard, 2025).

For a matching M, recall that its active set is

$$A(M) \coloneqq \{x_i : \exists\, y_j \text{ such that } (x_i, y_j) \in M\} \cup \{y_j : \exists\, x_i \text{ such that } (x_i, y_j) \in M\}.$$

Thus, if M has cardinality k, then $A(M)$ contains exactly k source atoms and k target atoms.

Theorem D.1 (Nested optimal extension)**.** *Let M_k be any optimal cardinality-k matching, with $k < K$. Then there exists an optimal cardinality-$(k+1)$ matching M_{k+1} such that*

$$A(M_{k+1}) = A(M_k) \cup \{x, y\}, \tag{D.1}$$

where $x \in A(M_k)^c$ is a source atom and $y \in A(M_k)^c$ is a target atom.

Proof. We represent the cardinality-constrained matching problem (B.5) as a unit-capacity minimum-cost flow problem. Introduce a super-source node s, a super-sink node t, one node for each source atom x_i, and one node for each target atom y_j. The network contains the arcs

$$s \longrightarrow x_i, \qquad x_i \longrightarrow y_j, \qquad y_j \longrightarrow t,$$

all with unit capacity. The arcs adjacent to s and t have zero cost, while the source-to-target arc $x_i \to y_j$ has cost

$$c_{ij} = w\, d_{\mathbb{S}^1}(x_i, y_j).$$

Because every arc has integral capacity, an integral flow of value k selects exactly k source-to-target arcs, with at most one selected arc incident to each source or target atom. It therefore corresponds exactly to a cardinality-k matching. Conversely, every cardinality-k matching defines an integral s-t flow of value k. Hence the minimum-cost flow problem of value k is equivalent to (B.5), and the given optimal matching M_k corresponds to an optimal integral flow f_k of value k.

Consider the residual network of f_k. For every unused source-to-target arc $x_i \to y_j$, the residual network contains a forward arc of cost c_{ij}. For every matched pair $(x_i, y_j) \in M_k$, it contains the reverse residual arc

$$y_j \longrightarrow x_i$$

of cost $-c_{ij}$. The latter arc represents the possibility of removing the currently selected matching edge during an augmentation.

Since f_k is optimal, its residual network contains no negative-cost directed cycle. Moreover, because $k < K$, there exist an unmatched source atom x_i and an unmatched target atom y_j; the path $s \to x_i \to y_j \to t$ is then residual, so an s–t residual path exists and a minimum-cost one is well defined. Equivalently, one may introduce feasible node potentials and compute the path using nonnegative reduced costs. By the successive shortest augmenting-path theorem, augmenting one unit of flow along such a path produces an optimal flow f_{k+1} of value $k+1$ (Ahuja et al., 1993). Importantly, this conclusion requires only that f_k be optimal, and not that it have been obtained through any particular sequence of previous augmentations. We may choose the shortest augmenting path to be simple, so it does not revisit the super-source or super-sink.

The first atom visited after s must be an unmatched source atom. Indeed, the arc $s \to x_i$ has forward residual capacity only when x_i is not already matched. Similarly, the final atom visited before t must be an unmatched target atom. Denote these two endpoints by x and y, respectively.

Between these endpoints, the augmenting path alternates between unused forward matching edges and reversed currently matched edges. Its source–target portion therefore has the form

$$x \longrightarrow y_1 \longrightarrow x_1 \longrightarrow y_2 \longrightarrow x_2 \longrightarrow \cdots \longrightarrow y,$$

where each forward edge $x_r \to y_{r+1}$ is inserted into the matching and each reverse edge $y_r \to x_r$ removes an edge of M_k. Augmenting along the path is thus equivalent to toggling the matching status of all edges on this alternating path.

Every internal source or target atom on the path loses one matched edge but simultaneously gains another. Consequently, it remains matched after the augmentation. Atoms that are not on the path are unaffected. Only the two endpoints x and y gain a matched edge without losing one, because they were unmatched before the augmentation. Therefore, the resulting optimal matching M_{k+1} satisfies

$$A(M_{k+1}) = A(M_k) \cup \{x, y\},$$

which proves (D.1). □

Corollary D.1 (Nested optimal sequence)**.** *There exist optimal matchings*

$$M_0, M_1, \ldots, M_K, \qquad K = \min\{n, m\},$$

such that

$$A(M_k) \subset A(M_{k+1}) \qquad \textit{for every } k < K,$$

where $A(M_{k+1}) \setminus A(M_k)$ consists of exactly one source atom and one target atom.

Proof. The empty matching M_0 is the unique, hence optimal, cardinality-0 matching. Applying Theorem D.1 successively for $k = 0, \ldots, K-1$ constructs the desired sequence. □

Remark D.1 (Active-set nestedness versus edge nestedness)**.** The nested-extension property concerns the active atoms, not necessarily the matching edges. Although

$$A(M_k) \subset A(M_{k+1}),$$

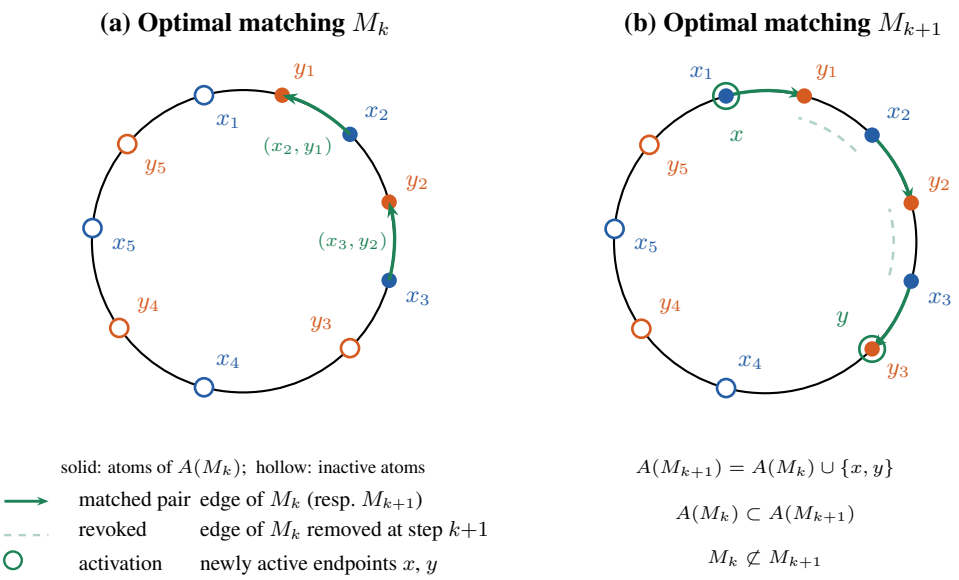


Figure D.1: Nested optimal active sets (Theorem D.1). **(a)** An optimal cardinality-k matching M_k on the circle, with matched pairs (x_2, y_1) and (x_3, y_2) transported along geodesic arcs; solid atoms form the active set $A(M_k)$ and hollow atoms are inactive. **(b)** An optimal cardinality-$(k{+}1)$ matching obtained by augmenting along a minimum-cost alternating path: exactly one source atom $x = x_1$ and one target atom $y = y_3$ are activated (circled), and they are cyclic neighbors among the inactive atoms since every atom between them is already active. The augmentation revokes the edges (x_2, y_1) and (x_3, y_2) of M_k (dashed) and rematches the previously active atoms as (x_1, y_1), (x_2, y_2), (x_3, y_3), so that $A(M_{k+1}) = A(M_k) \cup \{x, y\}$ while $M_k \not\subset M_{k+1}$ (Remark D.1).

one need not have

$$M_k \subset M_{k+1}.$$

Indeed, the augmenting path in the proof of Theorem D.1 may remove several edges of M_k and replace them with different edges; Figure D.1 shows an instance in which two edges of M_k are revoked and three new ones created, while every previously active atom remains active. Every internal atom on the augmenting path nevertheless remains matched, since it loses one incident matching edge and gains another. Consequently, all previously active source and target atoms remain active, while only the two endpoints of the augmenting path become newly active.

This distinction is important for the developments below: the local rematching of previously active atoms is precisely what gives rise to the chain and cell marginal costs. Moreover, Theorem D.1 is stronger than the mere existence of one nested optimal sequence, since it applies to any prescribed optimal cardinality-k matching. This stronger extension form is essential for the greedy analysis, where the theorem must be applied to the particular optimal active set produced by the algorithm at iteration k.

Remark D.2 (Independence from the circular geometry)**.** The proof of Theorem D.1 uses only the bipartite matching and minimum-cost flow structure; it does not rely on any special property of the geodesic cost or of the circle. Hence, the nested optimal-extension property holds for an arbitrary bipartite cost matrix. The circular geometry becomes essential only in the next step, where we show that the two newly activated atoms are cyclic neighbors among the inactive atoms and that the resulting cost increment can be evaluated locally.

D.2 A PAIRWISE COMMON-CUT LEMMA

The next lemma is the bridge between nested circular optima and the line neighbor theorem.

Lemma D.1 (Common optimal cut)**.** *Let A be balanced and let*

$$A' = A \cup \{x, y\},$$

where $x \notin A$ is a source and $y \notin A$ is a target. There exists a support gap G that carries zero flow in an optimal circular W_1 solution for both A and A'. Equivalently, cutting in G realizes both circular costs as line costs.

Proof. Let $I = [x, y]_{\circlearrowright}$ be the clockwise arc from the new source x to the new target y. Up to adding a constant to the cumulative imbalance of A', which is absorbed by the minimization in (C.3), we may write

$$H_{A'} = H_A + \mathbf{1}_I. \tag{D.2}$$

Let a be an attained weighted median of H_A, so

$$|\{H_A < a\}| \le L/2, \qquad |\{H_A > a\}| \le L/2, \tag{D.3}$$

where $|\cdot|$ denotes circular arc length. Set $E = \{H_A = a\}$.

If a is also a median of $H_{A'}$ and $|E \setminus I| > 0$, choose a gap in $E \setminus I$. On that gap, both cumulative functions equal a.

Otherwise, we show that $a + 1$ is a median of $H_{A'}$ and that $|E \cap I| > 0$. If a is not a median of $H_{A'}$, then because $H_{A'} \ge H_A$, only the upper median condition can fail. Hence $|\{H_{A'} > a\}| > L/2$, and the excess over $\{H_A > a\}$ must come from $E \cap I$, which therefore has positive length. Since $H_{A'}$ is integer-valued,

$$\{H_{A'} < a + 1\} = \{H_{A'} \le a\}$$

has length below $L/2$, while

$$\{H_{A'} > a + 1\} \subseteq \{H_A > a\}$$

has length at most $L/2$. Thus $a + 1$ is a median.

The remaining case is that a is a median of $H_{A'}$ but $|E \setminus I| = 0$. Then $|E \cap I| > 0$. Moreover,

$$\{H_{A'} < a + 1\} \subseteq \{H_A < a\} \cup (E \setminus I), \qquad \{H_{A'} > a + 1\} \subseteq \{H_A > a\},$$

so (D.3) again shows that $a + 1$ is a median of $H_{A'}$.

Choose a positive-length gap in $E \cap I$. On it, $H_A = a$ and $H_{A'} = a + 1$. Centering each cumulative imbalance at its respective median gives zero flow for both active sets on the chosen gap. Figure D.2 shows this second case: the two cumulative imbalances sit at different medians, and the gap chosen from $E \cap I$ is flow-free for both. □

D.3 Cyclic neighbor addition

Nestedness alone does not yet yield an efficient update rule. Although Theorem D.1 guarantees that an optimal cardinality-$(k + 1)$ active set can be obtained from an optimal cardinality-k active set by adding one source atom and one target atom, a direct search would still have to examine all opposite-type pairs in A_k^c, resulting in a quadratic number of candidates. The circular order provides a much stronger restriction: the two new atoms may be chosen adjacent after the currently active atoms are removed. Thus, at iteration k, it is sufficient to consider only opposite-type pairs that are consecutive in the cyclic ordering of A_k^c. This reduces the candidate family to linear size and, more importantly, identifies each candidate with a single circular cell whose interior contains only active atoms. The result is the circular counterpart of the locality property underlying PAWL on the line.

Theorem D.2 (Cyclic-neighbor property of optimal extensions)**.** *Let M_k be any circularly optimal cardinality-k matching, with $k < K$, and write $A_k := A(M_k)$. Let M_{k+1} be any circularly optimal cardinality-$(k + 1)$ matching satisfying*

$$A(M_{k+1}) = A_k \cup \{x, y\}, \tag{D.4}$$

where $x \in A_k^c$ is a source atom and $y \in A_k^c$ is a target atom. Then x and y are consecutive in the cyclic ordering of the inactive set A_k^c.

In particular, by Theorem D.1, every circularly optimal cardinality-k matching admits a circularly optimal cardinality-$(k + 1)$ extension whose two newly active atoms are cyclic neighbors.

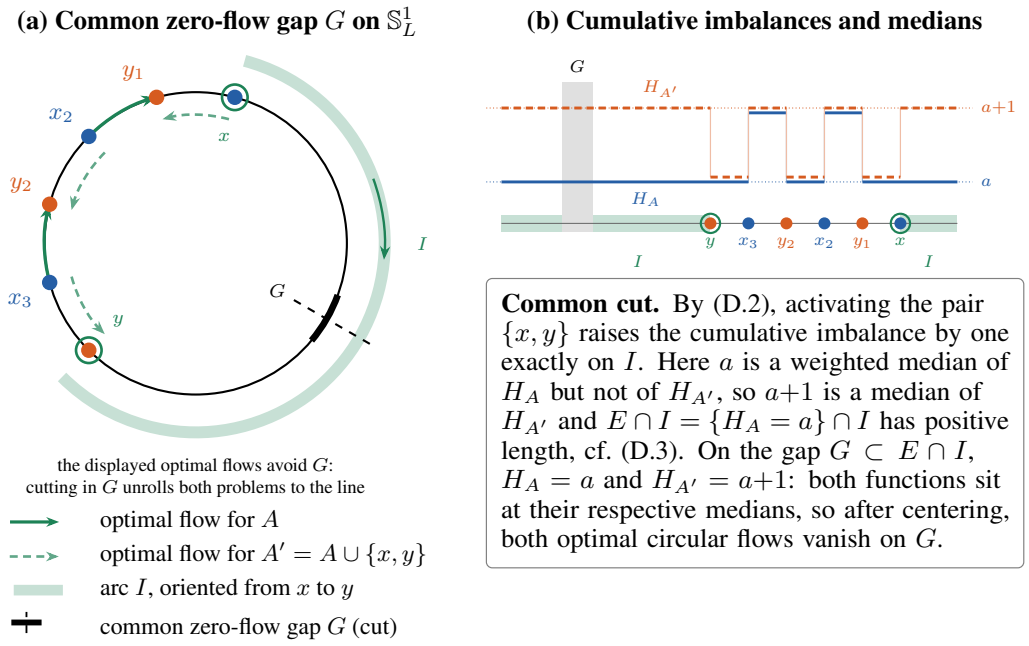


Figure D.2: Common optimal cut (Lemma D.1). **(a)** A balanced active set $A = \{x_2, y_1, x_3, y_2\}$ and its extension $A' = A \cup \{x, y\}$ on the circle. The arc I, oriented from x to y, supports the one-interval update $H_{A'} = H_A + \mathbf{1}_I$. The support gap G carries zero flow in an optimal circular W_1 solution for both A (solid arcs) and A' (dashed arcs), so cutting in G realizes both circular costs as line costs. **(b)** The cumulative imbalances, unrolled at a point of the large support gap. H_A is flat at its median a on $E = \{H_A = a\}$, while $H_{A'} = H_A + \mathbf{1}_I$ is flat at its median $a+1$ on $E \cap I$; the gap $G \subset E \cap I$ is at median level for both, hence flow-free for both. The cumulative functions are shown up to additive constants; the dashed curve has a small vertical display offset for visibility.

Proof. Let M_k and M_{k+1} be as in the statement, and set

$$A_{k+1} := A(M_{k+1}) = A_k \cup \{x, y\}.$$

It remains to show that x and y are cyclically adjacent among the atoms that are inactive at stage k.

By Lemma D.1, there exists an original support gap G_r such that cutting the circle at any point $\theta_r \in G_r$ and unwrapping it onto an interval of length L realizes the circular transport costs of both active sets as line transport costs:

$$W^{\text{line}}_{1,r}(A_k) = W_{1,\circ}(A_k), \qquad W^{\text{line}}_{1,r}(A_{k+1}) = W_{1,\circ}(A_{k+1}).$$

Since M_k and M_{k+1} are circularly optimal,

$$W_{1,\circ}(A_k) = C^\circ_k, \qquad W_{1,\circ}(A_{k+1}) = C^\circ_{k+1},$$

and therefore

$$W^{\text{line}}_{1,r}(A_k) = C^\circ_k, \qquad W^{\text{line}}_{1,r}(A_{k+1}) = C^\circ_{k+1}.$$

We next verify that A_k and A_{k+1} are optimal for the corresponding cardinality-constrained partial-transport problems on the unwrapped line. For clarity, let

$$\text{cost}^\circ(M) = w \sum_{(x_i, y_j) \in M} d_{\mathbb{S}^1}(x_i, y_j)$$

and let $\text{cost}^{\text{line}}_r(M)$ denote the analogous matching cost computed using the distances in the representation obtained by cutting at G_r.

Consider any cardinality-k matching M'. For each matched pair, the circular geodesic distance is no larger than its distance in the fixed unwrapped representation. Hence

$$\text{cost}^{\text{line}}_r(M') \geq \text{cost}^\circ(M').$$

Circular optimality of M_k, together with the fact that the common cut realizes its circular cost, gives

$$\mathrm{cost}_r^{\mathrm{line}}(M') \geq \mathrm{cost}^{\circ}(M') \geq C_k^{\circ} = W_{1,r}^{\mathrm{line}}(A_k).$$

Thus no cardinality-k line matching is cheaper than the optimal line matching supported on A_k. Therefore A_k is optimal for the cardinality-k partial-transport problem induced by the cut. The same argument, with $k+1$ in place of k, proves that A_{k+1} is optimal for the corresponding cardinality-$(k+1)$ line problem.

We may now apply the universal line-neighbor property of Lemma C.1 to the nested pair of line-optimal active sets

$$A_k \subset A_{k+1}.$$

It follows that the two newly activated atoms x and y are consecutive in the linear ordering of the atoms in A_k^c induced by the cut.

Finally, the linear inactive order is obtained by breaking the cyclic inactive order at a single location. This operation can remove the one cyclic adjacency crossing the cut, but it cannot create any new adjacency. Hence every pair that is consecutive in the linear inactive order is also consecutive in the original cyclic inactive order. Therefore x and y are cyclic neighbors in A_k^c. □

E CIRCULAR CELLS AND LOCAL MARGINALS

The previous section established the combinatorial structure of an optimal cardinality-by-cardinality construction. In particular, optimal active sets can be chosen to form a nested sequence, and the transition from A_k to A_{k+1} may be realized by activating a source atom and a target atom that are consecutive in the cyclic ordering of the inactive set A_k^c. This reduces the admissible updates from all opposite-type inactive pairs to a linear-size family of cyclic-neighbor candidates.

For an efficient algorithm, we need a local quantity for each candidate that can be evaluated without solving a new circular transport problem. Consecutive inactive points partition the circle into *current cells*, whose interiors contain only active atoms. For an opposite-type candidate, we define its local marginal as the line transport cost of the full cell minus that of its active interior. This quantity is intrinsic to the cell's directed arc and is computable from its atoms alone.

Its relation to a global cost requires a compatible cut: the cut must avoid both the candidate and every previously selected cell. Under that condition, activating the candidate changes the sorted line matching only inside its cell, as proved in Lemma F.2. The resulting line matching is feasible on the circle with no greater geodesic cost. An arbitrary candidate marginal need not equal the increment in optimal circular cost; Theorem F.1 will show that the *minimum* candidate marginal is exactly the next optimal circular increment. The chain preprocessing of PAWL then permits constant-time marginal evaluation.

Let $U_k = A_k^c$ be the current inactive set, written in cyclic order as

$$u_1, u_2, \ldots, u_r.$$

Each directed pair of consecutive inactive points defines a *cell*

$$\mathcal{C}_i = [u_i, u_{i+1}]_{\circlearrowright}, \qquad u_{r+1} = u_1, \tag{E.1}$$

where the arc is traversed clockwise and includes both endpoints. The open cell interior contains only active points. With two inactive atoms, the two complementary directed cells are distinct. With one inactive atom, its sole cell traverses the full circle back to that atom; this self-cell is not a candidate. With no inactive atoms, there are no current cells.

Definition E.1 (Candidate cell)**.** A current cell is a *candidate* if its two inactive endpoints come from opposite measures.

A *selection* activates the two endpoints of a cell, and it is *valid* if the cell is a current candidate at the moment it is selected. The algorithm makes only valid selections; the lemmas below that assume a history of valid selections require nothing about which candidate was chosen at each step.

Lemma E.1 (Balanced cell interiors)**.** *At every iteration of the algorithm, the active atoms in the interior of each current cell contain the same number of source and target atoms.*

Proof. Initially all cell interiors are empty. Suppose a candidate cell $[u, v]_{\circlearrowleft}$ is selected. If at least one inactive atom remains, let p precede u and q follow v in the old inactive order. Removing u, v merges the three cells

$$[p, u]_{\circlearrowleft}, \qquad [u, v]_{\circlearrowleft}, \qquad [v, q]_{\circlearrowleft}$$

into $[p, q]_{\circlearrowleft}$. Their old interiors are balanced by induction, and u, v contribute one source and one target. The new interior is therefore balanced; all other cells are unchanged. This includes $p = q$, when three inactive atoms become one and the merged cell is a full-circle self-cell. If only u, v were inactive, no current cells remain after activation and the assertion is vacuous. □

For a current cell

$$\mathcal{C} = [u, v]_{\circlearrowleft},$$

the endpoints u and v are consecutive inactive atoms in the clockwise order. Hence, by definition of a cell, there is no inactive atom in the open arc $\mathcal{C}^\circ = (u, v)_{\circlearrowleft}$; every support point in its interior already belongs to the current active set A_k. If $\mathcal{C}$ is a candidate cell, then u and v come from opposite distributions. Moreover, Lemma E.1 implies that the active atoms in $\mathcal{C}^\circ$ contain the same number of sources and targets. It follows that both the interior $\mathcal{C}^\circ$ and the closed cell

$$\mathcal{C} = \mathcal{C}^\circ \cup \{u, v\}$$

are balanced: the former represents the local transport configuration before activating u and v, while the latter represents the local configuration after their activation.

To compare these two configurations, choose a cut outside $\mathcal{C}$ and unwrap the circle onto an interval of length L. Because the cut does not intersect $\mathcal{C}$, the clockwise arc $\mathcal{C}$ becomes a contiguous interval on the real line. Let

$$\mu_{\mathcal{C}^\circ} \quad \text{and} \quad \nu_{\mathcal{C}^\circ}$$

denote the source and target measures supported on the active interior atoms, and let

$$\mu_{\mathcal{C}} \quad \text{and} \quad \nu_{\mathcal{C}}$$

denote the corresponding measures after including the two endpoints. Both are balanced, so (C.1) applies to each: $c_{\mathbb{R}}(\mathcal{C}^\circ)$ is the cost of matching $\mu_{\mathcal{C}^\circ}$ against $\nu_{\mathcal{C}^\circ}$ in increasing unwrapped order, and $c_{\mathbb{R}}(\mathcal{C})$ the same for $\mu_{\mathcal{C}}$ against $\nu_{\mathcal{C}}$.

Definition E.2 (Circular cell marginal)**.** The *local marginal cost* associated with a candidate cell $\mathcal{C}$ is

$$\mathfrak{m}(\mathcal{C}) = c_{\mathbb{R}}(\mathcal{C}) - c_{\mathbb{R}}(\mathcal{C}^\circ). \tag{E.2}$$

Thus, $\mathfrak{m}(\mathcal{C})$ measures the increase in the local transportation cost caused by activating the two inactive endpoints of $\mathcal{C}$ and recomputing the optimal matching inside the cell.

The distance entering both terms in (E.2) is the ordinary absolute distance in the unwrapped interval, equivalently the clockwise arc length measured inside $\mathcal{C}$. It is not the circular geodesic distance $d_{\mathbb{S}^1}$. Indeed, once a cut is fixed outside the cell, $\mathcal{C}$ is treated as a contiguous line interval, and the local matching is constrained to remain within that interval. Using the geodesic distance at this stage could allow a pair of atoms in $\mathcal{C}$ to be connected through the complementary arc, thereby crossing the cut and destroying the desired locality.

The local marginal equals the increment in the *global line cost* when the cut avoids $\mathcal{C}$ and every cell selected earlier from the initially empty active set. Under these hypotheses, Lemma F.2 proves that the old sorted matching treats the interior and both exterior portions as separate balanced pieces. Activating u, v replaces only the interior matching, and hence

$$W_{1,\theta}^{\mathrm{line}}(A_k \cup \{u, v\}) - W_{1,\theta}^{\mathrm{line}}(A_k) = c_{\mathbb{R}}(\mathcal{C}) - c_{\mathbb{R}}(\mathcal{C}^\circ) = \mathfrak{m}(\mathcal{C}).$$

A cut outside the candidate alone need not satisfy this identity: it can cross a previously selected cell and alter the exterior matching. Figure 2 depicts the rematching under a compatible cut. The local formula itself remains valid independently of that global interpretation, and uses only the atoms of the candidate cell.

F FREE GAPS, THE EXACT GREEDY THEOREM AND THE SIMULTANEOUS CUT

Let the *original gaps* be the N arcs between consecutive points in the fully sorted union support. A gap is called *used* once it lies inside a selected candidate cell, and *free* otherwise. While inactive atoms remain, the current cells partition the original gaps: every gap lies inside exactly one cell, and which cell that is changes as cells merge.

Lemma F.1 (Free-gap invariant)**.** *At every stage with at least one inactive atom, every current cell contains at least one free original gap.*

Proof. The labels *free* and *used* attach to the N original gaps, which are fixed once and for all, and not to the cells, which are not: a gap acquires the label *used* when it first lies inside a selected cell and never loses it. Merging changes which cell a gap lies in; it never changes the gap's label. This is the point on which the induction turns, since a gap stays free while the cell containing it grows.

Initially, each current cell is exactly one original gap, hence free. Selecting $[u, v]_{\circlearrowleft}$ marks used exactly the gaps inside that cell and removes u, v from the inactive list. If inactive atoms remain, let p precede u and q follow v in the old inactive order. The three cells $[p, u]_{\circlearrowleft}$, $[u, v]_{\circlearrowleft}$, $[v, q]_{\circlearrowleft}$ merge into $[p, q]_{\circlearrowleft}$. Each exterior cell contained a free gap by induction. These gaps lie outside the selected cell, so they remain free in the merged cell; the algorithm need only store one (Figure 3). Every unaffected cell retains its free gap. This includes $p = q$, the full-circle self-cell left when only one inactive atom remains. If only u, v were inactive, there is no merged cell; the complementary unselected cell still supplies a free gap that can be retained for the final cut. □

Lemma F.2 (Cut compatibility)**.** *Start from $A_0 = \varnothing$ and let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be any valid sequence of candidate-cell selections, each activating its two opposite-type inactive endpoints. Let θ be an original gap contained in none of these cells. No greedy choice or intermediate optimality is assumed. After cutting at θ, the optimal line cost of the resulting active set A_k is*

$$W^{\mathrm{line}}_{1,\theta}(A_k) = \sum_{r=1}^{k} \mathfrak{m}(\mathcal{C}_r). \tag{F.1}$$

If a current candidate cell $\mathcal{C}$ also avoids θ, then activating its endpoints gives the sorted line matching of cost

$$W^{\mathrm{line}}_{1,\theta}(A_k) + \mathfrak{m}(\mathcal{C}). \tag{F.2}$$

Proof. No selected cell crosses the cut, so each becomes an ordinary interval between consecutive inactive atoms in the linear order. We prove inductively that every maximal active block in that order is balanced. Initially there are no active atoms. At a selection $[u, v]$, the interior is empty or a balanced active block. Activation merges it with any active blocks immediately outside u and v, together with one new source and one new target. The resulting block is balanced, and other blocks are unchanged. This argument also covers a final activation leaving no inactive atom.

Before each selection, the active portions strictly before u, inside (u, v), and strictly after v are unions of whole balanced blocks. Sorted matching treats these portions independently. The new sorted matching uses the same exterior pairs and the sorted matching on $[u, v]$, so its cost increases by exactly $c_{\mathbb{R}}([u, v]) - c_{\mathbb{R}}((u, v)) = \mathfrak{m}(\mathcal{C})$. Summing from the zero cost of A_0 gives (F.1). The same induction applies to one additional candidate avoiding the cut. No global optimality is used. □

Lemma F.3 (Optimality of free cuts before full activation)**.** *Suppose an active set A_k is globally circularly optimal and is generated from the empty set by valid candidate-cell selections; the history enters only through the labels* free *and* used*, which it defines. If at least one inactive atom remains, every free original gap θ realizes its circular cost as a line cost:*

$$W^{\mathrm{line}}_{1,\theta}(A_k) = C^{\circ}_k.$$

Proof. Choose an optimal matching supported on A_k and a shortest geodesic arc for each pair. No chosen arc can contain an inactive atom z in its open interior. Otherwise, replacing the endpoint of the same type as z by z gives a feasible matching of the same cardinality. The remaining subarc is strictly shorter than the original arc and bounds the new geodesic distance from above, so the cost decreases strictly. This contradicts optimality; the argument also covers antipodal pairs.

Take an original gap ϕ incident to an inactive atom z. A chosen matching arc cannot cross this gap: no support endpoint lies inside it, so crossing it would require the arc to pass through z. Cutting in ϕ therefore preserves every chosen geodesic edge length. The optimal line cost is at most C_k°, and it is at least C_k° because a line matching has no greater geodesic cost on the circle. Thus $W_{1,\phi}^{\mathrm{line}}(A_k) = C_k^\circ$.

The gap ϕ is also free. A previous selected cell containing it would have contained z in its interior or used z as an endpoint; in either case z would now be active. Applying Lemma F.2 at ϕ gives

$$\sum_{r=1}^{k} \mathfrak{m}(\mathcal{C}_r) = W_{1,\phi}^{\mathrm{line}}(A_k) = C_k^\circ.$$

Compatibility at any other free gap gives this same sum, proving the claim. At $k = 0$, the matching and sum are empty and every gap has zero cost, so the argument applies unchanged. □

Theorem F.1 (Exact greedy choice)**.** *Let $0 \le k < K$. Assume that the current active set A_k is circularly optimal and was generated from the empty set by valid candidate-cell selections. Then*

$$C_{k+1}^\circ - C_k^\circ = \min_{\mathcal{C} \in \mathfrak{C}_k} \mathfrak{m}(\mathcal{C}), \tag{F.3}$$

where $\mathfrak{C}_k$ is the set of current candidate cells whose endpoints have opposite types. Moreover, activating the endpoints of any minimizing cell yields a circularly optimal active set A_{k+1}.

Proof. Write

$$\delta_{k+1} := C_{k+1}^\circ - C_k^\circ.$$

We first show that every candidate marginal is at least δ_{k+1}. Fix an arbitrary current candidate cell $\mathcal{C} \in \mathfrak{C}_k$. Choose a current cell different from $\mathcal{C}$ and, by Lemma F.1, select a free support gap θ contained in that cell. If only two inactive atoms remain, there are two complementary directed cells between them, and we choose the cell different from $\mathcal{C}$. Thus, in either case, θ lies outside $\mathcal{C}$ and belongs to none of the cells selected during the preceding iterations.

Because $k < K$, inactive atoms remain. Lemma F.3 therefore gives

$$W_{1,\theta}^{\mathrm{line}}(A_k) = C_k^\circ.$$

The cut avoids all previous selections and $\mathcal{C}$, so Lemma F.2 applies to the additional activation. It produces a feasible cardinality-$(k+1)$ line matching of cost

$$C_k^\circ + \mathfrak{m}(\mathcal{C}).$$

The same matching is feasible on the circle, and its circular geodesic cost is no larger than its cost in the fixed unwrapped representation. Consequently,

$$C_{k+1}^\circ \le C_k^\circ + \mathfrak{m}(\mathcal{C}),$$

or equivalently,

$$\delta_{k+1} \le \mathfrak{m}(\mathcal{C}). \tag{F.4}$$

Since $\mathcal{C}$ was arbitrary, this inequality holds for every candidate cell in $\mathfrak{C}_k$.

We next establish the reverse inequality by exhibiting a candidate cell whose marginal equals δ_{k+1}. Let M_k be an optimal cardinality-k matching supported on the given active set A_k. Applying Theorem D.1 to M_k yields a circularly optimal cardinality-$(k+1)$ matching $M_{k+1}^\star$ with active set

$$A_{k+1}^\star = A_k \cup \{x^\star, y^\star\},$$

where $x^\star$ is a previously inactive source atom and $y^\star$ is a previously inactive target atom. By Theorem D.2, these two newly activated atoms are consecutive in the cyclic ordering of A_k^c. They therefore delimit a current opposite-endpoint candidate cell, which we denote by $\mathcal{C}^\star$.

By Lemma D.1, there exists an original support gap at which the circle can be cut so that the circular costs of both A_k and $A_{k+1}^\star$ are realized as line costs. As established in the proof of Theorem D.2, both active sets are optimal for the corresponding cardinality-constrained line problems under this common cut.

When more than two inactive atoms remain, the common cut cannot lie in the interior of $\mathcal{C}^\star$. Indeed, such a cut would place $x^\star$ and $y^\star$ at the two opposite ends of the linear inactive ordering, whereas the line-neighbor property requires them to be consecutive. If only two inactive atoms remain, the two complementary directed cells have the same endpoints, and we take $\mathcal{C}^\star$ to be the one whose interior does not contain the common cut.

Thus, in all cases, the common cut lies outside $\mathcal{C}^\star$. Both A_k and $A^\star_{k+1}$ are line-optimal at this cut, so Lemma C.1 applies to the nested pair: the linear interval between $x^\star$ and $y^\star$ is $\mathcal{C}^\star$ itself, and the difference between the two optimal line costs is exactly the local marginal of this cell:

$$\mathfrak{m}(\mathcal{C}^\star) = W_1^{\mathrm{line}}(A^\star_{k+1}) - W_1^{\mathrm{line}}(A_k) = C^\circ_{k+1} - C^\circ_k = \delta_{k+1}.$$

It follows that

$$\min_{\mathcal{C}\in\mathfrak{C}_k} \mathfrak{m}(\mathcal{C}) \le \mathfrak{m}(\mathcal{C}^\star) = \delta_{k+1}.$$

Combining this inequality with (F.4) proves (F.3).

Finally, suppose that $\mathcal{C}$ attains the minimum in (F.3). The construction in the first part of the proof gives a feasible circular cardinality-$(k+1)$ matching whose cost is at most

$$C^\circ_k + \mathfrak{m}(\mathcal{C}) = C^\circ_k + \delta_{k+1} = C^\circ_{k+1}.$$

Since C°_{k+1} is the minimum cost among all feasible cardinality-$(k+1)$ matchings, this matching must attain C°_{k+1}. Therefore, activating the endpoints of any minimizing candidate cell produces a circularly optimal active set A_{k+1}. □

Corollary F.1 (Optimal nested sequence)**.** *Starting from $A_0 = \varnothing$ and repeatedly selecting a minimum-marginal candidate cell produces an optimal active set for every cardinality $k = 0, \dots, K$.*

F.1 A CONSTRUCTIVE SIMULTANEOUS CUT

Theorem F.2 (Simultaneous optimal cut)**.** *There exists an original support gap $\theta^\star$ such that, for every $k = 0, \dots, K$,*

$$C^{\mathrm{line}}_{k,\theta^\star} = C^\circ_k. \tag{F.5}$$

Moreover, the greedy algorithm constructs such a gap.

Proof. For every current cell, maintain one representative free gap. Initially this is the cell's unique original gap. When a selected cell and its two neighbors merge, let the new cell inherit a free-gap representative from either unselected neighboring cell, as guaranteed by Lemma F.1. Figure F.1 traces the resulting gap through a complete run: one original gap is avoided by every selection, so the same cut serves all cardinalities.

If $n \neq m$, the algorithm terminates with inactive atoms from only the larger measure. Choose a representative free gap from any remaining current cell. If $n = m$, the final step starts with two inactive atoms and two complementary cells. Select one cell and retain the free-gap representative of the other. In both cases, the resulting gap $\theta^\star$ belongs to no selected cell.

By Lemma F.2, cutting at $\theta^\star$ gives

$$W^{\mathrm{line}}_{1,\theta^\star}(A_k) = \sum_{r=1}^{k} \mathfrak{m}(\mathcal{C}_r) = C^\circ_k$$

for every greedy active set A_k, where the last equality follows by summing the optimal increments from Theorem F.1. This includes the balanced terminal state $k = K$: no inactive atom is needed here, and we do not invoke Lemma F.3 at that state. Therefore

$$C^{\mathrm{line}}_{k,\theta^\star} \le C^\circ_k.$$

Conversely, any line matching is feasible on the circle with no larger geodesic cost, so

$$C^\circ_k \le C^{\mathrm{line}}_{k,\theta^\star}.$$

The two inequalities imply (F.5). □

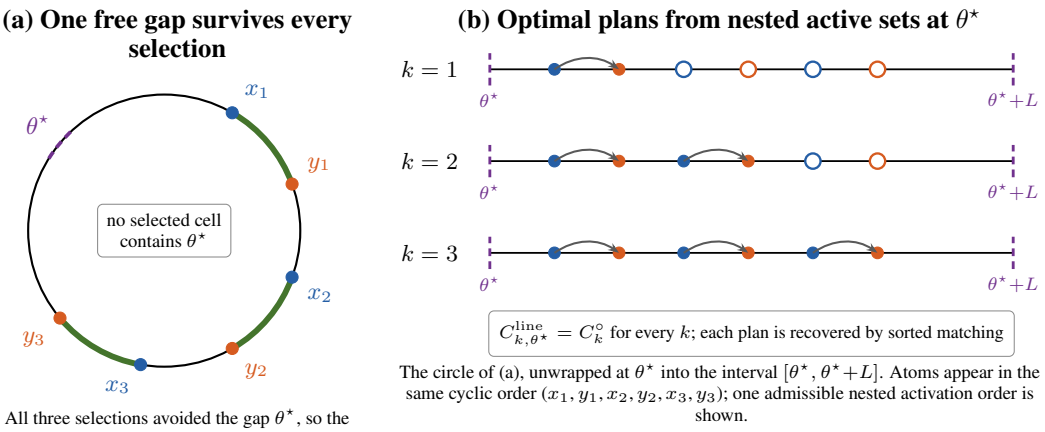


Figure F.1: Constructive simultaneous cut. The free gap $\theta^\star$ is contained in no selected cell, so the complete nested sequence can be cut and unwrapped at the same location. On the resulting line, the active sets grow by local sorted-matching updates and $C^{\text{line}}_{k,\theta^\star} = C^\circ_k$ for every transported cardinality k.

Remark F.1 (Existence is not uniqueness)**.** The theorem asserts existence of at least one simultaneous cut. It does not claim that every cut optimal for the full-mass problem is optimal for all smaller masses. The distinction is not academic, and it is what rules out the obvious shortcut: cut once at a gap optimal for $k = K$, run line PAWL there, and read off the whole profile. That procedure is $O(N \log N)$ and uses none of the structure above. Example F.1 shows that its error is unbounded.

Example F.1 (A cut optimal at $k = K$ and unbounded at $k = 1$)**.** Take $n = m = 2$ on a circle of circumference L, with $0 < \delta < L/8$:

$$x = \{0,\ L/4\}, \qquad y = \{\delta,\ 3L/4\}. \tag{F.6}$$

Write $G_1 = (0, \delta)$ and $G_3 = (L/4, 3L/4)$ for the two gaps whose cumulative imbalance sits one level above their neighbours. Together they carry $\delta + L/2 > L/2$ of the circumference, so that level is the weighted median in (C.3) and, by Proposition C.1,

$$1, 3 \in \Theta_K := \left\{r :\ C^{\text{line}}_{K,r} = C^\circ_K\right\}, \qquad K = 2. \tag{F.7}$$

Both are therefore admissible choices for a rule that cuts where the full-mass problem is solved, and G_1 is the first of them in index order.

At $k = 1$ the two disagree. The closest opposite-type pair is $(0, \delta)$, so $C^\circ_1 = w\delta$, and cutting in G_3 recovers it. Cutting in G_1 separates that pair: their unwrapped distance becomes $L - \delta$, and the sorted line matching falls back on the runner-up $d_2 = L/4 - \delta$. Hence

$$\frac{C^{\text{line}}_{1,1} - C^\circ_1}{C^\circ_1} = \frac{d_2 - \delta}{\delta} = \frac{L/4 - 2\delta}{\delta} \xrightarrow[\delta \to 0]{} \infty. \tag{F.8}$$

The mechanism is that C°_1 is a *geodesic* nearest-pair distance whereas $C^{\text{line}}_{1,r}$ is an *unwrapped* one, and nothing couples the location of the closest pair to the median that fixes Θ_K. So no bound of the form $C^{\text{line}}_{k,r} \le c\, C^\circ_k$ holds uniformly over $r \in \Theta_K$, for any constant c. Full-mass optimality alone therefore does not certify simultaneous optimality. Theorem F.2 constructs a single cut with the stronger guarantee.

G Constant-time cell-marginal queries

We now adapt the minimal-chain preprocessing of PAWL (Chapel & Tavenard, 2025) to cyclic intervals.

G.1 Doubled cyclic sequence

Let

$$z_1 < z_2 < \cdots < z_N < z_1 + L$$

be the sorted union support, and define labels

$$\sigma_i = \begin{cases} +1, & z_i \in \operatorname{supp}\mu, \\ -1, & z_i \in \operatorname{supp}\nu. \end{cases} \tag{G.1}$$

Form the doubled sequence

$$\widetilde{z}_i = z_i, \quad \widetilde{z}_{i+N} = z_i + L, \quad \widetilde{\sigma}_i = \widetilde{\sigma}_{i+N} = \sigma_i, \qquad i = 1, \dots, N. \tag{G.2}$$

Every candidate cell can be represented by a standard interval $[a, b] \subset \{1, \dots, 2N\}$ with $b - a < N$.

Define prefix differential ranks and signed coordinate sums

$$R_0 = 0, \qquad R_t = \sum_{i=1}^{t} \widetilde{\sigma}_i, \tag{G.3}$$

$$S_0 = 0, \qquad S_t = \sum_{i=1}^{t} \widetilde{\sigma}_i \widetilde{z}_i. \tag{G.4}$$

An interval $[a, b]$ is balanced if and only if

$$R_b = R_{a-1}. \tag{G.5}$$

For each t, let

$$p_t = \max\{0 \le s < t : R_s = R_t\}, \tag{G.6}$$

when this set is nonempty. Then $[p_t + 1, t]$ is the minimal balanced interval ending at t.

Lemma G.1 (Minimal-chain cost)**.** *If p_t exists, then*

$$c_{\mathbb{R}}([p_t + 1, t]) = w\,|S_t - S_{p_t}|\,. \tag{G.7}$$

Proof. By the maximality of p_t, the relative rank $R_s - R_{p_t}$ does not return to zero for $p_t < s < t$. It therefore has a constant strict sign throughout the interval. Equivalently, every pair in the increasing-order line matching has the same orientation. The sum of absolute pair displacements is consequently the absolute difference between the sum of source coordinates and the sum of target coordinates, which is $|S_t - S_{p_t}|$. Multiplying by w gives (G.7). This is the minimal-chain identity used by PAWL for the Manhattan cost (Chapel & Tavenard, 2025). □

Figure G.1 illustrates the two ingredients: a balanced interval decomposing into adjacent minimal chains, and the differential-rank walk whose first returns identify them. Define the maximal-chain prefix cost Q_t recursively by

$$Q_0 = 0, \qquad Q_t = \begin{cases} Q_{p_t} + w|S_t - S_{p_t}|, & p_t \text{ exists}, \\ 0, & p_t \text{ does not exist.} \end{cases} \tag{G.8}$$

Proposition G.1 (Balanced interval query)**.** *For every balanced interval $[a, b]$ in the doubled sequence,*

$$c_{\mathbb{R}}([a, b]) = Q_b - Q_{a-1}. \tag{G.9}$$

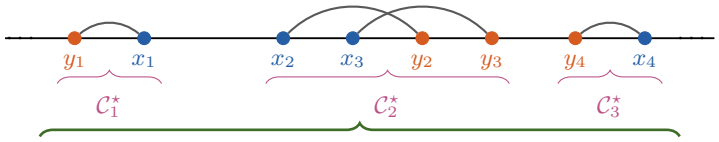
Proof. Repeatedly applying p_t decomposes every balanced interval into a unique sequence of adjacent minimal balanced intervals. The increasing-order line matching never crosses a boundary between adjacent balanced blocks, so their costs add. Because $R_b = R_{a-1}$, the maximal-chain decomposition ending at $a-1$ is exactly the prefix of the maximal-chain decomposition ending at b. Subtracting their accumulated costs gives (G.9). This is the same dynamic-programming principle as the maximal-chain cost identity in PAWL (Chapel & Tavenard, 2025). □

Corollary G.1 (Constant-time candidate marginal)**.** *Let a candidate cell be represented by the balanced doubled interval $[a, b]$. Then*

$$\mathfrak{m}([a, b]) = \big(Q_b - Q_{a-1}\big) - \big(Q_{b-1} - Q_a\big), \tag{G.10}$$

with the second term interpreted as zero when the interior is empty. Hence a candidate marginal is evaluated in $O(1)$ time after preprocessing.

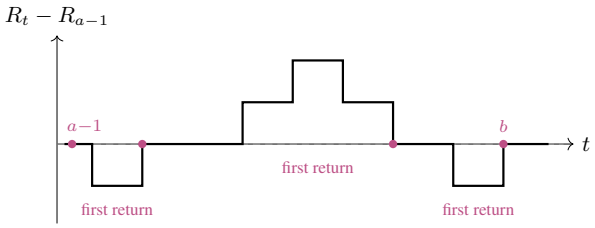
(a) A balanced interval in the doubled sequence decomposes into minimal balanced chains



$$[a,b] = \mathcal{C}_1^\star \cup \mathcal{C}_2^\star \cup \mathcal{C}_3^\star$$

$$\widetilde{z}_{i+N} = \widetilde{z}_i + L, \qquad c_{\mathbb{R}}([a,b]) = \sum_r c_{\mathbb{R}}(\mathcal{C}_r^\star) = Q_b - Q_{a-1}.$$

(b) Differential-rank walk



A balanced interval satisfies $R_b = R_{a-1}$. For each endpoint t, the most recent previous occurrence

$$p_t = \max\{0 \le s < t : R_s = R_t\}$$

identifies the minimal chain ending at t. Its cost is

$$c_{\mathbb{R}}([p_t + 1, t]) = w|S_t - S_{p_t}|.$$

When $[a,b]$ represents a candidate cell, as above (opposite-type endpoints, balanced interior), its marginal

$$\mathfrak{m}([a,b]) = (Q_b - Q_{a-1}) - (Q_{b-1} - Q_a)$$

is obtained from four table lookups.

Figure G.1: Constant-time cell-cost preprocessing on the doubled cyclic sequence. A balanced interval decomposes uniquely into adjacent minimal balanced chains, analogous to the chain decomposition in PAWL (Chapel & Tavenard, 2025). Repeated differential-rank values identify the minimal chains; their costs are signed-coordinate prefix differences. The accumulated values Q_t then provide both the cell cost and its interior cost with constant-time table lookups.

Proof. The outer interval $[a,b]$ and the interior $[a+1, b-1]$ are balanced by Lemma E.1. Apply Proposition G.1 to both and use (E.2). □

All values R_t, S_t, p_t, Q_t for $t = 0, \dots, 2N$ are computed in $O(N)$ time after sorting. The predecessor p_t is found by storing the most recent index of each integer differential-rank value in a hash table or array indexed over the rank range $[-2N, 2N]$.

H Correctness and Complexity

The active/inactive cyclic order is maintained by a circular doubly linked list. Each directed current cell stores its endpoints and a representative free original gap. Candidate cells are stored in a min-heap keyed by (G.10). Heap deletion is lazy: an entry is valid only if both endpoints remain inactive and the second endpoint is still the clockwise successor of the first.

Theorem H.1 (Correctness and complexity)**.** *Under Assumption B.1, Algorithm 1 returns the exact values C_k° and activation ranks encoding an optimal active set for every $k = 0, \dots, K$, together with a gap $\theta^\star$ satisfying* (F.5)*. Its time and memory complexities are*

$$\boxed{\textit{time} = O(N \log N), \qquad \textit{memory} = O(N).} \tag{H.1}$$

Proof. Correctness follows inductively from Theorem F.1; the simultaneous cut is guaranteed by Theorem F.2. Sorting costs $O(N \log N)$. The doubled-sequence preprocessing and linked-list initialization cost $O(N)$. There are at most N initial heap entries and at most one new candidate insertion per iteration. Thus only $O(N)$ heap entries are ever created. Every entry is popped at most

once, valid or stale, and each heap operation costs $O(\log N)$. The inner pop loop terminates because the heap always holds a valid entry: every current candidate cell was inserted when it was created, and by Theorem F.1 the cell of an optimal extension is a current candidate. Candidate-key evaluation, linked-list updates, and free-gap inheritance are $O(1)$. The heap, linked list, preprocessing arrays, and activation records all use $O(N)$ memory. □

I Plan recovery and arbitrary transported masses

I.1 Integer cardinalities

The activation rank of each atom determines the optimal active set A_k. Cut the circle at the simultaneous gap $\theta^\star$, retain the atoms activated by iteration k, and match the retained source and target atoms in increasing unwrapped order. By Theorem F.2, this line matching has value C_k° and is therefore circularly optimal.

Filtering activation ranks in the already sorted circular order recovers a plan represented by its k matched pairs in $O(N)$ time. Alternatively, retaining the activation-order lists gives the selected atoms in $O(k)$ time; sorting those atoms yields an $O(k \log k)$ recovery method. Activation ranks and the cut encode all plans in $O(N)$ space; explicitly listing the plans for every k requires $\Theta(K^2)$ pairs. A dense $n \times m$ matrix additionally requires $O(nm)$ output space.

I.2 Noninteger transported mass

At $s = Kw$, use the terminal plan π_K and cost C_K° directly. For $s \in [0, Kw)$, write

$$k = \left\lfloor \frac{s}{w} \right\rfloor, \qquad \lambda = \frac{s - kw}{w} \in [0, 1]. \tag{I.1}$$

Let π_k and π_{k+1} be consecutive optimal transport plans associated with the nested sequence. Then

$$\pi_s = (1 - \lambda)\pi_k + \lambda \pi_{k+1} \tag{I.2}$$

is feasible for transported mass s.

Proposition I.1 (Interpolation)**.** *The plan* (I.2) *is optimal, and*

$$\mathrm{PW}_\circ(s) = (1 - \lambda)C_k^\circ + \lambda C_{k+1}^\circ. \tag{I.3}$$

Proof. After dividing masses by w, the partial matching problem is a min-cost flow problem with integral capacities. Its optimal value as a function of the required flow is convex and piecewise linear, with breakpoints at integer flow values; successive shortest augmentations give nondecreasing marginal costs (Ahuja et al., 1993). Therefore the value on $[k, k + 1]$ is the linear interpolation of C_k° and C_{k+1}°. The feasible plan (I.2) attains exactly that value. Equivalently, after cutting at $\theta^\star$, this is the interpolation principle used by PAWL on the line (Chapel & Tavenard, 2025). □

J Spherical slicing and partial transport on $\mathbb{S}^{d-1}$

The construction so far is exact but one-dimensional. This section uses it as a primitive: we recall the spherical slicing of Bonet et al. (2023) and Liu et al. (2025), and then define a sliced discrepancy that inherits PAWC's partial structure, so that the whole transported-mass profile on $\mathbb{S}^{d-1}$ is available at once, and the two measures need not carry the same mass.

J.1 Great circles and the geodesic projection

Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ with the geodesic distance $d_{\mathbb{S}^{d-1}}(x, y) = \arccos\langle x, y\rangle$. The geodesics of $\mathbb{S}^{d-1}$ are its great circles, obtained by intersecting the sphere with a two-dimensional linear subspace. Slicing on $\mathbb{S}^{d-1}$ therefore replaces the lines of Euclidean sliced transport by great circles, and the one-dimensional problem it leaves behind is a transport problem on a circle rather than on a line – which is precisely the problem Algorithm 1 solves.

Following Bonet et al. (2023), subspaces are parameterised by the Stiefel manifold

$$V_{d,2} = \{U \in \mathbb{R}^{d\times 2} : U^\top U = I_2\}, \tag{J.1}$$

which covers the Grassmannian $G_{d,2}$ of two-planes through $U \mapsto UU^\top$ (Bendokat et al., 2020). Each $U \in V_{d,2}$ gives the great circle $\mathcal{C}_U = \mathrm{span}(U) \cap \mathbb{S}^{d-1}$, and the natural map onto it is the geodesic projection

$$\Pi_U(x) = \operatorname*{arg\,min}_{y\in\mathcal{C}_U} d_{\mathbb{S}^{d-1}}(x, y) = \frac{UU^\top x}{\|U^\top x\|_2} \ \in\ \mathcal{C}_U \subset \mathbb{R}^d, \tag{J.2}$$

defined when $U^\top x \neq 0$. For each fixed atom x, the exceptional frames have uniform measure zero. A finite union over the input atoms is still null, so all empirical pushforwards below are defined for almost every frame.

The columns of U form an orthonormal basis of $\mathrm{span}(U)$, so $z \mapsto Uz$ is an isometry from $\mathbb{S}^1$ onto $\mathcal{C}_U$ and $\Pi_U = U\,P^U$, where

$$P^U(x) = U^\top \Pi_U(x) = \frac{U^\top x}{\|U^\top x\|_2} \ \in\ \mathbb{S}^1 \tag{J.3}$$

is the coordinate representation of (J.2) in the basis U (Bonet et al., 2023, Lemma 1). The two are kept apart deliberately: Π_U lands on the great circle in $\mathbb{R}^d$, whereas P^U is its coordinate vector in $\mathbb{R}^2$, and it is P^U that the one-dimensional solver consumes. Concretely, Algorithm 1 is run on the circular coordinate

$$\theta_U(x) = \mathrm{atan2}\big((U^\top x)_2,\ (U^\top x)_1\big) \bmod 2\pi \ \in\ [0, 2\pi), \tag{J.4}$$

that is, on $\mathbb{S}^1_L$ with circumference $L = 2\pi$; because $z \mapsto Uz$ is an isometry, arc length along $\mathcal{C}_U$ agrees with arc length in the coordinate θ_U, so no rescaling is needed. Figure 4(a) illustrates the projection.

Two features of this construction matter here. First, the image of P^U is a circle of circumference 2π, so the slice inherits the cyclic structure with no distinguished boundary – the difficulty that motivated everything above, now arising for a second reason. Second, P^U is a pushforward of measures, so it preserves total mass and maps atoms to atoms: an empirical measure with n atoms of common weight w becomes an empirical measure on $\mathbb{S}^1$ with n atoms of the same weight w. The standing assumptions of Assumption B.1 are therefore met on almost every slice when the combined spherical support is pairwise distinct, as assumed below: two distinct, non-antipodal points project to the same circle coordinate only on a null set of frames, while antipodal points project to distinct antipodal coordinates. Here and below, ω denotes the uniform probability measure on $V_{d,2}$.

J.2 Sliced spherical partial Wasserstein

Let ω be the uniform probability measure on $V_{d,2}$. For probability measures $\mu, \nu \in \mathcal{P}(\mathbb{S}^{d-1})$, Bonet et al. (2023) define the spherical sliced-Wasserstein discrepancy by averaging the circular Wasserstein distance over slices,

$$\mathrm{SSW}_p^p(\mu, \nu) = \int_{V_{d,2}} W_{p,\circ}^p\big(P^U_\#\mu,\ P^U_\#\nu\big)\, d\omega(U). \tag{J.5}$$

This presumes μ and ν are probability measures, hence of equal mass. We drop that requirement.

Let μ and ν be finite positive measures on $\mathbb{S}^{d-1}$ with uniformly weighted atoms,

$$\mu = w\sum_{i=1}^n \delta_{x_i}, \qquad \nu = w\sum_{j=1}^m \delta_{y_j}, \qquad x_i, y_j \in \mathbb{S}^{d-1},\ w > 0, \tag{J.6}$$

Assume the combined spherical support is pairwise distinct. We allow $n \neq m$, so $\mu(\mathbb{S}^{d-1}) = nw$ and $\nu(\mathbb{S}^{d-1}) = mw$ may differ. Because each P^U preserves mass, the admissible transported masses $s \in [0, Kw]$ with $K = \min\{n, m\}$ are the same on every slice; a single s is thus meaningful across all of $V_{d,2}$ simultaneously.

Definition J.1 (Sliced spherical partial Wasserstein)**.** For $s \in [0, Kw]$ and μ, ν as in (J.6), define

$$\mathrm{SSPW}(s; \mu, \nu) = \int_{V_{d,2}} \mathrm{PW}_\circ\big(s;\, P^U_\# \mu,\, P^U_\# \nu\big)\, d\omega(U), \tag{J.7}$$

where $\mathrm{PW}_\circ(s; \cdot, \cdot)$ is the partial transport cost (B.3) on the circle, i.e. the quantity Algorithm 1 returns for every admissible s in a single run.

At integer cardinalities the integrand is C°_k of (B.5) evaluated on the projected measures; between them it is the interpolation of Proposition I.1. Four remarks fix the scope of the definition.

Remark J.1 (Scope)**.** Two standing restrictions are worth stating plainly. First, $d \geq 2$, and $d = 2$ is degenerate: there $\mathrm{span}(U) = \mathbb{R}^2$, so $\mathcal{C}_U$ is all of $\mathbb{S}^1$, P^U is an isometry, and SSPW collapses to Algorithm 1 itself – slicing does nothing. The construction is of interest for $d \geq 3$. Second, (J.6) gives μ and ν a *common* atom weight w, inherited from Assumption B.1. What varies is the number of atoms, hence the total mass; the atom weight does not. In particular this does not cover two empirical probability measures with different sample counts, whose atoms would weigh $1/n$ and $1/m$: that is a nonuniform-weight problem, requiring the event-driven treatment of saturating atom capacities noted in the discussion, and is outside the present statement.

Remark J.2 (What is inherited slice-wise)**.** Several properties pass from the circle to the sphere because they hold for the integrand at every U and are preserved by integration.

1. *Symmetry and nonnegativity*: $\mathrm{SSPW}(s; \mu, \nu) = \mathrm{SSPW}(s; \nu, \mu) \geq 0$, since $d_{\mathbb{S}^1}$ is symmetric.
2. *Shape in s*: $s \mapsto \mathrm{SSPW}(s; \mu, \nu)$ is nondecreasing, convex and piecewise linear with breakpoints contained in $\{kw\}_{k=0}^K$, because each slice profile is (Proposition I.1) and these properties survive averaging.
3. *Isometry invariance*: for $R \in O(d)$, $\mathrm{SSPW}(s; R_\# \mu, R_\# \nu) = \mathrm{SSPW}(s; \mu, \nu)$, since $P^U \circ R = P^{R^\top U}$ and ω is invariant under the action of $O(d)$ on $V_{d,2}$.
4. *Reduction*: if $n = m$ and $s = nw$, every slice problem is balanced. Since $W_{1,\circ}$ is positively homogeneous under a common rescaling of both masses,

$$\mathrm{SSPW}(nw; \mu, \nu) = nw\ \mathrm{SSW}_1\Big(\tfrac{\mu}{nw},\, \tfrac{\nu}{nw}\Big), \tag{J.8}$$

so (J.7) recovers SSW_1 of (J.5) exactly when $nw = 1$, and up to that mass factor otherwise.

Remark J.3 (What is not claimed)**.** SSPW is a discrepancy, not a metric, and we claim no more here. Two obstructions are already present one slice at a time. Partial transport at a *fixed* transported mass does not satisfy the triangle inequality even on the line, so no metric property can be inherited slice-wise; and $\mathrm{SSPW}(s; \mu, \nu) = 0$ for $s > 0$ only forces the projected measures to share s units of mass on ω-almost every slice, which is a statement about $P^U_\# \mu$ and $P^U_\# \nu$ rather than about μ and ν. The injectivity results that Bonet et al. (2023) obtain for the spherical Radon transform, and the metricity that Liu et al. (2025) establish for the linear embedding, both concern probability measures and equal-mass transport; whether an analogue holds once mass is allowed to differ is open, and settling it is the natural next question. Sample complexity, by contrast, is not open at the elementary level: the integrand of (J.7) is bounded, and Proposition J.1 below gives a uniform Monte-Carlo guarantee over the whole profile. What remains open there is sharper dimension- or distribution-dependent control, since that bound uses nothing about μ and ν beyond the range of the integrand.

Remark J.4 (Choice of s)**.** Because Algorithm 1 returns the entire profile, s need not be committed to in advance: one run per sampled slice yields $\widehat{\mathrm{SSPW}}(\cdot; \mu, \nu)$ on all of $[0, Kw]$. This is the practical payoff of the profile being free. When a single number is wanted, s can be selected afterwards, without further transport solves. This is a computational capability; the application experiments prescribe the mass and do not validate an automatic selection rule.

For comparisons across data sets of differing cardinality or total mass, the transported *fraction*

$$\rho = \frac{s}{w \min\{n, m\}} \in [0, 1] \tag{J.9}$$

is more interpretable than the absolute mass s, and puts profiles with different K on a common axis. When an elbow is wanted, it is usually easier to locate on the marginal profile $\Delta_k = C^\circ_k - C^\circ_{k-1}$ than

on the cumulative profile C_k°: the latter is convex and increasing, so the onset of expensive transport registers only as a change of curvature, whereas Δ_k is nondecreasing and registers the same event as a jump.

J.3 Estimation and cost

In practice (J.7) is estimated by Monte Carlo over slices; we write M for their number, reserving L for the circumference as elsewhere. Draw $U_1, \dots, U_M$ i.i.d. from ω, for instance by QR factorisation of a $d \times 2$ Gaussian matrix with the diagonal of R made positive, and set

$$\widehat{\mathrm{SSPW}}(s;\mu,\nu) = \frac{1}{M}\sum_{\ell=1}^{M} \mathrm{PW}_\circ\big(s;\, P^{U_\ell}_{\#}\mu,\, P^{U_\ell}_{\#}\nu\big). \tag{J.10}$$

The diagnostic implementation uses Gaussian QR without fixing these signs. It samples uniform two-dimensional subspaces; its frames need not themselves have the uniform Stiefel law. This gives the same distribution of slice costs: replacing U by UO, $O \in O(2)$, rotates or reflects both projected measures and leaves the circular partial-transport profile unchanged.

Proposition J.1 (Uniform concentration of the sliced estimate)**.** *Let μ, ν be as in* (J.6) *and let $U_1, \dots, U_M$ be i.i.d. ω, with $K \geq 1$. At $s = 0$ the estimation error is identically zero. For every fixed $s \in (0, Kw]$ and every $\varepsilon > 0$,*

$$\Pr\Big(\big|\widehat{\mathrm{SSPW}}(s;\mu,\nu) - \mathrm{SSPW}(s;\mu,\nu)\big| \geq \varepsilon\Big) \;\leq\; 2\exp\Big(-\frac{2M\varepsilon^2}{\pi^2 s^2}\Big). \tag{J.11}$$

Uniformly over the profile,

$$\Pr\Big(\sup_{s\in[0,Kw]} \big|\widehat{\mathrm{SSPW}}(s;\mu,\nu) - \mathrm{SSPW}(s;\mu,\nu)\big| \geq \varepsilon\Big) \;\leq\; 2(K+1)\exp\Big(-\frac{2M\varepsilon^2}{\pi^2 K^2 w^2}\Big), \tag{J.12}$$

so that $M \geq \big(\pi^2 K^2 w^2/2\varepsilon^2\big)\log\big(2(K+1)/\delta\big)$ slices suffice for ε-accuracy on the entire profile with probability at least $1-\delta$, for $0 < \delta < 1$.

Proof. Fix $s > 0$ and set $Z_\ell = \mathrm{PW}_\circ\big(s; P^{U_\ell}_{\#}\mu, P^{U_\ell}_{\#}\nu\big)$, so the Z_ℓ are i.i.d. with mean $\mathrm{SSPW}(s;\mu,\nu)$. A transport plan of mass s moves no mass further than the geodesic radius π of a circle of circumference 2π, whence $0 \leq Z_\ell \leq \pi s$; the upper end is attained when every transported pair in an optimal plan is antipodal. This occurs for a single antipodal source–target pair; distinct multi-atom supports concentrated near antipodal locations can approach the same ceiling arbitrarily closely. Thus no uniformly smaller range is available without further assumptions. Hoeffding's inequality for i.i.d. variables confined to an interval of length πs gives (J.11).

For (J.12), apply (J.11) at each breakpoint $s = kw$, $k = 1, \dots, K$, with the range relaxed to the uniform bound $\pi K w$; the $k = 0$ event has probability zero. A union bound, bounded loosely by $K+1$ terms, gives (J.12). By Proposition I.1 every slice profile is linear on $[kw, (k+1)w]$; hence so are $\widehat{\mathrm{SSPW}}$ and SSPW, being an average and an integral of such profiles over a common breakpoint set, and hence so is their difference. A linear function attains its maximum modulus on an interval at an endpoint, so control at the breakpoints controls the supremum. □

Each term of (J.10) costs one projection of the $N = n + m$ atoms, $O(dN)$, followed by a single call to Algorithm 1, $O(N \log N)$ by Theorem H.1. The estimator therefore costs

$$O\big(M\, N(d + \log N)\big) \tag{J.13}$$

in time, *for the entire profile in s*. Streaming the slice average uses $O(N)$ solver memory beyond the input, plus $O(d)$ for the current frame. Thus a streamed implementation uses $O(N + d)$ auxiliary memory. The supplied diagnostic implementation instead retains all frames and slice profiles, using $O(Md + MK + N)$ auxiliary memory. The slices are independent and parallelizable.

Remark J.5 (Relative accuracy over the full profile)**.** The range πs in (J.11) scales with s, so an *absolute* accuracy target treats small and large transported masses very differently, and (J.12) – which relaxes every breakpoint to the largest range $\pi K w$ – is correspondingly weak at small s

(Appendix K.6 measures by how much). Asking instead for accuracy *relative* to s removes the disparity: taking $\varepsilon = \eta s$ in (J.11) gives, for every $s = kw$ with $k \geq 1$,

$$\Pr\Big(\big|\widehat{\mathrm{SSPW}}(s;\mu,\nu) - \mathrm{SSPW}(s;\mu,\nu)\big| \geq \eta\, s\Big) \;\leq\; 2\exp\Big(-\frac{2M\eta^2}{\pi^2}\Big), \tag{J.14}$$

in which s has cancelled. The bound is the same at every breakpoint, so a union bound over the K of them costs a factor K and nothing else: $\Pr(\exists k \geq 1 :\ |\cdot| \geq \eta\, kw) \leq 2K\exp(-2M\eta^2/\pi^2)$. A relative guarantee over the whole profile therefore adds only a logarithmic dependence on K to the sample requirement. Indeed, linear interpolation of the breakpoint errors gives $|\widehat{\mathrm{SSPW}}(s) - \mathrm{SSPW}(s)| \leq \eta s$ between breakpoints as well. Here "relative" means relative to transported mass s, not to the possibly small transport cost. For an absolute error target, retaining the range πkw at each nonzero breakpoint gives the sharper union bound $2\sum_{k=1}^{K}\exp(-2M\varepsilon^2/(\pi^2k^2w^2))$.

Two points are worth isolating. First, the profile is free: obtaining $\widehat{\mathrm{SSPW}}(s;\cdot,\cdot)$ at every admissible s costs the same as obtaining it at one, because Algorithm 1 produces all cardinalities in a single sweep. A construction built instead on a per-s circular solver would pay a further factor of K, which is $\Theta(N)$ when the two sample counts are comparable. Second, replacing the circular solver by the $O(N^2\log N)$ cut-enumeration reference of Corollary C.1 would raise (J.13) to $O(M(dN + N^2\log N))$; the sliced construction is what makes the $O(N\log N)$ solver worth having, and the $O(N\log N)$ solver is what makes the sliced construction affordable at the sizes reported in Appendix K.4.

K Numerical validation of the solver

This section reports an implementation of Algorithm 1 and the programme used to verify it. The cut envelope (C.5) supplies a direct exact oracle – enumerate every original support gap, cut and unwrap there, run line PAWL once to obtain all $C^{\mathrm{line}}_{k,r}$, and compare the PAWC output to $\min_r C^{\mathrm{line}}_{k,r}$ for every k – and a cardinality-constrained assignment solver applied separately for each k supplies a second, slower oracle that does not use the circle-cut structure at all. The supplied code contains the core solvers, fixed-seed correctness and invariant tests, and walkthrough and scaling demos. The original timing records and the drivers and raw results for the one-cut and spherical-estimator studies below are not included in the current package. Those numerical summaries are retained as reported results; the available scaling demo is a separate run, not a reproduction of the reported timing table.

K.1 Solvers and the trust hierarchy

Four solvers are compared, ordered by how much structure they assume. A disagreement triggers checks of feasibility, numerical tolerance and the exact rational reference described below; no implementation is assumed infallible.

Baseline A (LP oracle). For each k separately, the cardinality-constrained problem (B.5) is solved as a linear program over the cardinality-k matching polytope $\mathcal{P}_k$, using the full circular cost matrix $d_{\mathbb{S}^1}(x_i, y_j)$. It uses no cut, no cell and no chain, so it is independent of every structural result proved above. Its extreme points are integral, and the returned vertex is checked to be a matching.

Baseline B (cut enumeration). The exact cut envelope (C.5): cut at each of the N original support gaps, unwrap, run line PAWL once per cut, and minimise over cuts for every k. This is the $O(N^2\log N)$ reference solver of Corollary C.1.

PAWC **(reference).** A transparent transcription of Algorithm 1 that mirrors the proof structure and carries the invariant assertions of Appendix K.3.

PAWC**.** The optimised implementation, in two variants – one in plain array code and one just-in-time compiled.

Instances are drawn from seven test families, illustrated in Figure K.1: *uniform* points; *clustered* supports with long free gaps; *near-antipodal* pairs perturbed away from equal-length geodesics (the standing assumptions also allow exact antipodal pairs); *tiny-arc* supports, where the circle reduces to the line; *unequal-size* supports with strongly unbalanced n and m; *perturbed grids*, which give many nearly equal candidate marginals; and *interleaved* supports with strict source/target alternation. The 504-instance battery contains 72 instances from each family.

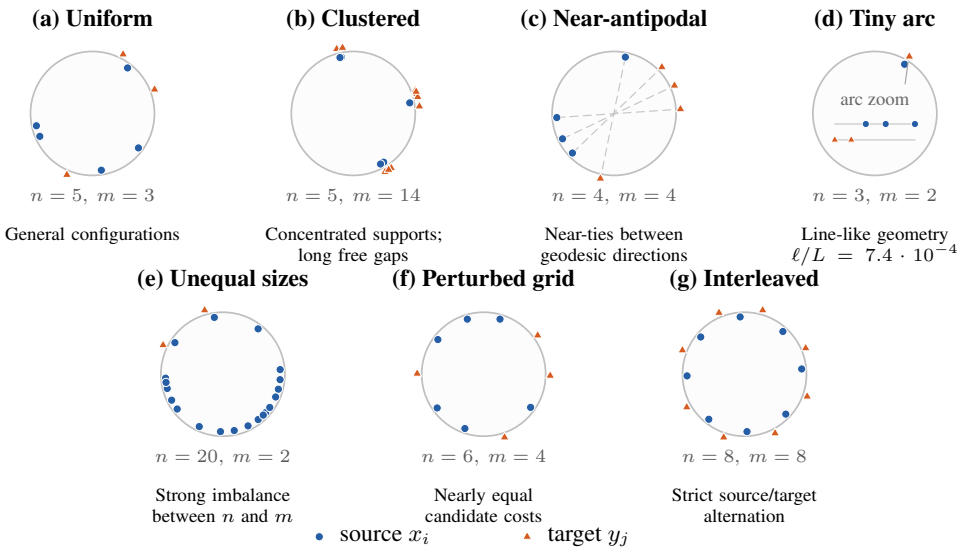


Figure K.1: The seven test-instance families used for numerical validation. Each panel shows one configuration from the fixed-seed battery and the feature it tests. Blue circles denote source atoms; orange triangles denote target atoms. Radial offsets separate the two measures visually; only angle carries data. Dashed chords in (c) connect the generated near-antipodal pairs, not a transport plan. Panel (d) includes a magnified, unwrapped view of its occupied arc of length ℓ. The generator and selected instance indices are supplied with the figure.

K.2 Exactness

On 504 random instances with $n, m \in [1, 40]$ spanning all seven families, the cut envelope and the LP oracle agree at every cardinality $k = 0, \ldots, K$ to 10^{-9}, confirming Proposition C.1 and Corollary C.1 numerically. Both PAWC variants reproduce these values on the same battery, and agree with the cut envelope on larger instances up to $N = 2000$.

Two further checks distinguish numerical agreement from exact arithmetic. First, a separate battery uses rational coordinates with denominator 2003 and seed 424242. The cut-envelope reference is recomputed with exact rational gap midpoints and an $O(nmK)$ dynamic program over non-crossing line matchings. The absolute-value cost satisfies the Monge condition, making this reference exact. PAWC and the LP still run in floating point and are compared with the rational values converted to floating point, using an absolute tolerance of 10^{-9}; the comparison is not an exact-arithmetic implementation of every solver. Second, tests require the two optimised PAWC variants to reproduce the reference implementation *bit for bit*: identical `float64` cost arrays and activation ranks under the same tie-breaking convention. This requirement applies to PAWC variants, not to the LP or PAWL baselines. Reassociating sums can break it, motivating the sequential evaluation of (G.8) (Remark K.1).

A separate check confirms the premise of Proposition C.1: cuts placed at uniformly random positions on the circle never beat the minimum over the N inter-point gaps, and moving the cut within a gap leaves the unwrapped profile unchanged.

K.3 Structural invariants

Cost agreement alone is a weak test, because optima need not be unique: a solver can return the right number from a structurally wrong state. The following are therefore checked directly by the debug implementation and dedicated invariant tests at the relevant induction steps. Floating-point cost comparisons use numerical tolerances; combinatorial properties are checked exactly.

1. *Nestedness* (Theorem D.1 and Corollary D.1): $A_k \subset A_{k+1}$, and the increment consists of exactly one source atom and one target atom.

2. *Cyclic-neighbor addition* (Theorem D.2): the two newly activated atoms are consecutive in the cyclic ordering of the inactive set A_k^c.
3. *Balanced cell interiors* (Lemma E.1): the active atoms strictly inside each current cell contain equally many sources and targets.
4. *Free-gap invariant* (Lemma F.1): the current cells are recomputed from scratch, and every one of them is confirmed to contain a gap belonging to no previously selected cell.
5. *Optimality of free cuts* (Lemma F.3): while inactive atoms remain, cutting at any free gap of any current cell reproduces C_k° to numerical tolerance. At balanced termination, the retained cut is checked using Theorem F.2.
6. *Cost consistency*: the incrementally maintained cost is compared with $W_{1,\circ}(A_k)$ recomputed from the circulation form (C.3) – a weighted median over gap lengths, which involves no cut, cell or chain and is therefore independent of how the algorithm arrived at its value.
7. *Constant-time marginal* (Corollary G.1): the four-lookup formula (G.10) is checked against Definition E.2 evaluated by explicit sorted matching inside the cell.

At the level of whole instances we further verify that $\theta^\star$ realises C_k° as a line cost simultaneously for every k (Theorem F.2) and lies in a gap contained in no selected cell; that $k \mapsto C_k^\circ$ is nondecreasing and convex; that recovered plans are feasible, with marginals at most w, total mass kw, and cost equal to the reported cost when re-evaluated with $d_{\mathbb{S}^1}$ directly; that Proposition I.1 holds at noninteger transported mass; and that the whole profile is invariant under rotation of the circle, under reflection, and under exchanging μ and ν. Rotation invariance is the sharpest of these in practice, since wrap-around handling is where circular implementations characteristically fail. Degenerate configurations are covered separately: $k = 0$; $k = 1$, where C_1° must equal w times the smallest geodesic distance over all pairs; $n = 1$ or $m = 1$; and all atoms confined to a short arc, where the profile must coincide with a single line PAWL call.

Remark K.1 (On evaluating (G.8))**.** The recursion $Q_t = Q_{p_t} + w|S_t - S_{p_t}|$ chains together exactly the indices sharing a differential rank, so it looks like a segmented cumulative sum and is tempting to vectorise as a global prefix sum minus its value at each segment start. That form is numerically inferior: it differences two partial sums whose magnitudes can be much larger than the local cost being recovered. Here $O(1)$ refers to the number of arithmetic operations in a query, not a uniform bound on the numerical value of a chain cost. Measured against the sequential recursion it loses about $1.6 \cdot 10^{-9}$ absolute at $N = 10^6$, which is enough to break exact agreement between implementations. Since the recursion is $O(N)$ either way, Theorem H.1 is indifferent to the choice, and the sequential form is preferable.

K.4 Complexity

Figure 5 and Table K.1 report the cost of computing the *complete* profile, i.e. all cardinalities $k = 0, \dots, K$, for N from 10^2 to 10^6, with five repetitions per size (two for Baseline A at $N \geq 600$) and just-in-time compilation excluded by an explicit warm-up. The reported environment is one core of an AMD Ryzen Threadripper PRO 9965WX, Python 3.11 and NumPy 2.4, as in the main paper. Baseline B is measured to $N = 3 \cdot 10^4$ and extrapolated beyond that along its theoretical $N^2 \log N$ slope, the constant being fitted at the largest measured size. Baseline A is measured to $N = 10^3$, beyond which one linear program per cardinality is no longer practical, and is extrapolated along N^3.

Fitting $\log(\text{time})$ against $\log N$ gives a slope of 2.00 over all sizes and 2.03 over the top decade for cut enumeration, matching its $O(N^2 \log N)$ cost. Baseline A fits at 3.05. A pure N^3 law fitted at the largest measured size reproduces every measured point to within 35% across a decade, motivating the reported cubic extrapolation. This is an empirical model of the measured timings, not a complexity guarantee for the implementation, which solves a separate LP at each cardinality. For PAWC the same fit gives 1.11 and 1.29 (array code), and 1.03 and 1.23 (compiled). The top-decade values exceed what one might naively expect, and the discrepancy deserves comment rather than rounding.

For comparison, $N \log N$ has log–log slope about 1.08 over the top decade. Restricted to $N \leq 3 \cdot 10^4$, the reported fits are 0.97 (array) and 0.91 (compiled). Memory access and fixed overhead can affect these finite-range timings, but the measurements do not identify a cache boundary or establish a particular cause. The reported line-PAWL fits on the same machine are 1.123 over all sizes and 1.251

N	PAWC (array)	PAWC (compiled)	Baseline B	Baseline A
10^2	0.000094 s	0.000042 s	0.0042 s	0.212 s
$3 \cdot 10^2$	0.000177 s	0.000052 s	0.0315 s	4.52 s
10^3	0.000555 s	0.000134 s	0.349 s	226 s
$3 \cdot 10^3$	0.00170 s	0.00048 s	3.35 s	—
10^4	0.00643 s	0.00178 s	38.5 s	$2.26 \cdot 10^5$ s‡
$3 \cdot 10^4$	0.0217 s	0.00574 s	357 s	$6.10 \cdot 10^6$ s‡
10^5	0.108 s	0.0217 s	$4.43 \cdot 10^3$ s†	—
$3 \cdot 10^5$	0.470 s	0.0777 s	$4.37 \cdot 10^4$ s†	—
10^6	2.13 s	**0.365** s	$5.32 \cdot 10^5$ s†	—

Table K.1: Time to compute the complete partial-transport profile for all k, mean of five repetitions (two for Baseline A at $N \geq 600$). †extrapolated along $N^2 \log N$, ‡along N^3, in both cases with the constant fitted at the largest measured size; a dash marks sizes past which no extrapolation is attempted. At $N = 10^6$ the compiled implementation returns every cardinality in 0.365 s; the extrapolation predicts roughly six days for cut enumeration. Baseline A already needs almost four minutes at $N = 10^3$, where PAWC needs $1.3 \cdot 10^{-4}$ s.

cut rule	instances failing	median	90th pct	worst
$\min \Theta_K$ (optimal at $k = K$)	13.5%	0.09	1.0	$1.8 \cdot 10^3$
widest support gap	22.6%	0.21	0.7	2.2
warm start near $K/2$	38.9%	0.28	28	$3.4 \cdot 10^4$

Table K.2: One-cut rules against the exact envelope, 5600 exact instances. A rule *fails* on an instance when its cut is strictly suboptimal at some k; the excess columns are $\max_k \varepsilon_k$ over the instances where it does, with $\varepsilon_k = (C^{\text{line}}_{k,r} - C^{\circ}_k)/C^{\circ}_k$ for $1 \leq k \leq K$; $k = 0$ is excluded because its cost is zero. The random-gap row is omitted because the original summary does not specify how expectation and failure were aggregated. Every cut is a valid upper bound (Corollary C.1).

over the top decade. These empirical slopes complement the operation-count argument; they do not prove complexity.

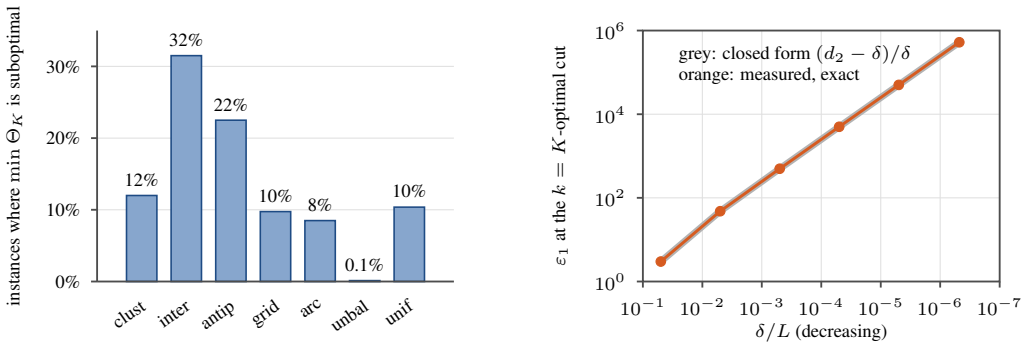
The machine-independent form of Theorem H.1 is cleaner to verify directly: the reported counts are about $0.81N$ heap pops and $0.31N$ subsequent pushes at $N = 10^4$, 10^5 and 10^6. The push count excludes the initial candidate list, which is built and heapified separately. That is, the total number of entries ever created is linear, each is removed at most once whether valid or stale, and at most one new candidate is inserted per iteration, exactly as Algorithm 1 prescribes. Profiling at $N = 10^5$ shows no stage growing faster than the others, and in particular no residual cost from candidate-list cleaning, which lazy deletion removes entirely.

K.5 Is the simultaneous cut necessary?

Example F.1 shows that a single cut chosen at $k = K$ can be arbitrarily wrong at $k = 1$. That is a worst case; this subsection measures the typical one. Everything below is a reduction of the table $T[r][k] = C^{\text{line}}_{k,r}$ that the cut envelope already computes, so no additional solver is involved, and all of it is computed in exact integer arithmetic: whether a cut is optimal is an equality test, and a tolerance would otherwise decide the answer. Write $\Theta_k = \{r : T[r][k] = C^{\circ}_k\}$ for the set of optimal gap indices at cardinality k.

We compare four rules for choosing one cut and reusing it for every k: the deterministic $\min \Theta_K$; the widest support gap; a uniformly random gap, whose expectation is evaluated exactly rather than sampled; and a warm start at an intermediate cardinality near $K/2$. The original records are needed to specify its integer-rounding convention. Table K.2 reports 5600 instances, 800 from each of the seven families of Appendix K.1; failure rates are stable to 1.1 percentage points against an independent replication of the same size. Figure K.2 breaks the deterministic rule down by family and places the adversarial family of Example F.1 against its closed form.

Three things are worth stating plainly, including the one that cuts against the result we set out to find.



(a) failure of the deterministic one-cut rule

(b) adversarial family: ε_1 against δ/L

Figure K.2: (a) Failure rate of the deterministic rule $\min \Theta_K$ by instance family, 800 instances each. (b) The family of Example F.1: measured relative excess at $k = 1$ of a cut that is exactly optimal at $k = K$, against the closed form (F.8), in exact arithmetic over six decades of δ.

The deterministic rule usually works. It is optimal at every k on 86.5% of instances. A reader who expects one-cut heuristics to fail routinely on random data is wrong, and we do not claim otherwise. What it lacks is a certificate: on the 13.5% where it fails there is nothing local distinguishing them, and the excess reaches three orders of magnitude. Its failure rate is also uneven – 0.1% on strongly unbalanced instances, 31.5% on strictly alternating ones – so it cannot be excused as a boundary effect. Two further caveats sharpen the comparison rather than soften it. First, obtaining Θ_K is only cheap when $n = m$, where the $k = K$ problem is balanced and its cut is a weighted median of (C.3); when $n \neq m$ the full-mass problem is itself partial, so the rule presupposes the machinery it claims to avoid. Restricted to $n = m$, where it is genuinely cheap, its failure rate rises to 27.3%. Second, the choice within Θ_K matters: $\min \Theta_K$ is one of several optimal cuts at $k = K$, and a different element of that same set fails on 46.5% of instances.

A simultaneously optimal cut is common, not rare. The set $\Theta_\cap = \bigcap_{k=0}^{K} \Theta_k$ of gap indices optimal at every k has median size between $0.17N$ and $0.92N$ depending on the family. Theorem F.2 therefore does not locate a needle in a haystack on random data, and we should not claim it does. Its content is that such a gap exists *always* and is produced constructively, not that it is hard to stumble on. Correspondingly, a uniformly random element of Θ_K is simultaneously optimal with median probability 1, but with probability as low as 0.14 on the worst instance encountered – which is the same point as before in another form: the typical case is benign and carries no guarantee.

The smallest optimal gap index can change with k. Tracking $\min \Theta_k$, the reported index changes on a majority of instances in every family except the strongly unbalanced one. This concerns one deterministic representative of a nonunique optimum; it does not mean that changing the cut is necessary, since $\Theta_\cap$ is nonempty by Theorem F.2. The intermediate-cardinality warm start is reported to fail on 38.9% of instances.

K.6 The spherical construction

Appendix J is a construction; this subsection is the evidence for it. Two things are measured: that the implementation reproduces an independent published one at the single point where the two theories must coincide, and how the Monte-Carlo estimator (J.10) actually behaves against Proposition J.1.

Reduction to the balanced case. Remark J.2(4) states that at $n = m$ and $s = nw$ every slice problem is balanced and $\mathrm{SSPW}(nw; \mu, \nu) = nw\,\mathrm{SSW}_1(\mu/nw, \nu/nw)$. This is the only quantity in Appendix J with an external reference, and reproducing it exercises the whole pipeline: Stiefel sampling, geodesic projection, the circular coordinate (J.4), Algorithm 1, and the slice average. We compare against the implementation shipped with POT (Flamary et al., 2021), which follows Bonet et al. (2023).

	what it isolates	vs. POT	vs. LP oracle
(a)	circular solver alone, no projection	$3.3 \cdot 10^{-15}$	$1.4 \cdot 10^{-15}$
(b)	per slice on the sphere, 384 slices	$7.5 \cdot 10^{-9}$	$5.6 \cdot 10^{-15}$
(c)	the aggregate (J.7) against SSW_1	$1.3 \cdot 10^{-9}$	–
(d)	the mass factor (J.8), four (n, w)	$1.5 \cdot 10^{-15}$	–
(e)	the $d = 2$ degeneracy of Remark J.1	$7.9 \cdot 10^{-16}$	–

Table K.3: Maximum relative discrepancy of the reduction cross-check, over $d \in \{3, 5, 10, 50\}$, $n = m \in \{10, 50, 200\}$ and both instance families, with slices shared between implementations. Check (e) additionally reports slice-to-slice variation of $6.7 \cdot 10^{-16}$, consistent with round-off: at $d = 2$ the projection is an isometry and every slice has the same mathematical value.

The comparison uses *the same slices* on both sides. Letting each implementation draw its own would compare two Monte-Carlo estimates, and agreement would then only be required in distribution – a weak test that a systematic bias of the size of the Monte-Carlo noise would pass. With shared slices both sides evaluate the same finite average of the same integrand, so agreement is required to solver tolerance. Table K.3 reports five checks, each isolating what the next would aggregate away.

Checks (b) and (c) report discrepancies of $7.5 \cdot 10^{-9}$ and $1.3 \cdot 10^{-9}$ against POT, respectively. POT provides a level-median solver for circular W_1, so these observations do not establish a binary-search tolerance floor. Their cause remains unresolved: diagnosing it requires rerunning the reduction checks with the recorded inputs, normalisations and backend settings. Table K.3 retains the observed discrepancies without attributing them to a particular implementation or numerical mechanism.

Concentration. Proposition J.1 is next. We take a reference at $M_{\mathrm{ref}} = 10^5$ slices as a numerical reference, drawn from a slice stream *disjoint* from the estimates under test – if the estimate's slices were a subset of the reference's, deviations would be biased downward and the convergence would flatter itself – and use 200 repetitions at each $M \in \{2^3, \ldots, 2^{12}\}$, over $d \in \{3, 10, 50\}$, three families and both $n = m$ and $n \neq m$.

The reported convergence slope is -0.497 across the eighteen configurations, with individual fits in $[-0.519, -0.476]$, consistent with $M^{-1/2}$ scaling. The observed exceedance frequencies lie below the right-hand side of (J.11) in all 540 configuration–threshold comparisons. These are empirical checks: the finite reference has sampling error, and finite repetitions do not verify a probability bound. The deviation on a fine s-grid also never exceeds the maximum breakpoint deviation, as predicted by the piecewise-linearity argument in Proposition J.1.

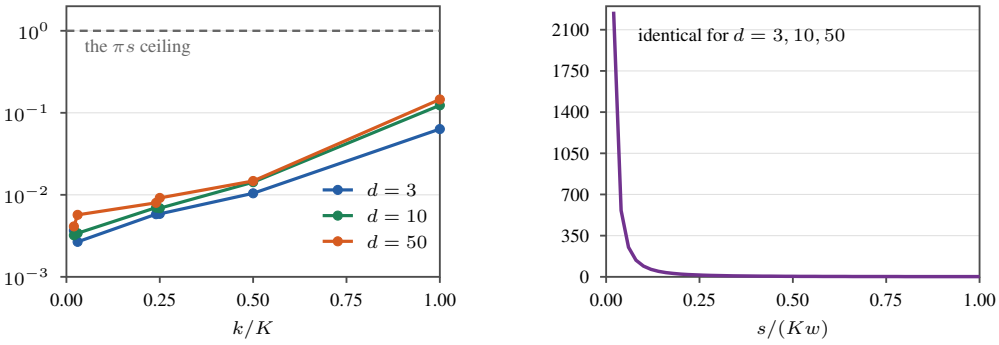
What the measurements add is a quantitative account of where Proposition J.1 is loose, and there are two distinct answers.

First, the range. The bound uses $0 \leq Z_\ell \leq \pi s$, attained only when every transported pair is antipodal. The observed range $\max_\ell Z_\ell - \min_\ell Z_\ell$ is between 0.3% and 15% of πs (Figure K.3(a)), and the ceiling itself is never approached: $\max_\ell Z_\ell$ stays below $0.29\,\pi s$ throughout. The observed range is a sample statistic, not a bound on the support of the random variable, so substituting it directly into Hoeffding's inequality would not give the same guarantee. Its median is 0.0067, 0.0091 and 0.0100 of πs at $d = 3, 10, 50$, respectively: it increases modestly with dimension in these configurations but remains far below the worst-case range. These results do not establish a general dimension-dependent refinement.

Second, the uniform bound (J.12) relaxes the range at every breakpoint from $\pi k w$ to $\pi K w$ and takes a union bound over $K + 1$ breakpoints. The ratio of its full right-hand side to the pointwise right-hand side at $s > 0$ is

$$(K+1) \exp\left[\frac{2M\varepsilon^2}{\pi^2}\left(\frac{1}{s^2} - \frac{1}{K^2 w^2}\right)\right].$$

For the settings of Figure K.3(b), this is about 10^{361} at $s = 0.05\,Kw$ and equals 51 at $s = Kw$. The ratio is independent of d. The practical reading is that (J.12) should be used for what it is – a statement about the whole profile – and (J.11) used when accuracy is wanted at a particular transported mass. Remark J.5 records the natural way to have both at once.



(a) observed range / πs

(b) $\log_{10}$(union bound / pointwise bound)

Figure K.3: (a) The observed slice-cost range $\max_\ell Z_\ell - \min_\ell Z_\ell$ as a fraction of the ceiling πs used in Proposition J.1, against transported cardinality; median over families, $M_{\text{ref}} = 10^5$. The ceiling is never approached, and the gap widens as s shrinks. (b) The ratio of the full union-bound right-hand side in (J.12) to the pointwise right-hand side in (J.11), before truncating either at 1. Here $K = 50$, $\varepsilon = 0.05\,Kw$ and $M = 2^{12}$; the ratio includes the union-bound factor $K + 1 = 51$, equals 51 at $s = Kw$, and diverges as $s \to 0$.

Finally, the slice count. Proposition J.1 demands $M \geq (\pi^2 K^2 w^2 / 2\varepsilon^2) \log(2(K+1)/\delta)$; the smallest tested M at which no repetition out of 200 exceeded ε is smaller than that by a reported median factor of 470. Zero exceedances in 200 repetitions do not certify a tail probability below δ, especially after selecting M from the tested grid. This comparison describes the tested configurations; the proposition supplies the distribution-free guarantee.

Remark K.2 (On the two PAWC variants)**.** The array implementation and the compiled one produce bit-identical output; the latter is about 2–6 times faster over the sizes in Table K.1. Both are reported because Baseline B calls line PAWL, which is itself compiled: comparing against the array implementation understates PAWC's advantage by a constant factor, while the compiled variant is the like-for-like comparison. The asymptotic separation visible in Figure 5 does not depend on the choice.

L Application Detail

This appendix records the protocols and reported results behind Sections 6.2 and 6.3. The supplied solver package does not contain the original application drivers, query manifests, training checkpoints or per-run results. The remaining protocol details identified below require those records; no new application experiment is reported here.

L.1 Shape Matching: Protocol

Contours are traced at the 0.5 level of each MPEG-7 mask, resampled to 128 points uniformly in arclength, and converted to outward normal angles by a centred difference over five samples. Pixel quantisation can produce identical normals, violating the distinct-support assumption. The reported implementation adds jitter of size 10^{-7} in its circular coordinate, seeded by a checksum of the file name. The coordinate convention must be recovered before converting this value to degrees. A finite jitter changes the numerical instance; it is not literally symbolic perturbation of the original one.

A query keeps a contiguous fraction v of the arclength starting at a uniformly random sample, replaces a fraction c of the survivors by uniformly drawn normals, and applies a uniformly random rotation. Every method scores the same 360 rotations. Write $\rho = k/K$, where K is the smaller support size. Orientation and retrieval in Table 1 use $\rho = 1 - c$, assuming known clutter; Figure 6 instead fixes $\rho = 0.8$. Maximum-cardinality transport ("full OT") uses $k = K$ and is balanced only when the support sizes agree. The integer-rounding convention for v, c and ρK needs the original code.

Orientation succeeds when the recovered rotation error is less than 5°. Orientation and correspondence use 212 queries per cell; retrieval uses 138 queries against all 1400 database shapes. The query's own database entry is excluded, leaving 1399 candidates and 19 other shapes from its class. The bullseye score is based on same-class hits among the top 40; whether the reported scores divide by 19 or the conventional class size 20 remains to be confirmed from the evaluation code. The query-selection manifests are also needed to reproduce these subsets.

For correspondence, alignment minimises the partial cost at $\rho = 0.8$. A recovered pair is correct when it matches the arclength index from which the query sample was built; clutter samples have no correct partner. The original records must establish whether the plan used for F_1 is extracted at this same mass. The reported retrieval improvement 0.231 ± 0.053 is a 95% interval, but its estimator and resampling unit are not documented in the available files.

The correlation baseline cross-correlates von Mises kernel densities of the normal-angle measures by FFT, using bandwidths 0.01, 0.02 and 0.04 in the reported circular coordinate and selecting the best bandwidth per cell. The correspondence baseline uses mutual nearest neighbours in normal angle after alignment. These comparisons hold the descriptor and rotation grid fixed.

L.2 Unbalanced transport and solver agreement

A separate twelve-cell comparison reports 864 queries. The earlier description "one shape per class" is incompatible with $70 \times 12 = 840$; the reported query count is retained pending its manifest and should not be identified with the 212-query cells of Table 1. A KL-relaxed unbalanced Sinkhorn cost (Peyré & Cuturi, 2019) replaces the partial cost, with marginal penalties $\tau \in \{0.01, 0.05, 0.2\}$ and entropic regularisation 10^{-2}. The best τ is selected separately in each cell. At $v = 0.35, c = 0.4$, reported orientation success is 0.71, versus 0.85 for PAWC using the known-clutter rule $\rho = 1 - c$. The reported rotation sweep costs 0.58 s per query per τ, versus 0.029 s for 360 PAWC solves, about 20 times longer per penalty. Neither hyperparameter protocol constitutes selection on unseen data.

On sampled rotations, PAWC's profile value is compared with POT's `partial_wasserstein2` at the same cardinality and unit atom weights. The reported largest absolute discrepancy over the 864 queries is 4.3×10^{-14}, demonstrating numerical agreement on the tested problems. For the separate timing comparison in Section 6.1, the rounded values 4.4 ms versus 0.09 ms at $n = m = 256$, and 1.54 s versus 0.56 ms at $n = m = 2048$, give speedups of about 49 and 2750, respectively, for PAWC's complete profile relative to POT at one mass. The original timing records are needed to recover the exact repetition and timing-boundary conventions.

L.3 Spherical fitting: gradients and protocol

For one slice U, write $p = U^\top x$ and $\theta_U(x) = \operatorname{atan2}(p_2, p_1)$. Away from $p = 0$,

$$\nabla_x \theta_U(x) = \frac{U(-p_2, p_1)^\top}{\|p\|^2}.$$

Away from coincident and antipodal angle pairs, a matched geodesic distance has angular derivative ± 1. The slice loss also includes the atom weight w and the slice average a factor $1/M$. Where the optimal matching is locally unchanged, differentiating its cost gives the value-function gradient. At matching changes or distance cusps, ordinary differentiability is not assumed. Gradients are projected onto the tangent space of each particle, and updates are retracted by renormalisation. The reported central-difference check has relative error 3×10^{-6}; its step sizes and test inputs are not included in the current package.

Training uses $n = 500$ particles from a uniform start, $M = 32$ slices redrawn per step, and 2000 Adam steps with a cosine learning-rate schedule from 8×10^{-2} to 4×10^{-3}. With target batch size $m = \gamma n$, $\gamma \geq 1$, it transports $k = K = n$: every model particle participates and a fraction $1 - 1/\gamma$ of target atoms is left unmatched. Thus $\rho = k/K = 1$ throughout. At $\gamma = 1$, the measures have equal mass and the loss reduces to spherical sliced Wasserstein, with the factor nw in (J.8); the training loss's weight/normalisation convention must be specified in the restored code.

A fraction ε of each target batch is replaced by uniform spherical points. Table 2 uses $\gamma = 1/(1 - \varepsilon)$ rounded to the tested grid, assuming known contamination; "best" is the largest gain on that grid,

without held-out selection. Figure 7 illustrates a different setting, $\gamma = 1.25$ at $\varepsilon = 0.3$. Increasing γ changes both target batch size and trimming, so these results do not isolate their effects.

The synthetic target has five von Mises–Fisher components with masses $0.28, 0.22, 0.18, 0.20, 0.12$ and concentrations $80, 200, 60, 120, 150$; their directions are needed to reproduce it. The real target contains $94\,953$ USGS epicentres of magnitude at least 4.5 from 2010–2024. It is split $80/20$ once; the model never sees the evaluation fifth. The catalogue snapshot, filtering rules and split seed are needed to reproduce that split. The reported coordinate tie perturbation has magnitude in $[10^{-9}, 2 \times 10^{-9})$ with random sign; the coordinate and renormalisation convention must be recovered.

Quality is evaluated by geodesic energy distance to 4000 clean target points, using eight seeds per table cell. The exact energy-distance normalisation, seed pairing, six tested γ values, batch-size rounding and per-seed records are needed for replication and uncertainty estimates. The "wins" counts and means in Table 2 do not determine the variance of the paired improvements.