

Homotopy Measures for Representative Trajectories

Erin Chambers^{*1}, Irina Kostitsyna^{†2}, Maarten Löffler^{‡3}, and Frank Staals^{§4}

- 1 Dept. of Math and Computer Science, Saint Louis University, Saint Louis, USA
echambe5@slu.edu
- 2 Dept. of Mathematics and Computer Science, TU Eindhoven, Eindhoven, The Netherlands
i.kostitsyna@tue.nl
- 3 Dept. of Computing and Information Sciences, Utrecht University, The Netherlands
m.loffler@uu.nl
- 4 MADALGO, Aarhus University, Aarhus, Denmark
f.staals@cs.au.dk

Abstract

An important task in trajectory analysis is defining a meaningful representative for a cluster of similar trajectories. Formally defining and computing such a representative r is a challenging problem. We propose and discuss two new definitions, both of which use only the geometry of the input trajectories. The definitions are based on the homotopy area as a measure of similarity between two curves, which is a minimum area swept by all possible deformations of one curve into the other. In the first definition we wish to minimize the maximum homotopy area between r and any input trajectory, whereas in the second definition we wish to minimize the sum of the homotopy areas between r and the input trajectories. For both definitions computing an optimal representative is NP-hard. However, for the case of minimizing the sum of the homotopy areas, an optimal representative can be found efficiently in a natural class of restricted inputs, namely, when the arrangement of trajectories forms a directed acyclic graph.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases trajectory analysis, representative trajectory, homotopy area

Digital Object Identifier 10.4230/LIPIcs.ESA.2016.27

1 Introduction

Motivated by GIS applications, the question of extracting a meaningful representative trajectory from a collection of similar trajectories has recently received considerable attention in the computational geometry community [3, 11, 12, 16, 19, 1, 10, 2]. In many trajectory analysis applications, only the locations (and not the corresponding time stamps) in the

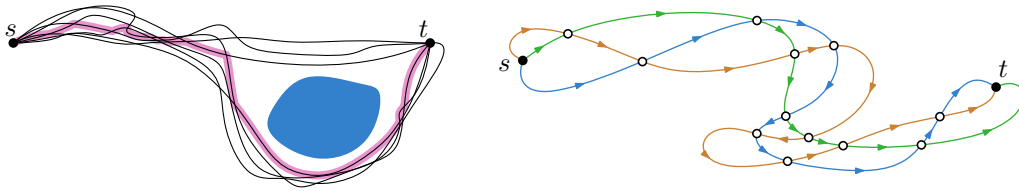
* E.C. is supported by the National Science Foundation under grant CCF-1054779 and IIS-1319573.

† I.K. is (partially) supported by the Netherlands Organisation for Scientific Research (NWO) under grant 639.023.208 and 639.021.123, respectively.

‡ M.L. is (partially) supported by the Netherlands Organisation for Scientific Research (NWO) under grant 639.023.208 and 639.021.123, respectively.

§ F.S. is supported by the Danish National Research Foundation under grant nr. DNRF84.





■ **Figure 1** Left: Curves representing hiking trajectories between two points, and a possible representative. A pointwise average trajectory would go through the lake. Right: The trajectory graph G . In this example, G is acyclic.

trajectories are relevant. Consider, for example, the case in the input trajectories originate from hikers that walked a similar trail, but possibly on different days. In such a setting the trajectories are just curves in the plane, and hence we wish to find a representative curve that captures important features shared by most of the input curves. It has been argued before that it is desirable that the representative uses only pieces of the input trajectories, so that it avoids obstacles in the underlying space [4]. See for example Fig. 1 (left). Hence, we will restrict our attention to representatives that consist of pieces of the input trajectories, and that ignore any temporal information available.

Buchin et al. investigate whether a reasonable notion of a median exists in such a setting that depends only on the intersections in a set of trajectories [4]. Their *simple median* is essentially not using the geometry. They also present a second definition, that incorporates a notion of the topology of the underlying space, by placing obstacles in large open regions and restricting the class of trajectories to the same *homotopy type*, that is, they require the representative trajectory to “wind around” the obstacles in the same way as the majority of the input trajectories. For example, in Fig. 1 (left), a user could for example mark the lake as an obstacle. Buchin et al. conclude that while computation of the median is possible to some extent, some notion of geometry and topology seems necessary to handle practical situations.

In this paper, we include some geometric and topological information in the selection of a representative trajectory (curve), namely, the area of the faces in the arrangement of trajectories. As a measure of similarity between two curves, we use the *homotopy area* from Chambers and Wang [7], which is the minimum area swept by a deformation of one non-self-intersecting curve into the other. More formally, if $H : [0, 1] \times [0, 1] \rightarrow \mathbb{R}^2$ is a continuous deformation of curve μ into τ , the homotopy area of H is

$$A(H) = \int_{s \in [0,1]} \int_{t \in [0,1]} \left| \frac{dH}{ds} \times \frac{dH}{dt} \right| ds dt.$$

The infimum $\text{HA}(\mu, \tau)$ of A over all continuous deformations between μ and τ is the homotopy area. The notion of homotopy area seems particularly attractive in our setting as it implicitly penalizes a representative trajectory for deviating from the bulk of the trajectories without making it necessary to artificially place obstacles in the ambient space, which was the solution used in prior work [4]. Homotopy area is defined only on curves which are non-self-intersecting, so we must also place this constraint on each of our input trajectories.

Problem Statement. We are given a set of *trajectories* $\mathcal{T} = \{\tau_1, \dots, \tau_n\}$, which are piecewise curves, each piece of low algebraic degree, in the plane. We wish to compute a single trajectory μ^* that best represents all trajectories in \mathcal{T} . As we will use homotopy area to measure the quality of μ^* we require that each individual trajectory τ_i is *simple*, that is, it has

no self-intersections (otherwise homotopy area is not well defined). Consider the arrangement of the trajectories \mathcal{T} in \mathbb{R}^2 , and orient each edge so that its direction corresponds to that of the trajectory defining it. We refer to this arrangement as the *trajectory graph* G . See Fig. 1.

Initially, we will assume that all trajectories start and end at the same points, say s and t respectively, and that s and t lie in the outer face of the arrangement of the trajectories. We will (partially) lift these restrictions in Sections 4.2 and 4.3.

For the output trajectory μ^* we require that it is a simple path in the trajectory graph; this means that it can consist only of segments of the input trajectories, that it is simple, and that it uses each segment in the same direction as used in the input trajectory.

Among all possible output trajectories (simple paths in G), we wish to construct one that represents \mathcal{T} best. We measure this by the distance between the (candidate) median μ and the trajectories in \mathcal{T} . Let $\text{HA}(\mu, \tau_i)$ be the minimum homotopy area between μ and a trajectory $\tau_i \in \mathcal{T}$. We consider two variants: minimizing the maximum distance $\text{HA}_{\max}(\mu, \mathcal{T}) = \max_{\tau_i \in \mathcal{T}} \text{HA}(\mu, \tau_i)$ between μ and the trajectories in \mathcal{T} , and the sum of the distances $\text{HA}_{\text{sum}}(\mu, \mathcal{T}) = \sum_{\tau_i \in \mathcal{T}} \text{HA}(\mu, \tau_i)$ between μ and the trajectories in \mathcal{T} . If \mathcal{T} is clear from the context we will write $\text{HA}_{\max}(\mu) = \text{HA}_{\max}(\mu, \mathcal{T})$ and $\text{HA}_{\text{sum}}(\mu) = \text{HA}_{\text{sum}}(\mu, \mathcal{T})$.

Results. We show that the first variant considered, minimizing the maximum distance, is NP-hard, even if we have only two trajectories, both of which are x -monotone (Section 2). In general, minimizing the sum of the distances, HA_{sum} , is also NP-hard, as we show in Section 3. However, the second hardness reduction is more involved and critically relies on cycles in the trajectory graph. If the trajectory graph is a directed acyclic graph (DAG), then we can compute a representative minimizing HA_{sum} efficiently, as we show in Section 4. Quite surprisingly, our results show that when the graph is a DAG and all trajectories share a start and end point on the outer face, the simple median from Buchin et al. [4] that does not incorporate areas in any way, remains the optimal choice for minimizing HA_{sum} . Hence, even though the best running time to compute homotopy area between two curves is $O(n + I^2 \log n)$ time, where n is the complexity of the input curves and I is the number of intersections between the two curves [7], we are able to calculate a median trajectory under homotopy area much more quickly using the simple median algorithm [4]. We also show that our approach generalizes to the case when the start and end points of the trajectories are in different, arbitrary faces of the DAG, although the simple median is no longer the curve minimizing HA_{sum} . Instead, a simple median must be computed between lifts of the trajectories in a particular covering space of the plane. Omitted proofs are in the full version.

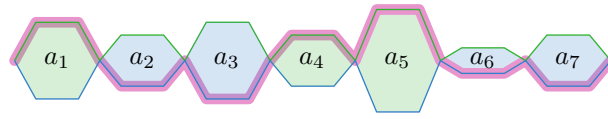
2 Minimizing the Maximum Distance HA_{\max} is NP-hard

In this section we consider computing a representative that minimizes the maximum distance to all other trajectories. Unfortunately, this problem is NP-hard, even for the case of a constant number of x -monotone input curves.

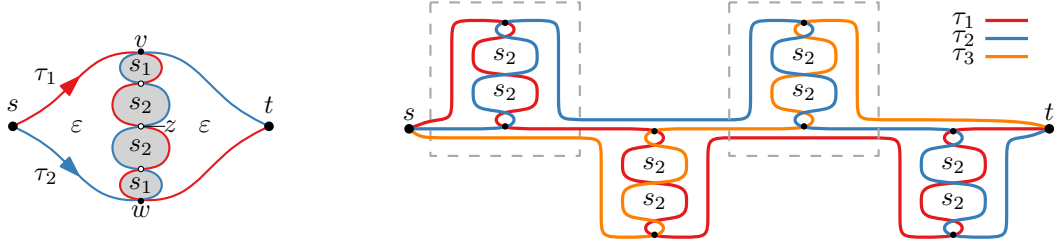
► **Theorem 1.** *Given a set of trajectories \mathcal{T} , computing a median μ that minimizes HA_{\max} is NP-hard, even if \mathcal{T} contains only two trajectories, both of which are x -monotone.*

Proof. We reduce from the Partition problem, which, given a set $A = \{a_1, \dots, a_n\}$ of positive integers, asks for a partition of A into sets A_1 and A_2 such that $\sum_{a_i \in A_1} a_i = \sum_{a_i \in A_2} a_i = \sum_{a_i \in A} a_i / 2$. Given the set A , we construct two x -monotone trajectories τ_1 and τ_2 such that the faces between successive intersections have area equal to some $a_i \in A$. See Fig. 2.

Any candidate trajectory μ corresponds to a partition of A into A_1 and A_2 : $a_i \in A_1$ if and only if μ uses the piece of τ_1 that bounds the face corresponding to a_i . It



■ **Figure 2** An illustration of the NP-hardness reduction from Partition. The purple curve represents the partition $B = \{a_2, a_3, a_6, a_7\}$ and $G = \{a_1, a_4, a_5\}$.



■ **Figure 3** Left: The *braid* construction—a basic building block for hardness proof gadgets. Optimal representative trajectory does not switch at intersections. $HA_{\text{sum}} = 4s_1 + 2\varepsilon$. Right: Four building blocks joint together. Optimal representative trajectory follows the red or the blue trajectory all the way from s to t and does not switch at intersections. $HA_{\text{sum}} = 5s_2 + O(s_1)$.

follows that the homotopy area between μ and τ_j is exactly $\sum_{a_i \in A_j} a_i$. Thus $HA_{\text{max}}(\mu) = \max \{ \sum_{a_i \in A_1} a_i, \sum_{a_i \in A_2} a_i \}$. Let μ^* be a trajectory minimizing HA_{max} . We have that $HA_{\text{max}}(\mu^*) = \sum_{a_i \in A} a_i / 2$ if and only if A can be partitioned such that $\sum_{a_i \in A_1} a_i = \sum_{a_i \in A_2} a_i$. Therefore minimizing HA_{max} is (weakly) NP-hard. ◀

3 Minimizing the Sum of Distances HA_{sum} is NP-hard

In this section we show that minimizing the total sum of the distances from the representative trajectory to all the trajectories in \mathcal{T} is NP-hard in general.

Before we describe the gadgets for variables and clauses, consider the two trajectories τ_1 and τ_2 in Fig. 3 (left). Let $\varepsilon \ll s_1 \ll s_2$ be the areas swept by the deformation of τ_1 into τ_2 between the intersection points, and let μ^* be a representative trajectory that minimizes HA_{sum} . We will call this construction a *braid* of τ_1 and τ_2 . We will show that in a braid $\mu^* = \tau_1$ or $\mu^* = \tau_2$, i.e., μ^* does not switch to another trajectory at any intersection point.

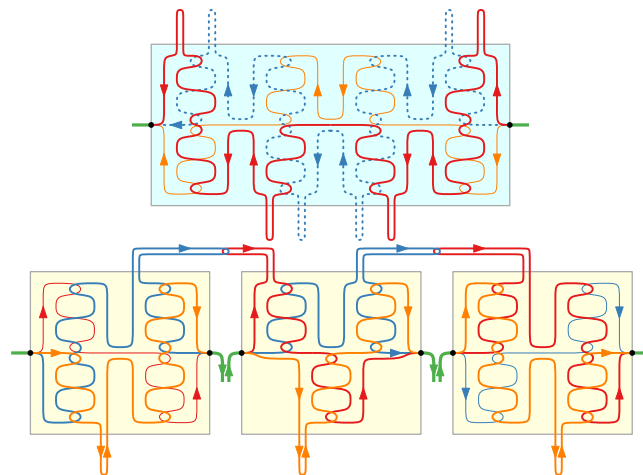
► **Lemma 2.** *If the areas of the faces of the arrangement of $\mathcal{T} = \{\tau_1, \tau_2\}$, for a braid construction of two trajectories τ_1 and τ_2 (depicted in Fig. 3 (left)), satisfy $\varepsilon \ll s_1 \ll s_2$, then the optimal representative trajectory $\mu^* = \tau_1$ or $\mu^* = \tau_2$.*

The braid construction is a crossing gadget, it allows two trajectories to cross while enforcing that μ^* does not switch to another trajectory at intersections. We will use it as a basic building element in the hardness proof gadgets.

Now consider an arrangement of three trajectories in Fig. 3 (right). There are four braids of pairs of trajectories used in this arrangement. The red trajectory, τ_1 , and the blue trajectory, τ_2 , are rotationally symmetrical. Let μ^* be a representative trajectory that minimizes HA_{sum} .

► **Lemma 3.** *For the arrangement of three trajectories $\mathcal{T} = \{\tau_1, \tau_2, \tau_3\}$ depicted in Fig. 3 (right) the optimal representative trajectory $\mu^* = \tau_1$ or $\mu^* = \tau_2$.*

► **Theorem 4.** *Minimizing HA_{sum} is NP-hard.*



■ **Figure 4** Left: An example of a variable gadget consisting of two building blocks. Variable gadgets are traversed by trajectories from right to left. Right: The clause gadget is traversed from left to right.

Proof. We prove that it is NP-hard to minimize HA_{sum} by a reduction from planar 3-SAT [17]: given an instance of a planar 3-SAT formula Φ with n variables and m clauses, and a rectilinear embedding¹ of its graph [15], we construct a set of three trajectories \mathcal{T} such that minimizing HA_{sum} for \mathcal{T} is equivalent to answering the question if Φ is satisfiable.

Variable gadget. The variable gadget (refer to Fig. 4 (left)) consists of a series of building blocks from Fig. 3 (right) with red and blue trajectories having two thin extensions (such that the area covered by them is $O(\varepsilon)$) that will serve as connectors to clauses. Up until entering the variable gadget all three trajectories follow the same path (shown in green in the figure), and they diverge after entering into the gadget. Selecting the red or the blue trajectory for the optimal representative trajectory μ^* to follow at this moment corresponds to setting the variable to true or false. As the variable gadget consists of building blocks that prevent μ^* from switching the color, the next color change can only occur once μ^* exits the variable gadget. One block of the variable gadget contributes $5s_2 + O(s_1)$ area to the total homotopy area between μ^* and the three trajectories.

Clause gadget. The clause gadget (shown in Fig. 4 (right)) consists of three blocks that will be connected to the corresponding variable gadgets: the leftmost and the rightmost blocks are the same as in Fig. 3 (right) (up to change of colors), and the middle block is a similar construction but consists of only three braids from Fig. 3 (left). The green trajectories in-between the blocks represent all three trajectories (red, blue, and yellow) following the same path that connect the current clause to the other clauses in hierarchical order (for more details refer to the next paragraph). The first block allows μ^* to follow the blue or the orange trajectory. In the second block all three trajectories will contribute the same amount to the area measure, thus any of the three trajectories can be chosen by μ^* . The third block allows μ^* to follow the red or the orange trajectory. Moreover, μ^* cannot choose

¹ Recall that in a rectilinear embedding of a graph of a 3-SAT formula, the variable-vertices are placed on a horizontal line, and the clause-vertices are placed above and below the horizontal line and connected to the corresponding variable-vertices with axis-aligned L-shape connectors.

the blue trajectory in the first block and the red trajectory in the second block at the same time, as this would cause a self-intersection. Similarly, μ^* cannot choose the blue trajectory in the second block and the red trajectory in the third block at the same time. Thus, μ^* has to choose the orange trajectory in at least one of the three blocks. Choosing the orange trajectory corresponds to satisfying the clause with the value of the corresponding variable. A clause gadget contributes $5s_2 + O(s_1) + 4s_2 + O(s_1) + 5s_2 + O(s_1) = 14s_2 + O(s_1)$ area to the total homotopy area between μ^* and the three trajectories.

Fig. 5 (left) shows an example of a clause $(\neg x \vee y \vee z)$ connected to the three corresponding variables. It is depicting the case when the clause is satisfied by setting the value of y to true.

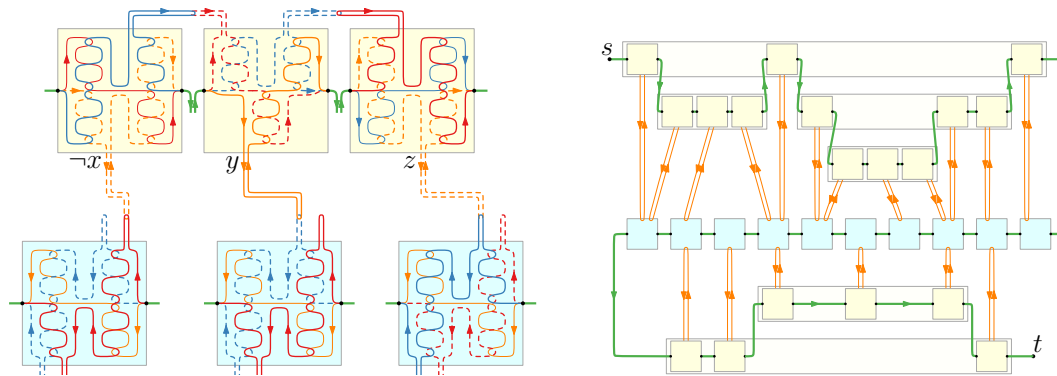
Putting all the building blocks together. Given the rectilinear embedding of the graph of the planar 3-SAT formula, we construct the gadgets for the variables and the clauses. The embedding provides a hierarchy of the clauses that leads to a natural order in which the clauses can be traversed (refer to Fig. 5 (right)). The outermost clause gets traversed the first; the clauses that lie between the first two legs of the outermost clause get traversed after the first block and before the second block of that clause; analogously, the clauses that lie between the second and the third legs of the outermost clause get traversed after the second block and before the third block of that clause; sibling clauses that lie in the same level get traversed one after another. Thus, the three trajectories will start at the top left of the embedding, traversing all the clauses that lie above the horizontal line containing the variables, then they traverse all the variables in order of appearance on the horizontal line, and then they traverse the clauses that lie below the variable line. If formula Φ is satisfiable, the total homotopy area of μ^* is

$$\text{HA}_{\text{sum}}(\mu^*, \mathcal{T}) = \left(\sum_1^n 5k_i + 14m \right) s_2 + \left(\sum_1^n 5k_i + 14m \right) O(s_1) \leq 29ms_2 + 29mO(s_1),$$

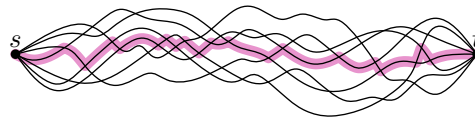
where k_i is the number of blocks in the variable x_i 's gadget, and since some of the blocks can be connected to multiple clauses, $\sum_1^n k_i \leq 3m$. Let $s_2 = 1$, and $s_1 = o(\frac{1}{m})$. If μ^* switches a trajectory at any intersection point inside of any gadget, the total area added as a penalty to HA_{sum} shall be $\gg 29m$. This can be easily achieved by increasing the space between the gadgets. Therefore, deciding if there exists a μ^* such that the total homotopy area $\text{HA}_{\text{sum}}(\mu^*, \mathcal{T})$ is not greater than $29m + o(1)$, is equivalent to deciding if Φ is satisfiable. The size of the construction is polynomial in size of the 3-SAT instance, therefore, it is NP-hard to find a representative trajectory that minimizes HA_{sum} . ◀

4 Minimizing the Sum of Distances HA_{sum} when G is a DAG

We now describe how to compute a representative trajectory that minimizes HA_{sum} for a set of trajectories \mathcal{T} whose trajectory graph is acyclic. For simplicity of presentation, we assume that n is odd. All our proofs can be extended to the case when n is even. As a warmup, we consider the case in which the trajectories in \mathcal{T} are x -monotone. Next, we expand to the case when s and t lie on the boundary of the outer face of G but the trajectories are no longer required to be x -monotone. Finally, we consider the most general case, when s and t lie in the interior faces of G .



■ **Figure 5** Left: An example of clause $(\neg x \vee y \vee z)$ connected to the three corresponding variables. Here, $x = \text{true}$, $y = \text{true}$ and $z = \text{false}$. Right: An order of traversal of the clause and variable gadgets is induced by the embedding of the planar graph of Φ .



■ **Figure 6** The simple median for a set of x -monotone trajectories.

4.1 Minimizing HA_{sum} for x -Monotone Trajectories

In this section we will show that for x -monotone trajectories, the *simple median*, as defined by Buchin et al. [4], minimizes the sum of the homotopy areas HA_{sum} . At the starting point s , the simple median starts at the $\lceil n/2 \rceil^{\text{th}}$ curve (ranking the trajectories by their y -coordinate just after s). It switches to the other trajectory at every intersection point it encounters, thus staying on the $\lceil n/2 \rceil^{\text{th}}$ trajectory. So, for x -monotone trajectories the simple median corresponds to the $\lceil n/2 \rceil$ -level in G . See Fig. 6.

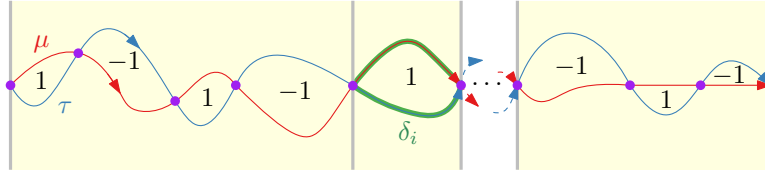
To show that the simple median μ^* minimizes HA_{sum} we now write $\text{HA}_{\text{sum}}(\mu)$ as an integral $\int f(x) dx$. The value $f(x)$ represents the sum of the lengths of a set of intervals along a vertical line with abscissa x . All intervals share a common endpoint $\mu(x)$. The total length of these intervals is minimal when μ has the same number of trajectories above and below it, that is, when it coincides with the simple median.

► **Lemma 5.** *The simple median μ^* minimizes $F(\mu) = \int_x \sum_{\tau_i \in \mathcal{T}} |\mu(x) - \tau_i(x)| dx$.*

Proof. Let y_1, \dots, y_n denote the (y -coordinates of the) intersection points of the trajectories with a vertical line ℓ with abscissa x . Any valid representative trajectory uses one of the points y_i , i.e., $\mu(x) \in \{y_1, \dots, y_n\}$. Note, that the median point $y_{\lceil n/2 \rceil}$ minimizes $f(y) = \sum_{i=1}^n |y - y_i|$. The simple median μ^* is on the $\lceil n/2 \rceil^{\text{th}}$ trajectory at any coordinate x . Therefore, μ^* minimizes $\int_x f(y) dx = \int_x \sum_{i=1}^n |\mu(x) - \tau_i(x)| dx$. ◀

► **Remark.** When n is even, there are two points $y_{\frac{n}{2}}$ and $y_{\frac{n}{2}+1}$ that minimize $f(y) = \sum_{i=1}^n |y - y_i|$. Therefore, any trajectory switching between the levels $\frac{n}{2}$ and $(\frac{n}{2} + 1)$ will minimize $F(\mu) = \int_x \sum_{\tau_i \in \mathcal{T}} |\mu(x) - \tau_i(x)| dx$.

Given a point p let $\omega(p, \delta)$ denote the winding number of p with respect to an oriented closed curve δ . We say that δ is *atomic* if $\omega(p, \delta)$ is either all non-negative, or all non-positive,



■ **Figure 7** There is a subsequence p_1, \dots, p_ℓ of intersection vertices (purple) that partition μ and τ (in red and blue, respectively) into subcurves $\delta_1, \dots, \delta_k$, such that all faces in $\delta_i = \text{loop}(\mu_i, \tau_i)$ (green) have either winding number one or minus one.

for all points $p \in \mathbb{R}^2$. Furthermore, let $W(\delta) = \int_{p \in \mathbb{R}^2} \omega(p, \delta) dp$ denote the *total winding number* of curve δ .

Let μ and τ be two curves from s to t , let $\delta = \text{loop}(\mu, \tau)$ denote the closed curve obtained by concatenating μ and the reverse of τ , and let $s = p_1, \dots, p_\ell = t$ denote the intersection points between μ and τ , ordered along μ . Chambers and Wang [7] show that there is a (not necessarily contiguous) subsequence of the intersection points $\{p_i\}$ that decompose δ into a set of atomic closed curves $\Delta(\mu, \tau) = \delta_1, \dots, \delta_k$, such that the minimum homotopy area $\text{HA}(\mu, \tau) = \sum_{i=1}^k |W(\delta_i)|$. See Fig. 7 for an illustration.

► **Observation 6.** *If μ and τ are x -monotone curves, the atomic curves in $\Delta(\mu, \tau)$ are pairwise disjoint (except for the subsequence of the intersection points $\{p_i\}$).*

► **Theorem 7.** *Let \mathcal{T} be a set of x -monotone trajectories. The simple median μ^* minimizes HA_{sum} .*

Proof. We will show that $\text{HA}_{\text{sum}}(\mu) = F(\mu)$. The theorem then follows from Lemma 5. Using the result of Chambers and Wang [7] we can then rewrite $\text{HA}_{\text{sum}}(\mu)$ as

$$\begin{aligned} \text{HA}_{\text{sum}}(\mu) &= \sum_{\tau_i \in \mathcal{T}} \text{HA}(\mu, \tau_i) = \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} |W(\delta)| = \\ &= \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \left| \int_{p \in \mathbb{R}^2} \omega(p, \delta) dp \right| = \\ &= \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} |\omega((x, y), \delta)| dy dx = \\ &= \int_{x \in \mathbb{R}} \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \int_{y \in \mathbb{R}} |\omega((x, y), \delta)| dy dx. \end{aligned}$$

A vertical line ℓ_x with x -coordinate x intersects (the faces of) G in a set of intervals $\mathcal{I}(x) = I_1, \dots, I_n$. So, for any curve δ that uses only edges of G , all points (values) in an interval I_i have the same winding number $\omega(I_i, \delta)$. So,

$$\begin{aligned} \text{HA}_{\text{sum}}(\mu) &= \int_{x \in \mathbb{R}} \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \sum_{I \in \mathcal{I}(x)} \int_{y \in I} |\omega((x, y), \delta)| dy dx = \\ &= \int_{x \in \mathbb{R}} \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \sum_{I \in \mathcal{I}(x)} |\omega(I, \delta)| \cdot |I| dx. \end{aligned}$$

Since the trajectories are x -monotone, so is μ . The curves $\delta \in \Delta(\mu, \tau_i)$ are built by concatenating a piece of μ and a reversed piece of τ_i . Hence, any vertical line ℓ_x intersects δ in exactly two points: $\mu(x)$ and $\tau_i(x)$. Therefore, any point p on ℓ_x that lies in between

these points has winding number one or minus one with respect to δ . Any point outside the interval defined by $\mu(x)$ and $\tau_i(x)$ has winding number zero. Thus, we get

$$\text{HA}_{\text{sum}}(\mu) = \int_{x \in \mathbb{R}} \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} |\mu(x) - \tau_i(x)| dx.$$

Since μ and all τ_i are x -monotone, Observation 6 gives us that all curves in $\Delta(\mu, \tau_i)$ are pairwise disjoint. This gives us

$$\text{HA}_{\text{sum}}(\mu) = \int_{x \in \mathbb{R}} \sum_{\tau_i \in \mathcal{T}} |\mu(x) - \tau_i(x)| dx = F(\mu). \quad \blacktriangleleft$$

4.2 Extending to Acyclic G with s and t on the Outer Face

The proof from the previous section consists of two steps: (i) show that the simple median minimizes the function F , which represents the sum of interval lengths along a sweep-line; and (ii) show that minimizing the sum of interval lengths along this sweep-line is equivalent to minimizing HA_{sum} . The two key ideas to extend the algorithm to the case in which the trajectory graph is a DAG that has s and t on the outer face (but is otherwise unconstrained), are that (a) we can generalize (i) to minimizing curve-intervals lengths along a sweep-curve, and, (b) a suitable sweep-curve exists for which minimizing the sum of curve-intervals lengths is again the same as minimizing HA_{sum} .

We say that a curve is *conforming* to trajectories \mathcal{T} if and only if it is simple and intersects all trajectories of \mathcal{T} exactly once. Let $\gamma : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^2$ be a continuous map such that for any $u \in [0, 1]$, $\gamma(u) = \bigcup_{z \in \mathbb{R}} \gamma(u, z)$ is an (open) conforming curve that separates s and t , such that for any u , $\gamma(u, -\infty)_y = -\infty$ and $\gamma(u, +\infty)_y = +\infty$, and for any $u \notin [0, 1]$, $\gamma(u) = \bigcup_{z \in \mathbb{R}} \gamma(u, z)$ is an open curve that does not intersect \mathcal{T} . We say that γ is a (conforming) sweep-curve. Assume, without loss of generality, that s lies to the left of $\gamma(u)$ and t to the right of $\gamma(u)$ for all $u \in (0, 1)$.

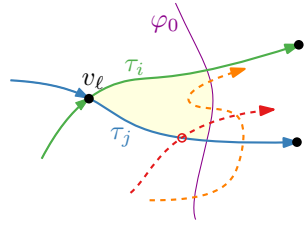
Let $c(\gamma, u, i)$ denote the i^{th} intersection point of $\gamma(u)$ with a trajectory in \mathcal{T} , and let μ_γ be the curve that for any value u corresponds to the $\lceil n/2 \rceil^{\text{th}}$ intersection point on $\gamma(u)$, i.e. $\mu_\gamma(u) = c(\gamma, u, \lceil n/2 \rceil)$. Note that μ_γ is simply connected.

► **Lemma 8.** *Let φ_0 and φ_1 be conforming curves. Furthermore, assume that the only point from $\bigcup \mathcal{T}$ to the left of φ_1 is s . There is a conforming sweep-curve γ that deforms $\gamma(0) = \varphi_0$ into $\gamma(1) = \varphi_1$.*

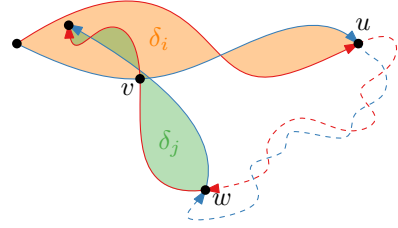
Proof. Let $k(\varphi)$ denote the number of vertices of the trajectory graph G that lie to the left of the conforming curve φ . We have $k(\varphi_1) = 1$, and $k(\varphi_0) = m + 1$, for some $m \in \mathbb{N}$. We now prove by induction on m that we can continuously deform φ_0 into φ_1 while remaining conforming. The lemma then follows.

The base case $m = 0$ is trivial, because two conforming curves without vertices of G between them must intersect exactly the same edges in exactly the same order. Hence, such curves are actually combinatorially equivalent.

For the induction step, let $V = v_0, v_1, \dots, v_z$ denote the vertices of G in topological order, let L be the set of vertices left of φ_0 , and let v_ℓ be the last vertex (with respect to order V) in L . Since φ_0 is conforming, it separates s from t . It follows that $v_\ell \neq t = v_z$, and thus $\ell < z$. Since $v = v_\ell$ is the last vertex in V that lies left of φ_0 , and $\ell < z$, both its outgoing edges cross φ_0 . Furthermore, the area enclosed by these edges and φ_0 is empty of other vertices: the trajectories that visit such a vertex would have to cross φ_0 twice, or they would have to intersect the outgoing edges of v (see Fig. 8). Since such an intersection point would



■ **Figure 8** The region enclosed by the outgoing edges of $v_\ell = v$ and φ_0 (yellow) must be empty.



■ **Figure 9** If δ_i and δ_j intersect in v then u, w , and v form a cycle.

have been a vertex in G both these cases cannot occur. Let τ_i and τ_j be the trajectories on the outgoing edges of v . Since φ_0 is conforming, it intersects τ_i and τ_j at most once, namely on the outgoing edges of v . Therefore, φ_0 does not intersect the incoming edges of v . It follows that we can continuously deform φ_0 into a conforming curve φ'_0 that (a) intersects the trajectories in the same order as φ_0 , with the exception of τ_i and τ_j ; they are swapped, and (b) has the set of vertices $L \setminus \{v\}$ to its left, by sweeping over vertex v , and while remaining conforming at any time. Since the number of vertices to the left of φ'_0 is only $m - 1$, the induction hypothesis gives us that there is a continuous deformation from φ'_0 into φ_1 . This completes the proof. ◀

► **Lemma 9.** *Let φ_0 and φ_1 be conforming curves. There is a conforming sweep-curve γ that deforms $\gamma(0) = \varphi_0$ into $\gamma(1) = \varphi_1$.*

► **Lemma 10.** *Let γ_1 and γ_2 be two conforming sweep-curves, with $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(1) = \gamma_2(1)$, and let $\mu_1 = \mu_{\gamma_1}(u)$ and $\mu_2 = \mu_{\gamma_2}(u)$ be their corresponding median curves for $u \in [0, 1]$. We have that $\mu_1 = \mu_2$.*

Proof. Let u_1, \dots, u_k and v_1, \dots, v_ℓ be the vertices of μ_1 and μ_2 , respectively. Since, $\gamma_1(0) = \gamma_2(0)$, the order in which $\gamma_1(0)$ and $\gamma_2(0)$ intersect the trajectories is the same. It follows that $\mu_1(0) = c(\gamma_0, 0, \lceil n/2 \rceil) = c(\gamma_1, 0, \lceil n/2 \rceil) = \mu_2(0)$, and thus $u_1 = v_1$.

Assume by contradiction that i is the index at which μ_1 and μ_2 diverge for the first time. So, μ_1 and μ_2 both arrive at $v = u_i = v_i$ on the same incoming edge, and leave on different outgoing edges of v . Clearly, μ_j , $j \in [1, 2]$, changes only if γ_j sweeps over a vertex of G . However, since γ_j is conforming, the number of curves intersected by γ_j before v does not change when it sweeps over a vertex $w \neq v$. This means that μ_1 and μ_2 also use the same outgoing edge of $v = v_i = u_i$. This contradicts the fact that i is the first index on which μ_1 and μ_2 diverge. ◀

Recall that μ_γ is the curve that for any value u corresponds to the $\lceil n/2 \rceil^{\text{th}}$ intersection point on $\gamma(u)$. Lemma 10 then implies:

► **Corollary 11.** *There is a unique curve μ^* connecting s to t , such that for any conforming sweep-curve γ , we have that $\mu_\gamma \subseteq \mu^*$.*

A conforming sweep-curve γ is *complete* if and only if the only point from $\bigcup \mathcal{T}$ left of $\gamma(0)$ is s , and the only point from $\bigcup \mathcal{T}$ right of $\gamma(1)$ is t . We then have:

► **Lemma 12.** *Let γ be a conforming sweep-curve that is complete, as defined above. The simple median μ^* minimizes $F_\gamma(\mu) = \int_{u \in [0, 1]} \sum_{\tau_i \in \mathcal{T}} \int_{z \in [z_{\mu(u)}, z_{\tau_i}(u)]} |J(u, z)| \, dz \, du$,*

where $J(u, z)$ is the Jacobian determinant of γ , and $z_{\mu(u)}$ and $z_{\tau_i(u)}$ are the z -coordinates of points $\mu(u)$ and $\tau_i(u)$ respectively.

Proof. The inner integral $\int_{z \in [z_{\mu(u)}, z_{\tau_i(u)}]} |J(u, z)| dz$ represents the length of the curve $\gamma(u)$ between the two intersection points with curves μ and τ_i . Analogous to Lemma 5 we note that $\mu_\gamma(u)$ minimizes $f(\mu) = \sum_{i=1}^n \text{len}_\gamma(\mu(u) - \tau_i(u))$, and therefore μ_γ minimizes $F_\gamma(\mu)$. The lemma follows from the fact that $\mu^* = \mu_\gamma$. ◀

As in Theorem 7 we now rewrite $\text{HA}_{\text{sum}}(\mu)$ as an integral over u . However, instead of directly mapping u to a vertical line we map it to a conforming curve. The resulting mapping is a conforming sweep-curve. Thus, we prove:

► **Lemma 13.** *For any pair of simple paths A and B in G from s to t , the atomic curves in $\Delta(A, B)$ are disjoint.*

Proof. Assume, by contradiction, that $\delta_i \in \Delta(A, B)$ and $\delta_j \in \Delta(A, B)$, with $i < j$, are not disjoint. Then there is an intersection vertex v between δ_i and δ_j . Let \prec denote the topological order of the vertices in G , let u be the ending vertex of δ_i and let w be the starting vertex of δ_j (see Fig. 9). Since $i < j$ we have that $u \prec w$, and since v lies on δ_i we have that $v \prec u$. However, v lies also on δ_j , so we have $w \prec v$, and thus $v \prec u \prec w \prec v$. Contradiction. ◀

► **Theorem 14.** *Let \mathcal{T} be a set of trajectories for which G is acyclic, and s and t are on the outer face of G . The simple median μ^* minimizes HA_{sum} .*

Proof. It is easy to see that there is a conforming curve φ_0 which, from the points in $\bigcup \mathcal{T}$, has just s to its left. Similarly, there exists a conforming curve φ_1 that has only t to its right. Therefore, by Lemma 9 there is a complete conforming sweep-curve γ . Lemma 12 then gives us that the simple median μ^* minimizes $F_\gamma(m)$. We now show that $\text{HA}_{\text{sum}}(\mu) = F_\gamma(m)$. Using the result of Chambers and Wang [7] we again rewrite $\text{HA}_{\text{sum}}(\mu)$ as

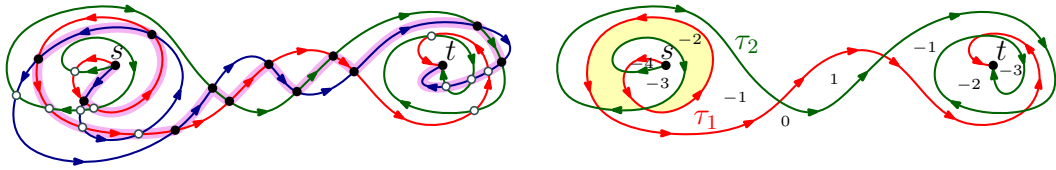
$$\begin{aligned} \text{HA}_{\text{sum}}(\mu) &= \sum_{\tau_i \in \mathcal{T}} \text{HA}(\mu, \tau_i) = \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} |W(\delta)| = \\ &= \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \iint_{(x,y) \in \mathbb{R}^2} |\omega(p(x, y), \delta)| dx dy = \\ &= \sum_{\tau_i \in \mathcal{T}} \sum_{\delta \in \Delta(\mu, \tau_i)} \int_{u \in \mathbb{R}} \int_{z \in \mathbb{R}} |\omega(\gamma(u, z), \delta)| |J(u, z)| dz du. \end{aligned}$$

Since for all $u \in [0, 1]$, $\gamma(u)$ intersects every atomic closed curve $\delta \in \Delta(\mu, \tau_i)$ in exactly two points, and due to Lemma 13 the curves in $\Delta(\mu, \tau_i)$ are pairwise disjoint, we get that

$$\sum_{\delta \in \Delta(\mu, \tau_i)} \int_{z \in \mathbb{R}} |\omega(\gamma(u, z), \delta)| |J(u, z)| dz = \int_{z \in [z_{\mu(u)}, z_{\tau_i(u)}]} |J(u, z)| dz,$$

where $z_{\mu(u)}$ and $z_{\tau_i(u)}$ are z -coordinates of the intersection points of $\gamma(u)$ with μ and τ_i respectively. And since $\gamma(u)$ does not intersect any trajectory in \mathcal{T} for $u \notin [0, 1]$,

$$\begin{aligned} \text{HA}_{\text{sum}}(\mu) &= \sum_{\tau_i \in \mathcal{T}} \int_{u \in \mathbb{R}} \int_{z \in [z_{\mu(u)}, z_{\tau_i(u)}]} |J(u, z)| dz du = \\ &= \int_{u \in [0, 1]} du \cdot \sum_{\tau_i \in \mathcal{T}} \int_{z \in [z_{\mu(u)}, z_{\tau_i(u)}]} |J(u, z)| dz = F_\gamma(\mu). \end{aligned} \quad \blacktriangleleft$$



■ **Figure 10** Left: A set of three trajectories with s and t lying on the boundary of an interior face. The optimal representative (light-purple) does not switch at every intersection. Right: The winding numbers for $loop(\tau_1, \tau_2)$. The highlighted face is swept twice by a minimal homotopy.

► **Observation 15.** *Note that the simple median μ^* minimizes F_γ , and thus HA_{sum} , among all curves from s to t , even ones that are not necessarily restricted to consist of pieces of the input trajectories.*

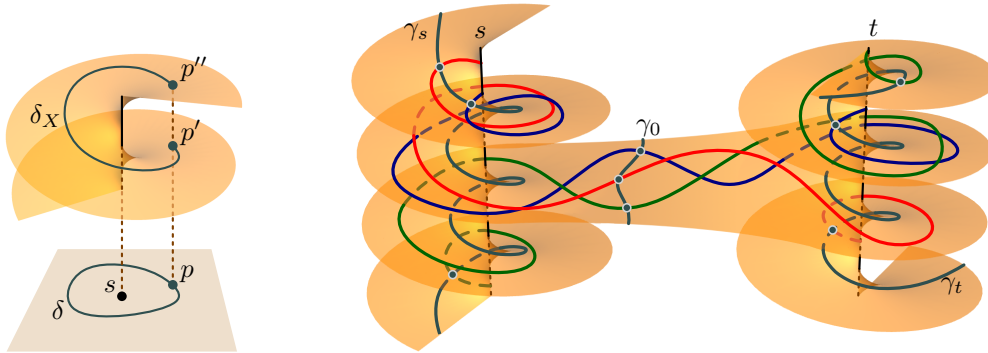
► **Remark.** When n is odd there is a unique curve that minimizes F_γ , and it is simple median μ^* . When n is even there can be multiple curves, not necessarily restricted to consist of pieces of the input trajectories, that all minimize F_γ . These curves are all bounded by the $\frac{n}{2}$ - and $(\frac{n}{2} + 1)$ -levels of the trajectories.

4.3 Extending to Acyclic G with Unrestricted s and t

In this section we extend our approach to compute an optimal representative trajectory when s and t can be anywhere in the DAG. However, unlike in the previous two sections, we can no longer start at the median trajectory from s and switch at every intersection point we encounter. Fig. 10 (left) shows an example of a set of trajectories in which any curve that always switches is not optimal, no matter where we start. The main reason why our argument breaks here is that the winding numbers between the individual pairs of curves are no longer just in the range $[-1, 1]$. Hence, an optimal homotopy may have to sweep over a face more than once.

Instead, we will lift the trajectories into a space X that we will construct from the covering spaces of $\mathbb{R}^2 \setminus s$ and $\mathbb{R}^2 \setminus t$; we refer the reader to a standard topology text for detailed definitions of covering spaces [14, 18]. The key in this setting is that we will be able to lift the trajectories into X in such a way that the trajectory graph will form a DAG with s and t on the outer face and the pairwise homotopy areas between lifted trajectories will be the same as the homotopy area in the plane. We then compute an optimal representative for the lifted trajectories, using our simple median algorithm, and show that its corresponding projection is an optimal representative for the original trajectories.

The space X . Intuitively, we start with a covering space of the space formed from the plane by cutting out small disks around s and t , where each of the boundaries is collapsed to a single point, and then obtain space X by adding the points s and t back. This means that X will cover $\mathbb{R}^2 \setminus \{s, t\}$ with infinitely many “layers” forming Riemann-like spirals around points s and t . Consider a simple cycle δ in \mathbb{R}^2 that goes through some point p and that encloses s or t . A walk along δ starting at p and ending at p in \mathbb{R}^2 corresponds to a curve in X that starts at the copy of p in some layer i and ends at the copy of p in layer $i + 1$ or $i - 1$. See Fig. 11 for an illustration. We formalize this more carefully (including the metric on the space X , which will be necessary in order to argue about the homotopy area) using a particular conforming curve, as follows:



■ **Figure 11** Lifting cycle δ from \mathbb{R}^2 into X . ■ **Figure 12** Input trajectories lifted into the space X .

► **Lemma 16.** *Given a set of simple trajectories \mathcal{T} that start at s and end in t , whose arrangement forms a directed acyclic graph G , there exists a conforming curve γ with endpoints at infinity that separates s and t .*

Proof. The graph G defines a partial order on the intersection points of the trajectories in \mathcal{T} . Consider a conforming 0-length cycle δ enclosing s . Let δ grow by sweeping over the vertices of G according to their partial order. As in Lemma 8 we can do this while maintaining conformity with respect to \mathcal{T} . Once γ crosses some intersection point that lies on the outer face of G , we can cut it at any point in the outer face and pull the endpoints toward infinity. The resulting open curve γ is conforming and separates s from t . ◀

Let $\gamma \subset \mathbb{R}^2$ be a conforming curve that separates s from t and has its end-points at infinity, and let p_0 be an arbitrary point on γ . Note that by Lemma 16 such a curve exists. Define space X_s corresponding to $\mathbb{R}^2 \setminus \{s\}$ in terms of polar coordinates, taking s to be the origin: Let r be the “radius” parameter, and let θ be the angular parameter, such that point $(0, \|sp_0\|) \in X_s$ corresponds to point $p_0 \in \mathbb{R}^2 \setminus \{s\}$, and a positive θ corresponds to a clockwise turn. We then have $X_s = \{(\theta, r) \mid (\theta, r \in \mathbb{R} \setminus \{s\}) \wedge r > 0\}$. Note that in the definition of X_s we explicitly do *not* limit the range of θ to $[0, 2\pi)$ (which would give us exactly \mathbb{R}^2 , parameterized around s). Instead, our space X_s allows us to “wind around” s an arbitrary number of times. Analogously, define X_t . Note, that X_s and X_t are the universal covers of $\mathbb{R}^2 \setminus \{s\}$ and $\mathbb{R}^2 \setminus \{t\}$, respectively.

We partition X_s into layers L_i , with $i \in \mathbb{Z}$. A point $(\theta, r) \in X_s$ lies in layer L_i , if and only if $\lfloor \theta/2\pi \rfloor = i$. We define layers analogously for X_t .

Recall that γ separates s and t in \mathbb{R}^2 , and hence we can consider a copy of γ in each of X_s and X_t . We cut each space along the copy of γ , and glue the part of X_s containing s and the part of X_t containing t together along each’s copy of γ . Furthermore, we again add the points s and t , and connect them to all the layers of X_s and X_t . Note, that we add the points s and t only so that all trajectories again start at s and end at t rather than arbitrarily close to s and t . Let X to be the resulting space.

Lifting trajectories. Next, we describe the image of trajectories \mathcal{T} in X . (See Fig. 12.) For each trajectory $\tau_i \in \mathcal{T}$, we construct its corresponding trajectory τ'_i in X by starting at the image of the intersection point of τ_i with γ , and moving along τ_i while continuously mapping the points to τ'_i . We call this process *lifting* trajectory τ_i to space X . Let $\mathcal{T}' = \{\tau'_i \mid \tau_i \in \mathcal{T}\}$ denote the set of resulting trajectories, and let G' be the corresponding trajectory graph.

We note that all trajectories cross the conforming curve γ and hence are fixed on a common reference, although not at a common lift of a base point as is more commonly seen in topology. However, we obtain that each trajectory lifts to a unique curve in X . In addition, since we have local homeomorphisms which lift everywhere (except right at s and t), we can also lift the definition of a winding number for any point p inside $loop(\tau'_i, \tau'_j)$ for any pair of lifted trajectories τ'_i and τ'_j . This leads to the following observation:

► **Observation 17.** *The points s and t lie on the outer face of G' . Thus, for any simple paths A and B in G' from s to t , and for any point $p \in X$, we have that $|\omega(p, loop(A, B))| \leq 1$.*

► **Lemma 18.** *For any two curves ϕ_1 and ϕ_2 in \mathbb{R}^2 that connect s to t . We have $HA(\phi_1, \phi_2) = HA(\phi'_1, \phi'_2)$, where ϕ'_1 and ϕ'_2 are the corresponding curves lifted into space X .*

Proof. It is easy to see that $HA(\phi_1, \phi_2) \leq HA(\phi'_1, \phi'_2)$: the covering map f_X is continuous, so a minimum homotopy between ϕ'_1 and ϕ'_2 defines a homotopy between ϕ_1 and ϕ_2 of cost $HA(\phi'_1, \phi'_2)$. Since $HA(\phi_1, \phi_2)$ is a minimum homotopy we have that $HA(\phi_1, \phi_2) \leq HA(\phi'_1, \phi'_2)$.

Next, we show that $HA(\phi_1, \phi_2) \geq HA(\phi'_1, \phi'_2)$. The lemma then follows. Let $\delta = loop(\phi_1, \phi_2)$ and $\delta' = loop(\phi'_1, \phi'_2)$. Now assume, by contradiction, that $HA(\phi_1, \phi_2) > HA(\phi'_1, \phi'_2)$. It follows that there is a point $p \in \mathbb{R}^2$, with $\omega(p, \delta) = k$ that is swept by a minimum homotopy H between ϕ'_1 and ϕ'_2 more than k times. Furthermore, assume without loss of generality that p lies left of γ , and thus the copies of p lie in X_s . Since all winding numbers in X are in the range $[-1, 1]$ (Observation 17) that means there must be more than k copies of point p swept by H . It follows that there is a point q' in layer L_ℓ or layer $L_{-\ell}$, with $\ell > k$, that lies on δ' , and has a larger r -coordinate than p' (otherwise we would not sweep over p'). Furthermore, note that δ' intersects γ (as the curves ϕ'_1 and ϕ'_2 connect s to t), and thus contains a point on in layer L_0 . Now consider traversing δ' , starting from point q' . It follows that the total turning angle is at least $2\ell\pi$ (since we must visit layer L_0). This means that the total turning angle of curve δ with respect to $f_X(p)$ is also at least $\ell 2\pi$. Therefore $\omega(p, \delta) \geq \ell > k$. Contradiction. ◀

► **Corollary 19.** *Let μ be a representative for the set of trajectories \mathcal{T} , and let μ' be its corresponding representative for \mathcal{T}' . We have that $HA_{sum}(\mu, \mathcal{T}) = HA_{sum}(\mu', \mathcal{T}')$.*

Any representative trajectory μ in X corresponds to some representative trajectory μ' in \mathbb{R}^2 . However, not every representative trajectory in \mathbb{R}^2 has a corresponding representative trajectory in X . The difference between the two cases comes from the fact, that some of the intersection points between trajectories \mathcal{T} that existed in \mathbb{R}^2 no longer exist once the trajectories are lifted to X . We call the intersection points of \mathcal{T} that remain in X *legal*, and the ones that disappear *illegal*. Following the projection to \mathbb{R}^2 of the median trajectory μ'_* corresponds to switching the trajectory at every legal intersection point. Next we will prove, that this projection of μ'_* to \mathbb{R}^2 gives an optimal representative trajectory for trajectories \mathcal{T} .

► **Theorem 20.** *Let \mathcal{T} be a set of trajectories for which G is acyclic, let μ'_* be the simple median on the lifted trajectories in X . The representative curve $\mu^* = f_X(\mu'_*)$ corresponding to μ'_* minimizes HA_{sum} with respect to \mathcal{T} .*

Proof. Suppose there exists some representative trajectory μ with the total homotopy area $HA_{sum}(\mu, \mathcal{T}) < HA_{sum}(\mu^*, \mathcal{T})$. Let curve μ' be μ lifted to X . If μ uses only legal intersections in G , then μ' is a candidate representative for the set of trajectories \mathcal{T}' . By Corollary 19 we then have that $HA_{sum}(\mu, \mathcal{T}) = HA_{sum}(\mu', \mathcal{T}') < HA_{sum}(\mu'_*, \mathcal{T}') = HA_{sum}(\mu^*, \mathcal{T})$. Contradiction. If μ uses also illegal intersections, then μ' does not consist of pieces of the trajectories \mathcal{T}' .

Theorem 14 and Observation 15 then implies that $\text{HA}_{\text{sum}}(\mu', \mathcal{T}') \geq \text{HA}_{\text{sum}}(\mu'_*, \mathcal{T}')$. Applying Corollary 19 on both sides then gives us $\text{HA}_{\text{sum}}(\mu, \mathcal{T}) \geq \text{HA}_{\text{sum}}(\mu^*, \mathcal{T})$. Contradiction. ◀

5 Computing a Representative Trajectory

From Theorem 14 it immediately follows that if s and t lie on the outer face of G , we can compute a representative μ^* trajectory that minimizes HA_{sum} using the algorithm of Buchin et al. [4]. Thus, we can compute μ^* in $O((N+k)\alpha(N)\log(N))$ time, where N is the total complexity of the input trajectories, and k is the output complexity. For an arbitrary DAG G , we have $k = O(N^2)$. If the trajectories are x -monotone, the simple median corresponds to the $\lceil n/2 \rceil$ -level in an arrangement of n curves, and thus bounds on the complexity of the $\lceil n/2 \rceil$ -level also bound the complexity of μ^* . In case our trajectories (curves) are all polylines with at most m vertices each, we have $k = O(mn^{4/3} \log^{1/3-\varepsilon} n)$, for some arbitrarily small constant $\varepsilon > 0$ [9]. Similarly, we can derive the results for more general types of curves.

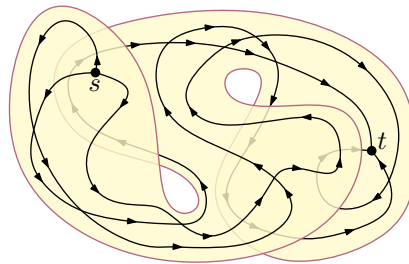
Unrestricted s and t . When s and t are not restricted to lie on the outer face we first construct a conforming curve γ that separates s from t and intersects the outer face. This allows us to find the $\lceil n/2 \rceil^{\text{th}}$ intersection p of γ with the trajectories \mathcal{T} , which is guaranteed to lie on the representative μ'_* that minimizes the homotopy area of the lifted trajectories \mathcal{T}' . We can now construct G' from G by walking along the trajectories, starting from their intersection points with γ . Similarly, we can trace μ'_* through G' , starting from p . The representative μ'_* then also gives us an optimal representative μ^* (Theorem 20). All that remains is to describe how to construct γ . We do this using the same procedure as used in the proof of Lemma 16: we explicitly construct G , sort the vertices in topological order, and add the vertices in this order to some set L . Once L contains a vertex v on the outer face of G , we can construct γ , starting on the outgoing edge of v incident to the outer face, and walking through G , while keeping exactly the set of vertices L to our left. It is easy to see that computing μ^* using this algorithm takes $O(|G|) = O(N^2)$ time. We summarize our results in the following theorem.

► **Theorem 21.** *Let \mathcal{T} be a set of trajectories that all start in s and end in t , and whose trajectory graph G is acyclic. If s and t lie on the outer face, a representative trajectory μ^* that minimizes HA_{sum} can be computed in $O((N+k)\alpha(N)\log(N))$ time, where N is the total complexity of the trajectories in \mathcal{T} , and k is the complexity of the resulting trajectory. If s and t can be anywhere in G , μ^* can be computed in $O(|G|) = O(N^2)$ time.*

6 Future Work

We have shown that computing a representative that minimizes HA_{sum} is NP-hard when the trajectory graph G may be an arbitrary graph, and we have presented an efficient algorithm for when G is a DAG. Hence, our results cover all cases. However, clearly there are situations in which the trajectories are similar, but for which the trajectory graph is not a DAG. Hence, we would like a more fine grained classification which kind of graphs allow us to find a representative efficiently.

We expect that we can extend our approach from Section 4.3 to cases in which the trajectories are have a similar “shape” but their trajectory graph contains cycles. In particular, we again lift the trajectories into a space, or *corridor*, that captures the global shape of the trajectories, and in which the trajectory graph forms a DAG. We then compute a concrete curve representing the trajectories in this space. The conceptual existence of a



■ **Figure 13** Even when trajectories do not form a DAG, the trajectories may lie in a “corridor”.

corridor is justified by the assumption that the input trajectories are similar. See Fig. 13 for an illustration.

While we focused on using the homotopy area to measure distance between the trajectories, there are other alternative measures that balance topology and geometry. Homotopy width (or homotopic Fréchet distance) [8] and homotopy height [5, 13] are obvious options, as is homology area [6], although it is unclear if any of these are tractable or useful in practice.

References

- 1 Pankaj K. Agarwal, Mark de Berg, Jie Gao, Leonidas J. Guibas, and Sariel Har-Peled. Staying in the middle: Exact and approximate medians in \mathbb{R}^1 and \mathbb{R}^2 for moving points. In *CCCG*, pages 43–46, 2005.
- 2 Riddhipratim Basu, Bhaswar B. Bhattacharya, and Tanmoy Talukdar. The projection median of a set of points in \mathbb{R}^d . *Discrete & Computational Geometry*, 47(2):329–346, 2012.
- 3 Kevin Buchin, Maike Buchin, Joachim Gudmundsson, Maarten Löffler, and Jun Luo. Detecting commuting patterns by clustering subtrajectories. *IJCGA*, 21(03):253–282, 2011. doi:10.1142/S0218195911003652.
- 4 Kevin Buchin, Maike Buchin, Marc Kreveld, Maarten Löffler, Rodrigo I. Silveira, Carola Wenk, and Lionov Wiratma. Median trajectories. *Algorithmica*, 66(3):595–614, 2013. doi:10.1007/s00453-012-9654-2.
- 5 Erin W. Chambers and David Letscher. On the height of a homotopy. In *CCCG*, pages 103–106, 2009.
- 6 Erin Wolf Chambers and Mikael Vejdemo-Johansson. Computing minimum area homologies. *Computer Graphics Forum*, 2014. doi:10.1111/cgf.12514.
- 7 Erin Wolf Chambers and Yusu Wang. Measuring similarity between curves on 2-manifolds via homotopy area. In *Proc. 29th Ann. Symp. on CG*, pages 425–434. ACM, 2013. doi:10.1145/2462356.2462375.
- 8 Erin Wolf Chambers, Éric Colin de Verdière, Jeff Erickson, Sylvain Lazard, Francis Lazarus, and Shripad Thite. Homotopic fréchet distance between curves or, walking your dog in the woods in polynomial time. *CG*, 43(3):295–311, 2010. doi:10.1016/j.comgeo.2009.02.008.
- 9 Timothy M Chan. On levels in arrangements of curves, iii: further improvements. In *Proc. of the 24th annual symposium on Computational geometry*, pages 85–93. ACM, 2008.
- 10 Stéphane Durocher and David Kirkpatrick. The projection median of a set of points. *CG*, 42(5):364–375, 2009.
- 11 S. Gaffney, A. Robertson, P. Smyth, S. Camargo, and M. Ghil. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics*, 29(4):423–440, 2007.

- 12 S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In *Proc. 5th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pages 63–72, 1999.
- 13 Sarel Har-Peled, Amir Nayyeri, Mohammad Salavatipour, and Anastasios Sidiropoulos. How to walk your dog in the mountains with no magic leash. In *Proc. 28th Ann. Symp. on CG*, pages 121–130. ACM, 2012. doi:10.1145/2261250.2261269.
- 14 Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2001. URL: <http://www.math.cornell.edu/~hatcher/>.
- 15 Donald E. Knuth and Arvind Raghunathan. The problem of compatible representatives. *SIAM Journal on Discrete Mathematics*, 5(3):422–427, 1992. doi:10.1137/0405033.
- 16 J.G. Lee, J. Han, and K.Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 593–604, 2007.
- 17 David Lichtenstein. Planar formulae and their uses. *SIAM J. Comput.*, 11(2):329–343, 1982. doi:10.1137/0211025.
- 18 James R. Munkres. *Topology*. Prentice-Hall, 2nd edition, 2000.
- 19 M. Vlachos, D. Gunopulos, and G. Kollios. Discovering similar multidimensional trajectories. In *Proc. 18th Int. Conf. Data Engin.*, pages 673–684, 2002.