

Untangling Graphs on Surfaces

Éric Colin de Verdière*

Vincent Despré†

Loïc Dubois‡

Abstract

Consider a graph drawn on a surface (for example, the plane minus a finite set of obstacle points), possibly with crossings. We provide an algorithm to decide whether such a drawing can be *untangled*, namely, if one can slide the vertices and edges of the graph on the surface (avoiding the obstacles) to remove all crossings; in other words, whether the drawing is homotopic to an embedding. While the problem boils down to planarity testing when the surface is the sphere or the disk (or equivalently the plane without any obstacle), the other cases have never been studied before, except when the input graph is a cycle, in an abundant literature in topology and more recently by Despré and Lazarus [SoCG 2017, J. ACM 2019], who gave a near-linear algorithm for this problem.

Our algorithm runs in $O(m + \text{poly}(g+b)n \log n)$ time, where $g \geq 0$ and $b \geq 0$ are the genus and the number of boundary components of the input orientable surface \mathcal{S} , and n is the size of the input graph drawing, lying on some fixed graph of size m cellularly embedded on \mathcal{S} .

We use various techniques from two-dimensional computational topology and from the theory of hyperbolic surfaces. Most notably, we introduce *reducing triangulations*, a novel discrete analog of hyperbolic surfaces in the spirit of *systems of quads* by Lazarus and Rivaud [FOCS 2012] and Erickson and Whittlesey [SODA 2013], which have the additional benefit that reduced paths are unique and stable upon reversal; they are likely of independent interest. Tailored data structures are needed to achieve certain homotopy tests efficiently on these triangulations. As a key subroutine, we rely on an algorithm to test the *weak simplicity* of a graph drawn on a surface by Akitaya, Fulek, and Tóth [SODA 2018, TALG 2019].

1 Introduction

In this paper, we study the following problem: Given a *drawing* δ of a graph G on a (compact, connected, orientable) surface \mathcal{S} , possibly with boundary, decide whether it is possible to *untangle* δ , that is, to make δ crossing-free by a continuous motion; more formally, whether δ is homotopic to an embedding, where the homotopy may move vertices and edges. See Figure 1.

We first remark that, in the case where the surface is topologically trivial (the sphere or the disk), this problem boils down to planarity testing, and is thus solvable in linear time (see Hopcroft and Tarjan [31]). Before stating our results in detail and presenting the main techniques, we survey related works, on curves and graphs on surfaces.

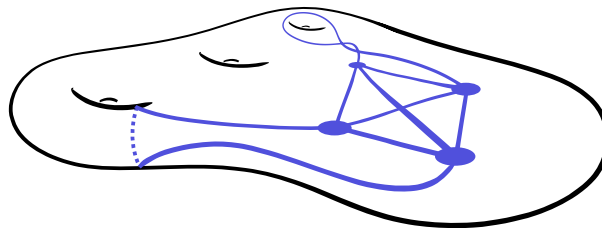


Figure 1: A drawing of a graph that cannot be untangled, on an orientable surface of genus three without boundary.

Related work on curves on surfaces. Topological questions on curves on surfaces have been an important field of study in the mathematical community for more than one hundred years. Dehn [19] gave combinatorial characterizations of whether a closed curve on a surface is contractible (can be continuously moved to a point) or whether two such closed curves are freely homotopic (can be continuously deformed into each other). Poincaré [40]

*LIGM, CNRS, Univ Gustave Eiffel, F-77454 Marne-la-Vallée, France.

†Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France.

‡LIGM, CNRS, Univ Gustave Eiffel, F-77454 Marne-la-Vallée, France.

provided similar characterizations to decide whether a closed curve on a surface is homotopic to a simple closed curve (in our language, can be untangled). Although Dehn and Poincaré did not provide detailed analyses of their constructions, it is clear that they can be transformed into effective algorithms; see, e.g., Stillwell [44, Chapter 6] for an overview.

Generally, surfaces of genus at least two without boundary turn out to be the most difficult case. For such a surface \mathcal{S} , a key insight, already present in these early works but reused in many other ones mentioned below, is the following: \mathcal{S} can be endowed with a hyperbolic metric (in the same way as the torus can be endowed with a Euclidean metric); under this metric, every closed curve is homotopic to a unique geodesic; moreover, any family of geodesic curves crosses minimally (has the least number of crossings among all curves in their respective homotopy classes).

The problem of determining whether a closed curve can be untangled has been extensively considered by the mathematical community in a long series of papers since the early 1960s; see Reinhart [41], Chillingworth [12, 13], and Birman and Series [5]. Later Cohen and Lustig [14] and Lustig [38] (see also Hass and Scott [30]) built upon these works to determine the geometric self-intersection number of a closed curve and the geometric intersection number of two closed curves (the number of crossings of homotopic curves crossing minimally). De Graaf and Schrijver [18] proved that it is possible to make curves cross minimally using homotopy (a.k.a. Reidemeister) moves that never increase the number of (self-)crossings.

These problems have been revisited under a more algorithmic lens by computational topologists since the 1990s. Contractibility and homotopy can be tested in linear time, as proved by Lazarus and Rivaud [37] and Erickson and Whittlesey [25], building upon earlier work by Dey and Guha [22]. One can decide whether a closed curve of length n , lying on a graph of size m itself embedded on \mathcal{S} , can be untangled in $O(m + n \log n)$ time, as proved by Despré and Lazarus [21], who also proved that the geometric self-intersection number can be computed in $O(m + n^2)$ time. Related works, by Chang and Erickson [10] and Chang and de Mesmay [9], focus on computing the number of non-increasing homotopy moves needed to make a set of closed curves cross minimally. Although all these algorithms are purely combinatorial, in many cases their proofs of correctness involve tools from hyperbolic geometry, suitably discretized, or at least rely on some intuition from hyperbolic geometry.

Related to deciding whether a curve can be untangled, the following problem has been studied recently: Decide whether a given closed curve drawn in a graph is *weakly simple*, namely, whether it can be untangled in a neighborhood of the graph by an arbitrarily small perturbation. Chang, Erickson, and Xu [11], building upon earlier work by Cortese, Di Battista, Patrignani, and Pizzonia [17], provided a near-linear algorithm.

Related work on graphs on surfaces. Many of the above questions can be extended from curves to graphs on surfaces. Surprisingly, the literature studying them for graphs is rather scarce, in stark contrast with the central importance of graphs in theoretical computer science.

In the mathematical community, the only work that we are aware of, due to Ladegaillerie [35], is a reduction from the test of existence of an isotopy (a continuous family of embeddings) between two embeddings of a graph on a surface, to a test of isomorphism between combinatorial maps and an infinite number of homotopy tests between closed curves. This characterization has been refined by Colin de Verdière and de Mesmay [15], leading to a polynomial time algorithm.

The well studied embeddability problem is to decide whether an input graph G can be embedded on an input surface \mathcal{S} . Our problem adds some homotopy constraints to the embeddability problem. The latter is NP-complete when \mathcal{S} is part of the input, as proved by Thomassen [45]; however, Mohar [39] has given a linear time algorithm when \mathcal{S} is fixed, later simplified and improved by Kawarabayashi, Mohar, and Reed [33]. Our result indicates that adding these homotopy constraints makes the problem solvable in polynomial time, even if \mathcal{S} is part of the input.

There is a vast interest in crossing numbers for graphs drawn in the plane or in surfaces (for a survey, see Schaefer [42]), including constraints on the parity of the number of crossings between edges, related to the Hanani-Tutte theorem (see Schaefer [43]). However, the constraint that we study, namely, fixing the homotopy of the drawing, appears to be entirely new. A geometric variant has been studied by Goac, Kratochvíl, Okamoto, Shin, Spillner, and Wolff [28]. Given a straight-line drawing of a graph in the plane, they proved that the minimum number of vertices that have to be moved in order to untangle the drawing (the edges remaining drawn as line segments) is NP-hard to compute and to approximate; they also provided upper and lower bounds on the number of required moves.

More directly related to our work, in a recent paper, Atikaya, Fulek, and Tóth [3] extended the weak simplicity

problem to graphs: Given a drawing δ of a graph G in another graph H , itself embedded on a surface \mathcal{S} , is it possible to make δ simple with an arbitrarily small perturbation? They solved this problem in near-linear time. We will heavily rely on their algorithm, and postpone a detailed description to Section 2.2. More recently, Fulek [27] gave a polynomial time algorithm to decide whether a graph drawn in the plane (without obstacles) can be perturbed by an arbitrarily small perturbation to turn it into an embedding in a pre-specified isotopy class.

Our results. The input to our problem is a fixed cellular embedding H of a graph on a surface \mathcal{S} , together with a drawing δ of a graph G on H , in the sense that δ maps vertices of G to vertices of H and edges of G to walks in H . We obtain the following result:

THEOREM 1.1. *Let H be a graph of size m cellularly embedded on an orientable surface \mathcal{S} of genus $g \geq 0$ with $b \geq 0$ boundary components. Let G be a graph and let $\delta : G \rightarrow H$ be a drawing of size n . We can decide whether there is an embedding of G on \mathcal{S} homotopic to δ in $O(m + (g + b)^2 n \log((g + b)n))$ time.*

(Again, the cases of the sphere and the disk boil down to planarity testing so we omit these from now on.) As in some previous papers in the area, e.g., [25, 37], we use the RAM model, in which pointers and integers bounded by a polynomial in the input size can be manipulated in constant time (see Aho, Hopcroft, and Ullman. [2] or Agarwal [1, Section 40.1]). If the drawing δ can be untangled, it is clear from our proofs that one can compute an embedding homotopic to δ in polynomial time; details are omitted from this version.

We also consider an alternative framework in the case where \mathcal{S} is the plane minus a finite set of obstacle points, and δ is a piecewise linear drawing of G avoiding the obstacles. In this framework, we prove:

THEOREM 1.2. *Let P be a set of p points in \mathbb{R}^2 . Let G be a graph and let $\delta : G \rightarrow \mathbb{R}^2 \setminus P$ be a piecewise linear drawing of size n . In time $O(p^{5/2} n \log(pn))$, we can decide whether there is an embedding homotopic to δ in $\mathbb{R}^2 \setminus P$.*

A new tool: reducing triangulations. As is typical, the most difficult case is when the input surface \mathcal{S} has genus at least two and no boundary, so we focus on this case. For this purpose, we introduce the concept of *reducing triangulation* of \mathcal{S} , a triangulation T of \mathcal{S} with all vertex degrees at least eight and whose dual graph is bipartite. Reducing triangulations form a new discrete analog of hyperbolic surfaces, and we define the notion of *reduced walk* in T , which behave similarly as geodesics in the continuous case. A (possibly closed) walk W in T is reduced if no “local” reduction rule can be applied to W . We prove that W is (possibly freely) homotopic to a unique reduced walk, which we can compute in linear time, and that reduced walks are stable upon reversal. *Systems of quads*, introduced by Lazarus and Rivaud [37], refined by Erickson and Whittlesey [25], and reused by Despré and Lazarus [21], have the same objective; in systems of quads, one can ensure either uniqueness in a given homotopy class or stability upon reversal, but (presumably) not both at the same time, which induces substantial technicalities in the previous papers [21, 25, 37]. Reducing triangulations are thus likely to be of independent interest. Incidentally, the definition of reduced walks depends on the choice of an orientation of the surface, which is the main reason why we state our results for orientable surfaces only.

Overview of the paper. After some preliminaries (Section 2), Sections 3 to 6 are devoted to a proof of Theorem 1.1 in the case where H is a reducing triangulation on a surface of genus at least two without boundary; more precisely:

THEOREM 1.3. *Let T be a reducing triangulation of an orientable surface \mathcal{S} of genus $g \geq 2$ without boundary. Let G be a graph and let $\delta : G \rightarrow T$ be a drawing of size n . We can determine whether there is an embedding of G on \mathcal{S} homotopic to δ in $O(gn \log(gn))$ time.*

In detail, we introduce reducing triangulations, reduced walks and their properties, and reduction algorithms in Section 3. We then prove Theorem 1.3 in a special case where G is a (sparse) loop graph—each connected component of G contains a single vertex (Proposition 4.1, in Section 4). The main idea, in the same spirit as Despré and Lazarus [21], is to use the fact that geodesics in hyperbolic surfaces cross minimally, and to prove that this remains true to some extent in our discrete analog. In our case, roughly but not exactly, reducing all loops of a loop graph makes it weakly simple unless it cannot be untangled, so our algorithm eventually applies the weak simplicity test by Akitaya, Fulek, and Tóth [3]; but the proof of correctness involves putting a particular hyperbolic metric on the surface obtained by puncturing \mathcal{S} and requires delicate arguments in the compactification of the universal cover of \mathcal{S} . Then, to solve the problem for an arbitrary graph G , at a high

level we contract a spanning forest of G and apply Proposition 4.1 to the resulting loop graph, after removing the contractible loops and identifying the homotopic loops. If the loop graph cannot be untangled, neither can the original graph. Otherwise, it turns out that the graph can be untangled in \mathcal{S} if and only if it is a weak embedding in (a neighborhood of) the loop graph, but proving this fact is rather subtle; we formalize this using the notion of *factorization* (informally, the contraction of the graph, and the removals and identifications of loops in the resulting loop graph), and conclude the proof of Theorem 1.3 by invoking once more the weak simplicity algorithm [3], all this in Section 5. A more efficient algorithm to compute a factorization, requiring tailored data structures to achieve certain homotopy tests efficiently, is deferred to Section 6.

After proving Theorem 1.3, we prove Theorem 1.1 for the case $g \geq 2$, $b = 0$ in Section 7, essentially by converting our input embedded graph H into a reducing triangulation. In Section 8, we prove our result for the remaining orientable surfaces, namely, the torus and the surfaces with non-empty boundary; while the overall strategy is the same, reducing triangulations are not needed anymore, which dramatically simplifies the algorithm. Finally, in Section 9, we prove our result on the plane with obstacles (Theorem 1.2).

2 Preliminaries

2.1 Graphs on surfaces, homotopies, and untangling In this paper, graphs are finite and undirected, but may have loops and parallel edges. We use standard notions of graph theory, in particular the notion of *walk* in a graph. In a walk of length k , $(v_0, e_0, v_1, e_1, \dots, v_{k-1}, e_{k-1}, v_k)$, the first and last occurrences of vertices, v_0 and v_k , are its *endpoints*, and the other occurrences, v_1, \dots, v_{k-1} , are the *interior vertices*. Vertices and edges may be repeated. The two endpoints of a walk may coincide; a walk can even be reduced to a single vertex. A *closed walk* is similar to a walk, but the vertices are ordered cyclically instead of linearly; every vertex of a closed walk is an interior vertex.

We need some basic topology of surfaces [4, 44]; we only provide the most basic definitions, sometimes only with an informal description. All the surfaces we consider are connected and orientable, so we omit these adjectives in the sequel. A compact *surface*, possibly with boundary, is determined up to homeomorphism by its *genus* (number of handles) and number of boundary components [4, Theorem 1.5]. We will occasionally consider *non-compact surfaces*, which will be either obtained by removing finitely many points (*punctures*) of a compact surface, or a universal cover, which are (usually) non-compact; see below.

A *path* on a surface \mathcal{S} is a continuous map $p : [0, 1] \rightarrow \mathcal{S}$; its *endpoints* are $p(0)$ and $p(1)$. A *loop* with *basepoint* b is a path whose both endpoints equal b . A path is *simple* if it is one-to-one, except, of course, that $p(0)$ and $p(1)$ coincide if p is a loop. A *closed curve* is a continuous map $c : S^1 \rightarrow \mathcal{S}$, where $S^1 = \mathbb{R}/\mathbb{Z}$ is the circle; it is *simple* if it is one-to-one. An *arc* on a surface with boundary is a path that intersects the boundary precisely at its endpoints. Paths and closed curves that differ only by their parameterizations are regarded as equal.

A *drawing* of a graph G on a surface \mathcal{S} maps every vertex of G to a point of \mathcal{S} , and every edge of G to a path with the appropriate endpoints. An *embedding* of G on \mathcal{S} is a “crossing-free” drawing (vertices are mapped to pairwise distinct points, edges are mapped to simple, interior-disjoint paths, and the relative interior of each edge does not contain the image of a vertex). See Figure 2. The *faces* of an embedding of G are the connected components of the complement of its image. A graph embedding is *cellular* if every face is homeomorphic to an open disk, and a *triangulation* if every face is homeomorphic to an open disk and is bounded by three sides of edges. The *rotation system* of an embedding of G is the cyclic ordering of the edges of G incident to each vertex in the embedding. We use any of the numerous and standard data structures to represent combinatorial maps of cellular graph embeddings on orientable surfaces and move around quickly, e.g., the doubly-connected edge list, the halfedge data structure, or the gem representation [23, 34].

A *homotopy* between two paths p_0 and p_1 with the same endpoints is a continuous family of paths $(p_t)_{t \in [0, 1]}$ with the same endpoints. A (*free*) *homotopy* between two closed curves c_0 and c_1 is a continuous family of closed curves $(c_t)_{t \in [0, 1]}$; this time, no point is required to be fixed. A loop or closed curve is *contractible* if it is homotopic to a constant loop or closed curve. A *homotopy* between two drawings of the same graph is a continuous family of drawings between them; we emphasize that the vertices may move during the deformation. A drawing of a graph on a surface can be *untangled* if it is homotopic to an embedding.

Let H be a graph embedded in a surface \mathcal{S} . Consider also an abstract graph G . A *drawing* φ of G in H is a drawing of G on \mathcal{S} such that each edge of G is mapped to a walk in H (possibly reduced to a single vertex). The *size* $|\varphi|$ of φ is the number of edges of G plus the sum of the lengths of the walks $\varphi(e)$, for all edges e of G .

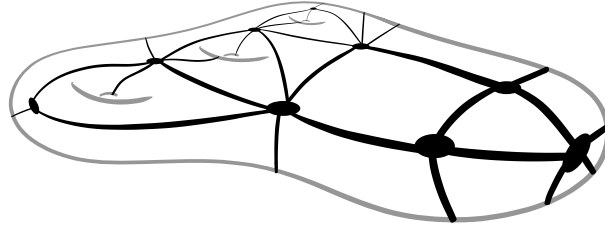


Figure 2: A graph embedded on an orientable surface of genus three without boundary.

2.2 Weak embeddings Let H be a graph embedded in a surface \mathcal{S} , and let G be an abstract graph. A drawing $\varphi : G \rightarrow H \subset \mathcal{S}$ is a **weak embedding** if there exist embeddings arbitrarily close to φ (viewed as a drawing on \mathcal{S}), or equivalently but more formally if φ is the limit of some sequence of embeddings $\psi : G \rightarrow \mathcal{S}$ in the compact-open topology. Akitaya, Fulek, and Tóth [3] provide an algorithm to decide whether such a φ is a weak embedding. We will heavily rely on this result, restated below, but we need some preparations. As noted in [3, Section 1] the property for φ to be a weak embedding does not depend on the precise embedding of H on \mathcal{S} , but only on G , φ , and the abstract graph H together with its rotation system. Akitaya et al. formulate an alternative, more combinatorial definition of weak embeddings (in terms of *strip systems*), and we present a trivially equivalent variation more suitable to our needs, the *patch system*.

Intuitively, the patch system of a graph H boils down to the dual graph of H , if H is cellularly embedded; but we cannot make this assumption in general. Formally, the patch system of the graph H (assumed here without loops or multiple edges, just for clarity of exposition, but this restriction can be dispensed of easily) equipped with its rotation system is defined as follows. Consider an oriented closed disk D_v for each vertex v of H ; consider pairwise disjoint closed segments along the boundary of D_v , one segment per edge incident to v , ordered along the boundary of D_v as prescribed by the rotation system. Now, for every edge uv of H , identify the corresponding segments of the disks D_u and D_v in a way that respects their orientations. These identifications result in a surface with boundary Σ , the **patch system** of H . The intersection of D_u and D_v is either a simple arc $a(uv)$ in Σ , if u and v are adjacent, or empty otherwise. If H is embedded in a surface \mathcal{S} and inherits its rotation system from the embedding, then its patch system Σ can be thought of as a neighborhood of H in \mathcal{S} , but its definition relies on combinatorial data only.

It follows from the considerations by Akitaya et al. [3, Section 1] that a map $\varphi : G \rightarrow H \subset \mathcal{S}$ is a weak embedding if and only if there exists an embedding $\psi : G \rightarrow \Sigma$ that satisfies the following: (1) ψ maps each vertex v of G inside D_v ; (2) ψ maps each edge e of G , traversing edges e_1, \dots, e_k of H in this order, to a path in Σ crossing the arcs exactly in the order $a(e_1), \dots, a(e_k)$. We say that such an embedding $\psi : G \rightarrow \Sigma$ **approximates** φ . Finally, here is the result that we will use:

THEOREM 2.1. (AKITAYA, FULEK, AND TÓTH [3]) *Given an abstract graph G , an embedded graph H , and a drawing $\varphi : G \rightarrow H$, we can compute an embedding approximating φ , or correctly report that no such embedding exists, in $O(|\varphi| \log |\varphi|)$ time.*

(We note that the authors require that φ be “simplicial”, but we can assume this trivially by subdividing the graph G . The running time, in their theorem $O(m \log m)$ where m is the number of edges, becomes $O(|\varphi| \log |\varphi|)$.)

2.3 Covering spaces, lifts, and a property on homotopy Here we recall basic properties of covering spaces; see, e.g., [4, Section 10.4]. A **covering space** of a surface \mathcal{S} is a (usually non-compact) surface $\hat{\mathcal{S}}$ equipped with a **projection** $\pi : \hat{\mathcal{S}} \rightarrow \mathcal{S}$ that is a local homeomorphism. The preimages of a point under π are its **lifts**. A **lift** of a path p in \mathcal{S} is a path \hat{p} on $\hat{\mathcal{S}}$ such that $\pi \circ \hat{p} = p$. Given any path p and any lift \hat{b} of $p(0)$, there is a unique **lift** \hat{p} , a path in $\hat{\mathcal{S}}$ starting at \hat{b} . A path p self-intersects if and only if either a lift of p self-intersects, or two lifts of p intersect. One can similarly lift **bi-infinite paths** $p : \mathbb{R} \rightarrow \mathcal{S}$, and even homotopies; see, e.g., [4, 10.11].

Closed curves have no basepoint, and thus typically lift to infinite paths in covering spaces. Given a closed curve $c : S^1 = \mathbb{R}/\mathbb{Z} \rightarrow \mathcal{S}$, let c' be the infinite path that “wraps around” c infinitely many times, namely $c' : \mathbb{R} \rightarrow \mathcal{S}$ is such that $c'(t) = c(t \bmod 1)$. A **lift** of the closed curve c is, by definition, a lift of the (bi-infinite) path c' .

We will mostly use universal covers: The **universal cover** $\tilde{\mathcal{S}}$ is a covering space in which every loop is contractible. A loop in \mathcal{S} is contractible if and only if it lifts to a loop in $\tilde{\mathcal{S}}$. If \mathcal{S} has positive genus and no

boundary, then \tilde{S} is homeomorphic to the open disk.

A closed curve is **primitive** if it is not homotopic to the k th power of some other closed curve for some $k \geq 2$ (a closed curve iterated k times). For future reference, we quote (see Farb and Margalit [26, Proposition 1.4], or Epstein [24, Theorem 4.2]):

LEMMA 2.1. *In an orientable surface possibly with boundary, every simple non-contractible closed curve is primitive.*

2.4 Hyperbolic surfaces and properties from Riemannian geometry We will use standard notions of hyperbolic geometry, described succinctly in Farb and Margalit [26, Chapter 1]. See also Cannon, Floyd, Kenyon, and Parry [7] for a thorough introduction to the properties of the hyperbolic plane.

One model of the **hyperbolic plane** \mathbb{H} , the Poincaré model, is the unit open disk endowed with a specific Riemannian metric of constant curvature -1 defined by $ds^2 = 4(dx^2 + dy^2)/(1 - x^2 - y^2)^2$ (intuitively, it is easier to move around at the center of the disk than close to its boundary). In this model, the maximal shortest paths are the subsets of circles (or, in the limit case, lines) touching the boundary of the disk orthogonally.

A **hyperbolic surface** is a metric surface locally isometric to the hyperbolic plane. A compact topological surface without boundary can be endowed with a hyperbolic metric if and only if its genus is at least two. The hyperbolic metric on such a surface lifts to its universal cover, which becomes isometric to the hyperbolic plane \mathbb{H} .

We need a few definitions and properties from Riemannian geometry that we will use in the realm of hyperbolic surfaces: A **crossing** between two paths is a point where they intersect in their relative interiors, and actually cross in a transverse manner. A **geodesic** is a path that is locally shortest. Given any tangent vector v at a point p , there is a unique maximal geodesic with tangent vector v at p . As a consequence, if any two geodesics intersect at p , they are either crossing at p , or they are tangent, which implies that they **overlap**: They are both part of the unique maximal geodesic passing through p with that common tangent vector.

Some non-compact hyperbolic surfaces can be constructed from **ideal** hyperbolic polygons, the sides of which are geodesics of infinite length. Pair the sides of a collection of ideal hyperbolic polygons, and identify the two sides in each pair in a way that respects the orientations of the polygons. The result is a hyperbolic surface whose topological type is that of a surface obtained by **puncturing** (removing points from) a compact surface without boundary, but punctures are relegated to infinity.

Both kinds of hyperbolic surfaces presented above enjoy the following specific properties: (1) There is a unique geodesic path homotopic to a given path; (2) there is a unique geodesic closed curve freely homotopic to a given closed curve, provided that curve is non-contractible and not homotopic to a neighborhood of a puncture. In both cases, the geodesic is the unique shortest path or closed curve in its (free) homotopy class.

3 Reducing triangulations and reduced walks

In this section, we introduce *reducing triangulations* and *reduced walks* and their properties. We start by defining these objects, then show the existence of a unique reduced (closed) walk in any given homotopy class and any non-trivial free homotopy class, and finally show how to compute them in linear time.

3.1 Reducing triangulations, turns, and reduced walks A triangulation of a surface is **reducing** if its dual is bipartite (each triangle is colored either red or blue, and adjacent triangles have different colors) and its vertices all have degree at least eight. A straightforward application of Euler's formula shows that any reducing triangulation of any surface of genus $g \geq 2$ without boundary must have size $O(g)$. It is easily seen that reducing triangulations with at most two vertices exist for every such surface (see Section 7, and in particular Figure 17). We will use reducing triangulations for these surfaces, and also for their covering spaces; indeed, note that any reducing triangulation on a surface naturally lifts to a reducing triangulation in its covering spaces, by lifting the colors.

Let W be a walk (closed or not) in a reducing triangulation T . Assume that a subwalk of W traverses the directed edge e , arrives at vertex v , and from there traverses the directed edge e' . Vertex v of W makes a **k -turn**, $k \geq 0$, if the walk of length two composed of edges e and e' , in this order, leaves exactly k triangles of T to its left in the cyclic ordering around v between e and e' . Similarly (and not exclusively), it makes a **$-k$ -turn**, $k \geq 1$, if that walk leaves exactly k triangles of T to its right. When a turn is not a 0-turn, the notation is ambiguous because it can be represented either by a positive integer or by a negative integer (whose absolute values sum up

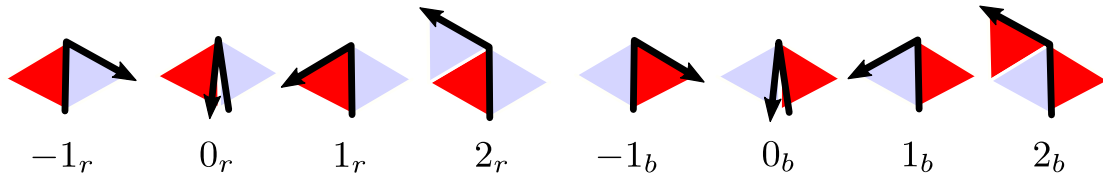


Figure 3: Some of the turns that a walk can make in a reducing triangulation.

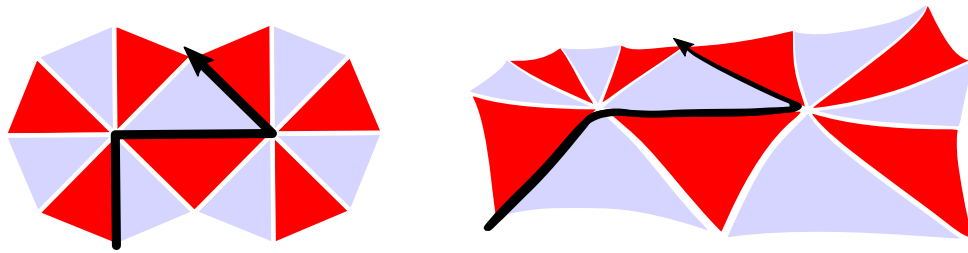


Figure 4: In a reducing triangulation, a portion of a walk that makes a -2_r -turn followed by a 1_b -turn (both are bad turns).

to the degree of the vertex in T). We always use an integer between -3 and 3 if possible, which is not ambiguous because each vertex of T has degree at least eight; otherwise, we choose the positive integer.

We sometimes need to be more precise and refine the notation: For $k \in \mathbb{Z}$, a k_b -turn, respectively a k_r -turn, is a k -turn such that e (with the previous notation) has a blue, respectively red, triangle to its left. See Figures 3 and 4.

We call **bad turn** any 0-turn, 1-turn, -1 -turn, 2_r -turn or -2_r -turn. A walk in T is **reduced** if none of its interior vertices makes a bad turn. A closed walk is **reduced** if none of its vertices makes a bad turn, not all of them make a 3_r -turn, and not all of them make a -3_b -turn. Intuitively, reduced walks are discrete geodesics in the triangulation T where all triangles are equilateral: A reduced walk leaves an angle at least π on both sides at each interior vertex, except when the vertex makes a 2_b -turn or a -2_b -turn, which corresponds to an angle of $2\pi/3$ on one side. The bipartiteness of the triangulation then “breaks ties” for determining the geodesic. We emphasize that the notion of reduced walks requires an orientation of the surface.

Here are a few immediate but crucial properties that we will use repeatedly. Any subwalk of a reduced walk (closed or not) is also reduced. The reversal of any reduced walk (closed or not) is also reduced, because reversing a walk exchanges 2_r -turns with -2_r -turns, and 3_r -turns with -3_b -turns.

The **turn sequence** of a (closed) walk is the list of turns made by the walk at its interior vertices; this is a cyclic sequence if the walk is closed. We use some straightforward notations for turn sequences: Exponents denote iterations, stars denote arbitrary nonnegative integers, and vertical bars denote “or”. For example, 23^*4 denotes a 2 followed by a nonnegative number of 3s, followed by a 4. $(23^*4)^*$ denotes a nonnegative number of concatenations of patterns of that form. $(2|4)$ denotes either a 2 or a 4.

3.2 Uniqueness of reduced walks

In this section, we prove:

PROPOSITION 3.1. *In a reducing triangulation T , any two homotopic reduced walks are equal.*

Intuitively, reduced walks are geodesics, and the proof of Proposition 3.1 goes by showing that any two distinct homotopic reduced walks would form a monogon or a bigon (a disk bounded by one or two subwalks, in the universal cover), and that this is impossible. It relies on the following two lemmas.

LEMMA 3.1. *Any walk in T whose turn sequence is of the form 23^*2 contains a 2_r -turn. Any walk in T whose turn sequence is of the form 23^*4 or 43^*2 contains a 2_r -turn or a 4_r -turn.*

Proof. By bipartiteness of the reducing triangulation T , the color of the triangle of T located to the left of the

walk changes when passing over an interior vertex that makes a 2-turn or 4-turn, and not when the interior vertex makes a 3-turn. \square

LEMMA 3.2. *Let C be a simple closed walk in T that bounds a (non-empty) disk to its left. For every $k \geq 1$, let m_k be the number of vertices of C at which C has exactly k triangles to its left. Then $2m_1 + m_2 \geq 6 + \sum_{k \geq 4} m_k$.*

The proof can be viewed as a consequence of a discrete version of the Gauss–Bonnet theorem [25, Section 2.3]:

Proof. Let D be the restriction of the reducing triangulation to the closed disk bounded by C . Consider the following discharging argument. Give initial weight 6 to each vertex and each triangle of D , and weight -6 to each edge of D . By Euler’s formula, the weights initially attributed sum up to 6. Discharge as follows. For each incidence between a vertex v and an edge e , transfer 3 from v to e . For each incidence between a vertex v and a triangle t , transfer 2 from t to v . Now, edges and triangles all have weight 0, while every vertex v has weight $\kappa(v) := 6 - 3 \deg(v) + 2 \deg'(v)$ where $\deg(v)$ and $\deg'(v)$ denote respectively the number of edge incidences and triangle incidences of v in D . We proved $6 = \sum_v \kappa(v)$.

Every vertex v that lies in the interior of D (not on C itself) satisfies $\deg(v) = \deg'(v)$ and thus $\kappa(v) = 6 - \deg(v) \leq 0$, since each vertex of T has degree at least eight. Every vertex v that lies on C satisfies $\deg(v) = \deg'(v) + 1$ and thus $\kappa(v) = 3 - \deg'(v)$. Thus, we have $6 = \sum_v \kappa(v) \leq \sum_{k \geq 1} (3 - k)m_k$, implying the result. \square

Proof of Proposition 3.1. The reducing triangulation T lifts to an infinite reducing triangulation \tilde{T} in its universal cover. Assume that there exist two distinct homotopic reduced walks in T (possibly one of them being a single vertex). Let W_1 and W_2 be lifts of these walks in \tilde{T} , with the same endpoints; they are also reduced. We claim that there exists, in \tilde{T} , a simple closed walk C with at most two vertices that make a bad turn.

Indeed, if one of W_1 and W_2 is not simple, it contains a non-empty subwalk with the same starting and ending vertex v , and otherwise not repeating any vertices; since subwalks of reduced walks are also reduced, the claim holds (only at v the walk can make a bad turn). Otherwise, W_1 and W_2 are simple, distinct, and have the same first and last vertices, which implies that they admit non-empty subwalks that have the same endpoints and are otherwise disjoint. (Indeed: up to exchanging W_1 and W_2 , some edge in W_1 does not belong to W_2 . Consider the subwalk of W_1 of minimum length containing that edge and intersecting W_2 at its endpoints. The subwalks of W_1 and W_2 with these endpoints have the desired property.) Let C be the simple closed walk that is the concatenation of these subwalks (or their reversals) W'_1 and W'_2 . The subwalks W'_1 and W'_2 are also reduced, since subwalks and reversals of reduced walks are reduced. This concludes the proof of the claim.

Without loss of generality, assume that C is directed so that the disk it bounds lies on its left. (Reversals of reduced walks are reduced.) Let S be the set of vertices of C that make a bad turn; we have $|S| \leq 2$. Using the notations of Lemma 3.2, we have $m_1 \leq |S| \leq 2$; indeed, any vertex that makes a 1-turn belongs to S .

We consider the subwalks with turning sequence of the form $2(1|3)^*2$. These subwalks may only share their first and last edges and are otherwise disjoint. Lemma 3.2 implies that $m_2 \geq (6 - 2m_1) + \sum_{k \geq 4} m_k$; thus, there are at least $6 - 2m_1$ such sequences. At most $|S| - m_1$ vertices in S make a 2-turn, and thus at most $2(|S| - m_1)$ such sequences start or end in S , because the sequences all start and end with a vertex that makes a 2-turn. There remains at least $(6 - 2m_1) - 2(|S| - m_1) = 6 - 2|S| \geq 1$ such sequences whose first and last elements are not in S . In such a sequence, exactly one of the two vertices that make a 2-turn actually makes a 2_r -turn by Lemma 3.1, which is impossible because the vertices of C that make a bad turn are all in S . \square

3.3 Uniqueness of reduced closed walks

PROPOSITION 3.2. *In a reducing triangulation T , any two freely homotopic, non-contractible, reduced closed walks are equal.*

The proof of Proposition 3.2 relies on the following lemma.

LEMMA 3.3. *Consider two disjoint simple closed walks C and C' that bound an annulus in T . Then C and C' are not both reduced as closed walks.*

Proof. We prove the claim by contradiction so assume that C and C' are both reduced as closed walks. Let A be the restriction of the reducing triangulation to the closed annulus bounded by C and C' . Let i be the number of

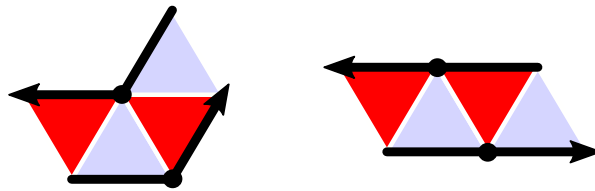


Figure 5: The cases leading to a contradiction at the end of the proof of Lemma 3.3.

vertices in the interior of A , and let B be the vertex set of the boundary of A . For every vertex v of A let $\deg(v)$ and $\deg'(v)$ be the number of incidences of v with respectively the edges and the triangles of A .

We first claim that $2i \leq \sum_{v \in B} (3 - \deg'(v))$. We prove this claim with the same discharging rules as in the proof of Lemma 3.2. Give initial weight 6 to each vertex and each triangle of A , and weight -6 to each edge of A . By Euler's formula, the weights initially attributed sum up to 0. Discharge as follows. For each incidence between a vertex v and an edge e , transfer 3 from v to e . For each incidence between a vertex v and a triangle t , transfer 2 from t to v . Now, edges and triangles all have weight 0, while each vertex v has weight $\kappa(v) := 6 - 3 \deg(v) + 2 \deg'(v)$. We proved $0 = \sum_v \kappa(v)$. Every vertex v of A satisfies $\kappa(v) = 3 - \deg'(v)$ if $v \in B$ and $\kappa(v) = 6 - \deg(v) \leq -2$ otherwise. Therefore, $0 \leq \sum_{v \in B} (3 - \deg'(v)) - 2i$, proving the claim.

We orient C and C' so that A lies to their left. Let $C_1 \in \{C, C'\}$, and let $B_1 \subset B$ be the set of vertices of C_1 . We remark that, because C_1 has no bad turn, every $v \in B_1$ satisfies $\deg'(v) \geq 2$. Moreover, because C_1 is reduced as closed walk, by Lemma 3.1, its turn sequence does not contain 23^*2 as a subword, and it is not of the form 23^* . Thus, $\sum_{v \in B_1} (3 - \deg'(v)) \leq 0$, with equality if and only if the turn sequence of C_1 has the form $(23^*43^*)^*$ or 3^* .

From the above claim and the conclusion of the previous paragraph, we deduce that A has no interior vertices and that each of the turn sequences of C and C' is of the form $(23^*43^*)^*$ or 3^* , which must actually be $(2_b 3^* 4_r 3^*)^*$ (by Lemma 3.1 and because there is no 2_r) or 3_b^* (by definition of a reduced closed walk).

Assume that the turn sequence of one boundary component of A has the form $(2_b 3^* 4_r 3^*)^*$. Let v be a vertex of B that makes a 2_b -turn. Then, the vertex across v in the other boundary component of A makes a 4_b -turn (Figure 5, left), which is a contradiction since no 4_b appears in the allowed forms above. So the turn sequences of C and C' are both of the form 3_b^* ; however, for a similar reason as above, if v is a vertex of B that makes a 3_b -turn, the vertices across v make a 3_r -turn (Figure 5, right), a contradiction. \square

Proof of Proposition 3.2. Consider a reducing triangulation T of a surface \mathcal{S} and two freely homotopic reduced closed walks C and C' in T .

We consider the *annular cover* $\hat{\mathcal{S}}$ of \mathcal{S} defined by C [16, Section 1.1]. This is a covering space homeomorphic to an open annulus, in which C lifts to a closed curve \hat{C} , and every simple closed curve is either contractible or homotopic to \hat{C} or its reverse. Lifting the homotopy from C to C' yields a lift \hat{C}' that is homotopic to \hat{C} . Both \hat{C} and \hat{C}' are reduced closed walks in \hat{T} , the lift of the reducing triangulation T .

We claim that \hat{C} (and, for the same reasons, also \hat{C}') is simple. Indeed, lifting \hat{C} to the universal cover $\tilde{\mathcal{S}}$ of \mathcal{S} yields a lift, which is a bi-infinite path \tilde{P} , and is actually reduced in the reducing triangulation \tilde{T} obtained from lifting T . By Proposition 3.1, \tilde{P} is simple. Since it is the *only* lift of \hat{C} , this implies that \hat{C} is simple.

Now there are two cases. If \hat{C} and \hat{C}' are disjoint, Lemma 3.3 implies a contradiction. If they intersect, at vertex v say, then we transform \hat{C} and \hat{C}' into loops with starting and ending vertex v ; these two loops have the same turning number, and are thus homotopic reduced walks, and equal by Proposition 3.1. Thus \hat{C} and \hat{C}' are equal, and so are C and C' by projection. \square

3.4 Reducing a walk

PROPOSITION 3.3. *Given a walk W of length n in a reducing triangulation T , we can compute a reduced walk homotopic to W in $O(n)$ time.*

The techniques of the proof of Proposition 3.3 will be reused in Section 6.1. In particular, we need a few data structures here, which will be extended later.

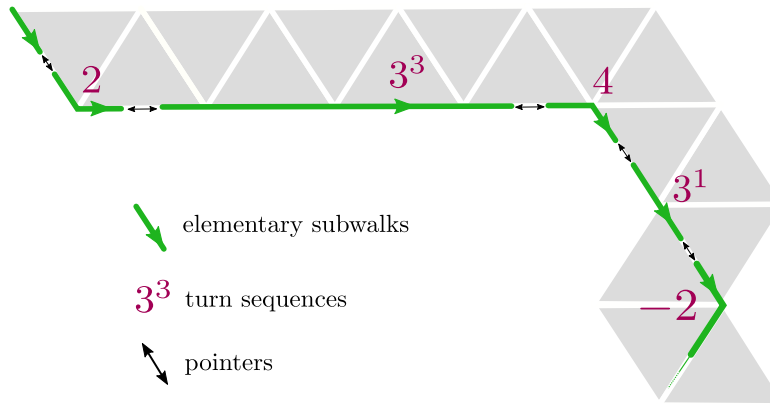


Figure 6: The compressed homotopy sequence data structure.

Our reducing triangulation T is stored as an embedded graph, but we need additional information. Specifically, for each vertex v of T , we number the directed edges with source v in clockwise order (starting at an arbitrary directed edge). This number is stored on each directed edge. Moreover, vertex v contains an array A_v of pointers such that $A_v[i]$ is the directed edge number i with source v . This array has size the degree of v , which is also stored in v . We note that this additional data can be computed in time linear in the size of the reducing triangulation, so we assume that the reducing triangulation incorporates this information. With this data structure, we can perform the following operations in constant time: (1) given two directed edges uv and vw , compute the integer i such that uvw forms an i -turn; (2) Given a directed edge uv and an integer i , compute the directed edge vw such that the walk uvw forms an i -turn. Again, this can be done as a preprocessing step in time linear in the size of the reducing triangulation.

We also need a data structure, called *compressed homotopy sequence*, that represents walks in a compact form, in the same spirit as Erickson and Whittlesey [25, Section 4.1] use run-length encoding to encode turn sequences. An *elementary subwalk* of W is an inclusionwise maximal subwalk of W whose turn sequence has the form 3^k or $(-3)^k$ for some $k \geq 1$, or is a single symbol $a \notin \{-3, 3\}$. The elementary subwalks of W are naturally ordered along W and cover W , with overlaps because the last edge of an elementary subwalk appears as the first edge of the following one. We represent W by storing the elementary subwalks of W in a doubly linked list in order along W ; each elementary subwalk is described by the images in T of its first and last directed edges, as well as its turn sequence in compressed form, namely, whether it is of the form 3^k , $(-3)^k$, or a , and the integer k or the symbol a . (Walks W of length zero or one cannot be encoded in this data structure, but can be handled separately in a trivial manner.)

Proof of Proposition 3.3. The *extended turn sequence* of a walk W is its usual turn sequence, with a new symbol E appended before its first symbol and after its last letter. We consider the *reduction moves* depicted in Figure 7 that modify a walk W by homotopy. Each of them can be applied when a specific subsequence appears in the extended turn sequence of W or of its reversal:

- *spur move*: 0 ,
- *spike move*: 1 ,
- *bracket move*: 23^k2 for some $k \geq 0$,
- *flip move*: $a3^k2, 3^\ell b$ for some $k, \ell \geq 0$ and some symbols $a, b \notin \{-1, 0, 1, 2, 3\}$. (In particular, a and/or b can be equal to E .)

We observe that if no reduction move can be applied to W , then W is reduced. Moreover, for each walk W , let $\varphi(W)$ be equal to three times the length of W , plus the number of bad turns of W . We observe that $\varphi(W)$ strictly decreases at each move; indeed, a spur, spike, or bracket move decreases the length of W (by at least one)

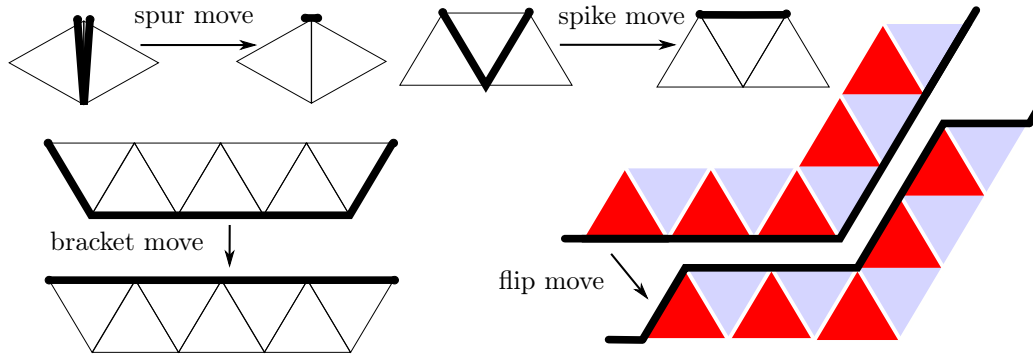


Figure 7: The reduction moves for the reduction of walks.

and creates at most two bad turns, while a flip move does not affect the length and removes at least one bad turn (no bad turn is created because the degree of each vertex is at least eight, and by the conditions on a and b). Thus, any sequence of moves has length $O(n)$.

The algorithm is as follows. In a first step, we walk along W to initialize its compressed homotopy sequence in $O(n)$ time (using Operation (1) above). Then, in a second step, we traverse this sequence in order, and reduce W as soon as a possible move is encountered, maintaining the compressed homotopy sequence as it evolves. Each possible move involves at most six consecutive elementary subwalks, which are replaced by at most seven elementary subwalks. Then, before resuming the algorithm and looking for a possible move, we need to backtrack by at most five elementary subwalks in the compressed homotopy sequence, in order not to skip any possible move created by performing the previous one. Finally, in a third step, when we reach the end of the list, no move is possible any more; we convert our compressed homotopy sequence back to a walk on T (using Operation (2) above). All of this takes $O(n)$ time because the sequence of moves has length $O(n)$, and because the length of a walk does not increase when reducing it, so that the compressed homotopy sequence always has length $O(n)$. \square

3.5 Reducing a closed walk

PROPOSITION 3.4. *Given a closed walk C of length n in a reducing triangulation T , we can compute a reduced closed walk freely homotopic to C in $O(n)$ time.*

Proof. We first remove the bad turns from C . For this purpose, we use the same reduction moves as in the proof of Proposition 3.3, but the turn sequence is now cyclic. We observe that, in each move, the part of C that is removed corresponds to an actual subwalk of C (and does not “wrap around” it). In particular, in a bracket move, the entire cyclic turn sequence of C cannot be 23^k , because the color of the triangle to the left of the walk changes after passing a 2. As a limit case, in a flip move, the symbols a and b in the sequence $a3^k2_r3^\ell b$ may correspond to the same vertex of the cyclic walk; see Figure 8, left for an illustration in the case where the turn at $a = b$ is a 4-turn.

As in the proof of Proposition 3.3, if no move is possible, then C has no bad turn. If $\varphi(C)$ is equal to three times the length of C , plus the number of bad turns of C , then $\varphi(C)$ decreases at each move, except in the very special “heart-like” case of Figure 8, corresponding to the case where the cyclic turn sequence of C (or its reversal) is $43^k2_r3^\ell$ for some $k, \ell \geq 0$. The flip move does not affect the length and removes the bad 2-turn, but both edges of C incident to the 4-turn are rotated by one triangle, which creates another bad 2-turn, see Figure 8. However, after this move, the only possible move flips the new bad 2-turn into a good one, at which point the closed walk is reduced. Thus, as in the proof of Proposition 3.3, any sequence of moves has length $O(n)$.

Finally, if the turn sequence of C is of the form $(3_r)^*$ or $(-3_b)^*$, then C can be reduced in $O(1)$ time using the new move depicted in Figure 9. \square

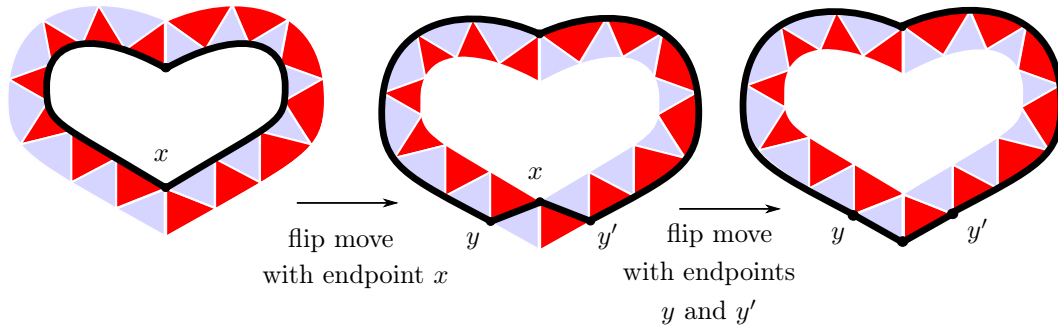


Figure 8: The “heart-like” situation in the proof of Proposition 3.4.

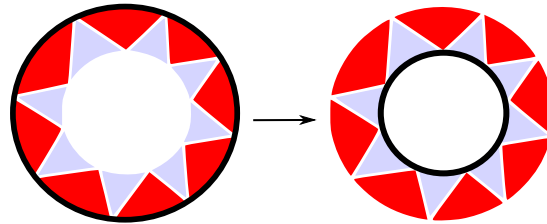


Figure 9: The last reduction move in the proof of Proposition 3.4.

4 Untangling a loop graph

A **loop graph** is a graph L whose connected components have a single vertex. A drawing λ of a loop graph L in a surface \mathcal{S} is **sparse** if, under λ , the edges of each connected component of L are non-contractible and pairwise non-homotopic. In this section, we prove:

PROPOSITION 4.1. *Consider a sparse drawing λ of a loop graph L in a reducing triangulation T of a surface of genus at least two without boundary. We can determine if λ can be untangled, and if so compute the rotation system of L in some embedding homotopic to λ , in time $O(|\lambda| \log |\lambda|)$.*

(Actually, if λ can be untangled, the rotation system turns out to be the same for all homotopic embeddings; see Corollary 5.1 below. Reducing triangulations exist for every surface of genus at least two; see Figure 17.) Throughout this section, \mathcal{S} has genus at least two and is without boundary.

4.1 Relation to weak embeddability of straightened loop graphs For the needs of this section, we say that a drawing λ of a loop graph L in a reducing triangulation T is **straightened** if it is sparse and satisfies each of the following on every connected component L_0 of L . Firstly, there is an edge e of L_0 that is mapped by λ to a reduced **closed** walk in T , where e is seen as a (non-directed) closed curve. (This is consistent because reduced (closed) walks are stable upon reversal.) Secondly, every other edge e' of L_0 is mapped by λ to a reduced walk in T , where e' is seen as a (non-directed) path this time. We say that e is a **major edge** and that e' is a **minor edge** of L (with respect to λ). (A similar idea has been used several times in the continuous setting [20, Section 5].)

LEMMA 4.1. *Consider a sparse drawing λ of a loop graph L in a reducing triangulation T of \mathcal{S} . We can compute a straightened drawing λ' homotopic to λ in time $O(|\lambda|)$.*

Proof. The algorithm works independently on the connected components of L , so assume that L is connected and let v be the vertex of L . We choose the major edge e of L as an edge whose image $\lambda(e)$ has minimal length $m \geq 1$. We reduce the closed walk $\lambda(e)$ with Proposition 3.4 to construct $\lambda'(e)$ in time $O(m)$. While reducing the closed walk $\lambda(e)$, we update homotopically the vertex $\lambda(v)$ so that $\lambda(v)$ remains on $\lambda(e)$ at all times, and we record the walk W performed by $\lambda(v)$ in T . In the end, the walk W has length $O(m)$. For every other (minor) edge e' of L , we concatenate the reverse of W with $\lambda(e')$ and then with W , and we reduce this walk with Proposition 3.3 to construct $\lambda'(e')$ in time linear in the length of $\lambda(e')$. \square

Here is the main technical ingredient of this section:

PROPOSITION 4.2. *Consider a straightened drawing λ of a loop graph L in a reducing triangulation T of a surface without boundary of genus at least two. If there is an embedding homotopic to λ , then λ is a weak embedding.*

This proposition implies immediately Proposition 4.1:

Proof of Proposition 4.1, assuming Proposition 4.2. Apply Lemma 4.1 to compute in $O(n)$ time a straightened drawing λ' of L homotopic to λ . Apply Theorem 2.1 to determine in $O(n \log n)$ time whether λ' is a weak embedding. If not, then conclude by Proposition 4.2 that there is no embedding homotopic to λ . Otherwise Theorem 2.1 produces an embedding homotopic to λ , from which we may compute the desired rotation system. \square

The rest of this section is devoted to the proof of Proposition 4.2. We need to start with some preliminaries.

4.2 The limit points of lifts in the universal cover Because \mathcal{S} has genus at least two and has no boundary, the universal cover $\tilde{\mathcal{S}}$ of \mathcal{S} can be compactified into a topological space $\tilde{\mathcal{S}} \cup \partial\tilde{\mathcal{S}}$, by adding a set $\partial\tilde{\mathcal{S}}$ of **limit points**, such that the compactified space is homeomorphic to the closed disk, and under this homeomorphism $\tilde{\mathcal{S}}$ is represented by the open disk and $\partial\tilde{\mathcal{S}}$ is represented by the circle. Moreover, every lift $\tilde{c} : \mathbb{R} \rightarrow \tilde{\mathcal{S}}$ of every non-contractible closed curve on \mathcal{S} admits well defined limit points at $+\infty$ and $-\infty$ in $\partial\tilde{\mathcal{S}}$. We will rely on the following topological properties.

LEMMA 4.2. *If $\tilde{c} : \mathbb{R} \rightarrow \tilde{\mathcal{S}}$ is a lift of a non-contractible closed curve on \mathcal{S} , then $\lim_{+\infty} \tilde{c}$ and $\lim_{-\infty} \tilde{c}$ exist (in $\tilde{\mathcal{S}} \cup \partial\tilde{\mathcal{S}}$) and are distinct points of $\partial\tilde{\mathcal{S}}$.*

LEMMA 4.3. *Lift a homotopy $c \simeq d$ between non-contractible closed curves on \mathcal{S} to a homotopy $\tilde{c} \simeq \tilde{d}$ between lifts of c and d . Then \tilde{c} and \tilde{d} have the same limit points.*

LEMMA 4.4. *Consider lifts \tilde{c} and \tilde{d} of non-contractible closed curves on \mathcal{S} . Assume either that \tilde{c} and \tilde{d} intersect exactly once, or that they are disjoint lifts of the same curve on \mathcal{S} . Then the four limit points of \tilde{c} and \tilde{d} are pairwise distinct.*

LEMMA 4.5. *Let c and d be two non-contractible closed curves on \mathcal{S} . If c and d admit lifts with the same pairs of limit points, there is a closed curve e such that c and d are homotopic to powers of e .*

For the sake of concision, we defer the (classical) construction of the limit points and the proofs of the above lemmas to Appendix A. Interestingly, although the properties of the limit points are expressed purely topologically above, they all follows from folklore arguments in hyperbolic geometry: The limit points are defined by endowing \mathcal{S} with a hyperbolic metric and lifting this metric to its universal cover $\tilde{\mathcal{S}}$; then $\tilde{\mathcal{S}}$ is isometric to the hyperbolic plane, which can be compactified as described in Appendix A.1 (by adding the boundary of the open disk in the Poincaré model), the limit points are naturally defined, and the properties follow. However, we stress that this hyperbolic structure is only used for the proofs in the appendix. We will need a different hyperbolic metric in the coming subsections.

4.3 Patch system and hyperbolic metric on the punctured surface To prove Proposition 4.2, we assume that there is an embedding homotopic to λ in \mathcal{S} , and we will prove that λ is a weak embedding. Recall that since the reducing triangulation T is a graph cellularly embedded on \mathcal{S} , its patch system corresponds to its dual. More precisely, to build (the interior of) the patch system Σ of T , we remove (puncture) a point p_t from the interior of each triangle t of T ; then, for each edge e of T , we consider the two distinct triangles t, t' incident to e in T , and we draw a simple arc between p_t and $p_{t'}$ that crosses e exactly once and does not intersect any other edge or any vertex of T . We assume without loss of generality that the resulting arcs are pairwise disjoint. From now on we see the patch system Σ as included in \mathcal{S} in this way. (Compared to Section 2.2, the patch system is open, not closed, but this does not change anything.) Now recall from Section 2.2 the notion of map $L \rightarrow \Sigma$ that

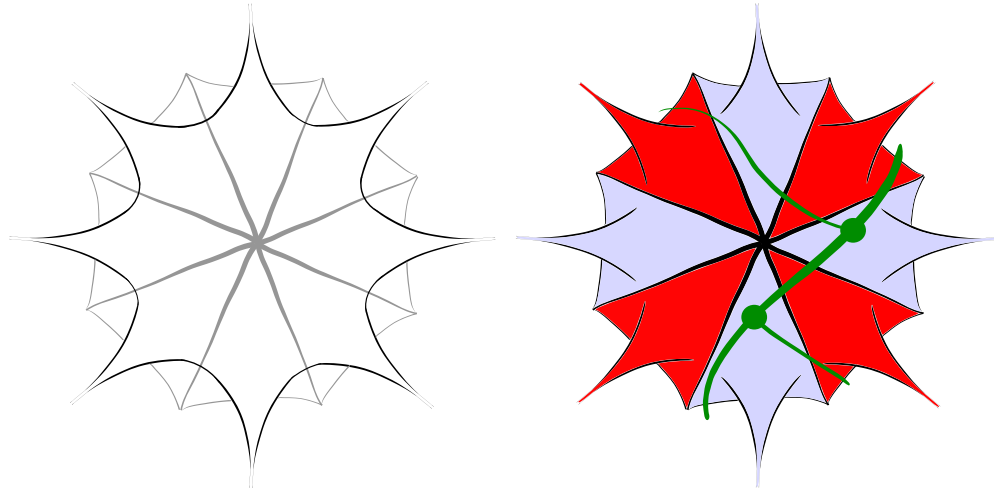


Figure 10: The map μ (in green) in the hyperbolic patch system Σ of the reducing triangulation T , in the proof of Proposition 4.2.

approximates λ . We will show that λ is a weak embedding by exhibiting an embedding $L \rightarrow \Sigma$ that approximates λ . We construct our candidate $\mu : L \rightarrow \Sigma$ for an embedding that approximates λ as follows.

Firstly, we endow Σ with a complete hyperbolic metric for which the arcs of Σ are geodesics, as follows. The arcs of Σ separate Σ into open disks. Each resulting connected component P has $k \geq 8$ incidences with arcs since the vertices of T have degree at least eight. We replace P by a hyperbolic polygon with k geodesic sides, with the particularity that these k geodesic sides all have *infinite* length (such polygons are called ideal polygons, see Section 2.4). Then we identify the sides of the polygons by pairs corresponding to the arcs of Σ .

Secondly, we see λ as a map from L to Σ and we construct μ by modifying λ homotopically on every connected component L_0 of L , as follows. Let v be the vertex of L_0 and let e be the major edge of L_0 . We start by replacing homotopically the image of e by the geodesic closed curve c_e in its free homotopy class, thus possibly changing the image of v . Here we make use of the fact that $\lambda(e)$ is neither contractible nor homotopic to the neighborhood of a puncture in Σ , since it is non-contractible in \mathcal{S} . We require that during this homotopy $\lambda \simeq \mu$ the image of v remains in the same disk of Σ (never belongs to an arc of Σ). We can do so since the reduced closed walk $\lambda(e)$ makes no 0-turn and thus has the same sequence of crossings with the arcs of Σ as the geodesic closed curve c_e . Then, we replace for each minor edge e' of L_0 the image of e' by the geodesic segment in its homotopy class, this time holding the image of v fixed. The map μ approximates λ since the images of the vertices of L remained in the same disks of Σ and since for every edge e'' of L the walk $\lambda(e'')$ makes no 0-turn.

Finally, we may ensure without loss of generality that the vertices of L are mapped by μ to distinct points of Σ , and belong to no arc of Σ . That can be achieved by moving infinitesimally the image of each vertex of L along the image of its major edge.

At this point μ is not necessarily an embedding; indeed, several major edges can overlap. The strategy is to prove that this is the only reason why μ may fail to be an embedding, and that such overlaps can be eliminated by a small perturbation of μ .

4.4 Proof of Proposition 4.2 In this section, we conclude the proof of Proposition 4.2. To ease the reading, we abuse the notation and write L to refer to the map μ on L (defined in the previous section).

Recall from Section 4.2 that the universal cover $\tilde{\mathcal{S}}$ of \mathcal{S} is homeomorphic to an open disk, which can be compactified into a closed disk $\tilde{\mathcal{S}} \cup \partial\tilde{\mathcal{S}}$, so that the properties of Section 4.2 can be applied. Note that, here, we view $\tilde{\mathcal{S}} \cup \partial\tilde{\mathcal{S}}$ as a purely topological space, and completely ignore any metric on it. In this section, we consider the subset $\tilde{\Sigma}$ of $\tilde{\mathcal{S}}$ made of the points that are lifts of Σ ; in other words, $\tilde{\Sigma}$ is obtained from $\tilde{\mathcal{S}}$ by removing the lifts of the punctures. (It turns out that $\tilde{\Sigma}$ is a covering space of Σ , but we will never use this property.) $\tilde{\Sigma}$ naturally inherits a hyperbolic metric, obtained by lifting the hyperbolic metric of Σ .

By construction, lifts of edges of L are geodesics under the metric of $\tilde{\Sigma}$. A *major lift* is a lift of a major

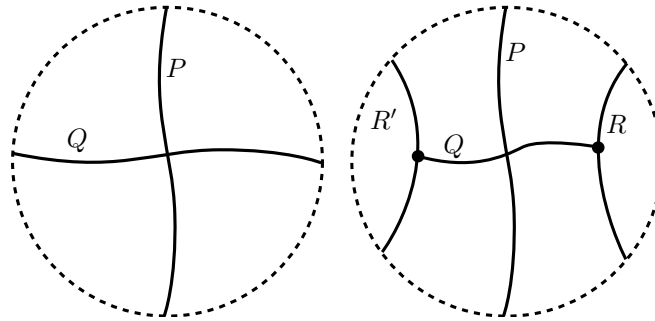


Figure 11: (Left) Proof of Lemma 4.7. (Right) Proof of Lemma 4.9.

edge; it is a bi-infinite geodesic path in $\tilde{\Sigma}$ with two distinct limit points on $\partial\tilde{\mathcal{S}}$, by Lemma 4.2 and sparsity. A **minor lift** is a lift of a minor edge; it is a geodesic path in $\tilde{\Sigma}$.

If P and Q are minor lifts of distinct minor edges, then they may share one of their endpoints, but not both, since L is sparse and since the vertices of L are distinct points of Σ . In particular, the minor lifts of distinct minor edges are distinct. However, the situation is slightly more complicated for major lifts. Two major lifts obtained from different major edges are considered different. Recall that major edges are primitive (by Lemma 2.1 and sparsity). So any two major lifts $\tilde{c}, \tilde{d} : \mathbb{R} \rightarrow \tilde{\Sigma}$ of the same major edge that differ only by a homeomorphism of \mathbb{R} actually differ by an integer translation; in that case, we see them as equal.

The proof of Proposition 4.2 combines a long series of relatively easy lemmas on the lifts of the edges of L in $\tilde{\Sigma}$. Recall that by assumption L can be untangled.

LEMMA 4.6. *Every lift of an edge is simple. Any two lifts of edges (possibly the same edge) that do not overlap intersect at most once, and if they intersect in their relative interiors, then they cross.*

Proof. If two portions of geodesic paths intersect in their relative interiors without overlapping, then they cross. Now assume, for a contradiction, that a lift is not simple; it contains a loop, which projects to a geodesic loop ℓ in Σ that is contractible in \mathcal{S} . The sequence of crossings of ℓ with the arcs of Σ is that of a reduced walk. Thus, by Proposition 3.3, the loop ℓ does not cross any arc of Σ , and so ℓ is contractible in Σ , which is impossible since ℓ is geodesic.

Similarly, if two lifts intersect twice without overlapping, they form two paths with the same endpoints and otherwise disjoint, which project to geodesic paths p and q in Σ that are homotopic in \mathcal{S} . The sequence of crossings of p and q with the arcs of Σ must be the same by Proposition 3.3, so p and q are homotopic in Σ , which is impossible since p and q are geodesics. \square

LEMMA 4.7. *Any two major lifts either have the same image or are disjoint.*

Proof. Assume that two major lifts P and Q intersect and do not have the same image. They cross exactly once by Lemma 4.6. But then, their limit points are pairwise distinct by Lemma 4.4, and interleaved on $\partial\tilde{\mathcal{S}}$ (see Figure 11 (left)), which implies that L cannot be untangled by Lemma 4.3. \square

LEMMA 4.8. *Let P be a major lift and Q a minor lift such that P and Q are lifts of edges in the same connected component of L , and such that an endpoint of Q belongs to the image of P . Then no other point of Q lies on the image of P .*

Proof. Otherwise, P and Q overlap by Lemma 4.6. Let P' be the part of P that starts and ends at the endpoints of Q (which are both lifts of the vertex of the corresponding connected component L_0 of L). By Lemma 2.1, the projection of P' is the major edge of L_0 . Moreover, the projection of Q is a minor edge of L_0 . This contradicts the sparsity of L . \square

Before treating the other cases, we need a definition. Let P be a minor lift and let v and v' be the endpoints of P (they are distinct). Let p be the minor edge of L that is the projection of P , and let q be the major edge

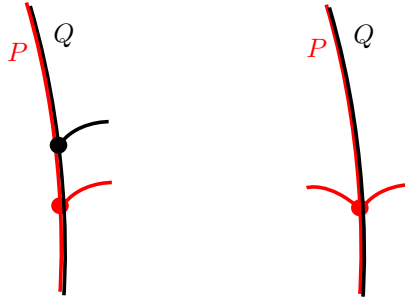


Figure 12: The two cases forbidden by Lemma 4.10.

of L that lies in the same connected component of L as p . Let Q and Q' be the major lifts, starting at v and v' respectively, that are lifts of q . We say that Q and Q' form the **H-block** of P . Indeed, Q , P , and Q' together form the shape of the embedded letter “H”, because they touch only at v and v' by the preceding lemmas, and moreover the four limit points of Q and Q' are pairwise distinct by Lemma 4.4.

LEMMA 4.9. *There is no intersection between a major lift P and the relative interior of a minor lift Q .*

Proof. Let R and R' be the major lifts forming the H-block of Q . Assume first that there is an intersection between P and the relative interior of Q that is not a crossing. Then P and Q overlap by Lemma 4.6. In particular, P intersects R and R' . However R and R' are disjoint, being part of the same H-block. This contradicts Lemma 4.7.

Now assume that there is an intersection between P and the relative interior of Q that is a crossing. By Lemma 4.6, there is exactly one. By Lemma 4.7, and since P does not have the same image as R or R' , the major lifts P , R , and R' are disjoint. See Figure 11 (right). Also, P does not have the same pair of limit points as R (or R'): for otherwise, by Lemma 4.5 and Lemma 2.1, P and R would project to closed curves freely homotopic in S (up to reversal), and thus freely homotopic in Σ by Proposition 3.2, but then P and R would project to the same geodesic closed curve, which is impossible since they are disjoint. Any homotopy of L induces a homotopy of the lifts P , R , and R' , preserving the limit points on $\partial\tilde{S}$. Thus, even after a homotopy, one crossing between P and Q , R , or R' must remain (even if P shares one endpoint with R and/or one endpoint with R'). This contradicts the assumption that L can be untangled. \square

Let v be a vertex of L , and let e be a directed major edge of L based at v . We say that e is **pulled to the right** if there exists a minor edge based at v that leaves v and/or arrives at v to the right of e ; in other words, in counterclockwise order around v , we do not see consecutively the target of e and the source of e . We say that a directed major lift is **pulled to the right** if its projection is.

LEMMA 4.10. *Let P and Q be two major lifts, projecting to different edges of L . Assume that they overlap; assume (up to reversing one of them) that they have the same direction. Then (1) at most one of P and Q is pulled to the right, and (2) none of them is pulled both to the left and to the right.*

Proof. We first prove (1); see Figure 12, left. Assume, for the sake of a contradiction, that both P and Q are pulled to the right. Let \bar{P} and \bar{Q} be these lifts after a homotopy that untangles L . (Generally, we use bars to denote lifts after this homotopy.) They must be disjoint (except at their limit points), so without loss of generality, up to exchanging P and Q , assume that \bar{P} lies to the left of \bar{Q} . Let R be the minor lift pulling P to the right; thus, P is part of the H-block of R ; let P' be the lift that forms the H-block of R together with P . Then, \bar{Q} intersects either \bar{R} , or \bar{P}' (twice), which is impossible.

We now prove (2); see Figure 12, right. Assume that P is pulled both to the left and to the right. After the homotopy, the lifts \bar{P} and \bar{Q} are disjoint, so without loss of generality, up to reversing the orientations of P and Q , assume that \bar{P} lies to the left of \bar{Q} . We are then in the same situation as the previous paragraph, and conclude similarly. \square

LEMMA 4.11. *The relative interiors of any two distinct minor lifts are disjoint.*

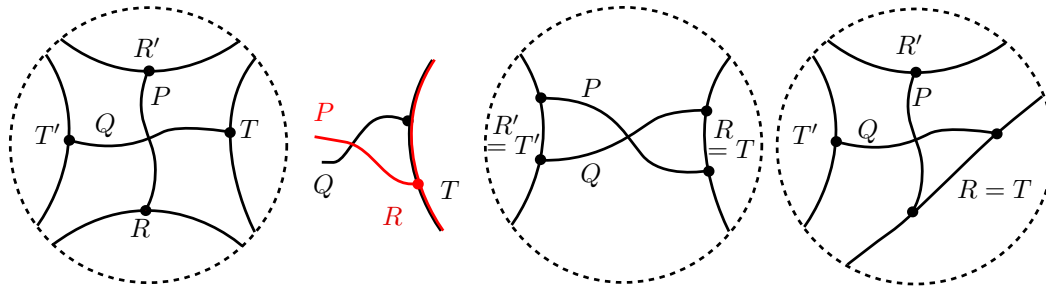


Figure 13: The four cases in the proof of Lemma 4.11.

Proof. If the relative interiors of two distinct minor lifts P and Q intersect without crossing, they overlap. Since P and Q are distinct they do not have the same pair of endpoints. Thus the relative interior of one of P and Q must intersect a major lift, a contradiction with Lemma 4.9.

If P and Q cross, they do so exactly once by Lemma 4.6. Let R and R' be the major lifts forming the H-block of P , and similarly let T and T' be the major lifts forming the H-block of Q . By Lemma 4.7, and up to exchanging notations, we distinguish between four cases, depicted in Figure 13:

- If $R, R', T,$ and T' are pairwise disjoint, the cyclic ordering of the limit points on $\partial\tilde{\mathcal{S}}$ is necessarily $r_1r_2t_1t_2r'_1r'_2t'_1t'_2$, with obvious notations (R has limit points r_1 and r_2 , and so on), with possible identifications of two consecutive limit points in this cyclic order if they come from different lifts. Any homotopy of L induces a homotopy of the lifts $P, Q, R, R', T,$ and T' , preserving the limit points on $\partial\tilde{\mathcal{S}}$. Thus, even after a homotopy, one crossing must remain. This contradicts the assumption that L can be untangled.
- If R is distinct from T but has the same image as T , these two lifts with the same image, when directed in the same way, are both pulled to the right, or both pulled to the left, which is impossible (Lemma 4.10(1)).
- If R is equal to T and R' is equal to T' , then any homotopy of L induces a homotopy of $P, Q, R, R',$ preserving the four limit points on $\partial\tilde{\mathcal{S}}$ and the relative orders of the endpoints of P and Q on R and R' . Thus, even after a homotopy, one crossing must remain.
- If R is equal to T and R' is disjoint from T' , then, again, a crossing must remain after lifting a homotopy of L . \square

Proof of Proposition 4.2. By Lemmas 4.6, 4.9, and 4.11, the only reason why L may fail to be an embedding is because two distinct major lifts intersect, which implies that they have the same image by Lemma 4.7, and thus come from distinct major edges.

So now, consider an inclusionwise maximal set A of at least two overlapping, simple, major edges, directed in the same way. By Lemma 4.10 A contains at most one edge pulled to the left (and not to the right), at most one edge pulled to the right (and not to the left), and possibly several edges which are pulled neither to the left nor to the right. We can slightly perturb these edges to make them disjoint: from left to right, the edge pulled to the left (if it exists), then, the edges not pulled at all, and finally, the edge pulled to the right (if it exists). After this operation we have an embedding of L that approximates λ . \square

5 Untangling on reducing triangulations: Proof of Theorem 1.3

In this section, we prove Theorem 1.3, by providing an algorithm to untangle a graph on a reducing triangulation. We already know how to do this for sparse loop graphs, so at a very high level, our strategy is to transform the input graph G into a sparse loop graph; but this hides many technicalities, both at the conceptual and at the computational level. We first compute a *graph factorization*, described in the next subsection. Then we need some auxiliary lemmas, mostly relating the possibility of untangling a graph to that of the corresponding sparse loop graph. Finally, we conclude the proof of Theorem 1.3, with a weaker complexity than announced; this is fixed in Section 6, which contains a better algorithm for computing a graph factorization.

5.1 Graph factorization To transform an arbitrary graph G into a sparse loop graph, the natural idea is to contract a spanning forest of G , to ignore the resulting contractible loops, and to identify the resulting homotopic loops. In this section, we describe this operation formally and provide a first, rather naïve, algorithm to compute it. Unless noted otherwise, \mathcal{S} is an arbitrary surface, with or without boundary.

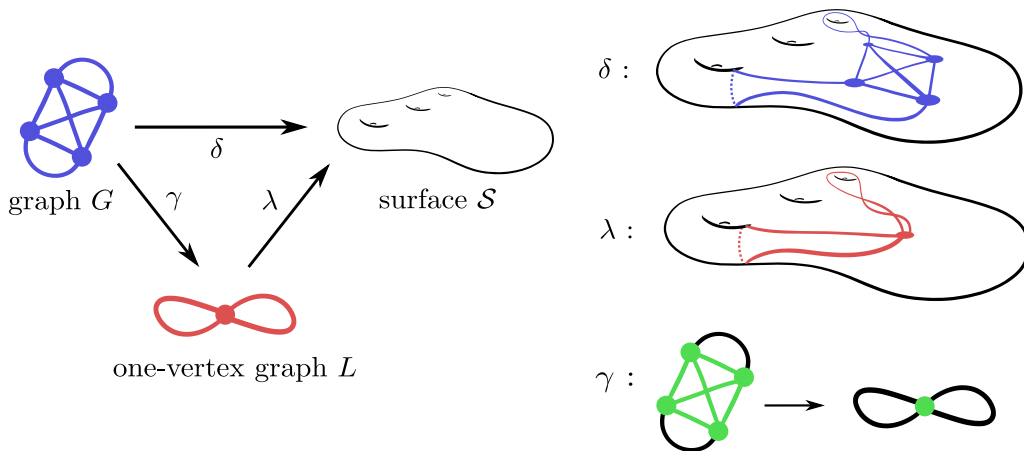


Figure 14: A factorization (L, λ, γ) of the drawing (G, δ) of Figure 1.

We need to carefully distinguish an abstract graph from its drawings (or embeddings). Let G_1 be a connected component of G and let δ be the input drawing of G_1 on \mathcal{S} ; moreover, let Y be a spanning tree of G_1 . A **factorization** of (G_1, δ) (with respect to Y) is obtained by the following process (see Figure 14). First, change the drawing δ of G_1 by contracting homotopically the spanning tree Y to a single point p . Let δ' be the new drawing of G_1 on \mathcal{S} . The non-tree edges of G_1 are now drawn as loops under δ' . Move these loops that are contractible to the constant loop based at p . Now, whenever there are several loops of G_1 in the same homotopy class (with basepoint p), select one of these loops, ℓ , arbitrarily, and redraw the other ones in the same way as ℓ . Let δ'' be the new drawing of G_1 on \mathcal{S} . We remark that, under δ'' , the graph G_1 is first drawn onto a one-vertex graph L , which is itself sparsely drawn in \mathcal{S} . The factorization of (G_1, δ) is given by the one-vertex graph L , its sparse drawing λ on \mathcal{S} , and the drawing γ of G_1 onto L . By construction, δ and $\lambda \circ \gamma$ are homotopic; see Figure 14, left.

Equivalently, but we will not need this equivalence, a factorization of (G_1, δ) (with respect to Y) is given by:

- a one-vertex graph L , and a *sparse* drawing λ of L on \mathcal{S} ,
- a drawing γ of G_1 *onto* L , mapping each edge of Y to the vertex of L , and each edge of $G_1 \setminus Y$ to either the vertex of L or a loop of L ,

such that the drawings δ and $\lambda \circ \gamma$ of G_1 on \mathcal{S} are homotopic.

A **factorization** of a drawing δ of a disconnected graph G is given by a factorization for each of the connected components of G . The **size** of a factorization (L, λ, γ) of (G, δ) is the sum of the sizes of λ and γ .

A crucial tool is the following proposition.

PROPOSITION 5.1. *Under the hypotheses of Theorem 1.3, we can, in $O(gn \log(gn))$ time, compute a factorization of (G, δ) of size $O(gn)$, or correctly report that δ cannot be untangled on \mathcal{S} .*

We first give a brute force algorithm to prove this proposition with a worse running time, namely $O(n^3)$, and output size, namely $O(n^2)$, in order to illustrate the general strategy. While this can be improved using simple arguments, to achieve the claimed running time of $O(gn \log(gn))$ requires several algorithmic ingredients. Most notably, we need a new data structure, possibly of independent interest, to store the contracted loops of G in a compressed form. We defer the full proof of Proposition 5.1 to Section 6.

Proof of Proposition 5.1, with a worse running time and output size. Without loss of generality, we may assume that G is connected. We take an arbitrary vertex v of an arbitrary spanning tree Y of G . We then contract Y

to v , thus obtaining a drawing δ' of G in which all edges of Y are mapped to v . The size of δ' is $O(n^2)$, because every non-tree edge is now drawn as a walk of length $O(n)$. We reduce all these walks in $O(n^2)$ total time using Proposition 3.3. and contract, in the drawing δ , the edges of Y . Now, by Proposition 3.1, this ensures that contractible loops are shrunk to the basepoint and that homotopic loops overlap, which implies that the image of G is that of a sparse drawing of a one-vertex graph.

We then compare pairwise these $O(n)$ walks, each of length $O(n)$, in $O(n^3)$ total time. From there, we can immediately compute the loop graph L , whose image is the union of these walks, its drawing λ , and the map γ . The size of the output is $O(n^2)$. \square

5.2 Auxiliary lemmas We present three auxiliary lemmas needed for the proof of Theorem 1.3. These lemmas are valid for arbitrary (compact, orientable) surfaces, with arbitrary genus, with or without boundary, which will be useful later.

The following lemma might not be new, but we could not find a reference, so we provide a proof. It may be of independent interest. We will only use its corollary (Corollary 5.1). An *isotopy* between two embeddings of a graph G is a continuous family of embeddings of G between them.

LEMMA 5.1. *On a surface \mathcal{S} with or without boundary, let each of $L = (\ell_1, \dots, \ell_k)$ and $L' = (\ell'_1, \dots, \ell'_k)$ be a set of simple, pairwise disjoint, pairwise non-homotopic, non-contractible loops with basepoint v . Assume that ℓ_i and ℓ'_i are homotopic (with basepoint fixed) for each i . Then L and L' are isotopic (with basepoint fixed).*

Proof. Let $\hat{\mathcal{S}}$ be obtained from \mathcal{S} by removing a small disk D around v . We can assume that all loops are piecewise linear with respect to a fixed triangulation of \mathcal{S} [24, Appendix], and thus that no crossing occurs in D , and that each loop crosses the boundary of D exactly twice. We let $\hat{\ell}_i$ and $\hat{\ell}'_i$ be the arcs in $\hat{\mathcal{S}}$ that are the pieces of ℓ_i and ℓ'_i obtained after removing D , and \hat{L} and \hat{L}' be the corresponding sets of arcs.

For each i , since ℓ_i and ℓ'_i are homotopic on \mathcal{S} , they are isotopic on \mathcal{S} , by a result of Epstein [24, Theorem 4.1]. So $\hat{\ell}_i$ and $\hat{\ell}'_i$ are homotopic on $\hat{\mathcal{S}}$, where the homotopy allows to slide the endpoints on $\partial\hat{\mathcal{S}}$.

Because \hat{L} and \hat{L}' are simple, whenever there is an embedded bigon or an embedded “half-bigon” [26, Section 1.2.7] between \hat{L} and \hat{L}' , there is an innermost embedded bigon or innermost embedded half-bigon, which we can remove using an isotopy of L (sliding endpoints on the boundary is allowed), decreasing the number of crossings. So without loss of generality we can assume that there is no embedded bigon or half-bigon between \hat{L} and \hat{L}' .

In particular, $\hat{\ell}_i$ and $\hat{\ell}'_i$ are in minimal position in their homotopy classes [26, Section 1.2.7], and since they are homotopic (allowing sliding on the boundary), they are disjoint. The corresponding loops ℓ_i and ℓ'_i bound a disk on \mathcal{S} . Moreover, because the loops in L are pairwise non-homotopic, and because there is no embedded (half-)bigon between \hat{L} and $\hat{\ell}'_i$, the disk does not meet L . We can thus, for each i , push ℓ_i to ℓ'_i by an isotopy of L . \square

COROLLARY 5.1. *Let λ and λ' be homotopic sparse embeddings of the same one-vertex graph L on a surface with or without boundary. Then the rotation systems of λ and λ' are the same.*

Proof. Assume first that the surface has no boundary. Note that the basepoint b may move during the homotopy between λ and λ' , following a (possibly non-simple) path p . By an ambient isotopy of the surface (a continuous family of self-homeomorphisms), starting from the embedding λ , we push the basepoint b along path p , obtaining an embedding λ'' isotopic to λ , and thus with the same rotation system as λ , that is also homotopic to λ' with basepoint fixed. We can then apply Lemma 5.1 to λ' and λ'' , obtaining the result. \square

We now present our two main technical lemmas concerned with factorizations. Here, our notations distinguish graphs from their drawings.

LEMMA 5.2. *On a surface with or without boundary, consider a factorization (L, λ, γ) of (G, δ) . If (G, δ) can be untangled, then (L, λ) can be untangled.*

Proof. Let δ' be an embedding of G homotopic to δ . Starting from δ' , we contract the edges of G that are part of the spanning trees used to define the factorization, obtaining a loop graph embedded on \mathcal{S} . We then remove contractible loops and keep only one loop whenever several loops are in the same homotopy class. In this way, we obtain an embedding λ' of L in \mathcal{S} that is homotopic to λ . Thus, (L, λ) can be untangled. \square

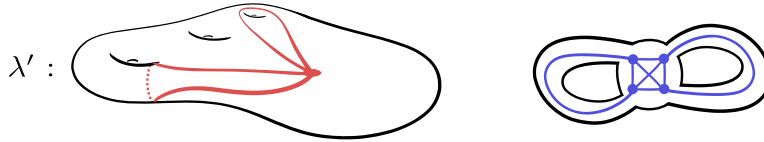


Figure 15: (Continuation of Figures 1 and 14) Although the map λ is homotopic to an embedding λ' (left), the map $\lambda' \circ \gamma$ is not a weak embedding (right), which implies by Lemma 5.3 that δ is not homotopic to an embedding.

LEMMA 5.3. *On a surface with or without boundary, consider a factorization (L, λ, γ) of (G, δ) . If δ and λ are embeddings, then $(G, \lambda \circ \gamma)$ is a weak embedding.*

See Figure 15 for an illustration of this lemma.

Proof. We first remark that, without loss of generality, we can assume that G is connected; this is because γ maps the connected components of G to those of L bijectively.

Then, we prove that we can isotope (G, δ) so that its new image in \mathcal{S} lies in the neighborhood of some sparsely embedded one-vertex graph. For this purpose, we almost-contract the spanning tree Y used to define the factorization, keeping the fact that we have an embedding of G . The resulting loops fall into homotopy classes. The loops in the trivial homotopy class can be isotoped close to the basepoint; indeed, in an embedded one-vertex graph, any contractible simple loop bounds a disk with only (possibly) contractible loops inside it [24, Theorem 1.7]. The loops in any other given homotopy class can be bundled together so that they are parallel; indeed, in an embedded one-vertex graph, any pair of homotopic loops bounds a disk with only (possibly) contractible or homotopic loops inside it. To summarize, we have an embedding δ' of G homotopic to δ whose image lies in a neighborhood of the image of some embedding λ' of a sparse loop graph L' .

Actually, there is a canonical isomorphism between L and L' , because each edge of L and L' corresponds to a particular set of non-tree edges of G (forming a single homotopy class after contracting Y); so we can identify L and L' via this isomorphism. Also, for use at the end of the proof, we remark that λ and λ' are (freely) homotopic.

By construction, the embedding δ' can be chosen to lie in any neighborhood of the map $\lambda' \circ \gamma$ (for the compact-open topology of maps from G to \mathcal{S}). Hence $\lambda' \circ \gamma$ is a weak embedding. Alternatively, it would now be easy to build the patch system of $\lambda'(L')$ (see Section 2.2) associated to λ' , so that δ and this patch system indeed witness that $(G, \lambda' \circ \gamma)$ is a weak embedding.

Now, we recall that λ and λ' are homotopic embeddings of L , so the rotation systems are the same in λ and λ' by Corollary 5.1. Also, as recalled in Section 2.2, the fact that $\lambda' \circ \gamma$ is a weak embedding depends only on the rotation system of (L, λ') , not on λ' itself. This implies that $(G, \lambda \circ \gamma)$ is a weak embedding as well. \square

5.3 End of proof of Theorem 1.3 We are now in a position to give the algorithm for Theorem 1.3. In a nutshell, we compute a factorization of G , untangle the corresponding loop graph L (or return that G cannot be untangled), determine whether G is a weak embedding in (a tubular neighborhood of) L , and declare that G can be untangled if and only if it is the case.

Proof of Theorem 1.3. Let T be a reducing triangulation of a surface \mathcal{S} of genus g , let G be a graph, and let δ be a drawing of G on T .

1. We apply Proposition 5.1 (proved with the announced running time in the following section), that is: In $O(gn \log(gn))$ time, we either determine that (G, δ) cannot be untangled on \mathcal{S} (and thus abort), or compute a factorization of (G, δ) of size $O(gn)$ given by the loop graph L , its sparse drawing λ on \mathcal{S} , and the drawing γ of G onto L , such that δ and $\lambda \circ \gamma$ are homotopic.

2. We decide whether (L, λ) can be untangled on \mathcal{S} , using Proposition 4.1, in $O(gn \log(gn))$ time. If it is not the case, we abort, because G cannot be untangled, by Lemma 5.2. Otherwise, (L, λ) can be untangled, and as indicated in Proposition 4.1, we can compute the rotation system of (L, λ') in some embedding λ' of L homotopic to λ . By Corollary 5.1, and since (L, λ) , and thus also (L, λ') , is sparse, we know that this rotation system is actually the same for all untanglings of (L, λ) .

3. Using the result of Akitaya, Fulek, and Tóth [3], restated in Theorem 2.1 above, we determine whether $(G, \lambda' \circ \gamma)$ is a weak embedding on \mathcal{S} . We do this as follows. While we have not computed the embedding λ'

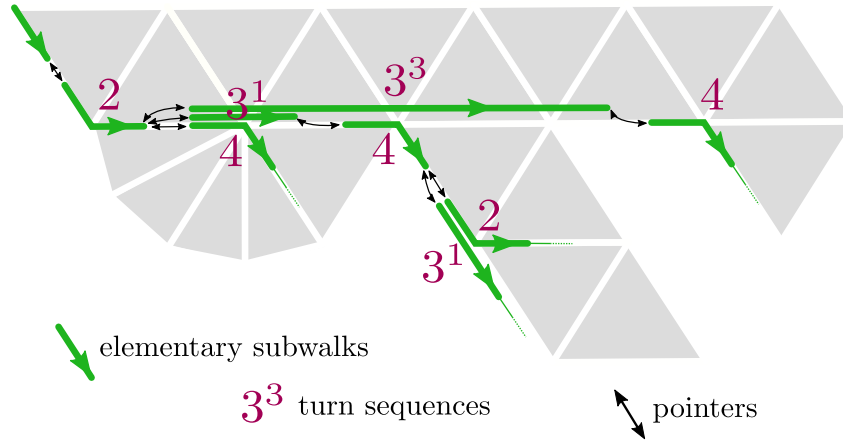


Figure 16: Homotopy classes of walks are represented by storing the elementary subwalks of their reduced walks in a tree-like fashion.

into \mathcal{S} , we have computed its rotation system in λ' , which is actually the information needed on λ' for the input of the algorithm of Akitaya et al. Thus, the algorithm of Akitaya et al. determines whether $(G, \lambda' \circ \gamma)$ is a weak embedding in $O(gn \log(gn))$ time, because the input has size $O(gn)$.

4. If $(G, \lambda' \circ \gamma)$ is a weak embedding, then the algorithm returns that (G, δ) can be untangled; this is correct because δ and $\lambda' \circ \gamma$ are homotopic on \mathcal{S} . Otherwise, the algorithm returns that (G, δ) cannot be untangled; this is correct by Lemma 5.3. We have proved the correctness at every step, and by the preceding considerations, the running time is $O(gn \log(gn))$. \square

6 Efficient graph factorization: Proof of Proposition 5.1

This section is devoted to the proof of Proposition 5.1 with the claimed running time. In particular, we fix a reducing triangulation T of an orientable surface \mathcal{S} of genus $g \geq 2$ without boundary. While a simple proof of Proposition 5.1 with a weaker complexity was presented in Section 5.1, obtaining a near-linear complexity requires more data structures.

6.1 Compressed homotopy trees Our main tool for the proof of Proposition 5.1 is a data structure, called *compressed homotopy tree* extending the *compressed homotopy sequence* from Section 3.4. While the compressed homotopy sequence is a linear list, the compressed homotopy tree is a tree-like structure. We will also extend the proof techniques from that section.

Let r be an arbitrary vertex of our reducing triangulation T . We need to support fast homotopy queries on an evolving set of walks starting at r . Intuitively, there is a map κ that associates a pointer to each walk starting at r , such that two walks W and W' are homotopic if and only if $\kappa(W) = \kappa(W')$. We call these $\kappa(W)$ **keys**. Our data structure, a **compressed homotopy tree**, provides efficient access to such a map κ in the following sense: If W' is the concatenation of W with a single edge, then we can compute $\kappa(W')$ from $\kappa(W)$ quickly (in time logarithmic in the number of operations already performed).

For this purpose, we remark that the homotopy class of a walk W is determined by the unique reduced walk $\rho(W)$ homotopic to W . We store the reduced walks $\rho(W)$ in compressed form, by enhancing the data structure from Section 3.4 in a tree-like fashion, instead of a linear list (in a way similar to shortest path trees); for each walk W starting at r , we let $\kappa(W)$ be a pointer to the last elementary subwalk of W . Each node of the tree corresponds to an elementary subwalk, and stores (see Figure 16):

- the first and last directed edges of T of the elementary subwalk;
- the turn sequence of the elementary subwalk, in compressed form (as in Section 3.4: the elementary subwalk has a turn sequence of the form 3^k , $(-3)^k$, or another symbol a ; we store its form together with the integer k or the symbol a);

- a pointer to its parent;
- the elementary subwalks of its children, stored in appropriate search trees (see below).

We maintain the following invariants:

- the children of any given node in the tree correspond to pairwise distinct turn sequences;
- the first edge of an elementary subwalk equals the last edge of its parent.

By these invariants, and because reduced paths are unique, two walks starting at r are homotopic if and only if they have the same keys. Thus, once the keys of two walks have been computed, we can, in $O(1)$ time, decide whether they are homotopic.

We now list a few operations that can be handled by our data structure. Two of them are easy:

LEMMA 6.1. *Compressed homotopy trees support the following operations:*

- TRIVIALKEY: *return the key corresponding to the empty walk at r , in $O(1)$ time;*
- REDUCEDWALK: *given the key of a walk W , return the compressed turn sequence of the reduced walk homotopic to W , in time linear in its length.*

Proof. The TRIVIALKEY operation is indeed trivial.

The REDUCEDWALK operation can be done by traversing the tree from $\kappa(W)$ up to the root, noting the labels of the nodes in order in which they are encountered, and then reading them in reverse order. \square

More interestingly, we can also extend the set of walks considered:

LEMMA 6.2. *Compressed homotopy trees support (with additional data structures described in the proof) the operation EXTEND: Given the key of a walk W , finishing at vertex v , and given a directed edge e of T starting at v , return the key of the concatenation of W and e , in $O(\log p)$ time, where p is the number of EXTEND operations already performed.*

Proof. Because we must maintain the invariant that the children of any given node in the tree correspond to pairwise distinct turn sequences, we require one more ingredient to our data structure. Given an elementary subwalk W , we need to be able to find the child of W that has a given turn sequence in $O(\log p)$ time, or certify that such a child does not exist. For this purpose, recall that each elementary subwalk has the turn sequence 3^k or $(-3)^k$ for some $k \geq 1$, or a for some $a \notin \{-3, 3\}$. We split the children of W according to each of these three categories. In a given category, a child of W is encoded by the nonnegative integer k or a . We use three red-black trees [29] (one for each category, indexed by either k or a) to decide whether the child of W with a specified compressed turn sequence exists, or to insert it if it does not exist. Each of these operations takes $O(\log p)$ time, because each red-black tree contains $O(p)$ elements.

The rest of the proof reuses tools and arguments from Section 3.4. Let W be a *reduced* walk finishing at vertex v ; let e be a directed edge starting at v ; let $W.e$ be the concatenation of W and e ; and let W' be the reduced walk homotopic to $W.e$. We claim that the compressed homotopy sequences of W and W' each have at most 49 elementary subwalks after their longest common prefix. To prove the claim it is enough to prove that the compressed homotopy sequences of $W.e$ and W' each have at most 48 symbols after their longest common prefix; indeed the compressed homotopy sequences of W and $W.e$ do not differ by more than one final elementary subwalk. Now recall from Section 3.4 that, for any walk W'' , $\varphi(W'')$ equals three times the length of W'' plus the number of bad turns of W'' , and that φ strictly decreases when performing a reduction move. In the same spirit as the proof of Proposition 3.3, observe that $\varphi(W.e) \leq \varphi(W) + 4$. Observe also that $\varphi(W) \leq \varphi(W') + 4$, since W can be obtained from the concatenation of W' and the reversal of e by performing reduction moves, which decrease the value of φ . So $\varphi(W.e) \leq \varphi(W') + 8$. Since W' is the unique walk obtained by reducing $W.e$, we have that any sequence of reductions on $W.e$ has length at most eight. Initially, only the last elementary subwalks of $W.e$ can yield a reduction. Each reduction replaces at most six consecutive elementary subwalks by at most seven ones, and then backtracks by at most six elementary subwalks (see the proof of Proposition 3.3). Thus, only the last 48 elementary subwalks of $W.e$ can be changed using successive reductions, which implies that the remaining first elementary subwalks of $W.e$ are not affected when reducing to W' . This proves the claim.

Let W and e be as in the statement of the lemma; we can assume that W is reduced. The claim implies that we can compute $\kappa(W.e)$ from $\kappa(W)$ in $O(\log p)$ time, for example as follows. Starting at the elementary subwalk representing W , in $O(1)$ time we go up in the tree by 49 levels, reaching an elementary subwalk that represents a walk W_1 . We can then compute, again in $O(1)$ time, the compressed homotopy sequence of W_2 , the walk obtained from W by removing its prefix W_1 and by concatenating e to the result. We reduce this compressed homotopy sequence in $O(1)$ time (see again the proof of Section 3.3); finally, using the first paragraph of this proof, we extend the elementary subwalk representing W with each elementary subwalk of this compressed homotopy sequence in the tree in turn, in $O(\log p)$ time, reaching eventually the elementary walk for W' . \square

Finally:

LEMMA 6.3. *Compressed homotopy trees support (with additional data structures described in the proof) the operation PARTITION: Given a set X of instances of some data structure containing, in particular, a key, we can, in time linear in the size of X , compute the partition of X such that two elements in X belong to the same part if and only if they contain the same key.*

Proof. This operation relies on easy bookkeeping techniques; in detail: Slightly extend the data structure of elementary subwalks with a pointer, initially NULL; for each element x of X in turn, if its key refers to an elementary subwalk whose pointer is NULL, create a new part of the partition, initially containing only x , and make the elementary subwalk refer to that part; otherwise, the key of x refers to an elementary subwalk W that has already been visited, and x can be added to the part thanks to the pointer in W . After computing this partition, restore all pointers to NULL. \square

One final detail: The root of the tree of our data structure must be handled in a special way. By convention, it is a virtual elementary subwalk whose only directed edge e does not belong to the triangulation T but points towards the root r , so that by convention the turns from e to each edge starting at r are ∞ , with the obvious modifications on the algorithm to reduce walks. We omit the tedious and trivial details.

6.2 Algorithm We can now give the algorithm for Proposition 5.1. The intuition is, again, to contract the edges of a spanning forest of G and to record the homotopy classes of the resulting loops. This computes the sparse drawing λ of the loop graph L . In a second step, we compute the drawing γ of G onto L .

Proof of Proposition 5.1. By definition of a factorization, we can assume in the whole proof that G is connected. Let r be an arbitrary root vertex of G and let Y be a spanning tree of G obtained by a *depth-first search* from r . Let us orient each edge e of $G \setminus Y$ so that its target $t(e)$ is an ancestor of its source $s(e)$. Let ℓ_e be the loop that is the concatenation of the (unique) path in Y from r to $s(e)$, edge e , and the (unique) path in Y from $t(e)$ to r . Given a vertex v on the path from $t(e)$ to r , we let ℓ_e^v be the loop ℓ_e where the last segment from v to r is truncated (thus, v always appears in ℓ_e^v).

In a first step, we compute the key of (the image in T of) the path in Y from r to each vertex of G . This can be done in $O(n \log n)$ time by traversing Y in top-down order and using the EXTEND operation $O(n)$ times. Similarly, in the same amount of time, we can compute, for each edge e of $G \setminus Y$, the key $\kappa(\ell_e^{t(e)})$ of the image of $\ell_e^{t(e)}$ in the triangulation T .

In a second step, using the keys already computed, we compute the keys of the entire loops ℓ_e , but to get an efficient algorithm we eliminate duplicate keys as we encounter them, and halt whenever there are too many distinct keys (i.e., too many homotopy classes, so that G cannot be untangled). We do this in a bottom-up way in the tree Y . For each vertex v of G , let $K(v)$ be the set of keys of the walks ℓ_e^v , for each edge e of $G \setminus Y$ such that v appears after e in ℓ_e .

Assume that we have computed the sets $K(v')$ for each child v' of a vertex v ; using the PARTITION operation, we can assume each $K(v')$ to be without duplicates. If one of them has size larger than $12g$, then the *directed* loops ℓ_e fall into at least $12g + 1$ distinct homotopy classes, which, by a folklore result (see, e.g., [8, Lemma 2.1]), implies that G cannot be untangled, so we abort. Otherwise, for each child v' of v in T and for each key $k \in K(v')$, we apply the EXTEND function to k and edge $v'v$. Let $K(v', v)$ be the resulting set of keys. We now observe that $K(v)$ is obtained by removing the duplicates in the union of the following sets of keys:

- the sets $K(v', v)$, for each child v' of v , and

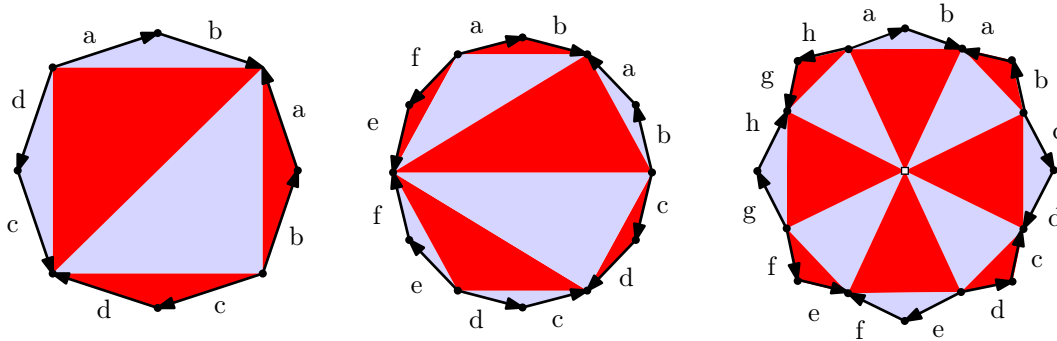


Figure 17: How to build a reducing triangulation from a canonical polygonal schema of \mathcal{S} . The cases $g = 2$ (left) and $g = 3$ (middle) use special constructions; the case $g \geq 4$ (right) trivially generalizes to higher genus. In all cases, there is a walk of length $O(1)$ between any two corners of the polygonal schema.

- the set of keys of the form $\kappa(\ell_e^v)$, for each edge e such that $t(e) = v$.

At the end of the recursion, unless the algorithm aborted, we have computed $K(r)$. At each vertex v , the number of times we apply EXTEND is $O(g \deg(v))$, where $\deg(v)$ is the degree of v in Y . So in total, EXTEND has been applied $O(gn)$ times, and the entire algorithm takes $O(gn \log(gn))$ time.

When $K(r)$ has been computed, we can also compute the compressed turn sequence of the reduced walks of $K(r)$ in time linear in their lengths, using the REDUCEDWALK operation. Actually, since $K(r)$ corresponds to $O(g)$ walks of length $O(n)$ in G , we may uncompress these turn sequences and return the explicit description of the $O(g)$ reduced walks in $O(gn)$ time, thus asymptotically without overhead in the running time. We have thus computed the sparse drawing λ of the one-vertex graph L on \mathcal{S} .

Finally, remember that we also need to compute the drawing γ of G onto L . In other words, we need to compute the mapping from each edge $e \in G \setminus Y$ to the loops of L (or to its basepoint), or, equivalently, the key of ℓ_e in $K(r)$. We now explain how to refine the second step above to achieve this. In this second step, every key considered is of the form $k = \kappa(\ell_e^v)$; whenever we encounter such a key k , we also store together with it a corresponding edge e such that $\kappa(\ell_e^v) = k$. Whenever we detect redundancies in the current set of keys (see the end of Section 6.1), we actually compute the set of edges $e_1, \dots, e_p \in G \setminus Y$, $p \geq 1$, whose keys share a common value k , which in particular implies $\kappa(\ell_{e_1}) = \dots = \kappa(\ell_{e_p})$. In that case, we declare that the edge corresponding to the common key k is e_1 , declare that the *leader* of e_2, \dots, e_p is e_1 , and then completely forget about the edges e_2, \dots, e_p . There is no overhead in the running time of the algorithm. At the very end, once the keys in $K(r)$ have been computed (each of them coming with an associated edge in $G \setminus Y$), we need to recover, for each edge $e \in G \setminus Y$, the corresponding key $\kappa(\ell_e) \in K_r$. For this purpose, we remark that the set of edges in $G \setminus Y$ is implicitly organized as a forest, in which the “leader” relation is actually a “parent” relation. Thus, our problem boils down to this: Given a forest, we need to compute, for each node, the root of its corresponding tree. This can easily be done in time linear in the size of the forest, and thus, in our case, in $O(n)$ time. \square

7 Proof of Theorem 1.1 when $g \geq 2$ and $b = 0$

In this section we prove Theorem 1.1 in the case where our input surface \mathcal{S} has genus at least two and is without boundary, using Theorem 1.3. To do so, we transform homotopically our input drawing δ into another one that lies in the 1-skeleton of a reducing triangulation. This is done in the following lemma.

Given a graph H embedded on \mathcal{S} , a *monogon* is face of H that is a disk bounded by a single edge of H (which is thus a loop). A *bignon* is a face of H that is a disk bounded by two edges of H (which are thus parallel edges).

LEMMA 7.1. *Let H be a graph of size m cellularly embedded on a surface \mathcal{S} of genus $g \geq 2$ without boundary. Let G be a graph, and let δ be a drawing of G on H . In $O(m + g(g + |\delta|))$ time, we can construct a reducing triangulation T of \mathcal{S} and a drawing δ' of G on T , where δ' is homotopic to δ and $|\delta'| = O(g|\delta|)$.*

Proof. We first claim that we can assume without loss of generality that H is a set of $O(g)$ loops with a common basepoint. Indeed, we can, in $O(m)$ time, contract an arbitrary spanning tree of H and transform the input drawing δ of G homotopically into a drawing of G in that new graph. Then we can iteratively remove edges forming monogons, and merge together edges forming bigons, all in $O(m)$ time. Euler's formula and double counting of the edge-face incidences then implies that H has $O(g)$ loops. This proves the claim.

A *canonical system of loops* L of \mathcal{S} is a set of pairwise disjoint simple loops with a common basepoint b , such that, when cutting \mathcal{S} along L , we obtain a *canonical polygonal schema*, a $4g$ -gon whose boundary reads $a_1 b_1 \bar{a}_1 \bar{b}_1 \dots a_g b_g \bar{a}_g \bar{b}_g$ in this order, where $a_1, \dots, a_g, b_1, \dots, b_g$ denote the loops in L , and bar denotes reversal. Using a result by Pocchiola, Lazarus, Vegter, and Verroust [36], we compute, in $O(g^2)$ time, a canonical system of loops L of \mathcal{S} in general position with respect to H such that each loop of L crosses each loop of H at most four times. We actually store the combinatorial map representing the overlay of H and L . Then, we merge the vertex of H with the basepoint b of L , so that we now have the overlay of L and H , with the same vertex b . Now, for each edge e of the system of loops L , if k edges of H cross e , thus splitting it into $k + 1$ subedges, we contract k of these subedges. The effect is that each edge of H is transformed into an ordered set of arcs in the polygonal schema, each connecting two corners. (H is not embedded any more, but this is irrelevant; only its homotopy class matters in our problem.)

It is not difficult to extend L into a reducing triangulation T of \mathcal{S} containing L ; see Figure 17. We can now push every arc of H into a walk of length $O(1)$ in T . Finally, this new drawing of H in T naturally yields a drawing δ' of G in T , homotopic to δ , which we can compute in $O(g|\delta|)$ time, of size $O(g|\delta|)$. \square

Proof of Theorem 1.1 in the case $g \geq 2, b = 0$. The result follows immediately from Lemma 7.1 and Theorem 1.3. \square

8 Proof of Theorem 1.1 for the remaining cases

In this section we solve our untangling problem for the (orientable) surfaces not handled in the previous section, namely when \mathcal{S} is the torus or has non-empty boundary.

Recall that the entire algorithm for surfaces of genus at least two works as follows. In Step (0), we transform the input graph H into a reducing triangulation (Section 7). Then we apply the four steps in the proof of Theorem 1.3 (Section 5.3): (1) We compute a factorization (L, λ, γ) of our graph drawing (G, δ) (or immediately return that (G, δ) cannot be untangled); (2) we untangle (L, λ) , obtaining an embedding (L, λ') (or return that (G, δ) cannot be untangled); (3) we decide whether $(G, \lambda' \circ \gamma)$ is a weak embedding; (4) we return the answer accordingly.

We follow the same general strategy. The arguments are still valid because they rely on auxiliary lemmas (Section 5.2) valid for arbitrary surfaces. Three ingredients must be adjusted: In Step (0), we must transform the input graph; in Step (1), we must compute a factorization; in Step (2), we must untangle a loop graph.

8.1 Untangling on the torus

We now prove Theorem 1.1 when the input surface \mathcal{S} is the torus.

The proof relies on three lemmas, for Steps (0), (1), and (2), respectively. We remark that we cannot apply the exact same strategy as in the case of genus at least two, because the torus neither admits a hyperbolic metric (by the Gauss–Bonnet theorem) nor a reduced triangulation (by Euler's formula). On the other hand, we can replace it with a simpler structure because of the Abelian structure on the homotopy group of the torus, see below. A *canonical system of loops* Q of the torus is a set of pairwise disjoint simple loops (q_1, q_2) with common basepoint that cross at the basepoint. Similar to Lemma 7.1, we have the following lemma, the proof of which is strictly easier than Lemma 7.1, and thus omitted.

LEMMA 8.1. *Let H be a graph of size m cellularly embedded on a torus \mathcal{S} . Let G be a graph, and let δ be a drawing of G on H . In $O(m + |\delta|)$ time, we can construct a canonical system of loops Q of \mathcal{S} and a drawing δ' of G on Q , where δ' is homotopic to δ and $|\delta'| = O(|\delta|)$.*

So henceforth we can assume that our input drawing is on Q . Let q_1 and q_2 be the two loops of Q . The homotopy group on the torus is isomorphic to \mathbb{Z}^2 ; this means that every closed walk in Q is homotopic to $q_1^{u_1} \cdot q_2^{u_2}$ (where “ \cdot ” denotes concatenation) for a unique $(u_1, u_2) \in \mathbb{Z}^2$, and we represent homotopy classes by such vectors. Moreover, walk concatenation translates, in homotopy, to vector addition, and free homotopy classes can also be represented by such vectors. We say that a walk in Q is *reduced* if it is of the form $q_1^{u_1} \cdot q_2^{u_2}$, for some integers u_1 and u_2 . Every walk in Q is homotopic to a unique reduced walk; we can reduce walks in linear time.

LEMMA 8.2. *We can, in time $O(|\delta'| \log |\delta'|)$, compute a factorization of (G, δ') , or correctly report that δ' cannot be untangled.*

Proof. The algorithm is an easy variation on the one described in Section 6.2. The only change concerns the data structure that stores homotopy classes of walks in Q , described in Section 6.1 for reducing triangulations. Here we replace the reduced walks in a reducing triangulation by reduced walks in Q . The compressed homotopy tree structure is replaced by a two-level tree-like structure, the first level for the integer u_1 and the second level for the integer u_2 in the notations above. The four operations introduced in Section 6.1 extend to this context, with the same complexity. Section 6.2 extends immediately, because it only uses these four operations. \square

LEMMA 8.3. *Let Q be a canonical polygonal schema of the torus. Let L be a loop graph and λ be a sparse drawing of L in Q . We can decide whether λ can be untangled in $O(|\lambda|)$ time. If so, we can compute the cyclic ordering of the loops around each vertex of an embedding λ' of L homotopic to λ .*

Proof. We claim that we can assume, without loss of generality, that L is connected and contains at most three loops. Indeed, if (L, λ) can be untangled, each connected component of L has at most three loops (by Euler's formula) and if it has at least two loops, then by sparsity every embedding of that connected component cuts the torus into a disk. Thus, either the conclusion of our claim holds, or L has several connected components, each made of a single loop. These loops must be pairwise freely homotopic, because otherwise they cannot be untangled; but then (L, λ) can be untangled if and only if its restriction to a single connected component of L can be untangled (and in that case the rotation systems are trivial). All these tests take linear time. This proves the claim.

Recall (see, e.g., Farb and Margalit [26, Section 1.2.3]) that the homotopy class (u_1, u_2) contains a simple closed curve if and only if $\gcd(|u_1|, |u_2|) = 1$, and that the homotopy classes (u_1, u_2) and (v_1, v_2) have geometric intersection number $|u_1 v_2 - u_2 v_1|$ (the minimum number of intersections of curves in their free homotopy classes). We discuss according to the number of loops in L .

If L is made of a single loop in homotopy class (u_1, u_2) , then (L, λ) can be untangled if and only if $\gcd(|u_1|, |u_2|) = 1$.

If L is made of two loops, in homotopy classes (u_1, u_2) and (v_1, v_2) , then, assuming (L, λ) can be untangled, the loops must be interleaved around the basepoint, and the corresponding closed curves must cross exactly once. Thus, (L, λ) can be untangled if and only if $|u_1 v_2 - u_2 v_1| = 1$.

If L is made of three loops, we first apply the previous test for the first two loops, in homotopy classes (u_1, u_2) and (v_1, v_2) . If it fails, then (L, λ) cannot be untangled. Otherwise, in an embedding λ' , these two loops together cut the torus into a square, and the third loop must be a diagonal of that square, so it must have signature $\pm(u_1, u_2) \pm (v_1, v_2)$, and the converse is obvious.

If the test passes, we need to recover the rotation system of L in some embedding homotopic to λ . We can actually build an explicit embedding of L homotopic to λ , as follows. Recall that the universal cover of the torus can be viewed by tiling infinitely many copies of the polygonal schema of (q_1, q_2) in a grid, identified to a square unit grid \mathbb{Z}^2 . We replace each loop of (L, λ) in homotopy class (u_1, u_2) by a loop that lifts, in the universal cover, from point $(0, 0)$ to point (u_1, u_2) . The above conditions are exactly an algebraic characterization for these loops to be simple and pairwise disjoint, except at their common basepoint. From this representation, we immediately obtain the rotation system in that embedding, by looking at the cyclic order around the origin. \square

The proof of Theorem 1.1 in the case of the torus is now straightforward.

Proof of Theorem 1.1 for the torus. We first apply Lemma 8.1. Then we apply the algorithm described in the proof of Theorem 1.3 in Section 5.3, with the only modifications that we apply Lemma 8.2 instead of Proposition 5.1 and Lemma 8.3 instead of Proposition 4.1. \square

8.2 Untangling on surfaces with non-empty boundary In this section, we prove Theorem 1.1 in the case where our input surface \mathcal{S} has non-empty boundary. An obvious strategy is to attach a handle to each boundary component and to apply Theorem 1.1 to the resulting surface without boundary. While this is certainly a valid approach, we show here that we can circumvent a large part of the technical machinery, even though this does not improve the asymptotic running time.

A **loop system** of \mathcal{S} is a set Q of pairwise disjoint simple loops with a common basepoint on \mathcal{S} such that each face of Q has genus zero and contains a single component of the boundary of \mathcal{S} .

LEMMA 8.4. *Let H be a graph of size m cellularly embedded on a surface \mathcal{S} of genus $g \geq 0$ with $b \geq 1$ boundary components. Let G be a graph, and let δ be a drawing of G on H . In $O(m + |\delta|)$ time, we can construct a loop system Q of \mathcal{S} and a drawing δ' of G on Q , where δ' is homotopic to δ and $|\delta'| = O((g + b)|\delta|)$.*

Proof. We represent \mathcal{S} as follows. First, we close each boundary component of \mathcal{S} with a disk, obtaining a surface without boundary $\bar{\mathcal{S}}$. The graph H is embedded on $\bar{\mathcal{S}}$. The faces of that graph on $\bar{\mathcal{S}}$ that correspond to boundary components of \mathcal{S} are marked as *punctured*. We keep the invariant that, at every step, every face contains at most one puncture.

We first contract the edges of a spanning tree of H , and in passing we remove these edges from the drawing δ . This preserves the homotopy class of δ . Now H has only one vertex and the edges of H are loops. We fix a set L of loops of H whose complement is an open disk in $\bar{\mathcal{S}}$. For example, we let L to be the complement of a spanning tree of the dual of H .

We construct Q from H by iteratively removing a loop ℓ of $H \setminus L$ whenever ℓ is incident to a face marked punctured and a face *not* marked punctured. These two faces merge into a single face marked punctured. In the end every face of Q is marked punctured so Q is indeed a loop system of \mathcal{S} . Removing these edges all at once can be done in linear time by taking a spanning forest of the dual of $(H \setminus L)$ rooted at the punctured faces.

There remains to reroute the edges of δ as needed. Every edge of H not in Q belongs to a unique face of Q , and can be rerouted homotopically to a subpath of that face, on the side of that face that does not contain the puncture. By Euler's formula Q contains $O(g + b)$ loops, so the faces have size $O(g + b)$ in total. Thus, the size of the drawing increases by a factor $O(g + b)$. \square

We remark that \mathcal{S} is homeomorphic to the patch system Σ of Q , and the arcs of Σ correspond to the dual of Q . Since \mathcal{S} is a surface with non-empty boundary, the homotopy group of \mathcal{S} is a free group; see, e.g., Stillwell [44, Chapter 2]. This implies two facts that we shall use later: (1) The homotopy classes of paths or loops in \mathcal{S} correspond to sequences of crossings with the arcs of Σ that are *reduced* (do not contain two consecutive crossings with the same arc in opposite directions); (2) the free homotopy classes of closed curves on \mathcal{S} correspond to *cyclically reduced* sequences of crossings.

(Such properties are folklore, and follow from their analogs for the homotopy of paths and closed walks in graphs. It is clear that two paths with the same reduced sequence of crossings are homotopic; conversely, any two homotopic paths lift to paths with the same endpoints in the universal cover, and their sequences of crossings, once reduced by homotopies, must be the same because Σ lifts to pairwise disjoint simple arcs, each of which separates the universal cover into two pieces. A similar argument holds for closed walks, where the homotopy between them must be lifted.)

LEMMA 8.5. *Let Q be a loop system of size m in a surface \mathcal{S} with non-empty boundary. Let G be a graph and δ be a drawing of G in Q . In $O(m|\delta| \log(m|\delta|))$ time, we can compute a factorization of (G, δ) of size $O(m|\delta|)$ or correctly report that δ cannot be untangled.*

Proof. The algorithm is the one described in Section 6.2 for the proof of Proposition 5.1. In the same way as for the torus the only change concerns the data structure that stores homotopy classes of walks. Here we replace the reduced walks in a reducing triangulation by reduced walks in Q . That is, we store these walks as lists of directed edges of Q in a tree-like fashion to implement the data-structure of Section 6.1. The four operations introduced in Section 6.1 extend to this context, with the same complexity. Section 6.2 extends immediately, because it only uses these four operations. \square

From the above factorization, we obtain a drawing λ of a loop graph L on Σ . As in Section 4, we choose an arbitrary edge of each connected component of L and declare it to be a **major edge**; the other edges are **minor edges**. We say that λ is **straightened** if, under λ , (1) the sequence of crossings of each major edge with the arcs of Σ is cyclically reduced, and (2) the sequence of crossings of each minor edge with the arcs of Σ is reduced (possibly not cyclically). The analog of Proposition 4.2 becomes:

LEMMA 8.6. *Let L be a loop graph and let $\lambda : L \rightarrow \Sigma$ be a drawing. Assume that λ is sparse and straightened. If there is an embedding homotopic to λ , then λ is a weak embedding.*

Proof. Since λ is straightened, $\lambda(e)$ is cyclically reduced for every major edge e of L , and $\lambda(e')$ is reduced for every minor edge e' of L . There is by assumption an embedding $\lambda' : L \rightarrow \Sigma$ freely homotopic to λ .

In a first step, we modify the embedding λ' by an ambient isotopy that fixes the boundary of Σ so that the major edges of L are cyclically reduced (with respect to the arcs of Σ). For this purpose, we remark that the major edges are, in λ' , pairwise disjoint simple closed curves. Whenever there is a bigon between the image of a major edge and an arc of Σ , there exists an innermost bigon, which we can remove by an ambient isotopy; the number of crossings with the arcs decreases. We repeat this operation until there is no more bigon, at which point the major edges are all cyclically reduced in λ' .

Now consider a connected component L_0 of L . We can make the major edge e of L_0 cross the arcs of Σ with the same sequences in λ' and in λ (not up to cyclic permutation, *exactly* the same sequence) simply by sliding the image of the basepoint v of L_0 along the image of e in λ' . We slide v by an ambient isotopy in the tubular neighborhood of e . We can do so since the sequence of crossings of $\lambda'(e)$ is a cyclic permutation of the one of $\lambda(e)$; here we make use of the fact (see above) that if two freely homotopic closed curves are cyclically reduced, then their sequences of crossings with the arcs of Σ are equal up to cyclic permutation. In λ' we slide the image of v along the image of e so that the two sequences become equal; λ' is still an embedding. We do this for every connected component L_0 of L in turn.

Consider again some connected component L_0 of L , with vertex v . We claim that we can modify λ' by sliding the image of v some finite number of times around the image of the major edge e of L_0 (each time, the image of v making one full loop around the image of e) so that, in the end, there is a free homotopy between $\lambda'(L_0)$ and $\lambda(L_0)$ in which the image of v does not leave its face of the patch system Σ . To prove this claim first observe that it would be possible to do so, not by sliding along e , but by some *free* homotopy of $\lambda'(L_0)$. During this homotopy the image of v makes a loop ℓ in Σ . The loop ℓ commutes, up to homotopy, with the loop $\lambda'(e)$ as $\lambda'(e) \simeq \lambda(e)$ by the previous paragraph and $\lambda(e) \simeq \ell\lambda'(e)\ell^{-1}$ by construction, where \simeq denotes homotopy of loops relatively to their basepoint. Moreover, the loop $\lambda'(e)$ is non-contractible since λ' is sparse, so it is also primitive by Lemma 2.1. Thus ℓ is a power (up to homotopy) of $\lambda'(e)$. Let us prove this. The fundamental group of Σ is a free group. It is known (and we prove) that in a free group if two elements x and y commute, then they are powers of some common element: indeed, the subgroup K generated by x and y is an Abelian subgroup, which is free by the Nielsen–Schreier theorem; but the only Abelian free group is \mathbb{Z} ; so K is cyclic. Now, as mentioned above, $\lambda'(e)$ is primitive.

Now in λ' any bigon between a minor edge of L_0 and an arc of Σ does not contain any vertex of L . Indeed, otherwise, the major edge incident to that vertex would not be cyclically reduced, a contradiction. So we can remove any innermost bigon by an ambient isotopy. When this is not possible anymore, by the preceding claim, the minor edges intersect the arcs of Σ with the same sequence in λ and λ' . Here we make use of the fact (see above) that if two loops are homotopic relatively to their basepoint (or via a free homotopy in which the basepoint does not leave its face of Σ) and reduced, then they intersect the arcs of Σ with the same sequence. \square

Proof of Theorem 1.1 in the case $b \geq 1$. We apply Lemma 8.4 to construct in time $O(m+n)$ a loop system Q and a drawing $\delta' : G \rightarrow Q$ of size $O((g+b)n)$ homotopic to δ . We apply Lemma 8.5 to compute in time $O((g+b)^2n \log((g+b)n))$ a factorization (L, λ, γ) of (G, δ') of size $O((g+b)^2n)$, or we correctly conclude that δ' (and thus δ) cannot be untangled. We straighten λ , by first cyclically reducing the major edges and then reducing the minor edges, similar to Lemma 4.1. Then we apply Lemma 8.6 and Theorem 2.1 in time $O((g+b)^2n \log((g+b)n))$ to determine if there is an embedding $\lambda' : L \rightarrow \mathcal{S}$ homotopic to λ . If not, we conclude by Lemma 5.2 that δ' (and thus δ) cannot be untangled. Otherwise, we retain the rotation system of such an embedding λ' , and we apply Theorem 2.1 to determine if $\lambda' \circ \gamma$ is a weak embedding. The result tells us whether δ can be untangled by Lemma 5.3. \square

9 Untangling piecewise linear drawings in the punctured plane

In this section we consider the model for the input where the surface is the punctured plane and the drawing is a piecewise linear map. More precisely we prove Theorem 1.2. (The *size* of a piecewise linear drawing δ of a graph G is the size of G plus the total number of segments comprising the edges of the images of G under δ .)

Proof of Theorem 1.2. We reduce to the combinatorial map model in a way very similar to Cabello, Liu, Mantler, and Snoeyink [6, Lemma 11], in the same spirit as Colin de Verdière and de Mesmay [15, Section 5.2]. More precisely we do the following. Firstly, we fix a closed box around each point of P that does not intersect the image of δ nor the other points of P . Denoting by \mathcal{P} the resulting collection of boxes we also fix a bounding box B that contains the image of δ and the boxes in \mathcal{P} in its interior. Secondly, we construct, in time $O(p^{1+\epsilon})$ for some

arbitrary ε , $0 < \varepsilon < 1/2$, a piecewise linear cellular decomposition H of $B \setminus \mathcal{P}$ of size $O(p)$ such that every line in the plane crosses at most $O(\sqrt{p})$ segments of H [6, Lemma 11]. Without loss of generality, we assume that H is in general position with respect to our drawing δ . Thirdly, for every segment s of the image of δ we compute the sequence of the $O(\sqrt{p})$ segments of H that are crossed by s , in the order in which they are crossed. Fourthly, we consider the dual graph H^* of H and compute a drawing δ' homotopic to δ that lies in H^* ; it has size $O(n\sqrt{p})$. Finally we apply Theorem 1.1 to δ' and H^* to determine if there exists an embedding homotopic to δ' in $B \setminus \mathcal{P}$. \square

Acknowledgments. We would like to thank Arnaud de Mesmay for many useful discussions, and the anonymous referees whose numerous comments improved the presentation of this paper.

References

- [1] Pankaj K. Agarwal. Range searching. In Jacob E. Goodman, Joseph O'Rourke, and Csaba Toth, editors, *Handbook of Discrete and Computational Geometry*, chapter 40, pages 1057–1092. CRC Press LLC, third edition, 2018.
- [2] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The design and analysis of computer programs*. Addison-Wesley, 1974.
- [3] Hugo A. Akitaya, Radoslav Fulek, and Csaba D. Tóth. Recognizing weak embeddings of graphs. *ACM Transactions on Algorithms*, 15(4):Article 50, 2019.
- [4] Mark A. Armstrong. *Basic Topology*. Springer New York, NY, 1983.
- [5] Joan S. Birman and Caroline Series. An algorithm for simple curves on surfaces. *Journal of the London Mathematical Society*, 2(2):331–342, 1984.
- [6] Sergio Cabello, Yuanxin Liu, Andrea Mantler, and Jack Snoeyink. Testing homotopy for paths in the plane. *Discrete & Computational Geometry*, 31:61–81, 2004.
- [7] James W. Cannon, William J. Floyd, Richard Kenyon, and Walter R. Parry. Hyperbolic geometry. In Silvio Levi, editor, *Flavors of geometry*. Cambridge University Press, 1997.
- [8] Erin W. Chambers, Éric Colin de Verdière, Jeff Erickson, Francis Lazarus, and Kim Whittlesey. Splitting (complicated) surfaces is hard. *Computational Geometry: Theory and Applications*, 41(1–2):94–110, 2008.
- [9] Hsien-Chih Chang and Arnaud de Mesmay. Tightening curves on surfaces monotonically with applications. *ACM Transactions on Algorithms*, 18(4):36:1–36:32, 2022.
- [10] Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Discrete & Computational Geometry*, 58(4):889–920, 2017.
- [11] Hsien-Chih Chang, Jeff Erickson, and Chao Xu. Detecting weakly simple polygons. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1655–1670, 2015.
- [12] David R.J. Chillingworth. Simple closed curves on surfaces. *Bulletin of the London Mathematical Society*, 1(3):310–314, 1969.
- [13] David R.J. Chillingworth. Winding numbers on surfaces. II. *Mathematische Annalen*, 199(3):131–153, 1972.
- [14] Marshall Cohen and Martin Lustig. Paths of geodesics and geometric intersection numbers: I. In *Combinatorial group theory and topology*, volume 111 of *Ann. of Math. Stud.*, pages 479–500. Princeton Univ. Press, 1987.
- [15] Éric Colin de Verdière and Arnaud de Mesmay. Testing graph isotopy on surfaces. *Discrete & Computational Geometry*, 51(1):171–206, 2014.
- [16] Éric Colin de Verdière and Jeff Erickson. Tightening nonsimple paths and cycles on surfaces. *SIAM Journal on Computing*, 39(8):3784–3813, 2010.
- [17] Pier Francesco Cortese, Giuseppe Di Battista, Maurizio Patrignani, and Maurizio Pizzonia. On embedding a cycle in a plane graph. *Discrete Mathematics*, 309(7):1856–1869, 2009.
- [18] Maurits de Graaf and Alexander Schrijver. Making curves minimally crossing by Reidemeister moves. *Journal of Combinatorial Theory, Series B*, 70(1):134–156, 1997.
- [19] Max Dehn. Transformation der Kurven auf zweiseitigen Flächen. *Mathematische Annalen*, 72:413–421, 1912.
- [20] Vincent Despré, Benedikt Kolbe, Hugo Parlier, and Monique Teillaud. Computing a Dirichlet domain for a hyperbolic surface. In *39th International Symposium on Computational Geometry (SoCG 2023)*, pages 27:1–27:15, 2023.
- [21] Vincent Despré and Francis Lazarus. Computing the geometric intersection number of curves. *Journal of the ACM*, 66(6):41:1–45:49, 2019.
- [22] Tamal K. Dey and Sumanta Guha. Transforming curves on surfaces. *Journal of Computer and System Sciences*, 58(2):297–325, 1999.
- [23] David Eppstein. Dynamic generators of topologically embedded graphs. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 599–608, 2003.

- [24] David B. A. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.
- [25] Jeff Erickson and Kim Whittlesey. Transforming curves on surfaces redux. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1646–1655, 2013.
- [26] Benson Farb and Dan Margalit. *A primer on mapping class groups*. Princeton university Press, 2012.
- [27] Radoslav Fulek. Embedding graphs into embedded graphs. *Algorithmica*, 82:3282–3305, 2020.
- [28] Xavier Goaoc, Jan Kratochvíl, Yoshio Okamoto, Chan-Su Shin, Andreas Spillner, and Alexander Wolff. Untangling a planar graph. *Discrete & Computational Geometry*, 42(4):542–569, 2009.
- [29] Leonidas J. Guibas and Robert Sedgwick. A dichromatic framework for balanced trees. In *Proceedings of the 20th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 8–21, 1979.
- [30] Joel Hass and Peter Scott. Intersections of curves on surfaces. *Israel Journal of mathematics*, 51:90–120, 1985.
- [31] John Hopcroft and Robert Tarjan. Efficient planarity testing. *Journal of the ACM*, 21(4):549–568, 1974.
- [32] Svetlana Katok. *Fuchsian groups*. University of Chicago press, 1992.
- [33] Ken-ichi Kawarabayashi, Bojan Mohar, and Bruce Reed. A simpler linear time algorithm for embedding graphs into an arbitrary surface and the genus of graphs of bounded tree-width. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 771–780. IEEE, 2008.
- [34] Lutz Kettner. Using generic programming for designing a data structure for polyhedral surfaces. *Computational Geometry: Theory and Applications*, 13:65–90, 1999.
- [35] Yves Ladegaillerie. Classes d’isotopie de plongements de 1-complexes dans les surfaces. *Topology*, 23(3):303–311, 1984.
- [36] Francis Lazarus, Michel Pocchiola, Gert Vegter, and Anne Verroust. Computing a canonical polygonal schema of an orientable triangulated surface. In *Proceedings of the 17th Annual ACM Symposium on Computational Geometry (SOCG)*, pages 80–89. ACM, 2001.
- [37] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 440–449, 2012.
- [38] Martin Lustig. Paths of geodesics and geometric intersection numbers: II. In *Combinatorial group theory and topology*, volume 111 of *Ann. of Math. Stud.*, pages 501–543. Princeton Univ. Press, 1987.
- [39] Bojan Mohar. A linear time algorithm for embedding graphs in an arbitrary surface. *SIAM Journal on Discrete Mathematics*, 12(1):6–26, 1999.
- [40] Henri Poincaré. Cinquième complément à l’analysis situs. *Rendiconti del Circolo Matematico di Palermo*, 18(1):45–110, 1904.
- [41] Bruce L. Reinhart. Algorithms for Jordan curves on compact surfaces. *Annals of Mathematics*, pages 209–222, 1962.
- [42] Marcus Schaefer. The graph crossing number and its variants: A survey. *Electronic Journal of Combinatorics*, Dynamic Surveys:Article 21, 2013.
- [43] Marcus Schaefer. Hanani–Tutte and related results. In Imre Bárány, Karoly J. Böröczky, and László Fejes Tóth, editors, *Geometry—intuitive, discrete, and convex. A tribute to László Fejes Tóth*, volume 24 of *Bolyai Society Mathematical Studies*. Springer-Verlag, 2014.
- [44] John Stillwell. *Classical topology and combinatorial group theory*. Springer-Verlag, New York, second edition, 1993.
- [45] Carsten Thomassen. The graph genus problem is np-complete. *Journal of Algorithms*, 10(4):568–576, 1989.

A Limit points in the universal cover

In this section, we present the (classical) details omitted from Section 4.2. A good overview is provided by Farb and Margalit [26, Chapter 1], but we need a few more properties, so we provide proofs.

A.1 Compactification of the hyperbolic plane and fixed points of translations Let \mathbb{H} be the hyperbolic plane, corresponding to the open unit disk in the Poincaré model. One can compactify \mathbb{H} by considering the set $\partial\mathbb{H}$ of “points at infinity”, corresponding to the unit circle (in the Poincaré model) with its usual topology. Equivalently [26, Chapter 1], the points in $\partial\mathbb{H}$ are the equivalence classes of unit speed geodesic rays, where two rays are equivalent if they stay at bounded distance from each other; the union $\bar{\mathbb{H}}$ of \mathbb{H} and $\partial\mathbb{H}$ is topologized via the basis containing the open sets of \mathbb{H} plus one open set U_P for each open half-plane P of \mathbb{H} , where $U_P \cap \mathbb{H} = P$ and $U_P \cap \partial\mathbb{H}$ contains the equivalence class ℓ of unit speed geodesic rays if all the rays in ℓ eventually end up in P .

Isometries of \mathbb{H} extend naturally to $\bar{\mathbb{H}}$, and the *hyperbolic translations* are those with exactly two fixed points on $\partial\mathbb{H}$. In particular the identity is not considered a hyperbolic translation here. Any hyperbolic translation f admits a unique geodesic line A , its *axis*, such that $f(A) = A$ and such that f is a real translation on A . A hyperbolic translation is uniquely determined by its axis and by the image of a point on this axis. Iterating any point of \mathbb{H} under f makes it converge to one of the two fixed points of f in $\partial\mathbb{H}$, the *fixed point*

at $+\infty$ of f , while iterating under f^{-1} makes it converge to the **fixed point at** $-\infty$ of f . See [32, p. 13-14] for more details.

The following lemma is standard and results from simple computations in the Poincaré model of the hyperbolic plane, so we omit the proof.

LEMMA A.1. *In the hyperbolic plane \mathbb{H} let $L \subset \mathbb{H}$ be a geodesic line and $a : \mathbb{R} \rightarrow \mathbb{H}$ be a unit speed geodesic ray. The distance between $a(t)$ and L either tends to $+\infty$ as $t \rightarrow +\infty$ or it tends to zero. In the latter case, there is a unit speed parametrization $c : \mathbb{R} \rightarrow L$ such that $c(t)$ and $a(t)$ remain at bounded distance over $t \geq 0$.*

We will need the following lemma in the next section.

LEMMA A.2. *Fix any $x \in \mathbb{H}$. Two hyperbolic translations $f, g : \mathbb{H} \rightarrow \mathbb{H}$ have the same fixed point at $+\infty$ if and only if they satisfy the following for some $D > 0$: there exist arbitrarily large values of $i, j \geq 0$ for which $f^i(x)$ and $g^j(x)$ are at distance less than D .*

Proof. Let $\ell \in \partial\mathbb{H}$ and $\ell' \in \partial\mathbb{H}$ be the fixed point at $+\infty$ of respectively f and g .

First assume the existence of some $D > 0$ such that there exist arbitrarily large values of $i, j \geq 0$ for which $f^i(x)$ and $g^j(x)$ are at distance less than D . To prove $\ell = \ell'$ we consider any open set O of $\bar{\mathbb{H}}$ that contains ℓ and we claim the existence of $i_0 \geq 0$ such that for every $i \geq i_0$ the ball of radius D centered at $f^i(x)$ is contained in O . This claim, combined with our assumption, implies that there exist arbitrarily large values of $j \geq 0$ for which $g^j(x) \in O$. Since $g^j(x)$ tends to ℓ' as x goes to $+\infty$, we have $\ell = \ell'$.

To prove this claim, and without loss of generality, we assume that O belongs to the basis described above to define the topology of $\bar{\mathbb{H}}$. Then, and since $\ell \in O$, there is an open half plane P of \mathbb{H} such that $O = U_P$. Parameterize the axis of f by some unit speed geodesic $a : \mathbb{R} \rightarrow \mathbb{H}$. By definition (up to reversing a) there is $\alpha > 0$ such that $f \circ a(t) = a(t + \alpha)$ on every $t \in \mathbb{R}$. Since f is an isometry the distance between $f^i(x)$ and $a(i\alpha)$ remains constant over $i \in \mathbb{Z}$. By construction a eventually ends up in P . Let L be the geodesic line that bounds P in \mathbb{H} . If the distance between $a(t)$ and L goes to infinity as $t \in \mathbb{R}$ goes to $+\infty$, then the claim is proved. Otherwise this distance goes to zero by Lemma A.1 and there is some unit speed parameterization $c : \mathbb{R} \rightarrow \mathbb{H}$ of L such that $c(t)$ and $a(t)$ remain at bounded distance over $t \geq 0$. Thus $c(t)$ tends to ℓ as $t \in \mathbb{R}$ goes to $+\infty$, contradicting the fact that $c(t) \notin P$ for every $t \in \mathbb{R}$.

Conversely, assume $\ell = \ell'$. Consider unit speed parametrizations $a : \mathbb{R} \rightarrow \mathbb{H}$ and $b : \mathbb{R} \rightarrow \mathbb{H}$ of their respective axes such that $a(t)$ and $b(t)$ tend to ℓ as t goes to $+\infty$. By definition of the limit points, there exists $D > 0$ such that for every $t \geq 0$ the distance between $a(t)$ and $b(t)$ is less than D . Let $\alpha > 0$ and $\beta > 0$ be the translation lengths of respectively f and g on their axes. There exists $D' > 0$ for which there exist $i, j \geq 0$ arbitrarily large such that $|\alpha i - \beta j| < D'$ and thus such that $a(\alpha i)$ and $b(\beta j)$ are at distance less than $D + D'$. Moreover the distance between $a(\alpha i) = f^i(a(0))$ and $f^i(x)$ does not depend on i since f is an isometry. The same holds for the distance between $b(\beta j)$ and $g^j(x)$. \square

A.2 Limit points of lifts of closed curves Every orientable surface \mathcal{S} of genus at least two without boundary is homeomorphic to the quotient of the hyperbolic plane \mathbb{H} by the action of some (actually, many) group Γ of isometries of \mathbb{H} . The elements of Γ other than the identity are hyperbolic translations [26, p. 22]. The action is free, in the sense that if $f \in \Gamma$ satisfies $f(x) = x$ on some $x \in \mathbb{H}$, then f is the identity. The action is also properly discontinuous in the sense that every $x \in \mathbb{H}$ admits a neighborhood whose intersection with the Γ -orbit of x is $\{x\}$. The surface \mathcal{S} then admits a unique hyperbolic metric for which the quotient map $\mathbb{H} \rightarrow \mathcal{S}$ is a local isometry. Also, the hyperbolic plane \mathbb{H} is a universal covering space of \mathcal{S} , where the quotient map $\mathbb{H} \rightarrow \mathcal{S}$ is the covering map.

In this appendix, when we write a surface \mathcal{S} as a quotient \mathbb{H}/Γ , we refer to the construction presented in the previous paragraph. We shall prove the four lemmas of Section 4.2. First we need two preliminary lemmas.

LEMMA A.3. *Let $\mathcal{S} = \mathbb{H}/\Gamma$. Consider a lift $\tilde{c} : \mathbb{R} \rightarrow \mathbb{H}$ of a non-contractible closed curve on \mathcal{S} . There is $f \in \Gamma \setminus \{1\}$ such that $\tilde{c}(t + 1) = f(\tilde{c}(t))$ on every $t \in \mathbb{R}$. Moreover $\lim_{+\infty} \tilde{c}$ and $\lim_{-\infty} \tilde{c}$ exist and are the fixed points of f in $\partial\mathbb{H}$, at respectively $+\infty$ and $-\infty$.*

Proof. For every $t \in \mathbb{R}$, there exists some $f_t \in \Gamma$ such that $f_t(\tilde{c}(t)) = \tilde{c}(t + 1)$. Moreover every such f_t is not the identity as $\tilde{c}(t) \neq \tilde{c}(t + 1)$ since \tilde{c} is a lift of a non-contractible closed curve. We claim that f_t does not depend

on t . This claim concludes the proof. We prove the claim by contradiction so assume the existence of some $t \in \mathbb{R}$ fixed and of some $t' \in \mathbb{R}$ arbitrarily close to t such that $f_t \neq f_{t'}$. By choosing t' close enough to t we make the distance between $f_{t'}^{-1} \circ f_t(\tilde{c}(t))$ and $\tilde{c}(t)$ go to zero, contradicting the fact that Γ acts properly discontinuously on \mathbb{H} . \square

LEMMA A.4. *Let $\mathcal{S} = \mathbb{H}/\Gamma$. Assume that $f, g \in \Gamma \setminus \{1\}$ have the same fixed point at $+\infty$. There are $h \in \Gamma \setminus \{1\}$ and $n, m \geq 1$ such that $f = h^n$ and $g = h^m$. In particular, they have the same fixed point at $-\infty$.*

Proof. Consider some arbitrary fixed $x \in \mathbb{H}$. We claim the existence of $a, b \geq 1$ such that $f^a(x) = g^b(x)$. Indeed by Lemma A.2 there is $D > 0$ that satisfies the following. There are $i, j \geq 0$ arbitrarily large such that $f^i(x)$ and $g^j(x)$ are at distance less than D . For every such i, j the point $f^{-i} \circ g^j(x)$ belongs to the closed ball B_x of radius D centered at x . Since Γ acts properly discontinuously on \mathbb{H} the Γ -orbit $\Gamma \cdot x$ of x intersects B_x in finitely many points. Indeed every such point $y \in B_x \cap \Gamma \cdot x$ admits a neighborhood whose intersection with $\Gamma \cdot x$ is $\{y\}$, and finitely many such neighborhoods suffice to cover the compact ball B_x . In particular there exist $i, j \geq 0$, $i' > i$ and $j' > j$ such that $f^{-i'} \circ g^{j'}(x) = f^{-i} \circ g^j(x)$. Letting $a := i' - i \geq 1$ and $b := j' - j \geq 1$ proves the claim.

Our claim implies that f and g have the same fixed points both at $+\infty$ and $-\infty$ and thus the same axis, say A , oriented from $-\infty$ to $+\infty$. Without loss of generality we assume that x was chosen so that $x \in A$. Recall that f and g are real translations on A and let $\alpha > 0$ and $\beta > 0$ be their respective periods.

We consider the hyperbolic translation h whose oriented axis is A and whose period γ is defined as follows. We proved $a\alpha = b\beta$ since $f^a(x) = g^b(x)$ and since $f^a(x)$ and $g^b(x)$ are the translations of x along A by a distance of respectively $a\alpha$ and $b\beta$. There are $n, m \geq 1$ relatively prime such that $na = mb$. By Bézout's theorem, there are $u, v \in \mathbb{Z}$ such that $un + vm = 1$. We let $\gamma = u\alpha + v\beta$.

We claim that $h = f^u \circ g^v$, and thus $h \in \Gamma$. To prove this claim observe that the point $h(x)$ is the translation of the point x by a distance γ along the oriented axis A of h . The point $g^v(x)$ is the translation of x by a distance $v\beta$ along the same oriented axis A , and the point $f^u(g^v(x))$ is the translation of $g^v(x)$ by a distance of $u\alpha$ along A . Thus $h(x) = f^u \circ g^v(x)$. That proves the claim since hyperbolic translations are defined by their axis, here A , and by the image of any point on this axis, here x .

In the same way we have $h^n = f$ and $h^m = g$ since these hyperbolic translations have the same oriented axis by construction, and since their periods satisfy $n\gamma = \alpha$ and $m\gamma = \beta$, as seen by a straightforward computation. \square

We can now provide the omitted proofs of the lemmas in Section 4.2, restated for convenience. From now on we fix a surface $\mathcal{S} = \mathbb{H}/\Gamma$. The universal cover $\tilde{\mathcal{S}}$ of \mathcal{S} is thus identified with \mathbb{H} .

LEMMA 4.2. *If $\tilde{c} : \mathbb{R} \rightarrow \tilde{\mathcal{S}}$ is a lift of a non-contractible closed curve on \mathcal{S} , then $\lim_{+\infty} \tilde{c}$ and $\lim_{-\infty} \tilde{c}$ exist (in $\tilde{\mathcal{S}} \cup \partial\tilde{\mathcal{S}}$) and are distinct points of $\partial\tilde{\mathcal{S}}$.*

Proof of Lemma 4.2. The result is given by Lemma A.3. \square

LEMMA 4.3. *Lift a homotopy $c \simeq d$ between non-contractible closed curves on \mathcal{S} to a homotopy $\tilde{c} \simeq \tilde{d}$ between lifts of c and d . Then \tilde{c} and \tilde{d} have the same limit points.*

Proof of Lemma 4.3. Consider the hyperbolic translations, say f and g , given by Lemma A.3 for \tilde{c} and \tilde{d} respectively. Since the homotopy $\tilde{c} \simeq \tilde{d}$ lifts the homotopy $c \simeq d$, the distance between the points $\tilde{c}(k) = f^k(\tilde{c}(0))$ and $\tilde{d}(k) = g^k(\tilde{d}(0))$ does not depend on $k \in \mathbb{Z}$. By Lemma A.2 f and g have the same fixed points. \square

LEMMA 4.4. *Consider lifts \tilde{c} and \tilde{d} of non-contractible closed curves on \mathcal{S} . Assume either that \tilde{c} and \tilde{d} intersect exactly once, or that they are disjoint lifts of the same curve on \mathcal{S} . Then the four limit points of \tilde{c} and \tilde{d} are pairwise distinct.*

Proof of Lemma 4.4. By Lemma 4.2, the two limit points of \tilde{c} are distinct, and similarly for \tilde{d} . Assume, for a contradiction, that \tilde{c} and \tilde{d} have the same limit point at $+\infty$ (up to reversing c or d). There are two cases.

First assume that \tilde{c} and \tilde{d} intersect exactly once in some point $x \in \mathbb{H}$. Without loss of generality assume $\tilde{c}(0) = \tilde{d}(0) = x$. Consider the hyperbolic translations, say f and g , given by Lemma A.3 for \tilde{c} and \tilde{d} respectively. Then f and g have the same fixed point at $+\infty$. Thus by Lemma A.4 there exist $i, j \geq 1$ such that $f^i = g^j$. Then

$\tilde{c}(i) = f^i(x) = g^j(x) = \tilde{d}(j)$ and this point is distinct from x since, for example, \tilde{c} is non-contractible. This is a contradiction.

Now assume that \tilde{c} and \tilde{d} are disjoint lifts of the same curve c in \mathcal{S} . There is a geodesic closed curve α homotopic to c [26, Proposition 1.3]. Lift the homotopy $c \simeq \alpha$ to homotopies $\tilde{c} \simeq \tilde{\alpha}$ and $\tilde{d} \simeq \tilde{\beta}$ for some lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of α . By the preceding lemmas, $\tilde{\alpha}$ and $\tilde{\beta}$ have the same limit points. Thus $\tilde{\alpha}$ and $\tilde{\beta}$ have the same image (a geodesic line), so they are equal up to homeomorphism $\mathbb{R} \rightarrow \mathbb{R}$. By the uniqueness part of the lifting property, \tilde{c} and \tilde{d} are then equal up to homeomorphism $\mathbb{R} \rightarrow \mathbb{R}$. We proved that \tilde{c} and \tilde{d} intersect, a contradiction. \square

LEMMA 4.5. *Let c and d be two non-contractible closed curves on \mathcal{S} . If c and d admit lifts with the same pairs of limit points, there is a closed curve e such that c and d are homotopic to powers of e .*

Proof of Lemma 4.5. Let \tilde{c} and \tilde{d} be lifts of respectively c and d and assume that \tilde{c} and \tilde{d} have the same limit points. In a first step, if \tilde{c} and \tilde{d} are disjoint, we apply a homotopy of c to make \tilde{c} intersect \tilde{d} ; this does not change the limit points of \tilde{c} by Lemma 4.3.

So we can assume without loss of generality that \tilde{c} and \tilde{d} intersect, and thus also that $\tilde{c}(0) = \tilde{d}(0) = x \in \mathbb{H}$. Consider the hyperbolic translations, say f and g , given by Lemma A.3 for \tilde{c} and \tilde{d} respectively. By Lemma A.4 there exist $n, m \geq 1$ and $h \in \Gamma \setminus \{1\}$ such that $f = h^n$ and $g = h^m$. Thus $f(x) = h^n(x)$ and $g(x) = h^m(x)$. Let ℓ be a loop on \mathcal{S} that lifts to a path from x to $h(x)$. Based at zero, c and d are homotopic as loops to respectively the n th power and the m th power of the loop ℓ . \square