Discrete Comput Geom 6:527–574 (1991)

Discrete & Computational
Geometry

Disjoint Homotopic Paths and Trees in a Planar Graph

A. Schrijver

Mathematical Centre, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands, and
University of Amsterdam, Amsterdam, The Netherlands

Abstract. In this paper we describe a polynomial-time algorithm for the following problem: *given*: a planar graph G embedded in $\mathbb{R}^2$, a subset $\{I_1, \dots, I_p\}$ of the faces of G, and paths $C_1, \dots, C_k$ in G, with endpoints on the boundary of $I_1 \cup \dots \cup I_p$; *find*: pairwise disjoint simple paths $P_1, \dots, P_k$ in G so that, for each $i = 1, \dots, k$, P_i is homotopic to C_i in the space $\mathbb{R}^2 \backslash (I_1 \cup \dots \cup I_p)$.

Moreover, we prove a theorem characterizing the existence of a solution to this problem. Finally, we extend the algorithm to disjoint homotopic trees. As a corollary we derive that, for each fixed p, there exists a polynomial-time algorithm for the problem: *given*: a planar graph G embedded in $\mathbb{R}^2$ and pairwise disjoint sets $W_1, \dots, W_k$ of vertices, which can be covered by the boundaries of at most p faces of G; *find*: pairwise vertex-disjoint subtrees $T_1, \dots, T_k$ of G where T_i covers W_i ($i = 1, \dots, k$).

1. Introduction

In this paper we describe a polynomial-time algorithm for the following *disjoint homotopic paths problem*:

given: a planar graph G embedded in the plane $\mathbb{R}^2$;
a subset $I_1, \dots, I_p$ of the faces of G (including the unbounded face);
paths $C_1, \dots, C_k$ in G, each with endpoints on the boundary of $I_1 \cup \dots \cup I_p$; (1.1)
find: pairwise disjoint simple paths $P_1, \dots, P_k$ in G so that, for each $i = 1, \dots, k$, P_i is homotopic to C_i in the space $\mathbb{R}^2 \backslash (I_1 \cup \dots \cup I_p)$.

We explain the terminology used here. By *embedding* we mean embedding without intersecting edges and with piecewise linear edges. We identify G with its image in $\mathbb{R}^2$. We consider edges as *open* curves (i.e., without endpoints) and faces as *open* subsets of $\mathbb{R}^2$.

Two curves C, $\tilde{C}$: $[0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ are called *homotopic in* $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ (in notation: $C \sim \tilde{C}$) if there exists a continuous function Φ: $[0, 1] \times [0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ so that

$$\Phi(0, x) = C(x), \qquad \Phi(1, x) = \tilde{C}(x), \qquad \Phi(x, 0) = C(0), \qquad \Phi(x, 1) = C(1) \tag{1.2}$$

for each $x \in [0, 1]$. (It implies that $C(0) = \tilde{C}(0)$ and $C(1) = \tilde{C}(1)$.) In this paper, by just *homotopic* we mean homotopic in $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$.

A *path* is a sequence $(v_0, e_1, v_1, \ldots, e_d, v_d)$ of not necessarily distinct vertices and edges, so that e_j connects v_{j-1} and v_j $(j = 1, \ldots, d)$. It is *simple* if $v_0, v_1, \ldots, v_d$ are all distinct. Vertices v_0 and v_d are called the *endpoints* of the paths. By identifying paths in G with curves in $\mathbb{R}^2$, homotopy extends to paths in G.

Thus we prove (in Section 3):

Theorem 1. *The disjoint homotopic paths problem* (1.1) *is solvable in polynomial time.*

The algorithm also yields the basis of a proof of the following theorem (Section 5) characterizing the existence of a solution to the disjoint homotopic paths problem (1.1) by means of "cut conditions" (conjectured by L. Lovász and P. D. Seymour):

Theorem 2. *Problem* (1.1) *has a solution if and only if*:

(i) *there exist pairwise disjoint simple curves* $\tilde{C}_1, \ldots, \tilde{C}_k$ *in* $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ *so that* $\tilde{C}_i$ *is homotopic to* C_i $(i = 1, \ldots, k)$;

(ii) *for each curve* D: $[0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ *with* $D(0)$, $D(1) \in \mathrm{bd}(I_1 \cup \cdots \cup I_p)$ *we have*

$$\mathrm{cr}(G, D) \geq \sum_{i=1}^{k} \mathrm{mincr}(C_i, D); \tag{1.3}$$

(iii) *for each doubly odd closed curve* D: $S_1 \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ *we have*

$$\mathrm{cr}(G, D) > \sum_{i=1}^{k} \mathrm{mincr}(C_i, D).$$

Here bd denotes boundary. For curves C, D: $[0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ we define

$$\begin{aligned} \mathrm{cr}(G, D) &:= |\{y \in [0, 1] \mid D(y) \in G\}|, \\ \mathrm{cr}(C, D) &:= |\{(x, y) \in [0, 1] \times [0, 1] \mid C(x) = D(y)\}|, \\ \mathrm{mincr}(C, D) &:= \min\{\mathrm{cr}(\tilde{C}, \tilde{D}) \mid \tilde{C} \sim C, \tilde{D} \sim D\}. \end{aligned} \tag{1.4}$$

(We take $c(G, D) := 1$ if D is a constant function with $D(0) \in G$.)

A *closed curve* is a continuous function $D: S_1 \to \mathbb{R}^2$ (where S_1 denotes the unit circle in $\mathbb{C}$). Two closed curves $D, \tilde{D}: S_1 \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ are called *freely* *homotopic in* $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$, or just *homotopic* (in notation: $D \sim \tilde{D}$), if there exists a continuous function $\Phi: [0, 1] \times S_1 \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ so that

$$\Phi(0, z) = D(z), \qquad \Phi(1, z) = \tilde{D}(z) \tag{1.5}$$

for all $z \in S_1$. (So there is no fixed point.) Again we denote (if $C: [0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ is a curve)

$$\begin{aligned} \operatorname{cr}(G, D) &:= |\{z \in S_1 \mid D(z) \in G\}|, \\ \operatorname{cr}(C, D) &:= |\{(x, z) \in [0, 1] \times S_1 \mid C(x) = D(z)\}|, \\ \operatorname{mincr}(C, D) &:= \min\{\operatorname{cr}(\tilde{C}, \tilde{D}) \mid \tilde{C} \sim C, \tilde{D} \sim D\}. \end{aligned} \tag{1.6}$$

If $D', D'': S_1 \to \mathbb{R}^2$ are closed curves with $D'(1) = D''(1)$, then the concatenation $D' \cdot D''$ (or just $D'D''$) is the closed curve given by

$$\begin{aligned} D' \cdot D''(z) &:= D'(z^2) \quad \text{if} \quad \operatorname{Im} z \geq 0, \\ &:= D''(z^2) \quad \text{if} \quad \operatorname{Im} z < 0. \end{aligned} \tag{1.7}$$

Call a point p a *fixed point* of a curve C if each curve homotopic to C traverses p. (In particular, the endpoints of C are fixed points of C.) A closed curve D is called *doubly odd* if:

(i) D does not traverse any fixed point of any $C_1, \ldots, C_k$;
(ii) $D = D' \cdot D''$ for some closed curves D', D'' with $D'(1) = D''(1) \notin G$ so that

$$\operatorname{cr}(G, D') + \sum_{i=1}^{k} \operatorname{kr}(C_i, D') \text{ is odd}$$

and (1.8)

$$\operatorname{cr}(G, D'') + \sum_{i=1}^{k} \operatorname{kr}(C_i, D'') \text{ is odd.}$$

Here $\operatorname{kr}(C, D)$ denotes the number of crossings of C and D (see Fig. 1).

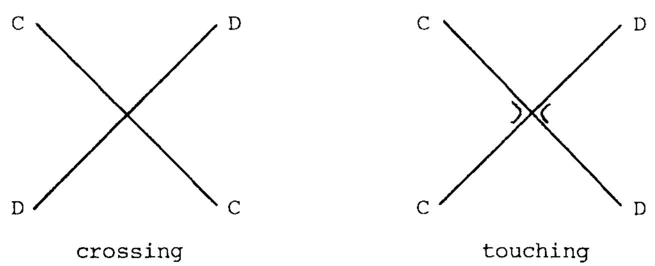

Fig. 1

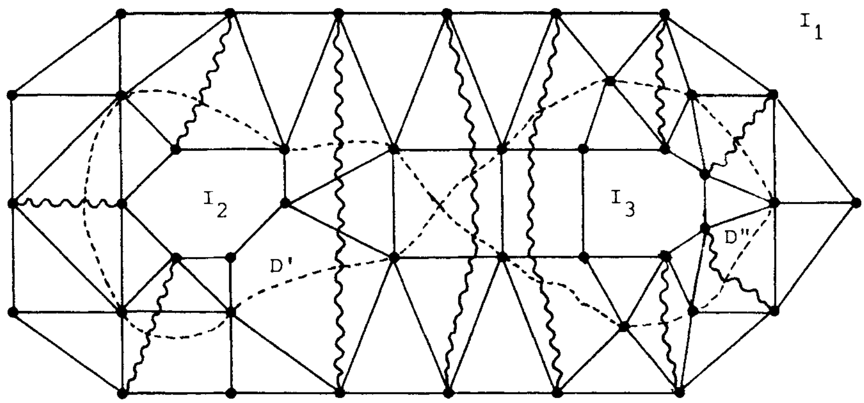

Fig. 2. Wavy curves represent $C_1, \ldots, C_{10}$ and the dashed curve represents the doubly odd closed curve D. Now $\mathrm{cr}(G, D') + \sum_{i=1}^{10} \mathrm{kr}(C_i, D') = 6 + 5 = 11$ and $\mathrm{cr}(G, D'') + \sum_{i=1}^{10} \mathrm{kr}(C_i, D'') = 7 + 8 = 15$, whereas $\mathrm{cr}(G, D) = 13 = \sum_{i=1}^{10} \mathrm{mincr}(C_i, D)$. So condition (1.3(iii)) is not satisfied, and hence (1.1) has no solution.

To clarify condition (1.3), we give a proof of necessity (see Fig. 2).

Proof of necessity of condition (1.3). Suppose problem (1.1) has a solution $P_1, \ldots, P_k$. Then condition (1.3(i)) is satisfied as we can take $\tilde{C}_i := P_i$ for $i = 1, \ldots, k$. Condition (1.3(ii)) follows from

$$\mathrm{cr}(G, D) \geq \sum_{i=1}^{k} \mathrm{cr}(P_i, D) \geq \sum_{i=1}^{k} \mathrm{mincr}(P_i, D) = \sum_{i=1}^{k} \mathrm{mincr}(C_i, D) \tag{1.9}$$

(the first inequality follows from the fact that the P_i are simple and disjoint).

To see condition (1.3(iii)), note that

$$\mathrm{cr}(G, D') \geq \sum_{i=1}^{k} \mathrm{cr}(P_i, D') \geq \sum_{i=1}^{k} \mathrm{kr}(P_i, D')$$

and (1.10)

$$\mathrm{cr}(G, D'') \geq \sum_{i=1}^{k} \mathrm{cr}(P_i, D'') \geq \sum_{i=1}^{k} \mathrm{kr}(P_i, D'').$$

Moreover, since the parity of $\mathrm{kr}(\cdot, \cdot)$ is invariant under homotopy, we have by (1.8(ii))

$$\begin{aligned} \mathrm{cr}(G, D') &\not\equiv \sum_{i=1}^{k} \mathrm{kr}(C_i, D') \equiv \sum_{i=1}^{k} \mathrm{kr}(P_i, D') \pmod 2, \\ \mathrm{cr}(G, D'') &\not\equiv \sum_{i=1}^{k} \mathrm{kr}(C_i, D'') \equiv \sum_{i=1}^{k} \mathrm{kr}(P_i, D'') \pmod 2. \end{aligned} \tag{1.11}$$

So we derive the following strict inequalities from (1.10):

$$\mathrm{cr}(G, D') > \sum_{i=1}^{k} \mathrm{kr}(P_i, D') \quad \text{and} \quad \mathrm{cr}(G, D'') > \sum_{i=1}^{k} \mathrm{kr}(P_i, D''). \tag{1.12}$$

Concluding,

$$\begin{aligned} \mathrm{cr}(G, D) = \mathrm{cr}(G, D') + \mathrm{cr}(G, D'') &> \sum_{i=1}^{k} (\mathrm{kr}(P_i, D') + \mathrm{kr}(P_i, D'')) \\ = \sum_{i=1}^{k} \mathrm{kr}(P_i, D) \geq \sum_{i=1}^{k} \mathrm{mincr}(P_i, D) &= \sum_{i=1}^{k} \mathrm{mincr}(C_i, D). \end{aligned} \tag{1.13}$$

(The last inequality follows from the fact that D does not traverse any fixed point of any C_i, so that any touching of D and P_i can be removed.) Therefore we have the strict inequality in (1.3(iii)). □

In Section 6 we describe a polynomial-time algorithm for the following *disjoint homotopic trees problem*, generalizing the disjoint homotopic paths problem (1.1):

given: a planar graph G embedded in $\mathbb{R}^2$;
a subset $I_1, \ldots, I_p$ of the faces of G (including the unbounded face);
paths $C_{11}, \ldots, C_{1t_1}, \ldots, C_{k1}, \ldots, C_{kt_k}$ in G, each with endpoints on the boundary of $I_1 \cup \cdots \cup I_p$, so that, for each $i = 1, \ldots, k$, $C_{i1}, \ldots, C_{it_i}$ have the same beginning vertex;
find: pairwise disjoint subtrees $T_1, \ldots, T_k$ of G so that, for each $i = 1, \ldots, k$ and $j = 1, \ldots, t_i$, T_i contains a path homotopic to C_{ij} in $\mathbb{R}^2 \setminus (I_1 \cup \cdots \cup I_p)$.

Theorem 3. *The disjoint homotopic trees problem* (1.14) *is solvable in polynomial time.*

Theorem 3 generalizes Theorem 1, since if $t_1 = \cdots = t_k = 1$, then problem (1.14) reduces to problem (1.1). However, for the sake of exposition we first restrict ourselves to studying problem (1.1). The algorithm for (1.14) arises from that for (1.1) by some direct modifications.

We do not formulate a theorem characterizing the existence of a solution to (1.14), analogous to Theorem 2, as we found only tedious inattractive conditions. Obviously, the fact that (1.14) is solvable in polynomial time implies that it has a "good characterization" (i.e., belongs to NP $\cap$ co-NP).

Finally, in Section 7 we consider the *disjoint trees problem*:

given: a graph G;
subsets $W_1, \ldots, W_k$ of $V(G)$;
find: pairwise disjoint subtrees $T_1, \ldots, T_k$ of G so that $W_i \subseteq V(T_i)$ for $i = 1, \ldots, k$. (1.15)

This problem is NP-complete. Robertson and Seymour showed that, for fixed $|W_1 \cup \cdots \cup W_k|$, problem (1.15) is solvable in polynomial time. We derive from Theorem 3 that if G is planar this can be extended to:

Theorem 4. *For each fixed p there exists a polynomial-time algorithm for the disjoint trees problem* (1.15) *when G is planar and $W_1 \cup \cdots \cup W_k$ can be covered by the boundaries of p faces of G.*

The reduction to Theorem 3 is based on enumerating homotopy classes of trees, taking the p faces as "holes."

Motivation for studying problems (1.1), (1.14), and (1.15) comes from two different sources. First, in their series of papers "Graph Minors," Robertson and Seymour study problem (1.1) for the case where $p = 1$ or 2 [6]. Moreover, they study a variant of problem (1.14) for graphs densely enough embedded on a compact surface [7], [8].

A second source of motivation is the design of very large-scale integrated (VLSI) circuits, where it is wished to interconnect sets of pins by disjoint sets of wires. Pinter [5] described a topological model for solving so-called "river-routing" problems. In consequence, Cole and Siegel [1] and Leiserson and Maley [3] proved the theorem above and gave a polynomial-time algorithm, respectively, for problem (1.1) in case G is part of the rectangular grid on $\mathbb{R}^2$, provided that each face not surrounded by exactly four edges belongs to $\{I_1, \ldots, I_p\}$ (then (1.3(iii)) is superfluous).

The algorithm for (1.1) is purely combinatorial. In [2] we described a polynomial-time algorithm for (1.1) based on the ellipsoid method (first a fractional solution to (1.1) is found with the ellipsoid method, next this fractional solution is "uncrossed," from which a solution to (1.1) is derived). The present algorithm extends to disjoint trees.

Another related result was published in [9], showing the necessity and sufficiency of conditions analogous to (1.3) for the existence of circuits of prescribed homotopy in a graph embedded on a compact surface. With some effort we may derive from this Theorem 2 above, by transforming the space $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ to a compact closed surface, by adding some "handles" between the "holes" $I_1, \ldots, I_p$, and by extending the graph and the curves over these handles.

Note 1.1. Analysis of our method would yield a running time bound of order $O(n^4 \log^2 n)$, where n is the number of vertices + edges of G, added with the lengths of the paths in the input. We do not however derive this bound. In fact, we conjecture that a sharpening of our methods gives a running time of order $O(n^2 \log^2 n)$.

Note 1.2. To apply the algorithm, it is not necessary to describe the embedding of G in $\mathbb{R}^2$. It suffices to specify the vertices, edges, and faces of G abstractly, and to give with each vertex and with each face the edges incident with it in clockwise orientation.

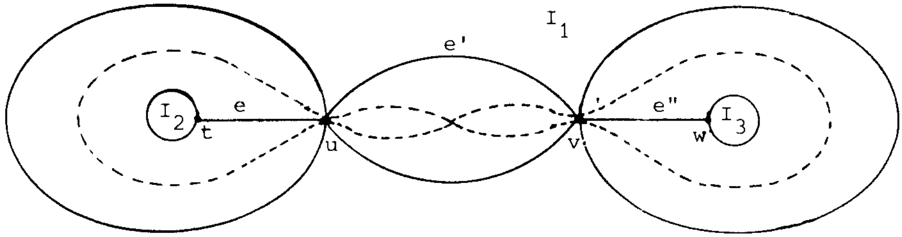

Fig. 3

Note 1.3. If we would delete from the definition of "double odd," condition (1.8(i)) that D should not traverse any fixed point of any C_i, condition (1.3(iii)) would not be a necessary condition. This is shown by the example in Fig. 3.

The graph in Fig. 3 has four vertices: t, u, v, w, with a loop attached at each of them, edges connecting t and u, and v and w, and two parallel edges connecting u and v. Let C_1 be the path from t to w following edges e, e', e''. So problem (1.1) has a solution (taking $k = 1$). Let D be the closed curve indicated by the dashed curve. D traverses the fixed points u and v of C_1. We easily check that D satisfies (1.8(ii)), but not the strict inequality in (1.3(iii)) (since $\mathrm{cr}(G, D) = 4 = \mathrm{mincr}(C_1, D)$).

2. The Universal Covering Space and Shortest Homotopic Paths

Before describing our method in Section 3, in this section we first discuss briefly the concept of universal covering space, and we describe a polynomial-time algorithm for finding a shortest path of given homotopy. One consequence of this algorithm is that we can check in polynomial time whether two given paths are homotopic. For background literature on the universal covering space, see Massey [4].

The *universal covering space* U of $\mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ can be defined set-theoretically as follows. Choose a point $u \in \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$. The underlying point set of U is the set of all homotopy classes of curves starting in u. A set $T \subseteq U$ is open if and only if, for each $\mu \in T$, say $\mu \in \mathrm{hom}(u, w)$, there exists a neighborhood N of w in $\mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ so that if P is a curve contained in N starting in w, then $\mu \cdot \langle P \rangle \in T$. [Here $\mathrm{hom}(u, w)$ denotes the collection of all homotopy classes of curves from u to w, and $\langle P \rangle$ denotes the homotopy class containing P.]

It is not difficult to see that the universal covering space is independent (up to homeomorphism) of the choice of u. With the universal covering space U a *projection function* π: $U \to \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ is given by $\pi(\mu) := w$ if $\mu \in \mathrm{hom}(u, w)$.

There is an alternative, combinatorial way of describing U. We can "cut open" $\mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ along $p - 1$ pairwise noncrossing simple curves, connecting the "holes" $I_1, \ldots, I_p$, in such a way that we obtain a simply connected region R, e.g., Fig. 4 becomes Fig. 5. We can deform R to a disk as in Fig. 6. If two of the I_j touch each other, we can obtain a region with cut points.

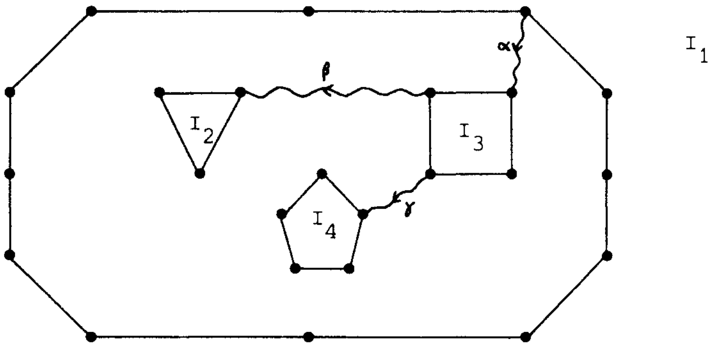

Fig. 4

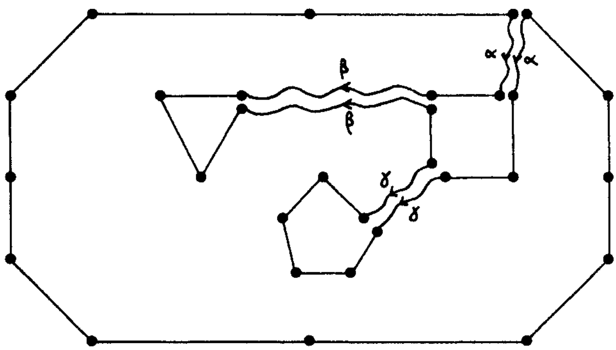

Fig. 5

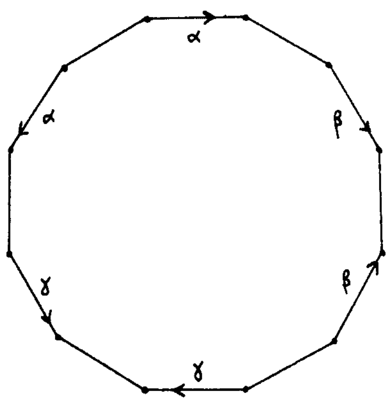

Fig. 6

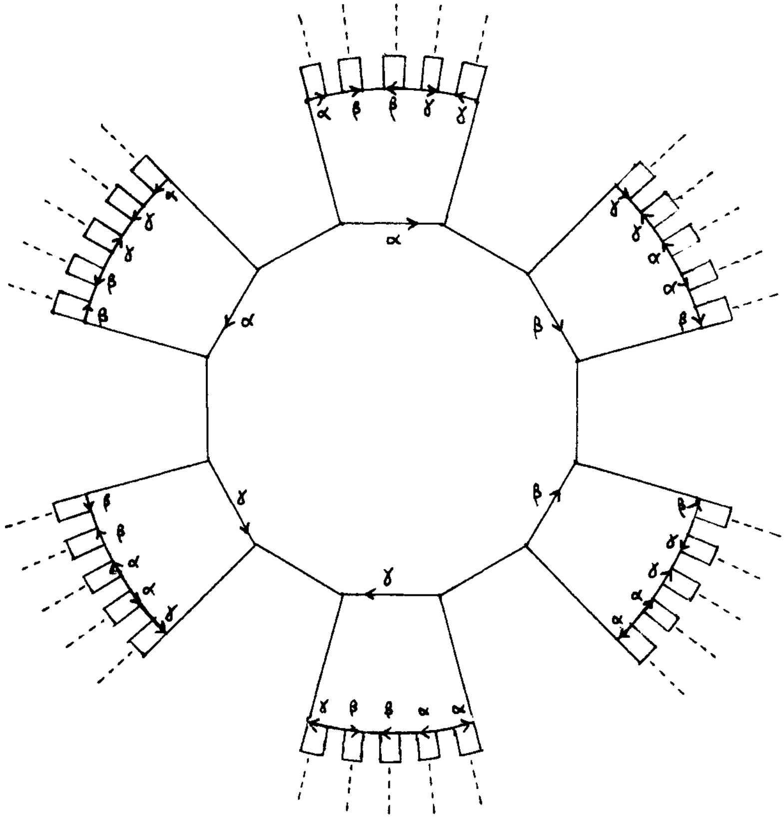

Fig. 7

We now take infinitely many copies of R, and glue them together along the cuts, in such a way that we obtain a simply connected space (see Fig. 7). This gives us the universal covering space U of $\mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$, with obvious projection function $\pi: U \to \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$.

The inverse image $G' := \pi^{-1}[G]$ of G is an infinite graph, embedded in U (assuming $p \geq 2$ here, the case $p = 1$ being trivial). In fact, G' is planar, and U can be identified with $\mathbb{R}^2\backslash\bigcup_{F\in\mathscr{F}} F$, where $\mathscr{F}$ is the collection of unbounded faces of G' (assuming G to be connected).

It is a fundamental property of the universal covering space that, for each curve $C: [0, 1] \to \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ and each choice of $v \in \pi^{-1}(C(0))$, there exists a unique curve $C': [0, 1] \to U$ satisfying $\pi \circ C' = C$ and $C'(0) = v$. Curve C' is called a *lifting* of C to U. Two curves $C, \tilde{C}: [0, 1] \to \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ are homotopic if and only if some lifting of C to U has the same endpoints as some lifting of $\tilde{C}$ to U. A point $u \in \mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$ is a fixed point of curve C if and only if, for some $u' \in \pi^{-1}(u)$ and some lifting C' of C to U, each curve in U connecting $C'(0)$ and $C'(1)$ traverses u'.

We now turn to the *shortest homotopic path problem*:

given: a planar graph $G = (V, E)$ embedded in $\mathbb{R}^2$;
a subset $\{I_1, \dots, I_p\}$ of the faces of G (including the unbounded face);
a path P in G; (2.1)
a "length" function $l\colon E \to \mathbb{Z}_+$;
find: a path $\tilde{P}$ in G homotopic to P in $\mathbb{R}^2 \backslash (I_1 \cup \dots \cup I_p)$ minimizing length $(\tilde{P})$.

Here by length($\tilde{P}$) we mean, if $\tilde{P} = (v_0, e_1, v_1, \dots, e_d, v_d)$,

$$\text{length}(\tilde{P}) := \sum_{i=1}^{d} l(e_i). \tag{2.2}$$

We do not require that $\tilde{P}$ is simple in (2.1).

To solve (2.1), consider a lifting P' of P to U. So P' is a path in G', say from u to w. Then, clearly, if Q is a shortest path in G' from u to w, then its projection $\pi \circ Q$ is a valid output for (2.1). (Taking the obvious length function on the edges of G'.)

Hence, the shortest homotopic path problem in G can be reduced to the shortest (nonhomotopic) path problem in G'. This would give us an algorithm if G' were not an infinite graph. However, it is clearly not necessary to consider G' completely. In fact, it suffices to consider a part of G' of polynomially bounded size, which implies that (2.1) is solvable in polynomial time.

To see this, we may assume that when cutting $\mathbb{R}^2 \backslash (I_1 \cup \dots \cup I_p)$ open to obtain the region R, we have done this along shortest paths in G. In fact, we can find shortest paths $Q_2, \dots, Q_p$ in G, where Q_j connects I_1 with I_j, so that $Q_2, \dots, Q_p$ are pairwise edge-disjoint and do not have crossings. (They can be found as follows. Choose vertices $v_1, \dots, v_p$ incident with $I_1, \dots, I_p$, respectively. With Dijkstra's algorithm, find a spanning tree T in G so that all simple path in T starting in v_1 are shortest paths. Let Q_j be the simple path in T from v_1 to v_j (for $j = 2, \dots, p$). Adding parallel edges gives $Q_2, \dots, Q_p$ as required.)

Now any lifting Q'_j of any Q_j to U is a shortest path in G'. So there exists a shortest path in G' from u to w not crossing any Q'_j which does not cross P'. That is, we have to consider only that part of U consisting of copies of R traversed by P'. This gives us a subgraph G'' of G' of size polynomially bounded by the size of G and the number of vertices in P'. For any shortest path Q in G'' from u to w, the path $\tilde{P} := \pi \circ Q$ is a shortest path homotopic to P.

Proposition 1. *The shortest homotopic path problem is solvable in polynomial time.*

Proof. See above. □

A consequence is:

Proposition 2. *It can be tested in polynomial time if two paths P and $\tilde{P}$ in a planar graph are homotopic in $\mathbb{R}^2 \backslash (I_1 \cup \dots \cup I_p)$ (where $I_1, \dots, I_p$ are faces).*

Proof. Paths P and $\tilde{P}$ are homotopic if and only if the shortest path homotopic to path $P \cdot \tilde{P}^{-1}$ has length 0 (where $\tilde{P}^{-1}$ denotes the path reverse to $\tilde{P}$ and taking length $l(e) = 1$ for each edge e). □

There is also a polynomial-time algorithm for finding a shortest path *not* homotopic to a given path. More generally, we have the following result. A *mapping* of a graph G to a space S is a—not necessarily one-to-one—continuous function from G to S. Call two paths in G *homotopic* if their images in S are homotopic in S.

The *shortest nonhomotopic path problem* is:

given: a graph $G = (V, E)$ mapped into a space S;
a path P in G, connecting, say, u and w;
a "length" function $l: E \to \mathbb{Z}_+$; (2.3)
find: a path Q in G from u to w, so that Q is not homotopic to P and so that Q has minimum length.

Proposition 3. *The shortest nonhomotopic path problem is solvable in polynomial time, provided we can decide in polynomial time if any given path Q is homotopic to P.*

[In fact, this last is the only thing we need to know about S and the mapping.]

Proof. First find for each vertex v *of* G a shortest path P_{uv} from u to v and a shortest path P_{vw} from v to w. Consider the following collection of paths in G:

$$\begin{array}{ll} P_{uv} \cdot P_{vw} & (v \in V), \\ P_{uv} \cdot e \cdot P_{v'w} & (e = vv' \in E). \end{array} \tag{2.4}$$

Select those paths Q from (2.4) which are not homotopic to P, and choose among these one of minimum length. We claim that this Q is a valid output for (2.3).

To see this, let

$$R = (u = v_0, e_1, v_1, \dots, e_d, v_d = w) \tag{2.5}$$

be a minimum-length path not homotopic to P. We must show length(Q) $\leq$ length(R).

Choose the largest t so that $P_{uv_t} \cdot (v_t, e_{t+1}, \dots, e_d, v_d)$ is not homotopic to P. Such a t exists, as R itself is not homotopic to P. If $t = d$, then P_{uw} is not homotopic to P. Moreover, $P_{uw} = P_{uw} \cdot P_{ww}$ occurs among (2.4), and hence length(Q) $\leq$ length(P_{uw}) $\leq$ length(R).

If $t < d$, by the maximality of t, path $P_{uv_t} \cdot e_{t+1} \cdot P_{v_{t+1}w}$ is not homotopic to path $P_{uv_{t+1}} \cdot P_{v_{t+1}w}$. Hence at least one of them is not homotopic to P. So one of them has length at least length(Q). On the other hand, each of them has length at most length(R) (since the P_{uv} and P_{vw} are shortest paths). Therefore, length(Q) $\leq$ length(R). □

3. The Method

We describe our method for solving the disjoint homotopic paths problem (1.1). Let input G, $I_1, \ldots, I_p$, $C_1, \ldots, C_k$ be given. The algorithm finds $P_1, \ldots, P_k$ as required, if conditions (1.3) are satisfied. It consists of four basic steps:

I. Uncrossing $C_1, \ldots, C_k$.
II. Constructing a system $Ax \leq b$ of linear inequalities.
III. Solving $Ax \leq b$ in integers.
IV. Shifting $C_1, \ldots, C_k$ (using the integer solution of $Ax \leq b$).

In order to facilitate the description, we make the following assumptions:

(i) each edge of G is traversed at most once by the C_i;
(ii) the endpoints of each C_i have degree 1 in G;
(iii) no edge traversed by any C_i, except for the first and last edge of C_i, is incident with any face in $\{I_1, \ldots, I_p\}$. (3.1)

These conditions can be fulfilled by adding new vertices and (parallel) edges. It follows from (3.1) that the endpoints of each C_i are not traversed by any other $C_1, \ldots, C_k$.

I. Uncrossing $C_l, \ldots, C_k$

This step modifies $C_1, \ldots, C_k$ so that they do not have (self-)crossings or null-homotopic parts. (A *part* is a subcurve.) Choose $i, i' \in \{1, \ldots, k\}$ with $i \neq i'$, and let

$$\begin{aligned} C_i &= (v_0, e_1, v_1, e_2, v_2, \ldots, e_m, v_m), \\ C_{i'} &= (v'_0, e'_1, v'_1, e'_2, v'_2, \ldots, e'_{m'}, v'_{m'}). \end{aligned} \tag{3.2}$$

Consider a pair (j, j') with $1 \leq j \leq m-1$ and $1 \leq j' \leq m'-1$. Call (j, j') a *crossing* if $v_j = v'_{j'}$, and the edges e_j, $e'_{j'}$, e_{j+1}, $e'_{j'+1}$ occur in this order cyclically at v_j, clockwise or anticlockwise (see Fig. 8).

Now there is the following easy proposition:

Proposition 4. *If (1.3(i)) is satisfied and $i \neq i'$, then for any crossing (j, j') of C_i and $C_{i'}$ there exists another crossing (h, h') of C_i and $C_{i'}$ so that*

$$\text{part } (v_j, \ldots, v_h) \text{ of } C_i \text{ is homotopic to part } (v'_{j'}, \ldots, v'_{h'}) \text{ of } C_{i'}. \tag{3.3}$$

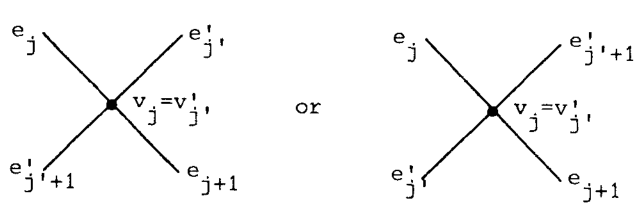

Fig. 8

[By part $(v_j, \ldots, v_h)$ of C_i we mean $(v_j, e_{j+1}, v_{j+1}, \ldots, e_h, v_h)$ if $j \le h$, and $(v_j, e_j, v_{j-1}, \ldots, e_{h+1}, v_h)$ if $j > h$. Similarly for $C_{i'}$.]

Proof. If (j, j') is a crossing of C_i and $C_{i'}$ there exist liftings

$$\bar{C}_i = (\bar{v}_0, \bar{e}_1, \bar{v}_1, \ldots, \bar{e}_m, \bar{v}_m) \quad \text{and} \quad \bar{C}_{i'} = (\bar{v}'_0, \bar{e}'_1, \bar{v}'_1, \ldots, \bar{e}'_{m'}, \bar{v}'_{m'}) \tag{3.4}$$

of C_i and $C_{i'}$ respectively to U so that $\bar{v}_j = \bar{v}'_{j'}$ and $\bar{e}_j, \bar{e}'_{j'}, \bar{e}_{j+1}, \bar{e}'_{j'+1}$ occur in this order cyclically at $\bar{v}_j$. By (1.3(i)) there exist $\tilde{C}_i \sim C_i$ and $\tilde{C}_{i'} \sim C_{i'}$ so that $\tilde{C}_i$ and $\tilde{C}_{i'}$ are disjoint. By considering liftings of $\tilde{C}_i$ and $\tilde{C}_{i'}$ it follows that $\bar{C}_i$ and $\bar{C}_{i'}$ have an even number of crossings. Hence $\bar{C}_i$ and $\bar{C}_{i'}$ must have a second crossing, say at $\bar{v}_h = \bar{v}'_{h'}$. This implies (3.3). □

By Proposition 2 we can test in polynomial time if two paths are homotopic. So if C_i and $C_{i'}$ have a crossing, we can find in polynomial time two distinct crossings (j, j') and (h, h') so that (3.3) holds. We now exchange parts $(v_j, \ldots, v_h)$ of C_i and $(v'_{j'}, \ldots, v'_{h'})$ of $C_{i'}$, e.g., if $j \le h$ and $j' \le h'$, we reset

$$\begin{aligned} C_i &:= (v_0, e_1, \ldots, e_{j-1}, v_j = v'_{j'}, e'_{j'+1}, \ldots, e'_{h'}, v'_{h'} = v_h, e_{h+1}, \ldots, e_m, v_m), \\ C_{i'} &:= (v'_0, e'_1, \ldots, e'_{j'-1}, v'_{j'} = v_j, e_{j+1}, \ldots, e_h, v_h = v'_{h'}, e'_{h'+1}, \ldots, e'_{m'}, v'_{m'}). \end{aligned}$$

This resetting reduces the total number of crossings of C_i and $C_{i'}$ (summing up over all pairs i, i'). Hence after a polynomial number of such modifications we are in the situation that no two distinct $C_i, C_{i'}$ have crossings.

Throughout this uncrossing process we remove null-homotopic parts of any C_i (they can be recognized again by Proposition 2). Since each such removal strictly decreases the total number of edges used by the C_i, this again can be done in polynomial time.

We still have to deal with self-crossings. A *self-crossing* of

$$C_i = (v_0, e_1, v_1, \ldots, e_m, v_m) \tag{3.5}$$

is a pair (j, j') with $j \ne j'$ and $v_j = v_{j'}$ so that $e_j, e_{j'}, e_{j+1}, e_{j'+1}$ occur in this order cyclically at v_j, clockwise or anticlockwise (see Fig. 9). (It follows that if (j, j') is a self-crossing, then (j', j) is also.) To remove self-crossings we can apply a similar

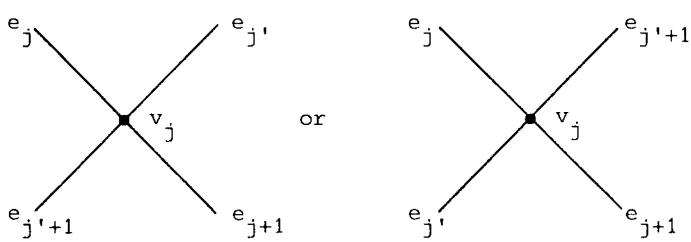

Fig. 9

approach as above, although we should be more careful: there are problems if we want to exchange parts $(v_j, \dots, v_h)$ and $(v_{j'}, \dots, v_{h'})$ of C_i if they "overlap," i.e., if they have at least one edge in common. The following proposition shows that we can avoid this situation:

Proposition 5. *If (1.3(i)) is satisfied and (j, j') is a self-crossing of C_i with j as small as possible, then there exists another self-crossing (h, h') of C_i so that:*

$$\begin{aligned} &\text{(i) parts } (v_j, \dots, v_h) \text{ and } (v_{j'}, \dots, v_{h'}) \text{ of } C_i \text{ are homotopic;} \\ &\text{(ii) } j \le h \le j' \le h' \text{ or } j \le h \le h' \le j'. \end{aligned} \tag{3.6}$$

Proof. By deforming C_i slightly, we may assume that C_i has no "self-touchings." (To allow this deformation we can add a little "space" at fixed points of C_i—this does not invalidate the conclusion of Proposition 5.) By (1.3(i)), there exists a simple curve $\tilde{C}_i \sim C_i$. Hence any two liftings of $\tilde{C}_i$ to the universal covering space U are disjoint and simple. So any two liftings of C_i to U have an even number of crossings.

Since C_i has no null-homotopic parts, each lifting of C_i to U is simple. Let us assume without loss of generality that (j, j') is a self-crossing where $e_j, e_{j'}, e_{j+1}, e_{j'+1}$ occur clockwise at $v_j = v_{j'}$ (so the first configuration in Fig. 9 applies). Consider a lifting

$$C_i' = (v_0', e_1', v_1', \dots, e_m', v_m') \tag{3.7}$$

of C_i. As (j, j') is a self-crossing, there exist liftings

$$C_i'' = (v_0'', e_1'', v_1'', \dots, e_m'', v_m'') \quad \text{and} \quad C_i''' = (v_0''', e_1''', v_1''', \dots, e_m''', v_m''') \tag{3.8}$$

of C_i so that $v_{j'}'' = v_j'$ and $v_{j'}''' = v_j''$ and so that

$$\begin{aligned} &e_j', e_{j'}'', e_{j+1}', e_{j'+1}'' \text{ occur clockwise at } v_{j'}'' = v_j' \\ \text{and} \\ &e_j'', e_{j'}''', e_{j+1}'', e_{j'+1}''' \text{ occur clockwise at } v_{j'}''' = v_j''. \end{aligned} \tag{3.9}$$

(see Fig. 10).

Now C_i' and C_i'' must have a second crossing. Choose the smallest h so that $h \neq j$ and $v_h' = v_{h'}''$ gives a crossing of C_i' and C_i'' for some h'. By the minimality of j we know $h > j$. Note that by the symmetry of the universal covering space, $v_h'' = v_{h'}'''$ gives a crossing of C_i'' and C_i'''. We consider two cases.

Case 1: $h' \ge j'$. Since, by the minimality of j, C_i''' cannot cross C_i'' at $v_0'', \dots, v_{j-1}''$ and cannot cross C_i' at $v_0', \dots, v_{j-1}'$, it follows that C_i''' crosses C_i'' in one of $v_{j+1}'', \dots, v_{j'}''$. Hence $h \le j'$, and we have (3.6).

Case 2: $h' < j'$. If $h \le h'$ we have (3.6), so assume $h > h'$. We show that this is not possible.

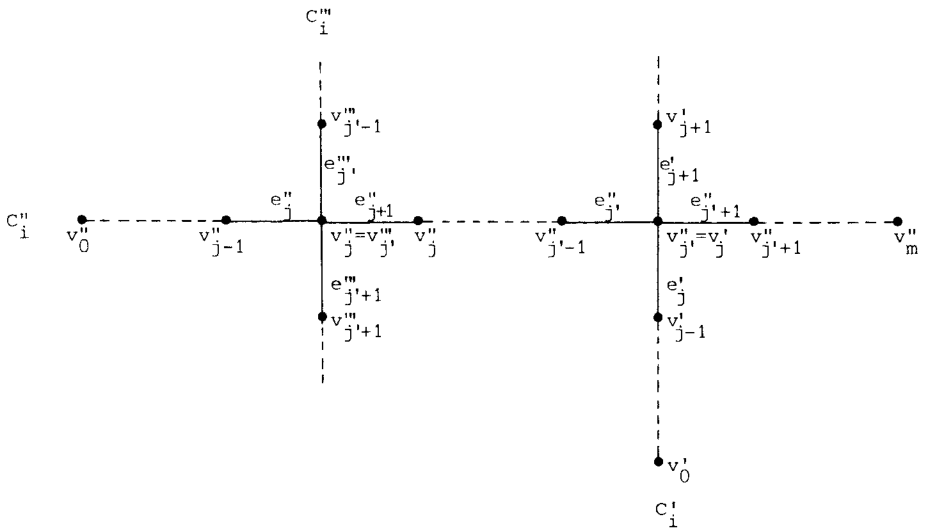

Fig. 10

Since $h > h'$, we have the situation shown in Fig. 11. Here parts $(v'_j, \ldots, v'_h)$ of C'_i and $(v'''_{h'}, \ldots, v'''_{j'})$ of C'''_i might have more than one crossing. We know however, by the minimality of h, that part $(v'_j, \ldots, v'_h)$ of C'_i does not intersect part $(v''_{h'}, \ldots, v''_{j'})$ of C''_i (except at the endpoints). Hence they enclose a simply connected (closed) region R. Similarly, parts $(v''_j, \ldots, v''_h)$ of C''_i and $(v'''_{h'}, \ldots, v'''_{j'})$ of C'''_i enclose a simply connected (closed) region R'. Moreover, by the symmetry of U, there exists a continuous function $\phi\colon R \to R'$, bringing $(v'_j, \ldots, v'_h)$ to $(v''_j, \ldots, v''_h)$ and $(v''_{h'}, \ldots, v''_{j'})$ to $(v'''_{h'}, \ldots, v'''_{j'})$ and not having any fixed point.

Furthermore, there exists a continuous function $\psi\colon U \to R$ so that:

(i) if $y \in R$, then $\psi(y) = y$;
(ii) if $y \in R' \backslash R$, then $\psi(y)$ belongs to the subcurve $(v'_j, \ldots, v'_h)$ of C'_i; (3.10)
(iii) if y belongs to subcurve $(v''_j, \ldots, v''_{h'})$ of C''_i, then $\psi(y) = v'_h$.

(This follows from the fact that C''_i divides U into two parts, and that R and R' are contained in one of these parts.)

Now consider the function $\psi \circ \varphi\colon R \to R$. Since R is simply connected, by Brouwer's fixed-point theorem there exists an $x \in R$ so that $\psi(\varphi(x)) = x$. Since φ has no fixed points, $\varphi(x) \neq x$. Hence $\varphi(x) \neq \psi(\varphi(x))$. So by (3.10(i)) $\varphi(x) \in R' \backslash R$.

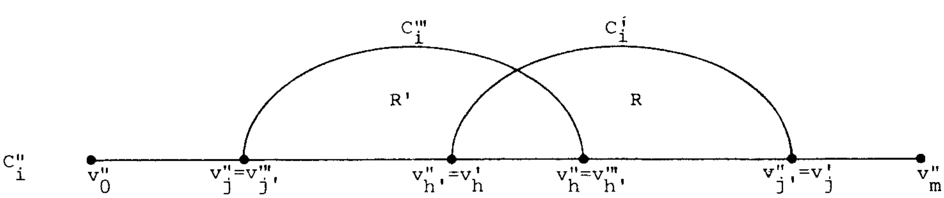

Fig. 11

Therefore, by (3.10(ii)), $x = \psi(\varphi(x))$ belongs to subcurve $(v'_j, \ldots, v'_h)$ of C'_i. So $\varphi(x)$ belongs to subcurve $(v''_j, \ldots, v''_{h'})$ of C''_i. This implies by (3.10(iii)) that $x = \psi(\varphi(x)) = v'_h$. However, $\psi(\varphi(v'_h)) = \psi(v''_h) = v''_h \neq v'_h$. □

Proposition 4 enables us to remove self-crossings. We choose a self-crossing with j as small as possible. Then with Proposition 2 we can find in polynomial time another self-crossing (h, h') satisfying (3.6), and we reset

$$C_i := (v_0, \ldots, v_j = v_{j'}, \ldots, v_{h'} = v_h, \ldots, v_{j'} = v_j, \ldots, v_h = v_{h'}, \ldots, v_m) \quad (3.11)$$

if $j' \leq h'$. Similarly if $j' > h'$.

After a polynomial number of such modifications we have that $C_1, \ldots, C_k$ have no (self-)crossings and no null-homotopic parts.

II. Constructing the System $Ax \leq b$ of Linear Inequalities

For each vertex of G, each time it is traversed by some C_i, we introduce a variable, indicating how far we should shift C_i in order to make $C_1, \ldots, C_k$ simple and pairwise disjoint. Figure 12 gives an impression.

More precisely, let, for each $i = 1, \ldots, k$,

$$C_i = (v_{i0}, e_{i1}, v_{i1}, \ldots, e_{im_i}, v_{im_i}). \quad (3.12)$$

We introduce a variable x_{ij} for each $i = 1, \ldots, k$ and $j = 1, \ldots, m_i - 1$. We put a number of linear constraints on the x_{ij} in order to make sure that the shifted C_i are (1) homotopic to the original C_i, (2) pairwise disjoint, and (3) simple. This divides the constraints into Classes 1, 2, and 3. It turns out that the full constraint system $Ax \leq b$ has an integer solution if and only if problem (1.1) has a solution.

We use the following notation. Let $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m_i - 1\}$, and consider $v_{i,j-1}, e_{ij}, v_{ij}$ as in Fig. 13. Then F^+_{ij} denotes the face incident with e_{ij} on the right-hand side when going from $v_{i,j-1}$ to v_{ij}, and F^-_{ij} denotes the face on the left-hand side.

Two faces F, F' are called *freely adjacent* at vertex v if v is incident both with F and with F', and either $F = F'$ or, when $e_1, \ldots, e_s, e_{s+1}, \ldots, e_d$ denote the edges incident with v in cyclic order as in Fig. 14, then there is no curve among $C_1, \ldots, C_k$ containing $\ldots, e_i, v, e_j, \ldots$ or $\ldots, e_j, v, e_i, \ldots$ with $1 \leq i \leq s$ and $s + 1 \leq j \leq t$. So roughly speaking, we can go from F to F' traversing v without crossing any $C_1, \ldots, C_k$. Note that at any vertex v, free adjacency forms an equivalence relation on the faces incident with v. (If a face has multiple incidences at v, we must be careful: each touch should be considered separately.)

To facilitate the construction of the system of inequalities, we define an auxiliary graph H, with length function on the edges, as follows. The vertices of H are the pairs (v, λ), where v is a vertex of G (not being one of the endpoints of $C_1, \ldots, C_k$) and where λ is an equivalence class of faces freely adjacent at v. If (v, λ) and (w, μ) are vertices of H, there is an edge of length 1 connecting them if λ and μ have a face

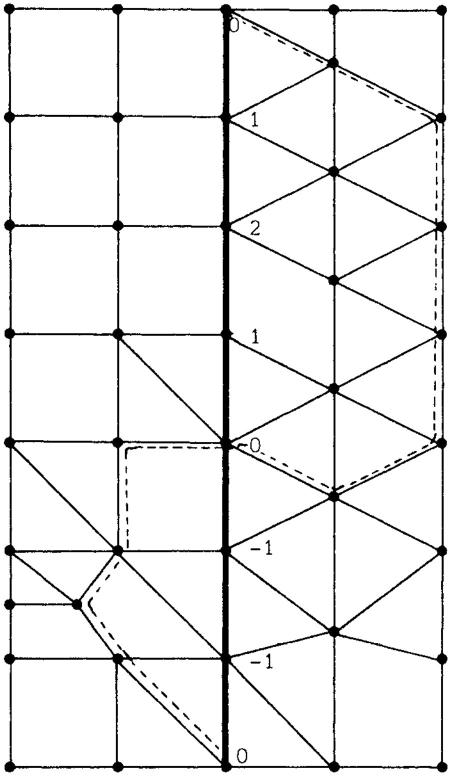

Fig. 12. Here the heavy line indicates the initial path, and the dashed line indicates the shifted path. A positive number t means shifting over a distance t to the right, and a negative number $-t$ means shifting over a distance t to the left (right and left with respect to the orientation of the initial path). For distance between vertices v, v' of G we take the minimum number of faces traversed by any curve connecting v and v'.

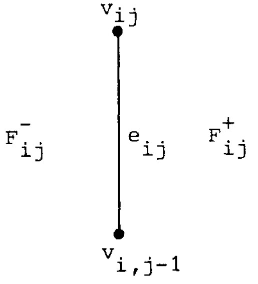

Fig. 13

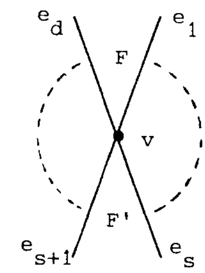

Fig. 14

$F \notin \{I_1, \ldots, I_p\}$ in common. In fact, we have an edge e_F for each face F in $\lambda \cap \mu \backslash \{I_1, \ldots, I_p\}$. Moreover, for each $i \in \{1, \ldots, k\}$ and $j, j' \in \{1, \ldots, m_i - 1\}$ there is an edge connecting $(v_{ij}, \langle F_{ij}^+ \rangle)$ and $(v_{ij'}, \langle F_{ij'}^+ \rangle)$ of length

$$\gamma_{i,j,j'} := \min_D(\mathrm{cr}(G, D) - 1), \tag{3.13}$$

where D ranges over all curves $D: [0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ homotopic to part $(v_{ij}, \ldots, v_{ij'})$ of C_i. (Here $\langle F \rangle$ denotes the equivalence class of F of free adjacency at the appropriate vertex.) Similarly, there is an edge connecting $(v_{ij'} \langle F_{ij}^- \rangle)$ and $(v_{ij'}, \langle F_{ij'}^- \rangle)$ of length $\gamma_{i,j,j'}$. Note that by Proposition 1, $\gamma_{i,j,j'}$ can be calculated in polynomial time.

There exist two "canonical" mappings φ and ψ of H in $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$. (A mapping is a continuous function, not necessarily one-to-one.) First, let

$$\varphi(v, \lambda) := \psi(v, \lambda) := v \tag{3.14}$$

for each vertex (v, λ) of H. The image, under φ as well as under ψ, of each edge e_F is a line segment contained in F connecting v and w. For the other edges, the images under φ and under ψ generally are different: the edge connecting $(v_{ij}, \langle F_{ij}^+ \rangle)$ and $(v_{ij'}, \langle F_{ij'}^+ \rangle)$ has as its φ-image a curve D attaining the minimum in (3.13). Its ψ-image is a curve traversing

$$v_{ij}, F_{i,j+1}^+, v_{i,j+1}, F_{i,j+2}^+, \ldots, F_{ij'}^+, v_{ij'}, \tag{3.15}$$

respectively (assuming without loss of generality $j \leq j'$). So the ψ-image is, informally speaking, parallel to part $(v_{ij}, \ldots, v_{ij'})$ of C_i and does not cross any $C_1, \ldots, C_k$ (since F_{ir}^+ and F_{ir+1}^+ are freely adjacent at v_{ir}). Similarly, the images of the edges connecting $(v_{ij}, \langle F_{ij}^- \rangle)$ and $(v_{ij'}, \langle F_{ij'}^- \rangle)$ are given.

Each path P in H gives two curves $\varphi \circ P$ and $\psi \circ P$, which are homotopic to each other. So we can speak of the homotopy of a path P in H.

We now describe the three classes of inequalities.

Class 1. This class of inequalities is meant to avoid that any C_i is shifted over any of the faces $I_1, \ldots, I_p$ (as we shall see below). Thus, for each $i = 1, \ldots, k$ and $j = 1, \ldots, m_i - 1$, we require:

$$\begin{aligned} &(\alpha)\ \ x_{ij} \le \min\nolimits_P \operatorname{length}(P),\\ &(\beta)\ \ -x_{ij} \le \min\nolimits_P \operatorname{length}(P). \end{aligned} \tag{3.16}$$

Here the minimum in (α) ranges over all paths P in H from $(v_{ij}, \langle F_{ij}^+\rangle)$ to any (w, λ) so that λ contains a face in $\{I_1, \ldots, I_p\}$. Similarly, the minimum in (β) ranges over all paths P in H from $(v_{ij}, \langle F_{ij}^-\rangle)$ to any (w, λ) so that λ contains a face in $\{I_1, \ldots, I_p\}$.

It is not difficult to see that such paths always exist, as $(v_{ij}, \langle F_{ij}^+\rangle)$ and $(v_{i1}, \langle F_{i1}^+\rangle)$ are connected by an edge of H, and as $\langle F_{i1}^+\rangle$ contains a face in $\{I_1, \ldots, I_p\}$. So the right-hand sides of (3.16) are finite. They can be calculated in polynomial time.

Note 3.1. The right-hand side of (3.16(α)) can be described equivalently as

$$\min_D(\operatorname{cr}(G, D) - 1), \tag{3.17}$$

where D ranges over all curves D for which there exists a curve $Q \sim D$ from v_{ij} to a vertex w on $\operatorname{bd}(I_1 \cup \cdots \cup I_p)$ so that:

$$\begin{aligned} &\text{(i) } Q \text{ does not cross any } C_1, \ldots, C_k;\\ &\text{(ii) } Q \text{ starts via a face freely adjacent at } v_{ij} \text{ to } F_{ij}^+;\\ &\text{(iii) } Q \text{ ends via a face freely adjacent at } w \text{ to some face in } \{I_1, \ldots, I_p\}. \end{aligned} \tag{3.18}$$

Here we say that Q *starts via* face F if $Q[(0, \varepsilon)] \subseteq F$ for some $\varepsilon > 0$. Similarly, Q *ends via* F if $Q[(1 - \varepsilon, 1)] \subseteq F$ for some $\varepsilon > 0$.

The fact that the right-hand side of (3.16(α)) is equal to (3.17) can be seen by observing that each path P in the range of (3.16(α)) gives a curve $D := \varphi \circ P$ in the range of (3.17), with $\operatorname{cr}(G, D) - 1 = \operatorname{length}(P)$. Conversely, for each curve D in the range of (3.17) there exists a path P in the range of (3.16(α)) with $\operatorname{length}(P) \le \operatorname{cr}(G, D) - 1$.

A similar formula holds for the right-hand side of (3.16(β)).

Class 2. This class of inequalities must accomplish that two different C_i and $C_{i'}$ do not intersect after shifting. Thus, for each $i, i' = 1, \ldots, k$ with $i \ne i'$, and for each $k = 1, \ldots, m_i - 1$ and $j' = 1, \ldots, m_{i'} - 1$, we require:

$$\begin{aligned} &(\alpha)\ \ x_{ij} + x_{i'j'} \le \operatorname{dist}_H((v_{ij}, \langle F_{ij}^+\rangle), (v_{i'j'}, \langle F_{i'j'}^+\rangle)) - 1,\\ &(\beta)\ \ x_{ij} - x_{i'j'} \le \operatorname{dist}_H((v_{ij}, \langle F_{ij}^+\rangle), (v_{i'j'}, \langle F_{i'j'}^-\rangle)) - 1,\\ &(\gamma)\ \ -x_{ij} - x_{i'j'} \le \operatorname{dist}_H((v_{ij}, \langle F_{ij}^-\rangle), (v_{i'j'}, \langle F_{i'j'}^-\rangle)) - 1, \end{aligned} \tag{3.19}$$

where dist_H denotes the distance in H (with respect to the length function given). Again, the right-hand sides of (3.19) are easily computed in polynomial time—they are allowed to be infinite.

Note 3.2. The right-hand side of (3.19(α)) can be described equivalently as

$$\min_D(\text{cr}(G, D) - 2), \tag{3.20}$$

where the minimum ranges over all curves D for which there exists a curve $Q \sim D$ from v_{ij} to $v_{i'j'}$ not crossing any $C_1, \ldots, C_k$, so that Q starts via a face freely adjacent at v_{ij} to F_{ij}^+ and ends via a face freely adjacent at $v_{i'j'}$ to $F_{i'j'}^+$. Similarly for (β) and (γ).

Class 3. The last class of inequalities must accomplish that each shifted C_i is simple. Thus, for each $i = 1, \ldots, k$ and $j, j' = 1, \ldots, m_i - 1$, we require:

$$\begin{aligned} &(\alpha)\ \ x_{ij} + x_{ij'} \leq \min\nolimits_P \text{length}(P) - 1,\\ &(\beta)\ \ x_{ij} - x_{ij'} \leq \min\nolimits_P \text{length}(P) - 1,\\ &(\gamma)\ \ -x_{ij} - x_{ij'} \leq \min\nolimits_P \text{length}(P) - 1. \end{aligned} \tag{3.21}$$

Here in (α) the minimum ranges over all paths P in H from $(v_{ij}, \langle F_{ij}^+\rangle)$ to $(v_{ij'}, \langle F_{ij'}^+\rangle)$ which are *not* homotopic to part $(v_{ij}, \ldots, v_{ij'})$ of C_i. Similarly for (β) and (γ). Again the right-hand sides of (3.21) can be infinite. If $j = j'$ we obtain bounds for $\pm 2x_{ij}$. The right-hand side of (3.21) can be calculated in polynomial time by Proposition 3.

Note 3.3. Again, the right-hand side of (3.21(α)) can be described equivalently as

$$\min_D(\text{cr}(G, D) - 2), \tag{3.22}$$

where D ranges over all curves D from v_{ij} to $v_{ij'}$ which are not homotopic to part $(v_{ij}, \ldots, v_{ij'})$ of C_i and for which there exists a curve $Q \sim D$ not crossing any $C_1, \ldots, C_k$, so that Q starts via a face freely adjacent at v_{ij} to F_{ij}^+ and ends via a face freely adjacent at $v_{ij'}$ to $F_{ij'}^+$. Similarly for (β) and (γ).

We denote the system of linear inequalities (3.16), (3.19), and (3.21) by $Ax \leq b$ (where A is a matrix and b is a column vector).

III. Solving $Ax \leq b$ in Integers

In general it is an NP-complete problem to solve a system of linear inequalities in integer variables. However, since matrix $A = (a_{ij})$ satisfies

$$\sum_{j=1}^{n} |a_{ij}| \leq 2 \qquad \text{for each} \quad i = 1, \ldots, m \tag{3.23}$$

(where A has order $m \times n$), it is quite easy to solve $Ax \leq b$ in integers, namely by "Fourier–Motzkin" elimination of variables. This recursively solves $Ax \leq b$ in integers, for any integer matrix satisfying (3.23) and any vector $b \in (\mathbb{Z} \cup \{\infty\})^m$.

Proposition 6. *There is a polynomial algorithm for solving $Ax \leq b$ in integers, for any integer $m \times n$-matrix A satisfying* (3.23) *and any vector $b \in (\mathbb{Z} \cup \{\infty\})^m$.*

Proof. We may assume that all rows of A are distinct, that A does not have any all-zero row, and that each integer row vector a^{T} with $1 \leq \|a\|_1 \leq 2$ occurs as a row of A.

We decompose the inequalities in $Ax \leq b$ as

$$
\begin{aligned}
x_1 \qquad &\leq \alpha, \\
2x_1 \qquad &\leq \beta, \\
x_1 + x_i &\leq \gamma_i \qquad (i = 2, \ldots, n), \\
x_1 - x_i &\leq \delta_i \qquad (i = 2, \ldots, n), \\
-x_1 \qquad &\leq \varepsilon, \\
-2x_1 \qquad &\leq \zeta, \\
-x_1 - x_i &\leq \eta_i \qquad (i = 2, \ldots, n), \\
-x_1 + x_i &\leq \theta_i \qquad (i = 2, \ldots, n), \\
A'x' &\leq b',
\end{aligned}
\tag{3.24}
$$

where $x' = (x_2, \ldots, x_n)^{\mathrm{T}}$ and where A' is a matrix with $n-1$ columns again satisfying (3.23).

We can replace (3.24) by the following equivalent conditions:

$$
\begin{aligned}
&\max\left\{ -\varepsilon,\ -\tfrac{1}{2}\zeta,\ \max_{2 \leq i \leq n} (-\eta_i - x_i),\ \max_{2 \leq i \leq n} (-\theta_i + x_i) \right\} \\
&\qquad \leq x_1 \leq \min\left\{ \alpha,\ \tfrac{1}{2}\beta,\ \min_{2 \leq i \leq n} (\gamma_i - x_i),\ \min_{2 \leq i \leq n} (\delta_i + x_i) \right\}, \qquad A'x' \leq b'.
\end{aligned}
\tag{3.25}
$$

Now if $\max\{-\varepsilon, -\tfrac{1}{2}\zeta\} > \min\{\alpha, \tfrac{1}{2}\beta\}$, then clearly (3.25) has no solution. Moreover, if $-\zeta = \beta$ and is odd, (3.25) has no integer value for x_1. Hence we may assume

$$
\max\{-\varepsilon, -\tfrac{1}{2}\zeta\} \leq \min\{\alpha, \tfrac{1}{2}\beta\} \qquad \text{and if} \quad -\zeta = \beta, \quad \text{then } \beta \text{ is even.} \tag{3.26}
$$

Eliminating x_1 from (3.25) gives

$$
\begin{aligned}
&\max\left\{ -\varepsilon,\ -\tfrac{1}{2}\zeta,\ \max_{2 \leq i \leq n} (-\eta_i - x_i),\ \max_{2 \leq i \leq n} (-\theta_i + x_i) \right\} \\
&\qquad \leq \min\left\{ \alpha,\ \tfrac{1}{2}\beta,\ \min_{2 \leq i \leq n} (\gamma_i - x_i),\ \min_{2 \leq i \leq n} (\delta_i + x_i) \right\}, \qquad A'x' \leq b'.
\end{aligned}
\tag{3.27}
$$

Equivalently,

$$\begin{array}{lll}
x_i \qquad\quad \le \gamma_i + \varepsilon & (i = 2, \ldots, n), & \\
x_i \qquad\quad \le \gamma_i + \tfrac{1}{2}\zeta & (i = 2, \ldots, n), & \\
-x_i \qquad\; \le \delta_i + \varepsilon & (i = 2, \ldots, n), & \\
-x_i \qquad\; \le \delta_i + \tfrac{1}{2}\zeta & (i = 2, \ldots, n), & \\
-x_i \qquad\; \le \eta_i + \alpha & (i = 2, \ldots, n), & \\
-x_i \qquad\; \le \eta_i + \tfrac{1}{2}\beta & (i = 2, \ldots, n), & \\
-x_i + x_j \le \eta_i + \gamma_j & (i, j = 2, \ldots, n), & (3.28)\\
-x_i - x_j \le \eta_i + \delta_j & (i, j = 2, \ldots, n), & \\
x_i \qquad\quad \le \theta_i + \alpha & (i = 2, \ldots, n), & \\
x_i \qquad\quad \le \theta_i + \tfrac{1}{2}\beta & (i = 2, \ldots, n), & \\
x_i + x_j \le \theta_i + \gamma_j & (i, j = 2, \ldots, n), & \\
x_i - x_j \le \theta_i + \delta_j & (i, j = 2, \ldots, n), & \\
A'x' \qquad \le b'. & &
\end{array}$$

This is a system of linear inequalities in the variables $x_2, \ldots, x_n$ again satisfying (3.23). We can reduce (3.28) so that we obtain an equivalent system $A''x' \le b''$ where A'' has no two equal rows. We next recursively solve $A''x' \le b''$ in integers. If it has no integer solution, then the original system $Ax \le b$ has neither. If $A''x' \le b''$ has an integer solution, we can insert it in (3.25), and determine an integer x_1 satisfying (3.25).

Such an integer x_1 does exist: the maximum in (3.25) is not more than the minimum. As both the maximum and the minimum are half-integers, an integer value for x_1 would not exist only if $-\frac{1}{2}\zeta = \frac{1}{2}\beta$ and is not an integer. But this is excluded by (3.26).

The case $n = 1$ being trivial completes the description of the algorithm. It has polynomially bounded running time since at each iteration we reduce the number of inequalities in (3.28) to $O(n^2)$. So we do not have exponential growth of the number of constraints (which would occur in ordinary Fourier–Motzkin elimination). $\square$

In Section 4 we show that if conditions (1.3) are satisfied, then the system $Ax \le b$ constructed in Step II indeed has an integer solution. For a direct proof of the fact that if (1.1) has a solution, then $Ax \le b$ has an integer solution, see Proposition 14 in Section 6.

IV. Shifting $C_1, \ldots, C_k$

Let $(x_{ij} \mid i = 1, \ldots, k; j = 1, \ldots, m_i - 1)$ form an integer solution of $Ax \le b$. These integers will determine the shifts of the C_i. We describe an iterative process, shifting the C_i by little steps, adapting the x_{ij} throughout.

If $x_{ij} = 0$ for all i, j, then $C_1, \ldots, C_k$ are pairwise disjoint and simple, as follows directly from the Class 2 and 3 inequalities, and from the fact that no C_i has null-homotopic parts.

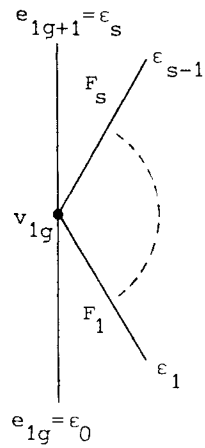

Fig. 15

Suppose next

$$M := \max\{|x_{ij}| \mid i = 1, \dots, k; j = 1, \dots, m_i - 1\} > 0. \tag{3.29}$$

First assume $x_{ig} = M$ for some i, g. Without loss of generality, $i = 1$. Consider e_{1g}, v_{1g}, $e_{1,g+1}$ and the faces and edges incident with it "at the right-hand side," as in Fig. 15. Note that $F_1, \dots, F_s \notin \{I_1, \dots, I_p\}$, by Class 1 inequalities. We claim that none of the edges $\varepsilon_1, \dots, \varepsilon_{s-1}$ is used by any C_i. For suppose $\varepsilon_t = e_{ij}$, $v_{1g} = v_{ij}$, and $\varepsilon_{t'} = e_{ij+1}$ for some i, j and some $t, t' \in \{1, \dots, s-1\}$. We may assume that $\varepsilon_{t''}$ is not traversed by $C_1, \dots, C_k$ if $1 \le t'' < \min\{t, t'\}$. If $t < t'$, then $x_{1g} - x_{ij} \le -1$, and hence $x_{ij} \ge x_{1g} + 1 = M + 1$, contradicting (3.29). Similarly, if $t > t'$, then $x_{1g} + x_{ij} \le -1$, and hence $-x_{ij} \ge x_{1g} + 1 = M + 1$, again contradicting (3.29).

Now let

$$(v_{1g-1} = w_0, f_1, w_1, f_2, w_2, \dots, f_r, w_r = v_{1g+1}) \tag{3.30}$$

be the vertices and edges on the path following the outer boundary of $F_1, \dots, F_s$ (see Fig. 16). More precisely, let $E(F)$ denote the set of edges incident with F. We take for path (3.30) any simple path from v_{1g-1} to v_{1g+1} with edges in the symmetric difference:

$$E(F_1)\Delta E(F_2)\Delta \cdots \Delta E(F_s)\Delta\{e_{1g}, e_{1g+1}\}. \tag{3.31}$$

Before proving the easy fact that path (3.30) thus obtained is homotopic to part $(v_{1g-1}, e_{1g}, v_{1g}, e_{1g+1}, v_{1g+1})$ of C_1, we show the following. Let, for each $h = 0, \dots, r$, Γ_h be some curve from v_{1g} to w_h contained in one of the faces $F_1, \dots, F_s$. Then, for each $h = 0, \dots, r$,

$$\Gamma_h \text{ is unique up to homotopy.} \tag{3.32}$$

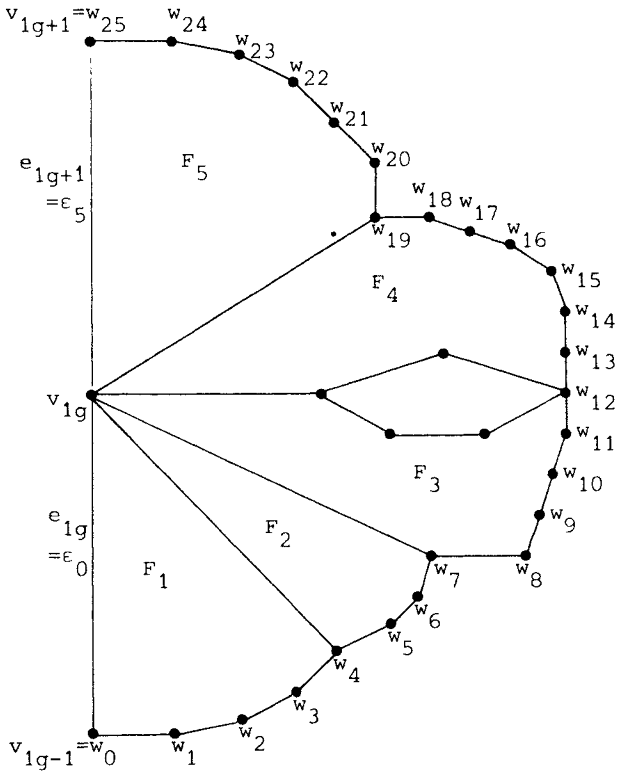

Fig. 16

For suppose there exists a curve Γ'_h from v_{1g} to w_h in one of $F_1, \ldots, F_s$ so that Γ'_h is not homotopic to Γ_h. Then we would have the contradiction $2 \le 2x_{1g} \le \mathrm{cr}(G, \Gamma'_h \cdot \Gamma_h^{-1}) - 2 = 1$, by a Class 3 inequality.

We derive:

Proposition 7. *Path* (3.30) *is homotopic to part* $(v_{1g-1}, e_{1g}, v_{1g}, e_{1g+1}, v_{1g+1})$ *of* C_1.

Proof. By (3.32), each one-edge path (w_{h-1}, f_h, w_h) is homotopic to $\Gamma_{h-1}^{-1} \cdot \Gamma_h$. Hence (3.30) is homotopic to $\Gamma_0^{-1} \cdot \Gamma_r$, which is by (3.32) homotopic to part $(v_{1g-1}, e_{1g}, v_{1g}, e_{1g+1}, v_{1g+1})$ of C_1. □

Let g' be the smallest index so that $v_{1g'} = w_{h'}$ for some $h' \in \{0, \ldots, r\}$ and so that part $(v_{1g'}, \ldots, v_{1g})$ of C_1 is homotopic to $\Gamma_{h'}^{-1}$. So $g' \le g - 1$. We can determine g' in polynomial time by Proposition 2.

Similarly, let g'' be the largest index so that $v_{1g''} = w_{h''}$ for some $h'' \in \{0, \ldots, r\}$ and so that part $(v_{1g}, \ldots, v_{1g''})$ of C_1 is homotopic to $\Gamma_{h''}$. So $g'' \ge g + 1$. Again g'' can be determined in polynomial time.

We easily check that $h' \le h''$ (using the fact that C_1 does not have null-homotopic parts). Now we obtain $\tilde{C}_1$ from C_1 by replacing part $(v_{1g'}, \ldots, v_{1g''})$ of

C_1 by part $(w_{h'}, \ldots, w_{h''})$ of (3.30). We add new edges to G so as to keep $\tilde{C}_1$, $C_2, \ldots, C_k$ pairwise edge-disjoint and without (self-) crossings.

Clearly, $\tilde{C}_1$ is homotopic to C_1(since $(v_{1g'}, \ldots, v_{1g''})$ is homotopic to $\Gamma_{h'}^{-1} \cdot \Gamma_{h''}$, which is homotopic to $(w_{h'}, \ldots, w_{h''})$). The new $\tilde{C}_1, C_2, \ldots, C_k$ give new variables $\tilde{x}_{ij}$. We set them equal to the original x_{ij} if $i \neq 1$, while $\tilde{x}_{1j}$ are set equal to $M - 1$ on the new part of $\tilde{C}_1$ and equal to the original values on the unchanged part of $\tilde{C}_1$.

To be more precise, note that $\tilde{C}_1$ passes $\tilde{m}_1 := m_1 - (g'' - g') + (h'' - h')$ edges. Let $\tilde{x}_{1j} := x_{1j}$ if $1 \leq j \leq g'$, $\tilde{x}_{1j} := M - 1$ if $g' < j < g' + (h'' - h')$, and $\tilde{x}_{1j} := x_{1, j+(g''-g')-(h''-h')}$ if $g' + (h'' - h') \leq j \leq \tilde{m}_1 - 1$. Moreover, $\tilde{x}_{ij} := x_{ij}$ for $i \neq 1$.

Proposition 8. *The $\tilde{x}_{ij}$ form an integer solution for the system of linear inequalities derived from $\tilde{C}_1, C_2, \ldots, C_k$.*

Proof. We only have to check those inequalities in the new system in which variables occur corresponding to the new trajectory of $\tilde{C}_1$ (i.e., $\tilde{x}_{1j}$ with $g' < j < g' + (h'' - h')$). This follows from the fact that for all other inequalities the values of the x_{ij} and the range for the minimum on the right-hand side are unchanged (see Notes 3.1–3.3).

Denote

$$\tilde{C}_1 = (\tilde{v}_{10}, \ldots, \tilde{v}_{1\tilde{m}}). \tag{3.33}$$

Consider some Class 2 inequality in the new system in which $\tilde{x}_{1j}$ occurs $(g' < j < g' + (h'' - h'))$, say,

$$\pm \tilde{x}_{1j} + \tilde{x}_{ij'} \leq \mathrm{cr}(G, D) - 2, \tag{3.34}$$

for some curve D in the range described in Note 3.2 (with $i \neq 1$). Let h be so that $w_h = \tilde{v}_{1j}$ (i.e., $h := h' + j - g'$).

If $\tilde{x}_{1j}$ has coefficient $+1$ in (3.34), then we can extend D to a curve $D' := \Gamma_h \cdot D$ from v_{1g} to $\tilde{v}_{1j} = w_h$. Then $\mathrm{cr}(G, D') = \mathrm{cr}(G, D) + 1$, and

$$x_{1g} + x_{ij'} \leq \mathrm{cr}(G, D') - 2 \tag{3.35}$$

(a Class 2 inequality in the original system). Therefore

$$\tilde{x}_{1j} + \tilde{x}_{ij'} = (M - 1) + x_{ij'} = x_{1g} + x_{ij'} - 1 \leq \mathrm{cr}(G, D') - 2 - 1 = \mathrm{cr}(G, D) - 2. \tag{3.36}$$

So we have (3.34).

If $\tilde{x}_{1j}$ has coefficient -1 in (3.34), the situation is slightly more complicated. We may assume D does not intersect edges of G. Now D is the concatenation $D' \cdot D''$ of two curves D' and D'' so that D' connects $\tilde{v}_{1j}$ with some vertex v_{1f} on C_1, in such a way that part $(v_{1g}, \ldots, v_{1f})$ of C_1 is homotopic to $\Gamma_h \cdot D'$. (This follows from the fact that D is homotopic to some curve Q starting at the negative side of $\tilde{C}_1$ at $\tilde{v}_{1j}$ and not crossing any of $\tilde{C}_1, C_2, \ldots, C_k$.)

Now $\mathrm{cr}(G, D') \geq 2$. (Otherwise, part $(v_{1g}, \ldots, v_{1f})$ of C_1 would be homotopic to Γ_h. However, $v_{1f} = \tilde{v}_{1j}$ belongs to $\{w_{h'+1}, \ldots, w_{h''-1}\}$, contradicting the choice of h' and h''.) Moreover, $x_{1f} \leq M$ and $-x_{1f} + x_{ij'} \leq \mathrm{cr}(G, D'') - 2$. Therefore,

$$\begin{aligned} -\tilde{x}_{1j} + \tilde{x}_{ij'} &= -M + 1 + x_{ij'} \\ &\leq -x_{1f} + x_{ij'} + 1 \\ &\leq \mathrm{cr}(G, D'') - 1 \\ &\leq \mathrm{cr}(G, D') + \mathrm{cr}(G, D'') - 3 \\ &= \mathrm{cr}(G, D) - 2. \end{aligned}$$

Again we have (3.34).

Other inequalities are proved similarly. □

The case $x_{ig} = -M$ is dealt with similarly. This describes an iterative process of adapting paths and variables. It is easy to see that it terminates, as at each iteration the number of variables x_{ij} with $|x_{ij}| = M$ strictly decreases. If all $|x_{ij}| = M$ have been removed, we can start to remove all $|x_{ij}| = M - 1$, and so on. We will end up with all $x_{ij} = 0$, i.e., the shifted $C_1, \ldots, C_k$ finally are simple and pairwise disjoint.

In fact, this is a polynomial-time procedure:

Proposition 9. *The number of iterations in the above algorithm is polynomially bounded.*

Proof. First, the number M is bounded by a polynomial in the size of the input, since, for each variable x_{ij}, we have $x_{ij} \leq j$ as consequence of Class 1 inequalities (since there is a curve D following v_{ij}, F^+_{ij}, v_{ij-1}, $F^+_{ij-1}, \ldots, v_{i1}$ successively, with $\mathrm{cr}(G, D) = j$). Similarly, $-x_{ij} \leq j$.

Moreover, the number of variables x_{ij} at any stage of the shifting process is bounded by $(2M + 1)\varepsilon$, where ε is the number of edges in the *initial* graph G, i.e., before adding parallel edges to G.

To see this upper bound, consider a parallel class of edges connecting, say, v and w. If e_{ij} belongs to this parallel class, let $z_{ij} := x_{ij}$ if $v = v_{ij}$ and $z_{ij} := -x_{ij-1}$ if $v = v_{ij-1}$. Now choose e_{ij} and $e_{i'j'}$ both in this parallel class, so that e_{ij} is left of $e_{i'j'}$ (when going from v to w), and so that no edge between e_{ij} and $e_{i'j'}$ is traversed by any $C_1, \ldots, C_k$. Then we have $z_{ij} - z_{i'j'} \leq -1$ (by Class 2 and 3 inequalities, since all faces between e_{ij} and $e_{i'j'}$ belong to the same free adjacency class at v).

So $z_{i'j'} \geq z_{ij} + 1$. Since each $|z_{ij}|$ is at most M, it follows that there are at most $2M + 1$ edges in the parallel class that are traversed by $C_1, \ldots, C_k$. Hence the sum of the lengths of the C_i is at most $(2M + 1)\varepsilon$. Therefore, there are at most $(2M + 1)\varepsilon$ variables, which proves the proposition. □

This finishes the description of the algorithm. In Section 5 we show that if condition (1.3) is satisfied, the system $Ax \leq b$ indeed has a solution. So if (1.3) holds, the algorithm yields a solution to the disjoint homotopic paths problem, thereby proving Theorems 1 and 2.

4. Integer Solutions to Systems $Ax \leq b$

We now give necessary and sufficient conditions for the existence of an integer solution for a general system $Ax \leq b$ of linear inequalities, where $A = (a_{ij})$ is any integer $m \times n$-matrix satisfying

$$\sum_{j=1}^{n} |a_{ij}| \leq 2 \qquad \text{for all} \quad i = 1, \ldots, m \tag{4.1}$$

In Section 5 we apply this characterization to the special system $Ax \leq b$ constructed in Step II of the algorithm in Section 3.

Thus let $A = (a_{ij})$ be an integer $m \times n$ matrix satisfying (4.1), and let $b \in \mathbb{Z}^m$. By (4.1), each row of A has at most two nonzeros. In characterizing if $Ax \leq b$ has an integer solution, we may assume that each row of A has at least one nonzero.

It is helpful to think of A as a *bidirected graph*: its vertices are the column indices and its edges are the row indices. If row i has nonzeros in positions j and j' with $j \neq j'$, it gives an edge connecting j and j', and can be represented as in Fig. 17, depending on whether $(a_{ij}, a_{ij'}) = (1, 1), (1, -1), (-1, 1)$ or $(-1, -1)$.

If row i has only one nonzero $a_{ij} = \pm 2$, it is represented by a *loop* as in Fig. 18 (where $a_{ij} = +2$ and -2, respectively). Moreover, there are edges called *ends*, with exactly one nonzero a_{ij} being ± 1. We represent them as shown in Fig. 19 (where $a_{ij} = +1$ and -1, respectively).

We consider certain types of paths in this bidirected graph A, which we call "links." A *link* is a sequence

$$(i_1, j_1, i_2, j_2, \ldots, j_{t-1}, i_t) \tag{4.2}$$

Fig. 17

Fig. 18

Fig. 19

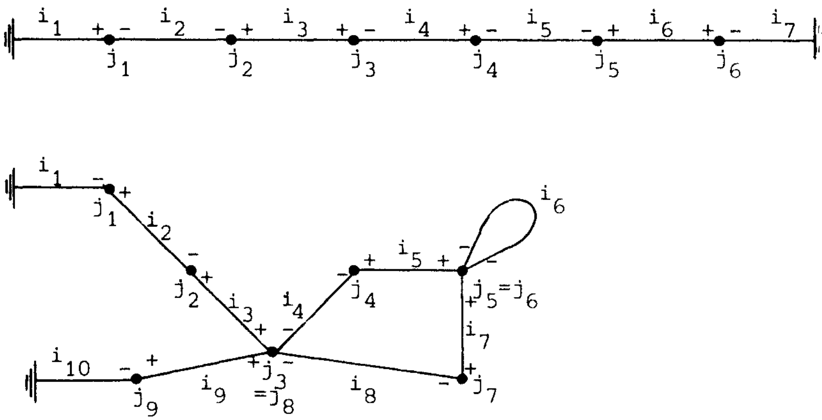

Fig. 20

(with $t \geq 2$) of rows $i_1, \ldots, i_t$ and columns $j_1, \ldots, j_{t-1}$ satisfying:

(i) i_1 is an end at j_1 and i_t is an end at j_{t-1};
(ii) for each $h = 2, \ldots, t-1$, either $j_{ij} \neq j_h$ and i_h is an edge connecting j_{h-1} and j_h, or $j_{h-1} = j_h$ and i_h is a loop at j_h; (4.3)
(iii) for each $h = 1, \ldots, t-1$,

$$a_{i_h j_h} \cdot a_{i_{h+1} j_h} < 0.$$

Condition (4.3(iii)) means that at each vertex j_h the sign flips. Examples of links are shown in Fig. 20.

Note that (4.3(iii)) implies that, for each vertex $j = 1, \ldots, n$,

$$\sum_{h=1}^{t} a_{i_h j} = 0. \tag{4.4}$$

That is, adding up the rows of A with indices $i_1, \ldots, i_t$ gives all zeros.

The *length* of link (4.2) is by definition

$$\sum_{h=1}^{t} b_{i_h}. \tag{4.5}$$

It follows directly from (4.4) that if $Ax \leq b$ has a solution x (integer or not), then each link has nonnegative length, since

$$\sum_{h=1}^{t} b_{i_h} \geq \sum_{h=1}^{t} \sum_{j=1}^{n} a_{i_h j} x_j = \sum_{j=1}^{n} x_j \sum_{h=1}^{t} a_{i_h j} = 0. \tag{4.6}$$

We next consider cycles. A *cycle* is a sequence

$$(j_0, i_1, j_1, \ldots, i_t, j_t) \tag{4.7}$$

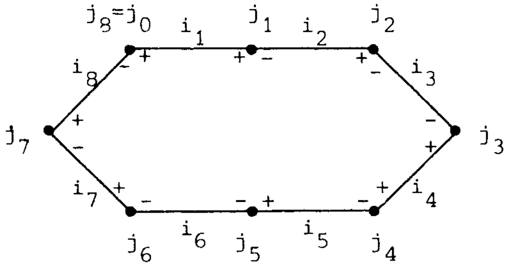

Fig. 21

(with $t \geq 1$) satisfying:

(i) $j_0 = j_t$;
(ii) for each $h = 1, \ldots, t$, either $j_{h-1} \neq j_h$ and i_h is an edge connecting j_{h-1} and j_h, or $j_{h-1} = j_h$ and i_h is a loop at j_h;
(iii) for each $h = 1, \ldots, t$ (taking $i_{t+1} := i_1$), (4.8)

$$a_{i_h j_h} \cdot a_{i_{h+1} j_h} < 0.$$

We give an example in Fig. 21 (in fact, vertices and edges may coincide).

Again, the *length* of cycle (4.7) is given by (4.5). Since (4.4) again holds, we know that if $Ax \leq b$ has a solution x (integer or not), then each cycle has nonnegative length. Actually, it can be shown that $Ax \leq b$ has a solution x, if and only if each link and each cycle has nonnegative length.

To characterize the existence of an *integer* solution, we need one further concept. A cycle (4.7) is called *doubly odd* if there exists an s with $0 < s < t$ so that

(i) $j_0 = j_s = j_t$ and $a_{i_1 j_0} \cdot a_{i_s j_s} > 0$;
(ii) $\sum_{h=1}^{s} b_{i_h}$ and $\sum_{h=s+1}^{t} b_{i_h}$ are odd numbers. (4.9)

An example of a cycle satisfying (4.9(i)) is given in Fig. 22.

Note that (4.9(i)) implies

$$\sum_{h=1}^{s} a_{i_h j} = \begin{cases} 0 & \text{if } j \neq j_0, \\ \pm 2 & \text{if } j = j_0. \end{cases} \tag{4.10}$$

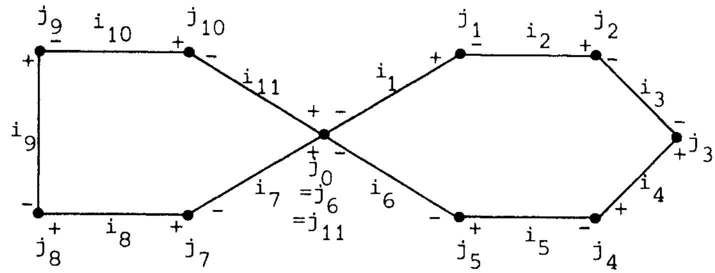

Fig. 22

This implies that if $Ax \leq b$ has an integer solution x, then any doubly odd cycle has *positive* length; since

$$\sum_{h=1}^{s} b_{i_h} \geq \sum_{h=1}^{s} \sum_{j=1}^{n} a_{i_h j} x_j = \sum_{j=1}^{n} x_j \sum_{h=1}^{s} a_{i_h j} = \pm 2x_{j_0} \tag{4.11}$$

and since the first term in (4.11) is odd, we should have strict inequality in (4.11) and hence also in (4.6).

We show that the necessary conditions mentioned are also sufficient:

Proposition 10. *A system $Ax \leq b$ satisfying* (4.1), *with $b \in \mathbb{Z}^m$, has an integer solution x, if and only if*:

(i) *each link has nonnegative length*;
(ii) *each cycle has nonnegative length*; (4.12)
(iii) *each doubly odd cycle has positive length.*

Proof. Above we showed the necessity of (4.12). We show sufficiency by induction on n, the case $n = 1$ being trivial. In fact, the inductive step follows from the algorithm (Fourier–Motzkin elimination) described in Proposition 6. To see this, let (4.12) be satisfied. This implies (3.26) (by applying (4.12(ii)) and (4.12(iii)) to cycles consisting of two loops at the same vertex). Moreover, (4.12) is maintained after elimination. This follows from the fact that each inequality in (3.28) is a combination of inequalities in (3.24), in such a way that each link and each (doubly odd) cycle for (3.28) comes from a link or (doubly odd) cycle for (3.24) with the same length. The induction hypothesis gives that (3.28) has an integer solution. Hence (3.24) also has an integer solution. □

In fact we have:

Proposition 11. *Let $Ax \leq b$ be a system satisfying* (4.1), *and $b \in \mathbb{Z}^m$, so that, for each $j = 1, \ldots, n$, the inequalities $x_j \leq \alpha_j$ and $-x_j \leq \beta_j$ occur among $Ax \leq b$ for some α_j, $\beta_j \in \mathbb{Z}$. Then condition* (4.12(ii)) *is implied by* (4.12(i)).

Proof. Suppose $(j_0, i_1, j_1, \ldots, i_t, j_t)$ is a cycle of length $-\lambda < 0$. Without loss of generality, $a_{i_1 j_0} < 0$ and $a_{i_t j_t} > 0$. By assumption, $x_{j_0} \leq \alpha$ and $-x_{j_0} \leq \beta$ occur among $Ax \leq b$, with finite α and β. We may assume that they are the first two inequalities in $Ax \leq b$. Let r be a natural number with $r > \alpha + \beta$. Consider the link

$$(1, j_0(i_1, j_1, \ldots, i_t, j_t = j_0), \ldots, (i_1, j_1, \ldots, i_t, j_t = j_0), 2), \tag{4.13}$$

where there are r repetitions of string $i_1, j_1, \ldots, i_t, j_t = j_0$. Link (4.13) has length $\alpha - r\lambda + \beta < 0$. This contradicts (4.12(i)). □

5. Proof of Theorems 1 and 2

We now apply the results described in Section 4 to the special system $Ax \leq b$ of linear inequalities constructed in Step II of the algorithm.

Proposition 12. *Let $Ax \leq b$ be the system of linear inequalities given by* (3.16), (3.19), *and* (3.21). *If condition* (1.3) *is satisfied, then $Ax \leq b$ has an integer solution x.*

Proof. Since the right-hand sides of (3.16) are finite, by Propositions 10 and 11 it suffices to show that conditions (4.12(i)) and (4.12(iii)) are satisfied. Observe that the column indices of A are now *pairs* (i, j), and that each row of A corresponds to a pair of curves $D \sim Q$ (see Notes 3.1–3.3).

I. Suppose $Ax \leq b$ contains a link of negative length. By construction of $Ax \leq b$ it means that there exist:

$$\begin{aligned}&\text{(i) pairs } (i_1, j_1), (i_2, j_2), \ldots, (i_t, j_t),\\ &\text{(ii) curves } D_0, D_1, \ldots, D_t\colon [0,1] \to \mathbb{R}^2\setminus(I_1 \cup \cdots \cup I_p),\\ &\text{(iii) curves } Q_0, Q_1, \ldots, Q_t\colon [0,1] \to \mathbb{R}^2\setminus(I_1 \cup \cdots \cup I_p),\end{aligned} \tag{5.1}$$

so that:

(i) D_h is homotopic to Q_h (for $h = 0, \ldots, t$),
(ii) $Q_0(0), Q_t(1) \in \mathrm{bd}(I_1 \cup \cdots \cup I_p)$,
(iii) $Q_{h-1}(1) = Q_h(0) = v_{i_h j_h}$ (for $h = 1, \ldots, t$),
(iv) Q_h does not cross any $C_1, \ldots, C_k$ ($h = 0, \ldots, t$),
(v) Q_0 starts via a face freely adjacent at $Q_0(0)$ to some face in $\{I_1, \ldots, I_p\}$,
(vi) Q_{h-1} ends via a face freely adjacent at $v_{i_h j_h}$ to $F^+_{i_h j_h}$ and Q_h starts via a face freely adjacent at $v_{i_h j_h}$ to $F^-_{i_h j_h}$, or conversely (i.e., $F^+_{i_h j_h}$ and $F^-_{i_h j_h}$ interchanged) (for $h = 1, \ldots, t$), (5.2)
(vii) Q_t ends via a face freely adjacent at $Q_t(1)$ to some face in $\{I_1, \ldots, I_p\}$,
(viii) if $i_h = i_{h+1}$, then Q_h is not homotopic to part $(v_{i_h j_h}, \ldots, v_{i_h j_{h+1}})$ of C_{i_h},

and so that

$$(\mathrm{cr}(G, D_0) - 1) + \left(\sum_{h=1}^{t-1} (\mathrm{cr}(G, D_h) - 2)\right) + (\mathrm{cr}(G, D_t) - 1) < 0. \tag{5.3}$$

Note that it follows from (5.2(vi)) that the concatenation $Q_{h-1}Q_h$ crosses C_{i_h} at $v_{i_h j_h}$.

Let D and Q be the concatenations $D_0 D_1 \cdots D_t$ and $Q_0 Q_1 \cdots Q_t$, respectively. So D and Q are homotopic (by (5.2(i))), and, moreover,

$$\mathrm{cr}(G, D) = 1 + \sum_{h=0}^{t} (\mathrm{cr}(G, D_h) - 1) < t \tag{5.4}$$

by (5.3). We show

$$\sum_{i=1}^{k} \operatorname{mincr}(C_i, D) \geq t, \tag{5.5}$$

thus contradicting (1.3(ii)). Since Q and D are homotopic, it is equivalent to show

$$\sum_{i=1}^{k} \operatorname{mincr}(C_i, Q) \geq t. \tag{5.6}$$

To this end, consider the universal covering space U of $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$. Each lifting Q' of Q to U is the concatenation of liftings $Q'_0, \ldots, Q'_t$ of $Q_0, \ldots, Q_t$, respectively. Now Q' connects two points on the boundary of U, and crosses, successively, t *different* liftings of $C_1, \ldots, C_k$ (by (5.2(viii))) (i.e., any two successive liftings of $C_1, \ldots, C_k$ met by Q' are different). Moreover, there are no further crossings of Q' with liftings of $C_1, \ldots, C_k$. Hence, if $\tilde{C}_1, \ldots, \tilde{C}_k, \tilde{Q}$ are homotopic to $C_1, \ldots, C_k, Q$, respectively, then any lifting of $\tilde{Q}$ to U intersects at least t liftings of $\tilde{C}_1, \ldots, \tilde{C}_k$. This implies (5.6).

II. It turns out that deriving condition (4.12(iii)) from (1.3) is less direct, due to the fact that fixed points are excluded from being traversed by doubly odd closed curves. To settle this, we first show a somewhat technical statement. Let $B = B_1 B_2$ be the concatenation of two closed curves $B_1, B_2 \colon S_1 \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$ so that $B_1(1) = B_2(1) \notin G$, $\operatorname{cr}(G, B)$ is finite, and

$$\begin{aligned} &\text{(i)} \ \operatorname{cr}(G, B_1) \not\equiv \textstyle\sum_{i=1}^{k} \operatorname{kr}(C_i, B_1) \pmod 2, \\ &\text{(ii)} \ \operatorname{cr}(G, B_2) \not\equiv \textstyle\sum_{i=1}^{k} \operatorname{kr}(C_i, B_2) \pmod 2, \end{aligned} \tag{5.7}$$

We show:

Claim. *There exists a natural number n so that, for each closed curve Q freely homotopic to $(B_1B_2)^n(B_1^{-1}B_2^{-1})^n$ with the property that each lifting of Q crosses each lifting of each C_i at most once, we have*

$$\operatorname{cr}(G, (B_1B_2)^n(B_1^{-1}B_2^{-1})^n) > \sum_{i=1}^{k} \operatorname{kr}(C_i, Q). \tag{5.8}$$

(Here for any closed curve D and $n \in \mathbb{Z}$, D^n denotes the closed curve with $D^n(z) := D(z^n)$ for all $z \in S_1$.)

Proof of the Claim. If B_1B_2 does not traverse any fixed point of any C_i, we can take $n = 1$; since $B_1B_2B_1^{-1}B_2^{-1}$ is doubly odd (with respect to the splitting into $B_1B_2B_1^{-1}$ and B_2^{-1}), we have by (1.3(iii))

$$\operatorname{cr}(G, B_1B_2B_1^{-1}B_2^{-1}) > \sum_{i=1}^{k} \operatorname{mincr}(C_i, Q) \geq \sum_{i=1}^{k} \operatorname{kr}(C_i, Q). \tag{5.9}$$

This implies (5.8).

Suppose next that B_1B_2 traverses some fixed point w of some C_i. Without loss of generality, $i = 1$ and B_1 traverses w. By condition (1.3(ii)), w cannot be a fixed point of any other C_i and is a fixed point of C_1 only once (i.e., C_1 is homotopic to a curve traversing w exactly once). So we can shift each C_i slightly in the neighborhood of w, so as to obtain curves $\tilde{C}_i \sim C_i$ so that

$$\text{no } \tilde{C}_i \text{ traverses } w, \text{ except for } \tilde{C}_1 \text{ traversing } w \text{ exactly once.} \tag{5.10}$$

We can decompose B_1 as the concatenation $B'_1B''_1$ of two (nonclosed) curves B'_1 and B''_1 with $B'_1(1) = B''_1(0) = w$.

Consider for $n \in \mathbb{N}$ the curve

$$A_1 := B''_1(B_2B_1)^n B_2(B_1^{-1}B_2^{-1})^n(B''_1)^{-1}, \tag{5.11}$$

taken as a nonclosed curve from w to w. (Here, for any curve D: $[0, 1] \to \mathbb{R}^2$, curve D^{-1} is given by $D^{-1}(x) := D(1 - x)$ for $x \in [0, 1]$.)

Let $\bar{A}_1$ be a lifting of A_1 to the universal covering space U of $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$. Then $\bar{A}_1$ connects liftings w_1 and w_2 of w, which are fixed points of liftings C_{11} and C_{12}, respectively, of $\tilde{C}_1$. Now we choose n so that C_{11} and C_{12} cross $\bar{A}_1$ the same number of times. (Such an n exists since, if n is large enough, C_{11} only crosses the beginning part (corresponding to $B''_1(B_2B_1)^n$) of $\bar{A}_1$, and C_{12} only crosses the end part (corresponding to $(B_1^{-1}B_2^{-1})^n(B''_1)^{-1}$) of $\bar{A}_1$. By the symmetry of the universal covering space and of A_1, it follows that the number of crossings are the same.)

Let A_2 be the following curve from w to w:

$$A_2 := (B'_1)^{-1}B_2^{-1}B'_1. \tag{5.12}$$

Let $\bar{A}_2$ be the lifting of A_2 to U with $\bar{A}_2(0) = w_2$. Let $w_3 := \bar{A}_2(1)$, which is again a lifting of w. Let C_{13} be the lifting of $\tilde{C}_1$ which has w_3 as a fixed point. Schematically we have Fig. 23.

Let $\mathscr{L}$ denote the collection of all liftings of all $\tilde{C}_1, \ldots, \tilde{C}_k$. Note that, except for C_{11} and C_{12}, no lifting in $\mathscr{L}$ traverses the endpoints w_1 and w_2 of $\bar{A}_1$ (by (5.10)). Similarly, except for C_{12} and C_{13}, no lifting in $\mathscr{L}$ traverses the endpoints w_2 and w_3 of $\bar{A}_2$.

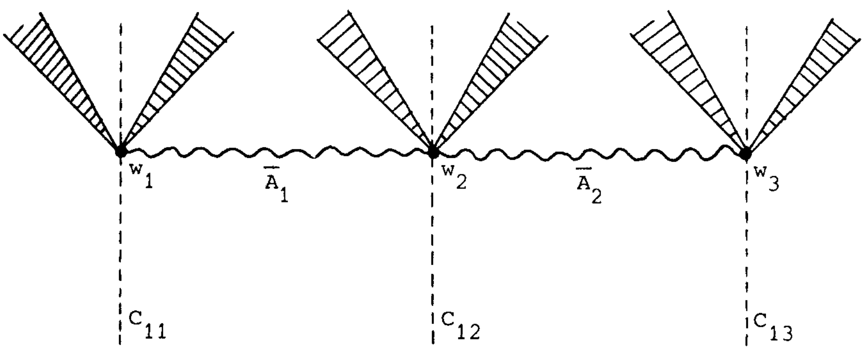

Fig. 23

Define

$$\begin{aligned}\alpha_1 &:= \text{number of } L \in \mathscr{L} \text{ with } \mathrm{kr}(L, \bar{A}_1) \text{ odd and } L \neq C_{11}, C_{12},\\ \alpha_2 &:= \text{number of } L \in \mathscr{L} \text{ with } \mathrm{kr}(L, \bar{A}_2) \text{ odd and } L \neq C_{12}, C_{13}.\end{aligned} \tag{5.13}$$

Then:

$$\begin{aligned}&\text{(i) } \mathrm{cr}(G, A_1) \geq \textstyle\sum_{i=1}^{k} \mathrm{mincr}(C_i, A_1) \geq \alpha_1 + 2,\\ &\text{(ii) } \mathrm{cr}(G, A_2) \geq \textstyle\sum_{i=1}^{k} \mathrm{mincr}(C_i, A_2) \geq \alpha_2 + 2.\end{aligned} \tag{5.14}$$

Moreover, since $\mathrm{kr}(C_{11}, \bar{A}_1) = \mathrm{kr}(C_{12}, \bar{A}_1)$, we have

$$\begin{aligned}\alpha_1 + 2 \equiv \alpha_1 &\equiv \text{number of } L \in \mathscr{L} \text{ with } \mathrm{kr}(L, \bar{A}_1) \text{ odd} \equiv \sum_{i=1}^{k} \mathrm{kr}(\tilde{C}_i, A_i)\\ &\equiv \sum_{i=1}^{k} \mathrm{kr}(C_i, B_2) \not\equiv \mathrm{cr}(G, B_2) \equiv \mathrm{cr}(G, A_1) \pmod 2.\end{aligned} \tag{5.15}$$

So we have strict inequality in (5.14(i)). Hence

$$\begin{aligned}\mathrm{cr}(G, (B_1B_2)^{n+1}(B_1^{-1}B_2^{-1})^{n+1}) &= \mathrm{cr}(G, A_1) + \mathrm{cr}(G, A_2) - 2 > \alpha_1 + \alpha_2 + 2\\ &\geq 1 + (\text{number of } L \in \mathscr{L} \text{ with } \mathrm{kr}(L, \bar{A}_1\bar{A}_2) \text{ odd})\\ &\geq \sum_{i=1}^{k} \mathrm{kr}(C_i, Q)\end{aligned} \tag{5.16}$$

for any closed curve Q freely homotopic to $(B_1B_2)^{n+1}(B_1^{-1}B_2^{-1})^{n+1}$ with the property that any lifting of Q crosses any $L \in \mathscr{L}$ at most once. □

III. We now show (4.12(iii)). Suppose to the contrary that $Ax \leq b$ has a doubly odd cycle of nonpositive length. Again it follows that there exist:

(i) pairs $(i_1, j_1), (i_2, j_2), \ldots, (i_t, j_t)$,
(ii) curves $D_1, \ldots, D_t$: $[0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$,
(iii) curves $Q_1, \ldots, Q_t$: $[0, 1] \to \mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$,
(iv) an index s with $0 < s < t$ and $(i_s, j_s) = (i_t, j_t)$, (5.17)

so that (taking $Q_{t+1} := Q_1$):

(i) D_h is homotopic to Q_h (for $h = 0, \ldots, t$),
(ii) $Q_h(1) = Q_{h+1}(0) = v_{i_hj_h}$ (for $h = 1, \ldots, t$),
(iii) Q_h does not cross any $C_1, \ldots, C_k$ ($h = 1, \ldots, t$),
(iv) Q_h ends via a face freely adjacent at $v_{i_hj_h}$ to $F^+_{i_hj_h}$ and Q_{h+1} starts via a face freely adjacent at $v_{i_hj_h}$ to $F^-_{i_hj_h}$, or conversely (i.e., $F^+_{i_hj_h}$ and $F^-_{i_hj_h}$ interchanged) (for $h = 1, \ldots, t$), (5.18)
(v) if $i_h = i_{h+1}$, then Q_h is not homotopic to part $(v_{i_{h-1}j_{h-1}}, \ldots, v_{i_hj_h})$ of C_{i_h} ($h = 1, \ldots, t$),
(vi) Q_s ends via a face freely adjacent at $v_{i_sj_s}$ to $F^+_{i_sj_s}$ and Q_t ends via a face freely adjacent at $v_{i_tj_t}$ to $F^-_{i_tj_t}$ or conversely,
(vii) $\sum_{h=1}^{s} (\mathrm{cr}(G, D_h) - 2)$ and $\sum_{h=s+1}^{t} (\mathrm{cr}(G, D_h) - 2)$ are odd,

and so that

$$\sum_{h=1}^{t} (\mathrm{cr}(G, D_h) - 2) \leq 0. \tag{5.19}$$

Define the closed curves

$$R_1 := D_1 \cdots D_s, \quad R_2 := D_{s+1} \cdots D_t, \qquad Y_1 := Q_1 \cdots Q_s, \quad Y_2 := Q_{s+1} \cdots Q_t. \tag{5.20}$$

We can decompose R_1 as $R'_1 R''_1$, where R'_1 and R''_1 are (nonclosed) curves with $R'_1(1) = R''_1(0) \notin G$. Let B_1 and B_2 be the closed curves given by

$$B_1 := R''_1 R_2 (R''_1)^{-1} \qquad \text{and} \qquad B_2 := (R'_1)^{-1} R_2^{-1} R'_1. \tag{5.21}$$

So $B_1(1) = B_2(1) = R'_1(1) \notin G$. By (5.18)(vii)) we have

$$\begin{aligned} \mathrm{cr}(G, B_1) &= 1 + 2(\mathrm{cr}(G, R''_1) - 1) + \mathrm{cr}(G, R_2) \equiv 1 + \mathrm{cr}(G, R_2) \\ &= 1 + \sum_{h=s+1}^{t} (\mathrm{cr}(G, D_h) - 1) \equiv t - s \pmod 2. \end{aligned} \tag{5.22}$$

Moreover, as each D_h crosses the C_i an even number of times,

$$\begin{aligned} \sum_{i=1}^{k} \mathrm{kr}(C_i, B_1) &= 2\left(\sum_{i=1}^{k} \mathrm{kr}(C_i, R''_1)\right) + \left(\sum_{i=1}^{k} \sum_{h=s+1}^{t} \mathrm{kr}(C_i, D_h)\right) + t - s + 1 \\ &\not\equiv t - s \pmod 2. \end{aligned} \tag{5.23}$$

So $\mathrm{cr}(G, B_1) \not\equiv \sum_{i=1}^{k} \mathrm{kr}(C_i, B_1) \pmod 2$. Similarly for B_2. Hence the Claim applies. Let n have the properties described. As

$$(B_1 B_2)^n (B_1^{-1} B_2^{-1})^n = (R''_1 R_2 R_1^{-1} R_2^{-1} R'_1)^n (R''_1 R_2^{-1} R_1^{-1} R_2 R'_1)^n \tag{5.24}$$

is freely homotopic to $(R_1 R_2 R_1^{-1} R_2^{-1})^n (R_1 R_2^{-1} R_1^{-1} R_2)^n$, it is also freely homotopic to

$$Q := (Y_1 Y_2 Y_1^{-1} Y_2^{-1})^n (Y_1 Y_2^{-1} Y_1^{-1} Y_2)^n. \tag{5.25}$$

By (5.18(iii)-(v)), any lifting of Q does not cross any lifting of any C_i more than once. So we have (5.8)

$$\mathrm{cr}(G, (B_1 B_2)^n (B_1^{-1} B_2^{-1})^n) > \sum_{i=1}^{k} \mathrm{kr}(C_i, Q). \tag{5.26}$$

Now

$$\begin{aligned} \mathrm{cr}(G, (B_1 B_2)^n (B_1^{-1} B_2^{-1})^n) &= 4n \cdot \mathrm{cr}(G, R_1 R_2) \\ &= 2n \cdot \sum_{h=1}^{t} (\mathrm{cr}(G, D_h) - 1) \leq 4nt \end{aligned} \tag{5.27}$$

by (5.19). On the other hand,

$$\sum_{i=1}^{k} \mathrm{kr}(C_i, Q) = 4nt, \tag{5.28}$$

contradicting (5.26). □

Proposition 12 shows the correctness of the algorithm, and proves Theorems 1 and 2.

6. Disjoint Homotopic Trees

In this section we extend the method described in Section 3 to the *disjoint homotopic trees problem*:

given: a planar graph G embedded in $\mathbb{R}^2$;
a subset $\{I_1, \ldots, I_p\}$ of the faces of G (including the unbounded face);
paths $C_{11}, \ldots, C_{1t_1}, \ldots, C_{k1}, \ldots, C_{kt_k}$ in G, each with endpoints on the boundary of $I_1 \cup \cdots \cup I_p$, so that, for each $i = 1, \ldots, k$, $C_{i1}, \ldots, C_{it_i}$ begin in the same vertex; (6.1)
find: pairwise vertex-disjoint subtrees $T_1, \ldots, T_k$ of G so that, for each $i = 1, \ldots, k$ and $j = 1, \ldots, t_i$, T_i contains a path homotopic to C_{ij} in $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$.

Theorem 3. *The disjoint homotopic trees problem* (6.1) *is solvable in polynomial time.*

The polynomial-time algorithm for (6.1) consists of four basic steps similar to those for solving the disjoint homotopic paths problem:

I. Uncrossing $C_{11}, \ldots, C_{kt_k}$.
II. Constructing a system $Ax \le b$ of linear inequalities.
III. Solving $Ax \le b$ in integers. (6.2)
IV. Shifting $C_{11}, \ldots, C_{kt_k}$ and deducing trees $T_1, \ldots, T_k$.

We make similar assumptions to those in Section 3 (assumptions (3.1)):

(i) each edge of G is traversed at most once by the C_{ij};
(ii) the beginning vertex of any C_{ij} has degree t_i in G, while the end vertex has degree 1 in G; (6.3)
(iii) no edge traversed by any C_{ij}, except for the first and last edge of C_{ij}, is incident with a face in $\{I_1, \ldots, I_p\}$.

These conditions can be attained by adding new vertices and (parallel) edges. From (6.3(ii)) it follows that the common beginning vertex of $C_{i1}, \ldots, C_{it_i}$ is not traversed by any other $C_{11}, \ldots, C_{kt_k}$. The end vertex of any C_{ij} is not traversed by any other $C_{11}, \ldots, C_{kt_k}$.

I. Uncrossing $C_{11}, \ldots, C_{kt_k}$

This step modifies $C_{11}, \ldots, C_{kt_k}$ so that they have no (self-)crossings and no null-homotopic parts. We can proceed similarly as in the uncrossing step of Section 3.1. We should however be a little more careful as now different curves can have the same beginning vertex. It means that in some cases we must exchange not the parts between two crossings, but the parts between the common beginning vertex and a crossing.

More precisely, let

$$\begin{aligned} C_{ij} &= (v_0, e_1, v_1, \ldots, e_m, v_m), \\ C_{i'j'} &= (v'_0, e'_1, v'_1, \ldots, e'_{m'}, v'_{m'}). \end{aligned} \tag{6.4}$$

Again, if $(i.j) \neq (i',j')$ we call a pair (h, h') (with $1 \leq h \leq m-1$ and $1 \leq h' \leq m'-1$) a *crossing* if $v_h = v'_{h'}$ and e_h, $e'_{h'}$, e_{h+1}, $e'_{h'+1}$ occur in this order cyclically at v_h (clockwise or anticlockwise). Then we have:

Proposition 13. *Let* (6.1) *have a solution, let* $(i,j) \neq (i',j')$, *and let* (h, h') *be a crossing of* C_{ij} *and* $C_{i'j'}$. *Then there exists* (g, g') *so that*

$$\textit{part } (v_g, \ldots, v_h) \textit{ of } C_{ij} \textit{ is homotopic to part } (v'_{g'}, \ldots, v'_{h'}) \textit{ of } C_{i'j'} \tag{6.5}$$

and so that $(g, g') = (0, 0)$ *or* (g, g') *is a crossing of* C_{ij} *and* $c_{i'j'}$.

Proof. Similar to the proof of Proposition 4 (consider the universal covering space of $\mathbb{R}^2 \setminus (I_1 \cup \cdots \cup I_p)$). □

So if C_{ij} and $C_{i'j'}$ have a crossing, we can find (by Proposition 2), in polynomial time, pairs (g, g') and (h, h') so that (6.5) holds. After exchanging the two parts we arrive at a situation with fewer crossings. Repeating this, finally no two different C_{ij} and $C_{i'j'}$ have any crossing.

Self-crossings and null-homotopic parts can be removed just as in Section 3 (see Proposition 5). So we end up with $C_{11}, \ldots, C_{kt_k}$ without (self-)crossings and null-homotopic parts.

II. Constructing the System $Ax \leq b$ of Linear Inequalities

Again we introduce a variable each time a curve C_{ij} traverses a vertex. More precisely, let, for each $i = 1, \ldots, k$ and $j = 1, \ldots, t_i$,

$$C_{ij} = (v_{ij0}, e_{ij1}, v_{ij1}, \ldots, e_{ijm_{ij}}, v_{ijm_{ij}}). \tag{6.6}$$

We introduce a variable x_{ijh} for each $i = 1, \ldots, k$, $j = 1, \ldots, t_i$, and $h = 1, \ldots, m_{ij} - 1$. The values of these variables are going to determine the shifts of the C_{ij}.

Again we put linear constraints on the x_{ijh} in order to accomplish that the shifted C_{ij} can be combined to trees as required.

We denote by F^+_{ijh} and F^-_{ijh} the faces on the right-hand side and on the left-hand side, respectively, of e_{ijh} when going from v_{ijh-1} to v_{ijh}. As in Section 3, the curves C_{ij} give us the free adjacency relation between faces at any vertex v (except at the end vertices of each C_{ij}). This yields the auxiliary graph H, with length function on the edges, and with two mappings φ and ψ to $\mathbb{R}^2\backslash(I_1 \cup \cdots \cup I_p)$.

The inequalities in Class 1 are similar to those in Section 3:

Class 1. For each $i = 1, \ldots, k, j = 1, \ldots, t_i$, and $h = 1, \ldots, m_{ij} - 1$ we require:

$$\begin{aligned}&(\alpha)\ x_{ijh} \le \min_P \text{length}(P),\\ &(\beta)\ -x_{ijh} \le \min_P \text{length}(P).\end{aligned} \tag{6.7}$$

The minimum in (α) ranges over all paths P in H from $(v_{ijh}, \langle F^+_{ijh}\rangle)$ to any vertex (w, λ) of H with $\lambda \cap \{I_1, \ldots, I_p\} \neq \varnothing$. The minimum in ($\beta$) ranges over all paths P in H from $(v_{ijh}, \langle F^-_{ijh}\rangle)$ to any vertex (w, λ) of H with $\lambda \cap \{I_1, \ldots, I_p\} \neq \varnothing$.

Class 2 falls apart into two subclasses. Class 2A will assure that curves C_{ij} and $C_{i'j'}$ with $i \neq i'$ do not intersect:

Class 2A. For each $i, i' = 1, \ldots, k, j = 1, \ldots, t_i, h = 1, \ldots, m_{ij} - 1, j' = 1, \ldots, t_{i'}$, and $h' = 1, \ldots, m_{i'j'} - 1$ with $i \neq i'$ we require:

$$\begin{aligned}&(\alpha)\ x_{ijh} + x_{i'j'h'} \le \text{dist}_H((v_{ijh}, \langle F^+_{ijh}\rangle), (v_{i'j'h'}, \langle F^+_{i'j'h'}\rangle)) - 1,\\ &(\beta)\ x_{ijh} - x_{i'j'h'} \le \text{dist}_H((v_{ijh}, \langle F^+_{ijh}\rangle), (v_{i'j'h'}, \langle F^-_{i'j'h'}\rangle)) - 1,\\ &(\gamma)\ -x_{ijh} - x_{i'j'h'} \le \text{dist}_H((v_{ijh}, \langle F^+_{ijh}\rangle), (v_{i'j'h'}, \langle F^-_{i'j'h'}\rangle)) - 1.\end{aligned} \tag{6.8}$$

If $i = i', j \neq j'$, then the shifted C_{ij} and $C_{i'j'}$ may touch, but may not cross. This gives the Class 2B inequalities:

Class 2B. For each $i = 1, \ldots, k,\ j, j' = 1, \ldots, t_i\ (j \neq j'),\ h = 1, \ldots, m_{ij} - 1$, and $h' = 1, \ldots, m_{ij'} - 1$ we require:

$$\begin{aligned}&(\alpha)\ x_{ijh} + x_{i'j'h'} \le \text{dist}_H((v_{ijh}, \langle F^+_{ijh}\rangle), (v_{i'j'h'}, \langle F^+_{ij'h'}\rangle)),\\ &(\beta)\ x_{ijh} - x_{ij'h'} \le \text{dist}_H((v_{ijh}, \langle F^+_{ijh}\rangle), (v_{ij'h'}, \langle F^-_{ij'h'}\rangle)),\\ &(\gamma)\ -x_{ijh} - x_{ij'h'} \le \text{dist}_H((v_{ijh}, \langle F^-_{ijh}\rangle), (v_{ij'h'}, \langle F^-_{ij'h'}\rangle)).\end{aligned} \tag{6.9}$$

Finally, Class 3 inequalities are intended to avoid that C_{ij} and $C_{ij'}$ (possible $j = j'$) intersect each other around one of the holes $I_1, \ldots, I_p$:

Class 3. For each $i = 1, \ldots, k,\ j, j' = 1, \ldots, t_i,\ h = 1, \ldots, m_{ij} - 1$, and $h' = 1, \ldots, m_{ij'} - 1$ we require:

$$\begin{aligned}&(\alpha)\ x_{ijh} + x_{ij'h'} \le \min_P \text{length}(P) - 1,\\ &(\beta)\ x_{ijh} - x_{ij'h'} \le \min_P \text{length}(P) - 1,\\ &(\gamma)\ -x_{ijh} - x_{ij'h'} \le \min_P \text{length}(P) - 1.\end{aligned} \tag{6.10}$$

Here in (α) the minimum ranges over all paths P in H from $(v_{ijh}, \langle F_{ijh}^+\rangle)$ to $(v_{ij'h'}, \langle F_{ij'h'}^+\rangle)$ which are *not* homotopic to the following part of $C_{ij}^{-1}C_{ij'}$:

$$(v_{ijh}, \ldots, v_{ij0} = v_{ij'0}, \ldots, v_{ij'h'}) \tag{6.11}$$

(if $j = j'$, (6.11) is homotopic to part $(v_{ijh}, \ldots, v_{ijh'})$ of C_{ij}). Similarly for (β) and (γ).

This defines the inequality system $Ax \leq b$. Note that the same left-hand sides may occur among (6.9) and (6.10)—we can restrict ourselves to the ones with smallest right-hand side.

III. Solving $Ax \leq b$ in Integers

Since matrix A again has the property that the sum of the absolute values in any row is at most 2, we can solve $Ax \leq b$ in integers in the same way as we did in Section 3. We show here:

Proposition 14. *If (6.1) has a solution, then $Ax \leq b$ has an integer solution.*

Proof. Suppose (6.1) has a solution, i.e., disjoint trees $T_1, \ldots, T_k$ as required exist. We describe an integer solution z for $Ax \leq b$. Let U be the universal covering space of $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$, with projection function π, and let $\bar{G}$ be the (infinite) graph $\pi^{-1}[G]$. Choose $i = 1, \ldots, k$, $j = 1, \ldots, t_i$, and $h = 1, \ldots, m_{ij} - 1$. Let $\bar{C}_{ij}$ be some lifting of C_{ij} to U. Denote

$$\bar{C}_{ij} = (\bar{v}_{ij0}, \ldots, \bar{v}_{ijm_{ij}}), \tag{6.12}$$

where $\bar{v}$ is a lifting of v. As C_{ij} has no null-homotopic parts, $\bar{C}_{ij}$ is a simple path in $\bar{G}$.

Let Q be the unique path in T_i connecting v_{ij0} and $v_{ijm_{ij}}$. So Q and C_{ij} are homotopic. Hence there exists a lifting $\bar{Q}$ of Q to U so that $\bar{Q}$ is a simple path from $\bar{v}_{ij0}$ to $\bar{v}_{ijm_{ij}}$.

$\bar{Q}$ splits U into two parts (as $\bar{v}_{ij0}$ and $\bar{v}_{ijm_{ij}}$ are on the boundary of U): a part to the left of $\bar{Q}$ and a part to the right of $\bar{Q}$. We consider three cases.

Case 1: $\bar{v}_{ijh}$ is on $\bar{Q}$. Then define

$$z_{ijh} := 0. \tag{6.13}$$

Case 2: $\bar{v}_{ijh}$ is to the left of $\bar{Q}$ (see Fig. 24). Then define

$$z_{ijh} := \min_{D} \operatorname{cr}(\bar{G}, D) - 1, \tag{6.14}$$

where the minimum ranges over all curves D in U connecting $\bar{v}_{ijh}$ and any point on $\bar{Q}$.

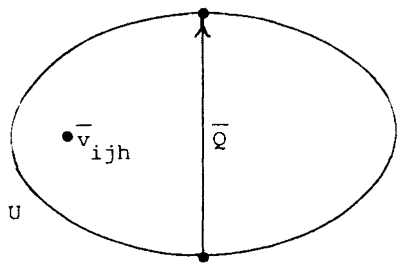

Fig. 24

Case 3. $\bar{v}_{ijh}$ is to the right of $\bar{Q}$ (see Fig. 25). Now define

$$z_{ijh} := -\left(\min_{D} \operatorname{cr}(\bar{G}, D) - 1\right), \tag{6.15}$$

where again the minimum ranges over all curves D in U connecting $\bar{v}_{ijh}$ and any point on $\bar{Q}$.

This defines the z_{ijh}. Note that by the symmetry of the universal covering surface, the values are independent of the choice of lifting $\bar{C}_{ij}$.

We show that the z_{ijh} form a solution to $Ax \leq b$.

Class 1 inequalities. By symmetry we need only check (6.7(α)). If $z_{ijh} \leq 0$ the inequality is trivially satisfied. If $z_{ijh} > 0$ we are in Case 2 above. Let P attain the minimum in (6.7(α)). Then $\psi \circ P$ is a curve from v_{ijh} to the boundary of $I_1 \cup \cdots \cup I_p$, starting via a face freely adjacent at v_{ijh} to F^+_{ijh} and not crossing any C_{ij}. Hence the lifting L of $\psi \circ P$ to U with $L(0) = \bar{v}_{ijh}$ has its endpoint on the boundary of U, on the right-hand side of $\bar{C}_{ij}$ or on $\bar{C}_{ij}$. Hence $L(0)$ is also on the right-hand side of $\bar{Q}$ or on $\bar{Q}$. So the lifting L' of $\varphi \circ P$ to U with $L'(0) = \bar{v}_{ijh}$ also has its endpoint on the boundary of U, on the right-hand side of $\bar{Q}$ or on $\bar{Q}$. So L' intersects Q. Therefore, by definition (6.14) of z_{ijh}:

$$z_{ijh} \leq \operatorname{cr}(\bar{G}, L') - 1 = \operatorname{length}(P). \tag{6.16}$$

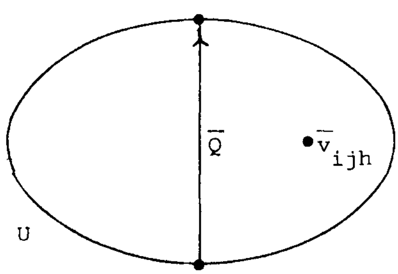

Fig. 25

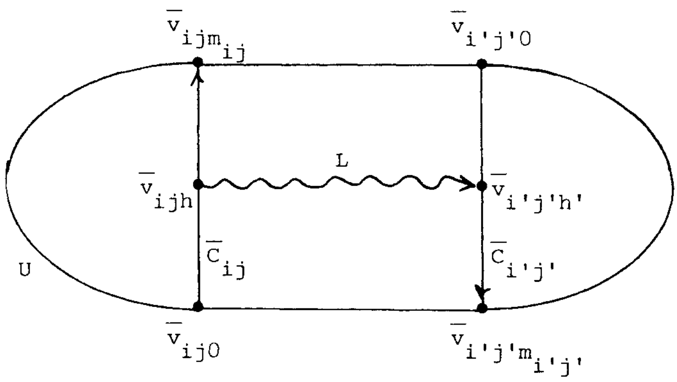

Fig. 26

Class 2A inequalities. By symmetry we need only check (6.8(α)). Let P be a shortest path in H from $(v_{ijh}, \langle F^+_{ijh}\rangle)$ to $(v_{i'j'h'}, \langle F^+_{i'j'h'}\rangle)$. Consider a lifting L of $\psi \circ P$ to U, connecting liftings $\bar{v}_{ijh}$ and $\bar{v}_{i'j'h'}$ say, of v_{ijh} and $v_{i'j'h'}$ respectively. Let $\bar{C}_{ij}$ and $\bar{C}_{i'j'}$ be liftings of C_{ij} and $C_{i'j'}$ so that the hth vertex of $\bar{C}_{ij}$ is $\bar{v}_{ijh}$ and the h'th vertex of $\bar{C}_{i'j'}$ is $\bar{v}_{i'j'h'}$.

Since $\psi \circ P$ starts via a face freely adjacent to F^+_{ijh} at v_{ijh} and ends via a face freely adjacent to $F^+_{i'j'h'}$ at $v_{i'j'h'}$ and since it does not cross any $C_1, \ldots, C_k$, we know that L runs on the right-hand side of $\bar{C}_{ij}$ and on the right-hand side of $\bar{C}_{i'j'}$ (see Fig. 26). Let Q be the simple path in T_i connecting v_{ij0} and $v_{ijm_{ij}}$ and let Q' be the simple path in $T_{i'}$ connecting $v_{i'j'0}$ and $v_{i'j'm_{i'j'}}$. Let $\bar{Q}$ and $\bar{Q}'$ be liftings of Q and Q' homotopic to $\bar{C}_{ij}$ and $\bar{C}_{i'j'}$ respectively. Again $\bar{Q}'$ is on the right-hand side of $\bar{Q}$, and $\bar{Q}$ is on the right-hand side of $\bar{Q}'$ (see Figure 27).

So U is decomposed into three regions A, B, and C as indicated, where we assume B to be open and A and C to be closed (so $\bar{Q}$ is in A and $\bar{Q}'$ is in C). We consider a number of cases depending on in which of the parts A, B, and C the

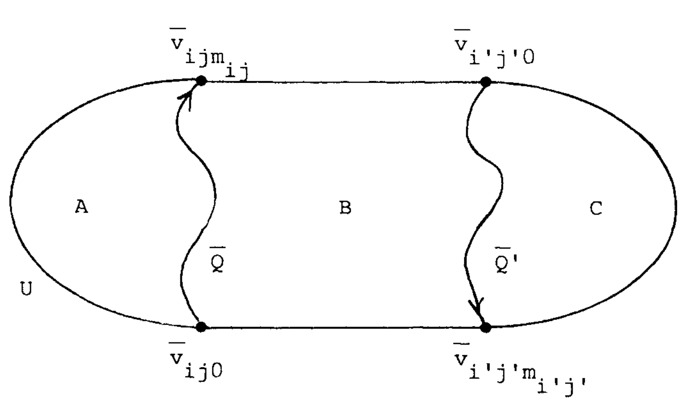

Fig. 27

points $\bar{v}_{ijh}$ and $\bar{v}_{i'j'h'}$ are located. The following fact is trivial but basic:

$$\text{for any curve } D \text{ in } U \text{ connecting } \bar{Q} \text{ and } \bar{Q}' \text{ we have} \quad \text{cr}(\bar{G}, D) \geq 2 \tag{6.17}$$

(since $\bar{Q}$ and $\bar{Q}'$ are disjoint, as T_i and $T_{i'}$ are disjoint). Let L' be the lifting of $\varphi \circ P$ connecting $\bar{v}_{ijh}$ and $\bar{v}_{i'j'h'}$.

Case A: $\bar{v}_{ijh} \in A$ *and* $\bar{v}_{i'j'h'} \in C$. Then L' can be decomposed as $D_1D_2D_3$ with $D_1(1) = D_2(0)$ on $\bar{Q}$ and $D_2(1) = D_3(0)$ on $\bar{Q}'$. By (6.17), $\text{cr}(\bar{G}, D_2) \geq 2$, and hence

$$\begin{aligned} z_{ijh} + z_{i'j'h'} &\leq (\text{cr}(\bar{G}, D_1) - 1) + (\text{cr}(\bar{G}, D_3) - 1) \\ &= \text{cr}(\bar{G}, L') - \text{cr}(\bar{G}, D_2) \leq \text{cr}(\bar{G}, L') - 2 = \text{length}(P) - 1. \end{aligned} \tag{6.18}$$

Case B: $\bar{v}_{ijh} \in B$ *and* $\bar{v}_{i'j'h'} \in C$. Then L' can be decomposed as D_2D_3 with $D_2(1) = D_3(0)$ on $\bar{Q}'$. Let D_1 attain the minimum in (6.15). Then, by (6.17), $\text{cr}(\bar{G}, D_1D_2) \geq 2$, and hence

$$\begin{aligned} z_{ijh} + z_{i'j'h'} &\leq (-\text{cr}(\bar{G}, D_1) + 1) + (\text{cr}(\bar{G}, D_3) - 1) \\ &= \text{cr}(\bar{G}, D_2D_3) - \text{cr}(\bar{G}, D_1D_2) \leq \text{cr}(\bar{G}, L') - 2 = \text{length}(P) - 1. \end{aligned} \tag{6.19}$$

Case C: $\bar{v}_{ijh} \in C$ *and* $\bar{v}_{i'j'h'} \in C$. Let D attain the minimum in (6.15). Then D can be decomposed as D_1D_2 with $D_1(1) = D_2(0)$ on $\bar{Q}'$. Then, by (6.17), $\text{cr}(\bar{G}, D_1) \geq 2$, and hence

$$\begin{aligned} z_{ijh} + z_{i'j'h'} &\leq (-\text{cr}(\bar{G}, D) + 1) + (\text{cr}(\bar{G}, D_2L') - 1) \\ &= \text{cr}(\bar{G}, L') - \text{cr}(\bar{G}, D_1) \leq \text{cr}(\bar{G}, L') - 2 = \text{length}(P) - 1. \end{aligned} \tag{6.20}$$

Case D: $\bar{v}_{ijh} \in A$ *and* $\bar{v}_{i'j'h'} \in B$. By symmetry similar to Case B.

Case E: $\bar{v}_{ijh} \in A$ *and* $\bar{v}_{i'j'h'} \in A$. By symmetry similar to Case C.

Case F: $\bar{v}_{ijh} \in B \cup C$ *and* $\bar{v}_{i'j'h'} \in A \cup B$. Then $z_{ijh} \leq 0$ and $z_{i'j'h'} \leq 0$, and hence trivially $z_{ijh} + z_{i'j'h'} \leq \text{length}(P) - 1$.

This shows (6.8(α)).

Class 2B inequalities. By symmetry we only consider (6.9(α)). They can be checked similarly to checking (6.8(α)) above. The only difference is that now Q and Q' can touch. So the liftings $\bar{Q}$ and $\bar{Q}'$ may also touch. Therefore instead of (6.17) we have

$$\text{for any curve } D \text{ in } U \text{ connecting } \bar{Q} \text{ and } \bar{Q}' \text{ we have} \quad \text{cr}(\bar{G}, D) \geq 1. \tag{6.21}$$

Hence we get $z_{ijh} + z_{i'j'h'} \leq \text{length}(P)$ instead of $\leq \text{length}(P) - 1$.

Class 3 inequalities. By symmetry we only consider (6.10(α)). Again checking this is similar to checking (6.8(α)). As path P attaining the minimum in (6.10(α)) is not homotopic to part $(v_{ijh}, \ldots, v_{ij0} = v_{ij'0}, \ldots, v_{ij'h'})$ of $C_{ij}^{-1}C_{ij'}$, we know that the lifting L of $\psi \circ P$ connects *disjoint* liftings $\bar{Q}$ and $\bar{Q}'$. So we can proceed as for Class 2A inequalities. □

IV. Shifting $C_{11}, \ldots, C_{kt_k}$ and Obtaining $T_1, \ldots, T_k$

We finally shift the C_{ij} using the integer solution x_{ijh} to $Ax \leq b$, and derive from the shifted C_{ij} the trees $T_1, \ldots, T_k$.

First assume $x_{ijh} = 0$ for all i, j, h. For each $i = 1, \ldots, k$, let T_i be any spanning tree in the subgraph of G made up by the vertices and edges occurring in $C_{i1}, \ldots, C_{it_i}$. We show:

Proposition 15, *$T_1, \ldots, T_k$ form a solution to the disjoint homotopic trees problem* (6.1).

Proof. First note that Class 2A and 3 inequalities imply that if C_{ij} and $C_{i'j'}$ have a vertex in common, say $v_{ijh} = v_{i'j'h'}$, then $i = i'$ and

$$\text{part } (v_{ij0}, \ldots, v_{ijh}) \text{ of } C_{ij} \text{ is homotopic to part } (v_{ij'0}, \ldots, v_{ij'h'}) \text{ of } C_{ij'}. \quad (6.22)$$

In particular, if moreover $j = j'$, then part $(v_{ijh}, \ldots, v_{ijh'})$ of C_{ij} is null-homotopic, and hence $h = h'$.

It follows that $T_1, \ldots, T_k$ are pairwise vertex-disjoint. Next we show that for each i, j the unique simple path in T_i from v_{ij0} to $v_{ijm_{ij}}$ is homotopic to C_{ij}. In fact we show that for each i, j, h the unique simple path P_{ijh} in T_i from v_{ij0} to v_{ijh} is homotopic to part $(v_{ij0}, \ldots, v_{ijh})$ of C_{ij}. This is done by induction on the number of edges in P_{ijh}.

If P_{ijh} has length 0, the statement is trivial. If P_{ijh} has at least one edge, consider the last edge e of P_{ijh}. As it is in one of the paths $C_{i1}, \ldots, C_{it_i}$ there exist j', h' so that

$$e = e_{ij'h'}, \qquad v_{ijh} = v_{ij'h'}. \quad (6.23)$$

Now $P_{ijh'-1}$ is shorter than P_{ijh} and hence by the induction hypothesis it is homotopic to part $(v_{ij'0}, \ldots, v_{ij'h'-1})$ of $C_{ij'}$. Therefore, P_{ijh} is homotopic to part $(v_{ij'0}, \ldots, v_{ij'h'})$ of $C_{ij'}$. Then by (6.22) we know that P_{ijh} is homotopic to part $(v_{ij0}, \ldots, v_{ijh})$ of C_{ij}. □

Suppose next

$$M := \max\{|x_{ijh}| \mid i = 1, \ldots, k; j = 1, \ldots, t_i; h = 1, \ldots, m_{ij}\} > 0, \quad (6.24)$$

and suppose $x_{ijh} = M$ for some i, j, h. Like in Section 3, consider $e_{ijh}, v_{ijh}, e_{ijh+1}$ and the faces and edges incident "on the right-hand side" (see Fig. 28).

Note that $F_1, \ldots, F_s \notin \{I_1, \ldots, I_p\}$ by Class 1 inequalities. We claim:

$$\textit{we may assume that } \varepsilon_1, \ldots, \varepsilon_{s-1} \textit{ are not used by } C_{11}, \ldots, C_{kt_k}. \quad (6.25)$$

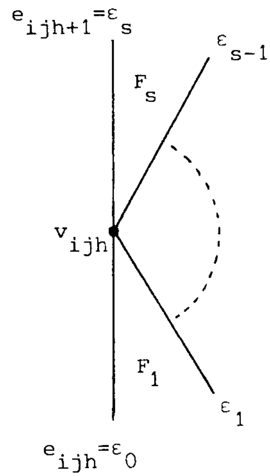

Fig. 28

Proof of (6.25). Suppose

$$\varepsilon_g = e_{i'j'h'}, \qquad v_{ijh} = v_{i'j'h'}, \qquad \varepsilon_{g'} = e_{i'j'h'+1} \tag{6.26}$$

for some i', j', h' and $g, g' \in \{i, \ldots, s-1\}$. We may assume that $\varepsilon_{g''}$ is not traversed by $C_{11}, \ldots, C_{kt_k}$ if $1 \leq g'' < \min\{g, g'\}$.

Suppose first $g > g'$. Then

$$x_{ijh} + x_{i'j'h'} \leq -1 \tag{6.27}$$

if $i \neq i'$, or if $i = i'$ and part $(v_{ij0}, \ldots, v_{ijh})$ of C_{ij} is not homotopic to part $(v_{ij'0}, \ldots, v_{ij'h'})$ of $C_{ij'}$ (by Class 2A and 3 inequalities). However, (6.27) implies $x_{i'j'h'} \leq -x_{ijh} - 1 = -M - 1$, contradicting (6.24).

Moreover, we have, if $i = i'$,

$$x_{ijh} + x_{ij'h'} \leq 0 \tag{6.28}$$

if part $(v_{ij0}, \ldots, v_{ijh})$ of C_{ij} is homotopic to part $(v_{ij'0}, \ldots, v_{ij'h'})$ of $C_{ij'}$ (by Class 2B inequalities). This however implies that either $v_{ijm_{ij}}$ or $v_{ij'm_{ij'}}$ is in the interior of the closed curve formed by $(v_{ij0}, \ldots, v_{ijh})$ and $(v_{ij'0}, \ldots, v_{ij'h'})$. As this closed curve is null-homotopic and as $v_{ijm_{ij}}$ and $v_{ij'm_{ij'}}$ are on the boundary of $I_1 \cup \cdots \cup I_p$, this is a contradiction.

So we know $g < g'$. Then $x_{ijh} - x_{i'j'h'} \leq 0$ (by Class 2A, 2B, and 3 inequalities). Hence also $x_{i'j'h'} = M$. Replacing i, j, h by i', j', h' decreases the "opening" (i.e., the number s of faces on the right-hand side in Fig. 28). After a finite number of such replacements we are in a situation as claimed in (6.25). □

Knowing (6.25), we can shift C_{ij} at v_{ijh} as in Section 3, and similarly if $x_{ijh} = -M$. As in Proposition 9 we show that the number of iterations is polynomially bounded, and hence we have a polynomial-time alagorithm. This proves Theorem 3.

7. Disjoint Trees

We finally consider the *disjoint trees problem*:

$$\begin{aligned}&\textit{given}: \text{a graph } G;\\ &\qquad\text{subsets } W_1, \dots, W_k \text{ of } V(G);\\ &\textit{find}: \text{pairwise vertex-disjoint subtrees } T_1, \dots, T_k \text{ in } G \text{ so that } W_i \subseteq V(T_i)\\ &\qquad\text{for } i = 1, \dots, k.\end{aligned} \tag{7.1}$$

This problem is NP-complete. Robertson and Seymour showed that for fixed $|W_1 \cup \cdots \cup W_k|$ there exists a polynomial-time algorithm for (7.1). We show that if G is planar, it suffices to fix the number of faces necessary to cover $W_1 \cup \cdots \cup W_k$. This is derived from Theorem 3, essentially by enumerating "homotopy classes" of trees.

For any connected planar graph G and any choice of faces $I_1, \dots, I_p$ of G, call two spanning trees T_1 and T_2 of G *equivalent* (with respect to $I_1, \dots, I_p$) if, for any two vertices u, w on $\text{bd}(I_1 \cup \cdots \cup I_p)$, the unique u–w path in T_1 is homotopic to the unique u–w path in T_2 in the space $\mathbb{R}^2 \setminus (I_1 \cup \cdots I_p)$.

We study enumerating equivalence classes of spanning trees. In fact, we enumerate *representatives* for these classes, i.e., we enumerate trees $B_1, \dots, B_N$ so that each equivalence class intersects $\{B_1, \dots, B_N\}$.

Proposition 16. *For each fixed p, we can enumerate, in polynomial time, representatives for the equivalence classes of spanning trees, for any connected planar graph G and any choice of faces $I_1, \dots, I_p$ of G.*

Proof. Let $W := V(G) \cap \text{bd}(I_1 \cup \cdots \cup I_p)$. Draw a graph G^* dual to G. So in each face F of G we put a vertex F^* of G^*. If F_1 and F_2 have an edge e in common, we connect F_1^* and F_2^* by an edge of G^* crossing e.

I. We first show that, for each $j = 2, \dots, p$, we can enumerate, in polynomial time, I_1^*-I_j^* paths $P_1, \dots, P_M$ in G^*, so that each simple I_1^*-I_j^* path in G^* is homotopic to at least one path among $P_1, \dots, P_M$ in the space $\mathbb{R}^2 \setminus W$. Without loss of generality, $j = 2$.

Consider the set E_1 of edges of G on $\text{bd}(I_1 \cup \cdots \cup I_p)$. Let E_2 be an inclusionwise minimal set of edges so that $E_1 \cup E_2$ forms a connected graph on the set V_0 of vertices covered by $E_1 \cup E_2$. Note that the edges in E_2 form a forest.

Let V_1 be the set of vertices that are not in $\text{bd}(I_1 \cup \cdots \cup I_p)$ and that are incident with at least three edges in E_2. Then the graph $(V_0, E_1 \cup E_2)$ is topologically homeomorphic to a graph H with vertex set $W \cup V_1$ and edge set $E_1 \cup \{q_1, \dots, q_r\}$ for some edges $q_1, \dots, q_r$ (which come from paths in E_2). So each vertex in V_1 has degree at least three in H. This implies $r \leq 2p - 3$. (To see this, contract the edges in E_1, making H a tree with r edges. Let the contracted $\mathrm{bd}(I_1 \cup \cdots \cup I_p)$ give p' vertices. So $p' \leq p$. Then $r = p' + |V_1| - 1$. On the other hand, all vertices in V_1 have degree at least 3. So $2r \geq 3|V_1| + p' = 3r - 2p' + 3$, implying $r \leq 2p' - 3 \leq 2p - 3$.)

Since H is connected, each face of H is simply connected. Note that $q_1, \ldots, q_r$ all are incident (at both sides) with only one face of H, call it F_0.

We enumerate representatives for the homotopy classes containing simple I_1^*-I_2^* curves, so that each face among $I_1, \ldots, I_p$ is traversed at most once, and so that face F_0 is traversed at most $m := |E(G)|$ times (homotopy in the space $\mathbb{R}^2 \setminus W$). This clearly includes all homotopy classes containing a simple I_1^*-I_2^* path in G^*.

To enumerate the curves, we first decide how often it crosses each of the edges of H. To this end, we decide, for $j = 3, \ldots, p$, whether I_j is traversed or not. If we decide I_j is traversed, then we choose two edges on $\mathrm{bd}(I_j)$ to be crossed by the curve. Moreover, we choose one edge on $\mathrm{bd}(I_1)$ and one edge on $\mathrm{bd}(I_2)$ to be crossed. These choices can be made in $O(m^{2p})$ ways. For each edge we decided is crossed, we consider a "little" line segment crossing this edge.

This fixes the crossings of the curves with the edges in E_1. To fix crossings with $q_1, \ldots, q_r$, we choose, for each $j = 1, \ldots, r$, a number α_j, indicating how often edge q_j is crossed. We take $0 \leq \alpha_j \leq m$. So this choice can be made in $O(m^r)$ ways. For each j we consider α_j "little" line segments crossing q_j.

We take all "little" line segments pairwise disjoint. Let $\mathscr{L}$ denote the set of all these line segments, and let R denote the set of endpoints of these line segments (so $|R| = 2|\mathscr{L}|$). Let R'_j and R''_j be the sets of endpoints of these line segments crossing q_j, at the two sides of q_j (see Fig. 29).

For $j = 1, 2$ we consider a curve in I_j connecting I_j^* with the unique point in $R \cap I_j$. For $j = 3, \ldots, p$, if $|R \cap I_j| = 2$, we consider a curve in I_j connecting the two points in $R \cap I_j$. In face F_0 we connect the points in $R \cap F_0$ pairwise, by pairwise disjoint curves (not crossing any line segment in $\mathscr{L}$), in such a way that no two points both in the same R'_j or both in the same R''_j are connected. Such a "matching" can be chosen in $O(m^{4r+4p})$ ways.

This bound can be seen as follows. Let $\mathscr{C}$ be the partition of $R \cap F_0$ with classes $R'_1, R''_1, \ldots, R'_r, R''_r$, together with singletons for the remaining points in $R \cap F_0$. Note that $|\mathscr{C}| \leq 2r + 2p$. For any two distinct classes γ, δ in $\mathscr{C}$ we choose a number $\beta_{\gamma\delta}$ indicating how many points in γ are to be matched to points in δ. We take $\beta_{\gamma\delta} \leq m$, and hence the choice can be made in $O(m^{4r+4p})$ ways. In fact, we consider

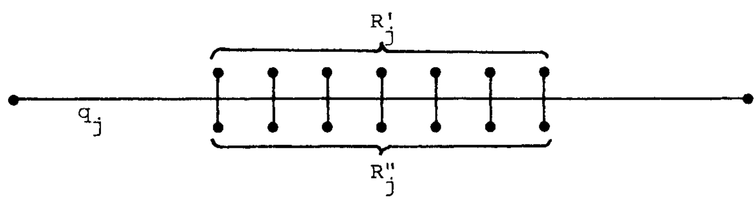

Fig. 29

only those choices for which

$$\sum_{\delta \neq \gamma} \beta_{\gamma\delta} = |\gamma| \tag{7.2}$$

holds for each $\gamma \in \mathscr{C}$. Then, for each γ, δ, we know which points in γ are matched to which points in δ. (This follows from the fact that F_0 is simply connected and that two distinct curves must be disjoint.) We only consider those choices $\beta_{\gamma\delta}$ for which this matching yields pairwise noncrossing curves.

Finally, in each face F of H with $F \notin \{I_1, \ldots, I_p, F_0\}$, we consider pairwise disjoint curves, pairwise connecting the points in $R \cap F$. Since $|R \setminus F_0| \leq 4p$, there is a constant number of such choices (as p is fixed).

Now all line segments and curves chosen yield a curve C from I_1^* to I_2^*, together with some (or none) closed curves. It is not difficult to replace C by a path P in G^* homotopic to C (path P need not be simple). All paths P thus generated, give our enumeration.

Clearly, each simple I_1^*-I_2^* path in G^* is homotopic to at least one of these paths.

II. We now enumerate spanning trees of G, covering all equivalence classes with respect to $I_1, \ldots, I_p$. By the first part of this proof we can enumerate, for each $j = 2, \ldots, p$, I_1^*-I_j^* paths $P_{j1}, \ldots, P_{jM_j}$ in G^* so that each simple I_1^*-I_j^* path is homotopic to at least one of them (in $\mathbb{R}^2 \setminus W$).

For each choice $i_2, \ldots, i_p$ with $1 \leq i_2 \leq M_2, \ldots, 1 \leq i_p \leq M_p$ we can find, in polynomial time (by Theorem 3), a tree T in G^* connecting $I_1^*, \ldots, I_p^*$ so that the simple I_1^*-I_j^* path in T is homotopic to P_{ji_j} ($j = 2, \ldots, p$), provided that such a tree exists. Choose an arbitrary spanning tree B in G not intersecting T.

We prove that the spanning trees thus obtained intersect all equivalence classes. Let B' be any spanning tree in G. Then, for each $j = 2, \ldots, p$, there exists a unique simple I_1^*-I_j^* path Q_j in G^* not intersecting B'. Without loss of generality, let Q_j be homotopic to P_{j1} in $\mathbb{R}^2 \setminus W$ ($j = 2, \ldots, p$). Let T_0 be the unique spanning tree in G^* not intersecting B'. So the choice $i_2 = 1, \ldots, i_p = 1$ indeed gives us a tree T in G^* connecting $I_1^*, \ldots, I_p^*$ so that the simple I_1^*-I_j^* path in T is homotopic to P_{j1} in $\mathbb{R}^2 \setminus W$ ($j = 2, \ldots, p$). Let B be the chosen spanning tree in G not intersecting T.

Then, for each u, $w \in W$, any u-w path not intersecting $Q_2 \cup \cdots \cup Q_p$ is homotopic to any u-w path not intersecting T, in $\mathbb{R}^2 \setminus (I_1 \cup \cdots \cup I_p)$. Hence B' is equivalent to B. □

We finally derive:

Theorem 4. *For each fixed p there exists a polynomial-time algorithm for the disjoint trees problem* (7.1) *when G is planar and $W_1 \cup \cdots \cup W_k$ can be covered by the boundaries of p faces of G.*

Proof. Let G be a planar graph, and let $W_1, \ldots, W_k$ be subsets of $V(G)$ so that $W_1 \cup \cdots \cup W_k \subseteq \text{bd}(I_1 \cup \cdots \cup I_p)$ for faces $I_1, \ldots, I_p$ of G. We may assume that the unbounded face is included in $\{I_1, \ldots, I_p\}$, that G is connected, and that $W_1, \ldots, W_k$ are nonempty and pairwise disjoint. Choose $w_1 \in W_1, \ldots, w_k \in W_k$ arbitrarily.

We enumerate spanning trees $B_1, \ldots, B_N$ of G covering all equivalence classes with respect to $I_1, \ldots, I_p$. By Proposition 16, this can be done in polynomial time.

For each tree B_j we do the following. For each $i = 1, \ldots, k$ and each $w \in W_i \backslash \{w_i\}$, let C_{iw} be the simple w_i-w path in B_j. With the algorithm of Theorem 3 we solve the problem:

find: pairwise vertex-disjoint subtrees $T_1, \ldots, T_k$ of G so that, for each $i = 1, \ldots, k$ and each $w \in W_i \backslash \{w_i\}$, T_i contains a w_i-w path homotopic to C_{iw} (in $\mathbb{R}^2 \backslash (I_1 \cup \cdots \cup I_p)$). (7.3)

If, for some B_j, (7.3) has a solution, it clearly is a solution to (7.1). We show that, conversely, if (7.1) has a solution, then (7.3) has a solution for at least one B_j. Let $T_1, \ldots, T_k$ be a solution to (7.1). Extend $T_1 \cup \cdots \cup T_k$ to a spanning tree B of G. Then B is equivalent to spanning tree B_j for at least one j. Then, for this j, problem (7.3) has a solution (namely $T_1, \ldots, T_k$). $\square$

Acknowledgements

I thank András Frank, Jeff Kahn, Laci Lovász, and Paul Seymour for very stimulating discussions on the topic of this paper. It was a conjecture of Lovász and Seymour that conditions (1.3) are sufficient. Moreover, I thank the referees for carefully reading the text and for giving several helpful suggestions improving the presentation.

References

1. R. Cole and A. Siegel, River routing every which way, but loose, *Proceedings of the 25th Annual Symposium on Foundations of Computer Science*, 1984, pp. 65–73.
2. A. Frank and A. Schrijver, Vertex-disjoint simple paths of given homotopy in a planar graph, in: *Polyhedral Combinatorics* (W. Cook and P. D. Seymour, eds), American Mathematical Society, Providence, RI, 1990, pp. 139–161.
3. C. E. Leiserson and F. M. Maley, Algorithms for routing and testing routability of planar VLSI layouts, *Proceedings of the 17th Annual ACM Symposium on Theory of Computing*, 1985, pp. 69–78.
4. W. S. Massey, *Algebraic Topology: An Introduction*, Springer-Verlag, New York, 1967.
5. R. Y. Pinter, River routing: methodology and analysis, *Third CalTech Conference on Very Large Scale Integration* (R. Bryant, ed.), Springer-Verlag, Berlin, 1983, pp. 141–163.
6. N. Robertson and P. D. Seymour, Graph minors, VI. Disjoint paths across a disc, *J. Combin. Theory Ser. B* **41** (1986), 115–138.
7. N. Robertson and P. D. Seymour, Graph minors, VII. Disjoint paths on a surface, *J. Combin. Theory Ser. B* **45** (1988), 212–254.
8. N. Robertson and P. D. Seymour, Graph minors, XV, to appear.
9. A. Schrijver, Disjoint circuits of prescribed homotopies in a graph on a compact surface, *J. Combin. Theory Ser. B* **51** (1991), 127–159.

Received March 24, 1989, and in revised form October 30, 1989, and January 24, 1990.