Annals of Mathematics and Artificial Intelligence, 3 (1991) 21–82 

# PLANNING CONSTRAINED MOTION

Steven FORTUNE and Gordon WILFONG

*AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA*

**Abstract**

We consider the motion planning problem for a point constrained to move along a path with radius of curvature at least one. The point moves in a two-dimensional universe with polygonal obstacles. We show the decidability of the reachability question: "Given a source placement (position and direction pair) and a target placement, is there a curvature-constrained path from source to target avoiding obstacles?" The decision procedure has time and space complexity $2^{O(poly(n,m))}$, where $n$ is the number of corners and $m$ is the number of bits required to specify the position of corners.

## 1. Introduction

Anyone who has ever tried to parallel park an automobile is familiar with the problem of planning constrained motion. The difficulty of planning such motions is twofold. While an automobile can move forward (or backward) and can turn, its turning radius has a lower bound defined by the steering mechanism. Thus the motion of the automobile must be collision-free and must conform to the automobile's turning radius constraint.

We study a very simple model of the motion of an automobile. The model is a directed point moving in a compact two-dimensional universe $U$. The directed point can move forward (in the sense of its associated direction) and it can change its direction at a rate (with respect to path length) bounded by some fixed maximum. We model the requirement that the path avoid obstacles by requiring that the path stay within $U$. We assume that $U$ has polygonal boundary; however, $U$ need not be simply connected, that is, $U$ may have two-dimensional holes. A *placement* of the point is a pair consisting of a position and a direction. We wish to answer the following reachability question: "Given a source placement and a target placement of the point is there a motion of the point within $U$ from the source placement to the target placement that satisfies the constraint on the rate of turning?". The main result of this paper is a decision procedure for this reachability question. The time and space complexity of the procedure is exponential ($2^{O(poly(n,m))}$) where $n$ is the number of vertices of the boundary of $U$, $m$ is the number of bits needed to specify the positions of the vertices, and $poly(n, m)$ is a polynomial in $n$ and $m$.

The bound on rate of turning is one kind of constraint on the possible motions of an object. Other types of constraints are bounds on velocity or acceleration. One reason for studying constraints on motion is to extend the class of solvable motion planning problems beyond what has been studied by classical motion planning.

Classical motion planning, as studied by Schwartz and Sharir [14] and others, addresses the problem of finding a collision-free path for an object. The object moves through a universe filled with obstacles. Both the object and the obstacles are assumed to be known ahead of time. A point in *configuration space* specifies a placement of the object; configuration space has as many dimensions as there are degrees of freedom of motion of the object. *Free space* is the subset of configuration space corresponding to placements of the object not intersecting any obstacle. Schwartz and Sharir [14] give an algorithm that constructs a representation of free space from a description of the object and the obstacles. Their algorithm is very general, allowing arbitrary rotations, translations, linkages, etc. The running time of the algorithm is doubly exponential in the dimension of configuration space. Recently Canny [3] obtained an algorithm that constructs a lower-dimensional "roadmap" of free space; his algorithm is (singly) exponential in the number of dimensions of free space. Much of the literature on motion planning can be viewed as efficient algorithms for special cases of the motion planning problem.

Construction of free space is essentially all that is required to solve motion planning problems if there are no constraints on the motion of the object. This is because there is a collision-free path of the object from one placement to another exactly if the corresponding points in free space are (topologically) path-connected. Testing path-connectness of two points is usually easy given the representation of free space (in fact actually constructing the path is also relatively easy).

Construction of free space is not sufficient to solve motion planning with constraints, however. For example, free space for the directed point is very simple. It is just the Cartesian product of $U$ with the space of all directions. There is a one–one correspondence between the path-connected components of free space and the path-connected components of $U$. Unfortunately, path connectedness in free space does not imply reachability because of the curvature constraint on motions. In fact two configurations in free space can be arbitrarily close but neither is reachable from the other.

Two recent papers have considered kinematic constraints on motion. They both construct a *time–configuration space*, with an extra dimension for time. Reif and Sharir [13] consider the problem of planning the motion of an object with a velocity bound. All paths must be monotonic in the time dimension; a velocity constraint corresponds to a bound on the slope of a path. O'Dunlaing [10] studies the problem of a point moving on a line constrained by an acceleration bound. He gives an algorithm to decide if the point can avoid obstacles moving with known trajectories. These two algorithms are obtained by showing that it is

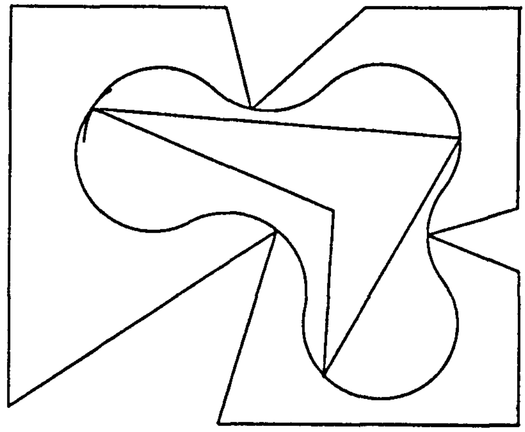

Fig. 1. Cycling behavior.

sufficient to consider certain "normalized" paths between "features" in time–configuration space. Notice that paths in time–configuration space never return to the same feature (because of time monotonicity). This is a substantial difference from the planning problem for the directed point. It is quite possible that the only path from source to target visits an obstacle more than once (see fig. 1), indeed arbitrarily many times. This substantially complicates the decision procedure.

We remark that we carefully chose to plan curvature-constrained motion for a point. Planning curvature-constrained motion for, say, a convex polygon is a substantially harder problem. Indeed it may not even be decidable. It is not sufficient to construct free space for the convex polygon and then plan a curvature-constrained motion through it. This is discussed in section 7. Other comments on the constrained motion planning problem also appear in section 7. These include planning shortest paths, planning paths with reversals, and other constraints.

### 1.1. OVERVIEW OF THE DECISION PROCEDURE

The decision procedure for curvature-constrained motion is quite complicated. We now give an overview of it, which provides an overview of the remainder of the paper. The following remarks are a somewhat simplified form of the decision procedure as it actually appears.

The first step in the decision procedure is a normalization theorem. We show that if there is a path from source placement to target placement, then there is a path that consists of a finite sequence of jumps. A jump is a path consisting of a unit circular arc, either left or right, a straight segment, and another unit circular arc. Furthermore, the start of the first jump is the source placement, the end of the last jump is the target placement, and every other endpoint of a jump is at an obstacle, either tangent to a wall or grazing a corner. (A *wall* is a line segment contained in the boundary of $U$, a *corner* is a vertex of the boundary of $U$.) An important consequence of the normalization theorem is that it suffices to compute just the placements reachable from the source placement at walls and corners. For then we can simply test if the target placement can be reached by a jump from some such reachable placement at a wall or corner.

The placements at a wall or corner can obviously be structured into intervals. An interval at a corner is the Cartesian product of the position of the corner with an interval of directions. An interval at a wall is the Cartesian product of a subsegment of the wall with one of the two directions along the wall. Clearly, if any placement at a wall is reachable, then any placement further along the wall in the same direction is also reachable. Hence there can be at most two intervals of reachable placements at a wall, one in each direction.

The second major step in the decision procedure is to bound the number of disjoint reachable intervals at corners. This is the heart of the decision procedure and indeed is by far the hardest part of the proof. We show that every reachable interval at a corner either contains a "root" or a "self-dual" interval. A "root" is a placement determined by the geometry of obstacles. Examples of roots are the final placement of a straight-line path connecting two corners and the final placement of a unit circular arc connecting two corners. The number of roots is polynomial in the number of corners. A "self-dual" interval at a corner $b$ is determined by a placement at a wall or corner $a$. The left endpoint of the self-dual interval is the placement at $b$ reached from the placement at $a$ by following a right–left hop, that is, a path consisting of a unit right arc followed by a unit left arc. Similarly the right endpoint of the self-dual interval is reached by a left–right hop (see fig. 6b). All arcs are constrained to be of length at most $\pi$, hence for some placements at $a$ there may not be a corresponding self-dual interval at $b$. The principal theorem is that in a collection of disjoint self-dual intervals at $b$ from $a$, the intervals can be ordered so that their size increases geometrically. We can show that the smallest interval must have size at least $2^{-\Omega(m)}$. Hence there can be at most $O(m)$ disjoint self-dual intervals at $b$ from $a$. Since there are only $O(n)$ possible choices of $a$ for each corner $b$, there can be at most polynomially many disjoint self-dual intervals. Since any reachable interval of placements at a corner contains a root or a self-dual interval, there are at most polynomially many disjoint reachable intervals.

The first step in attaining this bound is the proof that any reachable interval of placements at a corner contains a root or a self-dual interval. We define a procedure *stretch* that deforms a jump from a wall or corner $a$ to a corner $b$. The result is a family of paths from $a$ to $b$ that are homotopic to the jump, avoid obstacles, and satisfy the curvature constraint. Every path has the same starting placement at $a$; different paths have differing final directions at $b$. The goal of *stretch* is to produce as large an interval of placements at $b$ as possible. We can show that when *stretch* terminates, the interval of placements at $b$ contains either a root or a self-dual interval. Now suppose some placement at $b$ is reachable from the source placement. By the normalization theorem there is a path whose final portion is a jump from a wall or corner $a$ to $b$. By applying *stretch* to this last jump, we see that the placement at $b$ is in fact contained in a reachable interval containing a root or a self-dual interval. Procedure *stretch* is conceptually quite simple: a continuous deformation of the jump that avoids obstacles and satisfies the curvature constraint. Unfortunately, the details of the analysis of *stretch* are quite hard.

The second step in attaining the bound on the number of intervals is the proof that the size of disjoint self-dual intervals grows geometrically. Crucial to this proof is the function relating input direction to output direction of hops between two points at a fixed distance apart. The algebraic equations describing this function are quite complex, and it seems intractible to obtain and manipulate an explicit formula for it. We give a simple formula for the derivative of this function. With this formula we can carry out the required analysis: essentially we show that the function is convex, hence can be adequately approximated by a two-piece linear interpolant. We remark that the hop function is a special case of the equations arising from "four-bar linkages" in kinematics [15]. The derivative formula appears to be new.

Once the bound on the number of intervals is attained, there are two ways to obtain the decision procedure for the reachability question. The easiest is to encode the reachability question as a formula of $\mathbb{R}(+, \times)$, the first order theory of the reals with addition and multiplication. We can characterize the reachable intervals as the fixed point of the "one jump" mapping from sets of intervals to sets of intervals. Notice that because of the possibility of a cycle from a corner back to itself (fig. 1), it is necessary to use fixed points. The dominant term in the complexity of the formula is the number of variables. Representing each reachable interval requires a constant number of variables; since there are polynomially many reachable intervals, polynomially many variables are needed. This gives a double-exponential time decision procedure. Notice that some intervals of reachable placements may be open at an endpoint; this is because placements arbitrarily close to the endpoint are reachable, but the endpoint itself is not. However, it is easy to encode intervals with open endpoints in $\mathbb{R}(+, \times)$.

The resource bounds can be reduced by a second more complicated decision procedure. This procedure is essentially transitive closure of reachable intervals, using *stretch* to propagate intervals. We label each endpoint of an interval by the path that reaches it. The numerical value of the endpoint can be expressed using a formula of $\mathbb{R}(+, \times)$; hence manipulations of labels involve calls to the decision procedure for $\mathbb{R}(+, \times)$. The crucial point is that we need only logarithmically many variables of $\mathbb{R}(+, \times)$ in the formulas for endpoint labels. For technical reasons, the formulas have exponential length. However, deciding such a formula of $\mathbb{R}(+, \times)$ takes only exponential time. We show that the transitive closure terminates after only exponentially many steps (we need to test for cycles explicitly; the fixed point of a cycle can be represented as a formula of $\mathbb{R}(+, \times)$). Hence the improved decision procedure takes (singly) exponential time and space.

The next section, section 2, contains the definitions used throughout this paper. Section 3 is concerned with jumps. Section 3.1 contains the normalization theorem, proved in two steps: first showing that a curvature-constrained path can be modified to only contain straight segments and circular arcs, then further modified so that all circular arcs touch obstacles. Section 3.2 considers the collection of jumps from one wall or corner to another. These form a finite set of equivalence classes under a natural notion of restricted homotopy. In section 3.3 we characterize the reachable intervals as a fixed point of a mapping defined using jumps from sets of intervals to sets of intervals.

Section 4 describes procedure *stretch* and the analysis of the intervals resulting from *stretch*. Section 4.1 describes *stretch* informally; the formal definition and proofs about *stretch* are contained in section 4.2. Section 4.3 contains an easy classification of the intervals resulting from *stretch* into self-dual, saturated (i.e. containing roots) and "heterodual". Section 4.4 gives the proof of the main theorem about self-dual intervals, that they increase in size geometrically. Section 4.5 shows that any "heterodual" interval actually contains a self-dual interval, hence the size of heterodual intervals need not be analyzed separately.

Section 5 gives the easy reachability decision procedure. First we bound the number of disjoint reachable intervals, using the results proved in section 4. Then we construct a formula of $\mathbb{R}(+, \times)$ expressing the reachability question. Section 6 gives the improved decision procedure. Section 6.1 describes the labelling procedure and shows that a modified form of transitive closure terminates after an exponential number of steps. Section 6.2 discusses the implementation details of the decision procedure using formulas of $\mathbb{R}(+, \times)$. Finally, section 7 has more comments and some open questions.

## 2. Definitions

### 2.1. PATHS AND PLACEMENTS

This section contains many (though not all) of the definitions used in the paper. In appendix 2 we give a list of terms and symbols used in the paper, together with the section numbers where they are defined. When reading the paper it is probably useful to have this list easily accessible.

The *universe* $U$ is a closed, bounded subset of $\mathbb{R}^2$ with polygonal boundary. $U$ may have polygonal holes representing obstacles. The boundary of $U$ consists of vertices, called *corners*, and closed segments, called *walls*. There are two measures of the complexity of $U$: the number $n$ of corners and the size $m$ of the coordinates of corners (that is, the coordinates of any corner can be specified as a reduced rational $\pm p/q$ with $0 \leqslant p$, $q < 2^m$).

A *direction* $\theta$ is a point on $S^1$, the unit circle centered at the origin; we interpret $\theta$ as the direction of the ray through $\theta$ with endpoint at the origin. An *interval* $I$ of directions is a connected arc of $S^1$; its length is denoted $|I|$. If we write $I = [\theta_1, \theta_2]$ then $I$ is closed, $\theta_1$ is the clockwise endpoint of $I$ and $\theta_2$ is the counterclockwise endpoint of $I$; similar notation is used if $I$ is open or halfopen. We allow small reals to be added or subtracted to directions; the addition of a positive real corresponds to counterclockwise motion about $S^1$ and a negative real to clockwise motion. This extends to addition of intervals of reals: if $\theta$ is a direction and $[x, y]$ an interval of reals then $\theta + [x, y]$ is the interval $[\theta + x, \theta + y]$.

A function $p: [0, a] \to \mathbb{R}^2$ is a *smooth path* if $p(r) = (x_p(r), y_p(r))$ and $x_p, y_p: [0, a] \to \mathbb{R}$ are differentiable and the derivatives $x_p'$, $y_p'$ are continuous and not simultaneously zero. By elementary analysis ([6], p. 156), any smooth path has finite length; hence we assume $p$ is parameterized by arc length. Let $\phi_p(r)$ be the direction of the tangent to $p(r)$. Path $p$ is *curvature-constrained* if $|\phi_p(r_2) - \phi_p(r_1)| \leqslant |r_2 - r_1|$ for all $r_1, r_2 \in \text{dom}(p)$. A curvature-constrained path has curvature bounded above by 1 almost everywhere [4]; it may have finitely many points where, say, a straight segment meets a circular arc (the usual definition of curvature does not extend to such points). Path $p$ is *feasible* if it is curvature-constrained and $p(x) \in U$ for every $x \in \text{dom}(p)$.

A *placement* is a pair $(u, \theta)$ with $u \in \mathbb{R}^2$ and $\theta$ a direction; it is *feasible* if $u \in U$. If path $p$ is curvature-constrained and $r \in \text{dom}(p)$ we can define in the obvious way a placement at $p(r)$. We let $\Omega(p)$ be the initial placement of $p$ and $\Theta(p)$ be the last placement of $p$; we say $p$ is a path from $\Omega(p)$ to $\Theta(p)$. If there is a feasible path from placement $\omega$ to placement $\theta$, then $\theta$ is *reachable* from $\omega$.

A *corner contact* is a pair $(c, s)$ where $c$ is a corner and $s$ is "left" or "right". The *full contact interval* at $(c, s)$ is the set of pairs $(c, \theta)$ where $\theta$ is the direction of a line through $c$ so that near $c$ both walls incident to $c$ are to side $s$ of the line. Thus a placement in a contact interval at corner $c$ neither enters nor leaves the boundary of $U$. A full corner contact interval spans an interval of directions of size strictly less than $\pi$. A *wall contact* is a pair $(w, \theta)$, where $w$ is a wall and $\theta$ one of the two directions along the wall. The *full contact interval* at $(w, \theta)$ is the set of pairs $(u, \theta)$ with $u \in w$. An *interval at a corner contact* is a subinterval of the full contact interval at the corner contact (endpoints may be open or closed). An *interval at a wall contact* $(w, \omega)$ is $\{(u, \theta): u \in w'\}$, where $w'$ is a subsegment of $w$ containing the endpoint of $w$ in direction $\theta$ (the other endpoint of $w'$ may be open or closed). (If placement $(u, \theta)$ at a wall contact interval is reachable, then so is every placement along $w$ towards the endpoint in direction $\theta$.)

The *reachability question* is "Given $U$ of complexity $n$ and $m$, a *source* placement and a *target* placement, is there a feasible path from the source placement to the target placement?"

### 2.2. JUMPS, LEAPS, AND HOPS

An *oriented placement* $\sigma$ is a triple $(u, \theta, d)$ where $u \in U$, $\theta \in S^1$, and $d \in \{L, R\}$. Oriented placement $\sigma$ determines an oriented circle $C_\sigma$ of radius one; $C_\sigma$ is the circle that has tangent at $u$ in direction $\theta$ and orientation given by $d$ ($L$ is counterclockwise, $R$ is clockwise). The orientation is necessary to select one of the two possible circles with tangent at $u$ in direction $\theta$. The *position* of $\sigma$, written $u_\sigma$, is its location in $U$; $z_\sigma$ is the center of $C_\sigma$. If $d$ is $L$ or $R$ then $\bar{d}$ is $R$ or $L$, respectively; if $\sigma$ is $(u, \theta, d)$ then $\bar{\sigma}$ is $(u, \theta, \bar{d})$.

A *jump* is a pair of oriented placements. A jump $j = (\sigma, \sigma')$, $\sigma = (u, \theta, d)$, $\sigma' = (u', \theta', d')$ determines a path from $(u, \theta)$ to $(u', \theta')$ obtained by following an arc along $C_\sigma$, the segment tangent to $C_\sigma$ and $C_{\sigma'}$, and an arc along $C_{\sigma'}$. Notice that the tangent is unique because we require that it have direction consistent with the orientations of $C_\sigma$ and $C_{\sigma'}$. We use "jump" to refer both to a pair of oriented placements and to the path it specifies. (The path does not exist if $C_\sigma$ overlaps $C_{\sigma'}$ and they have opposite orientations; in general we will assume without explicit mention that $j$ is chosen so that the path exists.) We use $\Omega(j)$ and $\Theta(j)$ to denote the initial and final placements of $j$, as before. In addition $\Phi(j)$ is the direction of the tangent segment of $j$.

A *unit arc* is an arc of a circle of radius 1. A unit arc is *short* if it has length between 0 and $\pi$ and *long* if it has length 0 or length between $\pi$ and $2\pi$. The *type* of an oriented arc is $Tt$, where $T$ is $L$ or $R$ (the orientation of the arc) and $t$ is $s$ or $l$ (for short or long). The *type* of a jump is a concatenation of the types of the two arcs traversed in the jump. We define length 0 arcs as long so that, informally, jumps of a particular type are closed (otherwise the limit of a sequence of arcs with length approaching $2\pi$ would not be long). We use "$X$" or "$x$" if we do not care what the particular component of the type is; thus a jump of type $XsXs$ has short arcs of arbitrary orientation.

Jumps of a particular type define a relation on initial and final placements: in general, for a fixed initial placement many final placements may be reachable by some jump of that type, and conversely. We now impose restrictions on jumps to make this relation functional. A *leap* is a jump with one of the following restrictions: the initial arc has length 0, the final arc has length 0, the initial arc has length $\pi$, the final arc has length $\pi$, or the tangent segment has length 0 (i.e. the initial arc touches the final arc). The last kind of leap is a *hop*; necessarily the two arcs have opposite orientation.

A *leap type* is a similar to a jump type, with 0 or $\pi$ used in addition to $s$ and $l$. A leap type is *derived* from a jump type if it is obtained by possibly setting one $s$ or $l$ to 0 or $\pi$. Leap types are syntactically distinct from jump types with the exception of hop types: $L\pi Rs$ must be a leap type, but $LsRs$ could be a jump type or a hop type. Any leap type (except the type of a hop) can be derived from two different jump types; from a jump type there are four derived leap types if the arc orientations are the same (hop types are not possible) and five derived leap types otherwise.

For $a$ and $b$ contacts and $t$ a leap type, the *leap function* $g_{ab}^t$ is the function sending a placement at $a$ to the placement at $b$ that is reached from $a$ by a leap of type $t$. It can be verified that $g_{ab}^t$ is a function, indeed is one–one ($g_{ab}^t$ may not be defined for all placements at $a$). In the case that $a$ and $b$ are both corners, $g_{ab}^t$ is covariant if exactly one arc length in $t$ is $l$ (the others can be $s$, 0 or $\pi$), otherwise $g_{ab}^t$ is contravariant. "Covariant" (or "monotonically increasing") means that if $\omega$ is slightly counterclockwise of $\omega'$, then $g_{ab}^t(\omega)$ is slightly counterclockwise of $g_{ab}^t(\omega')$.

If we impose two restrictions on a jump, then there is at most one possible jump between two contacts with both restrictions. The (type 0) *roots* include the initial and final placements of every jump with two restrictions between two contacts. Thus placements at a corner with the following directions are roots: the straightline direction towards another corner, the direction along a unit circular arc to another corner or wall, the direction along a path consisting of a unit arc of length $\pi$ followed by a straight segment, and so forth. In fig. 4 (right-hand side), the source placement and final placement of every depicted jump is a root. We also include in type 0 roots the source placement and the endpoints of every corner contact interval. This implies that every interval at a wall contains a root. Since there are only a constant number of jump types with two restrictions, there are only a constant number of type 0 roots defined between a pair of contacts, so only $O(n)$ type 0 roots at a contact, and $O(n^2)$ altogether. In section 3.2 we define a larger class of (type 1) roots.

### 2.3. ALGEBRA

THEOREM 2.3.1. [5]

A formula of $\mathbb{R}(+, \times)$ (the first-order theory of the reals) of length $l$, $v$ variables, and maximum degree $d$ is decidable in time $(dl)^{2^{O(v)}}$.

In sections 5 and 6 we use the decision procedure for $\mathbb{R}(+, \times)$ in the decision procedure for the reachability question. We sketch here how to represent geometric primitives in $\mathbb{R}(+, \times)$. A placement $(u, \theta)$ can be represented using four real variables, $x$, $y$ for position and $v$, $w$ for direction, with the side constraint that $v^2 + w^2 = 1$. If $\sigma$ is the oriented placement $(u, \theta, d)$, then the center $z_\sigma$ of $C_\sigma$ can be obtained easily from $x$, $y$, $v$, $w$. The assertion $j = (\sigma, \tau)$ is a jump of type $t$ is "there is a directed line $l$ at distance $\pm 1$ from $z_\sigma$ and $z_\tau$ (sign chosen so that the direction of $l$ is consistent with the orientation of $\sigma$ and $\tau$) so that arc lengths along $C_\sigma$ and $C_\tau$ are consistent with $t$ (determined by the order of projections of $z_\sigma$, $u_\sigma$, $z_\tau$, $u_\tau$ on $l$)." Membership in $U$ can be expressed by a formula of length

$O(nm)$: for example, triangulate $U$ and express membership in some triangle. Hence feasibility of jumps is expressible with a formula of the same length. The construction of other primitives is similar and straightforward.

For the improved decision procedure in section 6, we will need a bound on the total number of possible fixed points of a cycle from a contact back to itself. This bound will follow from a bound on the number of simultaneous zeroes of a system of multivariate polynomial equations. The bound on the number of simultaneous zeroes is given by the degree of the univariate polynomial obtained by using resultants to eliminate all but one of the variables. Note that the degree counts both real and complex zeroes; we are really interested in real zeroes, but the bound on both real and complex zeroes is good enough. The following is a standard result, see for example [16].

LEMMA 2.3.2

Given two real polynomials $p(x, y)$ and $q(x, y)$ of degrees $d$ and $e$, there is a real polynomial $r(x)$, the *resultant of p and q* so that $r(x) = 0$ iff there is $y$ so that $p(x, y) = q(x, y) = 0$. (Note $x$ and $y$ may be real or complex, in fact $y$ may be complex even if $x$ is real.) The degree of $r$ is at most $de$ and the coefficients of $r$ have size at most $(d+e)^2 \log(d+e)s$, if $p$, $q$ have coefficients of size $s$.

For the material in section 6 it is convenient to use only a single real variable to represent a placement at a fixed contact. This is possible since we always know the contact interval containing the placement. If the contact is a wall, then clearly the direction of the placement is a constant and the position requires only a single variable. If the contact is a corner, then the position is a constant. Since the contact interval at the corner has size less than $\pi$, we can rotate the interval so that its directions are, say, the upper semicircle of $S'$ and represent the direction by a single variable constrained between $-1$ and 1.

LEMMA 2.3.3.

For each leap function $g^t_{ab}$ there is a real polynomial $p^t_{ab}(x, y)$ not identically zero of constant degree with coefficients of size $O(m)$ so that $g^t_{ab}(x) = y$ implies $p^t_{ab}(x, y) = 0$.

*Proof*

For each $g^t_{ab}$ there is a formula $F(x, y)$ of $\mathbb{R}(+, \times)$ so that $F(x, y)$ is satisfied exactly when $g^t_{ab}(x) = y$. It is possible to choose $F(x, y)$ in prenex normal form so that its prefix contains only existential quantifiers over a vector of variables $r$ and so that its matrix is the conjunction of polynomials $P_i(x, y, r) = 0$ and $Q_i(x, y, r) \geqslant 0$. Furthermore, each polynomial $P_i$ involves only two variables and there are at least as many polynomials $P_i$ as variables. We can use resultants to eliminate the vector of variables $r$, obtaining $p^t_{ab}$. □

Note that the lemma does not claim the converse statement, in particular $p_{ab}^t(x, y) = 0$ does not necessarily imply $g_{ab}^t(x) = y$. For the use that we will make of $p_{ab}^t$, that is, to bound the number of simultaneous fixed points of cycles, it does not matter that $p_{ab}^t$ may have spurious zeroes.

A final definition will be used in section 4. A real $r$ is of *constant algebraic complexity* if it is the zero of a real polynomial of constant degree with coefficients of size $cm$, $c$ a constant. Easily the direction of any root is of constant algebraic complexity. The following is a standard result; an extensive discussion appears in the paper by Schwartz and Sharir [14].

LEMMA 2.3.4

If $r$, $r'$ are distinct reals of constant algebraic complexity, then $|r - r'| \geqslant 2^{\Omega(-m)}$.

## 3. Reduction to jumps

### 3.1. NORMALIZATION

THEOREM 3.1.1

If $p$ is a feasible path, then there is a feasible path $p'$ consisting of a jump from $\Omega(p)$ to a contact, a finite sequence of jumps between contacts, and a final jump to $\Theta(p)$.

In particular, the path $p'$ consists only of straight segments and unit arcs, and every arc except possibly the first and last touches some point of the boundary of $U$. An immediate consequence is that the target placement is reachable exactly if it is reachable by a jump from the source placement or from some reachable placement at a wall or corner.

LEMMA 3.1.2 [7]

If $p$ is a curvature-constrained path of length at most $\pi/2$, then $p$ does not enter the interior of $C_\sigma$ or $C_{\bar{\sigma}}$, where $\sigma$ and $\bar{\sigma}$ are the oriented placements through $\Omega(p)$.

LEMMA 3.1.3

If $p$ is a feasible path, then there is a feasible path $p'$ with the same initial and final placements solely consisting of unit arcs and straight segments.

*Proof*

Recall that $p$ must have finite length (see section 2.1). Path $p$ can intersect each corner at most a finite number of times, since by lemma 3.1.2, the path from a corner back to itself must have length at least $\pi/2$. If $p$ intersects a wall $w$ for the first time at a placement $(u, \theta)$ then there must be a last time $p$ intersects $w$ at a placement $(u', \theta)$ with $u'$ in direction $\theta$ along $w$ from $u$; we can replace the section of $p$ from $(u, \theta)$ to $(u', \theta)$ with a segment along wall $w$. Hence we can assume that $p$ intersects each wall only a finite number of times (each time a point or segment), since again by lemma 3.1.2, the path from a placement $(u, \theta)$ on wall $w$ to a placement $(u', \theta)$ on $w$ in direction opposite $\theta$ from $u$ must have length at least $\pi/2$.

Say a *touch point* is a corner touched by $p$ or a point where $p$ just begins to touch a wall or a point where $p$ just finishes touching a wall. We claim that there is $\delta$ with $\pi/2 > \delta > 0$, so that if $p(x)$ ever gets within distance $\delta$ of the boundary of $U$, then for some $x'$ near $x$, $p(x')$ passes through a touch point. Let $\delta'$ be the minimum of $\pi/2$, the distance between any pair of corners, and the distance between a touch point on the interior of a wall and any other obstacle (including the corners at the end of the wall). Let the set of placements $p'$ be the placements in $p$ not including any placement whose distance along $p$ to a boundary point of $U$ is less than $\delta'$. Then the set of positions of placements in $p'$ is closed and does not intersect the boundary of $U$. Set $\delta > 0$ to be less than the minimum of $\delta'$ and the distance from the position of any placement in $p'$ to the boundary of $U$. Then the position of any placement in $p'$ is certainly at least $\delta$ from the boundary of $U$, and a placement on $p$ not in $p'$ lies on a subpath passing through a touch point.

We split $p$ into a finite number of subpaths, each of length at most $\delta/2$, so that each subpath $q$ satisfies one of the following: $q$ lies entirely along a wall, $q$ entirely avoids obstacles, $q$ avoids obstacles except that one endpoint touches a corner, or $q$ avoids obstacles except that one endpoint touches an interior point of a wall. In the last case we assume also that the projection of $q$ along the wall lies entirely within the wall. In each case we can assume all of $q$ is at least $\delta$ from any obstacle distinct from the obstacle that $q$ touches.

We claim that each path $q$ can be replaced by a feasible path $q'$ consisting of a unit arc, a straight segment, and another unit arc (one or two pieces may be empty). Let $r$ be the length of $q$. Say a *c-cone through placement* $\alpha = (u, \phi)$, $N_\alpha$, is the set of points of $\mathbb{R}^2$ within distance $r$ of $u$, lying on or outside $C_{(u,\phi,L)}$ and $C_{(u,\phi,R)}$, and inside the halfplane with bounding line through $u$ and interior normal in direction $\phi$. Then $N_\alpha$ is a "triangular" region with boundary a section of arc of $C_{(u,\phi,L)}$, a section of arc of $C_{(u,\phi,R)}$, and a circular arc of radius $r$ and center $u$ (see fig. 2). Let $\omega = \Omega(q)$, $\theta = \Theta(q)$, and $\theta'$ the placement with direction opposite $\theta$. By lemma 3.1.2, all of $q$ lies inside both $N_\omega$ and $N_{\theta'}$. In particular $N_\omega \cap N_{\theta'}$ is connected and contains both $u_\omega$ and $u_\theta$. Let $B_\omega$ be the portion of the boundary of $N_\omega$ that also bounds $N_\omega \cap N_{\theta'}$, and similarly for $B_{\theta'}$. Notice that no part of the arc of radius $r$ bounding $N_\omega$ can appear in $B_\omega$. $B_{\theta'}$ and $B_\omega$ intersect at two points $b$ and $t$. For $x \in B_\omega$ let $r(x)$ be the ray with endpoint $x$ tangent to $B_\omega$. We claim there is $x_0 \in B_\omega$ so that $r(x_0)$ is also tangent to $B_{\theta'}$; this demonstrates the required path $q'$. To see this claim let $\alpha(x)$ be the angle

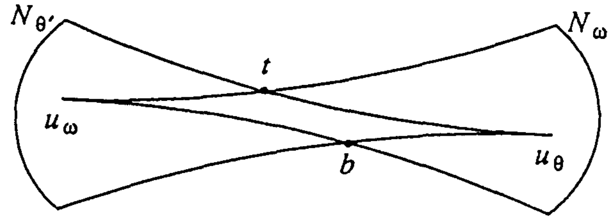


Fig. 2. Path $q$ is not shown but lies in $N_\omega \cap N_{\theta'}$.

between $r(x)$ and the tangent to $B_{\theta'}$ at the first point of intersection between $r(x)$ and $B_{\theta'}$. Then $\alpha(b)$ and $\alpha(t)$ must have opposite sign; since $\alpha(x)$ varies continuously as long as $r(x)$ is not tangent to $B_{\theta'}$, there must be some point at which $r(x)$ becomes tangent to $B_{\theta'}$. The constructed path $q'$ surely has length at most $2r$.

We must show $q'$ avoids obstacles (except possibly at $\omega$ and $\theta$); let $\omega$, $\theta$, $\theta'$ be as before. If $q$ avoids obstacles, then $q'$ must as well, since $q'$ has length at most $\delta$ and both $\omega$ and $\theta$ are distance more than $\delta$ from any obstacle. Suppose $\omega$ touches a wall (the case that $\theta$ touches a wall is similar). Then since the projection of $q$ stays within the endpoints of the wall, $q$ lies in halfplane $H$, where $H$ has bounding line through the wall and normal locally into $U$. Hence $q \subseteq N_\omega \cap N_{\theta'} \cap H$. It is easy to check that in this case $q' \subseteq H$ as well. Path $q'$ cannot intersect any other obstacle, since it is not long enough. The case that $q$ touches a corner is similar. □

*Proof of theorem 3.1.1*

By the previous lemma, we can assume $p$ consists of a finite sequence of unit arcs and straight segments. Say a *free arc* is an arc on $p$ that does not touch an obstacle and is not the initial or final arc. We give two transformations on paths; each reduces the number of free arcs by at least one. The theorem results by applying the transformations repeatedly until no free arcs remain.

Suppose $A_1$ is a free arc whose length is at least $\pi$ (but less than $2\pi$). Let $A_0$ and $A_2$ be the arcs preceding and following $A_1$ along $p$, and $C_0$, $C_1$, $C_2$ the corresponding circles (see fig. 3). The first transformation is to push $C_1$ in direction parallel to the tangent from $C_0$ to $C_1$. The transformed path then goes from $A_0$ to the new position of $C_1$ and then to $C_2$. Notice that as $C_1$ moves, the tangent from $A_0$ to $C_1$ simply gets longer, of course the tangent from $C_1$ to $C_2$ moves. The arc around $C_1$ gets shorter; however it is easy to check that it remains longer than $\pi$ (or stays at $\pi$ if it was originally). Hence the direction of the tangent from $C_1$ to $C_2$ changes by less than $\pi$. The arc around $C_2$ increases in length if $C_1$ and $C_2$ are on the same side of the tangent from $C_1$ to $C_2$ (they have the same orientation) and decreases if $C_1$ and $C_2$ are on opposite sides (they have opposite orientation). The transformation terminates when $C_1$ hits an obstacle (which must happen since $U$ is bounded). Possibly before then the following

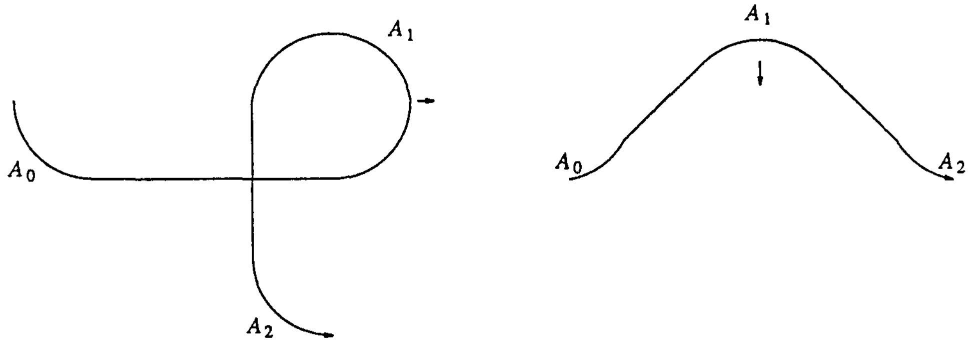


Fig. 3. Path deformations.

events could happen. First, the arc length around $C_2$ could go to zero. We continue with the arc $A_2'$ of $C_2'$ following $A_2$ on $P$ if it exists; if not we must be at $\Theta(p)$ and we create a unit circle of orientation opposite $C_2$ through $\Theta(p)$ and continue. Second, the arc length around $C_2$ could increase to $2\pi$; we continue the transformation with $C_2$ omitting the extra cycle around $C_2$. Finally, the tangent could touch a corner; we create a circle with the same orientation as $C_1$ through the placement at the corner, and continue with it instead of $C_2$. This introduces a new arc, but it is not free. The first two possibilities can happen at most once and the third at most once per corner since the direction of the tangent from $C_1$ to $C_2$ changes by at most $\pi$.

Now suppose $A_1$ is a free arc whose length is less than $\pi$. Let $C_0$, $C_1$ and $C_2$ be as before. Extend the tangents between $C_0$ and $C_1$ and between $C_1$ and $C_2$ until they meet; the transformation in this case is to push the center of $C_1$ along the ray bisecting the angle formed by the extended tangents. Notice that arc length along $C_1$ decreases as $C_1$ is pushed. The transformation terminates if arc length along $C_1$ becomes zero or the arc along $C_1$ touches an obstacle. The remaining details are similar to the previous case. □

### 3.2. CONFIGURATION SPACE FOR JUMPS

For this section, $a$ and $b$ are fixed contacts and $t_0$ is a fixed jump type. We define a "configuration space" $C$ for jumps of type $t_0$ from $a$ to $b$. Note that since the contacts are fixed, the initial and final placements of a jump can each be specified by a single parameter (direction in the case of corner contacts and position in the case of wall contacts). Hence $C$ is two-dimensional. $C$ is naturally partitioned by curves defined using obstacles; each region of the partition corresponds to feasible homotopic jumps. The purpose of the configuration space construction is to define $\Xi_{ab}^{t_0}(I)$ and $X(j, I)$, where $I$ is an interval at contact $a$. $\Xi_{ab}^{t_0}(I)$ is a collection of representative jumps, one chosen from each region that has a jump with source placement in $I$. For $j \in \Xi_{ab}^{t_0}(I)$, $X(j, I)$ is all feasible

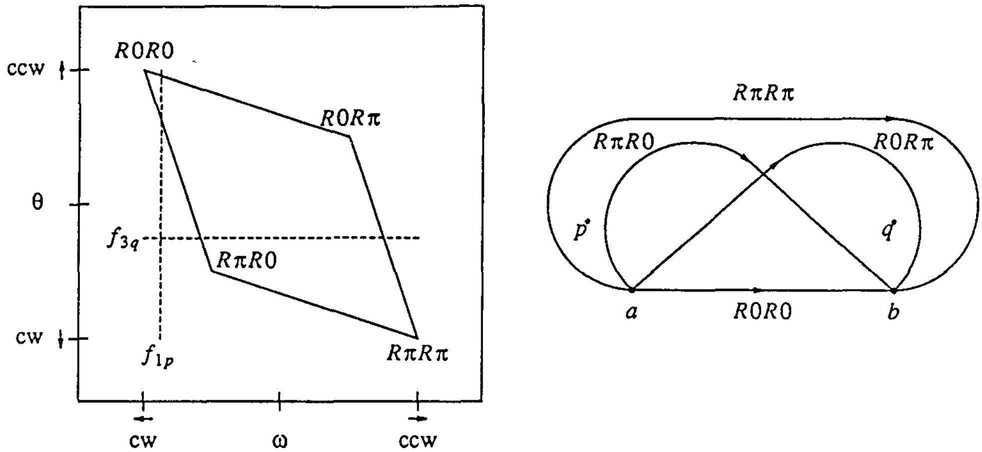


Fig. 4. Jumps of type *RsRs* from $a$ to $b$. Direction at $a$ is $\omega$; at $b$ it is $\theta$.

jumps homotopic to $j$ with source placement in $I$. The principal result is that $\Xi_{ab}^{t_0}(I)$ has size $O(n^2)$.

Define $c(j) = (\Omega(j), \Theta(j))$ and $C = \{c(j)$: $j$ is a jump from $a$ to $b$ of type $t_0\}$. (In this section only we use "$C$" for configuration space; in all other sections "$C$" is used in "$C_\sigma$" meaning the circle through oriented placement $\sigma$.) Clearly $C$ is contained in the Cartesian product of the contact interval at $a$ with the contact interval in $b$. We say a point in $C$ is *feasible* if it is the image under $c$ of a feasible jump; it can be checked that the feasible points in $C$ form a closed set.

Figure 4 depicts configuration space (on the left) and some corresponding jumps (on the right). The jumps are of type *RsRs* from a corner contact at $a$ to a corner contact at $b$ (the actual obstacles are not shown). The horizontal axis is the source direction $\omega = \Omega(j)$; the vertical axis is the target direction $\theta = \Theta(j)$. The zero direction, i.e. the middle tickmark, for $\omega$ is vertically upwards from $a$; the zero direction for $\theta$ is vertically downwards from $b$.

Let a *vertical segment through* $\omega$ be $\{(\omega, \theta)$: $\theta \in I\}$, where $I$ is a subinterval of the contact interval at $b$, and similarly for "horizontal segment". It is easy to check that the boundary of $C$ is contained in the union of curves $g_{ab}^t$ for $t$ a leap type derived from $t_0$ together with vertical segments through the endpoints of the contact interval at $a$ and horizontal segments through the endpoints of the contact interval at $b$. We split the boundary of $C$ into *boundary curves*; each is a maximal piece of a curve $g_{ab}^t$ or a horizontal or vertical segment. The boundary of $C$ actually forms a cycle, so $C$ is path-connected.

As depicted in fig. 4, the boundary of $C$ consists of the leap functions of type *R*0*Rs* (from the jump *R*0*R*0 to jump *R*0*R*$\pi$), type *RsR*$\pi$, *R*$\pi$*Rs*, and *RsR*0. Figure 4 is schematic in two ways. First, the boundary curves are not actually linear functions, so the boundary curves are not straight. Second, notice that the *R*0*R*0 and *R*$\pi$*R*$\pi$ jumps have initial directions that differ by $\pi$, as do the final directions. However, the size of the contact intervals at $a$ and $b$ must be strictly less than $\pi$. Hence the boundary of $C$ must also include vertical and horizontal segments defined using the endpoints of the contact intervals at $a$ and at $b$. These are not depicted in the figure.

We now define a collection of curves partitioning $C$. Informally, $f_{id}$ is the curve in $C$ defined by jumps that touch obstacle $d$ on the source arc ($i = 1$), tangent segment ($i = 2$), or final arc ($i = 3$). For $d$ a corner contact, let $f_{1d} = \{c(j)$: $j$ is a jump of type $t$ from $a$ to $b$ with source arc touching $d\}$; then $f_{1d}$ is a vertical segment through a root. One endpoint of $f_{1d}$ (corresponding to a jump with source arc as long as possible) must lie on a boundary curve; the other endpoint of $f_{1d}$ either lies on a boundary curve as well, or corresponds to a jump consisting of an arc to $d$ followed by a leap to $b$. In the second case the other endpoint lies on $f_{2d}$. The definition of $f_{1d}$ with $d$ a wall contact is similar, with source arc of $j$ tangent to $d$. The definition of $f_{3d}$ for $d$ a wall or corner contact is similar to that of $f_{1d}$, using the final arc of $j$; then $f_{3d}$ is a horizontal segment through a root. For $d$ a corner contact define $f_{2d}$ with tangent segment touching $d$; $f_{2d}$ is a monotonic curve with endpoints on the boundary or $f_{1d}$ or $f_{3d}$. If $d$ is a wall then no definition of $f_{2d}$ is needed since necessarily either an endpoint of $d$ touches the tangent segment of a jump or an arc of a jump touches $d$. In fig. 4, $f_{1p}$ corresponds to all jumps with source arc passing through point $p$ and $f_{3q}$ corresponds to all jumps with final arc passing through $q$. A curve $f_{2p}$ in general is a "diagonal" curve through configuration space.

Let $\mathscr{C}$ be all boundary curves and all curves $f_{id}$.

LEMMA 3.2.1

The intersection of two distinct curves in $\mathscr{C}$ is a point or a horizontal or vertical segment.

*Proof*

Every curve in $\mathscr{C}$ is horizontal, vertical, or monotonic. The lemma is immediate except in the case of two monotonic curves. There can be at most one leap with two types, hence $g^t_{ab} \cap g^{t'}_{ab}$ is at most a point. There can be at most one leap of a particular type with tangent constrained to go through a particular point, so $g^t_{ab} \cap f_{2d}$ is at most a point. Finally, there is at most one jump of a particular type with tangent constrained to go through two different points, so $f_{2d} \cap f_{2d'}$ is at most a point. □

A *vertex* is the point of intersection of two distinct curves in $\mathscr{C}$, if it is unique. Notice that the endpoint of every vertical or horizontal segment also intersects some other curve; hence the endpoint of a segment that is the intersection of two vertical or horizontal segments is a vertex. Since there are only $O(n)$ curves in $\mathscr{C}$, there are at most $O(n^2)$ vertices.

LEMMA 3.2.2

Let $(\omega, \theta)$ be a vertex. Then $\theta$ is a root or the composition of one or two leap functions applied to a root; $\omega$ is a root or the composition of one or two inverse leap functions applied to a root. In all cases roots are type 0 roots.

*Proof*

First suppose that vertex $(\omega, \theta) = c(j)$ is the unique point of intersection of two curves in $\mathscr{C}$. If one curve is a vertical segment $v$, then it must be through a root, so $\omega$ is a root. If $c(j) \in v \cap g_{ab}^t$, then $\theta = g_{ab}^t(\omega)$. If $c(j) \in v \cap f_{2d}$, then $j$ consists of a leap to $d$ followed by a leap to $b$. (Note that a special case is if $v = f_{1d}$, then $d$ is a corner at the intersection point of the source arc and tangent segment of $j$.) If $c(j) \in v \cap h$ for a horizontal segment $h$, then $h$ must be through a root, so $\theta$ is a root. The analysis of intersections involving a horizontal segment is similar, it requires the use of inverse leap functions. If $c(j) \in g_{ab}^t \cap f_{2d}$, then $j$ is a leap with tangent segment touching $d$; the placement along $j$ at $d$ must be a root as is one of $\omega$ and $\theta$; the other is obtained by a leap function or inverse leap function. If $c(j) \in f_{2d} \cap f_{2d'}$, then $j$ is a jump with tangent touching $d$ and $d'$; both placements along $j$ at $d$ and $d'$ must be roots, $\omega$ is obtained by an inverse leap function and $\theta$ by a leap function. Both coordinates of the unique point of intersection of two curves $g_{ab}^t$ and $g_{ab}^{t'}$ are roots. □

We now decree that the coordinates of all vertices are (type 1) roots. This gives $O(n^2)$ roots at contact $a$ for each contact $b$, hence a total of $O(n^3)$ roots at $a$ and $O(n^4)$ over all contacts.

The removal of the curves $\mathscr{C}$ splits $C$ into open path-connected regions; the removal of vertices from the curves $\mathscr{C}$ partitions them into open one-dimensional curves. We let $\mathscr{R}$ be the set whose elements are feasible vertices and the closure of curves or regions containing a feasible point. By planarity, $\mathscr{R}$ has at most $O(n^2)$ elements. The bound $O(n^2)$ is tight to within a constant factor (see fig. 5).

LEMMA 3.2.3

If $R \in \mathscr{R}$, then every point in $R$ is feasible.

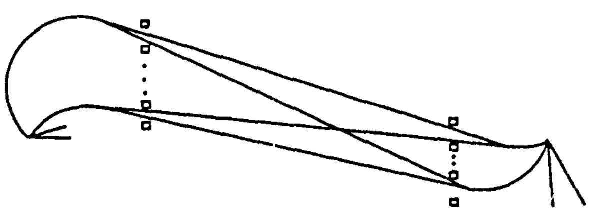

Fig. 5. Example showing $\mathscr{R}$ has $O(n^2)$ elements.

*Proof*

We claim that if $p$: $[0, 1] \to C$ is a path from $c(j)$ to $c(j')$, $j$ is feasible, and $p$ never leaves a curve $f_{ld}$ (that is, if $p$ intersects such a curve it must stay on it), then $j'$ is feasible. Clearly $h = c^{-1} \circ p$ is a homotopy from $j$ to $j'$ so that every intermediate path is a jump with source placement in $I$. Suppose $j$ is feasible and $j'$ is not. Then for some point $x$ jump $h(x)$ is feasible and $h(x')$ for all $x' > x$ near $x$ is not. Then for $x' > x$ near $x$, $h(x')$ must intersect the interior of an obstacle near a wall or corner $d$ and $h(x)$ must touch $d$. Hence $p(x)$ lies on $f_{ld}$ and $p(x')$ for $x' > x$ does not.

For the lemma, $R$ must be the closure of $R'$, where $R'$ is some open region or curve containing a feasible point. By the definition of $\mathscr{R}$, no path connecting two points in $R'$ can leave a curve $f_{ld}$, hence every point in $R'$ is feasible. Every point in $R$ is feasible since the set of feasible points is closed in $C$. □

Let $I$ be an interval contained in the contact interval at $a$ ($I$ can have open or closed endpoints) and set $C(I) = \{(\omega, \theta) \in C\colon \omega \in I\}$. Define $\mathscr{R}(I)$ restricted to $C(I)$ in the same fashion as $\mathscr{R}$ (the definition must use relative closure in $C(I)$). $\mathscr{R}(I)$ still has $O(n^2)$ elements, since the vertical segments through the endpoints of $I$ can intersect each curve in $\mathscr{C}$ at most once.

Construct a representative set of jumps $\Xi^{\prime t_0}_{ab}(I)$ by choosing a jump from the interior of every two-dimensional region in $\mathscr{R}(I)$, then a jump from the interior of any one-dimensional region not on the boundary of a two-dimensional region, and then any vertices not on the boundary of a one- or two-dimensional region. Then every jump $j \in \Xi^{\prime t_0}_{ab}(I)$ has $c(j)$ in a unique region of $\mathscr{R}(I)$, and every $R \in \mathscr{R}(I)$ has a representative jump (possibly lying in a region of which $R$ forms part of the boundary). For $j \in \Xi^{\prime t_0}_{ab}(I)$ we define $X(j, I) = \{j'\colon c(j') \in R$ where $c(j) \in R \in \mathscr{R}(I)\}$.

LEMMA 3.2.4

An endpoint of $\Theta(X(j, I))$ is a root or the composition of one or two leap functions applied to an endpoint of $I$. An endpoint of $\Omega(X(j, I))$ is an endpoint of $I$ or a root. If $\Omega(X(j, I)) \neq I$, then $\Theta(X(j, I))$ contains a root. If $a$ is a wall, then $\Theta(X(j, I))$ contains a root.

*Proof*

The proof of the first two statements is similar to lemma 3.2.2. For the third statement, if $j$ is the jump with $\Omega(j)$ attaining the endpoint of $\Omega(X(j, I))$ that is not an endpoint of $I$, then a similar case analysis shows that $\Theta(j)$ is a root. Finally, if $a$ is a wall, since motion along the wall is always possible, we can find a jump $j'$ so that $c(j')$ is a vertex of the same region $R$ of $\mathscr{R}(I)$ containing $j$. □

### 3.3. PROPAGATION USING JUMPS

For $I$ an interval at $a$, $b$ a corner contact, and $j \in \Xi^t_{ab}(I)$ we define $J(j, I) = \Theta(X(j, I))$. Thus $J(j, I)$ is the interval of placements at $b$ reachable from some placement in $I$ by jumps homotopic to $j$ and of the same type as $j$. If $b$ is a wall contact $(w, \theta)$, we define $J(j, I)$ to be the set of placements at $w$ reachable by some jump in $X(j, I)$ possibly followed by a straight segment along wall $w$. Then $J(j, I)$ is an interval at $b$. We now extend mapping $J$ to a mapping $\mathscr{J}$ propagating sets of intervals into sets of intervals. To do this we must discuss sets of intervals.

A set $\mathscr{I}$ of intervals is *well-formed* if each $I \in \mathscr{I}$ is a subset of some contact interval and if $I, I' \in \mathscr{I}$ are intervals at the same contact then $I \cup I'$ is not an interval. We write $\mathscr{I} \sqsubset \mathscr{I}'$, if for each $I \in \mathscr{I}$ there is $I' \in \mathscr{I}'$ with $I \subseteq I'$. The *merge* of $\mathscr{I}$ and $\mathscr{I}'$, written $\mathscr{I} \sqcup \mathscr{I}'$, is the smallest (under $\sqsubset$) well-formed set $\mathscr{I}''$ so that $\mathscr{I} \sqsubset \mathscr{I}''$ and $\mathscr{I}' \sqsubset \mathscr{I}''$. More explicitly $\mathscr{I}''$ is obtained from $\mathscr{I}$ and $\mathscr{I}'$ by merging together adjacent or overlapping intervals. Clearly $|\mathscr{I} \cup \mathscr{I}'| \leqslant |\mathscr{I}| + |\mathscr{I}'|$. We use script ($\mathscr{F}$, $\mathscr{I}$, $\mathscr{J}$, $\mathscr{R}$, $\mathscr{T}$) for sets of intervals (or functions with range a set of intervals); henceforth we assume without explicit mention that all sets of intervals are well-formed.

For $I$ an interval at contact $a$, define $\mathscr{J}^t_{ab}(I) = \sqcup_{j \in \Xi^t_{ab}} J(j, I)$. Thus $\mathscr{J}^t_{ab}(I)$ is all placements at $b$ reachable from a placement in $I$ by jumps of type $t$. Again for $I$ an interval at contact $a$, define $\mathscr{J}(I) = \{I\} \sqcup \bigsqcup_{b,t} \mathscr{J}^t_{ab}(I)$; thus $\mathscr{J}(I)$ is $I$ together with all placements at all contacts reachable by some jump from some placement in $I$. Finally extend $\mathscr{J}$ to sets of intervals: $\mathscr{J}(\mathscr{I}) = \sqcup_{I \in \mathscr{I}} \mathscr{J}(I)$.

Let $\mathscr{I}_0$ be the set containing the interval containing exactly the source placement. Let $\mathscr{R}$ be the set of all placements at contacts reachable by some feasible path from the source placement. By theorem 3.1.1, we have $\mathscr{R} = \sqcup_{k=0}^{\infty} \mathscr{J}^{(k)}(\mathscr{I}_0)$. Since $\mathscr{J}$ is a monotonic map, $\mathscr{R}$ is also the least fixed point of $\mathscr{J}$ containing $\mathscr{I}_0$.

For each $k$, $\mathscr{J}^{(k)}(\mathscr{I}_0)$ is a finite set of closed intervals, since $\mathscr{I}_0$ is closed and $\mathscr{J}$ maps a finite set of closed intervals into a finite set of closed intervals. Hence $\mathscr{R}$ is the countable union of closed intervals; this implies $\mathscr{R}$ is the countable union of disjoint intervals with open or closed endpoints. In section 5 we show that the number of intervals in $\mathscr{R}$ is in fact finite; however, intervals in $\mathscr{R}$ may well have open endpoints.

## 4. Procedure *stretch*

### 4.1. INTRODUCTION

We now consider a procedure *stretch* for generating reachable placements. The arguments to *stretch* are a source placement and a jump from the source placement to a corner. The result of *stretch* is an interval of placements at the corner reachable by feasible paths from the source placement.

The motivation for *stretch* is to propagate as large an interval of reachable placements as possible. The problem with simply propagating reachable placements using jumps is that there is no apparent way to bound the number and size of reachable intervals. With *stretch* and the tools developed in its analysis, good bounds on the number of disjoint reachable intervals can be obtained.

Suppose $j = (\sigma, \tau)$ is a jump to corner $c$. The goal of *stretch* is to deform jump $j$ into a family of paths with interval of final directions as large as possible. Each feasible path is homotopic to $j$. Procedure *stretch* can be described informally as follows. Jump $j$ traverses an arc of $C_\sigma$, a straight segment tangent to $C_\sigma$ and $C_\tau$, and an arc of $C_\tau$. Imagine rotating $C_\tau$ around corner $c$, say clockwise. Then the jump consisting of $C_\sigma$ and the rotating $C_\tau$ still traverses an arc of $C_\sigma$, a (moving) straight segment, and a (moving) arc of $C_\tau$. Several events can happen during this rotation. One event is that the final arc on $C_\tau$ hits an obstacle. Then the rotation stops, because it is not possible to deform the jump further. Alternatively, the

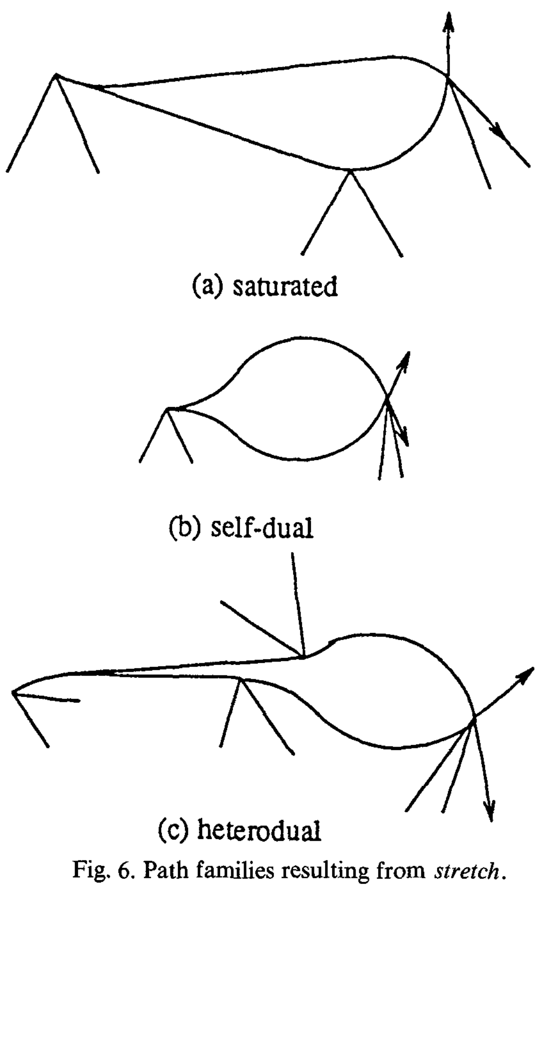


Fig. 6. Path families resulting from *stretch*.

straight segment of the jump could hit an obstacle. Continued rotation would result in an infeasible path. Instead, the path bends around the obstacle. Specifically, the final portion of the path becomes a jump with source placement at the obstructing obstacle; this jump is then deformed as before. (Notice bending is not possible if the final arc on $C_\tau$ hits an obstacle.) A third event is that the length of the straight segment between source arc and final arc goes to zero. Then rotation again stops, as further deformation is impossible. Other events are possible, for example, if the source or final arc length decreases to zero then the orientation of the corresponding circle is reversed. The whole procedure is discussed more fully in the next section. Once rotation stops in one direction, the procedure is repeated with rotation in the opposite direction. The output of *stretch* is the interval resulting from rotation in both directions.

The importance of *stretch* is that the resulting family of paths can be put into one of three categories. This categorization is used in section 5 to bound the number of disjoint reachable intervals. The first category is if one direction of deformation (or both) stops because of contact with an obstacle (see fig. 6a). Such path families are called "saturated"; clearly the resulting interval contains a root.

The second category is if deformation stops because each extreme path consists entirely of two oppositely oriented arcs (see fig. 6b). One extreme path must be a unit $L$-arc tangent to a unit $R$-arc; the other is a unit $R$-arc tangent to a unit $L$-arc. It is possible to prove that all arcs have length less than $\pi$. Such path families are called "self-dual" and the resulting interval a "self-dual" interval. Notice that a self-dual interval depends upon source direction; in particular there may be many disjoint self-dual intervals involving the same pair of corners.

The third category is if deformation stops because both extreme paths end with two oppositely oriented arcs, but at least one path also contains other arcs or segments. This happens if one of the paths had to bend around an obstacle (see fig. 6c). Such path families are called "heterodual".

The next section, section 4.2, describes procedure *stretch* in more detail and proves various properties of it, in particular termination. The details of the proofs are far from obvious. Section 4.3 discusses the categorization of path families. The actual theorem is that either the path family resulting from jump $j$ can be categorized as described, or there is a jump $j'$ with the same source placement whose path family can be categorized, and the interval of placements resulting from deforming jump $j$ overlaps the interval of placements resulting from deforming jump $j'$. Section 4.5 reduces the analysis of heterodual path families to self-dual path families: the main theorem is that any interval resulting from a heterodual path family contains a self-dual interval, defined from one of the corners intersected during the deformation. Self-dual intervals are discussed in section 4.4. If a path family from one corner to another is self-dual, then the source direction of the path family must lie in an interval of directions centered at the straight-line direction from the first corner to the second. The main theorem of the section is that the size of the self-dual interval grows quickly (in particular linearly) as the source direction moves away from an endpoint of the interval towards the center point of the interval. Together with a lower bound on the size of self-dual intervals, this theorem will be used to bound the number of possible disjoint self-dual intervals.

One final remark is that procedure *stretch* is somewhat *ad hoc*: the interval from *stretch* need not be the largest interval of placements obtained by feasible paths homotopic to the original jump. This happens because the extreme paths can contain straight segments. Such segments exist because *stretch* terminates as soon as the final arc becomes tangent to the preceding arc. A possible modification of *stretch* is to continue to deform paths in such cases. However, such a modification of *stretch* is not necessary for later sections and would further complicate its definition and analysis, which are already quite complicated enough. In retrospect, it seems it might be possible to analyze the class of all (curvature constrained) paths that are homotopic to a given jump and that have the same source direction as the jump. A more abstract proof that, say, extreme paths are either saturated or contain only unit circular arcs might be easier than the analysis of *stretch*. We do not know, however, how to carry out this approach.

### 4.2. PROCEDURE *STRETCH*

We now define procedure *stretch*$(j, d)$. The first parameter $j$ is a feasible jump $(\sigma_0, \tau_0)$ and the second parameter $d$ is a direction of rotation $L$ or $R$. Both $\sigma_0$ and $\tau_0$ must be feasible placements at corners.

Procedure *stretch* maintains a list $\Lambda = ((\rho_1, l_1), \ldots, (\rho_k, l_k))$, $k \geqslant 2$, where $\rho_i$ is an oriented placement and $l_i$ is a natural number. List $\Lambda$ specifies a path $P$ consisting of the sequence of jumps through the successive oriented placements, with an additional $l_i$ cycles around the circle $C_{\rho_i}$ at $\rho_i$. (We will eventually show that $l_i = 0$ or 1 and $l_i = 0$ if $i \neq 1, k$.) List $\Lambda$ is initialized to $((\sigma_0, 0), (\tau_0, 0))$, i.e. specifying a path identical to jump $j$.

Wet let $(\sigma, l_\sigma)$, $(\rho, l_\rho)$, and $(\tau, l_\tau)$ always denote the first, penultimate, and ultimate members of list $\Lambda$. $A_\tau$ is the arc of $C_\tau$ traversed in the jump from $\rho$ to $\tau$ and $A_\sigma$ and $A_\rho$ are the arcs of $C_\sigma$ and $C_\rho$ traversed in the jumps from $\sigma$ and $\rho$, respectively, to their successors. $T$ is the tangent segment of the jump from $\rho$ to $\tau$; $\phi$ is the direction of $T$.

We first define *stretch*$(j, L)$. One step of *stretch*$(j, L)$ is to minimally rotate $\tau$ in direction $L$ so that one of the events below happens. "Rotating" $\tau = (u_\tau, \theta, d)$ means changing $\theta$; thus $C_\tau$ rotates about $u_\tau$. As $\tau$ rotates, $A_\tau$ and $T$ move (and change length) and the length of $A_\rho$ changes. The rest of the path is fixed. When one of the events below happens, list $\Lambda$ is updated as described. This process continues until one of stopping conditions (A)–(C) happens. The events are depicted in fig. 7.

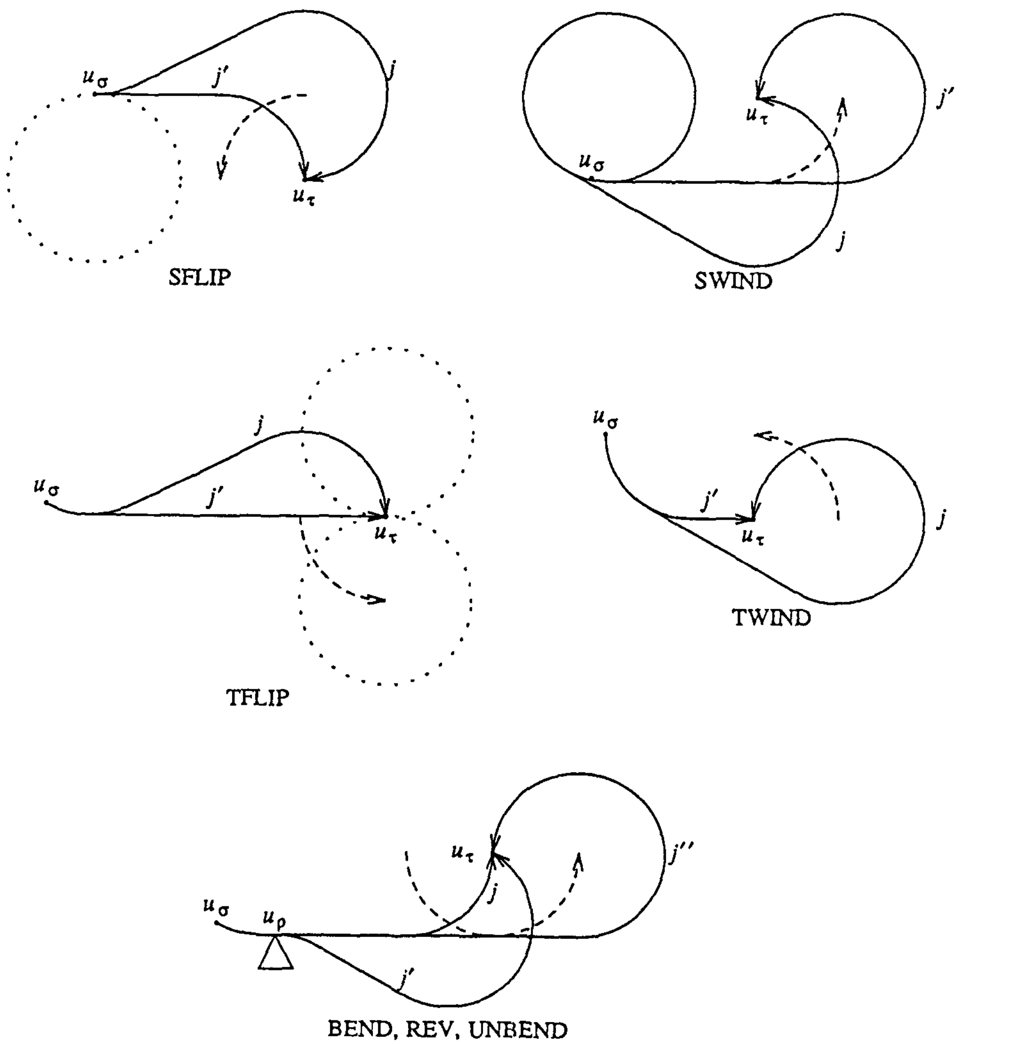


Fig. 7. Events in *stretch*. Jump $j$ is the initial jump; it is deformed until it reaches $j'$, at which point the indicated event happens. Dashed lines indicate motion of $z_\tau$, the center of $C_\tau$. SFLIP: After (SFLIP), the first arc of the jump traverses the dotted circle. SWIND: Initially jump $j$ traverses almost all of $C_\sigma$. Jump $j'$ traverses none of $C_\sigma$. TFLIP: At (TFLIP), the orientation of $C_\tau$ is flipped, i.e. from the lower dotted circle to the upper dotted circle. TWIND: Jump $j$ traverses almost all of $C_\tau$; jump $j'$ traverses none of $C_\tau$. BEND: The tangent segment of $j$ has hit an obstacle, so (BEND) adds a new penultimate oriented placement $\rho$ to $\Lambda$. At $j'$, (REV) occurs and at $j''$, (UNBEND) occurs, removing $\rho$.

(TWIND) There is a jump discontinuity of $|A_\tau|$ from $2\pi$ to 0. Increase $l_\tau$ by 1.

(TFLIP) $|A_\tau|$ decreases to 0. If $l_\tau > 0$ decrement $l_\tau$; otherwise replace $\tau$ with $\bar{\tau}$.

(REV) $|A_\tau|$ attains the value $\pi$. Do nothing; this event just exists for the analysis.

(SWIND) There is a jump discontinuity of $|A_\rho|$ from $2\pi$ to 0. Increase $l_\rho$ by 1.

(SFLIP) $|A_\sigma|$ decreases to 0. If $l_\sigma > 0$ decrement $l_\sigma$, otherwise replace $\sigma$ by $\bar{\sigma}$.

(BEND) Suppose tangent segment $T$ intersects a corner $c$, with the obstacle at $c$ lying on side $e$ of $T$. (If the segment intersects several corners simultaneously, or the segment overlaps a wall, $c$ is chosen to be the corner furthest along the path.) Set $\rho = (c, \phi, e)$ and insert $(\rho, 0)$ as the new penultimate element on the list, where $\phi$ is the direction of $T$. (The effect is that the path will now traverse an $e$-directed circle starting at corner $c$.)

(UNBEND) $|A_\rho|$ decreases to 0 and $\rho \neq \sigma$. If $l_\rho > 0$ decrement it; if $l_\rho = 0$, delete $(\rho, 0)$ from $L$.

The following are the stopping conditions.

(A) The direction of $\tau$ reaches the endpoint of the contact interval.
(B) $A_\tau$ intersects a wall or corner other than $u_\tau$.
(C) The length of $T$ decreases to 0 (so that jump $(\rho, \tau)$ becomes a hop) and $|A_\tau| \neq \pi$.

Comments:

(1) The deformation of *stretch* is continuous, hence all paths resulting from *stretch* are homotopic to the original jump.
(2) The definition of *stretch*$(j, R)$ is similar.
(3) The contact interval at $c$ has size strictly less than $\pi$, hence $\pi$ is an upper bound on the total change of direction stretching both $L$ and $R$.
(4) Corollary 4.2.3 shows that $|A_\rho|$ cannot increase to $2\pi$ if $\rho$ was created by (BEND). Hence $l_\rho = 0$ always for such $\rho$.

Figure 8 depicts two possible families of paths resulting from *stretch*. The path family on the left would arise from invoking *stretch*$(j, L)$ and *stretch*$(j, R)$ on any jump $j$ in the family. The path family on the right arises from applying *stretch*$(j, L)$ to the depicted jump with longest tangent segment.

A direction $\theta$ is *accessible* if only finitely many events happen before $\tau$ is rotated to direction $\theta$ (either for *stretch*$(j, L)$ or *stretch*$(j, R)$). An interval of directions is *accessible* if both its endpoints are accessible. Eventually we prove that only finitely many events happen during *stretch* before one of stopping conditions (A)–(C) happens. However, for now we need the assumption of accessibility to ensure that the list $\Lambda$ is well defined.

If $\theta$ is accessible and an event happens at $\theta$, we define $\Lambda(\theta)$ to be the list $\Lambda$ after updates have been performed. Suppose no event happens at $\theta$ and $\theta$ is accessible; let $\theta'$ be the last event to happen before $\theta$. Then $\Lambda(\theta)$ is $\Lambda(\theta')$ except that the direction of the last oriented placement is $\theta$ rather than $\theta'$. In either case

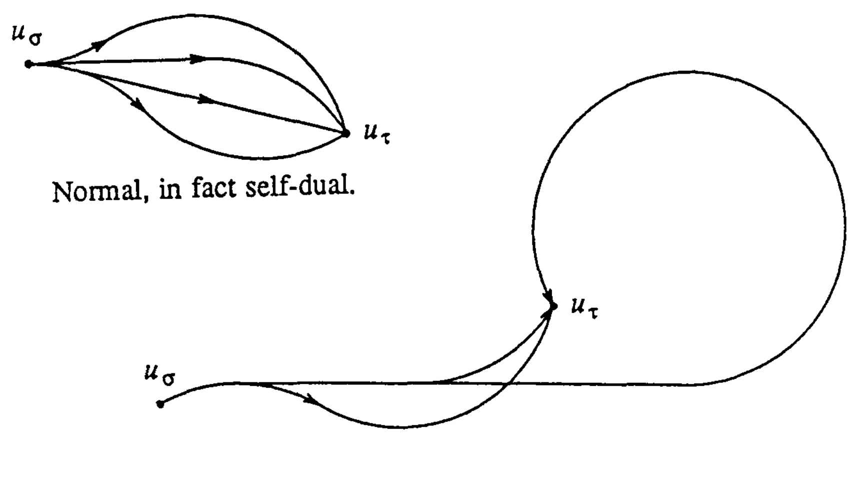


Fig. 8. Path families arising from *stretch*.

$P(\theta)$, $\tau(\theta)$, $\sigma(\theta)$, $\rho(\theta)$, $A_\tau(\theta)$, $A_\rho(\theta)$, $A_\sigma(\theta)$, $\phi(\theta)$ and $T(\theta)$ are defined in the obvious way from $\Lambda(\theta)$.

For the analysis of *stretch*, we often consider events that happen in an (accessible) interval of directions $I$. It is convenient to only have to consider rotation of $\tau$ in a single direction, specifically $L$. It is easy to check that *stretch* is reversible, i.e. *stretch*$(j, L)$ undoes the effect of *stretch*$(j, R)$. Hence we may analyze an interval $I = [\theta', \theta]$ by analyzing the effect of *stretch*$(j, L)$ starting on $\Lambda(\theta')$. If in fact direction $\theta'$ arose from *stretch*$(j, R)$, then the events that happened from $\theta'$ to $\theta_0$ can be deduced from the events that happen from $\theta_0$ to $\theta'$.

We define the *state* of $\Lambda$ from the type of its last jump: if $(\rho, \tau)$ is *XxRl*, then $\Lambda$ is in state (Rl), and similarly for *XxRs*, *XxLs*, and *XxLl*. The state does *not* depend upon the orientation of $\rho$ or the length of $A_\rho$ (see fig. 9). It is easy to

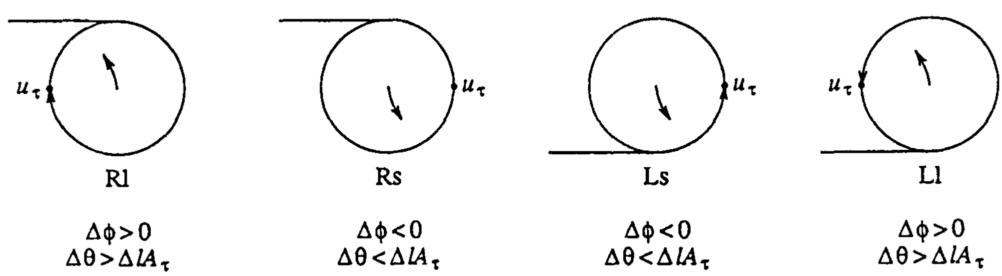


Fig. 9. $C_\tau$ is rotating in direction $L$ about $u_\tau$.

check that the state of $\Lambda(\theta)$ can change only according to the following diagram (using the assumption that $\tau$ rotates only left).

$$\mathrm{Rl} \xrightarrow{\mathrm{(REV)}} \mathrm{Rs} \xrightarrow{\mathrm{(TFLIP)}} \mathrm{Ls} \xrightarrow{\mathrm{(REV)}} \mathrm{Ll} \xrightarrow{\mathrm{(TWIND)}} \mathrm{Ls}$$

Let $lA_\tau$ be $-|A_\tau|$ if $\tau$ is oriented $R$ and $|A_\tau|$ if $\tau$ is oriented $L$. For an (implicit) interval $[\theta, \theta']$ we define $\Delta\theta = \theta' - \theta$, $\Delta\phi = \phi(\theta') - \phi(\theta)$, and $\Delta lA_\tau = lA_\tau(\theta') - lA_\tau(\theta)$. Then $\phi + lA_\tau = \theta$ and $\Delta\phi + \Delta lA_\tau = \Delta\theta$. In states (Rl) and (Ll) $\phi$ is an increasing function of $\theta$ (i.e. the direction of the tangent moves to the left), hence $\Delta\phi > 0$ and $\Delta\theta > \Delta lA_\tau$. In states (Rs) and (Ls) $\phi$ is a decreasing function of $\theta$, hence $\Delta\phi < 0$ and $\Delta\theta < \Delta lA_\tau$. This information is summarized in fig. 9.

Stopping conditions (A) and (B) can happen in any state. Suppose stopping condition (C) happens. Then $\rho$ and $\tau$ must have opposite orientations and the length of $A_\rho$ must be increasing. If $\rho$ is oriented $L$, then $\tau$ must be oriented $R$, $\phi$ must be increasing, and $\tau$ is in state (Rl). If $\rho$ is oriented $R$, then $\tau$ must be oriented $L$, $\phi$ must be decreasing, and $\tau$ is in state (Ls). Hence stopping condition (C) can arise only in states (Ls) and (Rl), with $|A_\tau|$ less than $\pi$ and greater than $\pi$, respectively.

List $\Lambda$ is *obstreperous* if $\rho$ and $\tau$ have the same orientation and $u_\tau$ lies inside $C_\rho$. (The name is chosen because this possibility adds many cases to the arguments that follow.) Necessarily $|A_\tau| > \pi$ if $\Lambda$ is obstreperous, so $\Lambda$ can be obstreperous only in states (Rl) and (Ll). If $\Lambda$ is obstreperous in state (Ll) it remains so until a stopping condition occurs (necessarily (A) or (B)); if $\Lambda$ is obstreperous in state (Rl) possibly also (SFLIP) can occur. Suppose $\Lambda$ is not obstreperous. Then $|A_\tau|$ is increasing if $\tau$ is oriented $L$ and decreasing if $\tau$ is oriented $R$; hence $lA_\tau$ is an increasing function of $\theta$ and $\Delta lA_\tau > 0$.

LEMMA 4.2.1

Consider the sequence of events in the accessible interval $[\theta, \theta']$. If (REV) causes a transition from state (Ls) to state (Ll), then (TWIND) does not subsequently happen.

*Proof*

In state (Ll), $\Delta\theta > \Delta lA_\tau$. If both (REV) and (TWIND) happen, then $\Delta lA_\tau \geqslant \pi$. But this is impossible, since $\Delta\theta$ is at most the size of the contact interval range at $c$, i.e. less than $\pi$. □

Lemma 4.2.1 implies that $\Lambda$ can go through at most the sequence of states (Rl),(Rs),(Ls),(Ll) or (Ll),(Ls).

An interval is *monotonic* if neither (TWIND) nor (REV) happens during the interval. During a monotonic interval, list $\Lambda$ is always in exactly one of states

(Rl), (Rs), (Ls), (L1), or is in state (Rs) followed by (Ls). Clearly $\phi$ is monotonic in the interval.

LEMMA 4.2.2

Suppose $I=[\theta_0, \theta_f]$ is an accessible monotonic interval and $\Lambda(\theta_0)$ describes a jump. Then $|\Delta\phi| < \pi$ if $\Lambda$ is never obstreperous and $|\Delta\phi| < 2\pi$ if $\Lambda$ is obstreperous.

*Proof*

First suppose list $\Lambda$ remains in a single state. We have $0 < \Delta\theta < \pi$ and $-\pi \leqslant \Delta lA_\tau \leqslant \pi$ since $\Lambda$ does not change state. If $\Lambda$ is never obstreperous then $\Delta lA_\tau > 0$; using $\Delta\phi + \Delta lA_\tau = \Delta\theta$ yields $-\pi < \Delta\phi < \pi$. If $\Lambda$ is obstreperous then $-\pi < \Delta\phi < 2\pi$. The case that $\Lambda$ is in state (Rs) followed by (Ls) is considerably harder, and the proof is deferred to appendix 1. □

COROLLARY 4.2.3

Suppose $I=[\theta_0, \theta_f]$ is an accessible monotonic interval and $\Lambda(\theta_0)$ describes a jump. Then (SWIND) does not follow (SFLIP) or (SWIND), hence $l_\sigma = 0$ or 1 always. If $\rho$ is created by (BEND) then $|A_\rho| < 2\pi$.

Let $r(\rho, \alpha)$ be the ray tangent to $C_\rho$ in direction $\alpha$ and for $J \subseteq S^1$ set $r(\rho, J) = \{r(\rho, \alpha): \alpha \in J)\}$. It is clear that if $|[\alpha_0, \alpha_1]| \leqslant \pi$, $\alpha' \in [\alpha_0, \alpha_1]$ and $\rho'$ is an oriented placement lying on $r(\rho, \alpha')$ with the same orientation as $\rho$, then $r(\rho', [\alpha', \alpha_1]) \subseteq r(\rho, [\alpha_0, \alpha_1])$ (if $\rho$ is oriented $L$) or $r(\rho', [\alpha_0, \alpha']) \subseteq r(\rho, [\alpha_0, \alpha_1])$ (if $\rho$ is oriented $R$).

LEMMA 4.2.4

Suppose $I=[\theta_0, \theta_f]$ is an accessible monotonic interval and $\Lambda(\theta_0)$ describes a jump. Then event (BEND) happens at most once per corner during $I$.

*Proof*

We prove the lemma first for states (Rs), (Ls), and (R1). Note that $\Lambda$ is never obstreperous in states (Rs) and (Ls), and is not obstreperous after the first instance of (BEND) in state (R1) since a created oriented placement in $\Lambda$ must have orientation $L$.

Suppose event (BEND) occurs at $\theta_i$ creating placement $\rho_i$, $i=1, 2, \ldots$. (As in the proof of lemma 4.2.2 above, we consider (SFLIP) an instance of (BEND).) Set $J_i = [\phi(\theta_i), \phi(\theta_1)+\pi]$ if the tangent is moving left and $J_i = [\phi(\theta_1)-\pi, \phi(\theta_i)]$ if the tangent is moving right. By lemma 4.2.2 $\phi$ changes by at most $\pi$, hence we have $J_{i+1} \subseteq J_i$. Also all the $\rho_i$ have the same orientation, so $r(\rho_{i+1}, J_{i+1}) \subseteq r(\rho_i, J_i)$. Since $u_{\rho_i} \in r(\rho_i, \phi(\theta_i))$ and $\phi(\theta_{i+1}) > \phi(\theta_i)$ we must have $u_{\rho_i} \notin r(\rho_{i+1}, J_{i+1})$. Hence all corners causing (BEND) must be distinct.

The remaining case is that $\Lambda$ is in state (L1). We show that $|u_{\rho_{i+1}} u_\tau| < |u_{\rho_i} u_\tau|$ as long as $|u_{\rho_i} u_\tau| > 2$; as soon as $|u_{\rho_i} u_\tau| \leqslant 2$ we show the direction from $u_\tau$ to

$u_{\rho_{i+1}}$ is to the left of the direction from $u_\tau$ to $u_{\rho_i}$. First suppose $|u_{\rho_i}u_\tau| > 2$. Let $k$ be the point of tangency between $T(\theta_i)$ and $C_{\tau(\theta_i)}$ and $k'$ the point of tangency between $T(\theta_{i+1})$ and $C_{\tau(\theta_{i+1})}$. Then $u_{\rho_{i+1}}$ lies inside the quadrilateral $Q = u_\tau k' k u_{\rho_i}$. Now $|u_\tau k'|$, $|u_\tau k| \leqslant 2$ and $2 < |u_\tau u_{\rho(\theta_i)}|$. Hence $u_{\rho_i}$ is the point furthest from $u_\tau$ inside the quadrilateral $Q$, and $|u_\tau u_{\rho_{i+1}}| < |u_\tau u_{\rho_i}|$.

Let $\rho_i$ be the first created oriented placement with $|u_{\rho_i}u_\tau| \leqslant 2$. The tangent from $C_{\rho_i}$ to $C_\tau(\theta)$ for $\theta \geqslant \theta_i$ lies entirely within the circle of radius 2 about $u_\tau$, hence also $|u_{\rho_{i+1}}u_\tau| \leqslant 2$, and in fact for all $j \geqslant i$, $|u_{\rho_j}u_\tau| \leqslant 2$. Choose $\theta_t$ so that $u_{\rho_i}$ lies on the semicircular arc of $C_{\tau(\theta_t)}$ from the point diametrically opposite $u_\tau$ to $u_\tau$. Certainly $\theta$ cannot move left past $\theta_t$, because $|A_\tau| \geqslant \pi$ implies stopping condition (B) must happen at $\theta_t$ or before.

Suppose (BEND) creates $\rho_{j+1}$, $j \geqslant i$. The direction from $u_\tau$ to $u_{\rho_{j+1}}$ must be to the left of the direction from $u_\tau$ to $u_{\rho_j}$, since $u_\tau$ must lie to the left of the line tangent to $P(\theta_{j+1})$ at every point between $u_{\rho_j}$ and $u_{\rho_{j+1}}$. However, the direction from $u_\tau$ to $u_{\rho_{j+1}}$ can be at most as far left as the direction from $u_\tau$ to the point diametrically opposite $u_\tau$ in $C_{\tau(\theta_f)}$. Hence $u_i$, $u_{i+1}, \ldots,$ must all be at distinct corners. □

COROLLARY 4.2.5

If $I = [\theta_0, \theta_f]$ is a monotonic interval and $\Lambda(\theta_0)$ describes a jump, then $I$ is accessible.

*Proof*

By lemmas 4.2.1 through 4.2.4, for each event there is a fixed bound on how many times it can happen in an accessible monotonic interval. If infinitely many events happen in $I$, then in some accessible prefix of $I$, some bound on a particular event must be violated. □

Suppose $\Lambda(\theta)$ is not obstreperous. Let $\tau(\theta) = (u_\tau, \theta, d)$. If $|A_\tau(\theta)| \neq 0, \pi$, define $\theta^* \neq \theta$, $\tau^* = (u_\tau, \theta^*, d)$ so that $C_{\tau^*}$ is tangent to the ray $r(\rho(\theta), \phi(\theta))$. Notice that if $\alpha$ is the length of the arc of $C_{\tau^*}$ on jump $(\rho(\theta), \tau^*)$, then $\alpha + |A_\tau| = 2\pi$. If $|A_\tau| = 0, \pi$ set $\theta^* = \theta$.

The next lemma shows that if (BEND) happens at $\theta$, then (UNBEND) happens at $\theta^*$. There is a difficulty that $\theta^*$ is not defined if $\Lambda(\theta)$ is obstreperous ($\tau^*$ can be chosen so that $C_{\tau^*}$ is tangent to the line containing $r(\rho(\theta), \phi(\theta))$ but not to $r(\rho(\theta), \phi(\theta))$ itself). However, (BEND) can happen with $\Lambda$ obstreperous only in state (L1). It is easy to see that subsequently $\Lambda$ must remain obstreperous in state (L1) and event (TWIND) cannot happen. Hence an arbitrary definition of $\theta^*$ in the obstreperous case is adequate for the following lemma, since the hypothesis that (TWIND) occurs is false.

LEMMA 4.2.6

Suppose $I = [\theta_0, \theta_f]$ is a monotonic interval, events (SWIND), (SFLIP), and (BEND) occur at $\theta_1, \theta_2, \ldots, \theta_k$ in $I$, and event (TWIND) or (REV) happens at $\theta_f$.

Then for $\theta$ occurring after $\theta_f$ and before $\theta_0^*$:

(1) If (BEND) occurred at $\theta_i = \theta^*$ creating $\rho_i$, then (UNBEND) occurs at $\theta$ deleting $\rho_i$.

(2) If (SWIND) or (SFLIP) occurred at $\theta_i = \theta^*$, then (SFLIP) occurs at $\theta$. The orientation of $\sigma$ changes at $\theta$ exactly if it did at $\theta_i$.

(3) No other events happen.

*Proof*

It suffices to show that if (1), (2), and (3) are true for $\theta' < \theta$, then (1), (2) and (3) are true at $\theta$ as well. So suppose (1), (2), and (3) are true for $\theta' < \theta$. Choose $i$ so that $\theta_i \leqslant \theta^* < \theta_{i+1}$ and let $\rho_i = \rho(\theta^*)$. A trivial induction using (1), (2), and (3) now shows that before events happen at $\theta$, if any, $\rho_i$ must also be the penultimate placement of $\Lambda(\theta)$.

We first show (1) and (2). Suppose $\theta = \theta_i^*$, then one of (SWIND), (SFLIP) or (BEND) occurred at $\theta_i$. Hence $|A_\rho(\theta')|$ increases from 0 as $\theta'$ increases from $\theta_i$, so $|A_\rho(\theta')|$ decreases to 0 as $\theta'$ increases towards $\theta = \theta_i^*$. If (BEND) happened at $\theta_i$, then $\rho_i \neq \sigma(\theta_i)$ and (UNBEND) happens at $\theta$. If (SWIND) happened at $\theta_i$, then $\rho_i = \sigma(\theta_i)$, $l_\sigma$ was incremented at $\theta_i$, and the orientation of $\sigma$ was not flipped at $\theta_i$. Hence (SFLIP) happens at $\theta$, $l_\sigma$ is decremented, and $\sigma$ is not flipped. Similarly if (SFLIP) happened at $\theta_i$, then $\rho_i = \sigma(\theta_i)$ and the orientation of $\sigma$ was flipped. At $\theta$ (SFLIP) happens again and again flips the orientation of $\sigma$.

Next we show (3). Event (TWIND) cannot happen at $\theta$ for then the event at $\theta_f$ would be (REV) changing state from (Ls) to (L1), contradicting lemma 4.2.1. Event (TFLIP) cannot happen, for then the event at $\theta_f$ would be (REV) changing state from (R1) and (Rs), $\Lambda(\theta_0)$ would be in state (R1), $|A_\tau(\theta_0^*)| > 0$, so $\theta > \theta_0^*$, contradicting $\theta < \theta_0^*$. Similarly (REV) cannot happen at $\theta$. If $\theta \neq \theta_i^*$, then $|A_\rho(\theta)| \neq 0$, so (SWIND), (SFLIP), and (UNBEND) do not happen at $\theta$.

Showing (BEND) does not happen is a little harder. We assume $|A_\tau(\theta)| > \pi$ and $|A_\tau(\theta^*)| < \pi$; the other case is similar but easier. This assumption implies $\tau$ is oriented $L$, $|A_\tau|$ is increasing, the tangent is moving right in $[\theta_0, \theta_f]$ and moving left after $\theta_f$. For simplicity we assume no event happens at $\theta^*$, hence there is a small interval $H = [\theta_h, \theta^*]$ in which $\rho$ has constant value $\rho_h$. Clearly $\rho(\theta) = \rho_h$. Let $k$ be the endpoint of $T(\theta)$ not on $C_{\rho_h}$. For $\alpha \in H$ define $s(\alpha)$ to be the segment extending $T(\alpha)$ to have endpoint on line $u_\tau k$. We claim $s(\alpha)$ does not intersect the interior of any obstacle. Now clearly $T(\alpha)$ forms an initial segment of $s(\alpha)$ and does not intersect the interior of any obstacle, since it is part of a feasible path. Let $a(\alpha) = s(\alpha) - T(\alpha)$. We claim every point of $a(\alpha)$ lies on $A_\tau(\alpha')$ for some $\alpha' \in H$. This follows from the observations that every point of $a(\alpha)$ is closer to $u_\tau$ than the length of the chord subtending $A_\tau(\alpha)$, and that $|A_\tau(\alpha^*)| \leqslant |A_\tau(\alpha')| \leqslant |A_\tau(\alpha)|$ for $\alpha' \in H$. Hence no point of $a(\alpha)$ lies in the interior of any obstacle.

Consider region $s(H) = \{s(\alpha)\colon \alpha \in H\}$. Now $T(\theta) = s(\theta^*)$ forms part of the boundary of $s(H)$. If some obstacle intersects the interior of $T(\theta)$ on the left side of $T(\theta)$, then it must also intersect the interior of $s(H)$. But this is impossible, since no point in the interior of $s(H)$ can intersect the interior of any obstacle. Hence (BEND) does not happen at $\theta$. □

THEOREM 4.2.7

Procedure *stretch* terminates after at most $O(n)$ steps.

*Proof*

The sequence of states of $\Lambda$ is either a subsequence of (R1),(Rs),(Ls),(L1), or (Ls),(L1),(Ls). We analyze the case that $\Lambda$ actually goes through all states of the first sequence. The other cases are similar.

Suppose $\Lambda$ is initially in state (R1) with direction $\theta_0$. By corollary 4.2.5, the interval until (REV) happens is accessible, and by lemmas 4.2.1–4.2.4 at most $O(n)$ events happen. By lemma 4.2.6, at most $O(n)$ events happen until $\theta_0^*$. At $\theta_0^*$, again by lemma 4.2.6, $\Lambda$ describes a jump. Hence the argument can be repeated for the remaining sequence of states. □

### 4.3. CATEGORIZING PATH FAMILIES

If $j$ is a jump, we let $S_j$ be the interval of final placements and $P_j$ (more explicit than simply $P$) be the family of paths resulting from *stretch*$(j, L)$ and *stretch*$(j, R)$. Thus $P_j$ is defined on $S_j$ and for $\theta \in S_j$, $P_j(\theta)$ is a path with final placement $\theta$. Notice $P_j$ contains enough information to reconstruct the sequence of events occurring during *stretch*.

If either *stretch*$(j, L)$ or *stretch*$(j, R)$ terminated because of stopping conditions (A) or (B), we say $P_j$ is *saturated*. Clearly if $P_j$ is saturated, one endpoint of $S_j$ is a root, and if $P_j$ is not saturated, stopping condition (C) terminates both *stretch*$(j, L)$ and *stretch*$(j, R)$. (Notice if $\Lambda$ is ever obstreperous, then necessarily $P_j$ is saturated.) If (TWIND) occurs during *stretch*, then $P_j$ is *winding*; if not, then $P_j$ is *normal* (see fig. 8). Suppose $P_j$ is normal and not saturated; if (BEND) does not occur during *stretch* then $P_j$ is *self-dual* and if (BEND) does occur then $P_j$ is *heterodual*. The main theorem of this section is the following.

THEOREM 4.3.1

Suppose $j$ is a feasible jump. Then there is a feasible jump $j'$ with the same source placement as $j$ so that $S_j$ overlaps $S_{j'}$ and $P_{j'}$ is either saturated, self-dual, or heterodual.

*Proof*

If $P_j$ is saturated or normal the theorem is trivial, since we can set $j' = j$. The other possibility is that $P_j$ is winding and not saturated. (An example of such a path family $P_j$ can be obtained from fig. 8. Add to it the jump obtained by rotating $C_\tau$ – the one with more than $3\pi/2$ of arc – clockwise until it hits $C_\sigma$.) By lemma 4.2.1, the sequence of states of $\Lambda$ rotating left from $\theta_r$ to $\theta_l$ must be a sequence of (R1),(Rs),(Ls),(L1) or (L1),(Ls). Stopping condition (C) can happen in state (R1) or (Ls) rotating left, and in state (L1) or (Rs) rotating right. Hence the sequence of states must be (Rs)(Ls) or (L1)(Ls). The former cannot happen since $P_j$ is winding. We cannot have $j$ be *XxLs*, since (TFLIP) must happen. Hence $j$ is *XxLl*.

Let $\theta$ be the final direction of $j$. We claim $\theta^* \in S_j$. (Note $\theta^*$ is defined, since the fact that (TWIND) happens implies $\Lambda$ is never obstreperous.) Stopping conditions (A) and (B) never happen since $P_j$ is not saturated. We show stopping condition (C) does not happen in $[\theta, \theta^*]$. First notice that the tangent moves left in the state (L1), hence any intermediate oriented placement $\rho$ has orientation $L$. Hence $C_\rho$ can never become tangent to ($L$-oriented) $C_\tau$. Similarly, $C_\tau$ cannot become tangent to $C_\sigma$ if $\sigma(\theta)$ is oriented $L$. If $\sigma(\theta)$ is oriented $R$, then $C_{\tau(\theta)}$ and $C_{\sigma(\theta)}$ are on opposite sides of the line through the tangent $T(\theta)$. Furthermore, $C_\tau$ remains in its halfplane until $\theta^*$, hence $C_\tau$ cannot become tangent to $C_\sigma$ before $\theta^*$.

Let $j' = (\sigma(\theta), \tau(\theta^*))$. Clearly $j'$ specifies the same path as $P_j(\theta^*)$, hence $j'$ is feasible. Now $j'$ is *XxLs*; this easily implies that $P_{j'}$ must be normal. Finally $S_j$ overlaps $S_{j'}$ since both contain $\theta^*$. □

The next easy lemma characterizes self-dual path families. The description of heterodual path families is more complicated, and appears in section 4.5.

LEMMA 4.3.2

If $P_j$ is self-dual then $P_j(\theta_r)$ is an *LsRs* hop and $P_j(\theta_l)$ is an *RsLs* hop, where $S_j = [\theta_r, \theta_l]$.

*Proof*

Since $P_j$ is self-dual, it is normal, not saturated, and never obstreperous. Hence stopping condition (C) terminates both *stretch*$(j, L)$ and *stretch*$(j, R)$. Thus (TFLIP) must occur, since (BEND) does not occur, stopping condition (C) happens because of contact between $\sigma$ and $\tau$ in both cases. The shortness of source hops follows from lemma 4.2.2; the target hops are short because (REV) does not occur. □

Eventually we will need to propagate intervals of reachable placements that have open endpoints. The next lemma answers the following question. Suppose interval $I$ with open endpoint $\omega$ is reachable at corner $a$, and $j$ is a feasible jump with $\Omega(j) = \omega$. What placements in $S_j$ are actually reachable from $I$? The lemma is not used until section 6.

LEMMA 4.3.3

Suppose $a$ and $b$ are corner contacts, $I$ is an interval at $a$ with open endpoint $\omega$, $j' \in \Xi^t_{ab}(I)$, and $j \in X(j', I \cup \{\omega\})$ with $\Omega(j) = \omega$. If $\theta \in S_j$ and $P_j(\theta)$ is not a hop, then placement $\theta$ is reachable from $I$.

*Proof*

We show that if $k = P_j(\theta)$ is a jump (but not a hop), then there is some placement on the tangent segment of $k$ that is reachable from $I$. In fact, it will be clear that any placement on the tangent segment of $k$ sufficiently close to the source arc of $k$ is reachable from $I$. This proves the lemma, since even if $P_j(\theta)$ is not a jump (because of (BEND)), there must be some jump $k = P_j(\theta')$ with tangent segment overlapping the initial straight segment of $P_j(\theta)$ and some placement in the overlap is reachable from $I$.

We can assume $j = (\sigma, \tau)$ is not a hop, otherwise we can perturb $j$ slightly, for example as done by *stretch*. Since neither $j$ nor $k$ is a hop, there is some minimum length $\delta > 0$ of the tangent segment of $P_j$ during *stretch* from $j$ to $k$. We let $F$ be the region of the plane swept out by the initial portion of length $\delta$ of the tangent segment during *stretch* from $j$ to $k$. Clearly the interior of $F$ avoids obstacles. If (SFLIP) does not occur, $F$ is a region bounded by an arc $A$ of $C_\sigma$, two segments of length $\delta$ tangent to $A$, and an arc $B$ of radius $\sqrt{1+\delta^2}$. If (SFLIP) does occur, $F$ has a more complicated but similar description. We henceforth assume (SFLIP) does not occur; the argument in the case it does is essentially the same as the case when $\sigma$ has orientation $R$, given below.

We assume that the direction of the tangent moves monotonically, specifically left, during *stretch* from $j$ to $k$. Hence the line segments of length $\delta$ forming part of the boundary of $F$ are initial portions of the tangent segments of $j$ and $k$. We label them $T_j$ and $T_k$, respectively. If the tangent segment does not move monotonically, it suffices to truncate $F$ so that the initial portions of the tangent segments of $j$ and $k$ do form part of its boundary.

We have $j' \in \Xi^t_{ab}(I)$ and $j \in X(j, I \cup \{\omega\})$, so there is a continuous deformation of $j'$ into $j$. In fact, there must be an interval $I' \subseteq I$ with open endpoint $\omega$ and a family $d$ so that for $\alpha \in I' \cup \{\omega\}$, $d(\alpha)$ is a feasible jump of the same type as $j$, the source placement of $d(\alpha)$ is $\alpha$, $d$ is continuous on $I' \cup \{\omega\}$, and $d(\omega) = j$. We let the source circle of $d(\alpha)$ be $C_\alpha$; clearly it has the same orientation as $\sigma$. Let $T_\alpha$ be the ray tangent to $C_\alpha$ with endpoint on $C_\alpha$ parallel to $T_k$, if $\sigma$ is oriented $L$, or parallel to $T_j$, if $\sigma$ is oriented $R$.

The proof strategy is as follows. We show that any placement on $T_k$ sufficiently close to $A$ is reachable from some placement in $I$ by a feasible path $p$. Path $p$ starts along one of the jumps in the family $d$, possibly diverges from the jump, enters $F$, and stays within $F$ until reaching $T_k$. The divergence from the jump happens in two ways: either by staying on the circle $C_\alpha$ until entering $F$, or by following the ray $T_\alpha$ to enter $F$. The argument is split into cases depending upon the orientation of $\sigma$.

The easier case is if $\sigma$ has orientation $L$. Then $A$ is an $L$-arc, $T_k$ is tangent to $A$ at its right endpoint and $T_j$ is tangent to $A$ at its left endpoint. Let $l$ be the arc length along $C_\sigma$ from corner $a$ to $T_k$. There are two possibilities; either circles $C_\alpha$ intersect $T_k$ or the rays $T_\alpha$ intersect $A$. Consider the first possibility: this happens if $l < \pi$ and $\omega$ is the left endpoint of $I$ or if $l > \pi$ and $\omega$ is the right endpoint of $I$. The desired path $p$ follows $C_\alpha$ until intersection with $T_k$, then follows a right-left path within $F$ to gain tangency with $T_k$. By choosing $\alpha$ sufficiently close to $\omega$ we can guarantee that $p$ is feasible: part of $p$ lies on feasible jump $d(\alpha)$, possibly diverges from $d(\alpha)$ by continuing on $C_\alpha$ until $T_k$ (but any such point must lie on $d(\alpha')$ for some $\alpha'$ sufficiently close to $\omega$), and the rest of the path lies in $F$. Since the angle between $p$ and $T_k$ at their intersection goes to zero as $\alpha$ approaches $\omega$, there is room inside $F$ to have a curvature-constrained path become tangent to $T_k$. The second possibility, that the rays $T_\alpha$ intersect $A$, happens if $l \geqslant \pi$ and $\omega$ is the left endpoint of $I$ or if $l \leqslant \pi$ and $\omega$ is the right endpoint of $I$. The required path $p$ is constructed using $C_\alpha$ and $T_\alpha$ and then a right arc and a left arc inside $F$ to gain tangency with $T_k$. The feasibility argument follows as before.

Now suppose $\sigma$ has orientation $R$. Then $A$ is an $R$-arc, $T_k$ is tangent to $A$ at its right endpoint and $T_j$ is tangent to $A$ at its left endpoint. Again there are two possibilities (determined by conditions on $l$ and $I$). The first is that for $\alpha \in I$, $C_\alpha$ intersects $T_j$. In this case, there is some $L$-oriented circle $D_\alpha$ tangent to both $C_\alpha$ and $C_\sigma$. By choosing $\alpha$ sufficiently close to $\omega$ we can guarantee that the point $D_\alpha \cap C_\alpha$ is past the point $T_j \cap C_\alpha$ on $C_\alpha$. If the point $T_k \cap C_\sigma$ is past the point $D_\alpha \cap C_\sigma$ on $C_\sigma$, then the required path $p$ traverses $C_\alpha$, $D_\alpha$, $C_\sigma$ and $T_k$. Otherwise $T_k$ intersects $D_\alpha$; we must instead choose a circle $D'_\alpha$ tangent to both $T_k$ and $C_\alpha$. For $\alpha$ sufficiently close to $\omega$ we can guarantee that the point $D'_\alpha \cap C_\alpha$ is past the point $T_j \cap C_\alpha$ on $C_\alpha$. Then the required path $p$ traverses $C_\alpha$, $D'_\alpha$, and $T_k$. The second possibility is that $T_\alpha$ intersects $A$. Consider an $R$-oriented unit circle $E_\alpha$ tangent to $T_\alpha$ at the point $T_\alpha \cap A$. This circle diverges from $A$, but never diverges by more than the distance between the center of $A$ and the center of $E_\alpha$, which can be made arbitrarily small by letting $\alpha$ approach $\omega$. Depending upon the arc length of $A$, and $L$-oriented circle can be tangent to either $E_\alpha$ and $T_k$ or $E_\alpha$ and $A$. Hence the required path $p$ can be constructed. For either possibility the feasibility argument is as before. □

### 4.4. SELF-DUAL INTERVALS AND HOP FUNCTIONS

Let $h_d$ be the function mapping source angle to target angle of *RsLs* hops between two points at distance $d$ apart, where the straightline direction between the two points is taken to be zero. More formally, $h_d(\omega) = \theta$ if $j = (\hat{\omega}, \hat{\theta}) = ((e, \mu + \omega, R), (f, \mu + \theta, L))$ is an *RsLs* hop from $e$ to $f$, the distance between $e$ and $f$ is $d$, and $\mu$ is the straightline direction from $e$ to $f$.

The single-valuedness of $h_d$ follows from the fact that only short hops are allowed; similarly $h_d$ is one–one. It is also clear that $h_d$ is continuous and monotonically decreasing. Notice $h_d$ is not defined for every angle; indeed if $d > 4$ then $h_d$ is nowhere defined. We analyze the domain of definition of $h_d$ below. In a similar fashion we define $\bar{h}_d$, the function mapping source angle to target angle of *LsRs* hops.

Set $H_d(\omega) = [\bar{h}_d(\omega), h_d(\omega)]$. An interval $I$ at $f$ is *self-dual* (*at f from e*) if there is an angle $\omega$ so that $I = H_d(\omega) + \mu$, where the distance from $e$ to $f$ is $d$ and $\mu$ is the straightline direction from $e$ to $f$. Clearly, if $j$ is a jump from $e$ to $f$ and $P_j$ is self-dual, then $S_j$ is a self-dual interval at $f$ from $e$; in particular, $S_j = H_d(\Omega(j) - \mu) + \mu$, where $\mu$ is the straightline direction from $e$ to $f$. The main theorem of this section is the following. (We show below that $\mathrm{dom}(H_d) = [-\alpha_d, \alpha_d]$ for some $\alpha_d < \pi$.)

THEOREM 4.4.1

Suppose $0 < d \leqslant 4$ and $\omega \in \mathrm{dom}(H_d) = [-\alpha_d, \alpha_d]$; then there is $\gamma_d > 0$ as follows. If $\omega \leqslant 0$, set $\Delta = \omega + \alpha_d$, then either $0 \in H_d(\omega)$ or $\gamma_d + [-2\Delta, -\Delta/2] \subseteq H_d(\omega)$. Similarly if $\omega \geqslant 0$, set $\Delta = \alpha_d - \omega$, then either $0 \in H_d(\omega)$ or $-\gamma_d + [\Delta/2, 2\Delta] \subseteq H_d(\omega)$.

Stated informally, either $H_d(\omega)$ contains zero, or $H_d(\omega)$ lies on the opposite side of 0 from $\omega$ and its size grows linearly with the distance of $\omega$ from the closer endpoint of $\mathrm{dom}(H_d)$. In terms of the jump $j = (\sigma, \tau)$ from $e$ to $f$ with $P_j$ self-dual, either $S_j$ contains $\mu$ or $S_j$ is on the opposite side of $\mu$ from $\Omega(j)$, and $S_j$ grows linearly in size as $\Omega(j)$ moves toward $\mu$.

The remainder of this section is devoted to the proof of theorem 4.4.1. We begin with the following simple proposition.

PROPOSITION 4.4.2

(1) If $h_d(\omega) = \theta$, then $h_d(\theta) = \omega$.
(2) If $h_d(\omega) = \theta$, then $\bar{h}_d(-\omega) = -\theta$.
(3) If $d' > d$ then $h_{d'}(\omega) > h_d(\omega)$ if both are defined.

*Proof*

Let $j = (\hat{\omega}, \hat{\theta})$ defined as before be an *RsLs* hop from $e$ to $f$ at distance $d$ apart. For (1), consider the path backwards. For (2), reflect the path about the line through $e$ and $f$. For (3), imagine translating $C_{\hat{\theta}}$ by a distance $d' - d$ in the direction from $e$ to $f$. It must then rotate to the left about the translated image of $f$ before it is tangent to $C_{\hat{\omega}}$. □

The domain of definition of $h_d$ can be inferred as follows. Fix $e$ at the origin and $f$ at $(d, 0)$. First suppose $d \leqslant 2$. Consider the left unit circular arc through $e$ and $f$ whose length is at most $\pi$, and let $\alpha_d$ be the angle between the $x$-axis and the tangent at $f$ (fig. 10). Explicitly we have $\alpha_d = \arcsin(d/2)$. It is easy to see that $\mathrm{dom}(h_d)$ is exactly $[-\alpha_d, \alpha_d]$ and that $h_d$ decreases monotonically from $\alpha_d$

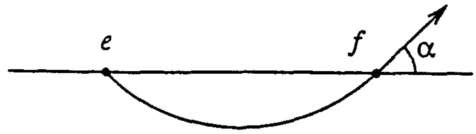


Fig. 10. Definition of $\alpha$.

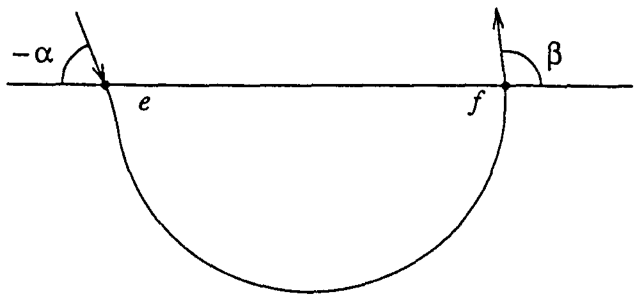


Fig. 11. Definition of $\beta$.

to $-\alpha_d$ in that interval. For symmetry with the case $d > 2$ we define $\beta_d = \alpha_d$, hence $\text{dom}(h_d) = [-\alpha_d, \beta_d]$.

Now suppose $4 \geqslant d > 2$. Consider the $RsL\pi$ hop $j$ from $e$ to $f$. Let the angle between the $x$-axis and $\Omega(j)$ be $-\alpha_d$ and the angle between the $x$-axis and $\Theta(j)$ be $\beta_d$ (fig. 11). Notice necessarily $\beta_d \geqslant \pi/2 \geqslant \alpha_d$, with both equalities holding only if $d = 4$. It is easy to see that $h_d$ is defined exactly on $[-\alpha_d, \beta_d]$ and that $h_d$ decreases monotonically from $\beta_d$ to $-\alpha_d$ over that interval.

In a similar fashion it follows that the domain of $\bar{h}_d$ is $[-\beta_d, \alpha_d]$ and $\bar{h}_d$ descends monotonically from $\alpha_d$ to $-\beta_d$ over this interval. Also, $\text{dom}(H_d) = \text{dom}(\bar{h}_d) \cap \text{dom}(h_d) = [-\beta_d, \alpha_d] \cap [-\alpha_d, \beta_d] = [-\alpha_d, \alpha_d]$. The domains of $h_d$ and $\bar{h}_d$ as function of $d$ are plotted in fig. 12. It can be checked that $\alpha_d = \beta_d$ increase from 0 to $\pi/2$ at $d = 2$. Then $\alpha_d$ decreases, crossing 0 at $\sqrt{8}$ and

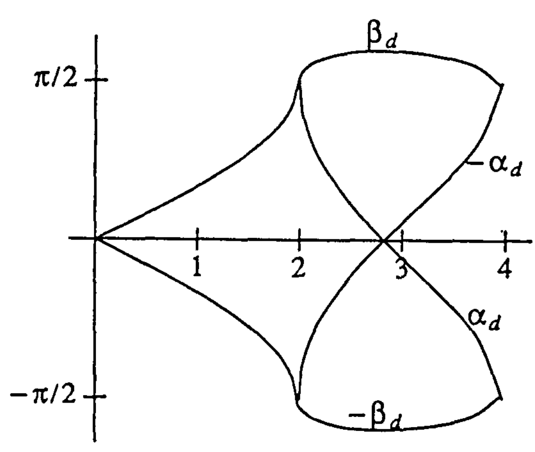


Fig. 12. $\text{dom}(h_d) = [-\alpha_d, \beta_d]$ and $\text{dom}(\bar{h}_d) = [-\beta_d, \alpha_d]$.

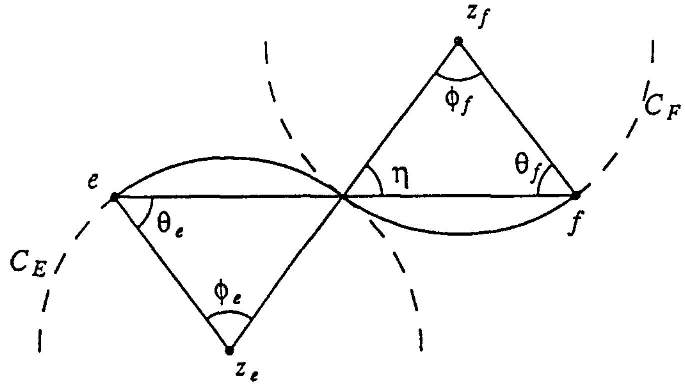


Fig. 13. Definition of angles.

reaching $-\pi/2$ at $d=4$; $\beta_d$ continues to increase to $\pi/2 + \arcsin 1/3$ at $d=\sqrt{8}$, then decreases to $\pi/2$ at $d=4$.

We now define a set of points and angles determined by $\omega \in \mathrm{dom}(h_d)$. As before set $e$ at the origin at $f$ at $(d, 0)$. Define oriented placements $E=(e, \omega, R)$ and $F=(f, h_d(\omega), L)$. Clearly $(E, F)$ is a short $RL$-hop from $e$ to $f$. Let $z_e$ be the center of circle $C_E$ and $z_f$ the center of $C_F$. Notice $|ez_e| = |fz_f| = 1$ and $|z_e z_f| = 2$ since $C_E$ and $C_F$ are tangent. Let $\phi_e$ be the angle $z_f z_e e$, $\phi_f$ the angle $z_e z_f f$, $\theta_e$ the angle $z_e ef$, $\theta_f$ the angle $z_f fe$, and $\eta$ the angle from line $ef$ to line $z_e z_f$ (fig. 13). (Figure 13 is drawn so that the midpoint of segment $ef$ coincides with the midpoint of segment $z_e z_f$; in general this is not true.) Notice that $\theta_e = \pi/2 - \omega$ and $\theta_f = \pi/2 - h_d(\omega)$; $\theta_e$ is negative if $z_e$ is above the $x$-axis and $\theta_f$ is negative if $z_f$ is below the $x$-axis. (Both these possibilities can happen only if $d>2$.) All of $\theta_e$, $\theta_f$, $\phi_e$, $\phi_f$, $z_e$, and $z_f$ depend functionally upon $\omega$. We occasionally write for example $\theta_e(\omega)$ to make this dependence explicit.

A few examples may illustrate the definitions of these angles. If $\omega = -\alpha_d$ then $h_d(\omega) = \beta_d$, and $\theta_e = \pi/2 + \alpha_d$. If $d \leqslant 2$ then $\beta_d = \alpha_d$, $\theta_f = \pi/2 - \alpha_d$, $\phi_e = 0$, and $\phi_f = 2\arcsin(d/2)$ (fig. 14a) and if $4 > d > 2$ then $\beta_d > \pi/2 > \alpha_d$, $\theta_f = \pi/2 - \beta_d >$

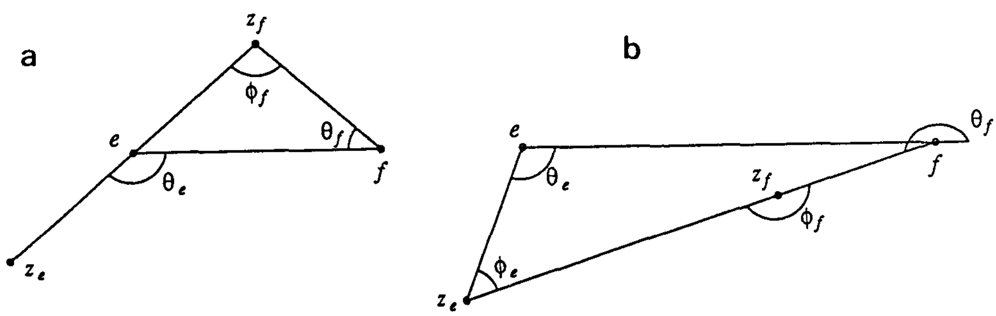


Fig. 14. Angles in extreme situations.

$\pi$, and $\phi_f = \pi$ (fig. 14b). If $\omega = \alpha_d$ then the figures are reflected about the point $(d/2, 0)$.

We define a final constant $\delta_d = \arcsin(d/4)$ for $d \in [0, 4]$; $\delta_d$ is used in the linear interpolations below. If $\omega = \delta_d$ then $\theta_e = \pi/2 - \delta = \arccos(d/4)$, segment $z_e z_f$ bisects segment $ef$, $\theta_f = \theta_e = \pi/2 - \delta_d$, and $h_d(\delta_d) = \delta_d$. Figure 14 is drawn with $\omega = \delta_d$.

LEMMA 4.4.3

If $\omega \in \mathrm{dom}(h_d)$, $\omega \neq -\alpha_d$, $\beta_d$, then $h'_d(\omega) = -\sin(\phi_e(\omega))/\sin(\phi_f(\omega))$, where $h'_d$ is the derivative of $h_d$.

*Proof*

We compute $h'_d$ at the specific point $\omega_0$. Set $\Delta\theta_e(\omega) = \theta_e(\omega) - \theta_e(\omega_0)$ and similarly for $\Delta\theta_f$ and $\Delta\eta$. Choose the coordinates so that $z_e(\omega_0)$ is at (0, 0) and $z_f(\omega_0)$ is at (2, 0). (For the following proof, it is helpful to draw a picture with $z_e(\omega_0)$ and $z_f(\omega_0)$ on the $x$-axis as indicated, with $e$ above the $x$-axis, $f$ below the $x$-axis, and $z_e(\omega)$ slightly to the right of $z_e(\omega_0)$, hence $z_f(\omega)$ slightly to the right of $z_f(\omega_0)$. Label the angles as in fig. 14; note for example that there are both $\theta_e(\omega)$ and $\theta_e(\omega_0)$.)

The segment from $e$ to $z_e(\omega)$ makes angle $\phi_e(\omega_0) - \Delta\theta_e(\omega)$ with the $x$-axis and the $x$ coordinate of $e$ is $\cos(\phi_e(\omega_0))$; hence the $x$-coordinate of $z_e(\omega)$ is

$$\cos(\phi_e(\omega_0)) - \cos(\phi_e(\omega_0) - \Delta\theta_e(\omega)).$$

Similarly the segment from $f$ to $z_f(\omega)$ makes angle $\phi_f(\omega_0) - \Delta\theta_f(\omega)$ with the $x$-axis and the $x$-coordinate of $f$ is $2 - \cos(\phi_f(\omega_0))$; hence the $x$-coordinate of $z_f(\omega)$ is

$$2 - \cos\big(\phi_f(\omega_0)\big) + \cos\big(\phi_f(\omega_0) - \Delta\theta_f(\omega)\big).$$

Since $z_e(\omega)$ and $z_f(\omega)$ are distance 2 apart and segment $z_e(\omega) z_f(\omega)$ makes angle $\Delta\eta(\omega)$ with the $x$-axis,

$$2\cos(\Delta\eta(\omega)) = 2 - \cos\big(\phi_f(\omega_0)\big) + \cos\big(\phi_f(\omega_0) - \Delta\theta_f(\omega)\big)$$
$$- \big[\cos(\phi_e(\omega_0)) - \cos(\phi_e(\omega_0) - \Delta\theta_e(\omega))\big].$$

Differentiate this with respect to $\omega$ (note terms in $\omega_0$ are constant and $\Delta$-terms are functions of $\omega$):

$$-2\sin(\Delta\eta(\omega))\frac{d\eta}{d\omega} = \sin\big(\phi_f(\omega_0) - \Delta\theta_f(\omega)\big)\frac{d\theta_f}{d\omega} + \sin\big(\phi_e(\omega_0) - \Delta\theta_e(\omega)\big)\frac{d\theta_e}{d\omega}.$$

Evaluate this at $\omega = \omega_0$ (so $\Delta$-terms disappear):

$$0 = \sin\big(\phi_f(\omega_0)\big)\frac{d\theta_f}{d\omega}(\omega_0) + \sin\big(\phi_e(\omega_0)\big)\frac{d\theta_e}{d\omega}(\omega_0).$$

Since $d\theta_e/d\omega = -1$ and $d\theta_f/d\omega = -h'_d$, we have $\sin(\phi_e(\omega_0)) = -\sin(\phi_f(\omega_0))h'_d(\omega_0)$. If $\omega_0 \neq -\alpha_d$, $\beta_d$ then $\sin(\phi_f(\omega_0)) \neq 0$ and the lemma is established. □

The values of $h'_d$ at $-\alpha_d$ and $\beta_d$ are not necessary for the development below. However, the proof of lemma 4.4.3 in fact shows that if $d < 2$ then $h'_d(-\alpha_d) = 0$ and $\lim_{\omega \to \beta_d} h'_d(\omega) = -\infty$ and if $d > 2$ then $\lim_{\omega \to -\alpha_d} h'_d(\omega) = -\infty$ and $h'_d(\beta_d) = 0$. If $d = 2$ then using a Taylor series expansion it can be shown that $h'_2(-\alpha_d) = -1/3$ and that $h'_d(\alpha_d) = -3$.

LEMMA 4.4.4

For any interior point of $\mathrm{dom}(h_d)$,

$$h''_d = \frac{-d(\sin\phi_f \cos\phi_e \sin\theta_f + \sin\phi_e \cos\phi_f \sin\theta_e)}{2\sin^3\phi_f}.$$

*Proof*

Notice that if $\omega \neq -\alpha$, $\beta$ then $\sin\phi_e$, $\sin\phi_f \neq 0$. By differentiating the formula for $h_d$ from the previous lemma,

$$h''_d = \frac{\left(\sin\phi_e \cos\phi_f \dfrac{d\phi_f}{d\omega} - \sin\phi_f \cos\phi_e \dfrac{d\phi_e}{d\omega}\right)}{\sin^2\phi_f}.$$

Let $s$ be the distance between $z_e$ and $f$. By the law of cosines $s^2 = d^2 + 1 - 2d\cos\theta_e$ and also $s^2 = 5 - 4\cos\phi_f$. Therefore $5 - 4\cos\phi_f = d^2 + 1 - 2d\cos\theta_e$; differentiating and using $d\theta_e/d\omega = -1$ yields

$$\frac{d\phi_f}{d\omega} = -\frac{d}{2}\frac{\sin\theta_e}{\sin\phi_f}.$$

By a similar argument using lemma 4.4.3 and $d\theta_f/d\omega = -h'_d$,

$$\frac{d\phi_e}{d\omega} = -\frac{d}{2}\frac{\sin\theta_f}{\sin\phi_e}\frac{d\theta_f}{d\omega} = \frac{d}{2}\frac{\sin\theta_f}{\sin\phi_e}\frac{\sin\phi_e}{\sin\phi_f} = \frac{d}{2}\frac{\sin\theta_f}{\sin\phi_f}.$$

The substitutions yield the lemma. □

LEMMA 4.4.5

If $0 < d \leqslant 2$ then $h_d$ is convex downwards.

*Proof*

We show $h''_d \leqslant 0$ for $\omega$ in the range $(-\alpha_d, \delta_d]$. Since $h_d$ is symmetric about the line $\theta = \omega$ (by proposition 4.4.2), it follows that $h_d$ is convex over $\mathrm{dom}(h_d) = [-\alpha_d, \alpha_d]$.

If $\omega \in (-\alpha_d, \delta_d]$ then $\theta_e \in [\pi/2 - \delta_d, \pi/2 + \alpha_d] \subseteq [0, \pi]$, $\theta_f \in [\pi/2 - \alpha_d, \pi/2 - \delta_d] \subseteq [0, \pi/2]$, $\phi_e \in (0, 2\delta] = (0, 2 \arcsin(d/4)] \subseteq (0, \pi/3]$, and $\phi_f \in [2\delta, 2\alpha) \subseteq (0, \pi)$.

We show $(*) = \sin \phi_f \cos \phi_e \sin \theta_f + \sin \phi_e \cos \phi_f \sin \theta_e \geqslant 0$, from which the lemma follows, since $\sin^3\phi_f > 0$. First note that if $\phi_f \leqslant \pi/2$, then every term in $(*)$ is nonnegative.

So suppose $\phi_f > \pi/2$, then $\cos \phi_f < 0$. Notice $\theta_e \geqslant \phi_f$ since triangle $z_f z_e f$ has sides of length 1, 2, and $|z_e f|$, and triangle $e z_e f$ has sides of length 1, $d \leqslant 2$, and $|z_e f|$. Since $\phi_f > \pi/2$, $\sin \theta_e \leqslant \sin \phi_f$. Similarly $\theta_f \geqslant \phi_e$ and $\sin \theta_f \geqslant \sin \phi_e$, since $\theta_f, \phi_e \in [0, \pi/2]$. Since $\cos \phi_f < 0$, we have $(*) \geqslant \sin \theta_e \sin \theta_f (\cos \phi_e + \cos \phi_f)$.

Now surely $|ef| = d \geqslant 2 - (\cos \phi_e + \cos \phi_f)$, since $2 - (\cos \phi_e + \cos \phi_f)$ is the distance between the projections of $e$ and $f$ on the line through $z_e$ and $z_f$. Hence $\cos \phi_e + \cos \phi_f \geqslant 2 - d \geqslant 0$. Since $\sin \theta_e, \sin \theta_f \geqslant 0$, the lemma is established. □

If $d > 2$ then $h_d$ is not convex downwards. This can be seen intuitively by considering the intersection $\iota$ between line $z_e z_f$ and line $ef$, as $\theta_e$ decreases from $\pi/2 + \alpha$. It is easy to see that $\sin \phi_e / \sin \phi_f = |e\iota| / |\iota f|$, hence $h_d'$ can be inferred from the position of $\iota$. First suppose $d$ is slightly bigger than 2. Then $\iota$ is initially at $f$, since $z_e$, $z_f$, and $f$ are collinear. As $\theta_e$ decreases $\iota$ moves to the left, eventually reversing direction at a value of $\theta_f < \pi$, then moving right to the midpoint of $ef$ when $\theta_e = \pi/2 - \delta$. As $\theta_e$ decreases further $\iota$ continues to the right, then reverses direction and ends up at $e$. Hence $h_d'$ is initially $-\infty$ (using the observation following lemma 4.4.3), increases rapidly, decreases, then increases to 0. If $d$ is sufficiently large, $\iota$ moves monotonically from $f$ to $e$, and $h_d$ is convex upwards. Fortunately, we can use proposition 4.4.2(3) to make analysis of $h_d$, $d > 2$, unnecessary.

For $d \leqslant 2$, let $l_d$: $[-\alpha_d, \alpha_d] \to [-\alpha_d, \alpha_d]$ be the two-piece linear interpolant through the points $(-\alpha_d, \alpha_d)$, $(\delta_d, \delta_d)$, and $(\alpha_d, -\alpha_d)$. Similarly let $\bar{l}_d$ be the two-piece linear interpolant through the points $(-\alpha_d, \alpha_d)$, $(-\delta_d, -\delta_d)$, and $(\alpha_d, -\alpha_d)$ (fig. 15).

COROLLARY 4.4.6

If $d \leqslant 2$, $h_d \geqslant l_d$ and $\bar{h}_d \leqslant \bar{l}_d$. If $d > 2$ then $h_d \geqslant h_2 \geqslant l_2$ and $\bar{h}_d \leqslant \bar{h}_2 \leqslant \bar{l}_2$.

*Proof*

Follows from the convexity of $h_d$ and proposition 4.4.3. □

PROPOSITION 4.4.7

If $d \leqslant 2$ then $(\delta_d - \alpha_d)/(\delta_d + \alpha_d) \geqslant -1/2$.

*Proof*

Notice $3\delta_d = 3 \arcsin(d/4) \geqslant \arcsin(d/2) = \alpha_d$, since if $d = 2$, $3 \arcsin(d/4) = 3\pi/6 = \arcsin(d/2) = \pi/2$, and the slope of arcsin is monotonically increasing

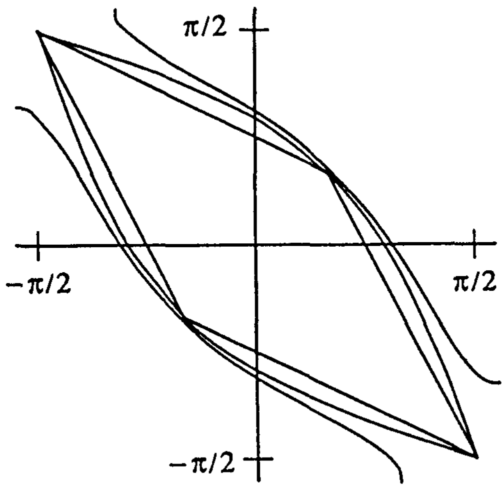


Fig. 15. The curves in order from bottom to top are $\bar{h}_{2,1}$, $\bar{h}_2$, $\bar{l}_2$, $l_2$, $h_2$, $h_{2,1}$.

over $[0, d/2] = [0, 1]$. Hence $2\delta_d - 2\alpha_d \geqslant -(\delta_d + \alpha_d)$, and the proposition follows. □

Thus the linear interpolants $l_d$ and $\bar{l}_d$ have slopes at least $-1/2$ over the intervals $[-\alpha_d, \delta_d]$ and $[-\delta_d, \alpha_d]$, respectively, and slopes at most $-2$ over the intervals $[\delta_d, \alpha_d]$ and $[-\delta_d, \alpha_d]$, respectively. Also notice that if $d = 2$, then $\alpha_d = \pi/2$ and $\delta_d = \pi/6$, so $h_2$ interpolates through $(-\pi/2, \pi/2)$, $(\pi/6, \pi/6)$ and $(\pi/2, -\pi/2)$.

*Proof of theorem 4.4.1*

Without loss of generality assume that $\omega < 0$. Set $\Delta = \omega - (-\alpha_d)$. Clearly $h_d(\omega) > 0$. If $\bar{h}_d(\omega) \leqslant 0$, then $0 \in H_d(\omega)$ and the theorem is established. Otherwise $\bar{h}_d(\omega) > 0$; since $h_d(-\delta_d) = -\delta_d$ and $\bar{h}_d$ is decreasing, $\omega < -\delta_d$.

First suppose that $d \leqslant 2$. Set $\gamma_d = \alpha_d$. Then clearly $\omega \in \text{dom}(H_d)$ and $H_d(\omega) \subseteq [-\alpha, \alpha]$. By corollary 4.4.6, $h_d(\omega) \geqslant l_d(\omega)$ and by proposition 4.4.7, $l_d$ has slope at least $-1/2$ in the interval $[-\alpha_d, -\delta_d] \subseteq [-\alpha_d, \delta_d]$. Hence $h_d(\omega) \geqslant \alpha_d - \Delta/2$. Similarly $\bar{h}_d(\omega) \leqslant \bar{l}_d(\omega)$ and $\bar{l}_d$ has slope at most $-2$ in the interval $[-\alpha_d, -\delta_d]$, so $\bar{h}_d(\omega) \leqslant \alpha_d - 2\Delta$.

Suppose $d > 2$. We use the interpolants $l_2$ and $\bar{l}_2$ to bound $h_d$ and $\bar{h}_d$. Notice $\text{dom}(l_2) = \text{dom}(\bar{l}_2) = [-\pi/2, \pi/2]$ and $\omega \in \text{dom}(H_2) = [-\alpha_d, \alpha] \subseteq [-\pi/2, \pi/2]$, hence $\omega \in \text{dom}(l_2) \cap \text{dom}(\bar{l}_2)$. We set $\gamma = \pi/2$; the rest of the proof is similar to the case $d \leqslant 2$. □

We finish with a quantitative version of proposition 4.4.2(3). This lemma is used in the next section. Let $\hat{f}(x) = 2 \arcsin((x/4)^2)$, then $\hat{f}(x)$ is $\Omega(x^2)$ for small $x$.

LEMMA 4.4.8

If $d' > d$ and both $h_d$, $h_{d'}$ are defined at $\omega$, then $h_{d'}(\omega) > h_d(\omega) + \hat{f}(d' - d)$.

*Proof*

Set $e$ at (0, 0), $f$ at $(d, 0)$, $f'$ at $(d', 0)$, and let $(E, F)$ and $(E, F')$ be short $RL$-hops starting at $e$ with source angle $\omega$ and ending at $f$ and $f'$, respectively. Let $z_e$, $z_f$, and $z_{f'}$ be the centers of circles $C_E$, $C_F$, and $C_{F'}$, respectively. It is easy to check that the line from $z_e$ to $z_f$ has positive slope or is vertical. Let $y$ be the point with the $x$-coordinate increased by $d' - d$ from $z_f$; then $z_f f$ and $yf'$ are parallel. Then $\pi >$ (angle $z_e z_f y$) $\geqslant \pi/2$ and we have $|z_e y| \geqslant \sqrt{4 + (d' - d)^2} > 2 + (d' - d)^2/8$, the last inequality using $d' - d < 4$. Since $|z_e z_{f'}| = 2$, $|z_{f'} y| > (d' - d)^2/8$. Since $|z_{f'} f'| = |yf'| = 1$ and $h_{d'}(\omega) - h_d(\omega)$ is angle $yf'z_{f'}$, $h_{d'}(\omega) - h_d(\omega) > 2 \arcsin(d' - d)^2/16$. □

4.5. HETERODUAL PATH FAMILIES

THEOREM 4.5.1

Suppose $j$ jumps from a corner $a$ to corner $b$ and $P_j$ is heterodual. Then $S_j$ contains an interval self-dual at $b$ from some corner (possibly distinct from $a$) and either $S_j$ contains a root or the size of $S_j$ is at least $\kappa_b$, where $\kappa_b$ is a constant of size at least $2^{-\Omega(m)}$.

LEMMA 4.5.2

Suppose interval $I$ at $b$ contains both an interval self-dual from a corner and a placement of constant algebraic complexity. Then either $I$ contains a root or $I$ has size at least $2^{-\Omega(m)}$.

*Proof*

Let $I = [\theta_r, \theta_l]$ be an interval at $b$ containing an interval self-dual from corner $a$. Let $d$ be the distance between $a$ and $b$ and $\mu$ be the straightline direction from $a$ to $b$. If $I$ contains one of $\mu$, $\mu - \alpha_d$, or $\mu + \alpha_d$, then $I$ contains a root. Otherwise we can assume $\mu - \alpha_d < \theta_r \leqslant \theta \leqslant \theta_l < \mu$, where $\theta$ is of constant algebraic complexity. By lemma 2.3.4, $\theta$ and $\mu - \alpha_d$ are separated by at least $2^{-\Omega(m)}$. Hence either $\theta_r$ and $\theta$ are separated by at least $2^{-\Omega(m)}$, and $I$ has size $2^{-\Omega(m)}$, or $\mu - \alpha_d$ and $\theta_r$ are separated by $2^{-\Omega(m)}$, and the size of $I$ is at least $2^{-\Omega(m)}$, using theorem 4.4.1. □

We now discuss the geometry of heterodual path families. For the rest of this section, let $j$ be a jump from corner $a$ to corner $b$, $j = (\sigma, \tau)$, $P_j$ be heterodual, and $S_j = [\theta_r, \theta_l]$. Choose $c$ as the last common point of $P_j(\theta_r)$ and $P_j(\theta_l)$ not including $b$; then $c$ is either $a$ if (SFLIP) occurred or a point on $C_\sigma$ if not. Let $\omega$ be the direction of the tangent to $P_j(\theta_r)$ (and $P_j(\theta_l)$) at $c$ and $l$ the line in direction $\omega$ through $c$. (Notice that for some $\theta \in S_j$, $P_j(\theta)$ is a jump ending at $b$

with tangent along $l$.) It is clear that $P_j(\theta_r)$ past $c$ consists of straight segments alternating with $L$-arcs ending with an $R$-arc to $b$. By lemma 4.2.2, the total arc length of the $L$-arcs is less than $\pi$, so all of $P_j(\theta_r)$ past $c$ except possibly for part of the $R$-arc lies to the left of $l$. Since the $R$-arc of $P_j(\theta_r)$ must have length less than $\pi$, $P_j(\theta_r)$ after $c$ ends with an $LsRs$ hop. Let $a_r$ be the beginning point of this hop; clearly $a_r$ is on or to the left of $l$. Also, either $a_r$ is a corner or $a_r = c$; the latter happens only if all of $P_j(\theta_r)$ is an $LxRs$ hop. By similar reasoning $P_j(\theta_l)$ consists of straight segments alternating with $R$-arcs ending with an $L$-arc. The final part of $P_j(\theta_l)$ is an $RsLs$ hop beginning at $a_l$; $a_l$ lies on or to the right of $l$ and $a_l \neq a_r$.

We let $\mu_l$ be the straightline direction from $a_l$ to $b$, $d_l$ be the distance between $a_l$ and $b$, $\phi_l$ the direction of the tangent to $P_j(\theta_l)$ at $a_l$, and $h_l = h_{d_l}$, and similarly for $\mu_r$, $d_r$, $\phi_r$, and $h_r$. Notice $\phi_l$ and $\phi_r$ are separated by less than $\pi$ by lemma 4.2.2, and clearly $\phi_l \leqslant \omega \leqslant \phi_r$.

LEMMA 4.5.3

Directions $\mu_l$ and $\mu_r$ are separated by at most $\pi/2$.

*Proof*

Let $C$ be a unit circle through $b$ so that for every $\theta \in S_j$ the ray from $b$ in direction $\theta$ intersects $C$ only at $b$; such $C$ exists since $|S_j| < \pi$. Let $R$ be the region swept out by translating $C$ in direction $-\omega$; then $R$ has as boundary two rays in direction $-\omega$ and a semicircle of $C$. Clearly $P_j(\theta_l)$ and $P_j(\theta_r) \subseteq R$. Let $R' \subseteq R$ be the region bounded by the two rays and the two segments from $u_\tau$ to the two points of tangency between the rays and $C$. Any point of $P_j(\theta_l)$ outside $R'$ necessarily has tangent to the left of $\omega$; since $\phi_l \leqslant \omega$, $a_l \in R'$. Similarly $a_r \in R'$. The angle formed by the two segments incident to $u_\tau$ has size $\pi/2$ since it subtends a diameter of $C$. The lemma follows. □

LEMMA 4.5.3

Distance to $b$ is monotonically decreasing along $P_j(\theta_r)$ and $P_j(\theta_l)$ past $c$.

*Proof*

For $x \in P_j(\theta_l)$ past $c$ let $\phi_x$ be the direction of the tangent to $P_j(\theta_l)$ at $x$ and $\mu_x$ be the direction from $x$ to $b$. We claim that the angle between $\phi_x$ and $\mu_x$ is strictly less than $\pi/2$; this implies the lemma. The claim is clearly true at $x = c$. It is also clearly true for any point on the $R$-arc of $P_j(\theta_l)$, since the $R$-arc has length less than $\pi$. In particular it is true for $f$, the first point of the $R$-arc. Now for any $x \in P_j(\theta_l)$ between $c$ and $f$, $\phi_c \leqslant \phi_x \leqslant \phi_f$ and $\mu_f \leqslant \mu_x \leqslant \mu_c$, hence the angle between $\phi_x$ and $\mu_x$ is less than $\pi/2$ as well. □

We henceforth assume that $d_r \leqslant d_l$. Notice that this assumption together with the previous lemma implies that $a_r$ is a corner; for if $a_r$ is not a corner, then $a_r = c$, $a_l$ appears on $P_j(\theta_l)$ past $c$, so $d_l < d_r$, a contradiction.

LEMMA 4.5.4

If $x \geqslant y$, $0 \leqslant s \leqslant t \leqslant 4$, $x \in \mathrm{dom}(h_s)$, $y \in \mathrm{dom}(h_t)$, then $h_t(y) \geqslant h_s(x)$.

*Proof*

We have $x \in \mathrm{dom}(h_s) = [-\alpha_s, \beta_s]$, $y \in \mathrm{dom}(h_t) = [-\alpha_t, \beta_t]$. We claim either $x \in \mathrm{dom}(h_t)$ or $y \in \mathrm{dom}(h_s)$. If $s \geqslant 2$ then $\beta_s \geqslant x \geqslant y \geqslant -\alpha_t \geqslant -\alpha_s$, the last inequality because $-\alpha_s$ increases in [2,4], so $y \in \mathrm{dom}(h_s)$; if $s < 2$ then $\beta_t \geqslant \beta_s \geqslant x \geqslant y \geqslant -\alpha_t$, the first inequality because $\beta_t$ increases up to $t = 2$ and then does not go below $\beta_2$, so $x \in \mathrm{dom}(h_t)$. In the first case $h_t(y) \geqslant h_s(y) \geqslant h_s(x)$ by lemma 4.4.2(3) and $h_s$ decreasing; in the second case similarly $h_t(y) \geqslant h_t(x) \geqslant h_s(x)$. □

*Proof of theorem 4.5.1*

We show that there is an *RsLs* hop from $a_r$ to $b$ with source direction $\phi_r$ and that $\theta_l$ is actually to the left of its final direction; explicitly we show $\phi_r - \mu_r \in \mathrm{dom}(h_r)$ and $\theta_l \geqslant h_r(\phi_r - \mu_r) + \mu_r$. (Note we do not explicitly show that the *RsLs* hop avoids obstacles.) We then show that either $S_j$ contains a root or $S_j$ has size at least $\kappa_b$ (defined below).

We first claim that without loss of generality we can assume $\mu_l \geqslant \mu_r$. Suppose $\mu_l < \mu_r$. Now $a_r$ must lie to the left of $l$. Since $\mu_l < \mu_r$, $b$ also lies to the left of $l$; furthermore, the line through $a_r$ and $b$ must intersect the *RsLs* hop from $a_l$ to $b$ at some point $i$, with $i$, $a_l$ and $b$ in that order. Let $\phi_i$ be the direction of the tangent to the *RsLs* hop at $i$ and let $d_i$ be the distance between $i$ and $b$. Notice that if $i$ lies on the $R$-arc of the hop, then $\phi_i < \phi_l \leqslant \phi_r$, and if $i$ lies on the $L$-arc, then $\phi_i \leqslant u_r - \alpha_{d_i} < u_r - \alpha_r \leqslant \phi_r$, the middle inequality following because $d_r \leqslant d_i \leqslant 2$. Hence we can simply replace $a_l$ with $i$ and use the hop starting at $i$.

Since $\phi_l \leqslant \phi_r$ and $\mu_l \geqslant \mu_r$, we have $\phi_r - \mu_r \geqslant \phi_l - \mu_l$. We claim $\phi_r - \mu_r \in \mathrm{dom}(h_r)$: if $d_r \leqslant 2$ then $\phi_r - \mu_r \in \mathrm{dom}(\bar{h}_r) = \mathrm{dom}(h_r)$ and if $d_r > 2$ then $\phi_r - \mu_r \geqslant \phi_l - \mu_l \geqslant -\alpha_l \geqslant -\alpha_r$ and $\phi_r - \mu_r \leqslant \alpha_r \leqslant \beta_r$. Since $d_l \geqslant d_r$ and $\phi_r - \mu_r \geqslant \phi_l - \mu_l$, by lemma 4.5.4, $h_l(\phi_l - \mu_l) \geqslant h_r(\phi_r - \mu_r)$. Hence $\theta_l = \mu_l + h_l(\phi_l - \mu_l) \geqslant \mu_r + h_r(\phi_r - \mu_r)$.

It remains to show $S_j$ has size $\kappa_b$ or that $S_j$ contains a root. First suppose that $a_l$ and $a_r$ are separated by at least $\kappa_c$ (the minimum distance between corners); necessarily this is the case if $a_l$ and $a_r$ are corners. We show that $S_j$ contains a root or that the inequality $\theta_l \geqslant \mu_r + h_r(\phi_r - \mu_r)$ is strict by an amount dependent upon $\kappa_c$. If in fact $\mu_l \geqslant \mu_r$ then either $\mu_l > \mu_r + \kappa_c/8$ or $d_l \geqslant d_r + \kappa_c/2$. The inequality is strict in the first case by the amount $\kappa_c/8$ and in the second by $\Omega(\kappa_c^2)$, using lemma 4.4.8. If $\mu_l < \mu_r$ then $i$ lies on the $L$-arc or the $R$-arc of the $RL$ hop ending $P_j(\theta_l)$. If $i$ lies on the $R$-arc then $S_j$ contains the root $\mu_r + \alpha_r$. Suppose $i$ lies on the $L$-arc. Let $l'$ be the line parallel to $l$ through $a_l$ and let $i'$ be the point of intersection of $l'$ with the line through $a_r$ and $b$. Then either $a_r$ and $i'$ or $a_l$ and $i$ are separated by at least $\kappa_c/2$. In the second case $i$ and $i'$

must be separated by $\Omega(\kappa_c^2)$; hence in either case $i$ and $a_r$ are separated by at least $\Omega(\kappa_c^2)$, so the inequality is strict by at least $\Omega(\kappa_c^4)$, using lemma 4.4.8 again.

Now suppose that $a_l$ and $a_r$ are closer than $\kappa_c$. We show that $S_j$ contains a placement of constant algebraic degree; the bound on the size of $S_j$ then follows from lemma 4.5.2. Since $a_l$ and $a_r$ are closer than $\kappa_c$, $a_l$ is not a corner and $a_l = c$; furthermore, $a_l$ lies on the circle $C$ through $a$ (the source corner of jump $j$). Circle $C$ is the source circle of both jump $j$ and hop $P_j(\theta_l)$. Notice $a_r$ lies on $l$, since there can be no corner along $P_j(\theta_r)$ from $c$ to $a_r$. Now let $l'$ be the line from $a_r$ to $C$ tangent to $C$ at a point $a_l'$; let $\omega'$ be its direction. Notice $a$ cannot lie on the arc of $C$ clockwise from $a_l$ to $a_l'$ since this arc is within $\kappa_c$ of $a_r$. Now imagine rotating $C$ about $a$ so that the direction of the line tangent from $C$ to $a_r$ (initially this line is $l$) moves right. If the arc from $a$ to $a_l$ is less than $\pi$, this is a counterclockwise rotation; if the arc is greater than $\pi$ it is a clockwise rotation. The rotation stops at circle $\hat{C}$ chosen so that the line $\hat{l}$ from $\hat{C}$ to $a_r$ has direction $\hat{\omega}$ as far right as possible; then either $\hat{C}$ touches $a_r$ or the arc along $\hat{C}$ from $a$ to the point of tangency with $\hat{l}$ has length $\pi$. Clearly $\hat{\omega}$ is of constant algebraic complexity. It can be checked that $\hat{\omega} \in [\omega', \omega]$. Now if there is no *LsRs* hop from $a_r$ to $b$ with final direction $\theta_l$, then $\theta_l$ is to the left of $\omega_r + \alpha_r$ and $S_j$ contains the root $\omega_r + \alpha_r$. Otherwise there is such a hop; it can be checked that its source direction must lie to the right of $\omega'$. Since $\hat{\omega} \geqslant \omega'$, $\bar{h}_r(\hat{\omega}) \geqslant \theta_l$ and $\bar{h}_r(\hat{\omega}) \in S_j$. The direction $\bar{h}_r(\hat{\omega})$ is the required direction of constant algebraic complexity.

Clearly we can choose $\kappa_b$ less than all the indicated bounds on the size of $S_j$ while still maintaining $\kappa_b$ of size $2^{-\Omega(m)}$. □

## 5. A simple reachability algorithm

We now give a bound on the number of disjoint reachable intervals. This bound is given in theorem 5.1.3 below. The proof of the bound uses the categorization of intervals developed in section 4 and in particular the analysis of self-dual intervals. As an application of the bound, we give an easy double-exponential time decision procedure for the reachability question. An improved decision procedure is given in the next section.

We begin by defining a way of propagating reachable intervals from one contact to another. Suppose $I$ is a closed interval of placements at $a$ and $j \in \Xi_{ab}^t(I)$. Define $T(j, I)$ (an interval of placements at $b$) as follows. If either $a$ or $b$ is a wall, then $T(j, I)$ is just $J(j, I)$. If both $a$ and $b$ are corners, then $T(j,I)$ is $J(j, I)$ merged with the following intervals. Choose $j_R$, $j_L \in X(j, I)$ with source directions as far right and left as possible. Merge $S_{j_R}$ with $T(j, I)$, and if $P_{j_R}$ is not saturated, self-dual, or heterodual then by theorem 4.3.1, there is $j_R'$ with the same source direction that is; merge $S_{j_R'}$ with $T(j, I)$ as well. Repeat this procedure with $j_L$.

It is clear that $T(j, I)$ is a closed interval and $J(j, I) \subseteq T(j, I)$. We claim $T(j, I)$ always contains a root or an interval self-dual from a corner. The claim is trivial if $b$ is a wall and follows from lemma 3.2.4 if $a$ is a wall. If $a$ and $b$ are both corners, then $T(j, I)$ contains $P_{j'}$ for some $j'$ which is saturated (hence $S_{j'}$ contains a root), or is self-dual, or is heterodual (hence $S_{j'}$ contains a self-dual interval by theorem 4.5.1).

In exactly the same way that $J$ (mapping a jump and an interval into an interval) was extended to $\mathscr{J}$ (mapping sets of intervals into sets of intervals) in section 3.3, we extend $T$ to $\mathscr{T}$. Since any path determined by *stretch* consists of at most $n$ jumps, it is clear that $\mathscr{J} \sqsubset \mathscr{T} \sqsubset \mathscr{J}^{(n)}$. Clearly $\mathscr{R} = \sqcup_{k=0}^{\infty} \mathscr{T}^{(k)}(\mathscr{I}_0)$.

Set $\kappa_s$ to be the minimum of $\kappa_b$ and the difference between distinct roots at the same contact. Then $\kappa_s$ is at least $2^{\Omega(-m)}$, using lemma 2.3.4.

LEMMA 5.1.1

Suppose $I \in \mathscr{T}^{(k)}(\mathscr{I}_0)$, for some $k$. Then either $I$ contains a root or $|I| \geqslant \kappa_s$.

*Proof*

By induction on $k$. If $k = 0$, then $I$ is the interval containing exactly the source placement, a root. If $I \in \mathscr{T}^{(k)}(\mathscr{I}_0)$, $k > 0$, then there is some $I_k \subseteq I$ so that $I_k = T(j, I_{k-1})$ for some $I_{k-1} \in \mathscr{T}^{(k-1)}(\mathscr{I}_0)$. If $I_k$ contains a root, then the lemma is established. If not, then $j$ must be a jump from a corner $a$ to a corner $b$, and there are $j'$, $j''$ so that $S_{j'}, S_{j''} \subseteq I_k$ and $P_{j'}$, $P_{j''}$ are heterodual or self-dual. If either are heterodual then the lemma follows from theorem 4.5.1. Otherwise both are self-dual and $S_{j'} = H_d(\Omega(j') - \mu) + \mu$, $S_{j''} = H_d(\Omega(j'') - \mu) + \mu$, where $d$ is the distance between $a$ and $b$ and $\mu$ is the straightline direction from $a$ to $b$. By lemma 3.2.4, $\Omega(j')$, $\Omega(j'')$ are the endpoints of $I_{k-1}$, since $I_k$ does not contain a root. Now $H_d(\omega)$ forms an interval whose endpoints are monotonic functions of $\omega$; hence $H_d(\theta - \mu) + \mu \subseteq I_k$ for any $\theta \in I_{k-1}$. We claim there is some $\theta \in I_{k-1}$ so that $\theta - \mu$ is at least distance $\kappa_s$ from both endpoints of $\text{dom}(H_d) = [-\alpha_d, \alpha_d]$. Notice $I_{k-1}$ contains none of $\mu - \alpha_d$, $\mu$, or $\mu + \alpha_d$, else $I_k$ contains a root. By the induction hypothesis $|I_{k-1}| \geqslant \kappa_s$ or $I_{k-1}$ contains a root $r$; the claim is immediate in the first case and follows from the definition of $\kappa_s$ in the second (since the root $r$ must be at least $\kappa_s$ from both the roots $\mu - \alpha_d$ and $\mu + \alpha_d$). By theorem 4.4.1, $|H_d(\theta - \mu)| \geqslant 3\kappa_s/2 > \kappa_s$, hence $|I_k| \geqslant \kappa_s$. □

A *self-dual marker at corner b from corner a* is one of the following placements at corner $b$. Suppose corners $a$ and $b$ are distance $d$ apart, the straightline direction from $a$ to $b$ is $\mu$, and $H_d$ has nonempty domain. Choose $\gamma_d$ as given by theorem 4.4.1. Then the markers are $\mu + \gamma_d$, $\mu - \gamma_d$, $\mu + \gamma_d - 2^i\kappa_s$, ($i = 0, 1, \ldots,$ so long as $\gamma_d - 2^i\kappa_s > 0$), and $\mu - \gamma_d + 2^i\kappa_s$, ($i = 0, 1, \ldots,$ so long as $-\gamma_d + 2^i\kappa_s < 0$). Since $\kappa_s$ is $2^{\Omega(-m)}$, there are only $O(m)$ self-dual markers at $a$ from $b$.

A *marker* is a root or a self-dual marker at some corner from some corner. Clearly there are $O(n^2(n^2 + m))$ markers.

LEMMA 5.1.2

Every interval in $\mathscr{R}$ contains a marker.

*Proof*

Let $I' \in \mathscr{R}$. Then there is some $I \in \mathscr{T}^{(k)}(\mathscr{I}_0)$, for some $k$, so that $I \subseteq I'$. If $I$ contains a root, then $I$ contains a marker. If $I$ does not contain a root, then $I$ is an interval at a corner $b$ containing a self-dual interval from a corner $a$ and $I$ has size at least $\kappa_s$. Let $\mu$ be the straightline direction from $a$ to $b$, $d$ the distance between $a$ and $b$, and choose $\gamma_d$ as in theorem 4.4.1. Then $\mu + H_d(\omega) \subseteq I$, for some $\omega$. Assume $\omega \leqslant 0$, the case $\omega \geqslant 0$ is similar. Set $\Delta = \alpha_d + \omega$, then by theorem 4.4.1, $-\gamma_d + [\Delta/2, 2\Delta] \subseteq H_d(\omega)$. Since $I$ has size at least $\kappa_s$, we either have $I$ containing marker $\mu - \gamma_d$, $I$ containing marker $\mu - \gamma_d + \kappa_s$, or $\kappa_s < \Delta/2$. We also have $2\Delta < \gamma_d$ since by assumption $I$ does not contain the root $\mu$. Hence we can choose natural number $i$ so that $\Delta/2 < 2^i\kappa_s < 2\Delta < \gamma_d$, and marker $\mu - \gamma_d + 2^i\kappa_s$ is in $I$. □

THEOREM 5.1.3

$\mathscr{R}$ contains at most $O(n(n^2 + m))$ intervals at each corner (and two per wall), for a total of $O(n^2(n^2 + m))$.

The theorem follows immediately from the previous lemma. We remark that the same proof shows that each $\mathscr{T}^{(k)}(\mathscr{I}_0)$ contains a marker. As an application, we give the following easy decision procedure for the reachability question.

THEOREM 5.1.4

The reachability question is decidable in time $2^{2^{poly(n,m)}}$, where $poly(n, m)$ is a polynomial in $n$ and $m$.

*Proof*

We show that the reachability question can be expressed as a formula of $\mathbb{R}(+, \times)$ with polynomial length (and hence polynomially many variables). The resource bounds on the decision procedure then follow from theorem 2.3.1. We have already discussed how placements and jumps can be expressed in $\mathbb{R}(+, \times)$; we now discuss more complex objects.

An interval of placements at a wall or corner contact can be represented as a pair of placements. A syntactic annotation indicates whether the endpoints are open or closed. Side formulas of size $O(m)$ constrain the pair of placements to lie within the contact interval. Formulas of constant size express membership of a placement in the interval and containment or equality of two intervals.

A set $\mathscr{I}$ of up to $k$ intervals, either all at one contact or at different contacts, can be represented with $4k$ "variables" ranging over intervals (hence $O(k)$ variables of $\mathbb{R}(+, \times)$). There are $k$ each of the four possible types of interval, open, closed, or halfopen; $3k$ of the intervals will always be empty. The predicate

$I \in \mathscr{I}$ is a disjunction (of size $O(k)$). Predicates expressing $\mathscr{I} = \mathscr{I}'$, $\mathscr{I} \sqsubset \mathscr{I}'$ and $\mathscr{I} = \mathscr{I}' \sqcup \mathscr{I}''$ require size proportional to the sum of the sizes of $\mathscr{I}$ and $\mathscr{I}'$ (and $\mathscr{I}''$).

The relation $\mathscr{I} = \mathscr{J}^t_{ab}(\mathscr{I})$ can be expressed as "$\forall \omega$ [($\exists I' \in \mathscr{I}$ and $\omega \in I'$) iff $\exists j$ a feasible jump of type $t$ from $a$ to $b$ with $\theta(j) \in I$ and $\omega = \Omega(j)$]". This formula is at most size $O(n^3 m)$. The relation $\mathscr{I} = \mathscr{J}_{ab}(\mathscr{I})$, requires sixteen copies of this formula, one for each type $t$, and the merge operator. The relation $\mathscr{I}' = \mathscr{J}(\mathscr{I})$ can be expressed as the merge at each contact $b$ of $\mathscr{J}_{ab}(\mathscr{I})$ for all $I$ at contact $a$ in $\mathscr{I}$.

$\mathscr{R}$ is the least fixed point of mapping $\mathscr{J}$: "$\mathscr{I}_0 \sqsubset \mathscr{R}$ and $\mathscr{J}(\mathscr{R}) = \mathscr{R}$ and $\forall \mathscr{I}$ ($\mathscr{I}_0 \sqsubset \mathscr{I}$ and $\mathscr{J}(\mathscr{I}) = \mathscr{I}$ implies $\mathscr{R} \sqsubset \mathscr{I}$)". By theorem 5.1.1, $\mathscr{R}$ contains at most $O(n^2(n^2 + m))$ intervals, hence needs only $O(n^2(n^2 + m))$ variables of $\mathbb{R}(+, \times)$ to represent it. Also, $\mathscr{R}$ is the intersection of all fixed points of $\mathscr{J}$ that have no more intervals than $\mathscr{R}$ does, so it suffices to use $O(n^2(n^2 + m))$ variables for $\mathscr{I}$ as well.

Finally the reachability question can be expressed as "$\exists j$ so that $\Theta(j) \in \mathscr{R}$ and $\Omega(j)$ is the target placement". □

The crucial step in theorem 5.1.4 is the fact that only $O(n^2(n^2 + m))$ disjoint intervals are reachable; this implies that reachability can be expressed using only a finite number of variables of $\mathbb{R}(+, \times)$. Note that though mapping $T$ was used in the proof of theorem 5.1.3, it is not necessary for the formula constructed in theorem 5.1.4. It will be used again in the next section. The construction in the proof makes no attempts to be economical in use of variables or formula length. The only way to make substantial improvements in the complexity of the decision procedure is to reduce the number of variables to $O(\log nm)$. This is the subject of the next section.

## 6. An improved decision procedure

### 6.1. LABELLING AND FIXED POINTS

This section and the next, section 6.2, together give a decision procedure for reachability running in single-exponential time. This section develops a more sophisticated way of propagating reachable intervals, extending the mapping $T$ of section 5. The main result of this section, theorem 6.1.7, is that the extended mapping finds all reachable intervals with only a singly exponential number of applications. Section 6.2 discusses the details of how to actually implement the mapping. The extended mapping requires a labelling strategy and a method to solve for fixed points of cycles. We begin with a technical point, extending mapping $T$ to intervals with open endpoints.

Mapping $T$ has been defined only for closed intervals. Now suppose $I$ is an interval with one or two open endpoints and let $j \in \Xi^t_{ab}(I)$. If $a$ or $b$ is a wall, then $T(j, I) = J(j, I)$, which is well-defined for open intervals $I$. So suppose $a$ and $b$ are corners. The propagation mapping depends upon the endpoints of $\Omega(X(j, I))$, handled separately for left and right endpoints. If the right endpoint of $\Omega(X(j, I))$ is closed, proceed as before. If $\Omega(X(j, I))$ has open right endpoint $\omega_R$, then $\omega_R$ must also be the open right endpoint of $I$. We choose $j_R \in X(j, I \cup \{\omega_R\})$ with $\Omega(j_R) = \omega_R$. By lemma 4.3.3, all of $S_{j_R}$ except possibly its endpoints are reachable from $I$. Hence we merge $S_{j_R}$ except possibly its endpoints into $T(j, I)$. (Notice we can explicitly test whether the endpoints of $S_{j_R}$ are reachable from $I$ by jumps; if so we can merge them into $T(j, I)$.). This is repeated as necessary for $j_{R'}$ and for the left endpoint of $\Omega(X(j, I))$.

A *leap sequence* $s = (a_1, t_1, a_2, \ldots, a_k)$ is a finite alternating sequence of contacts and leap types, starting and ending at a contact (contacts may be repeated). The *length* of a leap sequence is the number of leap types, i.e., one fewer than the number of contacts. The *concatenation* of leap sequences $s$ and $s'$, denoted $s \| s'$ is defined only if the ending contact of $s$ is the starting contact of $s'$ and is obtained by juxtaposing $s$ and $s'$ and deleting the duplicate contact. A path *follows s from placement* $\omega$ if the source placement of the path is $\omega$ and the path consists of leaps to successive placements in $s$ where the leaps have corresponding types given by $s$. A leap sequence $s$ defines a function $g_s$ from placements at its starting contact to placements at its ending contact: if $s = (a)$ then $g_s$ is the identity function with domain the contact interval at $a$; if $s = (a, t, b) \| s'$ then $g_s$ is $g_{s'} \circ g^t_{ab}$ with domain $\mathrm{dom}(g^t_{ab}) \cap (g^t_{ab})^{-1}(\mathrm{dom}(g_{s'}))$. (Recall $g^t_{ab}$ is the function from placements at $a$ to placements at $b$ obtained by following leaps of type $t$.) Clearly $\mathrm{dom}(g_s)$ is a closed interval, possibly empty, and if $\omega \in \mathrm{dom}(g_s)$ then the path obtained by following $s$ from $\omega$ has final placement $g_s(\omega)$. (The path is not required to avoid obstacles.)

Each function $g^t_{ab}$ is either monotonically increasing or monotonically decreasing. Hence $g_s$ is monotonically increasing or monotonically decreasing; furthermore, $g_s$ is monotonically decreasing exactly if it is the composition of an odd number of monotonically decreasing functions.

We define interval $D(\omega, s)$, the interval of placements containing $\omega$ so that all paths following $s$ from placements in the interval are homotopic and avoid obstacles. If $s = (a)$ then $D(\omega, s) = \mathrm{dom}(g_s)$, the contact interval at $a$. If $s = (a, t, b)$ then $D(\omega, s)$ is the (possibly empty) interval containing $\omega$ so that any leap of type $t$ from $a$ to $b$ with source placement in the interval is feasible. If $s = (a, t, b) \| s'$, then $D(\omega, s) = D(\omega, (a, t, b)) \cap (g^t_{ab})^{-1}(D(g^t_{ab}(\omega), s')$. Clearly $D(\omega, s) \subseteq \mathrm{dom}(g_s)$. If the path following $s$ from $\omega$ avoids obstacles then $\omega \in D(\omega, s)$ and if not then $D(\omega, s) = \varnothing$.

A *base label* at contact $a$ is a root at $a$ or a placement $\omega$ at $a$ so that $\omega = g_s(\omega)$ and $\omega \in D(\omega, s)$, where $s$ is a leap sequence starting and ending at $a$. The *length* of a base label is zero if it is a root and the length of $s$ if it is a fixed point of $g_s$. A *label* at contact $a$ is a pair $(\omega, s)$ where $\omega$ is a base label at some contact $a'$, $s$ a leap sequence starting at $a'$ and ending at $a$, and $\omega \in D(\omega, s)$. The *length* of a label is the length of its base label plus the length of its leap sequence.

LEMMA 6.1.1

If the endpoints of $I$ have label of length $k$, then the endpoints of $T(j, I)$ can be given labels of length no more than $\max(k+n, n+1)$.

*Proof*

Let $\theta$ be an endpoint of $T(j, I)$. If $\theta$ is an endpoint of $J(j, I)$, then by lemma 3.2.4, either $\theta$ is obtained by one or two leaps from an endpoint of $I$ or $\theta$ has a label of length zero. If $\theta$ is an endpoint of $S_j$, for some $j = j_L$, $j_L'$, $j_R$, or $j_R'$, then either $\Omega(j)$ is an endpoint of $I$ or again by lemma 3.2.4, $\Omega(j)$ has a label of length zero. Since extreme paths resulting from *stretch* consist of at most $n$ leaps, $\theta$ has a label of length at most $\max(k+n, n+1)$. □

An interval is *labelled* if its endpoints are labelled. We henceforth assume that all intervals are labelled. $T(j, I)$ maps labelled interval $I$ into an interval whose endpoints are labelled using the previous lemma. Similarly $\mathscr{T}$ maps a set of labelled intervals into a set of labelled intervals.

Suppose $s$ is a leap sequence from a contact back to itself, $g_s$ is monotonically increasing, $g_s(\omega) \neq \omega$, and $\omega \in D(\omega, s) = [\omega_R, \omega_L]$. If $\omega < g_s(\omega)$, define $\mathrm{fix}(\omega, s) = \min(g_s(\omega_L), \{\omega' : \omega' = g_s(\omega') \text{ and } \omega' \geqslant \omega\})$; if $\omega > g_s(\omega)$, define $\mathrm{fix}(\omega, s) = \max(g_s(\omega_R), \{\omega' : \omega' = g_s(\omega') \text{ and } \omega' \leqslant \omega\})$.

LEMMA 6.1.2

If $g_s(\omega) > \omega$ and $[\omega, g_s(\omega)]$ is reachable, then $[\omega, \mathrm{fix}(\omega, s))$ is reachable. Similarly if $g_s(\omega) < \omega$ and $[g_s(\omega), \omega]$ is reachable, then $(\mathrm{fix}(\omega, s), \omega]$ is reachable.

*Proof*

Suppose $g_s(\omega) > \omega$ and $\omega' \in [\omega, \mathrm{fix}(\omega, s))$. Then for some $k$, $g_s^{(k+1)}(\omega) > \omega' \geqslant g_s^{(k)}(\omega)$, i.e. for some $\omega'' \in [\omega, g_s(\omega)]$, $\omega' = g_s^{(k)}(\omega'')$. Since $[\omega, \mathrm{fix}(\omega, s)] \subseteq D(\omega, s)$, $\omega'$ can be reached by following $s$ $k$ times from $\omega''$. □

LEMMA 6.1.3

Placement $\mathrm{fix}(\omega, s)$ has a label of length at most the length of $s$.

*Proof*

If fix($\omega$, $s$) is a fixed point of $g_s$, it has a base label of length the same as $s$. Otherwise fix($\omega$, $s$) $= g_s(\omega')$ where $\omega'$ is an endpoint of $D(\omega, s)$. For some $s'$, $a$, $t$, $b$, $s = s' \| (a, t, b) \| s''$, $g_{s'}(\omega')$ is an endpoint of $\Omega(X(j, \text{dom}(g^t_{ab})))$. By lemma 3.2.4, $g^t_{ab}(g_{s'}(\omega'))$ has a label of length zero. Hence $g_s(\omega') = g_{s''}(g_{s'}(\omega'))$ has a label of length at most the length of $s''$, i.e., at most the length of $s$. □

We now use fix to define a mapping $F$ from labelled intervals at a contact to labelled intervals at the same contact. First suppose $I$ is an interval at corner $a$. Then $F(I)$ is $I$ possibly merged with other intervals as follows. Let the left endpoint of $I$ have label ($\omega$, $s$). If possible, choose $s'$ and $s''$ so that $s = s' \| s''$, $s''$ is a leap sequence from $a$ to $a$ of length at least one, $g_{s''}$ is monotonically increasing, and $\omega' = g_{s'}(\omega) \in I$, and so that among all such choices $s'$ is chosen as long as possible. Merge $[\omega'$, fix$(\omega', s''))$ with $F(I)$. Notice if the merge occurs, ($s$, $\omega$) is not an endpoint label of the resulting interval since $g_s(\omega)$ is no longer an endpoint. If there is no way to choose $s'$ and $s''$, do not merge anything into $I$. This procedure is repeated with the right endpoint of $I$. If $I$ is an interval at a wall, the procedure is similar; it only need be performed at the endpoint of $I$ in direction opposite to motion along the wall.

We let $\mathscr{F}$ be the extension of $F$ to be a mapping from sets of labelled intervals into sets of labelled intervals.

LEMMA 6.1.4

Every interval in $(\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$, any $k \geq 0$, contains a marker. Hence there are at most $O(n^2(n^2 + m))$ intervals in $(\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$.

*Proof*

Similar to the proof of lemma 5.1.2 and theorem 5.1.3. □

LEMMA 6.1.5

The label of every interval in $(\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$ has length at most $O(n^3(n^2 + m))$.

*Proof*

Suppose ($\omega$, $s$) is the endpoint label of some interval $I \in (\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$, some $k > 0$. The *direct predecessor* of ($\omega$, $s$) is the label ($\omega$, $s'$) of the interval $I' \in (\mathscr{F} \circ \mathscr{T})^{(k-1)}(\mathscr{I}_0)$ so that $T(j, I') \subseteq I$ has endpoint label ($\omega$, $s$). Possibly no direct predecessor exists; if several exist choose one arbitrarily. The *predecessors* of ($\omega$, $s$) are ($\omega$, $s$), the direct predecessor of ($\omega$, $s$), its direct predecessor, and so on until no direct predecessor exists. Clearly if ($\omega'$, $s'$) is a predecessor of

$(\omega, s)$, then $\omega' = \omega$ and $s'$ is a prefix of $s$. If $l$ is an endpoint label of interval $I \in (\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$, then the *marker* of $l$, $m(l)$, is the marker contained in $I$, chosen arbitrarily if there is more than one.

We claim that for every endpoint label $l$ of an interval in $(\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0)$ there are at most two predecessors $l_1$ and $l_2$ of $l$ with the same marker. The proof is by induction on $k$; it is trivial for $k = 0$. Notice that an endpoint label of $(\mathscr{F} \circ \mathscr{T})^{(k+1)}(\mathscr{I}_0)$ is either an endpoint label of $\mathscr{T}((\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0))$ or it is fix$(\omega, s)$ for some $\omega$ and $s$; in the latter case it has no direct predecessors. Hence it suffices to show that if $l = (\omega, s)$ is an endpoint label of some interval $I$ in $\mathscr{T}((\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0))$ with $m(l_1) = m(l_2) = m(l_3)$ for three distinct predecessors $l_1$, $l_2$, $l_3$ of $l$, then $l$ is not an endpoint label in $(\mathscr{F} \circ \mathscr{T})^{(k+1)}(\mathscr{I}_0)$. Now one of $l_1$, $l_2$, $l_3$, say $l_3$, must be $l = (\omega, s)$, otherwise the inductive hypothesis would be violated. We can assume $l_1$ is shorter than $l_2$, hence we can write $l_1 = (\omega, s_1)$, $l_2 = (\omega, s_1 \| s_2)$, $l_3 = (\omega, s_1 \| s_2 \| s_3) = (\omega, s)$. Now either $g_{s_2}$, $g_{s_3}$, or $g_{s_3} \circ g_{s_2}$ must be monotone increasing. Also $g_{s_1}(\omega)$, $g_{s_2}(g_{s_1}(\omega))$, $g_s(\omega)$ all lie in $I$ since they all have the same marker. Hence it is possible to write $s = s' \| s''$ with $g_{s''}$ monotone increasing and $g_{s'}(\omega) \in I$. Thus after applying $\mathscr{F}$, $g_s(\omega)$ is no longer an endpoint, and $(\omega, s)$ is no longer an endpoint label.

Since there are only $O(n^2(n^2 + m))$ markers, every endpoint label has at most $O(n^2(n^2 + m))$ predecessors. Since the direct predecessor of an endpoint label can be at most $n$ shorter than the label, and since every label with path sequence longer than $n$ has a direct predecessor, the path sequence portion of every label has length at most $O(n^3(n^2 + m))$. The length of a base label is also at most $O(n^3(n^2 + m))$, since a base label is either a root or a fixed point arising from some subsequence of the path sequence portion of some label. □

LEMMA 6.1.6

There are at most $n^{O(k)}$ distinct labels of length $k$.

*Proof*

Let $l = (\omega, s)$ be of length $k$. Either $\omega$ is a root (of length 0) or $\omega$ is a fixed point of $g_{s'}$, where $s'$ is some leap sequence of length $k' \leqslant k$. We show in the next section that there are at most $2^{O(k')}$ distinct fixed points of $g_{s'}$. Clearly there are at most $n^{O(k)}$ distinct choices of $s$ of length $k$, or of $s$ and $s'$ together of combined length $k$. Hence there are $2^{O(k')}n^{O(k)}$, i.e., $n^{O(k)}$ distinct labels of length $k$. □

THEOREM 6.1.7

There is $k \in 2^{O(n^3(m+n^2)\log n)}$ so that $(\mathscr{F} \circ \mathscr{T})^{(k)}(\mathscr{I}_0) = (\mathscr{F} \circ \mathscr{T})^{(k+1)}(\mathscr{I}_0)$.

*Proof*

There are at most $n^{O(n^3(n^2+m))} = 2^{O(n^3(n^2+m)\log n)}$ different possible labels, hence only $2^{O(n^3(n^2+m)\log n)}$ different possible intervals. Each successive application of $\mathscr{F} \circ \mathscr{T}$ must discover a new interval to be reachable, or discover nothing new. The former can happen only $2^{O(n^3 m \log n)}$ many times. □

### 6.2. IMPLEMENTATION DETAILS

We now discuss the implementation of mappings $\mathscr{F}$ and $\mathscr{T}$. We use as primitive an *elementary test*: given two labels at the same contact, determine the order of the corresponding placements. We show below how to construct a formula of $\mathbb{R}(+, \times)$ representing a placement specified by a label. Hence performing an elementary test involves deciding a formula of $\mathbb{R}(+, \times)$.

Mapping $\mathscr{T}$ can be computed using only elementary tests, in fact using only a number of tests polynomial in $m$ and $n$. In particular, elementary tests suffice for computing $J(j, I)$ (given labels for $I$), for computing the labels and placements implied by lemma 3.2.4, and for performing *stretch*. The jump guaranteed by theorem 4.3.1 is an intermediate jump in the computation of *stretch*, hence can be computed. The merge operator $\sqcup$ is straightforward given elementary tests. Mapping $\mathscr{F}$ is similar with the exception of the computation of fix, discussed below.

Let $l = (\omega, s)$ be a label. We construct a formula $F_l(v)$ of $\mathbb{R}(+, \times)$ that is satisfied exactly when $g_s(\omega)$ is substituted for free variable $v$. Formula $F_l(v)$ has number of bound variables proportional to the logarithm of the length of $l$ and length exponential in the length of $l$.

We begin by constructing a formula $G_s(\omega, \theta)$ satisfied exactly for pairs $\theta$ and $\omega$ so that $\theta = g_s(\omega)$. Formula $G_s(\omega, \theta)$ has length $O(rm)$, where $r$ is the length of $s$ and $O(\log r)$ bound variables. The construction uses the "path-doubling" trick. For each function $g^t_{ab}$ there is a formula with two free variables satisfied exactly by pairs $\theta$ and $\omega$ so that $\theta = g^t_{ab}(\omega)$. These formulas can be written in prenex normal form with common prenex $P$; there is a constant $c$ so that each formula is of size at most $cm$. For variables $b_1, \dots, b_i$, $0 \leqslant i \leqslant k = \log_2 r$, with each $b_j$ either 0 or 1, let $s[b_1, \dots, b_i]$ be the subsequence of $s$ of length $2^{k-i}$ and index $b_1 \dots b_i$, where $b_1 \dots b_i$ is interpreted as a number written in binary. We construct a formula $G^i_s(\omega, \theta, b_1, \dots, b_i)$ satisfied exactly when $b_1, \dots, b_i$ are each either 0 or 1 and $g_{s[b_1,\dots,b_i]}(\omega) = \theta$. We start with $i = k$. $G^k_s(\omega, \theta, b_1, \dots, b_k)$ can be constructed from the formulas representing the functions $g^t_{ab}$ by starting with the common prenex $P$ and using the variables $b_1, \dots, b_k$ to select the appropriate matrix. This formula can be constructed with length $O(mr)$ and having only the bound variables in the prenex $P$. Then $G^{i-1}_s(\theta, \omega, b_1, \dots, b_{i-1})$ is

$$\exists\gamma\forall\alpha\forall\beta\forall b((\alpha = \theta \wedge \beta = \gamma \wedge b = 0) \vee (\alpha = \gamma \wedge \beta = \omega \wedge b = 1)$$
$$\Rightarrow G^i_s(\alpha, \beta, b_1, \dots, b_{i-1}, b).$$

We set $G_s = G_s^0$.

We now construct a formula $B_\omega(x)$ satisfied exactly by $\omega$, where $\omega$ is a base label. For any root $\omega$, this is trivial; the formula has size $O(m)$ and constant number of bound variables. For any path sequence $s$ the formula $G_s(x, x)$ is satisfied exactly by the $\omega$ so that $g_s(\omega) = \omega$. Unfortunately there may be many such $\omega$, and we need a formula satisfied by a particular such $\omega$.

By lemma 2.3.3, for each function $g_{ab}^t$ there is a two-variable polynomial $p_{ab}^t$ so that $g_{ab}^t(\omega) = \theta$ implies $p_{ab}^t(\omega, \theta) = 0$. Furthermore, there is a constant $c$ bounding the degree of all such polynomials and a constant $d$ so that all coefficients have size less than $dm$. If $g_s(\omega) = \theta$, we can write $\omega_1 = g_1(\omega_0)$, $\omega_2 = g_2(\omega_1), \ldots, \omega_r = g_r(\omega_{r-1})$, where $\omega_0 = \omega$, $\omega_r = \theta$, $r$ is the length of $s$, and $g_1, \ldots, g_r$ are the leap functions whose composition is $g_s$. Hence $p_1(\omega_0, \omega_1) = 0$, $p_2(\omega_1, \omega_2) = 0, \ldots, p_r(\omega_{r-1}, \omega_r) = 0$. Using resultants, we can eliminate $\omega_1, \ldots, \omega_{r-1}$ and get the single polynomial $p(\omega_0, \omega_r)$. The degree of $p$ is $c^r$ and the size of its coefficients is bounded by $c^{2r}dm$. To get this bound we need to use a doubling trick: first eliminate $\omega_1$ from $p_1$ and $p_2$, $\omega_3$ from $p_3$ and $p_4$, $\omega_5$ from $p_5$ and $p_6$, and so forth. This squares the degree of the polynomials and cuts the number of variables in half, maintaining a diagonal system. Iterating $\log_2 r$ times gives the claimed bound. Let $p_s(\omega) = p(\omega, \omega)$. Then any $\omega$ so that $g_s(\omega) = \omega$ also satisfies $p_s(\omega) = 0$. Since the degree of $p_s$ is at most $c^r$, there are at most $c^r$ fixed points of $g_s$. (This completes the proof of lemma 6.1.6.)

Let $q = p/\gcd(p, p')$, then $q$ has the same roots as $p$ but of multiplicity one. Let $V(x)$ be the number of variations in sign of the Sturm sequence of $\hat{p}$ at $x$. By Sturm's Theorem, the number of zeroes of $q(x)$ in $[-\infty, x]$ is given by $V(-\infty) - V(x)$. Hence the sequence of signs of the Sturm sequence of $\hat{p}$ at a particular root $\omega$ is unique. We use this sequence to uniquely identify a particular $\omega$ satisfying $g_s(\omega) = \omega$: $B_\omega(x)$ is "$G_s(x, x)$ and $\bigwedge q^{(i)}(x) : 0$", where $q^{(i)}$ is the $i$th derivative of $q$ and the operator ":" is $>$, $<$, or $=$ as appropriate for $\omega$. For $s$ with length bounded by $O(n^3(n^2 + m))$, this formula has length $2^{O(n^3(n^2+m))}$, because both the degree of $q$ and the size of coefficients can be this big.

We can now construct the required formula $F_l$ for label $l = (\omega, s)$. $F_l(y)$ is "$\exists x\ B_\omega(x)$ and $G_s(x, y)$".

The remaining problem is to actually compute $F_l$ when needed. This is an issue only for the operation fix($\omega$, $s$). We can use elementary tests to decide if fix($\omega$, $s$) is a fixed point of $g_s$ or is $g_s$ applied to an endpoint of $D(\omega, s)$. The latter case is easily handled since fix($\omega$, $s$) has a label consisting of a root and a subsequence of $s$. In the former case, we compute $\hat{p}$ corresponding to $g_s$ and its derivatives. This takes time $2^{O(n^3(n^2+m))}$ as $s$ is bounded in length by $O(n^3(n^2 + m))$. We then evaluate the signs of the derivatives using the decision procedure for $\mathbb{R}(+, \times)$. This is possible since fix($\omega$, $s$) is given by the formula expressing "fix($\omega$, $s$) is the least fixed point of $g_s$ greater than $\omega$" (or "largest fixed point smaller than $\omega$"), and we have formulas for $g_s$ and $\omega$. Hence we can actually construct $B_\omega$. (Notice that we cannot simply use the formula for fix($\omega$, $s$)

involving the formula for $\omega$. We need a formula for fix($\omega$, $s$) whose size is independent of how it was computed, since its computation might be quite complex.)

THEOREM 6.2.1

The reachability question is decidable in time $2^{poly(n,m)}$ (hence space $2^{poly(n,m)}$).

*Proof*

An elementary test involves formulas of length $2^{O(n^3(n^2+m))}$ and only $O(\log nm)$ bound variables, hence can be decided in time $2^{poly(n,m)}$ and space $2^{poly(n,m)}$. The composition $\mathscr{F} \circ \mathscr{T}$ can be computed in essentially the same time bound. By theorem 6.1.7, $2^{poly(n,m)}$ applications of $\mathscr{F} \circ \mathscr{T}$ suffice. □

Notice that it is not evident how to give an accurate estimate of the degree of the polynomial in $n$ and $m$ that bounds the running time, since the degree is hidden in the asymptotic constant of the time bound for the decision procedure for $\mathbb{R}(+, \times)$.

## 7. Generalizations and open problems

Suppose we allow backward as well as forward motions. Laumond [9] shows that in this case, one placement is reachable from another if the two placements are in the same open path connected component of free space. In fact, Laumond proves the result for the more difficult case of a 2-dimensional model of a mobile cart. The idea of the proof is to approximate the path between the two placements by a sequence of short back and forth motions. However, there is no known bound on the number of times it is necessary to reverse motion from forward to backward or backward to forward in order to achieve reachability. Any physical mechanism will have some cost for reversing its motion, and so it would be interesting to find a motion that minimizes the number of reversals. It is easy to show that for the case of the directed point allowed to move backward or forward, it is sufficient to consider only normal paths and in fact we can restrict reversals in motion to take place only at contacts. Therefore the algorithms in this paper can be iterated $k+1$ items (where the source placements for an iteration are obtained by reversing the directions of the reachable placements computed by the previous iteration) to compute the placements reachable by paths with at most $k$ reversals.

Finding the shortest bounded curvature path between two placements in a universe containing obstacles remains an open problem. Even in the absence of obstacles this is not trivial; Dubins [7] shows how to compute shortest paths when only forward motions are allowed and Reeds and Shepp [11] answer the question when both forward and backward motions are permitted. We note that our approach is not adequate for shortest paths, since the transformations to obtain normal paths do not preserve shortest paths.

Suppose we wish to extend the model of a forward moving automobile from a point to a more complex object. Consider the following model of a bounded-radius-of-curvature constraint on the motion of a line segment. The path of the segment is specified by the path of some interior point $p$ of the segment; the segment is always tangent to the path of $p$. The path of $p$ has a fixed minimum radius of curvature.

If the segment turns right, then its trailing endpoint swings slightly to the left. To see this, note that the endpoint and $p$ both move in circular arcs with the same center, but the radius of the arc at the endpoint is larger. If there is an obstacle just to the left of the segment then the obstacle, rather than the curvature constraint, becomes the limiting factor on how quickly the segment can turn.

Suppose the obstacle is the $x$-axis, the segment is moving to the right, and the trailing endpoint is on the $x$-axis. The quickest the segment can turn is to have the trailing endpoint just drag along the $x$-axis. If point $p$ is at $(x, y)$ and the distance from $p$ to the endpoint is 1, then the path of $p$ satisfies the equation $\mathrm{d}x/\mathrm{d}y = \sqrt{(1-y^2)}/y$. This can be solved explicitly as $x(y) = \sqrt{(1-y^2)} - \ln|(1+\sqrt{1-y^2})/y|$; this gives the $x$-coordinate of point $p$ when $p$ is at height $y$, assuming that $x = 0$ when $y = 1$. The significance of this equation is that it involves both algebraic numbers and logarithms.

Can we plan motion for the segment? The path of the segment can consist of sections where an obstacle limits the radius of curvature and sections where the explicit bound limits the radius of curvature. The path of $p$ in the latter type of section can be described algebraically, while the former type requires the equations just derived. The two types of sections can alternate arbitrarily; hence the position of $p$ is described by an expression involving arbitrary nesting of algebraic numbers and logarithms. Unfortunately, it is not known how to manipulate such numbers. The strongest known result is that it is possible to decide whether a number of the form $\Sigma\alpha_i \log\beta_i$ is greater than zero, where the $\alpha$'s and $\beta$'s are algebraic numbers [13]. But in such an expression there is only a single level of nesting. Hence it seems unlikely that we can devise an exact motion planning algorithm for the segment, with the present state of knowledge.

Finally, we mention the following problem, suggested by Reif [12]. Consider a point moving in a two-dimensional space with fixed polygonal obstacles. The point is subject to unit bounds on its acceleration. The question is "Given a source placement (position plus velocity) and a target placement, is there a path from source to target avoiding all obstacles subject to the bounds on acceleration?" We believe this problem to be quite challenging. Note that planning motion with bounded radius of curvature is a special case, with acceleration always perpendicular to the direction of motion. Suppose both the source and target velocity are zero. If there is an arbitrary path between the placements then there is a path $p$

consisting of line segments and the point can follow $p$ for example by starting with zero velocity at the beginning of any segment of $p$, speeding up with unit acceleration to the midpoint of the segment and then slowing down with unit deceleration until stopped at the next endpoint. Therefore for arbitrary source and target velocities a sufficient (but not necessary) condition for reachability is to be able to bring the point to a halt from the source placement and from the target placement (with velocity reversed). However deciding if the point can be brought to a halt from a placement stills appears to be a difficult problem.

## Appendix 1

*Remainder of the proof of lemma 4.2.2*

Recall that $P(\theta)$ is the path as defined by *stretch* with final placement $\theta$, $\Lambda$ is the list of oriented placements defining $P(\theta)$, and $\phi(\theta)$ is the direction of the tangent of the last jump of $P(\theta)$. We wish to bound the change in $\phi$ while $\theta$ varies over monotonic interval $I$, given that $\Lambda$ is in state (Rs) followed by state (Ls), i.e., all final arcs of $P(\theta)$ are short. Notice $\phi$ is decreasing (moving to the right). Events (REV) and (TWIND) cannot happen in the interior of $I = [\theta_0, \theta_f]$ since $I$ is monotonic. Also (UNBEND) never occurs in $I$, because initially there are no intermediate oriented placements in $\Lambda$, and if any oriented placement is created then its arc length along path $P$ is increasing. We ignore event (SWIND). The remaining possible events are (SFLIP) from orientation $L$ to $R$, (TFLIP) from orientation $R$ to $L$, and (BEND) creating placements of orientation $R$. We view (SFLIP) as an instance of (BEND).

We find a function $B(\theta)$ that depends upon the final jump of $P(\theta)$. Function $B$ is defined to bound how much $\phi$ can decrease as $\theta$ increases, assuming (BEND) does not occur or occurs once, whether or not (TFLIP) occurs. We show that in fact $B$ does not decrease if (BEND) occurs, hence the original value $B(\theta_0)$ bounds $\phi$ throughout $I$. We also show that $B(\theta_0) \geqslant \phi(\theta_0) - \pi$, proving lemma 4.2.2. The bound $B$ is defined using an auxiliary function $b$, defined on $XxXs$ jumps to $u_\tau$. $B(\theta)$ is just $b(j)$, where $j$ is the final jump of path $P(\theta)$.

Let $j = (\rho, \tau)$ be an $XxXs$ jump to $u_\tau$. To define $b(j)$, we find a subpath $V$ of jump $j$. $V$ is the set of points along $j$ at which (BEND) can occur. If $j$ is $RX$, then $V$ is just the tangent segment of $j$. If $j$ is $LX$, then $V$ is the tangent segment of $j$ plus the arc of $C_\rho$ backwards to the first of $u_\rho$ or the point at which an $LxR\pi$ jump with source placement $\rho$ diverges from $C_\rho$.

For $v \in V$ let $\hat{\rho}(v)$ be the $R$-oriented placement tangent to $V$ at $v$ and $\hat{\tau}(v)$ the $L$-oriented placement through $u_\tau$ chosen so that among $RsLs$ jumps from $\hat{\rho}(v)$ to $u_\tau$, the jump $(\hat{\rho}(v), \hat{\tau}(v))$ has direction of tangent as small as possible. The possibilities for $(\hat{\rho}(v), \hat{\tau}(v))$ depend upon the distance between $z_{\hat{\rho}(v)}$ (the center of $C_{\hat{\rho}(v)}$) and $u_\tau$: if the distance is at least 3 then $(\hat{\rho}(v), \hat{\tau}(v))$ is an $RsL\pi$ jump (with nonempty tangent segment if the distance is strictly greater than 3),

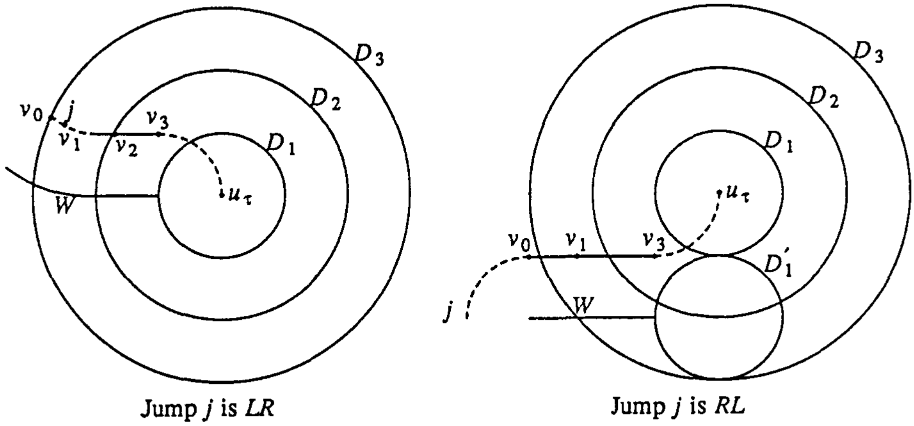


Fig. 16. Examples of definitions.

and if the distance is less than 3 then $(\hat{\rho}(v), \hat{\tau}(v))$ is an $RsLs$ hop. Let $\hat{\phi}(v)$ be the direction of the tangent segment of $(\hat{\rho}(v), \hat{\tau}(v))$. We define $b(j) = \min_{v \in V} \hat{\phi}(v)$.

Let $W$ be $\{z_{\hat{\rho}(v)}: v \in V\}$; then $W$ is the path that is distance one to the right of $V$. Manifestly $W$ consists of a straight segment $W_s$ and possibly an $L$-oriented arc $W_a$ of radius 2. Let $l$ be the line through $u_\tau$ parallel to $\Phi(j)$; we assume henceforth $l$ is horizontal. Let $D_i$, $i = 1, 2, 3$, be the $L$-oriented circles of radius $i$ centered at $u_\tau$. Let $D_1'$ be the circle of radius 1 distance two below $D_1$ (see fig. 16). Let $H$ be the halfplane with bounding line through $u_\tau$ perpendicular to $l$ contain the tangent segment of $j$.

LEMMA A1.1

(1) $W \subseteq H$ and $W$ has at most one point of intersection with $D_3$.

(2) The final endpoint of $W$ lies on $D_1$ (if $j$ is $XR$) or on $D_1'$ (if $j$ is $XL$).

(3) $W_a$ is at most a semicircle; furthermore if part of $W_a$ lies outside $D_3$, then $W_a$ extends at most as far as the point where a line is tangent to both $D_3$ and the circle through $W_a$.

*Proof*

Let $j = (\rho, \tau)$.

(2) Let $e$ be the final endpoint of $W$. If $j$ is $XR$, then $e = z_\tau$, so $e \in D_1$. If $j$ is $XL$, then $e$ is distance 2 below $z_\tau$, so $e \in D_1'$.

(3) Suppose $j$ is $LX$; $z_\rho$ lies distance two above the line through $W_s$, and must lie to the left of the final endpoint of $W$; using (2) it follows that $C_\rho$ cannot be contained in the interior of $D_2$. By definition of $V$, the arc of $V$ backwards around $C_\rho$ extends at most as far as the point of tangency of a line tangent to both $C_\rho$ and $D_2$ (since then the arc length around $C_{\hat{\tau}}$ would be $\pi$); this arc length is clearly at most $\pi$. Hence $W_a$ is at most a semicircle. Furthermore, it is easy to

see that the tangent to $D_3$ and the circle of radius 2 about $z_\rho$ is distance one from the tangent between $D_2$ and $C_\rho$. Hence $W_a$ extends at most as far as the point where a line is tangent to both $W_a$ and $D_3$.

(1) follows easily from (3). □

We say $j$ is *close* if $W$ is inside $D_3$. Let $v_0$ and $v_3$ be the initial and final endpoints of $V$ (considering $V$ as a directed subpath of $j$). If $j$ is not close, choose $v_1 \in V$ so that $z_{\hat{\rho}(v_1)}$ is the point of intersection of $W$ and $D_3$. If $j$ is close, set $v_1$ to $v_0$.

LEMMA A1.2

$b(j) = \min(\hat{\phi}(v_1), \Theta(j))$.

*Proof*

If $j$ is $XR$ we show $\hat{\phi}(v)$ is decreasing in $[v_0, v_1]$, increasing in $[v_1, v_2]$, decreasing in $[v_2, v_3]$, ($v_2$ is defined below), and $\hat{\phi}(v_3) = \Theta(j)$; hence $b(j) = \min(\hat{\phi}(v_1), \Theta(j))$. If $j$ is $XL$ we show $\hat{\phi}(v)$ is decreasing in $[v_0, v_1]$, increasing in $[v_1, v_3]$, and $\hat{\phi}(v_3) \leqslant \Theta(j)$; hence again $b(j) = \min(\hat{\phi}(v_1), \Theta(j))$.

First suppose $j$ is not close (of any type), and $v \in [v_0, v_1]$. Then $z_{\hat{\rho}(v)} \in W$ is on or outside $D_3$. Let $s(v)$ be the segment from $z_{\hat{\rho}(v)}$ tangent to $D_3$. We claim $\hat{\phi}(v)$ is the direction of $s(v)$: jump $(\hat{\rho}(v), \hat{\tau}(v))$ is an $RsL\pi$ jump, so its tangent segment is also tangent to $D_2$, and segment $s(v)$ is parallel to the tangent segment of $(\hat{\rho}(v), \hat{\tau}(v))$. Now as $v$ moves from $v_0$ to $v_1$, it is clear that the point of tangency of $s(v)$ with $D_3$ moves clockwise around $D_3$ (using lemma A1.1(3) in case $W$ contains a circular arc). Hence $\hat{\phi}(v)$ decreases in $[v_0, v_1]$.

Now suppose $v \in [v_1, v_3]$. Then $z_{\hat{\rho}(v)} \in W$ is on or inside $D_3$. Let $s(v)$ be the segment of length 2 from $z_{\hat{\rho}(v)}$ to $D_1$, chosen so that $s(v)$ lies to the left of the line from $z_{\hat{\rho}(v)}$ to $u_\tau$. We claim $\hat{\phi}(v)$ is $\pi/2$ to the right of the direction of $s(v)$: $(\hat{\rho}(v), \hat{\tau}(v))$ is an $RsLs$ hop, and the midpoint of $s(v)$ is the point of tangency of $C_{\hat{\rho}(v)}$ and $C_{\hat{\tau}(v)}$.

As $v$ moves from $v_1$ to $v_3$, how does $\hat{\phi}$ change? Let $\alpha(v)$ be the direction of the tangent to $W$ at $z_{\hat{\rho}(v)}$ and $\beta(v)$ the direction of the tangent to $D_1$ at $z_{\hat{\tau}(v)}$ (consider $D_1$ as oriented $R$). It is easy to check that $\alpha(v)$ and $\beta(v)$ never differ by more than $\pi/2$. If $\alpha(v)$ is to the right of $\beta(v)$, $s(v)$ rotates counterclockwise; if $\alpha(v)$ is to the left of $\beta(v)$, $s(v)$ rotates clockwise.

Suppose $j$ is $XR$, then $W_s$ is within one of $l$, either above or below. Assume for the moment that $v_1 \in D_3$. Clearly $\alpha(v_1)$ is to the right of $\beta(v_1)$ and $\hat{\phi}$ is increasing after $v_1$. Direction $\alpha$ remains to the right of $\beta$ until the tangent to $W$ at $z_{\hat{\rho}(v)}$ becomes parallel to the tangent to $D_1$ at $z_{\hat{\tau}(v)}$. This may happen when $z_{\hat{\tau}(v)}$ is the topmost point of $D_1$ and $z_{\hat{\rho}(v)} \in W_s$, or it may happen with $z_{\hat{\rho}(v)} \in W_a$ and $z_{\hat{\tau}(v)}$ past the topmost point of $D_1$. We define $v_2$ to be this value of $v$. Past $v_2$ direction $\alpha(v)$ is to the left of $\beta(v)$ and $\hat{\phi}$ is decreasing. If $v_1 \notin D_3$, then the analysis is similar. Possibly $W$ is so short that $\alpha$ is always to the left of $\beta$ and $\hat{\phi}$ is always decreasing; we set $v_2 = v_1$ in this case. Notice that at $v_3$, jump $(\hat{\phi}(v_3), \hat{\tau}(v_3))$ consists only of the arc of $C_\tau$ traversed in $j$; in particular $\hat{\phi}(v_3) = \Theta(j)$.

If $j$ is $XL$, the analysis is similar. The straight part of $W$ lies between distance 1 and distance 3 below $l$. Angle $\alpha$ remains to the right of $\beta$ until $v_3$ since $z_{\hat{\tau}(v)}$ cannot reach the topmost point of $D_1$. Hence $\hat{\phi}$ is increasing in $[v_1, v_3]$. At $v_3$, jump $(\hat{\phi}(v_3), \hat{\tau}(v_3))$ again consists only of the arc of $C_\tau$ traversed in $j$. However in this case $\hat{\phi}(v) = \Phi(j) < \Theta(j)$. □

LEMMA A1.3

$b(j) \geqslant \Phi(j) - \pi$.

*Proof*

By the previous lemma, $b(j) = \min(\hat{\phi}(v_1), \Theta(j))$. We have $\Theta(j) \geqslant \Phi(j) - \pi$, since in the worst case $j$ is an $XxRs$ jump, so its final arc is short. We must have $z_{\hat{\rho}(v_1)} \in H$, since $z_{\hat{\rho}(v_1)} \in W$ and $W_a$ is at most a semicircle, by lemma A1.1(3). If $j$ is not close, then $z_{\hat{\rho}(v_1)} \in D_3 \cap H$ and $\hat{\phi}(v_1)$ is the direction of the tangent of $D_3$ at $z_{\hat{\rho}(v_1)}$, clearly at least $\Phi(j) - \pi$. If $j$ is close, then $\hat{\phi}(v_1)$ is $\pi/2$ to the right of the direction of $s(v)$, but the direction of $s(v)$ is itself at most $\pi/2$ to the right of $\Phi(j)$. In either case we have $\hat{\phi}(v_1) \geqslant \Phi(j) - \pi$. □

For the following we write for example $V(j)$ or $v_1(j)$ to make the dependence upon $j$ explicit.

LEMMA A1.4

If $k$ is an $XxXs$ jump with the same source placement as $j$ and $\Theta(k) > \Theta(j)$, then $b(k) \geqslant b(j)$.

*Proof*

We have either $j$ and $k$ are both the same type, or the target placement of $j$ is oriented $R$ while the target placement of $k$ is oriented $L$. Also, we must have $\Phi(k) < \Phi(j)$. We claim $W(j)$ and $W(k)$ have the same initial endpoint (if $j$ and $k$ are $RX$) or have common initial arc and then diverge (if $j$ and $k$ are $LX$). This follows from the observation that $C_{\hat{\rho}(v_0(j))} = C_{\hat{\rho}(v_0(k))} = C_\rho$ if $j = (\rho, \tau)$ is $RX$ and $C_{\hat{\rho}(v_0(j))}$ and $C_{\hat{\rho}(v_0(k))}$ are both tangent to $C_\rho$ if $j$ is $LX$.

It suffices to show $\hat{\phi}(v_1(k)) \geqslant \hat{\phi}(v_1(j))$. First suppose $j$ is not close, then $k$ is not close either. It is clear that $W(k) \cap D_3$ is counterclockwise from $W(j) \cap D_3$ around $D_3$; hence $\hat{\phi}(v_1(k)) \geqslant \hat{\phi}(v_1(j))$. If $j$ is close, then $k$ is close as well. Since $W(j)$ and $W(k)$ have the same initial endpoints, we have $\hat{\phi}(v_1(k)) = \hat{\phi}(v_1(j))$. □

*Proof of lemma 4.2.2*

We show $B(\theta)$ does not decrease during *stretch*; the result then follows from lemma A1.3. Let $j = (\rho(\theta), \tau(\theta))$ be the final jump of $P(\theta)$. Suppose (BEND) occurs at $\theta' > \theta$; let $k$ be the final jump of $P(\theta')$ before processing (BEND) and $l$ be the jump newly created by processing (BEND). Clearly $l$ is a final portion of $k$ and the source placements of $j$ and $k$ are the same. By lemma A1.4, $b(k) \geqslant b(j)$, and $b(l) \geqslant b(k)$ since $V(l)$ is a subpath of $V(k)$. Hence $B(\theta') = b(l) \geqslant b(j) = B(\theta)$. □

**Appendix** 2

| Term | Symbol | Comment | Section |
|---|---|---|---|
| universe | $U$ | polygonal, closed, bounded | 2.1 |
| geometric complexity | $n$ | number of corners of $U$ | 2.1 |
| bit complexity | $m$ | number of bits to define $U$ | 2.1 |
| direction | $\theta$ | $\theta \in S^1$ | 2.1 |
| interval of directions | $I = [\theta_1, \theta_2]$ | arc of $S^1$ | 2.1 |
| position | $u$ | $u \in U$ | 2.1 |
| placement | $(u, \theta)$ | $u$ a position, $\theta$ a direction | 2.1 |
| path | $p$ | | 2.1 |
| initial placement | $\Omega(p)$, $\Omega(j)$ | initial placement of path $p$ or jump $j$ | 2.1 |
| final placement | $\Theta(p)$, $\Theta(j)$ | final placement of path $p$ or jump $j$ | 2.1 |
| orientation | $d$ | $d = L$ (counterclockwise) or $d = R$ (clockwise) | 2.2 |
| | $\bar{d}$ | orientation opposite to $d$ | 2.2 |
| oriented placement | $\sigma$, $\tau$, $\rho$ | $\sigma = (u_\sigma, \theta_\sigma, d_\sigma)$; $(u_\sigma, \theta_\sigma)$ a placement, $d_\sigma$ an orientation | 2.2 |
| | $\bar{\sigma}$ | orientation $\bar{d}$ | 2.2 |
| oriented circle | $C_\sigma$ | unit circle with orientation $d_\sigma$ and tangent direction $\theta_\sigma$ at $u_\sigma$ | 2.2 |
| | $z_\sigma$ | center of $C_\sigma$ | 2.2 |
| jump | $j$ | $j = (\sigma, \tau)$, $\sigma$ and $\tau$ oriented placements; also denotes resulting path | 2.2 |
| jump type | $XxXx$ | each $X$ is $L$ or $R$ and each $x$ is $s$ or $l$ indicating orientation and "length", respectively, of arcs in the jump | 2.2 |
| | $\Phi(j)$ | direction of tangent in jump $j$ | 2.2 |
| leap | | jump with arc of length 0 or $\pi$ or tangent with length 0 | 2.2 |
| hop | | leap with tangent length 0 | 2.2 |
| leap function | $g_{ab}^t$ | maps placement at $a$ to the placement at $b$ resulting from a leap of type $t$ | 2.3 |
| | $p_{ab}^t$ | polynomial encoding $g_{ab}^t$ | 2.3 |
| $c$-cone | $N_\alpha$ | $N_\alpha \subseteq \mathbb{R}^2$ | 3.1 |
| | $X(j, I)$ | jumps homotopic to $j$ with source in $I$ | 3.2 |
| jump representatives | $\Xi_{ab}^{t_0}(I)$ | collection of jumps of type $t_0$ | 3.2 |
| | $c(j) = (\Omega(j), \Theta(j))$ | $j$ a jump of type $t_0$ | 3.2 |
| configuration space | $C$ | $c(j)$'s for fixed contacts | 3.2 |
| | $f_{id}$ | curves partitioning $C$ | 3.2 |
| | $\mathscr{C}$ | collection of boundary curves and $f_{id}$'s | 3.2 |
| | $J(j, I)$ | interval of placements at $b$ | 3.3 |

Appendix 2 (continued)

| Term | Symbol | Comment | Section |
|---|---|---|---|
| | $\mathscr{I}^t_{ab}(I)$ | intervals of placements at $b$ | 3.3 |
| | $\mathscr{I}$ | sets of intervals of placements | 3.3 |
| | $\mathscr{I} \sqsubset \mathscr{I}'$ | each interval of $\mathscr{I}$ is contained in an interval of $\mathscr{I}'$ | 3.3 |
| merge | $\sqcup$ | smallest well-formed set $\mathscr{I}''$ so that $\mathscr{I} \sqsubset \mathscr{I}''$ and $\mathscr{I}' \sqsubset \mathscr{I}''$ | 3.3 |
| | $\mathscr{J}(\mathscr{I})$ | intervals of placements reachable by a jump from a placement in $\mathscr{I}$ | 3.3 |
| reachable placements | $\mathscr{R}$ | | 3.3 |
| | $\Lambda$ | list of oriented placement, natural number of pairs | 4.2 |
| | $\sigma, \rho, \tau$ | initial, penultimate and final oriented placements in $\Lambda$ | 4.2 |
| | $P$ | path corresponding to $\Lambda$ | 4.2 |
| | $\theta$ | final placement on $P$ | 4.2 |
| | $A_\sigma, A_\rho, A_\tau$ | initial, penultimate and final arcs traversed by $P$ | 4.2 |
| | $T$ | tangent segment of jump $(\rho, \tau)$ | 4.2 |
| | $\phi$ | direction of $T$ | 4.2 |
| | $u_\sigma, u_\rho, u_\tau$ | initial, penultimate and final contact on $P$ | 4.2 |
| | $r(\rho, \alpha)$ | ray tangent to $C_\rho$ in direction $\alpha$ | 4.2 |
| | $S_j$ | interval of final placements determined by *stretch(j)* | 4.3 |
| | $P_j$ | family of paths determined by *stretch(j)* | 4.3 |
| | $P_j(\theta)$ | path in $P_j$ with final placement $\theta$ | 4.3 |
| | $d$ | distance between contacts $a$ and $b$ | 4.4 |
| hop function | $h_d$ | maps initial to final angle of *RsLs* hop | 4.4 |
| hop function | $\bar{h}_d$ | maps initial to final angle of *LsRs* hop | 4.4 |
| | $H_d(\theta)$ | interval $[\bar{h}_d(\theta), h_d(\theta)]$ | 4.4 |
| | $\alpha_d, \beta_d$ | endpoints of dom$(h_d)$, i.e., dom$(h_d) = [-\alpha_d, \beta_d]$, dom$(\bar{h}_d) = [-\beta_d, \alpha_d]$ | 4.4 |
| | $\delta_d, \phi_e, \phi_f, \theta_e, \theta_f, \eta$ | angles in fig. 14 | 4.4 |
| | $\kappa_b$ | defined by theorem 4.5.1 | 4.5 |
| | $\kappa_c$ | minimum intercorner distance | 4.5 |
| | $\kappa_s$ | minimum of $\kappa_b$ and root separation | 5 |
| | $T(j, I)$ | interval of placements containing root or self-dual interval | 5 |
| | $\mathscr{T}(j, I)$ | extension of $T$ to sets of intervals | 5 |
| | $g_s$ | mapping from initial to final placement for a leap sequence | 6.1 |
| | $D(\omega, s)$ | initial placements of homotopic paths | 6.1 |
| | $F$ | maps labeled intervals to labeled intervals at the same contact | 6.1 |
| | $\mathscr{F}$ | extensions of $F$ to sets of intervals | 6.1 |

## References

[1] A. Baker, *Transcendental Number Theory* (Cambridge University Press, 1975).
[2] E. Bender, A driving hazard revisited, SIAM Rev. 21(1) (1979) 136–138.
[3] J. Canny, A new method for robot motion planning and real geometry, *Proc. 28th IEEE Symp. on Foundations of Computer Science* (1987) pp. 39–48.
[4] E.J. Cockayne and G.W.C. Hall, Plane motion of a particle subject to curvature constraints, SIAM J. Control 43(1) (1975) 197–220.
[5] G. Collins, Quantifier elimination for real closed fields by cylindrical algebraic decomposition, *Automata Theory and Formal Languages 2nd GI Conf.*, Lecture Notes in Computer Science, vol. 33 (1975) pp. 134–183.
[6] R. Cooper, *Functions of Real Variables* (D. Van Nostrand, London, 1966).
[7] L.E. Dubins, On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents, Am. J. Math. 79 (1957) 497–516.
[8] H.I. Freedman and S.D. Riemenschneider, Determining the path of the rear wheels of a bus, SIAM Rev. 25(4) (1983) 561–567.
[9] J.P. Laumond, Feasible trajectories for mobile robots with kinematic and environment constraints, *Proceedings of Intelligent Autonomous Systems* (1986) pp. 346–354.
[10] C. Ó'Dúnlaing, Motion planning with inertial constraints, Algorithmica 2(4) (1987) 431–475.
[11] J. Reeds and L. Shepp, Optimal paths for a car that goes both forwards and backwards, unpublished manuscript.
[12] J. Reif, private communication.
[13] J. Reif and M. Sharir, Motion planning in the presence of moving obstacles, *Proc. 26th Annual Symp. on Foundations of Computer Science* (1985) pp. 144–154.
[14] J.T. Schwartz and M. Sharir, On the piano movers' problem. II. General techniques for computing topological properties of real algebraic manifolds, Adv. Appl. Math. 4 (1983) 298–351.
[15] J.E. Shigley, *Kinematic Analysis of Mechanics* (McGraw-Hill, New York, 1959).
[16] L. Weisner, *Introduction to the Theory of Equations*, 6th ed. (MacMillan, 1959).
[17] C.K. Yap, Algorithmic motion planning, in: *Advances in Robotics, vol. 1: Algorithmic and Geometric Aspects*, J.T. Schwartz and C.K. Yap (eds.) (Lawrence Erlbaum Ass., Hillsdale, NJ, 1986).