OPTIMAL PATHS FOR A CAR THAT GOES BOTH FORWARDS AND BACKWARDS

J. A. Reeds and L. A. Shepp

The path taken by a car with a given minimum turning radius has a lower bound on its radius of curvature at each point, but the path has cusps if the car shifts into or out of reverse gear. What is the shortest such path a car can travel between two points if its starting and ending directions are specified? One need consider only paths with at most 2 cusps or reversals. We give a set of paths which is *sufficient* in the sense that it always contains a shortest path and *small* in the sense that there are at most 68, but usually many fewer paths in the set for any pair of endpoints and directions. We give these paths by explicit formula. Calculating the length of each of these paths and selecting the (not necessarily unique) path with smallest length yields a simple algorithm for a shortest path in each case. These optimal paths or geodesics may be described as follows: If C is an arc of a circle of the minimal turning radius and S is a line segment, then it is sufficient to consider only certain paths of the form $CCSCC$ where arcs and segments fit smoothly, one or more of the arcs or segments may vanish, and where reversals, or equivalently cusps, between arcs or segments are allowed. This contrasts with the case when cusps are not allowed, where Dubins (1957) has shown that paths of the form CCC and CSC suffice.

1. Introduction. We want to find a shortest path in the plane with specified initial and final points and directions and with the further constraint that at each point the radius of curvature should be ≥ 1. This problem arose in a simple model for a robot cart which moves under computer control. The cart can shift into reverse and so the path is allowed to have cusps.

In an elegant paper, Lester Dubins (1957) solved the problem when the car cannot reverse and cusps are not allowed. Even in this case it is apparently impossible to give an explicit formula for the shortest path. Instead Dubins gives a sufficient set of paths, i.e. a set which always contains what he called a *geodesic*, or optimal path. His sufficient set is so small that there are at most 6 contenders in the set for each case of specified endpoint conditions, and it is a simple matter to find the shortest of these 6, which gives an algorithm for the solution. He showed that any geodesic can be described by one of 6 words: $\ell r \ell$,

$\ell s\ell$, ℓsr , $r\ell r$, rsr , $rs\ell$ where ℓ , r , and s stand for "go left", "go right", and "go straight", respectively. Here left and right mean anticlockwise or clockwise around a *circle of unit radius*, i.e. a tightest possible circle, and of course one always goes less than 2π around any circle. More compactly, Dubins proved that a geodesic must be a smooth curve that is piecewise circular (radius 1) or linear, with at most 3 pieces, and always takes the form CCC or CSC where C is an arc of a unit circle and S is a line segment. A *word* notation like CCC or ℓsr thus stands for the corresponding class of paths. We use subscripts on a word (as in $\ell_t s_u r_v$) to specify the length of the corresponding arcs or segments involved. Note that one or more of these lengths may vanish. For example, to choose a path which returns to the initial point but in the opposite direction, two competing paths of Dubins type suggest themselves: $\ell_{3\pi/2} s_2 \ell_{3\pi/2}$ and $\ell_{\pi/3} r_{5\pi/3} \ell_{\pi/3}$ (see Figs. A, B). It is easy to verify that both of these accomplish the job
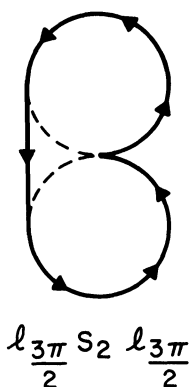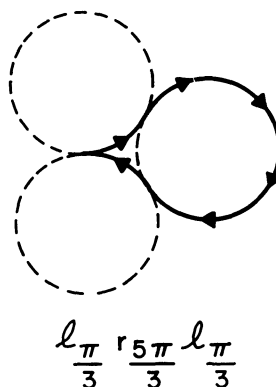

$\ell_{\frac{3\pi}{2}} s_2 \ell_{\frac{3\pi}{2}}$

FIGURE A

$\ell_{\frac{\pi}{3}} r_{\frac{5\pi}{3}} \ell_{\frac{\pi}{3}}$

FIGURE B

$r_{\frac{\pi}{3}} \ell_{\frac{5\pi}{3}} r_{\frac{\pi}{3}}$

FIGURE C

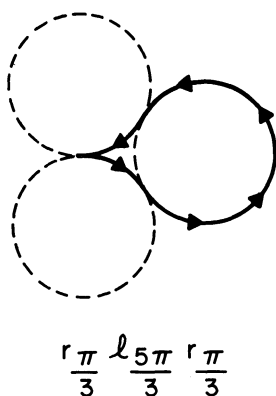$\ell^{+}_{\frac{\pi}{3}} r^{-}_{\frac{\pi}{3}} \ell^{+}_{\frac{\pi}{3}}$
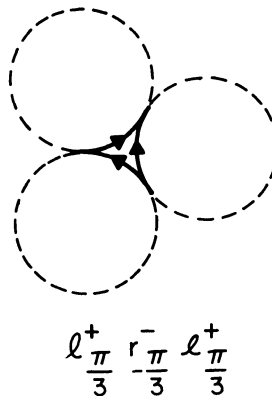
FIGURE D

of reversing in place but the second is shorter and turns out to be optimal by Dubins's theorem, as is the symmetrically reflected path, $r_{\pi/3}\ell_{5\pi/3}r_{\pi/3}$ (see Fig. C). Note that in a crude dimensional sense, the number of free parameters t, u, v exactly matches the number (3) of conditions of endpoint and direction. For each of the words $\ell s\ell$, ℓsr, rsr and $rs\ell$ it is not hard to show that there is at most one path obeying the end conditions.

Although there may actually be two distinct paths of form $\ell_t r_u \ell_v$ or of form $r_t \ell_u r_v$, Dubins shows [4, p. 513, Sublemma] that only one of them, with $u > \pi$, can be a geodesic.

With this fact it is not hard to show that there is at most one geodesic for each word, although there may be two different words which are both optimal, as in the reverse-in-place example.

Dubins has given an effective algorithm for the forward problem, but what if the car can reverse?

We remark that A. A. Markov in 1887 (Markov (1887), see also [Krein-Nudelman, p. 17]) considered and solved various related versions of these problems in his work on laying railroad track connecting already existing sections of track. Other papers of interest are Melzak (1961) and Dubins (1961).

If cusps are allowed in the path then we must consider words built from ℓ^+, ℓ^-, r^+, r^-, s^+, and s^-, where ℓ^+ means turning to the left while going forwards, ℓ^- means turning to the left while going backwards, etc. Note that a path of the form $\ell^+ r^-$, for example, has a cusp whereas $\ell^+ r^+$ or $\ell^- r^-$ has no cusp. (In car-driving terms, the letters ℓ, r, s refer to the steering wheel and the signs $+$ and $-$ refer to the gear shift.)

It is easy to see that cusps can sometimes shorten Dubins's paths; for example, in the reverse-in-place problem, the path $\ell_{\pi/3}^+ r_{\pi/3}^- \ell_{\pi/3}^+$ (see Fig. D) reverses in place and is shorter (in fact, it is optimal). Here the superscript indicates the direction taken along the corresponding arc or segment. We give a set of words in $\ell^\pm$, $r^\pm$, $s^\pm$ which give a solution to the reverse problem analogous to that of Dubins for the forward problem. These are more compactly given in C, S notation:

$$(1.1) \quad C^+ C^- C^+, \quad C^+ C^- C^-, \quad C^+ S^+ C^+, \quad C^+ C_u^+ C_u^- C^-,$$
$$C^+ C_u^- C_u^- C^+, \quad C^- C_{\pi/2}^+ S^+ C^+, \quad C^- C_{\pi/2}^+ S^+ C_{\pi/2}^+ C^-,$$
$$C^+ C^+ C^-, \quad C^- S^+ C_{\pi/2}^+ C^+$$

together with the words obtained by reversing all the signs. Here C

stands for either ℓ or r so CC means either ℓr or $r\ell$. A $C_{\pi/2}^+$ or $C_{-\pi/2}^-$ means the corresponding ℓ or r must be of length $\pi/2$, and the combination $C_u C_u$ in (1.1) means the two corresponding circular segments have equal lengths. For more precision we will use the somewhat redundant convention that in $C_t^\pm$, $\ell_t^\pm$, $r_t^\pm$, or $S_t^\pm$ the sign of t should match the direction, i.e. $t > 0$ when the car is going forward, $t < 0$ when it is reversing a distance $|t|$ along a left-circle, right-circle, or straight line. Note that in a family of paths such as $C_t^+ C_u^- C_v^- C_w^+$ the number (4) of free parameters t, u, v, w is one more than the number (3) of end conditions, and so there typically is a manifold of solutions, t, u, v, w for given end conditions. Optimizing among the paths of the family gives an extra equation such as $v = u$, or $v = \pi/2$, so that in each case in (1.1) only 3 free parameters remain. Of course any length not $\pi/2$ can vanish, so subwords are included as well. We will show (in the remark below the proof of Lemma 3) that there are paths—for example, in the class $C^+ C^- C^+ C^- C^+ C^-$— which are not of any of the above types and which are geodesics but in all such cases there are equally short geodesics inside our class (1.1). This phenomenon of geodesics that are not in the sufficient set does not occur in the simpler forward problem. In a still more compact notation that avoids $\pm$, we may write the list of sufficient special paths as

$$(1.2) \quad C|C|C, \quad CC|C, \quad CSC, \quad CC_u|C_uC, \quad C|C_uC_u|C,$$
$$C|C_{\pi/2}SC, \quad C|C_{\pi/2}SC_{\pi/2}|C, \quad C|CC, \quad CSC_{\pi/2}|C$$

where $|$ means reverse direction. There are 48 different words in $\ell^\pm$, $s^\pm$, $r^\pm$ when C in (1.1) or (1.2) is replaced by $C = \ell$ or $C = r$. Some of these 48 words have 2 formulas for an actual path of its word type. There are at most 68 formulas in any given case. Table 1 summarizes the 48 words and 68 formulas.

Dubin's proof is different from ours. He shows that there are geodesics for any endpoint conditions, i.e. the infimum is achieved, and then proves the lemma that any geodesic of length less than $\pi/8$ must be a CSC. It then follows easily that every geodesic must be a finite word in C and S, and then using a series of special arguments reduces all finite words to CCC or CSC.

We use advanced calculus to deduce our result from Dubin's theorem and then in §7, we use the same general method to outline a separate proof of Dubin's theorem itself. In fact, we do not see how

TABLE 1

This table lists the 48 words in our sufficient set, together with their shorthand names as used in (1.1) and (1.2). The last column gives the segment length formulas for the given word.

explicit	(1.1) form	(1.2) form	Section 8 formula
$l^+r^-l^+$	$C^+C^-C^+$	$C\mid C\mid C$	(8.3), two roots
$l^-r^+l^-$	$C^-C^+C^-$	$C\mid C\mid C$	(8.3), two roots
$r^+l^-r^+$	$C^+C^-C^+$	$C\mid C\mid C$	(8.3), two roots
$r^-l^+r^-$	$C^-C^+C^-$	$C\mid C\mid C$	(8.3), two roots
$l^+r^-l^-$	$C^+C^-C^-$	$C\mid CC$	(8.4), two roots
$l^-r^+l^+$	$C^-C^+C^+$	$C\mid CC$	(8.4), two roots
$r^+l^-r^-$	$C^+C^-C^-$	$C\mid CC$	(8.4), two roots
$r^-l^+r^+$	$C^-C^+C^+$	$C\mid CC$	(8.4), two roots
$l^-r^-l^+$	$C^-C^-C^+$	$CC\mid C$	(8.4), two roots
$l^+r^+l^-$	$C^+C^+C^-$	$CC\mid C$	(8.4), two roots
$r^-l^-r^+$	$C^-C^-C^+$	$CC\mid C$	(8.4), two roots
$r^+l^+r^-$	$C^+C^+C^-$	$CC\mid C$	(8.4), two roots
$l^+r_u^+l_{-u}^-r^-$	$C^+C_u^+C_{-u}^-C^-$	$CC_u\mid C_uC$	(8.7), two roots
$l^-r_{-u}^-l_u^+r^+$	$C^-C_{-u}^-C_u^+C^+$	$CC_u\mid C_uC$	(8.7), two roots
$r^+l_u^+r_{-u}^-l^-$	$C^+C_u^+C_{-u}^-C^-$	$CC_u\mid C_uC$	(8.7), two roots
$r^-l_{-u}^-r_u^+l^+$	$C^-C_{-u}^-C_u^+C^+$	$CC_u\mid C_uC$	(8.7), two roots
$l^+r_{-u}^-l_{-u}^-r^+$	$C^+C_{-u}^-C_{-u}^-C^+$	$C\mid C_uC_u\mid C$	(8.8)
$l^-r_u^+l_u^+r^-$	$C^-C_u^+C_u^+C^-$	$C\mid C_uC_u\mid C$	(8.8)
$r^+l_{-u}^-r_{-u}^-l^+$	$C^+C_{-u}^-C_{-u}^-C^+$	$C\mid C_uC_u\mid C$	(8.8)
$r^-l_u^+r_u^+l^-$	$C^-C_u^+C_u^+C^-$	$C\mid C_uC_u\mid C$	(8.8)
$l^+r_{-\pi/2}^-s^-l^-$	$C^+C_{-\pi/2}^-S^-C^-$	$C\mid C_{\pi/2}SC$	(8.9)
$l^-r_{\pi/2}^+s^+l^+$	$C^-C_{\pi/2}^+S^+C^+$	$C\mid C_{\pi/2}SC$	(8.9)
$r^+l_{-\pi/2}^-s^-r^-$	$C^+C_{-\pi/2}^-S^-C^-$	$C\mid C_{\pi/2}SC$	(8.9)
$r^-l_{\pi/2}^+s^+r^+$	$C^-C_{\pi/2}^+S^+C^+$	$C\mid C_{\pi/2}SC$	(8.9)
$l^-s^-r_{-\pi/2}^-l^+$	$C^-S^-C_{-\pi/2}^-C^+$	$CSC_{\pi/2}\mid C$	(8.9)
$l^+s^+r_{\pi/2}^+l^-$	$C^+S^+C_{\pi/2}^+C^-$	$CSC_{\pi/2}\mid C$	(8.9)
$r^-s^-l_{-\pi/2}^-r^+$	$C^-S^-C_{-\pi/2}^-C^+$	$CSC_{\pi/2}\mid C$	(8.9)
$r^+s^+l_{\pi/2}^+r^-$	$C^+S^+C_{\pi/2}^+C^-$	$CSC_{\pi/2}\mid C$	(8.9)
$l^+r_{-\pi/2}^-s^-r^-$	$C^+C_{-\pi/2}^-S^-C^-$	$C\mid C_{\pi/2}SC$	(8.10)
$l^-r_{\pi/2}^+s^+r^+$	$C^-C_{\pi/2}^+S^+C^+$	$C\mid C_{\pi/2}SC$	(8.10)
$r^+l_{-\pi/2}^-s^-l^-$	$C^+C_{-\pi/2}^-S^-C^-$	$C\mid C_{\pi/2}SC$	(8.10)
$r^-l_{\pi/2}^+s^+l^+$	$C^-C_{\pi/2}^+S^+C^+$	$C\mid C_{\pi/2}SC$	(8.10)
$r^-s^-r_{-\pi/2}^-l^+$	$C^-S^-C_{-\pi/2}^-C^+$	$CSC_{\pi/2}\mid C$	(8.10)
$r^+s^+r_{\pi/2}^+l^-$	$C^+S^+C_{\pi/2}^+C^-$	$CSC_{\pi/2}\mid C$	(8.10)
$l^-s^-l_{-\pi/2}^-r^+$	$C^-S^-C_{-\pi/2}^-C^+$	$CSC_{\pi/2}\mid C$	(8.10)
$l^+s^+l_{\pi/2}^+r^-$	$C^+S^+C_{\pi/2}^+C^-$	$CSC_{\pi/2}\mid C$	(8.10)
$l^+s^+r^+$	$C^+S^+C^+$	CSC	(8.2)
$l^-s^-r^-$	$C^-S^-C^-$	CSC	(8.2)
$r^+s^+l^+$	$C^+S^+C^+$	CSC	(8.2)
$r^-s^-l^-$	$C^-S^-C^-$	CSC	(8.2)
$l^+s^+l^+$	$C^+S^+C^+$	CSC	(8.1)
$l^-s^-l^-$	$C^-S^-C^-$	CSC	(8.1)
$r^+s^+r^+$	$C^+S^+C^+$	CSC	(8.1)
$r^-s^-r^-$	$C^-S^-C^-$	CSC	(8.1)
$l^+r_{-\pi/2}^-s^-l_{-\pi/2}^-r^+$	$C^+C_{-\pi/2}^-S^-C_{-\pi/2}^-C^+$	$C\mid C_{\pi/2}SC_{\pi/2}\mid C$	(8.11), two roots
$l^-r_{\pi/2}^+s^+l_{\pi/2}^+r^-$	$C^-C_{\pi/2}^+S^+C_{\pi/2}^+C^-$	$C\mid C_{\pi/2}SC_{\pi/2}\mid C$	(8.11), two roots
$r^+l_{-\pi/2}^-s^-r_{-\pi/2}^-l^+$	$C^+C_{-\pi/2}^-S^-C_{-\pi/2}^-C^+$	$C\mid C_{\pi/2}SC_{\pi/2}\mid C$	(8.11), two roots
$r^-l_{\pi/2}^+s^+r_{\pi/2}^+l^-$	$C^-C_{\pi/2}^+S^+C_{\pi/2}^+C^-$	$C\mid C_{\pi/2}SC_{\pi/2}\mid C$	(8.11), two roots

to formulate a result analogous to his short geodesic lemma for the reverse problem, so his methods do not seem to work in a straightforward way. We show instead that any curve can be approximated by a word in $C^\pm$ and $S^\pm$. Then we show that any word in $C^\pm$, $S^\pm$ can be reduced to one in at most 5 letters without increasing its length. Finally we reduce to a word on the list by using ideas similar to the last part of Dubin's proof.

Although we give a rigorous proof of our assertions, we used a computer to empirically determine a sufficient list of words as follows: Given a set W of words, we tested W for insufficiency by generating the endpoint conditions randomly and first finding the best path in W. If a shorter path can be found by concatenating two paths in W, then W is insufficient. Using this method and pruning, we eventually arrived at and convinced ourselves that we had a minimal sufficient set.

Once we had guessed at W, we used the computer again to help do the extensive algebra in the large number of cases involved to verify that a rigorous proof could be given by the method outlined above. Finally, we found that the proof could be simplified (§2), so that it can easily be followed by an ordinary human without a computer to check the details. But we think that we could never have found the right set of words without using a computer.

In §7 we outline a proof of Dubin's theorem by our method.

In §8 we give a list of formulas to compute the lengths of each of the 68 actual path-solutions for each of the 48 word types suitable for algorithmic implementation.

2. Admissible paths. For us, the state of a car at a given instant t, where t is arclength, is completely specified by its position $(x(t), y(t))$ in the plane. An *admissible path* or *curve* is a function $\gamma(t) = (x(t), y(t), \phi(t))$ for which we can find measurable functions ε and η for which

$$(2.1) \qquad x(t) = x(0) + \int_0^t \varepsilon(\tau) \cos \phi(\tau) \, d\tau,$$

$$y(t) = y(0) + \int_0^t \varepsilon(\tau) \sin \phi(\tau) \, d\tau, \quad \text{where}$$

$$\phi(t) = \phi(0) + \int_0^t \eta(\tau) \, d\tau$$

and where $\varepsilon(\tau) = \pm 1$ and $|\eta(\tau)| \leq 1$ for each τ. In words, a car can move only forwards or backwards in its own direction $\phi(\tau)$ with speed $\sqrt{\dot{x}^2 + \dot{y}^2}(\tau) = |\varepsilon(\tau)| = 1$ and cannot change its direction $\phi(\tau)$ faster than one radian per time unit so that its turning radius is at least one or the curvature of its path is at most one (since the curvature is the reciprocal of the turning radius).

Note that we must allow infinite acceleration at a cusp where $\varepsilon(t)$ changes sign instantly. The problem where the acceleration must satisfy $\ddot{x} + \ddot{y} \leq a < \infty$ is more difficult and is not treated here. However, for slowly moving vehicles, such as carts, this seems like a reasonable compromise to achieve tractability. How does one characterize the class of paths (x_t, y_t) which satisfy (2.1)? Any path with piecewise constant circles of radius ≥ 1 and/or line segments suffices, but one can take $\{t : \varepsilon(t) = 1\}$ to be an arbitrary measurable set. A somewhat complicated condition on (x_t, y_t), $0 \leq t \leq T$, is: It is assumed that t is length along the path, so that $\dot{x}_t^2 + \dot{y}_t^2 \equiv 1$. Then it is necessary and sufficient that for $\phi(t) = \tan^{-1}(\dot{y}_t / \dot{x}_t)$ we have

$$(2.2) \qquad |\phi(t + s) - \phi(t)| \leq s \quad \text{for } 0 \leq t \leq t + s \leq T.$$

This is an immediate consequence of the fact that a Lip 1 function ϕ as in (2.2) is the integral of its derivative η as in (2.1). Finally, $\varepsilon(\tau)$ is automatically uniquely defined by (2.1) since $\dot{x}_t^2 + \dot{y}_t^2 = 1$.

The *track* of an admissible curve $\gamma(t) = (x(t), y(t), \phi(t))$ is $\overline{\gamma}(t) = (x(t), y(t))$. By differentiating (2.1) we get

$$(2.3) \qquad \begin{aligned} \dot{x}(t) &= \varepsilon(t) \cos \phi(t), \\ \dot{y}(t) &= \varepsilon(t) \sin \phi(t), \\ \dot{\phi}(t) &= \eta(t), \end{aligned}$$

so both $\gamma(t)$ and $\overline{\gamma}(t)$ are rectifiable. If g is admissible with $g(t_0) = a$ and $g(t_1) = b$ with $t_0 < t_1$ we call its restriction $\gamma = g|_{[t_0, t_1]}$ to the domain $t_0 \leq t \leq t_1$ an *admissible path leading from a to b* and define *its length* $L(\gamma) = t_1 - t_0$. Our problem in short: given an arbitrary a and b in $\mathbb{R}^3$ find an admissible γ minimizing $L(\gamma)$.

3. Summary of results. A *word* is a finite string in the letters C, S, $|$ and with some abuse of notation is also thought of as a path or as a set of paths. Each path in $C|CS$, for example, starts somewhere, goes along a circle of radius 1 for some distance ≥ 0, then has a cusp, then goes along the other circle tangent to the first circle at the

cusp in the opposite direction for some distance ≥ 0, then continues along the straight line, for some distance ≥ 0, which is tangent to the second circle. A path in $C|CS$ is a member of $C^+C^-S^-$ or $C^-C^+S^+$ according as the initial direction is forward or reverse. A path of $C^+C^-S^-$ is a member of $\ell^+r^-s^-$ or $r^+\ell^-s^-$ according as the initial rotation is counterclockwise or clockwise. A word is said to have *proper* form if each of the durations of its segments or letters is nonzero.

A word of *special* form is one of those in (1.1), or (1.2). We say a word can be shortened to another word if for every path of the form of the first word there is a shorter path of the form of the second word with the same endpoint conditions of position and direction.

Our first result, proved in §4, is the hardest.

THEOREM 1. *Any word with 3 cusps can be shortened to a word with two or fewer cusps.*

The obvious induction then yields the following.

COROLLARY. *Any word can be shortened to a word with two or fewer cusps.*

By applying Dubins's theorem to the parts of the word between consecutive cusps or endpoints (where the path is always in the same direction) we see that any word with two or fewer cusps can be shortened to a word with two or fewer cusps and with 9 or fewer segments, or letters C and S. In fact even more can be done:

THEOREM 2. *Any word can be shortened to a word of special form* (1.3).

This is proved in §5. Finally a simple approximation argument in §6 shows these results apply to all admissible paths, not just to paths described by words. Thus we obtain:

THEOREM 3. *Any admissible path can be shortened to a word of special form* (1.2)

The general technique for proving Theorems 1 and 2 is as follows: Given a four segment word γ, that is, a word with 4 letters C and S, consider the four dimensional family of all words of the same type

as γ with the same starting point and starting direction. Typically a one dimensional subfamily will also share final endpoint and final direction with γ, as can be seen by counting constraint equations. Using calculus and setting derivatives to zero we can find the shortest path in this subfamily; its four durations will obey a critical point equation. Thus if γ is not strictly shortenable in its class its four durations or segment lengths must satisfy the appropriate critical point equations. If a word w with more than 4 segments is not strictly shortenable within its type, this process can be repeated on each 4 segment subword to yield a series of relations between the segment lengths. These relations typically force w to have a very constrained shape, which enable further deductions about w to be made.

The following notations are helpful. For $t \in \mathbb{R}$ let L_t, R_t, S_t

$$(3.1) \qquad\qquad L_t : \mathbb{R}^3 \to \mathbb{R}^3,$$
$$R_t : \mathbb{R}^3 \to \mathbb{R}^3,$$
$$S_t : \mathbb{R}^3 \to \mathbb{R}^3$$

be the position and direction at time t of a unit circle, ℓ_t or r_t, and a line segment s_t, respectively, starting from (x, y, ϕ), so that

$$(3.2) \quad L_t(x, y, \phi) = (x + \sin(\phi + t) - \sin\phi, y - \cos(\phi + t) + \cos\phi, \phi + t),$$
$$R_t(x, y, \phi) = (x - \sin(\phi - t) + \sin\phi, y + \cos(\phi - t) - \cos\phi, \phi - t),$$
$$S_t(x, y, \phi) = (x + t\cos\phi, y + t\sin\phi, \phi).$$

Then a curve $\ell_r^+ r_u^- s_v^- \ell_w^-$ starting at $(0, 0, 0)$ for instance must end at

$$(3.3) \quad (X(t, u, v, w), Y(t, u, v, w), \Phi(t, u, v, w))$$
$$= R_w(S_v(R_u(L_t(0, 0, 0))))$$

and its length is

$$(3.4) \quad L(t, u, v, w) = |t| + |u| + |v| + |w| = t - u - v - w.$$

Let ∇X be the gradient of $X(t, u, v, w)$, i.e.

$$\nabla X = \left(\frac{\partial X}{\partial t}, \frac{\partial X}{\partial u}, \frac{\partial X}{\partial v}, \frac{\partial X}{\partial w} \right)$$

and similarly for ∇Y, $\nabla \Phi$, and ∇L. For each word, formulas for ∇X, ∇Y, $\nabla \Phi$, and ∇L can be obtained.

For example, the word $\ell_t r_u s_v \ell_w$ starting at $(0, 0, 0)$ has, by a short calculation and regardless of signs of t, u, v, w

$$(3.5) \qquad X(t, u, v, w) = \sin(w - u + t) + v \cos(u - t)$$
$$+ 2 \sin(u - t) + 2 \sin t,$$
$$Y(t, u, v, w) = -\cos(w - u + t) - v \sin(u - t)$$
$$+ 2 \cos(u - t) - 2 \cos t + 1,$$
$$\Phi(t, u, v, w) = w - u + t,$$

and ∇X, ∇Y, $\nabla \Phi$ can be easily obtained.

Similarly, the word $\ell_t r_u s_v r_w$ starting at $(0, 0, 0)$ has

$$(3.6) \quad X(t, u, v, w) = \sin(w + u - t) + v \cos(u - t) + 2 \sin t,$$
$$Y(t, u, v, w) = \cos(w + u - t) - v \sin(u - t) - 2 \cos t + 1,$$
$$\Phi(t, u, v, w) = t - u - w.$$

4. At most 2 cusps suffice. Here we show that any word with 3 or more cusps can be shortened to a word with ≤ 2 cusps. We start with some lemmas:

LEMMA 1. *A word of proper form $C|S$ can always be strictly shortened to a word with zero or one cusps.*

Proof. Suppose γ is of the form $\ell_T^+ s_V^-$ where $T > 0$ and $V < 0$ and leads from $(0, 0, 0)$ to (x, y, ϕ). This is a special case of the 4-word $\ell_t^+ r_u^- s_v^- \ell_w^-$ where $u = w = 0$. We will show how to shorten γ by a word of this type. Consider the following system of ordinary differential equations with independent variable h:

$$(4.1) \qquad \dot{t}(h) = -\frac{1}{2 \cos u(h)},$$
$$\dot{u}(h) = -\frac{v(h) + 2 \sin u(h)}{2 v(h) \cos u(h)},$$
$$\dot{v}(h) = 1 + \frac{2 \sin u(h)}{v(h) \cos u(h)},$$
$$\dot{w}(h) = -\frac{\sin u(h)}{v(h) \cos u(h)}.$$

By Picard's existence and uniqueness theorems, [2, p. 12] this system has a solution for small enough values of h, which passes through the point $(t(0), u(0), v(0), w(0)) = (T, 0, V, 0)$. Let $\gamma(h)$ denote

the corresponding curve $\ell_{t(h)} r_{u(h)} s_{v(h)} \ell_{w(h)}$. It is easy to check (using (3.4) and (3.5)) that $\gamma(h)$ leads from $(0, 0, 0)$ to (x, y, ϕ) and that $L'(h) = -1$. It is also easy to see from (4.1) that $\dot{u}(h) < 0$ and $\dot{w}(h) \leq 0$ for all $h > 0$ sufficiently small, $t(h) > 0$, $u(h) \leq 0$, $v(h) < 0$, $w(h) \leq 0$ so that $\gamma(h)$ remains in the class $\ell^+ r^- s^- \ell^-$. For all h sufficiently small $\gamma(h)$ shortens γ (since $L'(h) < 0$) and $\gamma(h)$ has only one cusp. $\qquad\square$

LEMMA 2. *Suppose t, u, v, and w are all nonzero.*

(a) *Suppose $\lambda = C_t C_u | C_v S_w$ is not strictly shortenable in its class $CC|CS$. Then $u \equiv -v \pmod{2\pi}$ or $u = v + \pi \pmod{2\pi}$.*

(b) *Suppose $\lambda = S_t C_u | C_v S_w$ is not strictly shortenable in its class $SC|CS$. Then $u \equiv -v \pmod{2\pi}$ or $u = v + \pi \pmod{2\pi}$.*

(c) *Suppose $\lambda = C_t | C_u S_v C_w$ is not strictly shortenable in its class $C|CSC$. Then $v = 0$ or $u \equiv \pi/2 \pmod{2\pi}$ or $u \equiv -\pi/2 \pmod{2\pi}$. This also holds for $w = 0$, i.e., $\lambda = C_t | C_u S_v$ can be shortened inside the class $C|CSC$.*

(d) *Suppose $\lambda = C_t | C_u | C_v S_w$ is not strictly shortenable in its class $C|C|CS$. Then $u \equiv 0 \pmod{2\pi}$ or $u \equiv \pi \pmod{2\pi}$.*

(e) *Suppose $\lambda = C_t C_u C_v | C_w$ is not strictly shortenable in its class $CCC|C$. Then $u \equiv 2v + \pi \pmod{2\pi}$.*

(f) *Suppose $\lambda = C_t | C_u | C_v C_w$ is not strictly shortenable in its class $C|C|CC$. Then $u \equiv 0 \pmod{2\pi}$ or $u \equiv \pi \pmod{2\pi}$.*

(g) *Suppose $\lambda = C_t | C_u C_v | C_w$ is not strictly shortenable in its class $C|CC|C$. Then $u \equiv v \pmod{2\pi}$ or $u \equiv v + \pi \pmod{2\pi}$.*

(h) *Suppose $\lambda = C_t C_u | C_v C_w$ is not strictly shortenable in its class $CC|CC$. Then $u \equiv -v \pmod{2\pi}$ or $u \equiv v + \pi \pmod{2\pi}$.*

Proof. (a) Suppose λ is a path from $(0, 0, 0)$ to (x, y, ϕ) of type $\ell^+ r^+ \ell^- s^-$. Let F be the set of all paths of type $\ell^+ r^+ \ell^- s^-$ from $(0, 0, 0)$ to (x, y, ϕ); F can be identified with the set of all tuples (t, u, v, w) for which

$$(4.2) \qquad t \geq 0, \quad u \geq 0, \quad v \leq 0, \quad w \leq 0,$$

$$X(t, u, v, w) = x, \quad Y(t, u, v, w) = y, \quad \Phi(t, u, v, w) = \phi,$$

where

$$(4.3) \qquad (X(t, u, v, w), Y(t, u, v, w), \Phi(t, u, v, w))$$
$$= S_w(L_v(R_u(L_t(0, 0, 0)))).$$

Every point (t, u, v, w) in F obeys one or more of the following:

(1) One or more of the components of t, u, v, w vanishes.

(2) The linear space spanned by ∇X, ∇Y, and $\nabla \Phi$ has dimension less than 3.

(3) No component of (t, u, v, w) vanishes and the span of $(\nabla X, \nabla Y, \nabla \Phi)$ has dimension 3, so some neighborhood of (t, u, v, w) in F is a 1-manifold.

Now $\lambda = \ell_t^+ r_u^+ \ell_v^- s_w^-$ minimizes $L(\lambda) = t + u - v - w$ in F. By hypothesis, none of t, u, v, w vanish so either (2) or (3) holds. In case (3) the method of Lagrange multipliers ([3, p. 150], [6, p. 129]) shows there are reals λ_x, λ_y, and λ_ϕ so that

(4.4) $\nabla L = \lambda_x \nabla X + \lambda_y \nabla Y + \lambda_\phi \nabla \Phi.$

In both cases (2) and (3), then, the space spanned by ∇X, ∇Y, $\nabla \Phi$, and ∇L has dimension at most 3, and hence the Jacobian determinant

$$\Delta = \left| \frac{\partial(X, Y, \Phi, L)}{\partial(t, u, v, w)} \right|$$

vanishes. Elementary calculus shows that

(4.5) $\Delta = 4 \sin u + 4 \sin v$

so the conclusion of (a) follows.

Symmetry considerations show that if λ is of some other $CC|CS$ type (such as $\ell^- r^- \ell^+ s^+$ or $r^- l^- r^+ s^+$, etc) the same results hold.

The proof of the other claims (b)–(h) of Lemma 2 are mutatis mutandis the same. A trivial complication arises in (c), where both $\ell^+ r^- s^- r^-$ and $\ell^+ r^- s^- \ell^-$ must be checked.   □

LEMMA 3. *A word of the form $C|C|C|C$ can always be (not necessarily strictly) shortened to a word with at most two cusps.*

Proof. Let F be the set of all paths of the form $\ell_t^+ r_u^- \ell_v^+ r_w^-$ from $(0, 0, 0)$ to a given (x, y, ϕ). On connected components of F, the length function $L(t, u, v, w) = t - u + v - w$ is constant, congruent modulo 2π to ϕ. Consider the new problem of minimizing w on F. As in the proof of Lemma 2, either some component of (t, u, v, w) vanishes (yielding a word with at most 2 cusps) or the Jacobian determinant

$$\Delta = \left| \frac{\partial(X, Y, \Phi, w)}{\partial(t, u, v, w)} \right|$$

vanishes. Elementary calculation shows

$$(4.6) \qquad\qquad \Delta = 4 \sin u$$

so that $u = 0$ or π. If $u = 0$ the word is of the form $\ell_t^+ \ell_v^+ r_w^- = \ell_{t+v}^+ r_w^-$ which has ≤ 1 cusps. If $u = \pi$ then the sign of the second arc may be reversed since an arc of length π traversed in one direction is equivalent to a traversal of the other semicircle in the reverse direction (see Fig. 3). But this says that there is a curve of the form $\ell_t^+ r_\pi^+ \ell_v^+ r_w^-$ leading from $(0, 0, 0)$ to (x, y, ϕ) with fewer cusps. $\qquad\square$

We remark that if L is less then π then $C|C|C|C$ is optimal since $|\phi| = L$ and $|\Phi|$ continually increases along $C|C|C|C$. Similarly $C|C|C|\cdots|C$ is optimal for any number of C's so long as $L \leq \pi$.

We now can prove Theorem 1 itself:

THEOREM 1. *Any word with three cusps can be shortened to a word with two or fewer cusps.*

Proof. Let θ be a word with 3 cusps. Consider the family G of all words θ' of form $\theta' = \alpha|\beta|\gamma|\delta$ where each of α, β, γ, and δ is of form CCC or CSC, where each segment in θ' has length $\leq L(\theta)$, and where θ' has the same end conditions as θ. By Dubins's theorem applied to the cusp-free subwords of θ there is a word θ' in G with $L(\theta') \leq L(\theta)$, so we may assume without loss of generality that θ is already in G. Since G is compact we may further assume θ minimizes $L(\theta)$ in G, so that θ is not strictly shortenable among words with ≤ 3 cusps. The general idea is that no subword of such a θ can be strictly shortened without increasing the number of cusps. This is very much like Dubins's use of the principle (in the context of his forward problem) that a subpath of a geodesic is a geodesic.

By Dubins's theorem each of α, β, γ, δ is one of these seven proper forms

$$(4.7) \qquad C, \quad CC, \quad CCC, \quad S, \quad CS, \quad SC, \quad CSC.$$

Since θ cannot be strictly shortened with 3 or fewer cusps, the combination $S|S$ cannot occur in θ, so if $\alpha = S$ or CS then β cannot be S or SC for example. By Lemma 1 the combinations $C|S$ and $S|C$ cannot occur in θ. These restrictions alone force α to have one of the proper forms

$$(4.8) \qquad\qquad C, \quad CC, \quad CCC, \quad SC, \quad CSC$$

and force β and γ to have one of the proper forms

$$(4.9) \qquad C, \quad CC, \quad CCC, \quad CSC$$

and force δ to have one of the proper forms

$$(4.10) \qquad C, \quad CC, \quad CCC, \quad CS, \quad CSC$$

which reduces the overall number of possible forms from $7^4 = 2401$ to $5 \cdot 4^2 \cdot 5 = 400$.

We may suppose that $SC|CS$ does not occur in θ: if $S_t C_u | C_v S_w$ did occur then Lemma 2 (b) and (a) says $|v| = \pi/2$ and $u \equiv -v$ $(\mathrm{mod}\, 2\pi)$ or $u = v + \pi \ (\mathrm{mod}\, 2\pi)$. Since u and v have opposite signs and since $|u| < 2\pi$ we know $u = -v$. Since $s_t \ell_{\pi/2} r_{-\pi/2} = \ell_{\pi/2} r_{-\pi/2} s_{-t}$ (see Fig. 1) we see that in this case $S_t C_u | C_v S_w = S_t S_{-w} C_u | C_v = S_{t-w} C_u | C_v$. Since $\ell_{\pi/2} r_{-\pi/2} = r_{\pi/2} \ell_{-\pi/2}$ (again Fig. 1) we further see that C_v may be merged into a C_h curve. Thus without changing the length of θ we may repeatedly replace all instances of the form $SC|CS$ with $SC|C$, or by symmetry, by $C|CS$.

Similarly, we may suppose that w has no instance of $CC|CS$. Arguing as before, applying Lemma 2 (a) and (c) to $C_t C_u | C_v S_w$ we

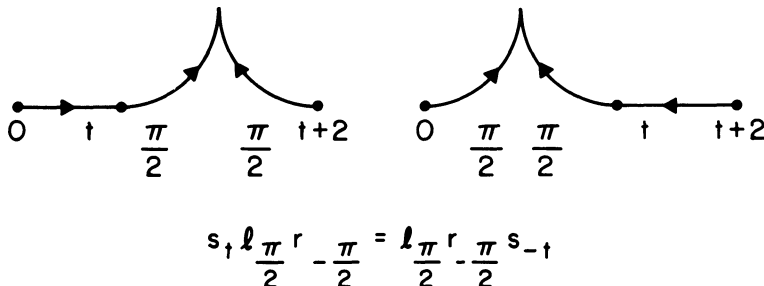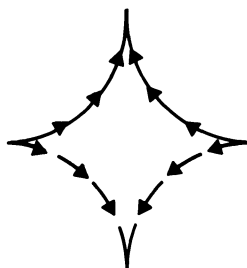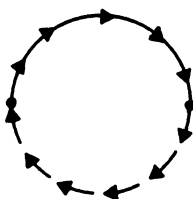$$s_t \ell_{\frac{\pi}{2}} r_{-\frac{\pi}{2}} = \ell_{\frac{\pi}{2}} r_{-\frac{\pi}{2}} s_{-t}$$

FIGURE 1

$$\ell_{\frac{\pi}{2}} r_{-\frac{\pi}{2}} = r_{\frac{\pi}{2}} \ell_{-\frac{\pi}{2}}$$

FIGURE 2

$$C_\pi = C_{-\pi}$$

FIGURE 3

see $u = \pm\pi/2$ and $v = -u$. Since $\ell_{\pi/2}r_{-\pi/2} = r_{\pi/2}\ell_{-\pi/2}$ (Fig. 2) we may reverse the handedness of $C_u|C_v$ which allows $C_tC_u|C_v$ to reduce to $C_{t+u}|C_v$. Thus all instances of $CC|CS$ and $CCC|CS$ get replaced by $C|CS$ and all instances of $SC|CC$ and $SC|CCC$ get replaced by $SC|C$ in θ.

Similarly, if $C_t|C_u|C_vS_w$ occurs in θ, then by Lemma 2 (d), $u \equiv \pi \pmod{2\pi}$. But $C_\pi = C_{-\pi}$ (Fig. 3) so in either case $C_t|C_u|C_vC_w = C_tC_{-u}C_vS_w$ which has strictly fewer cusps.

These simplifications together imply that we may replace θ by a word θ' with fewer cusps or by one in which neither β nor γ has the form CSC. In this last case, θ' contains one of these forms:
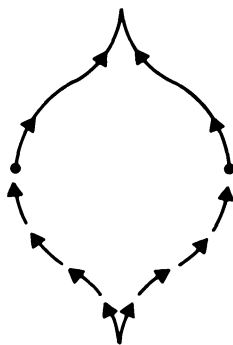
$$(4.11) \qquad C|C|CC, \quad CC|C|C, \quad C|CC|CC|C,$$
$$C|CCC|C, \quad C|C|C|C.$$

Each of *these* can be shortened or replaced by a word with fewer cusps, as follows:

By Lemma 2 (f), $C_t|C_u|C_vC_w$ is unshortenable only if $\sin u = 0$, i.e. $u = \pi$. Then $C_u = C_{-u}$ reduces the word as before. The case $CC|C|C$ is symmetrically dealt with.

The case $C_t|C_uC_v|C_wC_x|C_y$ is handled as follows. If $|u| > \pi$ the replacement $C_t|C_uC_v = C_tC_{2\pi-u}|C_v$ is a strict shortening with the same number of cusps. So we may assume $|u| \le \pi$ and similarly $|v| \le \pi$, $|w| \le \pi$, and $|x| \le \pi$. By Lemma 2 (g), strict shortening is possible unless $u \equiv v$ or $u = v+\pi$. But u and v have the same sign so $u = v$. Similarly, $w = x$ by Lemma 2 (g) and $v = -w$ by Lemma 2 (h). But as we see in Fig. 4, and is easily verified, $\ell_ur_u\ell_{-u}r_{-u} = \ell_{-u}r_{-u}\ell_ur_u$ so $C_t|C_uC_v|C_wC_x|C_y = C_t|C_uC_u|C_{-u}C_{-u}|C_y = C_tC_{-u}C_{-u}|C_uC_uC_y$ which has fewer cusps.

The pattern $C_t|C_uC_vC_w|C_x$ similarly can be supposed to have $|u| \le \pi$, $|w| \le \pi$. By Dubins's theorem [4, p. 513, Sublemma], $\pi < |v| < 2\pi$. By Lemma 2 (e), $v \equiv 2u + \pi \pmod{2\pi}$ and

$$\ell_u r_u \ell_{-u} r_{-u} = \ell_{-u} r_{-u} \ell_u r_u$$

FIGURE 4

$v \equiv 2w + \pi \pmod{2\pi}$, so $|u| < \pi/2$ and $|w| < \pi/2$. Because $\ell_u r_{2u+\pi} \ell_u = \ell_{-u} r_{\pi-2u} \ell_{-u}$ (see Fig. 5) we see $C_t | C_u C_v C_w | C_x$ is strictly shortenable to $C_t C_{-u} | C_h | C_{-w} C_x$, where $h = \pi - 2u$ since $\pi - 2|u| < \pi + 2|u|$.

Finally, Lemma 3 takes case of the $C|C|C|C$ case.   □

By applying Theorem 1 to subwords of an arbitrary word we can immediately obtain the following.

COROLLARY. *Any word can be shortened to a word with two or fewer cusps.*

REMARK. The reader may be wondering why we cannot simply get rid of any subword $C_u C_v C_w$ where u, v, w are nonzero and of the same sign. He may want to reason that by Dubins's lemma [4, p. 513, Sublemma], $\pi < |v| < 2\pi$ but reversing the middle segment lets us replace $C_u C_v C_w$ by $C_u | C_{2\pi - v} | C_w$ which is shorter. The trouble is this introduces two more cusps. But in Theorem 1 we want to keep reducing the number of cusps so such a cusp-increasing step is not permitted. However, now that Theorem 1 is proved we can use this lemma of Dubins to further reduce the sufficient set of words and eliminate CCC's and this will be done as part of the next step in the reduction to (1.2).

5. Words of special form suffice. Here we extend the conclusion of Theorem 1.

THEOREM 2. *Any word can be shortened to a word of the special form* (1.2).
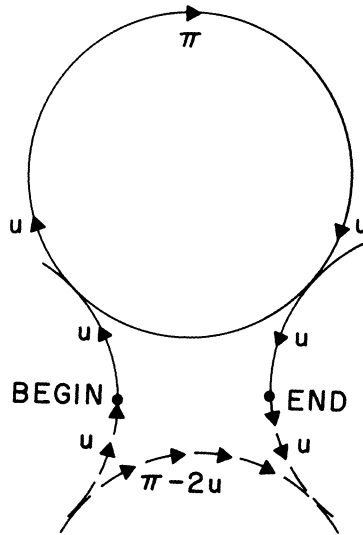
$$\ell_u{}^r 2u + \pi\, \ell_u = \ell_{-u}{}^r \pi - 2u\, \ell_{-u}$$

FIGURE 5

Proof. We observe that by Theorem 1 and Dubins's theorem, the set of words of the form $\alpha|\beta|\gamma$ where α, β, γ are each one of C, S, CC, CS, SC, CSC, CCC suffices. We narrow $\alpha|\beta|\gamma$ down to (1.1) or (1.2) using some lemmas.

LEMMA 4. *A proper word of the form $C_t|C_u S_v$ which is not shortenable to any word with ≤ 1 cusps must have $u = \pm\pi/2$.*

Proof. Suppose $\gamma = \ell_T r_U s_V$ where $T > 0$, $U < 0$, and $V < 0$ is not shortenable to any word with one or fewer cusps. Consider the system of ordinary differential equations used in the proof of Lemma 1, but this time with boundary condition

(5.1) $(t(0), u(0), v(0), w(0)) = (T, U, V, 0).$

If $V \neq 0$ and $\tan U < 0$ then Picard's theorem [2, p. 12] applies and since $\dot{w}(h)$ in (4.1) < 0, $w(h) < 0$ and so

(5.2) $\gamma(h) = \ell_{t(h)} r_{u(h)} s_{v(h)} \ell_{w(h)}$

for sufficiently small $h > 0$ strictly shortens γ (since $L'(h) = -1$) and has only one cusp (since $w(h) < 0$ and we remain inside the family $\ell^+ r^- s^- \ell^-$). If $\tan U \geq 0$, on the other hand, then the system

of ordinary differential equations, analogous to (4.1),

$$(5.3) \qquad \dot{t}(h) = -\frac{1}{2\cos u(h)},$$

$$\dot{u}(h) = -\frac{v(h) + 2\sin u(h)}{2v(h)\cos u(h)},$$

$$\dot{v}(h) = 1,$$

$$\dot{w}(h) = \frac{\sin u(h)}{v(h)\cos u(h)}$$

with initial conditions (5.1) has by Picard's theorem, a solution for sufficiently small h which satisfies $\dot{w}(h) < 0$ since $u(0) = U$ and $\tan U > 0$ and $v(0) = V < 0$ and so $w(h) < 0$ for small h, since $w(0) = 0$. Using (3.6) it is easy to verify that for all $h > 0$, the endpoint $(X(h), Y(h), \Phi(h))$ of the curve

$$(5.4) \qquad \gamma(h) = \ell_{t(h)}r_{u(h)}s_{v(h)}r_{w(h)}$$

is independent of h, i.e. $dX/dh = dY/dh = d\Phi/dh = 0$ and so $\gamma(h)$ remains with the family $\ell^+ r^- s^- r^-$. It may be verified that $L'(h) = -1$ where $L(h) = |t(h)| + |u(h)| + |v(h)| + |w(h)| = t(h) - u(h) - v(h) - w(h)$ and so $\gamma(h)$ for sufficiently small h has only one cusp and strictly shortens $C_T|C_U S_V$ and has the same endpoints, when $\tau U \geq 0$. The only other case is $U = \pm\pi/2$ which is excluded by the hypothesis of Lemma 4. $\qquad\square$

LEMMA 5. *A word of the form $CC|CC|C$ can always be strictly shortened to a word with at most 2 cusps.*

Proof. This is the only case where a second order analysis seems to be required, although in proving Dubins's theorem a second order analysis appears also to be required [4, Lemma 2] although it was not emphasized there that this was the method being used.

Thus consider the function $L(t, u, v, w, z)$ which is the length of an admissible $C_t C_u | C_v C_w | C_z$. We may assume without loss that the path is $\ell_t^+ r_u^+ \ell_v^- r_w^- \ell_z^+$ so that $L(t, u, v, w, z) = t + u - v - w + z$. From Lemma 2(h) we learn that $v = -u$ or else $\ell_t r_u \ell_v^- r_w^-$ can be shortened inside $CC|CC$ and so of course the full path can also be shortened. Similarly from Lemma 2(g) we learn $W = V$ or else $r_u^+ l_v^- r_w^- \ell_z^+$ can be shortened. We will show that for $0 < u < \pi/2$ that $v = w = -u$ is a *saddle point* rather than a *minimum* value

of L under the condition that t, u, v, w, z satisfy the endpoint conditions. For $\pi/2 \le u$ we will argue directly that the length can be shortened without increasing the number of cusps.

Case 1. $0 < u < \pi/2$.

The conditions on t, u, v, w, z under which the path $\ell_t r_u \ell_v r_w \ell_z$ arrives at (x, y, ϕ) are:

$$(5.5) \quad \begin{aligned} x &= \sin(z - w + v - u + t) + 2\sin(w - v + u - t) \\ &\quad + 2\sin(v - u + t) + 2\sin(u - t) + 2\sin t, \\ y &= 1 - \cos(z - w + v - u + t) + 2\cos(w - v + u - t) \\ &\quad - 2\cos(v - u + t) + 2\cos(u - t) - 2\cos t, \\ \phi &= z - w + v - u + t. \end{aligned}$$

Since we can assume all of t, u, v, w, z are nonzero, we can assume that at some $w = v = -u$, $L(t, u, v, w, z) = t + u - v - w + z$ has a *minimum* under all variations of t, u, v, w, z subject only to the constraints (5.5).

By the Lagrange multiplier theorem [6, p. 129] if (t, u, v, w, z) is a minimum of $L(t, u, v, w, z)$ under the constraints (5.5) then there are constant (multipliers) σ_x, σ_y, σ_ϕ such that

$$(5.6) \qquad L^* \overset{\Delta}{=} L(t, u, v, w, z) - \sigma_x X - \sigma_y Y - \sigma_\phi \Phi,$$

where X, Y, Φ are the right sides of (5.5) respectively, has an *unconstrained* minimum at (t, u, v, w, z). Indeed at the point $(t_0, u_0, v_0, w_0, z_0) = (t, u, -u, -u, z)$ we may take

$$(5.7) \qquad \sigma_x = \frac{\sin(t_0 + v_0)}{\sin v_0}, \quad \sigma_y = -\frac{\cos(t_0 + v_0)}{\sin v_0},$$

$$\sigma_\phi = \frac{\sin(z_0) + \sin v_0}{\sin v_0}$$

and verify that L^* has zero derivative with respect to each of t, u, v, w, z at the point $(t_0, u_0, v_0, w_0, z_0)$. There is a 2-dimensional tangent space to L^* under the constraints (5.5). It is easy to check that $\dot{\xi}_1 = (\dot{t}_1, \dot{u}_1, \dot{v}_1, \dot{w}_1, \dot{z}_1) = (1, 0, 0, 2 - 2\cos v, 1 - 2\cos v)$ and $\dot{\xi}_2 = (\dot{t}_2, \dot{u}_2, \dot{v}_2, \dot{w}_2, \dot{z}_2) = (0, 1, 1, -1, -1)$ are two linearly independent directions that each satisfy $0 = dX = dY = d\Phi$ if the differential is taken in the direction of $(\dot{t}_i, \dot{u}_i, \dot{v}_i, \dot{w}_i, \dot{z}_i)$, $i = 1, 2$. It is a consequence of the general theory of differentiable manifolds [3], that the (conditional) Hessian of the function $L(t, u, v, w, z)$

constrained by (5.5) is given by the 2×2 quadratic form $(\dot{\xi}_1, \dot{\xi}_2)^T \cdot d^2 L \cdot (\dot{\xi}_1, \dot{\xi}_2) = H$ where $d^2 L$ is the 5×5 unconditioned Hessian matrix of 2nd partials of L with t, u, v, w, z. It is seen that

$$(5.8) \qquad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix} \quad \text{where}$$

$$H_{11} = -\frac{4}{\sin v}(1 - 3\cos v + \cos^2 v),$$

$$H_{22} = -\frac{4}{\sin v}, \quad H_{12} = H_{21} = \frac{4}{\sin v}(1 - \cos v).$$

Since the determinant of H is

$$(5.9) \qquad \det(H) = -\frac{16}{\sin^2 v}\cos v(1 - \cos v)$$

and so is negative for $-\pi/2 < v < 0$, it follows that H has one negative and one positive eigenvalue for v in this range. Thus it follows that $(t, u, v, w, z) = (t, u, -u, -u, z)$ is a saddle point rather than a local minimum of L when $0 < u = -v < \pi/2$. So for u in this range the path $CC|CC|C$ can be shortened. An alternative proof avoiding the general theory can be based on computing the second order term in ε to be positive in the expansion of

$$L(t+\varepsilon, u+(1-\cos u)\varepsilon, -u+(1-\cos u)\varepsilon, -u+(1-\cos u)\varepsilon, z-\cos u\varepsilon),$$

in powers of ε checking that the first order term in ε vanishes, and the constraints all hold to order ε^2, so that along this line in εL has a local *maximum* rather than a minimum at $\varepsilon = 0$, i.e. at $(t, u, -u, -u, z)$.

Case 2. $\pi/2 < u < \pi$. By Fig. 1, $\ell_{\pi/2}^+ \ell_{-\pi/2}^- = \ell_{\pi/2}^+ r_{-\pi/2}^-$ so that in this case we have

$$\ell_t^+ r_u^+ \ell_{-u}^- r_{-u}^- \ell_z^+ = \ell_t^+ r_{u-\pi/2}^+ r_{\pi/2}^+ \ell_{-\pi/2}^- \ell_{-u+\pi/2}^- r_{-u}^- \ell_z^+$$

$$= \ell_t^+ r_{u-\pi/2}^+ \ell_{\pi/2}^+ r_{-\pi/2}^- \ell_{-u+\pi/2}^- r_{-u}^- \ell_z^+.$$

But by Dubins's lemma [4, p. 513, Sublemma], any $C_u^+ C_v^+ C_w^+$ can be strictly shortened if $u > 0$, $\pi \ge v > 0$, $w > 0$ and so the first 3 arcs of the last path can be shortened, and hence the whole path can be shortened.

Case 3. $u = \pi/2$. By the same argument as in Case 2, the path reduces to $\ell_t^+ \ell_{\pi/2}^+ r_{-\pi/2}^- r_{-u}^- \ell_z^+ = \ell_{t+\pi/2}^+ \ell_\pi^- + \ell_z^+ = \ell_{t+\pi/2}^+ \ell_\pi^+ \ell_z^+$ which can again be shortened by Dubins's lemma below. $\qquad \square$

This completes the proof of Lemma 5.

Resuming the proof of Theorem 2, by Theorem 1 and Dubins's theorem a sufficient set of words is those of the form

(5.3) α, $\alpha|\beta$, or $\alpha|\beta|\gamma$,

where α, β, $\gamma = C$, S, CC, CS, SC, CSC, CCC.

By Lemma 1 the triple $C|S$ or $S|C$ can be omitted and in the proof of Theorem 1 we have seen that $SC|CS$, $CC|CS$, $C|C|CS$, $C|C|CC$, $C|CCC$ (and their reverses by symmetry) can always be strictly shortened or replaced by a word with fewer cusps or fewer letters. By Lemma 5, $CC|CC|C$ and its reverse are eliminated. It is easy to see that the only words remaining are those in (1.2), (1.1), and their subwords. Note that $C_t C_u C_r$ alone is not needed because again by [**4**, p. 513, Sublemma], $u > \pi$ and then $C_t|C_{u-2\pi}|C_v$ is strictly shorter. □

6. Extension to admissible curves. Here we show that Theorem 2 applies to all admissible curves, not just finite words. This is a simple approximation argument.

THEOREM 3. *Any admissible curve can be shortened to a word of form* (1.2).

Proof. Let γ be any admissible curve. Let $w_n \to \gamma$ be a sequence of finite words converging to γ which exists by the usual approximations of integrals by finite sums in (2.1). Let w'_n be a shortening of w_n of the type described in Theorem 2. By the usual compactness argument, w'_n has a point of accumulation γ' which is also a word described in Theorem 2. By continuity, γ' has the same end points and end directions as γ. But $L(\gamma') = \lim L(w'_n) \leq \lim L(w_n) = L(\gamma)$. □

7. Outline of a direct proof of Dubins's theorem. By Theorem 3 it suffices to prove that any word, that is here any finite string of C's and S's reduces to either CCC or CSC.

Dubins's Lemmas 1 and 2 [**4**] can be stated in our terms as

LEMMA 6. *A proper word of the form SCS or SCC may be strictly shortened to a CSC or a CCC.*

LEMMA 7. *A proper word of the form $CCCC$ may be strictly shortened to a CCC, $CCSC$, or $CSCC$.*

Dubins [4] omits the proof in Lemma 6 that SCC is strictly short-enable, so we do this case. The case SCC is equivalent to CCS and it is no loss to assume the word is $\ell_t r_u s_v$. There are two cases according as $0 < u \leq \pi$, or $\pi < u < 2\pi$. In the first case consider the family ℓrsr and the homotopy of paths $P(h) = \ell_{t(h)} r_{u(h)} s_{v(h)} r_{w(h)}$ where $(t(0), u(0), v(0), w(0)) = (t, u, v, 0)$ and for $h \geq 0$

(7.1) $\dot{t}(h) = -\tfrac{1}{2} v(h), \quad \dot{u}(h) = -\tfrac{1}{2} v(h) - \sin u(h),$

          $\dot{v}(h) = \cos u(h), \quad \dot{w}(h) = \sin u(h).$

It is easy to see that the endpoint of the path is invariant using

(7.2) $(x, y, \phi) = (\sin(w + u - t) + v \cos(u - t) + w \sin t,$

          $\cos(w + u - t) - v \sin(u - t) - 2 \cos t + 1,$

          $-w - u + t)$

and since $\dot{w}(0) = \sin u \geq 0$, $P(h)$ remains in ℓrsr for small h. The length $L(h)$ of $P(h)$ decreases since

(7.3) $\dot{L}(h) = \dot{t}(h) + \dot{u}(h) + \dot{v}(h) + \dot{w}(h) = -v(h)(1 - \cos u(h)) < 0.$

Finally, the homotopy hits the boundary $t(h) = 0$ or $u(h) = 0$ or $v(h) = 0$ which are all of the form CSC or CCC.

In the second case, if $\pi < u < 2\pi$ we will shrink $\ell_t r_u s_v$ inside the family $\ell rs\ell$. The path $P(h) = \ell_{t(h)} r_{u(h)} s_{v(h)} \ell_{w(h)}$ which starts at $(t, u, v, 0)$ and, for $h \geq 0$

(7.4) $\dot{t}(h) = -\tfrac{1}{2} v(h), \quad \dot{u}(h) = -\tfrac{1}{2} v(h) - \sin u(h),$

          $\dot{v}(h) = 2 \sin u(h) + v(h) \cos u(h), \quad \dot{w}(h) = -\sin u(h)$

had endpoints invariant in h using the formula for $\ell rs\ell$,

(7.5) $(x, y, \phi) = (\sin(w - u + t) + v \cos(u - t) + 2 \sin(u - t)$

          $+ 2 \sin t, -\cos(w - u + t) - v \sin(u - t)$

          $+ 2 \cos(u - t) - 2 \cos t + 1, w - u + t).$

The length $L(h)$ of $P(h)$ is decreasing since (7.3) holds again and $\dot{w}(0) = -\sin u > 0$ so that $P(h)$ remains in $\ell rs\ell$. The boundary is finally hit at either $t(h) = 0$ or $u(h) = 0$ or $v(h) = 0$ and again we find a shorter CSC or CCC. This proves Lemma 6.

Now to prove Dubins's theorem, we see that any word can be simultaneously reduced both in path length and in the number of its letters unless the word contains no SCC, CCS, SCS, or $CCCC$. The only words containing none of these as subwords are subwords

of CSC and CCC. This proves Dubins's Theorem in a manner admittedly quite similar to his method except that we do not require his Proposition 11 that all short geodesics are CSC's.

8. Explicit formulas for geodesics. We give explicit formulas for each of the possible actual path-solutions for each of the 48 word types on the list (1.1). (See Table 1.) Note this section is unnecessary for the proof of the theorems but may be useful for someone who wants to implement the actual algorithm. Although some of the 48 words have 2 solutions of the minimizing equations and this gives rise to 68 actual formulas, we have observed empirically that only 48 (one for each word) formulas seem to be needed in any case: i.e. some of the formulas are no doubt spurious solutions of the equations. We have not bothered to actually prove this but in the discussion below merely indicate which solutions can probably be discarded.

To give explicit formulas it is simplest to use the $\ell^{\pm}$, $r^{\pm}$, $s^{\pm}$ notation in which there are 48 different words on the list (1.1) with C replaced by ℓ or r. However it is not necessary to give formulas for each word on the list because there are some elementary transforms:

The formulas for say $\ell^{-}r^{+}s^{+}\ell^{+}$ can be obtained from those for $\ell^{+}r^{-}s^{-}\ell^{-}$ using the "timeflip" transform which interchanges $+$'s and $-$'s. If the original path goes from $(0,0,0)$ to (x,y,ϕ) then it is clear that the timeflipped path goes from $(0,0,0)$ to $(-x,y,-\phi)$. Thus if we have a formula to find the arclengths t, u, v of a path of the form $\ell_{t}^{+}r_{\pi/2}^{-}s_{u}^{-}\ell_{v}^{-}$ from $(0,0,0)$ to an $(-x,y,-\phi)$ this may be used to find the corresponding arclengths for the $\ell^{-}r^{+}s^{+}\ell^{+}$ path to (x,y,ϕ). Thus timeflipping allows us to eliminate all words on the list whose first letter has a $-$ sign.

Similarly, the formulas for say $r^{+}\ell^{-}s^{-}r^{-}$ can be obtained from those for $\ell^{+}r^{-}s^{-}\ell^{-}$ using the "reflect" transform which interchanges r's and ℓ's along the path. The reflected path changes (x,y,ϕ) to $(x,-y,-\phi)$, so to find the t, u, v for which $r_{t}^{+}\ell_{\pi/2}^{-}s_{u}^{-}r_{v}^{-}$ reaches (x,y,ϕ) we find the corresponding quantities for $\ell_{t}^{+}r_{\pi/2}^{-}s_{u}^{-}\ell_{v}^{-}$ to reach $(x,-y,-\phi)$. Reflecting (and timeflipping) allows us to eliminate all words on the list except those beginning with ℓ^{+}.

Finally, the formulas for say $\ell^{-}s^{-}r^{-}\ell^{+}$ can be obtained from those for $\ell^{+}r^{-}s^{-}\ell^{-}$ using the "backwards" transform which follows the path in reverse order, but with a timeflip so that the individual segments are transversed in the same direction. This transforms the final point (x,y,ϕ) to $(x\cos\phi + y\sin\phi, x\sin\phi - y\cos\phi, \phi)$ and

so we should use the formula for $\ell^+ r^- s^- \ell^-$ to get to $(x\cos\phi + y\sin\phi,\ x\sin\phi - y\cos\phi,\ \phi)$ to obtain this formula for $\ell^- s^- r^- \ell^+$ to arrive at (x, y, ϕ). This eliminates a few more words on the list for which we need to give formulas. The remaining 9 words are discussed below.

In each formula the object is to move from $(0, 0, 0)$ to (x, y, ϕ). We write $(r, \theta) = R(x, y)$ for the polar transform $r\cos\theta = x$, $r\sin\theta = y$, $r \geq 0$, $-\pi \leq \theta < \pi$. We write $\phi = M(\theta)$ if $\phi \equiv \theta$ mod 2π and $-\pi \leq \phi < \pi$. The quantities T are discarded in the formulas below and may not be equal at different usages. In each case where there is a solution-path we give the unknown segment lengths t, u, v, and the overall length L. We say $L = \infty$ if there is no solution.

(8.1) $\ell_t^+ s_u^+ \ell_v^+$: Define $(u, t) = R(x - \sin\phi, y - 1 + \cos\phi)$,

and set

$$v = M(\phi - t), \quad L = |t| + |u| + |v|.$$

Of course this path cannot be optimal if t or v is outside $[0, \pi]$.

(8.2) $\ell_t^+ s_u^+ r_v^+$: Define $(u_1, t_1) = R(x + \sin\phi, y - 1 - \cos\phi)$.

If $u_1^2 < 4$, $L = \infty$; else let $u = \sqrt{u_1^2 - 4}$, $(T, \theta) = R(u, 2)$, $t = M(t_1 + \theta)$, $v = M(t - \phi)$, $L = |t| + |u| + |v|$.

(8.3) $\ell_t^+ r_u^- \ell_v^+$: Define $\xi = x - \sin\phi$, $\eta = y - 1 + \cos\phi$,

$$(u_1, \theta) = R(\xi, \eta).$$

If $u_1^2 > 4$, $L = \infty$; else let $A = \arcsin(u_1^2/4)$, $\pi/2 \leq A \leq \pi$,

$$u = M(A + \theta),$$
$$(T, v) = R(2 - \xi\sin u + \eta\cos u,\ \xi\cos u + \eta\sin u),$$
$$w = M(\phi + v - u), \quad L = |t| + |u| + |v|.$$

REMARK. There is another solution which we have never observed to be optimal and so doubtless can be discarded. It is the same as (8.3) above except that the other branch of arcsin is used, $0 \leq A \leq \pi/2$.

(8.4) $\ell_t^+ r_u^- \ell_v^-$: Use same formula as (8.3); same remark applies.

In each of the cases $\ell_t^+ r_u^+ \ell_{-u}^- r_v^-$ and $\ell_t^+ r_u^- \ell_u^- r_v^+$ getting to (x, y, ϕ) means solving the equations for t, u, u_1, v where $u_1 = \pm u$,

(8.5) $x + \sin\phi \stackrel{\Delta}{=} \xi = 2\sin(t - u + u_1) + 2\sin(u - t) + 2\sin t$,

$$y - 1 - \cos\phi \stackrel{\Delta}{=} \eta = -2\cos(t - u + u_1) + 2\cos(u - t) - 2\cos t,$$
$$\phi = t - u + u_1 - v.$$

We need the functions $\tau = \tau(u, v)$, $\omega = \Omega(u, v)$ defined for $u, v \in [-\pi, \pi)$

$$(8.6) \quad \delta = M(u - v), \quad A = \sin u - \sin \delta, \quad B = \cos u - \cos \delta - 1,$$
$$(T, t_1) = R(\xi A + \eta B, \eta A - \xi B),$$
$$t_2 = 2 \cos \delta - 2 \cos v - 2 \cos u + 3.$$

If $t_2 < 0$, $\tau = M(t_1 + \pi)$, otherwise $\tau = M(t_1)\omega = M(\tau - u + v - \phi)$,

$$(8.7) \qquad \ell_t^+ r_u^+ \ell_{-u}^- r_v^- : \text{ Let } \rho = \frac{2 + \sqrt{\xi^2 + \eta^2}}{4}.$$

If $0 \le \rho \le 1$, set $u = \arccos(\rho)$, $0 \le u \le \pi/2$, $t = \tau(u, -u)$, $u = w(u, -u)$, $L = |t| + 2|u| + |v|$; else $L = \infty$.

REMARK. There is another solution which we have never observed to be the unique optimum and so doubtless can be discarded. It is the same as (8.7) above except that $\rho = (2 - \sqrt{\xi^2 + \eta^2})/4$.

$$(8.8) \qquad \ell_t^+ r_u^- \ell_u^- r_v^+ : \text{ Let } \rho = (20 - \xi^2 - \eta^2)/16.$$

If $0 \le \rho \le 1$, set $u = -\arccos(\rho)$, $0 \le u \le \pi/2$, $t = \tau(u, u)$, $v = w(u, u)$; else $L = \infty$.

(8.9) $\ell_t^+ r_{-\pi/2}^- s_u^- \ell_v^- :$ Define $\xi = x + \sin\phi$, $\eta = y - 1 - \cos\phi$,

$$(\rho, \theta) = R(-\eta, \xi).$$

If $\rho < 2$, $L = \infty$. Else set $(T, \theta_1) = R(\sqrt{\rho^2 - 4}, -2)$ and $t - M(\theta - \theta_1)$, $u = 2 - \theta_1$, $v = M(\phi - \pi/2 - t)$, $L = |t| + \pi/2 + |u| + |v|$.

(8.10) $\ell_t^+ r_{-\pi/2}^- s_u^- r_v^- :$ Define ξ, η, ρ, θ as in (8.9) and set

$$t = \theta, \quad u = 2 - \rho, \quad v = M(t + \pi/2 - \phi),$$
$$L = |t| + \pi/2 + |u| + |v|.$$

(8.11) $\ell_t^+ r_{-\pi/2}^- r_u^- \ell_{-\pi/2}^- r_v^+ :$ Define $\xi = x + \sin\phi$, $\eta = y - 1 - \cos\phi$,

$$(\rho, \theta) = R(\xi, \eta).$$

If $\rho < 2$, $L = \infty$. Else set $t = M(\theta - \arccos(-2/\rho))$, $-\pi/2 \le t \le \pi/2$. If $t \le 0$, $L = \infty$. Else set $u = 4 - (\xi + 2\cos t)/\sin t$ and $w = M(t - \phi)$.

REMARK. There is another solution which we have never observed to be the unique optimum and so doubtless can be discarded. It is the same as (8.11) above except that $t = M(\theta + \arccos(-2\rho))$, $-\pi/2 \leq t \leq \pi/2$.

9. Additional remarks. It is perhaps more realistic to assume that in an application the cart will not have total freedom of motion, but there will be obstructions in its path. We have not considered the consequences of these. It does seem reasonable to conjecture that any optimal path γ in an obstructed plane, (i.e. where a certain open subset C of the plane is obstructed and cannot be entered) will have the property that any connected piece of γ disjoint from ∂C is of the form (1.2), or (1.1). However, we have not proved this nor are we even convinced that it is true. We leave it as an open problem.

Colin Mallows has formulated another more realistic problem of similar type in which the fact that the cart has two sets of wheels and nonzero length is respected, and obtains some partial results, but these are less complete.

These problems make sense in higher dimensions as well but we could solve neither the forward nor reverse case mainly because we could not explicitly solve the algebraic equations involved in finding the shortest CCC from a to b in $\mathbb{R}^5$, i.e. with specified initial and final coordinates and directions in 3 dimensions, and so we could not use the computer to guess the answer analogous to the way we used it here. We think this answer is CCC, CSC for the forward problem but without much evidence or hope of doing the reduction because the equations get so complicated. An interpretation for the 3-dimensional problem is that of a plumber wanting to connect two existing pipes with a pipe which can be bent but not too quickly so that its radius of curvature should always be ≥ 1. A similar problem is faced by glider pilots.

REFERENCES

[1]   E. J. Cockayne and G. W. C. Hall, *Plane motion of particle subject to curvature constraints*, SIAM J. Control, **13** (1975), 197–220.

[2]   E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[3]   J. Dieudonne, *Treatise on Analysis*, vol. III, Academic Press, New York, 1972.

[4]   L. E. Dubins, *On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents*, Amer. J. Math., **79** (1957), 497–516.

[5]   ____, *On plane curves with curvature*, Pacific J. Math., **11** (1961), 471–482.

[6]   W. H. Fleming, *Functions of Several Variables*, Addison-Wesley, Reading, Mass., 1966.

[7]   M. G. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*, American Mathematical Society, Providence, 1977.

[8]   A. A. Markov, *Some examples of the solution of a special kind of problem on greatest and least quantities*, (in Russian) Soobshch. Karkovsk. Mat. Obshch., **1** (1887), 250–276.

[9]   Z. A. Melzak, *Plane motion with curvature limitations*, J. Soc. Ind. Appl. Math., **9** (1961), 422–432.

AT&T Bell Laboratories
Murray, Hill, NJ 07974-2070