Dubins' problem is intrinsically three-dimensional

Dirk Mittenhuber*

March 7, 1997

Abstract

In his 1957 paper [1] L. Dubins considered the problem of finding shortest differentiable arcs in the plane with curvature bounded by a constant $\kappa > 0$ and prescribed initial and terminal positions and tangents. One can generalize this problem to non-euclidean manifolds as well as to higher dimensions, cf. [8].

Considering that the boundary data—initial and terminal position x_0, x_1 and tangents v_0, v_1—are genuinely three-dimensional, it seems natural to ask if the n-dimensional problem always reduces to the three-dimensional case. In this paper we will prove that this is true in the euclidean as well as in the noneuclidean case. At first glance one might consider this a trivial problem, but we will also give an example showing that this is not the case.

1 Introduction

Dubins' problem in $\mathbb{R}^n$ can be formulated as follows: Given points $x_0, x_1 \in \mathbb{R}^n$ and two vectors v_0, v_1 of unit length, find a curve $\gamma\colon [0, L] \to \mathbb{R}^n$ such that

(i) $\gamma(s)$ is a differentiable curve, parametrized by arclength,

(ii) γ has curvature $\kappa(s)$ almost everywhere, and the curvature is essentially bounded by some positive constant κ_0: $|\kappa(s)| \leq \kappa_0$ a.e.,

(iii) γ satisfies the boundary conditions:

$$\gamma(0) = x_0,\ \dot\gamma(0) = v_0,\ \gamma(L) = x_1,\ \dot\gamma(L) = v_1,$$

(iv) Among all possible arcs satisfying (i)–(iii), γ has shortest length, i.e., L is minimal.

*The author greatfully acknowledges the support by the Alexander von Humboldt-Foundation.

One of the well-known interpretations of this problem is to think of a car moving with constant speed in the plane (or n-space) subject to the constraint that it cannot make arbitrarily sharp turns, see [10], for example.

Dubins proved that optimal arcs in $\mathbb{R}^2$ are concatenations of circular arcs (with constant curvature κ_0) and straight line segments. Moreover he proves that optimal arcs consist of at most three pieces and that the line segment—if there is any—has to be in the middle, i.e., the optimal arcs follow what we now call *Dubins' pattern*, either CCC or CLC. This reduces finding the optimal arcs to a finite problem. There are at most six candidates for optimal arcs. So all one has to do is determine these arcs and compare their lengths.

It was an open conjecture whether optimal arcs in $\mathbb{R}^3$ are again concatenations of circles and straight lines, cf. [10], until this was disproved by Sussmann in [11].

In this paper Sussmann gave an example of a particular set of boundary conditions (or *Dubins' data*) x_0, v_0, x_1, v_1 and an arc γ matching these data, such that any concatenation of circles and lines satisfying these boundary conditions must be strictly longer than γ. The arc γ is a helicoidal arc, so it has constant curvature κ but also constant torsion $\tau \neq 0$. And this example sheds a new light on the problem. The result is no longer surprising because thinking of the curvature as a cost functional in our variational problem, we are free to add torsion at no cost at all.

Considering Dubins' problem as an optimal control problem on the manifold $\mathbb{R}^3 \times \mathbb{S}^2$ (position and unit tangent vector) Sussmann gives a detailed analysis of the structure of the extremals, i.e., curves satisfying the necessary conditions furnished by the Pontrjagin Maximum Principle (PMP). Of course Dubins' optimal arcs—concatenations of circles and lines—appear simply because the two-dimensional case is just a subcase, but there is another type of extremals characterized by having constant curvature κ_0 and their torsion τ satisfying a certain ODE, see [11, p. 3308].

Next we observe that the boundary data x_0, x_1, v_0, v_1 are intrinsically three-dimensional because the vector space $V := \operatorname{span}\{x_1 - x_0, v_0, v_1\}$ has dimension $\dim(V) \leq 3$, so the Dubins' data lie in the affine subspace $x_0 + V$. And since the group $\mathrm{SE}(n)$ of euclidean motions of $\mathbb{R}^n$ obviously leaves the problem invariant—carrying optimal arcs into optimal arcs—it seems as if one could immediately reduce the problem to the subspace V. How could an arc leave this subspace and be shorter than an arc staying inside of V? But the attempts to prove this immediately do not succeed.

So we look at the step from dimension 2 to 3 first, and suddenly we realize that this is a nontrivial question. We consider the following Dubins' data:

$$x_0 = 0, \quad x_1 = re_1, \quad \text{with } 0 < r \leq 4, \quad \text{and} \quad v_0 = v_1 = e_2.$$

Since we know the structure of optimal arcs in $\mathbb{R}^2$, it is not hard to verify that for small $r > 0$ the optimal arc in $\mathbb{R}^2$ is of type CLC and has length $L(r) = 2\pi + r$.

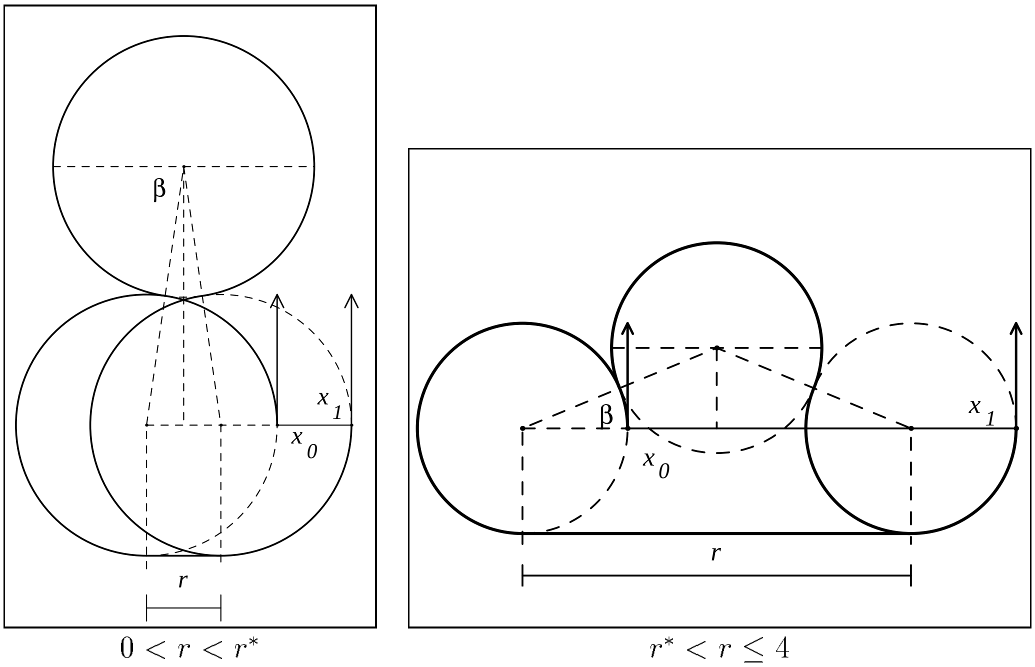


Figure 1: The optimal r-shift in $\mathbb{R}^2$, $r = 4\cos\beta$.

This is in fact true for $r \in (0, r^*]$ where β^* is the unique solution of the equation $\cos\beta = \beta$, and $r^* = 4\cos(\beta^*) \approx 2.95634$. Just for sake of completeness we mention that for $r \in [r^*, 4]$ the optimal arc has length $L(r) = 2\pi + 4\arccos(r/4)$. Figure 1 shows the competing CLC and CCC-arcs. But in $\mathbb{R}^3$ for any $r \in (0, 4]$ we find an admissible curve of length 2π, cf. Figure 2. Looking at the picture one can think of two semicircles tied together but free to rotate around the common axis at the joint. We pick up the joint lifting it into the vertical z-direction while the endpoints remain in the xy-plane. The height of the dashed triangles is $h = \frac{1}{2}\sqrt{16 - r^2}$. Even though I am convinced that these arcs are actually optimal I did not make any efforts to prove it. All we need here is that the three-dimensional arcs are shorter than the best possible planar arcs.

With this example in mind it is clear that it is not trivial to prove that shortest arcs are three-dimensional, i.e., stay in a three-dimensional affine subspace compatible with our Dubins' data. Leaving the subspace $V(x_0, x_1, v_0, v_1)$ may decrease the length of the arc! If we think of deforming an elastic rod in the shape of a circular arc, this will add (perhaps higher-dimensional) torsion but it can reduce the curvature or, if we leave the curvature unchanged, decrease the arclength.

Nevertheless we will prove that one need not go beyond dimension 3 in the euclidean as well as the noneuclidean case.

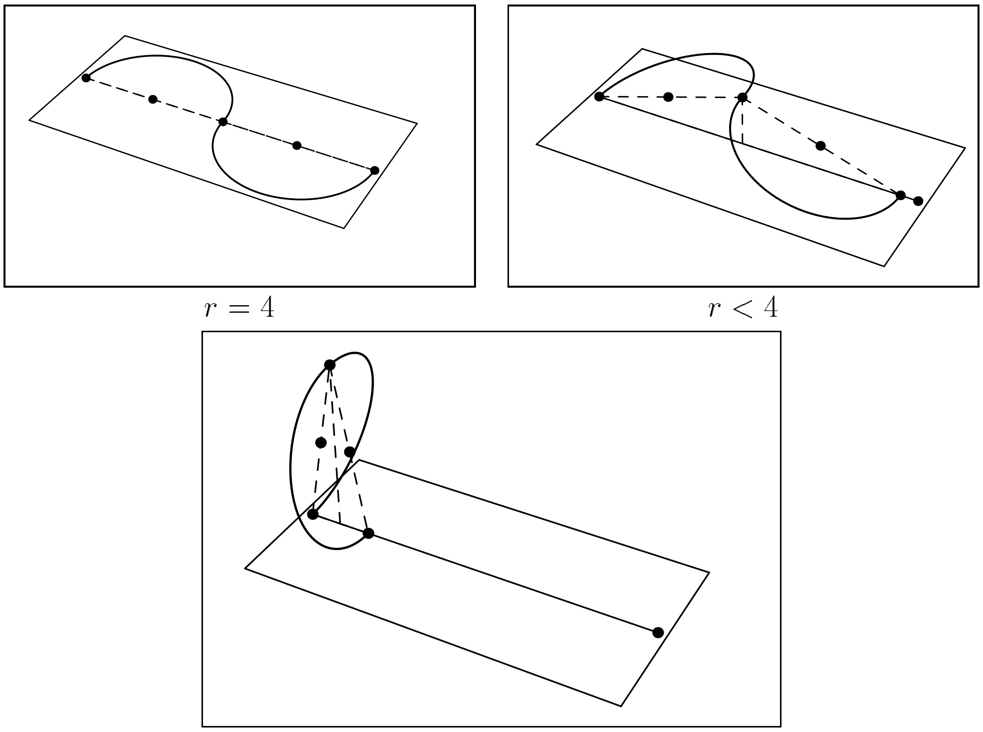


Figure 2: r-shifts in $\mathbb{R}^3$ of length 2π.

Theorem 1.1 *In order to solve Dubins' problem in $\mathbb{R}^n$, $\mathbb{S}^n$ or $\mathbb{H}^n$, it is sufficient to solve it in $\mathbb{R}^3$, $\mathbb{S}^3$, resp. $\mathbb{H}^3$.* □

Remark 1.2 In the euclidean case the previous statement may be reformulated as: Every optimal arc in $\mathbb{R}^n$ is contained in a three-dimensional affine subspace. And if the Dubins' data x_0, x_1, v_0, v_1 are truly three-dimensional, i.e., if $\det(x_1 - x_0, v_0, v_1) \neq 0$, then this subspace is unique, of course.

In the noneuclidean case this reformulation is a little bit more sophisticated. The abstract generalization of an *affine subspace* would be a *totally geodesic submanifold*. But we rather give a concrete description. If we consider $\mathbb{S}^n, \mathbb{H}^n$ appropriately embedded into $\mathbb{R}^{n+1}$, this totally geodesic submanifold is simply the intersection of $\mathbb{S}^n$, resp. $\mathbb{H}^n$ with a subspace $\mathbb{R}^4 \subseteq \mathbb{R}^{n+1}$. If we use the embeddings

$$\mathbb{S}^n = \left\{ y \in \mathbb{R}^{n+1} \mid \|y\| = 1 \right\}, \quad \mathbb{H}^n = \left\{ y \mid y_1^2 - y_2^2 - \cdots - y_{n+1}^2 = 1,\ y_1 > 0 \right\},$$

and our Dubins' data are $x_0, x_1, v_0, v_1 \in \mathbb{R}^{n+1}$, then the corresponding subsphere, resp. hyperbolic subspace are given as

$$\mathbb{S}^n \cap \operatorname{span}\{x_0, x_1, v_0, v_1\}, \quad \mathbb{H}^n \cap \operatorname{span}\{x_0, x_1, v_0, v_1\}.$$

The reason for this simple description is that for x_0, x_1 in $\mathbb{S}^n$, $x_1 \neq -x_0$ (not antipodal points), the geodesic through x_0, x_1 is actually $\mathbb{S}^n \cap \operatorname{span}\{x_0, x_1\}$. And a similar statement is true for $\mathbb{H}^n$, cf. [9, p. 68].

The proof of Theorem 1.1 is based on a detailed analysis of the structure of the extremals, i.e., arcs satisfying the necessary conditions provided by the (PMP).

2 Dubins' problem as an optimal control problem

Sussmann considers Dubins' problem as a control problem with state space the manifold $\mathbb{R}^n \times \mathbb{S}^{n-1}$ (position and unit tangent vector). We use the slightly different approach of Monroy in [8]. Monroy considers it as a control problem on the Lie group $\mathrm{SE}(n)$ of euclidean motions of $\mathbb{R}^n$. In block matrix notation one can write

$$\mathrm{SE}(n) = \left\{ \begin{pmatrix} 1 & 0 \\ x & Q \end{pmatrix} \mid x \in \mathbb{R}^n,\ Q \in \mathrm{SO}(n) \right\}.$$

The ODE of our optimal control problem is:

$$\dot{g} = g \cdot \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & u^T \\ 0 & -u & X \end{pmatrix}, \quad X \in \mathfrak{so}(n-1),\ u \in \mathbb{R}^{n-1},\ \|u\| \leq \kappa_0.$$

And the cost functional is time, i.e., we are looking for time-optimal paths steering from a given point g_0 to another point g_1.

This setting contains a little redundancy. In fact one can consider a group element $g = (x, Q) \in \mathrm{SE}(n)$ as a point in $\mathbb{R}^n$ with a positively oriented orthonormal frame attached to it. If $g(t)$ is a solution of the above ODE, then its first and second column are

$$g(t)e_1 = \begin{pmatrix} 1 \\ x(t) \end{pmatrix}, \text{ and } \begin{pmatrix} 0 \\ \dot{x}(t) \end{pmatrix} = \dot{g}(t)e_1 = g(t) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & u^T \\ 0 & -u & X \end{pmatrix} e_1 = g(t)e_2.$$

So the first two columns of $g(t)$ are in one-to-one correspondence with the arc $x(t)$ and its tangent vector $v(t) = \dot{x}(t)$. Monroy proves that the problems on $\mathrm{SE}(n)$ and $\mathbb{R}^n \times \mathbb{S}^{n-1}$ are equivalent in the sense that an optimal $g(t) \in G$ projects onto an optimal arc $x(t) \in \mathbb{R}^n$ and conversely every optimal $x(t)$ may be obtained this way, i.e., lifts to an optimal $g(t) \in \mathrm{SE}(n)$, cf. [8, Theorems 2.5.1 and 2.5.2].

This lifting is not unique in general. Of course, if $n = 2$ and v_0 is a unit vector, then there is only one possible choice of another unit vector w_0 such that (v_0, w_0) is a positively oriented orthonormal frame in $\mathbb{R}^2$. But for $n > 2$ there are infinitely many possibilities, of course. Any appropriately oriented orthonormal frame of $v_0^\perp$ will do the job.

In abstract terms the manifold $\mathbb{R}^n \times \mathbb{S}^{n-1}$ is a homogeneous space of the group $\mathrm{SE}(n)$. For $g \in \mathrm{SE}(n)$ let $\phi(g) = (g.e_1, g.e_2) = \begin{pmatrix} 1 & 0 \\ x & Qe_1 \end{pmatrix}$. Then ϕ maps $\mathrm{SE}(n)$ onto $\mathbb{R}^n \times \mathbb{S}^{n-1}$ because $\|Qe_1\| = \|e_1\| = 1$ for every $Q \in \mathrm{SO}(n-1)$. Let $K \subseteq \mathrm{SE}(n)$ denote the subgroup

$$K = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & Q \end{pmatrix} \mid Q \in \mathrm{SO}(n-1) \right\},$$

Then $k.e_1 = e_1$ and $k.e_2 = e_2$ for all $k \in K$, hence $\phi(gK) = \phi(g)$ for all $g \in \mathrm{SE}(n)$. Thus we have $\mathbb{R}^n \times \mathbb{S}^{n-1} \cong \mathrm{SE}(n)/K$.

Remark 2.1 Considering the problem on the group $\mathrm{SE}(n)$ rather than the homogeneous space $\mathrm{SE}(n)/K \cong \mathbb{R}^n \times \mathbb{S}^{n-1}$ is not only a national device. It also clarifies how to exploit the symmetry of our problem when determining the extremals. The integrals of motion we will use are derived from the group structure.

Another advantage of this approach is that it allows to treat the three cases (hyperbolical, euclidean, spherical) simultaneously. Let $\varepsilon \in \{-1, 0, 1\}$. Then the group G one has to consider is either $\mathrm{SO}(1,n)$ or $\mathrm{SE}(n)$ or $\mathrm{SO}(n+1)$ and the ODE of the problem is

$$\dot{g} = g \cdot \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & u^T \\ 0 & -u & X \end{pmatrix}, \quad X \in \mathfrak{so}(n-1),\ u \in \mathbb{R}^{n-1},\ \|u\| \leq \kappa_0. \tag{1}$$

With the same group $K \cong \mathrm{SO}(n-1)$ as above the objective is to find for given $g_0, g_1 \in G$ a time-optimal arc from g_0K to g_1K.

At first this seems to be more complicated than the original problem because now we must steer time-optimal from the submanifold g_0K to the target manifold g_1K. But we will show that the optimal time is the same for any choice of $h_0 \in g_0K$ and $h_1 \in g_1K$. Let $\mathfrak{g} = \mathbf{L}(G)$ denote the Lie algebra of G and $\mathcal{U} \subseteq \mathfrak{g}$ the set of admissible control values, i.e.,

$$\mathcal{U} = \left\{ \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & u^T \\ 0 & -u & X \end{pmatrix} \mid X \in \mathfrak{so}(n-1),\ u \in \mathbb{R}^{n-1},\ \|u\| \leq \kappa_0 \right\}.$$

Then $\mathcal{U}$ has the form $B + \mathfrak{k}$ where $\mathfrak{k} = \mathbf{L}(K) \cong \mathfrak{so}(n-1)$ is the Lie algebra of K and the set

$$B = \left\{ \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & u^T \\ 0 & -u & 0 \end{pmatrix} \mid u \in \mathbb{R}^{n-1},\ \|u\| \leq \kappa_0 \right\}$$

is compact and $\mathrm{Ad}(K)$-invariant because for $Q \in \mathrm{SO}(n-1)$, $Y = \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & u^T \\ 0 & -u & 0 \end{pmatrix}$ and $k = \mathrm{diag}(1,1,Q)$ we have

$$\mathrm{Ad}(k)Y = kYk^{-1} = \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & (Qu)^T \\ 0 & -Qu & 0 \end{pmatrix}.$$

Lemma 2.2 *Let $u(t)\colon [0,T] \to \mathcal{U}$ be an admissible control and $g(t)$ the corresponding trajectory. Let $k(t) \in K$ be absolutely continuous and $v(t) = k^{-1}\dot{k}$. Then $\widetilde{g}(t) := g(t)k(t)$ is a trajectory of the control system (1), generated by the control $\widetilde{u}(t) = \mathrm{Ad}(k^{-1})u(t) - v(t)$.*

Proof. We differentiate $\widetilde{g}(t)$. Since $\dot{g}(t) = g(t)\,u(t)$ and $\dot{k}(t) = k(t)\,v(t)$, we obtain

$$\begin{aligned}\dot{\widetilde{g}}(t) &= \dot{g}(t)\,k(t) + g(t)\,\dot{k}(t) = g(t)\,u(t)\,k(t) + g(t)\,k(t)\,v(t) \\ &= g(t)k(t)\Big(k^{-1}u(t)k - v(t)\Big) = \widetilde{g}(t)\Big(\mathrm{Ad}(k^{-1})u(t) - v(t)\Big) = \widetilde{g}(t)\,\widetilde{u}(t).\end{aligned}$$

Since $\mathrm{Ad}(K)\mathcal{U} = \mathcal{U}$ and $\mathcal{U} + \mathfrak{k} = \mathcal{U}$, $\widetilde{u}(t) \in \mathcal{U}$ follows. □

This allows to prove immediately that the minimal time is independent of a particular choice of points in g_0K, g_1K.

Corollary 2.3 *For $g_0, g_1 \in G$ let $T(g_0, g_1)$ denote the minimal time to steer from g_0 to g_1. Then for all $k_0, k_1 \in K$ we have $T(g_0k_0,\, g_1k_1) = T(g_0, g_1)$.*

Proof. Let $g(t)\colon [0,T] \to G$ denote a time-optimal path from g_0 to g_1. Since $K \cong \mathrm{SO}(n-1)$ is connected, we can find an absolutely continuous path $k(t)\colon [0,T] \to K$ from k_0 to k_1. Thus $\widetilde{g}(t) := g(t)k(t)$ is an admissible path steering from g_0k_0 to g_1k_1 in time $T = T(g_0, g_1)$, hence $T(g_0k_0, g_1k_1) \leq T(g_0, g_1)$ follows.

Applying this result to the case $g_i' := g_ik_i$ and $k_i' = k_i^{-1}$ yields the reverse inequality. Thus equality holds. □

Remark 2.4 The same argument works, of course, with *infimum* instead of *minimum time*. Since time-optimal arcs always exist, cf. [8, Theorem 3.1.1], for example, we do not need to make this distinction.

The reason why we chose this particular approach is that Lemma 2.2 allows a significant simplification when determining the optimal controls. If we write $u(t) = u_B(t) + u_{\mathfrak{k}}(t)$, then we will show that one may assume w.l.o.g. that $u_{\mathfrak{k}}(t) \equiv 0$.

3 Application of the Maximum Principle

For the discussion that follows we need to fix a basis of the Lie algebra $\mathfrak{g}$. Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{R}^n$ and let

$$\ell_i = \begin{pmatrix} 0 & -\varepsilon e_i^T \\ e_i & 0 \end{pmatrix}, \quad \text{and} \quad L_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & e_i e_j^T - e_j e_i^T \end{pmatrix}.$$

With the bracket being $[X, Y] = XY - YX$ we obtain

$$[\ell_i, \ell_j] = -\varepsilon L_{ij},\ [\ell_i, L_{jk}] = \delta_{ij}\ell_k - \delta_{ik}\ell_j,\ [L_{ij}, L_{kl}] = \delta_{jk}L_{il} - \delta_{ik}L_{jl} - \delta_{jl}L_{ik} + \delta_{il}L_{jk}.$$

(We note that Monroy uses the opposite sign, cf. [8, p.24]). We rewrite the ODE (1) as

$$\dot{g} = g\left(\ell_1 + \sum_{i<j} u_{ij} L_{ij}\right), \quad \sum_{j=2}^{n} u_{1j}^2 \leq \kappa_0^2. \tag{2}$$

In order to apply the (PMP) we must lift this control problem to the cotangent bundle T^*G. But the cotangent bundle of a Lie group has a very simple structure. If we let $\mathfrak{g}^*$ denote the vector space dual of $\mathfrak{g}$, then one can identify T^*G with $G \times \mathfrak{g}^*$. The set of smooth functions on T^*G with the Poisson bracket $\{\,\cdot\,,\,\cdot\,\}$ is a Lie algebra. Every $X \in \mathfrak{g}$ induces a smooth function on T^*G, namely $H_X(g, p) = \langle p, X\rangle$. If signs for the Lie bracket $[\,\cdot\,,\cdot\,]$ in $\mathfrak{g}$ and the Poisson bracket $\{\,\cdot\,,\,\cdot\,\}$ are chosen properly, then the map $(X \mapsto H_X)\colon \mathfrak{g} \to C^\infty(T^*G)$ is a Lie algebra homomorphism, i.e.,

$$\{H_X, H_Y\} = H_{[X,Y]} \quad \text{for all } X, Y \in \mathfrak{g}.$$

Now let $h_i, H_{ij}\colon G \times \mathfrak{g}^* \to \mathbb{R}$ denote the following functions corresponding to $\ell_i, L_{ij} \in \mathfrak{g}$:

$$h_i(g, p) = \langle p, \ell_i\rangle, \quad H_{ij}(g, p) = \langle p, L_{ij}\rangle.$$

Then the system Hamiltonian of our control problem is

$$\mathcal{H}_u(g, p) = \langle p, u\rangle = h_1 + u_{12}H_{12} + \cdots + u_{1n}H_{1n} + \sum_{1<i<j}^{n} u_{ij}H_{ij}.$$

Now suppose that $g(t)$ is an optimal arc generated by the control $u(t)$. Then application of the (PMP) yields the existence of a covector $p(t) \in \mathfrak{g}^*$ with the following properties, see [8]:

(i) $p(t)$ is not identically 0 and therefore $p(t) \neq 0$ for all t.

(ii) $p(t)$ satisfies a certain ODE. In implicit form, this ODE is

$$\left.\frac{d}{dt}\right|_{t=0} \mathrm{Ad}\left(g(t)^{-1}\right)^* p(t) \equiv 0,$$

so $\mathrm{Ad}\left(g(t)^{-1}\right)^* p(t) =: \Lambda \in \mathfrak{g}^*$ is constant.

(iii) The Hamiltonian is pointwise maximized by the control $u(t)$:

$$\mathcal{H}_{u(t)}(g(t),p(t)) = \max_{v\in\mathcal{U}} \mathcal{H}_v(g(t),p(t)),$$

and its value is constant, either 0 or 1:

$$\mathcal{H}_{u(t)}(g(t),p(t)) = \lambda \in \{0,1\}.$$

Since there is no restriction on u_{ij} with $1 < i < j$, an immediate consequence of (iii) is $H_{ij}(g(t),p(t)) \equiv 0$. As a drawback Condition (iii) does not give any information about the controls $u_{ij}(t)$ with $1 < i < j$. This is the reason why we proved Lemma 2.2. It allows to eliminate these controls.

Lemma 3.1 *We may assume w.l.o.g. that $u_{ij}(t) \equiv 0$ for $1 < i < j$.*

Proof. Since the set $\mathcal{U}$ has the form $B + \mathfrak{k}$ let us write $u(t) = u_B(t) + u_{\mathfrak{k}}(t)$. Then $u_{ij} \equiv 0$ for $1 < i < j$ is equivalent to $u_{\mathfrak{k}}(t) \equiv 0$.

Let $k(t) \in K$ denote the solution of the ODE $\dot{k}(t) = -u_{\mathfrak{k}}(t)\,k(t)$, $k(0) = \mathbf{1}$ (identity matrix), and set $\tilde{g}(t) := g(t)\,k(t)$. Then we obtain

$$\begin{aligned} \dot{\tilde{g}}(t) &= g(t)\,u(t)\,k(t) + g(t)\,(-u_{\mathfrak{k}}(t))\,k(t) \\ &= g(t)k(t)\,k(t)^{-1}(u(t) - u_{\mathfrak{k}}(t))\,k(t) = \tilde{g}(t)\ \mathrm{Ad}\left(k(t)^{-1}\right) u_B(t). \end{aligned}$$

Since the set B is $\mathrm{Ad}(K)$-invariant, we deduce that $\tilde{u}(t) := \mathrm{Ad}(k(t)^{-1})u_B(t) \in B$ for all t, or equivalently $\tilde{u}_{\mathfrak{k}} \equiv 0$. Of course $\tilde{g}(t)K = g(t)\,k(t)K = g(t)K$, so $\tilde{g}(t)$ and $g(t)$ project onto the same arc $x(t)$. □

We also note that because of Condition (ii) the covector $\tilde{p}$ that goes with $\tilde{g}, \tilde{u}$ is $\tilde{p}(t) = \mathrm{Ad}(k(t))^* p(t)$.

From now on, we assume $u_{ij} \equiv 0$ for $1 < i < j$. Since $H_{ij} \equiv 0$ for $1 < i < j$ many terms in the ODE for $p(t)$ simply disappear.

Let us write $h_i(t), H_{1j}(t)$ for $h_i(p(t)), H_{1j}(p(t))$. Then the derivatives $\dot{h}_i$, can be obtained through the Poisson bracket:

$$\dot{h}_i = \{\mathcal{H}_u, h_i\} = \{h_1 + u_{12}H_{12} + \cdots + u_{1n}H_{1n},\ h_i\},\ \text{etc.}$$

In particular for the functions h_i, H_{1j} we obtain the following Poisson brackets:

$$\begin{aligned} [\ell_1,\ell_j] &= -\varepsilon L_{1j}, & \{h_1,h_j\} &= -\varepsilon H_{1j}, \\ [\ell_1,L_{1j}] &= \ell_j, & \{h_1,H_{1j}\} &= h_j \\ [L_{1j},\ell_j] &= \ell_1, & \{H_{1j},h_j\} &= h_1 \\ [L_{1j},L_{1k}] &= L_{jk}, & \{H_{1i},H_{1j}\} &= H_{ij}. \end{aligned}$$

Let us define

$$h' = \begin{pmatrix} h_2 \\ \vdots \\ h_n \end{pmatrix},\ H' = \begin{pmatrix} H_{12} \\ \vdots \\ H_{1n} \end{pmatrix},\ \mu(t) = \|H'\| = \sqrt{H_{12}^2 + \cdots + H_{1n}^2}.$$

Since $H_{ij} \equiv 0$ for $i, j > 1$, we obtain

$$\begin{array}{rclrcl} \dot{h}_1 &=& -u_{12}h_2 - \cdots - u_{1n}h_n, && & \\ \dot{h}_2 &=& h_1u_{12} - \varepsilon H_{12}, & \dot{H}_{12} &=& h_2, \\ &\vdots& & &\vdots& \\ \dot{h}_n &=& h_1u_{1n} - \varepsilon H_{1n}, & \dot{H}_{1n} &=& h_n, \end{array}$$

i.e.,

$$\dot{h}_1 = -u_B^T h', \quad \dot{h}' = h_1 u - \varepsilon H', \quad \text{and} \quad \dot{H}' = h'.$$

The function μ is Monroy's *switching function.* The value of the Hamiltonian is

$$\mathcal{H}_{u(t)}(g(t), p(t)) = h_1 + \kappa_0 \|H'\| = h_1 + \kappa_0 \mu \equiv \lambda \in \{0, 1\}.$$

If $\mu(t) > 0$ then the optimal control $u_B(t)$ is uniquely determined:

$$u_B = \begin{pmatrix} u_{12} \\ \vdots \\ u_{1n} \end{pmatrix} = \frac{\kappa_0}{\mu} H'.$$

Thus if $H' \neq 0$, then our ODE for (h_1, h', H') is

$$\dot{h}_1 = -\frac{\kappa}{\mu}(h')^T H', \quad \dot{h}' = \left(\frac{\kappa h_1}{\mu} - \varepsilon\right) H', \quad \dot{H}' = h'. \tag{3}$$

Besides the Hamiltonian $\mathcal{H}_u = h_1 + \kappa_0 \mu$ there are additional integrals of motion, see [8, Theorem 4.3.1]:

$$\begin{aligned} J_2 &= \sum_{i=1}^n h_i^2 + \varepsilon \sum_{1 \le i < j \le n} H_{ij}^2, & (4) \\ J_3 &= \sum_{1 \le i < j < k \le n} (h_i H_{jk} - h_j H_{ik} + h_k H_{ij})^2 \\ &\quad + \varepsilon \sum_{1 \le i < j < k < l \le n} (H_{ij} H_{kl} - H_{jl} H_{ik} + H_{il} H_{jk})^2. & (5) \end{aligned}$$

Since $H_{ij} \equiv 0$ for $1 < i < j$ along $p(t)$, this simplifies significantly:

$$\begin{aligned} J_2 &= h_1^2 + \|h'\|^2 + \varepsilon \|H'\|^2 = h_1^2 + \|h'\|^2 + \varepsilon \mu^2, & (6) \\ J_3 &= \sum_{1 < j < k \le n} \det \begin{pmatrix} h_j & H_{1j} \\ h_k & H_{1k}^2 \end{pmatrix}^2. & (7) \end{aligned}$$

Remark 3.2 Unfortunately, Monroy does not prove that J_3 is an integral in the euclidean case. He only gives a proof for the non-euclidean cases.

Nevertheless J_3 is an integral in the euclidean case, too. We can prove a posteriori that not only J_3 is constant along $p(t)$ but even better each term

$$\det \begin{pmatrix} h_j & H_{1j} \\ h_k & H_{1k} \end{pmatrix} = h_j H_{1k} - h_k H_{1j}$$

is constant. We know that $\dot{h}_j = h_1 u_{1j} - \varepsilon H_{1j}$ and $\dot{H}_{1j} = h_j$. Moreover we know that the equality $H' = \mu \kappa_0^{-1} u$ holds, regardless whether $\mu > 0$ or $\mu = 0$. Hence we obtain

$$\begin{aligned} \frac{d}{dt}(h_j H_{1k} - h_k H_{1j}) &= \dot{h}_j H_{1k} + h_j \dot{H}_{1k} - \dot{h}_k H_{1j} - h_k \dot{H}_{1j} \\ &= (h_1 u_{1j} - \varepsilon H_{1j}) H_{1k} + h_j h_k - (h_1 u_{1k} - \varepsilon H_{1k}) H_{1j} - h_k h_j \\ &= h_1 (u_{1j} H_{1k} - u_{1k} H_{1j}) = h_1 \det \begin{pmatrix} u_{1j} & H_{1j} \\ u_{1k} & H_{1k} \end{pmatrix} = \frac{h_1 \mu}{\kappa_0} \det \begin{pmatrix} u_{1j} & u_{1j} \\ u_{1k} & u_{1k} \end{pmatrix} \equiv 0. \end{aligned}$$

Sussmann observed that for a really satisfying treatment (from an abstract point of view) one would have to consider the exterior product $\bigwedge \mathbb{R}^n$. Our previous computation shows that $h' \wedge H' \in \bigwedge \mathbb{R}^n$ is a constant of motion. And $\sqrt{J_3}$ is just the length of this vector in an appropriate norm. In dimension 3 one can avoid speaking of exterior products because $\bigwedge \mathbb{R}^3 \cong \mathbb{R}^3$, the isomorphism being essentially the vector product. But this does not work any more in higher dimensions.

In the sequel we will call a pair $(g(t), p(t))$ an extremal if it satisfies the necessary conditions of the (PMP). We do not need to specify the control $u(t)$ explicitly because $\dot{g} = g\, u$ a.e. implies $u = g^{-1}\dot{g}$.

4 Smooth extremals with torsion

Lemma 4.1 *Let (g,p) be an extremal with $J_3 > 0$. Then $\mu(t) > 0$ for all $t \in \mathbb{R}$ and $g(t), p(t), u(t)$ are analytic.*

Proof. If $\mu(t_0) = 0$ for some $t_0 \in \mathbb{R}$, then $H'(t_0) = 0$, whence $J_3(p(t_0)) = 0$. Since J_3 is constant along $p(t)$, this implies $J_3 \equiv 0$. Since we assumed $J_3 > 0$, this is impossible, hence $\mu(t) > 0$ for all t follows.

Since the righthand side of the ODE (3) is well-defined and analytic whenever $\mu > 0$, $p(t)$ must be analytic. As $\mu(t) > 0$ yields $u_B(t) = \kappa_0 \mu^{-1}(t) H'(t)$, the control u is analytic as well as $g(t)$. □

A short glance at the ODE (3) shows that all higher order derivatives of h', H' are linear combinations of h', H'. Thus we obtain:

Corollary 4.2 *The vectors $h'(t), H'(t) \in \mathbb{R}^{n-1}$ always stay in a fixed, at most two-dimensional subspace $V_0 \subseteq \mathbb{R}^{n-1}$.*

Proof. Take $t_0 \in \mathbb{R}$ arbitrary but fixed and let $V_0 = \mathbb{R}h'(t_0) + \mathbb{R}H'(t_0) \subseteq \mathbb{R}^{n-1}$. Take $v_0 \in V_0^\perp$ arbitrary but fixed. Since h', H' are analytic, the functions

$$f(t) = \langle v_0, h'(t)\rangle, \quad g(t) = \langle v_0, H'(t)\rangle$$

are analytic, too. In view of the ODE (3) all derivatives of h', H' at t_0 lie in V_0. Since v_0 is perpendicular to V_0, all derivatives of f and g at t_0 must vanish. Since f, g are analytic, they must vanish identically. Thus $h'(t), H'(t) \subseteq v_0^\perp$. Since $v_0 \in V_0^\perp$ was arbitrary, we deduce $h'(t), H'(t) \in V_0$ for all t. $\square$

Since the control $u(t)$ is given by $u_B = \kappa_0 \mu^{-1} H'$, this implies that $u_B(t)$ lies in a two-dimensional subspace of $\mathbb{R}^{n-1}$.

Lemma 4.3 *We may assume w.l.o.g. that* $u_B = (u_{12}, u_{13}, 0, \dots, 0)$.

Proof. Since $u_B(t)$ is contained in a fixed two-dimensional subspace $V_0 \subseteq \mathbb{R}^{n-1}$, we can find a $Q \in \mathrm{SO}(n-1)$ such that $QV_0 = \mathbb{R}^2 \times \{0\} \subseteq \mathbb{R}^{n-1}$. We have

$$\dot{g} = gX_{u_B}, \quad \text{with} \quad X_{u_B} = \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & u_B^T \\ 0 & -u_B & 0 \end{pmatrix}.$$

Let $k = \mathrm{diag}(1, 1, Q^T) \cong \mathrm{SO}(n-1) \cong K$ and set $\tilde{g} := gk$. Then by Lemma 2.2 $\tilde{g}$ projects onto the same curve as g, and $\dot{\tilde{g}} = \tilde{g}\,\mathrm{Ad}(k^{-1})X_{u_B}$. Since $\mathrm{Ad}(k^{-1})X_{u_B} = X_{Qu_B}$ and $Qu_B \in \mathbb{R}^2 \times \{0\} \subseteq \mathbb{R}^{n-1}$, our claim follows. $\square$

Thus we obtain:

Theorem 4.4 *If (g, p) is an extremal with $J_3 > 0$, then the arc $x(t) = g(t)e_1$ is three-dimensional.*

Proof. By the previous lemma we may assume w.l.o.g. that the controls $u_{14}, \dots, u_{1n}$ vanish identically. Thus $u(t) \in \mathrm{span}\{\ell_1, L_{12}, L_{13}\}$. Let $\mathfrak{g}_3$ denote either $\mathfrak{se}(3), \mathfrak{so}(1,3)$, or $\mathfrak{so}(4)$, i.e., the Lie algebra one has to consider for the three-dimensional Dubins' problem sitting in the upper 4 by 4-block of $\mathfrak{g}$, and let G_3 denote the corresponding subgroup of G. Since the Lie algebra generated by $\{\ell_1, L_{12}, L_{13}\}$ is $\mathfrak{g}_3$, we deduce that $g(t) \in g(0)G_3$. Thus $x(t)$ is a solution of the three-dimensional problem shifted by $g(0) \in G$ into a different position in $\mathbb{R}^n, \mathbb{H}^n$, resp. $\mathbb{S}^n$. $\square$

The extremals with $J_3 > 0$ correspond to smooth three-dimensional arcs with constant curvature κ_0 and nonvanishing torsion τ. Monroy has actually proved [8, Theorem 6.2.1]:

Theorem 4.5 (Monroy's Torsion Formula) *Let (g,p) be an extremal with $J_3 > 0$. Then the corresponding arc has constant curvature κ_0 and its torsion is*

$$\tau(t) = -\frac{\sqrt{J_3}}{\mu(t)^2}$$

where again $\mu(t) = \|H'(t)\|$ is the switching function.

For more information on these arcs we refer to [11] and [8]. Sussmann has proved that these arcs are actually strict local minimizers and that the torsion cannot be arbitrarily large. In $\mathbb{R}^3$ with $\kappa_0 = 1$ one can actually show that $\mu \equiv \mu^* > 0$ is possible. Thus the torsion is constant too, $\tau \equiv \tau^* < 0$. But further analysis shows that only $\tau^* \in [-1, 0)$ is possible. In particular $\tau^* = -1$ is achieved if and only if $\lambda = 0$ and $J_2^2 = 4J_3 > 0$. This seems to be noteworthy because it shows that in higher dimensions the abnormal extremals yield arcs that cannot be obtained (not even locally) from regular extremals.

Monroy solves the ODE for p as far as this is possible, i.e., up to quadratures. For $\mathbb{R}^3$ he even derives formulas expressing the arc $x(t) \in \mathbb{R}^3$ in terms of the switching function μ and $\int \mu(t)\, dt$. He also derives an ODE for μ thus proving that μ may be expressed in terms of elliptic functions.

Finally we observe that again the hyperbolic problem has an extraordinary property. If we consider $\varepsilon = -1$, $\kappa_0 = 1$ and $J_3 > 0$, $\lambda = 0$, then the ODE (3) yields $\dot{h}' = 0$, hence $h'(t)$ is constant, nonzero (because of $J_3 > 0$), and $H'(t) = a + tb$ with constant, linear independent vectors $a, b \in \mathbb{R}^{n-2}$, and $h_1 = \mu = \|H'\|$. In this case it is even possible to integrate the ODE for $g(t)$ in terms of elementary functions, see [4].

5 Extremals without torsion: the switching pattern

The function $\mu(t)$ is called the switching function because its zeros correspond to discontinuities of the controls, i.e., switches occur only when $\mu = 0$. The case $J_3 = 0$ corresponds to the planar Dubins' problem, it yields the arcs following Dubins' pattern.

Lemma 5.1 *If (g,p) is an extremal with $J_3 = 0$ and I is an interval where $\mu(t) > 0$, then $u(t)$ is constant in I.*

Proof. Since $\mu > 0$ in I we have $u = \mu^{-1}\kappa_0 H'$. Of course at least one component of H' does not vanish identically. We simply differentiate

$$\frac{d}{dt}\frac{H_{1j}}{H_{1k}} = \frac{\dot{H}_{1j}H_{1k} - H_{1j}\dot{H}_{1k}}{H_{1k}^2} = \frac{h_j H_{1k} - H_{1j}h_k}{H_{1k}^2} = \frac{1}{H_{1k}^2}\det\begin{pmatrix} h_j & H_{1j} \\ h_k & H_{1k}\end{pmatrix}.$$

The determinant vanishes because of $J_3 = 0$, so $H_{1j} = c_{jk} H_{1k}$ with a constant c_{jk}. Hence $H'(t) = H_{1k}(t)(c_{2k}, c_{3k}, \ldots, c_{n,k})^T$ and $u(t) = \kappa_0 \frac{1}{\|H'\|} H'$ is constant in I. Or equivalently we can write $H'(t) = \mu(t) c_I$ with a constant vector c_I of length $\|c_I\| = 1$, and $u_B \equiv \kappa_0 c_I$. □

The corresponding arc will be a circular arc. Geodesic arcs appear only in the following situation, cf. [8, Propositions 4.6.1 and 4.6.2].

Lemma 5.2 *If (g, p) is an extremal such that $\mu(t) \equiv 0$ in an open interval I, then $J_3 = 0$, $J_2 = 1$, $\lambda = 1$ and in I we have*

$$h_1 \equiv 1, \quad h' \equiv 0, \quad H' \equiv 0, \quad \textit{and} \quad u \equiv 0.$$

Proof. If $\mu \equiv 0$ in I, then $H' \equiv 0$ because $\|H'\| = \mu$. Thus $0 \equiv \dot{H}' = h'$ yields $h' \equiv 0$. Since $p(t) \not\equiv 0$ by the (PMP), $h_1 \neq 0$ follows. Therefore $0 \equiv \dot{h}' = h_1 u$ yields $u \equiv 0$. Finally $h_1 + \kappa_0 \mu - \lambda \equiv 0$ yields $h_1 \equiv \lambda \in \{0, 1\}$. As $h_1 \neq 0$, $h_1 = \lambda = 1$ follows. Evaluating J_2, J_3 at $(h_1, h', H') = (1, 0, 0)$ yields $J_3 = 0$ and $J_2 = 1$. □

Thus we obtain:

Corollary 5.3 *If (g, p) is an extremal with $J_3 = 0$, then the corresponding arc is a concatenation of circular arcs and geodesic segments.*

So locally these arcs are even planar. But since arbitrary concatenations will not be planar, in general, we need to determine the switching behavior of optimal controls. Monroy obtains the previous corollary from his torsion formula. But he does not determine the possible switching patterns. Besides he only considers $\kappa_0 = 1$. This does not affect the structure of optimal arcs in $\mathbb{R}^n$ and $\mathbb{S}^n$, but for $\mathbb{H}^n$ it makes a great difference. Since we also want to cover that case, we will have to adjust and refine some of his arguments. Since this requires a detailed analysis of the behavior of the switching function μ, we state the final result first. It is a slight generalization of Sussmann's three-dimensional euclidean results, cf. [11, Theorem 1].

Theorem 5.4 *Suppose that $\kappa_0^2 + \varepsilon > 0$. Let (g, p) be an extremal with $J_3 = 0$. Then the optimal control follows one of the following patterns:*

- *The control is $u(t) \equiv u_0$ is constant, $\|u_0\| = \kappa_0$. The corresponding arc is a circle and therefore genuinely planar.*
- *For a fixed u_0 with $\|u_0\| = \kappa_0$ the control $u(t)$ follows the pattern*

$$u_0 \vdash -u_0 \vdash u_0 \vdash -u_0 \vdash \ldots$$

and the time T^* between successive switches is always the same, between a half period and a full period:

$$T^* \in \left[\frac{\pi}{\sqrt{\kappa_0^2+\varepsilon}}, \frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}\right).$$

The corresponding arc $x(t)$ is a concatenation of circular arcs $C \ldots C$, all intermediate circles have the same length, they are at least semicircles, but shorter than a full circle, and all these arcs are in the same 'plane.' So the arc $x(t)$ is genuinely planar.

- *For arbitrary but fixed $u_0, u_1, u_2 \ldots$ with $\|u_i\| = \kappa_0$ the control $u(t)$ follows the pattern*
$$u_0 \vdash 0 \vdash u_1 \vdash 0 \vdash u_2 \vdash 0 \vdash u_3 \ldots$$
The intervals where $u(t) = 0$ may have arbitrary nonnegative length. The length of an interval between two switches where $u(t) \equiv u_i$, is an integer multiple of the full period $\frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}$. The corresponding arc is a concatenation of circles and geodesics $CLCLC\ldots$, and a C between two L's must be a full circle. Hence an optimal arc of this type has at most one L, so it must be CLC. Therefore it is at most three-dimensional

If $\kappa_0^2 + \varepsilon \leq 0$, the situation is very similar but actually simpler:

Theorem 5.5 *Suppose that $\kappa_0^2 + \varepsilon \leq 0$. Let (g, p) be an extremal with $J_3 = 0$. Then the optimal control $u(t)$ follows one of the following patterns:*

- *The control is $u(t) \equiv u_0$ is constant, $\|u_0\| = \kappa_0$. The corresponding arc is a circular arc C and therefore genuinely planar.*

- *There is exactly one switch. For a fixed u_0 with $\|u_0\| = \kappa_0$ there is $t_0 \in \mathbb{R}$ such that the control is*
$$u(t) = \begin{cases} u_0, & \text{if } t \in (-\infty, t_0) \\ -u_0, & \text{if } t \in (t_0, \infty). \end{cases}$$
The corresponding arc $x(t)$ is a concatenation of two circular arcs CC lying in the same 'plane'. So the arc $x(t)$ is genuinely planar.

- *There are two switches at $t_1 \leq t_2$. For arbitrary but fixed u_1, u_2 with $\|u_i\| = \kappa_0$ the control is*
$$u(t) = \begin{cases} u_1, & \text{if } t \in (-\infty, t_1), \\ 0, & \text{if } t \in (t_1, t_2), \\ u_2, & \text{if } t \in (t_2, \infty). \end{cases}$$
The corresponding arc is of the form CLC and therefore it is at most three-dimensional.

Remark 5.6 That a control following the alternating pattern $u_0 \vdash -u_0 \ldots$ yields a 'planar' arc follows from the fact that the subalgebra generated by $\{\ell_1, X_{u_0}\}$ is isomorphic to $\mathfrak{se}(2), \mathfrak{so}(1,2)$, resp. $\mathfrak{so}(3)$. So the same argument as in the proof of Theorem 4.4 shows that $x(t)$ is a planar curve.

The geometric meaning of the condition $\kappa_0^2 + \varepsilon > 0$, resp. ≤ 0 is quickly explained. Let

$$X_u = \begin{pmatrix} 0 & -\varepsilon & 0 \\ 1 & 0 & -u^T \\ 0 & u & 0 \end{pmatrix}, \quad \text{with } \|u\| = \kappa_0.$$

Then $X_u^3 = -(\kappa_0^2 + \varepsilon)X_u$. If $\kappa_0^2 + \varepsilon > 0$, then $\exp(\mathbb{R}X_u)$ is a circle group. With $w := \sqrt{\kappa_0^2 + \varepsilon}$ we obtain:

$$\exp(tX_u) \;=\; I + \frac{\sin(wt)}{w} X_u + \frac{1 - \cos(wt)}{w^2} X_u^2.$$

And the period is $T^* = \frac{2\pi}{w}$, i.e., $\exp(T^* X_u) = I$. The arc $\exp(tX_u)e_1$ is a circular arc. And an arc over the full period T^* is a full circle.

In the euclidean and spherical case, we always have $\kappa_0^2 + \varepsilon > 0$. In the hyperbolic case $0 < \kappa_0^2 + \varepsilon = \kappa_0^2 - 1$ is equivalent to $\kappa_0 > 1$. This is the controllable hyperbolic case, cf. [5].

In the noncontrollable hyperbolic case, i.e., $\varepsilon = -1$ and $\kappa_0 \leq 1$ we have $X_u^3 = (1 - \kappa_0^2)X_u$ and therefore with $w := \sqrt{1 - \kappa_0^2}$ we obtain:

$$\begin{aligned} \exp(tX_u) &= I + X_u + \frac{1}{2} X_u^2, && \text{if } \kappa_0 = 1, \\ \exp(tX_u) &= I + \frac{\sinh(wt)}{w} X_u + \frac{\cosh(wt) - 1}{w^2} X_u^2, && \text{if } \kappa_0 < 1. \end{aligned}$$

So in this case the 'circular arcs' are never closed. Thinking of traveling in a car at constant speed this means it is impossible to return to the starting point. It was proved in [5] that in this case Dubins' problem in $\mathbb{H}^n$ is not controllable, i.e., there are boundary conditions x_0, v_0, x_1, v_1 for which it is impossible to find an admissible arc.

6 Analyzing the switching behavior

In order to obtain this information about the extremals we must take a closer look at the switching function. Monroy has proved that the switching function μ satisfies a certain ODE, see [8, Theorem 4.5.1]. Since we consider $\kappa_0 > 0$ rather than just $\kappa_0 = 1$, we must modify his proof slightly (p. 39, first line, we have $h_1 = \lambda - \kappa_0 \mu$ rather than $h_1 = \lambda - \mu$) and obtain

$$(\mu\dot{\mu})^2 = -(\kappa_0^2 + \varepsilon)\mu^4 + 2\kappa_0 \lambda \mu^3 + (J_2 - \lambda^2)\mu^2 - J_3 =: f(\mu). \tag{8}$$

For $J_3 = 0$, this ODE simplifies to

$$\mu^2(\dot\mu)^2 = \mu^2\left(-(\kappa_0^2+\varepsilon)\mu^2 + 2\kappa_0\lambda\mu + J_2 - \lambda^2\right) = \mu^2 F(\mu). \tag{9}$$

If $\mu > 0$ in an interval I, then either $\dot\mu = +\sqrt{F(\mu)}$ or $\dot\mu = -\sqrt{F(\mu)}$ because μ is smooth in I. Since $F(\mu)$ is a quadratic polynomial, the ODE is integrable in terms of elementary functions, cf. [8, Corollary 4.5.1]. We omit integration of this ODE. We note that μ is defined for all t. For the control is locally constant, thus $g(t)$ is just a product of exponentials. So $(g(t), p(t))$ are defined for all t, hence so is μ.

One consequence of this ODE is that μ cannot assume arbitrary values. The righthand side $\mu^2\,F(\mu)$ has to be nonnegative. Therefore the possible range of μ is

$$\mathcal{R} = \left\{\mu \mid \mu \geq 0,\ \mu^2F(\mu)\geq 0\right\} = \{0\}\cup\left(F^{-1}(\mathbb{R}^+)\cap(0,\infty)\right).$$

The case $F(0) = J_2 - \lambda^2 < 0$ is the simplest case.

Lemma 6.1 *If (g,p) is an extremal with $J_3 = 0$ and $J_2 - \lambda^2 < 0$, then $\mu(t) > 0$ for all t, the corresponding control $u(t)$ is a constant u_0 with $\|u_0\| = \kappa_0$, and the corresponding arc is a circular arc.*

Proof. If $F(0) < 0$, then $\mu^2F(\mu) < 0$ for $\mu > 0$ sufficiently small. Since μ is continuous, either $\mu(t) > 0$ for all t or $\mu(t) \equiv 0$ follows. If $\mu \equiv 0$, then Lemma 5.2 yields $J_2 = 1$ and $\lambda = 1$, hence $F(0) = 0$, a contradiction, so this is impossible. Thus $\mu(t)$ never vanishes, the control $u(t)$ is constant, and the corresponding arc $x(t)$ is a circular arc. $\square$

Next we consider what kind of information we get about switches. We claim that if $F(0) > 0$, then we can only have an alternating switching pattern, i.e., switches of the form $u_0 \vdash -u_0$.

Lemma 6.2 *Let (g,p) be an extremal with $J_3 = 0$ and $J_2 - \lambda^2 > 0$. Let $t_l < t^* < t_r$ be points such that $\mu(t^*) = 0$ and $\mu(t) > 0$ in $(t_l, t^*)\cup(t^*, t_r)$. Let u_l, u_r denote the values of the control in these intervals. Then*

$$u_r = -u_l.$$

Proof. Let I_l, I_r denote the two intervals. Since $\mu > 0$ in each of these either $\dot\mu = +\sqrt{F(\mu)}$ or $\dot\mu = -\sqrt{F(\mu)}$. Thus μ is either increasing or decreasing. As $\mu(t^*) = 0$, it must be decreasing I_l and increasing in I_r, hence

$$\dot\mu(t) = \begin{cases} -\sqrt{F(\mu(t))}, & \text{in } I_l,\\ +\sqrt{F(\mu(t))}, & \text{in } I_r.\end{cases}$$

Since μ is smooth in I_l, I_r we look at the one-sided limits of $\dot\mu$:

$$\lim_{t\to t^*-} \dot\mu(t) = -\sqrt{F(0)} < 0, \quad \text{and} \quad \lim_{t\to t^*+} \dot\mu(t) = \sqrt{F(0)} > 0.$$

Now we recall that in I_l, I_r we have $\mu(t)u = \kappa_0 H'(t)$ and therefore $\dot\mu u = \kappa_0 \dot H' = \kappa_0 h'$. Now h' is continuous. So we compute the one-sided limits at t^* and get

$$-\sqrt{F(0)}u_l = \lim_{t\to t^*-} \dot\mu(t)u_l = \kappa_0 h'(t^*) = \lim_{t\to t^*+} \dot\mu(t)u_r = \sqrt{F(0)}u_r.$$

Since $F(0) > 0$, this proves $u_r = -u_l$. □

We notice that if $F(0) = 0$, then we do not obtain any information about the values of the control, besides being constant in I_l, I_r. So our final task is to determine the length of an interval where $\mu(t) > 0$. This also makes sure that there cannot be clusters of isolated zeros of μ, cf. [8, Proposition 4.6.3].

We distinguish between the two cases $\kappa_0^2 + \varepsilon > 0$ and $\kappa_0^2 + \varepsilon \leq 0$.

Lemma 6.3 *Suppose that $\kappa_0^2 + \varepsilon \leq 0$. Let (g, p) be an extremal with $J_3 = 0$ and $F(0) \geq 0$. If $F(0) > 0$, then there is at most one alternating switch. If $F(0) = 0$, the control has the form $u_1 \vdash 0 \vdash u_2$.*

Proof. We do not need to integrate the ODE (9). It suffices to analyze the behavior of its flow. Since $\kappa_0^2 + \varepsilon \leq 0$, the graph of $F(\mu)$ is a parabola open to the top or a line. Since $F'(0) = 2\lambda\kappa_0 \geq 0$, we obtain $\mathcal{R} = [0, \infty)$ and $F(\mu) > 0$ for $\mu \in (0, \infty)$.

Now let t_0 be a point with $\mu(t_0) > 0$. Let I denote the largest interval containing t_0 where $\mu(t) > 0$. Then we have two possibilities.

If $\dot\mu(t_0) > 0$, then $\dot\mu = \sqrt{F(\mu)} > 0$ in I, thus μ is increasing in I and therefore $\mu(t) > 0$ for all $t > t_0$ follows. Hence there are no switches after t_0.

If $\dot\mu(t_0) < 0$, then $\dot\mu = -\sqrt{F(\mu)}$ in I, so μ will decrease until it reaches 0, say at t^* and $\mu(t) > 0$ for all $t < t^*$. Thus there are no switches before t^*.

Hence we obtain that if $\mu(t_0) > 0$, then t_0 may lie either before or after a switch, but never between two switches. In other words, if $t_1 < t_2$ are zeros of μ, then $\mu \equiv 0$ in $[t_1, t_2]$. By Lemma 5.2 this can happen only if $J_2 = 1$, $\lambda = 1$, i.e., $F(0) = 0$.

So if $F(0) > 0$, then μ has at most one zero t_1. Now Lemma 6.2 applies proving that at t_1 there is an alternating switch: $u_0 \vdash -u_0$

If $F(0) = 0$, then let $t_1 = \inf\{t \mid \mu(t) = 0\}$ and $t_2 = \sup\{t \mid \mu(t) = 0\}$. Then $\mu \equiv 0$ in $[t_1, t_2]$ follows as well as $\mu > 0$ in $(-\infty, t_1) \cup (t_2, \infty)$. So the corresponding control is

$$\begin{cases} u_1, & \text{in } (-\infty, t_1), \\ 0 & \text{in } (t_1, t_2), \\ u_2, & \text{in } (t_2, \infty). \end{cases}$$

□

This finishes the proof of Theorem 5.5.

Finally let us consider the case $\kappa_0^2+\varepsilon>0$ and $F(0)\geq 0$. We must compute the length of a maximal interval where $\mu>0$. Since F is a quadratic function with negative leading coefficient, and $F(0)\geq 0$, it factorizes as

$$F(\mu)=(\kappa_0^2+\varepsilon)(\mu-a)(b-\mu), \quad \text{with } a\leq 0\leq b.$$

We observe that

$$\int \frac{d\mu}{\sqrt{(\mu-a)(b-\mu)}}=\arcsin\left(\frac{\mu-\frac{b+a}{2}}{\frac{b-a}{2}}\right).$$

Of course we could integrate the ODE for μ, but actually all we need is the *time between switches*, i.e., the maximal length of an interval I such that $\mu(t)>0$ in I.

Lemma 6.4 *Suppose that $\kappa_0^2+\varepsilon>0$ and let (g,p) be an extremal with $J_3=0$ and $J_2-\lambda^2\geq 0$. Then every maximal interval I where $\mu(t)>0$ has length*

$$\ell(I)\in\left[\frac{\pi}{\sqrt{\kappa_0^2+\varepsilon}},\frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}\right].$$

More precisely $\ell(I)=\frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}$ if and only if $J_2-\lambda^2=0$, and $\ell(I)=\frac{\pi}{\sqrt{\kappa_0^2+\varepsilon}}$ if and only if $\lambda=0$.

Proof. Since $\mu(t)>0\in I$ and $\mathcal{R}=[0,b]$, $b>0$ follows, and $\mu(t)\in(0,b]$ for $t\in I$. Let $t_0\in I$ be such that $\mu(t_0)=b$. Then in I μ must be increasing for $t<t_0$ and decreasing for $t>t_0$. The time to decrease from b to 0 is

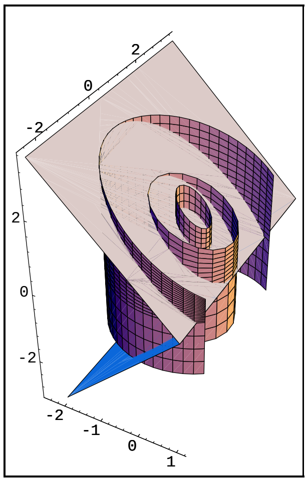
$$T^*=\int_b^0\frac{d\mu}{-\sqrt{F(\mu)}}=\int_0^b\frac{d\mu}{\sqrt{F(\mu)}}=\frac{1}{\sqrt{(\kappa_0^2+\varepsilon)}}\int_0^b\frac{d\mu}{\sqrt{(\mu-a)(b-\mu)}}.$$

Thus the length of I is

$$\ell(I) \;=\; 2T^*=\frac{2}{\sqrt{\kappa_0^2+\varepsilon}}\arcsin\left(\frac{\mu-\frac{b+a}{2}}{\frac{b-a}{2}}\right)\Bigg|_0^b=\frac{2}{\sqrt{\kappa_0^2+\varepsilon}}\left(\frac{\pi}{2}+\arcsin\frac{b+a}{b-a}\right).$$

From $F(\mu)=(\kappa_0^2+\varepsilon)(\mu-a)(b-\mu)=(\kappa_0^2+\varepsilon)(-\mu^2+(a+b)\mu-ab)$ we obtain $a+b=\frac{2\lambda\kappa_0}{\kappa_0^2+\varepsilon}\geq 0$. Thus

$$\ell(I)\in\left[\frac{\pi}{\sqrt{\kappa_0^2+\varepsilon}},\frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}\right].$$



the surfaces

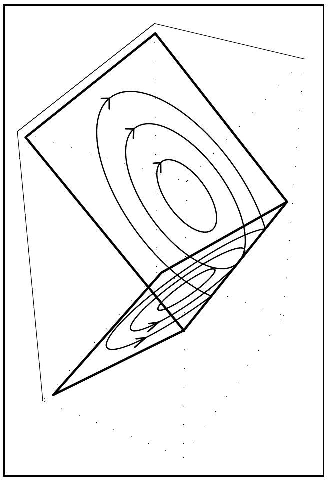

the flow

Figure 3: Coadjoint orbits and the flow of $p(t)$ in $\mathfrak{se}(2)^*$ for $\lambda = 1$.

The left endpoint is obtained if and only if $a+b=0$. Since $\kappa_0 > 0$, this happens if and only if $\lambda = 0$. The right endpoint is attained when $a = 0$. This is equivalent to $0 = F(0) = J_2 - \lambda^2$, of course. $\square$

In particular we obtain that the zeroes of μ are essentially isolated (cf. [8, Proposition 4.6.3]).

Corollary 6.5 *If (g,p) is an extremal with $J_3 = 0$ and $J_2 - \lambda^2 > 0$, then the corresponding control is alternating and the time between successive switches is a constant $T^* \in \left[\frac{\pi}{\sqrt{\kappa_0^2+\varepsilon}}, \frac{2\pi}{\sqrt{\kappa_0^2+\varepsilon}}\right)$.*

Proof. From the previous lemma we obtain that μ has only isolated zeros. So Lemma 6.2 shows that the control follows the alternating pattern $u_0 \vdash -u_0 \vdash u_0 \vdash \ldots$, etc. And the time between successive switches is given by the formula of the previous lemma. $\square$

Corollary 6.6 *If (g,p) is an extremal with $J_3 = 0$ and $J_2 = \lambda = 1$, and if the corresponding arc x is optimal, then it is of type CLC.*

Proof. The control u follows the pattern $u_0 \vdash 0 \vdash u_1 \vdash 0 \vdash u_2 \ldots$, and by Lemma 6.4 the time between two '0's is exactly a full period of X_{u_i}. So the corresponding arc is $CLCLC\ldots$, where each intermediate circle is a full circle. Hence an optimal arc of that type has at most one geodesic segment, i.e., it is of type CLC. $\square$

This finishes the proof of Theorem 5.4.

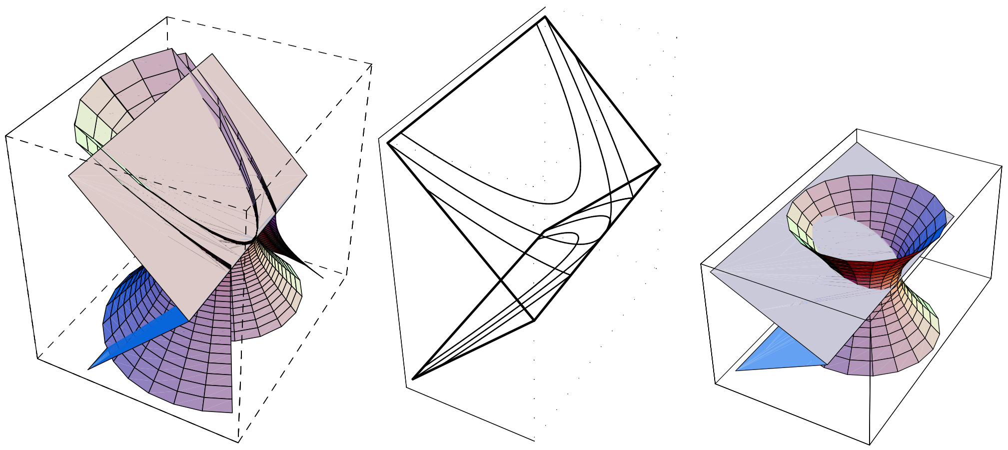


Figure 4: Coadjoint orbits and the flow of $p(t)$ in $\mathfrak{so}(1,2)^*$ for $\lambda = 1$.

Remark 6.7 Even though integration of the ODE for μ is not difficult, we refer to [8, Corollary 4.5.1] for the formulas (in the special case $\kappa_0 = 1$). The really interesting geometric information is to be found in the covector $p(t) \in \mathfrak{g}^*$. W.l.o.g. we only look at (h_1, h_2, H_{12})—this corresponds to the two-dimensional problem. The level surfaces of $J_2 = h_1^2 + h_2^2 + \varepsilon H_{12}^2$ are essentially the orbits of the coadjoint action of G on $\mathfrak{g}^*$. In $\mathfrak{se}(2)$ these are cylinders around the (vertical) H_{12}-axis, in $\mathfrak{so}(3)$ these are spheres, and in $\mathfrak{so}(1,2)$ these are one-sheeted hyperboloids, the upper or lower half of a two-sheeted hyperboloid or one half of a Lorentzian double cone (excluding the vertex at 0). The level surfaces of the optimal Hamiltonian

$$\mathcal{H}(g,p) = h_1 + \kappa_0 \mu - \lambda \equiv 0, \quad \lambda \in \{0,1\},$$

are the boundary of a (shifted) wedge in three-space with two faces. On each face the control is constant either $+\kappa_0$ or $-\kappa_0$. The switches occur when $p(t)$ hits the edge of this wedge. An alternating switch means that p passes from the upper to the lower face (and vice versa). Projecting into the (h_1, h_2)-plane one can interpret the time between switches as the euclidean angle between the origin and the two intercepts of p with the edge of the wedge. Figure 6 shows the coadjoint orbits and the flow of $p(t)$ in $\mathfrak{se}(2)^*$. In $\mathfrak{so}(3)^*$ one has to replace the cylinders by spheres, but the flow essentially looks the same.

In the hyperbolic case the geometric difference between the noncontrollable and the controllable case is neatly visualized. If $\kappa_0 \leq 1$, then the intersection of $\mathcal{H} = 1$ with a one-sheeted hyperboloid is never closed—so we never hit the edge of the wedge twice. On the other hand if $\kappa_0 >$, this intersection is always a closed curve, as in the euclidean and spherical case. And the flow of $p(t)$ looks alike.

References

[1] L. E. Dubins. On curves of minimal length with a constraint on average curvature and with presribed initial and terminal positions and tangents. *Am. J. Math.*, 79:497–516, 1957.

[2] Velimir Jurdjevic. Non Euclidean Elastica. *American Journal of Mathematics*, 117:93–124, 1995.

[3] Jerrold E. Marsden and Tudor Ratiu. *Introduction to Mechanics and Symmetry*, volume 17 of *Texts in Applied Mathematics*. Springer, New York, 1994.

[4] Dirk Mittenhuber. Dubins' problem in $\mathbb{H}^3$, integration of abnormal extremals. *in preparation*.

[5] Dirk Mittenhuber. Dubins' problem in hyperbolic space. *submitted*.

[6] Dirk Mittenhuber. Applications of the Maximum Principle to Problems in Lie Semigroups. In Karl H. Hofmann, Jimmie D. Lawson, and Ernest B. Vinberg, editors, *Semigroups in Algebra, Geometry, and Analysis*, pages 311–336. de Gruyter, Berlin, 1995.

[7] Felipe Monroy Pérez. Non-Euclidean Dubins' problem. *submitted*.

[8] Felipe Monroy Pérez. Non-Euclidean Dubins' problem: A control theoretic approach. PhD thesis, University of Toronto, 1995.

[9] John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, Berlin, New York, 1994.

[10] J.A. Reeds and L.A. Shepp. Optimal paths for a car that goes both forwards and backwards. *Pacific Journal of Mathematics*, 145:367–393, 1990.

[11] Héctor J. Sussmann. Shortest 3-dimensional paths with a prescribed curvature bound. In *Proceedings of the 1995 IEEE Conference on Decision and Control*, pages 3306–3312, 1995.

[12] Héctor J. Sussmann. Shortest paths with a prescribed bound on the curvature: the three-dimensional case. in preparation.