2020 European Control Conference (ECC)
May 12-15, 2020. Saint Petersburg, Russia

A novel formalisation of the Markov-Dubins problem

Bevilacqua, P., Frego, M., Fontanelli, D. *Member, IEEE,* Palopoli, L., *Member, IEEE*

***Abstract*— The Markov-Dubins problem requires to find the shortest path that connects an initial point and angle to a final point and angle with bounded turning radius. Formally, this is equivalent to solve an interpolation problem with continuity up to the first derivative and with bounded curvature. We propose a mathematical framework that models with a single equation the different cases that arise, i.e., we can represent with the same function an arc of circle or a line segment by smoothly blending from one to the other. This allows us to restate the problem as a standard Mixed Integer Nonlinear Programming (MINLP), which can be relaxed into a standard Nonlinear Programming (NLP) and therefore opens the way to solve it using off-the-shelf solvers. Moreover, our formalism captures the symmetries of the problem in a more intuitive way with respect to previous works, thanks to the considered conformal bipolar transform. This approach is suitable for an effective solution of the extended problem of connecting multiple points, that will be addressed in future research.**

I. Introduction

The Markov-Dubins path [1] is the shortest C^1 and piecewise C^2 curve $\gamma : [0, L] \to \mathbb{R}^2$ between two assigned points P_i and P_f and angles ϑ_i and ϑ_f, such that its curvature (in absolute value) is not greater than $\kappa > 0$ at almost every point, see [2]. Notice that the length $L > 0$ of the path is also an unknown of the problem.
Let $\gamma(s) = (x(s), y(s))$ be a C^1 and piecewise C^2 parametric curve. The Markov-Dubins curve is the path that solves the Markov-Dubins Problem:
Minimize $L > 0$ subject to

$$\begin{aligned} \gamma(0) &= P_i = (x_i, y_i), & \gamma(L) &= P_f = (x_f, y_f) \\ \gamma'(0) &= (\cos\vartheta_i, \sin\vartheta_i), & \gamma'(L) &= (\cos\vartheta_f, \sin\vartheta_f), \\ |k(\ell)| &\le \kappa, & ||\gamma'(\ell)|| &= 1, \quad \ell \in [0, L], \end{aligned}$$

where the (signed) curvature $k(\ell)$ can be written as

$$k(\ell) = \frac{x'(\ell)y''(\ell) - y'(\ell)x''(\ell)}{\sqrt{x'(\ell)^2 + y'(\ell)^2}}.$$

The condition $||\gamma'(\ell)|| = 1$ ensures that the curve is parametrised by arclength, thus simplifying the curvature to $k(\ell) = \theta'(\ell)$, with $\theta(\ell)$ the heading of the curve, i.e. the angle with respect to the x axis, see Figure 1.

The problem has relevance in robotics because it can be interpreted as the shortest path joining two postures for a wheeled nonholonomic robot with bounded steering radius and constant forward velocity. Modern applications employ the solution to the point-to-point Markov-Dubins problem as a building block to address several variants and extensions of the problem, like for example the interpolation of a sequence of points with constraints on the maximum curvature [3], [4], [5], used for advanced planning methods [6], [7] and waypoint-following [8]. Another application is the extension of classic Euclidean Orienteering Problems to nonholonomic Dubins Orienteering Problems [9], [10] and to Dubins Travelling Salesman [11].

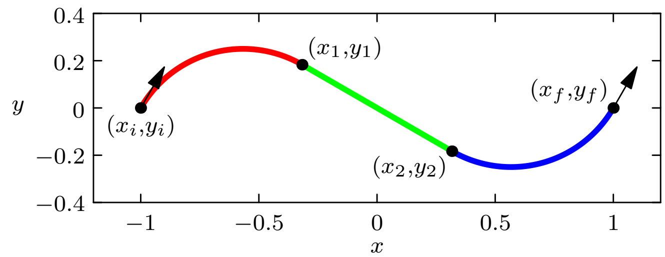


Fig. 1. The standard Markov-Dubins interpolation scheme between two configurations (x_i, y_i, ϑ_i) and (x_f, y_f, ϑ_f). The solution is a sequence of 3 segments with intermediate points (x_1, y_1) and (x_2, y_2).

The first complete solution to the problem was given by Dubins [12], who showed, by means of geometric arguments, that the optimal solution is a combination of at most three arcs of circle or line segments juxtaposed, see Figure 1. Since Markov and Dubins, the problem has received a lot of attention by a number of researchers in different fields and with different aims, from practical applications in science and engineering, to abstract theoretical studies. There are essentially two approaches to the problem, the first one based on purely geometrical arguments, the second one relying on optimal control theory. In the first group of papers, Cockayne and Hall [13] extended the results of Dubins studying the classes of trajectories of the problem. Shkel and Lumelsky [14] proposed explicit formulas to compute the solution of the Markov-Dubins problem. In the same paper, the authors go beyond the bruteforce method of testing the set of candidate solutions in order to find the optimal one, proposing a logical classification scheme to directly find the optimal manoeuvre, implemented in C++ in [15]. In the second group of papers, we mention (some of) the studies conducted by a team of researchers at INRIA [16], [17], [18], that looked at the Markov-Dubins problem from the perspective of the Pontryagin Maximum Principle (PMP) and the theory of Optimal Control (OCP). They opened the way to a series of papers that studied the properties of the problem under a new light [2], [19], [20].

P. Bevilacqua, M. Frego and L. Palopoli are with the Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento, Italy
D. Fontanelli and M. Frego are with the Dipartimento di Ingegneria Industriale, University of Trento, Italy

Paper contribution. As detailed below, the traditional formulation of the problem considers two different equations to model an arc of circle and a straight line segment, that is, the two main primitives compounding the optimal solution of the Markov-Dubins problem. In this paper, we show that *both* curve primitives can be modelled as instantiation of the *same* function. This fact brings two substantial advantages. The first one is that the resulting theoretical framework is more elegant and easier to treat. The second is represented by the smoothness of the transition between a circle and a line. This aspect is very important from a computational point of view because it is not handled with an "if" condition on the curvature, which becomes ill conditioned for very small values, e.g., when the circle is flat and tends to a line.
In view of these advantages, we can set up the Markov-Dubins problem within a standard optimisation framework, which paves the way for the application of standard off-the-shelf tools, as a Mixed Integer Nonlinear Programming (MINLP) or as its relaxed version, as a Nonlinear Programming (NLP). Importantly, this formulation scales well to the case of multipoint problems (i.e., designing a path that intersects a number of points in the configuration space). On the contrary, when the number of points grows, the brute-force solution *a la Dubins* becomes unfeasible because of the exponential number of cases to test.
As a final remark, our formulation captures the two symmetries of the optimal solution, thanks to the application of the conformal bipolar transform. This allows to reduce the search space for the optimal solution.

The paper is organised as follows: Section II introduces the problem from the point of view of optimal control. Section III discusses the proposed new formulation of the problem. Section IV formulates it as a standard MINLP or NLP optimisation. Section V illustrates the symmetries arising from the proposed formulation. Finally, Section VI validates the method by means of numerical experiments, while Section VII draws the conclusions, summarises the main results and gives new research perspectives.

II. Background

The OCP formulation of the Markov-Dubins problem is, [16], [17], [18], [2], [20]:

$$\text{Minimize } L = \int_0^L 1\,\mathrm{d}\ell \qquad \text{subject to: } |u(\ell)| \le \kappa,$$

$$\begin{aligned}
x'(\ell) &= \cos(\theta(\ell)), & y'(\ell) &= \sin(\theta(\ell)), & \theta'(\ell) &= u(\ell),\\
x(0) &= x_i, & y(0) &= y_i, & \theta(0) &= \vartheta_i,\\
x(L) &= x_f, & y(L) &= y_f, & \theta(L) &= \vartheta_f,
\end{aligned}$$

for the unknown control function $u(\ell)$ defined for the curvilinear abscissa $\ell \in [0, L]$. The choice of the differential equations for $x(s)$ and $y(s)$, introducing the heading function $\theta(\ell)$, ensures that $||\gamma'(\ell)|| = 1$. This definition of $x(\ell)$ and $y(\ell)$ yields $k(\ell) = \theta'(\ell)$, thus the requirement of bounded curvature $|k(\ell)| \le \kappa$ is equivalent to an optimal control $u(\ell) = k(\ell) = \theta'(\ell)$ that satisfies $|u(\ell)| \le \kappa$.

The Hamiltonian is $\mathcal{H} = \lambda_0 + \lambda_1 \cos\theta + \lambda_2 \sin\theta + \lambda_3 u$. The multipliers λ_1 and λ_2 are constant because the derivatives of $\mathcal{H}$ w.r.t. x and y are zero, thus we can re-write $\mathcal{H}$ as $\mathcal{H} = \lambda_0 + \rho\cos(\theta(\ell) - \phi) + \lambda_3 u(\ell)$, where $\rho := \sqrt{\lambda_1^2 + \lambda_2^2}$ and $\tan\phi := \lambda_2/\lambda_1$. The optimal control $u^\star$ is found minimising the Hamiltonian:

$$\begin{aligned}
u^\star &= \operatorname*{argmin}_u \mathcal{H} = \operatorname*{argmin}_u \lambda_3 u(\ell)\\
&= \begin{cases} \kappa & \text{if } \lambda_3(\ell) < 0,\\ -\kappa & \text{if } \lambda_3(\ell) > 0,\\ 0 & \text{if } \lambda_3(\ell) = 0 \text{ (singular case).} \end{cases}
\end{aligned}$$

The switching function is given by the derivative of $\mathcal{H}$ w.r.t. $\theta(\ell)$, i.e., $\lambda_3'(\ell) = \rho\sin(\theta(\ell) - \phi)$. If $\lambda_3(\ell) \ne 0$ then the control is $\pm\kappa$, which implies a path of constant curvature, e.g., a circle arc. If $\lambda_3(\ell) = 0$ over an interval, then also $\lambda_3'(\ell) = 0$ over that interval, being $\rho \ne 0$, then $\sin(\theta(\ell) - \phi) = 0$, which implies that $\theta(\ell)$ must be constant, thus $\theta'(\ell) = u(\ell) = 0$. Hence, in the singular case, the control is zero and the curve becomes a line segment.
The difficult part of the solution is to prove that the optimal solution is composed of *at most 3 pieces of curve*: a combination of circle arcs and line segments, see [12], [16], [17], [18], [2]. This sequences of curves can be written as words: the optimal solutions can be written as LSR, RLR, LSL, etc, where L, S and R represent a left turn, straight segment and right turn, respectively. A more compact notation is C for an arc of circle and S for a segment. Not all possible combinations of curves are optimal, it can be proved that the optimal words are only: $CSC, CS, SC, S, CCC, CC, C$. Thus there are 15 possible combinations: $LSL, LSR, RSL, RSR, RLR, LRL, LS, RS, SL, SR, S, LR, RL, L, R$. In practice, only the first six words are considered, the others can be obtained by setting to zero some of their arcs.

III. A new formulation

The Markov-Dubins problem is characterised by 7 parameters, namely the initial and final configurations (x_i, y_i, ϑ_i), (x_f, y_f, ϑ_f) and by the maximum allowed curvature κ. It is possible to reduce this number, without loss of generality, by applying a change of coordinates to a standard setting depending on only 3 parameters. In literature we can find two ways of achieving this. In [16], [17], [18], the Authors of Inria, using a classic change of variable in the controls community, formalise the problem so that $(x_i, y_i, \vartheta_i) = (0, 0, 0)$, i.e., the origin is the fixed initial configuration with maximum curvature $\kappa = 1$. Therefore, they study the problem for all the possible final configurations (x_f, y_f, ϑ_f). In [14], the Authors fix the initial point $(x_i, y_i) = (0, 0)$, rotate the problem to have the final point on the positive horizontal axis, and scale it in order to consider $\kappa = 1$. This implies that the final point becomes $(x_f, y_f) = (d, 0)$ for a certain $d > 0$. Their parameters are thus the initial and final angles ϑ_i, ϑ_f and the distance between the two points, d.
We propose yet another change of variable, a conformal

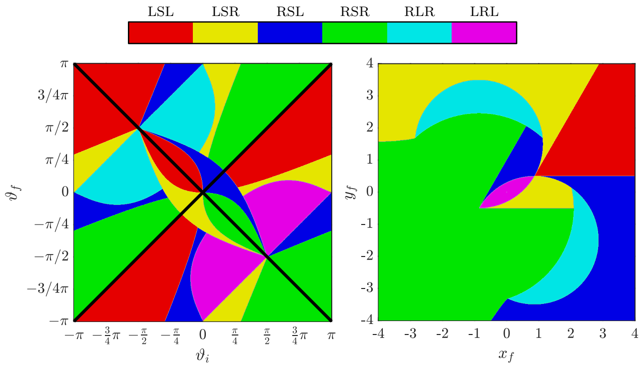


Fig. 2. Different symmetries between our approach (left) and Inria (right). Left: The proposed parameters space, e.g., the axes are the initial and final angles ϑ_i, ϑ_f for a fixed curvature ($\kappa = 1$ in this example). The different colours represent the 6 optimal solutions, see the colorbar for a specific reference. The black diagonal lines represent the axes of symmetry of our canonical space. Right: the parameters space proposed in [16], [17], [18], that is, the coordinates of the final point (x_f, y_f) for a fixed final angle ($\vartheta_f = \pi/3$ in this example).

bipolar coordinate system, i.e., a transform that does not change the measure of the angles. In particular we translate, rotate and scale the data of the problem so that the initial point is fixed to be $(x_i, y_i) = (-1, 0)$ and the final point is $(x_f, y_f) = (1, 0)$. Our problem parameters are thus the initial and the final angles ϑ_i, ϑ_f and the curvature $\kappa > 0$. This transform has the advantage of having a domain of the parameters that naturally exhibits the two symmetry axes of the problem, whereas, in the other approaches this becomes clear only after a deep analysis, see Figure 2 for a visual comparison. The symmetries of the problem are important to reduce the size of the parameters space.

A. Bipolar conformal coordinate system

The bipolar transform $\boldsymbol{B}$ is a composition of a translation, a rotation and a scaling. We introduce some notation to simplify the formulas, (see also [21], [22] for its application to related studies, e.g., Markov-Dubins problem with angular acceleration). Let

$$\begin{aligned} \Delta x &= x_f - x_i, & \lambda &= \frac{1}{2}\sqrt{\Delta x^2 + \Delta y^2}, & \hat{\vartheta}_i &= \vartheta_i - \varphi, \\ \Delta y &= y_f - y_i & \varphi &= \operatorname{atan2}(\Delta y, \Delta x), & \hat{\vartheta}_f &= \vartheta_f - \varphi, \end{aligned}$$

and $\hat{\kappa} = \lambda\kappa$. The transform $\boldsymbol{B}$ is invertible and is given by

$$\begin{aligned} \boldsymbol{B}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= \frac{1}{\lambda}\left(\begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}\right), \\ \boldsymbol{B}^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}\left(\lambda\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}\right). \end{aligned}$$

where $\operatorname{atan2}$ is the function that returns φ, i.e., the unique angle in $[-\pi, \pi)$ that satisfies $\bar{x} = x_i \cos\varphi + y_i \sin\varphi + \lambda$, $\bar{y} = -x_i \sin\varphi + y_i \cos\varphi$, where $\bar{x}$, $\bar{y}$ are the parameters of the translation along the axes.

B. The proposed formulation

A Markov-Dubins path is a piecewise defined smooth curve, with C^1 continuity at the junction points. Each piece is either a straight line or a circle arc. In literature, there are two different mathematical representations for these curves, here we propose a unified notation that allows us to model with the same equation both cases. Moreover, the transition between the two mathematical objects is handled smoothly, i.e., there are not switching conditions when the curvature goes to zero. This is accomplished by using the technique that we recently developed for the study of the biarcs [23], a curve closely related to the Markov-Dubins path. Our results are based on the introduction of the sinc function.

Definition 1: The function $\operatorname{sinc} : \mathbb{R} \to \mathbb{R}$ is defined as $\operatorname{sinc}(z) = \sin(z)/z$ for $z \neq 0$ and $\operatorname{sinc}(0) = 1$ for $z = 0$. This definition renders sinc a continuous function as we removed the singularity in zero. From a computational point of view, it is evaluated using its Taylor series

$$\operatorname{sinc}(z) = \frac{\sin z}{z} = 1 + \sum_{n=1}^{\infty}(-1)^n \frac{z^{2n}}{(2n+1)!},$$

for values around zero, and using the analytic expression far from zero. The numeric implementation near zero uses its Taylor approximation with a small number of terms. For example, to limit the error below $1.3 \cdot 10^{-20}$, we use

$$\left|\left(1 - \frac{z^2}{6}\left(1 - \frac{z^2}{20}\right)\right) - \frac{\sin z}{z}\right| \leq \frac{|z|^6}{5040}.$$

From the estimation of the remainder, this expression satisfies the above error for $|z| \leq 0.002$. Other properties needed in the sequel are given by the two following relations:

$$\frac{\sin z}{z} = \cos\frac{z}{2}\operatorname{sinc}\frac{z}{2}, \qquad \frac{1 - \cos z}{z} = \sin\frac{z}{2}\operatorname{sinc}\frac{z}{2}.$$

The proof is omitted as it follows from a direct application of the definition and of the formulas of bisection.

C. Modelling a line segment and a circle arc

With the tools introduced so far, we are able to state a first fundamental result that allows us to write with the same equation a segment of straight line or of a circle arc.

Lemma 1: The parametric equations of a line segment or of a circle arc, starting at the base point (x_0, y_0), with initial angle ϑ_0, constant curvature κ_0 (possibly zero) and length s_0 are given by

$$\begin{aligned} x(\ell) &= x_0 + \ell \operatorname{sinc}\left(\frac{\kappa_0 \ell}{2}\right)\cos\left(\vartheta_0 + \frac{\kappa_0 \ell}{2}\right) \\ &\equiv x_0 + f(\ell, \kappa_0, \vartheta_0), \\ y(\ell) &= y_0 + \ell \operatorname{sinc}\left(\frac{\kappa_0 \ell}{2}\right)\sin\left(\vartheta_0 + \frac{\kappa_0 \ell}{2}\right) \\ &\equiv y_0 + g(\ell, \kappa_0, \vartheta_0), \end{aligned} \qquad (1)$$

where the arclength parameter ℓ is in the range $[0, s_0]$.

Proof: The proof is a simple computation of the parametric equations of a clothoid, when, in the Fresnel integrals [21], the sharpness is equal to zero. Substituting the sinc function in the obtained expressions to remove the divisions by the curvature yields (1). ■

IV. Markov-Dubins Problem as MINLP or NLP

With the results of Lemma 1 we can write any Markov-Dubins path as the juxtaposition of three curve pieces that can represent either a circle arc or a line segment. In some cases the curve has less than three pieces but this is not a limitation for our formalism, as it is possible to smoothly set one or more of the lengths s_0, s_1, s_2 to zero. The j^{th} piece of a curve is characterised by the parameters: (x_j, y_j) the initial point, ϑ_j the initial angle, κ_j the curvature and s_j the length. The corresponding parametric equations are thus:

$$x(\ell) = x_j + f(\ell, \kappa_j, \vartheta_j), \quad y(\ell) = y_j + g(\ell, \kappa_j, \vartheta_j),$$

for $\ell \in [0, s_j]$ and $j = 0, 1, 2$. The curvatures are bounded by $|\kappa_j| \le \kappa$ for a (positive) maximum curvature κ, indeed it has been proved that the optimal path can have only $\kappa_j \in \{-\kappa, 0, \kappa\}$. Therefore, we introduce three new integer variables $\sigma_j \in \{-1, 0, +1\}$, so that $\kappa_j = \sigma_j \kappa$. Finding the optimal path that solves the Markov-Dubins problem, is equivalent to finding the solution of the MINLP in the variables σ_i and s_i that minimizes the total length of the curve subject to the C^0 and C^1 continuity conditions at the junction of each piece of the curve. The three equations are:

$$\begin{aligned} x_f &= x_i + f(s_0, \kappa\sigma_0, \vartheta_i) + f(s_1, \kappa\sigma_1, \vartheta_i + \kappa\sigma_0 s_0) \\ &\quad + f(s_2, \kappa\sigma_2, \vartheta_i + \kappa\sigma_0 s_0 + \kappa\sigma_1 s_1), \\ y_f &= y_i + g(s_0, \kappa\sigma_0, \vartheta_i) + g(s_1, \kappa\sigma_1, \vartheta_i + \kappa\sigma_0 s_0) \\ &\quad + g(s_2, \kappa\sigma_2, \vartheta_i + \kappa\sigma_0 s_0 + \kappa\sigma_1 s_1), \\ \vartheta_f &= \vartheta_i + \sigma_0 \kappa s_0 + \sigma_1 \kappa s_1 + \sigma_2 \kappa s_2, \end{aligned} \tag{2}$$

where (x_i, y_i), (x_f, y_f), ϑ_i, ϑ_f are, respectively, the given initial and final points and angles, finally $\kappa > 0$ is the assigned maximum curvature.

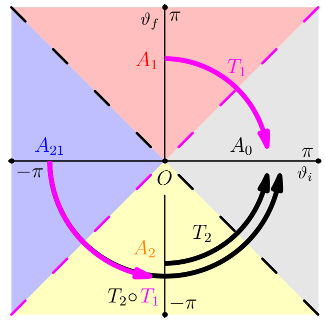
Remark 2: The third equation of (2) represents the C^1 continuity and must be intended modulo 2π. Therefore, for computational reasons, it is implemented with two constraints. Equality is enforced taking the cosine and sine of both members. See also Remark 6 of [20].

There are thus four equations in the six unknowns σ_j and s_j, for $j = 0, 1, 2$. The resulting MINLP is therefore

$$\begin{aligned} &\text{minimize} \quad s_0 + s_1 + s_2 \quad \text{subject to (2)} \\ &\text{with} \quad \sigma_i \in \{-1, 0, +1\},\ s_j \in \mathbb{R}^+,\ j = 0, 1, 2 \end{aligned} \tag{3}$$

and assigned boundary conditions (for $L = s_0 + s_1 + s_2$):

$$\begin{aligned} x(0) &= x_i, \quad y(0) = y_i, \quad \vartheta(0) = \vartheta_i; \\ x(L) &= x_f, \quad y(L) = y_f, \quad \vartheta(L) = \vartheta_f. \end{aligned} \tag{4}$$



A_0	A_1	A_2	A_{21}
LSL	RSR	LSL	RSR
LSR	LSR	RSL	RSL
RSL	RSL	LSR	LSR
RSR	LSL	RSR	LSL
RLR	LRL	RLR	LRL
LRL	RLR	LRL	RLR

Fig. 3. Left: the angle domain partitioned into the four regions A_0, A_1, A_{21}, A_2 and the reflection axes relative to T_1 and T_2, which are used to map the original data into the canonic domain. Right: look-up table to restore the original solution from the canonical.

The MINLP (3) can be relaxed to the NLP given in (5), by allowing the σ_j variables to belong to the convexified domain $\sigma_j \in [-1, 1]$. The optimisation problem becomes therefore:

$$\begin{aligned} &\text{minimize} \quad s_0 + s_1 + s_2 \quad \text{subject to (2)} \\ &\text{with} \quad \sigma_j \in [-1, 1],\ s_j \in \mathbb{R}^+,\ j = 0, 1, 2, \end{aligned} \tag{5}$$

with the same boundary conditions (4).

V. Symmetries and Domains

Figure 2 suggests that the parameters domain naturally splits into four regions, that are related by symmetry rules. For a fixed curvature κ, we partition the interval $[-\pi, \pi) \times [-\pi, \pi)$ along the diagonals $\vartheta_f = \pm\vartheta_i$. Formally, we identify with A_0 the triangular domain of $|\vartheta_f| \le \vartheta_i$, for $\vartheta_i \in [0, \pi)$; A_1 is the region $|\vartheta_i| \le \vartheta_f$, for $\vartheta_f \in [0, \pi)$; A_{21} is the region $|\vartheta_f| \le -\vartheta_i$, for $\vartheta_i \in [-\pi, 0]$; finally, A_2 is the region $|\vartheta_i| \le -\vartheta_f$, for $\vartheta_f \in [-\pi, 0]$, see Figure 3. One symmetry regards the fact that swapping the angles is equivalent to the original curve reflected along both axis. We refer to this transform as T_1. The second symmetry, T_2, swaps the angles with opposite signs, corresponding to a reflection along the y axis. The composition $T_2 \circ T_1$ is equivalent to the original curve reflected along the x axis. The matrices associated to these transforms are:

$$T_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad T_2 = -T_1, \quad T_2 \circ T_1 = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Using these relations it is possible to consider, instead of the full domain, the region A_0 only. The above transforms can be used to map the angles of the general problem to the canonic subset A_0, according to the scheme depicted in Figure 3 (left). To restore the canonical solution to the original problem we simply apply the look-up table presented in Figure 3 (right), that maps canonical solutions into the full domain, depending on the original data.

Modelling the parameters space with the angles and curvatures, a natural setting for viewing the sets of optimal Markov-Dubins words is as a ball of radius κ and coordinates ϑ_i and ϑ_f. In Figure 4 we show two spheres of radii $\kappa = 1$ and $\kappa = 4$. Explicitly determining the boundaries between