OPTIMAL ATTITUDE CONTROL WITH TWO ROTATION AXES

YULY BILLIG

Abstract. Euler proved that every rotation of a 3-dimensional body can be realized as a sequence of three rotations around two given axes. If we allow sequences of an arbitrary length, such a decomposition will not be unique. In this paper we solve an optimal control problem minimizing the total angle of rotation for such sequences. We determine the list of possible optimal patterns that give a decomposition of an arbitrary rotation. Our results may be applied to the attitude control of a spacecraft with two available axes of rotation.

1. Introduction

In this paper we investigate the problem of optimal attitude control of a 3-dimensional body which can be rotated around two fixed axes. The problem goes back to Euler [2] who proved in 1776 that an arbitrary rotation g of a 3-dimensional body may be factored as

$$g = R(t_1 Y) R(t_2 X) R(t_3 Y), \tag{1}$$

where $R(tX)$ (resp. $R(tY)$) is a rotation in angle t around X-axis (resp. Y-axis). The parameters t_1, t_2, t_3 are called the Euler's angles.

We could allow decompositions for g with more factors:

$$g = R(t_1 Y) R(t_2 X) \dots R(t_{n-1} Y) R(t_n X). \tag{2}$$

Clearly we will get infinitely many such decompositions for a fixed element g in the group $SO(3)$ of rotations. Thus it is natural to pose the question of finding a decomposition (2) that minimizes the total angle of rotation $|t_1| + |t_2| + \ldots + |t_n|$. It can happen that decompositions with more factors have a smaller total angle of rotation than the Euler's decomposition (1).

It turns out that this problem is not well-posed: for some g an optimal decomposition (2) does not exist. Instead, the infimum of the total angle of rotation is attained as a limit on a sequence of decompositions (2) with $n \to \infty$.

We can overcome this difficulty by noting that

$$R(aX + bY) = \lim_{n\to\infty} \left(R\left(\frac{a}{n} X\right) R\left(\frac{b}{n} Y\right)\right)^n,$$

hence it is natural to extend the set of controls from $\{\pm X, \pm Y\}$ to $\mathcal{C} = \left\{ aX + bY \;\middle|\; |a| + |b| = 1 \right\}$. Implementing a rotation $R(aX + bY)$ corresponds to carrying out rotations around axes X and Y simultaneously with the ratio $a : b$ of angular velocities.

Once we extend the control set, our optimization problem becomes well-posed and for every $g \in SO(3)$ there is an optimal decomposition with a finite number of factors.

We study this problem in a more general setting, where we allow an arbitrary angle $0 < \alpha \leq \frac{\pi}{2}$ between the axes X and Y. We also introduce a more general cost function to be minimized, where a rotation in angle t around Y-axis has the same cost as a rotation around X-axis in angle κt, $0 \leq \kappa \leq 1$.

We solve the optimization problem in this greater generality and determine possible patterns for the optimal decompositions. Each of these patterns has (at most) 3 independent time parameters, and it is fairly easy to find numerically the decompositions of a given element $g \in SO(3)$ according to each pattern. This produces a finite number of decompositions and we can immediately see which one of them is optimal.

It happens that our optimization problem has a bifurcation at $\kappa = \cos\alpha$. For the cases $\kappa > \cos\alpha$ and $\kappa < \cos\alpha$ we get different lists of optimal patterns. There are also special cases when $\kappa = 0$ or $\cos\alpha = 0$.

Let us present the list of optimal patterns in case when the axes X and Y are perpendicular to each other and $\kappa = 1$. Since in this case the problem is symmetric with respect to the dihedral group of order 8, generated by transformations $(X, Y) \mapsto (Y, X)$, $(X, Y) \mapsto (-X, Y)$, $(X, Y) \mapsto (X, -Y)$, the list of patterns will also be symmetric with respect to this group. We denote this group of 8 symmetries by $(X, Y) \mapsto \{\pm X, \pm Y\}$ and use it to present the list of patterns in a more compact form.

Theorem 1.1. *Let the angle between the axes X and Y be $\alpha = \frac{\pi}{2}$ and let $\kappa = 1$. For an element $g \in SO(3)$ there is an optimal decomposition with $t_1, t_2, t_3 \geq 0$ of one of the following types:*

$$
\begin{aligned}
&R(t_1X)R(t_2Y)R(-t_3X), && \text{with } \; t_1, t_3 \leq t_2 \leq \pi,\\
&R(t_1X)R(t_2Y)R(-t_2X)R(-t_3Y), && \text{with } \; t_1, t_3 \leq t_2 \leq \pi,\\
&R(t_1X)R\left(t_2\left(X+Y\right)/2\right)R(t_3X), && \text{with } \; t_1, t_3 \leq \pi,\; t_2 \leq \sqrt{2}\pi,\\
&R(t_1X)R\left(t_2\left(X+Y\right)/2\right)R(t_3Y), && \text{with } \; t_1, t_3 \leq \pi,\; t_2 \leq \sqrt{2}\pi,
\end{aligned}
$$

and symmetric to these under the group of transformations $(X, Y) \mapsto \{\pm X, \pm Y\}$.

Example 1.2. Suppose we would like to decompose a rotation $R(tZ)$ as a product of rotations around X- and Y-axes, where $\{X, Y, Z\}$ is the standard orthogonal basis of $\mathbb{R}^3$. The pattern for the optimal decompositions will depend on the value of t. If $0 \leq t \leq \frac{\pi}{2}$ then the following decomposition realizes the minimum of the total rotation angle:

$$R(tZ) = R(-t_1X)R(-t_2Y)R(t_2X)R(t_1Y),$$

where $$t_1 = \arccos\left(\frac{1}{\cos\left(\frac{t}{2}\right) + \sin\left(\frac{t}{2}\right)}\right),\; t_2 = \arccos\left(\cos\left(\frac{t}{2}\right) - \sin\left(\frac{t}{2}\right)\right).$$

For $\frac{\pi}{2} \leq t \leq \pi$ the Euler's decomposition (1) becomes optimal:

$$R(tZ) = R(-\frac{\pi}{2}Y)R(tX)R(\frac{\pi}{2}Y).$$

When $t = \frac{\pi}{2}$ both patterns are optimal. For $-\pi < t < 0$ the optimal decompositions may be obtained by switching X with Y in the above expressions.

In case when the axes X and Y are perpendicular to each other and $\kappa = 0$ (meaning that rotations around Y-axis have zero cost), the optimal decompositions are precisely those described by Euler (1).

In 2009 NASA launched a space telescope Kepler with a mission of finding planets outside the Solar system. This spacecraft was placed in an orbit around the Sun. To take images of stars, the telescope needs to be pointed in the target direction, with its solar panels facing the Sun. The attitude control of Kepler is done with reaction wheels, which are heavy disks mounted on electric motors. Once the reaction wheel is turned, the spacecraft will turn around the same axis in the opposite direction due to the angular momentum conservation law.

If we have three reaction wheels with linearly independent axes, by rotating them simultaneously with appropriate relative angular velocities, we can implement a continuous rotation of the spacecraft around an arbitrary axis. For redundancy, Kepler was equipped with four reaction wheels with their axes in a tetrahedral configuration, so that any three of them could provide an efficient attitude control. However by May 2013, two of the four reaction wheels failed, leaving Kepler with just two available axes of rotation [5]. The results of our paper provide optimal methods for attitude control with two rotation axes, like in situation with the Kepler space telescope.

This paper builds on our previous work [1], where we studied a similar problem for $SU(2)$, also with two available controls, but with a restriction that only a positive time evolution is allowed. That paper was motivated by the applications to quantum control in a 1-qubit system.

In the present paper we use the geometric control theory [3], which is an adaptation of the Pontryagin's Maximum Principle to the setting of Lie groups. The Maximum Principle provides only necessary conditions for optimality, which need not be sufficient. In Section 3 we identify decompositions that satisfy the necessary conditions of the Pontryagin's Maximum Principle. Then we go into a more detailed analysis in Section 4 by showing that decompositions with a large number of factors are not optimal, even when they satisfy the conditions of the Maximum Principle. Our main results are stated in Theorems 2.1 – 2.4 at the end of the next Section.

Acknowledgements. I thank Cornelius Dennehy, Ken Lebsock, Eric Stoneking and Alex Teutsch for the stimulating discussions. Support from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

2. Attitude control problem

For a unit vector $X \in \mathbb{R}^3$ denote by $R(tX)$ an operator of rotation of $\mathbb{R}^3$ in angle t around X, with the plane perpendicular to X turning counterclockwise when viewed from the endpoint of X. As a 3×3 matrix, $R(tX)$ is given by the formula:

$$R(tX) = \cos(t) I + \sin(t) \mathrm{ad} X + (1 - \cos(t)) X X^T,$$

where for $X = (a, b, c)^T$ with $a^2 + b^2 + c^2 = 1$, $\mathrm{ad}X$ is the adjoint matrix of X with respect to the cross product, so that $(\mathrm{ad}X)Y = X \times Y$:

$$\mathrm{ad}X = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \quad \text{and} \quad X X^T = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix}. \tag{3}$$

For $X \in \mathbb{R}^3$ with $|X| \neq 1$ we set $R(tX) = R(t'X')$, where $t' = t\,|X|$ and $X' = X/|X|$. The set of all rotations of $\mathbb{R}^3$ forms the group $SO(3)$.

For a fixed X the set $\{R(tX) | t \in \mathbb{R}\}$ is a 1-parametric subgroup in $SO(3)$.

It is well-known that for any two non-proportional unit vectors X, Y, the corresponding 1-parametric subgroups together generate the whole group $SO(3)$. This means that every element $g \in SO(3)$ may be decomposed into a product

$$g = R(t_1 C_1) R(t_2 C_2) \ldots R(t_n C_n) \tag{4}$$

with $C_j \in \{X, Y\}$. Decomposition (4) is of course not unique. It is then natural to consider the optimization problem of finding the infimum of $|t_1| + \ldots + |t_n|$ over all decompositions (4) with fixed $g \in SO(3)$. More generally, we may assign cost to each generator X, Y and minimize the total cost in (4).

Introduction of the cost parameters may be warranted in case when the body that we control has unequal momenta of inertia with respect to the axes X and Y, thus making it easier to rotate it around one of the axes.

Without loss of generality, we assume that $\mathrm{Cost}(X) \geq \mathrm{Cost}(Y)$ and renormalize the cost function by fixing $\mathrm{Cost}(X) = 1$, $\mathrm{Cost}(Y) = \kappa$ with $0 \leq \kappa \leq 1$.

For the rest of the paper we fix two non-proportional vectors $X, Y \in \mathbb{R}^3$ with $|X| = |Y| = 1$. An important parameter is the angle α between these vectors. Without loss of generality we assume $0 < \alpha \leq \frac{\pi}{2}$, otherwise we can replace Y with $-Y$. Throughout the paper we will the use parameter $c = \cos(\alpha)$, $0 \leq c < 1$. Let Z be a vector perpendicular to X and Y, $Z = X \times Y$, $|Z| = \sin(\alpha)$.

It could happen that the infimum of cost is not attained on any particular decomposition (4), but rather as a limit on a sequence of such decompositions with $n \to \infty$. It turns out that we can overcome this difficulty by enlarging the set of generators to be

$$\mathcal{C} = \left\{ aX + bY \,\middle|\, |a| + |b| = 1 \right\}.$$

Note that rotations corresponding to elements of $\mathcal{C}$ can be realized as limits of products of rotations with axes $\{X, Y\}$:

$$R(t(aX + bY)) = \lim_{n \to \infty} \left(R\left(\frac{ta}{n} X\right) R\left(\frac{tb}{n} Y\right) \right)^n. \tag{5}$$

From the point of view of the attitude control, this corresponds to turning on controls X and Y simultaneously with intensities a and b respectively.

We extend the definition of the cost function in such a way that the cost of both sides of (5) is the same:

$$\mathrm{Cost}(aX + bY) = |a| \,\mathrm{Cost}(X) + |b| \,\mathrm{Cost}(Y). \tag{6}$$

Our goal is to solve the following

Problem 1. For a given $g \in SO(3)$ find a decomposition $g = R(t_1 C_1) R(t_2 C_2) \ldots R(t_n C_n)$ with $C_1, \ldots, C_n \in \mathcal{C}$, $t_1, \ldots, t_n \geq 0$, realizing the the infimum of

$$t_1 \mathrm{Cost}(C_1) + \ldots + t_n \mathrm{Cost}(C_n).$$

It was shown in [1], Theorem 1.4, that the infimum cost in this problem problem is the same as for its more restricted version where the set of controls is taken to be $\{\pm X, \pm Y\}$ instead of $\mathcal{C}$.

In fact, we shall see that we would not need the whole set $\mathcal{C}$, but require in addition to controls $\{\pm X, \pm Y\}$ only the elements $\{\pm W_+, \pm W_-\}$, where W_+ (resp. W_-) is a linear combination of X and Y, which is orthogonal to $\kappa X + Y$ (resp. $\kappa X - Y$).

Since the cost of $R((\lambda t)C)$ and $R(t(\lambda C))$ is the same, we can rescale the generators without changing the cost of decompositions. We can thus drop the requirement $|a|+|b|=1$ for the generators $C=aX+bY$.

We fix $W_+=(1+\kappa c)X-(\kappa+c)Y$ and $W_-=(1-\kappa c)X+(\kappa-c)Y$. Taking into account that $(X,X)=(Y,Y)=1$ and $(X,Y)=c$, it is easy to check that $(W_+,\kappa X+Y)=0$ and $(W_-,\kappa X-Y)=0$.

Now we can state the main results of the paper. It turns out that the problem we consider has a bifurcation at $\kappa=c$, and we need to consider the cases $0\le c<\kappa\le 1$ and $0<\kappa\le c<1$ separately. There will be also a special case when $\kappa=0$.

We will give the solution of the above optimal control problem by specifying the patterns of optimal decomposition (4).

We begin with some elementary observations. Obviously we may restrict all angles of rotation to be less or equal to π.

If $g=R(t_1C_1)R(t_2C_2)\dots R(t_nC_n)$ is an optimal decomposition then a decomposition

$$R(t'_kC_k)R(t_{k+1}C_{k+1})\dots R(t_{m-1}C_{m-1})R(t'_mC_m) \tag{7}$$

with $1\le k\le m\le n$, $0\le t'_k\le t_k$, $0\le t'_m\le t_m$, is also optimal. We call (7) a *subword* of $R(t_1C_1)R(t_2C_2)\dots R(t_nC_n)$. We shall present the optimal decompositions as subwords of certain patterns.

Since the number of patterns can be fairly large, we shall use various symmetries in order to group several patterns together. For example, if we have an optimal decomposition

$$g=R(t_1C_1)R(t_2C_2)\dots R(t_nC_n)$$

with $C_j\in\mathcal{C}$, then

$$R(-t_1C_1)R(-t_2C_2)\dots R(-t_nC_n)$$

is also an optimal decomposition (for a different element of $SO(3)$). This follows from the fact that multiplication of controls by -1 is an automorphism of our problem. We denote this symmetry transformation on the set of patterns by $(X,Y)\mapsto(-X,-Y)$.

Whereas the set of optimal patterns is always invariant with respect to the symmetry $(X,Y)\mapsto(-X,-Y)$, other types of symmetries that we shall consider are not universal and are present only for some patterns. If we make the following schematic representation of the controls, all symmetries that we consider will be elements of the dihedral group of symmetries of a square:

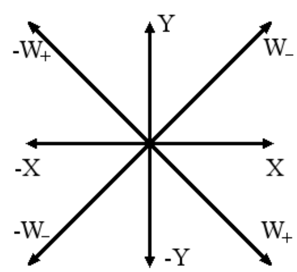


Figure 1

Consider a transformation $X\mapsto X$, $Y\mapsto -Y$, $W_+\mapsto W_-$, $W_-\mapsto W_+$. Together with the symmetry $(X,Y)\mapsto(-X,-Y)$ this generates a set of 4 transformations. We denote this set of symmetries by $(X,Y)\mapsto(\pm X,\pm Y)$. We assume that all symmetries we consider are compatible with multiplication by -1, even though they are not linear in general.

We also consider a transformation $X\mapsto Y$, $Y\mapsto X$, $W_+\mapsto -W_+$, $W_-\mapsto W_-$. Together with $(X,Y)\mapsto(-X,-Y)$, this generates a set of 4 transformations, which we denote by $(X,Y)\mapsto\{-X,-Y\}$.

Finally, if we consider all of the above transformations together, we generate a full set of 8 symmetries of the square in Fig.1, which we denote by $(X,Y)\mapsto\{\pm X,\pm Y\}$.

Theorem 2.1. *Let $c=0$, $0<\kappa\leq 1$. For an element $g\in SO(3)$ the infimum of the optimization Problem 1 is attained on a subword of one of the following patterns:*

(I) $R(t_X X)R(t_Y Y)R(-t_X X)R(-t_Y Y)$ where $\tan(t_X/2)=\kappa\tan(t_Y/2)$, $0<t_X,t_Y\leq\pi$, and symmetric to it under $(X,Y)\mapsto\{\pm X,\pm Y\}$.

(II) $R(\pi X)R(tW_+)R(\pi X)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{\pm X,\pm Y\}$.

(III) $R(\pi X)R(tW_+)R(-\pi Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{\pm X,\pm Y\}$.

When we apply symmetry transformations, e.g. $X\mapsto Y$, $Y\mapsto X$, we change the parameters t_X, t_Y accordingly, but the relation $\tan(t_X/2)=\kappa\tan(t_Y/2)$ in (I) is preserved. Under this symmetry transformation, the pattern (I) takes form $R(t_Y Y)R(t_X X)R(-t_Y Y)R(-t_X X)$ with $\tan(t_X/2)=\kappa\tan(t_Y/2)$.

Set

$$\widehat{t}_X=\arccos\left(\frac{c-\kappa}{c+\kappa}\right),\quad \widehat{t}_Y=\arccos\left(-\frac{1-\kappa c}{1+\kappa c}\right),\quad 0\leq\widehat{t}_X,\widehat{t}_Y\leq\pi. \tag{8}$$

Theorem 2.2. *Let $0<c<\kappa\leq 1$. For an element $g\in SO(3)$ the infimum of the optimization Problem 1 is attained on a subword of either pattern (I) or one of the following:*

(IV) $R(\widehat{t}_Y Y)R(\widehat{t}_X X)R(tW_+)R(\widehat{t}_X X)R(\widehat{t}_Y Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{-X,-Y\}$.

(V) $R(\widehat{t}_X Y)R(\widehat{t}_X X)R(tW_+)R(-\widehat{t}_Y Y)R(-\widehat{t}_X X)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{-X,-Y\}$.

(VI) $R(\pi Y)R(tW_-)R(\pi Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{-X,-Y\}$.

(VII) $R(\pi X)R(tW_-)R(\pi Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto\{-X,-Y\}$.

Theorem 2.3. *Let $0<\kappa\leq c<1$. For an element $g\in SO(3)$ the infimum of the optimization Problem 1 is attained on a subword of either patterns (I), (IV), (V) given above, or the following pattern*

(VIII) $R(\pi Y)R(tX)R(\pi Y)$, with $0\leq t\leq 2\widehat{t}_X$, and symmetric to it under $(X,Y)\mapsto(-X,-Y)$.

Theorem 2.4. *Let $\kappa=0$, $c\geq 0$. For an element $g\in SO(3)$ the infimum of the optimization Problem 1 is attained on a subword of one of the following two patterns*

(IX) $R(\pi Y)R(tW_+)R(\pi Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto(\pm X,\pm Y)$.

(X) $R(\pi Y)R(tW_+)R(-\pi Y)$, with $t\geq 0$, and symmetric to it under $(X,Y)\mapsto(\pm X,\pm Y)$.

Remark 2.5. When $\kappa=c>0$ we have W_- proportional to X, and the lists of patterns in Theorems 2.2 and 2.3 become equivalent.

3. Geometric optimization theory

In this section we will review the geometric optimization theory following [3], and apply it to our optimization problem.

The Lie algebra $so(3)$ of the Lie group $SO(3)$ is the tangent space to $SO(3)$ at identity and consists of skew-symmetric 3×3 matrices. The Lie bracket of two matrices in $so(3)$ is $[A, B] = AB - BA$. We may identify the space $so(3)$ with $\mathbb{R}^3$ via the map (3) $X \mapsto \mathrm{ad}(X)$. The corresponding Lie bracket of two vectors in $\mathbb{R}^3$ is the cross product.

Fix $g \in SO(3)$. A curve leading to g is an absolutely continuous function $U : [0, t_0] \to SO(3)$ such that $U(0) = I$ and $U(t_0) = g$. An absolutely continuous function has a measurable derivative $u : [0, t_0] \to so(3)$ such that $U'(t) = U(t)u(t)$ for almost all t. The derivative u is Lebesgue integrable [7].

Let us formulate a differential version of our optimization problem.

Problem 2. For an element $g \in SO(3)$ find the infimum of $\int_0^{t_0} \mathrm{Cost}(U^{-1}(t)U'(t))dt$ over all absolutely continuous curves $U : [0, t_0] \to SO(3)$ leading to g, satisfying $U^{-1}(t)U'(t) \in \mathcal{C} \subset \mathbb{R}^3 = so(3)$ for almost all t.

The parameter t_0 in Problem 2 is not fixed and when taking the infimum we consider the curves with all $t_0 \geq 0$.

It is clear that the restriction to the case of piecewise constant controls $u(t) = U^{-1}(t)U'(t)$ gives precisely Problem 1. On the other hand we shall see that the solutions of Problem 2 indeed have piecewise constant controls, which implies equivalence of Problems 1 and 2.

Proposition 3.1. *For any $g \in SO(3)$ there exists an absolutely continuous optimal solution U for Problem 2.*

Proof. The proof of this Proposition is based on the observation that the cost assigned to a curve $U : [0, t_0] \to SO(3)$ is independent of the choice of its parametrization. To prove this, we first note the cost function (6) satisfies $\mathrm{Cost}(\lambda u) = \lambda \mathrm{Cost}(u)$ for $\lambda \geq 0$. Consider an absolutely continuous increasing surjective reparametrization $f : [0, \tau_0] \to [0, t_0]$ and the corresponding reparametrized curve $\overline{U}(\tau) = U(f(\tau))$. Then $\overline{U}$ and U have the same cost:

$$\begin{aligned}
&\int_0^{\tau_0} \mathrm{Cost}\left(\overline{U}(\tau)^{-1}\frac{d}{d\tau}\overline{U}(\tau)\right) d\tau \\
&= \int_0^{\tau_0} \mathrm{Cost}\left(U(t)^{-1}\frac{d}{dt}U(t)\big|_{t=f(\tau)} f'(\tau)\right) d\tau \\
&= \int_0^{\tau_0} \mathrm{Cost}\left(U(t)^{-1}\frac{d}{dt}U(t)\big|_{t=f(\tau)}\right) f'(\tau) d\tau \\
&= \int_0^{t_0} \mathrm{Cost}\left(U(t)^{-1}U'(t)\right) dt.
\end{aligned}$$

This computation shows that rescaling of the set of controls $\mathcal{C}$ does not change the cost of a curve U leading to g with $U^{-1}(t)U'(t) \in \mathcal{C}$.

Let us modify Problem 2 by replacing the set $\mathcal{C}$ with its convex hull

$$\overline{\mathcal{C}} = \left\{ aX + bY \;\middle|\; |a| + |b| \leq 1 \right\}.$$

Once the control set is convex, we can apply Theorem 4.10 from [7] to obtain the existence of an absolutely continuous optimal solution $U : [0, t_0] \to SO(3)$ for the modified problem.

To go back to the setting of Problem 2, we note that every absolutely continuous curve admits a parametrization by the arc length, i.e., the natural parametrization (see for example Section 5.3 in [6]). Then it is easy to see that the curve U may also be reparametrized with $U^{-1}(t)U'(t) \in \mathcal{C}$. Since reparametrization does not change the cost, we see that an optimal solution of the modified problem with the set of controls $\overline{\mathcal{C}}$ yields an optimal solution for Problem 2. $\square$

Remark 3.2. Our optimization problem induces a left-invariant metric on $SO(3)$. It is possible to see that this metric does not correspond to any Riemannian structure on this Lie group.

The Hamiltonian function for Problem 2 is

$$\mathcal{H}(p,u) = p_0 \mathrm{Cost}(u) + (p,u), \quad u \in \mathcal{C}, p \in \mathbb{R}^3,$$

which involves a parameter $p_0 \leq 0$ (see Section 11.2.2 in [3] for details).

For each $p \in \mathbb{R}^3$ we define the maximal Hamiltonian

$$\mathcal{M}(p) = \max_{u \in \mathcal{C}} \mathcal{H}(p,u).$$

Theorem 3.3. *(Pontryagin's Maximum Principle, [3]) Let U be an optimal curve leading to $g \in SO(3)$ for Problem 2. Then there exists an absolutely continuous function $p : [0,t_0] \to \mathbb{R}^3 = so(3)$ and a constant $p_0 \leq 0$ such that for almost all $t \in [0,t_0]$ the following equations hold:*

$$\text{(i)} \qquad \mathcal{H}(p(t),u(t)) = \mathcal{M}(p(t)) = 0$$

and

$$\text{(ii)} \qquad \frac{dp}{dt} = p(t) \times u(t).$$

If $p_0 = 0$ then $p(t)$ is non-zero for almost all $t \in [0,t_0]$.

Lemma 3.4. *The quantity $|p(t)|$ is conserved.*

Proof.

$$\frac{d}{dt}|p(t)|^2 = 2\left(\frac{dp}{dt}, p(t)\right) = 2\left(p(t) \times u(t), p(t)\right) = 0.$$

$\square$

Note that for our problem the parameter p_0 can not be zero, otherwise condition (i) implies that $\max_{u \in \mathcal{C}}(p(t),u) = 0$, hence $p(t)$ is proportional to Z for almost all t, and so is $\frac{dp}{dt}$. However (ii) implies that $\left(\frac{dp}{dt}, p(t)\right) = 0$ and thus $\frac{dp}{dt} = 0$ and $p(t)$ is a constant multiple of Z. Inspecting (ii) again, we conclude that $p(t)$ must be zero for almost all t, which contradicts the last claim of the theorem.

In case when the parameter p_0 is non-zero, it can be rescaled to any negative value. A convenient choice for us is $p_0 = -\sin^2(\alpha)$.

Consider a second basis $\{S,Q\}$ of the XY-plane, where

$$S = Y \times Z = X - cY \quad \text{and} \quad Q = -X \times Z = Y - cX.$$

Then $(S,X) = (Q,Y) = (S,S) = (Q,Q) = \sin^2(\alpha)$, $(S,Y) = (Q,X) = 0$. In this basis $W_+ = (1+\kappa c)X - (\kappa + c)Y = S - \kappa Q$ and $W_- = (1-\kappa c)X + (\kappa - c)Y = S + \kappa Q$.

According to Theorem 3.3, the value of $p(t)$ determines the value of $u(t)$ via (i), while by (ii) the value of $u(t)$ determines the evolution of $p(t)$. Let us analyze (i) to see which values of $p(t)$ are admissible, and what are the corresponding controls u.

Let us write $u = aX + bY$ and $p = sS + qQ + zZ$. Then

$$\mathcal{H}(p, u) = \sin^2(\alpha)\,(-|a| - \kappa\,|b| + sa + qb)\,.$$

Since the set $\mathcal{C}$ is closed under symmetry $a \mapsto -a$, $b \mapsto -b$, we see that the maximum in $u \in \mathcal{C}$ of $H(p, u)$ is attained when a has the same sign as s and b has the same sign as q. Hence

$$\mathcal{M}(p)/\sin^2(\alpha) = \max_{|a|+|b|=1} (|s| - 1)|a| + (|q| - \kappa)|b| = \max\{|s| - 1, |q| - \kappa\}\,.$$

By property (i) of the Theorem, $\mathcal{M}(p(t)) = 0$, thus the admissible values of $p(t)$ satisfy either $|s| = 1$, $|q| \leq \kappa$ or $|q| = \kappa$, $|s| \leq 1$. We summarize this in the following Lemma, which describes controls in the resulting regions:

Lemma 3.5. *(a) Let $\kappa > 0$.*
(i) If $s = 1$, $-\kappa < q < \kappa$ then $a = 1$, $b = 0$, the control is $u = X$;
(ii) If $s = -1$, $-\kappa < q < \kappa$ then $a = -1$, $b = 0$, the control is $u = -X$;
(iii) If $q = \kappa$, $-1 < s < 1$ then $a = 0$, $b = 1$, the control is $u = Y$;
(iv) If $q = -\kappa$, $-1 < s < 1$ then $a = 0$, $b = -1$, the control is $u = -Y$.

(b) If $\kappa = 0$ then $q = 0$, $|s| \leq 1$. When $q = 0$ and $-1 < s < 1$ we could have either control $u = Y$ or $u = -Y$.

At the points where two regions meet, the whole segment joining the corresponding two controls is allowed. For example, when $s = 1$ and $q = \kappa > 0$ we could have any control $u = aX + bY$ with $a, b \geq 0$, $a + b = 1$. We will call such values of p *critical.*

If the curve $p(t)$ reaches a critical point, one of three things could happen: the curve $p(t)$ could cross the boundary of a region, in which case the control will switch; the curve $p(t)$ could return to the same region where it came from without a switch of control; or the curve $p(t)$ may stay inside the critical boundary for some positive time. Let us describe evolution of $p(t)$ inside the critical boundary.

Lemma 3.6. *(a) Suppose $p(t) = S - \kappa Q + z(t)Z$ for $t \in [t_1, t_2]$. Then $u(t) = W_+$ and $z(t) = 0$ for $t \in [t_1, t_2]$.*
(b) Suppose $p(t) = S + \kappa Q + z(t)Z$ for $t \in [t_1, t_2]$ and $\kappa > 0$. Then $\kappa \geq c$, $u(t) = W_-$ and $z(t) = 0$ for $t \in [t_1, t_2]$.

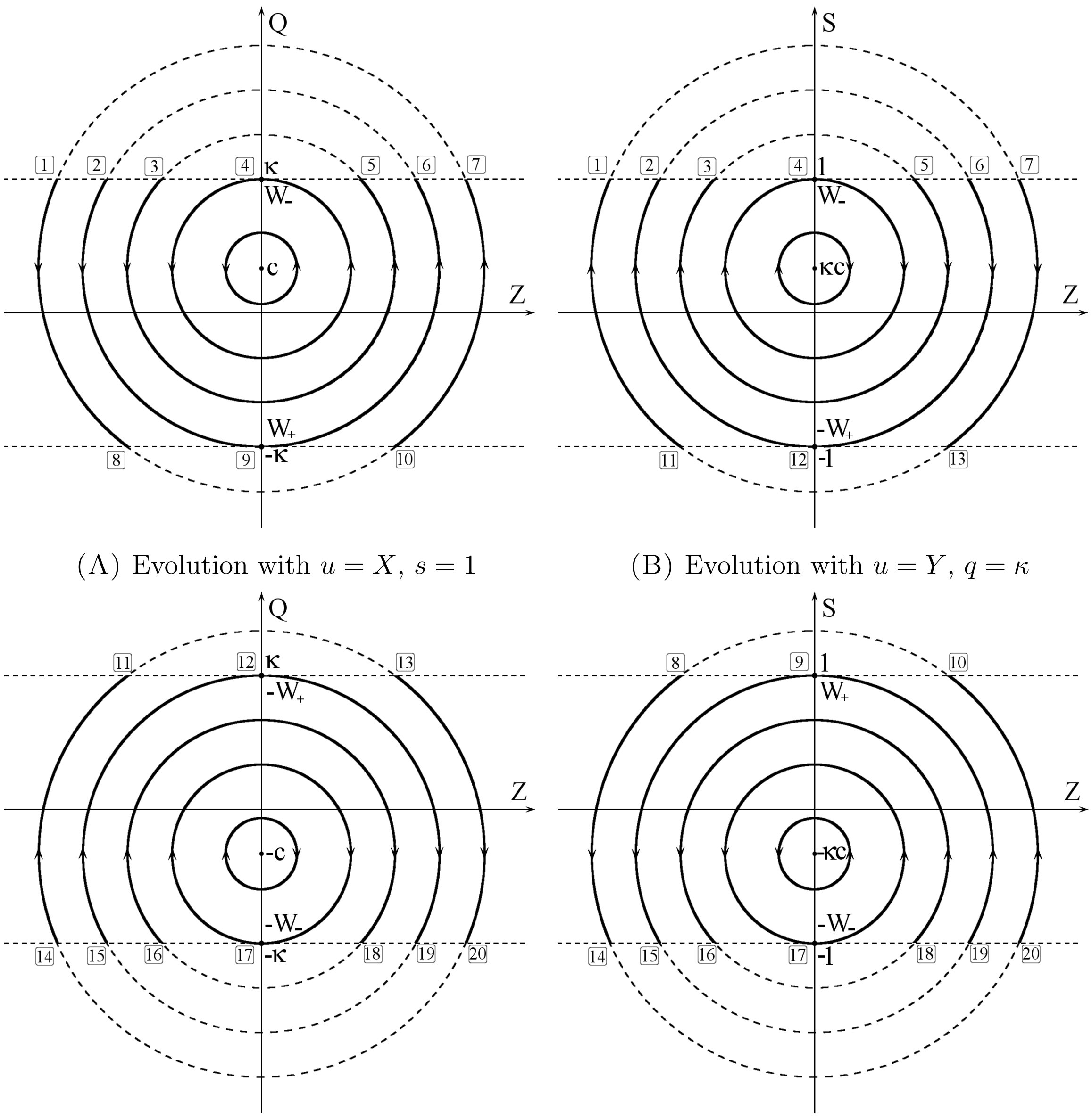
Cases $s(t) = -1$, $q(t) = \kappa$ and $s(t) = -1$, $q(t) = -\kappa$ are analogous, the controls are $u(t) = -W_+$ and $u(t) = -W_-$ respectively and $z(t) = 0$.

Proof. To prove (a) consider equation (ii) in Theorem 3.3. We get

$$\frac{d}{dt}\,(S - \kappa Q + z(t)Z) = (S - \kappa Q + z(t)Z) \times (aX + bY)\,.$$

Taking into account that

$$S \times X = cZ, \quad S \times Y = Z, \quad Q \times X = -Z, \quad Q \times Y = -cZ,$$



(A) Evolution with $u = X$, $s = 1$

(B) Evolution with $u = Y$, $q = \kappa$

(C) Evolution with $u = -X$, $s = -1$

(D) Evolution with $u = -Y$, $q = -\kappa$

FIGURE 2. Case $\kappa > c$.

we get that

$$\frac{dp}{dt} = az(t)Q - bz(t)S + (ca + \kappa a + b + c\kappa b)Z.$$

Since $q(t)$ and $s(t)$ are constant, this implies $z(t) = 0$ for $t \in [t_1, t_2]$. Then we get $(c+\kappa)a = -(1 + c\kappa)b$ and u is proportional to W_+.

Case (b) is analogous, except that for $\kappa < c$ the segment joining X and Y does not contain a vector proportional to W_-.

$\square$

Corollary 3.7. *An optimal solution of Problem 2 could only involve controls $\pm X$, $\pm Y$, $\pm W_+$ and $\pm W_-$. Moreover, controls $\pm W_-$ do not occur if $0 < \kappa < c$.*

Note that when $\kappa = c$ we get W_- proportional to X. When $\kappa = 0$ we get $W_+ = W_-$.

Next, let us study evolution of $p(t)$ under controls $\pm X$ and $\pm Y$.

As we have seen in Lemma 3.5, control X corresponds to the region $s = 1$, $-\kappa \le q \le \kappa$. Let $p(t) = S + q(t)Q + z(t)Z$. By part (ii) of Theorem 3.3, evolution of $p(t)$ is given by

$$\frac{dp}{dt} = (S + q(t)Q + z(t)Z) \times X = (c - q(t))Z + z(t)Q.$$

From this we get

$$q'(t) = z(t), \quad z'(t) = -(q(t) - c), \quad s'(t) = 0.$$

Setting $\tilde{q}(t) = q(t) - c$, we get the equations of the harmonic oscillator

$$\tilde{q}'(t) = z(t), \quad z'(t) = -\tilde{q}(t)$$

with solutions $q(t) = c + K\sin(t + \theta)$, $z(t) = K\cos(t + \theta)$. We plot the trajectories in QZ-plane in Fig.2A and 3A. Similarly, we plot the trajectories for the other regions described in Lemma 3.5.

This gives us the trajectories that satisfy the conditions of Theorem 3.3. For example, the path $\boxed{1} \mapsto \boxed{8} \mapsto \boxed{14} \mapsto \boxed{11} \mapsto \boxed{1}$ corresponds to the decomposition

$$R(t_1X)R(-t_2Y)R(-t_3X)R(t_4Y).$$

When a trajectory reaches a critical point, for example $\boxed{4}$, it could continue from $\boxed{4}$ either using evolution with controls X, Y or remain at this critical point for some positive time using control W_-.

The conservation law of Lemma 3.4 ensures that for the trajectory $\boxed{1} \mapsto \boxed{8} \mapsto \boxed{14}$ the points $\boxed{1}$ and $\boxed{14}$ have equal Z-coordinates. The same property holds in other similar cases, and in particular the trajectory that starts at a critical point $\boxed{9}$ and goes to $\boxed{15}$ will reach the critical point $\boxed{12}$.

It follows that for the trajectory $\boxed{1} \mapsto \boxed{8} \mapsto \boxed{14} \mapsto \boxed{11}$ evolution times for the parts $\boxed{1} \mapsto \boxed{8}$ and $\boxed{14} \mapsto \boxed{11}$ are the same, since the corresponding arcs are symmetric to each other.

Next we establish the relations between the time parameters in these trajectories (cf. Proposition 2.1 in [1]).

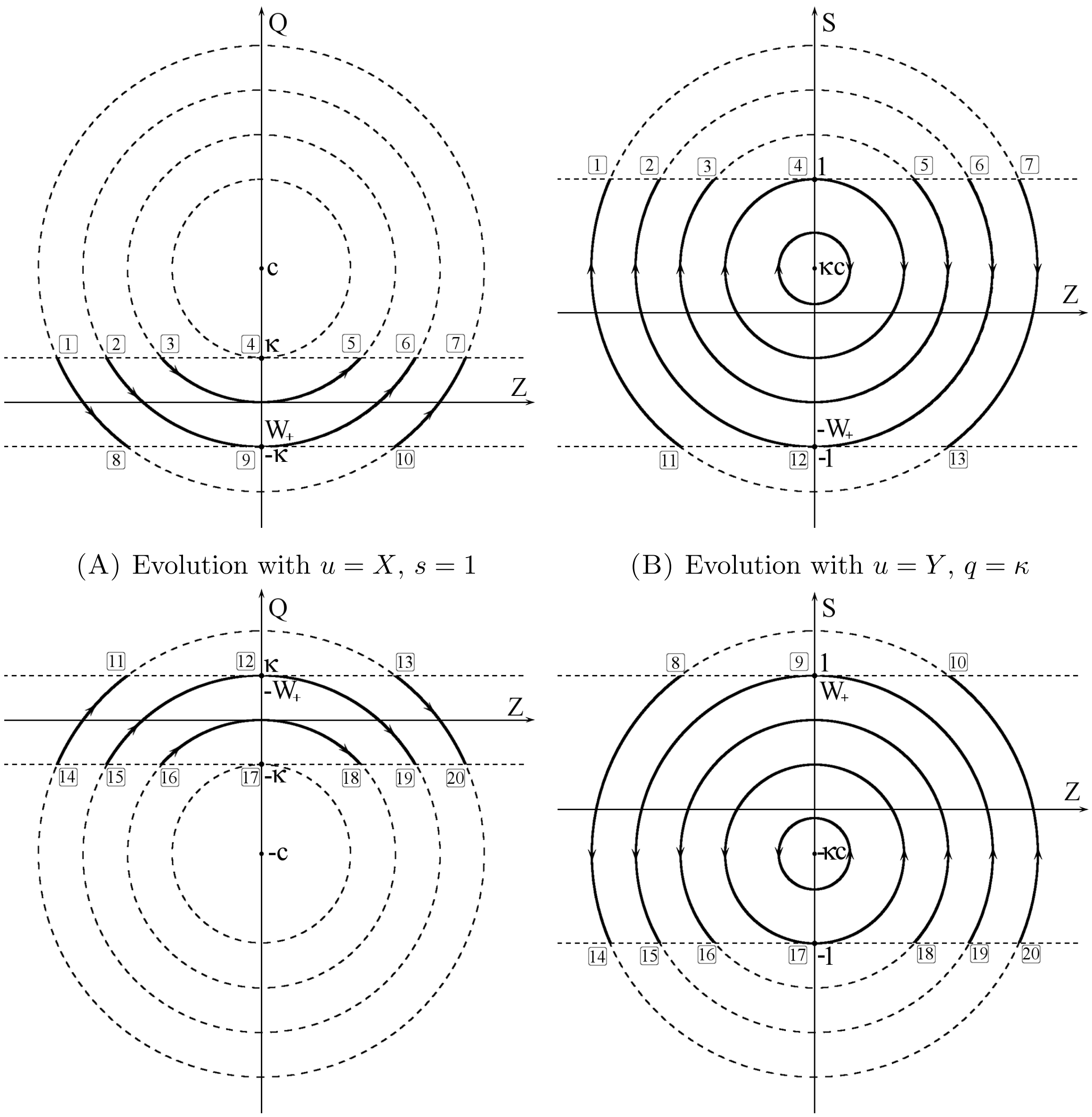
Proposition 3.8. *(a) Let t_X be the X-evolution time, and t_Y be $-Y$-evolution time for the trajectory $\boxed{1} \mapsto \boxed{8} \mapsto \boxed{14}$. Then*

$$\tan(t_X/2) = \kappa \tan(t_Y/2).$$

The same relation holds for the trajectories $\boxed{14} \mapsto \boxed{11} \mapsto \boxed{1}$, $\boxed{7} \mapsto \boxed{13} \mapsto \boxed{20}$, $\boxed{13} \mapsto \boxed{20} \mapsto \boxed{10}$, etc., with t_X being the time parameter for $\pm X$-evolution and t_Y for $\pm Y$-evolution.

(b) Let $\widehat{t}_X$ be the time of evolution for the trajectories involving critical points, $\boxed{9} \mapsto \boxed{6}$, $\boxed{2} \mapsto \boxed{9}$, $\boxed{15} \mapsto \boxed{12}$ or $\boxed{12} \mapsto \boxed{19}$.

Let $\widehat{t}_Y$ be the time of evolution for the trajectories $\boxed{12} \mapsto \boxed{2}$, $\boxed{6} \mapsto \boxed{12}$, $\boxed{19} \mapsto \boxed{9}$ or $\boxed{9} \mapsto \boxed{15}$.



(A) Evolution with $u = X$, $s = 1$

(B) Evolution with $u = Y$, $q = \kappa$

(C) Evolution with $u = -X$, $s = -1$

(D) Evolution with $u = -Y$, $q = -\kappa$

FIGURE 3. Case $0 < \kappa \leq c$.

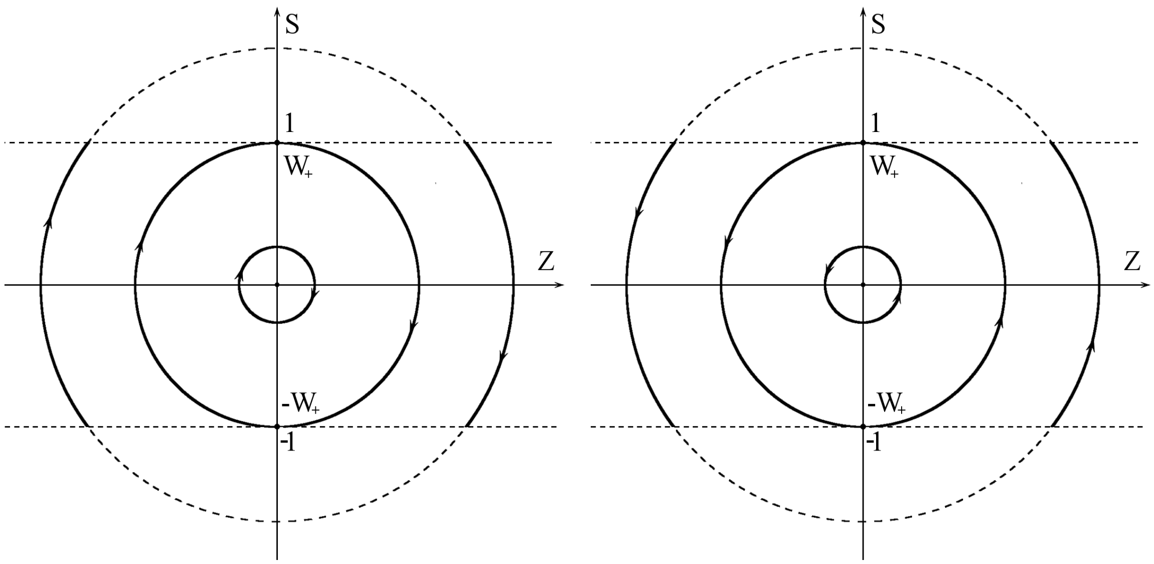
Then

$$\widehat{t}_X = \arccos\left(\frac{c-\kappa}{c+\kappa}\right), \quad \widehat{t}_Y = \arccos\left(-\frac{1-\kappa c}{1+\kappa c}\right). \tag{9}$$

Proof. Consider the trajectory $\boxed{1} \mapsto \boxed{8} \mapsto \boxed{14}$. Let z_1 and z_2 be Z-coordinates of the points $\boxed{1}$, and $\boxed{8}$ respectively. Then z_1 is also the Z-coordinate of the point $\boxed{14}$.

Since the points $\boxed{1}$ and $\boxed{8}$ lie on a circle with the center at $Z = 0$, $Q = c$, they satisfy the equation

$$z_1^2 + (\kappa - c)^2 = z_2^2 + (\kappa + c)^2. \tag{10}$$



(A) Evolution with $u = Y$, $q = 0$ (B) Evolution with $u = -Y$, $q = 0$

FIGURE 4. Case $\kappa = 0$.

Let b be the base of the isosceles triangle with vertices at $\boxed{1}$, $\boxed{8}$ and the center of the circle, and let h be the altitude in this triangle. Then

$$b^2 = (z_2 - z_1)^2 + (2\kappa)^2$$

and

$$h^2 = \left(\frac{z_1 + z_2}{2}\right)^2 + c^2.$$

Since t_X is the angle at the vertex of this triangle, we have

$$\tan^2\left(\frac{t_X}{2}\right) = \left(\frac{b}{2h}\right)^2 = \frac{(z_2 - z_1)^2 + 4\kappa^2}{(z_2 + z_1)^2 + 4c^2}.$$

Similarly,

$$\tan^2\left(\frac{t_Y}{2}\right) = \frac{(z_2 - z_1)^2 + 4}{(z_2 + z_1)^2 + 4\kappa^2 c^2}.$$

Since $\tan(t_X/2), \tan(t_Y/2) > 0$, in order to establish claim (a), we need to show that $\tan^2(t_X/2) = \kappa^2 \tan^2(t_Y/2)$. This equality however follows from (10):

$$\begin{aligned}\left((z_2 - z_1)^2 + 4\kappa^2\right)\left((z_2 + z_1)^2 + 4\kappa^2 c^2\right) - \kappa^2\left((z_2 - z_1)^2 + 4\right)\left((z_2 + z_1)^2 + 4c^2\right)& \\ = (\kappa^2 - 1)\left((z_2 - z_1)^2 (z_2 + z_1)^2 - 16\kappa^2 c^2\right) &= 0.\end{aligned}$$

The proof for the other cases in (a) is analogous.

Let us prove part (b). Consider the trajectory $\boxed{9} \mapsto \boxed{6}$. The time parameter $\widehat{t}_X$ is the angle corresponding to this arc of the circle with center at $Z = 0$, $Q = c$ and radius $\kappa + c$. Taking the projection to Q-axis we get

$$\cos(\widehat{t}_X) = \frac{c - \kappa}{c + \kappa}.$$

The derivation of the formula for $\widehat{t}_Y$ is analogous. $\square$

The patterns listed in Theorem 2.2 can be traced on the diagrams in Fig.2, while the patterns of Theorem 2.3 can be seen on Fig. 3. For example, the pattern given in part (I) of Theorem 2.2 corresponds to the trajectory $\boxed{10} \mapsto \boxed{7} \mapsto \boxed{13} \mapsto \boxed{20} \mapsto \boxed{10}$. The above Proposition describes the relations between the time parameters of the evolution.

To complete the proofs of Theorems 2.1 – 2.4 we need to show that decompositions with a large number of switches can not be optimal. We defer this to the next section.

4. Bounds on the number of control switches

In this section we are going to show that certain decompositions are not optimal, even though they satisfy the necessary conditions of the Pontryagin's Maximum Principle. This will give us constraints on the number of control switches in optimal decompositions.

Rather than doing computations in the group of rotations $SO(3)$, it is easier to carry them out in the unitary group $SU(2)$, which is a double cover of $SO(3)$:

$$\varphi : \; SU(2) \to SO(3). \tag{11}$$

Let us recall the construction of $SU(2)$ based on the quaternions. The algebra of quaternions $\mathbb{H}$ has a basis $\{1, i, j, k\}$ and relations $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. Similar to the complex numbers, we have the conjugation on $\mathbb{H}$, given by $\overline{1} = 1$, $\overline{i} = -i$, $\overline{j} = -j$, $\overline{k} = -k$, and the norm: $|ai + bj + ck + d| = \sqrt{a^2 + b^2 + c^2 + d^2}$. Every non-zero element of $\mathbb{H}$ has a multiplicative inverse given by $w^{-1} = \overline{w}/|w|^2$.

The unitary group $SU(2)$ may be realized as a unit sphere in the quaternion algebra $\mathbb{H}$:

$$SU(2) = \left\{ ai + bj + ck + d \mid a^2 + b^2 + c^2 + d^2 = 1 \right\}.$$

The Lie algebra $su(2)$ of the group $SU(2)$ is the tangent space at identity, it is a 3-dimensional subspace in $\mathbb{H}$ spanned by $\{i, j, k\}$. We are going to identify this Lie algebra with $\mathbb{R}^3$ via $i \mapsto e_1$, $j \mapsto e_2$, $k \mapsto e_3$, where $\{e_1, e_2, e_3\}$ is the standard basis of $\mathbb{R}^3$. Since $[i,j] = ij - ji = 2k$, $[j,k] = jk - kj = 2i$, $[k,i] = ki - ik = 2j$, we see that two Lie algebra structures on $\mathbb{R}^3$ coming from $so(3)$ and $su(2)$ differ by a factor of 2. For this reason there is a factor of 2 in the formula for the homomorphism φ:

$$\varphi\left(\exp(X)\right) = R(2X).$$

Here for a vector $X = (a, b, c)^T$ the exponential is computed in the algebra of quaternions $\exp(X) = \exp(ai + bj + ck) \in SU(2)$. Note that the rotation operator $R(X)$ is also an exponential: $R(X) = \exp(\mathrm{ad}(X))$.

The kernel of the homomorphism φ is $\{\pm 1\} \subset SU(2)$, so the map φ is 2 to 1.

The advantage of using $SU(2)$ instead of $SO(3)$ is that $SU(2)$ is embedded in a 4-dimensional vector space $\mathbb{H}$, while $SO(3)$ is embedded into the 9-dimensional space of 3×3 matrices.

In our computations we are going to use the Campbell-Hausdorff formula [4] (up to the second order terms):

$$\exp(\varepsilon A)\exp(\varepsilon B) = \exp(C), \quad \text{where} \;\; C = \varepsilon(A + B) + \frac{\varepsilon^2}{2}[A, B] + o(\varepsilon^2). \tag{12}$$

We will also need the conjugation formula: $C\exp(B)C^{-1} = \exp\left(CBC^{-1}\right)$.

Pontryagin's Maximum Principle that we use above is essentially a local first derivative test. In order to obtain stronger results, we need to either apply non-local transformations (those that do not come from a small variation of parameters) or use higher derivatives. In Proposition 4.4 we will be using the second derivative in order to show that certain decompositions are not optimal. An example of a non-local transformation is the identity $R(\pi X) = R(-\pi X)$ where $|X| = 1$. This trivial observation may be generalized in the following way. Suppose $R(t_1 X)R(t_2 Y)$ is a rotation in angle π. Then we get a relation $R(t_1 X)R(t_2 Y) = R(-t_2 Y)R(-t_1 X)$. Note that both sides of this equality have the same cost. This non-local relation and its consequences will be quite useful for our analysis.

Lemma 4.1. *Let $g = ai + bj + ck + d \in SU(2)$. The image of g in $SO(3)$ is a rotation in angle π if and only if $d = 0$.*

Proof. Clearly, $\varphi(g)$ is a rotation in angle π if and only if $\varphi(g)^2$ is the identity matrix, but $\varphi(g)$ is not identity. This is equivalent to $g^2 = \pm 1$, $g \neq \pm 1$ in $SU(2)$. It is easy to see that the only solutions to $g^2 = 1$ are $g = \pm 1$. Thus the preimages of rotations in angle π are precisely $g \in SU(2)$ with $g^2 = -1$, or equivalently, $g^{-1} = -g$. Since $|g| = 1$, this becomes $\overline{g} = -g$. For $g = ai + bj + ck + d$ this is equivalent to $d = 0$. □

Proposition 4.2. *Let $X, Y \in \mathbb{R}^3 = su(2)$. Suppose $|X| = |Y| = 1$ and let α be the angle between X and Y.*

(a) If $\tan(s_1)\tan(s_2) = \frac{1}{\cos(\alpha)}$ then the image of $\exp(s_1 X)\exp(s_2 Y)$ in $SO(3)$ is a rotation in angle π. In this case $\exp(s_1 X)\exp(s_2 Y) = -\exp(-s_2 Y)\exp(-s_1 X)$.

(b) Let $\tan\psi = \cos\alpha\tan(s_2)$, $-\frac{\pi}{2} < \psi \leq \frac{\pi}{2}$. Then

$$\exp(s_1 X)\exp(s_2 Y)\exp(s_3 X) = \exp(s_1' X)\exp(-s_2 Y)\exp(s_3' X),$$

where $s_1' = s_1 + \psi - \frac{\pi}{2}$, $s_3' = s_3 + \psi + \frac{\pi}{2}$.

Proof. The group $SU(2)$ acts on its Lie algebra $su(2)$ by conjugation, and its center $\{\pm 1\}$ acts trivially. This gives the action of $SO(3)$ on $su(2)$, which is the natural action of $SO(3)$ on $\mathbb{R}^3$. Since this action is transitive on pairs of unit vectors with a given angle between them, we may set without loss of generality $X = i$, $Y = i\cos\alpha + j\sin\alpha$. We complete this to a basis of $su(2)$ by setting $Z = \frac{1}{2}[X, Y] = k\sin\alpha$. We can easily verify that

$$\begin{aligned} XY &= -\cos\alpha + Z, & YX &= -\cos\alpha - Z, \\ XYX &= Y - 2\cos\alpha X, & YXY &= X - 2\cos\alpha Y, \\ [Z, X] &= 2Y - 2\cos\alpha X, & [Z, Y] &= -2X + 2\cos\alpha Y, \\ XZX &= Z, & YZY &= Z. \end{aligned} \tag{13}$$

We also note that $X^2 = -1$ and $\exp(sX) = \cos(s) + X\sin(s)$ and likewise for Y.

We have

$$\begin{aligned} \exp(s_1 X)\exp(s_2 Y) &= (\cos(s_1) + X\sin(s_1))(\cos(s_2) + Y\sin(s_2)) \\ &= (\cos(s_1)\cos(s_2) - \cos(\alpha)\sin(s_1)\sin(s_2)) \\ &\quad + X\sin(s_1)\cos(s_2) + Y\cos(s_1)\sin(s_2) + Z\sin(s_1)\sin(s_2). \end{aligned}$$

Applying Lemma 4.1 we establish the claim of part (a).

Using part (a), we get

$$\exp(s_2Y) = -\exp(-\tau X)\exp(-s_2Y)\exp(-\tau X),$$

where $\tan(\tau)\tan(s_2) = \frac{1}{\cos(\alpha)}$. Set $\psi = \frac{\pi}{2} - \tau$. Then $\tan\psi = \frac{1}{\tan\tau} = \cos\alpha\tan(s_2)$ and

$$\exp(s_2Y) = -\exp((\psi - \frac{\pi}{2})X)\exp(-s_2Y)\exp((\psi - \frac{\pi}{2})X)$$
$$= \exp((\psi - \frac{\pi}{2})X)\exp(-s_2Y)\exp((\psi + \frac{\pi}{2})X).$$

Multiplying both sides by $\exp(s_1X)$ on the left and $\exp(s_3X)$ on the right, we get the claim of part (b). $\square$

Proposition 4.3. *Let* $\tan\left|\frac{t_X}{2}\right| = \kappa\tan\left|\frac{t_Y}{2}\right|$. *Decompositions* $R(t_YY)R(t_XX)R(-t_YY)$ *with* $|t_Y| > \frac{\pi}{2}$ *and* $R(t_XX)R(t_YY)R(-t_XX)$ *with* $|t_X| > \frac{\pi}{2}$ *are not optimal.*

Proof. We may assume without loss of generality that $t_X, t_Y > 0$. Let us begin with the case of $R(t_YY)R(t_XX)R(-t_YY)$. We take its preimage under φ: $\exp(s_1Y)\exp(s_2X)\exp(-s_1Y) \in SU(2)$, where $s_1 = t_Y/2$, $s_2 = t_X/2$, $\tan(s_2) = \kappa\tan(s_1)$, $s_1 > \frac{\pi}{4}$. We claim that the decomposition $\exp(s_1'Y)\exp(-s_2X)\exp(s_3'Y)$ given by the previous proposition will have a lower cost. Since $|s_2| = |-s_2|$ we need to show that $|s_1'| + |s_3'| < 2|s_1|$, where

$$s_1' = s_1 + \psi - \frac{\pi}{2}, \quad s_3' = -s_1 + \psi + \frac{\pi}{2}. \tag{14}$$

We have $\psi > 0$ and $s_3' > 0$. If $s_1' < 0$ then

$$|s_1'| + |s_3'| = -(s_1 + \psi - \frac{\pi}{2}) + (-s_1 + \psi + \frac{\pi}{2}) = \pi - 2s_1 < \frac{\pi}{2} < 2s_1,$$

and we get that the new cost is lower. If $s_1' \geq 0$ then

$$|s_1'| + |s_3'| = (s_1 + \psi - \frac{\pi}{2}) + (-s_1 + \psi + \frac{\pi}{2}) = 2\psi. \tag{15}$$

Since $\tan\psi = \cos\alpha\tan s_2$ and $\tan s_2 = \kappa\tan s_1$, we get that $\psi < s_2 \leq s_1$, so the new cost is again lower.

We now apply the same approach to $R(t_XX)R(t_YY)R(-t_XX)$. We again take its preimage $\exp(s_1X)\exp(s_2Y)\exp(-s_1X)$ in $SU(2)$ and transform it into $\exp(s_1'X)\exp(-s_2Y)\exp(s_3'X)$ using Proposition 4.2. Here $\tan s_1 = \kappa\tan s_2$. The values of s_1', s_3' are still given by (14) with $\tan\psi = \cos\alpha\tan s_2$. We have $s_3' > 0$ and consider the sign of s_1'. The case $s_1' < 0$ is treated in the same way as before.

When $s_1' \geq 0$ we consider two subcases: $\kappa \leq \cos\alpha$ and $\kappa > \cos\alpha$. If $\kappa \leq \cos\alpha$, the claim of the Proposition follows from the observation that on the diagrams (A), (C) in Fig. 3 the arcs $\boxed{1} \mapsto \boxed{8}$, $\boxed{10} \mapsto \boxed{7}$, $\boxed{14} \mapsto \boxed{11}$ and $\boxed{13} \mapsto \boxed{20}$ correspond to an angle not exceeding $\frac{\pi}{2}$.

Let us assume $\kappa > \cos\alpha$. To show that the transformed expression has a lower cost, we need to prove that $\psi < s_1$. However $\tan\psi = \cos\alpha\tan s_2 = \frac{\cos\alpha}{\kappa}\tan s_1$. Since $\frac{\cos\alpha}{\kappa} < 1$, we get $\psi < s_1$, which completes the proof of the Proposition. $\square$

Proposition 4.4. *Let* $\delta > 0$ *be a small parameter and let* $\kappa \neq 0$. *Then the decompositions*

$$R(-\delta X)R(t_YY)R(t_XX)R(-t_YY)R(-\delta X) \tag{16}$$

with $t_X, t_Y > 0$, $\tan(t_X/2) = \kappa\tan(t_Y/2)$, *and those symmetric to it under* $(X, Y) \mapsto \{\pm X, \pm Y\}$, *are not optimal.*

Proof. Let us assume by contradiction that the given decomposition is optimal. As before, we take a preimage $\exp(-\varepsilon X)\exp(s_1 Y)\exp(s_2 X)\exp(-s_1 Y)\exp(-\varepsilon X)$, where $\varepsilon = \delta/2$, $s_1 = t_Y/2$, $s_2 = t_X/2$. We shall express the given decomposition in the following way:

$$\begin{aligned}\exp(-\varepsilon X)&\exp(s_1 Y)\exp(s_2 X)\exp(-s_1 Y)\exp(-\varepsilon X)\\ &= \exp((s_1+\varepsilon_1)Y)\exp((s_2+\varepsilon_2)X)\exp(-(s_1+\varepsilon_2')Y)\exp(-\varepsilon_3 X).\end{aligned} \tag{17}$$

We are going to solve for ε_1, ε_2, ε_2' and ε_3 in terms of ε, and show that the new decomposition has a lower cost. Note that the parameters ε_2 and ε_2' are bound by the relation $\tan(s_2+\varepsilon_2) = \kappa\tan(s_1+\varepsilon_2')$.

We will use the Campbell-Hausdorff formula (12) to rewrite both sides of (17) in the form

$$\exp(s_1 Y)\exp(s_2 X)\exp(L)\exp(-s_1 Y).$$

We shall calculate L up to the second order in ε. Applying (12) to the left hand side of (17), we get that

$$L = L_1 + L_2 + \frac{1}{2}[L_1, L_2] + o(\varepsilon^2), \tag{18}$$

where

$$L_1 = \exp(-s_2 X)\exp(-s_1 Y)(-\varepsilon X)\exp(s_1 Y)\exp(s_2 X),$$

and

$$L_2 = \exp(-s_1 Y)(-\varepsilon X)\exp(s_1 Y).$$

Let us carry out the detailed calculations. We shall use the basis $\{X, Y, Z\}$ and relations (13) as in the proof of the Proposition 4.2.

$$\begin{aligned}L_2 &= -\varepsilon\left(\cos(s_1) - Y\sin(s_1)\right)X\left(\cos(s_1) + Y\sin(s_1)\right)\\ &= -\varepsilon\left(X\cos(2s_1) + Z\sin(2s_1) + 2cY\sin^2(s_1)\right).\end{aligned}$$

Next,

$$\begin{aligned}L_1 &= \exp(-s_2 X)L_2\exp(s_2 X)\\ &= -\varepsilon\left(\cos(s_2) - X\sin(s_2)\right)\left(X\cos(2s_1) + Z\sin(2s_1) + 2cY\sin^2(s_1)\right)\left(\cos(s_2) + X\sin(s_2)\right)\\ &= -\varepsilon\Big(X\left(\cos(2s_1) - c\sin(2s_1)\sin(2s_2) + 4c^2\sin^2(s_1)\sin^2(s_2)\right)\\ &+Y\left(\sin(2s_1)\sin(2s_2) + 2c\sin^2(s_1)\cos(2s_2)\right) + Z\left(\sin(2s_1)\cos(2s_2) - 2c\sin^2(s_1)\sin(2s_2)\right)\Big).\end{aligned}$$

Doing the same calculations for the right hand side of (17), we get

$$L = L_3+L_4+L_5+L_6+\frac{1}{2}\left([L_3,L_4]+[L_3,L_5]+[L_3,L_6]+[L_4,L_5]+[L_4,L_6]+[L_5,L_6]\right)+o(\varepsilon^2), \tag{19}$$

where

$$\begin{aligned}L_3 &= \exp(-s_2 X)(\varepsilon_1 Y)\exp(s_2 X),\\ L_4 &= \varepsilon_2 X, \quad L_5 = -\varepsilon_2' Y,\\ L_6 &= \exp(-s_1 Y)(-\varepsilon_3 X)\exp(s_1 Y).\end{aligned}$$

We begin by solving (17) to the first order in ε. Equating (18) with (19) we get:

$$
\begin{aligned}
&-\varepsilon X\left(2\cos(2s_1) - c\sin(2s_1)\sin(2s_2) + 4c^2\sin^2(s_1)\sin^2(s_2)\right)\\
&-\varepsilon Y\left(\sin(2s_1)\sin(2s_2) + 4c\sin^2(s_1)\cos^2(s_2)\right) - \varepsilon Z\left(2\sin(2s_1)\cos^2(s_2) - 2c\sin^2(s_1)\sin(2s_2)\right)\\
&\quad = X\left(2\varepsilon_1 c\sin^2(s_2) + \varepsilon_2 - \varepsilon_3\cos(2s_1)\right) + Y\left(\varepsilon_1\cos(2s_2) - \varepsilon_2' - 2\varepsilon_3 c\sin^2(s_1)\right)\\
&\qquad + Z\left(-\varepsilon_1\sin(2s_2) - \varepsilon_3\sin(2s_1)\right).
\end{aligned}
$$

We divide both sides of this equation by $\cos^2(s_1)\cos^2(s_2)$, which allows us to express everything in terms of $\tan(s_1), \tan(s_2)$. Using the relation $\tan(s_2) = \kappa\tan(s_1)$, we further eliminate $\tan(s_2)$. To make the equations more compact we denote $\tan(s_1)$ by x. By Proposition 4.3 we have $0 < x \le 1$.

Since we also have the relation $\kappa\tan(s_1 + \varepsilon_2') = \tan(s_2 + \varepsilon_2)$, we use the Taylor expansion to find the relation between ε_2 and ε_2' to the first order:

$$\frac{\kappa\varepsilon_2'}{\cos^2(s_1)} = \frac{\varepsilon_2}{\cos^2(s_2)} + o(\varepsilon_2).$$

Expressing this in terms of x, we get

$$\varepsilon_2' = \varepsilon_2\frac{\kappa x^2 + \kappa^{-1}}{x^2 + 1} + o(\varepsilon_2).$$

Equating the coefficients at X, Y, Z, we get a system of equations

$$
\begin{aligned}
&2\varepsilon_1 c(1+x^2)\kappa^2x^2 + \varepsilon_2(1+x^2)(1+\kappa^2x^2) - \varepsilon_3(1+\kappa^2x^2)(1-x^2)\\
&\qquad = -\varepsilon\left(2(1-x^2)(1+\kappa^2x^2) - 4c\kappa x^2 + 4c^2\kappa^2x^4\right) + o(\varepsilon),\\
&\varepsilon_1(1+x^2)(1-\kappa^2x^2) - \varepsilon_2\kappa^{-1}(1+\kappa^2x^2)^2 - 2\varepsilon_3 cx^2(1+\kappa^2x^2) = -4\varepsilon(\kappa + c)x^2 + o(\varepsilon),\\
&-2\varepsilon_1\kappa x(1+x^2) - 2\varepsilon_3 x(1+\kappa^2x^2) = -\varepsilon\left(4x - 4c\kappa x^3\right) + o(\varepsilon).
\end{aligned}
$$

The determinant of this system is the jacobian of (17) and equals

$$4x(1+x^2)(1+\kappa^2x^2)^2(1 - \kappa^2x^4 + 2c\kappa x^2 + c\kappa x^4 + c\kappa^3x^4).$$

Since $0 < x \le 1$, $0 < \kappa \le 1$, $0 \le c < 1$, we see that the only case when the jacobian vanishes is $x = 1$, $\kappa = 1$, $c = 0$. We will consider this case separately below. In all other cases the jacobian is non-zero, hence by the Implicit Function Theorem, equation (17) has a unique solution for small ε.

Solving (17) to the second order in ε, we get that the cost of the right hand side of (17) is

$$
\begin{aligned}
&\kappa(s_1 + \varepsilon_1) + (s_2 + \varepsilon_2) + \kappa(s_1 + \varepsilon_2') + \varepsilon_3\\
&= 2\kappa s_1 + s_2 + 2\varepsilon - 2\varepsilon^2\frac{\kappa x(1 - c\kappa x^2)(2 - \kappa^2x^4 + \kappa^2x^2 + 3c^2x^2 + c^2x^4 + 2c^2\kappa^2x^4)}{(1+x^2)(1 - \kappa^2x^4 + 2c\kappa x^2 + c\kappa x^4 + c\kappa^3x^4)} + o(\varepsilon^2),
\end{aligned}
$$

which is lower than the cost of the left hand side $2\kappa s_1 + s_2 + 2\varepsilon$.

For the remaining case $x = 1$, $\kappa = 1$, $c = 0$, we have $s_1 = s_2 = \frac{\pi}{4}$.

Applying Proposition 4.2 we see that

$$\exp\left(\frac{\pi}{4}Y\right)\exp\left(\frac{\pi}{4}X\right)\exp\left(-\frac{\pi}{4}Y\right) = \exp\left(-\frac{\pi}{4}Y\right)\exp\left(-\frac{\pi}{4}X\right)\exp\left(\frac{\pi}{4}Y\right),$$

and thus

$$\exp\left(-\varepsilon X\right)\exp\left(\frac{\pi}{4}Y\right)\exp\left(\frac{\pi}{4}X\right)\exp\left(-\frac{\pi}{4}Y\right)\exp\left(-\varepsilon X\right)$$
$$=\exp\left(-\varepsilon X\right)\exp\left(-\frac{\pi}{4}Y\right)\exp\left(-\frac{\pi}{4}X\right)\exp\left(\frac{\pi}{4}Y\right)\exp\left(-\varepsilon X\right),$$

which is not optimal since it does not correspond to any trajectory in Fig. 2.

This completes the proof of the proposition for the decomposition $R(-\delta X)R(t_Y Y)R(t_X X)R(-t_Y Y)R(-\delta X)$. The cases of the decompositions obtained from this one by applying symmetries $(X,Y)\mapsto\{\pm X,\pm Y\}$ are analogous. For example, in the case of $R(-\delta Y)R(t_X X)R(t_Y Y)R(-t_X X)R(-\delta Y)$, we use the transformation

$$\exp(-\varepsilon Y)\exp(s_1X)\exp(s_2Y)\exp(-s_1X)\exp(-\varepsilon Y)$$
$$=\exp((s_1+\varepsilon_1)X)\exp((s_2+\varepsilon_2)Y)\exp(-(s_1+\varepsilon_2')X)\exp(-\varepsilon_3Y). \quad (20)$$

The cost of the right hand side is

$$2s_1+\kappa s_2+2\kappa\varepsilon-2\varepsilon^2\frac{\kappa x(1-c\kappa x^2)(2+x^2-\kappa^2x^4+3c^2\kappa^2x^2+2c^2\kappa^2x^4+c^2\kappa^4x^4)}{(1+\kappa^2x^2)(1-\kappa^2x^4+c\kappa x^4+2c\kappa x^2+c\kappa^3x^4)}+o(\varepsilon^2),$$

which is lower than the cost $2s_1+\kappa s_2+2\kappa\varepsilon$ of the left hand side. □

It follows from Proposition 4.4 that for the optimal decompositions corresponding to the trajectory $\ldots\mapsto\boxed{1}\mapsto\boxed{8}\mapsto\boxed{14}\mapsto\boxed{11}\mapsto\boxed{1}\mapsto\ldots$, and symmetric to it, the number of factors is at most 4. This corresponds to pattern (I) in Theorems 2.1 – 2.3.

Suppose $\kappa > c$. For the trajectory $\ldots\mapsto\boxed{3}\mapsto\boxed{5}\mapsto\boxed{3}\mapsto\ldots$, and symmetric to it, the number of factors is bounded by 2, since the evolution times $\boxed{3}\mapsto\boxed{5}$ and $\boxed{5}\mapsto\boxed{3}$ exceed π, and optimal decompositions can not have such time parameters. The case of two factors is incorporated in patterns (III) and (VII) with $t=0$ in Theorems 2.1 and 2.2.

In the case $0<\kappa\le c$, the decompositions corresponding to the trajectory $\boxed{5}\mapsto\boxed{3}\mapsto\boxed{5}$, could have up to 3 factors, since Y-evolution time $\boxed{5}\mapsto\boxed{3}$ exceeds π, but X-evolution time does not exceed $2\widehat{t}_X$, which is less than π. This corresponds to pattern (VIII) in Theorem 2.3.

To complete the proof of Theorems 2.1 – 2.4, we need to establish a bound on the number of factors for the trajectories that pass through the critical points. For the critical points $\pm W_-$ with $c\neq 0$ the existence of such a bound immediately follows from the diagrams in Fig. 2, since the trajectories connected to these points are full circles and require time evolution of 2π to complete the circle, while any evolution with time exceeding π is not optimal.

In the case of the critical points $\pm W_+$ we need to deal with a trajectory $\ldots\mapsto\boxed{9}\mapsto\boxed{6}\mapsto\boxed{12}\mapsto\boxed{12}\mapsto\boxed{19}\mapsto\boxed{9}\mapsto\boxed{9}\mapsto\boxed{15}\mapsto\ldots$, and other similar to it. Again, we will establish a bound on the number of switches.

The trajectory $\boxed{9}\mapsto\boxed{6}\mapsto\boxed{12}$ corresponds to the product

$$R(\widehat{t}_XX)R(\widehat{t}_YY)\quad\text{with }\widehat{t}_X=\arccos\left(\frac{c-\kappa}{c+\kappa}\right),\quad \widehat{t}_Y=\arccos\left(-\frac{1-\kappa c}{1+\kappa c}\right).$$

We are going to see that this element of $SO(3)$ is a rotation in angle π around an axis, which is orthogonal to W_+.

Proposition 4.5. *(a) The products $R(\widehat{t}_X X)R(\widehat{t}_Y Y)$ and $R(\widehat{t}_Y Y)R(\widehat{t}_X X)$ are both rotations in angle π.*

(b) The following relations hold:

$$
\begin{aligned}
R(\widehat{t}_X X)R(\widehat{t}_Y Y) &= R(-\widehat{t}_Y Y)R(-\widehat{t}_X X),\\
R(\widehat{t}_Y Y)R(\widehat{t}_X X) &= R(-\widehat{t}_X X)R(-\widehat{t}_Y Y),\\
R(\widehat{t}_Y Y)R(\widehat{t}_X X)R(tW_+) &= R(-tW_+)R(\widehat{t}_Y Y)R(\widehat{t}_X X),\\
R(\widehat{t}_X X)R(\widehat{t}_Y Y)R(-tW_+) &= R(tW_+)R(\widehat{t}_X X)R(\widehat{t}_Y Y).
\end{aligned}
$$

(c) Let $\kappa = 0$. Then $R(\pi Y)R(tW_+) = R(-tW_+)R(\pi Y)$.

Proof. Let us consider a preimage $h = \exp(\widehat{s}_X X)\exp(\widehat{s}_X Y)$ in $SU(2)$ for $R(\widehat{t}_X X)R(\widehat{t}_Y Y)$. Here $\widehat{s}_X = \widehat{t}_X/2$, $\widehat{s}_Y = \widehat{t}_Y/2$. It follows from (8) that

$$
\sin\widehat{s}_X = \sqrt{\frac{\kappa}{\kappa+c}}, \quad \cos\widehat{s}_X = \sqrt{\frac{c}{\kappa+c}}, \quad \sin\widehat{s}_Y = \frac{1}{\sqrt{1+\kappa c}}, \quad \cos\widehat{s}_Y = \sqrt{\frac{\kappa c}{1+\kappa c}}.
$$

Then

$$
\begin{aligned}
h = \exp(\widehat{s}_X X)\exp(\widehat{s}_X Y) &= (\cos\widehat{s}_X + X\sin\widehat{s}_X)(\cos\widehat{s}_Y + Y\sin\widehat{s}_Y)\\
&= \frac{1}{\sqrt{(\kappa+c)(1+\kappa c)}}\left(c\sqrt{\kappa} + X\kappa\sqrt{c} + Y\sqrt{c} + XY\sqrt{\kappa}\right)\\
&= \frac{1}{\sqrt{(\kappa+c)(1+\kappa c)}}\left(X\kappa\sqrt{c} + Y\sqrt{c} + Z\sqrt{\kappa}\right).
\end{aligned}
$$

By Lemma 4.1, the image of h in $SO(3)$ is a rotation in angle π and the first equality in part (b) holds. It is easy to see that h is orthogonal to W_+:

$$
\left(\kappa\sqrt{c}X + \sqrt{c}Y + \sqrt{\kappa}Z \middle| S - \kappa Q\right) = 0,
$$

which implies that the axis of rotation corresponding to h is orthogonal to W_+ and also that $hW_+h^{-1} = -W_+$. Taking the exponential of both sides, we get that $h\exp(sW_+)h^{-1} = \exp(-sW_+)$, from which the third claim of part (b) follows.

The argument for $R(\widehat{t}_Y Y)R(\widehat{t}_X X)$ is completely analogous.

For $\kappa = 0$ we have $(W_+, Y) = 0$, from which the claim (c) follows. ☐

Proposition 4.6. *Suppose an element $g \in SO(3)$ has an optimal decomposition containing factors $R(tW_+)$ or $R(tW_-)$. Then there is an optimal decomposition for g with a single factor of that type.*

Proof. First we consider the case $c \neq 0$ and $\kappa \neq 0$. We have pointed out above that the factor $R(tW_-)$ may only appear when $c < \kappa$ and there will be only one such a factor in that case. In case when $c = \kappa$, we have that W_- is proportional to X, and we do not need to consider the factors of the form $R(tW_-)$ at all.

Consider an optimal decomposition with factors $R(tW_+)$. Without loss of generality assume that the time parameter in the first such factor is positive. Then it will necessarily have the form

$$
h_0 R(t_1 W_+)h_1 R(-t_2 W_+)h_2 R(t_3 W_+)h_3 \dots h_{n-1}R((-1)^{n-1}t_n W_+)h_n,
$$

where $t_k \geq 0$ and for $1 \leq k \leq n-1$

$$h_k = \begin{cases} R(\widehat{t}_X X)R(\widehat{t}_Y Y) \text{ or } R(-\widehat{t}_Y Y)R(-\widehat{t}_X X) \text{ if } k \text{ is odd,} \\ R(\widehat{t}_Y Y)R(\widehat{t}_X X) \text{ or } R(-\widehat{t}_X X)R(-\widehat{t}_Y Y) \text{ if } k \text{ is even.} \end{cases}$$

By Proposition 4.5 we can combine all factors of type $R(tW_+)$ into one without changing the cost:

$$h_0 R((t_1 + t_2 + \ldots + t_n)W_+)h_1 h_2 \ldots h_{n-1} h_n.$$

Next suppose $\kappa = 0$. Using Proposition 4.5(c) and applying the above argument we see that there is an optimal decomposition with at most one factor of type $R(tW_+)$. This completes the proof of Theorem 2.4.

Now let us consider the case $c = 0$. Here we could have a decomposition that contains factors of both types, $R(tW_+)$ and $R(tW_-)$. Note that $\widehat{t}_X = \widehat{t}_Y = \pi$. Suppose that an optimal decomposition of $g \in SO(3)$ contains a factor $R(tW_+)$ with $t > 0$. This factor will be followed by either an X-evolution or $-Y$-evolution. Let us assume it is X-evolution that follows. If the time parameter for X-evolution is less than π, that will be the last factor in the decomposition, as the control switch can not occur. Otherwise, we get $R(tW_+)$ followed by a factor $R(\pi X)$. But this will imply optimality of the expression $R(tW_+)R(-\pi X)$, which gives a contradiction since $R(tW_+)R(-\varepsilon X)$ can not be optimal since it does not satisfy the necessary conditions for optimality of Theorem 3.3. All other cases are analogous and we conclude that in case $c = 0$ factors $R(tW_\pm)$ in optimal decompositions may be preceded or followed by just a single factor $R(t'X)$ or $R(t'Y)$ with $|t'| < \pi$, thus completing the proof of Theorem 2.1.

□

Finally, it remains to investigate the factors that could precede/follow $R(tW_+)$ in optimal decompositions. We are going to show that the number of such factors is at most two.

Proposition 4.7. *Let $c > 0$, $0 < \kappa \leq 1$. Suppose an optimal decomposition for $g \in SO(3)$ contains a factor $R(tW_+)$ with $t > 0$. Then there exists an optimal decomposition of g, which is a subword in one of the following:*

$$\begin{aligned} &R(\widehat{t}_Y Y)R(\widehat{t}_X X)R(tW_+)R(\widehat{t}_X X)R(\widehat{t}_Y Y), \\ &R(\widehat{t}_Y Y)R(\widehat{t}_X X)R(tW_+)R(-\widehat{t}_Y Y)R(-\widehat{t}_X X), \\ &R(-\widehat{t}_X X)R(-\widehat{t}_Y Y)R(tW_+)R(\widehat{t}_X X)R(\widehat{t}_X Y), \\ &R(-\widehat{t}_X X)R(-\widehat{t}_Y Y)R(tW_+)R(-\widehat{t}_Y Y)R(-\widehat{t}_X X). \end{aligned}$$

Proof. Let us show that in an optimal decomposition of g the number of factors following $R(tW_+)$ is at most two. Indeed, if it is followed by three or more factors, such evolution must begin with either $R(\widehat{t}_X X)R(\widehat{t}_Y Y)$ or $R(-\widehat{t}_Y Y)R(-\widehat{t}_X X)$. By Proposition 4.5, $R(\widehat{t}_X X)R(\widehat{t}_Y Y) = R(-\widehat{t}_Y Y)R(-\widehat{t}_X X)$. This will be followed by evolution with control $-X$ or Y. This would imply optimality of either $R(tW_+)R(\widehat{t}_X X)R((\widehat{t}_Y + \delta)Y)$ or $R(tW_+)R(-\widehat{t}_Y Y)R(-(\widehat{t}_X + \delta)X)$ for small $\delta > 0$. Let us show that these decompositions are not optimal. Consider a preimage $\exp(sW_+)\exp(\widehat{s}_X X)\exp((\widehat{s}_Y + \epsilon)Y)$ in $SU(2)$ for $R(tW_+)R(\widehat{t}_X X)R((\widehat{t}_Y + \delta)Y)$. Here $s = \frac{t}{2} > 0$, $\epsilon = \frac{\delta}{2} > 0$. Since $\kappa c > 0$ we get that $\widehat{s}_Y < \frac{\pi}{2}$

and we can assume that $\widehat{s}_Y + \epsilon < \frac{\pi}{2}$. Choose $0 < \tau < \frac{\pi}{2}$ such that $\tan\tau\tan(\widehat{s}_Y + \epsilon) = \frac{1}{c}$. Since $\tan\widehat{s}_X \tan\widehat{s}_Y = \frac{1}{c}$, we conclude that $0 < \tau < \widehat{t}_X$. Then by Proposition 4.2(a) we get

$$\exp(sW_+)\exp(\widehat{s}_X X)\exp((\widehat{s}_Y+\epsilon)Y) = -\exp(sW_+)\exp((\widehat{s}_X-\tau)X)\exp(-(\widehat{s}_Y+\epsilon)Y)\exp(-\tau X).$$

However the latter decomposition is not optimal since it does not correspond to a trajectory in Fig. 2, 3, yet both sides in the above equality have the same cost. This implies that $R(tW_+)R(\widehat{t}_X X)R((\widehat{t}_Y + \delta)Y)$ is not optimal. The argument for $R(tW_+)R(-\widehat{t}_Y Y)R(-(\widehat{t}_X + \delta)X)$ is analogous. This completes the proof of Proposition 4.7 and Theorems 2.2 and 2.3. $\square$

References

[1] Y. Billig, Time-optimal decompositions in $SU(2)$, Quantum Information Processing **12** (2013), 955-971.

[2] L. Euler, Formulae generales pro translatione quacunque corporum rigidorum, Novi Commentarii Academiae Scientiarum Petropolitanae **20** (1776), 189-207.

[3] V. Jurdjevic, Geometric control theory, Cambridge studies in advanced mathematics **51**, Cambridge University Press, New York, 1997.

[4] A. W. Knapp, Lie groups beyond an introduction, Progress in Mathematics **140**, Birkhäuser, Boston, 1996.

[5] NASA, (2013). NASA ends attempts to fully recover Kepler spacecraft, potential new missions considered. [online] Available at: http://www.nasa.gov/content/nasa-ends-attempts-to-fully-recover-kepler-spacecraft-potential-new-missions-considered [Accessed 14 Aug. 2014].

[6] P. Petersen, Riemannian geometry, Graduate Texts in Mathematics **171**, Springer, 2006.

[7] G. V. Smirnov, Introduction to the theory of differential inclusions, Graduate Studies in Mathematics **41**, Amer. Math. Soc., 2002.

School of Mathematics and Statistics, Carleton University, Ottawa, Canada
E-mail address: `billig@math.carleton.ca`