CrossMark

On Symmetries in Time Optimal Control, Sub-Riemannian Geometries, and the $K - P$ Problem

Francesca Albertini[1] · Domenico D'Alessandro[2]

Abstract The goal of this paper is to describe a method to solve a class of time optimal control problems which are equivalent to finding the sub-Riemannian minimizing geodesics on a manifold M. In particular, we assume that the manifold M is acted upon by a group G which is a symmetry group for the dynamics. The action of G on M is proper but not necessarily free. As a consequence, the *orbit space* M/G is not necessarily a manifold but it presents the more general structure of a *stratified space*. The main ingredients of the method are a reduction of the problem to the orbit space M/G and an analysis of the reachable sets on this space. We give general results relating the stratified structure of the orbit space, and its decomposition into *orbit types*, with the optimal synthesis. We consider in more detail the case of the so-called $K - P$ problem where the manifold M is itself a Lie group and the group G is determined by a Cartan decomposition of M. In this case, the geodesics can be explicitly calculated and are analytic. As an illustration, we apply our method and results to the complete optimal synthesis on $SO(3)$.

Keywords Geometric optimal control theory · Lie transformation groups · Symmetry reduction

Mathematics Subject Classification (2010) 93B27 · 58E10 · 58D19

✉ Francesca Albertini
  francesca.albertini@unipd.it

  Domenico D'Alessandro
  daless@iastate.edu

[1] Dipartimento di Tecnica e Gestione dei Sistemi Industriali, Università di Padova, Padua, Italy

[2] Department of Mathematics, Iowa State University, Ames, IA, USA

1 Introduction

In a recent paper [4], we have solved the time optimal control problem for a system on $SU(2)$ using a method which exploits the symmetries of the problem and provides an explicit description of the reachable sets at every time. In this paper, we formalize such methodology in general and give results (proved in Sections 3 and 4) linking the structure of a G-manifold[1] to the optimal synthesis. As an example of application, we provide the complete optimal synthesis for a minimum time problem on $SO(3)$, which complements some of the results of [7] obtained with a different method.

In order to introduce some of the ideas we shall explore, we provide a brief summary of the treatment of [4] for the problem on $SU(2)$, in its simplest formulation, from the point of view we will take in this paper. The problem is to control in minimum time the system

$$\dot{X} = -iu_x\sigma_x X - iu_y\sigma_y X, \qquad X(0) = \mathbf{1}, \tag{1}$$

to a desired final condition $X_f \in SU(2)$, subject to a bound on the L_2 norm of the control, i.e., $u_x^2 + u_y^2 \leq 1$. Here, σ_x and σ_y are *Pauli* matrices:

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \qquad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

The matrices, $i\sigma_x$ and $i\sigma_y$, span a subspace of $su(2)$ which is invariant under the operation of taking a similarity transformation using a *diagonal* matrix D in $SU(2)$, i.e., $A \in su(2) \to DAD^\dagger \in su(2)$. In this respect, the first observation is that if $X := X(t)$ is an optimal trajectory to go from the identity to X_f, then $DXD^\dagger := DX(t)D^\dagger$ is an optimal trajectory from the identity to $DX_f D^\dagger$. Therefore, once we have a minimizing geodesic leading to X_f, we also have a minimizing geodesic for every element $DX_f D^\dagger$ in the "orbit" of X_f and all such geodesics project to a unique curve in the space of orbits, $SU(2)/G$, where G denotes the subgroup of diagonal matrices. The second observation concerns the nature of the orbit space $SU(2)/G$. Since a general matrix in $SU(2)$ can be written as

$$X := \begin{pmatrix} x & y \\ -y^* & x^* \end{pmatrix}, \qquad |x|^2 + |y|^2 = 1 \tag{3}$$

and a similarity transformation by a diagonal matrix only affects the phase of the off-diagonal elements, an orbit is uniquely determined by the complex value x, with $|x| \leq 1$, i.e., an element of the unit disc in the complex plane which is therefore in one to one correspondence with the elements of $SU(2)/G$. With these facts, we studied in [4] the whole optimal synthesis in the unit disc. Since the problem has a $K - P$ structure (cf. Section 4), the candidate optimal trajectories can be explicitly expressed in terms of some parameters to be determined according to the desired final condition X_f. The number is reduced to only one if we consider the projection on the unit disc of these trajectories. Fixing the time t and varying such parameter, we obtained, as parametric curves, the boundary of the reachable in the unit disc, or, more precisely, the boundary of the projection of the reachable set onto the orbit space. Once an explicit description of the reachable sets is available, a method to determine the optimal controls is obtained as a consequence.

The study of the role of symmetries in optimal control problems is a fundamental subject in geometric control theory, important both from a conceptual point of view and a practical one as it allows us to reduce the problem to a smaller state (quotient) space. This *symmetry*

---

[1]That is, a manifold with the action of a Lie transformation group.

reduction in control problems has a long history (see, e.g., [10, 13–15, 18, 21, 26] and see, in particular, [27] for a recent account). It is obtained from the application of techniques in *geometric mechanics* such as in [19, 20]. However, typically, translation of these results of geometric mechanics in control theory has been restricted to the case where the action of the symmetry group G on the underlying manifold M is not only *proper* but also *free* (definitions are given in Section 2). In this case the orbit space M/G is guaranteed to be a manifold. In the case where such an action is not free, the orbit space M/G is a *stratified space* [8]. This is the case discussed here. One example is the abovementioned (closed) unit disc which is a manifold with boundary, a special case of a stratified space.

We have kept the paper as much as possible self-contained introducing several concepts from the beginning. In particular, the paper is organized as follows: In Section 2, we give the necessary background on sub-Riemannian geometry and how it connects with the time-optimal control problem (we refer to [1, 2] and [24] for a detailed treatment). This section also contains the basic facts on Lie transformation groups, in particular the decomposition of the orbit space into *orbit types* (see, e.g., [8]). In Section 3, we present results linking the geometry of the orbit space with the geometry of the optimal synthesis in optimal control. In Section 4, we apply and expand these results to the case where the problem has an underlying $K - P$ structure. As an example, we apply our results to determine the geometry of the optimal synthesis for a control system on $SO(3)$ in Section 5.

2 Background

In the next two subsections, we summarize some basic concepts in sub-Riemannian geometry and optimal control. We refer to [1, 2, 24] for introductory monographs on the subject.

2.1 Sub-Riemannian Structures and Minimizing Geodesics

Given a *Riemannian manifold*, M, a *sub-Riemannian structure* on M is given by a sub-bundle, Δ, of the tangent bundle TM. Letting $\pi_\Delta : \Delta \to M$ be the canonical projection, Δ is a vector bundle on M, whose fibers at $x \in M$, $\Delta_x := \pi_\Delta^{-1}(x) \subseteq T_x M$, are assumed to have constant dimension, i.e., $\dim \Delta_x := m$ independently of x. In the control theoretic setting, a sub-Riemannian structure is often described by giving a set of m, linearly independent, smooth vector fields (a *frame*) on M, $\mathcal{F} := \{X_1, X_2, \ldots, X_m\}$, such that at every point $x \in M$, $\mathrm{span}\{X_1(x), X_2(x), \ldots, X_m(x)\} = \Delta_x$. It is assumed that $\mathcal{F}$ is *bracket generating*: the smallest Lie algebra of vector fields containing $\mathcal{F}$, i.e., the Lie algebra *generated* by $\mathcal{F}$, $Lie\,\mathcal{F}$, is such that, at every point $x \in M$, $Lie\,\mathcal{F}(x) = T_x M$. Since M is a Riemannian manifold, by restricting the Riemannian metric to $\Delta_x \subseteq T_x M$ at every $x \in M$, we obtain a smoothly varying positive definite inner product for vectors in Δ_x, which we will denote by $\langle \cdot, \cdot \rangle$. We shall assume that the given frame $\mathcal{F}$ is *orthonormal* with respect to this inner product, that is, $\langle X_j(x), X_k(x) \rangle = \delta_{j,k}$, for every $x \in M$.

We shall consider *horizontal curves* on M. A curve $\gamma : [0, T] \to M$ is assumed to be *Lipschitz continuous* and therefore differentiable almost everywhere in $[0, T]$, with $\dot\gamma$ *essentially bounded*. That is, there exists a constant N and a map $H : [0, T] \to TM$, with $H(t) \in T_{\gamma(t)}M$, such that $\langle H(t), H(t) \rangle_R \leq N$, for every $t \in [0, T]$, and such that $H(t) = \dot\gamma(t)$, almost everywhere in $[0, T]$. Here, $\langle \cdot, \cdot \rangle_R$ denotes the original Riemannian metric on M from which the sub-Riemannian metric $\langle \cdot, \cdot \rangle$ is derived. We shall assume a curve γ to be *regular*, that is $\dot\gamma(t) \neq 0$, almost everywhere in $[0, T]$. A curve γ is said to

be *horizontal* if $\dot{\gamma}(t) \in \Delta_{\gamma(t)}$ almost everywhere in $[0, T]$. Given the orthonormal frame $\mathcal{F} := \{X_1, \ldots, X_m\}$, this implies that we can write, almost everywhere in $[0, T]$,

$$\dot{\gamma}(t) = \sum_{j=1}^{m} u_j(t) X_j(\gamma(t)), \tag{4}$$

with the functions u_j, $j = 1, \ldots, m$, given by $u_j(t) = \langle X_j(\gamma(t)), \dot{\gamma}(t) \rangle$. We remark that, because of the smoothness of the X_j's, the continuity of γ on the compact set $[0, T]$ and the fact that $\dot{\gamma}$ is essentially bounded, the functions u_j are also essentially bounded. Therefore, a horizontal curve determines m essentially bounded "control" functions, $u_1, \ldots, u_m$, satisfying Eq. 4 while, viceversa, given m essentially bounded control functions $u_1, \ldots, u_m$, the solution of Eq. 4 gives a horizontal curve.

A horizontal curve γ has a *length*, $l(\gamma)$, which is given by its length in the Riemannian geometry sense, i.e., (using Eq. 4)

$$l(\gamma) := \int_0^T \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle} dt = \int_0^T \sqrt{\sum_{j=1}^{m} u_j^2(t)} dt. \tag{5}$$

A horizontal curve γ, in the interval $[0, T]$, is said to be *parametrized by a constant* if $\langle \dot{\gamma}, \dot{\gamma} \rangle$ is constant, almost everywhere in $[0, T]$. It is said *parametrized by arclength* if such a constant is equal to one. The image of a curve γ in M as well as its length do not change if we *re-parametrize* the time t. A reparametrization is a Lipschitz, monotone and surjective map $\phi : [0, T'] \to [0, T]$, and a reparametrization of a curve γ is a curve $\gamma_\phi := \gamma \circ \phi : [0, T'] \to M$. Given a horizontal curve γ of length L and $\alpha > 0$, consider the increasing map $s : [0, T] \to [0, \alpha L]$,

$$s(t) := \int_0^T \alpha \|\dot{\gamma}(r)\| dr, \tag{6}$$

which is invertible. Let ϕ be the inverse map $\phi : [0, \alpha L] \to [0, T]$. Then a standard chain rule argument shows that the re-parametrization $\gamma_\phi := \gamma \circ \phi$ is parametrized by a constant $\frac{1}{\alpha}$, and in particular it is parametrized by arclength if $\alpha = 1$. Viceversa every horizontal curve is the reparametrization of a curve parametrized by a constant. We refer to [1] (Lemma 3.14 and Lemma 3.15) for details.

Given two points, q_0 and q_1, the *sub-Riemannian distance* between them, $d(q_0, q_1)$, is defined as the infimum of the lengths of all horizontal curves γ, such that $\gamma(0) = q_0$, and $\gamma(T) = q_1$. This is obviously greater or equal than the Riemannian distance between the two points where the infimum is taken among all the Lipschitz continuous curves, not necessarily horizontal. The *Chow-Raschevskii theorem* states that if M is connected, in the above described situation and in particular under the bracket generating assumption for $\mathcal{F}$, (M, d) is a *metric space* and its topology as a metric space is equivalent to the one of M. This theorem has several consequences including the fact that, for any two points q_0 and q_1 in M, the distance $d(q_0, q_1)$ is finite, i.e., there exists a horizontal curve γ joining q_0 and q_1 having finite length. Moreover, once q_0 is fixed $d(q_0, q_1)$ is continuous as a function of q_1. A *minimizing geodesic* γ joining q_0 and q_1, is a horizontal curve which realizes the sub-Riemannian distance $d(q_0, q_1)$. The *existence theorem* says that if M is a complete metric space, and in particular if it is compact, then there exists a minimizing geodesic for any pair of points q_0 and q_1 in M. We shall assume this to be the case in the following.

2.2 Time Optimal Control

The problem we shall consider will be, once $q_0 \in M$ is fixed, to characterize the minimizing geodesic connecting q_0 to q_1 for any $q_1 \in M$. This problem is related to the *minimum time optimal control problem* as described in the following theorem (cf., e.g., [1]).

Theorem 1 *The following two facts are equivalent:*

1. $\gamma : [0, T] \rightarrow M$ *is a minimizing sub-Riemannian geodesic joining q_0 and q_1, parametrized by constant speed L.*
2. $\gamma : [0, T] \rightarrow M$ *is a minimum time trajectory of* Eq. 4*, subject to $\gamma(0) = q_0$ and $\gamma(T) = q_1$, and subject to $\|\vec{u}\| \leq L$, almost everywhere.*

Proof The proof that $1 \rightarrow 2$ is obtained by contradiction. If 1 is true and 2 is not true, then there exists an essentially bounded control $\tilde{\vec{u}}$, with $\|\tilde{\vec{u}}\| \leq L$, and a corresponding solution of Eq. 4, $\tilde{\gamma}$, with $\tilde{\gamma}(0) = q_0$, and $\tilde{\gamma}(T_1) = q_1$, and $T_1 < T$. Calculate the length of $\tilde{\gamma}$,

$$\int_0^{T_1} \|\dot{\tilde{\gamma}}\| dt = \int_0^{T_1} \|\tilde{\vec{u}}\| dt \leq LT_1 < LT = \int_0^T \|\dot{\gamma}\| dt, \tag{7}$$

which contradicts the fact that γ is a minimizing geodesic.

Let us prove now that $2 \rightarrow 1$. First observe that γ must be indeed parametrized by constant speed. Since the vector fields X_j in Eq. 4 are orthonormal, we know that $\|\dot{\gamma}\| = \|\vec{u}\|$ almost everywhere. However, $\|\vec{u}\|$ (and therefore $\|\dot{\gamma}\|$) must be equal to L almost everywhere. In fact, assume $\|\vec{u}\| < L - \epsilon$, for some $\epsilon > 0$, on an interval of positive measure $[t_1, t_2]$, with $\gamma(t_1) := \bar{q}_1$, and $\gamma(t_2) := \bar{q}_2$. Direct computation shows that with the "re-scaled" control $\vec{u}_R := \frac{L}{L-\epsilon} \vec{u} \left(\frac{L}{L-\epsilon} t - \frac{\epsilon}{L-\epsilon} t_1 \right)$, the curve $\gamma_R := \gamma \left(\frac{L}{L-\epsilon} t - \frac{\epsilon t_1}{L-\epsilon} \right)$ is solution of Eq. 4 with $\gamma_R(t_1) = \gamma(t_1) = \bar{q}_1$ and $\gamma_R(t_1 + \frac{L-\epsilon}{L}(t_2 - t_1)) = \gamma(t_2) = \bar{q}_2$. Therefore, $\vec{u}_R$, which is an admissible control since its norm is bounded by L, achieves the transfer from $\bar{q}_1$ to $\bar{q}_2$ in time $\frac{L-\epsilon}{L}(t_2 - t_1) < (t_2 - t_1)$, which contradicts the optimality of $\vec{u}$. Moreover, γ has to be a minimizing geodesic with constant speed L. If there was another geodesic $\tilde{\gamma}$ with constant speed L, its length would be LT_1 which must be less than the length of γ, that is LT. This implies $T_1 < T$ and contradicts the optimality of the time T. □

In the following, we shall assume that our initial point q_0 is fixed and we shall look for the sub-Riemannian minimizing geodesics parametrized by constant speed L, or equivalently the minimum time trajectories (cf. Theorem 1) connecting q_0 to q_1, for any $q_1 \in M$. These curves describe the so called *optimal synthesis* on M. Two loci are important in the description of the optimal synthesis: The *critical locus* $\mathcal{CR}(M)$ is the set of points in M where minimizing geodesics loose their optimality, i.e., $p \in \mathcal{CR}(M)$ if and only if there exists a horizontal curve defined in $[0, T + \epsilon)$, with $T > 0$ and $\epsilon > 0$, such that $\gamma(0) = q_0$, $\gamma(T) = p$, γ is a minimizing geodesic joining q_0 and $\gamma(t)$, for every t in $(0, T)$ and it is not a minimizing geodesic for $t \in (T, T + \epsilon)$. The *cut locus* is the set of points $p \in M$ which are reached by two or more minimizing geodesics, i.e., $p \in \mathcal{CL}(M)$ if and only if there exists two horizontal curves γ_1 and γ_2, $[0, T] :\rightarrow M$ such that both γ_1 and γ_2 are optimal in $[0, T)$. Because of the existence of a minimizing geodesic, if $p \in \mathcal{CL}(M)$, at least one of the curves γ_1 and γ_2 is optimal for p, at time T. Points in the cut locus are called *cut points*.

Regularity of minimizing geodesics (cf. [25]) has consequences on the cut and critical locus. Next proposition proves that cut points are also critical points when analyticity is verified, this holds in the $K - P$ problem that will be treated in Section 4. We have

Proposition 2.1 *Assume that all minimizing geodesics are analytic functions of t defined in* $[0, \infty)$. *Then.*

$$\mathcal{CL}(M) \subseteq \mathcal{CR}(M).$$

Proof Assume $p \in \mathcal{CL}(M)$. Then beside the minimizing geodesic for p, $\gamma_1 : [0, \infty) \to M$, with $\gamma_1(T) = p$, there exists another horizontal curve $\gamma_2 : [0, \infty) \to M$, which is optimal on $[0, T)$ and satisfies $\gamma_2(T) = p$. At least one between γ_1 and γ_2 has to loose optimality at p. Therefore, $p \in \mathcal{CR}(M)$. If this is not the case, the concatenation of one of them until time T and the other after time T will also be optimal, which contradicts analyticity of the minimizing geodesics.   □

We shall also consider the *reachable sets* for system (4), with $\|\vec{u}\| \leq L$. The reachable set $\mathcal{R}(T)$ is the set of all points $p \in M$ such that there exists an essentially bounded function $\vec{u}$, with $\|\vec{u}\| \leq L$, a.e., such that the corresponding solution of Eq. 4, γ, satisfies, $\gamma(0) = q_0$ and $\gamma(T) = p$. We have that $T_1 \leq T_2$ implies $\mathcal{R}(T_1) \subseteq \mathcal{R}(T_2)$. Moreover if $\gamma = \gamma(t)$ is a time optimal trajectory on $[0, T]$ with $\gamma(T) = p$ then p belongs to the boundary of the reachable set $\mathcal{R}(T)$, which is a closed set since the set of values for the control is closed (cf. [12])

2.3 Symmetries and Lie Transformation Groups

In addition to the above sub-Riemannian structure, on the manifold M, we shall consider the *action of a Lie transformation group G* assuming that it is a *left* action,[2] and it is a *proper* action.[3] We shall also assume that the action map is smooth. We shall denote by M/G the *orbit space* of M under the action of G, i.e., the space of equivalence classes (orbits) $[p]$, where p_1 is equivalent to p_2, if and only if there exists a $g \in G$ such that $gp_1 = p_2$. $\pi : M \to M/G$ denotes the canonical projection, and M/G is endowed with the quotient topology. The study of the structure of M/G is part of the *theory of Lie transformation groups*. We now recall the main facts which are needed for our treatment. Details can be found in introductory monographs on the subject, such as, e.g., [8].

Two points x and y in M are said to be *of the same type* if their *isotropy groups* in G are conjugate. Recall, that the isotropy group of a point $x \in M$, G_x, is the subgroup of elements g of G, such that $gx = x$. Two subgroups H_1 and H_2 are conjugate if there exists a $g \in G$ such that the map $H_1 \to H_2$, $h \to ghg^{-1}$ is a group isomorphism. For any subgroup H of G, we denote by (H) the set of groups conjugate to H. The subset $M_{(H)} \subseteq M$ is the set of points of M whose isotropy group belongs to (H), or, in other words, the set of points whose isotropy group is conjugate to H. There will be only certain classes of groups (H) for which $M_{(H)}$ is not empty. These are called the *isotropy types*. It is known that $M_{(H)}$ is

---

[2]That is, for any $p \in M$, g_1 and g_2 in G, $(g_2 g_1)p = g_2(g_1 p)$. Every aspect of the theory goes through for right actions with minor modification, that is, $g_2(g_1 p) = (g_1 g_2)p$.

[3]That is, the action map $\alpha : G \times M \to M \times M$ defined by $\alpha(g, p) = (gp, p)$ is proper, that is, the preimage of any compact set is compact.

a *submanifold* of M (see, e.g., [23], 7.4). If two points x and y in M are on the same orbit, i.e., $y = gx$ and $h \in G_y$, then $hy = hgx = y = gx$, so that, $g^{-1}hg \in G_x$. This means that G_x and G_y are conjugate, and therefore, x and y both belong to $M_{(G_x)} = M_{(G_y)}$. A consequence of this is that $M_{(H)}$ is the inverse image of a set in M/G, $\pi(M_{(H)}) = M_{(H)}/G$, which is called the *isotropy stratum* of type (H). Isotropy strata have a smooth structure in M/G: They are smooth manifolds and the inclusion $M_{(H)}/G \to M/G$ is smooth (cf. e.g., [5]). We remark that M/G itself is not in general a smooth manifold. It is a smooth manifold if the action of G on M is *free* that is the isotropy group G_x is the trivial one given by the identity, for any $x \in M$. In that case, there exists only one possible (H) which contains only the trivial group composed of only the identity. Therefore, $M_{(H)}/G = M/G$ is a smooth manifold according to the above cited result. In general, both M and M/G have the structure of a *stratified space*.

Definition 2.2 A *stratification* of a topological space N is a partition of N into connected manifolds N_i, i.e., $N = \cup_i N_i$ which is *locally finite*, i.e., every compact set in N intersects only a finite number of N_i's. Moreover such a partition satisfies the *frontier condition*: If $N_i \cap \bar{N}_j \neq \emptyset$ then $N_i \subseteq \bar{N}_j$ and $\dim(N_i) < \dim(N_j)$.[4]

Consider $M := \cup_{(H)} M_{(H)}$, where the union is taken over all the isotropy types and further decompose each $M_{(H)}$ into its connected components, so as to obtain a partition of M, $M = \cup_i M_i$. Moreover, partition M/G as $M/G = \cup_i \pi(M_i) := \cup_i M_i/G$. Such partitions give a stratification of M and M/G, respectively (see, e.g., [22] Theorem 1.30).

On the sets of isotropy types (H), a *partial order* is established by saying that $(H_1) \leq (H_2)$ if H_1 is conjugate to a subgroup of H_2. This defines subsets in M and M/G: $M_{\leq H}$ is defined as

$$M_{\leq H} := \cup_{(H_1) \leq (H)} M_{(H_1)}, \tag{8}$$

with $M_{\leq (H)}/G = \cup_{(H_1) \leq (H)} M_{(H_1)}/G$. We have from the definition that $(H_1) \leq (H_2)$ implies $M_{\leq (H_1)} \subseteq M_{\leq (H_2)}$ and $M_{\leq (H_1)}/G \subseteq M_{\leq (H_2)}/G$. One of the fundamental results of the theory of transformation groups is the *theorem of existence of minimal orbit type*: *There exists a unique orbit type (K) such that $(K) \leq (H)$ for every orbit type (H). Moreover, $M_{(K)}/G$ is a connected, locally connected, open and dense set in M/G, which is a manifold of dimension* $\dim M_{(K)}/G = \dim M - \dim G + \dim K$ *(cf., e.g., [23])*. Notice in particular that if K is a discrete group the dimension of $M_{(K)}/G$ is $\dim M - \dim G$. The manifold $M_{(K)}/G$ $(M_{(K)})$ is called the *regular part* of M/G (M), while $M/G - M_{(K)}/G$ $(M - M_{(K)})$ is called the *singular part*.

Given the sub-Riemannian (and Riemannian) structure described in Section 2.1, with the initial point $q_0 \in M$, we shall say that the Lie transformation group G is a *group of symmetries* if the following conditions are verified: (Denote by Φ_g the smooth map on M which gives the action of G, $\Phi_g x := gx$)

1. q_0 is a fixed point for the action of G on M. That is,

$$\Phi_g(q_0) = q_0, \qquad \forall g \in G. \tag{9}$$

---

[4]The intuitive idea of the frontier connection is that smaller dimensional manifolds in the partition are either totally detached from higher dimensional manifolds (that is the intersection with the closure is empty) or they are part of the boundary.

2. If Δ denotes the distribution which defines the sub-Riemannian structure, the action of G satisfies the following invariance property, for every $p \in M$,

$$\Phi_{g*}\Delta_p = \Delta_{\Phi_g p}. \tag{10}$$

3. G is a group of *isometries* for the sub-Riemannian metric $\langle \cdot, \cdot \rangle$, that is, for every $p \in M$, and U, V in Δ_p

$$\langle U, V \rangle = \langle \Phi_{g*}U, \Phi_{g*}V \rangle. \tag{11}$$

In the next two sections of this paper, we investigate how the optimal synthesis is related to the orbit type decomposition in a sub-Riemannian structure where the group G is a group of symmetries for such a structure in the sense above specified.

3 Symmetries in the Time Optimal Control Problem

The following propositions clarify the role of symmetries and the corresponding orbit space decomposition in the optimal synthesis.

Proposition 3.1 *Let q_1 and q_2 be two points in M on the same orbit, i.e. $q_2 = gq_1$ for some $g \in G$. Let γ_1 be a minimizing sub-Riemannian geodesic parametrized by constant speed L (and therefore a minimum time trajectory for* Eq. 4 *subject to $\|\vec{u}\| \leq L$ (cf. Theorem 1), with $\gamma_1(0) = q_0$ and $\gamma_1(T) = q_1$. Then, $\gamma_2 := g\gamma_1$ is a minimizing sub-Riemannian geodesic parametrized by constant speed L (and therefore a minimum time trajectory) as well.*

Proof First notice that because of property (10), for almost every t in $[0, T]$, we have

$$\dot{\gamma}_2(t) = \Phi_{g*}\dot{\gamma}_1(t) \in \Delta_{g\gamma_1(t)} = \Delta_{\gamma_2(t)}, \tag{12}$$

so that γ_2 is horizontal. Moreover because of property (11), a.e.,

$$\langle \dot{\gamma}_2(t), \dot{\gamma}_2(t) \rangle = \langle \Phi_{g*}\dot{\gamma}_1, \Phi_{g*}\dot{\gamma}_1 \rangle = \langle \dot{\gamma}_1(t), \dot{\gamma}_1(t) \rangle = L, \tag{13}$$

so that γ_2 is also parametrized by constant speed L and $l(\gamma_2) = l(\gamma_1)$. It is the minimum length since a smaller length would contradict the minimality of γ_1. $\qquad\square$

As a consequence of the previous proposition, we have that the space M/G is a metric space with the distance $\bar{d}$ between the two orbits $\bar{q}_1$ and $\bar{q}_2$, defined as

$$\bar{d}(\bar{q}_1, \bar{q}_2) := \inf_{q_1 \in \bar{q}_1, q_2 \in \bar{q}_2} d(q_1, q_2), \tag{14}$$

where d is the sub-Riemannian distance (cf. the Chow-Rashevskii theorem). A geodesic connecting two points $\bar{q}_1$ and $\bar{q}_2$, in M/G is a curve that achieves such an infimum.

Corollary 3.2 *The distance $\bar{d}(\bar{q}_0, \bar{q}_1)$ is achieved by $\pi(\gamma)$ where γ is a minimizing sub-Riemannian geodesic connecting q_0 to any q_1, independently of the representative $q_1 \in \bar{q}_1$.*

Therefore, the optimal synthesis in M is the inverse image of the optimal synthesis in M/G. Furthermore, on M/G, we can define *critical locus*, *cut locus,* and *reachable sets*, in terms of geodesics in exactly the same way we defined them on M. These sets are the projections of the corresponding sets in M. We have:

Proposition 3.3

1.

$$\mathcal{R}(T) = \pi^{-1}(\pi(\mathcal{R}(T))). \tag{15}$$

2.

$$\mathcal{CR}(M) = \pi^{-1}(\pi(\mathcal{CR}(M))). \tag{16}$$

3.

$$\mathcal{CL}(M) = \pi^{-1}(\pi(\mathcal{CL}(M))). \tag{17}$$

Proof Analogously to the proof of Proposition 3.1, if q_1 and q_2 are on the same orbit and there is a horizontal curve γ_1 connecting q_0 to q_1, with control $\vec{u}_1$, $\|\vec{u}_1\| = \|\dot{\gamma}_1\| \leq L$ in time T, then the curve $\gamma_2 := g\gamma_1$, for some $g \in G$, corresponds to control $\vec{u}_2$, with $\|\vec{u}_2\| = \|\dot{\gamma}_2\| = \|\dot{\gamma}_1\| \leq L$, a.e., connecting q_0 to q_2, in the same time T. Therefore, q_1 is in $\mathcal{R}(T)$ if and only if q_2 is in $\mathcal{R}(T)$, which proves (15). Analogously, we can prove that if q_1 and q_2 are on the same orbit they are both in $\mathcal{CL}(M)$ and $\mathcal{CR}(M)$ or none of them is, that is, Eqs. 16 and 17 also hold. We illustrate the proof for $\mathcal{CL}(M)$ (the proof of $\mathcal{CR}(M)$ is similar). Assume by contradiction that $q_1 \in \mathcal{CL}(M)$ and $q_2 \notin \mathcal{CL}(M)$, with $q_2 = gq_1$, for $g \in G$. Since $q_1 \in \mathcal{CL}(M)$, there are two different minimizing sub-Riemannian geodesics with q_1 as their final point, γ and $\tilde{\gamma}$. Then, $g\gamma$ and $g\tilde{\gamma}$ will be two different geodesics with q_2 as their final point. In fact, $g\gamma(t) = g\tilde{\gamma}(t)$, for all $t \in [0, T]$ would imply $\gamma(t) = \tilde{\gamma}(t)$, which we have excluded. □

Proposition 3.4 *Assume that q_1 does not belong to the cut locus, i.e., $q_1 \notin \mathcal{CL}(M)$ and γ is a sub-Riemannian minimizing geodesic in $[0, T]$ for q_1. Then, for every $t \in (0, T)$, we have the following relation for the isotropy groups*

$$G_{q_1} \subseteq G_{\gamma(t)}. \tag{18}$$

Proof If Eq. 18 is not valid then there exists a $g \in G_{q_1}$ with $g \notin G_{\gamma(t)}$. Then, the curve $g\gamma$ is also a geodesic since it has the same length as γ (cf. Proposition 3.1). Moreover, since $g \in G_{q_1}$, the geodesic $g\gamma$ goes to q_1, but it is different from γ since $g \notin G_{\gamma(t)}$, for some $t \in (0, T)$. Therefore, q_1 must belong to the cut locus. Something we have excluded. □

Under the assumption that geodesics are analytic, we can obtain for general $q_1 \in M$ the converse inclusion to Eq. 18.

Proposition 3.5 *If all geodesics are analytic then, for any $q_1 \in M$, and any geodesic $\gamma : [0, T] \to M$, connecting q_0 to q_1, we have for any $t \in (0, T)$*

$$G_{\gamma(t)} \subseteq G_{q_1}. \tag{19}$$

Proof The proof uses some ideas of Lemma 3.5 in [5]. Assume there exists a $\bar{t} \in (0, T)$ and a $g \in G_{\gamma(\bar{t})}$ which is not in G_{q_1}. Then, the curve, which is equal to γ between q_0 and $\gamma(\bar{t})$ and is equal to $g\gamma$ between $\gamma(\bar{t})$ and gq_1 (which is different from q_1), has the same length as the curve $g\gamma$. Such a curve is a minimizing geodesic, according to Proposition 2.1, since it has the same length as γ, that is the minimal length. However, such a curve which is equal to γ in an interval of positive measure until time t and different from γ afterwards cannot be analytic, which contradicts the analyticity of all the geodesics. □

We collect in the following corollary some consequences of Propositions 3.4 and 3.5. The corollary describes how minimizing sub-Riemannian geodesics sit in the orbit type decomposition of M and M/G.

Corollary 3.6 *If all minimizing sub-Riemannian geodesics are analytic, then for every $q \in M$ ($\bar{q} \in M/G$), every minimizing geodesic is entirely contained in $M_{\leq(G_q)}$ ($M_{\leq(G_q)}/G$) except possibly for the initial point q_0. In particular, if $q \in M_{(K)}$ (the regular part of M) then the whole minimizing geodesic except possibly the initial point q_0 is in $M_{(K)}$. If, in addition, $q \notin \mathcal{CL}(M)$ then*

$$G_{\gamma(t)} = G_q, \qquad \forall t \in (0, T) \tag{20}$$

and the corresponding sub-Riemannian geodesic is all contained in $M_{(G_q)}$.

Similar "convexity" results in the Riemannian case are given in ([5] 3.4). Corollary 3.6 gives a general principle for the behavior of geodesics in the presence of a group of symmetries for the optimal control problem:

The geodesics can only go from lower ranked strata such as the lowest $M_{(K)}$ to higher ranked ones but not viceversa. If a geodesic touches a higher ranked point and then goes back to a lower ranked one, it means that it has lost optimality and therefore the point belongs to the critical locus.

Remark 3.7 The corollary suggests that the points in the singular part of M, $M_{sing} := M - M_{(K)}$, are, in general, good candidates to be in the cut locus $\mathcal{CL}(M)$.[5] In fact, any point $q \in M_{sing}$ which has a geodesic with points in $M_{(K)}$, which is an *open and dense set* in M, must be in $\mathcal{CL}(M)$. Points in M_{sing} which are *not* in $\mathcal{CL}(M)$ must be such that *every geodesic* leading to that point must be entirely contained in the singular part of M since we have that Eq. 20 is verified. In the $SU(2)$ example of [4], it holds that $\mathcal{CL}(M) = \mathcal{CR}(M) = M_{sing}$. However, this is not always the case and in general the situation changes by changing the group of symmetries we consider (cf. Remark 5.7 below).

4 The $K - P$ Problem

An example of a sub-Riemannian problem with symmetries is the $K - P$ *problem* discussed in [6, 7]. In this section, we shall focus on this type of problems.

In the $K - P$ problem, the manifold M is a *semisimple Lie group* with its Lie algebra of right invariant vector fields $\mathcal{L}$. The Lie algebra $\mathcal{L}$ has a *Cartan decomposition*, that is, a vector space decomposition

$$\mathcal{L} = \mathcal{K} \oplus \mathcal{P}, \tag{21}$$

with the commutation relations:[6]

$$[\mathcal{K}, \mathcal{K}] \subseteq \mathcal{K}, \quad [\mathcal{K}, \mathcal{P}] \subseteq \mathcal{P}, \qquad [\mathcal{P}, \mathcal{P}] \subseteq \mathcal{K}. \tag{22}$$

The Lie algebra $\mathcal{L}$ is endowed with a (pseudo)-inner product defined by the *Killing form* $\langle A, D \rangle := Kill(A, D) := Tr(ad_A ad_D)$, where ad_A is the linear operator on $\mathcal{L}$, given

---

[5] And therefore in the critical locus $\mathcal{CR}(M)$ cf. Proposition 2.1.

[6] It is true (see [6], Appendix A; see also Lemma 3.4 and Corollary 3.5 in [9]) that for semi-simple Lie Algebras the equality must hold in the second and the third of these inclusions.

$ad_A(X) = [A, X]$. A comprehensive introduction to notions of Lie theory can be found for example in [17]. In particular, since $\mathcal{L}$ is assumed semisimple, $Kill$ is non degenerate (this is the Cartan criterion for semisimplicity). Associated with a Cartan decomposition is a *Cartan involution*, that is, an automorphism θ of $\mathcal{L}$, such that θ^2 is the identity, and $\mathcal{K}$ and $\mathcal{P}$ above are the $+1$ and -1 eigenspaces of θ in $\mathcal{L}$. Moreover, $B(A, D) := -Kill(A, \theta D)$ is a *positive definite* bilinear form and therefore an inner product defined on all of $\mathcal{L}$. Notice that this, in particular, implies that $\mathcal{K}$ is a compact subalgebra of $\mathcal{L}$, (i.e., $Kill(A, A) < 0$, if $A \in \mathcal{K}$ and $A \neq 0$) and $\mathcal{K}$ and $\mathcal{P}$ are orthogonal with respect to such inner product.[7] Using the inner product B on the Lie algebra $\mathcal{L}$, one naturally defines a Riemannian metric on M. In fact, for any point $b \in M$ and any tangent vector $U \in T_b M$, one can associate a right invariant vector field X_U defined as $X_U|_a := R_{b^{-1}a*}U$ and this association is an isomorphism of vector spaces. Then, the Riemannian metric $\langle \cdot, \cdot \rangle_R$ is defined as (with $U, V \in T_b M$)

$$\langle U, V \rangle_R := B(X_U, X_V). \tag{23}$$

The $K - P$ *problem* is the minimum time problem for system (4) on a Lie group M with a Cartan decomposition as described above, where the vector fields X_j are right-invariant vector fields[8] on M spanning $\mathcal{P}$ and orthonormal with respect the inner product B. The initial point q_0 is the identity of the group M. The problem is to steer from q_0 to an arbitrary final condition in M in minimum time, subject to the condition $\|\vec{u}\| \leq L$.

The problem can be cast in the above sub-Riemannian setting with symmetries as follows: The distribution of vector fields $\mathcal{P}$ in $\mathcal{L}$ defines a sub-Riemmannian structure on M with the sub-Riemannian metric at any point b defined by the restriction of $\langle \cdot, \cdot \rangle_R$ to $\mathcal{P}|_b$ for any b in M. Consider now a Lie subgroup of M, G (not necessarily connected) with Lie algebra $\mathcal{K}$, which acts on M by conjugation, i.e., for $p \in M$, $g \in G$, $\Phi_g(p) := gp := g \times p \times g^{-1}$, where $\times$ is the group operation in M. Such (left) action induces a map on the Lie algebra $\mathcal{L}$ given by its differential Φ_{g*} which is a Lie algebra automorphism. We assume that the map Φ_g is an *isometry* and that for every connected component j of G there exists a g_j such that $\Phi_{g_j*}\mathcal{P} \subseteq \mathcal{P}$. This also implies, because of the (Killing) orthogonality of $\mathcal{K}$ and $\mathcal{P}$, $\Phi_{g_j*}\mathcal{K} \subseteq \mathcal{K}$. Moreover, because of Eq. 22, these properties are not restricted to g_j but are true for every $g \in G$.[9] A special case is when G is

---

[7]If $A \in \mathcal{K}$ and $D \in \mathcal{P}$

$$B(A, D) := -Kill(A, \theta D) = Kill(A, D) = Kill(\theta A, \theta D) = Kill(A, \theta D) = -B(A, D),$$

where we have used the property of the Killing form that for every Lie algebra automorphism ϕ, $Kill(A, D) = Kill(\phi A, \phi D)$.

[8]Notice that we could have as well set up the whole treatment for *right* invariant vector fields but we could have given an analogous treatment for *left* invariant vector fields.

[9]Consider a connected component of G. We know that there exists a number of right invariant vector fields $X_1, X_2, \ldots, X_m$ in $\mathcal{K}$ such that denoting by $\sigma_{1,t}, \sigma_{2,t}, \ldots, \sigma_{m,t}$ the corresponding flows, we have $\sigma_{m,t_m} \circ \sigma_{m-1,t_{m-1}} \circ \cdots \circ \sigma_{1,t_1}(g_j) = g$. For every $r = 1, \ldots m$ the map $\sigma_{r,t}$ is real analytic as a function of t. Denote by $\bar{g} := \sigma_{1,t_1}(g_j)$. We want to show that $\Phi_{\bar{g}*}\mathcal{P} \subseteq \mathcal{P}$ and applying this m times we have that $\Phi_{g*}\mathcal{P} \subseteq \mathcal{P}$. Consider K in $\mathcal{K}$ and $P \in \mathcal{P}$ and the Killing inner product $B(K, \Phi_{\sigma_{1,t}(g_j)*}P)$ which is a real analytic function of t at every point in M and it is zero for $t = 0$. By taking the k-th derivative of this function at $t = 0$, we obtain, using the definitions of Lie derivative

$$\frac{d^k}{dt^k}|_{t=0}B(K, \Phi_{\sigma_{1,t}(g_j)*}P) = B(K, \frac{d^k}{dt^k}|_{t=0}\Phi_{\sigma_{1,t}(g_j)*}P) = B(K, ad_{X_1}^k \Phi_{g_j*}P) = 0,$$

where $ad_{X_1}^k$ denotes the $k-$the repeated Lie bracket with X_1 and we have used Eq. 22.

the connected component containing the identity with Lie algebra $\mathcal{K}$, in which case g_j can be taken equal to the identity.

In the following, we shall restrict ourselves to *linear Lie groups* so that M and G will be the Lie groups of matrices.[10] In the standard coordinates (inherited from the standard ones of $Gl(n, \mathbf{R})$ or $Gl(n, \mathbf{C})$), the system (4) is written as

$$\dot{X}(t) = \sum_j B_j X u_j(t), \tag{24}$$

where the Lie algebra $\mathcal{L}$ of matrices of the Lie group M has a Cartan decomposition as in Eqs. 21 and 22 and the B_j's span an orthonormal basis of $\mathcal{P}$.[11] The $K - P$ problem is the minimum time problem, with initial condition $X_0 = \mathbf{1}$ equal to the identity matrix subject to $||\vec{u}|| \leq L$. The symmetries are given by the transformations $X \to KXK^{-1}$, for $K \in G$, which induce transformations on the matrices $B \in \mathcal{L}$, $B \to KBK^{-1}$. These are symmetries because they preserve $\mathcal{P}$ and $\mathcal{K}$, and the commutation relations (22).

As a special case of what we have seen in general, the minimum time control for system (24) is equivalent to that of finding minimizing geodesics on M and it can be treated on M/G. On the orbit space, we can describe the cut locus, the critical locus, and the reachable sets.

Remark 4.1 The knowledge of the reachable sets for a $K - P$ problem of the form Eq. 24 also gives the reachable sets for the larger class of systems

$$\dot{X} = AX + \sum_j^m B_j X u_j(t), \tag{25}$$

with the drift AX, with $A \in \mathcal{K}$ and the $B_j \in \mathcal{P}$. In fact, consider the change of coordinates $U(t) := e^{-At} X(t)$. A straightforward calculation gives

$$\dot{U} = \sum_{j=1}^m e^{-At} B_j e^{At} U u_j,$$

and since $B \to e^{-At} B e^{At}$ is assumed to be an isometry, there exists an orthogonal matrix $a_{j,k} := a_{j,k}(t)$ such that $e^{-At} B_j e^{At} = \sum_{k=1}^m a_{j,k}(t) B_k$, so that we have

$$\dot{U} = \sum_{j=1}^m \sum_{k=1}^m a_{j,k}(t) B_k U u_j = \sum_{k=1}^m B_k U v_k(t), \tag{26}$$

where $v_k(t) := \sum_{j=1}^m a_{j,k}(t) u_j(t)$, and $||\vec{v}|| \leq L$ if and only if $||\vec{u}|| \leq L$. Therefore, the reachable set of system (26) coincides with the one of Eq. 4 and knowledge of the reachable set for system (26), $\mathcal{R}_U(t)$, gives the reachable set for system (25), $\mathcal{R}_X(t)$, via the relation $\mathcal{R}_X(t) = e^{At} \mathcal{R}_U(t)$.

---

[10] The example of $SU(2)$ treated in [4] and the example of $SO(3)$ of the next section are $K - P$ problems of this type.

[11] Here, with minor abuse of notation, we identify $\mathcal{K}$, $\mathcal{L}$, and $\mathcal{P}$ with the spaces of matrices representing the corresponding vector fields.

In the $K - P$ problem, the equations of the Pontryagin maximum principle are *explicitly integrable* and give (cf. [6] and references therein) that the optimal control $\vec{u}$ is such that there exist matrices $A_k \in \mathcal{K}$ and $A_p \in \mathcal{P}$ with

$$\sum_{j=1}^{m} B_j u_j(t) = e^{A_k t} A_p e^{-A_k t}, \tag{27}$$

with $\|A_p\| = L$. Therefore, the optimal trajectories satisfy

$$\dot{X} = e^{A_k t} A_p e^{-A_k t} X, \quad X(0) = \mathbf{1}, \quad A_k \in \mathcal{K}, \quad A_p \in \mathcal{P},$$

and the solution can be written explicitly as

$$X(t) = e^{-A_k t} e^{(A_k + A_p)t}. \tag{28}$$

The geodesics (28) are *analytic* curves. Therefore, all the results on the geometry of the optimal synthesis in the previous section apply. Moreover, for every geodesic in the orbit space M/G (which is the projection of a geodesic in M), we can always take a representative (28) in M with $A_p := A_a \in \mathcal{A}$, with $\mathcal{A}$ a *maximal Abelian (Cartan) subalgebra* in $\mathcal{P}$. To see this, recall the known property of the Cartan decomposition that if $\mathcal{A} \subseteq \mathcal{P}$ is a maximal Abelian subalgebra in $\mathcal{P}$, then

$$\mathcal{P} = \bigcup_{K \in G} K \mathcal{A} K^{-1}. \tag{29}$$

Therefore, we can write (27), for $A_a \in \mathcal{A}$, as

$$\sum_{j=1}^{m} B_j u_j(t) = K e^{\bar{A}_k t} A_a e^{-\bar{A}_k t} K^{-1}, \tag{30}$$

for $K \in G$ and $\bar{A}_k \in \mathcal{K}$. Using $A_p := K A_a K$, with $K \in G$, we have (cf. (28))

$$\left[e^{-A_k t} e^{(A_k + A_p)t} \right] = \left[e^{-A_k t} e^{(A_k + K A_a K^{-1})t} \right] = \left[K e^{-\bar{A}_k t} e^{(\bar{A}_k + A_a)t} K^{-1} \right]$$
$$= \left[e^{-\bar{A}_k t} e^{(\bar{A}_k + A_a)t} \right], \tag{31}$$

with $\bar{A}_k := K^{-1} A_k K$.

The following proposition gives some restrictions on the pairs (A_k, A_p) for points that are not on the cut locus of M. This proposition can be used to prove that a certain point is in the cut locus.[12]

Proposition 4.2 *Let $X_f \notin \mathcal{CL}(M)$. Let H denote the isotropy group of X_f. Then, the pair (A_k, A_p) giving the minimizing geodesic are such that for every $\hat{H} \in H$,*

$$\hat{H} ad_{A_k}^n A_p \hat{H}^{-1} = ad_{A_k}^n A_p, \tag{32}$$

for any $n \geq 0$.

Proof We know from Propositions 3.4 and 3.5 and from formula (28) that the pairs (A_k, A_p) satisfy the invariance property

$$\hat{H} e^{-A_k t} e^{(A_k + A_p)t} \hat{H}^{-1} = e^{-A_k t} e^{(A_k + A_p)t}, \tag{33}$$

---

[12]This is done for example in the next section in Proposition 5.3.

for every $t \in [0, T]$, where T is the minimum time associated to X_f. Taking the n-th derivative and by induction it is seen that this implies that $X^n(t)$ also satisfies the invariance property with respect to $\hat{H}$, i.e.,

$$\hat{H} X^n(t) \hat{H}^{-1} = X^n(t), \tag{34}$$

where $X^n(t)$ is defined as

$$X^{(n)}(t) := e^{-A_k t} H_n e^{(A_k + A_p)t}.$$

with

$$\begin{aligned} H_0 &= \mathbf{1}, \\ H_{n+1} &= H_n A_p + [H_n, A_k]. \end{aligned} \tag{35}$$

Using the invariance (34) of $X^n(t)$ at $t = 0$, it follows that all the matrices H_n are also $\hat{H}$-invariant. We want to show that this implies the invariance of

$$L_n := ad_{A_k}^{n-1} A_p,$$

for each $n \geq 1$. For $n = 1$ $L_1 = A_p$ and it is clear that L_n is invariant since $L_1 = H_1 = A_p$. From this, we proceed by induction on n, for $n \geq 2$. We shall prove that each $H_n, n \geq 2$, can be written, with $l_n = 2^{n-1} - 1$, as:

$$H_n = \sum_{j=1}^{l_n} V_j^n W_j^n + L_n, \tag{36}$$

with V_j^n and W_j^n invariant and both addends of some H_s, with $s < n$. From this, since H_n is also invariant, we must have that L_n is invariant as well.

First notice that $H_2 = A_p^2 + [A_p, A_k] = A_p A_p + L_2$, so clearly the statement holds for $n = 2$.

Assume that the statement holds for H_n, then we have

$$H_{n+1} = H_n A_p + [H_n, A_k] = \left(\sum_{j=1}^{l_n} V_j^n W_j^n A_p + L_n A_p \right) + \left(\sum_{j=1}^{l_n} [V_j^n W_j^n, A_k] + [L_n, A_k] \right). \tag{37}$$

Letting

$$V_j^{n+1} = V_j^n W_j^n, \quad W_j^{n+1} = A_p, \quad V_{l_n+1}^{n+1} = L_n, \quad W_{l_n+1}^{n+1} = A_p,$$

we can write

$$\left(\sum_{j=1}^{l_n} V_j^n W_j^n A_p + L_n A_p \right) = \sum_{j=1}^{l_n+1} V_j^{n+1} W_j^{n+1},$$

where V_j^{n+1} are invariant, since product of invariant, and they are addend of H_n, and $W_j^{n+1} = A_p$ is also invariant and it is in H_1. Now we have in Eq. 37

$$[V_j^n W_j^n, A_k] = V_j^n [W_j^n, A_k] + [V_j^n, A_k] W_j^n.$$

If W_j^n is one of the addends of H_s with $s < n$, then $[W_j^n, A_k]$ is one of the addends of H_{s+1}, so it is also invariant, by inductive assumption, and $s + 1 < n + 1$, so $V_j^n [W_j^n, A_k]$ is the product of two invariant factors which are addends of two H_p for some $p < n + 1$. The same argument applies to $[V_j^n, A_k] W_j^n$. So also the sum in the second brackets can be rewritten in the desired form. Now it is sufficient to notice that $[L_n, A_k] = L_{n+1}$. $\quad\square$

The previous considerations suggests a *general methodology* to find the optimal synthesis for time optimal control problems with symmetries and in particular for $K - P$ problems.

The *first* step of the method is to identify a group of symmetries. There are in general several choices of groups, connected and not connected. In the $K-P$ case, the natural choice is the connected Lie group corresponding to the subalgebra $\mathcal{K}$ in the Cartan decomposition or a possible not connected Lie group having $\mathcal{K}$ as its Lie algebra. It is typically convenient to take the Lie group G as large as possible so as to have a finer orbit type decomposition of M/G, which we would like to have of as small dimension as possible.

The *second* step of the procedure is to determine the nature of M/G so that the problem is effectively reduced to a lower dimensional space. This is important both from a conceptual and practical point of view since a computer solution of the problem will have to consider a smaller number of parameters. This task typically requires some analysis since not all the quotient spaces are known in the literature.[13] An analysis of the various isotropy groups of the points in M reveals the stratified structure of M/G which, as we have seen in Section 3, has consequences for the optimal synthesis.

The *third step* is to obtain the boundaries of the reachable sets in M/G, that is, the projections of the boundaries of the reachable sets $\mathcal{R}(t)$ in M. In order to do this, if A_a is an element in the Cartan subalgebra $\mathcal{A} \subseteq \mathcal{P}$, we write a representative of a geodesic as (cf. (31)),

$$X(t) := e^{-\bar{A}_k t} e^{(\bar{A}_k + A_a)t}, \tag{38}$$

with $\bar{A}_k \in \mathcal{K}$ and $A_a \in \mathcal{A}$ and $\|A_a\| = L$. By fixing t and varying $\bar{A}_k \in \mathcal{K}$ and $A_a \in \mathcal{A}$, we obtain an hyper-surface in M/G, part of which is the boundary of the reachable set at time t. The determination of the sets in $\mathcal{K}$ and $\mathcal{A}$ which is mapped to this boundary is an analysis problem to be considered on a case by case basis, which is obviously simpler in low dimensional cases, and requires help from computer simulations in higher dimensional cases.

The *fourth step* is to find the first t such that $\pi(\mathcal{R}(t))$ contains $\pi(X_f)$. At this value of t, there are matrices A_k and A_a such that $[e^{-\bar{A}_k t} e^{(\bar{A}_k + A_a)t}] = \pi(X_f)$.

Finally, the *fifth step* is to find $K \in G$ such that

$$K e^{-\bar{A}_k t} e^{(\bar{A}_k + A_a)t} K^{-1} = X_f. \tag{39}$$

This gives the correct pair (A_k, A_p) to be used in the optimal control (27): $A_k = K\bar{A}_k K^{-1}$, $A_p := K\bar{A}_a K^{-1}$. From the last two steps, it follows that the problem is therefore effectively divided in two. Restricting ourselves to the orbit space we first find an optimal control to drive the state of the system to the desired orbit. Then, in the fifth step, we move inside the orbit to find exactly the final condition we desire.

The treatment of the optimal synthesis on $SO(3)$ in the following section gives an example of application of this method.

5 Optimal Synthesis for the $K-P$ Problem on $SO(3)$

A basis of the Lie algebra of skew-symmetric real 3×3 matrices, $so(3)$, is given by:

$$p_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad p_2 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad k := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

---

[13]Typical cases in the literature look at a Lie group M where the conjugation action on M is given by M itself and not by a subgroup G of M as in our case.

We consider the $K - P$ Cartan decomposition of $so(3)$ where $\mathcal{K} = \mathrm{span}\{k\}$, and $\mathcal{P} = \mathrm{span}\{p_1, \; p_2\}$. There are two possible maximal groups of symmetries with Lie algebra $\mathcal{K}$. A *maximal connected* Lie group, $\mathbf{K}^+$, which is the connected component containing the identity and consists of matrices of the form

$$K^+(r) := \begin{pmatrix} \cos(r) & \sin(r) & 0 \\ -\sin(r) & \cos(r) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{40}$$

So here the upper-left 2×2 block is in $SO(2)$. A *maximal not connected* Lie group, $\mathbf{K}^+ \cup \mathbf{K}^-$, is given by the matrices which are either of the previous type or of the type

$$K^-(r) := \begin{pmatrix} \cos(r) & -\sin(r) & 0 \\ -\sin(r) & -\cos(r) & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{41}$$

Therefore, in Eq. 41, the upper-left 2×2 block is in $O(2)$, with determinant equal to -1. We shall consider this second case, that is, $G = \mathbf{K}^+ \cup \mathbf{K}^-$. Remark 5.7 discusses what would change had we chosen $G = \mathbf{K}^+$.

5.1 Structure of the Orbit Spaces $SO(3)/G$

Following the second step of the procedure described in the previous section, we now describe the structure of $M/G = SO(3)/(\mathbf{K}^+ \cup \mathbf{K}^-)$ and its isotropy strata. We use the Euler decomposition of $SO(3)$ from which it follows that any matrix $X \in SO(3)$ can be written as $X = K^+(r_1)H(s)K^+(r_2)$, with $K^+(r_i)$ of the type (40), and $H(s) := e^{p_1 s}$, for some real s. Since $K^+(r) \subset G$, $[X] = [H(s)K^+(r_2)\left(K^+(r_1)\right)^T] = [H(s)K^+(r_2 - r_1)]$. So, we can always choose as representatives of the orbits matrices of the type:

$$\begin{aligned} H(s)K^+(r) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(s) & \sin(s) \\ 0 & -\sin(s) & \cos(s) \end{pmatrix} \begin{pmatrix} \cos(r) & \sin(r) & 0 \\ -\sin(r) & \cos(r) & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos(r) & \sin(r) & 0 \\ -\sin(r)\cos(s) & \cos(r)\cos(s) & \sin(s) \\ \sin(r)\sin(s) & -\cos(r)\sin(s) & \cos(s) \end{pmatrix} \end{aligned} \tag{42}$$

with $s, \; r \in [0, 2\pi)$. Moreover, we have,

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} H(s)K^+(r) \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(r) & \sin(r) & 0 \\ -\sin(r)\cos(s) & \cos(r)\cos(s) & -\sin(s) \\ -\sin(r)\sin(s) & \cos(r)\sin(s) & \cos(s) \end{pmatrix},$$

which changes the sign of $\sin(s)$ as compared with Eq. 42. Thus, we can assume $\sin(s) \geq 0$, so $s \in [0, \pi]$. Furthermore, we have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} H(s)K^+(r) \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos(r) & -\sin(r) & 0 \\ \sin(r)\cos(s) & \cos(r)\cos(s) & \sin(s) \\ -\sin(r)\sin(s) & -\cos(r)\sin(s) & \cos(s) \end{pmatrix},$$

so we can also assume $r \in [0, \pi]$. It follows that each equivalence class has an element of the form Eq. 42, with $r, s \in [0, \pi]$. By equating two matrices of the form Eq. 42 for different values of the pairs (r, s), one can see that such a correspondence is one to one unless $s = \pi$. In this case, all the matrices $H(\pi)K^+(r)$ (which give the set $\mathbf{K}^-$) are equivalent. So if $s \in [0, \pi)$ and $r \in [0, \pi]$, each $H(s)K^+(r)$ represents a unique orbit, while if $s = \pi$, since

they are all equivalent, the choice of r is irrelevant. We can therefore represent $SO(3)/G$ as the upper part of a disc of radius π, where if ρ and θ are the polar coordinate, we have $\rho \in [0, \pi]$ with $\rho = \pi - s$, and $\theta \in [0, \pi]$ with $\theta = r$ (see Fig. 1).

Remark 5.1 If $[X_1] = [X_2]$, then $(X_1)_{3,3} = (X_2)_{3,3}$, and also the trace is preserved. So, from any element X of a given equivalence class, we can compute the two parameters $s, r \in [0, \pi]$ of Eq. 42, by setting:

$$s := \arccos(X)_{3,3}, \quad \begin{array}{ll} r = \arccos\left(\frac{(X)_{1,1}+(X)_{2,2}}{1+X_{3,3}}\right) & \text{if } (X)_{3,3} \neq -1, \\ r = 0 & \text{if } (X)_{3,3} = -1 \end{array} \tag{43}$$

From these values, we have also the two values of $\rho = \pi - s$ and $\theta = r$. So there is a one to one, onto, readily computable correspondence between points in the half disc in Fig. 1 and orbits in $SO(3)/G$.

The point $\rho = \pi$ and $\theta = 0$ (B in Figure 1) represents the identity matrix, while the point $\rho = \pi$ and $\theta = \pi$ (A in Fig. 1) gives the matrix:

$$J := \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{44}$$

Both these matrices are fixed points for the action of G, so they are the only matrices in their orbit, and their isotropy group is the entire group G.

The points with $\rho = \pi$ and $\theta \in (0, \pi)$ give the matrices in $\mathbf{K}^+$, except for the identity $\mathbf{1}$ and the matrix J defined in Eq. 44. The matrices in $\mathbf{K}^+$ commute, and it holds that $K^-(v)K^+(r)(K^-(v))^T = K^+(-r)$; thus, the orbits of these elements contain two matrices, and we took as representative the one with $\sin(r) > 0$. Their isotropy group is $\mathbf{K}^+$.
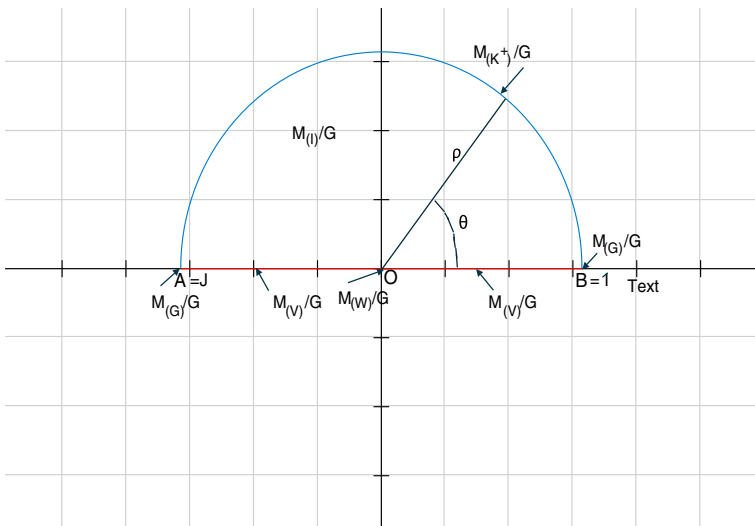

Fig. 1 The quotient space $SO(3)/G$

The origin, i.e., the point with $\rho = 0$ and $\theta = r$ arbitrary, corresponds to the matrices:

$$\begin{pmatrix} \cos(r) & \sin(r) & 0 \\ \sin(r) & -\cos(r) & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{45}$$

These matrices are all equivalent, and their isotropy groups are all conjugated to:

$$W = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, J, \mathbf{1} \right\} \tag{46}$$

which is the isotropy group of the matrix with $r = 0$.

The matrices with $\theta = \pi$ and $\rho \in (0, \pi)$ are the classes of the *symmetric matrices* in $SO(3)$. It can be seen that their isotropy group is conjugated to the one given by

$$V = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \mathbf{1} \right\} \tag{47}$$

The matrices with $\theta = 0$ and $\rho \in (0, \pi)$ correspond to matrices in $SO(3)$ of the type:

$$\begin{pmatrix} a & b & c \\ b & d & f \\ -c & -f & g \end{pmatrix}, \tag{48}$$

Their isotropy group is, again, conjugated to V, as in the symmetric case.

The matrices which are in the interior of the half disc have a trivial isotropy group, i.e., composed of only the identity matrix. This is the regular part of $SO(3)/G$ while the boundary of the half disc corresponds to the singular part.

Summarizing, the isotropy types of $SO(3)$ are given by $(\{\mathbf{1}\})$, (V), (W) in Eqs. 47 and 46, $(\mathbf{K}^+)$, and $(\mathbf{K}^+ \cup \mathbf{K}^-)$, with the partial ordering

$$(\{\mathbf{1}\}) \le (V) \le (W) \le (\mathbf{K}^+ \cup \mathbf{K}^-),$$

and

$$(\{\mathbf{1}\}) \le (\mathbf{K}^+) \le (\mathbf{K}^+ \cup \mathbf{K}^-).$$

$M_{(\mathbf{K}^+ \cup \mathbf{K}^-)}$ is composed by the matrices $\mathbf{1}$ and J, $M_{(W)}$ are the matrices in Eq. 45, $M_{(V)}$ are matrices which are either symmetric or of the form Eq. 48, $M_{(\mathbf{K}^+)}$ are the matrices in $\mathbf{K}^+$ except for $\mathbf{1}$ and J, $M_{(\{\mathbf{1}\})}$ are all the remaining matrices. The corresponding strata on the orbit space (half disc) are indicated in Fig. 1.

5.2 Cut Locus and Critical Locus

We shall now apply the results given in the previous two sections to determine the cut locus $\mathcal{CL}(SO(3))$ and the critical locus $\mathcal{CR}(SO(3))$. The cut locus was also described in [7] using a different method. Following what suggested in Remark 3.7, we analyze the singular points, first.

Proposition 5.2 *All the matrices that correspond to $\rho = \pi$ and $\theta \in (0, \pi]$ (these are all the matrices in $\mathbf{K}^+$ except the $\mathbf{1}$) are in $\mathcal{CL}(SO(3))$, and so also in the $\mathcal{CR}(SO(3))$ (cf. Proposition 2.1).*

Proof Fix a matrix $X_f \in \mathbf{K}^+$ and let $A_p = \alpha p_1 + \beta p_2$, be the matrix giving the minimizing geodesic that appear in Eq. 28 for X_f. If this matrix is not in $\mathcal{CL}(SO(3))$, then, using Proposition 4.2, it must hold:

$$[A_p, K^+] = 0,$$

for all $K^+ \in \mathbf{K}^+$, since $\mathbf{K}^+$ is contained in the isotropy group (indeed $\mathbf{K}^+$ *is the isotropy* group for all values of $\theta \in (0, \pi)$, while for $\theta = \pi$ the isotropy group is all G). The previous equality holds for all K^+ if and only if $A_p = 0$, which is not possible since $\|A_p\| = 1$. So X is in the cut locus and also in the critical locus. □

The next proposition proves that all the symmetric matrices (which correspond to the segment $O - A$ in Fig. 1) are in the cut locus.

Proposition 5.3 *The matrices corresponding to $\rho = 0$ and to $\rho \in (0, \pi)$ and $\theta = \pi$ (these are the matrices which correspond to the origin and to the segment (A, O) in the Fig. 1) are in $\mathcal{CL}(SO(3))$, and so also in $\mathcal{CR}(SO(3))$.*

Proof Fix a symmetric matrix X_f. Its isotropy group is conjugated either to W in Eq. 46 (if $\rho = 0$) or to V in Eq. 47. By continuity, the geodesic from $\mathbf{1}$ to X_f must contain matrices whose isotropy group is different from the one of X_f, so by using Corollary 3.6 we get that X_f lies in the cut locus. □

Now we will prove that all the remaining matrices, i.e., the ones corresponding to the open segment (OB) and the regular part (the interior of the disc) are neither on the $\mathcal{CL}(SO(3))$ nor in the critical locus $\mathcal{CR}(SO(3))$.

We know, that the geodesic are analytic curves given by Eq. 28. Here, we may choose as $\mathcal{A} = \text{span } \{p_1\}$, thus the geodesic are given by:[14]

$$[X(t)] = [e^{-\alpha kt} e^{(\alpha k + p_1)t}] = \tag{49}$$

$$\left[\begin{pmatrix} \frac{1+C_1\alpha^2}{1+\alpha^2}\cos(\alpha t) + C_2\alpha\sin(\alpha t) & C_1\sin(\alpha t) - C_2\alpha\cos(\alpha t) & C_3\cos(\alpha t) - C_2\sin(\alpha t) \\ -\frac{1+C_1\alpha^2}{1+\alpha^2}\sin(\alpha t) + C_2\alpha\cos(\alpha t) & C_1\cos(\alpha t) + C_2\alpha\sin(\alpha t) & -C_2\cos(\alpha t) - C_3\sin(\alpha t) \\ C_3 & C_2 & \frac{C_1+\alpha^2}{1+\alpha^2} \end{pmatrix} \right],$$

where

$$C_1 = \cos\left(\sqrt{(1+\alpha^2)}t\right), \quad C_2 = \frac{\sin\left(\sqrt{(1+\alpha^2)}t\right)}{\sqrt{1+\alpha^2}}, \quad C_3 = \frac{\alpha\left(1 - \cos\left(\sqrt{(1+\alpha^2)}t\right)\right)}{1+\alpha^2}.$$

The next proposition gives the optimal time to reach the matrices with $\rho = 0$, i.e., the ones corresponding to the origin of the half disc as in Eq. 45.

Proposition 5.4 *The optimal geodesic to reach any X_f such that $[X_f] = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right]$ must have the parameter α of Eq. 49 equal to 0, and the minimum time to reach X_f is π.*

---

[14]Here, we use the calculation of [7] section 3.2.1.

Proof Since the conjugation by elements of G does not change the 3,3 element, letting T the minimum time to reach X_f, we must have (see Eq. 49):

$$\frac{\cos\left(\sqrt{(1+\alpha^2)}T\right)+\alpha^2}{1+\alpha^2} = -1.$$

The previous equality can hold if and only if $\alpha = 0$. Moreover, we must have $\cos(T) = -1$. Thus, the minimum time T is equal to π. $\square$

The next proposition proves that the matrices in the singular part which correspond to the segment $O - B$ in Fig. 1 are neither on the cut locus nor on the critical locus. In particular, this implies that the projection of the geodesics reaching these matrices lies all in the segment, since each point of these trajectories has to have the same isotropy group.

Proposition 5.5 *Fix the matrix X_f that corresponds to $\theta = 0$ and $\rho = \pi - s$, with $s \in (0, \pi)$ as in Eq. 48. Then, this matrix is not on the cut locus nor on the critical locus, and the minimum time T to reach X_f from $\mathbf{1}$ is $T = s$.*

Proof Fix a matrix X_f that corresponds to $\theta = 0$ and $\rho = \pi - s$, with $s \in (0, \pi)$, i.e., such that

$$[X_f] = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(s) & \sin(s) \\ 0 & -\sin(s) & \cos(s) \end{pmatrix}\right].$$

These are matrices of the form Eq. 48. First, we prove that necessarily the geodesic reaching X_f must have $\alpha = 0$. Let $\gamma(t)$ be a geodesic with $\alpha = 0$. Then, by Proposition 5.4, its projection is optimal until $t = \pi$; thus, $\gamma(t)$ is optimal until $t = \pi$. Moreover, since its projection at time $t = s$ is equal to $H(s) := e^{P_1 s}$, we have $\gamma(s) = X_f$, and s is the minimum time, since the minimum time is the same for equivalent matrices (cf. Proposition 3.1). If there was another trajectory reaching optimally X_f, with $\alpha \neq 0$, and call this trajectory $\tilde{\gamma}(t)$, then the trajectory:

$$\eta(t) = \begin{cases} \tilde{\gamma}(t) & t \in [0, s) \\ \gamma(t) & t \in [s, \pi], \end{cases}$$

would also be an optimal trajectory to the origin, which contradicts the fact that all geodesics are analytic.

Assume now that X_f is on the cut locus. Then, there exist two optimal trajectories both with $\alpha = 0$, so $\gamma_i(t) = e^{A_p^i t}$ such that,

$$X_f = e^{A_p^1 s} = e^{A_p^2 s}.$$

Since every two Abelian subagebras in $\mathcal{P}$ are conjugate by an element of $\mathbf{K}^+$, there must exist a matrix $K^+ \in \mathbf{K}^+$ such that

$$\text{span}\{A_p^2\} = K^+ \text{span}\{A_p^1\}(K^+)^T,$$

However, since these spans are one dimensional, we must have

$$A_p^2 = \pm K^+ A_p^1 (K^+)^T. \tag{50}$$

Thus,

$$X_f = e^{A_p^2 s} = \begin{cases} K^+ e^{A_p^1 s}(K^+)^T & \text{if (50) is verified with } +1 \\ K^+ e^{-A_p^1 s}(K^+)^T & \text{if (50) is verified with } -1. \end{cases}$$

In the first case, we have that K^+ must be in the isotropy group of X_f. On the other hand, the isotropy group of X_f is conjugated to the group V of Eq. 47; thus, it contains two elements, one is the identity and the other must have -1 in the 3,3 position. Thus, necessarily since K^+ has $+1$ in the 3,3 position, we must have $K^+ = \mathbf{1}$, and so $A_p^2 = A_p^1$. In the second case, X_f is conjugate via an element of $\mathbf{K}^+$ to $X_f^{-1} = X_f^T$. Writing the third column of the relation $X_f K^+ = K^+ X_f^T$ using the formula (48) with $K^+ := \begin{pmatrix} K_1^+ & 0 \\ 0 & 1 \end{pmatrix}$ as

$$\begin{pmatrix} c \\ f \\ g \end{pmatrix} = \begin{pmatrix} -K_1^+ \begin{pmatrix} c \\ f \end{pmatrix} \\ g \end{pmatrix}, \tag{51}$$

we have that the 2×2 matrix $K_1^+ \in SO(2)$ has an eigenvalue in -1 (unless c and f are equal to zero which is to be excluded since $g \neq \pm 1$). Therefore,

$$K^+ = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using this in $A_p^2 = -K^+ A_p^1 (K^+)^T$ and the general expression for A_p^1, we find again $A_p^2 = A_p^1$.

Therefore, X_f is not on the cut locus. Moreover, since the projection of the trajectory is optimal until $t = \pi > s$, the matrix X_f is not on the critical locus either. □

5.3 The Optimal Synthesis

The last proposition has characterized the minimizing geodesics for points corresponding to the interval $O - B$ in Fig. 1, while Proposition 5.4 has given the minimizing geodesic and optimal time for points corresponding to the origin, i.e., matrices in $\mathbf{K}^-$, in Fig. 1. We now consider the geodesics leading to the remaining pieces of the singular part of $SO(3)/G$. Then, we put all things together to describe the full optimal synthesis.

The geodesic curves given in Eq. 49 depend on the parameter α which varies in $\mathbf{R}$. However, both parameters ρ and θ which characterize the points of the equivalence classes in the orbit space are even function of α (see Eq. 43), so in the analysis in the orbit space, we can restrict ourselves to values $\alpha \geq 0$.

The next proposition provides the optimal time to reach any matrix with $\rho = \pi$, i.e., all the matrices in $\mathbf{K}^+$.

Proposition 5.6 *Assume $X_f \in \mathbf{K}^+$, then $[X_f] = \{X_f, X_f^T\}$, and let $\theta \in (0, \pi]$ be the value of the parameter of Eq. 42, which together with $\rho = \pi$ gives the equivalence class $[X_f]$. Then, the minimum time T to reach X_f is given by*

$$T = \sqrt{\theta(4\pi - \theta)},$$

and the optimal value of the parameter α to reach $[X_f]$ is $\alpha = \frac{2\pi - \theta}{\sqrt{\theta(4\pi - \theta)}}$.

Proof First notice that necessarily $\alpha \neq 0$, since all the trajectories corresponding to $\alpha = 0$ have $\theta = 0$. Since the equivalence class of X_f consists of only two elements

(which coincide when $\theta = \pi$) and these elements have 0 in the 3,1 and 3,2 position, and 1 in the 3,3 position, for $t = T$ we must have in Eq. 49, $C_2 = C_3 = 0$ and $\frac{C_1 + \alpha^2}{1 + \alpha^2} = 1$, which implies:

$$C_2 = \sin\left(\sqrt{(1 + \alpha^2)}T\right) = 0 \text{ and } C_1 = \cos\left(\sqrt{(1 + \alpha^2)}T\right) = 1;$$

thus, we must have

$$\sqrt{(1 + \alpha^2)}T = 2m\pi, \tag{52}$$

for some $m \in \mathbb{N}$. Moreover, at time T, we have

$$[X(T)] = \left[\begin{pmatrix} \cos(\alpha T) & \sin(\alpha T) & 0 \\ -\sin(\alpha T) & \cos(\alpha T) & 0 \\ 0 & 0 & 1 \end{pmatrix}\right],$$

which implies

$$\cos(\alpha T) = \cos(\theta) \quad \Rightarrow \quad \alpha T = \pm\theta + 2p\pi, \tag{53}$$

for some $p \in \mathbb{N}$. We will treat the $\pm\theta$ sign separately.

Case +1 Assume that Eq. 53 holds with the $+1$ sign. Since $\sqrt{(1 + \alpha^2)}T > \alpha T$, we must have $p \leq m - 1$. From Eqs. 52 and 53, we have

$$\frac{2m\pi}{\sqrt{(1 + \alpha^2)}} = \frac{\theta + 2p\pi}{\alpha}.$$

The previous equality implies

$$\alpha = \frac{\theta + 2p\pi}{\sqrt{(4m^2\pi^2 - (\theta + 2p\pi)^2)}},$$

and consequently,

$$T_{m,p}^+ = \sqrt{(4m^2\pi^2 - (\theta + 2p\pi)^2)}.$$

The value of $T_{m,p}^+$, for each fixed m, is minimum when p is maximum, i.e. $p = m - 1$, and its minimum value is

$$T_{m,m-1}^+ = \sqrt{(2\pi - \theta)(4m\pi + \theta - 2\pi)},$$

which is minimum when $m = 1$ and we have $T_{1,0}^+ = \sqrt{(2\pi - \theta)(2\pi + \theta)}$.

Case −1 Assume that Eq. 53 holds with the -1 sign. Imposing again $\sqrt{(1 + \alpha^2)}T > \alpha T$, we now get $p \leq m$. From Eqs. 52 and 53 we have

$$\frac{2m\pi}{\sqrt{(1 + \alpha^2)}} = \frac{-\theta + 2p\pi}{\alpha}.$$

The previous equality implies

$$\alpha = \frac{-\theta + 2p\pi}{\sqrt{(4m^2\pi^2 - (-\theta + 2p\pi)^2)}},$$

and consequently,

$$T_{m,p}^- = \sqrt{(4m^2\pi^2 - (2p\pi - \theta)^2)}.$$

Again $T_{m,p}^-$, for each fixed m is minimum when p is maximum. Therefore, we now take $p = m$, and we get

$$T_{m,m}^- = \sqrt{\theta(4m\pi - \theta)},$$

which is minimum when $m = 1$ and we have $T_{1,1}^- = \sqrt{\theta(4\pi - \theta)}$.
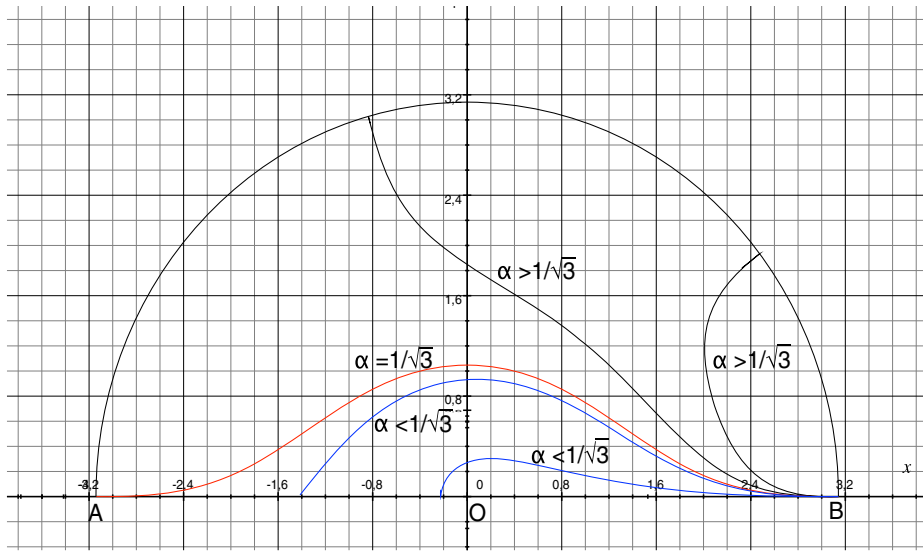
Fig. 2 Optimal trajectories

Since $\theta \leq \pi$, we have $T_{1,1}^- \leq T_{1,0}^+$, thus the minimum time is $T = \sqrt{\theta(4\pi - \theta)}$ with the corresponding $\alpha = \frac{2\pi - \theta}{\sqrt{\theta(4\pi - \theta)}}$. $\qquad\qquad\square$

From the previous proposition, since $\theta \in (0, \pi)$, we have that for $\alpha \geq \frac{1}{\sqrt{3}}$, all the geodesics are optimal until time $T = \frac{2\pi}{\sqrt{1+\alpha^2}}$, when they reach the boundary of the disc. It is clear that T is a decreasing function of α, with maximum equal to $\pi\sqrt{3}$, which corresponds to the trajectory reaching the matrix J. The trajectory corresponding to $\alpha = 0$ lies on the segment (O, B) and it is optimal until time $T = \pi$, when it reaches the origin. The
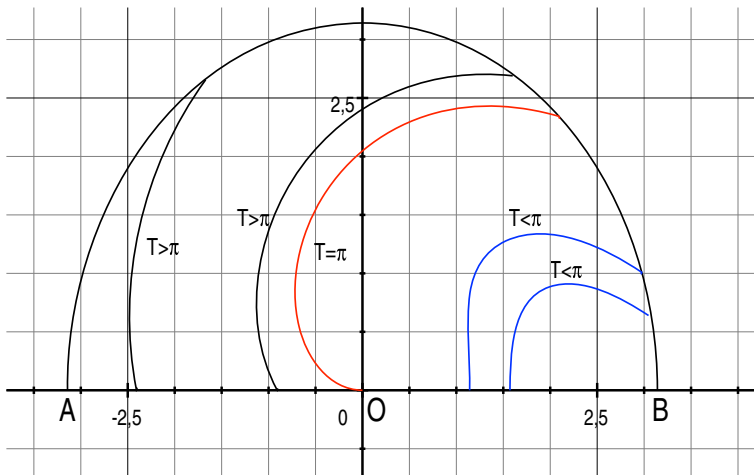

Fig. 3 Reachable sets

trajectories corresponding to $\alpha \in (0, \frac{1}{\sqrt{3}})$ are optimal until they reach the segment (A, O), which correspond to the symmetric matrices. We know from Proposition 5.3 that these matrices are on the cut locus. For a given α, the time T where the corresponding geodesic loses optimality can be numerally estimated, and it is always between π and $\sqrt{3}\pi$.

Thus, all elements are reached in time $T \leq \sqrt{3}\pi$. See Fig. 2 for the shape of the optimal trajectories, the red curve is the optimal curve with $\alpha = \frac{1}{\sqrt{3}}$, the black curves correspond to bigger values of α and loose optimality at the boundary of the circle, while the blue curves correspond to smaller values of α and loose optimality at the segment (A, O).

Figure 3 describes the optimal synthesis according to the third step of the procedure given in the previous section, that is, it gives the boundaries of the reachable sets at any time t. To draw these curves, for a given time T one finds the values of α such that the corresponding trajectory at time T lies on the boundary, and these are parametric curves with α as a parameter in the given interval. For $T < \pi$, the boundary is given varying α from 0, until the boundary of the circle is reached, for $T > \pi$, the parameter α has to be chosen from the values that correspond to the segment (A, O) until it again reaches the boundary of the circle. So the behavior changes at the curve in red corresponding to $T = \pi$.
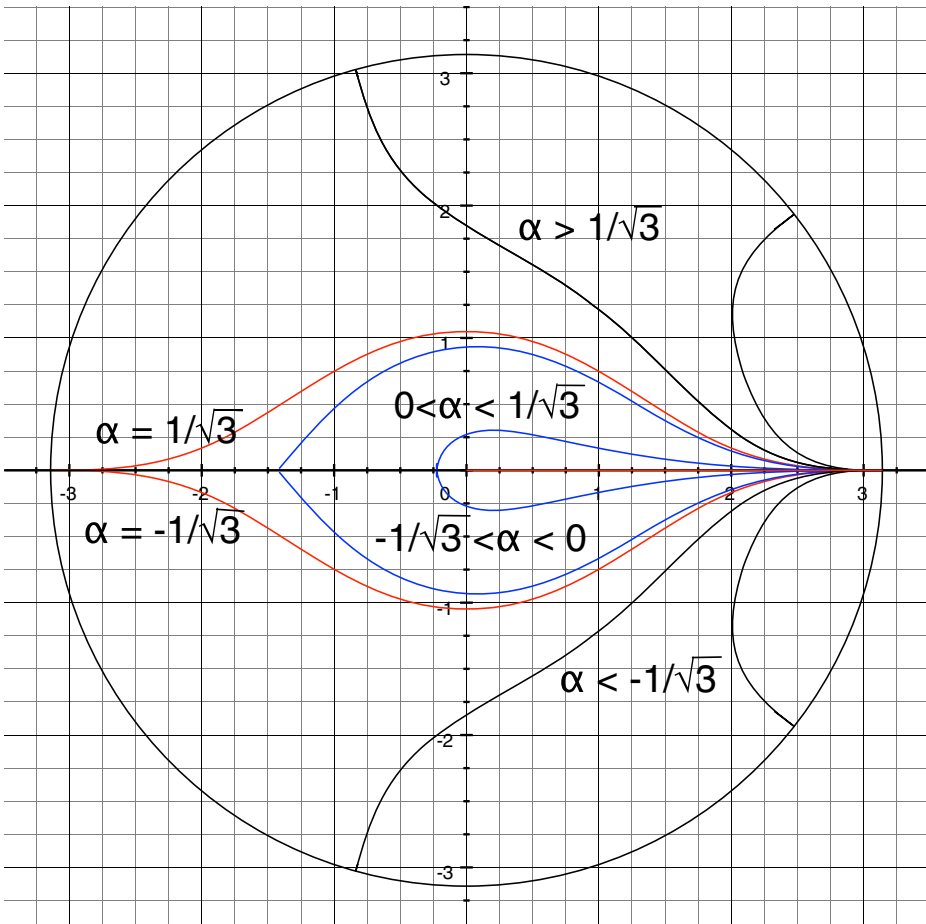

Fig. 4 Geodesics when $G = K^+$

Remark 5.7 To derive all the previous results, we have taken as symmetry group $G = K^+ \cup K^-$. We could have done a similar analysis, taken as a group of symmetries only the connected component containing the origin, i.e., $\widetilde{G} = K^+$. In this case, as representatives of equivalent classes, we could take again matrices of the type (42), but now, while $s \in [0, \pi]$, we may allow $r \in (-\pi, \pi]$. So the quotient space turns out to be the all disc of radius π, instead of only the upper part. Here, the boundary represents the matrices in K^+ that now are all fix points and the center are the matrices in K^-, which are again all equivalent. It is easy to see that this two sets give the singular part of $SO(3)/(K^+)$, while the interior of the disc is all in the regular part. The trajectories in the quotient space are given by the trajectories we have found previously and the one that are the symmetric with respect to the x-axis, this can be easily seen, since the two parameters s, r can be found using, as before, Remark 5.1, but while s is the same, for r we have two choices, the r given in Remark 5.1 and its opposite (see also Fig. 4).

References

1. Agrachev A, Barilari D, Boscain U. Introduction to Riemannian and sub-Riemannian geometry, Lecture Notes SISSA. Italy: Trieste; 2011.
2. Agrachev A, Sachkov Y. Control theory from the geometric viewpoint, Encyclopaedia of Mathematical Sciences, 87. Berlin: Springer; 2004.
3. Albertini F, D'Alessandro D. Minimum time optimal synthesis for two level quantum systems. J Math Phys. 2015;56:012106.
4. Albertini F, D'Alessandro D. Time optimal simultaneous control of two level quantum systems, submitted to Automatica.
5. Alekseevsky D, Kriegl A, Losik M, Michor P. The Riemannian geometry of orbit spaces. The metric, geodesics, and integrable systems. Publ Math Debrecen. 2003;62:247–276.
6. Boscain U, Chambrion T, Gauthier JP. On the K+P problem for a three-level quantum system: optiMality implies resonance. J Dyn Control Syst. 2002;8(4):547–572.
7. Boscain U, Rossi F. Invariant Carnot-Caratheodory metric on s^3, $SO(3)$ and $SL(2)$ and lens spaces. SIAM J Control Optim. 2008;47:1851–1878.
8. Bredon GE, Vol. 46. Introduction to compact transformation groups pure and applied mathematics. New York: Academic Press; 1972.
9. D'Alessandro D, Albertini F, Romano R. Exact algebraic conditions for indirect controllability of quantum systems. SIAM J Control Optim. 2015;53(3):1509–1542.
10. Echeverrìa-Enriquez A, Marìn-Solano J, Munõz Lecanda MC, Roman-Roy N. Geometric reduction in optimal control theory with symmetries. Rep Math Phys. 2003;52:89–113.
11. Bredon GE, Vol. 46. Introduction to compact transformation groups pure and applied mathematics. New York: Academic Press; 1972.
12. Filippov AF. On certain questions in the theory of optimal control. SIAM J on Control. 1962;1:78–84.
13. Grizzle J, Markus S. The structure of nonlinear control systems possessing symmetries. IEEE Trans Automat Control. 1985;30:248–258.
14. Grizzle J, Markus S. Optimal control of systems possessing symmetries. IEEE Trans Automat Control. 1984;29:1037–1040.
15. Ibort A, De la Peña TR, Salmoni R. Dirac structures and reduction of optimal control problems with symmetries, preprint; 2010.
16. Jacquet S. Regularity of the sub-Riemannian distance and cut locus, in nonlinear control in the year 2000. Lecture Notes Control Inf Sci. 2007;258:521–533.
17. Knapp A, Vol. 140. Lie groups beyond and introduction, Progress in Mathematics. Boston: Birkhäuser; 1996.

18. Koon WS, Marsden JE. The Hamiltonian and Lagrangian approaches to the dynamics of nonholonomic systems. Rep Math Phys. 1997;40:21–62.
19. Marsden JE, Ratiu TS. Introduction to mechanics and symmetry. New York: Springer; 1999.
20. Marsden JE, Weinstein A. Reduction of symplectic manifolds with symmetry. Rep Math Phys. 1974;5:121–130.
21. Martinez E. Reduction in optimal control theory. Rep Math Phys. 2004;53(1):79–90.
22. Meinrenken E. Group actions on manifolds (lecture notes) University of Toronto. 2003.
23. Michor P. Isometric actions of lie groups and invariants, Lecture Course at the University of Vienna. Unknown Month 1996.
24. Montgomery R. A Tour of sub-Riemannian geometries, their geodesics and applications, volume 91 of Mathematical Surveys and Monographs. RI: American Mathematical Society; 2002.
25. Monti R. The regularity problem for sub-Riemannian geodesics, in geometric control and sub-Riemannian geometry. Springer INdAM Series. In: Stefani G, Boscain U, Gauthier J-P, Sarychev A, and Sigalotti M, editors; 2014. p. 313–332.
26. Nijmeijer H, Van der Schaft A. Controlled invariance for nonlinear systems. IEEE Trans Automat Control. 1982;27:904–914.
27. Ohsawa T. Symmetry reduction of optimal control systems and principal connections. SIAM J Control Optim. 2013;51(1):96–120.