Introduction to Geometric Control Theory

Ugo Boscain, Dario Prandi, Mario Sigalotti

April 17, 2026

Contents

Notation

- The state space (a smooth differentiable manifold) is usually indicated with M. Points on M are usually indicated with q. Such q become x when a system of coordinates has been fixed.

- Curves on M are usually indicated with γ or with $q(\cdot)$. The presence of the "$(\cdot)$" is to avoid confusion between q (a point on the manifold) and $q(\cdot)$ (a curve).

- Vector fields are usually indicated with a capital letter as F, X or V. The flow at time t of a vector field F is indicated with the exponential notation e^{tF}.

- A control system is usually indicated as $\dot{q}(t) = F(q(t), u(t))$. Here $u(\cdot) : [0, \infty) \to U$ is the control. When we write simply u we usually mean a fixed value of the control $u \in U$.

Preface

This book presents material taught by the authors in graduate courses at several L3, M2 and doctoral courses (SISSA, Trieste; M2 Mathématiques de la Modélisation, Sorbonne Université; Master (M2) d'optimisation, Orsay; Polytechnique, Politecnico di Torino).

For further readings, see [ABB19; AS04; Arn06; BR05; BP04; BP21; BP07; Do 93; Jur97; Kha92; LM86; Lee13; SL12; Son98; Lib11].

1. Introduction

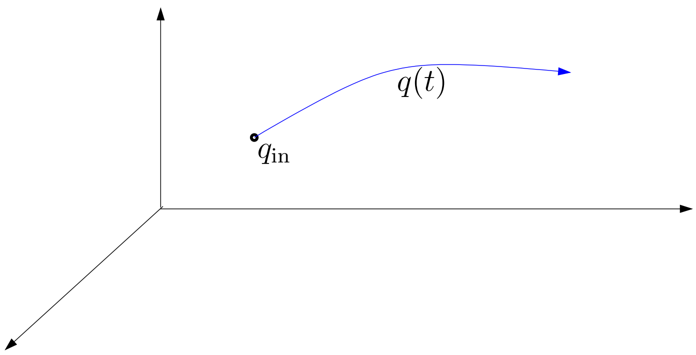
Consider a dynamical system on a smooth connected manifold M of dimension n.

$$\begin{cases} \dot{q}(t) = F(q(t)) \\ q(0) = q_{\text{in}} \end{cases} \tag{1.1}$$

The reader not used to the language of manifold can think of M as $\mathbb{R}^n$ or a smooth hypersurface embedded in $\mathbb{R}^{n+1}$. We are going to give the formal definition of manifold in Chapter 2. The function F is called a *vector field*, and when it is sufficiently regular (as, for instance, when F is smooth), the problem (1.1) admits a unique local solution. Such a solution is a smooth curve $q(\cdot) : I \to M$, where I is an open interval of $\mathbb{R}$ (here for simplicity we assume that $I = \mathbb{R}$). Hence, we can say that *the set of points that can be reached from* q_{in} *in positive time by solutions of* (1.1), *that is*

$$\{\bar{q} \in M \mid \exists\, T \geq 0 \ s.t. \ q(T) = \bar{q}\},$$

is a one dimensional object (actually, it is a one dimensional manifold).

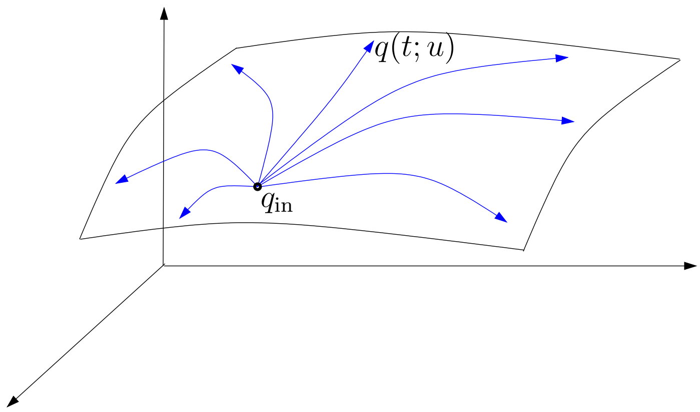


Assume now that F depends on an m-dimensional parameter $u = (u_1, \dots, u_m)$ belonging to a subset of U of $\mathbb{R}^m$. Here, we assume $m < n$, since this is the most common situation in practice and for clarity of presentation. Then, (1.1) becomes

$$\begin{cases} \dot{q}(t) = F(q(t), u), \quad u \in U \subset \mathbb{R}^m \\ q(0) = q_{\text{in}} \end{cases} \tag{1.2}$$

Since for every choice of u we have local existence and uniqueness of trajectories (that we denote by $q(t; u)$) we can expect that under suitable assumptions on U and on the dependence on u of F, the set of points that can be reached from q_{in} with solutions of (1.2), that is

$$\{\bar{q} \in M \mid \exists\, T \ \geq 0, \ u \in U \text{ s.t. } q(T; u) = \bar{q}\},$$



is an $m+1$ dimensional object.

We now let the parameter u depend on time. Then, (1.2) becomes

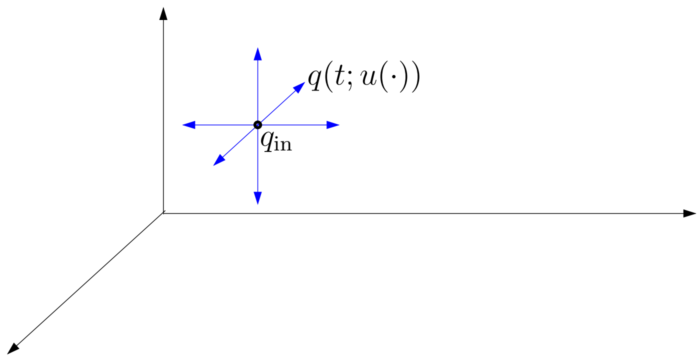
$$\begin{cases} \dot{q}(t) = F(q(t), u(t)), \quad u : [0,\infty) \to U \subset \mathbb{R}^m \\ q(0) = q_{\text{in}} \end{cases} \tag{1.3}$$

If we assume that $u(\cdot)$ has a reasonable dependence on time (e.g., it is locally bounded) then, for every fixed $u(\cdot)$ the problem (1.3) still has local existence and uniqueness of solutions. Such a solution (that we denote by $q(t, u(\cdot))$) is a locally Lipschitz curve. In this case, the set of points that can be reached from q_{in} by varying $u(\cdot)$ is

$$\mathcal{R}(q_{\text{in}}) := \{\bar{q} \in \mathbb{R}^n \mid \exists T \geq 0 \text{ and } u : [0,T] \to U, \text{ s.t. } q(T; u(\cdot)) = \bar{q}\}.$$

We can finally formulate the main question.

Q1 How large is the set $\mathcal{R}(q_{\text{in}})$? In particular, is it possible to reach every point of M?



Equation (1.3) is called a *control system* and question **Q1** is the main question to which the theory of *controllability*, developed in the first part of the book (Chapters 4 and 5), tries to answer. The function $u(\cdot)$ is called the *control*. The set $\mathcal{R}(q_{\text{in}})$ is called the *reachable set* from q_{in}.

The control usually models an external action on the dynamical system. For instance if F describes a mechanical system, $u(\cdot)$ could represent an external force. If F describes a quantum molecule $u(\cdot)$ could represent an external electric field.

Of course, since $u(\cdot)$ varies on a (typically infinite-dimensional) functional space, one can expect that the set of points that can be reached from q_{in} is quite large. However, as we will see this question is highly non-trivial. Actually, although we are allowed to let the control $u(\cdot)$ vary as a function of the time t, the equation $\dot{q}(t) = F(q(t), u(t))$ is a strong constraint on the *admissible* curves (i.e., locally Lipschitz curves for which there exists a control $u(\cdot)$ satisfying $\dot{q}(t) = F(q(t), u(t))$ for almost every t). Indeed, *such a constraint is acting on the velocity*: at almost every t we have that $\dot{q}(t)$ belongs to the set $\mathbf{F}(q(t))$ where $\mathbf{F}(q) := \{F(q, v) \mid v \in U\}$ is the set of admissible velocities at the point q, which is a subset of the T_qM, the tangent space at q of M. The constraint

$$\dot{q}(t) \in \mathbf{F}(q(t))$$

is called *non-holonomic* (in contrast with *holonomic constraints*, that are constraints on the position of the type $q(t) \in N$ where N is a subset of M). The case in which there are no constraints is the case in which $\mathbf{F}(q) = T_qM$ for every $q \in M$. In this case the control system is said to be a *trivial control system* and every trajectory in the chosen functional class (e.g. locally Lipschitz) is admissible. If for instance $M = \mathbb{R}^n$, a trivial constraint is obtained with $m = n$, $U = \mathbb{R}^n$, $F(q, u) = u$.

Once the controllability question has been settled by proving that it is possible to steer q_{in} to a point q_{fi}, then one could try to find the most efficient way of doing it. For instance, one could try to go from q_{in} to q_{fi} in the shortest possible time or by minimizing the energy fed to the system via the controls. Such type of problems are usually modeled as the problem of minimizing an integral cost. More precisely, one would like to answer to the following question.

Q2 Find a solution the following problem

$$\begin{cases} \dot{q}(t) = F(q(t), u(t)), & u : [0, \infty) \to U \subset \mathbb{R}^m \\ q(0) = q_{\text{in}} \\ q(T) = q_{\text{fi}} \\ \displaystyle\int_0^T L(q(t), u(t))\, dt \to \min \end{cases} \tag{1.4}$$

Here, L is a sufficiently regular function from $M \times \mathbb{R}^m$ to $\mathbb{R}$, and the final time $T > 0$ could be fixed or free. Notice that the problem of minimizing the time corresponds to the choice $L(q, u) = 1$ for all $q \in M$ and $u \in U$.

Problem (1.4), is referred to as an *optimal control problem*. The study of these problems is the main subject of the second part of this book, consisting of Chapters 8-13.

Observe that if $M = \mathbb{R}^n$ and the control system is trivial then problem (1.4) is a standard problem of calculus of variations:

$$\begin{cases} \dot{q}(t) = u(t), & u : [0,\infty) \to \mathbb{R}^n \\ q(0) = q_{\text{in}} \\ q(T) = q_{\text{fi}} \\ \displaystyle\int_0^T L(q(t), u(t))\, dt \to \min \end{cases} \tag{1.5}$$

More in general on a manifold, a problem of optimal control with a trivial control system (i.e., a problem of calculus of variations on a manifold) is

$$\begin{cases} \int_0^T L(q(t), \dot{q}(t))\, dt \to \min \\ q(0) = q_{\text{in}}, \\ q(T) = q_{\text{fi}} \end{cases} \tag{1.6}$$

A problem of optimal control can thus be seen as a problem of calculus of variations subject to non-holonomic constraints, since we have in addition the constraint $\dot{q}(t) = F(q(t), u(t))$.

It is interesting to notice that there are classical subjects of differential geometry that can be reduced to a problem of optimal control. For instance the problem of finding the geodesics in a Riemannian manifold can be formulated as an optimal control problem. An important generalization of Riemannian geometry (called sub-Riemannian geometry) is presented in Chapter 11.

There are many other questions that are treated in control theory that will not be treated in these lectures. For instance once that the controllability question has been solved and we know that the point q_{fi} is reachable from q_{in} one would like to have an explicit expression of a control function steering q_{in} to q_{fi}. This is usually called the *motion planning question*.

Another typical question that one would like to solve is the question of *stabilizability*. Consider the control system (1.3) and assume that there exists an equilibrium point, i.e., $(q_{\text{in}}, u_0) \in M \times U$ such that $F(q_{\text{in}}, u_0)$=0. Does it exist a function $K : M \to U$, with $K(q_{\text{in}}) = u_0$, such that the dynamical system

$$q(t) = F(q(t), K(q(t))), \tag{1.7}$$

is asymptotically stable at q_{in}? For this question one usually looks for a regular enough function K to avoid problems of existence and uniqueness of solutions of (1.7). The function K is called a *feedback control*.

We continue this introduction by providing two examples.

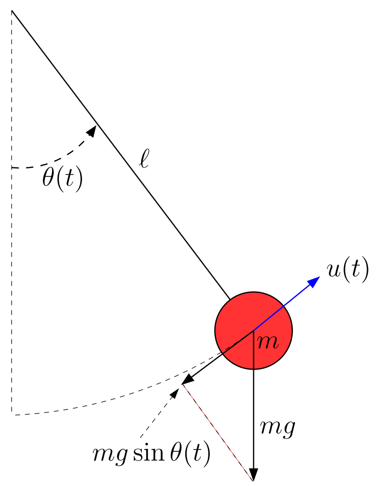
1.1. The controlled pendulum

Consider a stiff arm of length ℓ with attached a mass m, turning in a vertical plane around an horizontal axis. Assume moreover that the latter is equipped with an engine delivering a variable couple which we can choose *arbitrarily* at every time, but which is bounded. That is, our *control* acts on the mass via a (time-dependent) force

$$u(\cdot) : [0, \infty) \to [-u_{\max}, u_{\max}].$$

The geometric position of the system is described by an angle $\theta \in \mathbb{S}^1$, the unit circle. The dynamics of the system is obtained from the Newton's second law in the curvilinear coordinate $\ell\theta$. Letting g be the gravitational accelleration, this reads

$$m\ell\ddot{\theta}(t) = -mg \sin\theta(t) + u(t). \tag{1.8}$$



Although the constants intervening in this problem play a dominant role in practice, we shall suppose for simplicity that $m = \ell = g = 1$. The dynamics of the arm is thus reduced to

$$\ddot{\theta}(t) = -\sin\theta(t) + u(t). \tag{1.9}$$

To recast this second order differential equation as a first order system (as in (1.3)), we introduce a new variable $\omega = \dot{\theta}$. Equation (1.9) then becomes

$$\begin{cases} \dot{\theta} = \omega \\ \dot{\omega} = -\sin\theta + u, \end{cases} \tag{1.10}$$

The above equation has the form (1.3) with $M = \mathbb{S}^1 \times \mathbb{R}$, $q = (\theta, \omega)$ and

$$F(q, u) = \begin{pmatrix} \omega \\ -\sin\theta + u \end{pmatrix}.$$

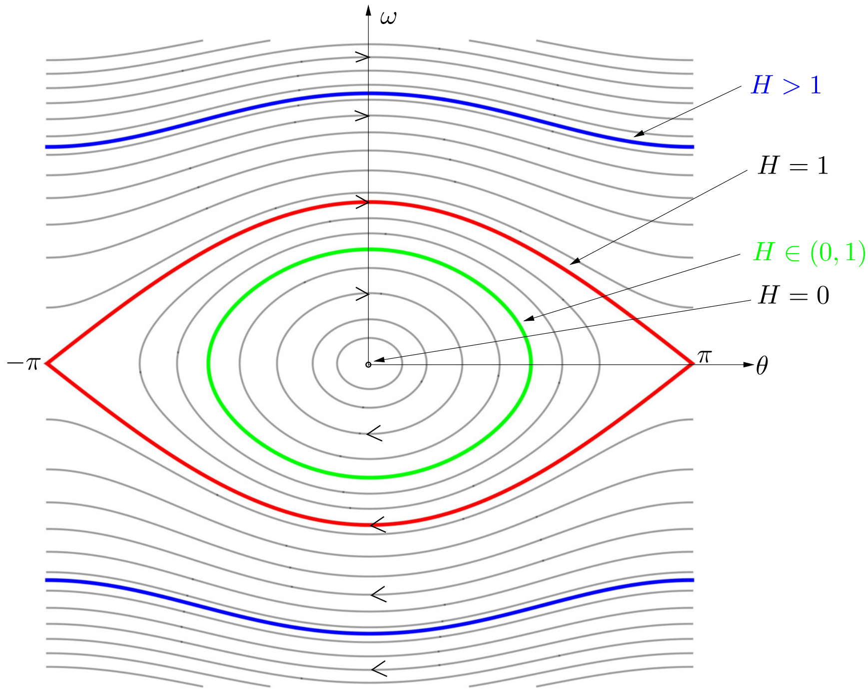


Figure 1.1.: Phase portrait of the pendulum with $u \equiv 0$.

The controllability question, when starting from the configuration of stable equilibrium $q_{\text{in}} = (0, 0)$, is then: "for every value $q_{\text{fi}} = (\bar{\theta}, \bar{\omega})$ is it possible to find a time T and control $u(\cdot) : [0, T] \to [-u_{\max}, u_{\max}]$ such that the solution of (**??**), that at time zero is in $q_{\text{in}} = (0, 0)$, reaches $q_{\text{fi}} = (\bar{\theta}, \bar{\omega})$ at time T?

The answer to this question is *YES*. However, as one can imagine, the proof is not trivial. It will be given in Chapter 6 (cf. Example 6.20) and it will use following two facts:

- when $u = 0$ almost every trajectory is periodic (this is a consequence of conservation of energy which implies that every trajectory should stay in a level sets of $H = \frac{1}{2}\omega^2 - \cos(\theta)$)

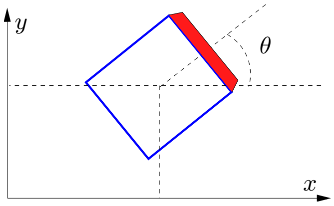
- evolving for time s with the control $u(t) \equiv u_{\max}$ and then for time s with the control $u(t) \equiv -u_{\max}$ is not the same than doing the evolution in opposite order. In this case, we say that the two vector fields $F(q, u_{\max})$ and $F(q, -u_{\max})$ *do not commute.*

An example of optimal control question on this system is: "find the control steering $(0, 0)$, to $(\bar{\theta}, \bar{\omega})$ in minimum time". Such a question is complicated. It can be solved with the techniques developed in Chapter 13, however the details will not be given in these notes.

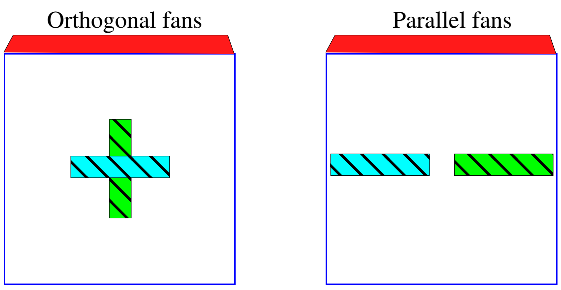
1.2. A vehicle sliding on ice, controlled by two fans

To illustrate why the presence of noncommuting dynamics is crucial in the controllabilitiy problem, we consider another example which is particularly intuitive.

Consider a vehicle sliding on ice. The state space of the vehicle is $\mathbb{R}^2 \times \mathbb{S}^1$: the coordinates (x, y) in $\mathbb{R}^2$ identify the center of mass of the vehicle; the coordinate $\theta \in \mathbb{S}^1$ identifies the orientation of the vehicle.



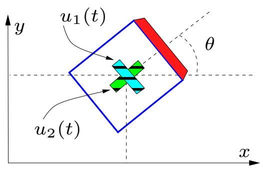
We control the vehicle by two fans (of which we can reverse the direction or rotation) that we will place on the vehicle in two different configurations as indicated in the following figure:



The reader should imagine that in the orthogonal configuration the two fans are paced one over the other. To simplify, we assume that the mass of the vehicle is negligible, so that using the fans allows to control directly the velocity of the vehicle rather than its acceleration. It is not hard to treat the more realistic case of the control on the acceleration.

Let us now write the two control system and study their controllability.

Orthogonal fans



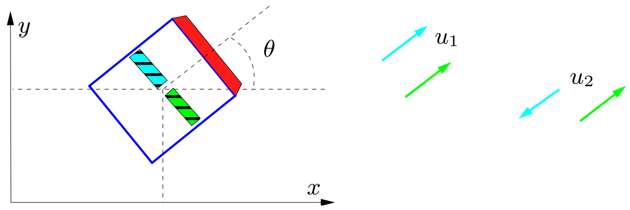
In this configuration, up to constants that we can fix to 1,the dynamics is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = u_1(t)F_1 + u_2(t)F_2, \tag{1.11}$$

$$\text{where} \qquad F_1(x,y,\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad \text{and} \quad F_2(x,y,\theta) = \begin{pmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{pmatrix}. \tag{1.12}$$

Since $\dot{\theta}$ is always zero, it is evident that with this configuration one cannot induce a rotation of the vehicle. Hence, from (x_0, y_0, θ_0) one can reach only (and actually all) configurations of the form $(\bar{x}, \bar{y}, \theta_0)$. That is, the reachable set is the plane $\mathbb{R}^2 \times \{\theta_0\}$ in $\mathbb{R}^2 \times \mathbb{S}^1$.

Parallel fans



In this configuration, up to constants that we can fix to 1, if $u_1(\cdot)$ is the control making the two fans rotate in the same direction and $u_2(\cdot)$ is the control making the two fans rotating in the opposite direction the dynamics is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = u_1(t)X_1 + u_2(t)F_2, \tag{1.13}$$

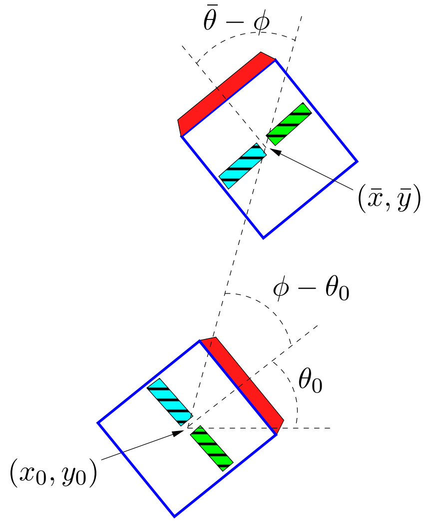
$$\text{where} \qquad X_1(x,y,\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad \text{and} \quad X_2(x,y,\theta) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{1.14}$$

In this case, starting from (x_0, y_0, θ_0) it is possible to reach any configuration $(\bar{x}, \bar{y}, \bar{\theta})$. Actually, if ϕ is the angle between the x-axis and the straight line connecting (x_0, y_0) to $(\bar{x}, \bar{y})$ it is possible to attain the desired final configuration by using a piecewise constant control composed of exactly 3 pieces (here by simplicity we assume $\theta_0 \leq \phi \leq \bar{\theta}$):

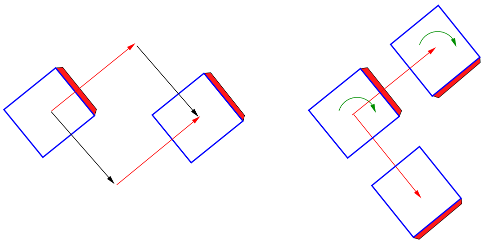
1. We let $u_1 = 0$ and $u_2 = 1$ for a time $t_1 = \phi - \theta_0$.

2. Then, $u_1 = 1$ and $u_2 = 0$ for a time t_2 that equals the Euclidean distance between (x_0, y_0) and $(\bar{x}, \bar{y})$.

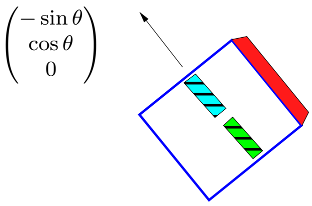
3. Finally, we let $u_1 = 0$ and $u_2 = 1$ for a time $t_3 = \bar{\theta} - \phi$.



What is the main difference between the vector fields for the orthogonal configuration (F_1 and F_2), and those for the parallel configuration (X_1 and X_2)? The point is that the two vector fields F_1 and F_2 generate two translations, which commute, while the vector fields X_1 and X_2 generate a translation and a rotation, which do not commute. It is exactly the non-commutativity of X_1 and X_2 that allows to approximate direc-



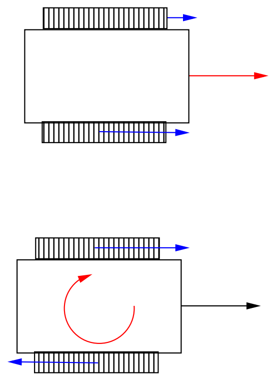
tions that are not directly admissible in the control system. In particular, the direction corresponding to a movement orthogonal to the wheels:



The concept of commutativity/non-commutativity of dynamics will be discussed in detail in Chapter 5.

Remark 1.1. The previous analysis is fairly independent on the set of controls. In particular, it is unchangend whether the controls are taken in the (relatively small) class of piecewise constant controls or in the (much larger) class of L^∞_{loc} (or even in the larger possible class, that is the class of L^1_{loc} controls). Actually, as it will be made clear in Chapter 5, the controllability analysis is not very sensitive to the regularity of the controls, and for this reason is usually done in the simple class of piecewise constant controls.

Notice that the problem of controlling in the "parallel configuration" is exactly the problem of controlling a tracked vehicle as the one in the following picture.



2. A primer in differential geometry

In this chapter we introduce the mathematical framework that we will use in the next chapters. We are going to develop the theory of control on smooth[1] differentiable manifolds since in many applications the state space is not $\mathbb{R}^n$, (think for instance to the example of the vehicle sliding on the ice of the previous chapter where the state space is $\mathbb{R}^2 \times S^1$).

This chapter should not be used for a systematic study of differential geometry. To this purpose there are many excellent books as for instance [Do 93; Arn06; Lee13]. The reader that is familiar with the concepts of smooth manifold, vector fields (including time-dependent vector fields), flows, and Lie brackets, can jump directly to the next chapter.

2.1. Smooth differentiable manifolds

Differentiable manifolds are objects that locally "look like" $\mathbb{R}^n$, but possibly not globally. Besides $\mathbb{R}^n$, examples that one meets very often are

- in dimension $n = 1$: the circle $\mathbb{S}^1$, an open interval $(-\varepsilon, \varepsilon)$;
- in dimension $n = 2$: any smooth surface in $\mathbb{R}^3$ (e.g., the sphere $\mathbb{S}^2$, or the torus $\mathbb{T}^2$), the Klein bottle (see Figure 2.1)

$$\{(x, y) \in [-\pi, \pi] \times [-\pi, \pi] \mid (x, -\pi) \sim (x, \pi) \text{ and } (-\pi, y) \sim (\pi, -y)\}.$$

- in dimension $n = 3$: the sphere $\mathbb{S}^3$.

Differentiable manifolds can be seen as the generalization of regular surface embedded in $\mathbb{R}^3$. In fact, any manifold of dimension n can be embedded in $\mathbb{R}^N$, with sufficiently big N, as the zero locus of a function $F \in \mathcal{C}^\infty(\mathbb{R}^N, \mathbb{R}^{N-n})$ whose differential is of maximal rank. For instance, a circle S^1 can be embedded in $\mathbb{R}^3$ as the set of zeros of the function $F(x, y, z) = (x^2 + y^2 - 1, z)$, see Figure 2.2.

However, it is much more instructive to think to differentiable manifolds of dimension n not as "regular" subset of $\mathbb{R}^N$, but as geometric objects that can be locally treated as $\mathbb{R}^n$ via a system of *charts*. We refer to Figure 2.3.

Definition 2.1. A smooth differentiable manifold of dimension n is a pair $(M, \mathcal{U})$ where

- M is a set;

[1]In this book, smooth will always refer to $\mathcal{C}^\infty$ regularity, albeit many of our statements apply to $\mathcal{C}^k$ for, say, $k \geq 2$.

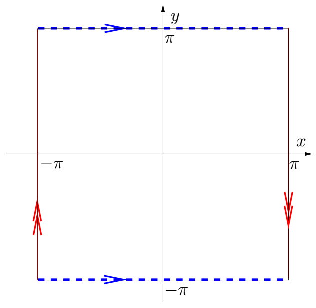


Figure 2.1.: The Klein bottle

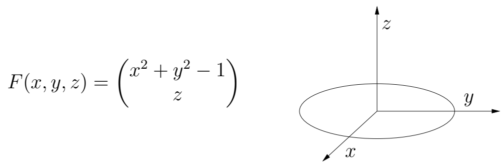


Figure 2.2.: A circle embedded in $\mathbb{R}^3$.

- $\mathcal{U} = \{(U_a, \varphi_a)\}_{a\in A}$ is a collection of *charts*, i.e., for any $a \in A$ we have that $U_a \subset M$ and $\varphi_a : U_a \to \mathbb{R}^n$ is an injective map such that $\varphi(U_a)$ is open.

Moreover, the family $\mathcal{U}$ is a *maximal atlas*, meaning that

1. $M = \bigcup_{a\in A} U_a$;

2. For any $a, b \in A$ such that $U_a \cap U_b = W \neq \emptyset$ it holds that:

  2.1. $\varphi_a(W)$ and $\varphi_b(W)$ are open sets in $\mathbb{R}^n$;

  2.2. the map $\varphi_a \circ \varphi_b^{-1} : \varphi_b(W) \to \varphi_a(W)$ is $\mathcal{C}^\infty$.

3. The family $\mathcal{U} = \{(U_a, \varphi_a)\}_{a\in A}$ is maximal[2] with respect to the previous conditions.

It is standard to refer to the couple $(M, \mathcal{U})$ as "the manifold M".

Condition **2.1**, allows to endow the set M with a natural topology by declaring $W \subset M$ to be open if and only $\varphi_a(W \cap U_a) \subset \mathbb{R}^n$ to be open for any $a \in A$. Henceforth, to avoid pathological situations, we will assume the following conditions on this natural topology:

- It is Hausdorff. That is, different points have different neighborhoods.
- It is second countable. That is, there exists an atlas $\mathcal{U}_0 = \{(U_a, \varphi_a)\}_{a\in A_0}$ which satisfies conditions **1.** and **2.**, and such that A_0 is at most countable.

Charts allows to consider *local coordinates* $(x^1, \ldots, x^n)$ on M. That is, when considering local properties it is always possible to fix a chart (φ_a, U_a) and use it to identify the region of interest with a subset of $\varphi_a(U_a) \subset \mathbb{R}^n$, which is then parametrized by the coordinates $(x^1, \ldots, x^n)$.

Remark 2.2. Since a differentiable manifold can be locally treated as $\mathbb{R}^n$, one could be tempted to assume that by fixing coordinates it is possible to essentially work as in $\mathbb{R}^n$. However, this is not true since, for example: (i) many properties are not stable under change of coordinates (for instance a curve $\gamma : [0,1] \to M$ can be a straight line in one system of coordinates and an arc of circle in another); (ii) the global topology can play

[2] This maximality condition is not always present in the definition of a manifold. However, any non-maximal atlas can be extended to a maximal one via Zorn Lemma.

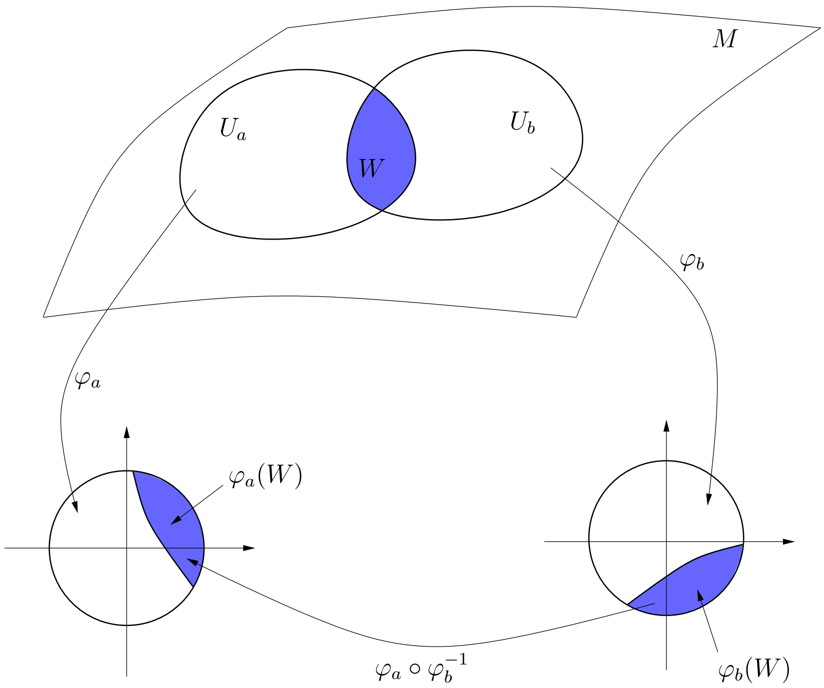


Figure 2.3.: Definition of a smooth manifold

an important role (e.g., continuous functions $f : \mathbb{S}^2 \to \mathbb{R}$ are forcibly bounded, which is clearly not the case for continuous functions $f : \mathbb{R}^2 \to \mathbb{R}$). See, for instance, [Arn06].

In general, a property on a manifold is said to be *intrinsic* if it does not depends on the choice of the system of coordinates. For instance, the property of a curve (se Definition 2.4) to be closed is an intrinsic property, while to have zero second order coefficient in his Taylor polynomial is not.

2.2. Curves, tangent vectors

Using charts, and thanks to assumption **2.2**, we can induce on M a differential structure.

Definition 2.3. Let M and N be two differentiable manifolds with maximal atlases $\mathcal{U} = \{(U_a, \varphi_a)\}_{a\in A}$ and $\mathcal{V} = \{(V_b, \psi_b)\}_{b\in B}$. A function $f : M \to N$ is smooth (resp., of class $\mathcal{C}^k$ or Lipschitz) if this is true *chart-wise*. That is, if $\psi_b \circ f \circ \varphi_a^{-1} : U_a \to \psi_b(V_b)$ is smooth (resp., of class $\mathcal{C}^k$ or Lipschitz) for any $a \in A$ and $b \in B$ such that $f(U_a) \cap V_a \neq \varnothing$.

The set of $\mathcal{C}^k$ functions (resp. Lipschitz) from M to N is denoted by $\mathcal{C}^k(M, N)$ (resp. $\mathrm{Lip}(M, N)$). When $N = \mathbb{R}$ we will omit it. In particular, the set of smooth functions $f : M \to \mathbb{R}$ is denoted by $\mathcal{C}^\infty(M)$.

A particular case of the above which is particularly interesting, is the case of curves.

Definition 2.4. A *curve* on a differentiable manifold M is a continuous map $\gamma : I \to M$, where $I \subset \mathbb{R}$ is an interval.

Once a system of coordinates is fixed, a curve γ is thus represented by an n-tuple $(\gamma_1, \dots, \gamma_n) : I \to \mathbb{R}^n$, where each $\gamma_i : I \to \mathbb{R}$ is a continuous function.

Given a smooth curve $\gamma : (-\varepsilon, \varepsilon) \to M$, we are interested in considering its tangent vector, say $\gamma'(0)$. It is then natural to try to define $\gamma'(0)$ by fixing a coordinate system around $\gamma(0)$ and computing[3]

$$\gamma'(0) = \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t}. \tag{2.1}$$

Unfortunately, since M has no *intrinsic*vector space structure, the result of the sum appearing in the above depends on the system of coordinates we chose, and hence so does the limit.

Remark 2.5. This problem arises even if the manifold is $M = \mathbb{R}^2$. Indeed, in this case, in order to obtain an intrinsic notion of derivative we need to associate to $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^2$ an object $\gamma'(0)$ that is *invariant* under the action of any chart, i.e., any local diffeomorphism $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$. That is, we need that $\gamma'(0) = (\varphi \circ \gamma)'(0)$. It is evident that the standard notion of derivative does not fullfill this property.

In order to obtain an intrinsic definition, we are then going to define the tangent vector to a smooth curve as a functional on real-valued smooth functions $f \in \mathcal{C}^\infty(M)$.

Definition 2.6. The tangent vector at 0 of a smooth curve $\gamma : (-\varepsilon, \varepsilon) \to M$ is the operator $\dot\gamma(0)$ defined by

$$\dot\gamma(0) f := \frac{d}{dt}\bigg|_{t=0} (f \circ \gamma). \tag{2.2}$$

This is a derivation at $\gamma(0)$, in the sense that it is a linear functional on $\mathcal{C}^\infty(M)$ that satisfies the Leibniz rule:

$$\dot\gamma(0)[fg] = [\dot\gamma(0) f]\, g(\gamma(0)) + f(\gamma(0))\, [\dot\gamma(0) g], \qquad \forall f, g \in \mathcal{C}^\infty(M). \tag{2.3}$$

This definition is intrinsic since $f \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}$ has "forgotten" the manifold M. We stress that it is possible to show that derivations at a point $q \in M$ are always obtained as derivatives of a curve $\gamma : (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = q$.

In coordinates $x = (x^1, \dots, x^n)$ such that $\gamma(t)$ is represented by $(\gamma_1(t), \dots, \gamma_n(t))$, we can compute (2.2) via the chain rule. This yields

$$\frac{d}{dt}\bigg|_{t=0} (f \circ \gamma) = \sum_{i=1}^n \partial_i f(x) \dot\gamma_i(0) = \Big(\sum_{i=1}^n \dot\gamma_i(0) \partial_i \Big|_x \Big) f. \tag{2.4}$$

Here, we denoted by $\partial_i|_x$ the derivation $\partial_i|_x f = \partial_i f(x)$. We have thus shown that, in coordinates, the derivation $\dot\gamma(0) : \mathcal{C}^\infty(M) \to \mathbb{R}$ is represented by

$$\dot\gamma(0) = \sum_{i=1}^n \dot\gamma_i(0) \partial_i \Big|_x \tag{2.5}$$

In particular, $\{\partial_1|_x, \dots, \partial_n|_x\}$ is a basis for the derivations at x.

[3] More precisely, if (U, φ) is a chart such that $\gamma(0) \in U$ the expression $\gamma'(0)$ in coordinates stands for $(\gamma \circ \varphi)'(0)$. To lighten the notation, when there is no ambiguity we identify geometric object with their expression in coordinates.

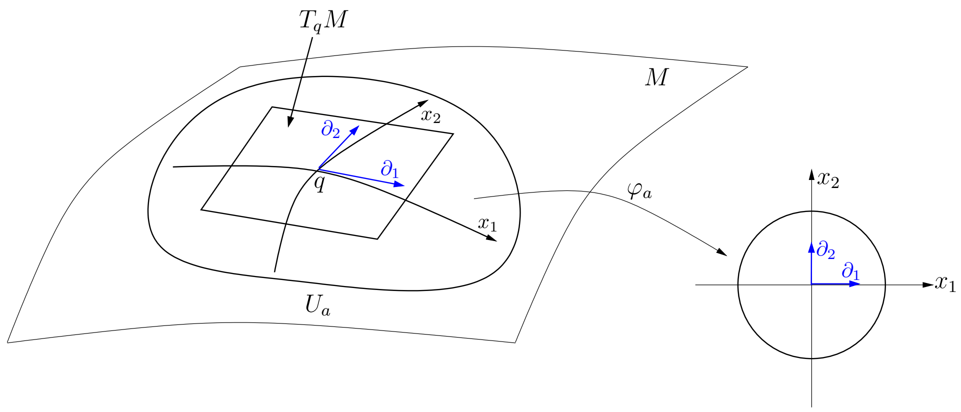


Figure 2.4.: Coordinates in the tangent space.

Definition 2.7. A *tangent vector at a point* $q \in M$ is the tangent vector at $t = 0$ to some smooth curve $\gamma : (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = q$. The set of all tangent vectors at q is the *tangent space of* M *at* q, and is denoted by T_qM.

By (2.5), a tangent vector $v \in T_qM$ is represented in coordinates by

$$\sum_{i=1}^{n} v_i \partial_i \Big|_x. \tag{2.6}$$

That is, v is represented by a $n-$tuple $(v_1, \ldots, v_n) \in \mathbb{R}^n$. Notice that $\partial_i\big|_x$ is the coordinate representation of the tangent vector at q of the "coordinate curve"

$$x_i(\cdot) : t \mapsto (0, \ldots, \underset{\substack{\uparrow \\ \text{position } i}}{t}, \ldots, 0)$$

It then follows that T_qM has the structure of an n-dimensional vector space, whose canonical basis in the system of coordinates $(x_1, \ldots, x_n)$ is given by $(\partial_1, \ldots, \partial_n)\big|_x$ (see Figure 2.4).

Definition 2.8. The tangent bundle to M, denoted by TM, is the disjoint union of the tangent spaces to M. That is,

$$TM = \bigsqcup_{q \in M} T_qM.$$

It can be proved that TM has the structure of a smooth differentiable manifold[4] of dimension $2n$, whose charts are induced by those of M. Indeed, as shown above, fixing a

[4]More is actually true: as its name suggests, TM carries the structure a vector bundle over M, meaning that there exists a continous surjection $\pi : TM \to M$ that associates to each tangent vector its *base* (i.e., the point to which it is tangent) and such that the *fiber* $\pi^{-1}(q)$ is an n-dimensional vector space. The structure satisfy additional compatibility conditions, guaranteeing that the locally TM looks like $U \times \mathbb{R}^n$ where $U \subset M$, and π looks like a the projection $\pi(x, v) = x$ for $x \in M$.

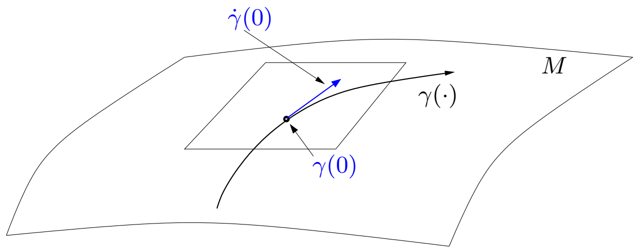


Figure 2.5.: Tangent vector to a curve.

chart (U, φ) on M fixes also a representation of the tangent vectors at each T_qM, $q \in U$. We stress that, although it can happen that $TM = M \times \mathbb{R}^n$ (as is the case if $M = \mathbb{R}^n$, $M = \mathbb{S}^1$, or $M = \mathbb{T}^2$), this is not true in general.

Remark 2.9. This point of view (i.e., tangent vectors as derivations) is not in contrast with the intuitive representation of a tangent vector on a surface embedded in $\mathbb{R}^3$ as an "arrow" on the corresponding tangent plane (see Figure (2.5)). As operator on functions such arrow represents the direction along which we compute the directional derivative. If our manifold is $\mathbb{R}^n$, v is a tangent vector at a point q and $f \in \mathcal{C}^\infty(\mathbb{R}^n)$ then

$$vf = \frac{d}{dt}\bigg|_{t=0} f(q + tv) = v \cdot \operatorname{grad} f(q).$$

Here, the "dot" represents the standard scalar product in $\mathbb{R}^n$ and "grad" is the standard gradient, i.e., $\operatorname{grad} f(q) = (\partial_{x_1} f(q), \dots, \partial_{x_n} f(q))$.

We conclude this section by generalizing the concept of differential of a map to the case of functions between manifolds.

Definition 2.10. Let $\psi : M \to N$ be a smooth function between differentiable manifolds. The tangent map of ψ at $q_0 \in M$ is

$$d\psi : TM \to TN, \tag{2.7}$$

acting at each point q_0 as $d_{q_0}\psi : T_{q_0}M \to T_{\psi(q_0)}N$. It is defined by

$$d_{q_0}\psi(v)f = \frac{d}{dt}\bigg|_{t=0} (f \circ \psi \circ \gamma), \qquad \forall f \in \mathcal{C}^\infty(N). \tag{2.8}$$

Here, $\gamma : (-\varepsilon, \varepsilon) \to M$ is any smooth curve such that $\gamma(0) = q_0$ and $\dot\gamma(0) = v$. The tangent map of ψ is also denoted by ψ_*.

2.3. Vector fields

A vector field is a smooth assignment of a tangent vector to each point of the manifold. More precisely, we have the following definition.

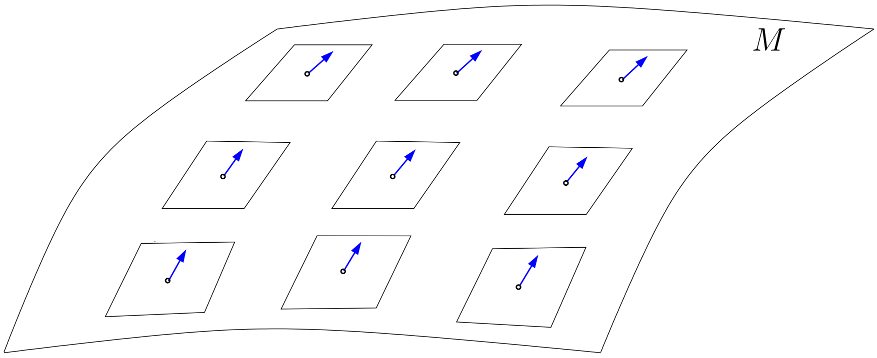


Figure 2.6.: A vector field.

Definition 2.11. A vector field on a manifold M is a smooth[5] function $X : M \to TM$ such that $X(q) \in T_qM$ for any $q \in M$. The set of vector fields on M is denoted by $\mathrm{Vec}(M)$, and is naturally endowed with a vector space structure.

Since for any $q \in M$ the tangent vector $X(q)$ is a derivation on $\mathcal{C}^\infty(M)$, it follows that a vector field X can be seen as a first order differential operator[6] on $\mathcal{C}^\infty(M)$ defined by

$$Xf(q) = X(q)f, \qquad \forall f \in \mathcal{C}^\infty(M). \tag{2.9}$$

One can then check that the smoothness of X is equivalent to the fact that $Xf \in \mathcal{C}^\infty(M)$ for any $f \in \mathcal{C}^\infty(M)$.

In coordinates, a vector field X is represented by an expression as (2.6) where coefficients v_i are replaced by smooth functions $X_i(x)$. Namely, letting ∂_{x_i} be the coordinate representation of the tangent vector at q of the coordinate curve $x_i(\cdot)$, we have that X is represented by

$$X(x) = \sum_{i=1}^n X_i(x)\partial_i\big|_x = \sum_{i=1}^n X_i(x)\partial_{x_i}. \tag{2.10}$$

It is easy to check that this is the local expression of a general first order differential operator on $\mathcal{C}^\infty(M)$. Hence, as per derivations and tangent vectors at a point, we can identify $\mathrm{Vec}(M)$ with the set of all first order differential operators on $\mathcal{C}^\infty(M)$.

To a vector field X is associated a dynamical system

$$\dot{q}(t) = X(q(t)). \tag{2.11}$$

A solution $q(\cdot)$ of (2.11) is called an *integral curve* of X. We have the following (see, e.g., [Arn06]).

[5]Here, smoothness is to be intended in the sense of functions between manifolds (i.e., the map $q \mapsto f(q)$ is smooth in local coordinates on M and TM).

[6]An operator $P : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M)$ is a differential operator if it is a linear map such that for any $f \in \mathcal{C}^\infty(M)$, the function Pf is also in $\mathcal{C}^\infty(M)$. It is of first order if it satisfies the Leibniz rule.

Theorem 2.12 (Cauchy-Lipschitz Theorem)**.** *Let X be a vector field on M and $q_0 \in M$. Then, the Cauchy problem*

$$\begin{cases} \dot{q}(t) = X(q(t)) \\ q(0) = q_0 \end{cases} \tag{2.12}$$

has local existence and uniqueness of solutions. More precisely, there exists $\varepsilon > 0$ such that there is a unique solution $q(\cdot; q_0) : (-\varepsilon, \varepsilon) \to M$ of the above problem. Moreover, the map $(t, q_0) \mapsto q(t; q_0)$ is smooth around $(0, q_0)$.

Remark 2.13. In (2.12) we have fixed the initial time at 0. There is no loss of generality in this choice. Actually, if $\tilde{q}(t; q_0)$ is solution of $\dot{\tilde{q}}(t) = X(\tilde{q}(t))$ with $\tilde{q}(t_0) = q_0$, we have $\tilde{q}(t; q_0) = q(t - t_0; q_0)$. This is due to the fact that the vector field X is *autonomous*, meaning that it does not depend explicitly on time.

Definition 2.14. The *flow* at time $t \in \mathbb{R}$ of the vector field X is the map

$$\phi_t : M \ni q_0 \mapsto q(t; q_0) \in M,$$

whenever the right hand side is well defined.

Notice that $\phi_0 : M \to M$ is the identity map. However, it may happen that the map ϕ_t is not defined on the whole manifold. This is the case, for instance, for $M = \mathbb{R}$ and the vector field $X(x) = x^2$. Indeed, the solution to $\dot{x} = x^2$ with $x(0) = x_0$ is

$$\phi_t(x_0) = \frac{x_0}{1 - t\, x_0}, \qquad t < \frac{1}{x_0}.$$

Hence, for $t > 0$, $\phi_t(x_0)$ is defined only for $x_0 < 1/t$. In particular, ϕ_t is never defined on the whole manifold $M = \mathbb{R}$ if $t > 0$.

This observation motivates the following definition.

Definition 2.15. A vector field X on M is *complete* if ϕ_t is defined on the whole manifold M for any $t \in \mathbb{R}$.

Notice that even constant vector fields are possibly non complete as for example a constant vector fields on the unit disc in $\mathbb{R}^2$.

Examples of complete vector fields are:

- any vector field on a compact manifold (e.g., $\mathbb{S}^1$, $\mathbb{S}^2$, $\mathbb{T}^2$);
- any vector field on $\mathbb{R}^n$ with sublinear growth (i.e., such that $\|X(x)\| \leq c(1 + \|x\|)$, where $c > 0$ is a constant and $\|\cdot\|$ is a norm on $\mathbb{R}^n$).

We then have the following.

Theorem 2.16 (see for instance [Arn06])**.** *If X is complete then ϕ_t is defined on the whole M and it is a diffeomorphism.*

It is immediate to verify the following properties of the flow of a complete vector field (for every $t, t_1, t_2 \in \mathbb{R}$):

$$\phi_0 = \mathrm{Id}, \tag{2.13}$$
$$\phi_{t_2} \circ \phi_{t_1} = \phi_{t_1+t_2}, \tag{2.14}$$
$$(\phi_t)^{-1} = \phi_{-t}. \tag{2.15}$$

That is, $\{\phi_t\}_{t\in\mathbb{R}}$ is a one parameter subgroup of the group of diffeomorphisms of M. Moreover, by construction we have[7]

$$\frac{\partial \phi_t}{\partial t} = X \circ \phi_t. \tag{2.16}$$

Due to the fact that the properties (2.13)–(2.16) remind those of the exponential function, usually the flow of a vector field X is denoted by e^{tX}. One should always remember that this is just a matter of notation (we cannot compute the exponential of a vector). However, for linear vector fields on $\mathbb{R}^n$, i.e. $X(x) = Ax$ where $A \in \mathbb{R}^{n\times n}$, $x \in \mathbb{R}^n$ we have

$$e^{tX}(x) = e^{tA}x.$$

Here, on the left hand side we have the flow of a vector field applied to the point x, and on the right hand side we have the standard matrix exponential. Notice that, in general and even for linear vector fields on $\mathbb{R}^n$, the flow of the sum of two vector fields is not the composition of the flows. That is,

$$e^{tX_1} \circ e^{tX_2} \neq e^{t(X_1+X_2)}.$$

A very important property of vector fields is the fact that, in a neighborhood of a point in which they are not zero, they can be rectified (see, e.g., [Arn06]).

Theorem 2.17 (Rectification of vector fields)**.** *Let X be a vector field on M such that $X(q) \neq 0$ for some $q \in M$. Then there exists a system of coordinates $(x_1, \dots, x^n)$ around q such that $X = \partial_{x_1}$.*

Proof. We prove the theorem in the case where M has dimension 2, the general case follows from similar arguments.

By continuity, we have that $X \neq 0$ on a neighborhood W of q. Up to reducing W, we can consider local coordinates x on W. Then, letting $X(x) = X_1(x)\partial_{x_1} + X_2(x)\partial_{x_2}$, the vector field $Y(x) = -X_2(x)\partial_{x_1} + X_1(x)\partial_{x_2}$ is well-defined on W, smooth, and such that $\{X(x), Y(x)\}$ are linearly independent. We now construct a different system of coordinates on W, still denoted by x, by letting

$$(x_1, x_2) \in U \subset \mathbb{R}^2 \mapsto e^{x_1 X} \circ e^{x_2 Y}(q) \in M. \tag{2.17}$$

By this definition, we see that U can be chosen to be a neighborhood of the origin, since $(0, 0) \mapsto q$.

[7]Here, we mean $\frac{\partial \phi_t(q)}{\partial t} = X(\phi_t(q))$ for every $q \in M$.

Let us show that (2.17) yields a system of coordinates (i.e., a bijection of U to some subset of W) by Theorem 2.12. Indeed, if both (x_1, x_2) and (x_1', x_2') map to the same point $p \in M$, then the curves $\gamma : t \mapsto e^{(x_1+t)X}(e^{x_2 Y}(q))$ and $\eta : t \mapsto e^{(x_1'+t)X}(e^{x_2' Y}(q))$ cross at $t = 0$. These curves both satisfy the same Cauchy problem:

$$\begin{cases} \dot{q}(t) = X(q(t)), \\ q(0) = p, \end{cases} \tag{2.18}$$

and thus must coincide. This implies that $(x_1, x_2) = (x_1', x_2')$.

To conclude the proof, we only need to show that $X(x) = \partial_{x_1}\big|_x$ in these coordinates. Let us compute

$$\partial_{x_1}\big|_x = \frac{d}{dt}\bigg|_{t=0}(x_1 + t, x_2) = \frac{d}{dt}\bigg|_{t=0} e^{(x_1+t)X} \circ e^{x_2 Y}(q) = X(e^{x_1 X} \circ e^{x_2 Y}(q)) = X(x).$$

This concludes the proof. □

2.3.1. Nonautonomous vector fields

In control theory one needs to work with vector fields that smoothly depends on a time dependent control function $u(\cdot)$. Namely, we consider dynamical systems of the form

$$\dot{q}(t) = F(q(t), u(t)) \tag{2.19}$$

where $U \subset \mathbb{R}^m$ is the set of admissible control values, $F : M \times \overline{U} \to TM$ is a smooth function such that $F(q, u) \in T_q M$ for any $q \in M$ and $u \in \overline{U}$, and $u(\cdot) : \mathbb{R} \to U$ is an L^∞_{loc} function. Once $u(\cdot)$ is fixed, if we define $X_t(q) := F(q, u(t))$ we obtain a dynamical system with a non-autonomous vector field

$$\dot{q}(t) = X_t(q(t)). \tag{2.20}$$

It is then necessary to extend the theory of vector fields introduced in the preceding section to non-autonomous vector fields that are regular w.r.t. q, but not necessarily w.r.t. t.

Definition 2.18. A *nonautonomous vector field* is a family of vector fields $\{X_t\}_{t \in \mathbb{R}}$ such that the map $X(q, t) = X_t(q)$ satisfies the following properties

(C1) For any $q \in M$, the map $t \mapsto X(q, t)$ is measurable;

(C2) For any $t \in \mathbb{R}$, the map $q \mapsto X(q, t)$ is smooth;

(C3) The map $(q, t) \mapsto X(q, t)$ is locally Lipschitz[8] in q uniformly w.r.t. t. That is, for any compact set $K \times I \subset M \times \mathbb{R}$ and any system of coordinates (U, φ), there exists $L > 0$ such that

$$\|X(x, t) - X(y, t)\| \leq L\|x - y\|, \qquad \forall x, y \in K \cap U, \quad \forall t \in I.$$

[8] This implies also that it is locally bounded.

One easily checks that the dynamical system (2.19) defines a non-autonomous vector field in the sense of Definition 2.18.

The local existence and uniqueness of integral curves of a nonautonomous vector field is guaranteed by the following theorem (see for instance [BP07]).

Theorem 2.19 (Carathéodory theorem)**.** *Assume that the nonautonomous vector field* $\{X_t\}_{t\in\mathbb{R}}$ *satisfies (C1)-(C3). Then the Cauchy problem*

$$\begin{cases} \dot{q}(t) = X_t(q(t)) & \textit{(for almost every } t\textit{)} \\ q(t_0) = q_0 \end{cases} \tag{2.21}$$

has local existence and uniqueness of solutions. Denoting the solution of (2.21) *by* $q(t;t_0,q_0)$*, the map* $(t,q_0)\mapsto q(t;t_0,q_0)$ *is locally Lipschitz with respect to* t *and smooth with respect to* q_0*.*

The flow of a non-autonomous vector fields is defined similarly to the autonomous case, but it depends on the initial time as well

$$\phi_{t_0,t} : M \ni q_0 \mapsto q(t;t_0,q_0)$$

The non-autonomous vector field X_t is said to be *complete*, if for all $t_0\in\mathbb{R}$ and $q_0\in M$ the solution $q(\cdot;t_0,q_0)$ to (2.21) is defined for every $t\in\mathbb{R}$.

For every $t_0,t\in\mathbb{R}$ the flow of a complete non-autonomous vector field is a diffeomorphism. However its dependence on t is only locally Lipschitz.

For the flow of a complete non-autonomous vector fields properties (2.13)–(2.16) become

$$\begin{aligned} \phi_{t_0,t_0} &= \mathrm{Id}, \\ \phi_{t_2,t_3}\circ\phi_{t_1,t_2} &= \phi_{t_1,t_3}, \\ (\phi_{t_1,t_2})^{-1} &= \phi_{t_2,t_1}, \\ \frac{\partial}{\partial t}\phi_{t_0,t} &= X_t\circ\phi_{t_0,t} \end{aligned}$$

Since the flow $\phi_{t_0,t}$ of a non-autonomous vector field depends both on t_0 and t, it is not convenient to use the exponential notation.

2.4. Lie bracket of vector fields

In this section we introduce a way to measure the non-commutativity of vector fields, which is a fundamental concept in control theory.

Definition 2.20. Let $X,Y\in\mathrm{Vec}(M)$ be two vector fields on M. The Lie bracket of X and Y is the vector field $[X,Y]\in\mathrm{Vec}(M)$ defined by

$$[X,Y]f = XYf - YXf, \qquad \forall f\in\mathcal{C}^\infty(M). \tag{2.22}$$

In coordinates, if X and Y are represented as in (2.10) by $X(x) = \sum_{i=1}^n X_i(x)\partial_{x_i}$ and $Y(x) = \sum_{i=1}^n Y_i(x)\partial_{x_i}$, then the Lie bracket $[X, Y]$ is represented by

$$\begin{aligned} [X,Y]f(x) &= \left[\sum_{i=1}^n X_i(x)\partial_{x_i}\right]\left[\sum_{j=1}^n Y_j(x)\partial_{x_j} f\right] - \left[\sum_{i=1}^n Y_i(x)\partial_{x_i}\right]\left[\sum_{j=1}^n X_j(x)\partial_{x_j} f\right] \\ &= \sum_{i,j=1}^n \left(X_i(x)\partial_{x_i} Y_j(x) - Y_i(x)\partial_{x_i} X_j(x)\right)\partial_{x_j} f(x). \end{aligned} \tag{2.23}$$

In matrix notation we can write the above as

$$[X,Y] = \frac{\partial Y}{\partial x}X - \frac{\partial X}{\partial x}Y, \tag{2.24}$$

In this formula vector fields are thought as column vectors and $\frac{\partial X}{\partial x}$ and $\frac{\partial Y}{\partial x}$ are the matrix of partial derivatives of the components of X and Y. More precisely,

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \qquad \frac{\partial X}{\partial x} = \begin{pmatrix} \partial_{x_1} X_1 & \dots & \partial_{x_n} X_1 \\ \vdots & \dots & \vdots \\ \partial_{x_1} X_n & \dots & \partial_{x_n} X_n \end{pmatrix}.$$

A similar expression holds for Y.

From expression (2.23) it follows that $[X, Y]$ is a first order operator on function and hence it is a vector field. Notice that $X \circ Y - Y \circ X$ is a vector field while $X \circ Y$ is not since this is a second order operator.

The Lie bracket of vector fields has the following properties:

- Bilinearity: $[a_1 X_1 + a_2 X_2, Y] = a_1[X_1, Y] + a_2[X_2, Y]$ for any $a_1, a_2 \in \mathbb{R}$ and $X_1, X_2, Y \in \mathrm{Vec}(M)$, and similarly for the second argument;
- Skew-symmetry: $[X, Y] = -[Y, X]$ for any $X, Y \in \mathrm{Vec}(M)$;
- Jacobi identity: $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ for any $X, Y, Z \in \mathrm{Vec}(M)$;

A vector space V endowed with an operation $V \times V \to V$ that is bilinear, antisymmetric and satisfying the Jacobi identity is said to be a *Lie algebra*. It follows that $(\mathrm{Vec}(M), [\cdot,\cdot])$ is a Lie algebra. This generalizes the notion of Lie algebra of matrices, which is the vector space $\mathbb{R}^{n\times n}$ endowed with the commutator $[A, B] = AB - BA$. We stress however that the Lie algebra of vector fields is infinite dimensional.

Remark 2.21. Due to the skew-symmetry of the Lie bracket, we have $[X, X] = 0$. Notice moreover that the value of $[X, Y]$ at a point q does not depend only on the values of X and Y at q, but also on their first order expansion at q. There is, however, the following special case

$$X(q) = Y(q) = 0 \implies [X, Y](q) = 0.$$

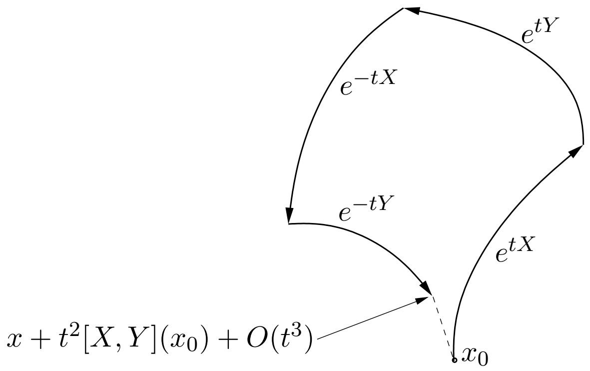


Figure 2.7.: Geometric meaning of the Lie bracket.

Example 2.22. Given two linear vector fields Ax and Bx, where $A, B \in \mathbb{R}^{n\times n}$, we have that $[Ax, Bx] = BAx - ABx = -[A,B]x$.

The following lemma clarifies the geometric meaning of the Lie bracket: $[X, Y]$ is actually a measure of the lack of commutation of the flows associated with X and Y. We refer to Figure 2.7.

Lemma 2.23. *Let $X, Y \in \mathrm{Vec}(M)$. Then, in coordinates*[9]*, we have that*

$$e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(x) = x + t^2[X,Y](x) + O(t^3), \qquad \text{as } t \to 0. \tag{2.25}$$

Proof. It is enough to compute for each flow the Taylor expansion at order 3. Observe that, for any vector field V, we have

$$e^{tV}(x) = x + tV(x) + \frac{t^2}{2}\frac{\partial V}{\partial x}(x)V(x) + O(t^3). \tag{2.26}$$

This implies in particular that

$$Y(e^{tX}(x)) = Y(x) + t^2\frac{\partial Y}{\partial x}(x)X(x) + O(t^3).$$

Using this, together with the expansion (2.26) applied to X and Y, we get

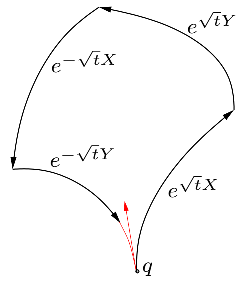
$$e^{tY}\circ e^{tX}(x) = x + t\big(X(x)+Y(x)\big) + \frac{t^2}{2}\frac{\partial X}{\partial x}(x)X(x) + t^2\frac{\partial Y}{\partial x}(x)X(x) + \frac{t^2}{2}\frac{\partial Y}{\partial x}(x)Y(x) + O(t^3).$$

Iterating this procedure yields

$$e^{-tX} \circ e^{tY} \circ e^{tX}(x) = x + tY(x) + t^2[X,Y](x) + \frac{t^2}{2}\frac{\partial Y}{\partial x}(x)Y(x) + O(t^3).$$

The result follows by using again (2.26) for $V = -Y$ on the above. □

[9]This expression is valid in local coordinates, since on a manifold we cannot sum points.



Remark 2.24. From Formula (2.25) it follows that an alternative definition of Lie brackets is

$$[X, Y](q) = \frac{d}{dt}\bigg|_{t=0} e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}(q).$$

In other words the Lie bracket $[X, Y](q)$ is the tangent vector at $t = 0$ to the curve $t \mapsto e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}(q)$.

In particular, we have the following immediate corollary.

Corollary 2.25. *Let N be a submanifold[10] of M, and assume that $X, Y \in \mathrm{Vec}(M)$ are tangent to N everywhere on N (that is, $X|_N, Y|_N$ belong to $\mathrm{Vec}(N)$). Then, $[X, Y]$ is also tangent to N and $[X, Y]|_N = [X|_N, Y|_N]$.*

Proof. It is enough to notice that, if X and Y are tangent to N, then the flow of X and Y preserves N. Hence, if $q \in N$, then the curve $t \mapsto e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}(q)$ is contained in N. By Remark 2.24, this implies that $[X, Y](q)$ is tangent to N at q. □

Remark 2.26. Lemma 2.23 and Remark 2.24 say that, by alternating between the dynamics of X and Y, it is possible to attain points that cannot be reached with the flow of linear combinations of X and Y. In particular, if $[X, Y](q) \notin \mathrm{Vec}(X(q), Y(q))$, then the curve $t \mapsto e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}(q)$ is tangent at $t = 0$ to a direction that is not spanned by $X(q)$ and $Y(q)$ (cf. the example in Section 1.2). This is the starting idea behind the conditions for controllability that we are going to study in the next chapter.

Example 2.27. Let us come back to the example of the vehicle driven by two orthogonal fans (see Section 1.2 and Figure 2.8). The dynamics of the system is given by

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ \theta \end{pmatrix} = u_1(t)X_1(x, y, \theta) + u_2(t)X_2(x, y, \theta), \quad \text{where } X_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The Lie bracket of X_1 and X_2 is given by

$$[X_1, X_2] = \frac{\partial X_2}{\partial x} X_1 - \frac{\partial X_1}{\partial x} X_2 = 0 - \begin{pmatrix} 0 & 0 & -\sin\theta \\ 0 & 0 & \cos\theta \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{pmatrix}.$$

In particular, $[X_1, X_2]$ is linearly independent from X_1 and X_2 at any point.

[10] That is, $N \subset M$ has the structure of a manifolds. It can be embedded or immersed, see Section 2.5.

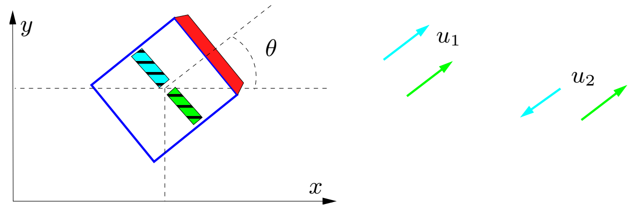


Figure 2.8.: Vehicle driven by two orthogonal fans.

Notice, however, that to be able to generate the Lie bracket $[X_1, X_2]$ one needs to be able to use, beside X_1 and X_2, also $-X_1$ and $-X_2$.

The following is an important consequence of Lemma 2.23.

Corollary 2.28. *Let $X, Y \in \mathrm{Vec}(M)$ be two vector fields. Then, their flows e^{tX} and e^{tY} commute for every $t \in \mathbb{R}$ if and only if their Lie bracket $[X, Y](q) = 0$ for every $q \in M$.*

Proof. Since Lemma 2.23 clearly implies that $[X, Y] \equiv 0$ if e^{tX} and e^{tY} commute, we are left to prove the converse.

To this aim, let us assume that $[X, Y](q) = 0$ for every $q \in M$. Let us fix $q \in M$ and notice that the curve $\gamma(t) := e^{tX} \circ e^{sY}(q)$ satisfies the Cauchy problem

$$\dot{\gamma}(t) = X(\gamma(t)), \qquad \gamma(0) = e^{sY}(q). \tag{2.27}$$

We are going to show that the same Cauchy problem is satisfied by $\eta(t) := e^{sY} \circ e^{tX}(q)$, which will prove that e^{tX} and e^{sY} commute for every $s, t \in \mathbb{R}$, and hence the statement.

The curve η satisfies the same initial condition at $t = 0$ as in (2.27). Moreover, denoting by $(e^{sY})_*$ the tangent map of e^{sY} (see Definition 2.10), we have

$$\dot{\eta}(t) = \frac{d}{dt} e^{sY} \circ e^{tX}(q) = (e^{sY})_* \frac{d}{dt} e^{tX}(q) = (e^{sY})_* X(e^{tX}(q)).$$

Hence, up to setting $\bar{q} = e^{sY} \circ e^{tX}(q)$, if we prove that

$$(e^{sY})_* X(e^{-sY}(\bar{q})) = X(\bar{q}) \tag{2.28}$$

we are done. Indeed, if (2.28) holds true, then

$$\dot{\eta}(t) = X(e^{sY} \circ e^{tX}(q)) = X(\eta(t)).$$

This concludes the proof that η satisfies (2.27).

We are then left to prove that (2.28) holds true at a general point $\bar{q} \in M$ and for every $s \in \mathbb{R}$ (under the assumption that $[X, Y] \equiv 0$). By looking at the Cauchy problem satisfied by $s \mapsto (e^{sY})_* X(e^{-sY}(\bar{q}))$, we reduce the problem to proving that

$$\frac{d}{ds} (e^{sY})_* X(e^{-sY}(\bar{q})) = 0, \qquad \forall s \in \mathbb{R}. \tag{2.29}$$

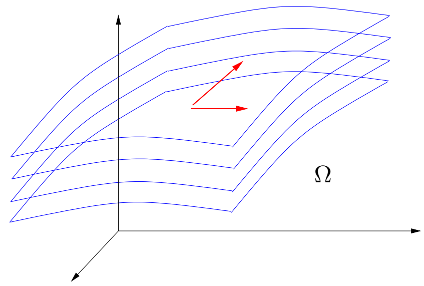


Figure 2.9.: Integral submanifolds of a family of vector fields.

We first show (2.29) at $s = 0$. Indeed, in coordinates,

$$\frac{d}{ds}(e^{sY})_* X(e^{-sY}(\bar{q}))\Big|_{s=0} = \frac{\partial Y}{\partial x}(\bar{q})X(\bar{q}) - \frac{\partial X}{\partial x}(\bar{q})Y(\bar{q}) = [X,Y](\bar{q}) = 0,$$

where we use that $(e^{sY})_* = \mathrm{Id} + s\frac{\partial Y}{\partial x}(\bar{q}) + O(s^2)$ for s small. We then conclude by fixing any $\bar{s} \in \mathbb{R}$, setting $\hat{q} = e^{-\bar{s}Y}(\bar{q})$, and noticing that

$$\frac{d}{ds}(e^{sY})_* X(e^{-sY}(\bar{q}))\Big|_{s=\bar{s}} = (e^{\bar{s}Y})_* \frac{d}{d\sigma}(e^{\sigma Y})_* X(e^{-\sigma Y}(\hat{q}))\Big|_{\sigma=0} = (e^{\sigma Y})_*[X,Y](\hat{q}) = 0.$$

This concludes the proof of the corollary. □

2.5. The Frobenius Theorem

In this section we present a fundamental result in differential geometry, which is the Frobenius Theorem. From the point of view of geometric control theory, this result highlights the importance of Lie brackets of vector fields, by showing that they encodes an obstruction to the controllability of a system.

Definition 2.29. The manifold N is an *immersed submanifold* of M if there exists an immersion map $i : N \to M$, i.e., a smooth map such that the differential i_* is injective for every $q \in N$.

Let $\mathcal{F} = \{X_1, \ldots, X_h\}$, $(h < n)$ be a family of vector fields defined on an open set $\Omega \subset M$, that are linearly independent at each point. Since $\dim \mathcal{F}_q = h$ is constant, we say that $\mathcal{F}$ has *constant rank*.

The following definition captures the notion of submanifold that is tangent to the family $\mathcal{F}$.

Definition 2.30. An immersed submanifold H of Ω of dimension h is said to be an *integral submanifold* of $\mathcal{F}$ if $T_qH = \mathrm{span}\{X_1(q), \ldots, X_h(q)\}$ for every $q \in H$. If for any $q \in \Omega$ there exists an integral submanifold of $\mathcal{F}$ passing through q, then $\mathcal{F}$ is said to be *completely integrable*.

Theorem 2.31 (Frobenius Theorem)**.** *The constant-rank family* $\mathcal{F} = \{X_1, \ldots, X_h\}$ *is completely integrable if and only if it is* involutive*, i.e.,*

$$[X_i, X_j](q) \in \operatorname{span}\{X_1(q), \ldots, X_h(q)\}, \qquad \forall i, j = 1, \ldots, h, \quad \forall q \in \Omega.$$

Proof. The fact that complete integrability implies involutivity is a direct consequence of Corollary 2.25. Indeed, if H is an integral submanifold of $\mathcal{F}$, then $[X_i, X_j]$ is tangent to H and hence it belongs to $\operatorname{span}\{X_1(q), \ldots, X_h(q)\}$ for every $q \in H$. Since this holds for every integral submanifold, we conclude that $\mathcal{F}$ is involutive.

The converse implication is more involved, and it will follow from a stronger result presented in Section 5.5, the Orbit Theorem. We refer to [Lee13, Theorem 19.12] for a self-contained proof of the Frobenius Theorem. □

Let us observe that this theorem says exactly that in order for a control system as (2.19) to be controllable in a neighborhood of a point q, it is necessary that the family $\mathcal{F} = \{F(\cdot, u) \mid u \in U\}$ is not involutive at q.

Part I.

Controllability

3. Controllability of nonlinear control systems: generalities

In this part, we study the controllability of the following non-linear system

$$\dot{q} = F(q, u(t)). \tag{3.1}$$

We assume that the state of the system q belongs to M, a smooth connected manifold of dimension n, $U \subset \mathbb{R}^m$ is the set of control values and $u(\cdot) : [0, \infty) \to U$ is the control. For the development of the theory it is not necessary to assume any structure on U. For example U could be a finite set of points, a polytope or the full $\mathbb{R}^m$.

We assume that F is a smooth function of its arguments and that $u(\cdot)$ belongs to a set $\mathcal{U}$ of sufficiently regular functions, so that equation (3.1) with initial condition $q(0) = q_{\text{in}} \in M$ has local existence and uniqueness of solutions. Such solutions is denoted by $q(\cdot; q_{\text{in}}, u)$. For instance, we can assume that $\mathcal{U} = L^1_{\text{loc}}(\mathbb{R}_+, U)$ or $\mathcal{U} = L^\infty_{\text{loc}}(\mathbb{R}_+, U)$. However, as it will be clear later, all conditions for controllability that we will get are actually sufficient conditions that are valid in the smaller class of piecewise constant controls.

Remark 3.1. To simplify some arguments and notations, in the following we assume that for every $u(\cdot)$ a solution $q(t; q_{\text{in}}, u)$ of (3.1) is defined for all $t \in [0, \infty)$. However, this hypothesis is not necessary for the validity of the theorems presented henceforth.

Definition 3.2. We consider the following *reachable* (or *attainable*) sets from $q_{\text{in}} \in M$:

- the *reachable set* from q_{in} at time $\tau \geq 0$ is

$$\mathcal{R}(\tau, q_{\text{in}}) = \{\bar{q} \in M \mid \exists\ u(\cdot) \in \mathcal{U} \text{ such that } q(\tau; q_{\text{in}}, u(\cdot)) = \bar{q}\};$$

- the *reachable set* from q_{in} within time $\tau \geq 0$ is

$$\mathcal{R}(\leq \tau, q_{\text{in}}) = \cup_{t\in[0,\tau]} \mathcal{R}(t, q_{\text{in}});$$

- the *reachable set* from q_{in} is

$$\mathcal{R}(q_{\text{in}}) = \cup_{t\in[0,+\infty)} \mathcal{R}(t, q_{\text{in}}).$$

Clearly, we have $\mathcal{R}(\tau, q_{\text{in}}) \subset \mathcal{R}(\leq \tau, q_{\text{in}}) \subset \mathcal{R}(q_{\text{in}})$ for every $\tau \geq 0$ and $q_{\text{in}} \in M$.

Given the control system (3.1), the purpose of the controllability theory is to characterize when these sets coincide with the entire state space.

Definition 3.3. System (3.1) is said to be

- *controllable* if for every $q_{\text{in}} \in M$ it holds that $\mathcal{R}(q_{\text{in}}) = M$;
- *controllable in time* $\tau > 0$ *at* q_{in} if it holds that $\mathcal{R}(\tau, q_{\text{in}}) = M$;
- *locally controllable* at $q_{\text{in}} \in M$ if q_{in} belongs to the interior of $\mathcal{R}(q_{\text{in}})$, i.e., $q_{\text{in}} \in \operatorname{int} \mathcal{R}(q_{\text{in}})$;
- *small-time controllable* if for every $q_{\text{in}} \in M$ and $\tau > 0$ we have $\mathcal{R}(\leq \tau, q_{\text{in}}) = M$;
- *small-time locally controllable at* q_{in} if it holds that $q_{\text{in}} \in \operatorname{int} \mathcal{R}(\leq \tau, q_{\text{in}})$ for every $\tau > 0$.

It is immediate to observe that small-time controllability implies controllability, and that small-time local controllability implies local controllability. Moreover, controllability implies local controllability at any point, and it turns out that the converse implication is also true (see Theorem 5.16).

We present here the following result, which will be useful in the sequel.

Lemma 3.4. *If the system is locally controllable, then* $\mathcal{R}(q_{\text{in}})$ *is open for every* $q_{\text{in}} \in M$*.*

Proof. If $\bar{q} \in \mathcal{R}(q_{\text{in}})$, then $\mathcal{R}(\bar{q}) \subset \mathcal{R}(q_{\text{in}})$. It follows that $\operatorname{int}(\mathcal{R}(\bar{q})) \subset \operatorname{int}(\mathcal{R}(q_{\text{in}}))$. But by local controllability we have that $\bar{q} \in \operatorname{int}(\mathcal{R}(\bar{q}))$. Hence $\bar{q} \in \operatorname{int}(\mathcal{R}(q_{\text{in}}))$. □

4. Controllability via linearization

Before delving into the controllability of nonlinear systems via the geometric techniques related to Lie brackets, we present here the classical theory of linear controllability. In particular, as expected, we will show that the controllability of the linearized system implies one of the nonlinear system. However, we will also show that this is a very restrictive condition, and that in many cases the linearized system is not controllable, while the nonlinear system is.

4.1. Linear control systems

Let us consider time-independent linear control systems on $\mathbb{R}^n$. These are systems of the form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ x(0) = x_0 \end{cases} \qquad A \in \mathbb{R}^{n\times n}, \quad B \in \mathbb{R}^{n\times m}. \tag{4.1}$$

Henceforth, we consider the set of controls to be $\mathcal{U} = L^\infty_{\mathrm{loc}}(\mathbb{R}_+, U)$, where $U \subset \mathbb{R}^m$ is a non-empty set of control values.

An essential tool for the study of these systems is the variation of constants formula (or Duhamel formula) for its solutions:

$$x(t; x_0, u) = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s)\,ds, \qquad \forall t \geq 0,\ u \in \mathcal{U}. \tag{4.2}$$

In the case of linear systems, it turns out that controllability can be verified via a purely algebraic condition.

Definition 4.1 (Kalman rank condition)**.** The pair $(A, B) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m}$ satisfies the *Kalman rank condition* if the Kalman matrix

$$K = [B, AB, \dots, A^{n-1}B] \in \mathbb{R}^{n\times nm}, \tag{4.3}$$

is of maximal rank n.

Theorem 4.2 (Kalman Theorem)**.** *Assume that $U = \mathbb{R}^m$. Then,* (4.1) *is controllable from $x_0 \in \mathbb{R}^n$ and in time $T > 0$ if and only if (A, B) satisfies the Kalman rank condition. In particular, if a linear system is controllable from x_0 in time $T > 0$, then it is small-time controllable from any initial point.*

Proof. By the variation of constants formula (4.2), we see that controllability is equivalent to the surjectivity of the linear operator

$$L : L^\infty([0, T], \mathbb{R}^m) \to \mathbb{R}^n, \qquad Lu = \int_0^T e^{-As}Bu(s)\,ds. \tag{4.4}$$

Here, we used that the exponential matrix e^{TA} is always invertible.

Let us prove that the fact that L invertible implies $\operatorname{rank} K = n$. We proceed by contradiction and assume that $\operatorname{rank} K < n$. That is, there exists $p \in \mathbb{R}^n$, $p \neq 0$, such that

$$p^\top K = 0 \iff p^\top A^i B = 0, \qquad \forall i \in \{1, \dots, n\}. \tag{4.5}$$

Recall that by Cayley-Hamilton Theorem[1], we can write A^j as a linear combination of $\operatorname{Id}, A, \dots, A^{n-1}$. That is, for any $j \in \mathbb{N}$, there exists $a_0, \dots, a_{n-1}$ such that

$$A^j = \sum_{i=0}^{n-1} a_i A^i. \tag{4.6}$$

This implies that (4.5) actually holds for any $i \in \mathbb{N}$, which yields

$$p^\top e^{-As} B = \sum_{j=0}^{+\infty} p^\top \frac{(-As)^j}{j!} B = 0. \tag{4.7}$$

In particular, this shows that $p^\top Lu = 0$ for any $u \in L^\infty([0,T], \mathbb{R}^m)$ proving that L is not surjective.

To prove the opposite implication, assume that there exists $p \in \mathbb{R}^n$, $p \neq 0$, such that

$$p^\top Lu = 0 \qquad \forall u \in L^\infty([0,T], \mathbb{R}^m). \tag{4.8}$$

Consider, for $i \in \{1, \dots, n\}$ and $\tau \in [0,T]$, the control

$$u(\tau) = \begin{cases} e_i, & \text{if } t \in [0, \tau], \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

Here, e_i is the i-th element of the canonical basis of $\mathbb{R}^n$. Thus, we have

$$Lu = \int_0^\tau e^{-As} Bu \, ds = \left[\frac{\operatorname{Id} - e^{-\tau A}}{A} \right] Bu, \qquad \text{where} \qquad \frac{\operatorname{Id} - e^{-\tau A}}{A} = \sum_{j=1}^{+\infty} \frac{(-1)^{j-1} \tau^j}{j!} A^{j-1}. \tag{4.10}$$

Assumption (4.8) then yields

$$0 = p^\top \left[\frac{\operatorname{Id} - e^{-\tau A}}{A} \right] Bu = \sum_{j=1}^{+\infty} \frac{(-1)^{j-1} \tau^j}{j!} p^\top A^{j-1} Bu, \qquad \forall \tau \in [0,T]. \tag{4.11}$$

By analyticity[2] w.r.t. τ of the right-hand side, this implies that $p^\top A^{j-1} Bu = 0$, that is $\operatorname{rank} K < n$. □

[1]Recall that the Cayley-Hamilton Theorem states that every square matrix satisfies its own characteristic polynomial.

[2]Equivalently, one can observe that

$$0 = \frac{d^k}{d\tau^k} \left[\sum_{j=1}^{+\infty} \frac{(-1)^{j-1} \tau^j}{j!} p^\top A^{j-1} Bu \right]_{\tau=0} = p^\top A^{k-1} Bu, \qquad \forall k \in \mathbb{N}.$$

4.2. Linearization principle

Consider now the nonlinear system (3.1), and assume that at q_{in} there exists $u^\star \in U$ inducing an equilibrium, i.e., such that $F(q_{\text{in}}, u^\star) = 0$. In this case, we can consider the linearized system at q_{in} and $u^\star$, which is given by

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad A = \frac{\partial F}{\partial q}(q_{\text{in}}, u^\star), \quad B = \frac{\partial F}{\partial u}(q_{\text{in}}, u^\star). \tag{4.12}$$

Our interest in linear systems is mainly due to the following result, which states that the controllability of the linearized system implies the local controllability of the nonlinear system.

Theorem 4.3 (Linearization principle)**.** *Assume that for* $q_{\text{in}} \in M$ *there exists* $u^\star \in \operatorname{int} U$ *such that* $(q_{\text{in}}, u^\star)$ *is an equilibrium of* (3.1) *(i.e.,* $F(q_{\text{in}}, u^\star) = 0$*). Assume moreover, that the linearized system at* $(q_{\text{in}}, u^\star)$ *is controllable, then* (3.1) *is locally controllable at* q_{in}.

Proof. Let us fix $v_1, \ldots, v_n \in \mathcal{U}$ such that the corresponding trajectories of the linearized system at time $T > 0$ are linearly independent. Say that $x_i(T; 0, v_i) = e_i$, where e_i is the i-th element of the canonical basis of $\mathbb{R}^n$.

For any $z \in \mathbb{R}^n$ let us define the control

$$u_z(\cdot) = u^\star + z_1 v_1(\cdot) + \cdots + z_n v_n(\cdot).$$

Observe that, since $u^\star \in \operatorname{int} U$, there exists $r > 0$ such that $u_z(t) \in U$ for every $t \in [0, T]$ and $z \in B_r(0) \subset \mathbb{R}^n$, so that $u_z(\cdot) \in \mathcal{U}$ for every $z \in B_r(0)$. Consider now the map $\Phi : B_r(0) \to \mathbb{R}^n$ defined as

$$\Phi(z) := q\big(T; q_{\text{in}}, u_z(\cdot)\big). \tag{4.13}$$

Clearly $\Phi(0) = q_{\text{in}}$, since $u_0(\cdot) = u^\star$. Thus, in order to prove the statement it suffices to show that Φ is a local diffeomorphism near 0. To this aim, by the Inverse Function Theorem, we now show that $\frac{\partial \Phi}{\partial z}(0)$ is invertible.

Let

$$W(t) := \frac{\partial}{\partial z}\Big|_{z=0} q(t; 0, u_z(\cdot)). \tag{4.14}$$

Observe that $W(0) = 0$, since $u_0(\cdot) = u^\star$. On the other hand, $W(T) = \frac{\partial \Phi}{\partial z}(0)$, since $\Phi(z) = q(T; q_{\text{in}}, u_z(\cdot))$. Let us compute the dynamics of W. Since the derivative w.r.t. z and t commute, and since $q(t; q_{\text{in}}, u_z(\cdot))$ is the solution of (3.1) with initial condition q_{in}, we have

$$\begin{aligned} \dot{W}(t) &= \frac{\partial}{\partial z}\Big|_{z=0} \dot{q}(t; q_{\text{in}}, u_z(\cdot)) \\ &= \frac{\partial}{\partial z}\Big|_{z=0} F\big(q(t; q_{\text{in}}, u_z(\cdot)), u_z(t)\big) \\ &= \frac{\partial F}{\partial q}\Big|_{(q_{\text{in}}, u^\star)} W(t) + \frac{\partial F}{\partial u}\Big|_{(q_{\text{in}}, u^\star)} V \\ &= AW(t) + BV. \end{aligned} \tag{4.15}$$

Here, we let $V(t) = [v_1(t), \dots, v_n(t)]$.

According to the above, each column W_i of W is the solution of the linearized system (4.12) with control $v_i(\cdot)$ and initial condition 0. Due to our choice of $v_1, \dots, v_n$ we then have that $W_i(T) = e_i$ for every $i \in \{1, \dots, n\}$, so that $W(T) = [e_1, \dots, e_n] = \mathrm{Id}$. This shows that $\frac{\partial \Phi}{\partial z}(0) = W(T)$ is invertible, and thus Φ is a local diffeomorphism near 0, concluding the proof. □

Example 4.4. Let us try to apply the linearization principle to the vehicle driven by two orthogonal fans (see Section 1.2). Recall that the dynamics of the system is given by

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ \theta \end{pmatrix} = F(x, y, \theta, u) = \begin{pmatrix} u_1 \cos\theta \\ u_1 \sin\theta \\ u_2 \end{pmatrix}. \tag{4.16}$$

We consider controls defined on $U = \mathbb{R}^2$, for simplicity.

Since the system is linear in the controls, at any point (x_0, y_0, θ_0) the constant control $u^\star = (0, 0)$ induces an equilibrium. It is easy to see that these are the only equilibria of the system.

To obtain the linearized system, we compute

$$\frac{\partial F}{\partial q} = \begin{pmatrix} 0 & 0 & -u_1 \sin\theta \\ 0 & 0 & u_1 \cos\theta \\ 0 & 0 & 0 \end{pmatrix}, \qquad \frac{\partial F}{\partial u} = \begin{pmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.17}$$

Evaluating the above at $q = (x_0, y_0, \theta_0)$ and $u^\star = (0, 0)$, we see that

$$A = \left.\frac{\partial F}{\partial q}\right|_{(q_{\mathrm{in}}, u^\star)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \left.\frac{\partial F}{\partial u}\right|_{(q_{\mathrm{in}}, u^\star)} = \begin{pmatrix} \cos\theta_0 & 0 \\ \sin\theta_0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.18}$$

In particular, the linearized system at $(q_{\mathrm{in}}, u^\star)$ is given by

$$\dot z = Bu, \quad \text{where } z \in \mathbb{R}^3.$$

Since $\operatorname{rank} B = 2$ and $A = 0$, the Kalman condition cannot hold, and thus the system is not controllable.

Nevertheless, we have shown in Section 1.2 that the nonlinear system (4.16) is indeed controllable (actually it is short-time locally controllable, if we consider $U = \mathbb{R}^2$).

It is not difficult to see that the above phenomenon (a controllabe nonlinear system whose linearization is never controllable) depends uniquely on the fact that $u \mapsto F(q, u)$ is linear and that the system is underactuated (i.e., $\dim \operatorname{span}\{F(q, u) \mid u \in U\} < \dim M$). Indeed, we have the following.

Proposition 4.5. *Assume that $u \mapsto F(q, u)$ is linear and that $0 \in \operatorname{int} U$. Then, for any $q_{\mathrm{in}} \in M$ it holds that $(q_{\mathrm{in}}, 0)$ is an equilibrium of (3.1). Moreover, the linearized system at $(q_{\mathrm{in}}, 0)$ is controllable if and only if the system is fully actuated, in the sense that*

$$\dim \operatorname{span}\{F(q_{\mathrm{in}}, u) \mid u \in U\} = \dim M. \tag{4.19}$$

Proof. The fact that $F(q_{\text{in}}, 0) = 0$ for any $q_{\text{in}} \in M$ is a consequence of the linearity assumption. Moreover, in coordinates, the linearity assumption yields that $F(x, u) = B(x)u$ where $B(x) \in \mathbb{R}^{n\times m}$ is such that

$$\operatorname{rank} B(x) = \dim \operatorname{span}\{F(q_{\text{in}}, u) \mid u \in U\}. \tag{4.20}$$

Then,

$$\left.\frac{\partial F}{\partial x}\right|_{(q_{\text{in}},u)} = \left.\frac{\partial B}{\partial x}\right|_{q_{\text{in}}} u \implies \left.\frac{\partial F}{\partial x}\right|_{(q_{\text{in}},0)} = 0. \tag{4.21}$$

On the other hand, letting x_0 be the coordinate representation of q_{in}, we have

$$\left.\frac{\partial F}{\partial u}\right|_{(q_{\text{in}},0)} = B(x_0). \tag{4.22}$$

Finally, the linearized system at $(q_{\text{in}}, 0)$ is

$$\dot{x} = B(x_0)u. \tag{4.23}$$

By Kalman Theorem, this system is controllable if and only if $\operatorname{rank} B(x_0) = n$. By (4.20) this completes the proof of the statement. □

5. The geometric approach to controllability

Recall that we consider a nonlinear control system on a manifold M of the form

$$\dot{q}(t) = F(q(t), u(t)), \tag{5.1}$$

where q, u, and F satisty the assumptions presented in Chapter 3.

In the following it will be useful to introduce the following objects.

Definition 5.1. The *family of vector fields* associated to (5.1) is the set $\mathcal{F} \subset \mathrm{Vec}(M)$ consisting of the vector fields obtained via constant controls $u \in U$. More precisely,

$$\mathcal{F} := \{F(\cdot, v) \mid v \in U\} \subset \mathrm{Vec}(M).$$

We denote by $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ the set of points that are reachable from $q_{\text{in}} \in M$ in arbitrary time via piecewise constant controls[1].

According to the assumptions in Chapter 3, all vector fields of the family $\mathcal{F}$ are complete, i.e., for every $F \in \mathcal{F}$, $q_{\text{in}} \in M$, the equation $\dot{q} = F(q)$ with initial condition $q(0) = q_{\text{in}}$ admits a solution in $(-\infty, \infty)$.

The techniques developed in this chapter will provide information on $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. Observe that since $\mathcal{R}(q_{\text{in}}) \supset \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$, showing that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = M$ will be enough to prove controllability of system (5.1).

Let us mention, however, that it might happen that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ and the reachable set $\mathcal{R}(q_{\text{in}})$ differ, as illustrated by the following example.

Example 5.2. Let $\Gamma = \{(t, t^2) \mid t \in [0, 1]\} \subset \mathbb{R}^2$ and consider any non-negative $\mathcal{C}^\infty$ function φ such that $\varphi(x, y) = 0$ if and only if $(x, y) \in \Gamma$. Consider the following control system in $\mathbb{R}^3$:

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = X + uY, \qquad X = \begin{pmatrix} 1 \\ 0 \\ \varphi(x,y) \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad u \in [-1, 1]. \tag{5.2}$$

Let us show that the point $q_{\text{fi}} = (1, 1, 0)$ belongs to the reachable set from the origin $\mathcal{R}(0)$ but not to $\mathcal{R}^{\mathcal{F}}(0)$.

[1] The notation is used to stress that the set depends only on the set of admissible vector fields, and not by their parameterization $v \mapsto F(\cdot, v)$.

The fact that $q_{\text{fi}} \in \mathcal{R}(0)$ is obtained by considering the control $\bar{u}(t) = t$. Indeed, with this choice we have

$$q(t;0,\bar{u}(\cdot)) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} t \\ \int_0^t \bar{u}(s)\,ds \\ \int_0^t \varphi(x(s),y(s))\,ds \end{pmatrix} = \begin{pmatrix} t \\ t^2 \\ 0 \end{pmatrix}. \tag{5.3}$$

Here, in the second equality we used the fact that $\varphi(t,t^2) = 0$. It follows that, $q_{\text{fi}} = q(1;0,\bar{u}(\cdot)) \in \mathcal{R}(0)$.

On the other hand, let us consider a piecewise constant control $u(\cdot)$. In particular, there exists $\tau > 0$ such that $u(t) = u_0 \in [-1,1]$ for all $t \in [0,\tau]$. This implies that, for $t \in [0,\tau]$ the corresponding trajectory has coordinates

$$q(\tau;0,\bar{u}(\cdot)) = \begin{pmatrix} \tau \\ u_0\tau \\ \int_0^\tau \varphi(t,tu_0)\,dt \end{pmatrix}. \tag{5.4}$$

Since $(t,tu_0) \notin \Gamma$ for all $t \notin \{0,1\}$, we have that $\varphi(t,tu_0) > 0$ and thus the coordinate $z(\tau)$ is strictly positive. Since the control does not act on this coordinate and the function φ is non-negative, this implies that $z(T) > 0$ and thus that $q_{\text{fi}} \notin \mathcal{R}^{\mathcal{F}}$.

Exercise 5.1. Under which conditions on g does the above example generalise to $\Gamma = \{(t,g(t)) \mid t \in [0,1]\}$?

5.1. The Lie bracket generating condition

Definition 5.3. Let $\mathcal{F}$ be a family of vector fields. The *Lie algebra generated by* $\mathcal{F}$, denoted by $\operatorname{Lie}\mathcal{F}$, is the smallest sub-algebra[2] of $\operatorname{Vec}(M)$ containing $\mathcal{F}$.

The following result allows to concretely compute $\operatorname{Lie}\mathcal{F}$.

Proposition 5.4. *The algebra* $\operatorname{Lie}\mathcal{F}$ *is the span of all vector fields of* $\mathcal{F}$ *and of their iterated Lie brackets of any order, that is,*

$$\operatorname{Lie}\mathcal{F} = \operatorname{span}\Big\{[f_1,[f_2,[\dots,f_k]]] \mid k \in \mathbb{N},\ f_1,\dots,f_k \in \mathcal{F}\Big\},$$

with the convention that, for $k = 1$, $[f_1,[f_2,[\dots,f_k]]] = f_1$.

Proof. Let $\mathcal{L} = \operatorname{span}\{[f_1,[f_2,[\dots,f_k]]] \mid k \in \mathbb{N},\ f_1,\dots,f_k \in \mathcal{F}\}$. Then $\mathcal{L} \subset \operatorname{Lie}(\mathcal{F})$, since the latter contains $\mathcal{F}$ and is closed under the Lie bracket.

[2] That is,

$$\operatorname{Lie}\mathcal{F} = \bigcap_{\mathcal{F}\subset E} E,$$

where $E \subset \operatorname{Vec}(M)$ is a vector sub-space of $\operatorname{Vec}(M)$ which is closed under the Lie bracket (i.e., $[f,g] \in E$ for any $f,g \in E$).

In order to prove the inclusion $\mathrm{Lie}(\mathcal{F}) \subset \mathcal{L}$, we are left to prove that the linear space $\mathcal{L}$ is a Lie algebra, i.e., that it is closed under the Lie bracket. Let then $\hat{f} = [f_1, [f_2, [\dots, f_k]]]$ and $\hat{g} = [g_1, [g_2, [\dots, g_h]]]$ with $k, h \in \mathbb{N}$ and $f_1, \dots, f_k, g_1, \dots, g_h \in \mathcal{F}$. We are going to prove by induction on k that $[\hat{f}, \hat{g}]$ is in $\mathcal{L}$. For $k = 1$, we have that $[\hat{f}, \hat{g}] = [f_1, \hat{g}] \in \mathcal{L}$ by definition. Let now $k \geq 2$ and set $\tilde{f} = [f_2, [f_3, [\dots, f_k]]]$, so that $\hat{f} = [f_1, \tilde{f}]$. By the inductive assumption, $[f_1, X]$ and $[\tilde{f}, X]$ belongs to $\mathcal{L}$ for every $X \in \mathcal{L}$. Hence, applying the Jacobi identity and using the fact that $\mathcal{L}$ is a linear space, we have

$$[\hat{f}, \hat{g}] = [[f_1, \hat{g}], \tilde{f}] + [f_1, [\tilde{f}, \hat{g}]] \in \mathcal{L}.$$

This concludes the proof. □

Definition 5.5. We say that the family $\mathcal{F}$ is *Lie bracket generating at a point* q if the dimension of $\mathrm{Lie}_q \mathcal{F} := \{f(q) \mid f \in \mathrm{Lie}\,\mathcal{F}\}$ is equal to n. We say that the family $\mathcal{F}$ is *Lie bracket generating* if this condition is verified for every $q \in M$.

Remark 5.6. Notice that in general $\mathrm{Lie}\,\mathcal{F}$ is an infinite-dimensional space, while $\mathrm{Lie}_q \mathcal{F}$ is a subspace of TM. In particular $\dim \mathrm{Lie}_q \mathcal{F} \leq n$.

5.1.1. Affine control systems

An affine control system is a system of the form

$$\dot{q} = f_0(q) + \sum_{i=1}^{m} u_i(t) f_i(q), \tag{5.5}$$

where $f_0, f_1, \dots, f_m$ belong to $\mathrm{Vec}(M)$ and $u(\cdot) = (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty) \to U \subset \mathbb{R}^m$ is the control. The vector fields f_0 is called *drift*. We notice that the drift represents the dynamics of the system "in absence of control", that is, when u is set to 0 (even if we do not necessarily assume that U contains the origin of $\mathbb{R}^m$).

Exercise 5.2. Let $\mathcal{F}$ be the family of vector fields associated with (5.5) and assume that the convex hull of U has nonemty interior in $\mathbb{R}^m$.

- Prove that if $\{f_0, f_1, \dots, f_m\}$ is Lie bracket generating then $\mathcal{F}$ is Lie bracket generating as well.
- Prove that if $\{f_1, \dots, f_m\}$ is Lie bracket generating then $\mathcal{F}$ is Lie bracket generating as well.

5.2. Krener Theorem: local accessibility

The fact that a control system is Lie bracket generating does not permit in general to conclude that it is controllable. Consider for instance the control system on the plane

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + u \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

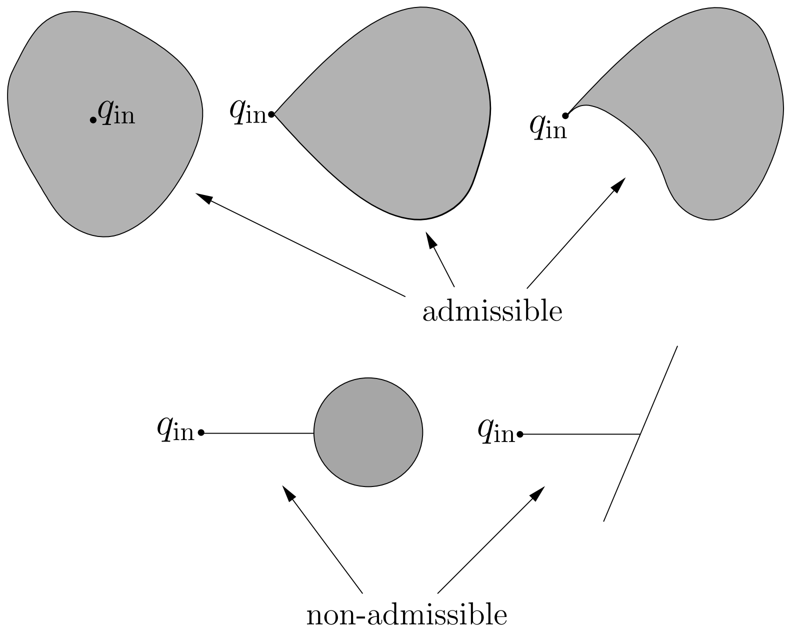


Figure 5.1.: Admissible and non-admissible reachable sets when the system is Lie bracket generating at q_{in}.

where $u(\cdot) : [0, \infty) \to [-1, 1]$. This system is trivially Lie bracket generating, indeed $\mathcal{F}$ contains the vector fields $\{(1,1),(1,-1)\}$ and thus $\operatorname{span}\mathcal{F} = \mathbb{R}^2$. However, starting from the origin one cannot reach any point whose first coordinate is negative. This is essentially due to the fact that the family $\mathcal{F}$ contains two vector fields but not their opposite (cf. Remark 2.26).

Nevertheless, the Lie bracket generating condition permits to say that a system is *locally accessible* in the following sense.

Theorem 5.7 (Krener Accessibility Theorem)**.** *If $\mathcal{F}$ is Lie bracket generating at q_{in}, then for every $\tau > 0$ we have that*

$$q_{\text{in}} \in \overline{\operatorname{int} \mathcal{R}(\leq \tau, q_{\text{in}})}. \tag{5.6}$$

The conclusion (5.6) of Krener's Theorem can be reformulated as follows:

- For any $\tau > 0$, the set $\mathcal{R}(\leq \tau, q_{\text{in}})$ has nonempty interior;
- The point q_{in} is a density point for such interior.

Krener's theorem says in particular that trajectories starting from a point at which the system is Lie bracket generating can reach (in an arbitrarily small time) a set having nonempty interior. Figure 5.1 shows what one can expect/non-expect from $\mathcal{R}(\leq \tau, q_{\text{in}})$, $\tau > 0$.

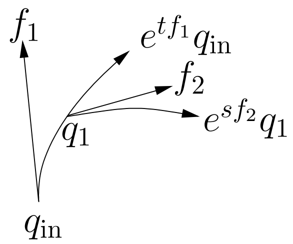


Figure 5.2.: Construction for the proof of Krener's theorem.

Proof Theorem 5.7. Let $\tau > 0$ and define $\mathbb{R}^n_+ := \{x_i \geq 0 \text{ for all } i = 1, \dots, n\}$. The idea is to find n vector fields $f_1, \dots, f_n \in \mathcal{F}$ such that the map

$$\Phi : x \in \mathbb{R}^n_+ \mapsto e^{x_n f_n} \circ \cdots \circ e^{x_1 f_1}(q_{\text{in}}) \in M,$$

is a local diffeomorphism at some $\bar{x} \in \mathbb{R}^n_+$ with $\|x\|_1 := \sum_{i=1}^n x_i < \tau$. This will prove the statement since $\Phi(\{x \in \mathbb{R}^n_+ \mid \|x\|_1 < \tau\}) \subset \mathcal{R}^{\mathcal{F}}(\leq \tau, q_{\text{in}})$ this proves the statement.

Recall that $[f, g](q) = 0$ if $f(q) = g(q) = 0$. Hence, the Lie bracket generating condition implies that there exists $f_1 \in \mathcal{F}$ such that $f_1(q_{\text{in}}) \neq 0$ (otherwise $\text{Lie}_{q_{\text{in}}} \mathcal{F} = \{0\}$). If $n = 1$ the conclusion follows with $\bar{x} = 0$ since

$$\left.\frac{\partial}{\partial x_1}\right|_{x_1=0} e^{x_1 f_1}(q_{\text{in}}) = f_1(q_{\text{in}}) \neq 0.$$

Let $n = 2$. In this case, $\Phi(x) = e^{x_2 f_2} \cdot e^{x_1 f_1}(q_{\text{in}})$ and, in coordinates, for any $\bar{x}, h \in \mathbb{R}^n_+$ such that $\bar{q} = \Phi(\bar{x})$ we have that (recall Definition 2.10):

$$\begin{aligned} \Phi_*|_{\bar{x}}(h) &= \left.\frac{d}{dt}\right|_{t=0} e^{(x_2+th_2)f_2} \cdot e^{(x_1+th_1)f_1}(q_{\text{in}}) \\ &= h_1 (e^{\bar{x}_2 f_2})_* f_1(e^{\bar{x}_1 f_1}(\bar{q})) + h_2 f_2(\bar{q}). \end{aligned}$$

The above expression suggests to consider $\bar{x} = (\bar{x}_1, 0)$, so that $(e^{\bar{x}_2 f_2})_* = \text{Id}$, by (2.13). Moreover, in this case $\Phi(\bar{x}) = e^{\bar{x}_1 f_1}(q_{\text{in}})$, and we have

$$\Phi_*|_{(\bar{x}_1,0)}(h) = h_1 f_1(\Phi(\bar{x})) + h_2 f_2(\Phi(\bar{x})).$$

Hence, we just need to find two vector fields $f_1, f_2 \in \mathcal{F}$ and $\bar{q} = \Phi(\bar{x})$ with $\|\bar{x}\| < \tau$ and such that $f_1(\bar{q})$ and $f_2(\bar{q})$ are linearly independent. Unfortunately, the Lie bracket generating condition does not allow to find such vector fields at q_{in} itself[3].

We bypass this issue as follows (we refer to Figure 5.2 for an illustration). Pick $f_1 \in \mathcal{F}$ such that $f_1(q_{\text{in}}) \neq 0$, and $\varepsilon < \tau$ to be fixed later. Assume by contradiction that for all

[3]Think, e.g., of the vector fields $X = \partial_x$ and $Y = x\partial_y$ in $\mathbb{R}^2$. They are Lie bracket generating, since $[X, Y] = \partial_y$, but $\dim \operatorname{span} \mathcal{F}_{(0,0)} = 1$.

$x_1 \in [0, \varepsilon]$ and any $f_2 \in \mathcal{F}$ it holds that $f_2(e^{x_1 f_1}(q_{\text{in}})) \parallel f_1(e^{x_1 f_1}(q_{\text{in}}))$. This means that, all vector fields in $\mathcal{F}$ are tangent to the onedimensional submanifold

$$N_1 = \{e^{x_1 f_1}(q_{\text{in}}) \mid 0 < x_1 < \varepsilon\} \subset M. \tag{5.7}$$

By Corollary 2.25, it follows that all Lie brackets of $\mathcal{F}$ are tangent to N_1, i.e., $\operatorname{Lie} \mathcal{F}_q \subset T_q N_1$ for all $q \in N_1$. In particular, since N_1 is onedimensional,

$$\dim \operatorname{Lie} \mathcal{F}_q \leq 1, \qquad \forall q \in N_1. \tag{5.8}$$

Recall now that the Lie bracket condition is an open condition[4]. Hence, we can fix ε sufficiently small so that $\mathcal{F}$ is Lie bracket generating at each point of N_1, i.e., $\dim \operatorname{Lie}_q \mathcal{F} = n$ for all $q \in N_1$. This contradicts (5.8). We have thus proven that there exist f_1, f_2 and $\bar{q} = \Phi(\bar{x}_1, 0) \in N_1$ such that $f_1(\bar{q})$ and $f_2(\bar{q})$ are linearly independent. Since we chose $\varepsilon < \tau$ we can also insure that $\|(\bar{x}_1, 0)\|_1 < \tau$.

By the previous discussion, this proves the statement for $n = 2$. More generally, this argument proves that the image of $(x_1, x_2) \mapsto e^{x_2 f_2} \circ e^{x_1 f_1}(q_{\text{in}})$ describe a 2-dimensional surface N_2 inside $\mathcal{R}(\leq \tau, q_{\text{in}})$.

In the general case, iterating this argument we construct an open subset N_n (where n is the dimension of M) that is contained in $\mathcal{R}(\leq n\varepsilon, q_{\text{in}})$, concluding the proof of the statement. □

Remark 5.8. Observe that the proof only uses constant controls. In particular, we have actually shown that q_{in} belongs to the closure of the interior of $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$.

Remark 5.9. From the proof of Krener's Theorem it follows that $\mathcal{R}(\leq \tau, q_{\text{in}})$, $\tau > 0$, contains an open set Ω having q_{in} in its closure whose points can be reached by trajectories of the type $e^{t_n f_n} \circ \ldots \circ e^{t_1 f_1} q_{\text{in}}$ where $t_1, \ldots, t_n > 0$ and $f_1, \ldots, f_n \in \mathcal{F}$, i.e., by trajectories corresponding to piecewise constant controls made by n pieces. Notice that the vector fields $f_1, \ldots, f_n$ could be repeated. For instance if $\mathcal{F} = \{f, g\}$ is a Lie bracket generating family and we are in dimension 3, we could have for instance $f_1 = f, \quad f_2 = g, \quad f_3 = f$.

5.3. Symmetric systems

The Lie bracket generating condition is not sufficient to conclude that the system is controllable. However, if we add the symmetry condition defined below, then we can conclude that the system is controllable. This is the content of Chow–Rashevskii Theorem.

Definition 5.10. A family of vector fields $\mathcal{F}$ is *symmetric* if $f \in \mathcal{F}$ if and only if $-f \in \mathcal{F}$.

The importance of this condition relies in the fact that it allows us to construct the Lie bracket $[f, g]$ (cf. Lemma 2.23). Indeed, we obtain the following.

[4] That is, it is satisfied at a neighborhood of q_{in}. Indeed, if n vector fields are linearly independent at a point, they stay linearly independt in a neighborhood of that point.

Theorem 5.11 (Chow–Rashevskii). *Assume that M is connected and that $\mathcal{F}$ is Lie bracket generating and symmetric. Then,*

$$\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) \subset \mathcal{R}(q_{\text{in}}) = M, \qquad \forall q_{\text{in}} \in M.$$

In particular, the system is controllable.

Proof. Step 1. Local controllability. We start by showing that the system is locally controllable. To this aim, let us fix $q_{\text{in}} \in M$ and $\tau > 0$, and show that $q_{\text{in}} \in \operatorname{int} \mathcal{R}^{\mathcal{F}}(\leq \tau, q_{\text{in}})$.

Since $\mathcal{F}$ is Lie bracket generating, for any $\tau > 0$, the set $\mathcal{R}(\leq \tau, q_{\text{in}})$ contains a nonempty open set Ω whose points can be reached by trajectories corresponding to piecewise controls made by n pieces (cf. Remark 5.9),. That is, there exist $f_1, \ldots, f_n \in \mathcal{F}$ such that the map

$$\Phi : x \in \mathbb{R}^n_+ \mapsto e^{x_n f_n} \circ \cdots \circ e^{x_1 f_1}(q_{\text{in}}) \in M,$$

is such that $\Omega \subset \Phi(\{x \in \mathbb{R}^n_+ \mid \|x\|_1 \leq \tau\}) \subset \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$.

Pick any $\bar{x}$ such that $\bar{q} = \Phi(\bar{x}) \in \Omega$. By symmetry of $\mathcal{F}$, we have that $-f_1, \ldots, -f_n \in \mathcal{F}$, and thus

$$V := e^{-\bar{x}_1 f_1} \circ \cdots \circ e^{-\bar{x}_n f_n}(\Omega) \subset \mathcal{R}^{\mathcal{F}}(\leq 2\tau, q_{\text{in}}).$$

Now, since Ω is open and $e^{-t_1 f_1} \circ \ldots \circ e^{-t_n f_n}$ is a diffeomorphism, we have that V is open. Moreover V is nonempty, since it contains q_{in}: $\bar{q} \in \Omega$ and $e^{-t_1 f_1} \circ \ldots \circ e^{-t_n f_n} \bar{q} = q_{\text{in}}$. It follows that, for every q_{in}, $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ contains a neighborhood of q_{in}.

Step 2. Global controllability. Although this will follows from Theorem 5.16, proven below, let us give a direct proof of this fact.

First of all, let us show that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is open for any $q_{\text{in}} \in M$. To this aim, let us fix $q_{\text{fi}} \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. By concatening the appropiate controls, we see that $\mathcal{R}(q_{\text{fi}}) \subset \mathcal{R}(q_{\text{in}})$. Since $\mathcal{R}(q_{\text{fi}})$ is a neighborhood of q_{fi}, this proves that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is a neighborhood of q_{fi}.

Now, we define the equivalence relation $\sim$ on M as follows:

$$q_1 \sim q_2 \quad \text{if and only if} \quad q_1 \in \mathcal{R}(q_2).$$

Let us show that this is an equivalence relation. The reflexivity follows since we are allowed to consider controls defined on time $T = 0$. The transitivity follows by concatenating the appropriate controls. Finally, the symmetry follows by reversing the controls, thanks to the symmetry of $\mathcal{F}$.

Let us now consider the quotient space $M/\sim$. Clearly, the equivalence class of any $q_{\text{in}} \in M$ coincides with its reachable set $\mathcal{R}(q_{\text{in}})$. In particular, by Lemma 3.4, this is open. Hence, the quotient space $M/\sim$ is a partition of M into open sets. Since M is connected, this implies that there is only one equivalence class, i.e., $\mathcal{R}(q_{\text{in}}) = M$ for any $q_{\text{in}} \in M$, concluding the proof. □

Remark 5.12. Notice that in the proof the time τ can be made arbitrary small. Hence, we have actually proved that the system is small-time locally controllable at every point q_{in}. Also, we have proved that every point of a neighborhood of q_{in} can be reached with trajectories made by $2n$ pieces.

Remark 5.13. The Chow–Rashevskii theorem can be used in more general situations than those fixed by the hypotheses stated here. For instance

- it is sufficient for the family $\mathcal{F}$ to contain a symmetric family of Lie bracket generating vector fields;
- if one can prove that $\mathcal{F}$ is symmetric and Lie bracket generating in a connected open set Ω of M then one get the system is controllable in Ω.

Since any neighborhood of the origin contains a symmetric neighborhood of the origin, by the previous remark we obtain the following corollary.

Corollary 5.14. *Let $\{f_1, \ldots, f_m\}$ be Lie bracket generating and assume that U is a neighborhood of the origin in $\mathbb{R}^m$. Then the system*

$$\dot{q} = \sum_{i=1}^{m} u_i f_i(q), \qquad (u_1(\cdot), \ldots, u_m(\cdot)) : [0, \infty) \to U,$$

is controllable.

Example 5.15 (Reed-Shepp car). The car driven by two orthogonal fans introduced in Section 1.2, which is usually called the Reed-Shepp car, is controllable by Chow-Rashevskii Theorem. Indeed, on can compute

$$[X_1, F_2] = [\cos\theta\partial_x + \sin\theta\partial_y, \partial_\theta] = -\sin\theta\partial_x + \cos\theta\partial_y =: X_2,$$

It is immediate to check that $\{X_1, X_2, F_2\}$ are linearly independent.

5.4. From local to global controllability

In this section we present show that global controllability and local controllability are equivalent, under the Lie bracket generating condition. We stress that this fact is actually true in general, cf. [Bos+23].

Theorem 5.16. *Assume that M is connected and that $\mathcal{F}$ is Lie bracket generating at every point of M. Then, controllability is equivalent to local controllability, i.e., $\mathcal{R}(q_{\text{in}}) = M$ for any $q_{\text{in}} \in M$ if and only if for any $q_{\text{in}} \in M$ it holds $q_{\text{in}} \in \operatorname{int} \mathcal{R}(q_{\text{in}})$.*

The proof of this theorem is based on the following lemmas.

Lemma 5.17. *Assume that M is connected and that the system is locally controllable. Then, the system is* approximate controllable*, i.e., $\mathcal{R}(q_{\text{in}})$ is dense in M for every $q_{\text{in}} \in M$.*

Proof. By connectedness of M, it suffices to show that $\overline{\mathcal{R}(q_{\text{in}})}$ is open for every $q_{\text{in}} \in M$. To this aim, for any $\bar{q} \in \overline{\mathcal{R}(q_{\text{in}})}$ we will show that $\mathcal{R}(\bar{q}) \subset \overline{\mathcal{R}(q_{\text{in}})}$. Since by local controllability $\mathcal{R}(\bar{q})$ is a neighborhood of $\bar{q}$, this will prove that $\overline{\mathcal{R}(q_{\text{in}})}$ is open.

The fact that $\bar{q} \in \overline{\mathcal{R}(q_{\text{in}})}$ implies that there exists a sequence of controls u_n steering the system from q_{in} to $\bar{q}$. Moreover, for any $\bar{p} \in \mathcal{R}(\bar{q})$ there exists a control $\bar{u}$ steering the system from $\bar{q}$ to $\bar{p}$. Then, by concatenating the control u_n with $\bar{u}$, we obtain a sequence of controls steering the system from q_{in} to $\bar{p}$ and thus $\bar{p} \in \overline{\mathcal{R}(q_{\text{in}})}$. □

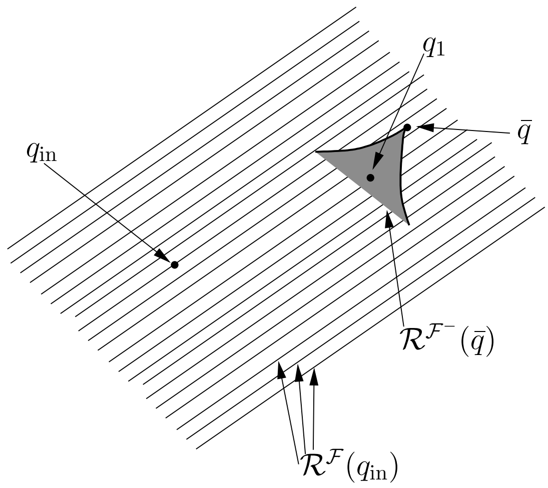


Figure 5.3.: Proof of Lemma 5.18.

Lemma 5.18. *If $\mathcal{F}$ is Lie bracket generating and $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is dense in M for some q_{in}, then $\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = M$.*

Proof. Let $\mathcal{F}^- := \{-f \mid f \in \mathcal{F}\}$. Since $\mathcal{F}$ is Lie bracket generating, then $\mathcal{F}^-$ is Lie bracket generating as well.

Let $\bar{q} \in M$. Thanks to Krener's theorem, $\mathcal{R}^{\mathcal{F}^-}(\bar{q})$ contains a nonempty open set. In particular it has nonempty intersection with $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. See Figure 5.3.

This means that $\bar{q} \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. Indeed from q_{in} one can reach a point $q_1 \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}}) \cap \mathcal{R}^{\mathcal{F}^-}(\bar{q})$ (since $q_1 \in \mathcal{R}(q_{\text{in}})$) and from q_1 one can reach $\bar{q}$ (since $q_1 = e^{t_k(-f_k)} \circ e^{t_1(-f_1)}(\bar{q})$ implies that $\bar{q} = e^{t_1 f_1} \circ e^{t_k f_k}(q_1)$). Being $\bar{q}$ arbitrary we have that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = M$. □

The same argument as above, also allows to prove the following.

We are now in position to prove Theorem 5.16.

Proof of Theorem 5.16. Clearly, every controllable system is locally controllable. Let us assume that the system is locally controllable and show that it is controllable.

By Lemma 5.17, the system is approximately controllable, i.e., $\mathcal{R}(q_{\text{in}})$ is dense in M for every $q_{\text{in}} \in M$. Hence, by Lemma 5.18, we have that $\mathcal{R}(q_{\text{in}}) = M$ for every $q_{\text{in}} \in M$, concluding the proof. □

Remark 5.19. The version of Theorem 5.16 presented in [Bos+23] does not require the Lie bracket generating condition. The proof is based on the equivalence relation argument we presented the proof of Theorem 5.11. In this case, however, the difficulty is to prove the symmetry of the relation ($q_1 \sim q_2 \iff q_1 \in \mathcal{R}(q_2)$), since the symmetry assumption on $\mathcal{F}$ is not available.

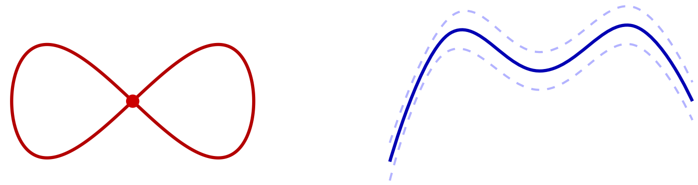


Figure 5.4.: *Left*: An *immersed* submanifold, highlighting the possible presence of self-intersections (marked in red). *Right*: An *embedded* submanifold, where it is always possible to find a tubular neighbourhood in M.

5.5. Orbits and necessary conditions for controllability

We have seen in the previous sections an important sufficient condition for the controllability of a nonlinear system, under the Lie bracket generating condition. In this section we discuss some necessary conditions that are a consequence of a very deep theorem of geometric nature: the *Orbit Theorem*. This theorem allows us to conclude that, besides pathological cases, the set of directions that one can hope to exploit starting from a point is measured precisely by $\mathrm{Lie}_q(\mathcal{F})$.

Definition 5.20. The *orbit* of the family $\mathcal{F}$ starting from $q_{\mathrm{in}} \in M$ is the set

$$\mathcal{O}(q_{\mathrm{in}}) = \{e^{t_k f_k} \circ \cdots \circ e^{t_1 f_1}(q_{\mathrm{in}}) \mid k \in \mathbb{N},\ t_1, \dots, t_k \in \mathbb{R},\ f_1, \dots, f_k \in \mathcal{F}\}. \tag{5.9}$$

Remark 5.21. $\mathcal{O}(q_{\mathrm{in}})$ can be interpreted as the reachable set, starting from q_{in}, of the family $-\mathcal{F} \cup \mathcal{F}$, using only piecewise constant controls.

Let us recall the following definitions.

Definition 5.22. An *immersed submanifold* of M is a subset of $\phi(N) \subset M$ such that N is a manifold and the inclusion map ϕ is an *immersion*. That is, ϕ is a smooth map and its tangent map $d_q\phi : T_qN \to T_{\phi(q)}M$ is injective at every point $q \in N$.

An *embedded submanifold* of M is an immersed submanifold such that the immersion ϕ is an *embedding*. That is, ϕ is an homeomorphism onto its image[5].

Figure 5.4 shows the difference between an embedded and an immersed submanifold. Another example of an immersed submanifold that is not embedded is the orbit of a vector field on the torus $\mathbb{T}^2$ with irrational slope. In this case, the orbit is dense in $\mathbb{T}^2$, and thus it cannot be homeomorphic to a one-dimensional manifold (cf. Figure 6.4).

We have the following result.

Theorem 5.23 (Orbit theorem)**.** *For every $q_{\mathrm{in}} \in M$, the set $\mathcal{O}(q_{\mathrm{in}})$ is an immersed submanifold of M. In particular, it has the same dimension at every point.*

Moreover, if $q \in \mathcal{O}(q_{\mathrm{in}})$ then $\mathrm{Lie}_q(\mathcal{F}) \subseteq T_q\mathcal{O}(q_{\mathrm{in}})$. The two spaces $\mathrm{Lie}_q(\mathcal{F})$ and $T_q\mathcal{O}(q_{\mathrm{in}})$ coincide if one of the following two conditions is verified:

[5]That is, the topology of the manifold N is the same as that of its image in M.

O1. Every element of $\mathcal{F}$ is an analytic vector field;

O2. The Lie algebra generated by $\mathcal{F}$ has constant rank, i.e., the dimension of $\mathrm{Lie}_q(\mathcal{F})$ is constant with respect to $q \in \mathcal{O}(q_{\mathrm{in}})$.

Before proving the Orbit Theorem, which is quite a feat, let us present some consequences.

First of all, observe that Frobenius Theorem (cf. Section 2.5) is a particular case of the Orbit Theorem, where $\mathrm{Lie}_q \mathcal{F} = \mathcal{F}$ and is of constant rank.

On the other hand, from the fact that $\mathrm{Lie}_q(\mathcal{F}) \subseteq T_q\mathcal{O}(q_{\mathrm{in}})$ it follows that every element of $\mathcal{F}$ is tangent to $\mathcal{O}(q_{\mathrm{in}})$. As a consequence we have the following.

Corollary 5.24. *For every $q_{\mathrm{in}} \in M$ we have that $\mathcal{R}(q_{\mathrm{in}}) \subseteq \mathcal{O}(q_{\mathrm{in}})$.*

Observe that, while $\mathcal{R}^{\mathcal{F}}(q_{\mathrm{in}}) \subset \mathcal{O}(q_{\mathrm{in}})$ by definition, the above result is not at all trivial: Indeed, $\mathcal{R}(q_{\mathrm{in}})$ is the set of points that one can reach using L^∞ controls (and not only piecewise constant ones).

Finally, we present an important consequence of the Orbit theorem on the controllability of nonlinear systems.

Corollary 5.25. *If $\mathcal{F}$ is not Lie bracket generating and it satisfies either O1. or O2. from Theorem 5.23, then $\mathcal{F}$ is not controllable.*

In particular, for symmetric families of vector fields, the Lie bracket generating condition is a necessary and sufficient condition for controllability whenever the family is analytic or has constant-dimensional Lie algebra.

In the following example, we show that the necessary part of this statement cannot be improved to encompass smooth vector fields.

Exercise 5.3. Let $\varphi \in \mathcal{C}^\infty(\mathbb{R}^2)$ be a nonnegative function such that $\operatorname{supp}\varphi = \{x \in \mathbb{R}^2 \mid x_1 \geq 0\}$. Consider the following system on $\mathbb{R}^2$:

$$\dot{q} = u_1 X(q) + u_2 Y(q), \qquad X(q) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad Y(q) = \begin{pmatrix} 0 \\ \varphi(x) \end{pmatrix}.$$

Clearly this system is not Lie bracket generating, since

$$\mathrm{Lie}_x \mathcal{F} = \mathbb{R} \times \{0\}, \qquad \text{if } x_1 < 0.$$

However, it is controllable. Indeed, to go from $q_{\mathrm{in}} = (x_1^0, x_2^0)$ to $q_{\mathrm{fi}} = (x_1^1, x_2^1)$ it suffices to:

1. Apply the control $u = (\pm 1, 0)$ for a certain amount of time t_1, so that $q(t_1) = (1, x_2^0)$;
2. Apply the control $u = (0, \pm 1)$ for a certain amount of time t_2, so that $q(t_1 + t_2) = (1, x_2^1)$.
3. Apply again the control $u = (\pm 1, 0)$ for a certain amount of time t_3, so that $q(t_1 + t_2 + t_3) = (x_1^1, x_2^1)$.

5.5.1. Proof of the Orbit Theorem

Before delving in the proof, let us introduce the following notion, strictly connected to the one of tangent map (cf. Definition 2.10).

Definition 5.26. Given a diffeomorphism $P : M \to M$ we can use it to *push-forward* vector fields. That is, we can define $P_* : \mathrm{Vec}(M) \to \mathrm{Vec}(M)$ letting

$$P_* X(q_0) := d_p P(X(p)), \qquad p = P^{-1}(q_0). \tag{5.10}$$

Equivalently,

$$P_* X(q_0) = \frac{d}{dt}\bigg|_{t=0} P(e^{tX} \circ P^{-1}(q_0)). \tag{5.11}$$

The push-forward satisfies the following properties.

- We have
$$e^{tP_*f} = P \circ e^{tf} \circ P^{-1}. \tag{5.12}$$
Indeed, we have that $\gamma(t) = P \circ e^{tf} \circ P^{-1}(q)$ satisfies $\gamma(0) = q$ and
$$\dot\gamma(t_0) = \frac{d}{dt}\bigg|_{t=t_0} P(e^{tf} \circ P^{-1}(q)) = P_* f(t_0) \tag{5.13}$$

- If $R : M \to M$ is another smooth map, we have
$$\begin{aligned} R_* P_* f(q) &= \frac{d}{dt}\bigg|_{t=0} R(e^{tP_*f}(q)) \\ &= \frac{d}{dt}\bigg|_{t=0} R \circ P \circ (e^{tf}) \circ P^{-1}(q) \\ &= \frac{d}{dt}\bigg|_{t=0} R \circ P \circ (e^{tf}) \circ (R \circ P)^{-1}(R(q)) \\ &= (R \circ P)_* f(R(q)) \end{aligned} \tag{5.14}$$

The idea We now delve into the proof, starting by some heuristics.

Let us denote by $\mathcal{P}$ the family of diffeomorphisms that can be obtained via the family $\mathcal{F}$. That is,

$$\mathcal{P} = \left\{ e^{t_k f_k} \circ \cdots \circ e^{t_1 f_1} \mid k \in \mathbb{N},\, f_1, \ldots, f_n \in \mathcal{F},\, t_1, \ldots, t_k > 0 \right\} \subset \mathrm{Diff}(M) \tag{5.15}$$

The intuition is that at the point $q \in M$ we can move either in the directions of $\mathcal{F}$, or in the directions that we can generate via these diffeomorphisms. Formally, this means that the candidate tangent space to the orbit at a point $q \in \mathcal{O}(q_{\mathrm{in}})$ is the set

$$\Pi_q = \mathrm{span}(\mathcal{P}_* \mathcal{F})_q := \mathrm{span}\,\{(P_* f)(q) \mid P \in \mathcal{P},\, f \in \mathcal{F}\} \subset T_q M. \tag{5.16}$$

Here, we use the push-forward $P_* : TM \to TM$, defined above, which encodes how a vector field is transformed under the action of the diffeomorphism P.

As we will see, it is not difficult to show that indeed, Π_q contains directions that are "tangent" to $\mathcal{O}_q$, in the sense that Π_q is the tangent space to a locally defined differentiable manifold $\Xi_q \subset \mathcal{O}_q$. The difficult part of the proof is to show that this set actually captures the whole of $\mathcal{O}_q$, i.e., that $X_q := \bigcup_{q' \in \mathcal{O}_q} \Xi_{q'} = \mathcal{O}_q$.

To do this one goes back to an idea already employed in the proof of Chow–Rashevskii Theorem, and considers the relation $q_1 \sim q_2$ if and only if $q_1 \in \mathcal{O}_{q_2}$. Since we are looking at orbits, which are symmetric by definition, this relation is an equivalence relation. As such M decomposes as the disjoint union of the disjoint orbits.

The key observation is then that exploiting the sets $\{\Xi_q\}_{q \in M}$ it is possible to define a new topology on M, which is finer than the original one, and such that X_q is automatically open-closed. Then, one shows that the connected components of the resulting topological space $M^{\mathcal{F}}$ are exactly the orbits, which entails that $X_q = \mathcal{O}_q$.

The proof We start by showing that Π_q has constant dimension.

Lemma 5.27. *It holds that* $\dim \Pi_q = \dim \Pi_{q_{\text{in}}}$ *for any* $q \in \mathcal{O}(q_{\text{in}})$.

Proof. Since $q \in \mathcal{O}(q_{\text{in}})$, there exists a diffeomorphism $Q \in \mathcal{P}$ such that $Q(q_{\text{in}}) = q$. We claim that

$$Q_*^{-1} \Pi_q \subset \Pi_{q_{\text{in}}}.$$

Observe that this implies that $\dim \Pi_q \leq \dim \Pi_{q_{\text{in}}}$. The statement will then follow by exchanging the role of q and q_{in}.

Let $P \in \mathcal{P}$ and $f \in \mathcal{F}$, so that $(P_* f)(q) \in \Pi_q$, and, using (5.14), compute

$$Q_*^{-1}[(P_* f)(q)] = [Q_*^{-1} \circ P_* f](Q^{-1} q) = [((Q^{-1} \circ P)_* f)](q_{\text{in}}) \in (\mathcal{P}_* \mathcal{F})_{q_{\text{in}}} \subset \Pi_{q_{\text{in}}}.$$

This completes the proof of the claim and of the statement. □

Thanks to the previous result, we define m as the common dimension of the sets Π_q.

We now want to construct a local parametrization of orbits $\mathcal{O}(q_{\text{in}})$ for any $q_{\text{in}} \in M$. To this aim, fix $q \in M$ and pick $g_1, \ldots, g_m \in \mathcal{P}_* \mathcal{F}$ such that

$$\Pi_q = \operatorname{span}\{g_1(q), \ldots, g_m(q)\}.$$

Define the map $\xi : \mathbb{R}^m \to M$ as follows

$$\xi_q(x) = e^{x_m g_m} \circ \cdots \circ e^{x_1 g_1}(q).$$

We have the following.

Lemma 5.28. *The map ξ_q is a local submersion at $x = 0$. More precisely, there exists a neighborhood of origin $V_q \subset \mathbb{R}^m$ such that $\Xi_q := \xi_q(V_q) \subset M$ is an immersed submanifold and, moreover, the following hold:*

- *The map ξ_q locally parametrizes the orbit of q, i.e., $\Xi_q \subset \mathcal{O}_q$;*

- *It holds that $T_p\Xi_q = \Pi_p$ for all $p \in \Xi_q$. This, in particular, implies that the definition of ξ_q is independent on the choice of the basis $\{g_1, \ldots, g_m\}$.*

Proof. Clearly $\xi_q(0) = q$. Moreover, for any $i = 1, \ldots, m$ it holds

$$\left.\frac{\partial \xi_q}{\partial x_i}\right|_{x=0} = (e^{x_m g_m} \circ \cdots \circ e^{x_{i+1} g_{i+1}})_* g_i(e^{x_i g_i} \circ \cdots \circ e^{x_1 g_1}(q))\Big|_{x=0} = g_i(q). \tag{5.17}$$

In particular, since $\{g_1(q), \ldots, g_m(q)\}$ are linearly independent, we have that $\left.\frac{\partial \xi_q}{\partial x}\right|_{x=0}$ has maximal rank m.

Letting $V_q \subset \mathbb{R}^m$ be a neighborhood of the origin such that $\xi_q|_{V_q}$ is a submersion, we have that $\Xi_q := \xi_q(V_q) \subset M$ is an immersed submanifold of M.

Let us show that $\Xi_q \subset \mathcal{O}_q$. We start by showing that $\xi_q(x_1, 0, \ldots, 0) \in \mathcal{O}_q$. Indeed, recall that $g_1 \in \mathcal{P}_*\mathcal{F}$, i.e., there exists $P \in \mathcal{P}$ (i.e., an admissible motion w.r.t. $\mathcal{F}$) and $f \in \mathcal{F}$ such that $g_1 = P_* f$. Then,

$$\xi_q(x_1, 0, \ldots, 0) = e^{x_1 g_1}(q) = e^{x_1 P_* f}(q) = P \circ e^{x_1 f} \circ P^{-1}(q) \in \mathcal{O}_q.$$

Here, we used (5.12). Since $e^{x_1 g_1}(q) \in \mathcal{O}_q$, we can repeat the above argument to show that $\xi_q(x_1, x_2, 0, \ldots, 0) \in \mathcal{O}_q$. Proceeding in this fashion proves the claim.

We are left to prove that $T_p\Xi_q = \Pi_p$ for all $p \in \Xi_q$. Observe that for $x \in V_q$ such that $p = \xi_q(x)$ it holds that $\frac{\partial \xi_q}{\partial x}$ has maximal rank m. Hence, $\dim T_p\Xi_q = m$ for any $p \in \Xi_q$. In particular, this dimension coincides with the dimension of Π_p, and thus we just need to prove that

$$\left.\frac{\partial \xi_q}{\partial x_i}\right|_x \in \Pi_{\xi_q(x)}, \qquad \forall x \in V_q.$$

The same computation as in (5.17) yields that

$$\left.\frac{\partial \xi_q}{\partial x_i}\right|_x = R_* g_i(e^{x_i g_i} \circ S(q)), \qquad \text{where we let} \qquad \begin{aligned} R &= e^{x_m g_m} \circ \cdots \circ e^{x_{i+1} g_{i+1}} \\ S &= e^{x_{i-1} g_{i-1}} \circ \cdots \circ e^{x_1 g_1}. \end{aligned}$$

Then, we conclude by computing

$$\left.\frac{\partial \xi_q}{\partial x_i}\right|_x = (R_* g_i)(R \circ e^{x_i g_i} \circ S(q)) = (R_* g_i)(\xi_q(x)) \in (\mathcal{P}_*\mathcal{F})_p \subset \Pi_p. \qquad \square$$

The above result suggests that a candidate differentiable structure on $\mathcal{O}_{q_{\text{in}}}$, $q_{\text{in}} \in M$ should be given by the atlas[6]

$$\mathcal{U} = \{(\xi_q(V), \xi_q^{-1})\}_{\substack{q \in \mathcal{O}(q_{\text{in}}) \\ V \subset V_q}}. \tag{5.18}$$

Indeed, the smoothness of the transitions maps

$$\xi_{\hat{q}}^{-1} \circ \xi_q : \xi_q^{-1}(W) \to \xi_{\hat{q}}^{-1}(W), \qquad \text{if } W := \xi_q(V) \cap \xi_{\hat{q}}(\hat{V}) \neq \emptyset, \tag{5.19}$$

[6]This is not maximal, of course, but it is easily completed to a maximal one via Zorn's Lemma.

is immediate, by the previous Lemma. Hence, letting

$$X_{q_{\rm in}} := \bigcup_{\substack{q\in\mathcal{O}_{q_{\rm in}}\\ V\subset V_q}} \xi_q(V), \tag{5.20}$$

we have that $X_{q_{\rm in}} \subset \mathcal{O}_{q_{\rm in}}$ is an m-dimensional manifold such that $T_q X = \Pi_q$.

We are left with the most difficult task, now: prove that indeed $X_{q_{\rm in}} = \mathcal{O}_{q_{\rm in}}$. To this aim, we need to exploit the fact that the orbits define a partition of the manifold M, via the equivalence relation $q_1 \sim q_2$ if and only if $q_1 \in \mathcal{O}_{q_2}$. The argument of proof goes as follows. We consider the covering family

$$\mathfrak{T} := \{\xi_q(V) \mid q \in M,\, V \subset V_q \text{ is open }\}. \tag{5.21}$$

We will show that $\mathfrak{T}$ is a basis for a topology on M (this is done in Lemma 5.29 below) whose connected components are exactly the orbits (this is Lemma 5.30). Since, by construction, $X_{q_{\rm in}} \subset \mathcal{O}(q_{\rm in})$ is open-closed in this topology, this yields the statement.

Lemma 5.29. *The family $\mathfrak{T}$ is a basis for a topology*[7] *on M.*

Proof. Let us show that for any $q \in M$ and any $\hat q \in \xi_q(V)$ there exists an arbitrarily small neighborhood of the origin $\hat V \subset V_{\hat q}$ such that $\xi_{\hat q}(\hat V) \subset \xi_q(V)$. The result will then follow.

Since $\hat q \in \xi_q(V)$, there exists $\hat x \in V$ such that $\xi_q(\hat x) = \hat q$. Then, since V is open, there exists a ball $B_\delta(\hat x) \subset V$ centered at $\hat x$ and of radius $\delta > 0$. The image $\xi_q(B_\delta(\hat x))$ is then a neighborhood of $\hat q$ in $\xi_q(V)$ and, up to reducing the radius δ, we can assume that $\xi_q(B_\delta(\hat x)) \subset \xi_{\hat q}(V_{\hat q})$. To conclude the proof it suffices to consider $\hat V = xi_{\hat q}^{-1}(\xi_q(B_\delta(\hat x)))$. Indeed, by construction $\hat V \subset V_{\hat q}$ and $\xi_{\hat q}(\hat V) \subset \xi_q(V)$, as required. Moreover, letting $\delta \to 0$ we can arbitrarily shrink $\hat V$. □

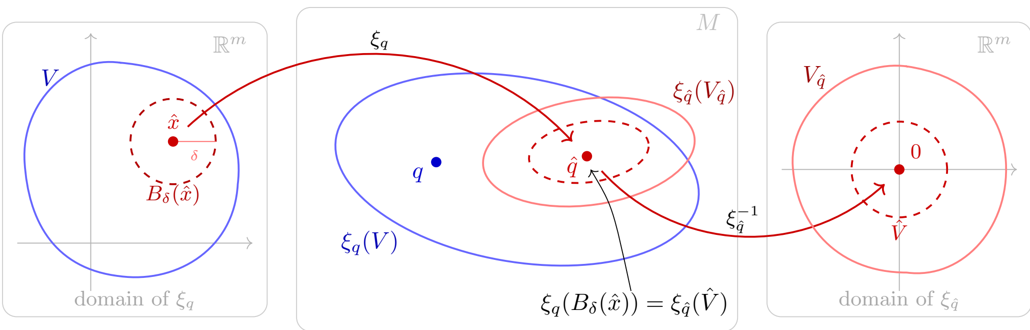


Figure 5.5.: Illustration of Lemma 5.29.

[7] The family $\mathcal{T}$ is a basis for a topology on M if and only if it is a covering of M for any $B_1, B_2 \in \mathfrak{T}$ and $p \in B_1 \cap B_2$ there exists $B_3 \in \mathfrak{T}$ such that $p \in B_3 \subset B_1 \cap B_2$. In this case, $\mathfrak{T}$ the associated topology is the smallest one containing $\mathfrak{T}$.

Lemma 5.30. *Let $M^{\mathcal{F}}$ be the topological space defined by $(M, \mathfrak{T})$. The orbits $\{\mathcal{O}_q\}_{q\in M}$ are exactly the connected components of $M^{\mathcal{F}}$.*

Proof. By Lemma 5.28, any point $\hat{q} \in \mathcal{O}_q$ is contained in $\Xi_{\hat{q}} \subset \mathcal{O}_{\hat{q}} = \mathcal{O}_q$. Since $\Xi_{\hat{q}} \in \mathfrak{T}$, it follows that $\mathcal{O}_q$ is open in $M^{\mathcal{F}}$. Since orbits are disjoint, we have that

$$\mathcal{O}_q = M \setminus \bigcup_{q'\neq q} \mathcal{O}_{q'}.$$

Hence, $\mathcal{O}_q$ is also closed.

We are left to prove that $\mathcal{O}_q$ is connected in $M^{\mathcal{F}}$. By definition of orbit (cf. (5.9)), it suffices to show that $t \mapsto e^{tf}(q)$ is continuous from $\mathbb{R}$ to $M^{\mathcal{F}}$ for any $q \in M$ and $f \in \mathcal{F}$. Consider $q \in M$ and $V \subset V_q$ such that there exists t_0 such that $e^{t_0 f}(q) \in \xi_q(V)$. Since $f \in \mathcal{F} \subset \mathcal{P}_*\mathcal{F}$, we have that the curve $\gamma(s) := e^{(t_0+s)f}(q)$ is tangent to $\Pi_{\gamma(s)}$ for any $s \in \mathbb{R}$. By Lemma 5.28 this implies that $\{s \in \mathbb{R} \mid \gamma(s) \in \xi_q(V)\}$ is an open interval, showing the required continuity. ☐

Summing up, we have shown the following.

Lemma 5.31. *For any $q_{\text{in}} \in M$, the orbit $\mathcal{O}(q_{\text{in}})$ is a manifold such that*

$$T_q\mathcal{O}(q_{\text{in}}) = \Pi_q.$$

Relations with the Lie algebra $\operatorname{Lie}\mathcal{F}$ We have the following immediate result.

Lemma 5.32. *It holds that $\operatorname{Lie}_q \mathcal{F} \subset T_q\mathcal{O}_q$.*

Proof. By Lemma 5.4 and Lemma 2.23 it follows that elements of $\operatorname{Lie}\mathcal{F}$ are tangent to the orbits. The statement follows by Corollary 2.25. ☐

The fact that $\operatorname{Lie}_q \mathcal{F} = T_q\mathcal{O}_q$ in the analytic or in the constant rank cases, follows from the following result, which is proven in [AS04, Lemma 5.2].

Lemma 5.33. *Assume that $\operatorname{Lie}\mathcal{F} \subset \operatorname{Vec}(M)$ is locally finitely generated as a $\mathcal{C}^\infty$-submodule*[8]*. Then,*

$$\mathcal{P}_*\mathcal{F} = \mathcal{F} \tag{5.22}$$

To conclude, one observes that the constant rank assumption trivially implies that $\operatorname{Lie}\mathcal{F}$ is locally finitely generated. The fact that analyticity implies the same fact is more involved and relies on the Nötherian property of the ring of germs of analytic functions [Mal66].

[8]That is, locally there always exists $f_1, \dots, f_k \in \operatorname{Lie}\mathcal{F}$ such that

$$\operatorname{Lie}\mathcal{F} = \left\{ \sum_{i=1}^{k} a_i f_i \mid a_i \in \mathcal{C}^\infty \right\}.$$

6. Controllability via compatible vector fields

In the previous chapter we proved the Chow–Rashevskii Theorem, which states the controllability of control systems associated with Lie bracket generating *symmetric* families of vector fields.

When a family of vector fields is Lie bracket generating but it is not symmetric, in general it is not easy to understand if the system is controllable or not. To bypass this problem, in this chapter we develop the theory of *compatible vector fields*.

This technique further develop the intuition on which the geometric theory of controllability is built on. Indeed, the Chow–Rashevskii Theorem can be interpreted as saying that, since the Lie brackets can be approximated by the flow of the vector fields in $\mathcal{F}$, it is as if these where actually present in $\mathcal{F}$. In the following we will generalize this idea by formally defining what it means that certain directions can be approximated, and showing that if this is the case, then they can be integrated to $\mathcal{F}$ for the sake of testing its controllability.

Definition 6.1. A vector field g is *compatible* with the family $\mathcal{F}$ if for every $q_{\text{in}} \in M$, every $q_{\text{in}} \in M$, and every $t \geq 0$, the point $e^{tg}(q_{\text{in}})$ is contained in the closure of $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$.

A compatible extension of $\mathcal{F}$ is a family $\hat{\mathcal{F}} \supset \mathcal{F}$ whose every element is compatible with $\mathcal{F}$.

Observe that, trivially: every vector field in $\mathcal{F}$ is compatible with $\mathcal{F}$, and if $\mathcal{F}$ is Lie bracjet generating, then so is every compatible extension of $\mathcal{F}$.

The main result of the theory of compatible vector fields is the following.

Theorem 6.2. *If $\mathcal{F}$ is a Lie bracket generating family of vector fields and $\hat{\mathcal{F}}$ is a compatible extension of $\mathcal{F}$, then*

$$\mathcal{R}^{\hat{\mathcal{F}}}(q_{\text{in}}) = M \quad \forall q_{\text{in}} \in M \iff \mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = M \quad \forall q_{\text{in}} \in M. \tag{6.1}$$

This theorem is typically put to use in the following way: if we want to test the controllability of $\mathcal{F}$, we can try to find a compatible extension $\hat{\mathcal{F}}$ of $\mathcal{F}$ that is symmetric. If we succeed, then $\mathcal{F}$ is controllable by the Chow–Rashevskii Theorem.

Proof of Theorem 6.2. By Lemma 5.3, it suffices that for any $q_{\text{in}} \in M$ it holds that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is dense in M. To this end, let us prove that $\mathcal{R}^{\hat{\mathcal{F}}}(q_{\text{in}}) \subset \overline{\mathcal{R}^{\mathcal{F}}(q_{\text{in}})}$. Since by assumption $\mathcal{R}^{\hat{\mathcal{F}}}(q_{\text{in}}) = M$, this will conclude the proof.

Fix $q \in \mathcal{R}^{\hat{\mathcal{F}}}(q_{\text{in}})$. Then there exists $X_1, \ldots, X_k \in \hat{\mathcal{F}}$ and $t_1, \ldots, t_k > 0$ such that

$$q = e^{t_k X_k} \circ \cdots \circ e^{t_1 X_1}(q_{\text{in}}).$$

Let us prove by induction on k that q belongs to the closure of $\mathcal{R}^{\mathcal{F}}(q_{\rm in})$. For $k=1$ the conclusion follows by definition of compatible vector fields. For $k>1$, it follows from the induction hypothesis that $\hat{q} = e^{t_{k-1}X_{k-1}} \circ \cdots \circ e^{t_1 X_1}(q_{\rm in})$ can be obtained as the limit of a sequence $(q_n)_{n\in\mathbb{N}}$ of points in $\mathcal{R}^{\mathcal{F}}(q_{\rm in})$. Moreover, since X_k is compatible with $\mathcal{F}$, there exists a sequence $(P_j)_{j\in\mathbb{N}}$ is in the semigroup of diffeomorphisms $\{e^{s_\ell f_\ell}\circ\cdots\circ e^{s_1 f_1} \mid \ell\in\mathbb{N},\ f_1,\dots,f_\ell\in\mathcal{F},\ s_1,\dots,s_k\geq 0\}$ such that $P_j(\hat{q})\to q$ as $j\to\infty$. By continuity of P_j, $\lim_{n\to\infty} P_j(q_n) = P_j(\hat{q})$ for every $j\in\mathbb{N}$. By a diagonal argument, there exists a subsequence $(q_{n_j})_{j\in\mathbb{N}}$ of $(q_n)_{n\in\mathbb{N}}$ such that $P_j(q_{n_j})\to q$ as $j\to\infty$. Since each point $P_j(q_{n_j})$ is in $\mathcal{R}^{\mathcal{F}}(q_{\rm in})$, this concludes the proof. □

Remark 6.3. It is useful to notice that the compatibility criteria can be used "in cascade", in the sense that, given $\mathcal{F},\mathcal{F}',\mathcal{F}''\subset \mathrm{Vec}(M)$, if $\mathcal{F}''$ is a compatible extension of $\mathcal{F}'$ and $\mathcal{F}'$ is a compatible extension of $\mathcal{F}$, then $\mathcal{F}''$ is a compatible extension of $\mathcal{F}$. This means that for checking the compatibility of a new vector field, we can use instead of $\mathcal{F}$ the family of all compatible vector fields already identified.

Remark 6.4. By the previous remark, the compatible extension property induces a partial order on the family of vector fields. In particular, if $\mathcal{F}$ is Lie bracket generating, then there exists a unique maximal compatible extension $\overline{\mathcal{F}}$ of $\mathcal{F}$. Theorem 6.2 implies that $\mathcal{F}$ is controllable if and only if $\overline{\mathcal{F}}$ is controllable. Moreover, observe that if $\mathcal{F}$ is controllable, then $\mathrm{Vec}(M)$ is a compatible extension of $\mathcal{F}$. Hence, $\mathcal{F}$ is controllable if and only if $\overline{\mathcal{F}} = \mathrm{Vec}(M)$.

The rest of this chapter is devoted to presenting some criteria for identifying compatible vector fields and deduce from them some important controllability result obtained using Theorem 6.2.

6.1. Compatibility by positive multiplication

Proposition 6.5. *Let $f\in\mathrm{Vec}(M)$ and $\varphi: M\to(0,+\infty)$ be a smooth function. Then φf is compatible with $\{f\}$.*

Proof. It suffices to show that the integral curves of φf are reparameterizations of those of f. To this aim, let $q:[0,T]\to M$ be a solution of $\dot{q}(t) = f(q(t))$ and set $\tilde{q}(t) = q(\eta(t))$ for some smooth increasing function $\eta:\mathbb{R}\to\mathbb{R}$, to be determined. Then,

$$\dot{\tilde{q}}(t) = \frac{d}{dt} q(\eta(t)) = \dot{q}(\eta(t))\eta'(t) = f(\tilde{q}(t))\eta'(t).$$

Choosing η as the unique solution of the Cauchy problem $\eta'(t) = \varphi(q(\eta(t)))$ with $\eta(0)=0$, we conclude that $\tilde{q}$ is a solution of $\dot{\tilde{q}}(t) = \varphi(\tilde{q}(t))f(\tilde{q}(t))$ whose support coincide with that of q, i.e., $\tilde{q}([0,\eta(T)]) = q([0,T])$. □

6.2. Compatibility by uniform convergence

Proposition 6.6. *Let $g\in\mathrm{Vec}(M)$ and $(f_k)_{k\in\mathbb{N}}$ be a sequence in $\mathrm{Vec}(M)$ such that $f_k\to g$ uniformly on compact sets as $k\to\infty$. Then, g is compatible with $\{f_k \mid k\in\mathbb{N}\}$.*

Proof. This follows from the standard fact that

$$\lim_{k\to\infty} e^{tf_k}(q) = e^{tg}(q), \qquad \forall q \in M, \forall t \geq 0. \tag{6.2}$$

This is easily proved in coordinates by letting $\gamma_k(t) = e^{tg}(q) - e^{tf_k}(q)$, and noticing that γ_k is a solution of the Cauchy problem

$$\begin{cases} \dot\gamma_k(t) = g(e^{tg}(q)) - f_k(e^{tf_k}(q)) \\ \gamma_k(0) = 0 \end{cases}$$

The uniform convergence of f_k to g implies that $\dot\gamma_k \to 0$ uniformly on compact sets, and hence $\gamma_k \to 0$ uniformly on compact sets, which proves (6.2). □

As a consequence of this compatibility criterion, we can deduce the following proposition, stating that an affine control system (cf. Section 5.1.1) is controllable if the controls are unbounded and if it is not necessary to use the drift to get a Lie algebra of full dimension at every point.

Theorem 6.7 (Controllability via strong Lie bracket generating property)**.** *Let $f_0, f_1 \dots, f_m$ be smooth vector fields on M, and consider the control system*

$$\dot q = f_0(q) + \sum_{i=1}^m u_i f_i(q), \qquad (u_1(\cdot), \dots, u_m(\cdot)) : [0,\infty) \to \mathbb{R}^m. \tag{6.3}$$

If the family $\{f_1, \dots, f_m\}$ is Lie bracket generating then (6.3) *is controllable.*

Remark 6.8. For an control system of the type (6.3), the condition that $\{f_1, \dots, f_m\}$ is Lie bracket generating is called the *strong Lie bracket generating condition.*

Proof. First notice that as a consequence of Exercise 2, being $\{f_1, \dots, f_m\}$ Lie bracket generating, then (6.3) is Lie bracket generating as well. Let $\mathcal{F}$ be the family of vector fields associated with (6.3). In the following we are going to prove that for every $(v_1, \dots, v_m) \in \mathbb{R}^m$ the vector field $\sum_{i=1}^m v_i f_i$ is compatible with $\mathcal{F}$. Once this is done, the controllability of (6.3) follows, since the family $\mathcal{F} \cup \{\sum_{i=1}^m v_i f_i \mid v_1, \dots, v_m \in \mathbb{R}\}$ contains a symmetric and Lie bracket generating sub-family (cf. Remark 5.13).

To show that $\sum_{i=1}^m v_i f_i$ is compatible with $\mathcal{F}$ for every $v_1, \dots, v_m \in \mathbb{R}$, remark that

$$g_n := \frac{1}{n}\left(f_0 + \sum_{i=1}^m (nv_i) f_i\right) \longrightarrow \sum_{i=1}^m v_i f_i \quad \text{uniformly on compact sets as } n \to \infty.$$

Moreover, every vector field g_n is compatible with $\mathcal{F}$ by Proposition 6.5 (it is the positive multiple of a vector field in $\mathcal{F}$). Hence, $\sum_{i=1}^m v_i f_i$ is compatible with $\mathcal{F}$ by Proposition 6.6. □

Remark 6.9. Under the strong bracket generating condition, system (6.3) is actually controllable in arbitrarily small time. Indeed, the trajectories of the controllable system $\dot{q} = \sum_{i=1}^{m} u_i f_i(q)$ can be reparameterized in such a way to reach their target in an arbitrarily small time. Hence, according to (6.2) and the convergence recalled in (6.2), system (6.3) is approximately controllable in arbitrarily small time. By following the proof of Corollary 5.18, one can actually deduce that (6.3) is controllable in arbitrarily small time.

6.3. Compatibility by convexification

A very useful application of the theory of compatible vector fields is based on the following result stating that a convex combination of vector fields of $\mathcal{F}$ is compatible with $\mathcal{F}$. It formalize the intuition that if one commutes quickly between the dynamics of two vector fields f and g, and one stays the same time on each dynamics, then the corresponding trajectory is close the trajectory described by the flow of $\frac{f+g}{2}$.

Proposition 6.10. *For every $\lambda_1, \dots, \lambda_k \geq 0$ and $f_1, \dots, f_k \in \mathcal{F}$, the vector field $\lambda_1 f_1 + \cdots + \lambda_k f_k$ is compatible with $\mathcal{F}$.*

The proof of this statement relies of the following averaging result.

Lemma 6.11 (Averaging)**.** *Let $f_1, \dots, f_k \in \mathrm{Vec}(M)$ and $T > 0$. Consider a bounded sequence $(u^n)_{n\in\mathbb{N}}$ in $L^\infty([0,T], \mathbb{R}^k)$. Assume that there exists $u \in L^\infty([0,T], \mathbb{R}^k)$ such that*

$$\int_0^t u^n(s)ds \overset{n\to\infty}{\longrightarrow} \int_0^t u(s)ds \qquad \text{for all } \ t \in [0,T]. \tag{6.4}$$

Then, denoting by $q(t; q_{\mathrm{in}}, v(\cdot))$ the solutions to

$$\begin{cases} \dot{q} = \displaystyle\sum_{i=1}^{k} u_i(t) f_i(q), \\ q(0) = q_{\mathrm{in}}, \end{cases}$$

we have that

$$q(t; q_0, u^n) \to q(t; q_0, u) \quad \text{uniformly w.r.t. } t \in [0,T] \text{ as } n \to \infty.$$

Proof. Up to splitting the curve $t \mapsto q(t; q_{\mathrm{in}}, u)$ in a finite number of pieces, we can assume it to be fully contained in a compact chart. We also let $q(t) := q(t; q_{\mathrm{in}}, u)$ and $q^n(t) := q(t; q_{\mathrm{in}}, u^n)$, so that, in coordinates,

$$q(t) = q_0 + \sum_{i=1}^{k} \int_0^t u_i(s) f_i(q(s))ds, \qquad q^n(t) = q_0 + \sum_{i=1}^{k} \int_0^t u_i^n(s) f_i(q^n(s))ds.$$

Observe that the second equation holds for n large enough, since we need $q^n(\cdot)$ to stay in the same coordinate chart as $q(\cdot)$. Moreover, since the vector fields f_i are Lipschitz in the coordinate patch (say with constant $L > 0$) and the controls u^n are bounded (say by $K > 0$), we have that there exists $c > 0$ such that

$$\|q^n(t) - q_{\text{in}}\| \leq KLt, \qquad \forall t \in [0, T].$$

Then, we have that for every $t \in [0, T]$,

$$\begin{aligned}
\xi^n(t) :=& \|q^n(t) - q(t)\| \\
\leq& \sum_{i=1}^{k} \left\| \int_0^t u_i^n(s)(f(q^n(s)) - f(q(s)))ds \right\| + \sum_{i=1}^{k} \left\| \int_0^t (u_i^n(s) - u_i(s)) f_i(q(s))ds \right\| \\
\leq& kKL \int_0^t \|q^n(s) - q(s)\| ds + \sum_{i=1}^{k} \left\| \int_0^t \left(\frac{d}{ds} \int_0^s (u_i^n(\tau) - u_i(\tau))d\tau \right) f_i(q(s))ds \right\|.
\end{aligned}$$

Integrating by parts and thanks to (6.4), we have that, for every $i = 1, \ldots, k$,

$$\int_0^t (u_i^n(s) - u_i(s)) f_i(q(s))ds = \int_0^t \left(\frac{d}{ds} \int_0^s (u_i^n(\tau) - u_i(\tau))d\tau \right) f(q(s)) \longrightarrow 0,$$

uniformly for $t \in [0, T]$, as $n \to +\infty$. Here, we use the fact that the convergence in (6.4) is uniform with respect to $t \in [0, T]$ because $\{\|u^n\|_\infty \mid n \in \mathbb{N}\}$ is bounded. Hence, for every $\varepsilon > 0$ and for n large enough, we have

$$\xi^n(t) \leq kKL \int_0^t \xi^n(s)ds + \varepsilon.$$

The conclusion can be deduced from Gröwnwall's lemma. We nevertheless present an explicit explicit estimates. Define $c = kKL$ and $w^n(t) = e^{-ct} \int_0^t \xi^n(s)ds$. Then,

$$\frac{d}{dt} w^n(t) \leq e^{-ct}\varepsilon \implies w^n(t) \leq \frac{1 - e^{-ct}}{c}\varepsilon.$$

Hence,

$$\int_0^t \xi^n(s)ds \leq \frac{e^{ct} - 1}{c}\varepsilon \leq \frac{e^{cT} - 1}{c}\varepsilon.$$

Since the functions $\xi^n(\cdot)$ are nonnegative and uniformly Lipschitz continuous with respect to n, by arbitrariness of $\varepsilon > 0$ we deduce that $\xi^n(t) \to 0$ as $n \to \infty$, uniformly with respect to $t \in [0, T]$. □

To obtain the result it suffices to apply the above lemma to piecewise constant controls u^n with values in the canonical basis of $\mathbb{R}^k$, switching faster and faster as $n \to \infty$.

Proof of Lemma 6.10. Let $g = \lambda_1 f_1 + \ldots + \lambda_k f_k$ and $t > 0$. Then, $e^{Tg}(q_{\text{in}})$ is the solution at time T of the control system

$$\dot{q} = \sum_{i=1}^{k} u_i(t) f_i(q), \qquad q(0) = q_{\text{in}} \tag{6.5}$$

with control $u_i(t) = \lambda_i$ for $t \in [0, T]$.

Let us consider the following controls $u^n \in L^\infty([0,T], \mathbb{R}^k)$ defined as follows. Each u^n is periodic of period T/n, i.e., it suffices to define it on the interval $[0, T/k]$. Here, denoting by $\{e_1, \dots, e_k\}$ the canonical basis of $\mathbb{R}^k$, we let

$$u^n(t) = kn\lambda_i e_\ell \qquad \text{for } t \in [\ell T/(kn), (\ell+1)T/(kn)], \qquad \ell = 0, \dots, k-1. \tag{6.6}$$

Let $q^n(\cdot)$ be the trajectory of system (6.5) with control u^n. Observe that the action of the control u^n w.r.t. system (6.5) can be written as

$$q^n(\ell T/n) = e^{\lambda_k f_k} \circ \cdots e^{\lambda_1 f_1}(q^n((\ell-1)T/n)), \qquad \ell = 1, \dots n. \tag{6.7}$$

Since $\lambda_i \geq 0$, we have that $\lambda_i f_i$ is compatible with $\{f_i\}$ by Proposition 6.5. By induction, this proves that every $q^n(T) \in \overline{\mathcal{R}(q_{\text{in}})}$.

To complete the proof, we are left to prove that $q^n(T) \to e^{Tg}(q_{\text{in}})$. By definition of u^n it follows that for any $\ell = 1, \dots, n$ it holds

$$\int_0^{T/\ell} u_i^n(s)\, ds = \lambda_i t. \tag{6.8}$$

This allows to show that

$$\int_0^t u_i^n(s)\, ds \longrightarrow \lambda_i t = \int_0^t u_i(s)\, ds. \tag{6.9}$$

Then, by Lemma 6.11 we have $q^n(T) \to e^{Tg}(q_{\text{in}})$. This completes the proof. □

From Proposition 6.10, using the theory of compatible vector fields, one obtains the following result

Corollary 6.12. *Let $\{f_1, \dots, f_m\}$ be Lie bracket generating and assume that the convex hull* $\text{co}(U)$ *of $U \subset \mathbb{R}^m$ is a neighborhood of the origin. Then the system*

$$\dot{q} = \sum_{i=1}^m u_i f_i(q), \qquad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty) \to U,$$

is controllable.

Another corollary of Lemma 6.10 is the following weakened version of Proposition 6.7.

Corollary 6.13. *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i f_i(q), \qquad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty) \to U \subset \mathbb{R}^m. \tag{6.10}$$

If the convex hull of U is $\mathbb{R}^m$ *and if $\{f_1, \dots, f_m\}$ is Lie bracket generating then* (6.10) *is controllable.*

Other applications of Lemma 6.10 are given later (see Theorem 6.17 and Figure 6.3).

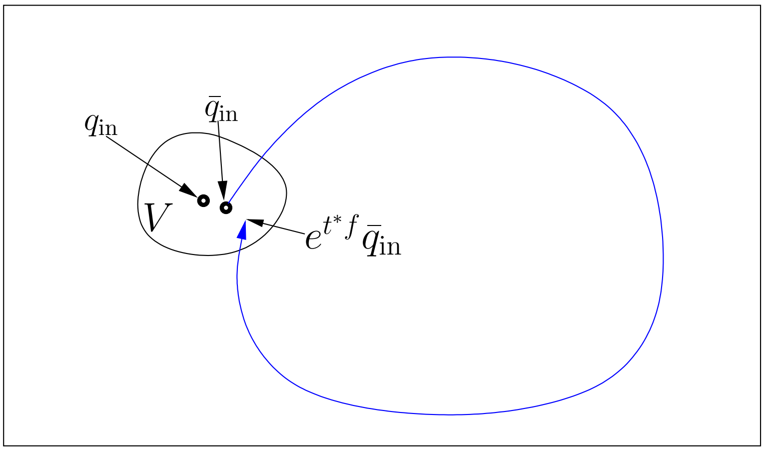


Figure 6.1.: Definition of recurrent vector field.

6.4. Compatibility by recurrence

In this section we apply the theory of compatible vector fields to affine control systems (cf. Section 5.1.1) that are Lie bracket generating and having a drift f_0 which is recurrent. We refer to Figure 6.1.

Definition 6.14 (Recurrent vector field)**.** A vector field f is said to be *recurrent* if for every open nonempty subset Ω of M, and every time $t > 0$, there exists $t^* > t$ such that $e^{t^* f}(\Omega) \cap \Omega \neq \emptyset$.

Remark 6.15. The definition given here is slightly more general than the one more commonly found, which requires that for every point $q_{\text{in}} \in M$, every neighborhood Ω of q_{in}, and every time $t > 0$, there exists $t^* > t$ such that $e^{t^* f}(q_{\text{in}}) \in \Omega$.

Notice that if the trajectories of f are periodic (possibly with period depending on the trajectory), then f is recurrent. We will discuss more general examples at the end of this section. First, let us show the relation between recurrent vector fields and the theory of compatible vector fields.

Lemma 6.16. *Assume that $\mathcal{F}$ is Lie bracket generating. If $f \in \mathcal{F}$ is recurrent, then $-f$ is compatible with $\mathcal{F}$.*

Proof. We have to prove that for every q_{in} and for every $t > 0$, $e^{-tf}q_{\text{in}}$ can be obtained as limit of points belonging to the reachable set $\mathcal{R}(q_{\text{in}}) = \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$.

We refer to Figure 6.2. By Krener theorem (thanks to the fact that $\mathcal{F}$ is Lie-bracket generating), there exists an arbitrarily small open set $W \subset \mathcal{R}(q_{\text{in}})$ such that $q_{\text{in}} \in \overline{W}$. Namely, in coordinates, for any $\varepsilon > 0$ we can pick $W \subset \mathcal{R}(q_{\text{in}}) \cap B(q_{\text{in}}, \varepsilon)$, where $B(q_{\text{in}}, \varepsilon)$ denotes the Euclidean ball.

The fact that e^{-tf} is a diffeomorphism and that $q_{\text{in}} \in \overline{W}$ imply that $e^{-tf}q_{\text{in}} \in \overline{e^{-tf}W}$. In particular the open set $e^{-tf}W$ is nonempty. Hence, by recurrence of f, there exists

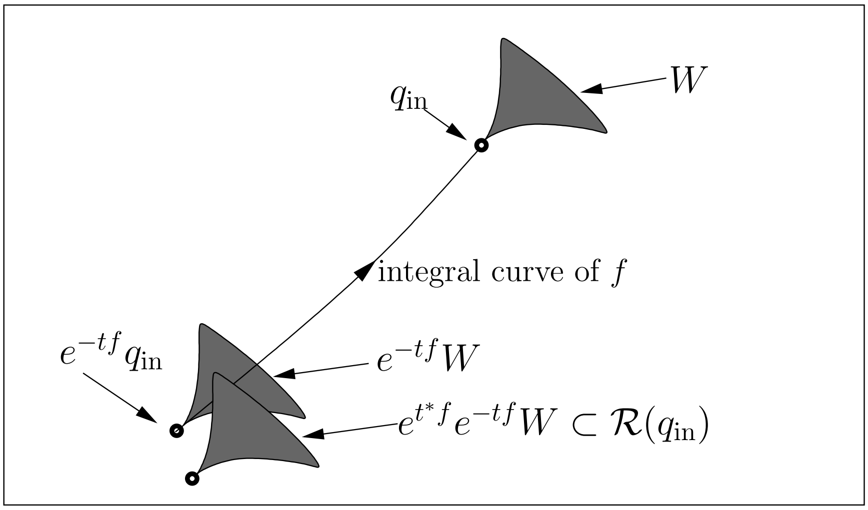


Figure 6.2.: Proof of Lemma 6.16.

$t^* > t$ such that

$$e^{t^*f}e^{-tf}W \cap e^{-tf}W \neq \emptyset,$$

or, equivalently,

$$e^{(t^*-t)f}W \cap e^{-tf}W \neq \emptyset.$$

But since $t^* - t > 0$ and $W \subset \mathcal{R}(q_{\text{in}})$ we have that $e^{(t^*-t)f}W \subset \mathcal{R}(q_{\text{in}})$. It then follows from the above that

$$\mathcal{R}(q_{\text{in}}) \cap e^{-tf}W \neq \emptyset.$$

Hence, since W was arbitrarily small, in any neighborhood of $e^{-tf}q_{\text{in}}$ there are points of $\mathcal{R}(q_{\text{in}})$. In other words $e^{-tf}q_{\text{in}} \in \overline{\mathcal{R}(q_{\text{in}})}$. □

As a consequence of the previous lemma we have the following.

Theorem 6.17. *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^{m} u_i(t) f_i(q), \qquad (u_1(\cdot), \ldots, u_m(\cdot)) : [0, \infty) \to U \subset \mathbb{R}^m. \tag{6.11}$$

*Assume that **(i)** 0 belongs to the interior of U, **(ii)** the control system* (6.11) *is Lie bracket generating, **(iii)** f_0 is recurrent. Then the system is controllable.*

Proof. Let $\mathcal{F}$ be the family of vector fields associated to 6.11. Notice that $f_0 \in \mathcal{F}$ since 0 belongs to U. Lemma 6.16 and Theorem 6.2 state the equivalence between the controllability of $\mathcal{F}$ and

$$\hat{\mathcal{F}} := \mathcal{F} \cup \{-f_0\} = \left\{ u_0 f_0 + \sum_{i=1}^{m} u_i f_i \mid u_0 \in \{-1, 1\}, (u_1, \ldots, u_m) \in U \right\}$$

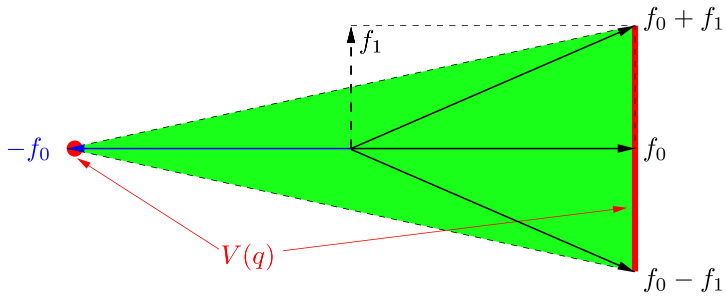


Figure 6.3.: For a control system of the type $f_0+u(t)f_1$ with $u(\cdot)$ taking values in $[-1,1]$, the picture represent the convexification of the set $V(q) = \{f_0(q) + uf_1(q) \mid u \in [-1,1]\} \cup \{-f_0(q)\}$ at a given point q. Notice that although $V(q)$ does not contain a symmetric set, its convexification does. For more general control systems as those of Corollary 6.11 the situation is exactly the same.

The conclusion follows from Corollary 6.12 (cf. see Figure 6.3). □

Remark 6.18. By using Lemma 6.10 in the proof of Theorem 6.17 it is clear that hypothesis **i)** can be weakened to merely asking that 0 belongs to the *convex hull* of U.

Example 6.19. On the sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\}$ consider the control system

$$\dot{x} = f_0(x) + uf_1(x), \qquad u(\cdot) : [0,\infty) \to (-1,1), \qquad x \in \mathbb{S}^2,$$

where

$$f_0(x) = \begin{pmatrix} -x_2 \\ x_1 \\ 0 \end{pmatrix}, \qquad f_1(x) = \begin{pmatrix} 0 \\ -x_3 \\ x_2 \end{pmatrix}.$$

Observe that the flows of f_0 and f_1 are rotations around the axes $(0,0,1)^T$ and $(1,0,0)^T$, respectively.

The previous results allows to conclude that this system is controllable since

- The system is Lie bracket generating: indeed the Lie bracket between f_0 and f_1 is given by

$$[f_0, f_1](x) = \begin{pmatrix} -x_3 \\ 0 \\ x_1 \end{pmatrix} \implies \dim(\text{span}\{f_0, f_1, [f_0, f_1]\}(x) = 2, \quad \forall x \in \mathbb{S}^2;$$

- The drift is recurrent: indeed the trajectories of f_0 are periodic;

Example 6.20 (Controlled pendulum). Consider the controlled pendulum system discussed in the Introduction. That is, we consider the control system on $M = \mathbb{S}^1 \times \mathbb{R}$ given by

$$\begin{pmatrix} \dot{\theta} \\ \dot{\omega} \end{pmatrix} = f_0(\theta, \omega) + u(t)f_1(\theta, \omega), \qquad u(\cdot) \in U := [-u_{\max}, u_{\max}].$$

Here,

$$f_0(\theta,\omega) = \begin{pmatrix} \omega \\ -\sin\theta \end{pmatrix}, \qquad f_1(\theta,\omega) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The system is Lie bracket generating. Indeed, it $\theta \neq \{0,\pi\}$, then f_0 and f_1 are linearly independent, and hence $\operatorname{span}\{f_0, f_1\}|_{(\theta,\omega)} = \mathbb{R}^2$. If $\theta \in \{0,\pi\}$, then f_0 and f_1 are linearly dependent. In this case, however, the Lie bracket $[f_0, f_1] = (0, \cos\theta)^T$ is linearly independent of f_0 and f_1. This shows that $\operatorname{span}\{f_0, f_1, [f_0, f_1]\}|_{(\theta,\omega)} = \mathbb{R}^2$ for every $(\theta,\omega) \in \mathbb{S}^1 \times \mathbb{R}$.

Since U is a neighborhood of the origin, to conclude that the system is controllable it suffices to show that f_0 is recurrent. The phase portrait of f_0 is depicted in Figure 1.1. In particular, we can distinguisgh two regions

$$M_1 = \{(\theta,\omega) \in \mathbb{S}^1 \times \mathbb{R} \mid \omega^2/2 - \cos\theta < 1\}, \qquad M_2 = \{(\theta,\omega) \in \mathbb{S}^1 \times \mathbb{R} \mid \omega^2/2 - \cos\theta > 1\},$$

such that all trajectories of f_0 starting in M_i, $i = 1, 2$, are periodic. That is, the only non-periodic trajectories are the heteroclinic trajectories connecting the unstable equilibrium points $(-\pi, 0)$ and $(\pi, 0)$.

In particular, f_0 is recurrent in M_1 and in M_2, which allows to conclude the controllability of the system in each of these two regions. Finally, the controllability of the system in the whole manifold M can be deduced by noticing that any initial condition the heteroclinic cycle $M \setminus (M_1 \cup M_2)$ can be steered to M_1 and to M_2 by using the control $u(\cdot)$.

Example 6.21 (Dubins car). The Dubins car is a modification of the car driven by two orthogonal fans introduced in Section 1.2. In this case, we assume that the car can only go forward with constant speed v ad that its steering radius is bounded by $\rho > 0$. Namely, for $v > 0$ fixed, the system reads

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = vX_1 + u(t)F_2, \qquad |u(\cdot)| \leq \frac{\rho}{v} \tag{6.12}$$

$$\text{where} \qquad X_1(x,y,\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad \text{and} \quad X_2(x,y,\theta) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{6.13}$$

We already saw in Example 5.15 that $\dim \operatorname{Lie}_q\{X_1, F_2\} = 3$ at all points $q \in \mathbb{R}^2 \times \mathbb{S}^1$. Moreover, the control set $U = [-\rho/v, \rho/v]$ is a neighborhood of the origin. However, the vector field X_2 is clearly not recurrent, since its orbit through (x_0, y_0, θ_0) is the line $\{(t\cos\theta_0 + x_0, t\sin\theta_0 + y_0, \theta) \mid t \in \mathbb{R}\}$.

Nevertheless, we can apply Theorem 6.17. Indeed, for any $0 < \bar{u} < v/\rho$, we can recast system (6.12) as

$$\dot{q} = f_0(q) + \tilde{u} f_1(q), \qquad u(\cdot) \in [-v/\rho - \bar{u}, v/\rho - \bar{u}],$$

where $f_0(q) = X_1(q) + \bar{u}F_2(q)$, and $f_1(q) = F_2$. It is immediate to check that f_0 is recurrent, since its orbits are circles of radius $v/\bar{u} > \rho$. Hence, system (6.12) is controllable by Theorem 6.17.

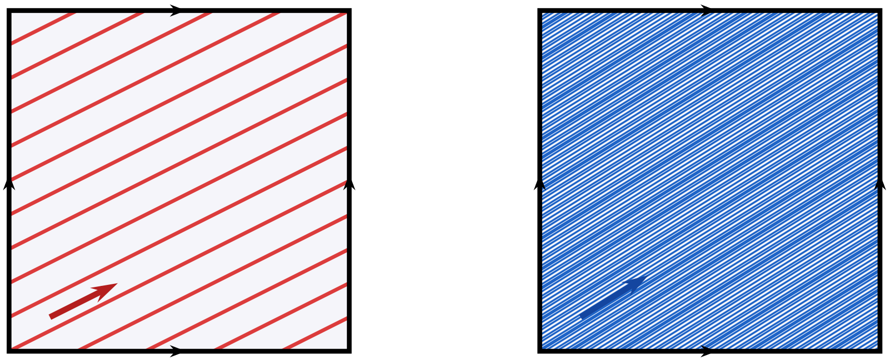


Figure 6.4.: The vector field f on the torus $\mathbb{R}^2/\mathbb{Z}$ is recurrent, even if its trajectories are not periodic.

6.4.1. Poincaré recurrence theorem

As we observed, vector fields with nonfixed periodic orbits are recurrent. Let us start by presenting an example of recurrent vector field with non-periodic orbits.

Example 6.22. Let $M = \mathbb{R}^2/\mathbb{Z}$ be the 2-dimensional torus and let f be the constant vector field on M given by $f(x, y) = (\alpha, \beta)$, where $\alpha, \beta \in \mathbb{R}$. We have that

$$e^{tf}(x, y) = (x + \alpha t, y + \beta t) \mod \mathbb{Z}.$$

In particular, one can show that the following dichotomy holds (see Figure 6.4):

- If $\alpha/\beta \in \mathbb{Q}$, then the trajectories of f are periodic, and hence f is recurrent.
- If $\alpha/\beta \notin \mathbb{Q}$, then the trajectories of f are dense in the torus, and hence f is recurrent as well.

The previous example depicts a typical situation for 2-dimensional compact manifolds. Indeed, in this case, trajectories of a vector field are either fixed points, nonfixed periodic orbits, or dense in the manifold (which then is homeomorphic to the torus). See, [Sch63].

Let us present a criterion for recurrence which covers the preceding example and other more general ones.

Definition 6.23. Let M be a manifold endowed with a measure μ. A vector field f is said to be *volume-preserving w.r.t.* μ if the flow of f preserves μ, i.e., if

$$\mu(e^{tf}(\Omega)) = \mu(\Omega) \quad \text{for every measurable set } \Omega \subset M \text{ and every } t \in \mathbb{R}.$$

Indeed, we have the following.

Theorem 6.24 (Poincaré recurrence theorem). *Let M be a manifold endowed with a finite measure μ with full support*[1], *and let $f \in \mathrm{Vec}(M)$ be a vector field such that the flow of f is volume-preserving w.r.t. μ. Then, f is recurrent.*

Proof. Suppose by contradiction that f is not recurrent. Then there exists an open nonempty set $\Omega \subset M$ and a time $t_0 > 0$ such that for all $t \geq t_0$,

$$e^{tf}(\Omega) \cap \Omega = \emptyset. \tag{6.14}$$

Define the sequence of sets $\Omega_n := e^{nt_0 f}(\Omega)$ for $n \in \mathbb{N}$. By the above assumption, for $n \neq m$ we have $\Omega_n \cap \Omega_m = \emptyset$. Indeed, if there exists $q \in \Omega_n \cap \Omega_m$ for some $n > m$, then there exists $p_n, p_m \in \Omega$ such that $q = e^{nt_0 f}(p_n) = e^{mt_0 f}(p_m)$. But this implies that $p_m = e^{(n-m)t_0 f}(p_n)$. In particular, $e^{(n-m)t_0 f}(\Omega) \cap \Omega \neq \emptyset$, which contradicts (6.14) since $(n-m)t_0 \geq t_0$.

Since f is measure-preserving w.r.t. μ and μ has full support, we have

$$\mu(\Omega_n) = \mu(e^{nt_0 f}(\Omega)) = \mu(\Omega) > 0, \qquad \text{for all } n \in \mathbb{N}.$$

Therefore,

$$\mu\left(\bigcup_{n=0}^{\infty} \Omega_n\right) = \sum_{n=0}^{\infty} \mu(\Omega_n) = \infty.$$

However, all the sets Ω_n are subsets of M, so $\bigcup_{n=0}^{\infty} \Omega_n \subset M$, which implies

$$\mu\left(\bigcup_{n=0}^{\infty} \Omega_n\right) \leq \mu(M) < \infty,$$

since μ is a finite measure. This is a contradiction, and hence f must be recurrent. □

We conclude by providing a standard differential characterization for volume-preservation, based on the concept of divergence. We say that the measure μ has a density, if for any coordinate chart there exists ρ such that

$$\mu(\Omega) = \int_\Omega \rho(x)\, dx \qquad \text{for any } \Omega \text{ in the chart.} \tag{6.15}$$

Definition 6.25. The divergence w.r.t. μ of $f \in \mathrm{Vec}(M)$ is the only function $\mathrm{div}_\mu f \in C^\infty(M)$ such that, for any a coordinate patch U it holds

$$\int_U \psi(x)\, \mathrm{div}_\mu f(x)\, \rho(x)\, dx = -\int_U f\psi(x)\, \rho(x)\, dx, \qquad \text{for any } \psi \in C_c^\infty(U).$$

Here, $f\psi \in \mathcal{C}^\infty(M)$ denotes the action of f on ψ as a differential operator. Equivalently, in coordinates,

$$\mathrm{div}_\mu f = \sum_{i=1}^n \rho^{-1} \partial_{x_i} \left(\rho f^i\right), \qquad \text{where} \qquad f(x) = \sum_{i=1}^n f^i \partial_{x_i}.$$

[1]That is, $\mu(U) > 0$ for every nonempty open set $U \subset M$.

We have the following result. The proof can seem very computational, but it is simply due to the fact that we avoided to introduce the theory of differential forms. We refer to [Lee13, Proposition 16.33, p.424] for a proof exploiting this theory.

Proposition 6.26. *Assume that the measure μ has a positive density ρ. Then, the vector field f is volume-preserving w.r.t. μ if and only if* $\operatorname{div}_\mu f = 0$.

Proof. It is convenient to use the notation $\phi_t = e^{tf}$. By additivity of the measure, it suffices to show the volume-preserving property a single chart U. Fix such a chart, and observe that since μ has a density, the requirement $\mu(\phi_t(\Omega)) = \mu(\Omega)$ for all $\Omega \subset U$ is equivalent to $\mu[\psi](t) = \mu[\psi](0)$, where

$$\mu[\psi](t) := \int_U \psi(\phi_t(x))\, \rho(x)\, dx, \qquad \forall \psi \in \mathcal{C}_c^\infty(U),\ t \geq 0. \tag{6.16}$$

Observe that we can differentiate under the integral sign in the definition of $\mu[\psi]$. This is useful, since recalling the definition of tangent map (cf. Definition 2.10), we can observe that

$$\left.\frac{d}{dt}\right|_{t=0} \psi \circ \phi_t = f(\psi) \circ \phi_t\Big|_{t=0} = f(\psi).$$

Here, $f(\psi)$ denotes the action of f as a differential operator on the function ψ. Hence,

$$\left.\frac{d}{dt}\right|_{t=0} \mu[\psi](t) = \int_U f(\psi)\, \rho(x)\, dx.$$

But, by definition of divergence, since $\psi \in \mathcal{C}^\infty(U)$ we have

$$\left.\frac{d}{dt}\right|_{t=0} \mu[\psi](t) = -\int_U \psi(x)\, \operatorname{div}_\mu f(x)\, \rho(x)\, dx. \tag{6.17}$$

If (6.16) holds, then the l.h.s. above is 0 for any $\psi \in \mathcal{C}_c^\infty(U)$. Since $\rho(x) > 0$ this yields $\operatorname{div}_\mu f(x) = 0$ for any $x \in U$.

Assume now that $\operatorname{div}_\mu f(x) = 0$ for any $x \in U$. Here, (6.17) is not enough, since it conly concerns the derivative at $t = 0$. However, one observes that replacing ψ by $\psi \circ \phi_{t_0+t}$ in the above computations yields

$$\left.\frac{d}{dt}\right|_{t=t_0} \mu[\psi](t) = -\int_U \psi(\phi_{t_0}(x))\, \operatorname{div}_\mu f(x)\, \rho(x)\, dx = 0, \qquad \forall \psi \in \mathcal{C}^\infty(M),\ t_0 \geq 0.$$

Hence, $\mu[\psi]'(t) = 0$ which implies the volume-preserving property. □

7. Worked out examples of controllability

In this chapter we present some examples of controllability problems, and we show how to apply the tools presented in the previous chapters to solve them. These examples are:

- In Section 7.1 we consider the car with n trailers, which is a generalization of the Reed-Shepp car. We will see that in order to prove its controllability, we need to consider a generalization of the Lie bracket generating conditions, where we consider Lie brackets of vector fields with coefficients in $\mathcal{C}^\infty(M)$, and not just Lie brackets of vector fields in $\mathcal{F}$.

- In Section 7.2 we consider the problem of controlling a satellite in orbit around the Earth, and we show how to exploit the recurrence of the drift to prove the controllability of the system. We will see that this can only be done when the satellite is not perfectly symmetric.

- In Section 7.3 we consider the problem of rolling two spheres on each other without slipping, and later without twisting. Controllability will follow from Chow-Rashevskii theorem, and we will observe again that a too symmetric configuration is not controllable.

- In Section 7.4 we show how the training problem for a ResNet can be recast as a controllability problem, which can ve solved by the tools developed in this notes.

7.1. The car with n trailers

In this section we consider a generalization of the Reed-Shepp car: the car with n trailers. Namely, we assume to have a Reed-Shepp car, with n trailers attached behind it.

There are many equivalen representations of the system, which leads to different degrees of complexities in the computations. Namely, when solving the general problem with n trailers, one needs for a simple form of induction to appear.

For mathematical simplicity, the car and the trailers are represented by two driving wheels connected by an axle, and all the rigid connections between the trailers have length equal to 1. We consider the following model, presented e.g., in [LT98]. In this model, we represent the position of the center of the *last trailer* by $(x, y) \in \mathbb{R}^2$, and the orientation w.r.t. the x axis of the i-th trailer by $\theta_i \in \mathbb{S}^1$. The orientation of the car pulling the trailers is $\theta_0 \in \mathbb{S}^1$. Namely, the state space is $M = \mathbb{R}^2 \times \mathbb{T}^{n+1}$ and the

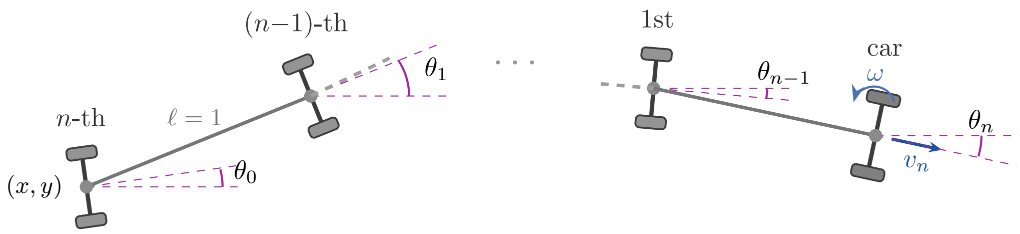


Figure 7.1.: The car with n trailers. The state is $q = (x, y, \theta_0, \theta_1, \dots, \theta_n) \in \mathbb{R}^2 \times \mathbb{T}^{n+1}$, where (x, y) is the position of the last trailer's axle, θ_i is the orientation of the i-th body, and the controls are the angular velocity ω and forward velocity v_n of the pulling car.

dynamics is

$$\begin{cases} \dot{x} = \cos\theta_0 v_0 \\ \dot{y} = \sin\theta_0 v_0 \\ \dot{\theta}_0 = \sin(\theta_1 - \theta_0) v_1 \\ \vdots \\ \dot{\theta}_{n-1} = \sin(\theta_n - \theta_{n-1}) v_n \\ \dot{\theta}_n = \omega \end{cases} \tag{7.1}$$

The control is $u = (v_n, \omega) \in \mathbb{R}^2$. That is, we can control the angular velocity ω and the tangential velocity v_n of the pulling car. Finally, r_i and v_i are quantities relative to the $n - i$ trailer: the length of its rigid connection with the $n - i + 1$ trailer (here, we let the car be the 0-th trailer), and its tangential velocity. In particular, the latter is given by

$$v_i = \varphi_i(\theta_i, \dots, \theta_n) v, \qquad \varphi_i(\theta_i, \dots, \theta_n) = \prod_{j=i+1}^{n} \cos(\theta_j - \theta_{j-1}), \qquad i = 0, \dots, n-1.$$

This system is affine in the control, indeed letting $q = (x, y, \theta_0, \dots, \theta_n)^\top \in M$ we have

$$\dot{q} = \omega X_1(q) + v X_2(q),$$

$$\text{where} \qquad X_1(q) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad X_2(q) = \begin{pmatrix} \cos\theta_0 \varphi_0(\theta_0, \dots, \theta_n) \\ \sin\theta_0 \varphi_0(\theta_0, \dots, \theta_n) \\ \sin(\theta_1 - \theta_0)\varphi_1(\theta_1, \dots, \theta_n) \\ \vdots \\ \sin(\theta_n - \theta_{n-1}) \\ 0 \end{pmatrix}$$

Clearly, the car with 0 trailers corresponds to the Reed-Shepp model studied before.

The car with 1 trailer In this case the state space is $M = \mathbb{R}^2 \times \mathbb{T}^2$, and the vector field read

$$X_1(q) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad X_2(q) = \begin{pmatrix} \cos\theta_0 \cos(\theta_1 - \theta_0) \\ \sin\theta_0 \cos(\theta_1 - \theta_0) \\ \sin(\theta_1 - \theta_0) \\ 0 \end{pmatrix}$$

Since we have two controls and the state space is 4 dimensional, we need to find two new directions by computing the Lie brackets. To perform this computation it is convenient to use the fact that $X_1 = \partial_{\theta_1}$, which yields

$$[X_1, X_2] = \begin{pmatrix} \cos\theta_0 \partial_{\theta_1} \cos(\theta_1 - \theta_0) \\ \sin\theta_0 \partial_{\theta_1} \cos(\theta_1 - \theta_0) \\ \partial_{\theta_1} \sin(\theta_1 - \theta_0) \\ 0 \end{pmatrix} = \begin{pmatrix} -\cos\theta_0 \sin(\theta_1 - \theta_0) \\ -\sin\theta_0 \sin(\theta_1 - \theta_0) \\ \cos(\theta_1 - \theta_0) \\ 0 \end{pmatrix}$$

This is indeed linearly independent of $\{X_1, X_2\}$. However, it should be evident that $[X_1, [X_1, X_2]]$ does not yield a new direction, since $(\sin\theta)'' = -\sin\theta$. Thus, we compute $[X_2, [X_1, X_2]]$. It is slightly convenient to use formula (2.24) that is,

$$[X_2, [X_1, X_2]] = \frac{\partial [X_1, X_2]}{\partial q} X_2 - \frac{\partial X_2}{\partial q} [X_1, X_2] = \begin{pmatrix} \sin\theta_0 \\ -\cos\theta_0 \\ -1 \\ 0 \end{pmatrix}.$$

It is then straightforward to check that $\{X_1, X_2, [X_1, X_2], [X_2, [X_1, X_2]]\}$ are linearly independent. Thus the system is controllable by Chow-Rashevskii.

The car with 2 trailers From the Reed-Shepp car (where one needed just $[X_1, X_2]$ to obtain the Lie bracket generating property), and the preceding example, where one needs to add also $[X_2, [X_1, X_2]]$, it would be tempting to infer that to control the car with n trailers we need to compute brackets of length $n + 1$. Unfortunately, this is not the case.

This is evident already in the case of the car with 2 trailers. Although we will not present the computations since they are quite long and tedious, let us synthetise what happens (recall that $\dim M = 5$):

- If $\theta_2 - \theta_1 \neq \pi/2$, then the above intuition holds. Namely,

$$\dim \operatorname{span} \{X_1, X_2, [X_1, X_2], [X_2, [X_1, X_2]], [X_2, [X_2, [X_1, X_2]]]\} \big|_q = 5,$$

- If $\theta_2 - \theta_1 = \pi/2$ the above dimension is only 4. In order to obtain the Lie bracket generating condition, we need the additional Lie bracket $[X_1, [X_2, [X_2, [X_1, X_2]]]]$, which has length 4.

Hence, the car with 2 trailers is controllable, but the structure of the Lie algebra $\operatorname{Lie} \mathcal{F}$ depends on the point.

The general case Let us prove the following.

Theorem 7.1. *The control system* (7.1) *is controllable.*

The fact that in general there are configurations where a different number of Lie brackets is required to obtain the Lie bracket generating condition represents a serious difficulty in producing a recursive argument for the controllability.

The simplest possible proof of the above theorem, which is due to Laumond, hinges on the following fact.

Lemma 7.2. *Let* $\mathfrak{F} \subset \mathrm{Vec}(M)$ *denote the* $\mathcal{C}^\infty$*-module of vector fields containing* $\{X_1, X_2\}$*. That is,*

$$\mathfrak{F} \subset \mathrm{Vec}(M) = \{aX_1 + bX_2 \mid a, b \in \mathcal{C}^\infty(M)\}.$$

In particular, $\mathfrak{F} \supset \mathcal{F}$ *and the same holds for the corresponding Lie algebras, i.e.,* $\mathrm{Lie}\,\mathfrak{F} \supset \mathrm{Lie}\,\mathcal{F}$*. Moreover, it holds that*

$$\mathrm{Lie}_q\,\mathfrak{F} = \mathrm{Lie}_q\,\mathcal{F}, \qquad \forall q \in M.$$

Proof. Recall the characterization of $\mathrm{Lie}\,\mathcal{F}$ presented in Proposition 5.4:

$$\mathrm{Lie}\,\mathcal{F} = \mathrm{span}\left\{[f_1, [f_2, [\dots, f_k]]] \mid k \in \mathbb{N},\ f_1, \dots, f_k \in \mathcal{F}\right\}.$$

It is easily checked that this holds also replacing $\mathcal{F}$ by $\mathfrak{F}$. Let us proceed by induction on the integer k in the above.

Let $k = 1$. Since any element $f_1 \in \mathfrak{F}$ reads $f_1(q) = a_1(q)X_1(q) + a_2(q)X_2(q)$ for $a_1, a_2 \in \mathcal{C}^\infty(M)$, we have that $f_1(q) \in \mathrm{span}\{X_1(q), X_2(1)\} \subset \mathrm{Lie}_q\,\mathcal{F}$.

Assume to have proven the statement for k, and let us consider $[f_1, \hat{f}]$ be such that f_1 and $\hat{f} = [f_2, [\dots, f_{k+1}]]$ where $f_1, \dots, f_{k+1} \in \mathfrak{F}$. By induction, it holds that $\hat{f}(q) \in \mathrm{Lie}_q\,\mathcal{F}$, and letting $f_1 = a_1X_1 + a_2X_2$ as above we have

$$[f_1, \hat{f}](q) = a_1(q)[X_1, \hat{f}](q) - \hat{f}(a_1)(q)X_1(q) + a_2(q)[X_2, \hat{f}](q) - \hat{f}(a_2)(q)X_2(q)$$

Thus, $[f_1, \hat{f}](q)$ is a linear combination of $[X_1, \hat{f}](q)$, $[X_2, \hat{f}](q)$, $X_1(q)$, and $X_2(q)$. This proves that it belongs to $\mathrm{Lie}_q\,\mathcal{F}$. □

We are in a position to conclude.

Proof of Theorem 7.1. Let us consider the following vector fields, for $i \geq 1$,

$$\begin{aligned} W_0 &= X_1, & W_{i+1} &= \sin(\theta_i - \theta_{i-1})V_i + \cos(\theta_i - \theta_{i-1})Z_i, \\ V_0 &= X_2, & V_{i+1} &= \cos(\theta_i - \theta_{i-1})V_i - \sin(\theta_i - \theta_{i-1})Z_i, \\ Z_0 &= [X_1, X_2], & Z_{i+1} &= [W_{i+1}, V_{i+1}]. \end{aligned}$$

Observe that, by construction, W_i, V_i, Z_i do not belong to the Lie algebra generated by $\mathcal{F}$, but only to the one generated by $\mathfrak{F}$.

The form of these vector fields can be obtained by induction, via some long computations. We present the important ones:

$$W_i = (\underbrace{0,\dots,0}_{n-i+2},1,\underbrace{0,\dots,0}_{i})^\top = \partial_{\theta_{n-i}}, \qquad i=0,\dots,n$$

$$V_n = (\cos(\theta_1-\theta_0),\sin(\theta_1-\theta_0),0,\dots,0)^\top = \cos(\theta_1-\theta_0)\partial_x + \sin(\theta_0-\theta_1)\partial_y,$$

$$Z_n = (-\sin(\theta_1-\theta_0),\cos(\theta_1-\theta_0),0,\dots,0)^\top = -\sin(\theta_1-\theta_0)\partial_x + \cos(\theta_0-\theta_1)\partial_y$$

Clearly the above $n+3$ vector fields are linearly independent, and thus $\mathrm{Lie}_q\,\mathfrak{F} = T_qM$ for any $q \in M$. Thanks to Lemma 7.2 this implies the Lie bracket generating condition and thus, by Chow-Rashevkii Theorem, the statement. □

It is intuitive that the minimum number of Lie brackets necessary to generate the full tangent space (the *degree of nonholonomy* of the control system) is deeply connected with the "easeness" to control the system. This can be seen, e.g., by observing that (by Lemma 2.23) to generate $[X_1,X_2]$ we need piecewise controls with 4 pieces, for $[X_2,[X_1,X_2]]$ we need 10 pieces, and so on.

The technique of proof presented above shows that the number of iterated Lie brackets necessary to control the system is at most 2^{n+1}. Indeed, we see that W_{i+1} and V_{i+1} require the same number of iterated Lie brackets as Z_i, and thus that Z_{i+1} doubles this number. Since Z_0 is an iterated Lie bracket of length 2 this proves the bound.

We mention that finer results are available in the literature, showing that the sharp bound is given by the $(n+3)$-rd Fibonacci number. This bound is attained at those points such that $\theta_i - \theta_{i-1} = \pm\pi/2$, $i=2,\dots,n$.

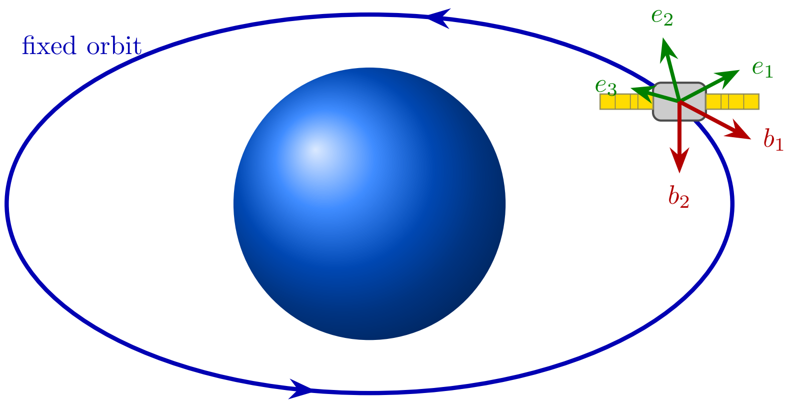


Figure 7.2.: Satellite on a fixed orbit. The body frame (e_1, e_2, e_3) encodes the orientation $R(t) \in SO(3)$; $\omega(t)$ is the angular velocity; b_1, b_2 are the reactor directions.

7.2. Recurrent vector fields: satellite

Given a satellite moving along a fixed orbit (cf. Figure 7.2), let us consider the problem of controlling its orientation, which is determined by an element of the special orthogonal group

$$SO(3) = \{R \in \mathbb{R}^{3\times 3} \mid R^\top R = \mathrm{Id},\, \det R = 1\}.$$

By the above definition, it is immediate to check that $SO(3)$ is a manifold. Moreover, we have the following.

Proposition 7.3. *The tangent space at $R \in SO(3)$ is*

$$T_R SO(3) = \{\Omega R \mid \Omega \in \mathfrak{so}(3)\}, \tag{7.2}$$

where $\mathfrak{so}(3)$ is the set of skew-symmetric matrices (i.e., such that $\Omega^\top = -\Omega$).

Proof. Let $\gamma : (-\varepsilon, \varepsilon) \to SO(3)$ be a smooth curve such that $\gamma(0) = R$, and let $\Omega = \dot\gamma(0)R^\top$. Then $\gamma(t)^\top\gamma(t) = \mathrm{Id}$ for any $t \in (-\varepsilon, \varepsilon)$. Differentiating w.r.t. time yields

$$0 = \frac{d}{dt}\bigg|_{t=0} \gamma(t) = \dot\gamma(0)^\top R + R^\top \dot\gamma(t) = \Omega + \Omega^\top.$$

This implies that $\Omega \in \mathfrak{so}(3)$. □

Any matrix $\Omega \in \mathfrak{so}(3)$ can be represented by the vector $\omega \in \mathbb{R}^3$, by decomposing it along the basis

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.3}$$

Namely,

$$\Omega = \begin{pmatrix} 0 & \omega^3 & \omega^2 \\ -\omega^3 & 0 & \omega^1 \\ -\omega^2 & -\omega^1 & 0 \end{pmatrix}. \tag{7.4}$$

If $R(t)$ is the orientation of the satellite at time t and $\dot{R}(t) = \Omega(t)R(t)$ is its velocity, the vector $\omega(t)$ associated with $\Omega(t)$ is called the *angular velocity* of the satellite. In particular, the infinitesimal motion of the satellite at time t corresponds to a rotation around the axis $\omega(t)$ at speed $|\omega(t)|$.

Letting $\mu^i = I_i\omega^i$, where $I = (I_1, I_2, I_3)$ are the (constant) *moments of intertia* of the satellite, Euler's rotation equations dictates that the dynamics of the system is

$$\dot{\mu} + \omega \wedge \mu = M,$$

where M is the *torque* applied to the system. Here, $\mu \wedge \omega$ denotes the cross product:

$$\omega \wedge \mu = \begin{pmatrix} (I_3 - I_2)\omega^2\omega^3 \\ (I_1 - I_3)\omega^3\omega^1 \\ (I_2 - I_1)\omega^1\omega^2 \end{pmatrix}. \tag{7.5}$$

In the case of a satellite, the control is induced via coupled reactors placed on its body. Each couple of reactors has a fixed orientation b_i and can be activated in one direction or its opposite, inducing a constant modulus torque along b_i or $-b_i$. That is, for n couple of reactors, we have

$$M = M_1 + \ldots + M_n, \qquad M_i = u_i b_i, \qquad u_i \in \{-1, 0, 1\}.$$

Controls of this type are called *bang-bang.* The resulting system reads

$$\dot{\mu} = f_0(\mu) + \sum_{i=1}^{n} u_i f_i(\mu), \qquad \mu \in \mathbb{R}^3,\ u \in U := \{-1, 0, 1\}^n. \tag{7.6}$$

Here, letting $a_1 = (I_2 - I_3)/(I_2 I_3)$, $a_2 = (I_3 - I_1)/(I_3 I_1)$, and $a_3 = (I_1 - I_2)/(I_1 I_1)$, we have

$$X_0(\mu) = \begin{pmatrix} a_1\mu^2\mu^3 \\ a_2\mu^3\mu^1 \\ a_3\mu^1\mu^2 \end{pmatrix}, \qquad X_i(\mu) = b_i, \qquad i = 1, \ldots, n.$$

Let us consider two different configurations, depending on the shape of the satellite.

Spherically symmetric case If the satellite is a uniform spherical ball, all moments of inertia coincide ($I_1 = I_2 = I_3$). Hence, $a_1 = a_2 = a_3 = 0$, $X_0(\mu) \equiv 0$, and thus (7.6) is a linear driftless system. In particular, it is controllable if and only if

$$\operatorname{rank}\{b_1, \ldots, b_n\} = 3.$$

As a consequence, at least 3 couples of reactors are needed to control it.

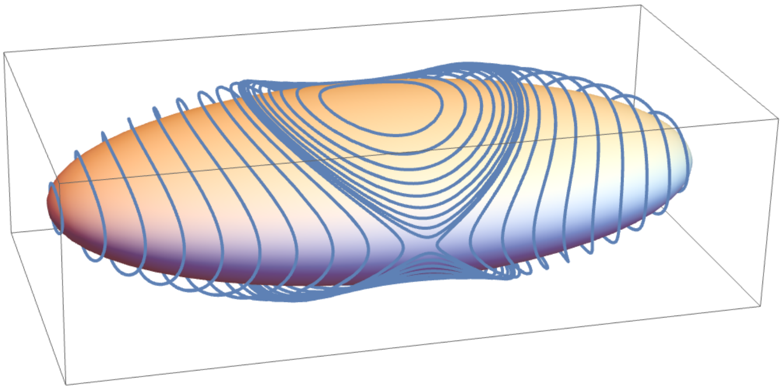


Figure 7.3.: Trajectories of the drift f_0 for generic moments of intertia.

Nonsymmetric case Let us assume that the satellite is nonsymmetric, i.e., that at least two inertia moments are different. In this case, at least two a_i's are nonzero and the drift is not trivially zero.

We then have the following.

Proposition 7.4. *Assume that $f_0(\mu) \not\equiv 0$. Then, it is recurrent.*

Proof. Let us consider the system

$$\dot{\mu} = f_0(\mu) = -\omega \wedge \mu. \tag{7.7}$$

Observe that this system has two constants of motions:

- The angular momentum norm

$$K(\mu) = \|\mu\|^2.$$

- The kinetic energy

$$E(\mu) = \frac{\|\omega\|^2}{2} = \frac{1}{2}\left[\frac{(\mu^1)^2}{I_1^2} + \frac{(\mu^2)^2}{I_2^2} + \frac{(\mu^3)^2}{I_3^2}\right].$$

Indeed, recalling that the cross product $a \wedge b$ is orthogonal to both a and b, we have

$$\begin{aligned}
\dot{K} &= 2\mu^\top \dot{\mu} = -2\mu^\top(\omega \wedge \mu) = 0,\\
\dot{E} &= \frac{1}{2}\omega^\top \dot{\omega} = -\frac{1}{2}\left[\frac{I_2 - I_3}{I_1} + \frac{I_3 - I_1}{I_2} + \frac{I_1 - I_2}{I_3}\right]\omega^1\omega^2\omega^3 = 0.
\end{aligned}$$

Hence, any trajectory $\mu(t) \in \mathbb{R}^3$ of (7.7) is contained in the intersection between a sphere $\{\|\mu\|^2 = R\}$ and an ellipsoid $\{E(\mu) = c\}$. Since these are all periodic, the statement follows (cf. Figure 7.3). □

By Theorem 6.17, to show controllability it suffices to prove that the vector fields $\{f_0, f_1, \dots, f_n\}$ are Lie bracket generating. Let us show that, actually, in this case it is sufficient to consider a single couple of reactors f_1.

Proposition 7.5. *Assume that $f_0(\mu) \not\equiv 0$. Then there exists $b \in \mathbb{R}^3$ such that the system*

$$\dot{\mu} = f_0(\mu) + ub, \qquad u \in \{-1, 0, 1\}, \tag{7.8}$$

is controllable.

Proof. We proceed in three steps:

Step 1. Compute Lie brackets. We have that

$$[f_1, f_0] = \frac{\partial f_0}{\partial \mu} f_1(\mu) - \frac{\partial f_1}{\partial \mu} f_0(\mu)$$
$$= \begin{pmatrix} 0 & a_1\mu^3 & a_1\mu^2 \\ a_2\mu^3 & 0 & a_2\mu^1 \\ a_3\mu^2 & a_3\mu^1 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1\left(\mu^3 b_2 + \mu^2 b_3\right) \\ a_2\left(\mu^3 b_1 + \mu^1 b_3\right) \\ a_1\left(\mu^2 b_1 + \mu^1 b_2\right) \end{pmatrix}.$$

We continue to compute,

$$[f_1, [f_1, f_0]] = \frac{\partial [f_1, f_0]}{\partial \mu} f_1(\mu) - \frac{\partial f_1}{\partial \mu} [f_1, f_0](\mu)$$
$$= \begin{pmatrix} 0 & a_1 b_3 & a_1 b_2 \\ a_2 b_3 & 0 & a_2 b_1 \\ a_3 b_2 & a_3 b_1 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 b_3 b_2 \\ a_2 b_3 b_1 \\ a_1 b_2 b_1 \end{pmatrix} = 2 f_0(b). \tag{7.9}$$

Step 2. Let us show that the statement is true if the set $\Pi := \{b, f_0(b)\}$ is 2-dimensional and not invariant under the flow of f_0.

The assumption on Π means that there exists $v \in \Pi$ such that

$$\dim \operatorname{span}\{b, f_0(b), f_0(v)\} = 3. \tag{7.10}$$

The fact that $v \in \Pi$ implies that $v = \alpha b + \beta f_0(b)$. By (7.9), we have that

$$v = \alpha b + \frac{\beta}{2}[f_1, [f_1, f_0]].$$

In particular, the constant vector field $f_2(\mu) = v$ satisfies $f_2 \in \operatorname{Lie}\{f_0, f_1\}$. Moreover, the same computation as in (7.9) shows that $f_0(v) = [f_2, [f_2, f_0]]$. Hence, the claim follows by (7.10).

Step 3. If $f_0(\mu) \not\equiv 0$, there exists b satisfying the assumptions of step 2.

Let $v \in \operatorname{span}\{b, f_0(b)\}$. Then, $v = \alpha b + \beta f_0(b)$, and hence

$$f_0(v) = \alpha^2 f_0(b) + \beta^2 (a_1 a_2 a_3 b_1 b_2 b_3) b + \alpha\beta d,$$

where

$$d = \begin{pmatrix} a_1(a_2 b_3^2 b_1 + a_3 b_1 b_2^2) \\ a_2(a_3 b_1^2 b_2 + a_1 b_2 b_3^2) \\ a_3(a_1 b_2^2 b_3 + a_2 b_3 b_1^2) \end{pmatrix}.$$

We then need to find a vector b such that $\{b, f_0(b), d\}$ are linearly indepent. This can be checked by computing

$$\det\left[b, f_0(b), d\right] = \det\begin{bmatrix} b_1 & a_1 b_2 b3 & a_1(a_2 b_3^2 b_1 + a_3 b_1 b_2^2) \\ b_2 & a_2 b_3 b_1 & a_2(a_3 b_1^2 b_2 + a_1 b_2 b_3^2) \\ b_3 & a_3 b_1 b_2 & a_3(a_1 b_2^2 b_3 + a_2 b_3 b_1^2) \end{bmatrix}.$$

Indeed, since at least two a_i's are different from zero, the above expression yields a polynomial expression in b with non-zero coefficients. As such, there exists b such that it is nonzero. $\square$

7.3. Rolling of a sphere onto another

Let us consider two spheres S_1 and S_2 of radii $r_1, r_2 > 0$, respectively, that roll one onto the other. That is, the two spheres are embedded in the same $\mathbb{R}^3$ and are in contact at exactly one point p.

The most convenient way to describe the dynamic of the system is the following: Up to a rigid change of coordinates, it is always possible to assume the point of contact p to be the origin of $\mathbb{R}^3$, the center of S_1 to be the point $c_1 = (0, 0, r_1)$, and the center of S_2 to be the point $c_2 = (0, 0, -r_2)$. In this way, the state space can be described by $M = SO(3) \times SO(3)$. Namely, a state is $(R_1, R_2) \in M$ such that R_1 and R_2 are the transition matrices between the fixed orthonormal frame at the origin and the frames attached to S_2 and S_1, respectively. (See Section 7.2 for some generalities on $SO(3)$.)

Then, the tangent space can be described as

$$T_{(R_1,R_2)}M = \{(R_1\Omega_1, R_2\Omega_2) \mid \Omega_1, \Omega_2 \in \mathfrak{so}(3)\}$$

A general motion of the system is then given by a curve $t \mapsto (R_1(t), R_2(t))$ satisfying

$$\begin{cases} \dot{R}_1(t) = R_1(t)\Omega_1(t) \\ \dot{R}_2(t) = R_2(t)\Omega_2(t) \end{cases} \tag{7.11}$$

for some $\Omega_1(t), \Omega_2(t) \in \mathfrak{so}(3)$.

Recall that each matrix Ω_i can be associated with its corresponding angular velocity $\omega_i \in \mathbb{R}^3$ such that $\Omega_i x = \omega_i \wedge x$ for any $x \in \mathbb{R}^3$ or, equivalently, $\Omega_i = \omega_1^1 A_1 + \omega_1^2 A_2 + \omega_1^3 A_3$, where A_1, A_2, A_3 are the standard basis elements of $\mathfrak{so}(3)$ given in (7.3).

7.3.1. Rolling without slipping

In the following we are interested in the case where the speres roll *without slipping*. That is, a rotation of S_1 induces (via the contact point p) a rotation of S_2 around the same point, and vice versa.

The no-slip condition states that the velocities of the two surfaces must coincide at the contact point p. Since p is the origin, the velocity of a point on the surface of S_1 at distance r_1 from its center c_1 is given by $\Omega_1(p - c_1) = -r_1\Omega_1 e_3$, and similarly the velocity of the corresponding point on S_2 is $\Omega_2(p - c_2) = r_2\Omega_2 e_3$. Here, we let $e_3 = (0, 0, 1)$ be the unit vector along the z-axis. The no-slip condition therefore reads

$$-r_1\,\Omega_1 e_3 = r_2\,\Omega_2 e_3 \iff (r_1\Omega_1 + r_2\Omega_2)\,e_3 = 0.$$

That is, the vector $r_1\omega_1 + r_2\omega_2$ is parallel to e_3, or equivalently, the vector ω_2 is obtained from ω_1 by a rotation around the z-axis. The component along the contact normal e_3 is unconstrained by the rolling condition: it corresponds to the two spheres spinning freely about their common normal without affecting the contact, and thus it can be chosen arbitrarily. Since this spinning motion does not affect the rolling, we can set it to zero without loss of generality (see Section 7.3.3 below). This yields the constraint

$$\omega_2 = -\frac{r_2}{r_1}\omega_1.$$

Substituting the constraint $\omega_2 = -\frac{r_2}{r_1}\omega_1$ into (7.11), the system reduces to a control system on $SO(3) \times SO(3)$ driven by the single angular velocity $\omega_1 \in \mathbb{R}^3$. To obtain a more symmetric system, we can reparametrize the control by considering $u \in \mathbb{R}^3$ and setting

$$\omega_1 = a\,u, \qquad \omega_2 = (a-1)\,u, \qquad a = \frac{r_2}{r_1 + r_2} \in (0,1).$$

This reparametrization is consistent with the no-slip constraint, since

$$r_1\,\omega_1 + r_2\,\omega_2 = [r_1 a + r_2(a-1)]\,\omega = 0.$$

Substituting into (7.11) and denoting by $\Omega_u \in \mathfrak{so}(3)$ the skew-symmetric matrix associated with u, we obtain the control system

$$\begin{cases} \dot R_1 = a\,R_1\Omega_u \\ \dot R_2 = (a-1)\,R_2\Omega_u \end{cases} \qquad u \in \mathbb{R}^3, \qquad a = \frac{r_2}{r_1 + r_2}. \tag{7.12}$$

Note that the opposite signs reflect the fact that the two spheres rotate in opposite directions at the contact point. In the symmetric case $r_1 = r_2$, we have $a = \frac{1}{2}$, and the two spheres respond symmetrically to the rolling motion.

We will now prove the following result.

Proposition 7.6. *The control system* (7.12) *is controllable.*

Recall the definition of the basis of $\mathfrak{so}(3)$ given in Section 7.2:

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then, system (7.14) can be rewritten as the linear control system

$$\dot q = u_1 f_1(q) + u_2 f_2(q) + u_3 f_3(q), \qquad f_i(R_1, R_2) = \begin{pmatrix} aR_1A_i \\ (a-1)R_2A_i \end{pmatrix}, \; i = 1,2,3. \tag{7.13}$$

Since the above system is symmetric, we just need to compute the Lie brackets of the vector fields f_i's and show that they generate the whole tangent space at any point. Since $f_i(R_1, R_2) = (ag_i(R_1), (a-1)g_i(R_2))$, where $g_i(R) = RA_i$, it suffices to compute $[g_i, g_j]$ for any $R \in SO(3)$. Indeed, we have

$$[f_i, f_j](R_1, R_2) = \begin{pmatrix} a^2[g_i, g_j](R_1) \\ (a-1)^2[g_i, g_j](R_2) \end{pmatrix}.$$

To this aim, we present the following, which encodes the fact that A_i's are generators of rotations around the three coordinate axes.

Proposition 7.7 (Rodriguez formula)**.** *It holds that*

$$e^{tA_i} = \mathrm{Id} + \sin(t)A_i + (1-\cos(t))A_i^2, \qquad i = 1,2,3, \qquad t \in \mathbb{R}.$$

Proof. We prove the statement for $i = 1$; the other cases are analogous. Observe that

$$A_1^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \text{and} \qquad A_1^3 = A_1 A_1^2 = -A_1.$$

Hence, we have that $A_1^{2k} = (-1)^{k+1} A_1^2$ and $A_1^{2k+1} = (-1)^k A_1$ for any $k \in \mathbb{N}$. As a consequence,

$$\begin{aligned} e^{tA_1} &= \sum_{k=0}^{\infty} \frac{t^k A_1^k}{k!} \\ &= \mathrm{Id} + \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1} A_1^{2k+1}}{(2k+1)!} + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{t^{2k} A_1^{2k}}{(2k)!} \\ &= \mathrm{Id} + \sin(t) A_1 + (1 - \cos(t)) A_1^2. \end{aligned}$$

□

Thanks to the above formula, we can now compute the Lie brackets of g_i's.

Proposition 7.8. *It holds that*

$$[g_i, g_j](R) = R[A_i, A_j], \qquad \textit{where} \qquad [A_i, A_j] = A_i A_j - A_j A_i, \qquad i, j = 1, 2, 3.$$

In particular, we have that

$$[g_1, g_2](R) = RA_3, \qquad [g_2, g_3](R) = RA_1, \qquad [g_3, g_1](R) = RA_2.$$

Proof. By Lemma 2.23, we have

$$[g_i, g_j](R) = \frac{1}{2} \ddot{\varphi}(0), \qquad \text{where} \qquad \varphi(t) := e^{-tg_j} \circ e^{-tg_i} \circ e^{tg_j} \circ e^{tg_i}(R).$$

Observe that $e^{tg_i}(R) = Re^{tA_i}$, since it is a solution of $\dot{R} = RA_i$ (pay attention to the fact that the l.h.s. is a flow on a manifold, while the r.h.s. is a matrix exponential). Hence, by the Rodriguez formula, we have that

$$e^{tg_i}(R) = R + \sin(t) A_i R + (1 - \cos t) A_i^2 R = R + tA_i R + \frac{t^2}{2} A_i^2 R + o(t^2).$$

This implies that

$$\varphi(t) = R + t^2 (A_i A_j - A_j A_i) R + o(t^2),$$

which yields the first part of the statement. The second part follows by direct computation. □

We are now in position to conclude.

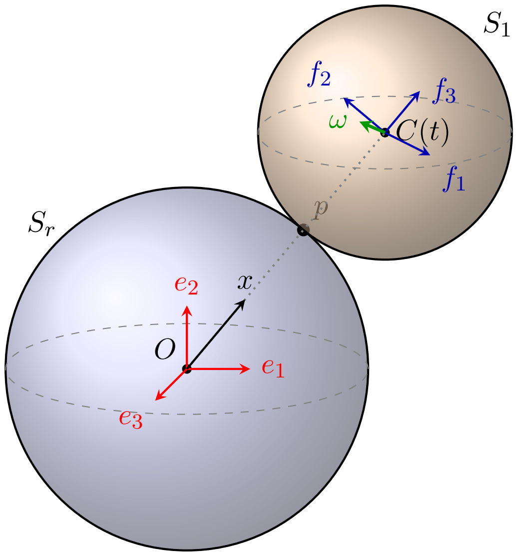


Figure 7.4.: Rolling of two spheres.

Proof of controllability. Since the system is linear in the control and the state, by the Orbit Theorem it is controllable if and only if the vector fields $\{f_1, f_2, f_3\}$ are Lie bracket generating. We have that

$$[f_i, f_j](R_1, R_2) = \begin{pmatrix} a^2 R_1 A_k \\ (a-1)^2 R_2 A_k \end{pmatrix}, \qquad \text{where} \qquad k = \begin{cases} 3 & \text{if } (i,j) = (1,2), \\ 1 & \text{if } (i,j) = (2,3), \\ 2 & \text{if } (i,j) = (3,1). \end{cases}$$

Since $a \neq \{0, 1\}$, the vector fields $\{f_1, f_2, f_3, [f_1, f_2], [f_2, f_3], [f_3, f_1]\}$ are linearly independent, and thus the system is controllable. □

7.3.2. Rolling without slipping and spinning

In the previous section we have shown that the rolling of two spheres without slipping is controllable. However, we have assumed that the spinning motion of one sphere around the common normal at the contact point (i.e., e_3) is also transmitted to the other sphere, which is not necessarily the case in practice (see Section 7.3.3). In this section, we thus introduce the additional constraint of no spinning. Namely, we consider the system

$$\begin{cases} \dot{R}_1 = a\, R_1 \Omega_u \\ \dot{R}_2 = (a-1)\, R_2 \Omega_u \end{cases} \qquad u \in \mathbb{R}^3, \qquad u \perp e_3, \qquad a = \frac{r_2}{r_1 + r_2}, \tag{7.14}$$

In this case we have the following.

Proposition 7.9. *The control system* (7.14) *is controllable if and only if* $r_1 \neq r_2$.

Proof. The system reads

$$\dot{q} = u_1 f_1(q) + u_2 f_2(q) \qquad f_i(R_1, R_2) = \begin{pmatrix} aR_1A_i \\ (a-1)R_2A_i \end{pmatrix}, \; i = 1,2,3. \tag{7.15}$$

Observe, that now we only have two control vector fields. In particular, we will forcibly need to iterate Lie brackets to generate the whole tangent space. Let us compute

$$f_{12}[f_1, f_2](R_1, R_2) = \begin{pmatrix} a^2 R_1 A_3 \\ (a-1)^2 R_2 A_3 \end{pmatrix}.$$

Since $a \neq \{0, 1\}$, the vector fields $\{f_1, f_2, f_{12}\}$ are linearly independent. Observe however that $f_{12} \neq f_3$, due to the sign of the second component. Moreover, we have that

$$f_{112} = [f_1, [f_1, f_2]](R_1, R_2) = \begin{pmatrix} a^3 R_1 A_2 \\ (a-1)^3 R_2 A_2 \end{pmatrix},$$
$$f_{212} = [f_2, [f_1, f_2]](R_1, R_2) = \begin{pmatrix} a^3 R_1 A_1 \\ (a-1)^3 R_2 A_1 \end{pmatrix}.$$

Here we have to pay attention: indeed the vector fields f_{112} and f_{212} are linearly independent of $\{f_1, f_2, f_{12}\}$, if and only if $a^2 \neq (a-1)^2$, which is equivalent to $r_1 \neq r_2$. Indeed, if $r_1 = r_2$, then $a = \frac{1}{2}$, and we have that $4f_{112} = f_2$ and $4f_{212} = f_1$.

To conclude, if $r_1 \neq r_2$, the vector fields $\{f_1, f_2, f_{12}, f_{112}, f_{212}\}$ are linearly independent, and thus the system is controllable while if $r_1 = r_2$, the system is not controllable since the Lie algebra generated by f_1 and f_2 is only 3-dimensional. □

In the case $r_1 = r_2$, we can actually push the analysis futher. Indeed, in this case we have that $a = \frac{1}{2}$, and thus the system reads

$$\begin{cases} \dot{R}_1 = \frac{1}{2} R_1 \Omega_u \\ \dot{R}_2 = -\frac{1}{2} R_2 \Omega_u, \end{cases} \qquad \Omega_u = u_1 A_1 + u_2 A_2, \qquad u \in \mathbb{R}^2. \tag{7.16}$$

The idea is that the system is not controllable because the spheres orientations are "locked": since they have the same radius, when they roll against each other the respective frames rotate in opposite directions at the same rate, and thus they keep the same relative position.

In order to formalize this intuition we introduce the operator $L : SO(3) \to SO(3)$ which reflects the orientation of the second sphere, i.e., $L(e_1) = e_1$, $L(e_2) = e_2$, and $L(e_3) = -e_3$. We then have the following.

Proposition 7.10. *For any* $(R_1^0, R_2^0) \in SO(3) \times SO(3)$, *it holds*

$$\mathcal{O}(R_1^0, R_2^0) = \{(R_1, R_2) \in SO(3) \times SO(3) \mid R_1 L R_2^\top L = R_1^0 L (R_2^0)^\top L\}.$$

In particular, $\dim \mathcal{O}(R_1^0, R_2^0) = 3$ *and the system is not controllable.*

Proof. This operator is an involution, i.e., $L^3 = \mathrm{Id}$, and it is an isometry, i.e., $L \in O(3)$. In particular, $L\Omega_u L = L(u_1A_1 + u_2A_2)L = -\Omega_u$, and thus $\frac{d}{dt}(LR_2L) = \frac{1}{2}R_2\Omega_u \in SO(3)$. Let us then consider the quantity $h(R_1, R_2) = R_1LR_2^\top L$. Then,

$$\dot{h} = \dot{R}_1LR_2^\top L + R_1L\dot{R}_2^\top L = \frac{1}{2}R_1\Omega_uR_2^\top - \frac{1}{2}R_1\Omega_uR_2^\top = 0.$$

This shows that the quantity $h(R_1, R_2)$ is a first integral of the system and thus that $H = \{h(R_1, R_2) = h(R_1^0, R_2^0)\}$ is contained in $\mathcal{O}(R_1^0, R_2^0)$.

To complete the proof, it suffices to show that H is a 3-dimensional submanifold of $SO(3) \times SO(3)$. To this aim, consider $S_i(t)$ such that $\dot{S}_i = A_iS_i$ and $S_i(0) = R_1^0$ for $i = 1, 2, 3$. Then, $\{S_1(t), S_2(t), S_3(t)\}$ is a basis of $SO(3)$ for any $t \in \mathbb{R}$, and for $i = 1, 2, 3$ we have that

$$B_i = \left.\frac{d}{dt}\right|_{t=0} h(S_i(t), R_2^0) = \dot{S}_i(0)L(R_2^0)^\top L = A_iS_i(0)L(R_2^0)^\top L = A_ih(R_1^0, R_2^0).$$

Since $h(R_1^0, R_2^0) \in SO(3)$ and $\{A_1, A_2, A_3\}$ is a basis of $\mathfrak{so}(3)$, the vectors $\{B_1, B_2, B_3\}$ are linearly independent, and thus H is a 3-dimensional submanifold of $SO(3) \times SO(3)$. □

7.3.3. Remark on the geometrical setting

A more natural way of describing the system is to consider the sphere S_1 as fixed (say with its center at the origin) and the sphere S_2 as moving. In this way, the state space can be described by $M = \mathbb{S}^2 \times SO(3)$, where the $\mathbb{S}^2$ component describes the position of the center of S_2 and the $SO(3)$ component describes the orientation of S_2. Observe that this state space is 5 dimensional, not 6 as in the previous description. The reason is that here we are disregarding the possible spinning motion of S_1 around the common normal at the contact point.

In order to pass from one description to the other, we can consider the projection $\pi : SO(3) \times SO(3) \to \mathbb{S}^2 \times SO(3)$ given by $\pi(R_1, R_2) = (R_1e_3, R_1R_2^\top)$. Indeed, letting $v = (R_1\Omega_1, R_2\Omega_2)$ be a tangent vector at $(R_1, R_2) \in SO(3) \times SO(3)$, given as $\dot{\gamma}(0) = v$ for some curve $\gamma : (-\epsilon, \epsilon) \to SO(3) \times SO(3)$, one computes

$$\pi_*v = \left.\frac{d}{dt}\right|_{t=0} \pi(\gamma(t)) = \left.\frac{d}{dt}\right|_{t=0} (R_1(t)e_3, R_1(t)R_2(t)^\top) = (R_1\Omega_1e_3, R_1\Omega_1R_2^\top - R_1\Omega_2R_2^\top).$$

In particular, letting $v = (aR_1\Omega_u, (1-a)R_2\Omega_u)$ as in system (7.12), we have that

$$\pi_*v = (aR_1\Omega_ue_3, aR_1\Omega_uR_2^\top - (1-a)R_1\Omega_uR_2^\top) = (aR_1\Omega_ue_3, R_1\Omega_uR_2^\top).$$

Namely, considering R_1 such that $R_1e_3 = x$ and considering as control $\Omega_v = R_1\Omega_uR_1^\top$, we have that the control system (7.12) on $SO(3) \times SO(3)$ is projected to the following control system on $\mathbb{S}^2 \times SO(3)$:

$$\begin{cases} \dot{x} = a\Omega_ux, \\ \dot{R} = \Omega_uR, \end{cases} \qquad v \in \mathbb{R}^3. \tag{7.17}$$

In this setting, adding the no spinning constraint corresponds to the additional constraint $v \perp x$.

7.4. Training problem for residual neural networks

In this section we show how the training problem for a ResNet can be recast as a controllability problem, which can ve solved by the tools developed in this notes. We follow [TG23], and we refer to [AS22; Zua25] for a more general discussion on the control-theoretic approach to machine learning.

A residual neural network (ResNet) is a particular architecture of feedforward neural networks, which has been introduced in [He+16]. Given an input $x(0) \in \mathbb{R}^n$, a ResNet with N layers produces a sequence $(x(k))_{k=0}^N \subset \mathbb{R}^n$ defined by the following recursion:

$$x(k+1) = x(k) + S(k)\sigma\left(W(k)x(k) + b(k)\right), \qquad k = 0, \dots, N. \tag{7.18}$$

The output of the network is $y = x(N)$. Here, $S(k), W(k) \in \mathbb{R}^{n\times n}$ and $b(k) \in \mathbb{R}^n$ are the parameters of the network, and $\sigma : \mathbb{R} \to \mathbb{R}$ is a non-linear activation function, which is applied componentwise to the vector $W(k)x(k) + b(k)$.

The number N is the depth of the network, and it is typically very large in practice. This fact led to the interpretation of ResNets as a discretization of a continuous-time system, where the depth N corresponds to the time horizon, and the parameters $(S(k), W(k), b(k))$ correspond to the control inputs. Indeed, if we let $h = \frac{1}{N}$ and we set $S(k) = hS(t_k)$, $W(k) = hW(t_k)$, and $b(k) = hb(t_k)$, where $t_k = kh$, then the above recursion can be rewritten as

$$\frac{x(k+1) - x(k)}{h} = S(t_k)\sigma\left(W(t_k)x(k) + b(t_k)\right).$$

In the limit $h \to 0$, we obtain the following continuous-time system:

$$\dot{x}(t) = S(t)\sigma\left(W(t)x(t) + b(t)\right) =: f(x(t), u(t)), \qquad t \in [0,1]. \tag{7.19}$$

This is a control system $\dot{x} = f(x,u)$, where the state is $x \in \mathbb{R}^n$ and the control input is $u = (S, W, b)$, which takes values in $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$.

The *training problem* for a ResNet consists in finding parameters such that the network approximates given samples. That is, we assume to be given a dataset $\{(x_0^i, y^i)\}_{i=1}^d \subset \mathbb{R}^n \times \mathbb{R}^n$, and we want to find parameters such that $x^i(N) \approx y^i$ for any $i = 1, \dots, d$. In the continuous-time limit, this corresponds to finding parameters such that $x^i(1) \approx y^i$ for any $i = 1, \dots, d$, where $x^i(\cdot)$ is the solution of (7.19) with initial condition $x^i(0)$. That is, letting $\Phi_u^t(x)$ be the flow of (7.19), the training problem can be recast as the following controllability problem:

$$\text{find } u(\cdot) \text{ such that } \Phi_u^1(x_0^i) = y^i, \qquad i = 1, \dots, d. \tag{7.20}$$

The crucial difference w.r.t. the controllability problems we have considered so far is that here we are looking for a single control input $u(\cdot)$ that steers *multiple* initial conditions to their corresponding targets. This problem is known as the *simultaneous controllability problem*, and it is a much more challenging problem than the standard controllability problem.

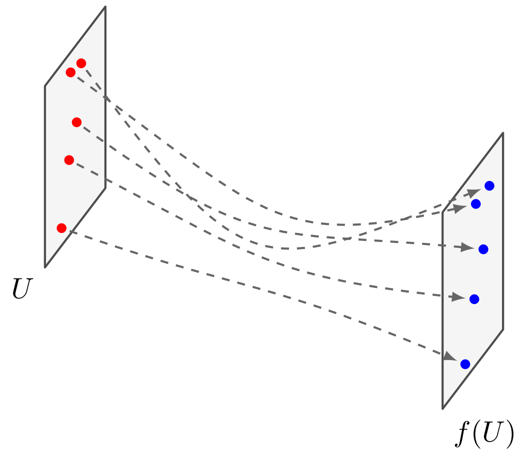


Figure 7.5.: Schematic representation of simultaneous controllability. The same control $u(\cdot)$ is used to steer all points x_0^i to the corresponding points y^i.

Let us introduce some notation. Given $X^1, \dots, X^d \in \mathbb{R}^n$, we let $X = (X^1, \dots, X^d)^\top \in \mathbb{R}^{n\times d}$ and define

$$F(X, u) = \begin{pmatrix} f(X^1, u) & \cdots & f(X^d, u) \end{pmatrix}.$$

Then, problem (7.20) is exactly equivalent to the controllability problem on $\mathbb{R}^{n\times d}$ of the system

$$\dot{X} = F(X, u) \iff \begin{cases} \dot{X}^1 = f(X^1, u), \\ \vdots \\ \dot{X}^d = f(X^d, u). \end{cases} \qquad f(X^i, u) = S\sigma(WX^i + b).$$

We need to introduce some assumptions on sigmoid.

Definition 7.11. The activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, injective. Moreover, there exists $a_0, a_1, a_2 \in \mathbb{R}$, with $a_2 \neq 0$ such that[1]

$$\sigma'(x) = a_0 + a_1\sigma(x) + a_2(\sigma(x))^2, \qquad \forall x \in \mathbb{R}.$$

We also assume[2] that holds that $\sigma(0) = 1$.

Finally, we define the following set

$$\begin{aligned} N &= \left\{X \in \mathbb{R}^{n\times d} \mid \text{for all } j \in \{1, \dots, n\} \text{ there exist } r \neq s \text{ such that } X_j^r = X_j^s\right\} \\ &= \left\{X \in \mathbb{R}^{n\times d} \mid \prod_{1\le r<s\le d} (X_j^r - X_j^s) = 0, \qquad j = 1, \dots, n\right\}. \end{aligned} \tag{7.21}$$

[1]This condition is satisfied, e.g., by the sigmoid activation $\sigma(x) = \frac{1}{1+e^{-x}}$, since $\sigma'(x) = \sigma(x)(1-\sigma(x))$, and by the activation $\sigma(x) = \tanh(x)$.

[2]This is just a normalization condition, which can be achieved by changing the control $(S, W, b) \mapsto (\tilde{S}, W, \tilde{b})$.

The rest of the section is devoted to the proof of the following result.

Theorem 7.12. *Under the stated assumptions, the control system* $\dot{X} = F(X, u)$ *is controllable on* $\mathbb{R}^{n\times d} \setminus N$ *whenever* $d \geq n \geq 2$.

Since there is no constraint on the matrices S, the family of vector fields $\mathcal{F}(X) = \{F(X, u) \mid u = (S, W, b)\}$ associated with the system is symmetric. Hence, we can apply Chow-Rashevskii theorem. To this aim we start by computing the Lie brackets of the vector fields to show that the system is Lie bracket generating.

It turns out that, to this aim, it is not necessary to consider the full family $\mathcal{F}$. More precisely, letting $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{R}^n$, we consider:

- $V_j^{\pm}(X) = F(X, u)$, where $u = (\pm \operatorname{Id}, 0, e_j)$. Namely,

$$V_j^{\pm}(X) = \pm V_j(X), \qquad V_j(X) := \sum_{i=1}^{d} \frac{\partial}{\partial X_j^i}, \qquad j = 1, \dots, n.$$

- $W_{jk}^{\pm}(X) = F(X, u)$, where $u = (\pm \operatorname{Id}, e_j e_k^\top, 0)$ and $e_j e_k^\top$ is the matrix with 1 in the (j, k)-th entry and 0 elsewhere. Namely,

$$W_{jk}^{\pm}(X) = \pm W_{jk}(X), \qquad W_{jk}(X) := \sum_{i=1}^{d} \sigma(X_k^i) \frac{\partial}{\partial X_j^i}, \qquad j, k = 1, \dots, n.$$

Let us condiser the family of vector fields

$$\mathcal{F}_0 = \{V_j, W_{jk} \mid j, k = 1, \dots, n\}.$$

Clearly, $\mathcal{F}_0 \subset \mathcal{F}$ and symmetric. Thus, if we show that $\mathcal{F}_0$ is Lie bracket generating, then the system is controllable. In particular, this amounts to proving that

$$\dim \operatorname{Lie}(\mathcal{F}_0)(X) = nd, \qquad \text{for any } X \in \mathbb{R}^{n\times d} \setminus N.$$

Let us compute the Lie brackets of the above vector fields.

$$[V_k, W_{jk}] = \sum_{i=1}^{d} \sigma'(X_k^i) \frac{\partial}{\partial X_j^i}. \tag{7.22}$$

Letting $\operatorname{ad}_{V_k}^0 W_{jk} = W_{jk}$ and $\operatorname{ad}_{V_k}^{m+1} W_{jk} = [V_k, \operatorname{ad}_{V_k}^m W_{jk}]$ for any $m \in \mathbb{N}_0$, we then have that

$$\operatorname{ad}_{V_k}^m W_{jk} = \sum_{i=1}^{d} \sigma^{(m)}(X_k^i) \frac{\partial}{\partial X_j^i}. \tag{7.23}$$

Since $X \notin N$, we can fix j such that $X_j^r \neq X_j^s$ for all $r \neq s$, and show that the following family of vector fields is linearly independent:

$$\{V_k \mid k = 1, \dots, n\} \cup \{\operatorname{ad}_{V_k}^m W_{jk} \mid m = 0, \dots, n-1\}. \tag{7.24}$$

That is, we introduce the operator vec : $\mathbb{R}^{n\times d} \to \mathbb{R}^{nd}$ which maps a matrix $A \in \mathbb{R}^{n\times d}$ to the vector obtained by concatenating the columns of A. For example,

$$\operatorname{vec}(V_j(X)) = (\underbrace{\underbrace{0,\ldots,0}_{=e_1}, \underbrace{1,0,\ldots,0}_{=e_1}, \ldots, \underbrace{1,0,\ldots,0}_{=e_1}}_{d \text{ times}})^\top.$$

Then, the linear independence of the family is equivalent to the invertibility of the following matrix:

$$G = \begin{pmatrix} \mathrm{Id}_n & \sigma(X_j^1)\,\mathrm{Id}_n & \sigma'(X_j^1)\,\mathrm{Id}_n & \cdots & \sigma^{(d-2)}(X_j^1) \\ \mathrm{Id}_n & \sigma(X_j^2)\,\mathrm{Id}_n & \sigma'(X_j^2)\,\mathrm{Id}_n & \cdots & \sigma^{(d-2)}(X_j^2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Id}_n & \sigma(X_j^d)\,\mathrm{Id}_n & \sigma'(X_j^d)\,\mathrm{Id}_n & \cdots & \sigma^{(d-2)}(X_j^d) \end{pmatrix} \in \mathbb{R}^{nd\times nd}.$$

It is easily observed that G is invertible if and only if the following Vandermonde-like matrix is invertible:

$$M = \begin{pmatrix} 1 & \sigma(X_j^1) & \sigma'(X_j^1) & \cdots & \sigma^{(d-2)}(X_j^1) \\ 1 & \sigma(X_j^2) & \sigma'(X_j^2) & \cdots & \sigma^{(d-2)}(X_j^d) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \sigma(X_j^d) & \sigma'(X_j^d) & \cdots & \sigma^{(d-2)}(X_j^d) \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

This follows from the following result.

Lemma 7.13. *Assume that σ satisfy the assumptions of Definition* (7.11). *Then, the matrix M is invertible if and only if $\sigma(X_j^r) \neq \sigma(X_j^s)$ for all $r \neq s$.*

Proof. Denote $P = P_1$ and set $k_1 = \deg(P_1) = 2$. Observe that $\sigma''(x) = P'(\sigma(x))\sigma'(x) = P'(\sigma(x))P(\sigma(x))$. In particular, $\sigma''(x) = P_2(\sigma(x))$ for some polynomial P_2 of degree $k_2 = k_1 + (k_1 - 1) = 3$. By iterating the above argument, a recursion argument shows that we have $\sigma^{(m)}(x) = P'_{m-1}(\sigma(x))P_1(\sigma(x)) =: P_m(\sigma(x))$ where P_m is a polynomial of degree $k_m = m + 1$.

In particular, setting $s_r = \sigma(X_j^r)$ for any $r = 1, \ldots, d$, we have the generalized Vandermonde matrix:

$$M = \begin{pmatrix} 1 & s_1 & P_1(s_1) & \cdots & P_{d-2}(s_1) \\ 1 & s_2 & P_1(s_2) & \cdots & P_{d-2}(s_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s_d & P_1(s_d) & \cdots & P_{d-2}(s_d) \end{pmatrix}.$$

Clearly, if $s_r = s_s$ for some $r \neq s$, then the r-th and s-th rows of M coincide, and thus M is not invertible.

Assume that $s_r \neq s_s$ for all $r \neq s$, and let us show that M is invertible. Assume by contradiction that the rows of M are not linearly independent. Then there exist $c_0, \ldots, c_{d-2} \in \mathbb{R}$, not all zero and such that the polynomial

$$Q(s) := c_0 + c_1 s + c_2 P_1(s) + c_3 P_2(s) + \cdots + c_{d-2} P_{d-2}(s),$$

satisfies

$$Q(s_r) = 0 \quad \text{for any } r = 1, \dots, d.$$

Observe that this implies that Q is a polynomial of degree at most $k_{d-2} = d - 1$ with at least d distinct roots. This implies that $Q \equiv 0$, which contradicts the fact that $c_0, \dots, c_{d-2}$ are not all zero. □

Thanks to the above lemma, we have that the family (7.4) is linearly independent, and thus $\dim \mathrm{Lie}(\mathcal{F}_0)(X) = nd$ for any $X \in \mathbb{R}^{n \times d} \setminus N$. In order to apply the Chow-Rashevskii Theorem, we are left to prove that the set $\mathbb{R}^{n \times d} \setminus N$ is connected. This is done in the following.

Proposition 7.14. *The set $\mathbb{R}^{n \times d} \setminus N$ is open, and dense in $\mathbb{R}^{n \times d}$. Moreover, if $d, n \geq 2$, it is connected.*

Proof. By its definition (see (7.21)), the set N is a finite union of algebraic varieties, and thus it is closed and has empty interior. Hence, $\mathbb{R}^{n \times d} \setminus N$ is open and dense in $\mathbb{R}^{n \times d}$.

To prove that $\mathbb{R}^{n \times d} \setminus N$ is connected, fix $X, Y \in \mathbb{R}^{n \times d} \setminus N$ and let us show that there exists a continuous curve $\gamma : [0,1] \to \mathbb{R}^{n \times d} \setminus N$ such that $\gamma(0) = X$ and $\gamma(1) = Y$. Given $j \in \{1, \dots, n\}$, we define the curve $\gamma_j : [0,1] \to \mathbb{R}^d$ by $\gamma_j(t) = X_j + t(Y_j - X_j)$, where X_j and Y_j are the j-th rows of X and Y, respectively. Then, γ_j is a continuous curve connecting X_j to Y_j. Consider now the map $\Gamma : [0,1]^n \to \mathbb{R}^{n \times d}$ defined by

$$\Gamma(t_1, \dots, t_n) = \begin{pmatrix} \gamma_1(t_1) \\ \vdots \\ \gamma_n(t_n) \end{pmatrix}.$$

Observe that $\Gamma(\bar{t}_1, \dots, \bar{t}_n) \in N$ if and only if $\gamma_j(\bar{t}_j) \in N_j$ for all j, where $N_j = \{X_j \in \mathbb{R}^d \mid X_j^r = X_j^s \text{ for some } r \neq s\}$ is a finite union of hyperplanes in $\mathbb{R}^d$. In particular, $Z_j := \{\bar{t}_j \in [0,1] \mid \gamma_j(\bar{t}_j) \in N_j\}$ is finite for any j, and thus $Z := \{\bar{t} \in [0,1]^n \mid \Gamma(\bar{t}) \in N\} = \prod_{j=1}^n Z_j$ is finite. Hence, we can fix a continuous curve $\varphi : [0,1] \to [0,1]^n$ connecting $(0, \dots, 0)$ to $(1, \dots, 1)$ such that $\varphi(t) \notin Z$ for any $t \in [0,1]$. Finally, this implies that the curve $\gamma = \Gamma \circ \varphi$ is a continuous curve connecting X to Y and lying entirely in $\mathbb{R}^{n \times d} \setminus N$. □

Part II.

Optimal control

8. Introduction to optimal control

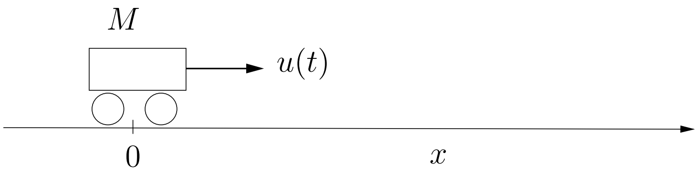
Consider the nonlinear control system

$$\dot{q} = f(q, u(t)). \tag{8.1}$$

Here $q \in M$ (a smooth connected manifold of dimension n, e.g., $\mathbb{R}^n$) is the state of the system, $U \subset \mathbb{R}^m$ is the set of control values and $u(\cdot) : [0, \infty[\to U$ is the control. We assume that f is a smooth function of its arguments and that $u(\cdot)$ is regular enough in such a way that equation (8.1) with the initial condition $q(0) = q_{\text{in}} \in M$ has local existence and uniqueness of solutions (e.g., $u(\cdot)$ is measurable and essentially bounded). Let us denote by $\mathcal{R}(\tau, q_{\text{in}})$ and by $\mathcal{R}(q_{\text{in}})$ the reachable set in time τ and the reachable set (cf. Chapter 5).

A typical problem that one meets in control theory is to find the "best" trajectory for a given criterion steering the system from one initial point q_{in} to a final "target" $\mathcal{T} \subset M$. When $\mathcal{T} = \{q_{\text{fi}}\}$, i.e., a single point, then the final condition is completely fixed. When $\mathcal{T} = M$ the final condition is completely free. All intermediate situations are possible.

Example 8.1. Consider a cart of mass m, moving on a rail (modelled as the real line $\mathbb{R}$), on which we act with an external force $u(t)$ such that $|u(t)| \leq F_{\text{max}}$, where $F_{\text{max}} > 0$.



The corresponding dynamics has the form $m\ddot{x} = u(t)$. Setting $x_1 = x$ and $x_2 = \dot{x}$ the control system becomes.

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = \frac{u}{m} \end{cases} \qquad |u| \leq F_{max}$$

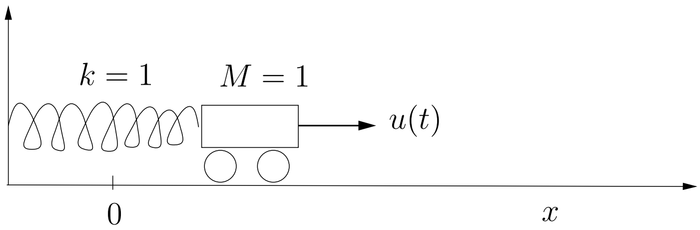
Given the initial position and velocity of the cart, say, $x_1(0) = a_1 \in \mathbb{R}$ and $x_2(0) = a_2 \in \mathbb{R}$, we are interested in bringing the cart to a stop at the origin in minimal time. That is, we look for the trajectory joining $(x_1, x_2) = (a_1, a_2)$ to $(x_1, x_2) = (0, 0)$ in minimum time. Notice that for this problem, initial and final points are fixed, but the final time is free. The criterion can be written in integral form since $T = \int_0^T 1\, dt$.

Example 8.2. Consider the Reed-Shepp car (i.e., the vehicle sliding on ice presented in the Introduction):

$$\begin{cases} \dot{x} = u_1(t)\cos(\theta) \\ \dot{y} = u_1(t)\sin(\theta) \\ \dot{\theta} = u_2(t) \end{cases} \qquad u_1, u_2 \in L^\infty([0,T],\mathbb{R})$$

Find the trajectory joining $(x, y, \theta) = (0, 0, 0)$ to $(x, y, \theta) = (\bar{x}, \bar{y}, \bar{\theta})$ minimizing an "energy-like cost" $\int_0^T (u_1^2 + u_2^2)\, dt$. Notice that for this problem the final time and the initial and final points are fixed. A variant of this problem is to start from the point $(x, y, \theta) = (0, 0, 0)$ and to reach $(x, y) = (\bar{x}, \bar{y})$ with any orientation and minimizing the same criterium.

Example 8.3. Consider an harmonic oscillator of mass $M = 1$ and elastic constant $k = 1$, on which we act with an external force $u(t)$ such that $|u(t)| \le 1$.



The corresponding dynamics has the form $\ddot{x} = -x + u(t)$. Setting $x_1 = x$ and $x_2 = \dot{x}$ the control system becomes,

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 + u(t) \end{cases} \qquad |u(t)| \le 1$$

Find the trajectory starting from $(x_1, x_2) = (0, 0)$ and maximizing $x_1^2(1)$ (i.e. minimizing $-x_1^2(1)$). Notice that for this problem the final time and the initial point are fixed, but the final point is not. Moreover the cost is not in integral form (it only depends on the final position).

All these problems can be written in the following form

Definition 8.4 (Optimal Control Problem (OCP))**.**

$$\begin{cases} \dot{q}(t) = f(q(t), u(t)) \\ q(0) = q_{\text{in}}, \quad q(T) \in \mathcal{T} \\ \int_0^T f^0(q(t), u(t))\, dt + \phi(q(T)) \to \min \\ T > 0 \text{ can be fixed or free.} \end{cases} \tag{OCP}$$

Here,

- M is a smooth n-dimensional connected manifold, $U \subset \mathbb{R}^m$,
- $\mathcal{T}$ is a smooth submanifold of M,
- f, f^0 are smooth functions of their arguments,
- $u(\cdot) \in L^\infty([0,T], U)$,
- $q(\cdot) : [0,T] \to M$, belongs to the set of Lipschitz curves.

Notice that for the problem (OCP) to make sense it is not necessary to assume that one can prove controllability of the system. Indeed, in applications, it happens often that one cannot guarantee the following condition to be satisfied:

Condition: $\mathcal{T} \cap \mathcal{R}(q_{\text{in}}) \neq \emptyset$ if T is free or $\mathcal{T} \cap \mathcal{R}^T(q_{\text{in}}) \neq \emptyset$ if T is fixed.

In this case usually one considers $\mathcal{T} = M$ and selects as ϕ something that represents the distance between the final point and the desired target.

8.1. General procedure to solve an optimal control problem

An optimal control problem is actually a minimization problem in infinite-dimensional space. Indeed, letting $\Omega(q_{\text{in}}, \mathcal{T})$ be the set of admissible controls steering the system from q_{in} to $\mathcal{T}$ (with time T fixed or free), we can write (OCP) as

$$\min_{u(\cdot)\in\Omega(q_{\text{in}},\mathcal{T})} J(u(\cdot)), \qquad J(u(\cdot)) := \int_0^T f^0(q(t;q_{\text{in}},u(\cdot)),u(t))\,dt + \phi(q(T;q_{\text{in}},u(\cdot))). \tag{8.2}$$

For this reason, the steps to determine a solution to (OCP) are quite similar to those used to find the minimum of a smooth function $f^0 : \mathbb{R} \to \mathbb{R}$. Namely:

0. **Find conditions which guarantee the existence of solutions.** Recall that among smooth functions $f^0 : \mathbb{R} \to \mathbb{R}$, it is easy to find examples not admitting a minimum (e.g., the function $x \mapsto e^{-x}$ and the function $x \mapsto x$ do not have minima). This step is crucial. If it is skipped, first-order conditions may give a wrong candidate for optimality (see below for details) and numerical optimization schemes may either not converge or converge towards a solution which is not a minimum.

   For optimal control problems, there exist several existence tests, but they are not always applicable or easy to use. In Section 8.2, we present the Filippov test.

1. **Apply first-order necessary conditions.** For a smooth function $f^0 : \mathbb{R} \to \mathbb{R}$, this means that if $\bar{x}$ is a minimum then $\frac{d}{dx} f^0(\bar{x}) = 0$. This condition only gives candidates for minima, i.e., it cannot distinguish between local minima, local maxima, and saddles. Note that if one does not verify a priori existence of minima, first-order conditions could give wrong candidates. Think for instance to the function

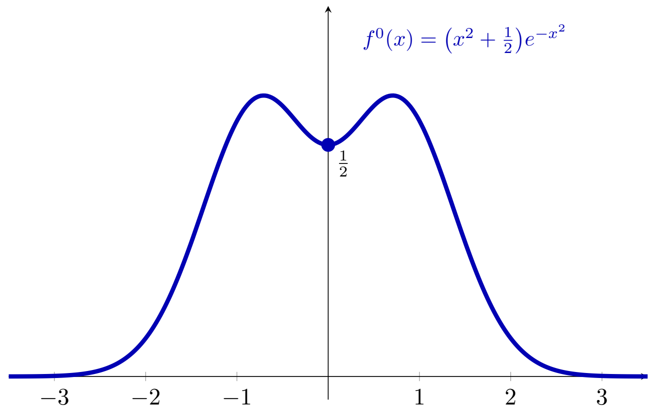


Figure 8.1.: Example of a function without a global minimum where the application of first-order conditions leads to a wrong candidate.

$x \mapsto (x^2 + 1/2)e^{-x^2}$. This function has a single local minimum, obtained at $x = 0$, whose value is $1/2$, which is well identified by first-order conditions. However its infimum is zero (for $x \to \pm\infty$, the function tends to zero). See Figure 8.1

For optimal control problems, first-order necessary conditions should be given in an infinite-dimensional space (a space of curves) and they are expressed by the Pontryagin Maximum Principle, which is presented in Section 9.1. Note that the condition that the system reaches exactly the target is a *constraint* leading to the appearance of *Lagrange multipliers (normal and abnormal).* This point is discussed in details in Section 8.3.1.

2. **Apply second-order conditions.** For instance, for a smooth function $f^0 : \mathbb{R} \to \mathbb{R}$, among the points for which we have $\frac{d}{dx} f^0(\bar{x}) = 0$, a necessary condition to have a minimum is $\frac{d^2}{dx^2} f^0(\bar{x}) \geq 0$. This step is generally used to reduce further the candidates for optimality.

   For optimal control problems, there are several second-order conditions as higher-order Pontryagin Maximum Principles or Legendre–Clebsch conditions (see for instance [AS04; BP04]). In some cases, this step is difficult and it could be more convenient to go directly to the next one.

3. **Selection of the best solution among all candidates.** Among the set of candidates for optimality identified in step 1 and (possibly) further reduced in step 2, one should select the best one. This step is often done by hand if the previous steps have identified a finite number of candidates for optimality.

   For optimal control problems, one often ends up with infinitely many candidates for optimality and this step is generally very difficult.

There are of course specific examples for which the solution is particularly simple. This is the case of convex problems, for which only first-order conditions should be applied, since the existence step is automatic and first-order conditions are both necessary and sufficient for optimality.

8.2. Existence of solutions for Optimal Control Problems: the Filippov test

The existence theory for optimal control is difficult and, unfortunately, there is no general procedure that can be applied in any situation. In this section, we present the most important technique, the Filippov test that allows to tackle several types of problems. We emphasize that it is fundamental to verify the existence of optimal controls before applying first-order conditions (i.e., the Pontryagin Maximum Principle). Otherwise, as discussed in the finite-dimensional case, it may occur that the Pontryagin Maximum Principle has solutions, but none of them is optimal.

To simplify the exposition, let us consider the problem (OCP) with without final cost ($\phi = 0$) and T fixed.

In order to tackle the existence problem, we define a new variable q^0 obtained as the value of the cost during the time-evolution, that is,

$$q^0(t) = \int_0^t f^0(q(s), u(s))\, ds,$$

and we denote $\hat{q} = (q^0, q)$. The dynamics of the new state $\hat{q}$ in $\mathbb{R} \times M$ are given by

$$\begin{aligned} \dot{\hat{q}} &= \begin{pmatrix} \dot{q}^0(t) \\ \dot{q}(t) \end{pmatrix} = \begin{pmatrix} f^0(q(t), u(t)) \\ f(q(t), u(t)) \end{pmatrix} =: \hat{f}(q(t), u(t)) \\ \dot{q}(0) &= (0, q_{\text{in}}) \\ \hat{q}(T) &\in \mathcal{T} \end{aligned}$$

This control system is called the *augmented system*. The importance of introducing this point of view is that it allows to convert the minimization problem in integral form to a problem of minimization of one of the coordinates at the final time. That is,

$$\min \int_0^T f^0(q(t), u(t))\, dt \quad \longleftrightarrow \quad \min q^0(T) \tag{8.3}$$

We denote by $\hat{\mathcal{R}}^T(0, q_{\text{in}})$ the reachable set in time T for the augmented system, starting from $(0, q_{\text{in}})$. (See Definition 3.2.) The key observation on which optimal control is based is expressed by the following proposition.

Proposition 8.5. *If $q(\cdot)$ is an optimal trajectory for problem* (OCP)*, then $\hat{q}(T) \in \partial\hat{\mathcal{R}}^T(0, q_{\text{in}})$.*

Proof. By contradiction, if $\hat{q}(T) = (q^0(T), q(T)) \in \operatorname{int} \hat{\mathcal{R}}^T(0, q_{\text{in}})$ then there exists a trajectory reaching a point $(\alpha, q(T))$ with $\alpha < q^0(T)$, i.e., arriving at the same point in M, but with a smaller cost. See Figure 8.2. □

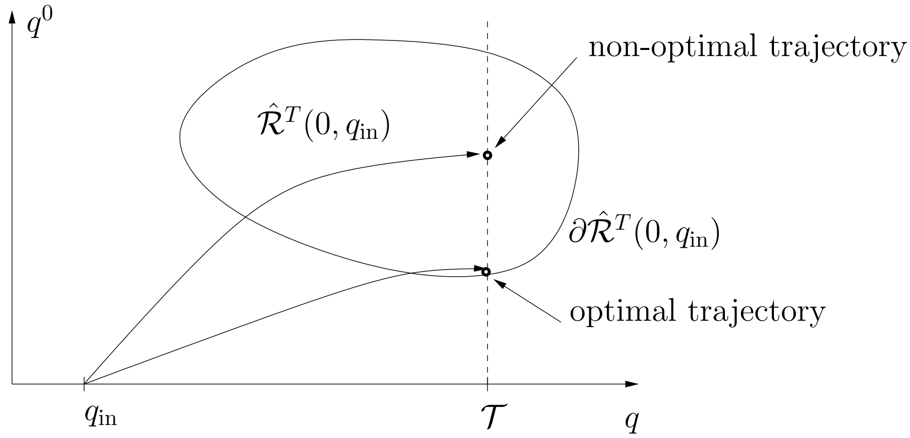


Figure 8.2.: The reachable set of the augmented system

When $\hat{\mathcal{R}}^T(0, q_{\text{in}}) \cap (\mathbb{R} \times \mathcal{T})$ is nonempty and compact, an optimal trajectory for problem (OCP) exists, as it can be deduced at once by taking any converging subsequence of a minimizing sequence. Hence we have the following.

Proposition 8.6. *If $\hat{\mathcal{R}}^T(0, q_{\text{in}})$ is compact, $\mathcal{T}$ is closed, and $\hat{\mathcal{R}}^T(0, q_{\text{in}}) \cap (\mathbb{R} \times \mathcal{T})$ is nonempty, then there exists a solution to problem* (OCP).

It follows that the compactness of $\hat{\mathcal{R}}^T(0, q_{\text{in}})$ is a key point. The following result gives a sufficient condition for compactness of the reachable set (see, e.g., [Lib11] for a proof).

Theorem 8.7 (Filippov)**.** *Consider the control system*

$$\dot{q}(t) = f(q(t), u(t)), \qquad q \in M.$$

Assume that $T > 0$ and $q_{\text{in}} \in M$ are fixed, and that the following conditions hold:

- *the controls are taken in $L^\infty([0, T], U)$, where $U \subset \mathbb{R}^m$ is a compact set,*
- *we have that $\mathcal{F}(q)$ is convex for every $q \in M$, where $\mathcal{F} := \{f(\cdot, u) \mid u \in U\} \subset \mathrm{Vec}(M)$ is the family associated with the control system,*
- *for every $u \in L^\infty([0, T], U)$, the solution of $\dot{q}(t) = f(q(t), u(t))$, $q(0) = q_{\text{in}}$, is defined on the whole interval $[0, T]$.*

Then, the sets $\mathcal{R}(T, q_{\text{in}})$ and $\mathcal{R}(\leq T, q_{\text{in}})$ are compact.

The conclusion of Filippov's theorem may fail to hold if we drop the convexity assumption on the set $\mathbf{F}(q)$ of admissible velocities, as illustrated by the following example.

Example 8.8. Take $M = \mathbb{R}^2$, $U = \{1, 2\}$, and

$$f(q, u) = A_u q, \qquad \text{with} \quad A_1 = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Notice that the corresponding control system satisfies all the assumptions of Th. 8.7 except for the convexity of $\mathbf{F}(q)$. Pick $q_{\text{in}} = (1, 0)$ and any $T > 0$. Then $q_T := e^{-T} q_{\text{in}}$ is in the closure of $\mathcal{R}^T(q_{\text{in}})$, since we can end up arbitrarily close to q_T at time T by applying a control that switches fast enough between 1 and 2. Indeed, by Theorem 6.10, we have

$$q_T = e^{-\frac{T}{2}(A_1+A_2)} q_{\text{in}} \in \overline{\mathcal{R}(T, q_{\text{in}})}, \qquad \frac{1}{2}(A_1 + A_2) = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Let us show that, nevertheless, $q_T \notin \mathcal{R}(T, q_{\text{in}})$. Indeed, for $i = 1, 2$ we have

$$\begin{pmatrix} x \\ y \end{pmatrix}^\top A_i \begin{pmatrix} x \\ y \end{pmatrix} = -x^2 \implies \frac{d}{dt}\frac{1}{2}\|q(t)\|^2 = q(t)^\top \dot{q}(t) = q(t)^\top A_{u(t)} q(t) \geq -\|q(t)\|^2, \quad (8.4)$$

with equality if and only if $q(t)$ is on the horizontal axis. Since every trajectory of the control system starting at q_{in} necessarily leaves the horizontal axis, we deduce that

$$\|q(T)\| > e^{-T}\|q_{\text{in}}\| = e^{-T}.$$

This proves that $q_T \notin \mathcal{R}^T(q_{\text{in}})$. Hence, the set $\mathcal{R}^T(q_{\text{in}})$ is not closed, and the conclusion of Theorem 8.7 does not hold.

Note that the third hypothesis of Theorem 8.7 is automatically satisfied when M is compact. By applying this theorem to the augmented system for problem (OCP), one obtains:

Proposition 8.9. *Fix $T > 0$ and $q_{\text{in}} \in M$. Assume that*

- *$\mathcal{T}$ is closed and $\mathcal{R}^T(q_{\text{in}}) \cap \mathcal{T} \neq \emptyset$,*
- *the set U is compact,*
- *the set $\hat{\mathcal{F}}(q)$ is convex for every $q \in M$, where*

$$\hat{\mathcal{F}}(q) = \left\{ \begin{pmatrix} f^0(q, u) \\ f(q, u) \end{pmatrix} \mid u \in U \right\}$$

  is the family associated with the augmented system,
- *for every $u(\cdot) \in L^\infty([0, T], U)$ the solution of $\dot{q}(t) = f(q(t), u(t))$, $q(0) = q_{\text{in}}$, is defined on the whole interval $[0, T]$.*

Then there exists a solution to problem (OCP)*.*

The idea of reducing the problem of existence of an optimal control to the compactness of the reachable set of the augmented system can be used for more general problems. For instance, if we add a terminal cost $\phi(q(T))$ to the cost $\int_0^T f^0(q(t), u(t))\,dt$, where ϕ is a smooth function (see Definition 8.4), we get a similar result adding to $f^0(q(t), u(t))$ the directional derivative of ϕ along $f(q(t), u(t))$, that is, replacing $f^0(q(t), u(t))$ by $f^0(q(t), u(t)) + \langle d\phi(q(t)), f(q(t), u(t))\rangle$. Here $d\phi(q(t))$ denotes the differential of ϕ evaluated at $q(t)$ and we recall that $\langle \cdot, \cdot \rangle$ is the duality product between covectors of T^*M

and vectors of TM (see Section 9.1 for the definition of T^*M). We leave the details as an exercise.

When the final time is free, it is more difficult to get the existence of optimal trajectories. However, the compactness of $\mathcal{R}(\leq T, q_{\text{in}})$ in Filippov Theorem can be used to find conditions for the existence of optimal controls in minimum time. We state this result in the case where M is compact and we leave its proof as an exercise. Note that the problem of minimizing time can be written in the form of problem (OCP) with T free and $f^0 = 1$.

Proposition 8.10. *Consider problem* (OCP) *with T free, $f^0 = 1$, and M compact. Assume that*

- *$\mathcal{T}$ is closed and $\mathcal{R}(q_{\text{in}}) \cap \mathcal{T} \neq \emptyset$,*
- *the set U is compact,*
- *the set $\mathcal{F}(q) = \{f(q,u) \mid u \in U\}$ is convex for every $q \in M$.*

Then there exists a solution to the problem.

Example 8.11. Consider the time evolution of the wave function of a N-level closed quantum system. In this case, the dynamics are governed by the Schrödinger equation (in units where $\hbar = 1$)

$$i\dot{\psi}(t) = \left(H_0 + \sum_{j=1}^{m} u_j(t) H_j \right) \psi(t), \tag{8.5}$$

where ψ, the wave function, belongs to the unit sphere in $\mathbb{C}^N$ and $H_0, \ldots, H_m$ are $N \times N$ Hermitian matrices. The control is $u(t) = (u_1(t), \ldots, u_m(t))$.

This control problem has the form $\dot{q} = f(q, u(t))$ where $M = \mathbb{S}^{2N-1} \subset \mathbb{C}^N$, $q = \psi$, and $f(\psi, u) = -i(H_0 + \sum_j^m u_j H_j)\psi$. Consider now the problem of steering an initial state ψ_{in} to a final state ψ_{fi} in minimum time with $U = [-1,1]^m$. We assume the controllability of the system. Since U and $M = \mathbb{S}^{2n-1}$ are compact and $\{(H_0 + \sum_{j=1}^m u_j H_j)\psi \mid u \in U\}$ is convex for every $\psi \in M$, we deduce the existence of an optimal solution. Observe that, in this case, the convexity of $\mathcal{F}(q)$ follows from the convexity of U.

8.3. First order conditions

For a smooth real-valued function of one variable $f^0 : \mathbb{R} \to \mathbb{R}$, first-order optimality conditions are obtained from the observation that, at points where $\frac{df^0}{dx} \neq 0$, the function f^0 is well approximated by its first-order Taylor series and hence cannot be optimal since it behaves locally as an affine (non-constant) function. In this way, one obtains the necessary condition: *If $\bar{x}$ is minimal for f^0 then $\frac{df^0}{dx}(\bar{x}) = 0$.* First-order conditions in optimal control are derived in the same way. We have to require that for a small control variation, there is no cost variation at first order.

More precisely, letting $J(u(\cdot))$ denote the value of the cost in (OCP) for a reference admissible control $u(\cdot)$ (see (8.2) in Section 8.1), and $v(\cdot)$ is another admissible control, one would like to consider a condition of the form

$$\left.\frac{\partial J(u(\cdot)+hv(\cdot))}{\partial h}\right|_{h=0} = 0. \tag{8.6}$$

But difficulties may arise for various reasons. The main ones are the following:

- *Infinite-dimensional space.* We work in an infinite-dimensional space (the space of controls) and hence condition (8.6) should be required for infinitely many $v(\cdot)$.
- *Control constraints.* It may very well happen that if $u(\cdot)$ and $v(\cdot)$ are admissible controls then $u(\cdot)+hv(\cdot)$ is not admissible for every h close to 0. Think, for instance, to the case in which $m = 1$ and $U = [a, b]$. If $u(t) \equiv b$ is the reference control, then $u(t) + hv(t)$ is not admissible for any non-zero perturbation $v(\cdot)$ when $hv(t)$ is strictly positive. Hence, one should be very careful in choosing the admissible variations in order to fulfill the control restrictions.
- *Target constraint.* One should restrict only to control variations for which the corresponding trajectory reaches the target. More precisely, if $\tilde{q}(\cdot)$ is the trajectory corresponding to the control $\tilde{u}(\cdot) := u(\cdot) + hv(\cdot)$, one should add the condition

$$\tilde{q}(T) \in \mathcal{T}, \tag{8.7}$$

  with T either free or constrained to be the fixed final time depending on the problem under study. Condition (8.7) is a *constraint* for the minimization problem, forces the introduction of *Lagrange multipliers* (normal and abnormal).

The occurrence of Lagrange multipliers in optimal control is therefore not due to the fact that the optimization takes place in an infinite-dimensional space, but is rather a general feature of constrained minimization problems, as we explain in the next section.

8.3.1. Why Lagrange multipliers appear in constrained problems

We first recall how to find the minimum of a function of n variables $f^0(x)$, where $x = (x_1, \ldots, x_n)$, under the constraint $f(x) = 0$, with the method of Lagrange multipliers. Here f^0 and f are two smooth functions $\mathbb{R}^n \to \mathbb{R}$. We have two cases.

1. If $\bar{x}$ is a point such that $f(\bar{x}) = 0$ with $\nabla f(\bar{x}) \neq 0$, then the implicit function theorem guarantees that $\{x \mid f(x) = 0\}$ is a smooth hypersurface in a neighborhood of $\bar{x}$. In this case, a necessary condition for f^0 to have a minimum at $\bar{x}$ is that the level set of f^0 (i.e., the set on which f^0 takes a constant value) is not transversal to the set $\{x \mid f(x) = 0\}$ at $\bar{x}$. See Figure 8.3.

   More precisely, this means that

$$\exists \lambda \in \mathbb{R} \text{ such that } \nabla f^0(\bar{x}) = \lambda \nabla f(\bar{x}). \tag{8.8}$$

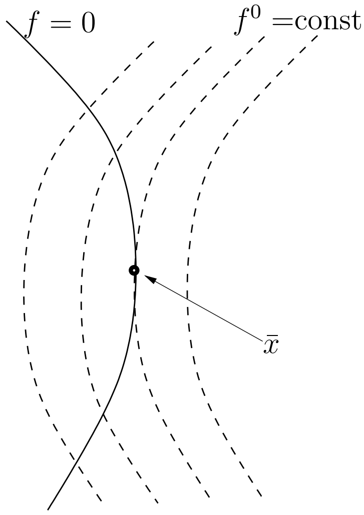


Figure 8.3.: Intersection of the set $f(x) = 0$ (solid line) with the level set of f^0 (dashed line). The two gradients $\nabla f(x)$ and $\nabla f^0(x)$ are parallel at $x = \bar{x}$.

This statement can be proved by assuming, for instance, that $\partial_{x_n} f(\bar{x}) \neq 0$. The set $\{x \mid f(x) = 0\}$ can then be expressed locally around $\bar{x}$ as $x_n = g(x_1, \ldots, x_{n-1})$. The requirement that

$$\begin{cases} \partial_{x_i} f(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) \equiv 0, & i = 1, \ldots, n-1, \\ \partial_{x_i} f^0(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1}))|_{x=\bar{x}} = 0, & i = 1, \ldots, n-1, \end{cases}$$

provides immediately condition (8.8) with $\lambda = \partial_{x_n} f^0(\bar{x}) / \partial_{x_n} f(\bar{x})$.

Notice that λ could be equal to zero. This case corresponds to the situation in which f^0 has a critical point at $\bar{x}$ even in absence of the constraint.

2. If $\bar{x}$ is a point such that $f(\bar{x}) = 0$ with $\nabla f(\bar{x}) = 0$ then the set $\{x \mid f(x) = 0\}$ could be very complicated in a neighborhood of $\bar{x}$ (typical examples are a single point, two crossing curves, ... but it could be any closed set). In general, the value of f^0 at these points cannot be compared with neighboring points by requiring that a certain derivative is zero (think for instance to the case in which $\{x \mid f(x) = 0\}$ is an isolated point). However, they are candidates to optimality. As an illustrative example, consider the case where $n = 2$, $f^0(x_1, x_2) = x_1^2 + (x_2 - 1/4)^2$, and $f(x_1, x_2) = (x_1^2 + x_2^2)(x_1^2 + x_2^2 - 1)$. The point $\bar{x}$ is an isolated point for which $f(\bar{x}) = 0$ and $\nabla f(\bar{x}) = 0$.

These results can be written in a unified way as follows.

Theorem 8.12 (Lagrange multiplier rule in $\mathbb{R}^n$). *Let f^0 and f be two smooth functions from $\mathbb{R}^n$ to $\mathbb{R}$. If f^0 has a minimum at $\bar{x}$ on the set $\{x \mid f(x) = 0\}$, then there exists*

$(\bar{\lambda}, \bar{\lambda}_0) \in \mathbb{R}^2 \setminus \{(0,0)\}$ *such that, setting* $\Lambda(x, \lambda, \lambda_0) = \lambda f(x) + \lambda_0 f^0(x)$*, we have*

$$\nabla_x \Lambda(\bar{x}, \bar{\lambda}, \bar{\lambda}_0) = 0, \;\; \nabla_\lambda \Lambda(\bar{x}, \bar{\lambda}, \bar{\lambda}_0) = 0. \tag{8.9}$$

To show that this statement is equivalent to what we just discussed, we observe that the second equality in (8.9) gives the constraint $f(\bar{x}) = 0$. For the first equation, we have two cases. If $\bar{\lambda}_0 \neq 0$ then we can normalize $\bar{\lambda}_0 = -1$ and we get $\bar{\lambda}\nabla_x f(\bar{x}) - \nabla_x f^0(\bar{x}) = 0$, i.e., Eq. (8.8) with the change of notation $\lambda \to \bar{\lambda}$. If $\bar{\lambda}_0 = 0$ then $\bar{\lambda} \neq 0$ and we get $\nabla_x f(\bar{x}) = 0$, that is, the second case studied above.

The quantities $\bar{\lambda}$ and $\bar{\lambda}_0$ are respectively called *Lagrange multiplier* and *abnormal Lagrange multiplier*. If $(\bar{x}, \bar{\lambda}_0, \bar{\lambda})$ is a solution of Eq. (8.9) with $\bar{\lambda}_0 \neq 0$ (resp., $\bar{\lambda}_0 = 0$) then $\bar{x}$ is called a *normal extremal* (resp., *abnormal extremal*). An abnormal extremal is a candidate for optimality and occurs, in particular, when we cannot guarantee (at first order) that the set $\{x \mid f(x) = 0\}$ is a smooth curve. Abnormal extremals are candidates for optimality regardless of cost f^0. Note that if $\bar{x}$ is such that $\nabla_x f(\bar{x}) = 0$ and $\nabla_x f^0(\bar{x}) = 0$ then $\bar{x}$ is both normal and abnormal. This is the case in which $\bar{x}$ satisfies the first-order condition for optimality even without the constraint, but we cannot guarantee that the constraint is a smooth curve.

In the (infinite-dimensional) case of an optimal control problem, normal and abnormal Lagrange multipliers appear in a very similar way. We are going to see this in the next chapter.

9. The Pontryagin Maximum Principle (PMP): Formulation and Use

In this chapter, we state the first-order necessary conditions for optimal control problems, namely the Pontryagin Maximum Principle (PMP for short).

Notation: the cotangent bundle. Recall that the tangent space T_qM at $q \in M$ is the set of all vectors v tangent to M at q. The set of all tangent spaces is called the tangent bundle and it is indicated as TM. A covector is a linear functional $p : T_qM \to \mathbb{R}$. The set of all covectors acting on vectors in T_qM is indicated as T_q^*M (the contangent space at $q \in M$). The disjoint union of all cotangent spaces is called the cotangent bundle and it is indicated as T^*M. More precisely, the action of $p \in T_q^*M$, on $v \in T_qM$, is called the *duality product* and it is denoted as

$$\langle p, v \rangle.$$

Locally, T^*M can be identified with couples $(q, p) \in M \times \mathbb{R}^n$ and a covector $p \in T_q^*M$ can be represented as a row vector. Then, we write

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \qquad p = (p_1, \ldots, p_n), \qquad \langle p\, v \rangle = pv = p_1v_1 + \ldots + p_nv_n.$$

9.1. Statement of the Pontryagin Maximum Principle

The basic idea is to define a new object (the pre-Hamiltonian, see Eq. (11.6) below) which allows to formulate the Langrange multiplier conditions in a simple and direct way.

Theorem 9.1 (Pontryagin Maximum Principle)**.** *Consider the optimal control problem*

$$\begin{cases} \dot{q}(t) = f(q(t), u(t)), \\ q(0) = q_{\text{in}}, \quad q(T) \in \mathcal{T}, \\ \displaystyle\int_0^T f^0(q(t), u(t))\; dt + \phi(q(T)) \longrightarrow \min, \end{cases} \tag{9.1}$$

where

- *M is a smooth connected manifold of dimension n, $U \subset \mathbb{R}^m$;*

- *$\mathcal{T}$ is a (non-empty) smooth submanifold of M. It can be reduced to a point (fixed terminal point) or coincide with M (free terminal point),*
- *f, f^0 are smooth, as is ϕ;*
- *$u(\cdot) \in L^\infty([0,T], U)$;*
- *$q : [0,T] \to M$ is a Lipschitz continuous curve;*
- *$T > 0$ is fixed or free.*

Define the pre-Hamiltonian *function $\mathcal{H}$ as*

$$\mathcal{H}(q, p, u, p^0) = \langle p, f(q,u)\rangle + p^0 f^0(q,u), \tag{9.2}$$

with

$$(q, p, u, p^0) \in T^*M \times U \times \mathbb{R}.$$

Assume that the pair $(q(\cdot), u(\cdot)) : [0,T] \to M \times U$ be optimal, then there exists a never vanishing Lipschitz continuous pair $(p(\cdot), p^0(\cdot)) : [0,T] \ni t \mapsto (p(t), p^0) \in T^*_{q(t)}M \times \mathbb{R}$ where $p^0(t) = p^0 \leq 0$ is constant and such that for almost every (a.e.) $t \in [0,T]$ we have

i) Hamiltonian equations*:*

$$\begin{cases} \dot{q}(t) = \dfrac{\partial \mathcal{H}}{\partial p}(q(t), p(t), u(t), p^0) \\ \dot{p}(t) = -\dfrac{\partial \mathcal{H}}{\partial q}(q(t), p(t), u(t), p^0) \end{cases}$$

ii) Maximization condition*: the function $\mathcal{H}_M(q(t), p(t), p^0) := \max_{v \in U} \mathcal{H}(q(t), p(t), v, p^0)$, called the* maximized Hamiltonian*, is well-defined and*

$$\mathcal{H}(q(t), p(t), u(t), p^0) = \mathcal{H}_M(q(t), p(t), p^0).$$

Moreover,

iii) Value of the Hamiltonian*: there exists a constant c such that $\mathcal{H}_M(q(t), p(t), p^0) = c$ on $[0,T]$, with $c = 0$ if the final time is free.*

iv) Transversality condition*: we have*

$$\langle p(T), v\rangle = p^0 \langle d\phi(q(T)), v\rangle, \qquad \text{for any } v \in T_{q(T)}\mathcal{T}. \tag{9.3}$$

Here, $d\phi$ is the differential of the function ϕ, defined in Definition 2.10.

Some comments are in order.

- The covector p is called *adjoint state* in control theory. The quantities $p(\cdot)$ and p^0 play the role of Lagrange multipliers for the constrained optimization problem. We point out the similarity between the expressions of $\mathcal{H}$ and of Λ in Theorem 8.12 (with the change of notation and $p \to \lambda$).

- A trajectory $q(\cdot)$ for which there exist $p(\cdot)$, $u(\cdot)$ and p^0 such that $(q(\cdot), p(\cdot), u(\cdot), p^0)$ satisfies all the conditions given by the PMP is called an *extremal trajectory* and the 4-uple $(q(\cdot), p(\cdot), u(\cdot), p^0)$ an *extremal* or, equivalently, an *extremal lift of* $q(\cdot)$. Such an extremal is called *normal* if $p^0 \neq 0$ and *abnormal* if $p^0 = 0$. It may happen that an extremal trajectory $q(\cdot)$ admits both a normal extremal lift $(q(\cdot), p_1(\cdot), u(\cdot), p^0)$ and an abnormal one $(q(\cdot), p_2(\cdot), u(\cdot), 0)$. In this case, we say that the extremal trajectory $q(\cdot)$ is a *non-strict abnormal trajectory.*

  Note that (as in the finite-dimensional case) abnormal trajectories are candidates for optimality regardless of the cost. In the finite-dimensional case they correspond to singularities of the constrain function, while here they correspond to singularities of the functional associating with a control $v(\cdot)$ the endpoint at time T of the solution of $\dot{q}(t) = f(q(t), v(t))$, $q(0) = q_{\text{in}}$.

- The PMP is only a necessary condition for optimality. It may very well happen that an extremal trajectory is not optimal. The PMP can therefore provide several candidates for optimality, only some of which are optimal (or even none of them if the step of existence has not been verified, see Section 8.2).

- Since the Hamiltonian equation for $p(\cdot)$ in the PMP is linear, if $(q(\cdot), p(\cdot), u(\cdot), p^0)$ is an extremal, then for every $\alpha > 0$, $(q(\cdot), \alpha p(\cdot), u(\cdot), \alpha p^0)$ is an extremal as well. As a consequence, some useful normalizations are possible. A typical normalization for normal extremals is to require $p^0 = -\frac{1}{2}$ but other choices are also possible.

- When there is no final cost ($\phi = 0$), the transversality condition **iv)** simplifies to:

$$\langle p(T), T_{q(T)}\mathcal{T} \rangle = 0. \tag{9.4}$$

  When the final point is fixed ($\mathcal{T} = \{q_{\text{fi}}\}$), $T_{q(T)}\mathcal{T}$ is a zero-dimensional manifold and hence condition (9.4) is empty. When the final point is free ($\mathcal{T} = M$) the transversality condition simplifies to $p(T) = p^0 d\phi(q(T))$. In local coordinates, we recover that $p(T)$ is proportional to the gradient of ϕ evaluated at the point $q(T)$. Notice that, since $(p(T), p^0) \neq 0$, in this case one necessarily has $p^0 \neq 0$.

9.2. Use of the Pontryagin Maximum Principle

The application of the PMP is not so straightforward. Indeed, there are many conditions to satisfy and all of them are coupled. This section is aimed at describing how to use it in practice.

The following points should be followed first for normal extremals ($p^0 < 0$) and then for abnormal extremals ($p^0 = 0$). Recall that in the normal case, p^0 can be normalized to $-1/2$, since p^0 is defined up to a multiplicative positive factor. In the different steps, several difficulties (that are briefly mentioned) may arise. Most of them should be solved case by case, since they can be of different nature depending on the problem under study.

Step 1. Use the maximization condition **ii)** to express, when possible, the control as a function of the state and of the covector, i.e., $u = w(q, p)$. Note that if we have m controls (e.g., if U is an open subset of $\mathbb{R}^m$) then the first-order maximality conditions yield m equations for m unknowns. When the maximization condition allows to express u as a function of q and p, we say that the control is *regular*, otherwise we say that it is *singular*. In T^*M, we may have regions where the control is regular and regions where it is singular. For singular controls, finer techniques have to be used to derive the expression of the control. This point is discussed in the examples.

Step 2. Insert the control found in the previous step into the Hamiltonian equations **i)** to get a closed-loop system for $q(\cdot)$ and $p(\cdot)$:

$$\begin{cases} \dot{q}(t) = \dfrac{\partial \mathcal{H}}{\partial p}(q(t), p(t), w(q(t), p(t)), p^0) \\ \dot{p}(t) = -\dfrac{\partial \mathcal{H}}{\partial q}(q(t), p(t), w(q(t), p(t)), p^0). \end{cases} \tag{9.5}$$

In case the previous step provides a smooth $w(\cdot, \cdot)$, this is a well-defined set of $2n$ equations for $2n$ unknown. Note, however, that the boundary conditions are given in a non-standard form since we know $q(0)$ but not $p(0)$. Instead of $p(0)$, we have a partial information on $q(T)$ and $p(T)$ depending on the dimension of $\mathcal{T}$ (see the next step to understand how these final conditions are shared between $q(T)$ and $p(T)$). We then solve Eq. (9.5) for *fixed* $q(0) = q_{\text{in}}$ and *any* $p(0) = p_{\text{in}} \in T^*_{q_{\text{in}}}M$. Let us denote the solution as

$$q(t; p_{\text{in}}, p^0), \qquad p(t; p_{\text{in}}, p^0). \tag{9.6}$$

We stress that when $w(\cdot, \cdot)$ is not regular enough, solutions to the Cauchy problem (9.5) with $q(0) = q_{\text{in}}$ and $p(0) = p_{\text{in}}$ may fail to exist or to be unique.

Step 3. Find p_{in} such that

$$q(T; p_{\text{in}}, p^0) \in \mathcal{T}. \tag{9.7}$$

Note that if $\mathcal{T}$ is reduced to a point and T is fixed, we get n equations for n unknown (the components of p_{in}). If T is free then an additional equation is needed. This condition is given by the relation **iii)** in the PMP. If $\mathcal{T}$ is a k-dimensional submanifold of M ($k \leq n$) then Eq. (9.7) provides only $n - k$ equations and the remaining ones correspond to the transversality condition **iv)** of the PMP.

Step 4. If Eq. (9.7) (together with the transversality condition and condition **iii)** of the PMP if T is free) has a unique solution p_{in} and if we have verified a priori the existence step, then the optimal control problem is solved. Unfortunately, in general there is no reason for Eq. (9.7) to provide a unique solution. Indeed, the PMP is only a necessary condition for optimality. If several solutions are found, one should choose among them the best one by a direct comparison of the value of the cost. This is, in general, a non-trivial step, complicated by the difficulty of solving explicitly Eq. (9.7). For this reason, several techniques have been developed to select the extremals. Among others, we mention the sufficient conditions for optimality given by Hamilton–Jacobi–Bellman theory and synthesis theory. We refer to [BP21] for a discussion.

9.2.1. Extremal and Optimal syntheses

Often in application it is interesting to solve an optimal control problem with fixed initial condition q_{in} and *any* final condition q_{fi}. Assume for simplicity that T is free. Let us assume that there is existence of optimal solutions for the problem (9.1) for every $q_{\text{fi}} \in M$. The collection of all extremal trajectories starting from q_{in} obtained in step 2 plus the condition that $\mathcal{H}_M = 0$:

$$\Gamma = \{q(\cdot; p_{\text{in}}, p^0) \mid p_{\text{in}} \in T^*_{q_{\text{in}}} M, p^0 \in \{0, 1/2\}, \quad \mathcal{H}_M(q_{\text{in}}, p_{\text{in}}, p^0) = 0\},$$

is called an *extremal synthesis*. Notice that for a trajectory to be in Γ is not required to satisfy **iii)** of the Pontryagin maximum principle since $\mathcal{T}$ is reduced to a point. Moreover we are not requiring that $q(T; p_{\text{in}}, p^0) = q_{\text{fi}}$.

A choice in Γ of one optimal trajectory for every $q_{\text{fi}} \in M$ is called an *optimal synthesis*.

$$\Gamma_{\text{opt}} = \{q_{q_{\text{fi}}}(\cdot) \mid q_{q_{\text{fi}}} \in M \text{ and } q_{q_{\text{fi}}}(\cdot) \text{ is a solution of (9.1) with final condition } q(T) = q_{\text{fi}}\}.$$

Notice that in general an extremal synthesis contains many more trajectories than an optimal synthesis. Actually to obtain Γ_{opt} from Γ one should do in the following way. For every q_{fi} one should consider all trajectories arriving at that point and eliminate all non-optimal ones. If there are several optimal trajectories arriving to that point, one should select only one of those.

The concepts of extremal and optimal synthesis will be clarified in the examples.

9.3. Proof of the PMP for Control Affine Systems with Quadratic Cost

We will present the proof of the Pontryagin Maximum Principle for a special, albeit very important, class of nonlinear optimal control problems, namely *control affine systems with quadratic cost*.

Definition 9.2 (Optimal control problem for affine systems with quadratic cost)**.**

$$\begin{cases} \dot{q} = F_0(q) + \sum_{i=1}^{m} u_i(t) F_i(q) \\ q(0) = q_{\text{in}}, \quad q(T) = q_{\text{fi}} \\ \int_0^T \sum_{i=1}^{m} u_i(t)^2 \, dt \end{cases} \tag{OCP-AQ}$$

Here,

- $T > 0$ is fixed,
- M is a smooth n-dimensional connected manifold,
- F_0, F_i are smooth vector fields,
- $u(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$,
- $q(\cdot) : [0,T] \to M$, belongs to the set of Lipschitz curves.

This problem is important since the control affine form (**??**) models most of the systems that one can find in applications and the quadratic cost here considered represents the *energy* given by the controls to the systems. Moreover, these problems are generalizations of the sub-Riemannian problem, which is a fundamental problem in geometry and control theory that we will study in more detail in Chapter 11.

In (OCP-AQ) it is very important that controls take values in $\mathbb{R}^m$ and that T is fixed. For simplicity we have decided to fix that $\mathcal{T}$ is a single point, but the generalization to the case in which $\mathcal{T}$ is a smooth submanifold of M is not hard.

In this chapter we do not assume that a solution to (OCP-AQ) exists. We only look for first order optimality conditions to recover the Pontryagin Maximum Principle in this case. The problem of existence of optimal solutions for (OCP-AQ) is hard since the control $u(\cdot) = (u_1(\cdot), \dots, u_m(\cdot))$ takes values in $\mathbb{R}^m$ which is non compact and hence we cannot apply the Filippov theorem. We will study the existence problem for the sub-Riemannian case in the next chapter.

We prove the following.

Theorem 9.3 (Pontryagin Maximum Principle for (OCP-AQ))**.** *Define the* pre-Hamiltonian *function* $\mathcal{H}$ *as*

$$\mathcal{H}(q,p,u,p^0) = \left\langle p, F_0(q) + \sum_{i=1}^{m} u_i F_i(q) \right\rangle + p^0 \sum_{i=1}^{m} u_i^2, \tag{9.8}$$

with

$$(q,p,u,p^0) \in T^*M \times \mathbb{R}^m \times \mathbb{R}.$$

If the pair $(q,u):[0,T]\to M\times\mathbb{R}^m$ *is optimal for* (OCP-AQ)*, then there exists a never vanishing Lipschitz continuous pair* $(p,p^0):[0,T]\ni t\mapsto(p(t),p^0)\in T^*_{q(t)}M\times\mathbb{R}$ *where* $p^0\le 0$ *is a constant and such that for almost every (a.e.)* $t\in[0,T]$ *we have*

i) Hamiltonian equations*:*

$$\begin{cases}\dot q(t)=\dfrac{\partial\mathcal{H}}{\partial p}(q(t),p(t),u(t),p^0)\\[2ex]\dot p(t)=-\dfrac{\partial\mathcal{H}}{\partial q}(q(t),p(t),u(t),p^0)\end{cases}\tag{9.9}$$

ii) Maximization condition*:*

$$\frac{\partial\mathcal{H}}{\partial u}(x(t),p(t),p^0,u(t))=0$$

iii) Value of the Hamiltonian*: there exists a constant* c *such that* $\mathcal{H}(q(t),p(t),u(t),p^0)=c$ *on* $[0,T]$*,*

Theorem 9.3 is exactly Theorem 9.1 written for the problem OCP-AQ. Observe that the Hamiltonian equations (condition **i)**) of the two theorems are identical. The maximization condition **ii)** of Theorem 9.1 applied to an Hamiltonian quadratic in u with negative coefficient in front of the quadratic part coincide with **ii)** of Theorem 9.3. Condition **iii)** of Theorem 9.1 is the same as condition **iii)** of Theorem 9.3 for T fixed. Condition **iv)** is empty since $\mathcal{T}$ coincide with a point.

To avoid difficulties we will however assume that the control system is complete starting from q_{in} in time T, namely we will assume that

Assumption For every $u(\cdot)=(u_1(\cdot),\dots,u_m(\cdot))\in L^\infty([0,T],\mathbb{R}^m)$ the solution of

$$\begin{cases}\dot q=F_0(q)+\displaystyle\sum_{i=1}^m u_i(t)F_i(q)\\[2ex] q(0)=q_{\text{in}}\end{cases}\tag{9.10}$$

is well-defined for $t\in[0,T]$.

9.3.1. Proof of the Pontryagin Maximum Principle

The idea of the proof is the following. As in Section 8.2, we work with the augmented system. Namely, we add a new variable $q^0(t)=\int_0^t\sum_{i=1}^m u(s)^2\,ds$ and we consider the dynamics for $\hat q=(q^0,q)^T$:

$$\dot{\hat q}(t)=\begin{pmatrix}\dot q^0(t)\\ \dot q(t)\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^m u_i(t)^2\\ F_0(q(t))+\sum_{i=1}^m u_i(t)F_i(q(t))\end{pmatrix}.\tag{9.11}$$

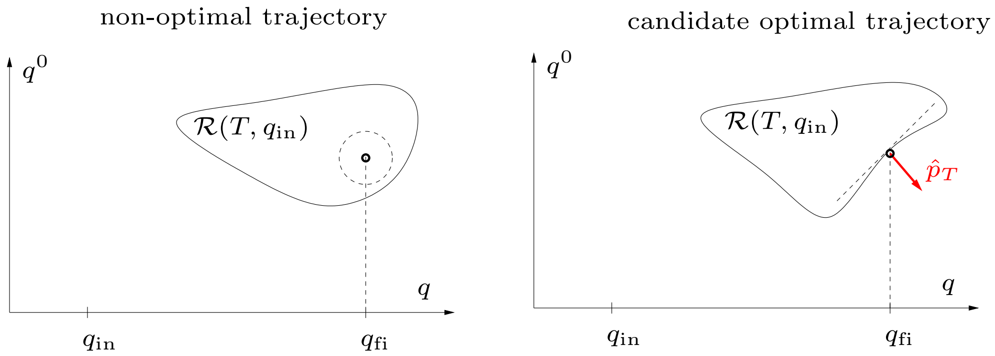


Figure 9.1.: If the differential of End at the optimal control $\tilde{u}(\cdot)$ was a surjection, then we could reach the same final point $q(T)$ with a smaller cost. Hence, there exists a covector $\hat{p}_T := (p^0, p_T) \neq 0$ which annihilates $\text{Im}(D_{\tilde{u}(\cdot)}(\text{End}))$.

We define the *end-point mapping* as the map that to a control associates the end point of the corresponding trajectory starting from q_{in}:

$$L^\infty([0,T],\mathbb{R}^m) \ni u(\cdot) \overset{\text{End}}{\longmapsto} \begin{pmatrix} q^0(T) \\ q(T) \end{pmatrix} \in \mathbb{R} \times M.$$

Consider $\tilde{u}(\cdot)$ to be optimal. Assume for a moment that End is Fréchet differentiable. In this case, this map cannot be a surjection at $\tilde{u}(\cdot)$ (i.e. its differential cannot be invertible). Indeed, if this was the case, the image of a ball in $L^\infty([0,T],\mathbb{R})$ around the optimal control $\tilde{u}(\cdot)$, would contain an open set in $\mathbb{R} \times M$ and one could reach the same final point $q(T)$ with a smaller cost (cf. Figure 9.1 and Section 8.2). It follows that it exists a covector $\hat{p}_T := (p^0, p_T) \neq 0$ which annihilates $\text{Im}(D_{\tilde{u}(\cdot)}(\text{End}))$. That is,

$$\langle \hat{p}_T, \text{Im}(D_{\tilde{u}(\cdot)}(\text{End})) \rangle = 0 \tag{9.12}$$

This is the main idea behind the proof of Theorem 9.3.

To avoid the difficulties inherent in working with differentials of maps on infinite dimensional spaces, in the following we will consider specific variations belonging to a finite dimensional set of controls. Moreover the condition (9.12) will be “transported back” the initial point. It is this that yields the equation for the covector.

Notation We use the following notation.

- We set $f_u(q) := F_0(q) + \sum_{i=1}^m u_i F_i(q)$.
- We let $\tilde{u}(\cdot)$ be the optimal control and $\tilde{q}(\cdot)$ be the optimal trajectory starting from q_{in}.
- We let ϕ_t be the *optimal flow*, i.e. the flow associated to the differential equation $\dot{q} = f_{\tilde{u}(t)}(q)$. Notice that:

  - Although $\tilde{q}(t) = \phi_t(q_{\text{in}})$, in general $\phi_t(\bar{q})$ is not an optimal trajectory for $\bar{q} \neq q_{\text{in}}$.
  - By definition of flow, we have $\frac{d}{dt}\phi_t = f_{\tilde{u}(t)} \circ \phi_t$.
  - Since the control system is complete, ϕ_t is a diffeomorphism.

- Let $\Phi : M \to M$ be a diffeomorphism (i.e. a smooth map which is invertible together with its inverse). Recall[1] that its differential, indicated as $\Phi_*(q)$ or $\Phi_*|_q$ or $\frac{\partial \Phi}{\partial q}$, at a point $q \in M$ is the map

$$T_q M \ni v \mapsto \Phi_*(q)v \in T_{\Phi(q)}M,$$

  defined by

$$\Phi_*(q)v = \left.\frac{\partial}{\partial t}\right|_{t=0} \Phi(\gamma(t)),$$

  where γ is a smooth curve such that $\gamma(0) = q$ and $\dot{\gamma}(0) = v$.

- Given a linear operator A on $T_q M$ we its adjoint A^* is the linear operator acting on $T_q^* M$ defined by

$$\langle A^* p, v\rangle = \langle p, Av\rangle \qquad \text{for every } v \in T_q M.$$

  In coordinates using our convention of representing vectors as column vectors and covectors as row vectors (cf. Section 9.1) we have that $\langle p, v\rangle = p\,v$ and $p(Av) = \langle p, Av\rangle = \langle A^* p, v\rangle = (pA)v$ meaning that A^* is represented by the same matrix as A but acting on the left.[2]

- The *codifferential* of Φ indicated as $\Phi^*(q)$ or $\Phi^*|_q$ at a point $q \in M$ is the map

$$T^*_{\Phi(q)}M \ni p \mapsto \Phi^*(q)p \in T_q^* M,$$

  defined by duality

$$\langle \Phi^*(q)p, v\rangle_q = \langle p, \Phi_*(q)v\rangle_{\Phi(q)}, \quad v \in T_q M, \quad p \in T^*_{\Phi(q)}M.$$

  In this formula we have indicated at which point the duality product between vectors and covectors is calculated. Alternatively (since Φ is invertible and hence $\Phi_*(q)$ and $\Phi^*(q)$ are invertible as well) we can define

$$\langle \Phi^{*-1}(q)p, \Phi_*(q)v\rangle_{\Phi(q)} = \langle p, v\rangle_q, \quad v \in T_q M, \quad p \in T_q^* M.$$

[1] See Definition 2.10.

[2] if we were using the convention of representing co-vectors as column vectors as well then $\langle p, v\rangle = p^T v$ and

$$p^T(Av) = \langle p, Av\rangle = \langle A^* p, v\rangle = (A^T p)^T v = (p^T A)v$$

meaning that A^* would be represented by its trasposition (acting on the right as A).

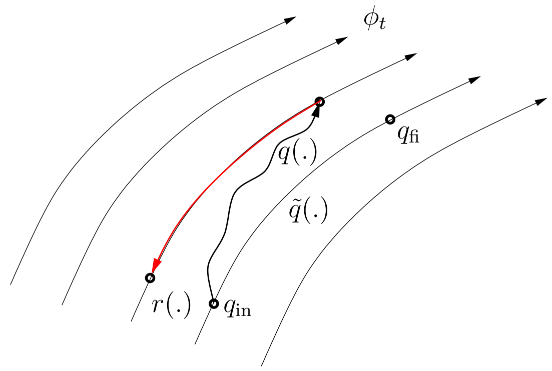


Figure 9.2.: The trajectory $r(\cdot)$ is the trajectory corresponding to the control $u(\cdot)$ brought back with the optimal flow.

Notice that when $M = \mathbb{R}^n$ or we are working in coordinates, then, with our conventions,

$$\Phi(x) = \begin{pmatrix} \Phi_1(x) \\ \vdots \\ \Phi_n(x) \end{pmatrix}, \tag{9.13}$$

$$\Phi_*(x) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \dots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \dots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \dots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix}, \tag{9.14}$$

$$\Phi^*(x) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \dots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \dots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \dots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix} \text{ acting on the left,} \tag{9.15}$$

$$\Phi^{*-1}(x) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \dots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \dots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \dots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix}^{-1} \text{ acting on the left.} \tag{9.16}$$

The variation

Let $q(\cdot)$ be the trajectory starting from $q_{\rm in}$ and corresponding to a control $u(\cdot) = \tilde{u}(\cdot) + v(\cdot)$. Let us define

$$r(t) = \phi_t^{-1}(q(t)).$$

This is the trajectory corresponding to the control $u(\cdot)$ brought back with the optimal flow (cf. Figure 9.2). Notice that $r(0) = q(0) = q_{\rm in}$. Moreover if $v(\cdot) \equiv 0$ then $r(\cdot) \equiv q_{\rm in}$.

Let us look for an equation for $r(t)$. Differentiating $q(t) = \phi_t(r(t))$, we get

$$\dot{q}(t) = \frac{d}{dt}\phi_t\Big|_{r(t)} + \frac{\partial \phi_t}{\partial q}\Big|_{r(t)}\dot{r}(t) = f_{\tilde{u}(t)}(\phi_t(r(t))) + \frac{\partial \phi_t}{\partial q}\Big|_{r(t)}\dot{r}(t) \tag{9.17}$$

Now $\dot{q}(t) = f_{u(t)}(q(t)) = f_{u(t)}(\phi_t(r(t)))$. Hence

$$\dot{r}(t) = \left[\frac{\partial \phi_t}{\partial q}\Big|_{r(t)}\right]^{-1} \Big(f_{u(t)}(\phi_t(r(t))) - f_{\tilde{u}(t)}(\phi_t(r(t)))\Big)$$
$$= \left[\frac{\partial \phi_t}{\partial q}\Big|_{r(t)}\right]^{-1} \sum_{i=1}^{m} v_i(t) F_i(\phi_t(r(t))) =: g_{v(t)}(r(t)).$$

Notice that if we set $v \equiv 0$ in the Cauchy problem

$$\begin{cases} \dot{r} = g_{v(t)}(r) \\ r(0) = q_{\text{in}} \end{cases}$$

then $g_v \equiv 0$ and $r(\cdot) \equiv q_{\text{in}}$. Notice moreover that $g_{sv} = s g_v$, for every $s \in \mathbb{R}$.

The crucial step

Fix $v(\cdot)$ and consider the map

$$s \mapsto \begin{pmatrix} q^0(T, \tilde{u}(\cdot) + sv(\cdot)) \\ r(T, \tilde{u}(\cdot) + sv(\cdot)) \end{pmatrix} \text{ starting from } \begin{pmatrix} 0 \\ q_{\text{in}} \end{pmatrix}$$

Lemma 9.4. *If* $(\tilde{q}(\cdot), \tilde{u}(\cdot))$ *is a solution to the problem* (OCP-AQ)*, then there exists* $\hat{p} \in \mathbb{R}^* \times T^*_{q_{\text{in}}} M$*,* $\hat{p} \neq 0$*, such that*

$$\left\langle \hat{p}, \Big(\frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0}, \frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0}\Big)^T \right\rangle = 0 \qquad \textit{for every } v(\cdot). \tag{9.18}$$

Proof. By contradiction, we assume that there exists $n+1$ variations $\eta^0(\cdot), \eta^1(\cdot), \dots, \eta^n(\cdot)$ such that

$$\begin{pmatrix} \frac{\partial q^0(T,\tilde{u}(\cdot)+s\eta^0(\cdot))}{\partial s}\Big|_{s=0} \\ \frac{\partial r(T,\tilde{u}(\cdot)+s\eta^0(\cdot))}{\partial s}\Big|_{s=0} \end{pmatrix}, \dots, \begin{pmatrix} \frac{\partial q^0(T,\tilde{u}(\cdot)+s\eta^n(\cdot))}{\partial s}\Big|_{s=0} \\ \frac{\partial r(T,\tilde{u}(\cdot)+s\eta^n(\cdot))}{\partial s}\Big|_{s=0} \end{pmatrix} \tag{9.19}$$

are linearly independent. It follows that the map

$$(s_0, \dots, s_n) \mapsto \begin{pmatrix} q^0(T, \tilde{u}(\cdot) + \sum_{j=0}^n s_j \eta^j(\cdot)) \\ r(T, \tilde{u}(\cdot) + \sum_{j=0}^n s_j \eta^j(\cdot)) \end{pmatrix} \tag{9.20}$$

is a local diffeomorphism in a neighborhood of the origin of $\mathbf{R}^{n+1}$ (indeed the vectors (9.19) are the components of the differential of the map (9.20)).

Let us prove that this is not possible. First, notice that the image of the origin through the map (9.20) is,

$$\begin{pmatrix} q^0(T, \tilde{u}(\cdot)) \\ r(T, \tilde{u}(\cdot)) \end{pmatrix} = \begin{pmatrix} q^0(T, \tilde{u}(\cdot)) \\ q_{\text{in}} \end{pmatrix}.$$

But if (9.20) were a local diffeomorphism there would exist $\tilde{v}(\cdot)$ such that

$$q^0(T, \tilde{u}(\cdot) + \tilde{v}(\cdot)) < q^0(T, \tilde{u}(\cdot))$$

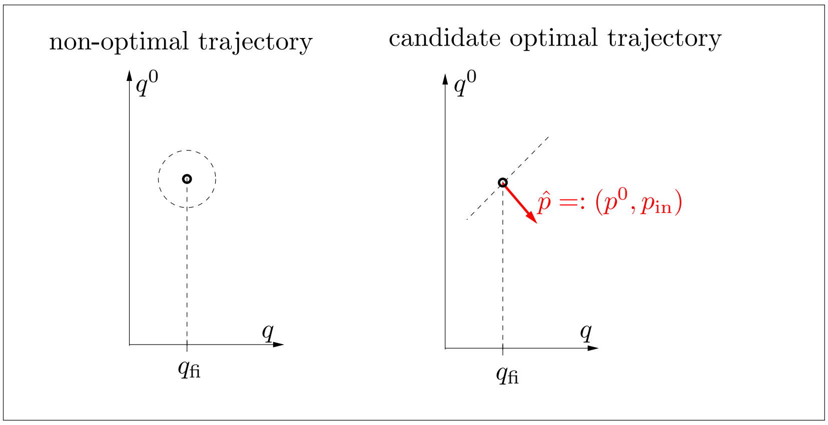


Figure 9.3.: At an optimal trajectory,there exists a covector $\hat{p} \in \mathbb{R}^* \times T^*_{q_{\text{in}}} M$, $\hat{p} \neq 0$, such that (9.18) holds.

and

$$r(T, \tilde{u}(\cdot) + \tilde{v}(\cdot)) = r(T, \tilde{u}(\cdot))$$

meaning that

$$q(T, \tilde{u}(\cdot) + \tilde{v}(\cdot)) = \phi_T(r(T, \tilde{u}(\cdot) + \tilde{v}(\cdot))) = \phi_T(r(T, \tilde{u}(\cdot)) = \phi_T(q_{\text{in}}) = q_{\text{fi}}.$$

Hence, $(\tilde{q}(\cdot), \tilde{u}(\cdot))$ would not be optimal which is a contradiction. □

Setting $\hat{p} = (p^0, p_{\text{in}})$ in (9.18), we have that

$$p^0 \frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0} + \left\langle p_{\text{in}}, \frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0} \right\rangle = 0 \quad \text{for every } v(\cdot) \tag{9.21}$$

Let us compute the different terms in the sum.

$$\begin{aligned}
\frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0} &= \frac{\partial}{\partial s}\Big|_{s=0} \int_0^T (\tilde{u}(t) + sv(t))(\tilde{u}(t) + sv(t))\, dt \\
&= 2 \int_0^T \sum_{i=1}^m \tilde{u}_i(t) v_i(t)\, dt \\
\frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s}\Big|_{s=0} &= \frac{\partial}{\partial s}\Big|_{s=0} \int_0^T g_{sv(t)}(r(t, \tilde{u}(\cdot) + sv(\cdot)))\, dt \\
&= \int_0^T g_{v(t)}(r(t, \tilde{u}(\cdot)))\, dt = \int_0^T \sum_{i=1}^m v_i(t) \phi_{t*}^{-1}(q_{\text{in}}) F_i(\tilde{q}(t))\, dt.
\end{aligned}$$

Here, we have used the notation ϕ_{t*} in place of $\frac{\partial \phi_t}{\partial q}$, the fact that $r(t, \tilde{u}(\cdot)) = q_{\text{in}}$ and that $\phi_t(r(t, \tilde{u}(\cdot))) = \tilde{q}(t)$. Then, from (9.21) we deduce that

$$\int_0^T \sum_{i=1}^m v_i(t) \Big(2p^0 \tilde{u}_i(t) + \langle p_{\text{in}}, \phi_{t*}^{-1}(q_{\text{in}}) F_i(\tilde{q}(t)) \rangle \Big)\, dt = 0.$$

Recall that $\langle p_{\mathrm{in}}, \phi_{t*}^{-1}(q_{\mathrm{in}})F_i(\tilde{q}(t))\rangle = \langle (\phi_t^{-1})^* p_{\mathrm{in}}, F_i(\tilde{q}(t))\rangle$. Hence, defining $p(t) := (\phi_t^{-1})^* p_{\mathrm{in}}$ we obtain

$$\int_0^T \sum_{i=1}^m v_i(t)\Big(2p^0\tilde{u}_i(t) + \langle p(t), F_i(\tilde{q}(t))\rangle\Big)\, dt = 0.$$

Finally, using the arbitrariness of $v(\cdot)$ we get

$$2p^0\tilde{u}_i(t) + \langle p(t), F_i(\tilde{q}(t))\rangle = 0, \text{ for almost every } t \in [0,T]$$

Hence, we have proven the following (here we remove the "tilde" from the $\tilde{q}(\cdot)$ and $\tilde{u}(\cdot)$).

Proposition 9.5. *If $(q(\cdot), u(\cdot))$ is a solution to the problem* (OCP-AQ)*, then there exists $\mathbb{R}^* \times T^*_{q_{\mathrm{in}}}M \ni (p^0, p_{\mathrm{in}}) \neq (0,0)$ such that*

$$2p^0u_i(t) + \langle p(t), F_i(q(t))\rangle = 0, \qquad \text{for almost every } t \in [0,T]. \tag{9.22}$$

Here, $p(t) := (\phi_t^{-1})^* p_{\mathrm{in}}$ and ϕ_t is the flow corresponding to the optimal control $u(\cdot)$.

The Hamiltonian form

It remains to show that Proposition 9.5 is equivalent to the Pontryagin Maximum Principle (Theorem (9.3)). Observe that:

- The first Hamiltonian equation (point **i)** of Theorem 9.3) is equivalent to $\dot{q}(t) = F_0(q(t)) + \sum_{i=1}^m u_i(t)F_i(q(t))$. This is the initial control system in problem (OCP-AQ).
- The second Hamiltonian equation immediately yields equation (9.22).
- it remains to show that $p(t) := (\phi_t^{-1})^* p_{\mathrm{in}}$ is a solution of the second Hamiltonian equation in point **i)** of Theorem 9.3. This is the content of the rest of this section.

We have the following

Lemma 9.6. *Consider the control system*

$$\dot{z} = \boldsymbol{F}(z(t), u(t)) \tag{9.23}$$

where

- *N is a smooth manifold, $U \subset \mathbb{R}^m$,*
- *$\boldsymbol{F}$ is a smooth function of its arguments,*
- *$u(\cdot) \in L^\infty([0,T], U)$,*
- *$z(\cdot) : [0,T] \to N$, belongs to the set of Lipschitz curves.*

Fix a control $u(\cdot)$, an initial condition $z(0) = z_{\mathrm{in}}$, and $\lambda_{\mathrm{in}} \in T^*_{z_{\mathrm{in}}}N$. Then, letting $\lambda(t) = (\Phi_t^{-1})^* \lambda_{in} \in T^*_{z(t)}N$ where Φ_t is the flow of (9.23)*, the pair $(z(\cdot), \lambda(\cdot))$ is solution of the Hamiltonian system*

$$\begin{cases} \dot{z} = \frac{\partial H}{\partial \lambda}(z, \lambda, u(t)) \\ \dot{\lambda} = -\frac{\partial H}{\partial q}(z, \lambda, u(t)), \end{cases} \qquad \text{where } H(z,\lambda,u) := \langle \lambda, \boldsymbol{F}(z,u)\rangle. \tag{9.24}$$

Remark 9.7. This Lemma tells us the following important fact. If $z(\cdot)$ is solution of a differential equation on a manifold N, we fix $\lambda_{\text{in}} \in T^*_{z_{\text{in}}}N$ and we evolve λ_{in} on T^*N according to the evolution of $z(\cdot)$ i.e., $\lambda(t) = (\Phi_t^{-1})^*\lambda_{in} \in T^*_{z(t)}N$ then the pair $(z(\cdot), \lambda(\cdot))$ is solution of the simplest possible Hamiltonain system, namely the one having the time-dependent Hamiltonian $H(z, \lambda, u(t)) := \langle \lambda, \mathbf{F}(z, u(t)) \rangle$.

Remark 9.8. We stated Lemma 9.23 for a control system. However in that lemma the control does not play any role and the result could be stated for a general time dependent ODE of the type $\dot{z} = \mathbf{F}(z, t)$ with suitable regularity conditions.

Let us first apply this Lemma to the augmented system (9.11) for the candidate optimal solution $(q(t), u(t))$, Namely we set

$$N = \mathbb{R} \times M \tag{9.25}$$

$$z = \begin{pmatrix} q^0 \\ q \end{pmatrix} \tag{9.26}$$

$$\mathbf{F}(z, u) = \begin{pmatrix} \sum_{i=1}^m u_i^2 \\ F_0(q) + \sum_{i=1}^m u_i F_i(q) \end{pmatrix} \tag{9.27}$$

for a fixed initial condition $z_{\text{in}} = \begin{pmatrix} 0 \\ q_{\text{in}} \end{pmatrix}$. Let now fix $T_{z_{\text{in}}}N \ni \lambda_{\text{in}} = (p^0, p_{\text{in}}) \in \mathbb{R}^* \times T^*_{q_{\text{in}}}M$ and let us write the equation satisfied by

$$\lambda(t) = (\Phi_t^{-1})^*\lambda_{in} \in T^*_{z(t)}N.$$

Notice that since the "zero" component of $\mathbf{F}$ is decoupled from the others, we have that

$$(\Phi_t^{-1})^* = \begin{pmatrix} \star & 0 \\ 0 & (\phi_t^{-1})^* \end{pmatrix}$$

where $\star$ is something that is not necessary to compute now. Hence $\lambda(t) = (\star p^0, p(t))$ where $p(t) = (\phi_t^{-1})^* p_{\text{in}}$.

The Hamiltonian is

$$H(z, \lambda, u) = \langle \lambda, \mathbf{F}(z, u) \rangle = \langle p, F_0(q) + \sum_{i=1}^m u_i F_i(q) \rangle + p^0 \sum_{i=1}^m u_i^2 = \mathcal{H}(q, p, u)$$

and the Hamiltonian equations (9.24) become

$$\dot{q}^0 = \frac{\partial \mathcal{H}}{\partial p^0} = \sum_{i=1}^m u_i^2(t) \tag{9.28}$$

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial p} \tag{9.29}$$

$$\dot{p}^0 = -\frac{\partial \mathcal{H}}{\partial q^0} = 0 \tag{9.30}$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial q} \tag{9.31}$$

Equation (9.31) is the desired equation **ii)** of Theorem (9.3). Equations (9.28) and (9.29) tell something that we already knew, Equation (9.30) says that $\star$ is actually the identity.

We are now left to prove the lemma, and then the proof of the PMP will be complete.

Proof of Lemma 9.6. Let us fix a system of coordinates and use matrix notation. Let us write the differential of the flow of $\mathbf{F}$ as

$$A(t,z) := \Phi_{t*} = \frac{\partial \Phi_t}{\partial z} \text{ then } \lambda(t) = (\Phi_t^{-1})^* \lambda_{in} = \lambda_{in} A^{-1}(t,z). \tag{9.32}$$

The operator A satisfies the famous *equation of variations*

$$\frac{\partial}{\partial t} A(t,z) = \frac{\partial \mathbf{F}}{\partial z}\Big|_{\Phi_t(z)} A(t,z) \tag{9.33}$$

i.e., the differential evolves with the linearized equation. Notice that this equation can be solved only together with $\dot{z} = \mathbf{F}(z, u(t))$

The proof of the equation of variation is immediate in the smooth case. Actually

$$\frac{\partial}{\partial t} A(t,z) = \frac{\partial}{\partial t}\frac{\partial \Phi_t}{\partial z} = \frac{\partial}{\partial z}\frac{\partial \Phi_t}{\partial t} = \frac{\partial}{\partial z}\mathbf{F}(\Phi_t(z), u(t)) = \frac{\partial \mathbf{F}}{\partial z}\Big|_{\Phi_t(z)} \frac{\partial \Phi_t}{\partial z} = \frac{\partial \mathbf{F}}{\partial z}\Big|_{\Phi_t(z)} A(t,z).$$

For the L^∞ case see for instance [BP07].

Now starting from (9.33) let us look for en equation for $A(t,z)^{-1}$. We have

$$AA^{-1} = \mathrm{id} \implies \dot{A}A^{-1} + A\dot{A}^{-1} = 0 \implies \dot{A}^{-1} = -A^{-1}\dot{A}A^{-1}$$

Hence,

$$\begin{aligned} \dot{\lambda} &= \frac{d}{dt}\left(\lambda_{in}A^{-1}\right) = \lambda_{in}\dot{A}^{-1} = -\lambda_{in}A^{-1}\dot{A}A^{-1} = -\underbrace{\lambda_{in}A^{-1}}_{\lambda}\frac{\partial \mathbf{F}}{\partial z}\Big|_{\Phi_t(z)} AA^{-1} \\ &= -\lambda\frac{\partial \mathbf{F}}{\partial z}\Big|_{\Phi_t(z)} = -\frac{\partial H}{\partial z}(z, \lambda, u(t)) \end{aligned}$$

Here, in the last equality we have used the fact that $H(z, \lambda, u) := \langle \lambda, \mathbf{F}(z,u)\rangle$. This concludes the proof of the lemma. □

10. First examples of optimal control problems

In this chapter we present some basic examples of optimal control problems, and we show how to apply the Pontryagin Maximum Principle to compute the optimal trajectories. These examples are:

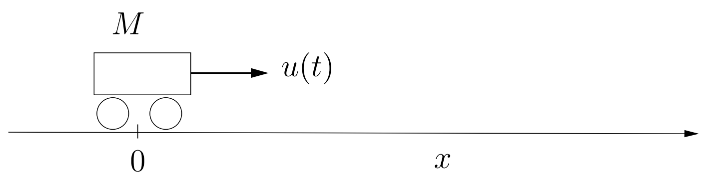
- In Section 10.1 we consider the problem of controlling a cart on a horizontal track. We solve both the minimum time and the minimum energy problems, obtaining in each case the complete optimal synthesis.

- In Section 10.2 we consider the problem of stopping a spring with a bounded external force. We solve two problems concerning the maximization of the final position, first without any constraint, and then by requiring the final velocity to be zero.

10.1. Stopping a cart

In this section we considert the problem of controlling a cart on an horizontal track, already introduced in Exemple 8.1. The state of the system is given by $x = (x_1, x_2)$, where x_1 is the position of the cart and x_2 its velocity. In order to simplify the computations, we consider the cart to have unit mass, and the control set to be $U = [-1, 1]$. The system is then described by the following equations:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u, \end{cases} \qquad |u| \leq 1. \tag{10.1}$$



We can write (10.1) as the following linear control system

$$\dot{x} = Ax + Bu, \qquad \text{where} \qquad A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{10.2}$$

10.1.1. Minimum time problem.

In this section we solve the problem presented in Exemple 8.1. The problem is to arrive at the origin and to stop as quickly as possible. That is, we want to steer the system from an initial state $x(0) = (a_1, a_2)$ to the origin in minimum time.

Existence of optimal controls. We apply Filippov Theorem to check for the existence of solution to the optimal control problem via Filippov Theorem. The family of controlled vector fields is

$$\mathcal{F}(x_1, x_2) = \{(x_2, u)^\top \mid u \in [-1, 1]\}, \tag{10.3}$$

which is convex.

Controllability of the system cannot be deduced from Kalman's condition, since the controls are bounded. It is possible to check that the system is indeed controllable by exhibiting explicit controls that drive the system to the origin. Since this will follow from the analysis of the optimal trajectories, we will not discuss this here.

All other assumptions are easily verified.

Optimal trajectories. Let us look for the extremals. The Hamiltonian is, for $p_0 \in \{0, -1\}$,

$$H(x, p) = p_0 + p(Ax + Bu) = p_0 + p_1 x_2 + p_2 u.$$

Here, $p = (p_1, p_2) \in \mathbb{R}^2$ is the adjoint variable. We start by considering the maximality condition,

$$p_2(t)u(t) = \max_{|v| \le 1} p_2(t)v \implies \begin{cases} u(t) = 1, & \text{if } p_2(t) > 0, \\ u(t) = -1, & \text{if } p_2(t) < 0. \end{cases} \tag{10.4}$$

Thus, from p_2 we can determine u, which is a.e. constant. It changes sign exactly when the extremal crosses the surface $\Sigma = \{(x, p) \mid p_2 = 0\} \subset \mathbb{R}^4 = T^*M$, which is called the *switching surface*.

The Hamiltonian system $\dot{x} = \partial H/\partial p$ and $\dot{p} = -\partial H/\partial x$, reads:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u, \\ \dot{p}_1 = 0, \\ \dot{p}_2 = -p_1. \end{cases} \tag{10.5}$$

That is, $p_1(t) \equiv p_1(0)$ for any $t \in [0, T]$ and $p_2(t) = -p_1(0)t + c$. This shows that either $p_2(t) \neq 0$ for any $t \in [0, T]$ (this happens if $p_1(0)c \le 0$) or there exists exactly one[1] *switching time* $t_s \in (0, T]$ where $p_2(t_s) = 0$ (this happens if $p_1(0)c > 0$). We have thus proved the following:

[1]Consider the case $c > 0$ and $p_1(0) > 0$. In this case, $p_2(0) > 0$. Moreover, if $p_2(t_s) = 0$ then $\dot{p}_2(t_s) = -p_1(0) < 0$. Hence, we have $p_2(t) < 0$ for all $t > t_s$.

Optimal controls are piece-wise constant functions taking values in $\partial U = \{-1, 1\}$, with at most 1 point of discontinuity.

Let us consider now the projection of an extremal. We have shown that this satisfies:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \pm 1. \end{cases} \tag{10.6}$$

On any open set where the control u is constant, the function x_2 is then strictly increasing or decreasing, and thus we can invert $t \mapsto x_2(t)$ and consider x_1 as a function of x_2 (i.e., the function $x_2 \mapsto x_1(t(x_2))$). This yields,

$$\frac{dx_1}{dx_2} = \frac{\dot{x}_1}{\dot{x}_2} = \pm x_2 \implies x_1 = \pm \frac{x_2^2}{2} + c, \qquad c \in \mathbb{R}. \tag{10.7}$$

That is, extremals are confined to the above parabolas for $u = +1$ and $u = -1$, respectively.

Observe that the only trajectories without switchings, i.e., those such that $u \equiv 1$ or $u \equiv -1$ for any $t \in [0, T]$ are the two semiparabolas:

$$x_1 = \frac{x_2^2}{2}, \qquad x_2 < 0, \qquad \dot{x}_2 = 1, \tag{10.8}$$

$$x_1 = -\frac{x_2^2}{2}, \qquad x_2 > 0, \qquad \dot{x}_2 = -1. \tag{10.9}$$

In particular, for any starting point $a = (a_1, a_2)$ such that $a_1 = a_2^2/2$ and $a_2 < 0$, or $a_1 = -a_2^2/2$ and $a_2 > 0$, there passes exactly one extremal that converges to the origin. This yields thus the time-optimal solution.

To obtain all the other optimal trajectories, observe that through any other starting point a pass exactly two parabolae (10.7). The corresponding trajectories will have a given sign for $\dot{x}_2$, and one can check that following these in the correct direction will always yield one trajectory that never crosses the non-switching trajectories (10.8)-(10.9) nor the origin and one that does. Thus, the only possible extremal steering the system to the origin, is the one obtained by following the second trajectory up to the crossing with the semiparabolas (10.8)-(10.9), and then switching to follow those. Since it is unique, it must be an optimal trajectory. See Figure 10.1.

Normal and abnormal extremals. In order for an extremal to be abnormal, it has to hold $p_0 = 0$, and thus. Since the final time is free, we have that $H(x(t), p(t)) = 0$ for any $t \in [0, T]$. In particular,

$$0 = H(x(t), p(t)) = p_1^0 x_2(t) + |p_2(t)|. \tag{10.10}$$

Here, we used the fact that $u(t) = \operatorname{sign} p_2(t)$, and let $p_1(0) = p_1^0$. In particular, since $x(T) = 0$, we have that $p_2(T) = 0$. That is, using this as terminal condition in (10.5),

$$p_2(t) = (T - t)p_1^0 \qquad t \in [0, T]. \tag{10.11}$$

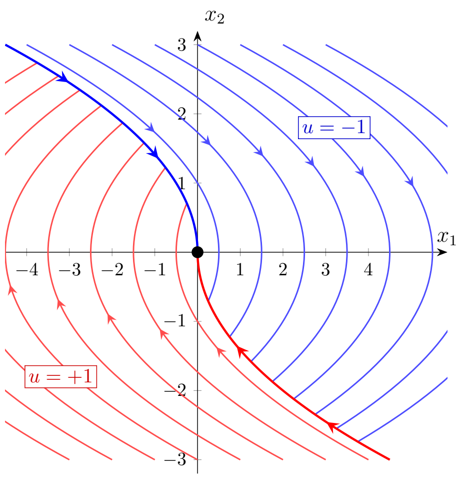


Figure 10.1.: Optimal synthesis for the minimal time problem of stopping a cart (10.1).

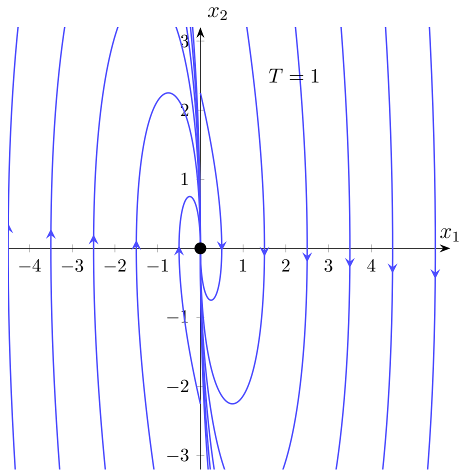


Figure 10.2.: Optimal synthesis for the minimal energy minimization problem of stopping a cart (10.1).

This immediately implies that $p_2(t) \neq 0$ for any $t \in [0, T)$, i.e., abnormal extremal never switch. In particular, the semiparabolas given by (10.8)-(10.9) are abnormals.

On the other hand, normal extremals satisfy

$$1 = p_0 = p_0 + H(x(t), p(t)) = p_1^0 x_2(t) + |p_2(t)|. \tag{10.12}$$

Here, we used that $H(x(t), p(t)) = 0$. Computations as above then yield

$$p_2(t) = (T - t)p_1^0 \pm 1. \tag{10.13}$$

This shows that normal extremals can have switches (and indeed we already know that switching extremals cannot be abnormal and are thus normal). However, also the two non-switching trajectories (10.8)-(10.9) are normal. Indeed, e.g., (10.8) is the projection of the normal extremals with

$$u(t) = 1, \qquad p_1^0 > 0 \qquad \text{and} \qquad p_2(t) = p_1^0(T - t) + 1, \qquad \forall t \in [0, T]. \tag{10.14}$$

One can check that for the above $p_2(t) \neq 0$ for $t \in [0, T]$. Thus, although an optimal trajectory is the projection of an extremal, this extremal **need not be unique**.

10.1.2. Minimal energy problem.

In order to highlight the importance of the choice of the cost, let us consider the problem of minimizing the energy $\int_0^T u^2(t)dt$ over all controls $u \in L^\infty([0, T], \mathbb{R})$ to steer the system from $x(0) = (a_1, a_2)$ to the origin in *fixed* time $T > 0$. Observe that we also removed the bound on the control.

Existence of optimal controls. The Filippov Theorem presented in Chapter 8 does not apply to this problem, since the set $\hat{\mathcal{F}}$ is not convex and the set U is not compact. The existence can be deduced by classical results for linear quadratic problems, but we will not discuss this here (see, e.g., [Jur97]).

Optimal trajectories. The Hamiltonian is, for $p^0 \in \{0, -1/2\}$,

$$H(x,p) = p^0 u^2 + p(Ax + Bu) = p^0 u^2 + p_1 x_2 + p_2 u.$$

We start by considering *normal extremals*, assuming $p^0 = -1/2$. The maximality condition, reads

$$p^0 u(t)^2 + p_2 u(t) = \max_{v \in \mathbb{R}} \left[-\frac{1}{2} v^2 + p_2 v \right] \implies u(t) = p_2(t).$$

This is a key difference w.r.t. the minimal time problem: the control is now a smooth function of the adjoint variable, and not a piece-wise constant function. Hence, the Hamiltonian reads

$$H(x,p) = p_1 x_2 + \frac{p_2^2}{2}. \tag{10.15}$$

The Hamiltonian system $\dot{x} = \partial H / \partial p$ and $\dot{p} = -\partial H / \partial x$, reads:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = p_2, \\ \dot{p}_1 = 0, \\ \dot{p}_2 = -p_1. \end{cases} \tag{10.16}$$

This is a chained system that can be easily solved, yielding,

$$\begin{cases} x_1(t) = -\frac{p_1^0}{6} t^3 + \frac{p_2^0}{2} t^2 + a_2 t + a_1, \\ x_2(t) = -\frac{p_1^0}{2} t^2 + p_2^0 t + a_2, \\ p_1(t) = p_1^0, \\ p_2(t) = -p_1^0 t + p_2^0. \end{cases} \tag{10.17}$$

Let us impose the final condition $x(T) = 0$. This yields a linear system in the unknowns p_1^0 and p_2^0, whose unique solution is given by

$$p_1^0 = -\frac{12}{T^3} a_1 - \frac{6}{T^2} a_2, \qquad p_2^0 = -\frac{6}{T^2} a_1 - \frac{4}{T} a_2. \tag{10.18}$$

Since there is no final cost, and the final point is a single point, the transversality condition is empty. Hence, the only remaining condition to check is the fact that $H(x(t), p(t))$ is constant along the extremal, but this is immediate, since

$$\frac{d}{dt} H(x(t), p(t)) = \dot{p}_1 x_2 + p_1 \dot{x}_2 + p_2 \dot{p}_2 = 0.$$

Abnormal extremals. If $p^0 = 0$, then the Hamiltonian reads

$$H(x, p) = p_1 x_2 + p_2 u. \tag{10.19}$$

In this case, since u is unbounded, for the maximisation condition to be satisfied it must hold $p_2(t) = 0$ for any $t \in [0, T]$. Since $\dot{p}_2 = -p_1$ and $\dot{p}_1 = 0$, this implies that $p_1(t) = p_1^0$ and, in turn, that $p_2^0 = 0$. Hence, $p(t) = 0$ for any $t \in [0, T]$, contradicting the non-triviality condition. Thus, there are no abnormal extremals.

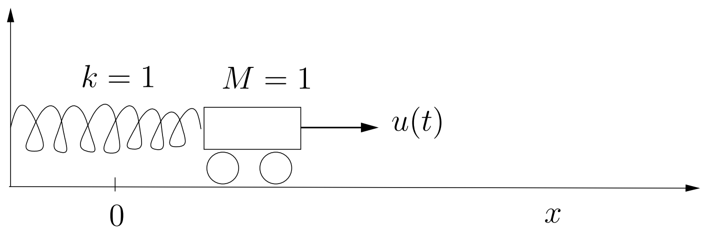
See Figure 10.2 for the optimal synthesis of this problem.

10.2. Controlling a spring

Let us consider the Example 8.3, where we have a spring of elastic constant $k = 1$ and mass $M = 1$, on which we act with an external force $u(t)$ such that $|u(t)| \leq 1$. The system is described by the following equations:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 + u(t) \end{cases} \qquad |u(t)| \leq 1$$

That is,



$$\dot{x} = Ax + Bu, \qquad \text{where} \qquad A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{10.20}$$

The problem is to find the he trajectory starting from $(x_1, x_2) = (0, 0)$ and maximizing $x_1^2(1)$ (i.e. minimizing $-x_1^2(1)$).

Existence of optimal controls. As before, we apply Filippov Theorem. Since now the final time is fixed ($T = 1$) we have to consider the family associated with the augmented system. Since the problem has only a terminal cost $\phi(x) = -x_1^2$, we have $f^0(x, u) = 0$ for any $x \in \mathbb{R}^2$ and $u \in U$. Thus, this family is given by

$$\hat{\mathcal{F}}(x_1, x_2) = \{(0, x_2, -x_1 + u)^\top \mid u \in [-1, 1]\}, \tag{10.21}$$

which is convex. Since the target is free, i.e., $\mathcal{T} = \mathbb{R}^2$, and U is compact, all other assumptions verified.

Optimal trajectories. Recall that $f^0(x, u) = 0$. Thus, the pre-Hamiltonian is independent of p^0 and reads

$$H(x, p) = p_1 x_2 + p_2(-x_1 + u). \tag{10.22}$$

The maximality condition reads as before:

$$p_2(t)u(t) = \max_{|v| \leq 1} p_2(t)v \implies \begin{cases} u(t) = 1, & \text{if } p_2(t) > 0, \\ u(t) = -1, & \text{if } p_2(t) < 0. \end{cases} \tag{10.23}$$

Thus, u is a.e. constant, and changes sign exactly when the extremal crosses the switching surface $\Sigma = \{(x, p) \mid p_2 = 0\}$.

The Hamiltonian system $\dot{x} = \partial H/\partial p$ and $\dot{p} = -\partial H/\partial x$, reads:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + u, \\ \dot{p}_1 = p_2, \\ \dot{p}_2 = -p_1. \end{cases} \tag{10.24}$$

The closed system of equations for (p_1, p_2) can be recast as $\ddot{p}_2 = -p_2$. That is,

$$p_2(t) = \alpha \cos(t + \beta), \quad p_1(t) = \alpha \sin(t + \beta), \qquad \alpha, \beta \in \mathbb{R}. \tag{10.25}$$

Namely, the covector p moves along a circle in the plane, with center at the origin and radius α. The switching times are given by

$$t_s = -\beta - \frac{\pi}{2} + k\pi, \quad k \in \mathbb{Z}. \tag{10.26}$$

In order to determine α and β, we need to use the transversality condition (9.3). Since the target is free, $T = 1$, and the terminal cost is $\phi(x) = -x_1^2$, this condition reads

$$p(1) = p^0 \nabla\phi(x(1)) = p^0 \begin{pmatrix} -2x_1(1) \\ 0 \end{pmatrix}$$

This excludes the case $p^0 = 0$, that is, the problem does not admit abnormal extremals. Indeed, if $p^0 = 0$ then $p(1) = 0$, and thus $p_1(t) = p_2(t) = 0$ for any $t \in [0, 1]$. This contradicts the non-triviality condition. Thus, we have $p^0 = -1$ and

$$p(1) = \nabla\phi(x(1)) = \begin{pmatrix} -2x_1(1) \\ 0 \end{pmatrix}. \tag{10.27}$$

Observe that $x_1(1) \neq 0$, since otherwise $p(1) = 0$ and we would contradict the non-triviality condition. Thus, we have $p_1(1) = -2x_1(1) \neq 0$ and $p_2(1) = 0$. By (10.25), we then obtain that $\alpha \neq 0$ and $1 + \beta = \pi/2 + k\pi$ for some $k \in \mathbb{Z}$.

Consequently, the switching times are given by $t_s = 1 + \ell\pi$ for $\ell \in \mathbb{N}$. Since the final time is $T = 1$, we have that there is no switch in the control, and we have that the optimal control is given by $u(t) = \operatorname{sign} p_2(0) = \operatorname{sign} \alpha \cos(1) = \operatorname{sign} \alpha$ for any $t \in [0, 1]$. In particular, if $\alpha > 0$ then $u(t) = 1$ for any $t \in [0, 1]$, while if $\alpha < 0$ then $u(t) = -1$ for any $t \in [0, 1]$. In both cases, the optimal trajectory is the projection of a normal extremal and is non-switching.

Since the controls take values only in ± 1, by (10.24) the corresponding trajectories satisfy

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 \pm 1. \end{cases} \tag{10.28}$$

Thus, the extremal trajectories are given by

$$\begin{cases} x_1(t) = 1 - \cos t, \\ x_2(t) = \sin t, \end{cases} \quad \text{or} \quad \begin{cases} x_1(t) = -1 + \cos t, \\ x_2(t) = -\sin t. \end{cases} \tag{10.29}$$

Since $\phi(x(1))$ is the same for both trajectories, both of them are optimal. In particular, the optimal control is not unique. (See Figure 10.3.)

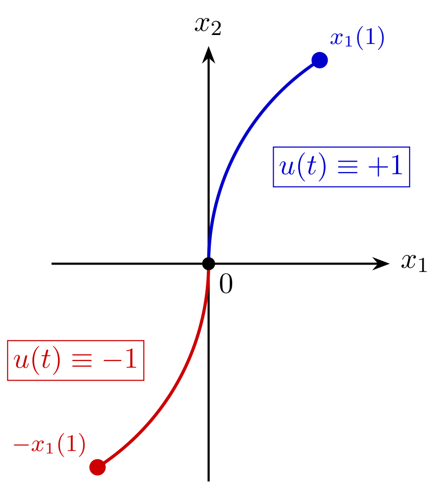


Figure 10.3.: Optimal trajectories for the spring problemx.

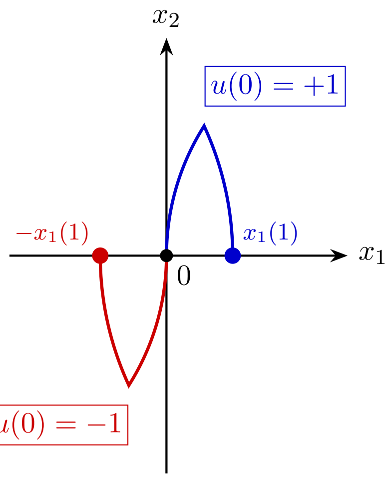


Figure 10.4.: Optimal trajectories for the spring problem with constraint $x_2(1) = 0$.

10.2.1. A variation on the problem.

The above derivation, albeit instructive, is not very surprising. It simply says that the best way to maximize $x_1^2(1)$ is to apply the maximum force in the direction of the desired motion.

Let us consider a more interesting problem by adding the constraint $x_2(1) = 0$ to the previous one. That is, we want to maximize $x_1^2(1)$ under the constraint that the final velocity is zero. In particular, we are changing the final target, from $\mathcal{T} = \mathbb{R}^2$ to

$$\mathcal{T} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\}.$$

The previous analysis is unchanged up to the moment when the transversality condition is applied. In this case, $T_x\mathcal{T} = \{(v_1, 0) \mid v_1 \in \mathbb{R}\}$ and thus the condition reads

$$\langle p(1), v\rangle = p^0\langle d\phi(x(1)), v\rangle, \quad \forall v \in T_{x(1)}\mathcal{T} \iff p_1(1) = p^0(-2x_1(1)).$$

In this case, we cannot exclude the case $p^0 = 0$. Let us consider separately the two possibilities:

Abnormal extremals. If $p^0 = 0$, then $p_1(1) = 0$. Since p is non-trivial, we have that $p_2(1) \neq 0$. By (10.25), this implies that $\alpha \neq 0$ and $1 + \beta = k\pi$ for some $k \in \mathbb{Z}$. Thus, the switching times are given by $t_s = k\pi$ for some $k \in \mathbb{N}$. Since the final time is $T = 1$, we have that there is no switch in the control, as before. Hence, the corresponding trajectories are (10.29), which do not satisfy the constraint $x_2(1) = 0$. Thus, there are no abnormal minimizers.

Normal extremals. If $p^0 = -1$, then $p_1(1) = -2x_1(1)$. Since $x_1(1) = 0$ implies that $\phi(x(1)) = 0$, it cannot be an optimal trajectory. Thus, we can assume $x_1(1) \neq 0$ and thus $\alpha \neq 0$. The observation done for abnormal extremals (i.e., that controls with no switchings never reach the target) still holds, and thus we know that the optimal trajectory has to switch at least once.

Recall that the switching times are given by (10.26), i.e., $t_s = -\beta - \pi/2 + k\pi$ for $k \in \mathbb{Z}$. Since the final time is $T = 1$, we have that there is at most one switching, and it happens if $-\beta - \pi/2 + k\pi \in (0, 1)$ for some $k \in \mathbb{Z}$.

Let us consider a trajectory starting with $u(0) = 1$ and with switching time $t_s \in (0, 1)$. By (10.24). The trajectory before the switching is given by

$$\begin{cases} x_1(t) = 1 - \cos t, \\ x_2(t) = \sin t, \end{cases} \qquad t \in [0, t_s). \tag{10.30}$$

After the switching, we have $u(t) = -1$ for any $t \in [t_s, 1]$, and thus

$$\begin{cases} x_1(t) = -1 + \cos t + 2\cos(t - t_s), \\ x_2(t) = \sin t - 2\sin(t - t_s), \end{cases} \qquad t \in [t_s, 1]. \tag{10.31}$$

The constraint $x_2(1) = 0$ then reads $\sin 1 - 2\sin(1 - t_s) = 0$, which has a unique solution $t_s \in (0, 1)$, given by

$$t_s = 1 - \arcsin\left(\frac{\sin 1}{2}\right). \tag{10.32}$$

This is the switching time of the only admissible extremal trajectory starting with $u(0) = 1$. Observe that this also determines β and thus the whole extremal trajectory. The only extremal trajectory starting with $u(0) = -1$ can be obtained by symmetry, and has the same switching time. (See Figure 10.4). Observe that the final value of the cost is the same for both trajectories, and is given by

$$\phi(x(1)) = -x_1^2(1) = -\left[\pm\left(-1 + 2\sqrt{1 - \frac{\sin^2(1)}{4}} - \cos(1)\right)\right]^2 \simeq 0.075.$$

11. Sub-Riemannian geometry

A very important type of optimal control problems is the so called *sub-Riemannian* problem. It has been already introduced at the beginning of Chapter 9.3 as an OCP-AQ without drift. It is a problem that is linear in the control, with controls taking values in $\mathbb{R}^m$. The cost is quadratic, depends only on the controls and the final time is fixed. The reason why it is called sub-Riemannian is explained in Section 11.2.

Recall that for a vector $v \in \mathbb{R}^m$ we denote by $|v|$ the Euclidean norm of v, i.e., $|v| = \sqrt{\sum_{i=1}^m v_i^2}$.

Definition 11.1 (Sub-Rienmannian problem)**.**

$$\dot{q} = \sum_{i=1}^{m} u_i(t) F_i(q) \quad (11.1)$$

$$q(0) = q_{\text{in}}, \quad q(T) = q_{\text{fi}} \quad (11.2)$$

$$E(u(\cdot)) := \int_0^T |u(t)|^2 \, dt \to \min \quad (11.3)$$

Here,

- $T > 0$ is fixed,
- M is a smooth n-dimensional connected manifold,
- $\mathcal{T}$ is a (non-empty) smooth submanifold of M. It can be reduced to a point (fixed terminal point) or coincide with M (free terminal point),
- F_i are smooth vector fields,
- $u(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$,
- $q(\cdot) : [0,T] \to M$, belongs to the set of Lipschitz curves.

11.1. Energy, length and existence of optimal trajectories

Observe that the Filippov test presented in Chapter 8 cannot be applied to the sub-Riemannian problem. Indeed, the controls take value in $U = \mathbb{R}^m$, which is not compact. Moreover, the cost is quadratic, and thus the family $\hat{F}$ associated with the augmented system is not convex.

To overcome this issue requires some preliminary observations on the dynamics (11.1) and on the cost (11.3).

11.1.1. Reparametrization of trajectories and controls

Let us start to to discuss a crucial property of the dynamics (11.1)

Lemma 11.2. *Let $q_{\text{in}} \in M$ and $u(\cdot) \in L^\infty([0,T],\mathbb{R}^m)$, and denote by $q : [0,T] \to M$ the associated trajectory of (11.1). Then, for every Lipschitz function $\tau : [0,\bar{T}] \to [0,T]$ such that $\dot{\tau}(t) > 0$ a.e. on $[0,\bar{T}]$, the curve $\bar{q}(\cdot) = q(\tau(\cdot))$ is an admissible trajectory corresponding to a control $\bar{u} \in L^\infty([0,\bar{T}],\mathbb{R}^m)$ given by*

$$\bar{u}_i(t) = u_i(\tau(t))\dot{\tau}(t), \qquad \forall t \in [0,\bar{T}], \qquad i = 1,\dots,m.$$

Proof. It suffices to compute:

$$\begin{aligned}\dot{\bar{q}}(t) &= \dot{q}(\tau(t))\dot{\tau}(t) \\ &= \Big(\sum_{i=1}^m u_i(\tau(t))F_i(q(\tau(t)))\Big)\dot{\tau}(t) \\ &= \sum_{i=1}^m \Big(u_i(\tau(t))\dot{\tau}(t)\Big)F_i(\bar{q}(t)).\end{aligned}$$ □

As a consequence of the previous lemma, for a sub-Riemannian problem it holds

$$\mathcal{R}(T,q_{\text{in}}) = \mathcal{R}(q_{\text{in}}), \qquad \forall q_{\text{in}} \in M, \forall T > 0,$$

Recall that the Chow-Rashevski theorem yields that a sufficient condition for controllability (i.e., to have $\mathcal{R}(q_{\text{in}}) = M$) is the Lie bracket generating condition:

$$\dim \text{Lie}_q\{F_1,\dots F_m\} = n, \qquad \forall q \in M.$$

We also have the following immediate consequence of Lemma 11.2.

Corollary 11.3. *If the final time $T > 0$ is free, the sub-Riemannian problem of Definition 11.1 does not admit minimizers.*

Proof. Consider the reparametrization $\tau_\lambda : [0,\lambda] \to [0,T]$ defined by $\tau_\lambda(t) = Tt/\lambda$, $\lambda > 0$. By Lemma 11.2, the curve $\bar{q}(\cdot) = q(\tau_\lambda(\cdot))$ is an admissible trajectory corresponding to a control $\bar{u} \in L^\infty([0,\lambda],\mathbb{R}^m)$ given by $\bar{u}(s) = \dot{\tau}(s)u(\tau(s))$. In particular, $\bar{q}(0) = q_{\text{in}}$ and $\bar{q}(\lambda) \in \mathcal{T}$. Then,

$$E(\bar{u}(\cdot)) = \int_0^\lambda |\bar{u}(s)|^2 ds = \int_0^\lambda \Big|\frac{T}{\lambda}u\Big(\frac{Ts}{\lambda}\Big)\Big|^2 ds = \frac{T}{\lambda}\int_0^T |u(t)|^2 dt = \frac{T}{\lambda}E(u(\cdot)).$$

Letting $\lambda \to +\infty$ we get $E(\bar{u}(\cdot)) \to 0$. Hence, if the final time is free, the infimum of the energy is zero, and there is no minimizer for the sub-Riemannian problem. □

We stress that the previous result is a consequence of the fact that the dynamics (11.1) is linear in the control and that the cost is quadratic and depends only on the controls. Considering a different cost allows to avoid this issue. Let us define the *length* of a control $u(\cdot) \in L^\infty([0,T],\mathbb{R}^m)$ as

$$L(u(\cdot)) = \int_0^T |u(t)|dt. \tag{Length}$$

Then, the following result holds.

Proposition 11.4. *Let $q : [0,T] \to M$ be a solution of (11.1) corresponding to a control $u(\cdot) \in L^\infty([0,T],\mathbb{R}^m)$. Then, for any $c > 0$, the curve $q(\cdot)$ is a Lipschitz reparameterization of an admissible trajectory defined on $[0,T/c]$, corresponding to a control $\bar{u} \in L^\infty([0,L],\mathbb{R}^m)$ of constant norm c. That is,*

$$|\bar{u}(t)| = c, \qquad \textit{for a.e. } t \in [0,T/c].$$

Proof. It suffices to consider the reparametrization $\tau_c : [0,T/c] \to [0,T]$ defined by $\tau_c(0) = 0$ and $\dot{\tau}_c(t) = c/|u(\tau_c(t))|$ for a.e. $t \in [0,T/c]$. By Lemma 11.2, the curve $\bar{q}(\cdot) = q(\tau_c(\cdot))$ is an admissible trajectory defined on $[0,T/c]$, corresponding to controls $\bar{u}$ with norm

$$|\bar{u}(s)| = |u(\tau_c(s))|\dot{\tau}_c(s) = c, \qquad \text{for a.e. } s \in [0,T/c].$$

□

Admissible trajectories for which the corresponding controls satisfies $|u(t)|^2 = 1$ a.e. are said to be *parametrized by arclength.* By Proposition 11.4, every admissible trajectory is a Lipschitz reparametrisation of a trajectory parameterized by arclength.

11.1.2. Existence of optimal trajectories

The sub-Riemannian problem defined in Definition 11.1 consists in considering the control system

$$\dot{q} = \sum_{i=1}^m u_i(t)F_i(q), \tag{sR}$$

$$T > 0, \quad q(0) = q_{\text{in}} \text{ and } q(T) \in \mathcal{T}, q_{\text{in}} \notin \mathcal{T}.$$

Here, $q_{\text{in}} \in M$ and $\mathcal{T}$ (a smooth submanifold of M) are fixed. Motivated by the previous observations, we now consider three different optimal control problems:

- **Energy.** In the sub-Riemannian problem, we are interested in minimizing the cost E given in (11.3), called *energy* of the control, which is

  $$E(u(\cdot)) = \int_0^T |u(t)|^2 dt. \tag{Energy}$$

  For this cost we have to assume the final time T to be fixed, otherwise there is no existence of minimizers by Corollary 11.3.

- **Length.** The *length* of the control $u(\cdot)$ is defined by

$$L(u(\cdot)) = \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2 dt}. \tag{Length}$$

For this cost, the final time T can be either fixed or free, since the cost is invariant by Lipschitz reparametrizations of the trajectory by Proposition 11.4.

- **Time.** The *minimal time* problem consists in considering controls $u \in L^\infty([0,T],\mathbb{R}^m)$ such that $|u(t)| \leq 1$ for a.e. $t \in [0,T]$, and the cost is given by

$$T \to \min\,. \tag{Time}$$

For this cost, the final time T is clearly free.

(See Section 11.2 for a discussion on the geometric meaning of these costs).

Indeed, it turns out that the three costs are related. Let us start by considering the energy cost E and the length cost L.

Proposition 11.5. *Consider the control problem* (sR)*. Then,*

- *Minimizers of E with T fixed, correspond to constant norm controls (i.e., $|u(t)| =$ const, for a.e. $t \in [0,T]$), and are minimizers of L,*
- *Minimizers of L with final time $T > 0$ and whose controls have constant norm, are minimizers of E with final time fixed to T.*

Proof. Recall the Cauchy-Schwarz inequality in $L^2([0,T],\mathbb{R})$, which states that for any $f, g \in L^2([0,T],\mathbb{R})$ it holds

$$\Big(\int_0^T f(t)g(t)dt\Big)^2 \leq \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt, \tag{11.4}$$

with equality holding if and only if f and g are proportional, i.e., there exists $\lambda \in \mathbb{R}$ such that $f(t) = \lambda g(t)$ for a.e. $t \in [0,T]$. Since $L^\infty([0,T],\mathbb{R}^m) \subset L^2([0,T],\mathbb{R}^m)$, the Cauchy-Schwarz inequality also holds for controls $u(\cdot) \in L^\infty([0,T],\mathbb{R}^m)$, and thus

$$L(u(\cdot))^2 \leq TE(u(\cdot)), \tag{11.5}$$

with equality holding if and only if $u(t)$ has constant norm for a.e. $t \in [0,T]$.

Let us start by showing that minimizers of E with T fixed, correspond to constant norm controls. Assume by contradiction that there exists a minimizer $q(\cdot)$ of E with T fixed, corresponding to controls $u(\cdot)$ such that $|u(t)|$ is not constant for a.e. $t \in [0,T]$. Then, by (11.5) we have

$$L(u(\cdot))^2 < TE(u(\cdot)).$$

On the other hand, by Proposition 11.4, there exists a Lipschitz reparametrization $\bar{q}(\cdot)$ of $q(\cdot)$, corresponding to controls $\bar{u}(\cdot)$ with constant norm, such that $L(\bar{u}(\cdot)) = L(u(\cdot))$. Again by (11.5) we have that $L(\bar{u}(\cdot))^2 = TE(\bar{u}(\cdot)) = TE(u(\cdot))$. Summing up, we get

$$TE(\bar{u}(\cdot)) = L(\bar{u}(\cdot))^2 = L(u(\cdot))^2 < TE(u(\cdot)).$$

This contradicts the fact that $q(\cdot)$ is a minimizer of E with T fixed.

Let us now show that minimizers of E with T fixed, are minimizers of L. Assume by contradiction that there exists a minimizer $q(\cdot)$ of E with T fixed, corresponding to controls $u(\cdot)$ such that there exists an admissible trajectory $\bar{q}(\cdot)$ corresponding to controls $\bar{u}(\cdot)$ such that $L(\bar{u}(\cdot)) < L(u(\cdot))$. By the previous part of the proof, we know that $u(\cdot)$ has constant norm. Then, by (11.5) we have that $L(u(\cdot))^2 = TE(u(\cdot))$. On the other hand, by Proposition 11.4, we can assume that $\bar{u}(\cdot)$ has constant norm. Then, again by (11.5) we have that $L(\tilde{u}(\cdot))^2 = TE(\tilde{u}(\cdot))$. Summing up, we get

$$TE(\tilde{u}(\cdot)) = L(\tilde{u}(\cdot))^2 = L(\bar{u}(\cdot))^2 < L(u(\cdot))^2 = TE(u(\cdot)).$$

This contradicts the fact that $q(\cdot)$ is a minimizer of E with T fixed.

Finally, let us show that minimizers of L with final time $T > 0$ and whose controls have constant norm, are minimizers of E with final time fixed to T. Assume by contradiction that there exists a minimizer of L with final time $T > 0$, corresponding to control $u(\cdot)$ with $|u(\cdot)| \equiv \text{const}$ which is not a minimiser of E. Namely, there exists an admissible control $\bar{u}(\cdot)$ such that $E(\bar{u}(\cdot)) < E(u(\cdot))$. Then,

$$L(\bar{u}(\cdot))^2 \leq TE(\bar{u}(\cdot)) < TE(u(\cdot)) = L(u(\cdot))^2.$$

This contradicts the fact that $u(\cdot)$ is a minimizer of L. □

Let us now consider the relation between the energy cost E and the time cost T.

Proposition 11.6. *For the control problem* (11.1)*, the minimal time problem is equivalent to the energy minimization problem with the additional constraint that* $|u(t)|^2 = 1$ *a.e. on* $[0, T]$*.*

Proof. Since $|u(t)|^2 = 1$ a.e. on $[0, T]$ we have

$$E(u(\cdot)) = \int_0^T |u(t)|^2 dt = \int_0^T 1 dt = T.$$

Hence the problem of minimizing E with the additional constraint that $|u(t)|^2 = 1$ a.e. on $[0, T]$ is equivalent to the problem of minimizing T with the constraint on the controls $|u(t)|^2 = 1$ a.e. on $[0, T]$.

To conclude the proof, let us show that a trajectory corresponding to controls for which the condition $|u(t)|^2 = 1$ a.e. on $[0, T]$ is not satisfied cannot be optimal for the minimal time problem. Actually, if $|u(t)|^2 = 1$ is not satisfied then $L = \int_0^T |u(t)| dt < T$. This implies that the arc-length reparametrization of the trajectory corresponding to $u(\cdot)$ reaches the target in time exactly $L < T$, contradicting the minimality of T. □

Let us now go back to the problem of existence of optimal trajectories for the sub-Riemannian problem. Thanks to Proposition 11.6, the sub-Riemannian problem with T fixed in such a way that trajectories are parametrized by arclength can be equivalently recast as a time-optimal control problem with controls in the convex and compact set $U = \{(u_1, \ldots, u_m) \in \mathbb{R}^m \mid u_1^2 + \ldots u_m^2 \leq 1\}$. We can then apply Proposition 8.10 and deduce the existence of an optimal trajectory for the sub-Riemannian problem. Of course nothing changes if T is fixed differently. We have then

Proposition 11.7. *Fix $T > 0$ and consider the control problem* (11.1)*. Assume that*

- *for every $u(\cdot) \in L^\infty([0,T], U)$, the solution of* (11.1)*, with $q(0) = q_{\text{in}}$ is defined on the whole interval* $[0,T]$*.*
- *$\mathcal{T}$ is closed and $\mathcal{R}(q_{\text{in}}) \cap \mathcal{T} \neq \emptyset$.*

Then the sub-Riemannian problem (i.e., the energy minimization problem), the length minimization and the time minimization problem are all equivalent and admit a solution.

Proof. Let us consider the minimal time problem with controls in $U = \{|u| \leq 1\}$. Observe that U is compact, and that $\mathcal{F}(q) = \{\sum_{i=1}^m u_i F_i(q) \mid u \in U\}$ is convex for every $q \in M$. Then, the existence of a solution $u^\star(\cdot)$ for the minimal time problem follows by Proposition 8.10. Let us denote such minimal time by $T^\star$.

Consider now the energy minimization problem, with fixed time $T = T^\star$. It follows by Proposition 11.6 that $u^\star(\cdot)$ is a solution for this problem as well. Indeed, if this was not the case, then there would exist a control $\bar{u}(\cdot)$ such that $E(\bar{u}(\cdot)) < E(u^\star(\cdot)) = T^\star$. By Proposition 11.5 we have that there exists $c > 0$ such that $|\bar{u}(t)| = c$ a.e. on $[0, T^\star]$. Then,

$$T^\star = E(u^\star(\cdot)) > E(\bar{u}(\cdot)) = cT^\star.$$

Hence, a reparametrisation of the trajectory corresponding to $\bar{u}(\cdot)$ reaches the target in time $cT^\star < T^\star$, contradicting the minimality of $T^\star$.

Finally, it follows by Proposition 11.5 that minimizers of E with $T = T^\star$ are minimizers of L and vice versa. □

11.2. Why sub-Riemannian?

The name sub-Riemannian for the problem that we are considering in this chapter comes from the following argument. Consider the simple case in which $m = n$, $M = \mathbb{R}^n$, $\mathcal{T} = \{q_{\text{fi}}\}$ and

$$F_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \ldots \qquad F_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

In this case the sub-Riemannian problem consist in minimizing

$$\int_0^T \sum_{i=1}^n u_i(t)^2 dt = \int_0^T \sum_{i=1}^n \dot{x}_i(t)^2 dt = \int_0^T \|\dot{x}(t)\|^2 dt,$$

where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^n$. We have seen that minimizing $\int_0^T \sum_{i=1}^m u_i(t)^2 dt$ is equivalent to minimize $\int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2 dt}$ up to parameterizing trajectories with constant velocity and fixing T (cf. Proposition 11.5, D. and E.). Hence the sub-Riemannian problem is equivalent to minimize

$$\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2 dt} = \int_0^T \sqrt{\sum_{i=1}^n \dot{x}_i(t)^2 dt} = \int_0^T \|\dot{x}(t)\| dt.$$

i.e., in finding the curve with shortest Euclidean length joining q_{in} in q_{fi}. Notice that the vectors $F_1, \dots, F_n$ play the role of an orthonormal frame for the Euclidean structure (i.e., they have norm 1 and they are mutually orthogonal).

In the more general case, when we are on a general manifold M, $m = n$, and $F_1, \dots F_n$ are (in general non constant) linearly independent vector fields, we can define a scalar product on each tangent space of M by declaring that $F_1(q), \dots F_n(q)$ is an orthonormal frame. That is, for each $q \in M$, we let $g : T_qM \times T_qM \to \mathbb{R}$ be the scalar product defined by

$$g_q\Big(\sum_{i=1}^n a_i F_i(q), \sum_{i=1}^n b_i F_i(q)\Big) = \sum_{i=1}^n a_i b_i, \qquad \forall a_i, b_i \in \mathbb{R}, \forall q \in M.$$

Such an object is called a *Riemannian metric*[1] on M. In this case, the cost

$$\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2 dt} = \int_0^T \sqrt{g_{q(t)}(\dot{q}(t), \dot{q}(t))} dt$$

is the length of the curve for the Riemannian metric g. Notice that in both Euclidean and Riemannian case, we have a control for every direction and hence all Lipschitz curves are admissible. In other words the equation $\dot{q} = \sum_{i=1}^n u_i(t) F_i(q)$ should be considered as a definition of the controls rather than a constraint on the possible dynamics.

In the case in which $m < n$ or, more in general, when $\text{span}\{F_1(q), \dots, F_n(q)\}$ is not of dimension n, the equation $\dot{q} = \sum_{i=1}^m u_i(t) F_i(q)$ is a constraint on the possible dynamics. For this reason the corresponding problem is called *sub-Riemannian.* In the case $m < n$ and $F_1(q), \dots, F_m(q)$ linearly independent we can think to a sub-Riemannian problem as a Riemannian problem in which the orthonormal frame is made by less vectors of the dimension of the space.

By what explained here is it then clear why the cost $\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2 dt}$ is called *length.* The cost $\int_0^T \sum_{i=1}^n u_i(t)^2 dt$ is called *energy* since it is often used to model the energy injected to the system via the controls.

[1] Notice that not all Riemannian metric come from problems in the form $\dot{q} = \sum_{i=1}^n u_i(t) F_i(q)$, $\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2 dt} \to \min$ (with fixed initial and final conditions). Actually the existence of an orthonormal frames is just local. However it is possible to prove that every Riemannian problem can be written in the form $\dot{q} = \sum_{i=1}^m u_i(t) F_i(q)$, $\int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2 dt} \to \min$ (with fixed initial and final conditions) if m is sufficiently large. See [ABB19].

11.3. Minimizers

Let us now apply the Pontryagin Maximum Principle to a sub-Riemannian problem to get information on minimizers. For a sub-Riemannian problem the Pontryagin maximum Principle (Theorem 9.1) becomes as follow.[2]

Theorem 11.8 (Pontryagin Maximum Principle for a sub-Riemannian problem)**.** *Define the* pre-Hamiltonian *function $\mathcal{H}$ as*

$$\mathcal{H}(q,p,u,p^0) = \sum_{i=1}^{m} u_i \langle p, F_i(q)\rangle + p^0 \sum_{i=1}^{m} u_i^2, \tag{11.6}$$

with

$$(q,p,u,p^0) \in T^*M \times \mathbb{R}^m \times \mathbb{R}.$$

If the pair $(q,u) : [0,T] \to M \times \mathbb{R}^m$ is optimal for (11.1)*, then there exists a never vanishing Lipschitz continuous pair $(p,p^0) : [0,T] \ni t \mapsto (p(t),p^0) \in T^*_{q(t)}M \times \mathbb{R}$ where $p^0 \leq 0$ is a constant and such that for almost every (a.e.) $t \in [0,T]$ we have*

i) Hamiltonian equations*:*

$$\begin{cases} \dot{q}(t) = \dfrac{\partial \mathcal{H}}{\partial p}(q(t),p(t),u(t),p^0) \\ \\ \dot{p}(t) = -\dfrac{\partial \mathcal{H}}{\partial q}(q(t),p(t),u(t),p^0) \end{cases} \tag{11.7}$$

ii) Maximization condition*:*

$$\frac{\partial \mathcal{H}}{\partial u}(x(t),p(t),p^0,u(t)) = 0$$

iii) Value of the Hamiltonian*: there exists a constant c such that $\mathcal{H}(q(t),p(t),u(t),p^0) = c$ on $[0,T]$,*

iv) *for every $v \in T_{q(T)}\mathcal{T}$, we have $\langle p(T), v\rangle = 0$ (transversality condition).*

We are going to apply the steps illustrated in Section 9.2.

Step 1. Use the maximization condition **ii)** to express, when possible, the control as a function of the state and of the covector. Considering the pre-Hamiltonian (11.6), we have two cases:

- **Abnormal extremals.** In this case, $p^0 = 0$ and we obtain

$$\langle p(t), F_i(q(t))\rangle = 0, \quad i = 1,\dots,m, \quad t \in [0,T] \tag{11.8}$$

  This condition does not permit to recover the control as function of the state and of the covector. Hence, for the sub-Riemannian problem, *abnormal extremals corresponds to singular controls.* In this case, some additional information can be recovered by deriving (11.8) with respect to time, cf. Section 11.4

[2] Equivalently, this is Theorem 9.3 with $F_0 = 0$ and the adding of transversality conditions since now the final point is constrained to belong to a smooth submanifold $\mathcal{T}$.

- **Normal extremals.** In this case $p^0 < 0$, and we can normalize it to be $p^0 = -1/2$. This gives

$$u_i(t) = \langle p(t), F_i(q(t))\rangle, \quad i = 1, \dots, m, \quad t \in [0, T]. \tag{11.9}$$

This condition expresses the control as function of the state and of the covector. More explicitly with the notations of Section 9.2 we obtain

$$u = w(q,p) := (\langle p, F_1(q)\rangle, \dots, \langle p, F_m(q)\rangle).$$

Hence, for sub-Riemannian problems *normal extremals corresponds to regular controls.*

Step 2. Insert the control found in the previous step into the Hamiltonian equations **i)**. We are going to develop this step for normal extremals only, since for abnormal ones the problem could be complicated and must be treated on a case by case basis (cf., e.g., Section 12.2). To this purpose, it is useful to introduce the maximized Hamiltonian, i.e. the pre-Hamiltonian computed on the controls obtained with the maximization condition. This reads

$$\begin{aligned}\mathcal{H}_M(q,p) := \mathcal{H}(q,p,w(q,p),-1/2) &= \langle p, \sum_{i=1}^m w_i(q,p)F_i(q)\rangle - 1/2\sum_{i=1}^m w_i(q,p)^2 \\ &= 1/2\sum_{i=1}^m w_i(q,p)^2 = 1/2\sum_{i=1}^m \langle p, F_i(q)\rangle^2.\end{aligned}$$

Notice that $\mathcal{H}_M(q(t),p(t)) = 1/2\sum_{i=1}^m w_i(q(t),p(t))^2 = 1/2|u(t)|^2$. This quantity is constant and equal to 1/2 if trajectories are parameterized by arclength. A useful fact is that

$$\begin{aligned}\frac{\partial \mathcal{H}_M}{\partial p}(q,p) &= \frac{\partial}{\partial p}\mathcal{H}(q,p,w(q,p),-1/2) \\ &= \frac{\partial \mathcal{H}}{\partial p}(q,p,w(q,p),-1/2) + \frac{\partial \mathcal{H}}{\partial u}(q,p,w(q,p),-1/2)\frac{\partial w}{\partial p}(q,p) \\ &= \frac{\partial \mathcal{H}}{\partial p}(q,p,w(q,p),-1/2).\end{aligned}$$

Here, we used the fact that $\frac{\partial \mathcal{H}}{\partial u}(q,p,w(q,p),-1/2) = 0$. Similarly

$$\frac{\partial \mathcal{H}_M}{\partial q}(q,p) = \frac{\partial \mathcal{H}}{\partial q}(q,p,w(q,p),-1/2).$$

Hence the Hamiltonian equations **i)** can be written in terms of the maximized Hamiltonian:

$$\begin{cases}\dot{q} = \dfrac{\partial \mathcal{H}_M}{\partial p}(q,p) \\ \dot{p} = -\dfrac{\partial \mathcal{H}_M}{\partial q}(q,p).\end{cases} \tag{11.10}$$

Observe that this allows to remove the control from the Hamiltonian equations.

These equations must be solved for fixed $q(0) = q_{\text{in}}$ and *any* $p(0) = p_{\text{in}} \in T^*_{q_{\text{in}}}M$.

If the final time is fixed in such a way that trajectories are parametrized by arclength, then it is convenient to normalize p_{in} in such a way that

$$\mathcal{H}_M(q_{\text{in}}, p_{\text{in}}) = \frac{1}{2}. \tag{11.11}$$

Remark 11.9. Notice that the set $\{p_{\text{in}} \in T_{q_{\text{in}}}M \mid \mathcal{H}_M(q_{\text{in}}, p_{\text{in}}) = 1/2\}$ is a ellipsoid if $\dim(\operatorname{span}\{F_1(q_{\text{in}}), \dots F_m(q_{\text{in}})\}) = n$ or a cylinder with ellipsoidal base otherwise.

The next steps are treated as a general OCP. If in the previous step we found $q(T; p_{\text{in}}, p^0)$ and $p(T; p_{\text{in}}, p^0)$ we have now to look for p_{in} and p^0 such that

$$\begin{cases} q(T; p_{\text{in}}, p^0) \in \mathcal{T} \\ \langle p(T; p_{\text{in}}, p^0), T_{q_{\text{in}}}\mathcal{T} \rangle = 0. \end{cases} \tag{11.12}$$

where the last condition is empty if $\mathcal{T}$ is a single point. If we can guarantee existence of optimal trajectories, and we are lucky because (11.12) provides a uniques pair (p^0, p_{in}) then if we have found the optimal solution. This however is not the general situation.

11.4. Abnormal extremals

In this section we discuss the existence of abnormal extremals. We recall that abnormal extremals do not depend on the cost.

We start with two examples where it is easy to guarantee the absence of abnormal extremals, and then we present a necessary condition for the existence of abnormal extremals.

The Riemannian case Let us consider the control system

$$\dot{q} = \sum_{i=1}^{n} u_i F_i(q), \qquad u \in L^\infty([0,T], \mathbb{R}^n), \tag{11.13}$$

where $n = \dim M$ and $\operatorname{span}\{F_1(q), \dots, F_n(q)\} = TM$ (in particular, we can freely move in every direction). That is, the system is a trivial control system, as discussed in the Introduction.

Let us show that, independently of the cost, this system does not admit abnormal extremals. Indeed, in (11.8) we observed that in the abnormal case it holds

$$\langle p(t), F_i(q(t)) \rangle = 0, \qquad i = 1, \dots, n.$$

Since $\operatorname{span}\{F_1(q), \dots, F_n(q)\} = T_qM$ this implies that $p(t) = 0$ a.e. $t \in [0,T]$, which contradicts the non-vanishing condition on the covector (p^0, p) in the PMP.

The 3D contact case Let $n = \dim M = 3$ and consider the control system

$$\dot{q} = u_1 F_1(q) + u_2 F_2(q), \qquad u \in L^\infty([0,T], \mathbb{R}^2). \tag{11.14}$$

We assume that this control system *contact*, meaning that the Chow-Hörmander condition is satisfied with Lie brackets of length at most 2. That is,

$$T_q M = \operatorname{span}\{F_1(q), F_2(q), [F_1, F_2](q)\}, \qquad \forall q \in M. \tag{11.15}$$

Proceeding as in the previous example, one has that the covector $p(\cdot)$ associated with an abnormal extremal should annihilate all the F_i's, that is,

$$\langle p(t), F_1(q(t))\rangle = \langle p(t), F_2(q(t))\rangle = 0, \qquad \forall t \in [0,T]. \tag{11.16}$$

We want to compute the derivative of the above equations with respect to time to get more information on $p(\cdot)$. Let us fix some coordinates, so that (11.16) reads $pf_i(x) = 0$ for $i = 1, 2$. Then, thanks to the Hamiltonian equations (11.10), we can compute

$$\dot{p} = -\frac{\partial \mathcal{H}_M}{\partial x}(x) = -p\left(u_1 \frac{\partial F_1}{\partial x}(x) + u_2 \frac{\partial F_2}{\partial x}(x)\right).$$

This allows to obtain the following expression for the derivative of $\langle p, f_1(x)\rangle$:

$$\begin{aligned}
0 &= \frac{d}{dt} p f_1(x) \\
&= \dot{p} f_1(x) + p \frac{\partial F_1}{\partial x}(x)\dot{x} \\
&= -p\left(u_1 \frac{\partial F_1}{\partial x}(x) + u_2 \frac{\partial F_2}{\partial x}(x)\right) F_1(x) + p\frac{\partial F_1}{\partial x}(x)\left(u_1 F_1(x) + u_2 F_2(x)\right) \\
&= u_2\, p\left(\frac{\partial F_1}{\partial x}(x) F_2(x) - \frac{\partial F_2}{\partial x}(x) F_1(x)\right) \\
&= u_2\, p[F_2, F_1](x).
\end{aligned}$$

The same computation with $i = 2$ yields $u_1\, p^\top [F_1, F_2](x) = 0$. That is,

$$u_1(t)\langle p(t), [F_1, F_2](x(t))\rangle = u_2(t)\langle p(t), [F_1, F_2](x(t))\rangle = 0, \qquad \forall t \in [0,T]. \tag{11.17}$$

Since $u_1(t) = u_2(t) = 0$ would yield a constant (trivial) extremal, the above and (11.16) prove that any nontrivial abnormal extremal must satisfy, for all $t \in [0,T]$,

$$\begin{cases} \langle p(t), F_1(x(t))\rangle, \\ \langle p(t), F_2(x(t))\rangle, \\ \langle p(t), [F_1, F_2](x(t))\rangle \end{cases} \tag{11.18}$$

Due to the contact assumption (11.15), this shows that $p(t) \equiv 0$, which contradicts condition i. of the PMP.

Actually, this argument works without any assumptions. In particular, we have the following necessary condition for the existence of abnormal extremals.

Theorem 11.10 (Goh condition)**.** *If a sub-Riemannian problem admits a nontrivial abnormal extremal with covector $p(\cdot)$, then $p(t) \neq 0$ for any $t \in [0, T]$ and it holds*

$$\langle p(t), F_i(q(t)) \rangle = 0, \qquad \langle p(t), [F_i, F_j](q(t)) \rangle = 0, \qquad \forall i, j = 1, \ldots, m, \quad t \in [0, T].$$

12. Examples of sub-Riemannian control problems

In this chapter we present some examples of sub-Riemannian optimal control problems, illustrating the general theory developed in Chapter 11. These examples are:

- In Section 12.1 we consider the sub-Riemannian problem on the Heisenberg group. We show how it can be interpreted as the classical isoperimetric problem in the plane, and we compute the geodesics explicitly using the Pontryagin Maximum Principle.
- In Section 12.2 we consider the Martinet sub-Riemannian structure, which is a prototypical example of a sub-Riemannian structure admitting non-trivial abnormal extremals.

12.1. The Heisenberg group and the isoperimetric problem.

Let us consider the isoperimetric problem in the plane:

Find the closed curve Γ in the plane that is shortest
among those enclosing a fixed area $A > 0$.

This problem is also knwon as *Dido's problem* from the name of the queen of Carthage who, according to Virgil's Eneid, was given by the king of Sicily as much land as she could enclose with a piece of rope. The name *isoperimetric* comes from the fact that the problem can be recast as follows: find the curve Γ in the plane that encloses a fixed area $A > 0$ and has fixed length $L > 0$ such that L is minimal.

Let us see how to recast this problem as an optimal control problem. A curve Γ in the plane can be parametrized by a function $\gamma : [0,1] \to \mathbb{R}^2$, $\gamma(t) = (x(t), y(t))$. The length of Γ is given by

$$\ell(\Gamma) = \int_0^1 \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} dt.$$

Observe that a change in parametrization (i.e., considering $\tilde{\gamma}(t) = \gamma(\alpha(t))$ for some increasing function $\alpha : [0,1] \to [0,1]$) does not change the length of the curve. Letting $\Omega \subset \mathbb{R}^2$ be the region enclosed by γ, Stokes Theorem allows to compute its (signed) area by the formula

$$\text{Area}(\Omega) = \int_\Omega dx \wedge dy = \frac{1}{2} \int_\Gamma (xdy - ydx) = \frac{1}{2} \int_0^1 (x\,\dot{y} - y\,\dot{x}) dt.$$

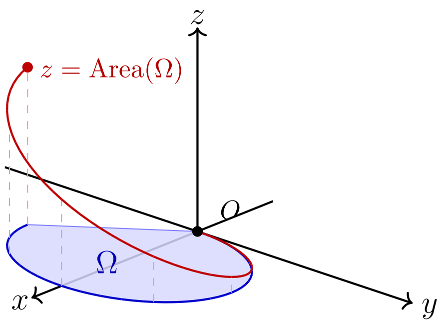


Figure 12.1.: Interpreting the isoperimetric problem as a sub-Riemannian problem on the Heisenberg group.

This suggests to introduce a third variable z by letting

$$z(t) = \frac{1}{2}\int_0^t (x(s)\dot{y}(s) - y(s)\dot{x}(s))ds. \tag{12.1}$$

Observe that $z(1) = \text{Area}(\Omega)$, and thus the constraint on the area can be recast as $z(1) = A$.

Since the area is invariant by translations, we can assume without loss of generality that $\gamma(0) = (0, 0)$. Considering $\dot{x} = u_1$ and $\dot{y} = u_2$, to be our controls, the isoperimetric problem is then recast as follows:

Among the curves $\gamma : [0, T] \to \mathbb{R}^2$ such that $\dot{\gamma} = (u_1, u_2 \in \mathbb{R}^2)$, $\gamma(0) = \gamma(T) = (0, 0)$ and $z(T) = A$, find the one minimizing the cost

$$\int_0^T \sqrt{u_1(t)^2 + u_2(t)^2}dt. \tag{12.2}$$

That is, we want to solve the following optimal control problem with final time T free:

$$\begin{cases} \dot{x} = u_1, \\ \dot{y} = u_2, \\ \dot{z} = \frac{1}{2}(xu_2 - yu_1), \gamma(0) = (0, 0, 0), \qquad \gamma(T) = (0, 0, A), \\ \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2}dt \to \min. \end{cases} \tag{12.3}$$

The above system is called the *Heisenberg system*, and the space $\mathbb{R}^3$ with coordinates (x, y, z) is called the *Heisenberg group*. By Chapter 11, this system is a sub-Riemannian system. In particular, we can consider the following equivalent formulation of the problem, where $T > 0$ is now fixed:

$$\begin{cases} \dot{q} = u_1F_1(q) + u_2F_2(q), \\ q(0) = (0, 0, 0), \qquad q(T) = (0, 0, A), \\ \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2}dt \to \min, \qquad u \in L^\infty([0, T], \mathbb{R}^2), \end{cases}$$

where, the vector fields driving the system are given by

$$F_1(x,y,z) = \begin{pmatrix} 1 \\ 0 \\ -\frac{y}{2} \end{pmatrix}, \qquad F_2(x,y,z) = \begin{pmatrix} 0 \\ 1 \\ \frac{x}{2} \end{pmatrix}. \tag{12.4}$$

Let us derive the optimal trajectories for this problem, for any $q(T) = q_{\rm fi}$.

Extremal trajectories. The existence of minimizers follows by Chapter 11, since the system satisfies the Lie bracket generating condition. Indeed,

$$[F_1, F_2](x,y,z) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

which implies that $\{F_1, F_2, [F_1, F_2]\}|_q$ is a basis of $T_q\mathbb{R}^3$ for any $q \in \mathbb{R}^3$. Observe, moreover, that the system is of contact type, and thus all nontrivial minimizers are normal extremals (see Chapter 11).

Since the problem is sub-Riemannian, we already know that controls corresponding to normal extremals are regular and satisfy

$$u_i(t) = \langle p(t), F_i(q(t))\rangle, \qquad i = 1, 2.$$

It follows that the maximized Hamiltonian reads

$$H(q,p) = \frac{1}{2}\left(\langle p, F_1(q)\rangle^2 + \langle p, F_2(q)\rangle^2\right) = \frac{1}{2}\left(\left(p_x - \frac{y}{2}p_z\right)^2 + \left(p_y + \frac{x}{2}p_z\right)^2\right). \tag{12.5}$$

Here, we denoted $p = (p_x, p_y, p_z)$ the covector. The Hamiltonian system $\dot q = \partial H/\partial p$ and $\dot p = -\partial H/\partial q$, reads:

$$\begin{cases} \dot x = \dfrac{\partial H}{\partial p_x} = p_x - \dfrac{y}{2}p_z, \\ \dot y = \dfrac{\partial H}{\partial p_y} = p_y + \dfrac{x}{2}p_z, \\ \dot z = \dfrac{\partial H}{\partial p_z} = \dfrac{h_2(q,p)x - h_1(q,p)y}{2}, \end{cases} \qquad \begin{cases} \dot p_x = -\dfrac{\partial H}{\partial x} = \left(p_x - \dfrac{y}{2}p_z\right)\dfrac{p_z}{2}, \\ \dot p_y = -\dfrac{\partial H}{\partial y} = -\left(p_y + \dfrac{x}{2}p_z\right)\dfrac{p_x}{2}, \\ \dot p_z = 0. \end{cases} \tag{12.6}$$

We immediately observe that p_z is constant. Thus, the sub-system for (x, y, p_x, p_y) is linear, and can be explicitly integrated.

To this purpose, observe that we have the following integral of motions:

- $p_z(t) = a \in \mathbb{R}$ for all $t \in [0, 1]$;
- $H(q,p) = c \in \mathbb{R}$, thanks to the PMP;
- $k_1 := ay/2 + p_x$;

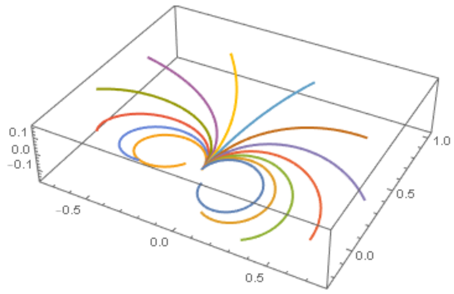


Figure 12.2.: Some extremals for the Heisenberg group

- $k_2 := ax/2 - p_y$.

The last two can be directly checked via (12.6):

$$\dot{k}_1 = \frac{a}{2}\dot{y} + \dot{p}_x = 0, \qquad \dot{k}_2 = \frac{a}{2}\dot{x} - \dot{p}_y = 0. \tag{12.7}$$

Recall that by (11.11) we can assume $H \equiv 1/2$. Since we have that $q(0) = 0$, and thus

$$\frac{1}{2} = H(q(0), p(0)) = \frac{1}{2}\left(p_x(0)^2 + p_y(0)^2\right). \tag{12.8}$$

That is, normal extremals are parametrised by points $p(0) = (\cos\theta, \sin\theta, a)$ on the cylinder $\mathbb{S}^1 \times \mathbb{R} \subset \mathbb{R}^3$.

We now use the constant of motions k_1 and k_2 to determine p_x and p_y. Indeed, $k_1(0) = p_x(0) = \cos\theta$ and $k_2(0) = p_y(0) = \sin\theta$, and

$$p_x(t) = -\frac{a}{2}y(t) + k_1(t) = -\frac{a}{2}y(t) + k_1(0) = -\frac{a}{2}y(t) + \cos\theta, \tag{12.9}$$

$$p_y(t) = \frac{a}{2}x(t) - k_2(t) = \frac{a}{2}x(t) - k_2(0) = \frac{a}{2}x(t) - \sin\theta. \tag{12.10}$$

Plugging these in the equations for $(\dot{x}, \dot{y})$ and recalling that $p_z(t) = a$, we get

$$\begin{cases} \dot{x} = -ay - \sin\theta, \\ \dot{y} = ax + \cos\theta. \end{cases} \tag{12.11}$$

That is, (x, y) moves along circular trajectories of radius $1/a$ and centered at $(-\frac{\cos\theta}{a}, -\frac{\sin\theta}{a})$:

$$\begin{cases} x(t) = \dfrac{\cos(at+\theta) - \cos\theta}{a}, \\ y(t) = \dfrac{\sin(at+\theta) - \sin\theta}{a}. \end{cases} \tag{12.12}$$

Finally, we can determine z by the original expression (12.1) of the control system (observe that $u_1 = \dot{x}$ and $u_2 = \dot{y}$):

$$z(t) = \int_0^t \dot{z}\, ds = \int_0^t \frac{x\dot{y} - y\dot{x}}{2}\, ds = \frac{at - \sin(at)}{2a^2}. \tag{12.13}$$

Some extremals are plotted in Figure 12.2.

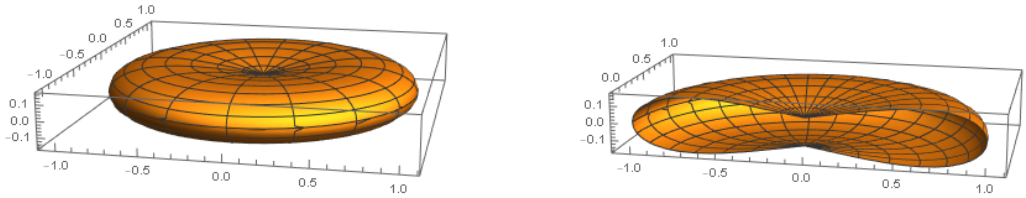


Figure 12.3.: Set of points joined by optimal trajectories with cost less than 1 in the Heisenberg group. *Left:* The full set. *Right:* The same set cut along the xz plane.

Optimal trajectories. Up to now we have determined a map associating to each $(\theta, a) \in \mathbb{S}^1 \times \mathbb{R}$ the normal extremal $\gamma_{\theta,a}(t) = (x(t), y(t), z(t))$.

Let now $q_{\text{fi}} = (x_1, y_1, z_1)$ be a point in $\mathbb{R}^3$. We want to determine the normal extremal $\gamma_{\theta,a}$ such that $\gamma_{\theta,a}(T) = q_{\text{fi}}$ for some $T > 0$. If $(x_1, y_1) \neq (0,0)$, this can be univoquely obtained by looking for the circle passing through $(0,0)$ and (x_1, y_1) and such that the (signed) area of the circular sector defined by the two points and the center of the circle is z_1. It is easy to check that such a circle exists and is unique, and thus there exists a unique normal extremal $\gamma_{\theta,a}$ such that $\gamma_{\theta,a}(T) = q_{\text{fi}}$ for some $T > 0$.

Assume now that $(x_1, y_1) = (0,0)$, and thus $q_{\text{fi}} = (0, 0, z_1)$ (observe that this is exactly the isoperimetric problem from which we started). In this case, there are infinitely many normal extremals $\gamma_{\theta,a}$ such that $\gamma_{\theta,a}(T) = q_{\text{fi}}$ for some $T > 0$. Indeed, any circle centered at the origin passes through $(0,0)$, and the area of the circular sector defined by the two points and the center of the circle is z_1 if T is chosen such that $aT - \sin(aT) = 2a^2 z_1$. Hence, there is a one-parameter family of normal extremals $\gamma_{\theta,a}$ such that $\gamma_{\theta,a}(T) = q_{\text{fi}}$ for some $T > 0$.

The previous observation implies that any normal extremal $\gamma_{\theta,a}$ is an optimal trajectory for all times $t \in (0, 2\pi/a]$. That is, after the time $t = 2\pi/a$, at which $\gamma_{\theta,a}$ arrives on the vertical axis $\{x = y = 0\}$, the trajectory loses optimality. It follows that there is no universal time $T_{\max}$ such that the normal extremals are optimal up to time $t = T_{\max}$: Conjugate times depend on the trajectory and become arbitrary small as $a \to +\infty$.

The boundary of the set of points joined by optimal trajectories with cost less than 1 is depicted in Figure 12.3.

12.2. The Martinet structure

We consider the following control system on $\mathbb{R}^3$:

$$\begin{cases} \dot{x} = u_1, \\ \dot{y} = u_2, \\ \dot{z} = u_1 \dfrac{y^2}{2}. \end{cases} \tag{12.14}$$

This is a system of the form $\dot{q} = u_1 f_1(q) + u_2 f_2(q)$, where $q = (x, y, z)$ and

$$f_1(x, y, z) = \begin{pmatrix} 1 \\ 0 \\ \frac{y^2}{2} \end{pmatrix}, \qquad \text{and} \qquad f_2(x, y, z) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{12.15}$$

One can compute

$$[f_1, f_2](x, y, z) = \begin{pmatrix} 0 \\ 0 \\ y \end{pmatrix}, \qquad [f_2, [f_1, f_2]](x, y, z) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{12.16}$$

This control system satisfy the Chow-Hörmander condition, and is thus controllable. However, the contact condition (11.15) is satisfied only outside the *Martinet surface* $\mathcal{M} = \{y = 0\}$. Indeed, we have

$$\operatorname{span}\{f_1(q), f_2(q), [f_1, f_2](q)\} = \begin{cases} \mathbb{R}^3, & \text{if } q \notin \mathcal{M} \\ \mathbb{R}^2 \times \{0\} & \text{if } q \in \mathcal{M}. \end{cases} \tag{12.17}$$

Indeed, on $\mathcal{M}$ we need the additional bracket $[f_2, [f_1, f_2]]$ to satisfy the Chow-Hörmander condition.

Since the structure is not contact, let us look for abnormal extremals. Since we are considering an abnormal extremal $q(\cdot)$, Goh condition (see Theorem 11.10) implies that for any $t \in [0, T]$ it holds

$$\langle p(t), v \rangle = 0 \qquad \forall v \in \operatorname{span}\{f_1(q(t)), f_2(q(t)), [f_1, f_2](q(t))\}. \tag{12.18}$$

In particular $q(\cdot)$ is entirely contained in the Martinet surface $\mathcal{M}$. Otherwise $p(t) = 0$ which contradicts the nontriviality of $(p(t), p^0)$.

Letting $p(t) = (p_x(t), p_y(t), p_z(t))$, condition (12.18) yields that

$$p_x(t) + \frac{y(t)^2}{2} p_y(t) = 0 \tag{12.19}$$

$$p_y(t) = 0. \tag{12.20}$$

Moreover, the same computations as in (11.4) can be used to differentiate the relation $y(t)p_z(t) = \langle p(t), [F_1, F_2](q(t))\rangle \equiv 0$. Using the fact that $[F_1, [F_1, F_2]] \equiv 0$, we obtain that

$$0 = \frac{d}{dt}\langle p(t), [F_1, F_2](q(t))\rangle = u_2(t)p[F_2, [F_1, F_2]](q(t)), \qquad \forall t \in [0, T]. \tag{12.21}$$

By the explicit expression of $[F_2, [F_1, F_2]]$, this yields

$$u_2(t)p_z(t) = 0, \qquad \forall t \in [0, T].$$

Putting this together with (12.19)-(12.20), we finally obtain that

$$p_x(t) = p_y(t) = u_2(t)p_z(t) = 0, \qquad \forall t \in [0, T]. \tag{12.22}$$

Since $p(t) \neq 0$ by the PMP, from the above equation we deduce that $u_2(t) = 0$ for all $t \in [0, T]$. Recallt that, using the fact that the Hamiltonian is constant, we can assume that $u_1(t)^2 + u_2(t) = 1$. It follows that $u_1(t) = \pm 1$.

Let us now determine the trajectories corresponding to the abnormal extremals. The pre-Hamiltonian with $p^0 = -1$ reads

$$H(q, p, u) = u_1\left(p_x + \frac{y^2}{2}p_y\right) + u_2 p_y - \frac{1}{2}(u_1^2 + u_2^2). \tag{12.23}$$

Hence, the Hamiltonian equations read

$$\begin{cases} \dot{x} = u_1, \\ \dot{y} = u_2, \\ \dot{z} = u_1\frac{y^2}{2}, \end{cases} \qquad \begin{cases} \dot{p}_x = 0, \\ \dot{p}_y = -u_1 p_z y, \\ \dot{p}_z = 0. \end{cases}$$

It follows that the abnormal extremals starting at the origin are given by

$$q(t) = (\pm t, 0, 0), \qquad p(t) = (0, 0, a), \qquad a \in \mathbb{R}. \tag{12.24}$$

We will not analyse in detail the normal extremals, since their structure is quite complicated (see Figure 12.4 for a depiction of the boundary of the set of points joined by optimal trajectories with cost less than 1). However, we remark that the Hamiltonian system associated with the normal Hamiltonian (i.e., with $u_1 = p_x + p_z\frac{y^2}{2}$ and $u_2 = p_y$) is

$$\begin{cases} \dot{x} = p_x + p_z\frac{y^2}{2}, \\ \dot{y} = p_y, \\ \dot{z} = (p_x + p_z\frac{y^2}{2})\frac{y^2}{2}, \end{cases} \qquad \begin{cases} \dot{p}_x = 0, \\ \dot{p}_y = -\left(p_x + p_z\frac{y^2}{2}\right)p_z y, \\ \dot{p}_z = 0. \end{cases} \tag{12.25}$$

This implies that the following are all normal extremals:

$$q(t) = (\pm t, 0, 0), \qquad p(t) = (\pm 1, 0, a), \qquad a \in \mathbb{R}. \tag{12.26}$$

In particular, the trajectories $t \mapsto (\pm t, 0, 0)$ given by the control $u(t) = (\pm 1, 0)$ are at the same time normal and abnormal.

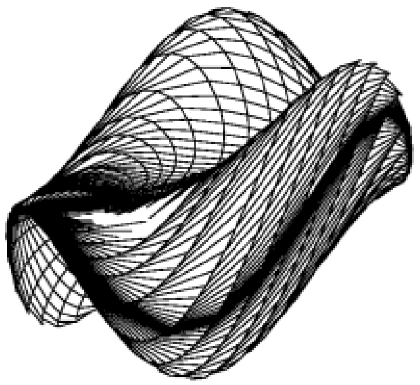


Figure 12.4.: Set of points joined by optimal trajectories with cost less than 1 for the Martinet structure. Notice the deeper singularities w.r.t. the Heisenberg case. Image from [Mon02].

13. Time-optimal control problem

In this chapter we are going to study an important class of optimal control problems namely minimum time problems for control affine systems on the plane with bounded controls. We already saw an example of such problems in Chapter 10, where we studied the minimum time problem for the linear system modelling a cart.

Definition 13.1 (Minimum time problem for 2D control affine systems)**.**

$$\begin{cases} \dot{x} = F(x) + u(t)G(x), \\ x(0) = x_{\text{in}}, \quad x(T) = x_{\text{fin}} \\ \\ T \to \min \end{cases} \tag{OCP-T}$$

Here,

- $T \geq 0$ is free,
- F, G are smooth vector fields on $\mathbb{R}^2$,
- $u(\cdot) \in L^\infty([0,T],[-1,1])$,
- $x(\cdot) : [0,T] \to \mathbb{R}^2$, belongs to the set of Lipschitz curves.

Notation In this chapter since we are working in $\mathbb{R}^2$, we represent covectors as row vectors and we indicate the duality product between covectors and vectors with a dot. Namely $\langle p, v\rangle = p \cdot v$ (here $p \in T_x^*\mathbb{R}^2 = \mathbb{R}^{*2}$ and $v \in T_x\mathbb{R}^2 = \mathbb{R}^2$ and x is the point on the plane where v and p are applied).

To avoid complicated notations, we assume in this chapter that for every $u(\cdot) \in L^\infty([0,T],[-1,1])$ the solution of (13.10) with $x(0) = x_{\text{in}}$ exists for any time. Moreover, we assume that a solution to (13.10) exists. This can be verified, e.g., by applying the Filippov test in Proposition 8.10.

13.1. Application of the Pontryagin Maximum Principle

In this case the pre-Hamiltonian (11.6) is

$$\mathcal{H}(x, p, u, p^0) = p \cdot F(x) + u\, p \cdot G(x) + p^0 \tag{13.1}$$

with

$$(x, p, u, p^0) \in \mathbb{R}^2 \times \mathbb{R}^{*2} \times [-1, 1] \times \mathbb{R}.$$

and the Pontryagin Maximum Principle (Theorem 9.1) tells the following.

Theorem 13.2. *If the pair $(x, u) : [0, T] \to \mathbb{R}^2 \times [-1, 1]$ is optimal, then there exists a never vanishing Lipschitz continuous pair $(p, p^0) : [0, T] \ni t \mapsto (p(t), p^0) \in \mathbb{R}^{*2} \times \mathbb{R}$ where $p^0 \leq 0$ is a constant and such that for almost every $t \in [0, T]$ we have*

i) Hamiltonian equations*:*

$$\dot{p}(t) = -p(t) \cdot \Big(\frac{\partial F}{\partial x} + u(t)\frac{\partial G}{\partial x}\Big)(x(t)) \tag{13.2}$$

ii) Maximization condition*:*

$$u(t)\ p(t) \cdot G(x(t)) = \max_{v \in [-1,1]} v\ p(t) \cdot G(x(t)). \tag{13.3}$$

iii) Zero value of the Hamiltonian*:*

$$p(t) \cdot F(x(t)) + u(t)\ p(t) \cdot G(x(t)) + p^0 = 0 \tag{13.4}$$

Equation (13.4) comes from the fact that T is free. A simple and important property is the following.

Lemma 13.3. *For the problem* (13.10) *we have $p(t) \neq 0$ for every t.*

Proof. Actually if $p(t) = 0$ for a time t then from (13.2) we would get $p(t) \equiv 0$. Then from (13.4) we would get $p^0 = 0$ which is impossible. □

For this kind of problems, instead of normalizing p^0 could be useful to normalize $p(0)$ for instance setting $\|p(0)\| = 1$.

We are going to apply the steps illustrated in Section 9.2.

Step 1. Use the maximization condition **ii)** to express, when possible, the control as a function of the state and of the covector. To this purpose we introduce the following.

Definition 13.4. The *switching function* associated with an extremal $(x(\cdot), p(\cdot), u(\cdot), p^0)$ defined on $[0, T]$ is the function $\phi(\cdot) : [0, T] \to \mathbb{R}$ defined as

$$\phi(t) := p(t) \cdot G(x(t)).$$

We already saw this object when we studied abnormal extremals in the sub-Riemannian case.

Then, equation (13.3) can be rewritten as the following:

Lemma 13.5. *If the switching function $\phi(\cdot)$ has only isolated zeros then $u(t) = sgn(\phi(t))$ on $[0, T]$.*

Definition 13.6. (bang-bang extremal) An extremal satisfying the hypotheses of Lemma 13.5 is called a bang-bang extremal. The times at which $u(\cdot)$ changes sign are called *switching times*.

Clearly bang-bang extremals are regular extremals. However there are situations in which the Pontryagin Maximum Principle does not permit to obtain directly the value of the control (think, e.g., of the case $\phi(t) \equiv 0$).

Definition 13.7. An extremal is *singular on* $[t_1, t_2] \subset [0, T]$ if the corresponding switching function is identically zero on $[t_1, t_2]$.

To obtain information concerning the control for singular extremals we proceed similarly to the sub-Riemannian case. Let us observe that $\phi(\cdot)$ is a Lipschitz function and hence we can differentiate it a.e.. Then, proceding as for the Goh condition in the sub-Riemannian case, we obtain the following.

Lemma 13.8. *The function $\phi(\cdot)$ is of class $\mathcal{C}^1$. More precisely, we have*

$$\dot{\phi}(t) = p(t) \cdot [F, G](x(t)), \qquad \forall t \in [0, T].$$

Proof. By smoothness of $[F, G]$ and continuity of $p(\cdot)$ and $x(\cdot)$, it suffices to prove the second part of the statement for a.e. $t \in [0, T]$. Using the PMP we have for almost every t such that ϕ is differentiable at t,

$$\begin{aligned}
\dot{\phi}(t) &= \frac{d}{dt}(p(t) \cdot G(x(t))) \\
&= \dot{p}(t) \cdot G(x(t)) + p(t) \cdot \dot{G}(x(t)) \\
&= -p(t) \left(\frac{\partial F}{\partial x} + u(t) \frac{\partial G}{\partial x} \right) (x(t)) \cdot G(x(t)) + p(t) \cdot \frac{\partial G}{\partial x}(x(t)) \, (F + u(t)G)(x(t)) \\
&= p(t) \cdot [F, G](x(t)). \quad \square
\end{aligned}$$

Let us introduce the following two functions

$$\begin{aligned}
\Delta_A(x) &:= \det(F(x), G(x)) = F_1(x)G_2(x) - F_2(x)G_1(x), && (13.5) \\
\Delta_B(x) &:= \det(G(x), [F, G](x)) = G_1(x)[F, G]_2(x) - G_2(x)[F, G]_1(x). && (13.6)
\end{aligned}$$

These two functions encode geometric information about the relationship of F and G. Namely,

- $\Delta_A^{-1}(0)$ is the set of points where F and G are parallel,
- $\Delta_B^{-1}(0)$ is the set of points where G is parallel to $[F, G]$.

The next result shows that $\Delta_B^{-1}(0)$ plays the same role as the Martinet surface in the sub-Riemannian case (cf. 12.2).

Lemma 13.9. *Singular extremals are contained in $\Delta_B^{-1}(0)$. More precisely, if an extremal is singular on $[t_1, t_2] \subset [0, T]$ we have that $x([t_1, t_2]) \subset \Delta_B^{-1}(0)$.*

Proof. Recall that $p(t) \neq 0$ for every t. Then, the singular assumption implies that for any $t \in [t_1, t_2]$ we have

$$0 = \phi(t) = p(t) \cdot G(x(t)) \implies 0 = \dot{\phi}(t) = p(t) \cdot [F, G](x(t)).$$

Since $p(t) \neq 0$ and we are in a two-dimensional space, this implies that G is parallel to $[F, G]$ along $x(\cdot)$ on $[t_1, t_2]$. □

The next Lemma provides the value of the control for a singular extremal. In particular, it turns out that such a control can be given in *feedback form*, i.e., as a function of the state x.

Lemma 13.10. *Consider an extremal that is singular on $[t_1, t_2] \subset [0, T]$ and assume that the differential[1] $\mathrm{d}\,\Delta_B$ of Δ_B does not vanish on $[t_1, t_2]$. Then, the corresponding control $u(\cdot)$ satisfies the following feedback relation on $[t_1, t_2]$:*

$$u(t) = \varphi(x(t)), \qquad \textit{where} \qquad \varphi(x) = -\frac{\mathrm{d}\,\Delta_B(x) \cdot F(x)}{\mathrm{d}\,\Delta_B(x) \cdot G(x)}. \tag{13.7}$$

Proof. By Lemma 13.9, we have $\Delta_B(x(t)) = 0$ for any $t \in [t_1, t_2]$. Differentiating this relation, and using the fact that $\dot{x}(t) = F(x(t)) + u(t)G(x(t))$, we get that for a.e. $t \in [t_1, t_2]$ it holds

$$0 = \frac{d}{dt}\Delta_B(x(t)) = \mathrm{d}\,\Delta_B(x(t)) \cdot \left(F(x(t)) + u(t)G(x(t))\right).$$

Since $\mathrm{d}\,\Delta_B(x(t)) \cdot G(x(t)) \neq 0$ on $[t_1, t_2]$, this implies that the control $u(t)$ has to satisfy the above relation. □

En example of switching function with the corresponding controls is provided in Figure 13.1.

Step 2. Insert the control found in the previous step into the Hamiltonian equations (13.10) and (13.2). This step in general is more complicated than in the sub-Riemannian case. The reason is that the previous step provided the control as a function of the state and of the covector that in general is not smooth.

The construction of the extremals is in general complicated. However one can obtain additional information using the functions Δ_A and Δ_B.

We start with the following technical result.

Lemma 13.11. *Let $x \notin \Delta_A^{-1}(0) \cup \Delta_B^{-1}(0)$. Then, $[F, G](x)$ can be written as a linear combination of $F(x)$ and $G(x)$ and it holds*

$$[F, G](x) = f_S(x)F(x) \mod G(x), \qquad \textit{where} \qquad f_S(x) = -\frac{\Delta_B(x)}{\Delta_A(x)}. \tag{13.8}$$

[1] Recall that $d\Delta_B(x) = (\frac{\partial \Delta_B}{\partial x_1}, \frac{\partial \Delta_B}{\partial x_2})$ is a covector, and that we are using the dot to indicate the duality product between covectors and vectors.

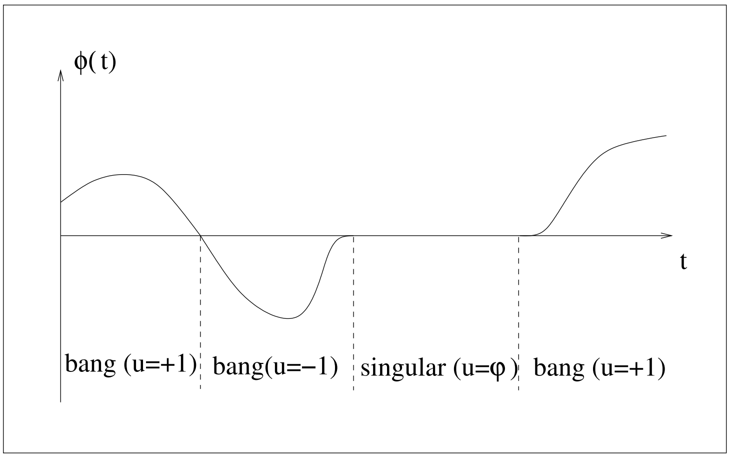


Figure 13.1.: Example of switching function $\phi(\cdot)$ with the corresponding control $u(\cdot)$. In the picture there are also the sets $\Delta_A^{-1}(0)$ and $\Delta_B^{-1}(0)$ and the corresponding extremal trajectory.

Proof. Let $[F,G](x) = f_S(x)F(x) + g_S(x)G(x)$ be the linear combination of F and G that gives $[F,G]$. We have:

$$\begin{aligned}\Delta_B(x) &= \det\big(G(x), [F,G](x)\big)\\ &= \det(G(x), f_S(x)F(x) + g_S(x)G(x))\\ &= f_S(x)\det\big(G(x), F(x)\big)\\ &= -f_S(x)\Delta_A(x).\end{aligned}$$

□

We are in a position to prove the following result that gives information on the structure of extremals in each connected component of $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$.

Theorem 13.12. *Let $\Omega \in \mathbb{R}^2$ be a connected component of $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$ and consider an extremal such that $x([t_1,t_2]) \subset \Omega$. Then $x(\cdot)|_{[t_1,t_2]}$ is bang-bang with at most one switching. Moreover, if $f_S > 0$ (resp. $f_S < 0$) in Ω then the corresponding control is either constant with value ± 1, or has a single switch $-1 \to +1$ switching (resp. has a single switch $+1 \to -1$).*

Proof. We prove the case $f_S > 0$ in Ω, the opposite case being similar. Let $\bar{t}$ be such that $\phi(\bar{t}) = p(\bar{t}) \cdot G(x(\bar{t})) = 0$. Then, denoting $\bar{p} := p(\bar{t})$ and $\bar{x} := x(\bar{t})$, we have:

$$\dot{\phi}(\bar{t}) = \bar{p} \cdot [F,G](\bar{x}) = \bar{p} \cdot (f_S(\bar{x})F(\bar{x}) + g_S(\bar{x})G(\bar{x}))(\bar{x}) = f_S(\bar{x})\ \bar{p} \cdot F(\bar{x}).$$

Now, the PMP implies that the Hamiltonian is constantly equal to zero along the extremal. Hence, we have that $p(t) \cdot F(x(t)) + u(t)\ p(t) \cdot G(x(t)) + p^0 = 0$ with $p^0 \leq 0$.

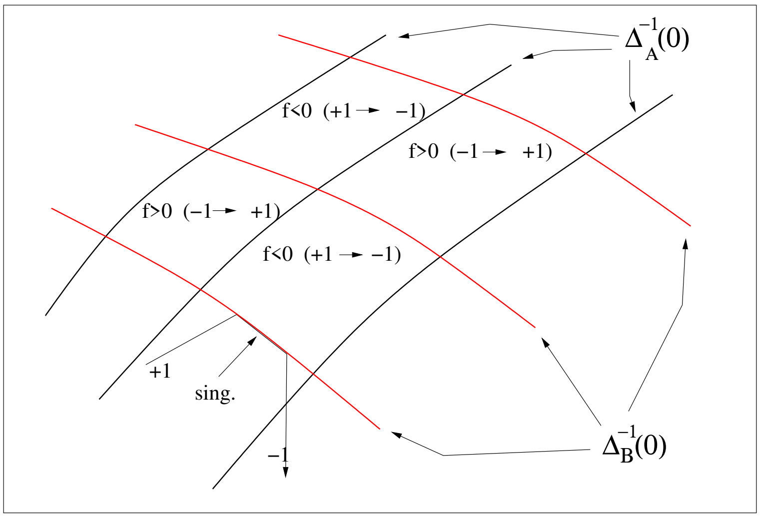


Figure 13.2.: Possible commutations in the connected components of $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$ and their relation with the sign of f_S (indicated as f in the picture). In the picture there is also an example of an extremal trajectory which is singular in a sub-interval.

Hence at $\bar{t}$ we have $\bar{p} \cdot F(\bar{x}) \geq 0$. Actually, $\dot{\phi}(\bar{t}) = \bar{p} \cdot F(\bar{x}) > 0$ otherwise $F(\bar{x})$ and $G(\bar{x})$ would be parallel, which is not possible since $\Delta_A(\bar{x}) \neq 0$.

We have thus proven that, on $[t_1, t_2]$, the switching function ϕ has positive derivative at any zero, which implies that it can vanish at most once on this interval. Moreover, such a zero corresponds to a switch from -1 to $+1$ for the control. □

The next steps are treated as a general OCP. If in the previous step we found $x(T; p_{\text{in}}, p^0)$ and $p(T; p_{\text{in}}, p^0)$ we have now to look for p_{in} and p^0 such that

$$x(T; p_{\text{in}}, p^0) = x_{\text{fi}} \tag{13.9}$$

The main difficulty for this class of problems is that in general for each p_{in} could correspond more than one trajectory $x(T; p_{\text{in}}, p^0)$. This is due to the non-smoothness of the dynamical system given by the Pontryagin maximum Principle.

13.2. Time optimal problems for linear systems

All these problems are easily solved for linear systems, i.e. for the problem

$$\dot{x} = Ax + u(t)Bx, \tag{13.10}$$
$$x(0) = x_{\text{in}}, \quad x(T) = x_{\text{fi}} \tag{13.11}$$
$$T \to \min \tag{13.12}$$

Here $A \in \mathbb{R}^{2\times 2}$, $B \in \mathbb{R}^2$ and $u(\cdot) \in L^\infty([0,T],[-1,1])$. In this case we have to assume that the matrix $[B, AB]$ is full rank otherwise the problem is trivial.

For these kind of problems one immediately verify:

- there is existence of optimal trajectories.
- $\Delta_B^{-1}(0) = \emptyset$ hence there are no singular trajectories.
- The equation (13.2) for $p(\cdot)$ involve neither $u(\cdot)$ nor $x(\cdot)$:

$$\dot{p}(t) = -p(t)A. \tag{13.13}$$

Hence for every p_{in} one solves (13.13), find $u(t)$ =sign$(p(t)\cdot B)$, and find $x(T; p_{\text{in}}, p^0)$. Notice that $x(T; p_{\text{in}}, p^0)$ is piecewise smooth since the zeros of $p(\cdot)$ cannot accumulate. In this case for every (p_{in}, p^0) there is a unique $x(T; p_{\text{in}}, p^0)$.

The problem is solved looking for all p_{in} for which $q(T; p_{\text{in}}, p^0) = x_{\text{fin}}$ and taking the one arriving at the smallest T.

Bibliography

[ABB19] Andrei Agrachev, Davide Barilari, and Ugo Boscain. *A Comprehensive Introduction to Sub-Riemannian Geometry*. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2019. DOI: 10.1017/9781108677325.

[Arn06] V. I. Arnold. *Ordinary Differential Equations*. 2nd printing, 2006. Universitext. Berlin: Springer, 2006. ISBN: 978-3-540-34563-3.

[AS04] Andrei A. Agrachev and Yuri L Sachkov. *Control Theory from the Geometric Viewpoint*. Control Theory and Optimization 2. Berlin: Springer, 2004. ISBN: 978-3-540-21019-1.

[AS22] Andrei Agrachev and Andrey Sarychev. "Control on the Manifolds of Mappings with a View to the Deep Learning". In: *Journal of Dynamical and Control Systems* 28.4 (Oct. 2022), pp. 989–1008. ISSN: 1079-2724, 1573-8698. DOI: 10.1007/s10883-021-09561-2. (Visited on 11/10/2025).

[Bos+23] Ugo Boscain et al. "Local Controllability Does Imply Global Controllability". In: *Comptes Rendus. Mathématique* 361.G11 (Dec. 2023), pp. 1813–1822. ISSN: 1631-073X, 1778-3569. DOI: 10.5802/crmath.538. (Visited on 02/12/2026).

[BP04] Ugo Boscain and Benedetto Piccoli. *Optimal Syntheses for Control Systems on 2-D Manifolds*. Mathématiques & Applications 43. Berlin ; New York: Springer, 2004. 261 pp. ISBN: 978-3-540-20306-3.

[BP07] Alberto Bressan and Benedetto Piccoli. *Introduction to the Mathematical Theory of Control: With 102 Figures and 107 Exercises*. AIMS Series on Applied Mathematics Vol. 2. Springfield, Mont: American Institute of Mathematical Sciences, 2007. 312 pp. ISBN: 978-1-60133-002-4.

[BP21] Ugo Boscain and Benedetto Piccoli. "Synthesis Theory in Optimal Control". In: *Encyclopedia of Systems and Control*. Ed. by John Baillieul and Tariq Samad. Cham: Springer International Publishing, 2021, pp. 2266–2275. ISBN: 978-3-030-44183-8 978-3-030-44184-5. DOI: 10.1007/978-3-030-44184-5_50. URL: https://link.springer.com/10.1007/978-3-030-44184-5_50 (visited on 11/30/2023).

[BR05] Frédéric Bonnans and Pierre Rouchon. *Commande et optimisation de systémes dynamiques*. Mathématiques appliquées. Palaiseau: Éd. de l'École Polytechnique, 2005. 280 pp. ISBN: 978-2-7302-1251-9.

[Do 93] Manfredo Perdigao Do Carmo. *Riemannian Geometry*. Mathematics. Boston Basel Berlin: Birkhäuser, 1993. ISBN: 978-3-7643-3490-1 978-0-8176-3490-2.

[He+16] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90. (Visited on 11/12/2025).

[Jur97] Velimir Jurdjevic. *Geometric Control Theory*. Cambridge Studies in Advanced Mathematics 51. Cambridge: Cambridge university press, 1997. ISBN: 978-0-521-49502-8.

[Kha92] Hassan K. Khalil. *Nonlinear Systems*. New York Oxford Singapore [etc.]: Maxwell MacMillan international, 1992. ISBN: 978-0-02-363541-0.

[Lee13] John M. Lee. *Introduction to Smooth Manifolds*. 2nd ed. Graduate Texts in Mathematics 218. New York ; London: Springer, 2013. 708 pp. ISBN: 978-1-4419-9981-8 978-1-4419-9982-5.

[Lib11] Daniel Liberzon. *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press, 2011.

[LM86] E. B. Lee and L. Markus. *Foundations of Optimal Control Theory*. Malabar, Fla: R.E. Krieger Pub. Co, 1986. 576 pp. ISBN: 978-0-89874-807-9.

[LT98] J. -P. Laumond and M. Thoma, eds. *Robot Motion Planning and Control*. Vol. 229. Lecture Notes in Control and Information Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. ISBN: 978-3-540-76219-5 978-3-540-40917-5. DOI: 10.1007/BFb0036069. (Visited on 12/03/2025).

[Mal66] Bernard Malgrange. *Ideals of Differentiable Functions*. Tata Institute Monographs on Mathematics & Physics. Oxford Univ. Press, 1966.

[Mon02] Richard Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics, and Applications*. Mathematical Surveys and Monographs 91. Providence, RI: American Mathematical Society, 2002. ISBN: 978-0-8218-4165-5 978-0-8218-1391-1.

[Sch63] Arthur J. Schwartz. "A Generalization of a Poincare-Bendixson Theorem to Closed Two-Dimensional Manifolds". In: *American Journal of Mathematics* 85.3 (July 1963), p. 453. ISSN: 00029327. DOI: 10.2307/2373135. JSTOR: 2373135. (Visited on 09/22/2025).

[SL12] Heinz M. Schättler and Urszula Ledzewicz. *Geometric Optimal Control: Theory, Methods and Examples*. Interdisciplinary Applied Mathematics 38. New York: Springer, 2012. 640 pp. ISBN: 978-1-4614-3833-5.

[Son98] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. 2. ed. Texts in Applied Mathematics 6. New York Heidelberg: Springer, 1998. 531 pp. ISBN: 978-0-387-98489-6.

[TG23] Paulo Tabuada and Bahman Gharesifard. "Universal Approximation Power of Deep Residual Neural Networks Through the Lens of Control". In: *IEEE Transactions on Automatic Control* 68.5 (May 2023), pp. 2715–2728. ISSN: 0018-9286, 1558-2523, 2334-3303. DOI: 10.1109/TAC.2022.3190051. (Visited on 11/03/2025).

[Zua25] Enrique Zuazua. *Machine Learning and Control: Foundations, Advances, and Perspectives*. Dec. 2025. DOI: 10.48550/arXiv.2510.03303. arXiv: 2510.03303 [math]. (Visited on 03/18/2026).