

Introduction to Geometric Control Theory

Ugo Boscain, Dario Prandi, Mario Sigalotti

February 16, 2022

Contents

1	Introduction	7
1.0.1	The controlled pendulum	10
1.0.2	A vehicle sliding on the ice, controlled by two fans	12
2	Manifolds, Vector fields, Brackets	17
2.1	Smooth differentiable manifolds	17
2.2	Curves, tangent vectors	20
2.3	Vector fields, flows, rectification	22
2.3.1	Nonautonomous vector fields	25
2.4	Lie Brackets	27
2.5	The Frobenius Theorem	30
3	Nonlinear controllability	31
3.1	Control systems as families of vector fields	32
3.1.1	Affine control systems	33
3.2	The Krener Theorem: local accessibility	33
3.3	Symmetric systems	35
3.4	Compatible vector fields	36
3.4.1	Convexification	37
3.4.2	Affine systems with recurrent drift	40
3.4.3	Affine systems with non-recurrent drift	43
3.5	Orbits and necessary conditions for controllability	44
4	Introduction to Optimal Control	47
4.1	The different steps to solve an optimal control problem	49
4.2	Existence of solutions for Optimal Control Problems: the Filippov test	50
4.3	First order conditions	54
4.3.1	Why Lagrange multipliers appear in constrained problems	55
5	The Pontryagin Maximum Principle (PMP): Formulation and Use	57
5.1	Statement of the Pontryagin Maximum Principle	57
5.2	Use of the Pontryagin Maximum Principle	59
5.2.1	Extremal and Optimal syntheses	61

6	Proof of the PMP for Control Affine Systems with Quadratic Cost	63
6.1	Statement of the PMP for control affine systems with quadratic cost . . .	64
6.2	Proof of the Pontryagin Maximum Principle	65
6.2.1	Notation	66
6.2.2	The variation	67
6.2.3	The crucial step	68
6.2.4	The Hamiltonian form	70
7	Sub-Riemannian Geometry	75
7.1	Energy, length and existence of optimal trajectories	76
7.2	Minimizers	81
8	Time-optimal Control on the Plane	85
8.1	Time optimal problems for linear systems	90
	Bibliographie	91

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 where we are fixing a system of coordinates.
- 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 $u(\cdot) : [0, \infty) \rightarrow U$ is the control. When we write simply u we usually mean a fixed value of the control $u \in U$.

Chapter 1

Introduction

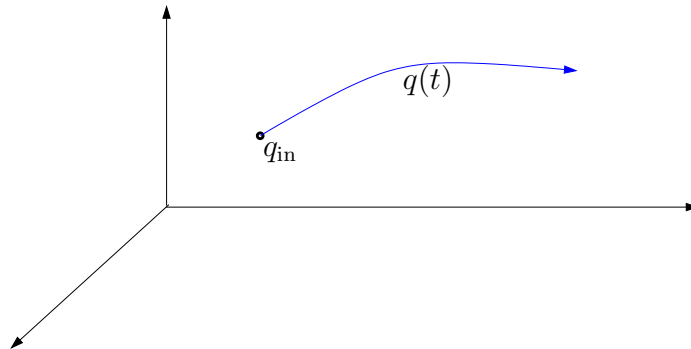
Consider a dynamical system on a smooth connected manifold M of dimension n .

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

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 $q(\cdot) : I \rightarrow M$, 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 with solutions of (1.1), that is*

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

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



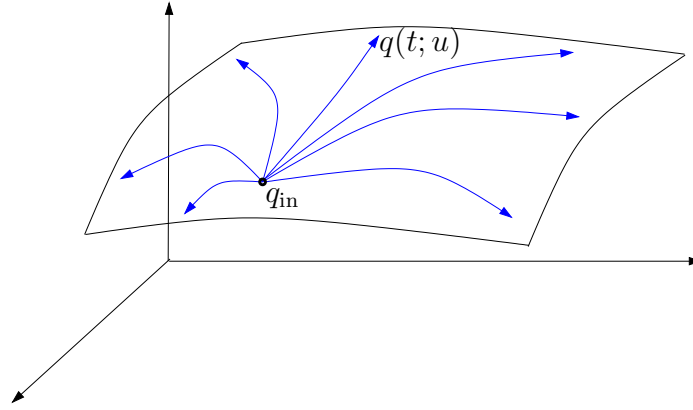
Assume now that F depends on an m -dimensional parameter $u = (u_1, \dots, u_m)$ belonging to a subset of U of $\mathbb{R}^m$. Here, we assume $m < n$, since this is the most common situation in practice and for clarity of presentation. Then, (1.1) becomes

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

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 to depend on time. Then, (1.2) becomes

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

If we assume that $u(\cdot)$ has a reasonable dependence on time (e.g., L_{loc}^∞) then, for every fixed $u(\cdot)$ the problem (1.3) has still local existence and uniqueness of solutions. Such a solution (that we denote by $q(t, u(\cdot))$) is a locally Lipschitz curve.

We can finally formulate the main question.

Q1 How large is the set of points that can be reached from q_{in} by varying $u(\cdot)$?

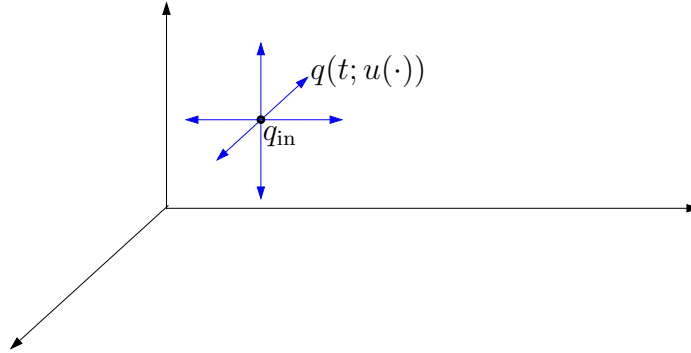
$$\mathcal{R}(q_{\text{in}}) := \{\bar{q} \in \mathbb{R}^n \mid \exists T \geq 0, u(\cdot) : [0, T] \rightarrow U, \text{ s.t. } q(T; u(\cdot)) = \bar{q}\}.$$

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 Chapter 3) tries to answer. The function $u(\cdot)$ is called the *control*. The set $\mathcal{R}(q_{\text{in}})$ is called the *reachable set* starting 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 (usually 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 $u(\cdot)$ to 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 $T_q M$, 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 $\mathbf{F}(q) = T_q M$ 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

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

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 Chapters 4-8.

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:

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

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

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

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 7.

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 $K : M \rightarrow U$, with $K(q_{\text{in}}) = u_0$, such that the dynamical system

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

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.

1.0.1 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, we are acting with on the mass with a (time-dependent) force

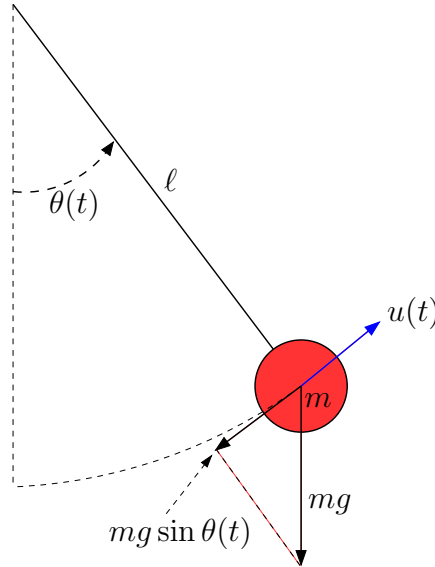
$$u(\cdot) : [0, \infty) \rightarrow [-u_{\max}, u_{\max}],$$

that is our *control*.

The geometric position of the system is described by an angle $\theta \in S^1$, the unit circle. The dynamics of the system is obtained from the second Newton law in the curvilinear coordinate $\ell\theta$:

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

where g the acceleration due to gravity.



Although the constants which intervene 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

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

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

$$\begin{cases} \dot{\theta}(t) &= \omega, \\ \dot{\omega}(t) &= -\sin \theta(t) + u(t). \end{cases} \quad (1.10)$$

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

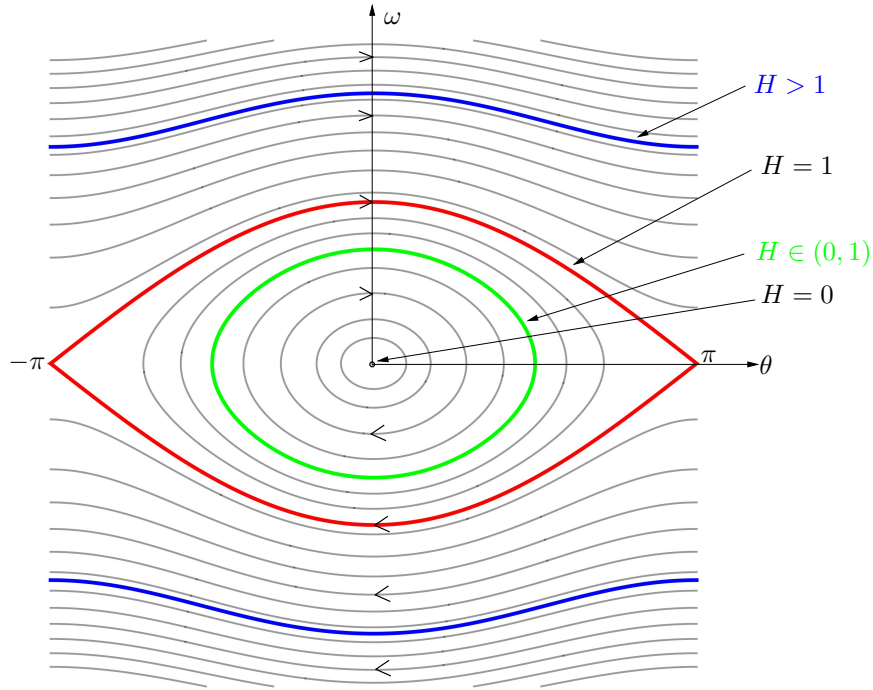
$$F(q, u) = \begin{pmatrix} \omega \\ -\sin(\theta) + u \end{pmatrix}.$$

When $u \equiv 0$, the points $(0,0)$ and $(\pi,0)$ are respectively the point of stable and unstable equilibrium of the pendulum.

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 $u(\cdot) : [0, T] \rightarrow [-u_{\text{max}}, u_{\text{max}}]$ such that the solution of (1.10), 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, its proof is not completely trivial. It will be given in Chapter (3) 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 $H = \frac{1}{2}\omega^2 - \cos(\theta)$)



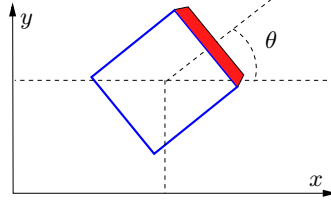
- evolving for time s with the control $u(t) \equiv u_{\text{max}}$ and then for time s with the control $u(t) \equiv -u_{\text{max}}$ is not the same than doing the evolution in opposite order. In this case, we say that the two vector fields $F(q, u_{\text{max}})$ and $F(q, -u_{\text{max}})$ *do not 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 8, however the details will not be given in these notes.

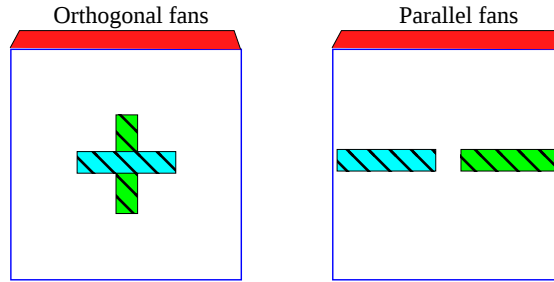
1.0.2 A vehicle sliding on the ice, controlled by two fans

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

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



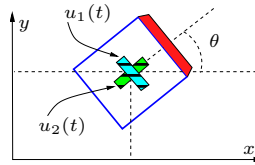
can reverse the direction or rotation) that we will place on the vehicle in two different configurations as indicated in the following figure:



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.



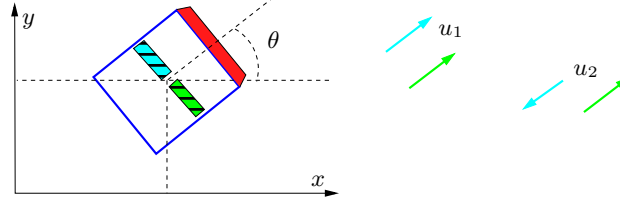
In this configuration, up to constants that we can fix to 1, the dynamics is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = u_1(t)F_1 + u_2(t)F_2, \quad \text{where } F_1 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}, \text{ and } F_2 = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \\ 0 \end{pmatrix}.$$

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 a plane in $\mathbb{R}^2 \times S^1$.

Parallel fans.

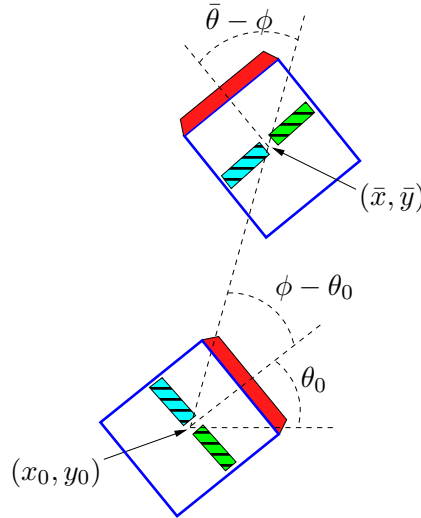


In this configuration, up to constants that we can fix to 1, if $u_1(\cdot)$ is the control making the two fans rotate in the same direction and $u_2(\cdot)$ is the control making the two fans rotating in the opposite direction the dynamics is

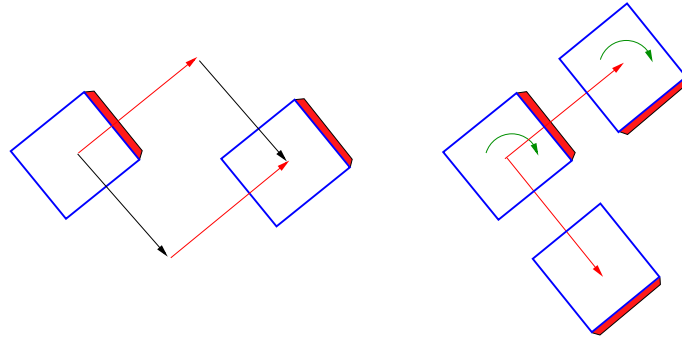
$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = u_1(t)X_1 + u_2(t)X_2, \quad \text{where} \quad X_1 = \begin{pmatrix} \cos \theta \\ \sin \theta \\ 0 \end{pmatrix}, \quad \text{and} \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

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 $\theta_0 \leq \phi \leq \bar{\theta}$):

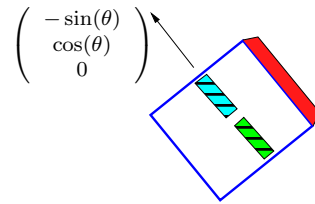
1. We let $u_1 = 0$ and $u_2 = 1$ for a time $t_1 = \phi - \theta_0$.
2. Then, $u_1 = 1$ and $u_2 = 0$ for a time t_2 that equals the (Euclidean) distance between (x_0, y_0) and $(\bar{x}, \bar{y})$.
3. Finally, we let $u_1 = 0$ and $u_2 = 1$ for a time $t_3 = \bar{\theta} - \phi$.



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.



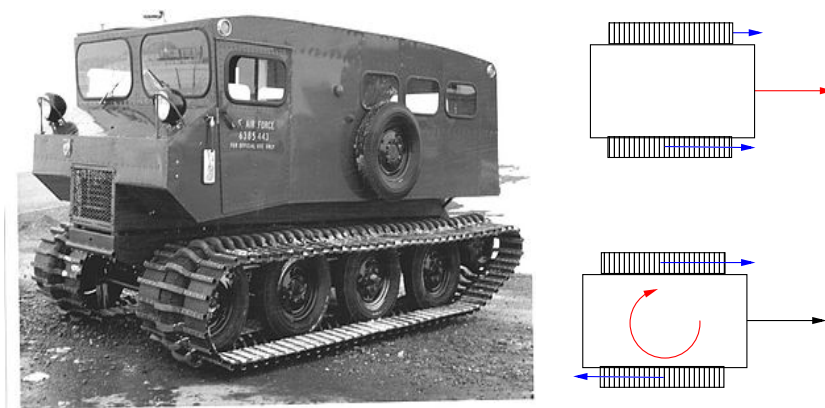
It is exactly the non-commutativity of X_1 and X_2 that permits to approximate directions that are not directly admissible in the control system. In particular, the direction corresponding to a movement orthogonal to the wheels:



The concept of commutativity/non-commutativity of dynamics will be discussed in detail in Chapter 3.

Remark 1.1. Notice that the previous analysis stays the same 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 3, 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.



Chapter 2

Manifolds, Vector fields, Brackets

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 smooth¹ 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 the first chapter of [DoCarmo] or [Arnold] or [Lee]. The reader that is familiar with the concepts of smooth manifold, vector fields (including time-dependent vector fields), flows, etc. can jump directly to the next chapter.

2.1 Smooth differentiable manifolds

Differentiable manifolds are object that locally “look like” $\mathbb{R}^n$ but possibly not globally. Example that one meets very often are

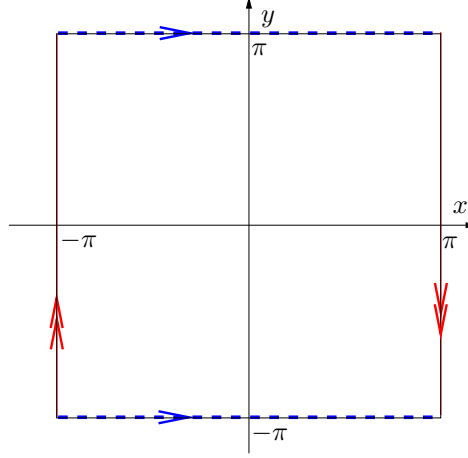
- in dimension $n = 1$: $\mathbb{R}$, the circle S^1 , an open interval $(-\varepsilon, \varepsilon)$;
- in dimension $n = 2$: $\mathbb{R}^2$, a smooth surface in $\mathbb{R}^3$ (e.g., the sphere S^2), the Klein bottle

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

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

Differentiable manifolds can always be seen as the generalization of regular surface embedded in $\mathbb{R}^3$. For instance a manifold of dimension n embedded in $\mathbb{R}^N$, ($N \geq n$) can be

¹i.e., $\mathcal{C}^\infty$



defined as the zero of function $F \in \mathcal{C}^\infty(\mathbb{R}^N, \mathbb{R}^{N-n})$ whose differential is of maximal rank.

$$\{q \in \mathbb{R}^N \mid F(q) = 0, F \in \mathcal{C}^\infty(\mathbb{R}^N, \mathbb{R}^{N-n}), \frac{\partial F}{\partial q} \text{ of maximal rank where } F(q) = 0\}.$$

For instance a circle S^1 can be embedded in $\mathbb{R}^3$ as the set of zeros of the function

$$F(x, y, z) = \begin{pmatrix} x^2 + y^2 - 1 \\ z \end{pmatrix}$$

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 2.1.

Definition 2.1. A smooth differentiable manifold of dimension n is a pair $(M, \{(U_a, \varphi_a)\}_{a \in A})$ where M is a set and for every $a \in A$ we have the following: U_a is a subset of M and φ_a is one to one map to an open subset $\varphi_a(U_a)$ of $\mathbb{R}^n$;

1. $M = \bigcup_{a \in A} U_a$;
2. if for any pairs $a, b \in A$ we have $U_a \cap U_b = W \neq \emptyset$ then:
 - 2.1. $\varphi_a(W)$ and $\varphi_b(W)$ are open sets in $\mathbb{R}^n$;
 - 2.2. the map $\varphi_a \circ \varphi_b^{-1} : \varphi_b(W) \rightarrow \varphi_a(W)$ is $\mathcal{C}^\infty$.
3. the family $\{(U_a, \varphi_a)\}_{a \in A}$ is maximal with respect to the previous conditions

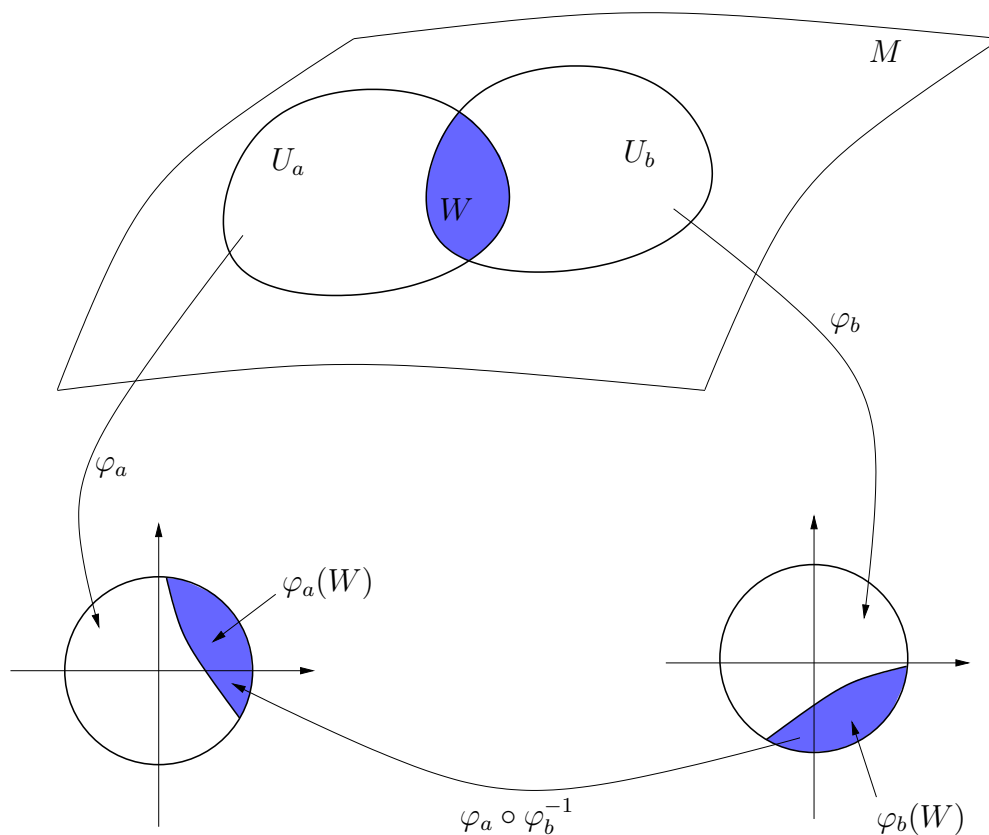


Figure 2.1: Definition of a smooth manifold

Remark 2.2. *Condition 3.* is included since one usually needs the maximal freedom of changing charts. *Condition 2.1* is important since we have not required that M is a topological space. Actually it is easy to verify that it is a consequence of the definition with $W \subset M$ open in M if and only if $\varphi_a(W \cap U_a)$ is open in $\mathbb{R}^n$ for every $a \in A$. To avoid pathological situations we will also make the additional requirements that **i)** different points have disjoint neighborhoods (i.e., that the topological structure given to M is Hausdorff); **ii)** conditions 1. and 2. can be realized with the set A finite or countable (i.e., that the topological structure given to M is second countable);

Often we will speak about system of coordinates $(x^1, \dots, x^n)$ in M , meaning that we are fixing a chart (φ_a, U_a) and we are describing $\varphi_a(U_a) \subset \mathbb{R}^n$ with coordinates $(x^1, \dots, x^n)$.

Since a differentiable manifold can be locally treated as $\mathbb{R}^n$, one could be tempted to assume that it is possible to essentially work as in $\mathbb{R}^n$. However it is important to take into account that there are plenty of properties that changes when changing system of coordinates. For instance a curve that in a system of coordinates is a straight line can become an arc of circle in another system of coordinates. Moreover, the global topology can play an important role. For instance a vector field (see Definition 2.9 below) on a

two dimensional sphere is obliged to have some zeros (thing that is not true on a two dimensional torus or on the plane). See for instance [Arnold].

In general, we will say that a property on a manifold is *intrinsic* if it does not depends on the choice of the system of coordinates. For instance the property of a curve (see Definition 2.14) to be closed is an intrinsic property. The property of a curve to have zero second order coefficient in his Taylor polynomial is not intrinsic.

2.2 Curves, tangent vectors

Definition 2.3. Let M and N be two differentiable manifolds. A map $f : M \rightarrow N$ is continuous (resp. smooth, piecewise smooth, Lipschitz) if it is continuous (resp. smooth, piecewise smooth, Lipschitz) chart by chart.

Definition 2.4. Let M be a smooth differentiable manifold and I an interval of $\mathbb{R}$. A curve on M is a continuous map $\gamma : I \ni t \mapsto M$.

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

Given a smooth curve $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ we can compute²

$$\dot{\gamma}(0) = \left. \frac{d\gamma}{dt} \right|_{t=0} = \lim_{h \rightarrow 0} \frac{\gamma(h) - \gamma(0)}{h} \quad \text{in coordinates.}$$

Obviously the quantity $\lim_{h \rightarrow 0} \frac{\gamma(h) - \gamma(0)}{h}$ has no meaning without fixing coordinates since M has no *intrinsic* vector space structure, and thus we are not allowed to sum points. Actually, it is not difficult to see that the above quantity changes when we change coordinate system.

We are then going to define the tangent vector to a smooth curve as an operator on smooth functions. Let us denote the set of smooth functions from M to $\mathbb{R}$ as $\mathcal{C}^\infty(M)$.

Definition 2.5. Let $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ be a smooth curve. The tangent vector $\dot{\gamma}(0)$ to the curve γ at $t = 0$ is the operator from $\mathcal{C}^\infty(M)$ to $\mathbb{R}$ given by

$$\dot{\gamma}(0)f = \left. \frac{d(f \circ \gamma)}{dt} \right|_{t=0}. \quad (2.1)$$

In coordinates, if q is represented by $x := (x_1, \dots, x_n)$ and γ is represented by $(\gamma_1, \dots, \gamma_n)$, we can compute (2.1) via the chain rule, obtaining³

$$\left. \frac{d(f(\gamma_1(t), \dots, \gamma_n(t)))}{dt} \right|_{t=0} = \sum_{i=1}^n \partial_i f(x) \dot{\gamma}_i(0) = \left(\sum_{i=1}^n \dot{\gamma}_i(0) \partial_i \Big|_x \right) f.$$

²To be precise if (U, φ) is a system of coordinates such that $\gamma(0) \in U$ the expression $\dot{\gamma}(0)$ in coordinates means $\left. \frac{d}{dt}(\gamma \circ \varphi) \right|_{t=0}$. To lighten the notation, we will however often identify geometric object with their expression in coordinates, specifying it if there is possibility of confusion.

³here we use the notation $\partial_i \Big|_x$ instead of ∂_{x_i} or $\frac{\partial}{\partial x_i}$ to stress that the point x is fixed.

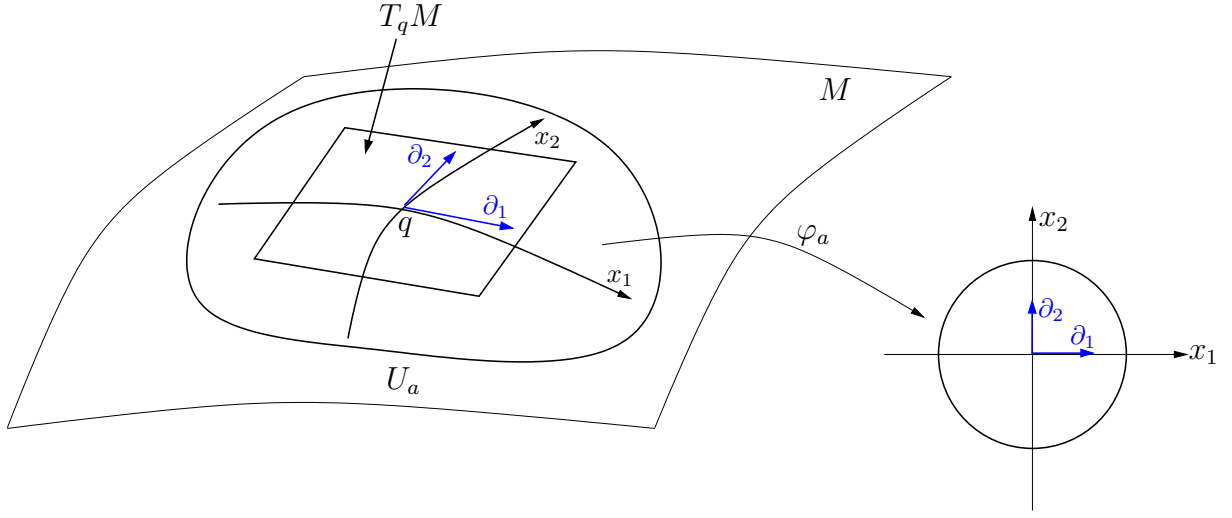


Figure 2.2: Coordinates in the tangent space.

Hence, $\dot{\gamma}(0)$ is represented by the first order operator from $\mathcal{C}^\infty(M) \rightarrow \mathbb{R}$

$$\sum_{i=1}^n \dot{\gamma}_i(0) \partial_i \Big|_x.$$

Definition 2.6. A tangent vector at a point $q \in M$ is the tangent vector at $t = 0$ to some smooth curve $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ 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 $T_q M$.

In coordinates, a general tangent vector $v \in T_q M$ is thus represented by

$$\sum_{i=1}^n v_i \partial_i \Big|_x, \quad (2.2)$$

hence by a n -tuple $(v_1, \dots, v_n) \in \mathbb{R}^n$. Notice that $\partial_i \Big|_x$ is the coordinate representation of the tangent vector at q of the “coordinate curve”

$$\begin{array}{c} x_i(\cdot) : t \mapsto (0, \dots, t, \dots, 0) \\ \uparrow \\ \text{position } i \end{array} \quad (2.3)$$

It then follows immediately that $T_q M$ has the structure of an n -dimensional vector space, whose canonical basis in the system of coordinates $(x_1, \dots, x_n)$ is given by $(\partial_1, \dots, \partial_n) \Big|_x$ (see Figure 2.2).

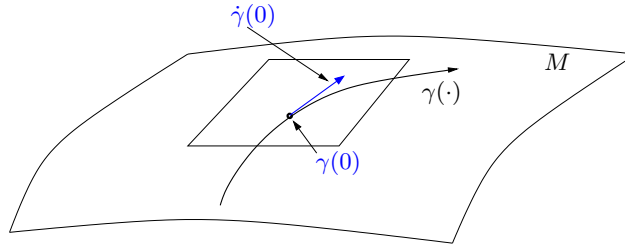


Figure 2.3: Tangent vector to a curve.

Definition 2.7. *The tangent bundle to M is the set of all tangent spaces to M i.e.,*

$$TM = \coprod_{q \in M} T_q M.$$

It is not difficult to prove that TM has the structure of a smooth differentiable manifold of dimension $2n$ with charts on TM induced by those of M .

Remark 2.8. *The point of view of representing a tangent vector to smooth curve as operator on functions is not in contrast with the intuitive representation of a tangent vector to a curve on a surface embedded in $\mathbb{R}^3$ as an “arrow” on the corresponding tangent plane (see Figure (2.3)). 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*

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

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

2.3 Vector fields, flows, rectification

A vector field is a smooth assignation of a tangent vector at each tangent space of a manifold (see Figure (2.4)).

Definition 2.9. *A vector field is a smooth map $X : M \ni q \rightarrow T_q M$. Since, for every $q \in M$, $X(q)$ is an operator from $\mathcal{C}^\infty(M)$ to $\mathbb{R}$, we can then interpret X as an operator from $\mathcal{C}^\infty(M)$ to $\mathcal{C}^\infty(M)$.*

In this definition, “smooth” means that the map $M \ni q \mapsto (q, X(q)) \in TM$ is smooth as map between two differentiable manifolds. One can easily check that that this concept

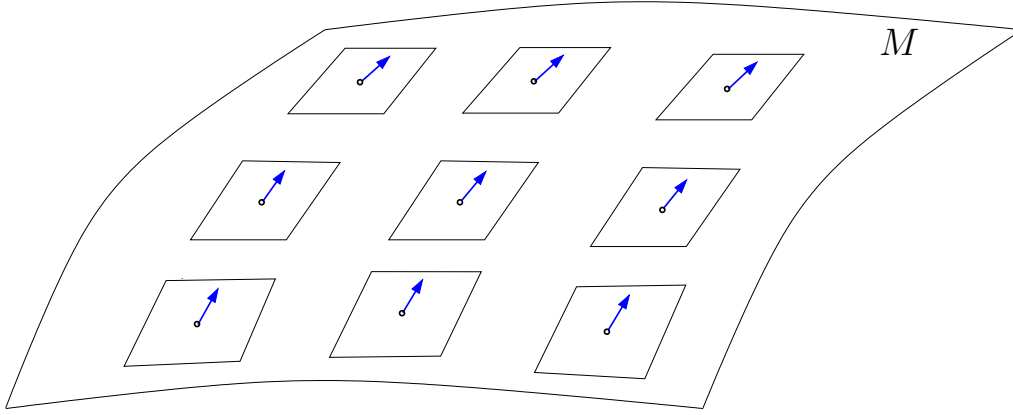


Figure 2.4: A vector field.

of smoothness is equivalent to the fact that for every $f \in \mathcal{C}^\infty(M)$ the function Xf is smooth.

The set of all vector fields on M is indicated with $\text{Vec}(M)$. One immediately verify that $\text{Vec}(M)$ has the structure of vector space.

For a fixed system of coordinates, a vector fields is represented by an expression as (2.2) but with coefficients smoothly depending on the point:

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

Definition 2.10. A first order operator on functions is a smooth map $Y : \mathcal{C}^\infty(M) \rightarrow \mathcal{C}^\infty(M)$ that in a system of coordinates can be written as

$$Y(x) = \sum_{i=1}^n Y_i(x) \partial_{x_i} \quad (2.4)$$

Notice that the property of a map of being a first order operator on functions is invariant by change of coordinates.

Clearly vector fields are in one to one correspondence with first order operators on functions. Actually on one side a vector field in coordinates has exactly an expression of the form (2.4); on the other side a first order operator of functions is a vector fields since it is a smooth map that to x associate a linear combination of the tangent vectors to the coordinate curves.

To a vector field X is associated a dynamical system

$$\dot{q}(t) = X(q(t)). \quad (2.5)$$

A solution $q(\cdot)$ of (2.5) is called an *integral curve* of X . We have the following.

Theorem 2.11 (see for instance [Arnold]). *Let X be a vector field on M and $q_0 \in M$. Then the problem (called Cauchy problem)*

$$\begin{cases} \dot{q}(t) = X(q(t)) \\ q(0) = q_0 \end{cases} \quad (2.6)$$

has local existence and uniqueness of solutions. If we indicate the solution of (2.6) as $q(t; q_0)$ then the map $(t, q_0) \mapsto q(t; q_0)$ is smooth around $(0, q_0)$.

Remark 2.12. *In (2.6) 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 2.13. *The map $\phi_t : M \ni q_0 \mapsto q(t; q_0) \in M$ is called the flow of X .*

Notice that $\phi_0 : M \rightarrow M$ is the identity map. However, in general it may happen that it does not exist $t > 0$ such that the map ϕ_t is defined in the whole manifold. Consider for instance $M = \mathbb{R}$ and the vector field $X(x) = x^2$. The solution of the equation $\dot{x} = x^2$ with $x(0) = x_0$ is

$$\phi_t(x_0) = \frac{x_0}{1 - tx_0}, \quad t < \frac{1}{x_0}.$$

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 $t > 0$.

Definition 2.14. *A vector field X is said to be complete if for every $q_0 \in M$ the solution $q(\cdot; q_0)$ of (2.6) exists for every positive and negative time.*

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:

- vector fields on compact manifolds,
- vector fields on $\mathbb{R}^n$ that are sublinear i.e., for which there exists $c_1, c_2 \geq 0$ such that $\|X(x)\| \leq c_1 + c_2\|x\|$ (here $\|\cdot\|$ is the standard norm in $\mathbb{R}^n$).

We have then the following.

Theorem 2.15 (see for instance [Arnold]). *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 $t, t_1, t_2 \in \mathbb{R}$):

$$\phi_0 = \text{id}, \quad (2.7)$$

$$\phi_{t_2} \circ \phi_{t_1} = \phi_{t_1+t_2}, \quad (2.8)$$

$$\phi_t^{-1} = \phi_{-t}. \quad (2.9)$$

That is, $\{\phi_t\}_{t \in \mathbb{R}}$ is a one parameter subgroup of the group of diffeomorphisms of M . Moreover, by construction we have⁴

$$\frac{\partial \phi_t}{\partial t} = X \circ \phi_t. \quad (2.10)$$

Because properties (2.7)–(2.10) look like those of the exponential function, usually the flow of a vector field X is denoted by e^{tX} , although one should always remember that this is not an exponential (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 used our notation for the flow of a vector field and on the right hand side we used the standard exponential of matrices. Notice that, in general, $e^{tX_1} \circ e^{tX_2}$ is not equal to $e^{t(X_1+X_2)}$ even for linear vector fields on $\mathbb{R}^n$.

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.

Theorem 2.16 (see for instance [Arnold]). *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 $X = \partial_{x_1}$.*

2.3.1 Nonautonomous vector fields

In control theory one needs to work with vector fields that smoothly depends on a control which depends (usually in a L_{loc}^∞ way) on time. Namely, we consider dynamical systems of the form

$$\dot{q} = F(q(t), u(t)) \quad (2.11)$$

where

H. $q \in M$ (a smooth connected manifold of dimension n), U is a subset of $\mathbb{R}^m$, $F : M \times U \ni (q, u) \mapsto F(q, u) \in T_q M$ is smooth and $u(\cdot) : \mathbb{R} \rightarrow U$ is a 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

$$\dot{q}(t) = X_t(q(t)). \quad (2.12)$$

It is then necessary to extend a bit the theory of vector fields to non-autonomous vector fields that are regular in q , but not necessarily in t .

⁴here we mean $\frac{\partial \phi_t(q)}{\partial t} = X(\phi_t(q))$ for every $q \in M$.

Definition 2.17. A nonautonomous vector field is 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) the map $t \mapsto X(q, t)$ is measurable, for every fixed $q \in M$,

(C2) the map $q \mapsto X(q, t)$ is smooth, for every fixed $t \in \mathbb{R}$,

(C3) for every system of coordinates defined in an open set $\Omega \subset M$ and every compact $K \subset \Omega$ and compact interval $I \subset \mathbb{R}$ there exists $C, L > 0$ such that for all $(x, t), (y, t) \in K \times I$

$$\|X(x, t)\| \leq C, \quad \|X(x, t) - X(y, t)\| \leq L\|x - y\|.$$

Remark 2.18. For non-autonomous vector fields of the form (2.11) under hypothesis **H.**, one verifies easily that (C1)-(C3) are satisfied.

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

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

$$\begin{cases} \dot{q}(t) = X_t(q(t)) & (\text{for almost every } t) \\ q(t_0) = q_0 \end{cases} \quad (2.13)$$

has local existence and uniqueness of solutions. If we denote the solution of (2.13) by $q(t; t_0, q_0)$, then 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 (2.13) is defined for every positive and negative time.

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 (2.7)–(2.10) becomes

$$\begin{aligned} \phi_{t_0, t_0} &= \text{id}, & \forall t_0 \in \mathbb{R} \\ \phi_{t_2, t_3} \circ \phi_{t_1, t_2} &= \phi_{t_1, t_3}, & \forall t_1, t_2, t_3 \in \mathbb{R}, \\ (\phi_{t_1, t_2})^{-1} &= \phi_{t_2, t_1}, & \forall t_1, t_2 \in \mathbb{R}, \\ \frac{\partial \phi_{t_0, t}}{\partial t} &= X_t \circ \phi_{t_0, t}, & \forall t_0, t \in \mathbb{R}, \end{aligned}$$

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 Brackets

Definition 2.20. Let X and Y be two vector fields on M . Their Lie brackets is the operator from $\mathcal{C}^\infty(M)$ to $\mathcal{C}^\infty(M)$ defined as

$$[X, Y] = X \circ Y - Y \circ X$$

Notice that in coordinates

$$\begin{aligned} [X, Y]f &= \left(\sum_{i=1}^n X_i(x) \partial_{x_i} \right) \left(\sum_{j=1}^n Y_j(x) \partial_{x_j} f \right) - \left(\sum_{i=1}^n Y_i(x) \partial_{x_i} \right) \left(\sum_{j=1}^n X_j(x) \partial_{x_j} f \right) \\ &= \sum_{i,j=1}^n \left(X_i(x) \partial_{x_i} Y_j(x) - Y_i(x) \partial_{x_i} X_j(x) \right) \partial_{x_j} f \end{aligned} \quad (2.14)$$

In matrix notation

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

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

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

Similar expressions hold for Y .

From the expression (2.14) 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.

One immediately verifies the following properties of the Lie bracket:

- bilinearity: for every $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$\begin{aligned} [X, \lambda_1 Y_1 + \lambda_2 Y_2] &= \lambda_1 [X, Y_1] + \lambda_2 [X, Y_2], \\ [\lambda_1 X_1 + \lambda_2 X_2, Y] &= \lambda_1 [X_1, Y] + \lambda_2 [X_2, Y]; \end{aligned}$$

- antisymmetry: $[Y, X] = -[X, Y]$;
- Jacobi identity: $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$.

A vector space V endowed with an operation $V \times V \rightarrow V$ that is bilinear, antisymmetric and satisfying the Jacobi identity is said to be a *Lie algebra*. It follows that $(\text{Vec}(M), [\cdot, \cdot])$ is a Lie algebra.

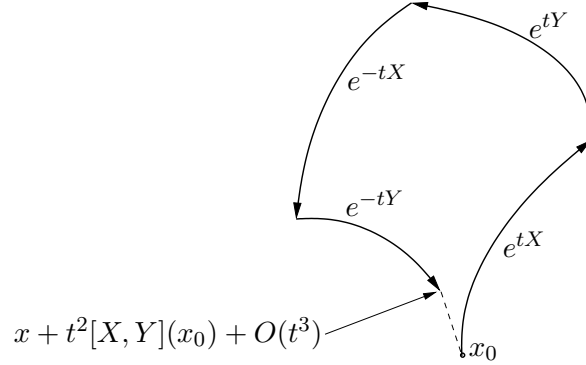


Figure 2.5: Geometric meaning of the Lie bracket.

Remark 2.21. As a consequence of the antisymmetry 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 . However, if at q_0 we have $X(q_0) = Y(q_0) = 0$, it follows from the definition that $[X, Y](q_0) = 0$.

Example 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 a measure of the lack of commutation of the flows associated with X and Y . We refer to Figure 2.5.

Lemma 2.22. For every $x \in M$ we have

$$e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(x) = x + t^2[X, Y](x) + O(t^3) \quad (\text{in coordinates}) \quad (2.15)$$

for t that tends to zero.

Notice that in this expression make sense only in coordinates, since in a manifold we do not have the right of summing points

Proof. It is enough to compute for each flow the Taylor expansion at order 3. We have

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

and

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

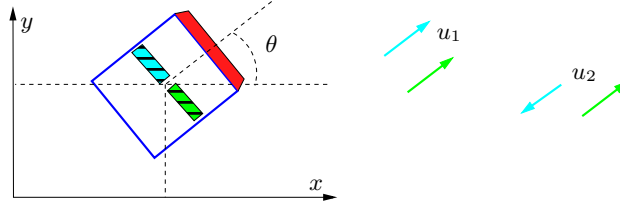
Then

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

At the next step the result follows. □

Remark 2.23. *The previous lemma says in particular that if $[X, Y](q) \notin \text{Vect}(X(q), Y(q))$, then it is possible, by alternating between the dynamics of X and Y , to attain points that cannot be reached with the flow of linear combinations of X and Y (cf. the example in Section 1.0.2). This is the starting idea behind the conditions for controllability that we are going to study in the next chapter.*

For instance for the vehicle driven by two fan in the parallel configuration



The dynamics is

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

The Lie bracket between X_1 and X_2 is

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

i.e. it is the non admissible direction that we can approximate.

Notice that to be able to approximate $[X, Y]$ one needs to switch among X , Y , $-X$, $-Y$.

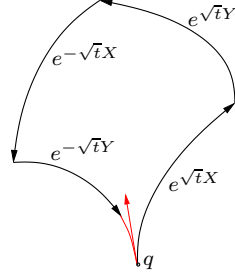
An important corollary of the previous lemma is the following.

Corollary 2.24. *Let X and Y be two complete 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 $q \in M$.*

Remark 2.25. *From Formula (2.15) it follows that an alternative definition of Lie brackets is*

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

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

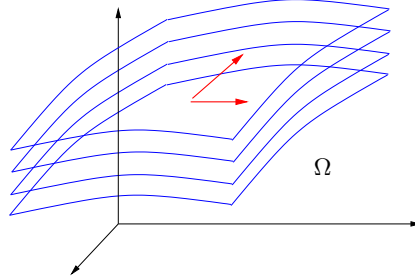


2.5 The Frobenius Theorem

Let M be a smooth manifold of dimension n

Let $\mathcal{F} = \{X_1, \dots, X_h\}$, ($h < n$) be a family of vector fields defined on an open set Ω that are linearly independent at each point.

Definition 2.26. An immersed submanifold⁵ H of Ω of dimension h is said to be an integral submanifold of $\mathcal{F}$ if $T_q H = \text{Span}\{X_1(q), \dots, X_h(q)\}$ for every $q \in H$. $\mathcal{F}$ is said to be completely integrable if through each point of M there exists an integral manifold of $\mathcal{F}$



Definition 2.27. $\mathcal{F}$ is said to be involutive if $[X_i, X_j](q) = \sum_k c_{ij}^k(q) X_k(q)$ for every $q \in \Omega$

Theorem 2.28. $\mathcal{F}$ is completely integrable iff it is involutive

In control theory we usually work with distributions that are *maximally non-integrable*.

⁵ $H \subset M$ is an immersed submanifold if the immersion map $H \ni q \mapsto q \in M$ is an immersion i.e., $\text{rank}(D_q i)$ is maximal $\forall q \in H$

Chapter 3

Nonlinear controllability

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

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

Here 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 $u(\cdot) : [0, \infty[\rightarrow U$ 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 smooth¹ function of its arguments and that $u(\cdot)$ is regular enough in such a way that equation (3.1) with the initial condition $q(0) = q_{\text{in}} \in M$ has local existence and uniqueness of solutions. For instance we can assume that $u(\cdot)$ is a L^1_{loc} or a L^∞_{loc} function of the time. 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. *For simplicity of notation in the following we also assume that for every $u(\cdot)$ a solution of (3.1) exists in $[0, \infty[$. However this hypothesis is not necessary for the validity of the theorems that we are going to prove.*

Denote by $q(t; q_{\text{in}}, u(\cdot))$ the solution at time t of (3.1) starting from q_{in} at $t = 0$ and corresponding to a control function $u(\cdot)$. We recall the definitions of the *reachable* (or *attainable*) sets starting from q_{in} :

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

$$\mathcal{R}(\tau, q_{\text{in}}) = \{\bar{q} \in M \mid \exists u(\cdot) : [0, \tau] \rightarrow U, q(\tau; q_{\text{in}}, u(\cdot)) = \bar{q}\};$$

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

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

¹i.e., $\mathcal{C}^\infty$

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

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

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 3.2. *The system (3.1) is said to be*

- *controllable if for every $q_{\text{in}} \in M$, $\mathcal{R}(q_{\text{in}}) = M$;*
- *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 q_{in} belongs to the interior of $\mathcal{R}(\leq \tau, q_{\text{in}})$ for every $\tau > 0$.*

3.1 Control systems as families of vector fields

In the following it will be useful to think to the system (3.1) as a *family of vector fields* parameterized by $u \in U$ (i.e., by constant controls). In other words we will often consider instead of the control system (3.1), the family of vector fields

$$\mathcal{F} = \{F(\cdot, v) \mid v \in U\}.$$

All vector fields of the family $\mathcal{F}$ are considered smooth and 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 $] -\infty, \infty[$.

Definition 3.3. *Let $\mathcal{F}$ be a family of vector fields. We call $\text{Lie}(\mathcal{F})$ the smallest sub-algebra of $\text{Vec}(M)$ containing $\mathcal{F}$. Namely, $\text{Lie}(\mathcal{F})$ is the span of all vector fields of $\mathcal{F}$ and of their iterated Lie brackets of any order:*

$$\text{Lie}(\mathcal{F}) = \text{span}\{f_1, [f_1, f_2], [f_1, [f_2, f_3]], [f_1, [f_2, [f_3, f_4]]], \dots \mid f_1, f_2, \dots \in \mathcal{F}\}.$$

Definition 3.4. *We say that the family $\mathcal{F}$ is Lie bracket generating at a point q if the dimension of $\text{Lie}_q(\mathcal{F}) := \{f(q) \mid f \in \text{Lie}(\mathcal{F})\}$ 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 3.5. *Notice that in general $\text{Lie}(\mathcal{F})$ is an infinite-dimensional space, while $\text{Lie}_q(\mathcal{F})$ is a subspace of M .*

Exercise 1. Let $\mathcal{F} = \{f_1, \dots, f_m\}$ and let A be an invertible $m \times m$ matrix. Define $f'_i = \sum_j A_{ij} f_j$ and $\mathcal{F}' = \{f'_1, \dots, f'_m\}$. Prove that $\mathcal{F}$ is Lie bracket generating if and only if $\mathcal{F}'$ is.

3.1.1 Affine control systems

An affine control system is a system of the form

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i(t) f_i(q), \quad (3.2)$$

where $f_0, f_1, \dots, f_m$ belong to $\text{Vec}(M)$ and $u(\cdot) = (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty[\rightarrow U \subset \mathbb{R}^m$ is the control. In an affine control system it is also assumed that U contains a neighborhood of the origin in $\mathbb{R}^m$. The vector field f_0 is called *drift*.

Exercise 2. Let $\mathcal{F}$ be the family of vector fields associated with (3.2).

- 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.

3.2 The 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

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + u(t) \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

where $u(\cdot) : [0, \infty[\rightarrow [-1, 1]$. It is Lie bracket generating (since the corresponding family $\mathcal{F}$ contains the vector fields $\{(1, 1), (1, -1)\}$, 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 2.23). The Lie bracket generating condition permits to say that a system is *locally accessible* in the following sense.

Theorem 3.6 (Krener). *If $\mathcal{F}$ is Lie bracket generating at q_{in} , then for every $\tau > 0$, q_{in} belongs to the closure of the interior of $\mathcal{R}(\leq \tau, q_{\text{in}})$.*

The conclusion of the Krener Theorem can be equivalently reformulated in the following way:

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

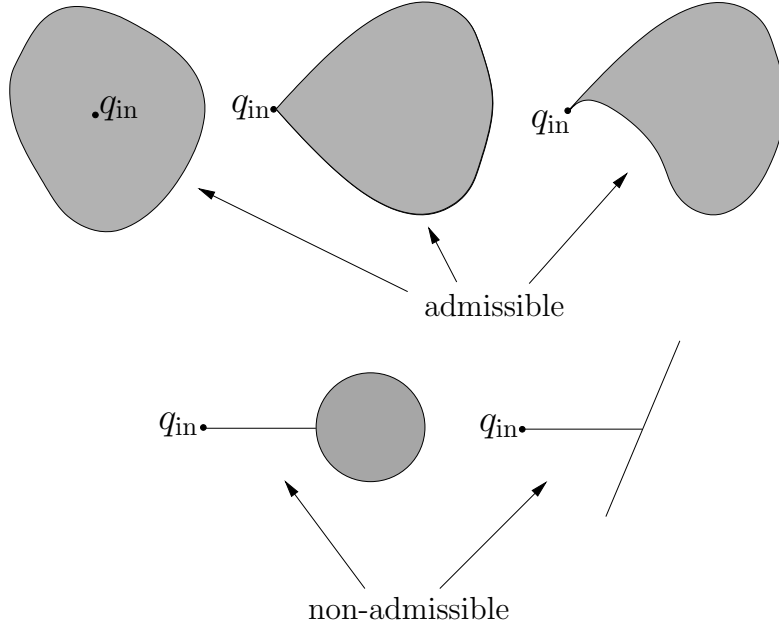


Figure 3.1: Admissible and non-admissible reachable sets when the system is Lie bracket generating at q_{in} .

Krener's theorem says in particular that the 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 3.1 shows what one can expect/non-expect from $\mathcal{R}(\leq \tau, q_{\text{in}})$, $\tau > 0$.

Proof of Theorem 3.6. First notice that if $\mathcal{F}$ is Lie bracket generating at a point, then the same property holds true in a neighborhood of that point. (If n vector fields are linearly independent at a point, then they are linearly independent in a neighborhood of that point.)

There exists $f \in \mathcal{F}$ such that $f(q_{\text{in}}) \neq 0$, otherwise $\text{Lie}_{q_{\text{in}}}(\mathcal{F}) = \{0\}$. If $n = 1$ the conclusion follows.

If $n > 1$ and all vector fields of $\mathcal{F}$ are tangent to the curve $t \mapsto e^{tf}(q_{\text{in}})$, $0 < t < \varepsilon$, then from Lemma 2.22 it follows that $\text{Lie}_{e^{tf}(q_{\text{in}})}(\mathcal{F})$ is also tangent to that curve and hence its dimension is less than 2. This contradicts the Lie bracket generating assumption. As a consequence, there exist $g \in \mathcal{F}$ and $0 < \bar{t} < \varepsilon$ such that f and g are linearly independent in a neighborhood of $q_1 = e^{\bar{t}f}(q_{\text{in}})$ (see Figure 3.2). Hence $(t, s) \mapsto e^{sg} \circ e^{tf}(q_{\text{in}})$, $0 < s < \varepsilon'$, $\bar{t} - \varepsilon' < t < \bar{t} + \varepsilon'$ has as image a surface of dimension 2. If $n = 2$ the conclusion follows.

Otherwise we iterate the same argument and we conclude by recurrence on n . $\square$

Remark 3.7. Notice that for this proof we have only used piecewise constant controls. In the following we will denote by $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ the set of points reachable from q_{in} using piecewise

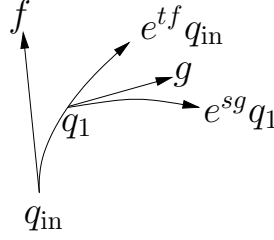


Figure 3.2: Proof of Krener's theorem.

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

Remark 3.8. From the proof of the Krener 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 $e^{t_n f_n} \circ \dots \circ e^{t_1 f_1} q_{\text{in}}$ where $t_1, \dots, t_n > 0$ and $f_1, \dots, f_n \in \mathcal{F}$, i.e., by trajectories corresponding to piecewise constant controls made by n pieces. Notice that the vector fields $f_1, \dots, f_n$ 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 $f_1 = f$, $f_2 = g$, $f_3 = f$.

3.3 Symmetric systems

Definition 3.9. A family of vector fields $\mathcal{F}$ is said to be symmetric if $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$.

When the family $\mathcal{F}$ is Lie bracket generating and symmetric one obtain that the system is controllable. This is the conclusion of the celebrated Chow–Rashevskii theorem.

Theorem 3.10 (Chow–Rashevskii). *If $\mathcal{F}$ is Lie bracket generating and symmetric, then for every $q_{\text{in}} \in M$ we have $\mathcal{R}(q_{\text{in}}) = M$.*

Proof. Step 1. Fix $q_{\text{in}} \in M$ and let us show that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ contains a neighborhood of q_{in} . Since $\mathcal{F}$ is Lie bracket generating, $\mathcal{R}(q_{\text{in}})$ contains a nonempty open set Ω whose points can be reached by trajectories corresponding to piecewise controls made by n pieces. Fix $t_1, \dots, t_n > 0$ and $f_1, \dots, f_n \in \mathcal{F}$ such that $\bar{q}_{\text{in}} := e^{t_n f_n} \circ \dots \circ e^{t_1 f_1} q_{\text{in}} \in \Omega$. Since $\mathcal{F}$ is symmetric, $-f_1, \dots, -f_n \in \mathcal{F}$ and we have that

$$e^{-t_1 f_1} \circ \dots \circ e^{-t_n f_n}(\Omega) \subset \mathcal{R}^{\mathcal{F}}(q_{\text{in}}).$$

In particular $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ contains the set

$$V = e^{-t_1 f_1} \circ \dots \circ e^{-t_n f_n}(\Omega).$$

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

Step 2. Let us show that $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is open. If $\bar{q} \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$, then $\mathcal{R}^{\mathcal{F}}(\bar{q}) \subset \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. It follows that $\text{int}(\mathcal{R}^{\mathcal{F}}(\bar{q})) \subset \text{int}(\mathcal{R}^{\mathcal{F}}(q_{\text{in}}))$. But from **Step 1** we have that $\bar{q} \in \text{int}(\mathcal{R}^{\mathcal{F}}(\bar{q}))$. Hence $\bar{q} \in \text{int}(\mathcal{R}^{\mathcal{F}}(q_{\text{in}}))$.

Step 3. From the fact that $\mathcal{F}$ is symmetric, it follows that $\bar{q} \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ if and only if $q_{\text{in}} \in \mathcal{R}^{\mathcal{F}}(\bar{q})$. Let us consider the equivalence classes $M/\sim$ where $\sim$ is the equivalence relation

$$\bar{q} \sim q_{\text{in}} \text{ if and only if } \bar{q} \in \mathcal{R}^{\mathcal{F}}(q_{\text{in}}).$$

Such equivalence classes are open and disjoint. Since M is connected, it follows that there is only one class. Hence, for every q_{in} we have $\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = \mathcal{R}(q_{\text{in}}) = M$. $\square$

Remark 3.11. Notice that in **Step 1** of the proof of Chow–Rashevskii theorem, since the times can be rendered arbitrarily small, we have proved that the system is small-time locally controllable in a neighborhood of 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 3.12. Notice that the Chow–Rashevskii theorem can be used in more general situations than those fixed by the hypotheses stated here. For instance

- to get controllability it is sufficient that the family $\mathcal{F}$ contains 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 .

3.4 Compatible vector fields

When a family of vector fields is Lie bracket generating but is not symmetric, in general it is not easy to understand if the system is controllable or not. A technique to study the controllability is the one of *compatible vector fields*.

Definition 3.13. A vector field g is said to be 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 $\mathcal{R}(q_{\text{in}})$.

A family of vector fields $\mathcal{G}$ is said to be compatible with $\mathcal{F}$ if every element of $\mathcal{G}$ is compatible with $\mathcal{F}$.

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

Theorem 3.14. If $\mathcal{F}$ is a Lie bracket generating family of vector fields, $\mathcal{G}$ is compatible with $\mathcal{F}$, and $\mathcal{F} \cup \mathcal{G}$ is controllable, then $\mathcal{F}$ is controllable as well.

This theorem should be used in the following way: one collects a family $\mathcal{G}$ of vector fields g that added to the family $\mathcal{F}$ do not change the closure of the reachable set and such that it is easy to prove the controllability of the family $\mathcal{F} \cup \mathcal{G}$.

The main ingredient to prove Theorem 3.14 is the following corollary of the Krener theorem.

Corollary 3.15. *If $\mathcal{F}$ is Lie bracket generating and $\mathcal{R}(q_{\text{in}})$ is dense in M for some q_{in} , then $\mathcal{R}(q_{\text{in}}) = M$.*

Proof. Consider the system

$$\dot{q} = -F(q, u(t)), \quad q \in M, \quad u(\cdot) : [0, \infty[\rightarrow U \subset \mathbb{R}^m, \quad (3.3)$$

which is obtained from (3.1) by reversing the time. Let $\mathcal{F}^-$ be the family of vector fields associated with (3.3). Since $\mathcal{F}$ is Lie bracket generating, then $\mathcal{F}^-$ is Lie bracket generating as well.

Let $\bar{q} \in M$ and $\mathcal{R}^-(\bar{q})$ be the reachable set for (3.3) starting from $\bar{q}$. Thanks to Krener's theorem, $\mathcal{R}^-(\bar{q})$ contains a nonempty open set. In particular it has nonempty intersection with $\mathcal{R}(q_{\text{in}})$ (being $\mathcal{R}(q_{\text{in}})$ dense). See Figure 3.3.

This means that $\bar{q} \in \mathcal{R}(q_{\text{in}})$. Indeed from q_{in} one can reach a point $q_1 \in \mathcal{R}(q_{\text{in}}) \cap \mathcal{R}^-(\bar{q})$ (since $q_1 \in \mathcal{R}(q_{\text{in}})$) and from q_1 one can reach $\bar{q}$ (since q_1 is reachable from $\bar{q}$ for the system with reverted time).

Being $\bar{q}$ arbitrary we have that $\mathcal{R}(q_{\text{in}}) = M$. □

Proof of Theorem 3.14. First notice that, since $\mathcal{G}$ is compatible with $\mathcal{F}$, then $\mathcal{R}^{\mathcal{F} \cup \mathcal{G}}(q_{\text{in}})$ is contained in the closure of $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$. Indeed, every point of $\mathcal{R}^{\mathcal{F} \cup \mathcal{G}}(q_{\text{in}})$ is of the type $e^{t_k g_k} \circ e^{t_1 g_1}(q_{\text{in}})$ with $g_1, \dots, g_k \in \mathcal{F} \cup \mathcal{G}$ and $t_1, \dots, t_k \geq 0$. For $k = 1$ the conclusion follows by definition of compatible vector fields, while the case $k > 1$ follows by an easy induction argument.

Hence $\mathcal{R}^{\mathcal{F}}(q_{\text{in}})$ is dense. Since the system is Lie bracket generating, the conclusion follows from Corollary 3.15. □

Next we present some important applications of the technique based on compatible vector fields.

3.4.1 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 of $\frac{f+g}{2}$ starting from the same point.



The lemma can be deduced from the following averaging result, taking piecewise constant controls u^n with values in the canonical basis of $\mathbb{R}^k$, switching faster and faster as $n \rightarrow \infty$.

$$\dot{q} = \sum_{i=1}^k u_i(t) f_i(q), \quad q(0) = q_0,$$
$$\int_0^t u^n(s)ds \xrightarrow{n \rightarrow \infty} \int_0^t u(s)ds \quad \text{for all } t \in [0, T]. \quad (3.4)$$

Proof. Given a compact coordinate chart Ω around q_0 and denoting by $L > 0$ a constant such that f_i is L -Lipschitz continuous on Ω for every $i = 1, \dots, k$, we have that, in

coordinates,

$$\|q(t; q_0, u^n) - q_0\| \leq LKt,$$

where $K = \sup_{n \in \mathbb{N}} \|u^n\|_{L^\infty[0, T]}$.

Up to splitting the curve $t \mapsto q(t; q_0, u)$ in a finite number of pieces, it is enough to prove the proposition in the case where $t \mapsto q(t; q_0, u)$ and each $t \mapsto q(t; q_0, u^n)$, $n \in \mathbb{N}$, stay in Ω .

Then $q(t) := q(t; q_0, u)$ and, for n large enough, $q^n(t) := q(t; q_0, u^n)$, can be written in coordinates as

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

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

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

Integrating by parts and thanks to (3.4), we have that, for every $i = 1, \dots, k$,

$$\int_0^t \left(\frac{d}{ds} \int_0^s (u_i^n(\tau) - u_i(\tau)) d\tau \right) f(q(s)) ds \xrightarrow{n \rightarrow \infty} 0 \quad \text{uniformly w.r.t. } t \in [0, T].$$

(Notice that the convergence in (3.4) 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, denoting $\xi^n(t) = \|q^n(t) - q(t)\|$, we have

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

Define then $c = mKL$ and $w^n(t) = e^{-ct} \int_0^t \xi^n(s) ds$. Then

$$\frac{d}{dt} w^n(t) \leq e^{-ct} \varepsilon,$$

yielding, by integration, that $w^n(t) \leq \frac{1-e^{-ct}}{c} \varepsilon$, hence

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

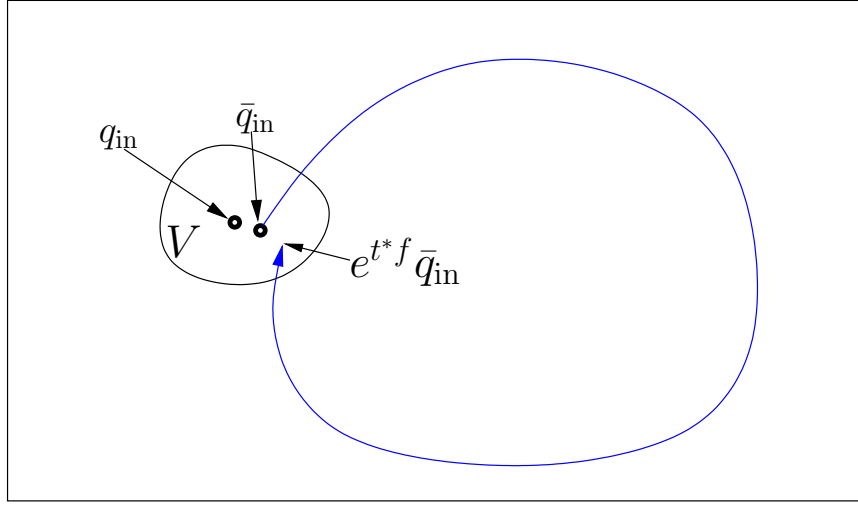


Figure 3.4: Definition of recurrent vector field.

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

From Lemma 3.16, using the theory of compatible vector fields, one get immediately a useful corollary of Theorem 3.10.

Corollary 3.18. *If $\mathcal{F}$ is Lie bracket generating and its convex hull in the space $\text{Vec}(M)$ is symmetric, then for every $q_{\text{in}} \in M$ we have $\mathcal{R}^{\mathcal{F}}(q_{\text{in}}) = M$.*

Other applications of Lemma 3.16 are given later.

3.4.2 Affine systems with recurrent drift

In this section we apply the theory of compatible vector fields to affine control systems (cf. Section 3.1.1) that are Lie bracket generating and having a drift f_0 which is recurrent.

We refer to Figure 3.4.

Definition 3.19 (Recurrent vector field). *A vector field f is said to be recurrent if for every point $q_{\text{in}} \in M$, every neighborhood V of q_{in} and every time $t > 0$, there exist $\bar{q}_{\text{in}} \in V$ and $t^* > t$ such that $e^{t^* f}(\bar{q}_{\text{in}}) \in V$.*

Notice that if the trajectories of f are periodic (possibly with period depending on the trajectory), then f is recurrent.

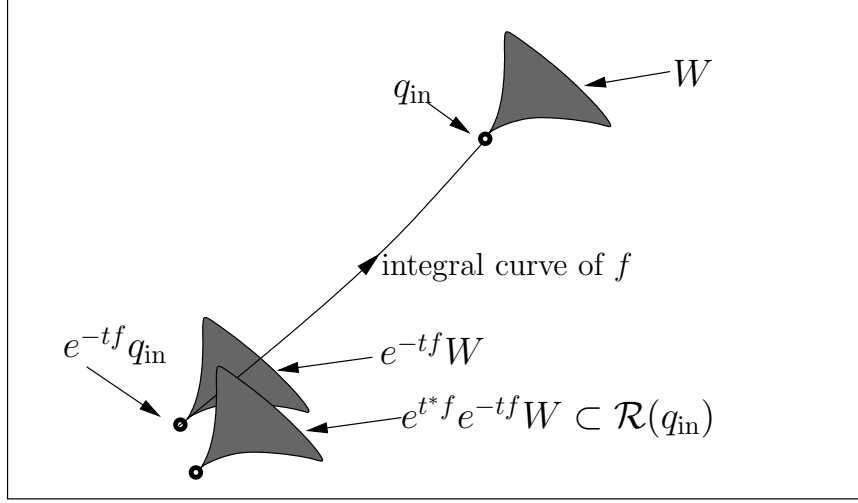


Figure 3.5: Proof of Lemma 3.20.

Lemma 3.20. *Let $\mathcal{F}$ be a Lie bracket generating family of vector fields and $f \in \mathcal{F}$. If 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 3.5. 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 $q_{\text{in}} \in \overline{W}$.

Now since e^{-tf} is a diffeomorphism this implies that $e^{-tf}q_{\text{in}} \in \overline{e^{-tf}W}$.

Since f is recurrent, there exists $t^* > t$ 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 follows 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 $e^{-tf}q_{\text{in}} \in \overline{\mathcal{R}(q_{\text{in}})}$. $\square$

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

Theorem 3.21. *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i(t) f_i(q), \quad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty[\rightarrow U \subset \mathbf{R}^m. \quad (3.6)$$

Assume that (i) 0 belongs to the interior of U , (ii) the control system (3.6) is Lie bracket generating, (iii) f_0 is recurrent. Then the system is controllable.

Proof. Let $\mathcal{F} = \{f_0 + \sum_{i=1}^m f_i \mid (u_1, \dots, u_m) \in U\}$ be the family of vector fields associated to 3.6. Notice that $f_0 \in \mathcal{F}$ since 0 belongs to U . Lemma 3.20 and Theorem 3.14 state the equivalence between the controllability of $\mathcal{F}$ and $\hat{\mathcal{F}} := \mathcal{F} \cup \{-f_0\}$. The family $\hat{\mathcal{F}}$ is not symmetric in general, but its convexification contains a symmetric set (see Figure ??). The conclusion follows from Corollary 3.18 (cf. also Remark 3.12). $\square$

By using Lemma 3.16 in the proof of Theorem 3.21 it is clear that hypothesis **i)** can be weakened. More precisely one gets

Corollary 3.22. *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i f_i(q), \quad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty[\rightarrow U \subset \mathbf{R}^m. \quad (3.8)$$

Assume that (i) 0 belongs to the interior of the convex hull of U , (ii) the control system (3.8) is Lie bracket generating, (iii) f_0 is recurrent. Then the system is controllable.

Example. Consider the control system on the sphere $S^2 = \{x \in \mathbf{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\}$ given by

$$\dot{x} = f_0(x) + u f_1(x), \quad u(\cdot) : [0, \infty[\rightarrow (-1, 1), \quad x \in S^2,$$

where

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

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

This system is controllable since

- the Lie bracket between f_0 and f_1 is given by

$$[f_0, f_1](x) = \begin{pmatrix} -x_3 \\ 0 \\ x_1 \end{pmatrix},$$

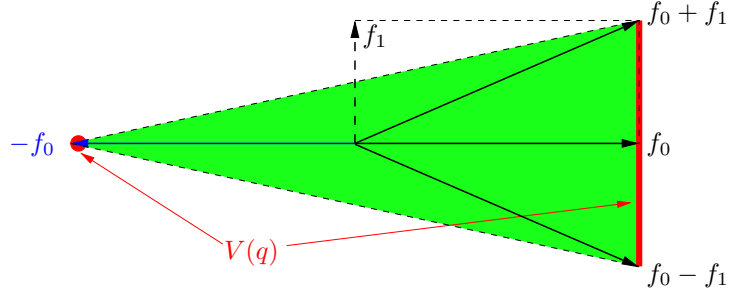


Figure 3.6: For a control system of the type $f_0 + u(t)f_1$ with $u(\cdot)$ taking values in $[-1, 1]$, the picture represents the convexification of the set $V(q) = \{f_0(q) + u f_1(q) \mid u \in [-1, 1]\} \cup \{-f_0(q)\}$ at a given point q . Notice that although $V(q)$ does not contain a symmetric set, its convexification does. For more general control systems as those of Corollary 3.6 the situation is exactly the same.

and hence the system is Lie bracket generating (for every $x \in S^2$, $\dim(\text{span}\{f_0, f_1, f_2\}(x)) = 2$);

- the trajectories of f_0 are periodic and hence f_0 is recurrent.

3.4.3 Affine systems with non-recurrent drift

When the drift is not recurrent one can still obtain that the system 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. More precisely we have the following.

Proposition 3.23 (Strong bracket generating). *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i f_i(q), \quad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty[\rightarrow \mathbb{R}^m. \quad (3.10)$$

If $\{f_1, \dots, f_m\}$ is Lie bracket generating then (3.10) is controllable.

Remark 3.24. For an affine control system as (3.10), the condition that $\{f_1, \dots, f_m\}$ is Lie bracket generating is called the strong bracket generating condition.

Proof. First notice that as a consequence of Exercise 2, being $\{f_1, \dots, f_m\}$ Lie bracket generating, then (3.10) is Lie bracket generating as well. Let $\mathcal{F}$ be the family of vector fields associated with (3.10). 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 (3.10) follows, since the family $\mathcal{F} \cup \{\sum_{i=1}^m v_i f_i, \quad v_i \in \mathbb{R}, i = 1, \dots, m\}$ contains a symmetric and Lie bracket generating sub-family (Remark 3.12).

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

$$\sum_{i=1}^m v_i f_i = \lim_{n \rightarrow \infty} \frac{1}{n} \left(f_0 + \sum_{i=1}^m (nv_i) f_i \right) \quad (3.11)$$

and that $\frac{1}{n}(f_0 + \sum_{i=1}^m (nv_i) f_i)$ is compatible with $\mathcal{F}$ since a trajectory of (3.10) corresponding to controls $\tilde{u}_i(\cdot)$, $i = 1, \dots, m$, is a time-reparameterisation of a trajectory of

$$\dot{q} = \frac{1}{n} \left(f_0(q) + \sum_{i=1}^m u_i f_i(q) \right)$$

corresponding to controls $n\tilde{u}_i(\cdot)$, $i = 1, \dots, m$.

The conclusion follows from the fact that if a vector field is the uniform limit on all compacts of a sequence of compatible vector fields, then it is compatible as well (from the continuity of solutions of ODEs with respect to the vector field). $\square$

The use of Lemma 3.16 permits to weaken a bit the hypothesis that the controls take values in $\mathbb{R}^m$. More precisely one gets.

Corollary 3.25. *Consider the control system*

$$\dot{q} = f_0(q) + \sum_{i=1}^m u_i f_i(q), \quad (u_1(\cdot), \dots, u_m(\cdot)) : [0, \infty[\rightarrow U \subset \mathbb{R}^m. \quad (3.13)$$

If the convex hull of U is $\mathbb{R}^m$ and if $\{f_1, \dots, f_m\}$ is Lie bracket generating then (3.13) is controllable.

3.5 Orbits and necessary conditions for controllability

We have seen in the previous sections several sufficient conditions for the controllability of a nonlinear system. In this section we discuss some necessary conditions that are consequence of a very deep theorem of geometric nature, the so-called *orbit theorem*. This theorem permits to conclude that, beside pathological cases, $\text{Lie}_q(\mathcal{F})$ measures precisely the dimension of the set of directions that can be used starting from a point.

We define the *orbit* of the family $\mathcal{F}$ starting from a point $q_{\text{in}} \in M$ as the set

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

Remark 3.26. $\mathcal{O}(q_{\text{in}})$ can be interpreted as the reachable set, starting from q_{in} , of the family $-\mathcal{F} \cup \mathcal{F}$, using only piecewise constant controls.

We have the following result

Theorem 3.27 (Orbit theorem). *For every $q_{\text{in}} \in M$, the set $\mathcal{O}(q_{\text{in}})$ has the structure of an immersed sub-manifold of M . In particular it has the same dimension at every point. Moreover if $q \in \mathcal{O}(q_{\text{in}})$ then $\text{Lie}_q(\mathcal{F}) \subseteq T_q\mathcal{O}(q_{\text{in}})$. The two spaces $\text{Lie}_q(\mathcal{F})$ and $T_q\mathcal{O}(q_{\text{in}})$ coincide if one of the following two conditions is verified:*

- *every element of $\mathcal{F}$ is an analytic vector field;*
- *the dimension of $\text{Lie}_q(\mathcal{F})$ is constant with respect to $q \in \mathcal{O}(q_{\text{in}})$.*

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

Lemma 3.28. *For every $q_{\text{in}} \in \mathbb{R}$ we have that $\mathcal{R}(q_{\text{in}}) \subseteq \mathcal{O}(q_{\text{in}})$.*

This result wouldn't be so obvious without the orbit theorem since $\mathcal{R}(q_{\text{in}})$ is the set of points that one can reach using L^∞ controls (and not only piecewise constant ones).

The following corollary gives some consequence of the Orbit theorem on the controllability of nonlinear systems.

Corollary 3.29. *If $\mathcal{F}$ is not Lie bracket generating and either every vector field of $\mathcal{F}$ is analytic or the dimension of $\text{Lie}_q(\mathcal{F})$ is constant with respect to $q \in \mathcal{O}(q_{\text{in}})$, then $\mathcal{F}$ is not controllable.*

Chapter 4

Introduction to Optimal Control

Consider the nonlinear control system

$$\dot{q} = f(q, u(t)). \quad (4.1)$$

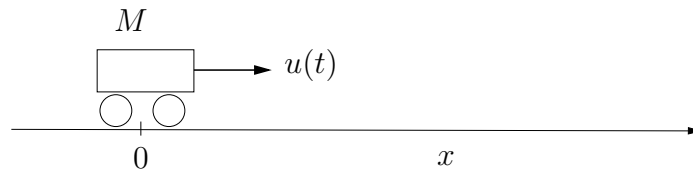
Here $q \in M$ (a smooth connected manifold of dimension n , e.g., $\mathbb{R}^n$) is the state of the system, $U \subset \mathbb{R}^m$ is the set of control values and $u(\cdot) : [0, \infty[\rightarrow U$ is the control. We assume that f is a smooth¹ function of its arguments and that $u(\cdot)$ is regular enough in such a way that equation (4.1) with the initial condition $q(0) = q_{\text{in}} \in M$ has local existence and uniqueness of solutions (e.g., it 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 3).

A typical problem that one meets in control theory is to find the “best” trajectory for a given criterium going from one initial point q_{in} to a final “target” $\mathcal{T} \subset M$. When $\mathcal{T} = \{q_{\text{f}}\}$, 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 1 Consider a cart of mass M 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.

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \frac{u}{M}, \quad |u| \leq F_{\text{max}} \end{aligned}$$

Find the trajectory joining $(x_1, x_2) = (0, 0)$ to $(x_1, x_2) = (a, 0)$ in minimum time. Notice that for this problem, initial and final points are fixed, but the final time is free. The criterium can be written in integral form since $T = \int_0^T 1 \, dt$.



¹i.e., $\mathcal{C}^\infty$

Example 2 Consider the robot $E = M6$:

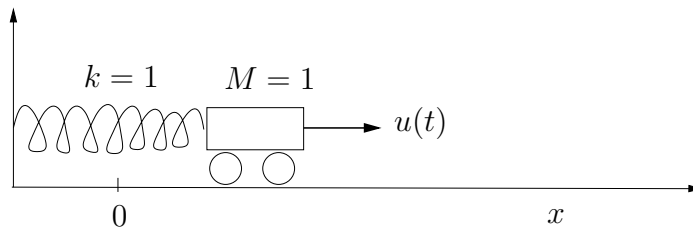
$$\begin{aligned}\dot{x} &= u_1(t) \cos(\theta) \\ \dot{y} &= u_1(t) \sin(\theta) \\ \dot{\theta} &= u_2(t), \quad u_1, u_2 \in L^\infty([0, T], \mathbb{R})\end{aligned}\tag{4.2}$$

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 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 $|u(t)| \leq 1$. The corresponding dynamics has the form $\ddot{x} = -x + u(t)$. Setting $x_1 = x$ and $x_2 = \dot{x}$ the control system becomes,

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 + u(t), \quad |u(t)| \leq 1\end{aligned}$$

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

Optimal Control Problem (OCP)

$$\begin{aligned}\dot{q}(t) &= f(q(t), u(t)), \\ q(0) &= q_{\text{in}}, \quad q(T) \in \mathcal{T}, \\ \int_0^T f^0(q(t), u(t)) dt + \phi(q(T)) &\rightarrow \min, \\ \text{here } T > 0 &\text{ can be fixed or free.}\end{aligned}$$

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,
- $u(\cdot) \in L^\infty([0, T], U)$,
- $q(\cdot) : [0, T] \rightarrow M$, 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. Often in applications one cannot guarantee that the following condition is 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 take $\mathcal{T} = M$ and as ϕ something that represent the distance between the point where one arrives and the desired target.

4.1 The different steps to solve an optimal control problem

The steps to determine a solution to the minimization problem OCP are similar to finding the minimum of a smooth function $f^0 : \mathbb{R} \rightarrow \mathbb{R}$.

0. **Find conditions which guarantee the existence of solutions.** We recall that among smooth functions $f^0 : \mathbb{R} \rightarrow \mathbb{R}$, 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 Sec. 4.2, we present the Filippov test.
1. **Apply first-order necessary conditions.** For a smooth function $f^0 : \mathbb{R} \rightarrow \mathbb{R}$, this means that if $\bar{x}$ is a minimum then $\frac{d}{dx} f^0(\bar{x}) = 0$. This condition gives candidates

for minima, i.e., identifies 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 $x \mapsto (x^2 + 1/2)e^{-x^2}$. 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 $x \rightarrow \pm\infty$, the function tends to zero). 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 Sec. 5.1. 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 Sec. 4.3.1.

2. **Apply second-order conditions.** For instance, for a smooth function $f^0 : \mathbb{R} \rightarrow \mathbb{R}$, 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 [Agr-Sa,Bo-Pi,Sc-Le]). 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.

4.2 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 fix the ideas, 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

$$\begin{aligned} \dot{\hat{q}}(t) &= \begin{pmatrix} \dot{q}^0(t) \\ \dot{q}(t) \end{pmatrix} = \begin{pmatrix} f^0(q(t), u(t)) \\ f(q(t), u(t)) \end{pmatrix} =: \hat{f}(q(t), u(t)), \\ \hat{q}(0) &= (0, q_{\text{in}}), \quad \hat{q}(T) \in \mathbb{R} \times \mathcal{T}. \end{aligned}$$

This control system is called the *augmented system*. The minimization problem in integral form, $\min \int_0^T f^0(q(t), u(t)) dt$, becomes a problem of minimization of one of the coordinates at the final time, i.e., $\min q^0(T)$.

We denote by $\hat{\mathcal{R}}^T(0, q_{\text{in}})$ the reachable set in time T for the augmented system, starting from $(0, q_{\text{in}})$. The key observation on which optimal control is based is expressed by the following proposition.

Proposition 4.0. *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 $\hat{q}(T) = (q^0(T), q(T)) \in \text{int} \hat{\mathcal{R}}^T(0, q_{\text{in}})$ 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 Fig. 4.1. □

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 4.1. *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.*

Hence the compactness of $\hat{\mathcal{R}}^T(0, q_{\text{in}})$ is a key point. A sufficient condition for compactness of the reachable set is given by the following theorem (see, e.g., [?]).

Theorem 4.2 (Filippov). *Fix $T > 0$. Consider the control system $\dot{q}(t) = f(q(t), u(t))$, $q \in M$, $u(\cdot) \in L^\infty([0, T], U)$, where M is a n -dimensional smooth connected manifold and $U \subset \mathbb{R}^m$. Fix an initial condition $q_{\text{in}} \in M$. Assume the following conditions:*

- *the set U is compact,*
- *the set $\mathbf{F}(q) = \{f(q, u) \mid u \in U\}$ is convex for every $q \in M$,*

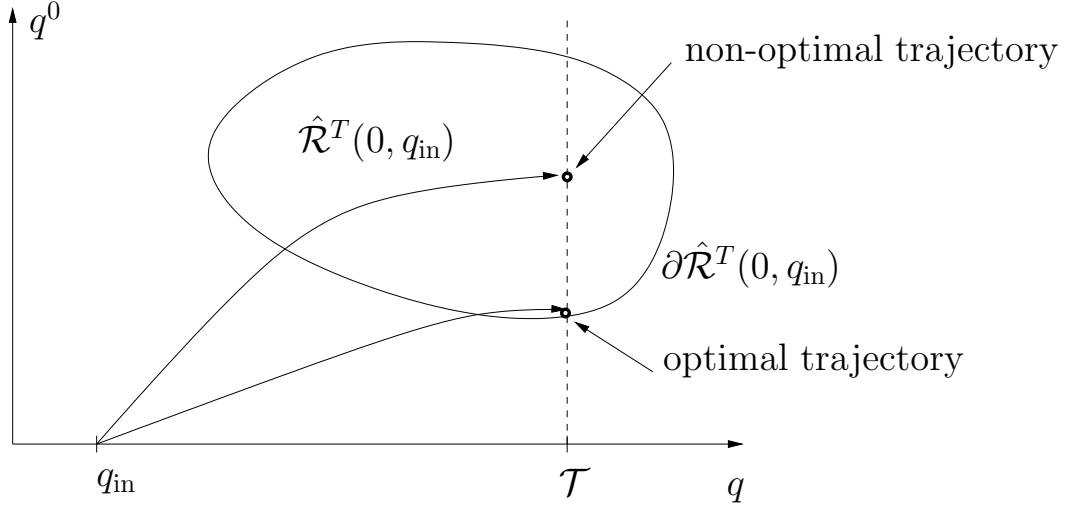


Figure 4.1: The reachable set of the augmented system

- 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.

Here

$$\begin{aligned} \mathcal{R}^{\leq T}(q_{\text{in}}) &:= \{\bar{q} \in M \mid \exists T' \in [0, T] \\ &\text{and an admissible trajectory } q : [0, T'] \rightarrow M \\ &\text{such that } q(0) = q_{\text{in}}, \quad q(T') = \bar{q}\}. \end{aligned}$$

Remark 1. 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. Take $M = \mathbb{R}^2$, $U = \{1, 2\}$, and $f(q, u) = A_u q$ with $A_1 = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$ and $A_2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}$. Notice that the corresponding control system satisfies all the assumptions of Th. 4.2 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. On the other hand, if $\dot{q}(t) = f(q(t), u(t))$ at time t and $q(t)$ has a nonzero vertical component, then $\frac{d}{dt} \|q(t)\| > -\|q(t)\|$. Since every trajectory of the control system starting at q_{in} necessarily leaves the horizontal axis, its final point at time T has norm larger than e^{-T} . Consequently $q_T \notin \mathcal{R}^T(q_{\text{in}})$. This proves that $\mathcal{R}^T(q_{\text{in}})$ is not closed, hence the conclusion of Th. 4.2 does not hold.

Note that the third hypothesis of Th. 4.2 is automatically satisfied when M is compact. By applying this theorem to the augmented system for problem OCP, one obtains:

Proposition 4.3. Fix $T > 0$. 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 $\hat{\mathbf{F}}(q) = \left\{ \begin{pmatrix} f^0(q, u) \\ f(q, u) \end{pmatrix} \mid u \in U \right\}$ is convex for every $q \in M$,
- 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 (as for instance in the general formulation given in Sec. 5.1), 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 5.1 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 the 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 4.4. *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 $\mathbf{F}(q) = \{f(q, u) \mid u \in U\}$ is convex for every $q \in M$.

Then there exists a solution to the problem.

Example. 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 $\hbar = 1$)

$$i\dot{\psi}(t) = \left(H_0 + \sum_{j=1}^m u_j(t) H_j \right) \psi(t), \quad (4.3)$$

where ψ , the wave function, belongs to the unit sphere in $\mathbb{C}^N$ and $H_0, \dots, H_m$ are $N \times N$ Hermitian matrices. The control parameters $u_j(t) \in \mathbb{R}$ are the components of the control $u(\cdot)$. This control problem has the form $\dot{q} = f(q, u(t))$ where $M = S^{2N-1} \subset \mathbb{C}^N$, $q = \psi$, 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_{fin} in minimum time with $U = [-1, 1]^m$. We assume the controllability of the system. Since U and $M = S^{2N-1}$ 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. In this case, note that the convexity of U leads to the one of the set of admissible velocities at each point.

4.3 First order conditions

For a smooth real-valued function of one variable $f^0 : \mathbb{R} \rightarrow \mathbb{R}$, 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, if $J(u(\cdot))$ is the value of the cost for a reference admissible control $u(\cdot)$ (for instance $J(u(\cdot)) = \int_0^T f^0(q(t), u(t)) dt + d(\mathcal{T}, q(T))$), and $v(\cdot)$ is another admissible control, one would like to consider a condition of the form

$$\left. \frac{\partial J(u(\cdot) + hv(\cdot))}{\partial h} \right|_{h=0} = 0. \quad (4.4)$$

But difficulties may arise for the following reasons.

We work in an infinite-dimensional space (the space of controls) and hence condition (4.4) should be required for infinitely many $v(\cdot)$. 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.

Moreover 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

$$\tilde{q}(T) \in \mathcal{T}, \quad (4.5)$$

with T either free or constrained to be the fixed final time depending on the problem under study. Condition (4.5) should be considered as a *constraint* for the minimization problem, which results in the use 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 explained in Sec. 4.3.1.

4.3.1 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 $x = (x_1, \dots, x_n)$, under the constraint $f(x) = 0$, with the method of Lagrange multipliers. Here f^0 and f are two smooth functions $\mathbb{R}^n \rightarrow \mathbb{R}$. We have two cases.

- 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 4.2.

More precisely, this means that

$$\exists \lambda \in \mathbb{R} \text{ such that } \nabla f^0(\bar{x}) = \lambda \nabla f(\bar{x}). \quad (4.6)$$

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 $x_n = g(x_1, \dots, x_{n-1})$. The requirement that

$$\begin{aligned} \partial_{x_i} f(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) &\equiv 0, \\ i &= 1, \dots, n-1, \\ \partial_{x_i} f^0(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1}))|_{x=\bar{x}} &= 0, \\ i &= 1, \dots, n-1, \end{aligned}$$

provides immediately condition (4.6) with $\lambda = \frac{\partial_{x_n} f^0(\bar{x})}{\partial_{x_n} f(\bar{x})}$.

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.

- 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 rewritten in the following form.

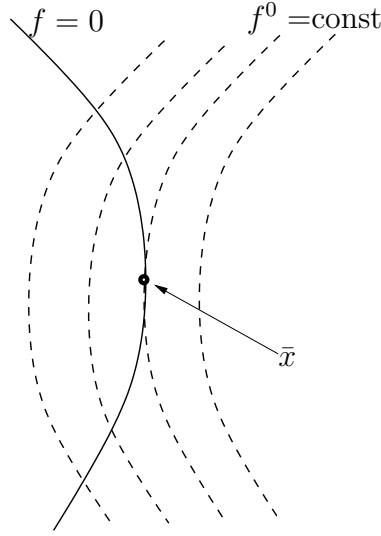


Figure 4.2: Intersection of the set $f(x) = 0$ (solid line) with the level set of f^0 (dashed line). The two gradients $\nabla f(x)$ and $\nabla f^0(x)$ are parallel at $x = \bar{x}$.

Theorem 4.5 (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*

$$\nabla_x \Lambda(\bar{x}, \bar{\lambda}, \bar{\lambda}_0) = 0, \quad \nabla_\lambda \Lambda(\bar{x}, \bar{\lambda}, \bar{\lambda}_0) = 0. \quad (4.7)$$

To show that this statement is equivalent to what we just discussed, we observe that the second equality in (4.7) 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. (4.6) with the change of notation $\lambda \rightarrow \bar{\lambda}$. 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. (4.7) 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.

Chapter 5

The Pontryagin Maximum Principle (PMP): Formulation and Use

5.1 Statement of the Pontryagin Maximum Principle

In this chapter, we state the first-order necessary conditions for optimal control problems, namely the Pontryagin Maximum Principle (PMP for short).

The basic idea is to define a new object (the pre-Hamiltonian, see Eq. (6.6) below) which allows to formulate the Langrange multiplier conditions in a simple and direct way.

Notation Given a smooth manifold M we indicate its tangent space at a point $q \in M$ as $T_q M$. The set of all tangent spaces is called the tangent bundle and it is indicated as TM .

A covector is a linear $p : T_q M \rightarrow \mathbb{R}$. The set of all covectors acting on vectors in $T_q M$ is indicated as $T_q^* M$ (the cotangent space at $q \in M$). The set of all cotangent spaces is called the cotangent bundle and it is indicated as $T^* M$.

The action of $p \in T_q^* M$, on $v \in T_q M$ is indicated with the symbol (called the *duality product*)

$$\langle p, v \rangle.$$

When $M = \mathbb{R}^n$ or we are working in coordinates, we write

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad p = (p_1, \dots, p_n), \quad \langle p, v \rangle = p \cdot v = p_1 v_1 + \dots + p_n v_n.$$

Theorem 5.1. *Consider the optimal control problem*

$$\begin{cases} \dot{q}(t) = f(q(t), u(t)), \\ q(0) = q_{\text{in}}, \quad q(T) \in \mathcal{T}, \\ \int_0^T f^0(q(t), u(t)) dt + \phi(q(T)) \longrightarrow \min, \end{cases} \quad (5.1)$$

where

Chapter 5 – The Pontryagin Maximum Principle (PMP): Formulation and Use

- 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,
- $u(\cdot) \in L^\infty([0, T], U)$,
- $q : [0, T] \rightarrow M$ belongs to the set of Lipschitz continuous curves.

Define the function (called pre-Hamiltonian)

$$\mathcal{H}(q, p, u, p^0) = \langle p, f(q, u) \rangle + p^0 f^0(q, u), \quad (5.2)$$

with

$$(q, p, u, p^0) \in T^*M \times U \times \mathbb{R}.$$

If the pair $(q, u) : [0, T] \rightarrow M \times U$ is optimal, 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) $\dot{q}(t) = \frac{\partial \mathcal{H}}{\partial p}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for q);
- ii) $\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial q}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for p);
- iii) the quantity $\mathcal{H}_M(q(t), p(t), p^0) := \max_{v \in U} \mathcal{H}(q(t), p(t), v, p^0)$ is well-defined and

$$\mathcal{H}(q(t), p(t), u(t), p^0) = \mathcal{H}_M(q(t), p(t), p^0)$$

which corresponds to the maximization condition.

Moreover,

- iv) there exists a constant $c \geq 0$ such that $\mathcal{H}_M(q(t), p(t), p^0) = c$ on $[0, T]$, with $c = 0$ if the final time is free (value of the Hamiltonian);
- v) for every $v \in T_{q(T)}\mathcal{T}$, we have $\langle p(T), v \rangle = p^0 \langle d\phi(q(T)), v \rangle$ (transversality condition), where $d\phi$ is the differential of the function ϕ .

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 Th. 4.5 (with the change of notation and $p \rightarrow \lambda$).
- 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 Sec. 4.2).
- Since the equation for $p(\cdot)$ at point **ii)** of 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 **v)** simplifies to:

$$\langle p(T), T_{q(T)}\mathcal{T} \rangle = 0. \quad (5.3)$$

When the final point is fixed ($\mathcal{T} = \{q_{\text{fin}}\}$), $T_{q(T)}\mathcal{T}$ is a zero-dimensional manifold and hence condition (5.3) 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 $p^0 \neq 0$.

5.2 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 ($p^0 = 0$). In the first 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 **iii)** 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 give m equations for m unknowns. When the maximization condition

permits to express u as a function of q and p , we say that the control is *regular*, otherwise the control is said to be *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)** and **ii)**:

$$\begin{cases} \dot{q}(t) = \frac{\partial \mathcal{H}}{\partial p}(q(t), p(t), w(q(t), p(t)), p^0) \\ \dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial q}(q(t), p(t), w(q(t), p(t)), p^0). \end{cases} \quad (5.4)$$

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. (5.4) 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

$$q(t; p_{\text{in}}, p^0), \quad p(t; p_{\text{in}}, p^0). \quad (5.5)$$

We stress that when $w(\cdot, \cdot)$ is not regular enough, solutions to the Cauchy problem (5.4) 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

$$q(T; p_{\text{in}}, p^0) \in \mathcal{T}. \quad (5.6)$$

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 **iv)** in the PMP. If $\mathcal{T}$ is a k -dimensional submanifold of M ($k \leq n$) then Eq. (5.6) provides only $n - k$ equations and the remaining ones correspond to the transversality condition **v)** of the PMP.

Step 4. If Eq. (5.6) (together with the transversality condition and condition **iv)** 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. (5.6) 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. (5.6). 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 [Bo-Pi20] for a discussion.

5.2.1 Extremal and Optimal syntheses

Often in application 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 (5.1) 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 $\mathcal{H}_M = 0$:

$$\Gamma = \{q(\cdot; p_{\text{in}}, p^0) \mid p_{\text{in}} \in T_{q_{\text{in}}}^* M, p^0 \in \{0, 1/2\}, \mathcal{H}_M(q_{\text{in}}, p_{\text{in}}, p^0) = 0\},$$

is called an *extremal synthesis*. Notice that for a trajectory to be in Γ is not required to satisfy **v)** 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*.

$$\Gamma_{\text{opt}} = \{q_{q_{\text{fi}}}(\cdot) \mid q_{q_{\text{fi}}} \in M \text{ and } q_{q_{\text{fi}}}(\cdot) \text{ is a solution of (5.1) with final condition } q(T) = q_{\text{fi}}\}.$$

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.

Chapter 6

Proof of the PMP for Control Affine Systems with Quadratic Cost

In this chapter we give the proof of the Pontryagin Maximum Principle for a very important class of nonlinear optimal control problems, namely *control affine systems with quadratic cost*.

Problem OCP-AQ

$$\dot{q} = F_0(q) + \sum_{i=1}^m u_i(t) F_i(q), \quad (6.3)$$

$$\begin{aligned} q(0) &= q_{\text{in}}, \quad q(T) = q_{\text{fi}}, \\ \int_0^T \sum_{i=1}^m u_i(t)^2 dt &\rightarrow \min \end{aligned} \quad (6.4)$$

Here

- $T > 0$ is fixed,
- M is a smooth n -dimensional connected manifold,
- F_i , $i = 0, 1, \dots, m$ are smooth vector fields,
- $u(\cdot) := (u_1(\cdot), \dots, u_m(\cdot)) \in L^\infty([0, T], \mathbb{R}^m)$,
- $q(\cdot) : [0, T] \rightarrow M$, belongs to the set of Lipschitz curves.

This problem is important since the control affine form (6.3) models most of the systems that one can find in applications and a cost of the type (6.4) represents the *energy* given by the controls to the systems.

In OCP-AQ it is very important that controls take values in $\mathbb{R}$ 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..

Definition 6.0. When $F_0 = 0$ the problem OCP-AQ is called sub-Riemannian.

We will study in more detail the sub-Riemannian problem in Chapter (7). 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.

6.1 Statement of the PMP for control affine systems with quadratic cost

We prove the following.

Theorem 6.1 (Pontryagin Maximum Principle for **OCP-AQ**). *Define the function (called pre-Hamiltonian)*

$$\mathcal{H}(q, p, u, p^0) = \langle p, F_0(q) + \sum_{i=1}^m u_i F_i(q) \rangle + p^0 \sum_{i=1}^m u_i^2, \quad (6.6)$$

with

$$(q, p, u, p^0) \in T^*M \times \mathbb{R}^m \times \mathbb{R}.$$

If the pair $(q, u) : [0, T] \rightarrow M \times \mathbb{R}^m$ 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 $p^0 \leq 0$ is a constant and such that for almost every (a.e.) $t \in [0, T]$ we have

- i) $\dot{q}(t) = \frac{\partial \mathcal{H}}{\partial p}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for q);
- ii) $\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial q}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for p);
- iii) $\frac{\partial \mathcal{H}}{\partial u}(x(t), p(t), p^0, u(t)) = 0$ (maximization condition);
- iv) there exists a constant $c \geq 0$ such that $\mathcal{H}(q(t), p(t), u(t), p^0) = c$ on $[0, T]$,

Theorem 6.1 is exactly Theorem 5.1 written for the problem OCP-AQ. Actually conditions i) and ii) of the two theorems are identical. The maximization condition iii) of Theorem 5.1 applied to an Hamiltonian quadratic in u with negative coefficient in front of the quadratic part coincide with iii) of Theorem 6.1. Condition iv) of Theorem 5.1 is the same as condition iv) of Theorem 6.1 for T fixed. Condition v) is empty since $\mathcal{T}$ coincide with a point.

6.2 Proof of the Pontryagin Maximum Principle

To avoid difficulties we will however assume that the control system (6.3) 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 (6.3) with $q(0) = q_{\text{in}}$ exists in $[0, T]$.

The idea of the proof is the following. We add a new variable $q^0(t) = \int_0^t \sum_{i=1}^m u_i(s)^2 ds$ and we consider the dynamics for $\hat{q} = (q^0, q)^T$:

$$\dot{\hat{q}}(t) = \begin{pmatrix} \dot{q}^0(t) \\ \dot{q}(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m u_i(t)^2 \\ F_0(q(t)) + \sum_{i=1}^m u_i(t) F_i(q(t)) \end{pmatrix}. \quad (6.7)$$

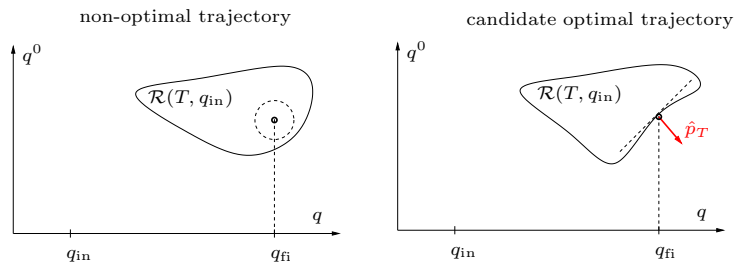
We define the *end-point mapping* as the map that to a control associates the end point of the corresponding trajectory starting from q_{in} :

$$L^\infty([0, T], \mathbb{R}^m) \ni u(\cdot) \xrightarrow{\text{End}} \begin{pmatrix} q^0(T) \\ q(T) \end{pmatrix} \in \mathbb{R} \times M.$$

Assume for a moment that End is Fréchet differentiable. On an optimal control $\tilde{u}(\cdot)$ this map cannot be a surjection (i.e. its differential cannot be invertible). Otherwise 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 (see the following picture and cf. Section s-existence). Then it should exist a covector $\hat{p}_T := (p^0, p_T) \neq 0$ which annihilate $\text{Im}(D_{\tilde{u}(\cdot)}(\text{End}))$.

$$\langle \hat{p}_T, \text{Im}(D_{\tilde{u}(\cdot)}(\text{End})) \rangle = 0 \quad (6.8)$$

This is the main idea that permit to prove Theorem 6.1.



To avoid the difficulties of making the differential of a map on an infinite dimensional space, in the following we will make only variations in a finite dimensional set of controls. Moreover the condition (6.8) will be “transported back” the initial point. This permits to get an equation for the covector.

6.2.1 Notation

In the following, for simplicity of notation, we assume that the control system is complete. If it is not complete, the proof can be made similarly. We use the following notation.

- We set $f_u(q) := F_0(q) + \sum_{i=1}^m u_i F_i(q)$
- We call $\tilde{u}(\cdot)$ the optimal control and $\tilde{q}(\cdot)$ the optimal trajectory starting from q_{in} .
- We call ϕ_t the *optimal flow* i.e. the flow associated to the differential equation $\dot{q} = f_{\tilde{u}(t)}(q)$. Notice that
 -) $\tilde{q}(t) = \phi_t(q_{\text{in}})$, but in general $\phi_t(\bar{q})$ is not an optimal trajectory for $\bar{q} \neq q_{\text{in}}$.
 -) $\frac{d}{dt}\phi_t = f_{\tilde{u}(t)} \circ \phi_t$.
 -) Since the control system is complete, ϕ_t is a diffeomorphism.
- Let $\Phi : M \rightarrow M$ be a diffeomorphism (i.e. a smooth math which is invertible together with its inverse). 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 define the adjoint A^* of A as the linear operator acting on $T_q^* M$ defined by

$$\langle A^* p, v \rangle = \langle p, Av \rangle \text{ for every } v \in T_q M.$$

In coordinates using our convention of representing vectors as column vectors and covectors as row vectors (cf. Section 5.1) 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.¹

The *codifferential* of Φ indicated as $\Phi^*(q)$ or $\Phi^*|_q$ at a point $q \in M$ is the map

$$T_{\Phi(q)}^* M \ni p \mapsto \Phi^*(q)p \in T_q^* M,$$

¹if 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).

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

$$\langle \Phi^{*-1}(q)p, \Phi_*(q)v \rangle_{\Phi(q)} = \langle p, v \rangle_q, \quad v \in T_q M, \quad p \in T_q^* M.$$

Notice that when $M = \mathbb{R}^n$ or we are working in coordinates then with our convention

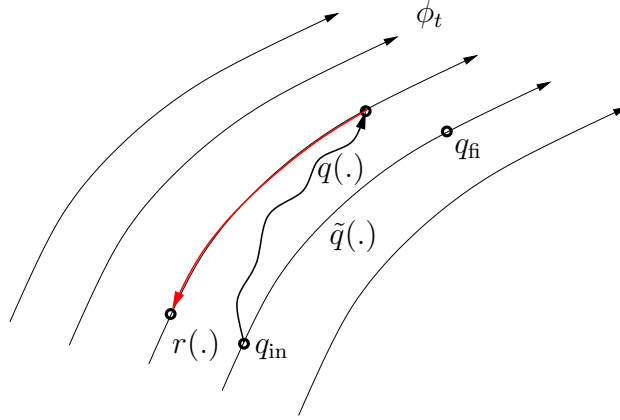
$$\begin{aligned} \Phi(x) &= \begin{pmatrix} \Phi_1(x) \\ \vdots \\ \Phi_n(x) \end{pmatrix}, \\ \Phi_*(x) &= \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \cdots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \cdots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix}, \\ \Phi^*(x) &= \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \cdots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \cdots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix} \text{ acting on the left,} \\ \Phi^{*-1}(x) &= \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1} & \cdots & \frac{\partial \Phi_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial \Phi_n}{\partial x_1} & \cdots & \frac{\partial \Phi_n}{\partial x_n} \end{pmatrix}^{-1} \text{ acting on the left.} \end{aligned}$$

6.2.2 The variation

Let $q(\cdot)$ be the trajectory starting from q_{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. Notice that $r(0) = q(0) = q_{\text{in}}$. Moreover if $v(\cdot) \equiv 0$ then $r(\cdot) \equiv q_{\text{in}}$.



Let us look for an equation for $r(t)$. Differentiating $q(t) = \phi_t(r(t))$, we get

$$\dot{q}(t) = \frac{d}{dt}\phi_t \Big|_{r(t)} + \frac{\partial \phi_t}{\partial q} \Big|_{r(t)} \dot{r}(t) = f_{\tilde{u}(t)}(\phi_t(r(t))) + \frac{\partial \phi_t}{\partial q} \Big|_{r(t)} \dot{r}(t)$$

Now $\dot{q}(t) = f_{u(t)}(q(t)) = f_{u(t)}(\phi_t(r(t)))$. Hence

$$\begin{aligned} \dot{r}(t) &= \left[\frac{\partial \phi_t}{\partial q} \Big|_{r(t)} \right]^{-1} \left(f_{u(t)}(\phi_t(r(t))) - f_{\tilde{u}(t)}(\phi_t(r(t))) \right) = \\ &= \left[\frac{\partial \phi_t}{\partial q} \Big|_{r(t)} \right]^{-1} \sum_{i=1}^m v_i(t) F_i(\phi_t(r(t))) =: g_{v(t)}(r(t)). \end{aligned} \quad (6.9)$$

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 $g_{sv} = sg_v$, for every $s \in \mathbb{R}$.

6.2.3 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 6.2. *If $(\tilde{q}(\cdot), \tilde{u}(\cdot))$ is a solution to the problem **P-AQ** then there exists $\hat{p} \in \mathbb{R}^* \times T_{q_{\text{in}}}^* M$, $\hat{p} \neq 0$, such that*

$$\left\langle \hat{p}, \left(\frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \Big|_{s=0}, \frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \Big|_{s=0} \right)^T \right\rangle = 0 \text{ for every } v(\cdot). \quad (6.10)$$

PROOF.

▷ By contradiction there exists $n + 1$ variations $\eta^0(\cdot), \eta^1(\cdot), \dots, \eta^n(\cdot)$ such that

$$\left(\left. \frac{\partial q^0(T, \tilde{u}(\cdot) + s\eta^0(\cdot))}{\partial s} \right|_{s=0} \right), \dots, \left(\left. \frac{\partial q^0(T, \tilde{u}(\cdot) + s\eta^n(\cdot))}{\partial s} \right|_{s=0} \right) \quad (6.11)$$

are linearly independent. It follows that the map

$$(s_0, \dots, s_n) \mapsto \begin{pmatrix} q^0(T, \tilde{u}(\cdot) + \sum_{j=0}^n s_j \eta^j(\cdot)) \\ r(T, \tilde{u}(\cdot) + \sum_{j=0}^n s_j \eta^j(\cdot)) \end{pmatrix} \quad (6.12)$$

is a local diffeomorphism in a neighborhood of the origin of $\mathbf{R}^{n+1}$ (indeed the vectors (6.11) are the components of the differential of the map (6.12)).

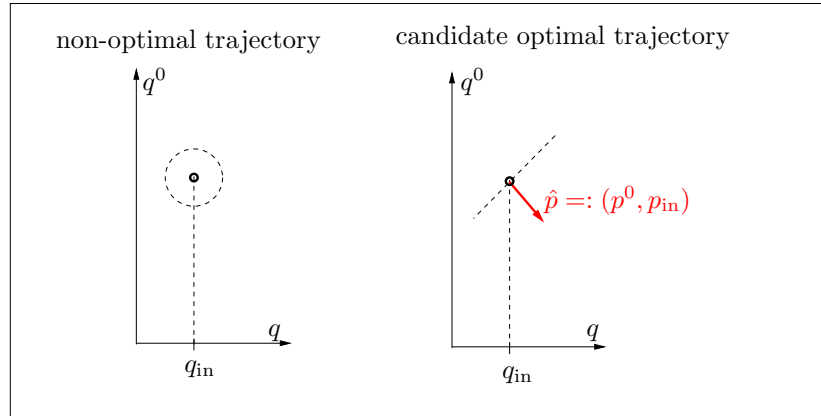
Let us prove that this is not possible. First notice that the image of the origin through the map (6.12) is,

$$\begin{pmatrix} q^0(T, \tilde{u}(\cdot)) \\ r(T, \tilde{u}(\cdot)) \end{pmatrix} = \begin{pmatrix} q^0(T, \tilde{u}(\cdot)) \\ q_{\text{in}} \end{pmatrix}.$$

Now if (6.12) were a local diffeomorphism there would exist $\tilde{v}(\cdot)$ such that

$$\begin{aligned} q^0(T, \tilde{u}(\cdot) + \tilde{v}(\cdot)) &< q^0(T, \tilde{u}(\cdot)) \\ q(T, \tilde{u}(\cdot) + \tilde{v}(\cdot)) &= \phi_T(r(T, \tilde{u}(\cdot) + \tilde{v}(\cdot))) = \phi_T(r(T, \tilde{u}(\cdot))) = \phi_T(q_{\text{in}}) = q_{\text{in}}, \end{aligned}$$

and $(\tilde{q}(\cdot), \tilde{u}(\cdot))$ would not be optimal.



□

Setting $\hat{p} = (p^0, p_{\text{in}})$ in (6.10), we have that

$$p^0 \left. \frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \right|_{s=0} + \langle p_{\text{in}}, \left. \frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \right|_{s=0} \rangle = 0 \text{ for every } v(\cdot) \quad (6.13)$$

Let us compute the different terms in the sum.

$$\begin{aligned}
 \left. \frac{\partial q^0(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \right|_{s=0} &= \left. \frac{\partial}{\partial s} \right|_{s=0} \int_0^T (\tilde{u}(t) + sv(t))(\tilde{u}(t) + sv(t)) dt \\
 &= 2 \int_0^T \sum_{i=1}^m \tilde{u}_i(t) v_i(t) dt \\
 \left. \frac{\partial r(T, \tilde{u}(\cdot) + sv(\cdot))}{\partial s} \right|_{s=0} &= \left. \frac{\partial}{\partial s} \right|_{s=0} \int_0^T g_{sv(t)}(r(t, \tilde{u}(\cdot) + sv(\cdot))) dt \\
 &= \int_0^T g_{v(t)}(r(t, \tilde{u}(\cdot))) dt = \int_0^T \sum_{i=1}^m v_i(t) \phi_{t*}^{-1}(q_{\text{in}}) F_i(\tilde{q}(t)) dt
 \end{aligned}$$

where 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 $\phi_t(r(t, \tilde{u}(\cdot))) = \tilde{q}(t)$. From (6.13) it follows that

$$\int_0^T \sum_{i=1}^m v_i(t) \left(2p^0 \tilde{u}_i(t) + \langle p_{\text{in}}, \phi_{t*}^{-1}(q_{\text{in}}) F_i(\tilde{q}(t)) \rangle \right) dt = 0.$$

Now $\langle p_{\text{in}}, \phi_{t*}^{-1}(q_{\text{in}}) F_i(\tilde{q}(t)) \rangle = \langle (\phi_t^{-1})^* p_{\text{in}}, F_i(\tilde{q}(t)) \rangle$. Hence defining $p(t) := (\phi_t^{-1})^* p_{\text{in}}$ we obtain

$$\int_0^T \sum_{i=1}^m v_i(t) \left(2p^0 \tilde{u}_i(t) + \langle p(t), F_i(\tilde{q}(t)) \rangle \right) dt = 0.$$

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 6.3. *If $(q(\cdot), u(\cdot))$ is a solution to the problem OCP-AQ then there exists $\mathbb{R}^* \times T_{q_{\text{in}}} M \ni (p^0, p_{\text{in}}) \neq (0, 0)$ such that*

$$2p^0 u_i(t) + \langle p(t), F_i(q(t)) \rangle = 0, \text{ for almost every } t \in [0, T], \quad (6.15)$$

where $p(t) := (\phi_t^{-1})^ p_{\text{in}}$ and ϕ_t is the flow corresponding to the optimal control $u(\cdot)$.*

6.2.4 The Hamiltonian form

It remains to show that Proposition 6.3 is equivalent to the Pontryagin Maximum Principle (Theorem (6.1)).

- equation **i)** of Theorem 6.1 is equivalent to $\dot{q}(t) = F_0(q(t)) + \sum_{i=1}^m u_i(t)F_i(q(t))$. Such an equation is our starting point in problem OCP-AQ.
- equation **iii)** of Theorem 6.1 gives immediately equation (6.15).
- it remains to show that $p(t) := (\phi_t^{-1})^* p_{\text{in}}$ is the solution of **ii)** of Theorem 6.1. This is the content of the rest of this section.

We have the following

Lemma 6.4. *Consider the control system*

$$\dot{z} = \mathbf{F}(z(t), u(t)) \quad (6.16)$$

where

- N is a smooth manifold, $U \subset \mathbb{R}^m$,
- $\mathbf{F}$ is a smooth function of its arguments,
- $u(\cdot) \in L^\infty([0, T], U)$,
- $z(\cdot) : [0, T] \rightarrow N$, belongs to the set of Lipschitz curves.

Fix a control $u(\cdot)$, an initial condition $z(0) = z_{\text{in}}$, and $\lambda_{\text{in}} \in T_{z_{\text{in}}}^* N$.

Define $\lambda(t) = (\Phi_t^{-1})^* \lambda_{\text{in}} \in T_{z(t)}^* N$ where Φ_t is the flow of (6.16).

Then the pair $(z(\cdot), \lambda(\cdot))$ is solution of the Hamiltonian system

$$\begin{cases} \dot{z} = \frac{\partial H}{\partial \lambda}(z, \lambda, u(t)) \\ \dot{\lambda} = -\frac{\partial H}{\partial q}(z, \lambda, u(t)), \end{cases} \quad (6.17)$$

where $H(z, \lambda, u) := \langle \lambda, \mathbf{F}(z, u) \rangle$.

Remark 6.5. *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 make evolving λ_{in} on T^*N according to the evolution of $z(\cdot)$ i.e., $\lambda(t) = (\Phi_t^{-1})^* \lambda_{\text{in}} \in T_{z(t)}^* N$ then the pair $(z(\cdot), \lambda(\cdot))$ is solution of the simplest possible Hamiltonian system, namely the one having the time-dependent Hamiltonian $H(z, \lambda, u(t)) := \langle \lambda, \mathbf{F}(z, u(t)) \rangle$.*

Remark 6.6. *We stated Lemma 6.16 for a control system. However in that lemma the control does not play any role and the lemma 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 (6.7) for the candidate optimal solution $(q(t), u(t))$, Namely we set

$$\begin{aligned} N &= \mathbb{R} \times M \\ z &= \begin{pmatrix} q^0 \\ q \end{pmatrix} \\ \mathbf{F}(z, u) &= \begin{pmatrix} \sum_{i=1}^m u_i^2 \\ F_0(q) + \sum_{i=1}^m u_i F_i(q) \end{pmatrix} \end{aligned}$$

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

$$\lambda(t) = (\Phi_t^{-1})^* \lambda_{\text{in}} \in T_{z(t)}^*N.$$

Notice that since the “zero” component of $\mathbf{F}$ is decoupled from the others, we have that

$$(\Phi_t^{-1})^* = \left(\begin{array}{c|c} \star & 0 \\ \hline 0 & (\phi_t^{-1})^* \end{array} \right)$$

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 (6.17) become

$$\dot{q}^0 = \frac{\partial \mathcal{H}}{\partial p^0} = \sum_{i=1}^m u_i^2(t) \quad (6.18)$$

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial p} \quad (6.19)$$

$$\dot{p}^0 = -\frac{\partial \mathcal{H}}{\partial q^0} = 0 \quad (6.20)$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial q} \quad (6.21)$$

Equations (6.21) is the desired equation **ii**) of Theorem (6.1). Equations (6.18) and (6.19) tell something that we already knew, Equation (6.18) says that $\star$ is actually the identity.

Proof of Lemma 6.4

Let us fix a system of coordinates and use matrix notation. Let us write the differential of the flow of $\mathbf{F}$ as

$$A(t, z) := \Phi_{t*} = \frac{\partial \Phi_t}{\partial z} \text{ then } \lambda(t) = (\Phi_t^{-1})^* \lambda_{\text{in}} = \lambda_{\text{in}} A^{-1}(t, z).$$

The operator A satisfies the famous *equation of variations*

$$\frac{\partial}{\partial t} A(t, z) = \frac{\partial \mathbf{F}}{\partial z} \Big|_{\Phi_t(z)} A(t, z) \quad (6.22)$$

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 [Bressan-Piccoli].

Now starting from (6.22) let us look for an equation for $A(t, z)^{-1}$. We have

$$AA^{-1} = \text{id} \Rightarrow \dot{A}A^{-1} + A\dot{A}^{-1} = 0 \Rightarrow \dot{A}^{-1} = -A^{-1}\dot{A}A^{-1}$$

Hence

$$\begin{aligned} \dot{\lambda} &= \frac{d}{dt}(\lambda_{in}A^{-1}) = \lambda_{in}\dot{A}^{-1} = -\lambda_{in}A^{-1}\dot{A}A^{-1} = -\underbrace{\lambda_{in}A^{-1}}_{\lambda} \frac{\partial \mathbf{F}}{\partial z} \Big|_{\Phi_t(z)} AA^{-1} \\ &= -\lambda \frac{\partial \mathbf{F}}{\partial z} \Big|_{\Phi_t(z)} = -\frac{\partial H}{\partial z}(z, \lambda, u(t)) \end{aligned}$$

(recall that $H(z, \lambda, u) := \langle \lambda, \mathbf{F}(z, u) \rangle$).

□

Chapter 7

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 6 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. Why it is called sub-Riemannian is explained in Remark (7.5).

Sub-Riemannian Problem

$$\begin{aligned}\dot{q} &= \sum_{i=1}^m u_i(t) F_i(q), \\ q(0) &= q_{\text{in}}, \quad q(T) \in \mathcal{T}, \\ \int_0^T \sum_{i=1}^m u_i(t)^2 dt &\rightarrow \min\end{aligned}\tag{7.2}$$

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),
- $F_i, i = 1, \dots, m$ are smooth vector fields,
- $u(\cdot) := (u_1(\cdot), \dots, u_m(\cdot)) \in L^\infty([0, T], \mathbb{R}^m)$,
- $q(\cdot) : [0, T] \rightarrow M$, belongs to the set of Lipschitz curves.

7.1 Energy, length and existence of optimal trajectories

Let us start to discuss a crucial property of the dynamics (7.2)

Lemma 7.0. *Let $q(\cdot) : [0, T] \rightarrow \mathbf{R}^n$, $T > 0$, be an admissible trajectory corresponding to controls $u_i(\cdot)$, $i = 1, \dots, m$. Let $\tau : [0, \bar{T}] \rightarrow [0, T]$, be a Lipschitz function such that $\frac{d}{dt}\tau(t) > 0$ a.e. on $[0, \bar{T}]$. Set $\bar{q}(\cdot) = q(\tau(\cdot))$ (this trajectory defined on $[0, \bar{T}]$ is called a Lipschitz reparametrization of $q(\cdot)$). Then $\bar{q}(\cdot)$ is an admissible trajectory (on $[0, \bar{T}]$) and corresponds to controls $u_i(\tau(\cdot))\dot{\tau}(\cdot)$, $i = 1, \dots, m$.*

Proof. We have

$$\dot{\bar{q}}(t) = \frac{d}{dt}q(\tau(t)) = \dot{q}(\tau(t))\dot{\tau}(t) = \left(\sum_{i=1}^m u_i(\tau(t))F_i(q(\tau(t))) \right) \dot{\tau}(t) = \sum_{i=1}^m \left(u_i(\tau(t))\dot{\tau}(t) \right) F_i(\bar{q}(t)).$$

□

As a consequence for a sub-Riemannian problem we have that for every $q_{\text{in}} \in M$ and $T > 0$, we have that $\mathcal{R}(q_{\text{in}}, T) = \mathcal{R}(q_{\text{in}})$. Recall that a sufficient condition to have $\mathcal{R}(q_{\text{in}}) = M$ is that for every $q \in M$ we have $\dim(\text{Lie}_q\{F_1, \dots, F_m\}) = n$.

Actually, with a little bit more work, one can prove the following.

Lemma 7.1. *Let $q(\cdot) : [0, T] \rightarrow \mathbf{R}^n$, $T > 0$, be an admissible trajectory corresponding to controls $u_i(\cdot)$, $i = 1, \dots, m$. Let $L = \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2} dt$. Then $q(\cdot)$ is a Lipschitz reparameterization of an admissible trajectory, defined on $[0, T]$ as well, corresponding to controls $\bar{u}_i(\cdot)$, $i = 1, \dots, m$ satisfying $\sqrt{\sum_{i=1}^m \bar{u}_i(t)^2} = L/T$, a.e. on $[0, T]$. Alternatively, $q(\cdot)$ is a Lipschitz reparameterization of an admissible trajectory defined on $[0, L]$, corresponding to controls satisfying $\sqrt{\sum_{i=1}^m u_i(t)^2} = 1$, a.e. on $[0, L]$.*

Admissible trajectories for which the corresponding controls satisfies $\sum_{i=1}^m u_i(t)^2 = 1$ a.e. are said to be *parametrized by arclength*. In other words Lemma 7.1 tells that every admissible trajectory is a Lipschitz reparametrization of a trajectory parameterized by arclength.

Now, for the control system (7.2), consider the two costs:

$$E(u(\cdot)) = \int_0^T \sum_{i=1}^m u_i(t)^2 dt, \quad (\text{energy}) \quad (7.3)$$

$$L(u(\cdot)) = \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2} dt \quad (\text{length}) \quad (7.4)$$

with the conditions

$$T > 0, \quad q(0) = q_{\text{in}} \text{ and } q(T) \in \mathcal{T}, q_{\text{in}} \notin \mathcal{T}. \quad (7.5)$$

Here $q_{\text{in}} \in M$ and $\mathcal{T}$ (a smooth submanifold of M) are fixed. Why these costs are called energy and length is explained in Remark (7.5).

Proposition 7.2. *Consider the optimal control problems (7.2), (7.5) and either (7.3) (energy) or (7.4) (length). We have the following.*

- A. Let $q(\cdot) : [0, T] \rightarrow \mathbf{R}^n$ be an admissible trajectory corresponding to controls $u_i(\cdot)$, $i = 1, \dots, m$ and $\bar{q}(\cdot) = q(\tau(\cdot))$ a Lipschitz reparametrization of $q(\cdot)$ as in Lemma 7.0. Let $\bar{u}(\cdot) = u(\tau(\cdot))\dot{\tau}(\cdot)$ defined on $[0, \bar{T}]$ be the controls corresponding to $\bar{q}(\cdot)$. We have

$$L(u(\cdot)) = \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2} dt = \int_0^{\bar{T}} \sqrt{\sum_{i=1}^m \bar{u}_i(t)^2} dt = L(\bar{u}(\cdot)).$$

Namely the cost L is invariant by Lipschitz reparametrizations of the trajectory.

- B. If T is not fixed, then there is no existence of minimizers for the cost E .
- C. Minimizers of E , with T fixed, satisfies $\sum_{i=1}^m u_i(t)^2 = \text{const}$, for almost every $t \in [0, T]$.
- D. Minimizers of E with T fixed are minimizers of L ,
- E. Minimizers of L defined on $[0, T]$ and parameterized in such that $\sum_{i=1}^m u_i(t)^2 = \text{const}$ a.e. are minimizers of E with final time fixed to T .

Proof. To prove A, simply notice that

$$L(\bar{u}(\cdot)) = \int_0^{\bar{T}} \sqrt{\sum_{i=1}^m \bar{u}_i(t)^2} dt = \int_0^{\bar{T}} \sqrt{\sum_{i=1}^m u_i(\tau(t))^2 \dot{\tau}(t)^2} dt = \int_0^T \sqrt{\sum_{i=1}^m u_i(s)^2} ds,$$

where in the last equality we made the change of variable $s = \tau(t)$.

To prove B assume by contradiction that $q(\cdot)$, defined in $[0, T_0]$, and corresponding to controls $u_i(\cdot)$, $i = 1, \dots, m$ is a minimizer of E for T free. Let $c \in (0, 1)$. The trajectory corresponding to controls $\bar{u}_i(\cdot) = u_i(c \cdot)c$ is a reparameterization of $q(\cdot)$ reaching the final point at time $\bar{T} = T_0/c$. Its cost is

$$E(\bar{u}(\cdot)) = \int_0^{T_0/c} \sum_{i=1}^m u_i(ct)^2 c^2 dt = cE(u(\cdot)) < E(u(\cdot)).$$

Contradiction.

To prove C, D, E, we are going to use the Cauchy-Schwarz inequality $(f, g)^2 \leq \|f\|^2 \|g\|^2$ (with equality holding if and only if f and g are proportional) which holds in any Hilbert space.¹ In particular this holds in $L^2([0, T], \mathbf{R})$. Namely

$$\left(\int_0^T f(t)g(t) dt \right)^2 \leq \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt \quad (\text{with equality holding iff } f \text{ is a.e. proportional to } g).$$

Now let $f(t) = \sqrt{\sum_{i=1}^m u_i(t)^2}$ and $g = 1$, $t \in [0, T]$. Notice that $f \in L^\infty([0, T], \mathbf{R}) \subset L^2([0, T], \mathbf{R})$. We have

$$L(u(\cdot))^2 \leq E(u(\cdot))T \quad (\text{with equality holding iff } \sum_{i=1}^m u_i(t)^2 \text{ is a.e. constant}).$$

To prove C let $u(\cdot)$ be a minimizer of E defined on $[0, T]$. By contradiction assume that $\sum_{i=1}^m u_i(t)^2$ is not a.e. constant. Then $L(u(\cdot))^2 < E(u(\cdot))T$. Let $q_R(\cdot)$ be an admissible trajectory defined on $[0, T]$ corresponding to controls satisfying $\sum_{i=1}^m u_i(t)^2 = \text{const}$ a.e., of which $q(\cdot)$ is a reparametrization as in Lemma 7.1. For this trajectory we have

$$E(u_R(\cdot))T = L(u_R(\cdot))^2 = L(u(\cdot))^2 < E(u(\cdot))T,$$

contradicting the fact that $u(\cdot)$ is a minimizer of E .

To prove D let $u(\cdot)$ be the minimizer of E defined on $[0, T]$. By C we have $\sum_{i=1}^m u_i(t)^2 = \text{const}$. By contradiction assume that $u(\cdot)$ is not a minimizer of L . Then there exists $\bar{u}(\cdot)$ going from q_{in} to $\mathcal{T}$ such that $L(\bar{u}(\cdot)) < L(u(\cdot))$. We can assume that $\sum_{i=1}^m \bar{u}_i(t)^2 = \text{const}$. Then

$$E(\bar{u}(\cdot))T = L(\bar{u}(\cdot))^2 < L(u(\cdot))^2 = E(u(\cdot))T,$$

contradicting the fact that $u(\cdot)$ is a minimizer of E .

To prove E, let $u(\cdot)$ be a minimizer of L parameterized in such that $\sum_{i=1}^m u_i(t)^2 = \text{const}$. By contradiction assume that $u(\cdot)$ is not a minimizer of E . Then there exists $\bar{u}(\cdot)$ going from q_{in} to $\mathcal{T}$ in time T such that $E(\bar{u}(\cdot)) < E(u(\cdot))$. We have

$$L(\bar{u}(\cdot))^2 \leq E(\bar{u}(\cdot))T < E(u(\cdot))T = L(u(\cdot))^2$$

contradicting the fact that $u(\cdot)$ is a minimizer of L . □

Proposition 7.3. *For the control system (7.2), the two optimal control problems:*

- P1. *minimize E with the boundary conditions (7.5) and T fixed in such a way that $\sum_{i=1}^m u_i(t)^2 = 1$ a.e.,*
- P2. *minimizing time with the same boundary conditions and the constraint on the controls $\sum_{i=1}^m u_i(t)^2 \leq 1$,*

¹This inequality simply tells that the scalar product of two vectors is less or equal to the product of the norm of the two vectors (with equality holding iff the two vectors are collinear).

are equivalent.

Proof. Since $\sum_{i=1}^m u_i(t)^2 = 1$ a.e. we have $\int_0^T (\sum_{i=1}^m u_i(t)^2) dt = T$. Hence P1. is equivalent to the problem of minimizing T with the constraint on the controls $\sum_{i=1}^m u_i(t)^2 = 1$. To conclude the proof, let us show that a trajectory corresponding to controls for which the condition

$$\sum_{i=1}^m u_i(t)^2 = 1 \quad \text{a.e.} \quad (7.6)$$

is not satisfied cannot be optimal for [P2]. Actually if (7.6) is not satisfied then $L = \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2} < T$ and an arc length reparameterization of the trajectory reaches the target in time exactly L . $\square$

Let us now go back to the problem of existence of optimal trajectories for the sub-Riemannian problem. Thanks to Proposition 7.3, 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 $U = \{(u_1, \dots, u_m) \in \mathbb{R}^m \mid u_1^2 + \dots + u_m^2 \leq 1\}$. We can then apply Proposition 4.4 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 7.4. *Fix $T > 0$ and consider the optimal control problems (7.2), (7.5) and either (7.3) (energy) or (7.4) (length). Assume that*

- *for every $u(\cdot) \in L^\infty([0, T], U)$, the solution of (7.2), 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 these two optimal control problems admit a solution.

Remark 7.5. *The name sub-Riemannian for the problem that we are considering in this chapter comes from the following argument.*

Consider the very simple case in which $m = n$, $M = \mathbb{R}^n$, $\mathcal{T} = \{q_{\text{fin}}\}$ and

$$F_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, F_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

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 7.2 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 (they have norm 1 and they are mutually orthogonal).

When we are on a general manifold M , we have $m = n$, and $F_1, \dots, F_n$ are (in general non constant) linearly independent vector fields, the cost

$$\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2} dt$$

is the length of the curve for the Riemannian metric for which $F_1, \dots, F_n$ is an orthonormal frame.² 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 $\text{span}\{F_1(q), \dots, F_m(q)\}$ 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 and 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.

²A Riemannian metric can be thought as a scalar product depending on the point assigned on each tangent space of a manifold. Equivalently it can be thought as the assignment of a point-dependent orthonormal frame. Notice that not all Riemannian problems can be written in the form $\dot{q} = \sum_{i=1}^n u_i(t) F_i(q)$, $\int_0^T \sqrt{\sum_{i=1}^n u_i(t)^2} dt \rightarrow \min$ (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 $\dot{q} = \sum_{i=1}^m u_i(t) F_i(q)$, $\int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2} dt \rightarrow \min$ (with fixed initial and final conditions) if m is sufficiently large. See [Agr-Ba-Bo].

7.2 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 5.1) becomes as follow.³

Theorem 7.6 (Pontryagin Maximum Principle for a sub-Riemannian problem). *Define the function (called pre-Hamiltonian)*

$$\mathcal{H}(q, p, u, p^0) = \langle p, \sum_{i=1}^m u_i F_i(q) \rangle + p^0 \sum_{i=1}^m u_i^2, \quad (7.8)$$

with

$$(q, p, u, p^0) \in T^*M \times \mathbb{R}^m \times \mathbb{R}.$$

*If the pair $(q, u) : [0, T] \rightarrow M \times \mathbb{R}^m$ is optimal for the sub-Riemannian problem, 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) $\dot{q}(t) = \frac{\partial \mathcal{H}}{\partial p}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for q);
- ii) $\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial q}(q(t), p(t), u(t), p^0)$ (Hamiltonian equation for p);
- iii) $\frac{\partial \mathcal{H}}{\partial u}(x(t), p(t), p^0, u(t)) = 0$ (maximization condition);
- iv) there exists a constant $c \geq 0$ such that $\mathcal{H}(q(t), p(t), u(t), p^0) = c$ on $[0, T]$,
- v) 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 5.2.

Step 1. Use the maximization condition **iii)** to express, when possible, the control as a function of the state and of the covector.

Condition **iii)** on (7.8) gives the following.

- for abnormal extremals ($p^0 = 0$) we obtain

$$\langle p(t), F_i(q(t)) \rangle = 0, \quad i = 1, \dots, m, \quad t \in [0, T] \quad (7.9)$$

This condition does not permit to obtain the controls as function of the state and of the covector. Hence, for the sub-Riemannian problem, *abnormal extremals*

³or equivalently Theorem 6.1 with $F_0 = 0$ and the adding of transversality conditions since now the final point is constrained to belong to a smooth submanifold $\mathcal{T}$.

corresponds to singular controls. To obtain the controls one should derive with respect to time equation (7.9).

Consider for instance the case $m = n$ and $F_1, \dots, F_n$ linearly independent at each point. In this case the condition (7.9) gives $p \equiv 0$ which is impossible since $p^0 = 0$. this is the reason why in the Riemannian case there are no abnormal extremals (cf. remark 7.5) .

As another example, consider the case in which $n = 3$, $m = 2$ and $F_1, F_2, [F_1, F_2]$ linearly independent at each point. This is called the 3D contact sub-Riemannian case (cf. [Agr-Ba-Bo]).

If we differentiate with respect to time (7.9) for $i = 1, 2$ we get a.e.:

$$u_2(t) \langle p(t), [F_1, F_2](q(t)) \rangle = 0,$$

$$u_1(t) \langle p(t), [F_1, F_2](q(t)) \rangle = 0.$$

These equations, together with $p(t) \neq 0$ for every t and

$$\langle p(t), F_1(q(t)) \rangle = 0, \quad (7.10)$$

$$\langle p(t), F_2(q(t)) \rangle = 0, \quad (7.11)$$

imply $u_1(t) = u_2(t) = 0$ for a.e. t . Hence in the 3D contact case the only possible abnormal trajectory is the trivial trajectory $q(t) \equiv q_{\text{in}}$. Non-trivial abnormal extremals appear if we do not require that $F_1, F_2, [F_1, F_2]$ are linearly independent at each point or if the difference between n and m is big enough.

- for normal extremals ($p^0 \neq 0$) we can normalize $p^0 = -\frac{1}{2}$ and we obtain

$$u_i(t) = \langle p(t), F_i(q(t)) \rangle, \quad i = 1, \dots, m, \quad t \in [0, T]. \quad (7.12)$$

This condition expresses the control as function of the state and of the covector. More explicitly with the notations of Section 5.2 we obtain

$$u = w(q, p) := (\langle p, F_1(q) \rangle, \dots, \langle p, F_m(q) \rangle).$$

Hence for the sub-Riemannian problem *normal extremals corresponds to regular controls.*

Step 2. Insert the control found in the previous step into the Hamiltonian equations **i)** and **ii)**.⁴ 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.

$$\begin{aligned} \mathcal{H}_M(q, p) &:= \mathcal{H}(q, p, w(q, p), -\frac{1}{2}) = \langle p, \sum_{i=1}^m w_i(q, p) F_i(q) \rangle - \frac{1}{2} \sum_{i=1}^m w_i(q, p)^2 \\ &= \frac{1}{2} \sum_{i=1}^m w_i(q, p)^2 = \frac{1}{2} \sum_{i=1}^m \langle p, F_i(q) \rangle^2. \end{aligned} \quad (7.13)$$

⁴We are going to develop this step for normal extremals only . For abnormal extremals the problem could be complicated and will be treated case by case in the examples.

Notice that $\mathcal{H}_M(q(t), p(t)) = \frac{1}{2} \sum_{i=1}^m w_i(q(t), p(t))^2 = \frac{1}{2} \sum_{i=1}^m u_i(t)^2$. This quantity is constant and equal to $\frac{1}{2}$ if trajectories are parameterized by arclength. A usefull fact is that

$$\begin{aligned} \frac{\partial \mathcal{H}_M}{\partial p}(q, p) &= \frac{\partial}{\partial p} \mathcal{H}(q, p, w(q, p), -\frac{1}{2}) \\ &= \frac{\partial \mathcal{H}}{\partial p}(q, p, w(q, p), -\frac{1}{2}) + \frac{\partial \mathcal{H}}{\partial u}(q, p, w(q, p), -\frac{1}{2}) \frac{\partial w}{\partial p}(q, p) \\ &= \frac{\partial \mathcal{H}}{\partial p}(q, p, w(q, p), -\frac{1}{2}). \end{aligned}$$

where we used the fact that $\frac{\partial \mathcal{H}}{\partial u}(q, p, w(q, p), -\frac{1}{2}) = 0$. Similarly

$$\frac{\partial \mathcal{H}_M}{\partial q}(q, p) = \frac{\partial \mathcal{H}}{\partial q}(q, p, w(q, p), -\frac{1}{2}).$$

Hence the Hamiltonian equations **i)** and **ii)** can be written with the maximized Hamiltonian in which the control is not present anymore.

$$\begin{cases} \dot{q} = \frac{\partial \mathcal{H}_M}{\partial p}(q, p) \\ \dot{p} = -\frac{\partial \mathcal{H}_M}{\partial q}(q, p). \end{cases} \quad (7.14)$$

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

$$\mathcal{H}_M(q_{\text{in}}, p_{\text{in}}) = \frac{1}{2}.$$

Remark 7.7. Notice that the set $\{p_{\text{in}} \in T_{q_{\text{in}}}^* M \mid \mathcal{H}_M(q_{\text{in}}, p_{\text{in}}) = \frac{1}{2}\}$ is a ellipsoid if $\dim(\text{span}\{F_1(q_{\text{in}}), \dots, F_m(q_{\text{in}})\}) = n$ 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

$$\begin{cases} q(T; p_{\text{in}}, p^0) \in \mathcal{T} \\ \langle p(T; p_{\text{in}}, p^0), T_{q_{\text{in}}} \mathcal{T} \rangle = 0. \end{cases} \quad (7.15)$$

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 (7.15) provides a unques pair (p^0, p_{in}) then if we have found the optimal solution. This however is not the general situation.

Chapter 8

Time-optimal Control on the Plane

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.

Problem Min-T-2D

$$\dot{x} = F(x) + u(t)G(x), \quad (8.3)$$

$$x(0) = x_{\text{in}}, \quad x(T) = x_{\text{fi}} \quad (8.4)$$

$$T \rightarrow \min$$

Here

- $T \geq 0$ is free,
- F, G , are smooth vector fields on $\mathbb{R}^2$,
- $u(\cdot) \in L^\infty([0, T], [-1, 1])$,
- $x(\cdot) : [0, T] \rightarrow \mathbb{R}^2$, 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 (8.15) with $x(0) = x_{\text{in}}$ exists for any time.

In this case the pre-Hamiltonian (6.6) is

$$\mathcal{H}(x, p, u, p^0) = p \cdot F(x) + u p \cdot G(x) + p^0 \quad (8.5)$$

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 5.1) tells the following.

Theorem 8.0. *If the pair $(x, u) : [0, T] \rightarrow \mathbb{R}^2 \times [-1, 1]$ 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*

$$\dot{p}(t) = -p(t) \cdot \left(\frac{\partial F}{\partial x} + u(t) \frac{\partial G}{\partial x} \right)(x(t)), \quad (\text{Hamiltonian equation for } p) \quad (8.6)$$

$$u(t) \, p(t) \cdot G(x(t)) = \max_{v \in [-1, 1]} v \, p(t) \cdot G(x(t)) \quad (\text{maximization condition}). \quad (8.7)$$

$$p(t) \cdot F(x(t)) + u(t) \, p(t) \cdot G(x(t)) + p^0 = 0 \quad (\text{zero value of the Hamiltonian}) \quad (8.8)$$

Equation (8.8) comes from the fact that T is free. A simple and important property is the following.

Lemma 8.1. *For the problem Min- T -2D we have $p(t) \neq 0$ for every t .*

Proof. Actually if $p(t) = 0$ for a time t then from (8.6) we would get $p(t) \equiv 0$. Then from (8.8) we would get $p^0 = 0$ which is impossible. $\square$

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 5.2.

Step 1. Use the maximization condition **iii)** to express, when possible, the control as a function of the state and of the covector. To this purpose it is useful the following.

Definition 8.2. (switching function) *Let $(x(\cdot), p(\cdot), u(\cdot), p^0) : [0, T] \rightarrow \mathbb{R}^2 \times \mathbb{R}^{*2} \times [-1, 1] \times \mathbb{R}$ be an extremal. The corresponding switching function is $\phi(\cdot) := p(\cdot) \cdot G(x(\cdot)) : [0, T] \rightarrow \mathbb{R}$.*

Equation (8.7) gives the following.

Lemma 8.3. *Consider an extremal and let $\phi(\cdot)$ the corresponding switching function. If $\phi(\cdot)$ has only isolated zeros then $u(t) = \text{sgn}(\phi(t))$ on $[0, T]$.*

Definition 8.4. (bang-bang extremal) *An extremal satisfying the hypotheses of Lemma 8.3 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 as for instance when $\phi(t) \equiv 0$.

Definition 8.5. (singular extremal) *An extremal for which the corresponding switching function is identically zero on some interval $[t_1, t_2] \subset [0, T]$ is called singular on $[t_1, t_2]$.*

To obtain information concerning the control for singular extremals let us observe that $\phi(\cdot)$ is a Lipschitz function and hence we can differentiate it a.e. In this way we obtain the following.

Lemma 8.6. *We have*

$$\dot{\phi}(t) = p(t) \cdot [F, G](x(t)).$$

hence $\phi(\cdot)$ is of class $\mathcal{C}^1$.

Proof. Using the Pontryagin Maximum principle we have for almost every t ,

$$\begin{aligned} \dot{\phi}(t) &= \frac{d}{dt}(p(t) \cdot G(x(t))) = \dot{p}(t) \cdot G(x(t)) + p(t) \cdot \dot{G}(x(t)) \\ &= -p(t) \left(\frac{\partial F}{\partial x} + u(t) \frac{\partial G}{\partial x} \right)(x(t)) \cdot G(x(t)) + p(t) \cdot \frac{\partial G}{\partial x}(x(t)) (F + u(t)G)(x(t)) \\ &= p(t) \cdot [F, G](x(t)). \end{aligned} \tag{8.9}$$

□

Let us introduce the following two functions of which we are interested to the set of zeros.

$$\Delta_A(x) := \det(F(x), G(x)) = F_1(x)G_2(x) - F_2(x)G_1(x), \tag{8.10}$$

$$\Delta_B(x) := \det(G(x), [F, G](x)) = G_1(x)[F, G]_2(x) - G_2(x)[F, G]_1(x). \tag{8.11}$$

The set of zeros $\Delta_A^{-1}(0), \Delta_B^{-1}(0)$, of these two functions, are respectively the set of points where F and G are parallel and the set of points where G is parallel to $[F, G]$.

Using Lemma 8.6, the next Lemma gives information on where singular extremals could be.

Lemma 8.7. *Let $(x(\cdot), p(\cdot), u(\cdot), p^0) : [0, T] \rightarrow \mathbb{R}^2 \times \mathbb{R}^{*2} \times [-1, 1] \times \mathbb{R}$ be an extremal which is singular on $[t_1, t_2] \subset [0, T]$. Then $x([t_1, t_2]) \subset \Delta_B^{-1}(0)$.*

Proof. Since $\phi(t) = p(t) \cdot G(x(t)) \equiv 0$ and $\dot{\phi}(t) = p(t) \cdot [F, G](x(t)) \equiv 0$, it follows 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.

Lemma 8.8. *Let $(x(\cdot), p(\cdot), u(\cdot), p^0) : [0, T] \rightarrow \mathbb{R}^2 \times \mathbb{R}^{*2} \times [-1, 1] \times \mathbb{R}$ be an extremal which is singular on $[t_1, t_2] \subset [0, T]$. Assume that $\nabla \Delta_B(x(t)) \cdot G(x(t)) \neq 0$ on $[t_1, t_2]$. Then on $[t_1, t_2]$ it corresponds to the control $u(t) = \varphi(x(t))$ where*

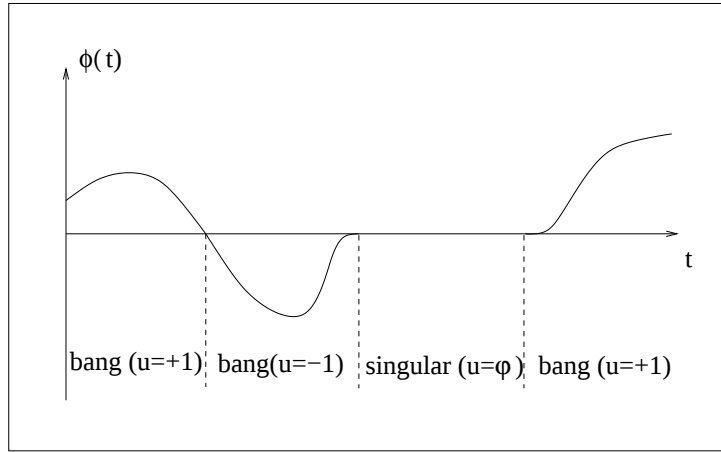
$$\varphi(x) = -\frac{\nabla \Delta_B(x) \cdot F(x)}{\nabla \Delta_B(x) \cdot G(x)}. \tag{8.12}$$

Proof. On $[t_1, t_2]$ we have a.e. $\dot{x}(t) = F(x(t)) + u(t)G(x(t))$ and $\Delta_B(x(t)) = 0$. Differentiating this last equation we obtain for a.e. $t \in [t_1, t_2]$,

$$0 = \frac{d}{dt}\Delta_B(x(t)) = \nabla\Delta_B \cdot (F(x(t)) + u(t)G(x(t))).$$

This means that at the point $x(t)$ we have to use control $\varphi(x(t))$, where φ is given nby given by (8.12). $\square$

En example of switching function with the corresponding controls is provided in the following picture.

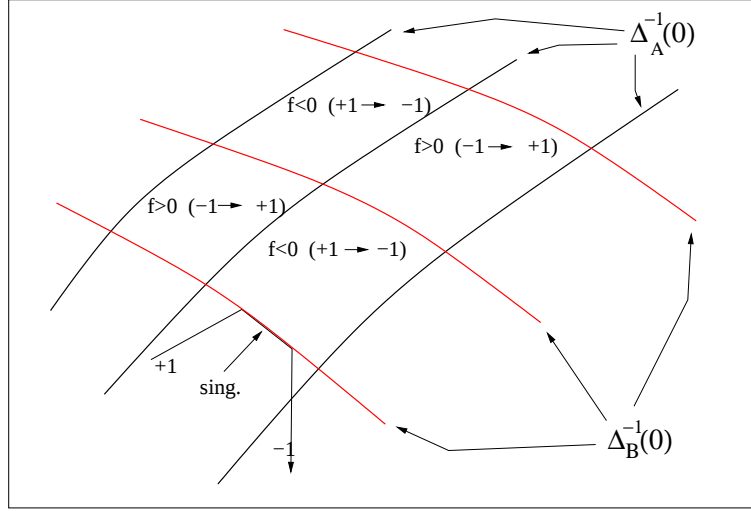


Step 2. Insert the control found in the previous step into the Hamiltonian equations (8.15) and (8.6). This step in general is more complicated than in the sub-Riemannian case. The reason is that the previous step provided the control as a function of the state and of the covector that in general is not smooth.

The construction of the extremals is in general complicated. However one can obtain additional information using the function the sets $\Delta_A^{-1}(0), \Delta_B^{-1}(0)$. We are going to prove that on each connected component of $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$, every extremal trajectory is bang-bang with at most one switching. Moreover, at a switching time, the value of the control switches from -1 to $+1$ if $-\Delta_B^{-1}(0) \setminus \Delta_A^{-1}(0) > 0$ and from $+1$ to -1 if $-\Delta_B^{-1}(0) \setminus \Delta_A^{-1}(0) < 0$. To prove this result first we need the following.

Lemma 8.9. *Let $x \notin \Delta_A^{-1}(0) \cup \Delta_B^{-1}(0)$. Then $F(x), G(x)$ form a basis of $\mathbb{R}^2$ and we define the scalar functions f_S, g_S to be the coefficients of the linear combination: $[F, G](x) = f_S(x)F(x) + g_S(x)G(x)$. We have*

$$f_S(x) = -\frac{\Delta_B(x)}{\Delta_A(x)}. \quad (8.13)$$



Possible commutations dans les composantes connexes de $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$ en relation avec the sign of f_S (in the picture indicated as f). In the picture there is also an example of extremal trajectory which is singular in a sub-interval.

Proof. We have:

$$\begin{aligned} \Delta_B(x) &= \text{Det}(G(x), [F, G](x)) = \text{Det}(G(x), f_S(x)F(x) + g_S(x)G(x)) \\ &= f_S(x)\text{Det}(G(x), F(x)) = -f_S(x)\Delta_A(x). \end{aligned}$$

□

Theorem 8.10. *Let $\Omega \in \mathbb{R}^2$ be a connected component of $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$. Let $[t_1, t_2] \subset [0, T]$ and $(x(\cdot), p(\cdot), u(\cdot), p^0) : [0, T] \rightarrow \mathbb{R}^2 \times \mathbb{R}^{*2} \times [-1, 1] \times \mathbb{R}$ be 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 $x(\cdot)|_{[t_1, t_2]}$ corresponds to control $+1, -1$ or has a $-1 \rightarrow +1$ switching (resp. has a $+1 \rightarrow -1$ switching).*

Proof. Assume $f_S > 0$ 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:

$$\dot{\phi}(\bar{t}) = p(\bar{t}) \cdot [F, G](x(\bar{t})) = p(\bar{t}) \cdot (f_S F + g_S G)(x(\bar{t})) = f_S(x(\bar{t})) p(\bar{t}) \cdot F(x(\bar{t})).$$

Now from (8.8) of the Pontryagin Maximum Principle we have that $p(t) \cdot F(x(t)) + u(t) p(t) \cdot G(x(t)) + p^0 = 0$ with $p^0 \leq 0$. Hence at $\bar{t}$ we have $p(\bar{t}) \cdot F(x(\bar{t})) \geq 0$. Actually $p(\bar{t}) \cdot F(x(\bar{t})) \neq 0$ otherwise F and G are parallel at $x(\bar{t})$ which is not possible since $\Delta_A(x(\bar{t})) \neq 0$. Hence $\dot{\phi}(\bar{t}) > 0$. This proves that ϕ has at most one zero with positive derivative at the switching time and gives the desired conclusion. □

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

$$x(T; p_{\text{in}}, p^0) = x_{\text{fi}} \quad (8.14)$$

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.

8.1 Time optimal problems for linear systems

All these problems are easily solved for linear systems, i.e. for the problem

$$\dot{x} = Ax + u(t)Bx, \quad (8.15)$$

$$x(0) = x_{\text{in}}, \quad x(T) = x_{\text{fin}} \quad (8.16)$$

$$T \rightarrow \min$$

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 (8.6) for $p(\cdot)$ involve neither $u(\cdot)$ nor $x(\cdot)$:

$$\dot{p}(t) = -p(t)A. \quad (8.17)$$

Hence for every p_{in} one solves (8.17), find $u(t) = \text{sign}(p(t) \cdot B)$, 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 .

Bibliographie

[Agr-Sa] A. Agrachev, Y. Sachkov, Control Theory from the Geometric Viewpoint, vol. 87 of Encyclopaedia of Mathematical Sciences. Control Theory and Optimization, II. Springer, Berlin (2004)

[Agr-Ba-Bo] A. Agrachev, D. Barilari, U. Boscain. A comprehensive introduction to sub-Riemannian geometry. Cambridge Studies in Advanced Mathematics, 181. Cambridge University Press, Cambridge, 2020.

[Rou-Bo] F. Bonnans, P. Rouchon, Commande et optimisation de systèmes dynamiques. Les Édition de L'École Polytechnique 2005.

[Arnold] V.I. Arnol'd, Ordinary differential equations. Translated from the third Russian edition by Roger Cooke. Springer Textbook. Springer-Verlag, Berlin, 1992.

[Bre-Pic] A. Bressan, B. Piccoli, Introduction to the mathematical theory of control. AIMS Series on Applied Mathematics, 2. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.

[Bo-Pi] U. Boscain, B. Piccoli, Optimal Synthesis for Control Systems on 2-D Manifolds, Springer, SMAI, Vol.43, 2004.

[Bo-Pi20] Boscain, B. Piccoli, Synthesis theory in optimal control in Encyclopedia of Systems and Control, Springer. 2020 (second edition).

[DoCarmo] M.P. do Carmo, Riemannian geometry. Translated from the second Portuguese edition by Francis Flaherty. Mathematics: Theory & Applications. Birkhauser Boston, Inc., Boston, MA, 1992.

[Jurd] V. Jurdjevic, Geometric Control Theory, vol. 52 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1997.

[Khal] H.K. Khalil, Nonlinear Systems, 3rd Edition, Pearson, 2001.

[Lee] J.M. Lee, Introduction to smooth manifolds. Second edition. Graduate Texts in Mathematics, 218. Springer, New York, 2013.

[Lee-Mark] Lee, E.B., Markus, L.: Foundations of optimal control theory, John Wiley, New York, 1967.

[Sc-Le] H. Schaettler and U. Ledzewicz, Geometric optimal control, vol. 38 of Interdisciplinary Applied Mathematics (Springer, New York, 2012).

[So] E. D. Sontag, Mathematical Control Theory: Deterministic Finite Dimensional Systems, no. 6 in Texts in Applied Mathematics, Springer-Verlag, New York/Heidelberg/Berlin, 2 ed., 1998.