A. A. Agrachev[1]   **Yu. L. Sachkov**[2]

Lectures on Geometric Control Theory

S.I.S.S.A. Ref. 38/2001/M (May 2001)

[1]Steklov Mathematical Institute, 8 ul. Gubkina, Moscow 117966, Russia
& S.I.S.S.A.–I.S.A.S., 2-4 Via Beirut, Trieste 34014, Italy
E-mail: agrachev@sissa.it
[2]Program Systems Institute, Pereslavl-Zalessky 152140, Russia
& S.I.S.S.A.–I.S.A.S., 2-4 Via Beirut, Trieste 34014, Italy
E-mail: sachkov@sissa.it

These are lecture notes based on the course given by the first coauthor in SISSA in Spring 2000. The potential reader could be a graduate student with a reasonable level of the mathematical culture and without any preliminary knowledge on Control Theory.

The main topics in Part I of the lecture notes are controllability and the equivalence of smooth systems under state and feedback transformations. The central result here is the Orbit Theorem of Nagano and Sussmann.

Part II is devoted to Optimal Control; the principal result is the Pontryagin Maximum Principle.

Geometric Control Theory is now a broad subject and many important topics are not even touched in our lecture notes. In particular, we do not study the feedback stabilization problem and the huge theory of control systems with outputs including fundamental concepts of observability and realization. For this and other material, see books [8], [9], [10], [11], [16].

The second author thanks SISSA, Trieste, Italy, and Program Systems Institute of Russian Academy of Sciences, Pereslavl-Zalessky, Russia, for support during the work on these lecture notes.

Contents

Chapter 1

Vector fields and control systems on smooth manifolds

We give just a brief outline of basic notions related to the smooth manifolds. For a consistent presentation, see an introductory chapter to any textbook on analysis on manifolds, e. g. [17].

In the sequel, "smooth" (manifold, mapping, vector field etc.) means C^∞.

1.1 Smooth manifolds

Definition 1.1. A subset $M \subset \mathbb{R}^n$ is called a *smooth k-dimensional submanifold* of $\mathbb{R}^n$, $k \leq n$, if any point $x \in M$ has a neighborhood O_x in $\mathbb{R}^n$ in which M is described in one of the following ways:

(1) there exists a smooth vector-function

$$F \; : \; O_x \to \mathbb{R}^{n-k}, \qquad \operatorname{rank} \left. \frac{dF}{dx} \right|_x = n - k$$

such that

$$O_x \cap M = F^{-1}(0);$$

(2) there exists a smooth vector-function

$$f \; : \; V_0 \to \mathbb{R}^n$$

from a neighborhood of the origin $0 \in V_0 \subset \mathbb{R}^k$ with

$$f(0) = x, \qquad \operatorname{rank} \left. \frac{df}{dx} \right|_0 = k$$

such that

$$O_x \cap M = f(V_0)$$

and $f : V_0 \to O_x \cap M$ is a homeomorphism.

(3) there exists a smooth vector-function

$$\Phi \ : \ O_x \to O_0 \subset \mathbb{R}^n$$

onto a neighborhood of the origin $0 \in O_0 \subset \mathbb{R}^n$ with

$$\operatorname{rank} \left. \frac{d\,\Phi}{d\,x} \right|_x = n$$

such that

$$\Phi(O_x \cap M) = \mathbb{R}^k \cap O_0.$$

Exercise 1.1. Prove that three local descriptions of a smooth submanifold given in (1)–(3) are mutually equivalent.

Remarks. (1) There are two topologically different one-dimensional manifolds: the line $\mathbb{R}^1$ and the circle S^1. The sphere S^2 and the torus $T^2 = S^1 \times S^1$ are two-dimensional manifolds. The torus can be viewed as a sphere with a handle. Any compact orientable two-dimensional manifold is topologically a sphere with p handles, $p = 0, 1, 2, \dots$.

(2) Smooth manifolds appear naturally already in the basic analysis. For example, the circle S^1 and the torus T^2 are natural domains of periodic and doubly periodic functions respectively. On the sphere S^2, it is convenient to consider restriction of homogeneous functions of 3 variables.

So a smooth submanifold is a subset in $\mathbb{R}^n$ which can locally be defined by a regular system of smooth equations and by a smooth regular parametrization.

In spite of the intuitive importance of the image of manifolds as subsets of a Euclidean space, it is often convenient to consider manifolds independently of any embedding in $\mathbb{R}^n$. An abstract manifold is defined as follows.

Definition 1.2. A *smooth k-dimensional manifold M* is a Hausdorff paracompact topological space endowed with a smooth structure: M is covered by a system of open subsets

$$M = \cup_\alpha O_\alpha$$

called *coordinate neighborhoods*, in each of which is defined a homeomorphism

$$\Phi_\alpha \ : \ O_\alpha \to \mathbb{R}^k$$

called a *local coordinate system* such that all compositions

$$\Phi_{\alpha_1} \circ \Phi_{\alpha_2}^{-1} \ : \Phi_{\alpha_2}(O_{\alpha_1} \cap O_{\alpha_2}) \subset \mathbb{R}^k \to \Phi_{\alpha_1}(O_{\alpha_1} \cap O_{\alpha_2}) \subset \mathbb{R}^k$$

are diffeomorphisms.

Remark. For a smooth submanifold in $\mathbb{R}^n$, the abstract definition holds.

Conversely, any connected smooth abstract manifold can be considered as a smooth submanifold in $\mathbb{R}^n$. Before precise formulation of this statement, we give two definitions.

Definition 1.3. Let M and N be k- and l-dimensional smooth manifolds respectively. A mapping

$$f \ : \ M \to N$$

is called *smooth* if it is smooth in coordinates. That is, let $M = \cup_\alpha O_\alpha$ and $N = \cup_\beta V_\beta$ be coverings of M and N by coordinate neighborhoods and

$$\Phi_\alpha \ : \ O_\alpha \to \mathbb{R}^k, \qquad \Psi_\beta \ : \ V_\beta \to \mathbb{R}^l$$

the corresponding coordinate mappings. Then all compositions

$$\Psi_\beta \circ f \circ \Phi_\alpha^{-1} \ : \ \Phi_\alpha(O_\alpha \cap f^{-1}(V_\beta)) \subset \mathbb{R}^k \to \Psi_\beta(f(O_\alpha) \cap V_\beta) \subset \mathbb{R}^l$$

should be smooth.

Definition 1.4. A smooth manifold M is called *diffeomorphic* to a smooth manifold N if there exists a homeomorphism

$$f \ : \ M \to N$$

such that both f and its inverse f^{-1} are smooth mappings. Such mapping f is called a *diffeomorphism*.

The set of all diffeomorphisms $f \ : \ M \to M$ of a smooth manifold M is denoted by $\operatorname{Diff} M$.

A smooth mapping $f \ : \ M \to N$ is called an *embedding* of M into N if $f \ : \ M \to f(M)$ is a diffeomorphism. A mapping $f \ : \ M \to N$ is called *proper* if $f^{-1}(K)$ is compact for any compactum $K \Subset N$.

Theorem 1.1 (Whitney). *Any smooth connected k-dimensional manifold can be properly embedded into $\mathbb{R}^{2k+1}$.*

Summing up, we may say that a smooth manifold is a space which looks locally like a linear space but without fixed linear structure, so that all smooth coordinates are equivalent. The manifolds, not linear spaces, form an adequate framework for the modern nonlinear analysis.

1.2 Vector fields on smooth manifolds

The tangent space to a smooth manifold at a point is a linear approximation of the manifold in the neighborhood of this point.

Definition 1.5. Let M be a smooth k-dimensional submanifold of $\mathbb{R}^n$ and $x \in M$ its point. Then the *tangent space* to M at the point x is a k- dimensional linear subspace

$$T_x M \subset \mathbb{R}^n$$

defined as follows for cases (1)–(3) of Definition 1.1:

(1) $T_x M = \operatorname{Ker} \left.\dfrac{dF}{dx}\right|_x,$

(2) $T_x M = \operatorname{Im} \left.\dfrac{df}{dx}\right|_0,$

(3) $T_x M = \left(\left.\dfrac{d\Phi}{dx}\right|_x \right)^{-1} \mathbb{R}^k.$

Remark. The tangent space is a coordinate-invariant object since smooth changes of variables lead to linear transformations of the tangent space.

In an abstract way, the tangent space to a manifold at a point is the set of velocity vectors to all smooth curves in the manifold that start from this point.

Definition 1.6. Let $\gamma(\cdot)$ be a smooth curve in a smooth manifold M starting from a point $x \in M$:

$$\gamma \,:\, (-\varepsilon, \varepsilon) \to M \text{ a smooth mapping}, \qquad \gamma(0) = x.$$

The *tangent vector*

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

to the curve $\gamma(\cdot)$ at the point x is the equivalence class of all smooth curves in M starting from x and having the same 1-st order Taylor polynomial as $\gamma(\cdot)$, for any coordinate system in a neighborhood of x.

Definition 1.7. The *tangent space* to a smooth manifold M at a point $x \in M$ is the set of all tangent vectors to all smooth curves in M starting at x:

$$T_x M = \left\{ \left.\frac{d\gamma}{dt}\right|_{t=0} \mid \gamma \,:\, (-\varepsilon, \varepsilon) \to M \text{ smooth}, \gamma(0) = x \right\}.$$

Exercise 1.2. Let M be a smooth k-dimensional manifold and $x \in M$. Show that the tangent space $T_x M$ has a natural structure of a linear k-dimensional space.

Definition 1.8. A *smooth vector field* on a smooth manifold M is a smooth mapping

$$x \in M \mapsto V(x) \in T_x M$$

that associates to any point $x \in M$ a tangent vector $V(x)$.

In the sequel we denote by $\operatorname{Vec} M$ the set of all smooth vector fields on a smooth manifold M.

Definition 1.9. A *smooth dynamical system*, or an *ordinary differential equation* (*ODE*), on a smooth manifold M is an equation of the form

$$\frac{dx}{dt} = V(x), \qquad x \in M,$$

or, equivalently,

$$\dot{x} = V(x), \qquad x \in M,$$

where $V(x)$ is a smooth vector field on M. A *solution* to this system is a smooth mapping

$$\gamma : I \to M,$$

where $I \subset \mathbb{R}$ is an interval, such that

$$\frac{d\gamma}{dt} = V(\gamma(t)) \quad \forall\, t \in I.$$

Definition 1.10. Let $\Phi : M \to N$ be a smooth mapping between smooth manifolds M and N. The *differential* of Φ at a point $x \in M$ is a linear mapping

$$D_x\Phi : T_xM \to T_{\Phi(x)}N$$

defined as follows:

$$D_x\Phi \left(\left. \frac{d\gamma}{dt} \right|_{t=0} \right) = \left. \frac{d}{dt} \right|_{t=0} \Phi(\gamma(t)),$$

where

$$\gamma : (-\varepsilon, \varepsilon) \subset \mathbb{R} \to M, \qquad \gamma(0) = x,$$

is a smooth curve in M starting at x.

Now we apply smooth mappings to vector fields. Let $V \in \operatorname{Vec} M$ be a vector field on M and

$$\dot{x} = V(x) \tag{1.1}$$

the corresponding ODE. To find the action of a diffeomorphism

$$\Phi : M \to N, \qquad \Phi : x \mapsto y = \Phi(x)$$

on the vector field $V(x)$, take a solution $x(t)$ of (1.1) and compute the ODE satisfied by the image $y(t) = \Phi(x(t))$:

$$\dot{y}(t) = \frac{d}{dt}\Phi(x(t)) = (D_x\Phi)\,\dot{x}(t) = (D_x\Phi)\,V(x(t)) = (D_{\Phi^{-1}(y)}\Phi)\,V(\Phi^{-1}(y(t))).$$

So the required ODE is

$$\dot{y} = \left(D_{\Phi^{-1}(y)}\Phi\right) V(\Phi^{-1}(y)). \tag{1.2}$$

The right-hand side here is the transformed vector field on N induced by the diffeomorphism Φ:

$$(\Phi_* V)(y) \stackrel{\text{def}}{=} \left(D_{\Phi^{-1}(y)}\Phi\right) V(\Phi^{-1}(y)).$$

The notation Φ_{*x} is used, along with $D_x\Phi$, for differential of a mapping Φ at a point x.

Remark. In general, a smooth mapping Φ induces transformation of tangent vectors, not of vector fields. In order that $D\Phi$ transform vector fields to vector fields, Φ should be a diffeomorphism.

1.3 Smooth differential equations and flows on manifolds

Theorem 1.2. *Consider a smooth ODE*

$$\dot{x} = V(x), \qquad x \in M \subset \mathbb{R}^n, \tag{1.3}$$

on a smooth submanifold M of $\mathbb{R}^n$. For any initial point $x_0 \in M$, there exists a unique solution

$$x(t, x_0), \qquad t \in (a, b), \quad a < 0 < b,$$

of equation (1.3) *with the initial condition*

$$x(0, x_0) = x_0,$$

defined on a sufficiently short interval (a, b). The mapping

$$(t, x_0) \mapsto x(t, x_0)$$

is smooth. In particular, the domain (a, b) of the solution $x(\cdot, x_0)$ can be chosen smoothly depending on x_0.

Proof. We prove the theorem by reduction to its classical analog in $\mathbb{R}^n$.

The statement of the theorem is local. We rectify the submanifold M in the neighborhood of the point x_0:

$$\Phi \,:\, O_{x_0} \subset \mathbb{R}^n \to O_0 \subset \mathbb{R}^n,$$
$$\Phi(O_{x_0} \cap M) = \mathbb{R}^k.$$

Consider the restriction $\varphi = \Phi|_M$. Then a curve $x(t)$ in M is a solution to (1.3) if and only if its image $y(t) = \varphi(x(t))$ in $\mathbb{R}^k$ is a solution to the induced system:

$$\dot{y} = \Phi_* V(y), \qquad y \in \mathbb{R}^k.$$

□

Theorem 1.3. *Let $M \subset \mathbb{R}^n$ be a smooth submanifold and let*

$$\dot{x} = V(x), \qquad x \in \mathbb{R}^n, \tag{1.4}$$

be a system of ODEs in $\mathbb{R}^n$ such that

$$x \in M \;\Rightarrow\; V(x) \in T_x M.$$

Then for any initial point $x_0 \in M$, the corresponding solution $x(t, x_0)$ to (1.4) *with $x(0, x_0) = x_0$ belongs to M for all sufficiently small $|t|$.*

Proof. Consider the restricted vector field:

$$f = V|_M.$$

By the existence theorem for M, the system

$$\dot{x} = f(x), \qquad x \in M,$$

has a solution $x(t, x_0)$, $x(0, x_0) = x_0$, with

$$x(t, x_0) \in M \quad \text{for small } |t|. \tag{1.5}$$

On the other hand, the curve $x(t, x_0)$ is a solution of (1.4) with the same initial condition. Then inclusion (1.5) proves the theorem. □

Definition 1.11. A vector field $V \in \operatorname{Vec} M$ is called *complete*, if for all $x_0 \in M$ the solution $x(t, x_0)$ of the Cauchy problem

$$\dot{x} = V(x), \qquad x(0, x_0) = x_0 \tag{1.6}$$

is defined for all $t \in \mathbb{R}$.

Example 1.1. The vector field $V(x) = x$, $x \in \mathbb{R}$, is complete, but the vector field $V(x) = x^2$, $x \in \mathbb{R}$, is not complete.

Proposition 1.1. *Suppose that there exists $\varepsilon > 0$ such that for any $x_0 \in M$ the solution $x(t, x_0)$ to (1.6) is defined for $t \in (-\varepsilon, \varepsilon)$. Then the vector field $V(x)$ is complete.*

Remark. In this proposition it is required that there exists $\varepsilon > 0$ common for all initial points $x_0 \in M$. In general, ε may be not bounded away from zero for all $x_0 \in M$. E.g., for the vector field $V(x) = x^2$ we have $\varepsilon \to 0$ as $x_0 \to \infty$.

Proof. Suppose that the hypothesis of the proposition is true. Then we can introduce the following family of mappings in M:

$$P^t : M \to M, \qquad t \in (-\varepsilon, \varepsilon),$$
$$P^t : x_0 \mapsto x(t, x_0).$$

$P^t(x_0)$ is the shift of a point $x_0 \in M$ along the trajectory of the vector field $V(x)$ for time t.

By Theorem 1.2, all mappings P^t are smooth. Moreover, the family $\{ P^t \mid t \in (-\varepsilon, \varepsilon) \}$ is a smooth family of mappings.

A very important property of this family is that it forms a local one-parameter group, i.e.,

$$P^t(P^s(x)) = P^s(P^t(x)) = P^{t+s}(x), \quad x \in M, \quad t,\ s,\ t+s \in (-\varepsilon, \varepsilon).$$

Indeed, the both curves in M:

$$t \mapsto P^t(P^s(x)) \quad \text{and} \quad t \mapsto P^{t+s}(x)$$

satisfy the ODE $\dot{x} = V(x)$ with the same initial value $P^0(P^s(x)) = P^{0+s}(x) = P^s(x)$. By uniqueness, $P^t(P^s(x)) = P^{t+s}(x)$. The equality for $P^s(P^t(x))$ is obtained by switching t and s.

So we have the following local group properties of the mappings P^t:

$$\begin{aligned}
&P^t \circ P^s = P^s \circ P^t = P^{t+s}, \qquad t,\ s,\ t+s \in (-\varepsilon, \varepsilon),\\
&P^0 = \mathrm{Id},\\
&P^{-t} \circ P^t = P^t \circ P^{-t} = \mathrm{Id}, \qquad t \in (-\varepsilon, \varepsilon),\\
&P^{-t} = \left(P^t\right)^{-1}, \qquad t \in (-\varepsilon, \varepsilon).
\end{aligned}$$

In particular, all P^t are diffeomorphisms.

Now we extend the group properties of P^t for all $t \in \mathbb{R}$. Any $t \in \mathbb{R}$ can be represented as

$$t = \frac{\varepsilon}{2}K + \tau, \qquad 0 \le \tau < \frac{\varepsilon}{2}, \quad K = 0, \pm 1, \pm 2, \dots .$$

We set

$$P^t \stackrel{\mathrm{def}}{=} P^\tau \circ \underbrace{P^{\varepsilon/2} \circ \cdots \circ P^{\varepsilon/2}}_{K \text{ times}}.$$

Then the curve

$$t \mapsto P^t(x_0), \qquad t \in \mathbb{R},$$

is a solution to Cauchy problem (1.6). □

Definition 1.12. For a complete vector field $V \in \operatorname{Vec} M$, the mapping

$$t \mapsto P^t, \qquad t \in \mathbb{R},$$

is called the *flow* generated by V.

Remark. It is useful to imagine a vector field $V \in \operatorname{Vec} M$ as a field of velocity vectors of a moving liquid in M. Then the flow P^t takes any particle of the liquid from a position $x \in M$ and transfers it for a time $t \in \mathbb{R}$ to the position $P^t(x) \in M$.

Simple sufficient conditions for completeness of a vector field are given in terms of compactness.

Proposition 1.2. *Let $K \subset M$ be a compact subset, and let $V \in \operatorname{Vec} M$. Then there exists $\varepsilon_K > 0$ such that for all $x_0 \in K$ the solution $x(t, x_0)$ to Cauchy problem* (1.6) *is defined for all $t \in (-\varepsilon_K, \varepsilon_K)$.*

Proof. By Theorem 1.2, domain of the solution $x(t, x_0)$ can be chosen continuously depending on x_0. The diameter of this domain has a positive infimum $2\varepsilon_K$ for x_0 in the compact set K. □

Corollary 1.1. *If a smooth manifold M is compact, then any vector field $V \in \operatorname{Vec} M$ is complete.*

Corollary 1.2. *Suppose that a vector field* $V \in \operatorname{Vec} M$ *has a compact support:*

$$\operatorname{supp} V \stackrel{\text{def}}{=} \overline{\{ x \in M \mid V(x) \neq 0 \}} \text{ is compact.}$$

Then V *is complete.*

Proof. Indeed, by Proposition 1.2, there exists $\varepsilon > 0$ such that all trajectories of V starting in $\operatorname{supp} V$ are defined for $t \in (-\varepsilon, \varepsilon)$. But $V|_{M \setminus \operatorname{supp} V} = 0$, and all trajectories of V starting outside of $\operatorname{supp} V$ are constant, thus defined for all $t \in \mathbb{R}$. By Proposition 1.1, the vector field V is complete. □

Remark. If we are interested in the behavior of (trajectories of) a vector field $V \in \operatorname{Vec} M$ in a compact subset $K \subset M$, we can suppose that V is complete. Indeed, take an open neighborhood O_K of K with the compact closure $\overline{O_K}$. We can find a function $a \in C^\infty(M)$ such that

$$a(x) = \begin{cases} 1, & x \in K, \\ 0, & x \in M \setminus O_K. \end{cases}$$

Then the vector field $a(x)V(x) \in \operatorname{Vec} M$ is complete since it has a compact support. On the other hand, in K the vector fields $a(x)V(x)$ and $V(x)$ coincide, thus have the same trajectories.

1.4 Control systems

For dynamical systems, the future $x(t, x_0)$, $t > 0$, is completely determined by the present state $x_0 = x(0, x_0)$. The law of transformation $x_0 \mapsto x(t, x_0)$ is the flow P^t, i.e., dynamics of the system

$$\dot{x} = V(x), \qquad x \in M, \tag{1.7}$$

it is determined by one vector field $V(x)$.

In order to be able to affect dynamics, to control it, we consider a family of dynamical systems

$$\dot{x} = V_u(x), \qquad x \in M, \quad u \in U, \tag{1.8}$$

with a family of vector fields V_u parametrized by a parameter $u \in U$. A system of the form (1.8) is called a *control system*. The variable u is a *control parameter*, and the set U is the *space of control parameters*. A priori we do not impose any restrictions on U, it is an arbitrary set, although, typically U will be a subset of a smooth manifold. The variable x is the *state*, and the manifold M is the *state space* of control system (1.8).

In control theory we can change dynamics of control system (1.8) at any moment of time by changing values of $u \in U$. For any $u \in U$, the corresponding vector field $V_u \in \operatorname{Vec} M$ generates the flow, which is denoted by P_u^t.

A typical problem of control theory is to find the set of points that can be reached from an initial point $x_0 \in M$ by choosing various values of $u \in U$ and switching from one value to another time to time (for dynamical system (1.7), this reachable set is just the semitrajectory $x(t, x_0) = P^t(x_0)$, $t \geq 0$). Suppose that we start from a point $x_0 \in M$ and use the following control strategy for control system (1.8): first we choose some control parameter $u_1 \in U$, then we switch to another control parameter $u_2 \in U$. Which points in M can be reached with such control strategy? With the control parameter u_1, we can reach points of the form

$$\{ P_{u_1}^{t_1}(x_0) \mid t_1 \geq 0 \},$$

and the whole set of reachable points has the form

$$\{ P_{u_2}^{t_2} \circ P_{u_1}^{t_1}(x_0) \mid t_1, t_2 \geq 0 \},$$

a piece of a 2-dimensional surface.

A natural next question is: what points can be reached from x_0 by any kind of control strategies?

Before studying this question, consider a particular control system that gives a simplified model of a car.

Example 1.2. We suppose that the state of a car is determined by the position of its center of mass $x = (x^1, x^2) \in \mathbb{R}^2$ and orientation angle $\theta \in S^1$ relative to the positive direction of the axis x^1. Thus the state space of our system is a nontrivial 3-dimensional manifold, a solid torus

$$M = \{ q = (x, \theta) \mid x \in \mathbb{R}^2, \ \theta \in S^1 \} = \mathbb{R}^2 \times S^1.$$

Suppose that two kinds of motion are possible: we can drive the car forward and backwards with some fixed linear velocity $u_1 \in \mathbb{R}$, and we can turn the car around its center of mass with some fixed angular velocity $u_2 \in \mathbb{R}$. We can combine these two kinds of motion in an admissible way.

The dynamical system that describes the linear motion with a velocity $u_1 \in \mathbb{R}$ has the form

$$\begin{cases} \dot{x}^1 = u_1 \cos\theta, \\ \dot{x}^2 = u_1 \sin\theta, \\ \dot{\theta} = 0. \end{cases} \tag{1.9}$$

Rotation with an angular velocity $u_2 \in \mathbb{R}$ is described as

$$\begin{cases} \dot{x}^1 = 0, \\ \dot{x}^2 = 0, \\ \dot{\theta} = u_2. \end{cases} \tag{1.10}$$

The control parameter $u = (u_1, u_2)$ can take any values in the given subset $U \subset \mathbb{R}^2$. If we write ODEs (1.9) and (1.10) in the vector form:

$$\dot{q} = u_1 V_1(q), \qquad \dot{q} = u_2 V_2(q),$$

where

$$V_1(q) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \qquad V_2(q) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{1.11}$$

then our model reads

$$\dot{q} = V_u(q) = u_1 V_1(q) + u_2 V_2(q), \qquad q \in M, \quad u \in U.$$

This model can be rewritten in the complex form:

$$\begin{aligned} &z = x^1 + ix^2 \in \mathbb{C}, \\ &\dot{z} = u_1 e^{i\theta}, \\ &\dot{\theta} = u_2, \\ &(u_1, u_2) \in U, \qquad (z, \theta) \in \mathbb{C} \times S^1. \end{aligned}$$

Remark. Control system (1.8) is often written in another form:

$$\dot{x} = f(x, u), \qquad x \in M, \quad u \in U.$$

We prefer the notation $V_u(x)$, which stresses that for a fixed $u \in U$, V_u is a single object — a vector field on M.

Now we return to the study of the points reachable by trajectories of a control system from an initial point.

Definition 1.13. The *attainable set* (or *reachable set*) of control system (1.8) with piecewise-constant controls from a point $x_0 \in M$ for a time $t \geq 0$ is defined as follows:

$$\mathcal{A}_{x_0}(t) = \{\, P_{u_k}^{\tau_k} \circ \ldots \circ P_{u_1}^{\tau_1}(x_0) \mid \tau_i \geq 0,\ \sum_{i=1}^{k} \tau_i = t,\ u_i \in U,\ k \in \mathbb{N}\}.$$

The attainable set from x_0 for arbitrary nonnegative time of motion has the form

$$\mathcal{A}_{x_0} = \cup_{t \geq 0} \mathcal{A}_{x_0}(t)$$

For simplicity, consider first the smallest nontrivial space of control parameters consisting of two indices:

$$U = \{1, 2\}$$

(even this simple case shows essential features of the reachability problem). Then the attainable set for arbitrary nonnegative times has the form:

$$\mathcal{A}_{x_0} = \{\, P_2^{\tau_k} \circ P_1^{\tau_{k-1}} \circ \ldots \circ P_2^{\tau_2} \circ P_1^{\tau_1}(x_0) \mid \tau_i \geq 0,\ k \in \mathbb{N}\}.$$

This expression suggests that the attainable set $\mathcal{A}_{x_0}$ depends heavily upon commutator properties of the flows P_1^t and P_2^s.

Consider first the trivial commutative case, i.e., suppose that the flows commute:

$$P_1^t \circ P_2^s = P_2^s \circ P_1^t \quad \forall t,\ s \in \mathbb{R}.$$

Then the attainable set can be evaluated precisely: since

$$P_2^{\tau_k} \circ P_1^{\tau_{k-1}} \circ \ldots \circ P_2^{\tau_2} \circ P_1^{\tau_1} = P_2^{\tau_k+\ldots+\tau_2} \circ P_1^{\tau_{k-1}+\ldots+\tau_1},$$

then

$$\mathcal{A}_{x_0} = \{\, P_2^s \circ P_1^t(x_0) \mid t,\ s \geq 0 \,\}.$$

So in the commutative case the attainable set by two control parameters is a piece of a smooth two-dimensional surface, possibly with singularities. It is easy to see that if the number of control parameters is $k \geq 2$ and the corresponding flows $P_1^{t_1}, \ldots, P_k^{t_k}$ commute, then $\mathcal{A}_{x_0}$ is, in general, a piece of a k-dimensional manifold, and, in particular, $\dim \mathcal{A}_{x_0} \leq k$.

But this commutative case is exceptional and occurs almost never in real control systems.

Example 1.3. In the model of a car considered above the control dynamics is defined by two vector fields (1.11) on the 3-dimensional manifold $M = \mathbb{R}^2_x \times S^1_\theta$. It is obvious that from any initial configuration $q_0 = (x_0, \theta_0) \in M$ we can drive the car to any terminal configuration $q_1 = (x_1, \theta_1) \in M$ by alternating linear motions and rotations (both with fixed velocities). So any point in the 3-dimensional manifold M can be reached by means of 2 vector fields V_1, V_2. This is due to noncommutativity of these fields (i.e., of their flows).

Given an arbitrary pair of vector fields $V_1, V_2 \in \operatorname{Vec} M$, how can one recognize their commuting properties without finding the flows P_1^t, P_2^s explicitly, i.e., without integration of the ODEs $\dot{x} = V_1(x)$, $\dot{x} = V_2(x)$?

If the flows P_1^t, P_2^s commute, then the curve

$$\gamma(s,t) = P_1^{-t} \circ P_2^s \circ P_1^t(x) = P_2^s(x), \qquad t,\ s \in \mathbb{R}, \tag{1.12}$$

does not depend on t. It is natural to suggest that a lower-order term in the Taylor expansion of (1.12) at $t = s = 0$ is responsible for commuting properties of flows of the vector fields V_1, V_2 at the point x. The first-order derivatives

$$\left.\frac{\partial\,\gamma}{\partial\,t}\right|_{s=t=0} = 0, \qquad \left.\frac{\partial\,\gamma}{\partial\,s}\right|_{s=t=0} = V_2(x)$$

are obviously useless, as well as the pure second-order derivatives

$$\left.\frac{\partial^2\gamma}{\partial t^2}\right|_{s=t=0} = 0, \qquad \left.\frac{\partial^2\gamma}{\partial s^2}\right|_{s=t=0} = \left.\frac{\partial}{\partial\,s}\right|_{s=0} V_2(P_2^s(x)).$$

The required derivative should be the mixed second-order one

$$\left.\frac{\partial^2\gamma}{\partial t \partial s}\right|_{s=t=0}.$$

It turns out that this derivative is a tangent vector to M. It is called the *Lie bracket* of the vector fields V_1, V_2 and is denoted by $[V_1, V_2](x)$:

$$[V_1, V_2](x) \stackrel{\text{def}}{=} \left.\frac{\partial^2}{\partial t \partial s}\right|_{t=s=0} P_1^{-t} \circ P_2^s \circ P_1^t(x) \in T_x M. \tag{1.13}$$

The vector field $[V_1, V_2] \in \operatorname{Vec} M$ determines commuting properties of V_1 and V_2 (it is often called *commutator* of vector fields V_1, V_2).

An effective way to compute Lie bracket of vector fields in local coordinates is given in the following statement.

Proposition 1.3. *Let $M = \mathbb{R}^n$ and $V_1, V_2 \in \operatorname{Vec} M$. Then*

$$[V_1, V_2](x) = \frac{d V_2}{d x} V_1(x) - \frac{d V_1}{d x} V_2(x). \tag{1.14}$$

The proof is left to the reader as an exercise.

Another way to define Lie bracket of vector fields V_1, V_2 is to consider the path

$$\gamma(t) = P_2^{-t} \circ P_1^{-t} \circ P_2^t \circ P_1^t(x).$$

Exercise 1.3. Show that

$$\gamma(t) = x + [V_1, V_2](x) t^2 + o(t^2), \qquad t \to 0,$$

i.e., $[V_1, V_2](x)$ is the velocity curve of the C^1 curve $\gamma(\sqrt{t})$. In particular, this proves that $[V_1, V_2](x)$ is indeed a tangent vector to M:

$$[V_1, V_2](x) \in T_x M.$$

Later we will develop an efficient algebraic way to do similar calculations without any coordinates.

In the commutative case, the set of reachable points does not depend on the number of switches of a control strategy used. In the general noncommutative case, on the contrary, the greater number of switches, the more points can be reached.

Suppose that we can move along vector fields $\pm V_1$ and $\pm V_2$. Then, infinitesimally, we can move in the new direction $\pm[V_1, V_2]$, which is in general linearly independent of the initial ones $\pm V_1$, $\pm V_2$. Using the same switching control strategy with the vector fields $\pm V_1$ and $\pm[V_1, V_2]$, we add one more infinitesimal direction of motion $\pm[V_1, [V_1, V_2]]$. Analogously, we can obtain $\pm[V_2, [V_1, V_2]]$. Iterating this procedure with the new vector fields obtained at previous steps, we can have a Lie bracket of arbitrarily high order as an infinitesimal direction of motion with a sufficiently large number of switches.

Example 1.4. Compute the Lie bracket of the vector fields

$$V_1(q) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad V_2(q) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad q = \begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} \in \mathbb{R}^2_{(x_1, x_2)} \times S^1_\theta$$

appearing in the model of a car. Recall that the field V_1 generates the forward motion, and V_2 the counterclockwise rotation of the car. By (1.14), we have

$$\begin{aligned}[V_1, V_2](q) &= \frac{d\,V_2}{d\,q} V_1(q) - \frac{d\,V_1}{d\,q} V_2(q) = - \begin{pmatrix} 0 & 0 & -\sin\theta \\ 0 & 0 & \cos\theta \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{pmatrix}.\end{aligned}$$

The vector field $[V_1, V_2]$ generates the motion of the car in the direction perpendicular to orientation of the car. This is a typical maneuver in parking a car: the sequence of 4 motions with the same small amplitude of the form

motion forward $\to$ rotation counterclockwise $\to$ motion backward $\to$
$\to$ rotation clockwise

results in motion to the right (in the main term).

We show this explicitly by computing the Lie bracket $[V_1, V_2]$ as in Example 1.3:

$$\begin{aligned}P_2^{-t} \circ P_1^{-t} \circ P_2^t \circ P_1^t \begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} &= \begin{pmatrix} x_1 + t(\cos\theta - \cos(\theta + t)) \\ x_2 + t(\sin\theta - \sin(\theta + t)) \\ \theta \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} + t^2 \begin{pmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{pmatrix} + o(t^2), \qquad t \to 0,\end{aligned}$$

and we have once more

$$[V_1, V_2](q) = \begin{pmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{pmatrix}. \tag{1.15}$$

Of course, we can also compute this Lie bracket by definition as in (1.13):

$$\begin{aligned}P_1^{-t} \circ P_2^s \circ P_1^t \begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} &= \begin{pmatrix} x_1 + t(\cos\theta - \cos(\theta + s)) \\ x_2 + t(\sin\theta - \sin(\theta + s)) \\ \theta + s \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} + s \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + ts \begin{pmatrix} \sin\theta \\ -\cos\theta \\ 0 \end{pmatrix} + O(t^2 + s^2)^{3/2}, \quad t, s \to 0,\end{aligned}$$

and the Lie bracket (1.15) follows.

Chapter 2

Elements of Chronological Calculus

We introduce an operator calculus that will allow us to work with nonlinear systems and flows as with linear ones, at least at the formal level. The idea is to replace a nonlinear object, a smooth manifold M, by a linear, although infinite-dimensional one: the commutative algebra of smooth functions on M (for details, see [1], [2]). For basic definitions and facts of functional analysis used in this chapter, one can consult e.g. [14].

2.1 Points, diffeomorphisms, and vector fields

In this section we identify points, diffeomorphisms, and vector fields on the manifold M with functionals and operators on the algebra $C^\infty(M)$.

Addition, multiplication, and product with constants are defined in the algebra $C^\infty(M)$, as usual, pointwise: if $a, b \in C^\infty(M)$, $q \in M$, $\alpha \in \mathbb{R}$, then

$$
\begin{aligned}
(a+b)(q) &= a(q) + b(q), \\
(a \cdot b)(q) &= a(q) \cdot b(q), \\
(\alpha \cdot a)(q) &= \alpha \cdot a(q).
\end{aligned}
$$

Any point $q \in M$ defines a linear functional

$$\widehat{q} : C^\infty(M) \to \mathbb{R}, \qquad \widehat{q}a = a(q), \ a \in C^\infty(M).$$

The functionals $\widehat{q}$ are homomorphisms of the algebras $C^\infty(M)$ and $\mathbb{R}$:

$$
\begin{aligned}
&\widehat{q}(a+b) = \widehat{q}a + \widehat{q}b, \qquad a, \ b \in C^\infty(M), \\
&\widehat{q}(a \cdot b) = (\widehat{q}a) \cdot (\widehat{q}b), \qquad a, \ b \in C^\infty(M), \\
&\widehat{q}(\alpha \cdot a) = \alpha \cdot \widehat{q}a, \qquad \alpha \in \mathbb{R}, \ a \in C^\infty(M).
\end{aligned}
$$

So to any point $q \in M$, there corresponds a nontrivial homomorphism of algebras $\widehat{q} : C^\infty(M) \to \mathbb{R}$. It turns out that there exists an inverse correspondence.

Proposition 2.1. *Let $\varphi : C^\infty(M) \to \mathbb{R}$ be a nontrivial homomorphism of algebras. Then there exists a point $q \in M$ such that $\varphi = \widehat{q}$.*

We prove this proposition in Appendix A.

Remark. Not only the manifold M can be reconstructed as a set from the algebra $C^\infty(M)$. One can recover topology on M from the weak topology in the space of functionals on $C^\infty(M)$:

$$\lim_{n\to\infty} q_n = q \quad \text{if and only if} \quad \lim_{n\to\infty} \widehat{q}_n a = \widehat{q} a \quad \forall a \in C^\infty(M).$$

Moreover, the smooth structure on M is also recovered from $C^\infty(M)$, actually, "by definition": a real function on the set $\{\widehat{q} \mid q \in M\}$ is smooth if and only if it has a form $\widehat{q} \mapsto \widehat{q}a$ for some $a \in C^\infty(M)$.

Any diffeomorphism $P : M \to M$ defines an automorphism of the algebra $C^\infty(M)$:

$$\widehat{P} : C^\infty(M) \to C^\infty(M), \qquad \widehat{P} \in \mathrm{Aut}(C^\infty(M)),$$
$$(\widehat{P}a)(q) = a(P(q)), \qquad q \in M, \quad a \in C^\infty(M),$$

i.e., $\widehat{P}$ acts as a change of variables in a function a. Conversely, any automorphism of $C^\infty(M)$ has such a form.

Proposition 2.2. *Any automorphism $A : C^\infty(M) \to C^\infty(M)$ has a form of $\widehat{P}$ for some $P \in \mathrm{Diff}\, M$.*

Proof. Let $A \in \mathrm{Aut}(C^\infty(M))$. Take any point $q \in M$. Then the composition

$$\widehat{q} \circ A : C^\infty(M) \to \mathbb{R}$$

is a nonzero homomorphism of algebras, thus it has the form $\widehat{q_1}$, $q_1 \in M$. We denote $q_1 = P(q)$ and obtain

$$\widehat{q} \circ A = \widehat{P(q)} = \widehat{q} \circ \widehat{P} \qquad \forall q \in M,$$

i.e.,

$$A = \widehat{P},$$

and P is the required diffeomorphism. □

Now we characterize tangent vectors to M as functionals on $C^\infty(M)$. Tangent vectors to M are velocity vectors to curves in M, and points of M are identified with linear functionals on $C^\infty(M)$; thus we should obtain linear functionals on $C^\infty(M)$, but not homomorphisms into $\mathbb{R}$. To understand, which functionals on $C^\infty(M)$ correspond to tangent vectors to M, take a smooth curve $q(t)$ of points in M. Then the corresponding curve of functionals $\widehat{q}(t) = \widehat{q(t)}$ on $C^\infty(M)$ satisfies the multiplicative rule

$$\widehat{q}(t)(a \cdot b) = \widehat{q}(t)a \cdot \widehat{q}(t)b, \qquad a,\ b \in C^\infty(M).$$

We differentiate this equality at $t = 0$ and obtain that the velocity vector to the curve of functionals

$$\xi \stackrel{\text{def}}{=} \left.\frac{d\,\widehat{q}}{d\,t}\right|_{t=0}, \qquad \xi : C^\infty(M) \to \mathbb{R},$$

satisfies the Leibniz rule:

$$\xi(ab) = \xi(a)b(q(0)) + a(q(0))\xi(b).$$

Consequently, to each tangent vector $v \in T_qM$ we should put into correspondence a linear functional

$$\xi : C^\infty(M) \to \mathbb{R}$$

such that

$$\xi(ab) = (\xi a)b(q) + a(q)(\xi b), \qquad a,\ b \in C^\infty(M). \tag{2.1}$$

But there is a linear functional $\xi = \widehat{v}$ naturally related to any tangent vector $v \in T_qM$, the directional derivative along v:

$$\widehat{v}a = \left.\frac{d}{d\,t}\right|_{t=0} a(q(t)), \qquad q(0) = q, \quad \dot{q}(0) = v,$$

and such functional satisfies Leibniz rule (2.1).

Now we show that this rule characterizes exactly directional derivatives.

Proposition 2.3. *Let $\xi : C^\infty(M) \to \mathbb{R}$ be a linear functional that satisfies Leibniz rule* (2.1) *for some point $q \in M$. Then $\xi = \widehat{v}$ for some tangent vector $v \in T_qM$.*

Proof. Notice first of all that any functional ξ that meets Leibniz rule (2.1) is local, i.e., it depends only on values of functions in an arbitrarily small neighborhood O_q of the point q:

$$\tilde{a}|_{O_q} = a \quad \Rightarrow \quad \xi\tilde{a} = \xi a, \qquad a,\ \tilde{a} \in C^\infty(M).$$

So the statement of the proposition is local, and we prove it in coordinates.

Let $(x_1, \dots, x_n)$ be local coordinates on M centered at the point q. We have to prove that there exist $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ such that

$$\xi = \sum_{i=1}^n \alpha_i \left.\frac{\partial}{\partial\,x_i}\right|_0 .$$

First of all,

$$\xi(1) = \xi(1 \cdot 1) = (\xi 1) \cdot 1 + 1 \cdot (\xi 1) = 2\xi(1),$$

thus $\xi(1) = 0$. By linearity, $\xi(\text{const}) = 0$.

In order to find the action of ξ on an arbitrary smooth function, we expand it by Hadamard Lemma:

$$a(x) = a(0) + \sum_{i=1}^{n} \int_0^1 \frac{\partial a}{\partial x_i}(tx) x_i \, dt = a(0) + \sum_{i=1}^{n} b_i(x) x_i,$$

where

$$b_i(x) = \int_0^1 \frac{\partial a}{\partial x_i}(tx)\, dt$$

are smooth functions. Now

$$\xi a = \sum_{i=1}^{n} \xi(b_i x_i) = \sum_{i=1}^{n} ((\xi b_i) x_i(0) + b_i(0)(\xi x_i)) = \sum_{i=1}^{n} \alpha_i \frac{\partial a}{\partial x_i}(0),$$

where we denote $\alpha_i = \xi x_i$ and make use of the equality $b_i(0) = \frac{\partial a}{\partial x_i}(0)$. □

So tangent vectors $v \in T_q M$ can be identified with directional derivatives $\widehat{v} : C^\infty(M) \to \mathbb{R}$, i.e., linear functionals that meet Leibniz rule (2.1).

Now we characterize vector fields on M. A smooth vector field on M is a family of tangent vectors $v_q \in T_q M$, $q \in M$, such that for any $a \in C^\infty(M)$ the mapping $q \mapsto v_q a$, $q \in M$, is a smooth function on M.

To a smooth vector field $V \in \operatorname{Vec} M$ there corresponds a linear operator

$$\widehat{V} : C^\infty(M) \to C^\infty(M)$$

that satisfies the Leibniz rule

$$\widehat{V}(ab) = (\widehat{V} a) b + a(\widehat{V} b), \qquad a, \ b \in C^\infty(M),$$

the directional derivative (Lie derivative) along V.

A linear operator on an algebra meeting the Leibniz rule is called a *derivation* of the algebra, so the Lie derivative $\widehat{V}$ is a derivation of the algebra $C^\infty(M)$. We show that the correspondence between smooth vector fields on M and derivations of the algebra $C^\infty(M)$ is invertible.

Proposition 2.4. *Any derivation of the algebra $C^\infty(M)$ is the directional derivative along some smooth vector field on M.*

Proof. Let $D : C^\infty(M) \to C^\infty(M)$ be a derivation. Take any point $q \in M$. We show that the linear functional

$$d_q \stackrel{\text{def}}{=} \widehat{q} \circ D : C^\infty(M) \to \mathbb{R}$$

is a directional derivative at the point q, i.e., satisfies Leibniz rule (2.1):

$$\begin{aligned} d_q(ab) &= \widehat{q}(D(ab)) = \widehat{q}((Da)b + a(Db)) = \widehat{q}(Da)b(q) + a(q)\widehat{q}(Db) = \\ & (d_q a) b(q) + a(q)(d_q b), \qquad a, \ b \in C^\infty(M). \end{aligned}$$

□

So we can identify points $q \in M$, diffeomorphisms $P \in \operatorname{Diff} M$, and vector fields $V \in \operatorname{Vec} M$ with nontrivial homomorphisms $\widehat{q} : C^\infty(M) \to \mathbb{R}$, automorphisms $\widehat{P} : C^\infty(M) \to C^\infty(M)$, and derivations $\widehat{V} : C^\infty(M) \to C^\infty(M)$ respectively.

For example, we can write a point $P(q)$ in the operator notation as $\widehat{q} \circ \widehat{P}$. Moreover, in the sequel we omit hats and write $q \circ P$. This does not cause ambiguity: if q is to the right of P, then q is a point, P a diffeomorphism, and $P(q)$ is the value of the diffeomorphism P at the point q. And if q is to the left of P, then q is a homomorphism, P an automorphism, and $q \circ P$ a homomorphism of $C^\infty(M)$. Similarly, $V(q) \in T_qM$ is the value of the vector field V at the point q, and $q \circ V : C^\infty(M) \to \mathbb{R}$ is the directional derivative along the vector $V(q)$.

2.2 Seminorms and $C^\infty(M)$-topology

We introduce seminorms and topology on the space $C^\infty(M)$.

By Whitney's Theorem, a smooth manifold M can be properly embedded into a Euclidean space $\mathbb{R}^N$ for sufficiently large N. Denote by h_i, $i = 1, \dots, N$, the smooth vector field on M that is the orthogonal projection from $\mathbb{R}^N$ to M of the constant basis vector field $\frac{\partial}{\partial x_i} \in \operatorname{Vec}(\mathbb{R}^N)$. So we have N vector fields $h_1, \dots, h_N \in \operatorname{Vec} M$ that span the tangent space T_qM at each point $q \in M$.

We define the family of seminorms $\|\cdot\|_{s,K}$ on the space $C^\infty(M)$ in the following way:

$$\|a\|_{s,K} = \sup \left\{ |h_{i_1} \circ \cdots \circ h_{i_l} a(q)| \mid q \in K,\ 1 \le i_1, \dots, i_l \le N,\ 1 \le l \le s \right\}, \qquad a \in C^\infty(M), \quad s \ge 0, \quad K \Subset M,$$

This family of seminorms defines a topology on $C^\infty(M)$. A local base of this topology is given by the subsets

$$\left\{ a \in C^\infty(M) \mid \|a\|_{n,K_n} < \frac{1}{n} \right\}, \qquad n \in \mathbb{N},$$

where K_n, $n \in \mathbb{N}$, is a chained system of compacta that cover M: $K_n \subset K_{n+1}$, $\cup_{n=1}^\infty K_n = M$.

This topology on $C^\infty(M)$ does not depend on embedding of M into $\mathbb{R}^N$. It is called the *topology of uniform convergence of all derivatives on compacta*, or just *$C^\infty(M)$-topology*. This topology turns $C^\infty(M)$ into a Fréchet space (a complete, metrizable, locally convex topological vector space).

A sequence of functions $a_k \in C^\infty(M)$ converges to $a \in C^\infty(M)$ as $k \to \infty$ if and only if

$$\lim_{k \to \infty} \|a_k - a\|_{s,K} = 0 \quad \forall\ s \ge 0,\ K \Subset M.$$

For vector fields $V \in \operatorname{Vec} M$, we define the seminorms

$$\|V\|_{s,K} = \sup \left\{ \|Va\|_{s,K} \mid \|a\|_{s+1,K} = 1 \right\}, \qquad s \ge 0, \quad K \Subset M.$$

One can prove estimates of the action of a vector field $V \in \operatorname{Vec} M$ and a diffeomorphism $P \in \operatorname{Diff} M$ on a function $a \in C^\infty(M)$:

$$\begin{aligned} \|Va\|_{s,K} &\le C^1_{s,V}\|a\|_{s+1,K}, \\ \|Pa\|_{s,K} &\le C^2_{s,P}\|a\|_{s,P(K)}, \qquad s \ge 0, \quad K \Subset M. \end{aligned}$$

Thus vector fields and diffeomorphisms are linear continuous operators on the topological vector space $C^\infty(M)$.

2.3 Families of functionals and operators

In the sequel we will often consider one-parameter families of points, diffeomorphisms, and vector fields that satisfy various regularity properties (e.g. differentiability or absolute continuity) with respect to the parameter. Since we treat points as functionals, and diffeomorphisms and vector fields as operators on $C^\infty(M)$, we can introduce regularity properties for them in the weak sense, via the corresponding properties for one-parameter families of functions

$$t \mapsto a_t, \qquad a_t \in C^\infty(M), \quad t \in \mathbb{R}.$$

So we start from definitions for families of functions.

Continuity and *differentiability* of a family of functions a_t w.r.t. parameter t are defined in a standard way since $C^\infty(M)$ is a topological vector space. A family a_t is called *measurable* w.r.t. t if the real function $t \mapsto a_t(q)$ is measurable for any $q \in M$. A measurable family a_t is called *locally integrable* if

$$\int_{t_0}^{t_1} \|a_t\|_{s,K}\, dt < \infty \qquad \forall\, s \ge 0, \quad K \Subset M, \quad t_0,\ t_1 \in \mathbb{R}.$$

A family a_t is called *absolutely continuous* w.r.t. t if

$$a_t = a_{t_0} + \int_{t_0}^{t} b_\tau\, d\tau$$

for some locally integrable family of functions b_t. A family a_t is called *Lipshizian* w.r.t. t if

$$\|a_t - a_\tau\|_{s,K} \le C_{s,K}|t-\tau| \qquad \forall s \ge 0, \quad K \Subset M, \quad t,\ \tau \in \mathbb{R},$$

and *locally bounded* w.r.t. t if

$$\|a_t\|_{s,K} \le C_{s,K,I}, \qquad \forall\, s \ge 0, \quad K \Subset M, \quad I \Subset \mathbb{R}, \quad t \in I,$$

where $C_{s,K}$ and $C_{s,K,I}$ are some constants depending on s, K, and I.

Now we can define regularity properties of families of functionals and operators on $C^\infty(M)$. A family of linear functionals or linear operators on $C^\infty(M)$

$$t \mapsto A_t, \qquad t \in \mathbb{R},$$

has some property (i.e., is *continuous*, *differentiable*, *measurable*, *locally integrable*, *absolutely continuous*, *Lipschizian*, *locally bounded* w.r.t. t) if the family

$$t \mapsto A_t a, \qquad t \in \mathbb{R},$$

has the same property for any $a \in C^\infty(M)$.

A locally integrable w.r.t. t family of vector fields

$$t \mapsto V_t, \qquad V_t \in \operatorname{Vec} M, \quad t \in \mathbb{R},$$

is called a *nonautonomous vector field*, or simply a *field*, on M. An absolutely continuous w.r.t. t family of diffeomorphisms

$$t \mapsto P^t, \qquad P^t \in \operatorname{Diff} M, \quad t \in \mathbb{R},$$

is called a *flow* on M. So, for a nonautonomous vector field V_t, the family of functions $t \mapsto V_t a$ is locally integrable for any $a \in C^\infty(M)$. Similarly, for a flow P^t, the family of functions $(P^t a)(q) = a(P^t(q))$ is absolutely continuous w.r.t. t for any $a \in C^\infty(M)$.

Integrals of measurable locally integrable families, and derivatives of differentiable families are also defined in the weak sense:

$$\int_{t_0}^{t_1} A_t \, dt \ : a \mapsto \int_{t_0}^{t_1} (A_t a) \, dt, \qquad a \in C^\infty(M),$$

$$\frac{d}{dt} A_t \ : a \mapsto \frac{d}{dt} (A_t a), \qquad a \in C^\infty(M).$$

One can show that if A_t and B_t are continuous families, which are differentiable at t_0, then the family $A_t \circ B_t$ is continuous, moreover, differentiable at t_0, and satisfies the Leibniz rule:

$$\left.\frac{d}{dt}\right|_{t_0} (A_t \circ B_t) = \left(\left.\frac{d}{dt}\right|_{t_0} A_t \right) \circ B_{t_0} + A_{t_0} \circ \left(\left.\frac{d}{dt}\right|_{t_0} B_t \right),$$

see the proof in Appendix A.

If families A_t and B_t are absolutely continuous, then the composition $A_t \circ B_t$ is absolutely continuous as well. For an absolute continuous family of functions a_t, the family $A_t a_t$ is also absolutely continuous, and the Leibniz rule holds for it as well.

2.4 Chronological exponential

In this section we consider a *nonautonomous ODE* of the form

$$\dot{q} = V_t(q), \qquad q(0) = q_0, \tag{2.2}$$

where V_t is a locally bounded nonautonomous vector field on M, and study the flow determined by this field.

2.4.1 ODEs with discountinous right-hand side

To obtain local solutions to Cauchy problem (2.2) on a manifold M, we reduce it to a Cauchy problem in a Euclidean space. For details about nonautonomous differential equations in $\mathbb{R}^n$ with right-hand side discontinuous in t, see e.g. [7].

Choose local coordinates $x = (x^1, \dots, x^n)$ in a neighborhood O_{q_0} of the point q_0:

$$\Phi : O_{q_0} \subset M \to O_{x_0} \subset \mathbb{R}^n, \qquad \Phi : q \mapsto x,$$
$$\Phi(q_0) = x_0.$$

In these coordinates, the field V_t reads

$$(\Phi_* V_t)(x) = \widetilde{V}_t(x) = \sum_{i=1}^n v_i(t,x)\frac{\partial}{\partial x^i}, \qquad x \in O_{x_0}, \quad t \in \mathbb{R}, \tag{2.3}$$

and problem (2.2) takes the form

$$\dot{x} = \widetilde{V}_t(x), \quad x(0) = x_0, \qquad x \in O_{x_0} \subset \mathbb{R}^n. \tag{2.4}$$

Since the nonautonomous vector field $V_t \in \operatorname{Vec} M$ is locally bounded, the components $v_i(t,x)$, $i = 1, \dots, n$, of its coordinate representation (2.3) are:

(1) measurable and locally integrable w.r.t. t for any fixed $x \in O_{x_0}$,

(2) smooth w.r.t. x for any fixed $t \in \mathbb{R}$,

(3) differentiable in x with locally bounded partial derivatives:

$$\left| \frac{\partial v_i}{\partial x}(t,x) \right| \le C_{I,K}, \qquad t \in I \Subset \mathbb{R},\ x \in K \Subset O_{x_0},\ i = 1, \dots, n.$$

By the classical Carathéodory Theorem, Cauchy problem (2.4) has a unique solution, i.e., a vector-function $x(t, x_0)$, Lipshizian w.r.t. t and smooth w.r.t. x_0, and such that:

(1) ODE (2.4) is satisfied for almost all t,

(2) initial condition holds: $x(0, x_0) = x_0$.

Then the pull-back of this solution from $\mathbb{R}^n$ to M

$$q(t, q_0) = \Phi^{-1}(x(t, x_0)),$$

is a solution to problem (2.2) in M. The mapping $q(t, q_0)$ iz Lipshizian w.r.t. t and smooth w.r.t. q_0, it satisfies almost everywhere the ODE and the initial condition in (2.2).

For any $q_0 \in M$, the solution $q(t, q_0)$ to Cauchy problem (2.2) can be continued to a maximal interval $t \in J_{q_0} \subset \mathbb{R}$ containing the origin and depending on q_0.

We will assume that the solutions $q(t, q_0)$ are defined for all $q_0 \in M$ and all $t \in \mathbb{R}$, i.e., $J_{q_0} = \mathbb{R}$ for any $q_0 \in M$. Then the nonautonomous field V_t is called *complete*. This holds, e.g., when all the fields V_t, $t \in \mathbb{R}$, vanish outside of a common compactum in M (in this case we say that the nonautonomous vector field V_t has a *compact support*).

2.4.2 Definition of the right chronological exponential

Equation (2.2) rewritten as a linear equation for Lipshizian w.r.t. t families of functionals on $C^\infty(M)$:

$$\dot{q}(t) = q(t) \circ V_t, \qquad q(0) = q_0, \tag{2.5}$$

is satisfied for the family of functionals

$$q(t, q_0) \,:\, C^\infty(M) \to \mathbb{R}, \qquad q_0 \in M, \quad t \in \mathbb{R}.$$

We prove later that this Cauchy problem has no other solutions (see Proposition 2.5). Thus the flow defined as

$$P^t \;:\; q_0 \mapsto q(t, q_0) \tag{2.6}$$

is a unique solution of the operator Cauchy problem

$$\dot{P}^t = P^t \circ V_t, \qquad P^0 = \mathrm{Id}, \tag{2.7}$$

(where Id is the identity operator) in the class of Lipshizian flows on M. The flow P^t determined in (2.6) is called the *right chronological exponential* of the field V_t and is denoted as

$$P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau.$$

Now we develop an asymptotic series for the chronological exponential, which justifies such a notation.

2.4.3 Formal series expansion

We rewrite differential equation in (2.5) as an integral one:

$$q(t) = q_0 + \int_0^t q(\tau) \circ V_\tau \, d\tau \tag{2.8}$$

then substitute this expression for $q(t)$ into the right-hand side

$$\begin{aligned} &= q_0 + \int_0^t \left(q_0 + \int_0^{\tau_1} q(\tau_2) \circ V_{\tau_2} \, d\tau_2 \right) \circ V_{\tau_1} \, d\tau_1 \\ &= q_0 \circ \left(\mathrm{Id} + \int_0^t V_\tau \, dt \right) + \iint_{0 \le \tau_2 \le \tau_1 \le t} q(\tau_2) \circ V_{\tau_2} \circ V_{\tau_1} \, d\tau_2 \, d\tau_1 , \end{aligned}$$

repeat this procedure iteratively, and obtain the decomposition:

$$\begin{aligned}q(t) &= q_0 \circ \left(\mathrm{Id} + \int_0^t V_\tau \, d\tau + \iint_{0 \le \tau_2 \le \tau_1 \le t} V_{\tau_2} \circ V_{\tau_1} \, d\tau_2 \, d\tau_1 + \ldots + \right. \\ & \left. \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 \right) + \\ & \int \cdots \int_{0 \le \tau_{n+1} \le \ldots \le \tau_1 \le t} q(\tau_{n+1}) \circ V_{\tau_{n+1}} \circ \cdots \circ V_{\tau_1} \, d\tau_{n+1} \ldots d\tau_1 .\end{aligned}$$

Purely formally passing to the limit $n \to \infty$, we obtain a formal series for the solution $q(t)$ to problem (2.5):

$$q_0 \circ \left(\mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 \right),$$

thus for the solution P^t to problem (2.7):

$$\mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 . \tag{2.9}$$

We obtained the previous series expansion under the condition $t > 0$, although the chronological exponential is defined for all values of t. But the flow P^{-t} is a solution to the Cauchy problem

$$\frac{d\,P^{-t}}{d\,t} = P^{-t} \circ (-V_{-t}), \qquad P^{-t}\big|_{t=0} = \mathrm{Id},$$

thus

$$P^{-t} = \overrightarrow{\exp} \int_0^t (-V_{-\tau}) \, d\tau .$$

So

$$\overrightarrow{\exp} \int_0^t V_\tau \, d\tau = \overrightarrow{\exp} \int_0^{-t} (-V_{-\tau}) \, d\tau, \qquad t < 0,$$

whence one can easily obtain the series expansion for the flow $\overrightarrow{\exp} \int_0^t V_\tau \, d\tau$, $t < 0$:

$$\mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le -t} (-V_{\tau_n}) \circ \cdots \circ (-V_{\tau_1}) \, d\tau_n \ldots d\tau_1 .$$

This series is similar to (2.9), so in the sequel we restrict ourselves by the study of the case $t > 0$.

2.4.4 Estimates and convergence of the series

Unfortunately, these series never converge on $C^\infty(M)$ in the weak sense (if $V_t \not\equiv 0$): there always exists a smooth function on M, on which they diverge.

Although, one can show that (2.9) is an asymptotic series for the chronological exponential $P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$. There holds the following bound of the remainder term: denote the m-th partial sum of series (2.9) as

$$S_m(t) = \mathrm{Id} + \sum_{n=1}^{m-1} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 ,$$

then for any $a \in C^\infty(M)$, $s \ge 0$, $K \Subset M$

$$\begin{aligned} &\left\| \left(\overrightarrow{\exp} \int_0^t V_\tau \, d\tau - S_m(t) \right) a \right\|_{s,K} \\ &\qquad \le C_1 e^{C_2 \int_0^t \|V_\tau\|_{s,K'} \, d\tau} \left(\int_0^t \|V_\tau\|_{s+m-1,K'} \, d\tau \right)^m \|a\|_{s+m,K'} \\ &\qquad = O(t^m), \qquad t \to 0, \end{aligned} \tag{2.10}$$

where $K' \Subset M$ is some neighborhood of compactum K. Moreover, it follows from estimate (2.10) that

$$\left\| \left(\overrightarrow{\exp} \int_0^t \varepsilon V_\tau \, d\tau - S_m^\varepsilon(t) \right) a \right\|_{s,K} = O(\varepsilon^m), \qquad \varepsilon \to 0,$$

where $S_m^\varepsilon(t)$ is the m-th partial sum of series (2.9) for the field εV_t.

Thus we have an asymptotic series expansion:

$$\overrightarrow{\exp} \int_0^t V_\tau \, d\tau \approx \mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 . \tag{2.11}$$

In the sequel we will use terms of the zeroth, first, and second orders of the series obtained:

$$\overrightarrow{\exp} \int_0^t V_\tau \, d\tau \approx \mathrm{Id} + \int_0^t V_\tau \, d\tau + \iint_{0 \le \tau_2 \le \tau_1 \le t} V_{\tau_2} \circ V_{\tau_1} \, d\tau_2 \, d\tau_1 + \cdots .$$

We prove that the asymptotic series converges to the chronological exponential on linear functions in the case when $M = \mathbb{R}^n$ and V_t is a linear vector field, i.e., the space of linear functions $\mathbb{R}^{n*} \subset C^\infty(\mathbb{R}^n)$ is invariant for the family of operators V_t. Fix any norm $\| \cdot \|$ on $\mathbb{R}^{n*}$. For any t the linear operator on a finite-dimensional vector space $V_t|_{\mathbb{R}^{n*}}$ is bounded. The norm of the operator $V_t|_{\mathbb{R}^{n*}}$ is defined as usual:

$$\|V_t\| = \sup \{ \|V_t a\| \mid a \in \mathbb{R}^{n*}, \ \|a\| \le 1 \} .$$

We apply operator series (2.11) to any $a \in \mathbb{R}^{n*}$ and bound terms of the series obtained:

$$\left(\mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \ldots d\tau_1 \right) a. \tag{2.12}$$

We have

$$\left\| \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \, \ldots \, d\tau_1 \; a \right\|$$
$$\le \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} \|V_{\tau_n}\| \cdot \cdots \cdot \|V_{\tau_1}\| \, d\tau_n \, \ldots \, d\tau_1 \cdot \|a\|$$

by symmetry w.r.t. permutations of indices $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$

$$= \int \cdots \int_{0 \le \tau_{\sigma(n)} \le \ldots \le \tau_{\sigma(1)} \le t} \|V_{\tau_n}\| \cdot \cdots \cdot \|V_{\tau_1}\| \, d\tau_n \, \ldots \, d\tau_1 \cdot \|a\|$$

passing to the integral over cube

$$= \frac{1}{n!} \int_0^t \cdots \int_0^t \|V_{\tau_n}\| \cdot \cdots \cdot \|V_{\tau_1}\| \, d\tau_n \, \ldots \, d\tau_1 \cdot \|a\|$$
$$= \frac{1}{n!} \left(\int_0^t \|V_\tau\| \, d\tau \right)^n \cdot \|a\|.$$

So series (2.12) is majorized by the exponential series, thus the operator series (2.11) converges on $\mathbb{R}^{n*}$.

Series (2.12) can be differentiated termwise, thus it satisfies the same ODE as the function $P^t a$:

$$\dot{a}_t = V_t a_t, \qquad a_0 = a.$$

Consequently,

$$P^t a = \left(\mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} V_{\tau_n} \circ \cdots \circ V_{\tau_1} \, d\tau_n \, \ldots \, d\tau_1 \right) a.$$

So in the linear case the asymptotic series converges, moreover, there holds the bound

$$\|P^t a\| \le e^{\int_0^t \|V_\tau\| \, d\tau} \|a\|, \qquad a \in \mathbb{R}^{n*}.$$

Notice that this bound and convergence hold not only on $\mathbb{R}^{n*}$, but also on any invariant for V_t subspace of $C^\infty(\mathbb{R}^n)$ where V_t is bounded with respect to some norm (and V_t is not necessarily linear). Moreover, the bound and convergence hold not only for a locally bounded, but also for integrable on $[0, t]$ vector fields:

$$\int_0^t \|V_\tau\| \, d\tau < \infty.$$

If M and V_t are real analytic, then series (2.11) converges for sufficiently small t (we leave this without proof).

2.4.5 Left chronological exponential

Consider the inverse operator $Q^t = (P^t)^{-1}$ to the right chronological exponential $P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$. We find an ODE for the flow Q^t by differentiation of the identity

$$P^t \circ Q^t = \mathrm{Id}\,.$$

Leibniz rule yields

$$\dot{P}^t \circ Q^t + P^t \circ \dot{Q}^t = 0,$$

thus, in view of ODE (2.7) for the flow P^t,

$$P^t \circ V_t \circ Q^t + P^t \circ \dot{Q}^t = 0.$$

We multiply this equality by Q^t from the left and obtain

$$V_t \circ Q^t + \dot{Q}^t = 0.$$

That is, the flow Q^t is a solution of the Cauchy problem

$$\frac{d}{dt} Q^t = -V_t \circ Q^t, \qquad Q^0 = \mathrm{Id}, \tag{2.13}$$

which is dual to Cauchy problem (2.7) for P^t. The flow Q^t is called the *left chronological exponential* and is denoted as

$$Q^t = \overleftarrow{\exp} \int_0^t (-V_\tau)\, d\tau.$$

We find an asymptotic expansion for the left chronological exponential in the same way as for the right one, by successive substitutions into the right-hand side:

$$\begin{aligned}
Q^t &= \mathrm{Id} + \int_0^t (-V_\tau) \circ Q^\tau \, d\tau \\
&= \mathrm{Id} + \int_0^t (-V_\tau)\, d\tau + \iint_{0 \le \tau_2 \le \tau_1 \le t} (-V_{\tau_1}) \circ (-V_{\tau_2}) \circ Q^{\tau_2} \, d\tau_2 \, d\tau_1 = \cdots \\
&= \mathrm{Id} + \sum_{n=1}^{m-1} \int \cdots \int_{0 \le \tau_n \le \cdots \le \tau_1 \le t} (-V_{\tau_1}) \circ \cdots \circ (-V_{\tau_n})\, d\tau_n \, \ldots \, d\tau_1 \\
&\quad + \int \cdots \int_{0 \le \tau_m \le \cdots \le \tau_1 \le t} (-V_{\tau_1}) \circ \cdots \circ (-V_{\tau_m}) \circ Q^{\tau_m} \, d\tau_m \, \ldots \, d\tau_1
\end{aligned}$$

For the left chronological exponential holds an estimate of the remainder term as (2.10) for the right one, and the series obtained is asymptotic:

$$\begin{aligned}
&\overleftarrow{\exp} \int_0^t (-V_\tau)\, d\tau \\
&\qquad \approx \mathrm{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \ldots \le \tau_1 \le t} (-V_{\tau_1}) \circ \cdots \circ (-V_{\tau_n})\, d\tau_n \, \ldots \, d\tau_1 .
\end{aligned}$$

Remarks. (1) Notice that the reverse arrow in the left chronological exponential $\overleftarrow{\exp}$ corresponds to the reverse order of the operators $(-V_{\tau_1}) \circ \cdots \circ (-V_{\tau_n})$, $\tau_n \leq \ldots \leq \tau_1$.

(2) The right and left chronological exponentials satisfy the corresponding differential equations:

$$\frac{d}{dt} \overrightarrow{\exp} \int_0^t V_\tau \, d\tau = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau \circ V_t,$$
$$\frac{d}{dt} \overleftarrow{\exp} \int_0^t (-V_\tau) \, d\tau = -V_t \circ \overleftarrow{\exp} \int_0^t (-V_\tau) \, d\tau.$$

The directions of arrows correlate with the direction of appearance of operators V_t, $-V_t$ in the right-hand side of these ODEs.

(3) If the initial value is prescribed at a moment of time $t_0 \neq 0$, then the lower limit of integrals in the chronological exponentials is t_0.

(4) There holds the following obvious rule for composition of flows:

$$\overrightarrow{\exp} \int_{t_0}^{t_1} V_\tau \, d\tau \circ \overrightarrow{\exp} \int_{t_1}^{t_2} V_\tau \, d\tau = \overrightarrow{\exp} \int_{t_0}^{t_2} V_\tau \, d\tau.$$

Exercise 2.1. Prove that

$$\overrightarrow{\exp} \int_{t_0}^{t_1} V_\tau \, d\tau = \left(\overrightarrow{\exp} \int_{t_1}^{t_0} V_\tau \, d\tau \right)^{-1} = \overleftarrow{\exp} \int_{t_1}^{t_0} (-V_\tau) \, d\tau. \tag{2.14}$$

2.4.6 Uniqueness for functional and operator ODEs

We saw that equation (2.5) for Lipshizian families of functionals has a solution $q(t) = q_0 \circ \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$. We can prove now that this equation has no other solutions.

Proposition 2.5. *Let V_t be a complete locally bounded nonautonomous vector field on M. Then Cauchy problem* (2.5) *has a unique solution in the class of Lipshizian families of functionals on $C^\infty(M)$.*

Proof. Let a Lipshizian family of functionals q_t be a solution to problem (2.5). Then

$$\frac{d}{dt} \left(q_t \circ (P^t)^{-1} \right) = \frac{d}{dt} \left(q_t \circ Q^t \right) = q_t \circ V_t \circ Q^t - q_t \circ V_t \circ Q^t = 0.$$

thus $q_t \circ Q^t \equiv \text{const}$. But $Q_0 = \text{Id}$, consequently, $q_t \circ Q^t \equiv q_0$, hence

$$q_t = q_0 \circ P^t = q_0 \circ \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$$

is a unique solution of Cauchy problem (2.5). □

Similarly, the both operator equations $\dot{P}^t = P^t \circ V_t$ and $\dot{Q}^t = -V_t \circ Q^t$ have no other solutions in addition to the chronological exponentials.

2.4.7 Autonomous vector fields

In the special case of *autonomous vector fields*,

$$V_t \equiv V \in \operatorname{Vec} M,$$

the flow generated by a complete field is called the *exponential* and is denoted as e^{tV}. The asymptotic series for the exponential takes the form

$$e^{tV} \approx \sum_{n=0}^{\infty} \frac{t^n}{n!} V^n = \operatorname{Id} + tV + \frac{t^2}{2} V \circ V + \cdots \ ,$$

i.e, it is the standard exponential series.

The exponential of an autonomous vector field satisfies the ODEs

$$\frac{d}{dt} e^{tV} = e^{tV} \circ V = V \circ e^{tV}, \qquad e^{tV}\big|_{t=0} = \operatorname{Id}.$$

We apply the asymptotic series for exponential to find the Lie bracket of autonomous vector fields $V, W \in \operatorname{Vec} M$. We compute the first nonconstant term in the asymptotic expansion at $t = 0$ of the curve:

$$\begin{aligned}
q(t) &= q \circ e^{tV} \circ e^{tW} \circ e^{-tV} \circ e^{-tW} \\
&= q \circ \left(\operatorname{Id} + tV + \frac{t^2}{2} V^2 + \cdots\right) \circ \left(\operatorname{Id} + tW + \frac{t^2}{2} W^2 + \cdots\right) \\
&\qquad \circ \left(\operatorname{Id} - tV + \frac{t^2}{2} V^2 + \cdots\right) \circ \left(\operatorname{Id} - tW + \frac{t^2}{2} W^2 + \cdots\right) \\
&= q \circ \left(\operatorname{Id} + t(V + W) + \frac{t^2}{2}(V^2 + 2V \circ W + W^2) + \cdots\right) \\
&\qquad \circ \left(\operatorname{Id} - t(V + W) + \frac{t^2}{2}(V^2 + 2V \circ W + W^2) + \cdots\right) \\
&= q \circ \left(\operatorname{Id} + t^2 (V \circ W - W \circ V) + \cdots\right).
\end{aligned}$$

So the Lie bracket of the vector fields as operators (directional derivatives) in $C^\infty(M)$ is

$$[V, W] = V \circ W - W \circ V.$$

This proves the formula in local coordinates: if

$$V = \sum_{i=1}^{n} a_i \frac{\partial}{\partial x_i}, \qquad W = \sum_{i=1}^{n} b_i \frac{\partial}{\partial x_i}, \qquad a_i,\ b_i \in C^\infty(M),$$

then

$$[V, W] = \sum_{i,j=1}^{n} \left(a_j \frac{\partial b_i}{\partial x_j} - b_j \frac{\partial a_i}{\partial x_j} \right) \frac{\partial}{\partial x_i} = \frac{dW}{dx} V - \frac{dV}{dx} W.$$

Similarly,

$$\begin{aligned} q \circ e^{tV} \circ e^{sW} \circ e^{-tV} &= q \circ (\mathrm{Id} + tV + \cdots) \circ (\mathrm{Id} + sW + \cdots) \circ (\mathrm{Id} - tV + \cdots) \\ &= q \circ (\mathrm{Id} + sW + ts[V, W] + \cdots), \end{aligned}$$

and

$$q \circ [V, W] = \left. \frac{\partial^2}{\partial s \partial t} \right|_{s=t=0} q \circ e^{tV} \circ e^{sW} \circ e^{-tV}.$$

2.5 Action of diffeomorphisms on vector fields

We have already found counterparts to points, diffeomorphisms, and vector fields among functionals and operators on $C^\infty(M)$. Now we consider actions of diffeomorphisms on vector fields.

Take a tangent vector $v \in T_q M$ and a diffeomorphism $P \in \mathrm{Diff}\, M$. The tangent vector $P_* v \in T_{P(q)} M$ is the velocity vector of the image of a curve starting from q with the velocity vector v. We claim that

$$P_* v = v \circ P, \qquad v \in T_q M, \quad P \in \mathrm{Diff}\, M, \tag{2.15}$$

as functionals on $C^\infty(M)$. Take a curve

$$q(t) \in M, \qquad q(0) = q, \qquad \left. \frac{d}{dt} \right|_{t=0} q(t) = v,$$

then

$$\begin{aligned} P_* v\, a &= \left. \frac{d}{dt} \right|_{t=0} a(P(q(t))) = \left(\left. \frac{d}{dt} \right|_{t=0} q(t) \right) \circ P a \\ &= v \circ P a, \qquad a \in C^\infty(M). \end{aligned}$$

Now we find expression for $P_* V$, $V \in \mathrm{Vec}\, M$, as a derivation of $C^\infty(M)$. We have

$$\begin{aligned} q \circ P \circ P_* V &= P(q) \circ P_* V = (P_* V)(P(q)) = P_*(V(q)) = V(q) \circ P \\ &= q \circ V \circ P, \quad q \in M, \end{aligned}$$

thus

$$P \circ P_* V = V \circ P,$$

i.e.,

$$P_* V = P^{-1} \circ V \circ P, \qquad P \in \mathrm{Diff}\, M, \ V \in \mathrm{Vec}\, M.$$

So diffeomorphisms act on vector fields as similarities. In particular, diffeomorphisms preserve compositions:

$$P_*(V \circ W) = P^{-1} \circ (V \circ W) \circ P = (P^{-1} \circ V \circ P) \circ (P^{-1} \circ W \circ P) = P_* V \circ P_* W,$$

and Lie brackets of vector fields:

$$P_*[V,W] = P_*(V \circ W - W \circ V) = P_*V \circ P_*W - P_*W \circ P_*V = [P_*V, P_*W].$$

If $B : C^\infty(M) \to C^\infty(M)$ is an automorphism, then the standard algebraic notation for the corresponding similarity is $\operatorname{Ad} B$:

$$(\operatorname{Ad} B)V \stackrel{\text{def}}{=} B \circ V \circ B^{-1}.$$

That is,

$$P_* = \operatorname{Ad} P^{-1}, \qquad P \in \operatorname{Diff} M.$$

Now we find an infinitesimal version of the operator Ad. Let P^t be a flow on M,

$$P^0 = \operatorname{Id}, \qquad \left.\frac{d}{dt}\right|_{t=0} P^t = V \in \operatorname{Vec} M.$$

Then

$$\left.\frac{d}{dt}\right|_{t=0} \left(P^t\right)^{-1} = -V,$$

so

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=0} (\operatorname{Ad} P^t)W &= \left.\frac{d}{dt}\right|_{t=0} (P^t \circ W \circ (P^t)^{-1}) = V \circ W - W \circ V \\ &= [V,W], \qquad W \in \operatorname{Vec} M. \end{aligned}$$

Denote

$$\operatorname{ad} V = \operatorname{ad} \left(\left.\frac{d}{dt}\right|_{t=0} P^t \right) \stackrel{\text{def}}{=} \left.\frac{d}{dt}\right|_{t=0} \operatorname{Ad} P^t,$$

then

$$(\operatorname{ad} V)W = [V,W], \qquad W \in \operatorname{Vec} M.$$

Differentiation of the equality

$$\operatorname{Ad} P^t\,[X,Y] = [\operatorname{Ad} P^t\, X, \operatorname{Ad} P^t\, Y] \qquad X, Y \in \operatorname{Vec} M,$$

at $t = 0$ gives *Jacobi identity* for Lie bracket of vector fields:

$$(\operatorname{ad} V)[X,Y] = [(\operatorname{ad} V)X, Y] + [X, (\operatorname{ad} V)Y],$$

which may also be written as

$$[V,[X,Y]] = [[V,X],Y] + [X,[V,Y]], \qquad V, X, Y \in \operatorname{Vec} M,$$

or, in a symmetric way

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0, \qquad X, Y, Z \in \operatorname{Vec} M. \tag{2.16}$$

The set $\operatorname{Vec} M$ is a vector space with an additional operation — Lie bracket, which has the properties:

(1) bilinearity:

$$[\alpha X + \beta Y, Z] = \alpha[X, Z] + \beta[Y, Z],$$
$$[X, \alpha Y + \beta Z] = \alpha[X, Y] + \beta[X, Z], \qquad X, Y, Z \in \operatorname{Vec} M, \quad \alpha, \beta \in \mathbb{R},$$

(2) skew-symmetry:

$$[X, Y] = -[Y, X], \qquad X, Y \in \operatorname{Vec} M,$$

(3) Jacobi identity (2.16).

In other words, the set $\operatorname{Vec} M$ of all smooth vector fields on a smooth manifold M forms a *Lie algebra*.

Consider the flow $P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$ of a nonautonomous vector field V_t. We find an ODE for the family of operators $\operatorname{Ad} P^t = (P^t)_*^{-1}$ on the Lie algebra $\operatorname{Vec} M$.

$$\begin{aligned} \frac{d}{dt}(\operatorname{Ad} P^t)X &= \frac{d}{dt}\left(P^t \circ X \circ (P^t)^{-1}\right) \\ &= P^t \circ V_t \circ X \circ (P^t)^{-1} - P^t \circ X \circ V_t \circ (P^t)^{-1} \\ &= (\operatorname{Ad} P^t)[V_t, X] = (\operatorname{Ad} P^t) \operatorname{ad} V_t \, X, \qquad X \in \operatorname{Vec} M. \end{aligned}$$

Thus the family of operators $\operatorname{Ad} P^t$ satisfies the ODE

$$\frac{d}{dt} \operatorname{Ad} P^t = (\operatorname{Ad} P^t) \circ \operatorname{ad} V_t \tag{2.17}$$

with the initial condition

$$\operatorname{Ad} P^0 = \operatorname{Id}. \tag{2.18}$$

So the family $\operatorname{Ad} P^t$ is an invertible solution for the Cauchy problem

$$\dot{A}_t = A_t \circ \operatorname{ad} V_t, \quad A_0 = \operatorname{Id}$$

for operators $A_t : \operatorname{Vec} M \to \operatorname{Vec} M$. We can apply the same argument as for the analogous problem (2.7) for flows to derive the asymptotic expansion

$$\begin{aligned} \operatorname{Ad} P^t \approx \operatorname{Id} + \int_0^t \operatorname{ad} V_\tau \, d\tau + \cdots \\ + \int \cdots \int_{0 \le \tau_n \le \cdots \le \tau_1 \le t} \operatorname{ad} V_{\tau_n} \circ \cdots \circ \operatorname{ad} V_{\tau_1} \, d\tau_n \, \ldots \, d\tau_1 + \cdots \end{aligned} \tag{2.19}$$

then prove uniqueness of the solution, and justify the following notation:

$$\overrightarrow{\exp} \int_0^t \operatorname{ad} V_\tau \, d\tau \stackrel{\text{def}}{=} \operatorname{Ad} P^t = \operatorname{Ad}\left(\overrightarrow{\exp} \int_0^t V_\tau \, d\tau\right).$$

Similar identities for the left chronological exponential are

$$\overleftarrow{\exp} \int_0^t \operatorname{ad}(-V_\tau)\, d\tau \overset{\text{def}}{=} \operatorname{Ad} \left(\overleftarrow{\exp} \int_0^t (-V_\tau)\, d\tau \right)$$
$$\approx \operatorname{Id} + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le \tau_n \le \dots \le \tau_1 \le t} (-\operatorname{ad} V_{\tau_1}) \circ \cdots \circ (-\operatorname{ad} V_{\tau_n})\, d\tau_n \, \dots \, d\tau_1.$$

For the asymptotic series (2.19), there holds an estimate of the remainder term similar to estimate (2.10) for the flow P^t. Denote the partial sum

$$T_m = \operatorname{Id} + \sum_{n=1}^{m-1} \int \cdots \int_{0 \le \tau_n \le \dots \le \tau_1 \le t} \operatorname{ad} V_{\tau_n} \circ \cdots \circ \operatorname{ad} V_{\tau_1}\, d\tau_n \, \dots \, d\tau_1,$$

then for any $X \in \operatorname{Vec} M$, $s \ge 0$, $K \Subset M$

$$\left\| \left(\operatorname{Ad} \overrightarrow{\exp} \int_0^t V_\tau \, d\tau - T_m \right) X \right\|_{s,K}$$
$$\le C_3 e^{C_4 \int_0^t \|V_\tau\|_{s+1,K'}\, d\tau} \left(\int_0^t \|V_\tau\|_{s+m,K'}\, d\tau \right)^m \|X\|_{s+m,K'} \tag{2.20}$$
$$= O(t^m), \qquad t \to 0,$$

where $K' \Subset M$ is some neighborhood of compactum K.

For autonomous vector fields, we denote

$$e^{t \operatorname{ad} V} \overset{\text{def}}{=} \operatorname{Ad} e^{tV},$$

thus the family of operators $e^{t \operatorname{ad} V} : \operatorname{Vec} M \to \operatorname{Vec} M$ is the unique solution to the problem

$$\dot{A}_t = A_t \circ \operatorname{ad} V, \qquad A_0 = \operatorname{Id},$$

which admits the asymptotic expansion

$$e^{t \operatorname{ad} V} \approx \operatorname{Id} + t \operatorname{ad} V + \frac{t^2}{2} \operatorname{ad}^2 V + \cdots .$$

In the sequel we will need the following properties of the operator Ad and its infinitesimal version ad.

Exercise 2.2. Let $X, Y \in \operatorname{Vec} M$, $a \in C^\infty(M)$, $P \in \operatorname{Diff} M$, and let V_t be a nonautonomous vector field on M. Prove the equalities:

$$(\operatorname{ad} X)(aY) = (Xa)Y + a(\operatorname{ad} X)Y,$$
$$P \circ \overrightarrow{\exp} \int_0^t V_\tau \, d\tau \circ P^{-1} = \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P\, V_\tau)\, d\tau, \tag{2.21}$$
$$(\operatorname{Ad} P)(aX) = (Pa) \operatorname{Ad} P\, X.$$

2.6 Commutation of flows

Let $V_t \in \operatorname{Vec} M$ be a nonautonomous vector field and $P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$ the corresponding flow. We are interested in the question: under what conditions the flow P^t preserves a vector field $W \in \operatorname{Vec} M$:

$$P^t{}_* W = W \qquad \forall t,$$

or, which is equivalent,

$$(P^t)_*^{-1} W = W \qquad \forall t.$$

Proposition 2.6.

$$P_*^t W = W \quad \forall t \quad \Leftrightarrow \quad [V_t, W] = 0 \quad \forall t.$$

Proof. We have

$$\begin{aligned} \frac{d}{dt} (P_t)_*^{-1} W &= \frac{d}{dt} \operatorname{Ad} P^t W = \left(\frac{d}{dt} \overrightarrow{\exp} \int_0^t \operatorname{ad} V_\tau \, d\tau \right) W \\ &= \left(\overrightarrow{\exp} \int_0^t \operatorname{ad} V_\tau \, d\tau \circ \operatorname{ad} V_\tau \right) W = \left(\overrightarrow{\exp} \int_0^t \operatorname{ad} V_\tau \, d\tau \right) [V_t, W] \\ &= (P^t)_*^{-1} [V_t, W], \end{aligned}$$

thus $(P^t)_*^{-1} W \equiv W$ if and only if $[V_t, W] \equiv 0$. □

In general, flows do not commute, neither for nonautonomous vector fields V_t, W_t:

$$\overrightarrow{\exp} \int_0^{t_1} V_\tau \, d\tau \circ \overrightarrow{\exp} \int_0^{t_2} W_\tau \, d\tau \neq \overrightarrow{\exp} \int_0^{t_2} W_\tau \, d\tau \circ \overrightarrow{\exp} \int_0^{t_1} V_\tau \, d\tau,$$

nor for autonomous vector fields V, W:

$$e^{t_1 V} \circ e^{t_2 W} \neq e^{t_2 W} \circ e^{t_1 V}.$$

In the autonomous case, commutativity of flows is equivalent to commutativity of vector fields:

$$e^{t_1 V} \circ e^{t_2 W} = e^{t_2 W} \circ e^{t_1 V}, \quad t_1, t_2 \in \mathbb{R}, \qquad \Leftrightarrow \qquad [V, W] = 0.$$

We already showed that commutativity of vector fields is necessary for commutativity of flows. Let us prove that it is sufficient. Indeed,

$$\left(\operatorname{Ad} e^{t_1 V} \right) W = e^{t_1 \operatorname{ad} V} W = W.$$

Taking into account equality (2.21), we obtain

$$e^{t_1 V} \circ e^{t_2 W} \circ e^{-t_1 V} = e^{t_2 \left(\operatorname{Ad} e^{t_1 V} \right) W} = e^{t_2 W}.$$

2.7 Variations formula

Consider an ODE of the form

$$\dot q = V_t(q) + W_t(q). \tag{2.22}$$

We think of V_t as an initial vector field and W_t as its perturbation. Our aim is to find a formula for the flow Q^t of the new field $V_t + W_t$ as a perturbation of the flow $P^t = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau$ of the initial field V_t. In other words, we wish to have a decomposition of the form

$$Q^t = \overrightarrow{\exp} \int_0^t (V_\tau + W_\tau) \, d\tau = C_t \circ P^t.$$

We proceed as in the method of variation of parameters; we substitute the previous expression to ODE (2.22):

$$\begin{aligned} \frac{d}{dt} Q^t &= Q^t \circ (V_t + W_t) \\ &= \dot C_t \circ P^t + C_t \circ P^t \circ V_t \\ &= \dot C_t \circ P^t + Q^t \circ V_t, \end{aligned}$$

cancel the common term $Q^t \circ V_t$:

$$Q^t \circ W_t = \dot C_t \circ P^t,$$

and write down the ODE for the unknown flow C_t:

$$\begin{aligned} \dot C_t &= Q^t \circ W_t \circ \left(P^t\right)^{-1} \\ &= C_t \circ P^t \circ W_t \circ \left(P^t\right)^{-1} \\ &= C_t \circ \left(\operatorname{Ad} P^t\right) W_t \\ &= C_t \circ \left(\overrightarrow{\exp} \int_0^t \operatorname{ad} V_\tau \, d\tau \right) W_t, \\ C_0 &= \operatorname{Id}. \end{aligned}$$

This operator Cauchy problem is of the form (2.7), thus it has a unique solution:

$$C_t = \overrightarrow{\exp} \int_0^t \left(\overrightarrow{\exp} \int_0^\tau \operatorname{ad} V_\theta \, d\theta \right) W_\tau \, d\tau.$$

Hence we obtain the required decomposition of the perturbed flow:

$$\overrightarrow{\exp} \int_0^t (V_\tau + W_\tau) \, d\tau = \overrightarrow{\exp} \int_0^t \left(\overrightarrow{\exp} \int_0^\tau \operatorname{ad} V_\theta \, d\theta \right) W_\tau \, d\tau \circ \overrightarrow{\exp} \int_0^t V_\tau \, d\tau. \tag{2.23}$$

This equality is called the *variations formula*. It can be written as follows:

$$\overrightarrow{\exp} \int_0^t (V_\tau + W_\tau)\, d\tau = \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P^\tau)\, W_\tau\, d\tau \;\circ\; P^t.$$

So the perturbed flow is a composition of the initial flow P^t with the flow of the perturbation W_t twisted by P^t.

Now we obtain another form of the variations formula, with the flow P^t to the left of the twisted flow. We have

$$\begin{aligned}
\overrightarrow{\exp} \int_0^t (V_\tau + W_\tau)\, d\tau &= \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P^\tau)\, W_\tau\, d\tau \;\circ\; P^t \\
&= P^t \circ \left(P^t\right)^{-1} \circ\; \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P^\tau)\, W_\tau\, d\tau \circ P^t \\
&= P^t \circ\; \overrightarrow{\exp} \int_0^t \left(\operatorname{Ad} \left(P^t\right)^{-1} \circ \operatorname{Ad} P^\tau\right) W_\tau\, d\tau \\
&= P^t \circ\; \overrightarrow{\exp} \int_0^t \left(\operatorname{Ad} \left(\left(P^t\right)^{-1} \circ P^\tau\right)\right) W_\tau\, d\tau.
\end{aligned}$$

Since

$$\left(P^t\right)^{-1} \circ P^\tau = \overrightarrow{\exp} \int_t^\tau V_\theta\, d\theta,$$

we obtain

$$\begin{aligned}
\overrightarrow{\exp} \int_0^t (V_\tau + W_\tau)\, d\tau &= P^t \circ\; \overrightarrow{\exp} \int_0^t \left(\overrightarrow{\exp} \int_t^\tau \operatorname{ad} V_\theta\, d\theta\right) W_\tau\, d\tau \\
&= \overrightarrow{\exp} \int_0^t V_\tau\, d\tau \circ\; \overrightarrow{\exp} \int_0^t \left(\overrightarrow{\exp} \int_t^\tau \operatorname{ad} V_\theta\, d\theta\right) W_\tau\, d\tau.
\end{aligned} \tag{2.24}$$

For autonomous vector fields $V, W \in \operatorname{Vec} M$, the variations formulas (2.23), (2.24) take the form:

$$e^{t(V+W)} = \overrightarrow{\exp} \int_0^t e^{\tau \operatorname{ad} V} W\, d\tau \circ e^{tV} = e^{tV} \circ\; \overrightarrow{\exp} \int_0^t e^{(\tau - t) \operatorname{ad} V} W\, d\tau.$$

In particular, for $t = 1$ we have

$$e^{V+W} = \overrightarrow{\exp} \int_0^1 e^{\tau \operatorname{ad} V} W\, d\tau \circ e^V.$$

2.8 Derivative of flow with respect to parameter

Let $V_t(s)$ be a nonautonomous vector field depending smoothly on a real parameter s. We study dependence of the flow of $V_t(s)$ on the parameter s.

We write

$$\overrightarrow{\exp}\int_0^t V_\tau(s+\varepsilon)\,d\tau = \overrightarrow{\exp}\int_0^t \left(V_\tau(s)+\delta_{V_\tau}(s,\varepsilon)\right)\,d\tau \qquad (2.25)$$

with the perturbation $\delta_{V_\tau}(s,\varepsilon) = V_\tau(s+\varepsilon) - V_\tau(s)$. By the variations formula (2.23), the previous flow is equal to

$$\overrightarrow{\exp}\int_0^t \left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\delta_{V_\tau}(s,\varepsilon)\,d\tau \circ \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau.$$

Now we expand in ε:

$$\begin{aligned}
\delta_{V_\tau}(s,\varepsilon) &= \varepsilon\frac{\partial}{\partial s}V_\tau(s) + O(\varepsilon^2), \qquad \varepsilon\to 0,\\
W_\tau(s,\varepsilon) &\stackrel{\text{def}}{=} \left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\delta_{V_\tau}(s,\varepsilon)\\
&= \varepsilon\left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\frac{\partial}{\partial s}V_\tau(s) + O(\varepsilon^2), \qquad \varepsilon\to 0,
\end{aligned}$$

thus

$$\begin{aligned}
\overrightarrow{\exp}\int_0^t W_\tau(s,\varepsilon)\,d\tau &= \operatorname{Id} + \int_0^t W_\tau(s,\varepsilon)\,d\tau + O(\varepsilon^2)\\
&= \operatorname{Id} + \varepsilon\int_0^t \left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\frac{\partial}{\partial s}V_\tau(s)\,d\tau + O(\varepsilon^2).
\end{aligned}$$

Finally,

$$\begin{aligned}
&\overrightarrow{\exp}\int_0^t V_\tau(s+\varepsilon)\,d\tau = \overrightarrow{\exp}\int_0^t W_{s,\tau}(\varepsilon)\,d\tau \circ \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau\\
&\quad = \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau\\
&\qquad + \varepsilon\int_0^t \left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\frac{\partial}{\partial s}V_\tau(s)\,d\tau \circ \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau + O(\varepsilon^2),
\end{aligned}$$

that is,

$$\begin{aligned}
&\frac{\partial}{\partial s}\,\overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau\\
&\qquad = \int_0^t \left(\overrightarrow{\exp}\int_0^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\frac{\partial}{\partial s}V_\tau(s)\,d\tau \circ \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau. \qquad (2.26)
\end{aligned}$$

Similarly, we obtain from the variations formula (2.24) the equality

$$\begin{aligned}
&\frac{\partial}{\partial s}\,\overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau\\
&\qquad = \overrightarrow{\exp}\int_0^t V_\tau(s)\,d\tau \circ \int_0^t \left(\overrightarrow{\exp}\int_t^\tau \operatorname{ad} V_\theta(s)\,d\theta\right)\frac{\partial}{\partial s}V_\tau(s)\,d\tau. \qquad (2.27)
\end{aligned}$$

For an autonomous vector field depending on a parameter $V(s)$, formula (2.26) takes the form

$$\frac{\partial}{\partial s} e^{tV(s)} = \int_0^t e^{\tau \operatorname{ad} V(s)} \frac{\partial V}{\partial s} \, d\tau \circ e^{tV(s)},$$

and at $t = 1$:

$$\frac{\partial}{\partial s} e^{V(s)} = \int_0^1 e^{\tau \operatorname{ad} V(s)} \frac{\partial V}{\partial s} \, d\tau \circ e^{V(s)}. \tag{2.28}$$

Proposition 2.7. *Assume that*

$$\left[\int_0^t V_\tau \, d\tau, V_t \right] = 0 \qquad \forall t. \tag{2.29}$$

Then

$$\overrightarrow{\exp} \int_0^t V_\tau \, d\tau = e^{\int_0^t V_\tau \, d\tau} \qquad \forall t.$$

That is, we state that under the commutativity assumption (2.29), the chronological exponential $\overrightarrow{\exp} \int_0^t V_\tau \, d\tau$ coincides with the flow $Q^t = e^{\int_0^t V_\tau \, d\tau}$ defined as follows:

$$Q^t = Q_1^t,$$
$$\frac{\partial Q_s^t}{\partial s} = \int_0^t V_\tau \, d\tau \circ Q_s^t, \qquad Q_0^t = \operatorname{Id}.$$

Proof. We show that the exponential in the right-hand side satisfies the same ODE as the chronological exponential in the left-hand side. By (2.28), we have

$$\frac{d}{dt} e^{\int_0^t V_\tau \, d\tau} = \int_0^1 e^{\tau \operatorname{ad} \int_0^t V_\theta \, d\theta} V_t \, d\tau \circ e^{\int_0^t V_\tau \, d\tau}.$$

In view of equality (2.29),

$$e^{\tau \operatorname{ad} \int_0^t V_\theta \, d\theta} V_t = V_t,$$

thus

$$\frac{d}{dt} e^{\int_0^t V_\tau \, d\tau} = V_t \circ e^{\int_0^t V_\tau \, d\tau}.$$

By equality (2.29), we can permute operators in the right-hand side:

$$\frac{d}{dt} e^{\int_0^t V_\tau \, d\tau} = e^{\int_0^t V_\tau \, d\tau} \circ V_t.$$

Notice the initial condition

$$\left. e^{\int_0^t V_\tau \, d\tau} \right|_{t=0} = \operatorname{Id}.$$

Now the statement follows since the Cauchy problem for flows

$$\dot{A}_t = A_t \circ V_t, \qquad A_0 = \mathrm{Id}$$

has a unique solution:

$$A_t = e^{\int_0^t V_\tau \, d\tau} = \overrightarrow{\exp} \int_0^t V_\tau \, d\tau.$$

□

Chapter 3

Linear systems

Linear control systems have the form

$$\dot{x} = Ax + c + \sum_{i=1}^{m} u_i b_i, \qquad x \in \mathbb{R}^n, \quad u = (u_1, \dots, u_m) \in \mathbb{R}^m, \tag{3.1}$$

where A is a constant real $n \times n$ matrix and c, $b_1, \dots, b_m$ are constant vectors in $\mathbb{R}^n$.

3.1 Cauchy's formula for linear systems

Let $u(t) = (u_1(t), \dots, u_m(t))$ be locally integrable functions. Then the solution of (3.1) corresponding to this control and satisfying the initial condition

$$x(0, x_0) = x_0$$

is given by *Cauchy's formula*:

$$x(t, x_0) = e^{tA} \left(x_0 + \int_0^t e^{-\tau A} \left(\sum_{i=1}^{m} u_i(\tau) b_i + c \, d\tau \right) \right), \quad t \in \mathbb{R}.$$

Here we use the standard notation for the matrix exponential:

$$e^{tA} = \mathrm{Id} + tA + \frac{t^2}{2!} A^2 + \cdots + \frac{t^n}{n!} A^n + \cdots .$$

Cauchy's formula is verified by differentiation. In view of uniqueness, it gives the solution to the Cauchy problem.

Linear system (3.1) is a particular case of an *affine in control system*:

$$\dot{x} = x \circ \left(f_0 + \sum_{i=1}^{m} u_i f_i \right), \tag{3.2}$$

in order to obtain (3.1) from (3.2), one should just take

$$f_0(x) = Ax + c, \qquad f_i(x) = b_i, \quad i = 1, \dots, m. \tag{3.3}$$

Let us show that Cauchy's formula is actually a special case of the general variations formula. We restrict ourselves with the case $c = 0$.

The variations formula for (3.2) takes the form

$$\begin{aligned} &\overrightarrow{\exp} \int_0^t \left(f_0 + \sum_{i=1}^m u_i(\tau) f_i \right) d\tau \\ &= \overrightarrow{\exp} \int_0^t \left(\left(\overrightarrow{\exp} \int_0^\tau \operatorname{ad} f_0 \, d\theta \right) \circ \sum_{i=1}^m u_i(\tau) f_i \right) d\tau \circ \overrightarrow{\exp} \int_0^t f_0 \, d\tau \\ &= \overrightarrow{\exp} \int_0^t \left(\sum_{i=1}^m u_i(\tau) e^{\tau \operatorname{ad} f_0} f_i \right) d\tau \circ e^{t f_0}. \end{aligned} \tag{3.4}$$

We assume that $c = 0$, i.e., $f_0(x) = Ax$. Then

$$x \circ e^{t f_0} = e^{tA} x. \tag{3.5}$$

Further, since $(\operatorname{ad} f_0) f_i = [f_0, f_i] = [Ax, b] = -Ab$ then

$$\begin{aligned} e^{\tau \operatorname{ad} f_0} f_i &= f_i + \tau (\operatorname{ad} f_0) f_i + \frac{\tau^2}{2!} (\operatorname{ad} f_0)^2 f_i + \dots + \frac{\tau^n}{n!} (\operatorname{ad} f_0)^n f_i + \cdots \\ &= b_i - \tau A b_i + \frac{\tau^2}{2!} (-A)^2 b_i + \dots + \frac{\tau^n}{n!} (-A)^n b_i + \cdots \\ &= e^{-\tau A} b_i. \end{aligned}$$

In order to compute the left flow in (3.4), recall that the curve

$$x_0 \circ \overrightarrow{\exp} \int_0^t \left(\sum_{i=1}^m u_i(\tau) e^{\tau \operatorname{ad} f_0} f_i \right) d\tau = x_0 \circ \overrightarrow{\exp} \int_0^t \left(\sum_{i=1}^m u_i(\tau) e^{-\tau A} b_i \right) d\tau \tag{3.6}$$

is the solution to the Cauchy problem

$$\dot{x}(t) = \sum_{i=1}^m u_i(\tau) e^{-\tau A} b_i, \qquad x(0) = x_0,$$

thus (3.6) is equal to

$$x(t) = x_0 + \int_0^t \left(e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \right) d\tau.$$

Taking into account (3.5), we obtain Cauchy's formula:

$$
\begin{aligned}
x(t) &= x_0 \circ \overrightarrow{\exp} \int_0^t \left(f_0 + \sum_{i=1}^m u_i(\tau) f_i \right) d\tau \\
&= \left(x_0 + \int_0^t \left(e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \right) d\tau \right) \circ e^{t f_0} \\
&= e^{tA} \left(x_0 + \int_0^t \left(e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \right) d\tau \right).
\end{aligned}
$$

Notice that in the general case ($c \neq 0$) Cauchy's formula can be written as follows:

$$
\begin{aligned}
x(t, x_0) &= e^{tA} x_0 + e^{tA} \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau + e^{tA} \int_0^t e^{-\tau A} c \, d\tau \\
&= e^{tA} x_0 + e^{tA} \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau + \frac{e^{tA} - \mathrm{Id}}{A} c, \qquad (3.7)
\end{aligned}
$$

where

$$
\frac{e^{tA} - \mathrm{Id}}{A} c = t \, \mathrm{Id} + \frac{t^2}{2!} A + \frac{t^3}{3!} A^2 + \cdots + \frac{t^n}{n!} A^{n-1} + \cdots .
$$

3.2 Controllability of linear systems

Cauchy's formula (3.7) yields that the mapping

$$
u(\cdot) \mapsto x(t, x_0, u(\cdot)),
$$

which sends a locally inegrable control to the endpoint of the corresponding trajectory, is affine. Thus the attainable set $\mathcal{A}_{x_0}(t)$ of linear system (3.1) for a fixed time $t > 0$ is an affine subspace in $\mathbb{R}^n$.

Definition 3.1. A control system on a state space M is called *completely controllable* for time $t > 0$ if

$$
\mathcal{A}_{x_0}(t) = M \qquad \forall x_0 \in M.
$$

This definition means that for any pair of points $x_0, x_1 \in M$ exists an admissible control $u(\cdot)$ such that the corresponding solution $x(\cdot, x_0, u(\cdot))$ of the control system steers x_0 to x_1 in t units of time:

$$
x(0, x_0, u(\cdot)) = x_0, \quad x(t, x_0, u(\cdot)) = x_1.
$$

The study of complete controllability of linear systems is facilitated by the following observation. The affine mapping

$$
u(\cdot) \mapsto e^{tA} x_0 + \frac{e^{tA} - \mathrm{Id}}{A} c + e^{tA} \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau
$$

is surjective if and only if its linear part

$$u(\cdot) \mapsto e^{tA} \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau \tag{3.8}$$

is onto. Moreover, (3.8) is surjective iff the following mapping is:

$$u(\cdot) \mapsto \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau. \tag{3.9}$$

Theorem 3.1. *The linear system* (3.1) *is completely controllable for a time* $t > 0$ *if and only if*

$$\operatorname{span}\{A^j b_i \mid j = 0, \dots, n-1, \; i = 1, \dots, m\} = \mathbb{R}^n. \tag{3.10}$$

Proof. Necessity. Assume, by contradiction, that condition (3.10) is violated. Then there exists a covector $p \in \mathbb{R}^{n*}$, $p \neq 0$, such that

$$pA^j b_i = 0, \qquad j = 0, \dots, n-1, \; i = 1, \dots, m. \tag{3.11}$$

By Cayley theorem,

$$A^n = \sum_{j=0}^{n-1} \alpha_j A^j$$

for some real numbers $\alpha_0, \dots, \alpha_{n-1}$, thus

$$A^k = \sum_{j=0}^{n-1} \beta_j^k A^j$$

for any $k \in \mathbb{N}$ and some $\beta_j^k \in \mathbb{R}$. Now we obtain from (3.11):

$$pA^k b_i = \sum_{j=0}^{n-1} \beta_j^k pA^j b_i = 0, \qquad k = 0, 1, \dots, \; i = 1, \dots, m.$$

That is why

$$pe^{-\tau A} b_i = 0, \qquad i = 1, \dots, m,$$

and finally

$$p \int_0^t e^{-\tau A} \sum_{i=1}^m u_i(\tau) b_i \, d\tau = \int_0^t \sum_{i=1}^m u_i(\tau) p e^{-\tau A} b_i \, d\tau = 0,$$

i.e., mapping (3.9) is not surjective. The contradiction proves necessity.

Sufficiency. By contradiction, suppose that mapping (3.9) is not surjective. Then there exists a covector $p \in \mathbb{R}^{n*}$, $p \neq 0$, such that

$$p \int_0^t \sum_{i=1}^m u_i(\tau) e^{-\tau A} b_i \, d\tau = 0 \qquad \forall u(\cdot) = (u_1(\cdot), \dots, u_m(\cdot)). \tag{3.12}$$

Choose a control of the form:

$$u(\tau) = (0, \dots, 0, v_s(\tau), 0, \dots, 0),$$

where the only nonzero i-th component is

$$v_s(\tau) = \begin{cases} 1, & 0 \le \tau \le s, \\ 0, & \tau > s. \end{cases}$$

Then equality (3.12) gives

$$p \int_0^s e^{-\tau A} b_i \, d\tau = 0, \qquad s \in \mathbb{R}, \ i = 1, \dots, m,$$

thus

$$p e^{-sA} b_i = 0, \qquad s \in \mathbb{R}, \ i = 1, \dots, m.$$

We differentiate this equality repeatedly at $s = 0$ and obtain

$$p A^k b_i = 0, \qquad k = 0, 1, \dots, \ i = 1, \dots, m,$$

a contradiction with (3.10). Sufficiency follows. □

Chapter 4

State linearizability of nonlinear systems

The aim of this chapter is to characterize nonlinear systems

$$\dot{q} = f_0(q) + \sum_{i=1}^{m} u_i f_i(q), \qquad u = (u_1, \dots, u_m) \in \mathbb{R}^m, \ q \in M \tag{4.1}$$

that are equivalent, locally or globally, to controllable linear systems. That is, we seek conditions on vector fields $f_0, f_1, \dots, f_m$ that guarantee existence of a diffeomorphism (global $\Phi : M \to \mathbb{R}^n$ or local $\Phi : O_{q_0} \subset M \to O_0 \subset \mathbb{R}^n$) which transforms nonlinear system (4.1) into a controllable linear one (3.1).

4.1 Local linearizability

We start with the local problem. A natural language for conditions of local linearizability is provided by Lie brackets, which are invariant under diffeomorphisms:

$$\Phi_*[V, W] = [\Phi_* V, \Phi_* W], \qquad V, W \in \operatorname{Vec} M.$$

The controllability condition (3.10) can easily be rewritten in terms of Lie brackets: since

$$(-A)^j b_i = (\operatorname{ad} f_0)^j f_i = [\underbrace{f_0, [\dots [f_0}_{j \text{ times}}, f_i] \dots]]$$

for vector fields (3.3), then the controllability test for linear systems (3.10) reads

$$\operatorname{span}\{x_0 \circ (\operatorname{ad} f_0)^j f_i \mid j = 0, \dots, n-1, \ i = 1, \dots, m\} = T_{x_0}\mathbb{R}^n.$$

Further, one can see that the following equality is satisfied for linear vector fields (3.3):

$$[(\operatorname{ad} f_0)^{j_1} f_{i_1}, (\operatorname{ad} f_0)^{j_2} f_{i_2}] = [(-A)^{j_1} b_{i_1}, (-A)^{j_2} b_{i_2}] = 0,$$
$$0 \le j_1, \ j_2, \quad 1 \le i_1, \ i_2 \le m.$$

It turns out that the two conditions found above give a precise local characterization of controllable linear systems.

Theorem 4.1. *Let M be a smooth n-dimensional manifold, and let $f_0, f_1, \dots, f_m \in \operatorname{Vec} M$. There exists a diffeomorphism*

$$\Phi : O_{q_0} \to O_0$$

of a neighborhood $O_{q_0} \subset M$ of a point $q_0 \in M$ to a neighborhood $O_0 \subset \mathbb{R}^n$ of the origin $0 \in \mathbb{R}^n$ such that

$$(\Phi_* f_0)(x) = Ax + c, \qquad x \in O_0,$$
$$(\Phi_* f_i)(x) = b_i, \qquad x \in O_0, \quad i = 1, \dots, m,$$

for some $n \times n$ matrix A and $c, b_1, \dots, b_m \in \mathbb{R}^n$ that satisfy the controllability condition (3.10) *if and only if the following conditions hold:*

$$\operatorname{span}\{q_0 \circ (\operatorname{ad} f_0)^j f_i \mid j = 0, \dots, n-1, \ i = 1, \dots, m\} = T_{q_0} M. \tag{4.2}$$
$$q \circ [(\operatorname{ad} f_0)^{j_1} f_{i_1}, (\operatorname{ad} f_0)^{j_2} f_{i_2}] = 0, \tag{4.3}$$
$$q \in O_{q_0}, \quad 0 \le j_1, \ j_2 \le n, \quad 1 \le i_1, \ i_2 \le m.$$

Remark. In other words, the diffeomorphism Φ from the theorem transforms a nonlinear system (4.1) to a linear one (3.1).

Before proving the theorem, we consider the following proposition, which we will need later.

Lemma 4.1. *Let M be a smooth n-dimensional manifold, and let $Y_1, \dots, Y_k \in \operatorname{Vec} M$. There exists a diffeomorphism*

$$\Phi : O_0 \to O_{q_0}$$

of a neighborhood $O_0 \subset \mathbb{R}^n$ to a neighborhood $O_{q_0} \subset M$, $q_0 \in M$, such that

$$\Phi_* \left(\frac{\partial}{\partial x_i} \right) = Y_i, \qquad i = 1, \dots, k,$$

if and only if the vector fields $Y_1, \dots, Y_k$ commute:

$$[Y_i, Y_j] \equiv 0, \qquad i, \ j = 1, \dots, k,$$

and are linearly independent:

$$\dim \operatorname{span}(q_0 \circ Y_1, \dots, q_0 \circ Y_k) = k.$$

Proof. Necessity is obvious since Lie bracket and linear independence are invariant with respect to diffeomorphisms.

Sufficiency. Choose $Y_{k+1}, \dots, Y_n \in \operatorname{Vec} M$ that complete $Y_1, \dots, Y_k$ to a basis:

$$\operatorname{span}(q \circ Y_1, \dots, q \circ Y_n) = T_q M, \qquad q \in O_{q_0}.$$

The mapping

$$\Phi(s_1, \dots, s_n) = q_0 \circ e^{s_n Y_n} \circ \dots \circ e^{s_1 Y_1}$$

is defined on a sufficiently small neighborhood of the origin in $\mathbb{R}^n$. We have

$$\left.\frac{\partial}{\partial s_i}\right|_{s=0} \Phi(s) \stackrel{\text{def}}{=} \Phi_* \left(\left.\frac{\partial}{\partial s_i}\right|_{s=0} \right) = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} q_0 \circ e^{\varepsilon Y_i} = q_0 \circ Y_i.$$

Hence $\Phi_*|_{s=0}$ is surjective and Φ is a diffeomorphism of a neighborhood of 0 in $\mathbb{R}^n$ and a neighborhood of q_0 in M, according to the implicit function theorem.

Now we prove that Φ rectifies the vector fields $Y_1, \dots, Y_k$. First of all, notice that since these vector fields commute, then their flows also commute, thus

$$e^{s_k Y_k} \circ \dots \circ e^{s_1 Y_1} = e^{\sum_{i=1}^k s_i Y_i}$$

and

$$\Phi(s_1, \dots, s_n) = q_0 \circ e^{s_n Y_n} \circ \dots \circ e^{s_{k+1} Y_{k+1}} \circ e^{\sum_{i=1}^k s_i Y_i}.$$

Then for $i = 1, \dots, k$

$$\begin{aligned}
\left.\Phi_* \left(\frac{\partial}{\partial s_i} \right)\right|_{\Phi(s)} &= \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \Phi(s_1, \dots, s_i + \varepsilon, \dots, s_n) \\
&= \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} q_0 \circ e^{s_n Y_n} \circ \dots \circ e^{s_{k+1} Y_{k+1}} \circ e^{\sum_{j=1}^k s_j Y_j} \circ e^{\varepsilon Y_i} \\
&= q_0 \circ e^{s_n Y_n} \circ \dots \circ e^{s_{k+1} Y_{k+1}} \circ e^{\sum_{j=1}^k s_j Y_j} \circ \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} e^{\varepsilon Y_i} \\
&= \Phi(s) \circ Y_i.
\end{aligned}$$

□

Now we can prove Theorem 4.1 on local equivalence of nonlinear systems with linear ones.

Proof. Necessity is obvious since Lie brackets are invariant with respect to diffeomorphisms, and for linear systems conditions (4.2), (4.3) hold.

Sufficiency. Select a basis of the space $T_{q_0} M$ among vectors of the form $(\operatorname{ad} f_0)^j f_i(q_0)$:

$$Y_\alpha = (\operatorname{ad} f_0)^{j_\alpha} f_{i_\alpha}, \qquad \alpha = 1, \dots, n, \quad 0 \le j_\alpha \le n-1, \quad 1 \le i_\alpha \le m,$$
$$\operatorname{span}(q_0 \circ Y_1, \dots, q_0 \circ Y_n) = T_{q_0} M.$$

By Lemma 4.1, there exists a rectifying diffeomorphism:

$$\Phi : O_{q_0} \to O_0, \qquad \Phi_* Y_\alpha = \frac{\partial}{\partial x_\alpha}, \qquad \alpha = 1, \dots, n.$$

We show that Φ is the required diffeomorphism.

(1) First we check that the vector fields $\Phi_* f_i$, $i = 1, \ldots, m$ are constant. That is, we show that in the decomposition

$$\Phi_* f_i = \sum_{\alpha=1}^{n} \beta_\alpha^i(x) \frac{\partial}{\partial x_\alpha}. \qquad i = 1, \ldots, m$$

the functions $\beta_\alpha^i(x)$ are constant. We have

$$[\frac{\partial}{\partial x_\alpha}, \Phi_* f_i] = \sum_{\alpha=1}^{n} \frac{\partial \beta_\alpha^i}{\partial x_j} \frac{\partial}{\partial x_\alpha}, \tag{4.4}$$

on the other hand

$$[\frac{\partial}{\partial x_\alpha}, \Phi_* f_i] = [\Phi_* Y_\alpha, \Phi_* f_i] = \Phi_*[Y_\alpha, f_i] = \Phi_*[(\mathrm{ad}\, f_0)^{j_\alpha} f_{i_\alpha}, f_i] = 0 \tag{4.5}$$

by hypothesis (4.3). Now we compare (4.4) and (4.5) and obtain

$$\frac{\partial \beta_\alpha^i}{\partial x_j} \frac{\partial}{\partial x_\alpha} \equiv 0 \quad \Rightarrow \quad \beta_\alpha^i = \mathrm{const}, \quad i = 1, \ldots m, \quad \alpha = 1, \ldots, n,$$

i.e., $\Phi_* f_i$, $i = 1, \ldots, m$, are constant vector fields b_i, $i = 1, \ldots, m$.

(2) Now we show that the vector field $\Phi_* f_0$ is linear. We prove that in the decomposition

$$\Phi_* f_0 = \sum_{i=1}^{n} \beta_i(x) \frac{\partial}{\partial x_i}$$

all functions $\beta_i(x)$, $i = 1, \ldots, n$ are linear. Indeed,

$$\begin{aligned}
\sum_{\alpha=1}^{n} \frac{\partial^2 \beta_i}{\partial x_\alpha \partial x_\beta} \frac{\partial}{\partial x_i} &= [\frac{\partial}{\partial x_\alpha}, [\frac{\partial}{\partial x_\beta}, \Phi_* f_0]] \\
&= [\Phi_* Y_\alpha, [\Phi_* Y_\beta, \Phi_* f_0]] = \Phi_*[Y_\alpha, [Y_\beta, f_0]] \\
&= \Phi_*[(\mathrm{ad}\, f_0)^{j_\alpha} f_{i_\alpha}, [(\mathrm{ad}\, f_0)^{j_\beta} f_{i_\beta}, f_0]] \\
&= -\Phi_*[(\mathrm{ad}\, f_0)^{j_\alpha} f_{i_\alpha}, [f_0, (\mathrm{ad}\, f_0)^{j_\beta} f_{i_\beta}]] \\
&= -\Phi_*[(\mathrm{ad}\, f_0)^{j_\alpha} f_{i_\alpha}, (\mathrm{ad}\, f_0)^{j_\beta + 1} f_{i_\beta}] \\
&= 0, \qquad \alpha,\ \beta = 1, \ldots, n,
\end{aligned}$$

by hypothesis (4.3). Thus

$$\frac{\partial^2 \beta_i}{\partial x_\alpha \partial x_\beta} \frac{\partial}{\partial x_i} \equiv 0, \qquad i,\ \alpha,\ \beta = 1, \ldots, n,$$

i.e., $\Phi_* f_0$ is a linear vector field $Ax + c$.

For the linear system $\dot{x} = Ax + c + \sum_{i=1}^{m} u_i b_i$, hypothesis (4.2) implies the controllability condition (3.10) □

4.2 Global linearizability

Now we prove the following statement on global equivalence.

Theorem 4.2. *Let M be a smooth connected n-dimensional manifold, and let $f_0, f_1, \dots, f_m \in \operatorname{Vec} M$. There exists a diffeomorphism*

$$\Phi : M \to T^k \times \mathbb{R}^{n-k}$$

of M to the product of a k-dimensional torus with $\mathbb{R}^{n-k}$ for some $k \le n$ such that

$$(\Phi_* f_0)(x) = Ax + c, \qquad x \in T^k \times \mathbb{R}^{n-k},$$
$$(\Phi_* f_i)(x) = b_i, \qquad x \in T^k \times \mathbb{R}^{n-k} \quad i = 1, \dots, m,$$

for some $n \times n$ matrix A with zero first k rows:

$$Ae_i = 0, \qquad i = 1, \dots, k, \tag{4.6}$$

and $c, b_1, \dots, b_m \in \mathbb{R}^n$ that satisfy the controllability condition (3.10) *if and only if the following conditions hold:*

$$(\operatorname{ad} f_0)^j f_i, \qquad j = 0, 1, \dots, n-1, \quad i = 1, \dots, m, \quad \textit{are complete vector fields} \tag{4.7}$$

$$\operatorname{span}\{q \circ (\operatorname{ad} f_0)^j f_i \mid j = 0, \dots, n-1, \ i = 1, \dots, m\} = T_q M. \tag{4.8}$$

$$q \circ [(\operatorname{ad} f_0)^{j_1} f_{i_1}, (\operatorname{ad} f_0)^{j_2} f_{i_2}] = 0, \quad q \in M, \quad 0 \le j_1, \ j_2 \le n, \quad 1 \le i_1, \ i_2 \le m. \tag{4.9}$$

Remarks. (1) If M is additionally supposed simply connected, then it is diffeomorphic to $\mathbb{R}^n$, i.e., $k = 0$.

(2) If, on the contrary, M is compact, i.e., diffeomorphic to T^n and $m < n$, then there are no globally linearizable controllable systems on M. Indeed, then $A = 0$, and the controllability condition (3.10) is violated.

Proof. Sufficiency. Fix a point $q_0 \in M$ and find a basis in $T_{q_0} M$ of vectors of the form

$$Y_\alpha = (\operatorname{ad} f_0)^{j_\alpha} f_{i_\alpha}, \qquad \alpha = 1, \dots, n,$$
$$\operatorname{span}(q_0 \circ Y_1, \dots, q_0 \circ Y_n) = T_{q_0} M.$$

(1) First we show that the vector fields $Y_1, \dots, Y_n$ are linearly independent everywhere in M. The set

$$O = \{q \in M \mid \operatorname{span}(q \circ Y_1, \dots, q \circ Y_n) = T_q M\}$$

is obviously open. We show that it is closed. In this set we have a decomposition

$$q \circ (\operatorname{ad} f_0)^j f_i = q \circ \sum_{\alpha=1}^{n} a_\alpha^{ij} Y_\alpha, \qquad q \in O, \quad j = 0, \dots, n-1, \quad i = 1, \dots, m, \tag{4.10}$$

for some functions $a_\alpha^{ij} \in C^\infty(O)$. We prove that actually all a_α^{ij} are constant. We have

$$\begin{aligned} 0 &= [Y_\beta, \sum_{\alpha=1}^{n} a_\alpha^{ij} Y_\alpha] \\ &\quad \text{(by Leibniz rule } [X, aY] = (Xa)Y + a[X,Y]) \\ &= \sum_{\alpha=1}^{n} a_\alpha^{ij} [Y_\beta, Y_\alpha] + \sum_{\alpha=1}^{n} (Y_\beta a_\alpha^{ij}) Y_\alpha \\ &= \sum_{\alpha=1}^{n} (Y_\beta a_\alpha^{ij}) Y_\alpha, \qquad \beta = 1, \dots, n, \quad j = 0, \dots, n-1, \quad i = 1, \dots, m, \end{aligned}$$

thus

$$Y_\beta a_\alpha^{ij} = 0 \quad \Rightarrow \quad a_\alpha^{ij}\big|_O = \text{const}, \qquad \alpha = 1, \dots, n, \quad j = 0, \dots, n-1, \quad i = 1, \dots, m.$$

That is why equality (4.10) holds in the closure $\overline{O}$. Thus the vector fields $Y_1, \dots, Y_n$ are linearly independent in $\overline{O}$ (if this is not the case, then the whole family $(\operatorname{ad} f_0)^j f_i$, $j = 0, \dots, n-1$, $i = 1, \dots, m$, is not linearly independent in $\overline{O}$). Hence the set O is closed. Since it is simultaneously open and M is connected,

$$O = M,$$

i.e., the vector fields $Y_1, \dots, Y_n$ are linearly independent in M.

(2) We define the "inverse" Ψ of the required diffeomorphism as follows:

$$\begin{aligned} \Psi(x_1, \dots, x_n) &= q_0 \circ e^{x_1 Y_1} \circ \dots \circ e^{x_n Y_n} \\ &\quad \text{(since the vector fields } Y_\alpha \text{ commute)} \\ &= q_0 \circ e^{\sum_{\alpha=1}^{n} x_\alpha Y_\alpha}, \qquad x = (x_1, \dots, x_n) \in \mathbb{R}^n. \end{aligned}$$

(3) We show that the (obviously smooth) mapping $\Psi : \mathbb{R}^n \to M$ is regular, i.e., its differential is surjective. Indeed,

$$\begin{aligned} \frac{\partial \Psi}{\partial x_\alpha}(x) &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \Psi(x_1, \dots, x_\alpha + \varepsilon, \dots, x_n) \\ &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} q_0 \circ e^{\sum_{\beta=1}^{n} x_\beta Y_\beta + \varepsilon Y_\alpha} \\ &= q_0 \circ e^{\sum_{\beta=1}^{n} x_\beta Y_\beta} \circ Y_\alpha \\ &= \Psi(x) \circ Y_\alpha, \qquad \alpha = 1, \dots, n, \end{aligned}$$

thus

$$\Psi_{*x}(\mathbb{R}^n) = T_{\Psi(x)}M.$$

The mapping Ψ is regular, thus a local diffeomorphism. In particular, $\Psi(\mathbb{R}^n)$ is open.

(4) We prove that $\Psi(\mathbb{R}^n)$ is closed. Take any point $q \in \overline{\Psi(\mathbb{R}^n)}$. Since the vector fields $Y_1, \ldots, Y_n$ are linearly independent, the image of the mapping

$$(y_1, \ldots, y_n) \mapsto q \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha}, \qquad y = (y_1, \ldots, y_n) \in \mathbb{R}^n$$

contains a neighborhood of the point q. Thus there exists $y \in \mathbb{R}^n$ such that

$$q \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha} \in \Psi(\mathbb{R}^n),$$

i.e.,

$$q \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha} = q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha}$$

for some $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} q &= q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} \circ e^{-\sum_{\alpha=1}^n y_\alpha Y_\alpha} = q_0 \circ e^{\sum_{\alpha=1}^n (x_\alpha - y_\alpha) Y_\alpha} \\ &= \Psi(x-y). \end{aligned}$$

In other words, $q \in \Psi(\mathbb{R}^n)$.

That is why the set $\Psi(\mathbb{R}^n)$ is closed. Since it is open and M is connected,

$$\Psi(\mathbb{R}^n) = M.$$

(5) It is easy to see that the preimage

$$\Psi^{-1}(q_0) = \{x \in \mathbb{R}^n \mid \Psi(x) = q_0\}$$

is a subgroup of the Abelian group $\mathbb{R}^n$. Indeed, let

$$\Psi(x) = q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} = \Psi(y) = q_0 \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha} = q_0,$$

then

$$\Psi(x+y) = q_0 \circ e^{\sum_{\alpha=1}^n (x_\alpha + y_\alpha) Y_\alpha} = q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha} = q_0.$$

Analogously, if

$$\Psi(x) = q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} = q_0,$$

then

$$\Psi(-x) = q_0 \circ e^{-\sum_{\alpha=1}^n x_\alpha Y_\alpha} = q_0.$$

Finally,

$$\Psi(0) = q_0.$$

(6) Moreover, $\Psi^{-1}(q_0)$ is a discrete subgroup G_0 of $\mathbb{R}^n$, i.e., there are no nonzero elements of $\Psi^{-1}(q_0)$ in some neighborhood of the origin in $\mathbb{R}^n$, since Ψ is a local diffeomorphism.

(7) The mapping Ψ is well-defined on the quotient $\mathbb{R}^n/G_0$. Indeed, let $y \in G_0$. Then

$$\begin{aligned}\Psi(x+y) &= q_0 \circ e^{\sum_{\alpha=1}^n (x_\alpha + y_\alpha) Y_\alpha} = q_0 \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha} \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} \\ &= q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} = \Psi(x).\end{aligned}$$

So

$$\Psi : \mathbb{R}^n/G_0 \to M. \tag{4.11}$$

(8) The mapping (4.11) is one-to-one: if

$$\Psi(x) = \Psi(y), \qquad x,\ y \in \mathbb{R}^n,$$

then

$$q_0 \circ e^{\sum_{\alpha=1}^n x_\alpha Y_\alpha} = q_0 \circ e^{\sum_{\alpha=1}^n y_\alpha Y_\alpha},$$

thus

$$q_0 \circ e^{\sum_{\alpha=1}^n (x_\alpha - y_\alpha) Y_\alpha} = q_0,$$

i.e., $x - y \in G_0$.

(9) That is why mapping (4.11) is a diffeomorphism. By Lemma 4.2 (see below), the discrete subgroup G_0 of $\mathbb{R}^n$ is a lattice:

$$G_0 = \left\{ \sum_{i=1}^k n_i e_i \mid n_i \in \mathbb{Z} \right\},$$

thus the quotient is a cylinder:

$$\mathbb{R}^n/G_0 = T^k \times \mathbb{R}^{n-k}.$$

Hence we constructed a diffeomorphism

$$\Phi = \Psi^{-1} : M \to T^k \times \mathbb{R}^{n-k}.$$

Equalities (4.8) and (4.9) follow exactly as in Theorem 4.1.

The vector field $\Phi_* f_0 = Ax + c$ is well-defined on the quotient $T^k \times \mathbb{R}^{n-k}$, that is why equalities (4.6) hold.

Necessity. For a linear system on a cylinder $T^k \times \mathbb{R}^{n-k}$, conditions (4.7) and (4.9) obviously hold. If a linear system is controllable on the cylinder, then it is also controllable on $\mathbb{R}^n$, thus the controllability condition (4.8) is also satisfied. □

Lemma 4.2. *Let Γ be a discrete subgroup in $\mathbb{R}^n$. Then it is a lattice, i.e., there exist linearly independent vectors $e_1, \dots, e_k \in \mathbb{R}^n$ such that*

$$\Gamma = \left\{ \sum_{i=1}^{k} n_i e_i \mid n_i \in \mathbb{Z} \right\}.$$

Proof. We prove by induction on dimension n of the ambient group $\mathbb{R}^n$.

(1) Let $n = 1$. Since the subgroup $\Gamma \subset \mathbb{R}$ is discrete, it contains an element $e_1 \neq 0$ closest to the origin $0 \in \mathbb{R}$. By the group property, all multiples $\pm e_1 \pm e_1 \pm \cdots \pm e_1 = \pm n e_1$, $n = 0, 1, 2, \dots$, are also in Γ. We prove that Γ contains no other elements.

By contradiction, assume that there is an element $x \in \Gamma$ such that $n e_1 < x < (n+1) e_1$, $n \in \mathbb{Z}$. Then the element $y = x - n e_1 \in \Gamma$ is in the interval $(0, e_1) \subset \mathbb{R}$. So $y \neq 0$ is closer to the origin than e_1, a contradiction. Thus $\Gamma = \mathbb{Z} e_1 = \{ n e_1 \mid n \in \mathbb{Z} \}$, q.e.d.

(2) We prove the inductive step: let the statement of the lemma be proved for some $n - 1 \in \mathbb{N}$, we prove it for n.

Choose an element $e_1 \in \Gamma$, $e_1 \neq 0$, closest to the origin $0 \in \mathbb{R}^n$. Denote by l the line $\mathbb{R} e_1$, and by Γ_1 the lattice $\mathbb{Z} e_1 \subset \Gamma$. We suppose that $\Gamma \neq \Gamma_1$ (otherwise everything is proved).

Now we show that there is an element $e_2 \in \Gamma \setminus \Gamma_1$ closest to l:

$$\operatorname{dist}(e_2, l) = \min\{ \operatorname{dist}(x, l) \mid x \in \Gamma \setminus l \}. \tag{4.12}$$

Take any segment $I = [n e_1, (n+1) e_1] \subset l$, and denote by $\pi : \mathbb{R}^n \to l$ the orthogonal projection from $\mathbb{R}^n$ to l along the orthogonal complement to l in $\mathbb{R}^n$. Since the segment I is compact and the subgroup Γ is discrete, the n-dimensional strip $\pi^{-1}(I)$ contains an element $e_2 \in \Gamma \setminus l$ closest to I:

$$\operatorname{dist}(e_2, I) = \min\{ \operatorname{dist}(x, I) \mid x \in (\Gamma \setminus l) \cap \pi^{-1}(I) \}.$$

Then the element e_2 is the required one: it satisfies equality (4.12) since any element that satisfies (4.12) can be translated to the strip $\pi^{-1}(I)$ by elements of the lattice Γ_1.

That is why a sufficiently small neighborhood of l is free of elements of $\Gamma \setminus \Gamma_1$. Thus the quotient group Γ / Γ_1 is a discrete subgroup in $\mathbb{R}^n / l = \mathbb{R}^{n-1}$. By the inductive hypothesis, Γ / Γ_1 is a lattice, hence Γ is also a lattice. ◻

Chapter 5

The Orbit Theorem and its applications

5.1 Formulation of the Orbit Theorem

Let $\mathcal{F} \subset \operatorname{Vec} M$ be any set of smooth vector fields. In order to simplify notations, we assume that all fields from $\mathcal{F}$ are complete. Actually, all further definitions and results have clear generalizations to the case of noncomplete fields; we leave them to the reader.

We return to the study of attainable sets: we study the structure of the attainable sets of $\mathcal{F}$ by piecewise constant controls

$$\mathcal{A}_{q_0} = \{q_0 \circ e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k} \mid t_i \geq 0,\ f_i \in \mathcal{F},\ k \in \mathbb{N}\}, \qquad q_0 \in M.$$

But first we consider a greater set — the *orbit* of the family $\mathcal{F}$ through a point:

$$\mathcal{O}_{q_0} = \{q_0 \circ e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k} \mid t_i \in \mathbb{R},\ f_i \in \mathcal{F},\ k \in \mathbb{N}\}, \qquad q_0 \in M.$$

In an orbit $\mathcal{O}_{q_0}$, it is allowed to move along vector fields f_i both forward and backwards, while in an attainable set $\mathcal{A}_{q_0}$ only the forward motion is possible.

Although, if the family $\mathcal{F}$ is *symmetric*: $\mathcal{F} = -\mathcal{F}$ (i.e., $f \in \mathcal{F} \Rightarrow -f \in \mathcal{F}$), then attainable sets coincide with orbits: $\mathcal{O}_{q_0} = \mathcal{A}_{q_0}$, $q_0 \in M$.

In general, orbits have more simple structure that attainable sets. It is described in the following fundamental proposition.

Theorem 5.1 (Orbit Theorem, Nagano–Sussmann). *Let* $\mathcal{F} \subset \operatorname{Vec} M$ *and* $q_0 \in M$. *Then:*

(1) $\mathcal{O}_{q_0}$ *is an immersed submanifold of* M.

(2) $T_q\mathcal{O}_{q_0} = \operatorname{span}\{q \circ (\operatorname{Ad} P)f \mid P \in \mathcal{P},\ f \in \mathcal{F}\}$, $q \in \mathcal{O}_{q_0}$.

Here we denote by $\mathcal{P}$ the group of diffeomorphisms of M generated by flows in $\mathcal{F}$:

$$\mathcal{P} = \{e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k} \mid t_i \in \mathbb{R},\ f_i \in \mathcal{F},\ k \in \mathbb{N}\} \subset \operatorname{Diff} M.$$

5.2 Immersed submanifolds

Now we discuss an important notion that appears in the Orbit Theorem.

Definition 5.1. A subset W of a smooth n-dimensional manifold is called an *immersed k-dimensional submanifold* of M, $k \le n$, if there exists a one-to-one immersion

$$\Phi : N \to M, \qquad \operatorname{Ker} \Phi_{*x} = 0 \quad \forall\, x \in N$$

of a k-dimensional smooth manifold N such that

$$W = \Phi(N).$$

Remark. An immersed submanifold W of M can also be defined as a manifold contained in M such that the inclusion mapping

$$i : W \to M, \qquad i : q \mapsto q$$

is an immersion.

Sufficiently small neighborhoods O_x in an immersed submanifold W of M are submanifolds of M, but the whole W is not necessarily a submanifold of M in the sense of Definition 1.1. In general, the own topology of W can be stronger than the topology induced on W by the topology of M.

Example 5.1. Let $\Phi : \mathbb{R} \to \mathbb{R}^2$ be a one-to-one immersion of the line into the plane such that $\lim_{t\to+\infty} \Phi(t) = \Phi(0) = 0 \in \mathbb{R}^2$, Then $W = \Phi(\mathbb{R})$ is an immersed one-dimensional submanifold of $\mathbb{R}^2$. The own topology of W (inherited from $\mathbb{R}$) is stronger than the topology induced by $\mathbb{R}^2$. The intervals $\Phi(-\varepsilon, \varepsilon)$, $\varepsilon > 0$ small enough, are open in the first topology, but not open in the second one.

The notion of immersed submanifold appears inevitably in the description of orbits of families of vector fields. Already the orbit of one vector field (i.e., its trajectory) is an immersed submanifold, but may fail to be a submanifold in the sense of Definition 1.1.

Example 5.2. Oscillator with 2 degrees of freedom is described by the equations:

$$\begin{aligned} \ddot{x} + \alpha^2 x &= 0, \qquad x \in \mathbb{R}, \\ \ddot{y} + \beta^2 y &= 0, \qquad y \in \mathbb{R}. \end{aligned}$$

In the complex variables

$$z = x - i\dot{x}/\alpha, \qquad w = y - i\dot{y}/\beta$$

these equations read

$$\begin{aligned} \dot{z} &= i\alpha z, \qquad z \in \mathbb{C}, \\ \dot{w} &= i\beta w, \qquad w \in \mathbb{C}, \end{aligned} \tag{5.1}$$

and their solutions have the form

$$z(t) = e^{i\alpha t} z(0),$$
$$w(t) = e^{i\beta t} w(0).$$

Any solution $(z(t), w(t))$ to equations (5.1) belongs to an invariant torus

$$T^2 = \{(z, w) \in \mathbb{C}^2 \mid |z| = \text{const}, \ |w| = \text{const}\}.$$

Any such torus is parametrized by arguments of z, w, thus it is a group: $T^2 \simeq \mathbb{R}^2/\mathbb{Z}^2$.

We introduce a new parameter $\tau = \alpha t$, then trajectories (z, w) become images of the line $\{(\tau, (\beta/\alpha)\tau) \mid \tau \in \mathbb{R}\}$ under the immersion

$$(\tau, (\beta/\alpha)\tau) \mapsto (\tau + \mathbb{Z}, (\beta/\alpha)\tau + \mathbb{Z}) \in \mathbb{R}^2/\mathbb{Z}^2,$$

thus immersed submanifolds of the torus.

If the ratio β/α is irrational, then trajectories are everywhere dense in the torus: they form the irrational winding of the torus. In this case, trajectories, i.e., orbits of a vector field, are not submanifolds, but just immersed submanifolds.

Remark. Immersed submanifolds inherit many local properties of submanifolds. In particular, the tangent space to an immersed submanifold $W = \operatorname{Im}\Phi \subset M$, Φ an immersion, is given by

$$T_{\Phi(x)} W = \operatorname{Im} \Phi_{*x}.$$

Further, it is easy to prove the following property of a vector field $V \in \operatorname{Vec} M$:

$$V(q) \in T_q W \quad \forall\, q \in W \quad \Rightarrow \quad q \circ e^{tV} \in W \quad \forall\, q \in W,$$

for all t close enough to 0.

5.3 Corollaries of the Orbit Theorem

Before proving the Orbit Theorem, we obtain several its corollaries.

Let $\mathcal{O}_{q_0}$ be an orbit of a family $\mathcal{F} \subset \operatorname{Vec} M$.

First of all, if $f \in \mathcal{F}$, then $f(q) \in T_q\mathcal{O}_{q_0}$ for all $q \in \mathcal{O}_{q_0}$. Indeed, the trajectory $q \circ e^{tf}$ belongs to the orbit $\mathcal{O}_{q_0}$, thus its velocity vector $f(q)$ is in the tangent space $T_q\mathcal{O}_{q_0}$.

Further, if $f_1, f_2 \in \mathcal{F}$, then $[f_1, f_2](q) \in T_q\mathcal{O}_{q_0}$ for all $q \in \mathcal{O}_{q_0}$. This follows since the vector $[f_1, f_2](q)$ is tangent to the trajectory

$$t \mapsto q \circ e^{tf_1} \circ e^{tf_2} \circ e^{-tf_1} \circ e^{-tf_2} \in \mathcal{O}_{q_0}.$$

Given three vector fields $f_1, f_2, f_3 \in \mathcal{F}$, we have $[f_1, [f_2, f_3]](q) \in T_q\mathcal{O}_{q_0}$, $q \in \mathcal{O}_{q_0}$. Indeed, it follows that $[f_2, f_3](q) \in T_q\mathcal{O}_{q_0}$, $q \in \mathcal{O}_{q_0}$, then all trajectories of the field $[f_2, f_3]$ starting in the immersed submanifold $\mathcal{O}_{q_0}$ do not leave it. Then we repeat the argument of the previous items.

We can go on and consider Lie brackets of arbitrarily high order

$$[f_1, [\dots [f_{k-1}, f_k] \dots]](q)$$

as tangent vectors to $\mathcal{O}_{q_0}$ if $f_i \in \mathcal{F}$. These considerations can be summarized in terms of the Lie algebra of vector fields generated by $\mathcal{F}$:

$$\operatorname{Lie} \mathcal{F} = \operatorname{span}\{[f_1, [\dots [f_{k-1}, f_k] \dots]] \mid f_i \in \mathcal{F},\ k \in \mathbb{N}\} \subset \operatorname{Vec} M,$$

and its evaluation at a point $q \in M$:

$$\operatorname{Lie}_q \mathcal{F} = \{q \circ V \mid V \in \operatorname{Lie} \mathcal{F}\} \subset T_q M.$$

We obtain the following statement.

Corollary 5.1.

$$\operatorname{Lie}_q \mathcal{F} \subset T_q \mathcal{O}_{q_0} \tag{5.2}$$

for all $q \in \mathcal{O}_{q_0}$.

Remark. We show soon that in many important cases inclusion (5.2) turns into equality. In the general case, we have the following estimate:

$$\dim \operatorname{Lie}_q \mathcal{F} \leq \dim \mathcal{O}_{q_0}, \qquad q \in \mathcal{O}_{q_0}.$$

Another important corollary of the Orbit Theorem is the following proposition often used in control theory.

Theorem 5.2 (Rashevsky–Chow). *Let M be a connected smooth manifold, and let $\mathcal{F} \subset \operatorname{Vec} M$. If the family $\mathcal{F}$ is completely nonholonomic:*

$$\operatorname{Lie}_q \mathcal{F} = T_q M \qquad \forall\, q \in M, \tag{5.3}$$

then

$$\mathcal{O}_{q_0} = M \quad \forall\, q_0 \in M. \tag{5.4}$$

Definition 5.2. A family $\mathcal{F} \subset \operatorname{Vec} M$ that satisfies property (5.3) is called *completely nonholonomic* or *bracket-generating*.

Proof. By Corollary 5.1, equality (5.3) means that any orbit $\mathcal{O}_{q_0}$ is an open set in M.

Further, consider the following equivalence relation in M:

$$q_1 \sim q_2 \ \Leftrightarrow q_2 \in \mathcal{O}_{q_1}, \qquad q_1,\ q_2 \in M. \tag{5.5}$$

The manifold M is the union of (naturally disjoint) equivalence classes. Each class is an open subset of M and M is connected. Hence there is only one nonempty class. That is, M is a single orbit $\mathcal{O}_{q_0}$. □

For symmetric families attainable sets coincide with orbits, thus we have the following statement.

Corollary 5.2. *Symmetric bracket-generating families are completely controllable on connected manifolds.*

5.4 Proof of the Orbit Theorem

Introduce the notation:

$$(\operatorname{Ad}\mathcal{P})\mathcal{F} \stackrel{\text{def}}{=} \{(\operatorname{Ad}P)f \mid P \in \mathcal{P},\ f \in \mathcal{F}\} \subset \operatorname{Vec} M.$$

Consider the following subspace of T_qM:

$$\Pi_q \stackrel{\text{def}}{=} \operatorname{span}\{q \circ (\operatorname{Ad}\mathcal{P})\mathcal{F}\}.$$

This space is a candidate for the tangent space $T_q\mathcal{O}_{q_0}$.

Lemma 5.1. $\dim \Pi_q = \dim \Pi_{q_0}$ *for all* $q \in \mathcal{O}_{q_0}$, $q_0 \in M$.

Proof. If $q \in \mathcal{O}_{q_0}$, then $q = q_0 \circ Q$ for some diffeomorphism $Q \in \mathcal{P}$.

Take an arbitrary element $q_0 \circ (\operatorname{Ad}P)f$ in Π_{q_0}, $P \in \mathcal{P}$, $f \in \mathcal{F}$. Then

$$\begin{aligned} Q_*(q_0 \circ (\operatorname{Ad}P)f) &= q_0 \circ (\operatorname{Ad}P)f \circ Q = q_0 \circ P \circ f \circ P^{-1} \circ Q \\ &= (q_0 \circ Q) \circ (Q^{-1} \circ P \circ f \circ P^{-1} \circ Q) \\ &= q \circ \operatorname{Ad}(Q^{-1} \circ P)f \in \Pi_q \end{aligned}$$

since $Q^{-1} \circ P \in \mathcal{P}$.

We have $Q_*\Pi_{q_0} \subset \Pi_q$, thus $\dim \Pi_{q_0} \leq \dim \Pi_q$. But q_0 and q can be switched, that is why $\dim \Pi_q \leq \dim \Pi_{q_0}$. Finally, $\dim \Pi_q = \dim \Pi_{q_0}$. □

Now we prove the Orbit Theorem.

Proof. The manifold M is divided into disjoint equivalence classes of relation (5.5) — orbits $\mathcal{O}_q$. We introduce a new "strong" topology on M in which all orbits are connected components.

For any point $q \in M$, denote $m = \dim \Pi_q$ and pick elements $V_1, \dots, V_m \in (\operatorname{Ad}\mathcal{P})\mathcal{F}$ such that

$$\operatorname{span}(V_1(q), \dots, V_m(q)) = \Pi_q. \tag{5.6}$$

Introduce a mapping:

$$G_q \ : \ (t_1, \dots, t_m) \mapsto q \circ e^{t_1 V_1} \circ \cdots \circ e^{t_m V_m}, \qquad t_i \in \mathbb{R}.$$

We have

$$\left.\frac{\partial G_q}{\partial t_i}\right|_0 = V_i(q),$$

thus in a sufficiently small neighborhood O_0 of the origin $0 \in \mathbb{R}^m$

$$\left.\frac{\partial G_q}{\partial t_1} \wedge \cdots \wedge \frac{\partial G_q}{\partial t_m}\right|_{O_0} \neq 0,$$

i.e., $G_q|_{O_0}$ is an immersion.

The sets of the form $G_q(O_0)$, $q \in M$, are candidates for elements of a topology base on M. We prove several properties of these sets.

(1) Since the mappings G_q are regular, the sets $G_q(O_0)$ are m-dimensional submanifolds of M, may be, for smaller neighborhoods O_0.

(2) We show that $G_q(O_0) \subset \mathcal{O}_q$. Any element of base (5.6) has the form $V_i = (\operatorname{Ad} P_i) f_i$, $P_i \in \mathcal{P}$, $f_i \in \mathcal{F}$. Then

$$e^{tV_i} = e^{t(\operatorname{Ad} P_i) f_i} = e^{t P_i \circ f_i \circ P_i^{-1}} = P_i \circ e^{t f_i} \circ P_i^{-1} \in \mathcal{P},$$

thus

$$G_q(t) = q \circ e^{tV_i} \in \mathcal{O}_q, \qquad t \in O_0.$$

(3) We show that $G_{*t}(T_t \mathbb{R}^m) = \Pi_{G(t)}$, $t \in O_0$. Since $\operatorname{rank} G_{*t}\big|_{O_0} = m$ and $\dim P_{G(t)}\big|_{O_0} = m$, it remains ro prove that $\frac{\partial G_q}{\partial t_i}\Big|_t \in \Pi_{G_q(t)}$ for $t \in O_0$. We have

$$\begin{aligned}
\frac{\partial}{\partial t_i} G_q(t) &= \frac{\partial}{\partial t_i} q \circ e^{t_1 V_1} \circ \cdots \circ e^{t_m V_m} \\
&= q \circ e^{t_1 V_1} \circ \cdots \circ e^{t_i V_i} \circ V_i \circ e^{t_{i+1} V_{i+1}} \circ \cdots \circ e^{t_m V_m} \\
&= q \circ e^{t_1 V_1} \circ \cdots \circ e^{t_i V_i} \circ e^{t_{i+1} V_{i+1}} \circ \cdots \circ e^{t_m V_m} \\
&\qquad \circ\, e^{-t_m V_m} \circ \cdots \circ e^{-t_{i+1} V_{i+1}} \circ V_i \circ e^{t_{i+1} V_{i+1}} \circ \cdots \circ e^{t_m V_m}
\end{aligned}$$

(introduce the notation $Q = e^{t_{i+1} V_{i+1}} \circ \cdots \circ e^{t_m V_m} \in \mathcal{P}$)

$$= G_q(t) \circ Q^{-1} \circ V_i \circ Q = G_q(t) \circ \operatorname{Ad} Q^{-1} V_i \in \Pi_{G_q(t)}.$$

(4) We prove that sets of the form $G_q(O_0)$, $q \in M$, form a topology base in M. It is enough to prove that any nonempty intersection $G_q(O_0) \cap G_{\tilde{q}}(\widetilde{O}_0)$ contains a set of the form $G_{\hat{q}}(\widehat{O}_0)$.

Let a point $\hat{q}$ belong to $G_q(O_0)$. Then $\dim \Pi_{\hat{q}} = \dim \Pi_q = m$. Consider the mapping

$$\begin{aligned}
&G_{\hat{q}} \;:\; (t_1, \ldots, t_m) \mapsto \hat{q} \circ e^{t_1 \widehat{V}_1} \circ \cdots \circ e^{t_m \widehat{V}_m}, \\
&\operatorname{span}(\hat{q} \circ \widehat{V}_1, \ldots, \hat{q} \circ \widehat{V}_m) = \Pi_{\hat{q}}.
\end{aligned}$$

It is enough to prove that for small enough $(t_1, \ldots, t_m)$

$$G_{\hat{q}}(t_1, \ldots, t_m) \in G_q(O_0),$$

then we can replace $G_q(O_0)$ by $G_{\hat{q}}(\widehat{O}_0)$. We do this step by step. Consider the curve $t_1 \mapsto \widehat{q} \circ e^{t_1 \widehat{V}_1}$. By property (3) above, $\widehat{V}_1(q) \in \Pi_q$ for $q \in G(O_0)$ and sufficiently close to $\widehat{q}$. Since $G_q(O_0)$ is a submanifold of M and $\Pi_q = T_q G_q(O_0)$, the curve $\widehat{q} \circ e^{t_1 \widehat{V}_1}$ belongs to $G_q(O_0)$ for sufficiently small $|t_1|$.

We repeat this argument and show that

$$(\widehat{q} \circ e^{t_1 \widehat{V}_1}) \circ e^{t_2 \widehat{V}_2} \in G_q(O_0)$$

for small $|t_1|$, $|t_2|$. We continue this procedure and obtain the inclusion

$$(\widehat{q} \circ e^{t_1 \widehat{V}_1} \circ \cdots \circ e^{t_{m-1} \widehat{V}_{m-1}}) \circ e^{t_m \widehat{V}_m} \in G_q(O_0)$$

for $(t_1, \ldots, t_m)$ sufficiently close to $0 \in \mathbb{R}^m$.

Property (4) follows, and the sets $G_q(O_0)$, $q \in M$, form a topology base on M. We denote by $M^{\mathcal{F}}$ the topological space obtained, i.e., the set M endowed with the "strong" topology just introduced.

(5) For any $q_0 \in M$, the orbit $\mathcal{O}_{q_0}$ is connected, open, and closed in the "strong" topology.

Connectedness: all mappings $t \mapsto q \circ e^{tf}$, $f \in \mathcal{F}$, are continuous in the "strong" topology, thus any point $q \in \mathcal{O}_{q_0}$ can be connected with q_0 by a path continuous in $M^{\mathcal{F}}$.

Openness: for any $q \in \mathcal{O}_{q_0}$, a set of the form $G_q(O_0) \subset \mathcal{O}_{q_0}$ is a neighbourhood of the point q in $M^{\mathcal{F}}$.

Closedness. Let a sequence of points $q_n \in \mathcal{O}_{q_0}$, $n = 1, 2, \ldots$, converge in $M^{\mathcal{F}}$ to a point $q \in M$. We have to show that the limit q is in the orbit $\mathcal{O}_{q_0}$. Fix a base neighbourhood $G_q(O_0)$, then it contains a point of the sequence q_n for a sufficiently large n. We have:

$$q_n \in \mathcal{O}_{q_0}, \quad q_n \in \mathcal{O}_q \qquad \Rightarrow \qquad q \in \mathcal{O}_{q_0},$$

and the closedness follows.

So each orbit $\mathcal{O}_{q_0}$ is a connected component of the topological space $M^{\mathcal{F}}$.

(6) A smooth structure on each orbit $\mathcal{O}_{q_0}$ is defined by choosing $G_q(O_0)$ to be coordinate neighborhoods and G_q^{-1} coordinate mappings. Since $G_q|_{O_0}$ are immersions, then each orbit $\mathcal{O}_{q_0}$ is an immersed submanifold of M. Notice that dimension of these submanifolds may vary for different q_0.

(7) By property (3) above, $T_q \mathcal{O}_{q_0} = \Pi_q$, $q \in \mathcal{O}_{q_0}$.

The Orbit Theorem is proved. □

The description of the tangent space of an orbit given by the Orbit Theorem:

$$T_q \mathcal{O}_{q_0} = q \circ (\operatorname{Ad} \mathcal{P}) \mathcal{F},$$

is rather implicit since the structure of the group $\mathcal{P}$ is quite complex. However, we already obtained the lower estimate

$$\operatorname{Lie}_q \mathcal{F} \subset \operatorname{span}(q \circ (\operatorname{Ad} \mathcal{P}) \mathcal{F}). \tag{5.7}$$

from the Orbit Theorem. Notice that this inclusion can easily be proved directly. We make use of the asymptotic expansion of the field $\operatorname{Ad} e^{tf}\widehat{f} = e^{t \operatorname{ad} f}\widehat{f}$. Take an arbitrary element $\operatorname{ad} f_1 \circ \cdots \circ \operatorname{ad} f_k \widehat{f} \in \operatorname{Lie}\mathcal{F}$, $f_i, \widehat{f} \in \mathcal{F}$. We have $\operatorname{Ad}(e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k})\widehat{f} \in (\operatorname{Ad}\mathcal{P})\mathcal{F}$, thus

$$\begin{aligned} q \circ \left.\frac{\partial^k}{\partial t_1 \cdots \partial t_k}\right|_0 & \operatorname{Ad}(e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k})\widehat{f} \\ &= q \circ \left.\frac{\partial^k}{\partial t_1 \cdots \partial t_k}\right|_0 (e^{t_1 \operatorname{ad} f_1} \circ \cdots \circ e^{t_k \operatorname{ad} f_k})\widehat{f} \\ &= q \circ \operatorname{ad} f_1 \circ \cdots \circ \operatorname{ad} f_k \ \widehat{f} \in \operatorname{span}(q \circ (\operatorname{Ad}\mathcal{P})\mathcal{F}). \end{aligned}$$

Now we consider a situation where inclusion (5.7) is strict.

Example 5.3. Let $M = \mathbb{R}^2$, $\mathcal{F} = \left\{\frac{\partial}{\partial x^1},\ a(x^1)\frac{\partial}{\partial x^2}\right\}$, where the function $a \in C^\infty(\mathbb{R})$, $a \not\equiv 0$, has a compact support.

It is easy to see that the orbit $\mathcal{O}_x$ through any point $x \in \mathbb{R}^2$ is the whole plane $\mathbb{R}^2$. Indeed, the family $\mathcal{F} \cup (-\mathcal{F})$ is completely controllable in the plane. Given an initial point $x_0 = (x_0^1, x_0^2)$ and a terminal point $x_1 = (x_1^1, x_1^2)$, we can steer x_0 to x_1: first we go from x_0 by a field $\pm\frac{\partial}{\partial x^1}$ to a point $(\tilde{x}^1, x_0^2)$ with $a(\tilde{x}^1) \neq 0$, then we go by a field $\pm a(x^1)\frac{\partial}{\partial x^2}$ to a point $(\tilde{x}^1, x_1^2)$, and finally we reach (x_1^1, x_1^2) along $\pm\frac{\partial}{\partial x^1}$.

On the other hand, we have

$$\dim \operatorname{Lie}_{(x^1,x^2)}(\mathcal{F}) = \begin{cases} 1, & x^1 \notin \operatorname{supp} a, \\ 2, & a(x^1) \neq 0. \end{cases}$$

That is, $x \circ (\operatorname{Ad}\mathcal{P})\mathcal{F} = T_x\mathbb{R}^2 \neq \operatorname{Lie}_x \mathcal{F}$ if $x^1 \notin \operatorname{supp} a$.

Although, such example is essentially non-analytic. In the analytic case, inclusion (5.7) turns into equality. We prove this statement in the next section.

5.5 Analytic case

The set $\operatorname{Vec} M$ is not just a Lie algebra (i.e., a vector space close under the operation of Lie bracket), but also a *module* over $C^\infty(M)$: any vector field $V \in \operatorname{Vec} M$ can be multiplied by a function $a \in C^\infty(M)$, and the resulting vector field $aV \in \operatorname{Vec} M$. If vector fields are considered as derivations of $C^\infty(M)$, then the product of a function a and a vector field V is the vector field

$$(aV)b = a \cdot (Vb), \qquad b \in C^\infty(M).$$

In local coordinates, each component of V at a point $q \in M$ is multiplied by $a(q)$.

A submodule $\mathcal{V} \subset \operatorname{Vec} M$ is called *finitely generated* over $C^\infty(M)$ if it has a global basis of vector fields:

$$\exists\ V_1, \ldots, V_k \in \operatorname{Vec} M \quad \text{such that} \quad \mathcal{V} = \left\{ \sum_{i=1}^k a_i V_i \mid a_i \in C^\infty(M) \right\}.$$

Lemma 5.2. *Let $\mathcal{V} \subset \operatorname{Vec} M$ be a finitely generated submodule over $C^\infty(M)$. Assume that*

$$(\operatorname{ad} X)\mathcal{V} = \{(\operatorname{ad} X)V \mid V \in \mathcal{V}\} \subset \mathcal{V}$$

for a vector field $X \in \operatorname{Vec} M$. Then

$$\operatorname{Ad} e^{tX} \mathcal{V} = \mathcal{V}.$$

Proof. Let $V_1, \dots, V_k$ be a basis of $\mathcal{V}$. By the hypothesis of the lemma,

$$[X, V_i] = \sum_{j=1}^{k} a_{ij} V_j \tag{5.8}$$

for some functions $a_{ij} \in C^\infty(M)$. We have to prove that the vector fields

$$V_i(t) = (\operatorname{Ad} e^{tX}) V_i = e^{t \operatorname{ad} X} V_i, \qquad t \in \mathbb{R},$$

can be expressed as a linear combination of the fields V_i with coefficients from $C^\infty(M)$.

We define an ODE for $V_i(t)$:

$$\begin{aligned} \dot V_i(t) &= e^{t \operatorname{ad} X}[X, V_i] = e^{t \operatorname{ad} X} \sum_{j=1}^{k} a_{ij} V_j \\ &= \sum_{j=1}^{k} \left(e^{tX} a_{ij}\right) V_j(t). \end{aligned}$$

For a fixed $q \in M$, define the $k \times k$ matrix:

$$A(t) = (a_{ij}(t)), \qquad a_{ij}(t) = e^{tX} a_{ij}, \quad i,\ j = 1, \dots, k.$$

Then we have a linear system of ODEs:

$$\dot V_i(t) = \sum_{j=1}^{k} a_{ij}(t) V_j(t). \tag{5.9}$$

Find a fundamental matrix Γ of this system:

$$\dot\Gamma = A(t)\Gamma, \qquad \Gamma(0) = \operatorname{Id}.$$

Since $A(t)$ smoothly depends on q, then Γ depends smoothly on q as well:

$$\Gamma(t) = (\gamma_{ij}(t)), \qquad \gamma_{ij}(t) \in C^\infty(M), \quad i,\ j = 1, \dots, k, \quad t \in \mathbb{R}.$$

Now solutions of the linear system (5.9) can be written as follows:

$$V_i(t) = \sum_{j=1}^{k} \gamma_{ij}(t) V_j(0).$$

But $V_i(0) = V_i$ are the generators of the module, and the required decomposition of $V_i(t)$ along the generators is obtained. □

A submodule $\mathcal{V} \subset \operatorname{Vec} M$ is called *locally finitely generated* over $C^\infty(M)$ if any point $q \in M$ has a neighborhood $O \subset M$ in which the restriction $\mathcal{F}|_O$ is finitely generated over $C^\infty(O)$, i.e., has a basis of vector fields.

Theorem 5.3. *Let $\mathcal{F} \subset \operatorname{Vec} M$. Suppose that the module $\operatorname{Lie} \mathcal{F}$ is locally finitely generated over $C^\infty(M)$. Then*

$$T_q \mathcal{O}_{q_0} = \operatorname{Lie}_q \mathcal{F}, \qquad q \in \mathcal{O}_{q_0} \tag{5.10}$$

for any orbit $\mathcal{O}_{q_0}$, $q_0 \in M$, of the family $\mathcal{F}$.

We prove this theorem later, but now obtain from it the following consequence.

Corollary 5.3. *If M and $\mathcal{F}$ are real analytic, then equality* (5.10) *holds.*

Proof. In the analytic case, $\operatorname{Lie} \mathcal{F}$ is locally finitely generated. Indeed, any module generated by analytic vector fields is locally finitely generated. This is Nötherian property of the ring of germs of analytic functions. □

Now we prove Theorem 5.3.

Proof. By the Orbit Theorem,

$$T_q \mathcal{O}_{q_0} = \operatorname{span} \left\{ q \circ \operatorname{Ad} \left(e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k} \right) \widehat{f} \mid f_i,\ \widehat{f} \in \mathcal{F},\ t_k \in \mathbb{R},\ k \in \mathbb{N} \right\}. \tag{5.11}$$

By definition of the Lie algebra $\operatorname{Lie} \mathcal{F}$,

$$(\operatorname{ad} f) \operatorname{Lie} \mathcal{F} \subset \operatorname{Lie} \mathcal{F} \qquad \forall\ f \in \mathcal{F}.$$

Apply Lemma 5.2 for the locally finitely generated $C^\infty(M)$-module $\mathcal{V} = \operatorname{Lie} \mathcal{F}$. We obtain

$$\left(\operatorname{Ad} e^{tf} \right) \operatorname{Lie} \mathcal{F} \subset \operatorname{Lie} \mathcal{F} \qquad \forall\ f \in \mathcal{F}.$$

That is why

$$\operatorname{Ad} \left(e^{t_1 f_1} \circ \cdots \circ e^{t_k f_k} \right) \widehat{f} = \operatorname{Ad} e^{t_1 f_1} \circ \cdots \circ \operatorname{Ad} e^{t_k f_k} \widehat{f} \in \operatorname{Lie} \mathcal{F}$$

for any f_i, $\widehat{f} \in \mathcal{F}$, $t_k \in \mathbb{R}$. In view of equality (5.11),

$$T_q \mathcal{O}_{q_0} \subset \operatorname{Lie}_q \mathcal{F}.$$

But the reverse inclusion (5.7) was already obtained. Thus $T_q \mathcal{O}_{q_0} = \operatorname{Lie}_q \mathcal{F}$.

Another proof of the theorem can be obtained via local convergence of the exponential series in the analytic case. □

5.6 Frobenius Theorem

We apply the Orbit Theorem to obtain the classical Frobenius Theorem as a corollary.

Definition 5.3. A *distribution* $\Delta \subset TM$ on a smooth manifold M is a family of linear subspaces $\Delta_q \subset T_qM$ smoothly depending on a point $q \in M$. Dimension of the subspaces Δ_q, $q \in M$, is assumed constant.

Geometrically, at each point $q \in M$ there is attached a space $\Delta_q \subset T_qM$, i.e., we have a field of tangent subspaces on M.

Definition 5.4. A distribution Δ on a manifold M is called *integrable* if for any point $q \in M$ there exists an immersed submanifold $N_q \subset M$, $q \in N_q$, such that

$$T_{q'} N_q = \Delta_{q'} \qquad \forall\, q' \in N_q.$$

The submanifold N_q is called an *integral manifold* of the distribution Δ through the point q.

In other words, integrability of a distribution $\Delta \subset TM$ means that through any point $q \in M$ we can draw a submanifold N_q whose tangent spaces are elements of the distribution Δ.

Remark. If $\dim \Delta_q = 1$, then Δ is integrable by Theorem 1.2 on existence and uniqueness of solutions of ODEs. Indeed, in a neighborhood of any point in M, we can find a base of the distribution Δ, i.e., a vector field $V \in \operatorname{Vec} M$ such that $\Delta_q = \operatorname{span}(V(q))$, $q \in M$. Then trajectories of the ODE $\dot{q} = V(q)$ are one-dimensional submanifolds with tangent spaces Δ_q.

But in the general case ($\dim \Delta_q > 1$), a distribution Δ may be nonintegrable. Indeed, consider the family of vector fields tangent to Δ:

$$\overline{\Delta} = \{V \in \operatorname{Vec} M \mid V(q) \in \Delta_q \quad \forall\, q \in M\}.$$

Assume that the distribution Δ is integrable. Any vector field from the family $\overline{\Delta}$ is tangent to integral manifolds N_q, thus the orbit $\mathcal{O}_q$ of the family $\overline{\Delta}$ restricted to a small enough neighborhood of q is contained in the integral manifold N_q. Moreover, since $\dim \mathcal{O}_q \geq \dim \Delta_q = \dim N_q$, then locally $\mathcal{O}_q = N_q$: we can go in N_q in any direction along vector fields of the family $\overline{\Delta}$. By the Orbit Theorem, $T_q\mathcal{O}_q \supset \operatorname{Lie}_q \overline{\Delta}$, that is why

$$\operatorname{Lie}_q \overline{\Delta} = \Delta_q.$$

This means that

$$[V_1, V_2] \in \overline{\Delta} \quad \forall\, V_1,\ V_2 \in \overline{\Delta}. \tag{5.12}$$

Let $\dim \Delta_q = k$. In a neighborhood O_{q_0} of a point $q_0 \in M$ we can find a base of the distribution Δ:

$$\Delta_q = \operatorname{span}(f_1(q), \dots, f_k(q)) \quad \forall\, q \in O_{q_0}.$$

Then condition (5.12) reads as *Frobenius condition*:

$$[f_i, f_j] = \sum_{l=1}^{k} c_{ij}^l f_l, \quad c_{ij}^l \in C^\infty(O_{q_0}). \tag{5.13}$$

We have shown that integrability of a distribution Δ implies Frobenius condition for its base.

Conversely, if condition (5.13) holds in a neighborhood of any point $q_0 \in M$, then $\operatorname{Lie}(\overline{\Delta}) = \overline{\Delta}$. Thus $\operatorname{Lie}(\overline{\Delta})$ is a locally finitely generated module over $C^\infty(M)$. By Theorem 5.3,

$$T_q \mathcal{O}_{q_0} = \operatorname{Lie}_q \overline{\Delta}, \qquad q \in \mathcal{O}_{q_0}.$$

So

$$T_q \mathcal{O}_{q_0} = \Delta_q, \qquad q \in \mathcal{O}_{q_0},$$

i.e., the orbit $\mathcal{O}_{q_0}$ is an integral manifold of Δ through q_0. We proved the following proposition.

Theorem 5.4 (Frobenius). *A distribution $\Delta \subset TM$ is integrable if and only if Frobenius condition* (5.13) *holds for any base of Δ in a neighborhood of any point $q_0 \in M$.*

Remarks. (1) In view of the Leibniz rule

$$[f, ag] = (fa)g + a[f, g], \qquad f,\ g \in \operatorname{Vec} M, \quad a \in C^\infty(M),$$

Frobenius condition is independent on the choice of a base $f_1, \dots, f_k$: if it holds in one base, then it also holds in any other base.

(2) One can prove analogously the generalized Frobenius theorem for "distributions with variable $\dim \Delta_q$". If we do not demand that $\dim \Delta_q$ is constant, then Frobenius condition implies integrability; but dimension of integrable manifolds becomes, in general, different, although it stays constant along orbits of $\overline{\Delta}$. This is a generalization of phase portraits of vector fields. Although, notice once more that in general distributions with $\dim \Delta_q > 1$ are nonintegrable.

5.7 State equivalence of control systems

In this section we consider one more application of the Orbit Theorem — to the problem of equivalence of control systems (or families of vector fields).

Let U be an arbitrary index set. Consider two families of vector fields on smooth manifolds M and N parametrized be the same set U:

$$f_U = \{f_u \mid u \in U\} \subset \operatorname{Vec} M,$$
$$g_U = \{g_u \mid u \in U\} \subset \operatorname{Vec} N.$$

Take any pair of points $x_0 \in M$, $y_0 \in N$, and assume that the families f_U, g_U are bracket-generating:

$$\operatorname{Lie}_{x_0} f_U = T_{x_0} M, \qquad \operatorname{Lie}_{y_0} g_U = T_{y_0} N.$$

Definition 5.5. Families f_U and g_U are called *locally state equivalent* if there exists a local diffeomorphism

$$\Phi : O_{x_0} \subset M \to O_{y_0} \subset N,$$
$$\Phi : x_0 \mapsto y_0,$$

that transforms one family to another:

$$\Phi_* f_u = g_u \qquad \forall\, u \in U.$$

Notation: $(f_U, x_0) \simeq (g_U, y_0)$.

Remark. Here we consider only smooth transformations of state $x \mapsto y$, while the controls u do not change. That is why this kind of equivalence is called state equivalence. We already studied state equivalence of nonlinear and linear systems, both local and global, see Chapter 4.

Now, we first try to find necessary conditions for local equivalence of systems f_U and g_U. Assume that

$$f_U \simeq g_U.$$

By invariance of Lie bracket, we get

$$\Phi_*[f_{u_1}, f_{u_2}] = [\Phi_* f_{u_1}, \Phi_* f_{u_2}] = [g_{u_1}, g_{u_2}], \qquad u_1,\ u_2 \in U,$$

i.e., relations between Lie brackets of vector fields of the equivalent families f_U and g_U must be preserved. We collect all relations between these Lie brackets at one point: define the systems of tangent vectors

$$\xi_{u_1 \dots u_k} = [f_{u_1}, [\dots, f_{u_k}] \dots](x_0) \in T_{x_0}M,$$
$$\eta_{u_1 \dots u_k} = [g_{u_1}, [\dots, g_{u_k}] \dots](y_0) \in T_{y_0}N.$$

Then we have

$$\Phi_*\big|_{x_0} \xi_{u_1 \dots u_k} = \eta_{u_1 \dots u_k} \quad u_1, \dots, u_k \in U,\ k \in \mathbb{N}.$$

Now we can state a necessary condition for local equivalence of families f_U and g_U in terms of the linear isomorphism

$$\Phi_*\big|_{x_0} = A : T_{x_0}M \leftrightarrow T_{y_0}N.$$

If $f_U \simeq g_U$, then there exists a linear isomorphism

$$A : T_{x_0}M \leftrightarrow T_{y_0}N$$

that maps the configuration of vectors $\{\xi_{u_1 \dots u_k}\}$ to the configuration $\{\eta_{u_1 \dots u_k}\}$. It turns out that in the analytic case this condition is sufficient. I.e., in the analytic case the combinations of partial derivatives of vector fields f_u, $u \in U$, that enter $\{\xi_{u_1 \dots u_k}\}$, form a complete system of state invariants of a family f_U.

Theorem 5.5. *Let f_U and g_U be real analytic and bracket-generating families of vector fields on real analytic manifolds M and N respectively. Let $x_0 \in M$, $y_0 \in N$. Then $(f_U, x_0) \simeq (g_U, y_0)$ if and only if there exists a linear isomorphism*

$$A \ : \ T_{x_0}M \leftrightarrow T_{y_0}N$$

such that

$$A\{\xi_{u_1\dots u_k}\} = \{\eta_{u_1\dots u_k}\} \qquad \forall\ u_1, \dots, u_k \in U, \quad k \in \mathbb{N}. \tag{5.14}$$

Remark. If in addition M, N are simply connected and the fields f_U, g_U are complete, then we have the global equivalence.

Before proving Theorem 5.5, we reformulate condition (5.14) and provide a method to check it.

Let a family f_U be bracket-generating:

$$\operatorname{span}\{\xi_{u_1\dots u_k} \mid u_1, \dots, u_k \in U,\ k \in \mathbb{N}\} = T_{x_0}M.$$

We can choose a basis:

$$\operatorname{span}(\xi_{\bar\alpha_1}, \dots, \xi_{\bar\alpha_n}) = T_{x_0}M, \qquad \bar\alpha_i = (u_{1i}, \dots, u_{ki}), \quad i = 1, \dots, n, \tag{5.15}$$

and express all vectors in the configuration ξ through the base vectors:

$$\xi_{u_1\dots u_k} = \sum_{i=1}^{n} c^i_{u_1\dots u_k} \xi_{\bar\alpha_i}. \tag{5.16}$$

If there exists a linear isomorphism $A \ : \ T_{x_0}M \leftrightarrow T_{y_0}N$ with (5.14), then the vectors

$$\eta_{\bar\alpha_i}, \qquad i = 1, \dots, n,$$

should form a basis of $T_{y_0}N$:

$$\operatorname{span}(\eta_{\bar\alpha_1}, \dots, \eta_{\bar\alpha_n}) = T_{y_0}N, \tag{5.17}$$

and all vectors of the configuration η should be expressed through the base vectors with the same coefficients as the configuration ξ, see (5.16):

$$\eta_{u_1\dots u_k} = \sum_{i=1}^{n} c^i_{u_1\dots u_k} \eta_{\bar\alpha_i}. \tag{5.18}$$

It is easy to see the converse implication: if we can choose bases in $T_{x_0}M$ and $T_{y_0}N$ from the configurations ξ and η as in (5.15) and (5.17) such that decompositions (5.16) and (5.18) with the same coefficients $c^i_{u_1\dots u_k}$ hold, then there exists a linear isomorphism A with (5.14). Indeed, we define then the isomorphism on the bases:

$$A \ : \ \xi_{\bar\alpha_i} \mapsto \eta_{\bar\alpha_1}, \qquad i = 1, \dots, n.$$

We can obtain one more reformulation via the following agreement. Configurations $\{\xi_{u_1\dots u_k}\}$ and $\{\eta_{u_1\dots u_k}\}$ are called equivalent if the sets of relations $K(f_U)$ and $K(g_U)$ between elements of these configurations coincide: $K(f_U) = K(g_U)$. We denote here by $K(f_U)$ the set of all systems of coefficients such that the corresponding linear combinations vanish:

$$K(f_U) = \left\{ (b_{u_1\dots u_k}) \mid \sum_{u_1\dots u_k} b_{u_1\dots u_k} \xi_{u_1\dots u_k} = 0 \right\}.$$

Then Theorem 5.5 can be expressed in the following form.

Nagano Principle. *All local information about bracket-generating families of analytic vector fields is contained in Lie brackets.*

Notice, although, that the configuration $\xi_{u_1\dots u_k}$ or the system of relations $K(f_U)$ are, in general, immense and cannot be easily characterized. Thus Nagano Principle cannot usually be applied directly to describe properties of control systems, but it is an important guiding principle.

Now we prove Theorem 5.5.

Proof. Necessity was already shown. We prove sufficiency by reduction to the Orbit Theorem. For this we construct an auxiliary system on the Cartesian product

$$M \times N = \{(x, y) \mid x \in M,\ y \in N\}.$$

For vector fields $f \in \operatorname{Vec} M$, $g \in \operatorname{Vec} N$, define their direct product $f \times g \in \operatorname{Vec}(M \times N)$ as the derivation

$$(f \times g)a\big|_{(x,y)} = \left\langle \frac{\partial a}{\partial x}, f(x) \right\rangle + \left\langle \frac{\partial a}{\partial y}, g(y) \right\rangle, \tag{5.19}$$

so projection of $f \times g$ to M is f, and projection to N is g. Finally, we define the direct product of systems f_U and g_U as

$$f_U \times g_U = \{f_u \times g_u \mid u \in U\} \subset \operatorname{Vec}(M \times N).$$

We suppose that there exists a linear isomorphism $A : T_{x_0}M \leftrightarrow T_{y_0}N$ that maps the configuration ξ to η as in (5.14), and construct the local equivalence $f_U \simeq g_U$.

In view of definition (5.19), Lie bracket in the family $f_U \times g_U$ is computed as

$$[f_{u_1} \times g_{u_1}, f_{u_2} \times g_{u_2}] = [f_{u_1}, f_{u_2}] \times [g_{u_1}, g_{u_2}], \qquad u_1,\ u_2 \in U,$$

thus

$$\begin{aligned}
&[f_{u_1} \times g_{u_1}, [\dots, f_{u_k} \times g_{u_k}] \dots](x_0, y_0) \\
&\quad = [f_{u_1}, [\dots, f_{u_k}] \dots](x_0) \times [g_{u_1}, [\dots, g_{u_k}] \dots](y_0) \\
&\quad = \xi_{u_1\dots u_k} \times \eta_{u_1\dots u_k} = \xi_{u_1\dots u_k} \times A\xi_{u_1\dots u_k}, \qquad u_1, \dots, u_k \in U,\ k \in \mathbb{N}.
\end{aligned}$$

That is why

$$\dim \mathrm{Lie}_{(x_0,y_0)}(f_U \times g_U) = n,$$

where $n = \dim M$. By the analytic version of the Orbit Theorem (Corollary 5.3) for the family $f_U \times g_U \subset \mathrm{Vec}(M \times N)$, the orbit $\mathcal{O}$ of $f_U \times g_U$ through the point (x_0, y_0) is an n-dimensional immersed submanifold (thus, locally a submanifold) of $M \times N$. The tangent space of the orbit is

$$\begin{aligned} T_{(x_0,y_0)}\mathcal{O} &= \mathrm{span}(\xi_{u_1\ldots u_k} \times A\xi_{u_1\ldots u_k}) \\ &= \mathrm{span}\{v \times Av \mid v \in T_{x_0}\} \subset T_{(x_0,y_0)}M \times N = T_{x_0}M \times T_{y_0}N, \end{aligned}$$

i.e., the graph of the linear isomorphism A. Consider the canonical projections onto the factors:

$$\begin{aligned} \pi_1 &: M \times N \to M, \qquad \pi_1(x,y) = x, \\ \pi_2 &: M \times N \to N, \qquad \pi_2(x,y) = y. \end{aligned}$$

The restrictions $\pi_1|_{\mathcal{O}}$, $\pi_2|_{\mathcal{O}}$ are local diffeomorphisms since the differentials

$$\begin{aligned} \pi_{1*}|_{(x_0,y_0)} &: (v, Av) \mapsto v, \qquad v \in T_{x_0}M, \\ \pi_{2*}|_{(x_0,y_0)} &: (v, Av) \mapsto Av, \qquad v \in T_{x_0}M, \end{aligned}$$

are one-to-one.

Now $\Phi = \pi_2 \circ (\pi_1|_{\mathcal{O}})^{-1}$ is a local diffeomorphism from M to N with the graph $\mathcal{O}$, and

$$\Phi_* = \pi_{2*} \circ (\pi_1|_{\mathcal{O}})_*^{-1} : f_u \mapsto g_u, \qquad u \in U.$$

Consequently, $(f_U, x_0) \simeq (g_U, y_0)$. □

Chapter 6

Rotations of the rigid body

In this chapter we consider rotations of a rigid body around a fixed point. That is, we study motions of a body in the three-dimensional space such that:

- distances between all points in the body remain fixed (rigidity), and
- there is a point in the body that stays immovable during motion (fixed point).

We consider both free motions (in the absence of external forces) and controlled motions (when external forces are applied in order to bring the body to a desired state).

Such system is a very simplified model of a satellite in the space rotating around its center of mass.

For details about ODEs describing rotations of the rigid body, see [3].

6.1 State space

The state of the rigid body is determined by its position and velocity.

We fix an orthonormal frame attached to the body at the fixed point (the moving frame), and an orthonormal frame attached to the ambient space at the fixed point of the body (the fixed frame). The set of positions of the rigid body is the set of all orthonormal frames in the three-dimensional space with positive orientation. This set can be identified with SO(3), the group of linear orthogonal orientation-preserving transformations of $\mathbb{R}^3$, or, equivalently, with the group of 3×3 orthogonal unimodular matrices:

$$\begin{aligned} \mathrm{SO}(3) &= \{Q : \mathbb{R}^3 \to \mathbb{R}^3 \mid (Qx, Qy) = (x, y),\ \det Q = 1\} \\ &= \{Q : \mathbb{R}^3 \to \mathbb{R}^3 \mid QQ^* = \mathrm{Id},\ \det Q = 1\}. \end{aligned}$$

The mapping $Q : \mathbb{R}^3 \to \mathbb{R}^3$ transforms the moving frame to the fixed frame.

Remark. We denote above the standard inner product in $\mathbb{R}^3$ by $(\cdot,\cdot)$. If a pair of vectors $x, y \in \mathbb{R}^3$ have coordinates $x = (x_1, x_2, x_3)$, $y = (y_1, y_2, y_3)$ in some orthonormal frame, then $(x, y) = x_1y_1 + x_2y_2 + x_3y_3$.

Notice that the set of positions of the rigid body $\mathrm{SO}(3)$ is not a linear space, but a nontrivial smooth manifold.

Now we describe velocities of the rigid body. Let $Q_t \in \mathrm{SO}(3)$ be position of the body at a moment of time t. Since the operators $Q_t : \mathbb{R}^3 \to \mathbb{R}^3$ are orthogonal, then

$$(Q_t x, Q_t y) = (x, y), \qquad x,\ y \in \mathbb{R}^3, \quad t \in \mathbb{R}.$$

We differentiate this equality w.r.t. t and obtain

$$(\dot{Q}_t x, Q_t y) + (Q_t x, \dot{Q}_t y) = 0. \tag{6.1}$$

The matrix

$$\Omega_t = Q_t^{-1}\dot{Q}_t$$

is called the *angular velocity* of the body in the moving frame. Since

$$\dot{Q}_t = Q_t\Omega_t,$$

then equality (6.1) reads

$$(Q_t\Omega_t x, Q_t y) + (Q_t x, Q_t\Omega_t y) = 0,$$

whence by orthogonality

$$(\Omega_t x, y) + (x, \Omega_t y) = 0,$$

i.e.,

$$\Omega_t^* = -\Omega_t,$$

the matrix Ω_t is antisymmetric. So velocities of the rigid body have the form

$$\dot{Q}_t = Q_t\Omega_t, \qquad \Omega_t^* = -\Omega_t.$$

In other words, we found the tangent space

$$T_Q\,\mathrm{SO}(3) = \{Q\Omega \mid \Omega^* = -\Omega\}, \qquad Q \in \mathrm{SO}(3).$$

The space of antisymmetric 3×3 matrices is denoted by $\mathrm{so}(3)$, it is the tangent space to $\mathrm{SO}(3)$ at the identity:

$$\mathrm{so}(3) = \{\Omega : \mathbb{R}^3 \to \mathbb{R}^3 \mid \Omega^* = -\Omega\} = T_{\mathrm{Id}}\,\mathrm{SO}(3).$$

The space $\mathrm{so}(3)$ is the Lie algebra of the Lie group $\mathrm{SO}(3)$.

To each antisymmetric matrix $\Omega \in \mathrm{so}(3)$, we associate a vector $\omega \in \mathbb{R}^3$:

$$\Omega \sim \omega, \qquad \Omega = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}. \tag{6.2}$$

Then the action of the operator Ω on a vector $x \in \mathbb{R}^3$ can be represented via the cross product in $\mathbb{R}^3$:

$$\Omega x = \omega \times x, \qquad x \in \mathbb{R}^3.$$

Let x be a point in the rigid body. Then its position in the ambient space $\mathbb{R}^3$ is $Q_t x$. Further, velocity of this point is

$$\dot{Q}_t x = Q_t \Omega_t x = Q_t(\omega_t \times x).$$

ω_t is the vector of angular velocity of the point x in the moving frame: if we fix the moving frame Q_t at one moment of time t, then the instantaneous velocity of the point x at the moment of time t in the moving frame is $Q_t^{-1}\dot{Q}_t x = \Omega_t x = \omega_t \times x$, i.e., the point x rotates around the line through ω_t with the angular velocity $\|\omega_t\|$.

Introduce the following scalar product of matrices $\Omega = (\Omega_{ij}) \in \mathrm{so}(3)$:

$$\langle \Omega^1, \Omega^2 \rangle = -\frac{1}{2}\operatorname{tr}(\Omega^1\Omega^2) = \frac{1}{2}\sum_{i,j=1}^{3} \Omega^1_{ij}\Omega^2_{ij} = \sum_{i<j} \Omega^1_{ij}\Omega^2_{ij}.$$

This product is compatible with identification of 3×3 antisymmetric matrices and 3-dimensional vectors (6.2):

$$\begin{aligned} &\langle \Omega^1, \Omega^2 \rangle = (\omega^1, \omega^2), \\ &\Omega^i \sim \omega^i, \qquad \Omega^i \in \mathrm{so}(3), \quad \omega^i \in \mathbb{R}^3, \quad i = 1, 2. \end{aligned}$$

Moreover, this product is invariant in the following sense:

$$\langle (\operatorname{Ad} Q)\Omega^1, (\operatorname{Ad} Q)\Omega^2 \rangle = \langle \Omega^1, \Omega^2 \rangle, \qquad Q \in \mathrm{SO}(3), \quad \Omega^1,\ \Omega^2 \in \mathrm{so}(3), \tag{6.3}$$

i.e., $\operatorname{Ad} Q : \mathrm{so}(3) \to \mathrm{so}(3)$ is an orthogonal transformation w.r.t. $\langle \cdot, \cdot \rangle$. Indeed:

$$\operatorname{tr}((\operatorname{Ad} Q)\Omega^1(\operatorname{Ad} Q)\Omega^2) = \operatorname{tr}(Q\Omega^1 Q^{-1} Q \Omega^2 Q^{-1}) = \operatorname{tr}(Q\Omega^1\Omega^2 Q^{-1}) = \operatorname{tr}(\Omega^1\Omega^2)$$

by invariance of trace.

Now we derive the infinitesimal version of invariance (6.3). Take an arbitrary $\Omega \in \mathrm{so}(3)$ and consider a smooth curve $Q_t \in \mathrm{SO}(3)$ that starts from identity with the velocity Ω:

$$\dot{Q}_0 = \Omega, \qquad Q_0 = \mathrm{Id}.$$

Then

$$\left.\frac{d}{dt}\right|_0 \operatorname{Ad} Q_t = \operatorname{ad}\Omega,$$

and differentiation of (6.3) w.r.t. t at $t = 0$ yields the equality:

$$\langle (\operatorname{ad}\Omega)\Omega^1, \Omega^2 \rangle + \langle \Omega^1, (\operatorname{ad}\Omega)\Omega^2 \rangle = 0, \qquad \Omega,\ \Omega^1,\ \Omega^2 \in \mathrm{so}(3),$$

i.e., $\operatorname{ad}\Omega : \mathrm{so}(3) \to \mathrm{so}(3)$ is antisymmetric w.r.t. $\langle \cdot, \cdot \rangle$.

Equality (6.4) can be rewritten in terms of cross product:

$$(\omega \times \omega^1, \omega^2) + (\omega^1, \omega \times \omega^2) = 0, \qquad \omega,\ \omega^1, \omega^2 \in \mathbb{R}^3. \tag{6.4}$$

6.2 Euler equations

We derive equations of motion of the rigid body from the least action principle.

Let the distribution of mass in the rigid body have density $\rho(x)$, where $\rho : \mathbb{R}^3 \to \mathbb{R}_+$ is an integrable nonnegative function with compact support. Let $Q_t \in \mathrm{SO}(3)$ be position and $\Omega_t \in \mathrm{so}(3)$ angular velocity of the body so that

$$\dot{Q}_t = Q_t \Omega_t. \tag{6.5}$$

Take a point x in the body. Then position of this point in the ambient space is $Q_t x$, and velocity of this point is $\dot{Q}_t x$. Distribution of the kinetic energy in the body has density $\frac{1}{2}\rho(x)(\dot{Q}_t x, \dot{Q}_t x)$, thus the total kinetic energy of the body at a moment of time t is

$$j(\Omega_t) = \frac{1}{2}\int_{\mathbb{R}^3} \rho(x)(Q_t\Omega_t x, Q_t\Omega_t x)\, dx = \frac{1}{2}\int_{\mathbb{R}^3} \rho(x)(\Omega_t x, \Omega_t x)\, dx,$$

i.e., a quadratic form $j = j(\Omega_t)$ on the space $\mathrm{so}(3)$. The corresponding bilinear form can be written as

$$\int_{\mathbb{R}^3} \rho(x)(\Omega^1 x, \Omega^2 x)\, dx = \langle A\Omega^1, \Omega^2\rangle, \qquad \Omega^1,\ \Omega^2 \in \mathrm{so}(3)$$

for some linear symmetric positive definite operator

$$A : \mathrm{so}(3) \to \mathrm{so}(3), \qquad A = A^* > 0,$$

called inertia tensor of the rigid body. Finally, the functional of action has the form

$$J(\Omega_\cdot) = \int_0^{t_1} j(\Omega_t)\, dt = \frac{1}{2}\int_0^{t_1} \langle A\Omega_t, \Omega_t\rangle\, dt,$$

where 0 and t_1 are the initial and terminal moments of motion.

Let Q_0 and Q_{t_1} be the initial and terminal positions of the moving body. By the least action principle, the motion Q_t, $t \in [0, t_1]$, of the body should be an extremal of the following problem:

$$\begin{aligned} &J(\Omega_\cdot) \to \min, \\ &\dot{Q}_t = Q_t\Omega_t, \qquad Q_0,\ Q_{t_1} \text{ fixed.} \end{aligned} \tag{6.6}$$

We find these extremals.

Let Ω_t be angular velocity along the reference trajectory Q_t, then

$$Q_0^{-1} \circ Q_{t_1} = \overrightarrow{\exp} \int_0^{t_1} \Omega_t\, dt.$$

Consider an arbitrary small perturbation of the angular velocity:

$$\Omega_t + \varepsilon U_t + O(\varepsilon^2), \qquad \varepsilon \to 0.$$

In order that such perturbation was admissible, the starting point and endpoint of the corresponding trajectory should not depend on ε:

$$Q_0^{-1} \circ Q_{t_1} = \overrightarrow{\exp} \int_0^{t_1} \left(\Omega_t + \varepsilon U_t + O(\varepsilon^2)\right) \, dt,$$

thus

$$0 = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} Q_0^{-1} \circ Q_{t_1} = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \overrightarrow{\exp} \int_0^{t_1} \left(\Omega_t + \varepsilon U_t + O(\varepsilon^2)\right) \, dt. \tag{6.7}$$

By formula (2.26) of derivative of a flow w.r.t. parameter, the right-hand side above is equal to

$$\begin{aligned}
&\int_0^{t_1} \operatorname{Ad}\left(\overrightarrow{\exp} \int_0^t \Omega_\tau \, d\tau\right) U_t \, dt \circ \overrightarrow{\exp} \int_0^{t_1} \Omega_t \, dt \\
&\quad = \int_0^{t_1} \operatorname{Ad}\left(Q_0^{-1} \circ Q_t\right) U_t \, dt \circ Q_0^{-1} \circ Q_{t_1} \\
&\quad = Q_0^{-1} \int_0^{t_1} \operatorname{Ad} Q_t U_t \, dt \circ Q_{t_1}.
\end{aligned}$$

Taking into account (6.7), we obtain

$$\int_0^{t_1} \operatorname{Ad} Q_t U_t \, dt = 0.$$

Denote

$$V_t = \int_0^t \operatorname{Ad} Q_\tau U_\tau \, d\tau, \tag{6.8}$$

then admissibility condition of a variation U_t takes the form

$$V_0 = V_{t_1} = 0. \tag{6.9}$$

Now we find extremals of problem (6.6).

$$0 = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} J(\Omega_{\cdot\varepsilon}) = \int_0^{t_1} \langle A\Omega_t, U_t \rangle \, dt$$

by (6.3)

$$= \int_0^{t_1} \langle (\operatorname{Ad} Q_t) A\Omega_t, (\operatorname{Ad} Q_t) U_t \rangle \, dt$$

by (6.8)

$$= \int_0^{t_1} \langle (\operatorname{Ad} Q_t) A\Omega_t, \dot{V}_t \rangle \, dt$$

integrating by parts with the admissibility condition (6.9)

$$= -\int_0^{t_1} \left\langle \frac{d}{dt}(\operatorname{Ad} Q_t) A\Omega_t, V_t \right\rangle dt.$$

So the previous integral vanishes for any admissible operator V_t, thus

$$\frac{d}{dt}(\operatorname{Ad} Q_t) A\Omega_t = 0, \qquad t \in [0, t_1].$$

Hence

$$\operatorname{Ad} Q_t([\Omega_t, A\Omega_t] + A\dot{\Omega}_t) = 0, \qquad t \in [0, t_1],$$

that is why

$$A\dot{\Omega}_t = [A\Omega_t, \Omega_t], \qquad t \in [0, t_1]. \tag{6.10}$$

Introduce the operator

$$M_t = A\Omega_t,$$

called kinetic momentum of the body, and denote

$$B = A^{-1}.$$

We combine equations (6.10), (6.5) and come to *Euler equations* of rotations of a free rigid body:

$$\begin{cases} \dot{M}_t = [M_t, BM_t], & M_t \in \mathrm{so}(3), \\ \dot{Q}_t = Q_t B M_t, & Q_t \in \mathrm{SO}(3). \end{cases}$$

Remark. The presented way to derive Euler equations can be applied to the curves on the group $\mathrm{SO}(n)$ of orthogonal orientation-preserving $n \times n$ matrices with an arbitrary $n > 0$. Then we come to equations of rotations of a generalized n-dimensional rigid body.

Now we rewrite Euler equations via isomorphism (6.2) of so(3) and $\mathbb{R}^3$, which is essentially 3-dimensional and does not generalize to higher dimensions. Recall that for an antisymmetric matrix

$$M = \begin{pmatrix} 0 & -\mu_3 & \mu_2 \\ \mu_3 & 0 & -\mu_1 \\ -\mu_2 & \mu_1 & 0 \end{pmatrix} \in \mathrm{so}(3),$$

the corresponding vector $\mu \in \mathbb{R}^3$ is

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \qquad M \sim \mu.$$

Now Euler equations read as follows:

$$\begin{cases} \dot{\mu}_t = \mu_t \times \beta\mu_t, & \mu_t \in \mathbb{R}^3, \\ \dot{Q}_t = Q_t \widehat{\beta\mu_t}, & Q_t \in \mathrm{SO}(3), \end{cases}$$

where $\beta : \mathbb{R}^3 \to \mathbb{R}^3$ and $\widehat{\beta} : \mathbb{R}^3 \to \mathrm{so}(3)$ are the operators corresponding to $B : \mathrm{so}(3) \to \mathrm{so}(3)$ via the isomorphism $\mathrm{so}(3) \leftrightarrow \mathbb{R}^3$ (6.2).

Eigenvectors of the symmetric positive definite operator $\beta : \mathbb{R}^3 \to \mathbb{R}^3$ are called principal axes of inertia of the rigid body. In the sequel we assume that the rigid body is asymmetric, i.e., the operator β has 3 distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3$. We order the eigenvalues of β:

$$\lambda_1 > \lambda_2 > \lambda_3,$$

and choose an orthonormal frame e_1, e_2, e_3 of the corresponding eigenvectors, i.e., principal axes of inertia. In the basis e_1, e_2, e_3, the operator β is diagonal:

$$\beta \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \lambda_1 \mu_1 \\ \lambda_2 \mu_2 \\ \lambda_3 \mu_3 \end{pmatrix},$$

and the equation $\dot{\mu}_t = \mu_t \times \beta \mu_t$ reads as follows:

$$\begin{cases} \dot{\mu}_1 = (\lambda_3 - \lambda_2)\mu_2\mu_3, \\ \dot{\mu}_2 = (\lambda_1 - \lambda_3)\mu_1\mu_3, \\ \dot{\mu}_3 = (\lambda_2 - \lambda_1)\mu_1\mu_2. \end{cases} \tag{6.11}$$

6.3 Phase portrait

Now we draw the phase portrait of the first of Euler equations:

$$\dot{\mu}_t = \mu_t \times \beta \mu_t, \qquad \mu_t \in \mathbb{R}^3. \tag{6.12}$$

This equation has two integrals: energy

$$(\mu_t, \mu_t) = \mathrm{const}$$

and moment of momentum

$$(\mu_t, \beta \mu_t) = \mathrm{const}\,.$$

Indeed:

$$\begin{aligned} \frac{d}{dt}(\mu_t, \mu_t) &= 2(\mu_t \times \beta \mu_t, \mu_t) = -2(\beta \mu_t, \mu_t \times \mu_t) = 0, \\ \frac{d}{dt}(\mu_t, \beta \mu_t) &= (\mu_t \times \beta \mu_t, \beta \mu_t) + (\mu_t, \beta(\mu_t \times \beta \mu_t)) = 2(\mu_t \times \beta \mu_t, \beta \mu_t) \\ &= -2(\mu_t, \beta \mu_t \times \beta \mu_t) = 0 \end{aligned}$$

by the invariance property (6.4) and symmetry of β.

So all trajectories μ_t of equation (6.12) satisfy the restrictions

$$\begin{cases} \mu_1^2 + \mu_2^2 + \mu_3^2 = \mathrm{const}, \\ \lambda_1 \mu_1^2 + \lambda_2 \mu_2^2 + \lambda_3 \mu_3^2 = \mathrm{const}, \end{cases} \tag{6.13}$$

i.e., belong to intersection of spheres with ellipsoids. Moreover, since ODE (6.12) is homogeneous, we draw its trajectories on one sphere — the unit sphere

$$\mu_1^2 + \mu_2^2 + \mu_3^2 = 1, \tag{6.14}$$

and all other trajectories are obtained by homotheties.

First of all, intersections of the unit sphere with the principal axes of inertia, i.e., the points

$$\pm e_1, \ \pm e_2, \ \pm e_3$$

are equilibria, and there are no other equilibria, see equations (6.11).

Further, the equilibria $\pm e_1$, $\pm e_3$ corresponding to the maximal and minimal eigenvalues λ_1, λ_3 are stable, more precisely, they are centers, and the equilibria $\pm e_2$ corresponding to λ_2 are instable — saddles. This is obvious from the geometry of intersections of the unit sphere with ellipsoids

$$\lambda_1 \mu_1^2 + \lambda_2 \mu_2^2 + \lambda_3 \mu_3^2 = C.$$

Indeed, for $C < \lambda_3$ the ellipsoids are inside the sphere and do not intersect it. For $C = \lambda_3$, the ellipsoid touches the unit sphere from inside at the points $\pm e_3$. Further, for $C > \lambda_3$ and close to λ_3, the ellipsoids intersect the unit sphere by 2 closed curves surrounding e_3 and $-e_3$ respectively. The behavior of intersections is similar in the neighborhood of $C = \lambda_1$. If $C > \lambda_1$, then the ellipsoids are big enough and do not intersect the unit sphere; for $C = \lambda_1$, the small semiaxis of the ellipsoid becomes equal to radius of the sphere, so the ellipsoid touches the sphere from outside at $\pm e_1$; and for $C < \lambda_1$ and close to λ_1 the intersection consists of 2 closed curves surrounding $\pm e_1$. If $C = \lambda_2$, then the ellipsoid touches the sphere at the endpoints of the medium semiaxes $\pm e_2$, and in the neighborhood of each point e_2, $-e_2$, the intersection consists of four separatrix branches tending to this point. Equations for the separatrices are derived from the system

$$\begin{cases} \mu_1^2 + \mu_2^2 + \mu_3^2 = 1, \\ \lambda_1 \mu_1^2 + \lambda_2 \mu_2^2 + \lambda_3 \mu_3^2 = \lambda_2. \end{cases}$$

We multiply the first equation by λ_2 and subtract it from the second equation:

$$(\lambda_1 - \lambda_2)\mu_1^2 - (\lambda_2 - \lambda_3)\mu_3^2 = 0.$$

Thus the separatrices belong to intersection of the unit sphere with two planes

$$\Pi_\pm \stackrel{\text{def}}{=} \{(\mu_1, \mu_2, \mu_3) \in \mathbb{R}^3 \mid \sqrt{\lambda_1 - \lambda_2}\, \mu_1 = \pm \sqrt{\lambda_2 - \lambda_3}\, \mu_3\},$$

thus they are arcs of great circles.

It turns out that separatrices are the only planar curves (i.e., curves belonging to intersection of the unit sphere with affine planes in $\mathbb{R}^3$) in the phase portrait of Euler equation (6.12) on the sphere. Moreover, all other trajectories (excluding equilibria) satisfy the following condition:

$$\mu \notin \Pi_\pm, \mu \notin \mathbb{R} e_i \quad \Rightarrow \quad \mu \wedge \dot{\mu} \wedge \ddot{\mu} \neq 0. \tag{6.15}$$

Indeed, take any trajectory μ_t on the unit sphere. All trajectories homothetic to the chosen one form a cone of the form

$$C(\mu_1^2 + \mu_2^2 + \mu_3^2) = \lambda_1\mu_1^2 + \lambda_2\mu_2^2 + \lambda_3\mu_3^2, \qquad \lambda_3 \leq C \leq \lambda_1. \tag{6.16}$$

But a second-order cone in $\mathbb{R}^3$ is either degenerate or elliptic. The conditions $\mu \notin \Pi_\pm$, $\mu \notin \mathbb{R}e_i$ mean that $C \neq \lambda_i$, $i = 1, 2, 3$, i.e., cone (6.16) is elliptic and not circular, thus strongly convex. Thus inequality (6.15) follows.

In view of ODE (6.12), the convexity condition (6.15) for the cone generated by the curve is rewritten as follows:

$$\mu \notin \Pi_\pm, \mu \notin \mathbb{R}e_i \Rightarrow \mu \wedge (\mu \times \beta\mu) \wedge ((\mu \times \beta\mu) \times \beta\mu + \mu \times \beta(\mu \times \beta\mu)) \neq 0. \tag{6.17}$$

The planar separatrix curves in the phase portrait are regular curves on the sphere, hence

$$\mu \in \Pi_\pm, \mu \notin \mathbb{R}e_2 \Rightarrow \mu \wedge \dot{\mu} \neq 0,$$

or, by ODE (6.12),

$$\mu \in \Pi_\pm, \mu \notin \mathbb{R}e_2 \Rightarrow \mu \wedge (\mu \times \beta\mu) \neq 0. \tag{6.18}$$

6.4 Controlled rigid body: orbits.

Assume that we can control rotations of the rigid body by applying a torque along a line that is fixed in the body. We can change the direction of torque to the opposite one in any moment of time.

Then the control system for the angular velocity is written as

$$\dot{\mu}_t = \mu_t \times \beta\mu_t \pm l, \qquad \mu_t \in \mathbb{R}^3, \tag{6.19}$$

and the whole control system for the controlled rigid body is

$$\begin{cases} \dot{\mu}_t = \mu_t \times \beta\mu_t \pm l, & \mu_t \in \mathbb{R}^3, \\ \dot{Q}_t = Q_t\widehat{\beta\mu_t}, & Q_t \in \mathrm{SO}(3), \end{cases} \tag{6.20}$$

where $l \neq 0$ is a fixed vector along the chosen line.

Now we describe orbits and attainable sets of the 6-dimensional control system (6.20). But before that we study orbits of the 3-dimensional system (6.19).

6.4.1 Orbits of the 3-dimensional system

System (6.19) is analytic, thus dimension of the orbit through a point $\mu \in \mathbb{R}^3$ coincides with dimension of the space

$$\mathrm{Lie}_\mu(\mu \times \beta\mu \pm l) = \mathrm{Lie}_\mu(\mu \times \beta\mu, l).$$

Denote the vector fields:

$$f(\mu) = \mu \times \beta\mu, \qquad g(\mu) \equiv l,$$

and compute several Lie brackets:

$$[g,f](\mu) = \frac{d\,f}{d\,\mu} g(\mu) - \frac{d\,g}{d\,\mu} f(\mu) = l \times \beta\mu + \mu \times \beta l,$$
$$[g,[g,f]](\mu) = l \times \beta l + l \times \beta l = 2l \times \beta l,$$
$$\frac{1}{2}[[g,[g,f]],[g,f]](\mu) = l \times \beta(l \times \beta l) + (l \times \beta l) \times \beta l.$$

We apply (6.17) with $l = \mu$ and obtain that three constant vector fields g, $[g,f]$, $[[g,[g,f]],[g,f]]$ are linearly independent:

$$g(\mu) \wedge \frac{1}{2}[g,f](\mu) \wedge \frac{1}{2}[[g,[g,f]],[g,f]](\mu)$$
$$= l \wedge l \times \beta l \wedge ((l \times \beta l) \times \beta l + l \times \beta(l \times \beta l)) \neq 0$$

if $l \notin \Pi_\pm$, $l \notin \mathbb{R}e_i$.

We obtain the following statement for the generic

Case 1. $l \notin \Pi_\pm$, $l \notin \mathbb{R}e_i$.

Proposition 6.1. *Assume that $l \notin \Pi_\pm$, $l \notin \mathbb{R}e_i$. Then* $\mathrm{Lie}_\mu(f,g) = \mathbb{R}^3$ *for any $\mu \in \mathbb{R}^3$. System* (6.19) *has one* 3*-dimensional orbit,* $\mathbb{R}^3$.

Now consider special dispositions of the vector l.

Case 2. Let $l \in \Pi_+$, $l \notin \mathbb{R}e_2$. Since the plane Π_+ is invariant for the free body (6.12) and $l \in \Pi_+$, then the plane Π_+ is also invariant for the controlled body (6.19), i.e., the orbit through any point of Π_+ is contained in Π_+. On the other hand, implication (6.18) yields

$$l \wedge (l \times \beta l) \neq 0.$$

But the vectors $l = g(\mu)$ and $l \times \beta l = \frac{1}{2}[g,[g,f]](\mu)$ form a basis of the plane Π_+, thus Π_+ is in the orbit through any point $\mu \in \Pi_+$. Consequently, the plane Π_+ is an orbit of (6.19). If an initial point $\mu_0 \notin \Pi_+$, then the trajectory μ_t of (6.19) through μ_0 is not flat, thus

$$(\mu_t \times \beta\mu_t) \wedge l \wedge (l \times \beta l) \not\equiv 0.$$

So the orbit through μ_0 is 3-dimensional. We proved the following statement.

Proposition 6.2. *Assume that $l \in \Pi_+ \setminus \mathbb{R}e_2$. Then system* (6.19) *has one* 2*-dimensional orbit, the plane Π_+, and two* 3*-dimensional orbits, connected components of* $\mathbb{R}^3 \setminus \Pi_+$.

The case $l \in \Pi_- \setminus \mathbb{R}e_2$ is completely analogous, and there holds a similar proposition with Π_+ replaced by Π_-.

Case 3. Now let $l \in \mathbb{R}e_1 \setminus \{0\}$, i.e., $l = ce_1$, $c \neq 0$. First of all, the line $\mathbb{R}e_1$ is an orbit. Indeed, if $\mu \in \mathbb{R}e_1$, then $f(\mu) = 0$, and $g(\mu) = l$ is also tangent to the line $\mathbb{R}e_1$.

To find other orbits, we construct an integral of the control system (6.19) from two integrals (6.13) of the free body. Since $g(\mu) = l = ce_1$, we seek for a linear combination of the integrals in (6.13) that does not depend on μ_1. We multiply the first integral by λ_1, subtract from it the second integral and obtain an integral for the controlled rigid body:

$$(\lambda_1 - \lambda_2)\mu_2^2 + (\lambda_1 - \lambda_3)\mu_3^2 = C. \tag{6.21}$$

Since $\lambda_1 > \lambda_2 > \lambda_3$, this is an elliptic cylinder in $\mathbb{R}^3$.

So each orbit of (6.19) is contained in a cylinder (6.21). On the other hand, the orbit through any point $\mu_0 \in \mathbb{R}^3 \setminus \mathbb{R}e_1$ must be at least 2-dimensional. Indeed, if $\mu_0 \notin \mathbb{R}e_2, \mathbb{R}e_3$, then the free body has trajectories not tangent to the field g; and if $\mu_0 \in \mathbb{R}e_2$ or $\mathbb{R}e_3$, this can be achieved by a small translation of μ_0 along the field g. Thus all orbits outside of the line $\mathbb{R}e_1$ are elliptic cylinders (6.21).

Proposition 6.3. *Let $l \in \mathbb{R}e_1 \setminus \{0\}$. Then all orbits of system* (6.19) *have the form* (6.21)*: there is one* 1*-dimensional orbit — the line $\mathbb{R}e_1$ ($C = 0$), and an infinite number of* 2*-dimensional orbits — elliptic cylinders* (6.21) *with $C > 0$.*

The case $l \in \mathbb{R}e_3 \setminus \{0\}$ is completely analogous to the previous one.

Proposition 6.4. *Let $l \in \mathbb{R}e_3 \setminus \{0\}$. Then system* (6.19) *has one* 1*-dimensional orbit — the line $\mathbb{R}e_3$, and an infinite number of* 2*-dimensional orbits — elliptic cylinders*

$$(\lambda_1 - \lambda_3)\mu_1^2 + (\lambda_2 - \lambda_3)\mu_2^2 = C, \qquad C > 0.$$

Case 4. Finally, consider the most complicated case: let $l \in \mathbb{R}e_2 \setminus \{0\}$. As above, we obtain an integral of control system (6.19):

$$(\lambda_1 - \lambda_2)\mu_1^2 - (\lambda_2 - \lambda_3)\mu_3^2 = C. \tag{6.22}$$

If $C \neq 0$, this equation determines a hyperbolic cylinder. By an argument similar to that used in Case 3, we obtain the following description of orbits.

Proposition 6.5. *Let $l \in \mathbb{R}e_2 \setminus \{0\}$. Then there is one* 1*-dimensional orbit — the line $\mathbb{R}e_2$, and an infinite number of* 2*-dimensional orbits of the following form:*

(1) *connected components of hyperbolic cylinders* (6.22) *for $C \neq 0$;*

(2) *half-planes — connected components of the set $(\Pi_+ \cup \Pi_-) \setminus \mathbb{R}e_2$.*

So we considered all possible dispositions of the vector $l \in \mathbb{R}^3 \setminus \{0\}$, and in all cases described orbits of the 3-dimensional system (6.19). Now we study orbits of the full 6-dimensional system (6.20).

6.4.2 Orbits of the 6-dimensional system

The vector fields in the right-hand side of the 6-dimensional system (6.20) are

$$f(Q,\mu) = \begin{pmatrix} Q\widehat{\beta\mu} \\ \mu \times \beta\mu \end{pmatrix}, \qquad g(Q,\mu) = \begin{pmatrix} 0 \\ l \end{pmatrix}, \qquad (Q,\mu) \in \mathrm{SO}(3) \times \mathbb{R}^3.$$

Notice the commutation rule for vector fields of the form that appear in our problem:

$$f_i(Q,\mu) = \begin{pmatrix} Q\widehat{\beta w_i(\mu)} \\ v_i(\mu) \end{pmatrix} \in \mathrm{Vec}(\mathrm{SO}(3) \times \mathbb{R}^3),$$

$$[f_1,f_2](Q,\mu) = \begin{pmatrix} Q[\widehat{\beta w_1}, \widehat{\beta w_2}]_{\mathrm{so}(3)} + Q\widehat{\beta\left(\dfrac{\partial w_2}{\partial \mu}v_1 - \dfrac{\partial w_1}{\partial \mu}v_2\right)} \\ \\ \dfrac{\partial v_2}{\partial \mu}v_1 - \dfrac{\partial v_1}{\partial \mu}v_2 \end{pmatrix}.$$

We compute first the same Lie brackets as in the 3-dimensional case:

$$[g,f] = \begin{pmatrix} Q\widehat{\beta l} \\ l \times \beta\mu + \mu \times \beta l \end{pmatrix},$$

$$\frac{1}{2}[g,[g,f]] = \begin{pmatrix} 0 \\ l \times \beta l \end{pmatrix},$$

$$\frac{1}{2}[[g,[g,f]],[g,f]] = \begin{pmatrix} 0 \\ l \times \beta(l \times \beta l) + (l \times \beta l) \times \beta l \end{pmatrix}.$$

Further, for any vector field $X \in \mathrm{Vec}(\mathrm{SO}(3) \times \mathbb{R}^3)$ of the form

$$X = \begin{pmatrix} 0 \\ x \end{pmatrix}, \qquad x \text{ — a constant vector field on } \mathbb{R}^3, \tag{6.23}$$

we have

$$[X,f] = \begin{pmatrix} Q\widehat{\beta x} \\ * \end{pmatrix}. \tag{6.24}$$

To study the orbit of the 6-dimensional system (6.20) through a point $(Q,\mu) \in \mathrm{SO}(3) \times \mathbb{R}^3$, we follow the different cases for the 3-dimensional system (6.19) in Subsec. 6.4.1.

Case 1. $l \notin \Pi_\pm$, $l \notin \mathbb{R}e_i$. We can choose 3 linearly independent vector fields in $\mathrm{Lie}(f,g)$ of the form (6.23):

$$X_1 = g, \qquad X_2 = \frac{1}{2}[g,[g,f]], \qquad X_3 = \frac{1}{2}[[g,[g,f],[g,f]].$$

By the commutation rule (6.24), we have 6 linearly independent vectors in $\mathrm{Lie}_{(Q,\mu)}(f,g)$:

$$X_1 \wedge X_2 \wedge X_3 \wedge [X_1,f] \wedge [X_2,f] \wedge [X_3,f] \neq 0.$$

Thus the orbit through (Q, μ) is 6-dimensional.

Case 2. $l \in \Pi_\pm \setminus \mathbb{R}e_2$.

Case 2.1. $\mu \notin \Pi_\pm$. First of all, $\mathrm{Lie}(f, g)$ contains 2 linearly independent vector fields of the form (6.23):

$$X_1 = g, \qquad X_2 = \frac{1}{2}[g, [g, f]].$$

Since the trajectory of the free body in $\mathbb{R}^3$ through μ is not flat, we can assume that the vector $v = \mu \times \beta\mu$ is linearly independent of l and $l \times \beta l$. Now our aim is to show that $\mathrm{Lie}(f, g)$ contains 2 vector fields of the form

$$Y_1 = \begin{pmatrix} QM_1 \\ v_1 \end{pmatrix}, \qquad Y_2 = \begin{pmatrix} QM_2 \\ v_2 \end{pmatrix}, \qquad M_1 \wedge M_2 \neq 0, \tag{6.25}$$

where the vector fields v_1 and v_2 vanish at the point μ. If this is the case, then $\mathrm{Lie}_{(Q,\mu)}(f, g)$ contains 6 linearly independent vectors:

$$\begin{aligned}
&X_1(Q, \mu), \quad X_2(Q, \mu), \quad f(Q, \mu), \\
&Y_1(Q, \mu) = \begin{pmatrix} QM_1 \\ 0 \end{pmatrix}, \quad Y_2(Q, \mu) = \begin{pmatrix} QM_2 \\ 0 \end{pmatrix}, \\
&[Y_1, Y_2](Q, \mu) = \begin{pmatrix} Q[M_1, M_2] \\ 0 \end{pmatrix},
\end{aligned}$$

and the orbit through the point (Q, μ) is 6-dimensional.

Now we construct 2 vector fields of the form (6.25) in $\mathrm{Lie}(f, g)$. Taking appropriate linear combinations with the fields X_1, X_2, we project the second component of the fields $[g, f]$ and $\frac{1}{2}[f, [g, [g, f]]$ to the line $\mathbb{R}v$, thus we obtain the vector fields

$$\begin{pmatrix} Q\widehat{\beta l} \\ k_1 v \end{pmatrix}, \quad \begin{pmatrix} Q\widehat{\beta(l \times \beta l)} \\ k_2 v \end{pmatrix} \in \mathrm{Lie}(f, g). \tag{6.26}$$

If both k_1 and k_2 vanish at μ, these vector fields can be taken as Y_1, Y_2 in (6.25). And if k_1 or k_2 does not vanish at μ, we construct such vector fields Y_1, Y_2 taking appropriate linear combinations of fields (6.26) and f with the fields g, $[g, [g, f]]$.

So in Case 2.1 the orbit is 6-dimensional.

Case 2.2. $\mu \in \Pi_\pm$. There are 5 linearly independent vectors in $\mathrm{Lie}_{(Q,\mu)}(f, g)$:

$$X_1 = g, \quad X_2 = \frac{1}{2}[g, [g, f]], \quad [X_1, f], \quad [X_2, f], \quad [[X_1, f], [X_2, f]].$$

Since the orbit in $\mathbb{R}^3$ is 2-dimensional, the orbit in $\mathrm{SO}(3) \times \mathbb{R}^3$ is 5-dimensional.

Case 3. $l \in \mathbb{R}e_1 \setminus \{0\}$.

Case 3.1. $\mu \notin \mathbb{R}e_1$. The argument is similar to that of Case 2.1. We can assume that the vectors l and $v = \mu \times \beta\mu$ are linearly independent. The orbit in $\mathbb{R}^3$ is 2-dimensional and the vectors l, v span the tangent space to this orbit, thus we can find vector fields in $\mathrm{Lie}(f, g)$ of the form:

$$Y_1 = [g, f] - C_1 g - C_2 f = \begin{pmatrix} Q\widehat{\beta} l + C_3 Q\widehat{\beta}\mu \\ 0 \end{pmatrix},$$
$$Y_2 = [Y_1, f] = \begin{pmatrix} Q[\widehat{\beta} l, \widehat{\beta}\mu] + C_4 Q\widehat{\beta}\mu \\ 0 \end{pmatrix}$$

for some real functions C_i, $i = 1, \ldots, 4$. Then we have 5 linearly independent vectors in $\mathrm{Lie}_{(Q,\mu)}(f, g)$:

$$g, \quad f, \quad Y_1, \quad Y_2, \quad [Y_1, Y_2].$$

So the orbit of the 6-dimensional system (6.20) is 5-dimensional (it cannot have dimension 6 since the 3-dimensional system (6.19) has a 2-dimensional orbit).

Case 3.2. $\mu \in \mathbb{R}e_1$. The vectors

$$f(Q, \mu) = \begin{pmatrix} Q\widehat{\beta}\mu \\ 0 \end{pmatrix}, \qquad [g, f](Q, \mu) = \begin{pmatrix} Q\widehat{\beta} l \\ 0 \end{pmatrix},$$

are linearly dependent, thus $\dim \mathrm{Lie}_{(Q,\mu)}(f, g) = \dim \mathrm{span}(f, g)|_{(Q,\mu)} = 2$. So the orbit is 2-dimensional.

The cases $l \in \mathbb{R}e_i \setminus \{0\}$, $i = 1, 2$, are similar to Case 3.

We completed the study of orbits of the controlled rigid body (6.20) and now summarize it.

Proposition 6.6. *Let (Q, μ) be a point in $\mathrm{SO}(3) \times \mathbb{R}^3$. If the orbit $\mathcal{O}$ of the 3-dimensional system (6.19) through the point μ is 3- or 2-dimensional, then the orbit of the 6-dimensional system (6.20) through the point (Q, μ) is $\mathrm{SO}(3) \times \mathcal{O}$, i.e., respectively 6- or 5-dimensional. If $\dim \mathcal{O} = 1$, then the 6-dimensional system has a 2-dimensional orbit.*

We will describe attainable sets of this system in Section 7.3 after acquiring some general facts on attainable sets.

Chapter 7

Attainable sets

Let M be a smooth manifold and $\mathcal{F} \subset \operatorname{Vec} M$ a bracket-generating family of vector fields on M:

$$\operatorname{Lie}_q \mathcal{F} = T_q M \qquad \forall\, q \in M. \tag{7.1}$$

If a family $\mathcal{F} \subset \operatorname{Vec} M$ is not bracket-generating, and M and $\mathcal{F}$ are real analytic, we can pass from $\mathcal{F}$ to a bracket-generating family $\mathcal{F}|_{\mathcal{O}}$, where $\mathcal{O}$ is an orbit of $\mathcal{F}$ (see the analytic version of the Orbit Theorem, Corollary 5.3). Thus in the analytic case requirement (7.1) is not restrictive in essence.

7.1 Krener's theorem

The following proposition describes important properties of attainable sets $\mathcal{A}_{q_0}$, $q_0 \in M$, for arbitrary nonnegative time.

Theorem 7.1 (Krener). *Let $\mathcal{F} \subset \operatorname{Vec} M$ be a bracket-generating system. Then $\mathcal{A}_{q_0} \subset \overline{\operatorname{int} \mathcal{A}_{q_0}}$ for any $q_0 \in M$.*

Remark. In particular, attainable sets for arbitrary time have nonempty interior:

$$\operatorname{int} \mathcal{A}_{q_0} \neq \emptyset.$$

Attainable sets may be:

- open sets,
- manifolds with smooth boundary,
- manifolds with boundary having singularities (corner or cuspidal points).

One can easily construct control systems (e.g. in the plane) that realize these possibilities.

On the other hand, Krener's theorem prohibits an attainable set $\mathcal{A}_{q_0}$ of a bracket-generating family to be:

- a lower-dimensional subset of M,
- a set where boundary points are isolated from interior points.

Now we prove Krener's theorem.

Proof. Fix an arbitrary point $q_0 \in M$ and take a point $q' \in \mathcal{A}_{q_0}$. We show that

$$q' \in \overline{\operatorname{int} \mathcal{A}_{q_0}}. \tag{7.2}$$

(1) There exists a vector field $f_1 \in \mathcal{F}$ such that $f_1(q') \neq 0$, otherwise $\operatorname{Lie}_{q'}(\mathcal{F}) = 0$ and $\dim M = 0$. The curve

$$s_1 \mapsto q' \circ e^{s_1 f_1}, \qquad s_1 \in (0, \varepsilon) \tag{7.3}$$

is a 1-dimensional submanifold of M for small enough $\varepsilon > 0$.

If $\dim M = 1$, then $q' \circ e^{s_1 f_1} \in \operatorname{int} \mathcal{A}_{q_0}$ for sufficiently small $s_1 > 0$, and inclusion (7.2) follows.

(2) Assume that $\dim M > 1$. Then arbitrarily close to q' we can find a point q_1 on curve (7.3) and a field $f_2 \in \mathcal{F}$ such that $f_2(q_1)$ is not tangent to manifold (7.3):

$$q_1 = q' \circ e^{t_1^1 f_1}, \qquad t_1^1 \text{ sufficiently small},$$
$$(q_1 \circ f_1) \wedge (q_1 \circ f_2) \neq 0,$$

otherwise $\dim \operatorname{Lie}_q \mathcal{F} = 1$ for q on curve (7.3) with small s_1. Then the mapping

$$(s_1, s_2) \mapsto q' \circ e^{s_1 f_1} \circ e^{s_2 f_2}, \tag{7.4}$$
$$(s_1, s_2) \in O(t_1^1, 0), \qquad s_1 > 0,\ s_2 > 0,$$

is an immersion in a small neighborhood $O(t_1^1, 0) \subset \mathbb{R}^2_{s_1, s_2}$, thus its image is a 2-dimensional submanifold of M.

If $\dim M = 2$, inclusion (7.2) is proved.

(3) Assume that $\dim M > 2$. We can find a vector $f_3(q)$, $f_3 \in \mathcal{F}$, not tangent to surface (7.4) sufficiently close to q': there exist $t_2^1, t_2^2 > 0$ and $f_3 \in \mathcal{F}$ such that the vector field f_3 is not tangent to surface (7.4) at a point $q_2 = q' \circ e^{t_2^1 f_1} \circ e^{t_2^2 f_2}$. Otherwise the family $\mathcal{F}$ is not bracket-generating.

The mapping

$$(s_1, s_2, s_3) \mapsto q' \circ e^{s_1 f_1} \circ e^{s_2 f_2} \circ e^{s_3 f_3},$$
$$(s_1, s_2, s_3) \in O(t_2^1, t_2^2, 0), \qquad s_i > 0,\ i = 1, 2, 3,$$

is an immersion in a small neighborhood $O(t_1^1, t_2^2, 0) \subset \mathbb{R}^3_{s_1, s_2, s_3}$, thus its image is a smooth 3-dimensional submanifold of M.

If $\dim M = 3$, inclusion (7.2) follows. Otherwise we continue this procedure.

(4) For $\dim M = n$, inductively, we find $(t^1_{n-1}, t^2_{n-1}, \dots, t^{n-1}_{n-1}) \in \mathbb{R}^{n-1}$, $t^i_{n-1} > 0$, such that the mapping

$$(s_1, \dots, s_n) \mapsto q' \circ e^{s_1 f_1} \circ \dots \circ e^{s_n f_n},$$
$$(s_1, \dots, s_n) \in O(t^1_{n-1}, t^2_{n-1}, \dots, t^{n-1}_{n-1}, 0),$$

is an immersion. The image of this immersion is an n-dimensional submanifold of M, thus an open set. This open set is contained in $\mathcal{A}_{q_0}$ and can be chosen as close to the point q' as we wish. Inclusion (7.2) is proved, and the theorem follows. □

We obtain the following proposition from Krener's theorem.

Corollary 7.1. *Let $\mathcal{F} \subset \operatorname{Vec} M$ be a bracket-generating system. If $\overline{\mathcal{A}_{q_0}(\mathcal{F})} = M$ for some $q_0 \in M$, then $\mathcal{A}_{q_0}(\mathcal{F}) = M$.*

Proof. Take an arbitrary point $q \in M$. We show that $q \in \mathcal{A}_{q_0}(\mathcal{F})$.
Consider the system

$$-\mathcal{F} = \{-V \mid V \in \mathcal{F}\} \subset \operatorname{Vec} M.$$

This system is bracket-generating, thus by Theorem 7.1

$$\mathcal{A}_q(-\mathcal{F}) \subset \overline{\operatorname{int} \mathcal{A}_q(-\mathcal{F})} \qquad \forall q \in M.$$

Take any point $\widehat{q} \in \operatorname{int} \mathcal{A}_q(-\mathcal{F})$ and a neighborhood of this point $O_{\widehat{q}} \subset \mathcal{A}_q(-\mathcal{F})$. Since $\mathcal{A}_{q_0}(\mathcal{F})$ is dense in M, then

$$\mathcal{A}_{q_0}(\mathcal{F}) \cap O_{\widehat{q}} \neq \emptyset.$$

That is why $\mathcal{A}_{q_0}(\mathcal{F}) \cap \mathcal{A}_q(-\mathcal{F}) \neq \emptyset$, i.e., there exists a point

$$q' \in \mathcal{A}_{q_0}(\mathcal{F}) \cap \mathcal{A}_q(-\mathcal{F})$$

Thus the point q' can be represented as follows:

$$q' = q_0 \circ e^{t_1 f_1} \circ \dots \circ e^{t_k f_k}, \qquad f_i \in \mathcal{F},\ t_i > 0,$$
$$q' = q \circ e^{-s_1 g_1} \circ \dots \circ e^{-s_l g_l}, \qquad g_i \in \mathcal{F},\ s_i > 0.$$

We multiply both decompositions from the right by $e^{s_l g_l} \circ \dots \circ e^{s_1 g_1}$ and obtain $q = q_0 \circ e^{t_1 f_1} \circ \dots \circ e^{t_k f_k} \circ e^{s_l g_l} \circ \dots \circ e^{s_1 g_1}$, i.e., $q \in \mathcal{A}_{q_0}(\mathcal{F})$. □

The sense of the previous proposition is that in the study of controllability, we can replace the attainable set of a bracket-generating system by its closure. Further, now we try to add new vector fields to a system so that the closure of its attainable set do not change.

Definition 7.1. A vector field $f \in \operatorname{Vec} M$ is called *compatible* with a system $\mathcal{F} \subset \operatorname{Vec} M$ if

$$\mathcal{A}_q(\mathcal{F} \cup f) \subset \overline{\mathcal{A}_q(\mathcal{F})} \qquad \forall q \in M.$$

Easy compatibility condition is given by the following statement.

Proposition 7.1. *Let $\mathcal{F} \subset \operatorname{Vec} M$. For any vector fields $f_1, f_2 \in \mathcal{F}$, and any functions $a_1, a_2 \in C^\infty(M)$, $a_1, a_2 \geq 0$, the vector field $a_1 f_1 + a_2 f_2$ is compatible with $\mathcal{F}$.*

In view of Corollary 5.2, the following proposition holds.

Corollary 7.2. *If $\mathcal{F} \subset \operatorname{Vec} M$ is a bracket-generating system such that the positive convex cone generated by $\mathcal{F}$*

$$\operatorname{cone}(\mathcal{F}) = \left\{ \sum_{i=1}^{k} a_i f_i \mid f_i \in \mathcal{F},\ a_i \in C^\infty(M),\ a_i \geq 0,\ k \in \mathbb{N} \right\} \subset \operatorname{Vec} M$$

is symmetric, then $\mathcal{F}$ is controllable.

Proposition 7.1 is a corollary of the following general and strong statement.

Theorem 7.2. *Let X_τ, Y_τ, $\tau \in [0, t_1]$, be nonautonomous vector fields, which are bounded w.r.t. τ and have a compact support. Let $0 \leq \alpha(\tau) \leq 1$ be a measurable function. Then there exists a sequence of nonautonomous vector fields $Z_\tau^n \in \{X_\tau, Y_\tau\}$, i.e., $Z_\tau^n = X_\tau$ or Y_τ for any τ and n, such that the flow*

$$\overrightarrow{\exp} \int_0^t Z_\tau^n \, d\tau \to \overrightarrow{\exp} \int_0^t (\alpha(\tau) X_\tau + (1 - \alpha(\tau)) Y_\tau) \, d\tau, \qquad n \to \infty,$$

uniformly w.r.t. $(t, q) \in [0, t_1] \times M$ and uniformly with all derivatives w.r.t. $q \in M$.

Now Proposition 7.1 follows: in the case $a_1(x) + a_2(x) = 1$ it is a corollary of Theorem 7.2, for the case $a_1(x), a_2(x) > 0$ we generalize by multiplication of control parameters by arbitrary positive function (this does not change attainable set for all nonnegative times), and the case $a_1(x), a_2(x) \geq 0$ is obtained by passage to limit.

Theorem 7.2 follows from the next two lemmas.

Lemma 7.1. *Under conditions of Theorem 7.2, there exists a sequence of nonautonomous vector fields $Z_\tau^n \in \{X_\tau, Y_\tau\}$ such that*

$$\int_0^t Z_\tau^n \, d\tau \to \int_0^t (\alpha(\tau) X_\tau + (1 - \alpha(\tau)) Y_\tau) \, d\tau$$

uniformly w.r.t. $(t, q) \in [0, t_1] \times M$ and uniformly with all derivatives w.r.t. $q \in M$.

Proof. Fix an arbitrary positive integer $n \in \mathbb{N}$. We can choose a covering of the segment $[0, t_1]$ by subsets

$$[0, t_1] = \bigcup_{i=1}^{N} E_i$$

such that

$$\forall i = 1, \ldots, N \ \exists X_i, Y_i \in \operatorname{Vec} M \quad \text{s.t.} \quad \|X_\tau - X_i\|_{C^n} \leq \frac{1}{n}, \quad \|Y_\tau - Y_i\|_{C^n} \leq \frac{1}{n}.$$

Indeed, the fields X_τ, Y_τ are bounded in C^{n+1}-topology, thus they form a precompact set in C^n-topology.

Then divide E_i into n subsets of equal measure:

$$E_i = \cup_{j=1}^n E_{ij}, \quad |E_{ij}| = \frac{1}{n}|E_i|, \qquad i, j = 1, \ldots, n.$$

In each E_{ij} pick a subset F_{ij} so that

$$F_{ij} \subset E_{ij}, \qquad |F_{ij}| = \int_{E_{ij}} \alpha(\tau)\, d\tau.$$

Finally, define the following vector field:

$$Z_\tau^n = \begin{cases} X_\tau, & \tau \in F_{ij}, \\ Y_\tau, & \tau \in E_{ij} \setminus F_{ij}. \end{cases}$$

Then the sequence of vector fields Z_τ^n is the required one. □

Now we prove the second part of Theorem 7.2.

Lemma 7.2. *Let Z_τ^n, $n = 1, 2, \ldots$, and Z_τ, $\tau \in [0, t_1]$, be nonautonomous vector fields on M, bounded w.r.t. τ, and let these vector fields have a compact support. If*

$$\int_0^t Z_\tau^n \, d\tau \to \int_0^t Z_\tau \, d\tau, \qquad n \to \infty,$$

then

$$\overrightarrow{\exp} \int_0^t Z_\tau^n \, d\tau \to \overrightarrow{\exp} \int_0^t Z_\tau \, d\tau, \qquad n \to \infty,$$

the both convergences being uniform w.r.t. $(t, q) \in [0, t_1] \times M$ and uniform with all derivatives w.r.t. $q \in M$.

Proof. (1) First we prove the statement for the case $Z_\tau = 0$. Denote the flow

$$P_t^n = \overrightarrow{\exp} \int_0^t Z_\tau^n \, d\tau.$$

Then

$$P_t^n = \operatorname{Id} + \int_0^t P_\tau^n \circ Z_\tau^n \, d\tau$$

integrating by parts

$$= \operatorname{Id} + P_t^n \circ \int_0^t Z_\tau^n \, d\tau - \int_0^t \left(P_\tau^n \circ Z_\tau^n \circ \int_0^\tau Z_\theta^\tau \, d\theta \right) d\tau.$$

Since $\int_0^t Z_\tau^n \, d\tau \to 0$, the last two terms above tend to zero, thus

$$P_t^n \to \mathrm{Id},$$

and the statement of the lemma in the case $Z_\tau = 0$ is proved.

(2) Now we consider the general case. Decompose vector fields in the sequence as follows:

$$Z_\tau^n = Z_\tau + V_\tau^n, \qquad \int_0^t V_\tau^n \, d\tau \to 0, \qquad n \to \infty.$$

Denote $P_t^n = \overrightarrow{\exp} \int_0^t V_\tau^n \, d\tau$. From the variations formula, we have

$$\overrightarrow{\exp} \int_0^t Z_\tau^n \, d\tau = \overrightarrow{\exp} \int_0^t (V_\tau^n + Z_\tau) \, d\tau = \overrightarrow{\exp} \int_0^t \mathrm{Ad}\, P_\tau^n \, Z_\tau \, d\tau \circ P_t^n.$$

Since $P_t^n \to \mathrm{Id}$ by part (1) of this proof and thus $\mathrm{Ad}\ P_t^n \to \mathrm{Id}$, we obtain the required convergence:

$$\overrightarrow{\exp} \int_0^t Z_\tau^n \, d\tau \to \overrightarrow{\exp} \int_0^t Z_\tau \, d\tau.$$

□

So we proved Theorem 7.2 and thus Proposition 7.1.

7.2 Poisson stability and compatibility of vector fields

Now we return to the study of controllability.

Definition 7.2. Let $f \in \mathrm{Vec}\, M$ be a complete vector field. A point $q \in M$ is called *Poisson stable* for f if for any $t > 0$ and any neighborhood O_q of q there exists a point $q' \in O_q$ and a time $t' > t$ such that $q' \circ e^{t'f} \in O_q$.

In other words, all trajectories cannot leave a neighborhood of a Poisson stable point forever, some of them must return to this neighborhood for arbitrarily large times.

Remark. If a trajectory $q \circ e^{tf}$ is periodic, then q is Poisson stable for f.

Definition 7.3. A complete vector field $f \in \mathrm{Vec}\, M$ is *Poisson stable* if all points of M are Poisson stable for f.

The condition of Poisson stability seems to be rather restrictive, but nevertheless there are surprisingly many Poisson stable vector fields in applications, see Poincaré's theorem below.

But first we prove a consequence of Poisson stability for controllability.

Proposition 7.2. *Let $\mathcal{F} \subset \operatorname{Vec} M$ be a bracket-generating system. If a vector field $f \in \mathcal{F}$ is Poisson stable, then the field $-f$ is compatible with $\mathcal{F}$.*

Proof. Choose an arbitrary point $q_0 \in M$ and a moment of time $t > 0$. To prove the statement, we should approximate the point $q_0 \circ e^{-tf}$ by reachable points.

Since $\mathcal{F}$ is bracket-generating, we can choose an open set $W \subset \operatorname{int} \mathcal{A}_{q_0}(\mathcal{F})$ arbitrarily close to q_0. Then the set $W \circ e^{-tf}$ is close enough to $q_0 \circ e^{-tf}$.

By Poisson stability, there exists $t' > t$ such that

$$\emptyset \neq \left(W \circ e^{-tf}\right) \circ e^{t'f} \cap W \circ e^{-tf} = W \circ e^{(t'-t)f} \cap W \circ e^{-tf}.$$

But $W \circ e^{(t'-t)f} \subset \mathcal{A}_{q_0}(\mathcal{F})$, thus

$$A_{q_0}(\mathcal{F}) \cap W \circ e^{-tf} \neq \emptyset.$$

So in any neighborhood of $q_0 \circ e^{-tf}$ there are points of the attainable set $\mathcal{A}_{q_0}(\mathcal{F})$, i.e., $q_0 \circ e^{-tf} \in \overline{\mathcal{A}_{q_0}(\mathcal{F})}$. □

Theorem 7.3 (Poincaré). *Let M be a smooth manifold with a volume form* Vol. *Let a vector field $f \in \operatorname{Vec} M$ be complete and its flow e^{tf} preserve volume. Let $W \subset M$, $W \subset \overline{\operatorname{int} W}$, be a subset of finite volume, invariant for f:*

$$\operatorname{Vol}(W) < \infty, \qquad W \circ e^{tf} \subset W \quad \forall t > 0.$$

Then all points of W are Poisson stable for f.

Proof. Take any point $q \in W$ and any its neighborhood $O \subset M$ of finite volume. The set $V = W \cap O$ contains an open nonempty subset $\operatorname{int} W \cap O$, thus $\operatorname{Vol}(V) > 0$. In order to prove the theorem, we show that

$$V \circ e^{t'f} \cap V \neq \emptyset \quad \text{for some large } t'.$$

Fix any $t > 0$. Then all sets

$$V \circ e^{ntf}, \qquad n = 0, 1, 2, \dots,$$

have the same positive volume, thus they cannot be disjoint. Indeed, if

$$V \circ e^{ntf} \cap V \circ e^{mtf} = \emptyset \qquad \forall n,\ m = 0, 1, 2, \dots,$$

then $\operatorname{Vol}(W) = \infty$ since all these sets are contained in W. Consequently, there exist nonnegative integers $n > m$ such that

$$V \circ e^{ntf} \cap V \circ e^{mtf} \neq \emptyset.$$

We multiply this inequality by e^{-mtf} from the right and obtain

$$V \circ e^{(n-m)tf} \cap V \neq \emptyset.$$

Thus the point q is Poisson stable for f. Since $q \in W$ is arbitrary, the theorem follows. □

A vector field that preserves volume is called *conservative*.

Remark. Recall that on $M = \mathbb{R}^n = \{(x_1, \dots, x_n)\}$, a flow e^{tf} is conservative, i.e., preserves the standard volume $\mathrm{Vol}(V) = \int_V dx_1 \dots dx_n$ iff the field $f = \sum_{i=1}^n f_i \frac{\partial}{\partial x_i}$ is divergence-free:

$$\mathrm{div}_x f = \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} = 0.$$

7.3 Controlled rigid body: attainable sets

We apply preceding general results on controllability to the control system that governs rotations of the rigid body, see (6.20):

$$\begin{pmatrix} \dot Q \\ \dot\mu \end{pmatrix} = f(Q,\mu) \pm g(Q,\mu), \qquad f = \begin{pmatrix} Q\widehat{\beta}\mu \\ \mu \times \beta\mu \end{pmatrix}, \qquad g = \begin{pmatrix} 0 \\ l \end{pmatrix}, \qquad (7.5)$$
$$(Q,\mu) \in \mathrm{SO}(3) \times \mathbb{R}^3.$$

The vector field $f = \frac{1}{2}(f+g) + \frac{1}{2}(f-g)$ is compatible with system (7.5).

We show now that the field f is Poisson stable on $\mathrm{SO}(3) \times \mathbb{R}^3$.

Consider first the vector field $f(Q,\mu)$ on the larger space $\mathbb{R}^9_Q \times \mathbb{R}^3_\mu$, where $\mathbb{R}^9_Q$ is the space of all 3×3 matrices. Since $\mathrm{div}_{(Q,\mu)} f = 0$, the field f is conservative on $\mathbb{R}^9_Q \times \mathbb{R}^3_\mu$.

Further, since the first component of the field f is linear in Q, it has the following right-invariant property in Q:

$$e^{tf}\begin{pmatrix} Q \\ \mu \end{pmatrix} = \begin{pmatrix} Q_t \\ \mu_t \end{pmatrix} \quad \Rightarrow \quad e^{tf}\begin{pmatrix} PQ \\ \mu \end{pmatrix} = \begin{pmatrix} PQ_t \\ \mu_t \end{pmatrix}, \qquad (7.6)$$
$$Q,\ Q_t,\ P \in \mathbb{R}^9_Q, \quad \mu,\ \mu_t \in \mathbb{R}^3_\mu.$$

In view of this property, the field f has compact invariant sets in $\mathbb{R}^9_Q \times \mathbb{R}^3_\mu$ of the form

$$W = K \cdot \mathrm{SO}(3) \times \{ (\mu,\mu) \le C \}, \quad K \Subset \mathbb{R}^9_Q,\ K \subset \overline{\mathrm{int}\, K},\ C > 0,$$

so that

$$W \subset \overline{\mathrm{int}\, W}.$$

By Poincaré's theorem, the field f is Poisson stable on all such sets W, thus on $\mathbb{R}^9_Q \times \mathbb{R}^3_\mu$. In view of the invariance property (7.6), the field f is Poisson stable on $\mathrm{SO}(3) \times \mathbb{R}^3$.

Since f is compatible with (7.5), then $-f$ is also compatible. The vector fields $\pm g = (f \pm g) - f$ are compatible with (7.5) as well. The vector fields $f \pm g$ are contained in the convex hull of the compatible vector fields found: $\mathrm{span}(f,g)$, a symmetric system. Thus if system (7.5) is bracket-generating, then its attainable set coincides with the orbit, i.e., the whole state space $\mathrm{SO}(3) \times \mathbb{R}^3$. So in the bracket-generating case system (7.5) is completely controllable.

In the non-bracket-generating cases, the structure of attainable sets is more complicated. If l is a principal axis of inertia, then the orbits of system (7.5) coincide with attainable sets. If $l \in \Pi_\pm \setminus \mathbb{R}e_2$, they do not coincide. This is easy to see from the phase portrait of the vector field $f(\mu) = \mu \times \beta\mu$ in the plane $\Pi_\pm$: the line $\mathbb{R}e_2$ consists of equilibria of f, and in the half-planes $\Pi_\pm \setminus \mathbb{R}e_2$ trajectories of f are semicircles centered at the origin.

Chapter 8

Feedback and state equivalence of control systems

8.1 Feedback equivalence

Consider control systems of the form

$$\dot{q} = f(q, u), \qquad q \in M, \ u \in U. \tag{8.1}$$

We suppose that not only M, but also U is a smooth manifold. For the right-hand side, we suppose that for all fixed $u \in U$, $f(q, u)$ is a smooth vector field on M, and, moreover, the mapping

$$(u, q) \mapsto f(q, u)$$

is smooth. Admissible controls are measurable locally bounded mappings

$$t \mapsto u(t) \in U$$

(for simplicity, one can consider piecewise continuous controls). If such a control $u(t)$ is substituted to control system (8.1), one obtains a nonautonomous ODE

$$\dot{q} = f(q, u(t)), \tag{8.2}$$

with the right-hand side smooth in q and measurable, locally bounded in t. For such ODEs, there holds a standard theorem on existence and uniqueness of solutions, at least local. Solutions $q(\cdot)$ to ODEs (8.2) are Lipschitzian curves in M.

In Section 5.7 we already considered state transformations of control systems, i.e., diffeomorphisms of M. State transformations map trajectories of control systems to trajectories, with the same control. Now we introduce a new class of feedback transformations, which also map trajectories to trajectories, but possibly with a new control.

Denote the space of new control parameters by $\widehat{U}$. We assume that it is a smooth manifold.

Definition 8.1. Let $\varphi : M \times \widehat{U} \to U$ be a smooth mapping. A transformation of the form

$$f(q,u) \mapsto f(q, \varphi(q, \widehat{u})), \qquad q \in M, \quad u \in U, \quad \widehat{u} \in \widehat{U},$$

is called a *feedback transformation*.

Remark. A feedback transformation reparametrizes control u in a way depending on q.

It is easy to see that any admissible trajectory $q(\cdot)$ of the system $\dot{q} = f(q, \varphi(q, \widehat{u}))$ corresponding to a control $\widehat{u}(\cdot)$ is also admissible for the system $\dot{q} = f(q,u)$ with the control $u(\cdot) = \varphi(q(\cdot), \widehat{u}(\cdot))$, but, in general, not vice versa.

In order to consider feedback equivalence, we consider invertible feedback transformations with

$$\widehat{U} = U, \qquad \varphi|_{q \times U} \in \operatorname{Diff} U.$$

Such mappings $\varphi : M \times U \to U$ generate feedback transformations

$$f(q,u) \mapsto f(q, \varphi(q,u)).$$

The corresponding control systems

$$\dot{q} = f(q,u) \quad \text{and} \quad \dot{q} = f(q, \varphi(q,u))$$

are called *feedback equivalent*.

Our aim is to simplify control systems with state and feedback transformations.

Remark. In mathematical physics, feedback transformations are often called gauge transformations.

Consider affine in control systems

$$\dot{q} = f(q) + \sum_{i=1}^{k} u_i g_i(q), \qquad u = (u_1, \dots, u_k) \in \mathbb{R}^k, \ q \in M. \tag{8.3}$$

To such systems, it is natural to apply feedback transformations affine in control:

$$\varphi = (\varphi_1, \dots, \varphi_k) : M \times \mathbb{R}^k \to \mathbb{R}^k,$$
$$\varphi_i(q,u) = c_i(q) + \sum_{j=1}^{k} d_{ij}(q) u_j, \qquad i = 1, \dots, k. \tag{8.4}$$

Our aim is to characterize affine in control systems (8.3) which are locally equivalent to linear controllable systems w.r.t. state and feedback transformations (8.4) and to classify them w.r.t. this class of transformations.

First we consider linear controllable systems

$$\dot{x} = Ax + \sum_{i=1}^{k} u_i b_i, \qquad x \in \mathbb{R}^n, \ u = (u_1, \ldots, u_k) \in \mathbb{R}^k, \tag{8.5}$$

where A is an $n \times n$ matrix and b_i, $i = 1, \ldots, k$, are vectors in $\mathbb{R}^n$. We assume that the vectors $b_1, \ldots, b_k$ are linearly independent:

$$\dim \operatorname{span}(b_1, \ldots, b_k) = k.$$

If this is not the case, we eliminate some b_i's. We find normal forms of linear systems w.r.t. linear state and feedback transformations.

To linear systems (8.5) we apply feedback transformations which have the form (8.4) and, moreover, preserve the linear structure:

$$\begin{aligned} c_i(x) &= \langle c_i, x \rangle, \qquad & c_i \in \mathbb{R}^{n*}, \ i = 1, \ldots, k, \\ d_{ij}(x) &= d_{ij} \in \mathbb{R}, \qquad & i, j = 1, \ldots, k. \end{aligned} \tag{8.6}$$

Denote by $D : \operatorname{span}(b_1, \ldots, b_k) \to \operatorname{span}(b_1, \ldots, b_k)$ the linear operator with the matrix (d_{ij}) in the base $b_1, \ldots, b_k$. Linear feedback transformations (8.4), (8.6) map the vector fields in the right-hand side of the linear system (8.5) as follows:

$$(Ax, b_1, \ldots, b_k) \mapsto \left(Ax + \sum_{i=1}^{k} \langle c_i, x \rangle b_i, Db_1, \ldots, Db_k \right). \tag{8.7}$$

Such mapping should be invertible, so we assume that the operator D (or, equivalently, its matrix (d_{ij})) is invertible.

Linear state transformations act on linear systems as follows:

$$(Ax, b_1, \ldots, b_k) \mapsto \left(CAC^{-1}x, Cb_1, \ldots, Cb_k \right), \tag{8.8}$$

where $C : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear operator. State equivalence of linear systems means that these systems have the same coordinate representation in suitably chosen bases in the state space $\mathbb{R}^n$.

8.2 Linear systems with scalar control

Consider a simple model linear control system — scalar high-order control:

$$x^{(n)} + \sum_{i=0}^{n-1} \alpha_i x^{(i)} = u, \qquad u \in \mathbb{R}, \ x \in \mathbb{R}, \tag{8.9}$$

where $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{R}$. We rewrite this system in the standard form in the variables $x_i = x^{(i-1)}$, $i = 1, \dots, n$:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dots \\ \dot{x}_{n-1} = x_n, \\ \dot{x}_n = -\sum_{i=0}^{n-1} \alpha_i x_{i+1} + u, \end{cases} \qquad u \in \mathbb{R}, \ x = (x_1, \dots, x_n) \in \mathbb{R}^n. \tag{8.10}$$

It is easy to see that if we take $-\sum_{i=1}^{n-1} \alpha_i x_{i+1} + u$ as a new control, i.e., apply the feedback transformation (8.4), (8.6) with

$$k = 1, \quad c = (-\alpha_0, \dots, -\alpha_{n-1}), \quad d = 1,$$

then system (8.10) maps into the system

$$\begin{cases} \dot{x}_1 = x_2, \\ \dots \\ \dot{x}_{n-1} = x_n, \\ \dot{x}_n = u, \end{cases} \qquad u \in \mathbb{R}, \ x = (x_1, \dots, x_n) \in \mathbb{R}^n, \tag{8.11}$$

which is written in the scalar form as

$$x^{(n)} = u, \qquad u \in \mathbb{R}, \ x \in \mathbb{R}. \tag{8.12}$$

So system (8.10) is feedback equivalent to system (8.11).

It turns out that the simple systems (8.10) and (8.11) are normal forms of linear controllable systems with scalar control under state transformations and state-feedback transformations respectively.

Proposition 8.1. *Any linear controllable system with scalar control*

$$\dot{x} = Ax + ub, \qquad u \in \mathbb{R}, \ x \in \mathbb{R}^n, \tag{8.13}$$

$$\operatorname{span}(b, Ab, \dots, A^{n-1}b) = \mathbb{R}^n, \tag{8.14}$$

is state equivalent to a system of the form (8.10)*, thus state-feedback equivalent to system* (8.11)*.*

Proof. We find a basis $e_1, \dots, e_n$ in $\mathbb{R}^n$ in which system (8.13) is written in the form (8.10). Coordinates $y_1, \dots, y_n$ of a point $x \in \mathbb{R}^n$ in a basis $e_1, \dots, e_n$ are found from the decomposition

$$x = \sum_{i=1}^{n} y_i e_i.$$

In view of the desired form (8.10), the vector b should have coordinates $b = (0, \dots, 0, 1)^*$, thus the n-th basis vector is uniquely determined:

$$e_n = b.$$

Now we find the rest basis vectors $e_1, \dots, e_{n-1}$. We can rewrite our linear system (8.13) as follows:

$$\dot{x} = Ax \mod \mathbb{R}b,$$

then we obtain in coordinates:

$$\dot{x} = \sum_{i=1}^{n} \dot{y}_i e_i = \sum_{i=1}^{n} y_i A e_i \mod \mathbb{R}b,$$

thus

$$\sum_{i=1}^{n-1} \dot{y}_i e_i = \sum_{i=0}^{n-1} y_{i+1} A e_{i+1} \mod \mathbb{R}b.$$

The required differential equations:

$$\dot{y}_i = y_{i+1}, \qquad i = 1, \dots, n-1,$$

are fulfilled in a basis $e_1, \dots, e_n$ if and only if the following equalities hold:

$$Ae_{i+1} = e_i + \beta_i b, \qquad i = 1, \dots, n-1, \tag{8.15}$$

$$Ae_1 = \beta_0 b \tag{8.16}$$

for some numbers $\beta_0, \dots, \beta_{n-1} \in \mathbb{R}$.

So it remains to show that we can find basis vectors $e_1, \dots, e_{n-1}$ which satisfy equalities (8.15), (8.16). We rewrite equality (8.15) in the form

$$e_i = Ae_{i+1} - \beta_i b, \qquad i = 1, \dots, n-1, \tag{8.17}$$

and obtain recursively:

$$\begin{aligned} &e_n = b, \\ &e_{n-1} = Ab - \beta_{n-1} b, \\ &e_{n-2} = A^2 b - \beta_{n-1} Ab - \beta_{n-2} b, \\ &\dots \\ &e_1 = A^{n-1} b - \beta_{n-1} A^{n-2} b - \dots - \beta_1 b. \end{aligned} \tag{8.18}$$

So equality (8.16) yields

$$Ae_1 = A^n b - \beta_{n-1} A^{n-1} b - \dots - \beta_1 Ab = \beta_0 b.$$

The equality

$$A^n b = \sum_{i=0}^{n-1} \beta_i A^i b \tag{8.19}$$

is satisfied for a unique n-tuple $(\beta_0, \dots, \beta_{n-1})$ since the vectors $b, Ab, \dots, A^{n-1}b$ form a basis of $\mathbb{R}^n$ (in fact, β_i are coefficients of the characteristic polynomial of A).

With these numbers β_i, the vectors $e_1, \dots, e_n$ given by (8.18) form the required basis. Indeed, equalities (8.15), (8.16) hold by construction. The vectors $e_1, \dots, e_n$ are linearly independent by the controllability condition (8.14). □

Remark. The basis $e_1, \dots, e_n$ constructed in the previous proof is unique, thus the state transformation that maps a controllable linear system with scalar control (8.13) to the normal form (8.10) is also unique.

8.3 Linear systems with vector control parameters

Now consider general controllable linear systems:

$$\dot{x} = Ax + \sum_{i=1}^{k} u_i b_i, \qquad x \in \mathbb{R}^n, \ u = (u_1, \dots, u_k) \in \mathbb{R}^k, \tag{8.20}$$

$$\operatorname{span}\{A^j b_i \mid j = 0, \dots, n-1, \ i = 1, \dots, k\} = \mathbb{R}^n. \tag{8.21}$$

Recall that we assume vectors $b_1, \dots, b_k$ linearly independent.

In the case $k = 1$, all controllable linear systems in $\mathbb{R}^n$ are state-feedback equivalent to the normal form (8.11), thus there are no state-feedback invariants in a given dimension n.

If $k > 1$, this is not the case, and we start from description of state-feedback invariants.

Consider the following subspaces in $\mathbb{R}^n$:

$$D^m = \operatorname{span}\{A^j b_i \mid j = 0, \dots, m-1, \ i = 1, \dots, k\}, \qquad m = 1, \dots, n. \tag{8.22}$$

Invertible linear state transformations (8.8) preserve dimension of these subspaces, thus the numbers

$$\dim D^m, \qquad m = 1, \dots, n,$$

are state invariants.

Now we show that invertible linear feedback transformations (8.7) preserve the spaces D^m, $m = 1, \dots, n$. Any such transformation can be decomposed into two feedback transformations of the form:

$$(Ax, b_1, \dots, b_k) \mapsto (Ax + \sum_{i=1}^{k} \langle c_i, x \rangle b_i, b_1, \dots, b_k), \tag{8.23}$$

$$(Ax, b_1, \dots, b_k) \mapsto (Ax, Db_1, \dots, Db_k). \tag{8.24}$$

Transformations (8.24), i.e., changes of b_i, obviously preserve the spaces D^m. Consider transformations (8.23). Denote the new matrix:

$$\widehat{A}x = Ax + \sum_{i=1}^{k} \langle c_i, x \rangle b_i.$$

We have:

$$\widehat{A}^j x = A^j x \mod D^j, \qquad j = 1, \dots, n-1.$$

But $D^{m-1} \subset D^m$, $m = 2, \dots, n$, thus feedback transformations (8.23) preserve the spaces D^m, $m = 1, \dots, n$.

So the spaces D^m, $m = 1, \dots, n$, are invariant under feedback transformations, and their dimensions are state-feedback invariants.

Now we express the numbers $\dim D^m$, $m = 1, \dots, n$, through other integers — Kronecker indices. Construct the following $n \times k$ matrix:

$$\begin{pmatrix} b_1 & \cdots & b_k \\ Ab_1 & \cdots & Ab_k \\ \vdots & \vdots & \vdots \\ A^{n-1}b_1 & \cdots & A^{n-1}b_k \end{pmatrix}. \tag{8.25}$$

Replace each vector $A^j b_i$, $j = 0, \dots, n-1$, $i = 1, \dots, k$, in this matrix by a sign: cross $\times$ or circle $\circ$, by the following rule. We go in matrix (8.25) by raws, i.e., order its elements as follows:

$$b_1, \dots, b_k, Ab_1, \dots, Ab_k, \dots, A^{n-1}b_1, \dots, A^{n-1}b_k. \tag{8.26}$$

A vector $A^j b_i$ in matrix (8.25) is replaced by $\times$ if it is linearly independent of the previous vectors in chain (8.26), otherwise it is replaced by $\circ$. After this procedure we obtain a matrix of the form:

$$\Sigma = \begin{pmatrix} \times & \times & \times & \times & \cdots & \times \\ \times & \circ & \times & \times & \cdots & \circ \\ \times & \circ & \circ & \times & \cdots & \circ \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \circ & \circ & \circ & \times & \cdots & \circ \end{pmatrix}.$$

Notice that there are some restrictions on appearance of crosses and circles in matrix Σ. The total number of crosses in this matrix is n (by the controllability condition (8.21)), and the first row is filled only with crosses (since $b_1, \dots, b_k$ are linearly independent). Further, if a column of Σ contains a circle, then all elements below it are circles as well. Indeed, if a vector $A^j b_i$ in (8.25) is replaced by circle in Σ, then

$$A^j b_i \in \operatorname{span}\{A^j b_\alpha \mid \alpha < i\} + \operatorname{span}\{A^\beta b_\alpha \mid \beta < j, \ \alpha = 1, \dots, k\}.$$

Then the similar inclusions hold for all vectors $A^{j+1}b_i, \dots, A^{n-1}b_i$, i.e., below circles are only circles. So each column in the matrix Σ consists of a column of crosses over a column of circles (the column of circles can be absent).

Denote by n_1 the height of the highest column of crosses in the matrix Σ, by n_2 the height of the next highest column of crosses, ... , and by n_k the height of the lowest column of crosses in Σ. The positive integers obtained:

$$n_1 \geq n_2 \geq \cdots \geq n_k$$

are called *Kronecker indices* of the linear control system (8.20). Since the total number of crosses in matrix Σ is equal to dimension of the state space, then

$$\sum_{i=1}^{k} n_i = n = \dim \mathbb{R}^n .$$

Moreover, by the construction, we have

$$\operatorname{span}(b_1, Ab_1, \dots, A^{n_1-1}b_1; \dots ; b_k, Ab_k, \dots, A^{n_k-1}b_k) = \mathbb{R}^n . \tag{8.27}$$

Now we show that Kronecker indices n_i are expressed through the numbers $\dim D^i$. We have:

$$\begin{aligned}
&\dim D^1 = k = \text{number of crosses in the first row of } \Sigma, \\
&\dim D^2 = \text{number of crosses in the first 2 rows of } \Sigma, \\
&\dots \\
&\dim D^i = \text{number of crosses in the first } i \text{ rows of } \Sigma,
\end{aligned}$$

so that

$$\Delta(i) = \dim D^i - \dim D^{i-1} = \text{ number of crosses in the } i\text{-th row of } \Sigma.$$

Permute columns in matrix Σ, so that the first column become the highest one, the second column becomes the next highest one, etc. We obtain an $n\times k$-matrix in the "block-triangular" form. This matrix rotated at the angle $\pi/2$ gives the subgraph of the function $\Delta : \{1, \dots, n\} \to \{1, \dots, k\}$. It is easy to see that the values of the Kronecker indices is equal to the points of jumps of the function Δ, and the number of Kronecker indices for each value is equal to the height of the corresponding jump of Δ.

So Kronecker indices are expressed through $\dim D^i$, $i = 1, \dots, k$, thus are state-feedback invariants.

Now we show that the set of Kronecker indices n_i, $i = 1, \dots, k$, is a complete set of state-feedback invariants of controllable linear systems (8.20).

We show first that any linear controllable system (8.20) can be written, in a suitable basis in $\mathbb{R}^n$:

$$e_1^1, \dots, e_{n_1}^1; \dots ; e_1^k, \dots, e_{n_k}^k \tag{8.28}$$

in the following canonical form:

$$\begin{cases} \dot y_1^1 = y_2^1, \\ \dots \\ \dot y_{n_1-1}^1 = y_{n_1}^1, \\ \dot y_{n_1}^1 = -\sum_{i=0}^{n_1-1} \alpha_i^1 y_{i+1}^1 + u_1, \end{cases} \quad , \quad \dots , \quad \begin{cases} \dot y_1^k = y_2^k, \\ \dots \\ \dot y_{n_k-1}^k = y_{n_k}^k, \\ \dot y_{n_k}^k = -\sum_{i=0}^{n_k-1} \alpha_i^k y_{i+1}^k + u_k, \end{cases} \tag{8.29}$$

where

$$x = \sum_{\substack{1 \le i \le k \\ 1 \le j \le n_i}} y_j^i e_j^i. \tag{8.30}$$

We proceed exactly as in the scalar-input case (Section 8.2). The required canonical form (8.29) determines uniquely the last basis vectors in all k groups:

$$e_{n_1}^1 = b_1, \dots, e_{n_k}^k = b_k. \tag{8.31}$$

Denote the space $B = \operatorname{span}(b_1, \dots, b_k)$. Then our system

$$\dot{x} = Ax \mod B$$

reads in coordinates as follows:

$$\dot{x} = \sum_{\substack{1 \le i \le k \\ 1 \le j \le n_i}} \dot{y}_j^i e_j^i = \sum_{\substack{1 \le i \le k \\ 1 \le j \le n_i}} y_j^i A e_j^i \mod B.$$

In view of the required equations

$$\dot{y}_j^i = y_{j+1}^i, \qquad 1 \le i \le k, \quad 1 \le j < n_i,$$

we have

$$\sum_{\substack{1 \le i \le k \\ 1 \le j < n_i}} y_{j+1}^i e_j^i = \sum_{\substack{1 \le i \le k \\ 1 \le j \le n_i}} y_j^i A e_j^i \mod B,$$

or, equivalently,

$$\sum_{\substack{1 \le i \le k \\ 2 \le j \le n_i}} y_j^i e_{j-1}^i = \sum_{\substack{1 \le i \le k \\ 1 \le j \le n_i}} y_j^i A e_j^i \mod B.$$

So the following relations should hold for the required basis vectors:

$$Ae_j^i = e_{j-1}^i \mod B, \qquad 1 \le i \le k, \quad 2 \le j \le n_i, \tag{8.32}$$

$$Ae_1^i = 0 \mod B, \qquad 1 \le i \le k. \tag{8.33}$$

We resolve equations (8.32) recursively starting from (8.31), for all $i = 1, \dots, k$:

$$\begin{aligned}
e_{n_i}^i &= b_i, \\
e_{n_i-1}^i &= Ab_i - \sum_{\alpha=1}^{n} \beta_{i,n_i-1}^{\alpha} b_\alpha, \\
e_{n_i-2}^i &= A^2 b_i - \sum_{\alpha=1}^{n} \beta_{i,n_i-1}^{\alpha} A b_\alpha - \sum_{\alpha=1}^{n} \beta_{i,n_i-2}^{\alpha} b_\alpha, \\
&\dots \\
e_1^i &= A^{n_i-1} b_i - \sum_{\alpha=1}^{n} \beta_{i,n_i-1}^{\alpha} A^{n_i-2} b_\alpha - \dots - \sum_{\alpha=1}^{n} \beta_{i,1}^{\alpha} b_\alpha,
\end{aligned}$$

while (8.33) yields

$$Ae_1^i = \sum_{\alpha=1}^{n} \beta_{i,0}^{\alpha} b_{\alpha}$$

for some constants $\beta_{i,j}^{\alpha}$, $1 \le i \le k$, $0 \le j \le n_i$, $1 \le \alpha \le n$. We obtain the equation

$$A^{n_i} b_i = \sum_{\alpha=1}^{n} \beta_{i,n_i-1}^{\alpha} A^{n_i-1} b_{\alpha} + \dots + \sum_{\alpha=1}^{n} \beta_{i,0}^{\alpha} b_{\alpha},$$

which has a unique solution in $\beta_{i,j}^{\alpha}$ in view of (8.27).

So we proved that there exists a unique linear state transformation that maps a linear controllable system (8.20) to the canonical form (8.29).

Choosing $-\sum_{i=0}^{n_j-1} \alpha_i^j y_{i+1}^j + u_j$, $j = 1, \dots, k$, as new controls, we see that each of the k subsystems in (8.29) is feedback equivalent to a system of the form (8.11), or, equivalently, (8.12).

Thus the whole system (8.20) is state-feedback equivalent to the following normal form:

$$\begin{cases} y_1^{(n_1)} = u_1, \\ \dots \\ y_k^{(n_k)} = u_k, \end{cases} \tag{8.34}$$

called *Brunovsky normal form.*

We proved the following statement.

Theorem 8.1. *Any controllable linear system* (8.20), (8.21) *with k control parameters is state equivalent to a system of the form* (8.29) *and state-feedback equivalent to a system in Brunovsky normal form* (8.34), *where n_i, $i = 1, \dots, k$, are Kronecker indices of system* (8.20).

So the action of the group of invertible state-feedback transformations on controllable linear systems with k control parameters has only discrete invariants — Kronecker indices $n_1, \dots, n_k$.

8.4 State-feedback linearizability

Consider a nonlinear affine in control system:

$$\dot{q} = f(q) + \sum_{j=1}^{k} u_j g_j(q), \qquad u = (u_1, \dots, u_k) \in \mathbb{R}^k, \ q \in M. \tag{8.35}$$

We are interested, when such a system is locally state-feedback equivalent to a controllable linear system.

Definition 8.2. System (8.35) is called locally state-feedback equivalent to a linear system (8.20) in a neighborhood of a point $q_0 \in M$, if there exist a state transformation — a diffeomorphism

$$\Phi : O_{q_0} \to \widehat{O} \subset \mathbb{R}^n$$

from a neighborhood O_{q_0} of q_0 in M onto an open subset $\widehat{O} \subset \mathbb{R}^n$, and a feedback transformation

$$\varphi : O_{q_0} \times \mathbb{R}^k \to \mathbb{R}^k,$$

$$\varphi(q,u) = \begin{pmatrix} a_1(q) \\ \cdots \\ a_k(q) \end{pmatrix} + D(q)u, \tag{8.36}$$

with an invertible and smooth in q matrix

$$D(q) = (d_{ij}(q)), \qquad i,\ j = 1,\dots,k,$$

such that the state-feedback transformation (Φ, φ) maps system (8.35) restricted to O_{q_0} to a linear system (8.20) restricted to $\widehat{O}$.

We can generalize the construction of the subspaces D^m (8.22) for the case of nonlinear systems (8.35): consider the families of subspaces

$$D_q^m = \operatorname{span}\{(\operatorname{ad} f)^j g_i(q) \mid j = 0,\dots,m-1,\ i = 1,\dots,k\} \subset T_q M.$$

Notice that, in general, $\dim D_q^m \neq \text{const}$, thus D^m is not a distribution.

Observe that for controllable linear systems (8.20), the following properties hold for the family $D_x^m \equiv D^m$, $x \in \mathbb{R}^n$:

1. $\dim D_x^m = \text{const}$,
2. $D_x^n = T_x \mathbb{R}^n$,
3. the distributions D^m, $m = 1,\dots,n$, are integrable (since they are spanned by the constant vector fields $A^j b_i$).

Before formulating conditions for state-feedback linearizability of nonlinear systems, which are given in terms of the families D_q^m, we prove the following property of these families.

Lemma 8.1. *If the families D^m, $m = 1,\dots,n$, are involutive, then they are feedback-invariant.*

Proof. Notice first that feedback transformations (8.36) can be decomposed into transformations of the two kinds:

$$(f, g_1,\dots,g_k) \mapsto (f + a_j g_j, g_1,\dots,g_k), \tag{8.37}$$

$$(f, g_1,\dots,g_k) \mapsto (f, Dg_1,\dots,Dg_k), \tag{8.38}$$

where $D(q) = (d_{ij}(q))$, $i, j = 1, \dots, k$, is invertible and smooth w.r.t. q. We prove the lemma by induction on m.

Let $m = 1$. The family

$$D^1 = \operatorname{span}\{g_i \mid i = 1, \dots, k\}$$

is obviously preserved by the both transformations (8.37) and (8.38).

Induction step: we assume that the statement is proved for $m - 1$ and prove it for m. The family

$$D^m = \{[f, X] \mid X \in D^{m-1}\} + D^{m-1}$$

is preserved by transformation (8.38). Consider transformation (8.37). We have

$$[f + a_j g_j, X] = [f, X] - [X, a_j g_j] = [f, X] - (X a_j) g_j - a_j [X, g_j].$$

Further:

$$\begin{aligned}
&X \in D^{m-1} \;\Rightarrow\; [f, X] \in D^m,\\
&(X a_j) g_j \in D^1 \subset D^m,\\
&X \in D^{m-1},\; g_j \in D^1 \subset D^{m-1} \;\Rightarrow\; [X, g_j] \in D^{m-1} \subset D^m,
\end{aligned}$$

thus

$$[f + a_j g_j, X] \in D^m \quad \forall\, X \in D^{m-1}.$$

So D^m is preserved by feedback transformation (8.37). □

Theorem 8.2. *System* (8.35) *is locally state-feedback equivalent to a controllable linear system* (8.20) *if and only if:*

(1) $\dim D_q^m$, $m = 1, \dots, n$, *does not depend on* q, *i.e.,* D^m *are distributions,*

(2) $D_q^n = T_q M$,

(3) *the distributions* D^m, $m = 1, \dots, n$, *are involutive.*

Conditions (1)–(3) are necessary for local state-feedback linearizability, see discussion before Lemma 8.1.

We prove sufficiency in Theorem 8.3 below only in the case of scalar control parameter. For $k = 1$ we have the system

$$\dot q = f(q) + u g(q), \qquad u \in \mathbb{R},\; q \in M, \tag{8.39}$$

and the corresponding families of subspaces

$$D_q^m = \operatorname{span}\{(\operatorname{ad} f)^i g(q) \mid i = 0, 1, \dots, m - 1\}, \qquad m = 1, \dots, n,\; q \in M.$$

In this case it happens that involutivity of D^{n-1} implies involutivity of D^m with smaller m.

Theorem 8.3. *System* (8.39) *is locally state-feedback equivalent to a controllable linear system* (8.13) *if and only if:*

(1) $D_q^n = T_q M$,

(2) *the distribution* D^{n-1} *is involutive.*

First we prove the following proposition of general interest: integral manifolds of integrable distributions can be smoothly parametrized.

Lemma 8.2. *Let* $\Delta = \operatorname{span}\{X_1, \dots, X_k\}$ *be an integrable distribution on a smooth* n*-dimensional manifold* M, $\dim \Delta_q = k$. *Then for any point* $q_0 \in M$ *there exist a neighborhood* $q_0 \in O_{q_0} \subset M$ *and a smooth vector-function*

$$\varphi \,:\, O_{q_0} \to \mathbb{R}^{n-k}$$

such that:

(1) $\operatorname{rank} \varphi_{*q} = n - k$, $q \in O_{q_0}$, *and*

(2) $\varphi^{-1}(y)$ *is an integral manifold of* Δ *for any* $y \in \varphi(O_{q_0})$, *or, equivalently,*

(2′) $\ker \varphi_{*q} = \Delta_q$, $q \in O_{q_0}$.

Proof. Complete the vector fields $X_1, \dots, X_k$ to a basis:

$$\operatorname{span}\{Y_1, \dots, Y_{n-k}, X_1, \dots, X_k\} = \operatorname{Vec} O_{q_0},$$

for a sufficiently small neighborhood $q_0 \in O_{q_0} \subset M$. Consider the mapping

$$\psi \,:\, (t, s) \mapsto q_0 \circ e^{t_1 Y_1} \circ \dots \circ e^{t_{n-k} Y_{n-k}} \circ e^{s_1 X_1} \circ \dots \circ e^{s_k X_k},$$
$$t = (t_1, \dots, t_{n-k}) \in \mathbb{R}^{n-k}, \; s = (s_1, \dots, s_k) \in \mathbb{R}^k.$$

We have

$$\left.\frac{\partial \psi}{\partial t_i}\right|_0 = Y_i, \qquad i = 1, \dots, n-k,$$
$$\left.\frac{\partial \psi}{\partial s_i}\right|_0 = X_i, \qquad i = 1, \dots, k,$$

thus ψ is a local diffeomorphism in a neighborhood of $0 \in \mathbb{R}^n$.

Further, for fixed $t = t^0$, the set

$$\{\psi(t^0, s) \mid s \in \mathbb{R}^k\}$$

is an integral manifold of Δ.

Finally, locally, by the implicit function theorem, there exists a well-defined smooth mapping

$$\varphi \,:\, \psi(t, s) \mapsto t.$$

It is the required vector-function. □

Now we prove Theorem 8.3.

Proof. Necessity is already known since for linear controllable systems both conditions (1), (2) hold, see discussion before Lemma 8.1.

To prove sufficiency, we construct coordinates in which our system (8.39) is simplified, and then apply a feedback transformation which maps this system to the normal form (8.11).

Since the distribution D^{n-1} is integrable, then by Lemma 8.2 there exists a smooth function

$$\varphi_1 : O_{q_0} \to \mathbb{R}$$

such that

$$d_q\varphi_1 \neq 0, \quad \langle d_q\varphi_1, D_q^{n-1}\rangle = 0, \qquad q \in O_{q_0}. \tag{8.40}$$

Define the following functions in the neighborhood O_{q_0}:

$$\begin{aligned}
&\varphi_2 = f\varphi_1 = \langle d\varphi_1, f\rangle, \\
&\varphi_3 = f\varphi_2 = f^2\varphi_1, \\
&\dots \\
&\varphi_n = f\varphi_{n-1} = f^{n-1}\varphi_1
\end{aligned}$$

(iterated directional derivatives along the vector field f).

We claim that the functions $\varphi_1, \dots, \varphi_n$ (which will be the coordinates that simplify (8.39)) have the following property:

$$(\operatorname{ad} f)^j g\varphi_l = \begin{cases} 0, & j+l < n, \\ \pm(\operatorname{ad} f)^{n-1} g\varphi_1 \neq 0, & j+l = n. \end{cases} \tag{8.41}$$

First of all, notice that $b = (\operatorname{ad} f)^{n-1} g\varphi_1\big|_{O_{q_0}} \neq 0$. Indeed, we have

$$\begin{aligned}
&D_q^{n-1} = \operatorname{span}\{g(q), \dots, (\operatorname{ad} f)^{n-2} g(q)\}, \\
&T_qM = \operatorname{span}\{g(q), \dots, (\operatorname{ad} f)^{n-1} g(q)\} = \operatorname{span}\{D_q^{n-1}, (\operatorname{ad} f)^{n-1} g(q)\},
\end{aligned}$$

thus the equality $(\operatorname{ad} f)^{n-1} g\varphi_1(q) = 0$ is incompatible with properties (8.40).

Now we prove (8.41) by induction on l. If $l = 1$, there is nothing to prove.

Assume that equality (8.41) is proved for $l-1$ and prove it for l. We have

$$\begin{aligned}
(\operatorname{ad} f)^j g\varphi_l &= \left((\operatorname{ad} f)^j g \circ f\right) \varphi_{l-1} \\
&= \left((\operatorname{ad} f)^j g \circ f - f \circ (\operatorname{ad} f)^j g + f \circ (\operatorname{ad} f)^j g\right) \varphi_{l-1} \\
&= \left(-[f, (\operatorname{ad} f)^j g] + f \circ (\operatorname{ad} f)^j g\right) \varphi_{l-1} \\
&= \left(-(\operatorname{ad} f)^{j+1} g + f \circ (\operatorname{ad} f)^j g\right) \varphi_{l-1}.
\end{aligned}$$

If $j+l \leq n$, then $j+l-1 < n$, and $(\operatorname{ad} f)^j g\varphi_{l-1} = 0$ by the induction assumption. Thus

$$(\operatorname{ad} f)^j g\varphi_l = -(\operatorname{ad} f)^{j+1} g\varphi_{l-1} \qquad \text{for } j+l \leq n,$$

and equality (8.41) for l follows from this equality for $l-1$.

So equality (8.41) is proved for all l. Since the vectors $g(q), \dots, (\operatorname{ad} f)^{n-1} g(q)$ span the tangent space $T_q M$ for $q \in O_{q_0}$, the mapping

$$\Phi = \begin{pmatrix} \varphi_1 \\ \cdots \\ \varphi_n \end{pmatrix} : O_{q_0} \to \mathbb{R}^n$$

is a local diffeomorphism: the differentials $d_q\varphi_1, \dots, d_q\varphi_n$ form a basis of $T_q^* M$ dual to $g(q), \dots, (\operatorname{ad} f)^{n-1} g(q) \in T_q M$.

Take Φ as a coordinate mapping, then coordinates of a point $q \in M$ are

$$x_l = \varphi_l(q), \qquad l = 1, \dots, n.$$

Now we write our system $\dot{q} = f(q) + ug(q)$ in these coordinates: we differentiate x_l with respect to this system.

$$\frac{d}{dt} x_l = \frac{d}{dt} \varphi_l(q(t)) = (f + ug)\varphi_l = f\varphi_l + ug\varphi_l.$$

If $l < n$, then $g\varphi_l = 0$ by equality (8.41), thus

$$\frac{d}{dt} x_l = f\varphi_l = \varphi_{l+1} = x_{l+1}, \qquad l = 1, \dots, n-1.$$

And if $l = n$, then

$$\frac{d}{dt} x_n = f\varphi_n + ug\varphi_n = f\varphi_n \pm ub, \qquad b = g\varphi_n \neq 0.$$

So in coordinates $x_1, \dots, x_n$ our system (8.39) takes the form

$$\begin{cases} \dot{x}_1 = x_2, \\ \cdots \\ \dot{x}_{n-1} = x_n, \\ \dot{x}_n = f\varphi_n \pm ub. \end{cases}$$

Now consider the feedback transformation

$$u \mapsto \mp \frac{f\varphi_n - u}{b}.$$

After this transformation the n-th component of our system reads

$$\dot{x}_n = f\varphi_n \pm \left(\mp \frac{f\varphi_n - u}{b} \right) b = f\varphi_n - f\varphi_n + u = u,$$

i.e., the whole system takes the required form (8.11). □

Chapter 9

Optimal control problem

9.1 Problem statement

Consider a control system of the form

$$\dot{q} = f_u(q), \qquad q \in M, \quad u \in U \subset \mathbb{R}^m. \tag{9.1}$$

Here M is, as usual, a smooth manifold, and U an arbitrary subset of $\mathbb{R}^m$. For the right-hand side of the control system, we suppose that:

$$q \mapsto f_u(q) \text{ is a smooth vector field on } M \text{ for any fixed } u \in U, \tag{9.2}$$

$$(q, u) \mapsto f_u(q) \text{ is a continuous mapping for } q \in M,\ u \in \overline{U}, \tag{9.3}$$

and moreover, in any local coordinates on M

$$(q, u) \mapsto \frac{\partial f_u}{\partial q}(q) \text{ is a continuous mapping for } q \in M,\ u \in \overline{U}. \tag{9.4}$$

Admissible controls are measurable locally bounded mappings

$$u : t \mapsto u(t) \in U.$$

Substitute such a control $u = u(t)$ for control parameter into system (9.1), then we obtain a nonautonomous ODE $\dot{q} = f_u(q)$. By the classical Carathéodory's Theorem, for any point $q_0 \in M$, the Cauchy problem

$$\dot{q} = f_u(q), \qquad q(0) = q_0, \tag{9.5}$$

has a unique solution, see Subsec. 2.4.1. We will often fix the initial point q_0 and then denote the corresponding solution to problem (9.5) as $q_u(t)$.

In order to compare admissible controls one with another on a segment $[0, t_1]$, introduce a *cost functional*:

$$J(u) = \int_0^{t_1} \varphi(q_u(t), u(t))\, dt \tag{9.6}$$

with an integrand

$$\varphi : M \times U \to \mathbb{R}$$

satisfying the same regularity assumptions as the right-hand side f, see (9.2)–(9.4).

Take any pair of points $q_0, q_1 \in M$. We consider the following *optimal control problem*:

MINIMIZE THE FUNCTIONAL J AMONG ALL ADMISSIBLE CONTROLS $u = u(t)$, $t \in [0, t_1]$, FOR WHICH THE CORRESPONDING SOLUTION $q_u(t)$ OF CAUCHY PROBLEM (9.5) SATISFIES THE BOUNDARY CONDITION

$$q_u(t_1) = q_1. \tag{9.7}$$

We study two types of problems, with fixed t_1 and free t_1. A solution u of this problem is called an *optimal control*, and the corresponding curve $q_u(t)$ is the *optimal trajectory*.

So the optimal control problem is the minimization problem for $J(u)$ with constraints on u given by control system and the fixed endpoints conditions (9.5), (9.7). These constraints cannot usually be resolved w.r.t. u, thus solving optimal control problems requires special techniques.

9.2 Reduction to study of attainable sets

Fix an initial point $q_0 \in M$. *Attainable set* of control system (9.1) for time $t \geq 0$ from q_0 with measurable locally bounded controls is defined as follows:

$$\mathcal{A}_{q_0}(t) = \{q_u(t) \mid u \in L^\infty([0,t], U)\}.$$

Similarly, one can consider the attainable sets for time not greater than t:

$$\mathcal{A}^t_{q_0} = \bigcup_{0 \leq \tau \leq t} \mathcal{A}_{q_0}(\tau)$$

and for arbitrary nonnegative time:

$$\mathcal{A}_{q_0} = \bigcup_{0 \leq \tau < \infty} \mathcal{A}_{q_0}(\tau).$$

It turns out that optimal control problems on the state space M can be essentially reduced to the study of attainable sets of some auxiliary control systems on the extended state space

$$\widehat{M} = \mathbb{R} \times M = \{\widehat{q} = (y, q) \mid y \in \mathbb{R},\ q \in M\}.$$

Namely, consider the following extended control system on $\widehat{M}$:

$$\frac{d\widehat{q}}{dt} = \widehat{f}_u(\widehat{q}), \qquad \widehat{q} \in \widehat{M},\ u \in U, \tag{9.8}$$

with the right-hand side

$$\widehat{f}_u(\widehat{q}) = \begin{pmatrix} \varphi(q,u) \\ f_u(q) \end{pmatrix}, \qquad q \in M, \quad u \in U,$$

where φ is the integrand of the cost functional J, see (9.6). Then solutions $\widehat{q}_u(t)$ of the extended system (9.8) with the initial conditions

$$\widehat{q}_u(0) = \begin{pmatrix} y(0) \\ q(0) \end{pmatrix} = \begin{pmatrix} 0 \\ q_0 \end{pmatrix}$$

are expressed through solutions $q_u(t)$ of the original system (9.1) as

$$\widehat{q}_u(t) = \begin{pmatrix} J_t(u) \\ q_u(t) \end{pmatrix},$$

where

$$J_t(u) = \int_0^t \varphi(q_u(\tau), u(\tau))\, d\tau.$$

Thus attainable sets of the extended system (9.8) from the point $(0, q_0)$ have the form

$$\widehat{\mathcal{A}}_{(0,q_0)}(t) = \{(J_t(u), q_u(t)) \mid u \in L^\infty([0,t], U)\}.$$

Let $q(t)$, $t \in [0, t_1]$, be an optimal trajectory for the optimal control problem in M. Consider the corresponding trajectory

$$\widehat{q}(t) = \begin{pmatrix} J_t \\ q(t) \end{pmatrix}, \qquad t \in [0, t_1],$$

of the extended control system in $\widehat{M}$. The endpoint $\widehat{q}(t_1)$ must belong to the boundary of the attainable set $\widehat{\mathcal{A}}_{(0,q_0)}(t_1)$; moreover, this set should not intersect the ray

$$\{(y, q_1) \in \widehat{M} \mid y < J_{t_1}\}.$$

Indeed, if there exist points

$$(y, q_1) \in \widehat{\mathcal{A}}_{(0,q_0)}(t_1), \quad y < J_{t_1},$$

then the trajectory of the extended system

$$\widehat{q'}(t) = \begin{pmatrix} J'_t \\ q'(t) \end{pmatrix}$$

that steers $(0, q_0)$ to (y, q_1):

$$\widehat{q'}(0) = \begin{pmatrix} 0 \\ q_0 \end{pmatrix}, \qquad \widehat{q'}(t_1) = \begin{pmatrix} y \\ q_1 \end{pmatrix},$$

gives a trajectory $q'(t)$, $q'(0) = q_0$, $q'(t_1) = q_1$, with a smaller value of the cost functional:

$$J'_{t_1} = y < J_{t_1},$$

a contradiction with optimality of $q(\cdot)$.

So optimal trajectories (more precisely, their lift to the extended state space $\widehat{M}$) must come to the boundary of the attainable set $\widehat{\mathcal{A}}_{(0,q_0)}(t_1)$. In order to find optimal trajectories, we find those coming to the boundary of $\widehat{\mathcal{A}}_{(0,q_0)}(t_1)$, and then select optimal among them. The first step is much more important than the second one, so solving optimal control problems essentially reduces to the study of dynamics of boundary of attainable sets.

9.3 Compactness of attainable sets

Due to the reduction of optimal control problems to the study of attainable sets, existence of optimal solutions to these problems is reduced to compactness of attainable sets.

For control system (9.1), sufficient conditions for compactness of the attainable sets $\mathcal{A}_{q_0}(t)$ for time t and $\mathcal{A}_{q_0}^t$ for time not greater than t are given in the following proposition.

Theorem 9.1 (Filippov). *Let the space of control parameters $U \Subset \mathbb{R}^m$ be compact. Let there exist a compact $K \Subset M$ such that $f_u(q) = 0$ for $q \notin K$, $u \in U$. Moreover, let the velocity sets*

$$f_U(q) = \{f_u(q) \mid u \in U\} \subset T_qM, \qquad q \in M,$$

be convex. Then the attainable sets $\mathcal{A}_{q_0}(t)$ and $\mathcal{A}_{q_0}^t$ are compact for all $q_0 \in M$, $t > 0$.

Remark. The condition of convexity of the velocity sets $f_U(q)$ is natural in view of Theorem 7.2: the flow of the ODE

$$\dot{q} = \alpha(t) f_{u_1}(q) + (1 - \alpha(t)) f_{u_2}(q), \qquad 0 \le \alpha(t) \le 1,$$

can be approximated by flows of the systems of the form

$$\dot{q} = f_v(q), \quad \text{where} \quad v(t) \in \{u_1(t), u_2(t)\}.$$

Now we give a sketch of the proof of Theorem 9.1.

Proof. Notice first of all that all nonautonomous vector fields $f_u(q)$ with admissible controls u have a common compact support, thus are complete. Further, under hypotheses of the theorem, velocities $f_u(q)$, $q \in M$, $u \in U$, are uniformly bounded, thus all trajectories $q(t)$ of control system (9.1) starting at q_0 are Lipschitzian with the same Lipschitz constant. Thus the set of admissible trajectories is precompact in the topology of uniform convergence. (We can embed the manifold M into a Euclidean space $\mathbb{R}^N$, then the space of continuous curves $q(t)$ becomes endowed with the uniform topology of continuous mappings from $[0, t_1]$ to $\mathbb{R}^N$.) For any sequence $q_n(t)$ of admissible trajectories:

$$\dot{q}_n(t) = f_{u_n}(q_n(t)), \qquad 0 \le t \le t_1, \quad q_n(0) = q_0,$$

there exists a uniformly converging subsequence, we denote it again by $q_n(t)$:

$$q_n(\cdot) \to q(\cdot) \text{ in } C[0,t_1] \text{ as } n \to \infty.$$

Now we show that $q(t)$ is an admissible trajectory of control system (9.1).

Fix a sufficiently small $\varepsilon > 0$. Then

$$\begin{aligned} \frac{1}{\varepsilon}(q_n(t+\varepsilon) - q_n(t)) &= \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_{u_n}(q_n(\tau))\, d\tau \\ &\in \operatorname{conv} \bigcup_{\tau\in[t,t+\varepsilon]} f_U(q_n(\tau)) \subset \operatorname{conv} \bigcup_{q\in O_{q(t)}(c\varepsilon)} f_U(q) \end{aligned}$$

where c is the doubled Lipschitz constant of admissible trajectories. Then we pass to the limit $n \to \infty$ and obtain

$$\frac{1}{\varepsilon}(q(t+\varepsilon) - q(t)) \in \operatorname{conv} \bigcup_{q\in O_{q(t)}(c\varepsilon)} f_U(q).$$

Now let $\varepsilon \to 0$. If t is a point of differentiability of $q(t)$, then

$$\dot{q}(t) \in f_U(q)$$

since $f_U(q)$ is convex.

In order to show that $q(t)$ is an admissible trajectory of control system (9.1), we should find a measurable selection $u(t) \in U$ that generates $q(t)$. We do this via the lexicographic order on the set $U = \{(u_1, \dots, u_m)\} \subset \mathbb{R}^m$.

The set

$$V_t = \{v \in U \mid \dot{q}(t) = f_v(q(t))\}$$

is a compact subset of U, thus of $\mathbb{R}^m$. There exists a vector $v_{\min}(t) \in V_t$ minimal in the sense of lexicographic order: to find $v_{\min}(t)$, we minimize the first coordinate v_1 among all $v = (v_1, \dots, v_m) \in V_t$, then minimize the second coordinate v_2 on the compact set found at the first step, etc. The control $u(t) = v_{\min}(t)$ is measurable, thus $q(t)$ is an admissible solution of control system (9.1).

The proof of compactness of the attainable set $\mathcal{A}_{q_0}(t)$ is complete. Compactness of $\mathcal{A}_{q_0}^t$ is proved by a slightly modified argument. □

Remark. In Filippov's theorem, the hypothesis of common compact support of the vector fields in the right-hand side is essential to ensure the uniform boundedness of velocities and completeness of vector fields. On a manifold, sufficient conditions for completeness of a vector field cannot be given in terms of boundedness of the vector field and its derivatives: a constant vector field is not complete on a bounded domain in $\mathbb{R}^n$. Nevertheless, one can prove compactness of attainable sets for many systems without the assumption of common compact support. If for such a system we have a priori bounds on solutions, then we can multiply its right-hand side by a cut-off function, and obtain a system with vector fields having compact support. We can apply Filippov's theorem to the new system. Since trajectories of the initial and new systems coincide in a domain of interest for us, we obtain a conclusion on compactness of attainable sets for the initial system.

For control systems on $M = \mathbb{R}^n$, there exist well-known sufficient conditions for completeness of vector fields: if the right-hand side grows at infinity not faster than a linear field, i.e.,

$$|f_u(x)| \leq C(1 + |x|), \qquad x \in \mathbb{R}^n, \quad u \in U, \tag{9.9}$$

for some constant C, then the nonautonomous vector fields $f_u(x)$ are complete (here $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$ is the norm of a point $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$).

These conditions provide an a priori bound for solutions: any solution $x(t)$ of the control system

$$\dot{x} = f_u(x), \qquad x \in \mathbb{R}^n, \quad u \in U, \tag{9.10}$$

with the right-hand side satisfying (9.9) admits the bound

$$|x(t)| \leq e^{2Ct} \left(|x(0)| + 1\right), \qquad t \geq 0.$$

So Filippov's theorem plus the previous remark imply the following sufficient condition for compactness of attainable sets for systems in $\mathbb{R}^n$.

Corollary 9.1. *Let system* (9.10) *have a compact space of control parameters* $U \Subset \mathbb{R}^m$ *and convex velocity sets* $f_U(x)$, $x \in \mathbb{R}^n$. *Suppose moreover that the right-hand side of the system satisfies a bound of the form* (9.9). *Then the attainable sets* $\mathcal{A}_{x_0}(t)$ *and* $\mathcal{A}^t_{x_0}$ *are compact for all* $x_0 \in \mathbb{R}^n$, $t > 0$.

9.4 Time-optimal problem

Given a pair of points $q_0 \in M$, $q_1 \in \mathcal{A}_{q_0}$, the *time-optimal problem* consists in minimizing the time of motion from q_0 to q_1 via admissible controls of control system (9.1):

$$\min_u \{t_1 \mid q_u(t_1) = q_1\}. \tag{9.11}$$

That is, we consider the optimal control problem described in Sec. 9.1 with the integrand $\varphi(q, u) \equiv 1$ and free terminal time t_1.

Reduction of optimal control problems to the study of attainable sets and Filippov's Theorem yield the following existence result.

Corollary 9.2. *Under hypotheses of Theorem* 9.1, *time-optimal problem* (9.1), (9.11) *has a solution for any points* $q_0 \in M$, $q_1 \in \mathcal{A}_{q_0}$.

9.5 Relaxations

Consider a control system of the form (9.1) with a compact set of control parameters U. There is a standard procedure called *relaxation* of control system (9.1), which extends the velocity set $f_U(q)$ of this system to its convex hull $\operatorname{conv} f_U(q)$.

Recall that the *convex hull* $\operatorname{conv} S$ of a subset S of a linear space is the minimal convex set that contains S. A constructive description of convex hull is given by the following classical proposition: any point in the convex hull of a set S in the n-dimensional linear space is contained in the convex hull of some $n+1$ points in S.

Lemma 9.1 (Carathéodory). *For any subset $S \subset \mathbb{R}^n$, its convex hull has the form*

$$\operatorname{conv} S = \left\{ \sum_{i=0}^{n} \alpha_i x_i \mid x_i \in S,\ \alpha_i \geq 0,\ \sum_{i=0}^{n} \alpha_i = 1 \right\}.$$

For the proof of this lemma, one can consult e.g. [13].

Relaxation of control system (9.1) is constructed as follows. Let $n = \dim M$ be dimension of the state space. The set of control parameters of the relaxed system is

$$V = \Delta^n \times \underbrace{U \times \cdots \times U}_{n+1 \text{ times}},$$

where

$$\Delta^n = \left\{ (\alpha_0, \ldots, \alpha_n) \mid \alpha_i \geq 0,\ \sum_{i=0}^{n} \alpha_i = 1 \right\} \subset \mathbb{R}^{n+1}$$

is the standard n-dimensional simplex. So the control parameter of the new system has the form

$$v = (\alpha, u_0, \ldots, u_n) \in V, \qquad \alpha = (\alpha_0, \ldots, \alpha_n) \in \Delta^n,\ u_i \in U.$$

If U is compact, then V is compact as well.

The *relaxed system* is

$$\dot{q} = g_v(q) = \sum_{i=0}^{n} \alpha_i f_{u_i}(q), \qquad v = (\alpha, u_0, \ldots, u_n) \in V, \quad q \in M. \tag{9.12}$$

By Carathéodory's lemma, the velocity set $g_V(q)$ of system (9.12) is convex, moreover,

$$g_V(q) = \operatorname{conv} f_U(q).$$

If all vector fields in the right-hand side of (9.12) have a common compact support, we obtain by Filippov's theorem that attainable sets for the relaxed system are compact. By Theorem 7.2, any trajectory of relaxed system (9.12) can be uniformly approximated by families of trajectories of initial system (9.1). Thus attainable sets of the relaxed system coincide with closure of attainable sets of the initial system.

Chapter 10

Elements of Exterior Calculus and Symplectic Geometry

In order to state necessary conditions of optimality for optimal control problems on smooth manifolds — Pontryagin Maximum Principle, see Chapter 11 — we make use of some standard technique of Symplectic Geometry. In this chapter we develop such a technique. Before this we recall some basic facts on calculus of exterior differential forms on manifolds. The exposition in this chapter is rather explanatory than systematic, it is not a substitute to a regular textbook. For a detailed treatment of the subject, see e.g. [17], [3], [5].

10.1 Differential 1-forms

10.1.1 Linear forms

Let E be a real vector space of finite dimension n. The set of linear forms on E, i.e., of linear mappings $\xi : E \to \mathbb{R}$, has a natural structure of a vector space called the *dual space* to E and denoted by E^*. If vectors $e_1, \dots, e_n$ form a basis of E, then the corresponding *dual basis* of E^* is formed by the covectors $e_1^*, \dots, e_n^*$ such that

$$\langle e_i^*, e_j \rangle = \delta_{ij}, \qquad i,\ j = 1, \dots n$$

(we use the angle brackets to denote the value of a linear form $\xi \in E^*$ on a vector $v \in E$: $\langle \xi, v \rangle = \xi(v)$). So the dual space has the same dimension as the initial one:

$$\dim E^* = n = \dim E.$$

10.1.2 Cotangent bundle

Let M be a smooth manifold and T_qM its tangent space at a point $q \in M$. The space of linear forms on T_qM, i.e., the dual space $(T_qM)^*$ to T_qM, is called the *cotangent space* to M at q and is denoted as T_q^*M. The disjoint union of all cotangent spaces is called the *cotangent bundle* of M:

$$T^*M \stackrel{\text{def}}{=} \bigcup_{q \in M} T_q^*M.$$

The set T^*M has a natural structure of a smooth manifold of dimension $2n$, where $n = \dim M$. Local coordinates on T^*M are constructed from local coordinates on M.

Let $O \subset M$ be a coordinate neighborhood and let

$$\Phi : O \to \mathbb{R}^n = \{(x_1, \dots, x_n)\}$$

be a local coordinate system. Let $e_1, \dots, e_n \in \operatorname{Vec} \mathbb{R}^n$ be the standard basis vector fields on $\mathbb{R}^n$:

$$e_i = \begin{pmatrix} 0 \\ \cdots \\ 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

with the only identity in the i-th row. Then the pull-back vectors

$$\left.\frac{\partial}{\partial x_i}\right|_q \stackrel{\text{def}}{=} \Phi_*^{-1} e_i \in T_qM, \qquad i = 1, \dots, n, \quad q \in O,$$

form a basis of the tangent space T_qM, $q \in O$.

Recall that if $F : M \to N$ is a smooth mapping between smooth manifolds, then the differential

$$F_* : T_qM \to T_{F(q)}N$$

has the adjoint mapping

$$F^* \stackrel{\text{def}}{=} (F_*)^* : T_{F(q)}^*N \to T_q^*M$$

defined as follows:

$$F^*\xi = \xi \circ F_*, \qquad \xi \in T_{F(q)}^*N,$$
$$\langle F^*\xi, v\rangle = \langle \xi, F_*v\rangle, \qquad v \in T_qM.$$

A vector $v \in T_qM$ is pushed forward by the differential F_* to the vector $F_*v \in T_{F(q)}N$, while a form $\xi \in T_{F(q)}^*N$ is pulled back to the form $F^*\xi \in T_q^*M$.

The linear forms

$$dx_i\big|_q \overset{\text{def}}{=} \Phi^* e_i^* \in T_q^* M, \qquad i = 1, \dots, n, \quad q \in O,$$

make up the basis in the cotangent spaces $T_q^* M$, $q \in O$, dual to the basis $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$ in the tangent spaces $T_q M$:

$$\left\langle dx_i, \frac{\partial}{\partial x_j} \right\rangle = \delta_{ij}, \qquad i,\ j = 1, \dots, n.$$

Any linear form $\xi \in T_q^* M$ can be decomposed via the basis forms:

$$\xi = \sum_{i=1}^{n} \xi_i \, dx_i.$$

So any covector $\xi \in T^* M$ is characterized by n coordinates $(x_1, \dots, x_n)$ of the point $q \in M$ where ξ is attached, and by n coordinates $(\xi_1, \dots, \xi_n)$ of the linear form ξ in the basis $dx_1, \dots, dx_n$. Mappings of the form

$$\xi \mapsto (\xi_1, \dots, \xi_n; x_1, \dots, x_n)$$

define local coordinates on the cotangent bundle. Consequently, $T^* M$ is an $2n$-dimensional manifold. Coordinates of the form (ξ, x) are called *canonical coordinates* on $T^* M$.

10.1.3 Differential 1-forms

A *differential* 1-*form* on M is a smooth mapping

$$q \mapsto \omega_q \in T_q^* M, \qquad q \in M,$$

i.e, a family $\omega = \{\omega_q\}$ of linear forms on the tangent spaces $T_q M$ smoothly depending on the point $q \in M$. The set of all differential 1-forms on M has a natural structure of an infinite-dimensional vector space denoted as $\Lambda^1(M)$.

Like linear forms on a vector space are dual objects to vectors of the space, differential forms on a manifold are dual objects to smooth curves in the manifold. The pairing operation is the *integral* of a differential 1-form $\omega \in \Lambda^1(M)$ along a smooth oriented curve $\gamma : [t_0, t_1] \to M$, defined as follows:

$$\int_\gamma \omega \overset{\text{def}}{=} \int_{t_0}^{t_1} \langle \omega_{\gamma(t)}, \dot{\gamma}(t) \rangle \, dt.$$

The integral of a 1-form along a curve does not change under orientation-preserving smooth reparametrizations of the curve and changes its sign under change of orientation.

10.2 Differential k-forms

A differential k-form on M is an object to integrate over k-dimensional surfaces in M. Infinitesimally, a k-dimensional surface is presented by its tangent space, i.e., a k-dimensional subspace in T_qM. We thus need a dual object to the set of k-dimensional subspaces in the linear space. Fix a linear space E. A k-dimensional subspace is defined by its basis $v_1, \dots, v_k \in E$. The dual objects should be mappings

$$(v_1, \dots, v_k) \mapsto \omega(v_1, \dots, v_k) \in \mathbb{R}$$

such that $\omega(v_1, \dots, v_k)$ depend only on the linear hull $\operatorname{span}\{v_1, \dots, v_k\}$ and the oriented volume of the k-dimensional parallelepiped generated by $v_1, \dots, v_k$. Moreover, the dependence on the volume should be linear. Recall that the ratio of volumes of the parallelepipeds generated by vectors $w_i = \sum_{j=1}^k \alpha_{ij} v_j$, $i = 1, \dots, k$, and the vectors $v_1, \dots, v_k$, equals $\det(\alpha_{ij})_{i,j=1}^k$, and that determinant of a $k \times k$ matrix is a multilinear skew-symmetric form of the columns of the matrix. This is why the following definition of the "dual objects" is quite natural.

10.2.1 Exterior k-forms

Let E be a finite-dimensional real vector space, $\dim E = n$, and let $k \in \mathbb{N}$. An *exterior k-form* on E is a mapping

$$\omega : \underbrace{E \times \cdots \times E}_{k \text{ times}} \to \mathbb{R},$$

which is multilinear:

$$\begin{aligned} &\omega(v_1, \dots, \alpha_1 v_i^1 + \alpha_2 v_i^2, \dots, v_k) \\ &\qquad = \alpha_1 \omega(v_1, \dots, v_i^1, \dots, v_k) + \alpha_2 \omega(v_1, \dots, v_i^2, \dots, v_k), \qquad \alpha_1, \ \alpha_2 \in \mathbb{R}, \end{aligned}$$

and skew-symmetric:

$$\omega(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -\omega(v_1, \dots, v_j, \dots, v_i, \dots, v_k), \quad i, \ j = 1, \dots, k.$$

The set of all exterior k-forms on E is denoted by $\Lambda^k E$. By the skew-symmetry, any exterior form of order $k > n$ is zero, thus $\Lambda^k E = \{0\}$ for $k > n$.

Exterior forms can be multiplied by real numbers, and exterior forms of the same order can be added one with another, so each $\Lambda^k E$ is a vector space. We construct a basis of $\Lambda^k E$ after we consider another operation between exterior forms — the exterior product. The exterior product of two forms $\omega_1 \in \Lambda^{k_1} E$, $\omega_2 \in \Lambda^{k_2} E$ is an exterior form $\omega_1 \wedge \omega_2$ of order $k_1 + k_2$.

Given linear 1-forms $\omega_1, \omega_2 \in \Lambda^1 E$, we have a natural (tensor) product for them:

$$\omega_1 \otimes \omega_2 \ : \ (v_1, v_2) \mapsto \omega_1(v_1)\omega_2(v_2), \qquad v_1, v_2 \in E.$$

The result is a bilinear but not a skew-symmetric form. The *exterior product* is the anti-symmetrization of the tensor one:

$$\omega_1 \wedge \omega_2 \ : \ (v_1, v_2) \mapsto \omega_1(v_1)\omega_2(v_2) - \omega_1(v_2)\omega_2(v_1), \qquad v_1, v_2 \in E.$$

Similarly, the tensor and exterior products of forms $\omega_1 \in \Lambda^{k_1} E$ and $\omega_2 \in \Lambda^{k_2} E$ are the following forms of order $k_1 + k_2$:

$$\begin{aligned}
&\omega_1 \otimes \omega_2 \ : \ (v_1, \dots, v_{k_1+k_2}) \mapsto \omega_1(v_1, \dots, v_{k_1})\omega_2(v_{k_1+1}, \dots, v_{k_1+k_2}),\\
&\omega_1 \wedge \omega_2 \ : \ (v_1, \dots, v_{k_1+k_2}) \mapsto\\
&\quad \frac{1}{k_1!\,k_2!} \sum_\sigma (-1)^{\nu(\sigma)} \omega_1(v_{\sigma(1)}, \dots, v_{\sigma(k_1)})\omega_2(v_{\sigma(k_1+1)}, \dots, v_{\sigma(k_1+k_2)}), \qquad (10.1)
\end{aligned}$$

where the sum is taken over all permutations σ of order $k_1 + k_2$ and $\nu(\sigma)$ is parity of a permutation σ. The factor $\frac{1}{k_1!\,k_2!}$ normalizes the sum in (10.1) since it contains $k_1!\,k_2!$ identically equal terms: e.g., if permutations σ do not mix the first k_1 and the last k_2 arguments, then all terms of the form

$$(-1)^{\nu(\sigma)} \omega_1(v_{\sigma(1)}, \dots, v_{\sigma(k_1)})\omega_2(v_{\sigma(k_1+1)}, \dots, v_{\sigma(k_1+k_2)})$$

are equal to

$$\omega_1(v_1, \dots, v_{k_1})\omega_2(v_{k_1+1}, \dots, v_{k_1+k_2}).$$

This guarantees the associative property of the exterior product:

$$\omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3, \qquad \omega_i \in \Lambda^{k_i} E,$$

Further, the exterior product is skew-commutative:

$$\omega_2 \wedge \omega_1 = (-1)^{k_1 k_2} \omega_1 \wedge \omega_2, \qquad \omega_i \in \Lambda^{k_i} E.$$

Let $e_1, \dots, e_n$ be a basis of the space E and $e_1^*, \dots, e_n^*$ the corresponding dual basis of E^*. If $1 \le k \le n$, then the following $\binom{n}{k}$ elements form a basis of the space $\Lambda^k E$:

$$e_{i_1}^* \wedge \dots \wedge e_{i_k}^*, \qquad 1 \le i_1 < i_2 < \dots < i_k \le n.$$

The equalities

$$\begin{aligned}
&(e_{i_1}^* \wedge \dots \wedge e_{i_k}^*)(e_{i_1}, \dots, e_{i_k}) = 1,\\
&(e_{i_1}^* \wedge \dots \wedge e_{i_k}^*)(e_{j_1}, \dots, e_{j_k}) = 0, \quad \text{if } (i_1, \dots, i_k) \ne (j_1, \dots, j_k)
\end{aligned}$$

for $1 \le i_1 < i_2 < \dots < i_k \le n$ imply that any k-form $\omega \in \Lambda^k E$ has a unique decomposition of the form

$$\omega = \sum_{1 \le i_1 < i_2 < \dots < i_k \le n} \omega_{i_1 \dots i_k} e_{i_1}^* \wedge \dots \wedge e_{i_k}^*$$

with

$$\omega_{i_1 \dots i_k} = \omega(e_{i_1}, \dots, e_{i_k}).$$

Exercise 10.1. Show that for any 1-forms $\omega_1, \dots \omega_p \in \Lambda^1 E$ and any vectors $v_1, \dots, v_p \in E$ there holds the equality

$$(\omega_1 \wedge \dots \wedge \omega_p)(v_1, \dots, v_p) = \det\left(\langle \omega_i, v_j \rangle\right)_{i,j=1}^{p}. \tag{10.2}$$

Notice that the space of n-forms of an n-dimensional space E is one-dimensional. Any nonzero n-form on E is a volume form. For example, the value of the standard volume form $e_1^* \wedge \dots \wedge e_n^*$ on an n-tuple of vectors $(v_1, \dots, v_n)$ is

$$(e_1^* \wedge \dots \wedge e_n^*)(v_1, \dots, v_n) = \det\left(\langle e_i^*, v_j \rangle\right)_{i,j=1}^{n},$$

the oriented volume of the parallelepiped generated by the vectors $v_1, \dots, v_n$.

10.2.2 Differential k-forms

A *differential k-form* on M is a mapping

$$\omega \ : \ q \mapsto \omega_q \in \Lambda^k T_q M, \qquad q \in M,$$

smooth w.r.t. $q \in M$. The set of all differential k-forms on M is denoted by $\Lambda^k M$. It is natural to consider smooth functions on M as 0-forms, so $\Lambda^0 M = C^\infty(M)$.

In local coordinates $(x_1, \dots, x_n)$ on a domain $O \subset M$, any differential k-form $\omega \in \Lambda^k M$ can be uniquely decomposed as follows:

$$\omega_x = \sum_{i_1 < \dots < i_k} a_{i_1 \dots i_k}(x) dx_{i_1} \wedge \dots \wedge dx_{i_k}, \qquad x \in O, \quad a_{i_1 \dots i_k} \in C^\infty(O). \tag{10.3}$$

Any smooth mapping

$$F \ : \ M \to N$$

induces a mapping of differential forms

$$F^* \ : \ \Lambda^k N \to \Lambda^k M$$

in the following way: given a differential k-form $\omega \in \Lambda^k N$, the k-form $F^*\omega \in \Lambda^k M$ is defined as

$$(F^*\omega)_q(v_1, \dots, v_k) = \omega_{F(q)}(F_* v_1, \dots, F_* v_k), \qquad q \in M, \ v_i \in T_q M.$$

For 0-forms, pull-back is a substitution of variables:

$$F^* a = F a, \qquad a \in C^\infty(M).$$

The pull-back F^* is linear w.r.t. forms and preserves the exterior product:

$$F^*(\omega_1 \wedge \omega_2) = F^*\omega_1 \wedge F^*\omega_2.$$

Exercise 10.2. Prove the composition law for pull-back of differential forms:

$$(F_1 \circ F_2)^* = F_1^* \circ F_2^*, \tag{10.4}$$

where $F_1 \ : \ M_1 \to M_2$ and $F_2 \ : \ M_2 \to M_3$ are smooth mappings.

Notice that in our notation (when points are written to the left of mappings as $q \circ F$) the star does not reverse the order of mappings F_1, F_2 in composition, unlike the classical notation $F(q)$.

Now we can define the integral of a k-form over an oriented k-dimensional surface. Let $\Pi \subset \mathbb{R}^k$ be a k-dimensional open oriented domain and

$$\Phi \, : \, \Pi \to \Phi(\Pi) \subset M$$

a diffeomorphism. Then the integral of a k-form $\omega \in \Lambda^k M$ over the k-dimensional oriented surface $\Phi(\Pi)$ is defined as follows:

$$\int_{\Phi(\Pi)} \omega \overset{\text{def}}{=} \int_{\Pi} \Phi^* \omega,$$

it remains only to define the integral over Π in the right-hand side. Since $\Phi^* \omega \in \Lambda^k(\mathbb{R}^k)$ is a k-form on $\mathbb{R}^k$, it is expressed via the standard volume form $dx_1 \wedge \ldots \wedge dx_k \in \Lambda^k(\mathbb{R}^k)$:

$$(\Phi^* \omega)_x = a(x)\, dx_1 \wedge \cdots \wedge dx_k, \qquad x \in \Pi.$$

We set

$$\int_{\Pi} \Phi^* \omega \overset{\text{def}}{=} \int_{\Pi} a(x)\, dx_1 \ldots dx_k,$$

a usual multiple integral.

The integral $\int_{\Phi(\Pi)} \omega$ is defined correctly with respect to orientation-preserving reparametrizations of the surface $\Phi(\Pi)$. Although, if a parametrization changes orientation, then the integral changes sign.

The notion of integral is extended to arbitrary submanifolds as follows. Let $N \subset M$ be a k-dimensional submanifold and let $\omega \in \Lambda^k M$. Consider a covering of N by coordinate neighborhoods $O_i \subset M$:

$$N = \bigcup_i (N \cap O_i).$$

Take a partition of unity subordinated to this covering:

$$\alpha_i \in C^\infty(M), \quad \operatorname{supp} \alpha_i \subset O_i, \quad 0 \leq \alpha_i \leq 1,$$
$$\sum_i \alpha_i \equiv 1.$$

Then

$$\int_N \omega \overset{\text{def}}{=} \sum_i \int_{N \cap O_i} \alpha_i \omega.$$

The integral thus defined does not depend upon choice of partition of unity.

Remark. Another possible approach to definition of integral of a differential form over a submanifold is based upon triangulation of the submanifold.

10.3 Exterior differential

Exterior differential of a function (i.e., a 0-form) is a 1-form: if $a \in C^\infty(M) = \Lambda^0 M$, then its differential

$$d_q a \in T_q^* M$$

is the functional (directional derivative)

$$\langle d_q a, v \rangle = va, \qquad v \in T_q M, \tag{10.5}$$

so

$$da \in \Lambda^1 M.$$

By the Newton-Leibniz formula, if $\gamma \subset M$ is a smooth oriented curve starting at a point $q_0 \in M$ and terminating at $q_1 \in M$, then

$$\int_\gamma da = a(q_1) - a(q_0).$$

The right-hand side can be considered as the integral of the function a over the oriented boundary of the curve: $\partial\gamma = q_1 - q_0$, thus

$$\int_\gamma da = \int_{\partial\gamma} a. \tag{10.6}$$

In the exposition above, Newton-Leibniz formula (10.6) comes as a consequence of definition (10.5) of differential of a function. But one can go the reverse way: if we postulate Newton-Leibniz formula (10.6) for any smooth curve $\gamma \subset M$ and pass to the limit $q_1 \to q_0$, we necessarily obtain definition (10.5) of differential of a function.

Such approach can be realized for higher order differential forms as well. Let $\omega \in \Lambda^k M$. We define the *exterior differential*

$$d\omega \in \Lambda^{k+1} M$$

as the differential $(k+1)$-form for which Stokes formula holds:

$$\int_N d\omega = \int_{\partial N} \omega \tag{10.7}$$

for $(k+1)$-dimensional submanifolds with boundary $N \subset M$ (for simplicity, one can take here N equal to a diffeomorphic image of a $(k+1)$-dimensional polytope). The boundary ∂N is oriented by a frame of tangent vectors $e_1, \dots e_k \in T_q(\partial N)$ in such a way that the frame $e_n, e_1, \dots, e_k \in T_q N$ define a positive orientation of N, where e_n is the outward normal vector to N at q.

The existence of a form $d\omega$ that satisfies Stokes formula (10.7) comes from the fact that the mapping $N \mapsto \int_{\partial N} \omega$ is additive w.r.t. domain: if $N = N_1 \cup N_2$, $N_1 \cap N_2 = \partial N_1 \cap \partial N_2$, then

$$\int_{\partial N} \omega = \int_{\partial N_1} \omega + \int_{\partial N_2} \omega$$

(notice that orientation of the boundaries is coordinated: ∂N_1 and ∂N_2 have mutually opposite orientations at points of their intersection). Thus the integral $\int_{\partial N} \omega$ is a kind of measure w.r.t. N, and one can recover $(d\omega)_q$ passing to limit in (10.7) as the submanifold N contracts to a point q.

We recall some basic properties of exterior differential. First of all, it is obvious from the Stokes formula that $d : \Lambda^k M \to \Lambda^{k+1} M$ is a linear operator. Further, if $F : M \to N$ is a diffeomorphism, then

$$dF^* \omega = F^* d\omega, \qquad \omega \in \Lambda^k N. \tag{10.8}$$

Indeed, if $W \subset M$, then

$$\int_{F(W)} \omega = \int_W F^* \omega, \qquad \omega \in \Lambda^k N,$$

thus

$$\begin{aligned} \int_W dF^* \omega &= \int_{\partial W} F^* \omega = \int_{F(\partial W)} \omega = \int_{\partial F(W)} \omega = \int_{F(W)} d\omega \\ &= \int_W F^* d\omega, \end{aligned}$$

and equality (10.8) follows.

Another basic property of exterior differential is given by the equality

$$d \circ d = 0,$$

which follows since $\partial(\partial N) = \emptyset$ for any submanifold with boundary $N \subset M$.

Exterior differential is an antiderivation:

$$d(\omega_1 \wedge \omega_2) = (d\omega_1) \wedge \omega_2 + (-1)^{k_1} \omega_1 \wedge d\omega_2, \qquad \omega_i \in \Lambda^{k_i} M.$$

In local coordinates exterior differential is computed as follows: if

$$\omega = \sum_{i_1 < \cdots < i_k} a_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}, \qquad a_{i_1 \ldots i_k} \in C^\infty,$$

then

$$d\omega = \sum_{i_1 < \cdots < i_k} (da_{i_1 \ldots i_k}) \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}.$$

10.4 Lie derivative of differential forms

The "infinitesimal version" of the pull-back P^* of a differential form by a flow P is given by the following operation.

Lie derivative of a differential form $\omega \in \Lambda^k M$ along a vector field $f \in \operatorname{Vec} M$ is the differential form $L_f \omega \in \Lambda^k M$ defined as follows:

$$L_f \omega \stackrel{\text{def}}{=} \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \left(e^{\varepsilon f} \right)^* \omega. \tag{10.9}$$

Since

$$\left(e^{tf}\right)^*(\omega_1 \wedge \omega_2) = \left(e^{tf}\right)^* \omega_1 \wedge \left(e^{tf}\right)^* \omega_2,$$

Lie derivative L_f is a derivation of the algebra of differential forms:

$$L_f(\omega_1 \wedge \omega_2) = (L_f\omega_1) \wedge \omega_2 + \omega_1 \wedge L_f\omega_2.$$

Further, we have

$$\left(e^{tf}\right)^* \circ d = d \circ \left(e^{tf}\right)^*,$$

thus

$$L_f \circ d = d \circ L_f.$$

For 0-forms, Lie derivative is just the directional derivative:

$$L_f a = fa, \qquad a \in C^\infty(M),$$

since

$$\left(e^{tf}\right)^* a = e^{tf}\, a$$

as substitution of variables.

Now we obtain a useful formula for the action of Lie derivative on differential forms of an arbitrary order.

Consider, along with exterior differential

$$d : \Lambda^k M \to \Lambda^{k+1} M$$

the *interior product* of a differential form ω with a vector field $f \in \operatorname{Vec} M$:

$$i_f : \Lambda^k M \to \Lambda^{k-1} M,$$
$$(i_f\omega)(v_1, \dots, v_{k-1}) \stackrel{\text{def}}{=} \omega(f, v_1, \dots, v_{k-1}), \qquad \omega \in \Lambda^k M,\ v_i \in T_q M,$$

which acts as substitution of f for the first argument of ω. By definition, for 0-order forms

$$i_f a = 0, \qquad a \in \Lambda^0 M.$$

Interior product is an antiderivation, as well as the exterior differential:

$$i_f(\omega_1 \wedge \omega_2) = (i_f\omega_1) \wedge \omega_2 + (-1)^{k_1}\omega_1 \wedge i_f\omega_2, \qquad \omega_i \in \Lambda^{k_i} M.$$

Now we prove that Lie derivative of a differential form of an arbitrary order can be computed by the following formula:

$$L_f = d \circ i_f + i_f \circ d \tag{10.10}$$

called *Cartan's formula*, for short "$L = di + id$". Notice first of all that the right-hand side in (10.10) has the required order:

$$d \circ i_f + i_f \circ d : \Lambda^k M \to \Lambda^k M.$$

Further, $d \circ i_f + i_f \circ d$ is a derivation as it is obtained from two antiderivations. Moreover, this derivation commutes with differential:

$$d \circ (d \circ i_f + i_f \circ d) = d \circ i_f \circ d,$$
$$(d \circ i_f + i_f \circ d) \circ d = d \circ i_f \circ d.$$

Now we check formula (10.10) on 0-forms: if $a \in \Lambda^0 M$, then

$$(d \circ i_f)a = 0,$$
$$(i_f \circ d)a = \langle da, f \rangle = fa = L_f a.$$

So equality (10.10) holds for 0-forms. The properties of the mappings L_f and $d \circ i_f + i_f \circ d$ established and the coordinate representation (10.3) reduce the general case of k-forms, $k > 0$, to the case of 0-forms. Formula (10.10) is proved.

The differential definition (10.9) of Lie derivative can be integrated, i.e., there holds the following equality on $\Lambda^k M$:

$$\left(\overrightarrow{\exp} \int_0^t f_\tau \, d\tau \right)^* = \overrightarrow{\exp} \int_0^t L_{f_\tau} \, d\tau, \tag{10.11}$$

in the following sense. Denote the flow

$$P_{t_0}^{t_1} = \overrightarrow{\exp} \int_{t_0}^{t_1} f_\tau \, d\tau.$$

The family of operators on differential forms

$$(P_0^t)^* \, : \, \Lambda^k M \to \Lambda^k M$$

is a unique solution of the Cauchy problem

$$\frac{d}{dt}(P_0^t)^* = (P_0^t)^* \circ L_{f_t}, \qquad (P_0^t)^*\big|_{t=0} = \mathrm{Id}, \tag{10.12}$$

compare with Cauchy problems for the flow P_0^t (2.7) and for the family of operators $\operatorname{Ad} P_0^t$ (2.17), (2.18), and this solution is denoted as

$$\overrightarrow{\exp} \int_0^t L_{f_\tau} \, d\tau \overset{\text{def}}{=} \left(\overrightarrow{\exp} \int_0^t f_\tau \, d\tau \right)^*.$$

In order to verify the ODE in (10.12), we prove first the following equality for operators on forms:

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} (P_t^{t+\varepsilon})^* \omega = L_{f_t} \omega, \qquad \omega \in \Lambda^k M. \tag{10.13}$$

This equality is straightforward for 0-order forms:

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} (P_t^{t+\varepsilon})^* a = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P_t^{t+\varepsilon} a = f_t a = L_{f_t} a, \qquad a \in C^\infty(M).$$

Further, the both operators $\frac{d}{d\varepsilon}\big|_{\varepsilon=0}(P_t^{t+\varepsilon})^*$ and L_{f_t} commute with d and satisfy the Leibnitz rule w.r.t. product of a function with a differential form. Then equality (10.13) follows for forms of arbitrary order, as in the proof of Cartan's formula.

Now we easily verify the ODE in (10.12):

$$\frac{d}{dt}(P_0^t)^* = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}(P_0^{t+\varepsilon})^* = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}(P_0^t \circ P_t^{t+\varepsilon})^*$$

by the composition rule (10.4)

$$\begin{aligned} &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}(P_0^t)^* \circ (P_t^{t+\varepsilon})^* = (P_0^t)^* \circ \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}(P_t^{t+\varepsilon})^* \\ &= (P_0^t)^* \circ L_{f_t}. \end{aligned}$$

Exercise 10.3. Prove uniqueness for Cauchy problem (10.12).

For an autonomous vector field $f \in \operatorname{Vec} M$, equality (10.11) takes the form

$$\left(e^{tf}\right)^* = e^{tL_f}.$$

Notice that the Lie derivatives of differential forms L_f and vector fields $(-\operatorname{ad} f)$ are in a certain sence dual one to another, see equality (10.14) below. That is, the function

$$\langle \omega, X \rangle \; : \; q \mapsto \langle \omega_q, X(q) \rangle, \qquad q \in M,$$

defines a pairing of $\Lambda^1(M)$ and $\operatorname{Vec}(M)$ over $C^\infty(M)$. Then the equality

$$\langle P^*\omega, X \rangle = P\langle \omega, P_* X \rangle, \qquad P \in \operatorname{Diff} M, \; X \in \operatorname{Vec} M, \; \omega \in \Lambda^1(M),$$

has an infinitesimal version of the form

$$\langle L_Y \omega, X \rangle = Y\langle \omega, X \rangle - \langle \omega, (\operatorname{ad} Y) X \rangle, \qquad X, \; Y \in \operatorname{Vec} M, \quad \omega \in \Lambda^1(M). \tag{10.14}$$

Taking into account Cartan's formula, we immediately obtain the following important equality:

$$d\omega(Y, X) = Y\langle \omega, X \rangle - X\langle \omega, Y \rangle - \langle \omega, [Y, X] \rangle, \qquad X, \; Y \in \operatorname{Vec} M, \quad \omega \in \Lambda^1(M).$$

10.5 Elements of Symplectic Geometry

We have already seen that the cotangent bundle $T^*M = \cup_{q \in M} T_q^*M$ of an n-dimensional manifold M is a $2n$-dimensional manifold. Any local coordinates $x = (x_1, \ldots, x_n)$ on M determine canonical local coordinates on T^*M of the form $(\xi, x) = (\xi_1, \ldots, \xi_n; x_1, \ldots, x_n)$ in which any covector $\lambda \in T_{q_0}^*M$ has the decomposition $\lambda = \sum_{i=1}^n \xi_i \, dx_i\big|_{q_0}$.

The *"tautological" 1-form* (or *Liouville 1-form*) on the cotangent bundle

$$s \in \Lambda^1(T^*M)$$

is defined as follows. Let $\lambda \in T^*M$ be a point in the cotangent bundle and $w \in T_\lambda(T^*M)$ a tangent vector to T^*M at λ. Denote by π the canonical projection from T^*M to M:

$$\pi \; : T^*M \to M,$$
$$\pi \;:\; \lambda \mapsto q, \qquad \lambda \in T_q^*M.$$

Differential of π is a linear mapping

$$\pi_* \;: T_\lambda(T^*M) \to T_qM, \qquad q = \pi(\lambda).$$

The tautological 1-form s at the point λ acts on the tangent vector w in the following way:

$$\langle s_\lambda, w\rangle \stackrel{\text{def}}{=} \langle \lambda, \pi_* w\rangle.$$

That is, we project the vector $w \in T_\lambda(T^*M)$ to the vector $\pi_* w \in T_qM$, and then act by the covector $\lambda \in T_q^*M$. So

$$s_\lambda \stackrel{\text{def}}{=} \lambda \circ \pi_*.$$

The title "tautological" is explained by the coordinate representation of the form s. In canonical coordinates (ξ, x) on T^*M, we have:

$$\lambda = \sum_{i=1}^n \xi_i dx_i, \tag{10.15}$$
$$w = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial \xi_i} + \beta_i \frac{\partial}{\partial x_i}.$$

The projection written in canonical coordinates

$$\pi \;:\; (\xi, x) \mapsto x$$

is a linear mapping, its differential acts as follows:

$$\pi_* \left(\frac{\partial}{\partial \xi_i}\right) = 0, \qquad i = 1, \dots, n,$$
$$\pi_* \left(\frac{\partial}{\partial x_i}\right) = \frac{\partial}{\partial x_i}, \qquad i = 1, \dots, n.$$

Thus

$$\pi_* w = \sum_{i=1}^n \beta_i \frac{\partial}{\partial x_i},$$

consequently,

$$\langle s_\lambda, w\rangle = \langle \lambda, \pi_* w\rangle = \sum_{i=1}^{n} \xi_i \beta_i.$$

But $\beta_i = \langle dx_i, w\rangle$, so the form s has in coordinates (ξ, x) exactly the same expression

$$s_\lambda = \sum_{i=1}^{n} \xi_i dx_i \tag{10.16}$$

as the covector λ, see (10.15). Although, definition of the form s does not depend on any coordinates.

Remark. In mechanics, the tautological form s is denoted as $p\,dq$.

Consider the exterior differential of the 1-form s:

$$\sigma \stackrel{\text{def}}{=} ds.$$

The differential 2-form $\sigma \in \Lambda^2(T^*M)$ is called the *canonical symplectic structure* on T^*M. In canonical coordinates, we obtain from (10.16):

$$\sigma = \sum_{i=1}^{n} d\xi_i \wedge dx_i. \tag{10.17}$$

This expression shows that the form σ is nondegenerate, i.e., the bilinear skew-symmetric form

$$\sigma_\lambda \;:\; T_\lambda(T^*M) \times T_\lambda(T^*M) \to \mathbb{R}$$

has no kernel:

$$\sigma(w, \cdot) = 0 \quad \Rightarrow \quad w = 0, \qquad\quad w \in T_\lambda(T^*M).$$

In the following basis in the tangent space $T_\lambda(T^*M)$

$$\frac{\partial}{\partial x_1}, \frac{\partial}{\partial \xi_1}, \;\ldots\;, \frac{\partial}{\partial x_n}, \frac{\partial}{\partial \xi_n},$$

the form σ_λ has the block matrix

$$\begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix}.$$

The form σ is closed:

$$d\sigma = 0$$

since it is exact: $\sigma = ds$, and $d \circ d = 0$.

Remarks. (1) A closed nondegenerate exterior differential 2-form on a $2n$-dimensional manifold is called a *symplectic structure*. A manifold with a symplectic structure is called a *symplectic manifold*. The cotangent bundle T^*M with the canonical symplectic structure σ is the most important example of a symplectic manifold.

(2) In mechanics, the 2-form σ is known as the form $dp \wedge dq$.

Due to the symplectic structure $\sigma \in \Lambda^2(T^*M)$, we can develop the Hamiltonian formalism on T^*M. A *Hamiltonian* is an arbitrary smooth function on the cotangent bundle:

$$h \in C^\infty(T^*M).$$

To any Hamiltonian h, we associate the *Hamiltonian vector field*

$$\vec{h} \in \mathrm{Vec}(T^*M)$$

by the rule:

$$\sigma_\lambda(\cdot, \vec{h}) = d_\lambda h, \qquad \lambda \in T^*M. \tag{10.18}$$

In terms of the interior product $i_v\omega(\cdot,\cdot) = \omega(v,\cdot)$, the Hamiltonian vector field is a vector field $\vec{h}$ that satisfies

$$i_{\vec{h}}\sigma = -dh.$$

Since the symplectic form σ is nondegenerate, the mapping

$$w \mapsto \sigma_\lambda(\cdot, w)$$

is a linear isomorphism

$$T_\lambda(T^*M) \to T^*_\lambda(T^*M),$$

thus the Hamiltonian vector field $\vec{h}$ in (10.18) exists and is uniquely determined by the Hamiltonian function h.

In canonical coordinates (ξ, x) on T^*M we have

$$dh = \sum_{i=1}^n \left(\frac{\partial h}{\partial \xi_i} d\xi_i + \frac{\partial h}{\partial x_i} dx_i \right),$$

then in view of (10.17)

$$\vec{h} = \sum_{i=1}^n \left(\frac{\partial h}{\partial \xi_i} \frac{\partial}{\partial x_i} - \frac{\partial h}{\partial x_i} \frac{\partial}{\partial \xi_i} \right), \tag{10.19}$$

So the *Hamiltonian system* of ODEs corresponding to h

$$\dot{\lambda} = \vec{h}(\lambda), \qquad \lambda \in T^*M,$$

reads in canonical coordinates as follows:

$$\dot{x}_i = \frac{\partial h}{\partial \xi_i}, \qquad i = 1, \dots, n,$$
$$\dot{\xi}_i = -\frac{\partial h}{\partial x_i}, \qquad i = 1, \dots, n.$$

The Hamiltonian function can depend on a parameter: h_t, $t \in \mathbb{R}$. Then the nonautonomous Hamiltonian vector field $\vec{h}_t$, $t \in \mathbb{R}$ is defined in the same way as in the autonomous case.

The flow of a Hamiltonian system preserves the symplectic form σ.

Proposition 10.1. *Let $\vec{h}_t$ ba a nonautonomous Hamiltonian vector field on T^*M. Then*

$$\left(\overrightarrow{\exp} \int_0^t \vec{h}_\tau \, d\tau\right)^* \sigma = \sigma.$$

Proof. In view of equality (10.11), we have

$$\left(\overrightarrow{\exp} \int_0^t \vec{h}_\tau \, d\tau\right)^* = \overrightarrow{\exp} \int_0^t L_{\vec{h}_\tau} \, d\tau,$$

thus the statement of this proposition can be rewritten as

$$L_{\vec{h}_t} \sigma = 0.$$

But this Lie derivative is easily computed by Cartan's formula:

$$L_{\vec{h}_t} \sigma = i_{\vec{h}_t} \circ \underbrace{d\sigma}_{=0} + d \circ \underbrace{i_{\vec{h}_t} \sigma}_{=-dh_t} = -d \circ dh_t = 0.$$

□

Moreover, there holds a local converse statement: if a flow preserves σ, then it is locally Hamiltonian. Indeed,

$$\left(\overrightarrow{\exp} \int_0^t f_\tau \, d\tau\right)^* \sigma = \sigma \quad \Leftrightarrow \quad L_{f_t} \sigma = 0,$$

further

$$L_{f_t} \sigma = i_{f_t} \circ \underbrace{d\sigma}_{=0} + d \circ i_{f_t} \sigma,$$

thus

$$L_{f_t} \sigma = 0 \quad \Leftrightarrow \quad d \circ i_{f_t} \sigma = 0.$$

If the form $i_{f_t} \sigma$ is closed, then it is locally exact (Poincaré's Lemma), i.e., there exists a Hamiltonian h_t such that locally $f_t = \vec{h}_t$.

Essentially, only Hamiltonian flows preserve σ (globally, "multi-valued Hamiltonians" can appear). If a manifold M is simply connected, then there holds a global statement: a flow on T^*M is Hamiltonian if and only if it preserves the symplectic structure.

The *Poisson bracket* of Hamiltonians $a, b \in C^\infty(T^*M)$ is a Hamiltonian

$$\{a, b\} \in C^\infty(T^*M)$$

defined in one of the following equivalent ways:

$$\{a, b\} = \vec{a} b = \langle db, \vec{a} \rangle = \sigma(\vec{a}, \vec{b}) = -\sigma(\vec{b}, \vec{a}) = -\vec{b} a.$$

It is obvious that Poisson bracket is bilinear and skew-symmetric:

$$\{a, b\} = -\{b, a\}.$$

In canonical coordinates (ξ, x) on T^*M,

$$\{a, b\} = \sum_{i=1}^{n} \left(\frac{\partial a}{\partial \xi_i} \frac{\partial b}{\partial x_i} - \frac{\partial a}{\partial x_i} \frac{\partial b}{\partial \xi_i} \right). \tag{10.20}$$

Leibniz rule for Poisson bracket easily follows from definition:

$$\{a, bc\} = \{a, b\} c + b \{a, c\}$$

(here bc is the usual pointwise product of functions b and c).

Moreover, there holds Jacobi identity for Poisson bracket.

Proposition 10.2.

$$\{a, \{b, c\}\} + \{b, \{c, a\}\} + \{c, \{a, b\}\} = 0, \qquad a, b, c \in C^\infty(T^*M). \tag{10.21}$$

Proof. It is easy to see from the coordinate representation (10.20) that each iterated bracket $\{a, \{b, c\}\}$, $\{b, \{c, a\}\}$, $\{c, \{a, b\}\}$ is a sum of products of second- and first-order derivatives of the functions a, b, c. Now we compute all terms in the left-hand side of (10.21) that contain second-order derivatives of a. The bracket $\{a, \{b, c\}\}$ contains only first-order derivatives of a, so remain the last two brackets:

$$\{b, \{c, a\}\} + \{c, \{a, b\}\} = \{b, \{c, a\}\} - \{c, \{b, a\}\} = \vec{b} \circ \vec{c}\, a - \vec{c} \circ \vec{b}\, a = [\vec{b}, \vec{c}]\, a.$$

The Lie bracket $[\vec{b}, \vec{c}]$ is a first-order differential operator, thus $[\vec{b}, \vec{c}]a$ does not contain second-order derivatives of a. So the sum in the left-hand side of (10.21) contains no second-order derivatives of a. Similarly, it contains no second-order derivatives of b and c. Jacobi identity (10.21) follows. □

So the space of all Hamiltonians $C^\infty(T^*M)$ forms a Lie algebra with Poisson bracket as a product. The correspondence

$$a \mapsto \vec{a}, \qquad a \in C^\infty(T^*M), \tag{10.22}$$

is a homomorphism from the Lie algebra of Hamiltonians to the Lie algebra of Hamiltonian vector fields on M. This follows from the next statement.

Corollary 10.1. $\overrightarrow{\{a,b\}} = [\vec{a}, \vec{b}]$ *for any Hamiltonians* $a, b \in C^\infty(T^*M)$.

Proof. Jacobi identity can be rewritten as

$$\{\{a,b\},c\} = \{a,\{b,c\}\} - \{b,\{a,c\}\},$$

i.e.,

$$\overrightarrow{\{a,b\}}\; c = \vec{a} \circ \vec{b}\, c - \vec{b} \circ \vec{a}\, c = [\vec{a}, \vec{b}]\, c, \qquad c \in C^\infty(T^*M).$$

□

Symplectomorphisms of cotangent bundle preserve the Lie algebra of Hamiltonian vector fields; the action of a symplectomorphism $P \in \mathrm{Diff}(T^*M)$ on a Hamiltonian vector field $\vec{h}$ reduces to the action of P on the Hamiltonian function as substitution of variables:

$$(\mathrm{Ad}\, P)\, \vec{h} = \overrightarrow{Ph}\,.$$

This follows from the chain

$$\begin{aligned}\sigma\left(X, (\mathrm{Ad}\, P)\vec{h}\right) &= \sigma\left(X, P_*^{-1}\vec{h}\right) = (P^*\sigma)\left(X, P_*^{-1}\vec{h}\right) = P\sigma\left(P_*X, \vec{h}\right) \\ &= P\langle dh, P_*X\rangle = X(Ph), \qquad X \in \mathrm{Vec}(T^*M).\end{aligned}$$

In particular, the action of a Hamiltonian flow on a Hamiltonian vector field gives another Hamiltonian vector field

$$\left(\overrightarrow{\exp} \int_0^t \vec{a}_\tau \, d\tau\right)_*^{-1} \vec{b}_\tau = \left(\overrightarrow{\exp} \int_0^t \mathrm{ad}\, \vec{a}_\tau \, d\tau\right) \vec{b}_\tau = \vec{c}_t \tag{10.23}$$

with the Hamiltonian function

$$c_t = \left(\overrightarrow{\exp} \int_0^t \vec{a}_\tau \, d\tau\right) b_t.$$

Corollary 10.1 can be viewed as an infinitesimal version of equality (10.23).

It is easy to see from the coordinate representation (10.19) that the kernel of the mapping $a \mapsto \vec{a}$ consists of constant functions, i.e., this is isomorphism up to constants. On the other hand, this homomorphism is far from being onto all vector fields on T^*M. Indeed, a general vector field on T^*M is locally defined by arbitrary $2n$ smooth real functions of $2n$ variables, while a Hamiltonian vector field is determined by just one real function of $2n$ variables, a Hamiltonian.

Theorem 10.1 (Nöther). *A function* $a \in C^\infty(T^*M)$ *is an integral of a Hamiltonian system of ODEs*

$$\dot{\lambda} = \vec{h}(\lambda), \qquad \lambda \in T^*M, \tag{10.24}$$

i.e.,

$$e^{t\vec{h}} a = a \qquad t \in \mathbb{R},$$

if and only if it Poisson-commutes with the Hamiltonian:

$$\{a, h\} = 0.$$

Proof. $e^{t\vec{h}}a \equiv a \Leftrightarrow 0 = \vec{h}a = \{h, a\}$. □

Corollary 10.2. *$e^{t\vec{h}}h = h$, i.e., any Hamiltonian $h \in C^\infty(T^*M)$ is an integral of the corresponding Hamiltonian system* (10.24).

Further, Jacobi identity for Poisson brackets implies that the set of integrals of the Hamiltonian system (10.24) forms a Lie algebra with respect to Poisson brackets.

Corollary 10.3. $\{h, a\} = \{h, b\} = 0 \Rightarrow \{h, \{a, b\}\} = 0$.

Remark. The Hamiltonian formalism developed generalizes for arbitrary symplectic manifolds.

Now we introduce a construction that works only on T^*M. Given a vector field $X \in \operatorname{Vec} M$, we define a Hamiltonian function

$$X^* \in C^\infty(T^*M),$$

which is linear on fibers T_q^*M, as follows:

$$X^*(\lambda) = \langle \lambda, X(q) \rangle, \qquad \lambda \in T^*M, \quad q = \pi(\lambda).$$

In canonical coordinates (ξ, x) on T^*M we have:

$$\begin{aligned} X &= \sum_{i=1}^n a_i(x) \frac{\partial}{\partial x_i}, \\ X^* &= \sum_{i=1}^n \xi_i a_i(x). \end{aligned} \tag{10.25}$$

This coordinate representation implies that

$$\{X^*, Y^*\} = [X, Y]^*, \qquad X, Y \in \operatorname{Vec} M,$$

i.e., Poisson brackets of Hamiltonians linear on fibers in T^*M contain usual Lie brackets of vector fields on M.

The Hamiltonian vector field $\overrightarrow{X^*} \in \operatorname{Vec}(T^*M)$ corresponding to the Hamiltonian function X^* is called the *Hamiltonian lift* of the vector field $X \in \operatorname{Vec} M$. It is easy to see from the coordinate representations (10.25), (10.19) that

$$\pi_*\left(\overrightarrow{X^*}\right) = X.$$

Now we pass to nonautonomous vector fields. Let X_t be a nonautonomous vector field and

$$P_{\tau,t} = \overrightarrow{\exp} \int_\tau^t X_\theta \, d\theta$$

the corresponding flow on M. The action of this flow on covectors defines the flow $P^*_{\tau,t} : T^*M \to T^*M$. Let V_t be the nonautonomous vector field on T^*M that generates the flow $P^*_{\tau,t}$:

$$V_t = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P^*_{t,t+\varepsilon}.$$

Then

$$\frac{d}{dt} P^*_{\tau,t} = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P^*_{\tau,t+\varepsilon} = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P^*_{t,t+\varepsilon} \circ P^*_{\tau,t} = V_t \circ P^*_{\tau,t},$$

so the flow $P^*_{\tau,t}$ is a solution to the Cauchy problem

$$\frac{d}{dt} P^*_{\tau,t} = V_t \circ P^*_{\tau,t}, \qquad P^*_{\tau,\tau} = \mathrm{Id},$$

i.e., it is the left chronological exponential:

$$P^*_{\tau,t} = \overleftarrow{\exp} \int_\tau^t V_\theta \, d\theta.$$

It turns out that the nonoautonomous field V_t is simply related with the Hamiltonian vector field corresponding to the Hamiltonian X^*_t:

$$V_t = -\overrightarrow{X^*_t}. \tag{10.26}$$

Indeed, the flow $P^*_{\tau,t}$ preserves the tautological form s, thus

$$L_{V_t} s = 0.$$

By Cartan's formula,

$$i_{V_t} \sigma = -d\langle s, V_t \rangle,$$

i.e., the field V_t is Hamiltonian:

$$V_t = \overrightarrow{\langle s, V_t \rangle}.$$

But $\pi_* V_t = -X_t$, consequently,

$$\langle s, V_t \rangle = -X^*_t,$$

and equality (10.26) follows. Taking into accound relation (2.14) between the left and right chronological exponentials, we obtain

$$P^*_{\tau,t} = \overleftarrow{\exp} \int_\tau^t -\overrightarrow{X^*_\theta} \, d\theta = \overrightarrow{\exp} \int_t^\tau \overrightarrow{X^*_\theta} \, d\theta.$$

We proved the following statement.

Proposition 10.3. *Let X_t be a complete nonautonomous vector field on M. Then*

$$\left(\overrightarrow{\exp} \int_\tau^t X_\theta \, d\theta \right)^* = \overrightarrow{\exp} \int_t^\tau \overrightarrow{X^*_\theta} \, d\theta.$$

Chapter 11

Pontryagin Maximum Principle

In this chapter we prove the fundamental necessary condition of optimality for optimal control problems — Pontryagin Maximum Principle (PMP). In order to obtain a coordinate-free formulation of PMP on manifolds, we apply the technique of Symplectic Geometry developed in the previous chapter. The first classical version of PMP was proved for optimal control problems in $\mathbb{R}^n$ by L. S. Pontryagin and his collaborators [12].

11.1 Geometric statement of PMP and discussion

Consider the optimal control problem stated in Sec. 9.1 for a control system

$$\dot{q} = f_u(q), \qquad q \in M, \quad u \in U \subset \mathbb{R}^m, \tag{11.1}$$

with the initial condition

$$q(0) = q_0. \tag{11.2}$$

Define the following family of Hamiltonians:

$$h_u(\lambda) = \langle \lambda, f_u(q) \rangle, \qquad \lambda \in T_q^*M,\ q \in M,\ u \in U.$$

In terms of the previous section,

$$h_u(\lambda) = f_u^*(\lambda).$$

Fix an arbitrary instant $t_1 > 0$.

In Sec. 9.2 we reduced the optimal control problem to the study of boundary of attainable sets. Now we give a necessary optimality condition in this geometric setting.

Theorem 11.1 (PMP). *Let $\tilde{u}(t)$, $t \in [0, t_1]$, be an admissible control and $\tilde{q}(t) = q_{\tilde{u}}(t)$ the corresponding solution of* (11.1), (11.2)*. If*

$$\tilde{q}(t_1) \in \partial \mathcal{A}_{q_0}(t_1),$$

then there exists a Lipschitz curve in the cotangent bundle

$$\lambda_t \in T^*_{\tilde{q}(t)} M, \qquad 0 \le t \le t_1,$$

such that

$$\lambda_t \ne 0, \tag{11.3}$$

$$\dot{\lambda}_t = \vec{h}_{\tilde{u}(t)}(\lambda_t), \tag{11.4}$$

$$h_{\tilde{u}(t)}(\lambda_t) = \max_{u \in U} h_u(\lambda_t) \tag{11.5}$$

for almost all $t \in [0, t_1]$.

If $u(t)$ is an admissible control and λ_t a Lipschitz curve in T^*M such that conditions (11.3)–(11.5) hold, then the pair $(u(t), \lambda_t)$ is said to satisfy PMP. In this case the curve λ_t is called an *extremal*, and its projection $q(t) = \pi(\lambda_t)$ is called an *extremal trajectory*.

Remark. If a pair $(\tilde{u}(t), \lambda_t)$ satisfies PMP, then

$$h_{\tilde{u}(t)}(\lambda_t) = \text{const}, \qquad t \in [0, t_1]. \tag{11.6}$$

Indeed, since the admissible control $\tilde{u}(t)$ is bounded, we can take maximum in (11.5) over the compact $\overline{\{\tilde{u}(t) \mid t \in [0, t_1]\}} = \widetilde{U}$. Further, the function

$$\varphi(\lambda) = \max_{u \in \widetilde{U}} h_u(\lambda)$$

is Lipschitzian w.r.t. $\lambda \in T^*M$. We show that this function has zero derivative. For any admissible control $u(t)$,

$$\varphi(\lambda_t) \ge h_{u(\tau)}(\lambda_t), \qquad \varphi(\lambda_\tau) = h_{u(\tau)}(\lambda_\tau),$$

thus

$$\frac{\varphi(\lambda_t) - \varphi(\lambda_\tau)}{t - \tau} \ge \frac{h_{u(\tau)}(\lambda_t) - h_{u(\tau)}(\lambda_\tau)}{t - \tau}, \qquad t > \tau.$$

Consequently,

$$\left.\frac{d}{dt}\right|_{t=\tau} \varphi(\lambda_t) \ge \{h_{u(\tau)}, h_{u(\tau)}\} = 0$$

if τ is a differentiability point of $\varphi(\lambda_t)$. Similarly,

$$\frac{\varphi(\lambda_t) - \varphi(\lambda_\tau)}{t - \tau} \le \frac{h_{u(\tau)}(\lambda_t) - h_{u(\tau)}(\lambda_\tau)}{t - \tau}, \qquad t < \tau,$$

thus

$$\left.\frac{d}{dt}\right|_{t=\tau} \varphi(\lambda_t) \le 0.$$

So

$$\frac{d}{dt}\varphi(\lambda_t) = 0,$$

and identity (11.6) follows.

The Hamiltonian system of PMP

$$\dot{\lambda}_t = \vec{h}_{u(t)}(\lambda_t) \tag{11.7}$$

is an extension of the initial control system (11.1) to the cotangent bundle. Indeed, in canonical coordinates $\lambda = (\xi, x) \in T^*M$, the Hamiltonian system yields

$$\dot{x} = \frac{\partial h_{u(t)}}{\partial \xi} = f_{u(t)}(x).$$

That is, solutions λ_t to (11.7) are Hamiltonian lifts of solutions $q(t)$ to (11.1):

$$\pi(\lambda_t) = q_u(t).$$

Before proving Pontryagin Maximum Principle, we discuss its statement.

First we give an euristic expalantion of the way the covector curve λ_t appears naturally in the study of trajectories coming to boundary of the attainable set. Indeed, let

$$q_1 = \tilde{q}(t_1) \in \partial \mathcal{A}_{q_0}(t_1).$$

Consider a local convex approximation of the attainable set $\mathcal{A}_{q_0}(t_1)$ in the neighborhood of the point q_1, which is a convex cone in $T_{q_1}M$. Then $0 \in T_{q_1}M$ must belong to the boundary of the convex approximation of $\mathcal{A}_{q_0}(t_1)$. Thus the convex approximation has a hyperplane of support at q_1 determined by a covector $\lambda_{t_1} \in T^*M$, $\lambda_{t_1} \neq 0$ (the covector λ_{t_1} is an analog of Lagrange multipliers). In order to construct the whole curve λ_t, $t \in [0, t_1]$, consider the flow generated by the control $\tilde{u}(\cdot)$:

$$P_\tau^{t_1} = \overrightarrow{\exp} \int_\tau^{t_1} f_{\tilde{u}(t)}\, dt, \qquad \tau \in [0, t_1].$$

It is easy to see that

$$P_\tau^{t_1}(\mathcal{A}_{q_0}(\tau)) \subset \mathcal{A}_{q_0}(t_1), \qquad \tau \in [0, t_1].$$

Indeed, if a point $q \in \mathcal{A}_{q_0}(\tau)$ is reachable from q_0 by a control $u(t)$, $t \in [0, \tau]$, then the point $P_\tau^{t_1}(q)$ is reachable from q_0 by the control

$$v(t) = \begin{cases} u(t), & t \in [0, \tau], \\ \tilde{u}(t), & t \in [\tau, t_1]. \end{cases}$$

Further, the flow $P_\tau^{t_1} : M \to M$ satisfies the condition

$$P_\tau^{t_1}(\tilde{q}(\tau)) = \tilde{q}(t_1) = q_1, \qquad \tau \in [0, t_1].$$

Thus if $\tilde{q}(\tau) \in \operatorname{int} \mathcal{A}_{q_0}(\tau)$, then $q_1 \in \operatorname{int} \mathcal{A}_{q_0}(t_1)$. By contradiction, we obtain

$$\tilde{q}(\tau) \in \partial \mathcal{A}_{q_0}(\tau), \qquad \tau \in [0, t_1].$$

Consequently, we can find a hyperplane of support to the convex approximation of $\mathcal{A}_\tau(q_0)$ and the corresponding covector at any instant τ:

$$\lambda_\tau \in T^*_{\tilde{q}(\tau)}, \qquad \tau \in [0, t_1].$$

The covectors λ_t are defined up to nonzero factors. They can be renormalized so that satisfy the Hamiltonian system $\dot{\lambda}_t = \vec{h}_{\tilde{u}(t)}(\lambda_t)$.

So the covector curve λ_t in Pontryagin Maximum Principle appears naturally from hyperplanes of support to convex approximations of attainable sets.

Now we show the power of PMP by the following statement.

Proposition 11.1. *Assume that the maximized Hamiltonian of PMP*

$$H(\lambda) = \max_{u \in U} h_u(\lambda), \qquad \lambda \in T^*M,$$

is defined and C^2-smooth on T^*M.

If a pair $(\tilde{u}(t), \lambda_t)$, $t \in [0, t_1]$, satisfies PMP, then

$$\dot{\lambda}_t = \vec{H}(\lambda_t), \qquad t \in [0, t_1]. \tag{11.8}$$

Conversely, if a Lipshitzian curve $\lambda_t \neq 0$ is a solution to the Hamiltonian system (11.8)*, then one can choose an admissible control $\tilde{u}(t)$, $t \in [0, t_1]$, such that the pair $(\tilde{u}(t), \lambda_t)$ satisfy PMP.*

That is, in the favorable case when the maximized Hamiltonian H is C^2-smooth, PMP reduces the problem to the study of solutions to just one Hamiltonian system (11.8). From the point of view of dimension, this reduction is the best one we can expect. Indeed, for a full-dimensional attainable set ($\dim \mathcal{A}_{q_0}(t_1) = n$) we have $\dim \partial\mathcal{A}_{q_0}(t_1) = n - 1$, i.e., we need an $(n-1)$-parameter family of curves to describe the boundary $\partial\mathcal{A}_{q_0}(t_1)$. On the other hand, the family of solutions to Hamiltonian system (11.8) with the initial condition $\pi(\lambda_0) = q_0$ is n-dimensional. Taking into account that the Hamiltonian H is homogeneous:

$$H(c\lambda) = cH(\lambda), \qquad c > 0,$$

thus

$$e^{t\vec{H}}(c\lambda_0) = ce^{t\vec{H}}(\lambda_0), \qquad \pi \circ e^{t\vec{H}}(c\lambda_0) = \pi \circ e^{t\vec{H}}(\lambda_0),$$

we obtain the required $(n-1)$-dimensional family of curves.

Now we prove Proposition 11.1.

Proof. We show that if an admissible control $\tilde{u}(t)$ satisfies the maximality condition (11.5), then

$$\vec{h}_{\tilde{u}(t)}(\lambda_t) = \vec{H}(\lambda_t), \qquad t \in [0, t_1]. \tag{11.9}$$

By definition of the maximized Hamiltonian H,

$$H(\lambda) - h_{\tilde{u}(t)}(\lambda) \geq 0 \qquad \lambda \in T^*M, \quad t \in [0, t_1].$$

On the other hand, by the maximality condition of PMP (11.5), along the extremal λ_t this inequality turns into equality:

$$H(\lambda_t) - h_{\tilde{u}(t)}(\lambda_t) = 0, \qquad t \in [0, t_1].$$

That is why

$$d_{\lambda_t} H = d_{\lambda_t} h_{\tilde{u}(t)}, \qquad t \in [0, t_1].$$

But a Hamiltonian vector field is obtained from differential of the Hamiltonian by a standard linear transformation, thus equality (11.9) follows.

Conversely, let $\lambda_t \neq 0$ be a trajectory of the Hamiltonian system $\dot{\lambda}_t = \vec{H}(\lambda_t)$. In the same way as in the proof of Filippov's theorem, one can choose an admissible control $\tilde{u}(t)$ that realizes maximum along λ_t:

$$H(\lambda_t) = h_{\tilde{u}(t)}(\lambda_t) = \max_{u \in U} h_u(\lambda_t).$$

As we have shown above, then there holds equality (11.9). So the pair $(\tilde{u}(t), \lambda_t)$ satisfies PMP. □

11.2 Proof of PMP

We start from two auxiliary propositions.

Denote the positive orthant in $\mathbb{R}^m$ as

$$\mathbb{R}^m_+ = \{(x_1, \dots, x_m) \in \mathbb{R}^m \mid x_i \geq 0, \ i = 1, \dots, m\}.$$

Lemma 11.1. *Let a vector-function $F : \mathbb{R}^m \to \mathbb{R}^n$ be Lipshitzian, $F(0) = 0$, and differentiable at 0:*

$$\exists \ F'_0 = \left.\frac{dF}{dx}\right|_0 .$$

Assume that

$$F'_0(\mathbb{R}^m_+) = \mathbb{R}^n.$$

Then for any neighborhood of the origin $O_0 \subset \mathbb{R}^m$

$$0 \in \operatorname{int} F(O_0 \cap \mathbb{R}^m_+).$$

Remark. The statement of the previous lemma holds if the orthant $\mathbb{R}^m_+$ is replaced by an arbitrary convex cone $C \subset \mathbb{R}^m$. In this case the proof given below works without any changes.

Proof. Choose points $y_0, \dots, y_n \in \mathbb{R}^n$ that generate an n-dimensional simplex centered at the origin:

$$\frac{1}{n+1} \sum_{i=0}^{n} y_i = 0.$$

Since the mapping $F_0' : \mathbb{R}^m_+ \to \mathbb{R}^n$ is surjective and the positive orthant $\mathbb{R}^m_+$ is convex, it is easy to show that restriction to the interior $F_0'|_{\operatorname{int} \mathbb{R}^m_+}$ is also surjective:

$$\exists\ v_i \in \operatorname{int} \mathbb{R}^m_+ \quad \text{such that} \quad F_0' v_i = y_i, \quad i = 0, \dots, n.$$

The points $y_0, \dots, y_n$ are affinely independent in $\mathbb{R}^n$, thus their preimages $v_0, \dots, v_n$ are also affinely independent in $\mathbb{R}^m$. The mean

$$v = \frac{1}{n+1} \sum_{i=0}^{n} v_i$$

belongs to $\operatorname{int} \mathbb{R}^m_+$ and satisfies the equality

$$F_0' v = 0.$$

Further, the subspace

$$W = \operatorname{span}\{v_i - v \mid i = 0, \dots, n\} \subset \mathbb{R}^m$$

is n-dimensional. Since $v \in \operatorname{int} \mathbb{R}^m_+$, we can find an n-dimensional ball $B_\delta \subset W$ of a sufficiently small radius δ centered at the origin such that

$$v + B_\delta \subset \operatorname{int} \mathbb{R}^m_+.$$

Since $F_0'(v_i - v) = F_0' v_i$, then $F_0' W = \mathbb{R}^n$, i.e., the linear mapping $F_0' : W \to \mathbb{R}^n$ is invertible.

Consider the following family of mappings:

$$\begin{aligned} &G_\alpha : B_\delta \to \mathbb{R}^n, \qquad \alpha \in [0, \alpha_0), \\ &G_\alpha(w) = \frac{1}{\alpha} F(\alpha(v + w)), \qquad \alpha > 0, \\ &G_0(w) = F_0' w. \end{aligned}$$

By the hypotheses of the proposition,

$$F(x) = F_0' x + o(x), \qquad x \in \mathbb{R}^m, \ x \to 0,$$

thus

$$G_\alpha(w) = F_0' w + o(1), \qquad \alpha \to 0, \quad w \in B_\delta. \tag{11.10}$$

Since the mapping F is Lipschitz, all mappings G_α are Lipschitz with a common constant. Thus the family G_α is equicontinuous. Equality (11.10) means that

$$G_\alpha \to G_0, \qquad \alpha \to 0,$$

pointwise, thus uniformly.

But $0 \in \operatorname{int} G_0 B_\delta$, and it is easy to show that $0 \in \operatorname{int} G_\alpha B_\delta$, $\alpha > 0$. Indeed, the mapping

$$w \mapsto \frac{G_\alpha(w)}{\|G_\alpha(w)\|}, \qquad w \in \partial B_\delta \tag{11.11}$$

from the $n-1$-dimensional sphere ∂B_δ to the unit sphere S^{n-1} has degree ± 1 for $\alpha = 0$, thus for small $\alpha > 0$. Then $0 \in \operatorname{int} G_\alpha B_\delta$, $\alpha > 0$. Consequently, $0 \in \operatorname{int} F(\alpha B_\delta)$ for small $\alpha > 0$. □

Now we start to compute a convex approximation of the attainable set $\mathcal{A}_{q_0}(t_1)$ at the point $q_1 = \tilde{q}(t_1)$. Take any admissible control $u(t)$ and express the endpoint of a trajectory via Variations Formula (2.24):

$$\begin{aligned}
q_u(t_1) &= q_0 \circ \overrightarrow{\exp} \int_0^{t_1} f_{u(\tau)}\, d\tau = q_0 \circ \overrightarrow{\exp} \int_0^{t_1} f_{\tilde{u}(\tau)} + (f_{u(\tau)} - f_{\tilde{u}(\tau)})\, d\tau \\
&= q_0 \circ \overrightarrow{\exp} \int_0^{t_1} f_{\tilde{u}(\tau)}\, d\tau \circ \overrightarrow{\exp} \int_0^{t_1} \left(P_\tau^{t_1}\right)_* (f_{u(\tau)} - f_{\tilde{u}(\tau)})\, d\tau \\
&= q_1 \circ \overrightarrow{\exp} \int_0^{t_1} \left(P_\tau^{t_1}\right)_* (f_{u(\tau)} - f_{\tilde{u}(\tau)})\, d\tau.
\end{aligned}$$

Introduce the following vector field depending on two parameters:

$$g_{\tau,u} = \left(P_\tau^{t_1}\right)_* (f_u - f_{\tilde{u}(\tau)}), \qquad \tau \in [0, t_1], \quad u \in U. \tag{11.12}$$

We showed that

$$q_u(t_1) = q_1 \circ \overrightarrow{\exp} \int_0^{t_1} g_{\tau,u(\tau)}\, d\tau. \tag{11.13}$$

Notice that

$$g_{\tau,\tilde{u}(\tau)} \equiv 0, \qquad \tau \in [0, t_1].$$

Lemma 11.2. *Let $\mathcal{T} \subset [0, t_1]$ be the set of Lebesgue points of the control $\tilde{u}(\cdot)$. If*

$$T_{q_1} M = \operatorname{cone}\{g_{\tau,u}(q_1) \mid \tau \in \mathcal{T},\ u \in U\},$$

then

$$q_1 \in \operatorname{int} \mathcal{A}_{q_0}(t_1).$$

Remark. The set $\operatorname{cone}\{g_{\tau,u}(q_1) \mid \tau \in \mathcal{T},\ u \in U\} \subset T_{q_1} M$ is a local convex approximation of the attainable set $\mathcal{A}_{q_0}(t_1)$ at the point q_1.

Recall that a point $\tau \in [0, t_1]$ is called a *Lebesgue point* of a function $u \in L^1[0, t_1]$ if

$$\lim_{t \to \tau} \frac{1}{|t-\tau|} \int_\tau^t |u(\theta) - u(\tau)| \, d\theta = 0.$$

At Lebesgue points of u, the integral $\int_0^t u(\theta)\, d\theta$ is differentiable and

$$\frac{d}{dt}\left(\int_0^t u(\theta)\, d\theta\right) = u(t).$$

The set of Lebesgue points has the full measure in the domain $[0, t_1]$. For details on this subject, see e.g. [15].

Now we prove Lemma 11.2.

Proof. We can choose vectors

$$g_{\tau_i, u_i}(q_1) \in T_{q_1} M, \qquad \tau_i \in \mathcal{T}, \quad u_i \in U, \quad i = 1, \dots, k,$$

that generate the whole tangent space as a positive convex cone:

$$\operatorname{cone}\{g_{\tau_i, u_i}(q_1) \mid i = 1, \dots, k\} = T_{q_1} M,$$

moreover, we can choose points τ_i distinct: $\tau_i \neq \tau_j$, $i \neq j$. Indeed, if $\tau_i = \tau_j$ for some $i \neq j$, we can find a sufficiently close Lebesgue point $\tau_j' \neq \tau_j$ such that the difference $g_{\tau_j', u_j}(q_1) - g_{\tau_j, u_j}(q_1)$ is as small as we wish. This is possible since for any $\tau \in \mathcal{T}$ and any $\varepsilon > 0$

$$\frac{1}{|t-\tau|} \operatorname{meas}\{t' \in [\tau, t] \mid |u(t') - u(\tau)| \leq \varepsilon\} \to 1 \text{ as } t \to \tau.$$

We suppose that $\tau_1 < \tau_2 < \cdots < \tau_k$.

We define a family of variations of controls that follow the control $\tilde{u}(\cdot)$ everywhere except neighborhoods of τ_i, and follow u_i near τ_i (such variations are called *needle-like*). More precisely, for any $s = (s_1, \dots, s_k) \in \mathbb{R}_+^k$ consider a control of the form

$$u_s(t) = \begin{cases} u_i, & t \in [\tau_i, \tau_i + s_i], \\ \tilde{u}(t), & t \notin \cup_{i=1}^k [\tau_i, \tau_i + s_i]. \end{cases} \tag{11.14}$$

For small s, the segments $[\tau_i, \tau_i + s_i]$ do not overlap since $\tau_i \neq \tau_j$, $i \neq j$. In view of formula (11.13), the endpoint of the trajectory corresponding to the control constructed is expressed as follows:

$$\begin{aligned} q_{u_s}(t_1) &= q_0 \circ \overrightarrow{\exp} \int_0^{t_1} f_{u_s(t)}\, dt \\ &= q_1 \circ \overrightarrow{\exp} \int_{\tau_1}^{\tau_1 + s_1} g_{t, u_1}\, dt \circ \overrightarrow{\exp} \int_{\tau_2}^{\tau_2 + s_2} g_{t, u_2}\, dt \circ \cdots \\ &\qquad \circ \overrightarrow{\exp} \int_{\tau_k}^{\tau_k + s_k} g_{t, u_k}\, dt. \end{aligned}$$

The mapping

$$F \ : \ s = (s_1, \ldots, s_k) \mapsto q_{u_s}(t_1)$$

is Lipschitz, differentiable at $s = 0$, and

$$\left.\frac{\partial F}{\partial s_i}\right|_{s=0} = g_{\tau_i, u_i}(q_1).$$

By Lemma 11.1,

$$F(0) = q_1 \in \operatorname{int} F(O_0 \cap \mathbb{R}^k_+)$$

for any neighborhood $O_0 \subset \mathbb{R}^k$. But the curve $q_{u_s}(t)$, $t \in [0, t_1]$, is an admissible trajectory for small $s \in \mathbb{R}^k_+$, thus $F(O_0 \cap \mathbb{R}^k_+) \subset \mathcal{A}_{q_0}(t_1)$ and $q_1 \in \operatorname{int} \mathcal{A}_{q_0}(t_1)$. □

Now we can prove the geometric statement of Pontryagin Maximum Principle, Theorem 11.1.

Proof. Let the endpoint of the reference trajectory

$$q_1 = \tilde{q}(t_1) \in \partial \mathcal{A}_{q_0}(t_1).$$

By Lemma 11.2, the origin $0 \in T_{q_1}M$ belongs to the boundary of the convex set $\operatorname{cone}\{g_{t,u}(q_1) \mid t \in \mathcal{T}, \ u \in U\}$, so this set has a hyperplane of support at the origin:

$$\exists \ \lambda_{t_1} \in T^*_{q_1}M, \quad \lambda_{t_1} \neq 0,$$

such that

$$\langle \lambda_{t_1}, g_{t,u}(q_1) \rangle \leq 0 \qquad \forall \text{ a.e. } t \in [0, t_1], \quad u \in U.$$

Taking into account definition (11.12) of the field $g_{t,u}$, we rewrite this inequality as follows:

$$\langle \lambda_{t_1}, \left(P^{t_1}_{t*} f_u\right)(q_1) \rangle \leq \langle \lambda_{t_1}, \left(P^{t_1}_{t*} f_{\tilde{u}(t)}\right)(q_1) \rangle,$$

i.e.,

$$\langle \left(P^{t_1}_t\right)^* \lambda_{t_1}, f_u(\tilde{q}(t)) \rangle \leq \langle \left(P^{t_1}_t\right)^* \lambda_{t_1}, f_{\tilde{u}(t)}(\tilde{q}(t)) \rangle.$$

The action of the flow $P^{t_1}_t$ on covectors defines the curve in the cotangent bundle:

$$\lambda_t \stackrel{\text{def}}{=} \left(P^{t_1}_t\right)^* \lambda_{t_1} \in T^*_{\tilde{q}(t)}M, \qquad t \in [0, t_1].$$

In terms of this covector curve, the inequality above reads

$$\langle \lambda_t, f_u(\tilde{q}(t)) \rangle \leq \langle \lambda_t, f_{\tilde{u}(t)}(\tilde{q}(t)) \rangle$$

Thus the maximality condition of PMP (11.5) holds along the reference trajectory:

$$h_u(\lambda_t) \leq h_{\tilde{u}(t)}(\lambda_t) \qquad \forall \ u \in U \quad \forall \text{ a.e. } t \in [0, t_1].$$

By Proposition 10.3, the curve λ_t is a trajectory of the nonautonomous Hamiltonian flow with the Hamiltonian function $f^*_{\tilde{u}(t)} = h_{\tilde{u}(t)}$:

$$\lambda_t = \lambda_{t_1} \circ \left(\overrightarrow{\exp} \int_t^{t_1} f_{\tilde{u}(\theta)} \, d\theta \right)^* = \lambda_{t_1} \circ \overrightarrow{\exp} \int_{t_1}^{t} \vec{h}_{\tilde{u}(\theta)} \, d\theta,$$

thus it satisfies the Hamiltonian equation of PMP (11.4)

$$\dot{\lambda}_t = \vec{h}_{\tilde{u}(t)}(\lambda_t).$$

□

11.3 Geometric statement of PMP for free time

In the previous section we proved Pontryagin Maximum Principle for the case of fixed terminal time t_1. Now we consider the case of free t_1.

Theorem 11.2. *Let $\tilde{u}(\cdot)$ be an admissible control for control system* (11.1) *such that*

$$\tilde{q}(t_1) \in \partial \left(\cup_{|t-t_1|<\varepsilon} \mathcal{A}_{q_0}(t) \right)$$

for some $t_1 > 0$ and $\varepsilon \in (0, t_1)$. Then there exists a Lipschitz curve

$$\lambda_t \in T^*_{\tilde{q}(t)} M, \quad \lambda_t \neq 0, \qquad 0 \leq t \leq t_1,$$

such that

$$\begin{aligned} &\dot{\lambda}_t = \vec{h}_{\tilde{u}(t)}(\lambda_t), \\ &h_{\tilde{u}(t)}(\lambda_t) = \max_{u \in U} h_u(\lambda_t), \\ &h_{\tilde{u}(t)}(\lambda_t) = 0 \end{aligned} \tag{11.15}$$

for almost all $t \in [0, t_1]$.

Remark. In problems with free time, there appears one more variable, the terminal time t_1. In order to eliminate it, we have one additional condition — equality (11.15). This condition is indeed scalar since the previous two equalities imply that $h_{\tilde{u}(t)}(\lambda_t) = \text{const}$, see remark after formulation of Theorem 11.1.

Proof. We reduce the case of free time to the case of fixed time by extension of the control system via substitution of time. Admissible trajectories of the extended system are reparametrized admissible trajectories of the initial system (the positive direction of time on trajectories is preserved).

Let a new time be a smooth function

$$\varphi : \mathbb{R} \to \mathbb{R}, \quad \dot{\varphi} > 0.$$

We find an ODE for a reparametrized trajectory:

$$\frac{d}{dt} q_u(\varphi(t)) = \dot{\varphi}(t) f_{u(\varphi(t))}(q_u(\varphi(t))),$$

so the required equation is

$$\dot{q} = \dot{\varphi}(t) f_{u(\varphi(t))}(q).$$

Now consider along with the initial control system

$$\dot{q} = f_u(q), \qquad u \in U,$$

an extended system of the form

$$\dot{q} = v f_u(q), \qquad u \in U, \ |v - 1| < \delta, \tag{11.16}$$

where $\delta = \varepsilon / t_1 \in (0, 1)$. Admissible controls of the new system are

$$w(t) = (v(t), u(t)),$$

and the reference control corresponding to the control $\tilde{u}(\cdot)$ of the initial system is

$$\tilde{w}(t) = (1, \tilde{u}(t)).$$

It is easy to see that since $\tilde{q}(t_1) \in \partial \left(\cup_{|t - t_1| < \varepsilon} \mathcal{A}_{q_0}(t) \right)$, then the trajectory of the new system through the point q_0 corresponding to the control $\tilde{w}(\cdot)$ comes at the moment t_1 to the boundary of the attainable set of the new system for time t_1. Thus $\tilde{w}(t)$ satisfies PMP with fixed time. We apply Theorem 11.1 to the new system (11.16). The Hamiltonian for the new system is $v h_u(\lambda)$. Then the maximality condition (11.5) reads

$$1 \cdot h_{\tilde{u}(t)}(\lambda_t) = \max_{u \in U, \ |v-1| < \delta} v h_u(\lambda_t).$$

We take $u = \tilde{u}(t)$ under the maximum and obtain

$$h_{\tilde{u}(t)}(\lambda_t) = 0,$$

then we restrict the maximum to the set $v = 1$ and come to

$$h_{\tilde{u}(t)}(\lambda_t) = \max_{u \in U} h_u(\lambda_t).$$

The Hamiltonian systems along $\tilde{w}(\cdot)$ and $\tilde{u}(\cdot)$ coincide one with another, thus the proposition follows. □

11.4 PMP for optimal control problems

Now we apply PMP in geometric form to optimal control problems, starting from problems with fixed time.

For a control system

$$\dot{q} = f_u(q), \qquad q \in M, \quad u \in U, \tag{11.17}$$

with the boundary conditions

$$q(0) = q_0, \quad q(t_1) = q_1, \quad q_0, q_1 \in M \text{ fixed}, \tag{11.18}$$

$$t_1 > 0 \text{ fixed}, \tag{11.19}$$

and the cost functional

$$J(u) = \int_0^{t_1} \varphi(q_u(t), u(t))\, dt \tag{11.20}$$

we consider the optimal control problem

$$J(u) \to \min. \tag{11.21}$$

We transform the problem as in Sec. 9.2. We extend the state space:

$$\widehat{q} = \begin{pmatrix} y \\ q \end{pmatrix} \in \mathbb{R} \times M,$$

define the extended vector field $\widehat{f}_u \in \mathrm{Vec}(\mathbb{R} \times M)$:

$$\widehat{f}_u(q) = \begin{pmatrix} \varphi(q,u) \\ f_u(q) \end{pmatrix},$$

and come to the new control system:

$$\frac{d\widehat{q}}{dt} = \widehat{f}_u(q) \quad \Leftrightarrow \quad \begin{cases} \dot{y} = \varphi(q,u), \\ \dot{q} = f_u(q) \end{cases} \tag{11.22}$$

with the boundary conditions

$$\widehat{q}(0) = \widehat{q}_0 = \begin{pmatrix} 0 \\ q_0 \end{pmatrix}, \qquad \widehat{q}(t_1) = \begin{pmatrix} J(u) \\ q_1 \end{pmatrix}.$$

If a control $\tilde{u}(\cdot)$ is optimal for problem (11.17)–(11.21), then the trajectory $\widehat{q}_{\tilde{u}}(t)$ of the extended system (11.22) starting from $\widehat{q}_0$ satisfies the condition

$$\widehat{q}_{\tilde{u}}(t_1) \in \partial \widehat{\mathcal{A}}_{\widehat{q}_0}(t_1),$$

where $\widehat{\mathcal{A}}_{\widehat{q}_0}(t_1)$ is the attainable set of system (11.22) from the point $\widehat{q}_0$ for time t_1. So we can apply Theorem 11.1.

But the geometric form of PMP applied to the extended system (11.22) does not distinguish minimum and maximum of the cost $J(u)$. In order to have conditions valid only for minimum, we introduce a new control parameter v and consider a new system of the form

$$\begin{cases} \dot y = \varphi(q,u) + v, \\ \dot q = f_u(q), \end{cases} \qquad v \geq 0, \quad u \in U. \tag{11.23}$$

Now the trajectory of system (11.23) corresponding to the controls $\tilde v(t) \equiv 0$, $\tilde u(t)$, comes to the boundary of the attainable set of this system at time t_1. We apply Theorem 11.1 to system (11.23). We have

$$T_{(y,q)}(\mathbb{R} \times M) = \mathbb{R} \oplus T_q M,$$
$$T^*_{(y,q)}(\mathbb{R} \times M) = \mathbb{R} \oplus T^*_q M = \{(\nu, \lambda)\}.$$

The Hamiltonian function for system (11.23) has the form

$$\widehat{h}_{(v,u)}(\nu, \lambda) = \langle \lambda, f_u \rangle + \nu(\varphi + v),$$

and the Hamiltonian system of PMP is

$$\begin{cases} \dot\nu = \frac{\partial \widehat{h}}{\partial y} = 0, \\ \dot y = \varphi(q,u), \\ \dot\lambda = \vec{h}_{\tilde u(t)}(\nu, \lambda_t). \end{cases} \tag{11.24}$$

Here $\vec{h}_u(\nu, \lambda)$ is the Hamiltonian vector field with the Hamiltonian function

$$h_u(\nu, \lambda) = \langle \lambda, f_u \rangle + \nu\varphi.$$

The first of equations (11.24) means that

$$\nu = \text{const}$$

along the reference trajectory.

The maximality condition has the form

$$\langle \lambda_t, f_{\tilde u(t)} \rangle + \nu\varphi(\tilde q(t), \tilde u(t)) = \max_{u \in U,\ v \geq 0} \left(\langle \lambda_t, f_u \rangle + \nu\varphi(\tilde q(t), u) + \nu v \right).$$

Since the previous maximum is attained, we have

$$\nu \leq 0,$$

thus $v = 0$ and

$$\langle \lambda_t, f_{\tilde u(t)} \rangle + \nu\varphi(\tilde q(t), \tilde u(t)) = \max_{u \in U} \left(\langle \lambda_t, f_u \rangle + \nu\varphi(\tilde q(t), u) \right).$$

So we proved the following statement.

Theorem 11.3. *Let* $\tilde{u}(t)$, $t \in [0, t_1]$, *be an optimal control for problem* (11.17)–(11.21)*:*

$$J(\tilde{u}) = \min\{J(u) \mid q_u(t_1) = q_1\}.$$

Define a Hamiltonian function

$$h_u^\nu(\lambda) = \langle \lambda, f_u \rangle + \nu\varphi(q, u), \qquad \lambda \in T_q^* M, \quad u \in U, \quad \nu \in \mathbb{R}.$$

Then there exists a nontrivial pair:

$$(\nu, \lambda_t) \neq 0, \qquad \nu \in \mathbb{R}, \quad \lambda_t \in T_{\tilde{q}(t)}^* M,$$

such that the following conditions hold:

$$\begin{aligned}
&\dot{\lambda}_t = \vec{h}_{\tilde{u}(t)}^\nu(\lambda_t), \\
&h_{\tilde{u}(t)}^\nu(\lambda_t) = \max_{u \in U} h_u^\nu(\lambda_t) \quad \forall \text{ a.e. } t \in [0, t_1], \\
&\nu \leq 0.
\end{aligned}$$

Remarks. (1) If we have a maximization problem instead of minimization problem (11.21), then the preceding inequality for ν should be reversed:

$$\nu \geq 0.$$

(2) For the problem with free time t_1: (11.17), (11.18), (11.20), (11.21), necessary optimality conditions of PMP are the same as in Theorem 11.3 plus one additional scalar equality $h_{\tilde{u}(t)}^\nu(\lambda_t) \equiv 0$.

There are two distinct possibilities for the constant parameter ν in Theorem 11.3:

(a) if $\nu \neq 0$, then the curve λ_t is called a *normal extremal*. Since the pair (ν, λ_t) can be multiplied by any positive number, we can normalize $\nu < 0$ and assume that $\nu = -1$ in the normal case;

(b) if $\nu = 0$, then λ_t is an *abnormal extremal*.

So we can always assume that $\nu = -1$ or 0.

Now consider the time-optimal problem:

$$\begin{aligned}
&\dot{q} = f_u(q), \qquad q \in M, \quad u \in U, \\
&q(0) = q_0, \quad q(t_1) = q_1, \quad q_0, q_1 \text{ fixed}, \\
&t_1 = \int_0^{t_1} 1\, dt \to \min.
\end{aligned}$$

For the time-optimal problem, Pontryagin Maximum Principle takes the following form.

Corollary 11.1. *Let an admissible control* $\tilde{u}(t)$, $t \in [0, t_1]$, *be time-optimal. Define a Hamiltonian function*

$$h_u(\lambda) = \langle \lambda, f_u \rangle, \qquad \lambda \in T_q^* M, \quad u \in U.$$

Then there exists a Lipschitz curve

$$\lambda_t \in T^*M, \quad \lambda_t \neq 0, \qquad t \in [0, t_1],$$

such that the following conditions hold for almost all $t \in [0, t_1]$*:*

$$\begin{aligned} \dot{\lambda}_t &= \vec{h}_{\tilde{u}(t)}(\lambda_t), \\ h_{\tilde{u}(t)}(\lambda_t) &= \max_{u \in U} h_u(\lambda_t), \\ h_{\tilde{u}(t)}(\lambda_t) &\geq 0. \end{aligned} \tag{11.25}$$

Proof. Apply Theorem 11.3 and the second remark after it, taking $\varphi \equiv 1$. Then the Hamiltonian system and the maximality condition follow. Inequality (11.25) is equivalent to conditions $h_{\tilde{u}(t)}(\lambda_t) + \nu = 0$ and $\nu \leq 0$.

The inequality $\lambda_t \neq 0$ is obtained as follows: if $\lambda_t = 0$, then $h_{\tilde{u}(t)}(\lambda_t) = 0$, thus $\nu = 0$. But the pair (ν, λ_t) must be nontrivial, consequently, $\lambda_t \neq 0$. □

In all previous problems, boundary conditions for a trajectory $q(t)$ were of the form $q(0) = q_0$, $q(t_1) = q_1$. Consider more general boundary conditions:

$$q(0) \in N_0, \qquad q_u(t_1) \in N_1,$$

where $N_0, N_1 \subset M$ are smooth submanifolds. It is easy to see that optimal solutions in the new problem are optimal for the problem with fixed $q(0)$, $q(t_1)$ as well. So all conditions of Pontryagin Maximum Principle should be satisfied. In addition to them, we need $(\dim N_1 + \dim N_2)$ extra conditions for the initial and terminal points. They are called *transversality conditions*: the adjoint covector λ_t must be orthogonal to the submanifolds N_0 and N_1 at the moments of time t_0 and t_1 respectively:

$$\begin{aligned} \lambda_0 \perp T_{q_0} N_0 \quad &\Leftrightarrow \quad \langle \lambda_0, T_{q_0} N_0 \rangle = 0, \\ \lambda_{t_1} \perp T_{q_1} N_1 \quad &\Leftrightarrow \quad \langle \lambda_{t_1}, T_{q_1} N_1 \rangle = 0. \end{aligned}$$

We leave this statement without proof.

Chapter 12

Examples of optimal control problems

In this chapter we apply Pontryagin Maximum Principle to solve concrete optimal control problems.

12.1 The fastest stop of a train at a station

Consider a train moving on a railway. The problem is to drive the train to a station and stop it there in a minimal time.

Describe position of the train by a coordinate x_1 on the real line; the origin $0 \in \mathbb{R}$ corresponds to the station. Assume that the train moves without friction, and we can control acceleration of the train by applying a force bounded by absolute value. Using rescaling if necessary, we can assume that absolute value of acceleration is bounded by 1.

We obtain the control system

$$\ddot{x}_1 = u, \qquad x_1 \in \mathbb{R}, \quad |u| \leq 1,$$

or, in the standard form,

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u, \end{cases} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2, \quad |u| \leq 1.$$

The time-optimal control problem is

$$\begin{aligned} &x(0) = x^0, \quad x(t_1) = 0, \\ &t_1 \to \min. \end{aligned}$$

First we verify existence of optimal controls by Filippov's theorem. The set of control parameters $U = [-1, 1]$ is compact, the vector fields in the right-hand side

$$f(x,u) = \begin{pmatrix} x_2 \\ u \end{pmatrix}, \qquad |u| \leq 1,$$

are linear, and the set of admissible velocities at a point

$$f(x,U) = \{f(x,u) \mid |u| \leq 1\}$$

is convex. By Corollary 9.2, the time-optimal control problem has a solution if the origin $0 \in \mathbb{R}^2$ is attainable from the initial point x^0. We will show that any point $x \in \mathbb{R}^2$ can be connected with the origin by an extremal curve.

Now we apply Pontryagin Maximum Principle. Introduce canonical coordinates on the cotangent bundle:

$$M = \mathbb{R}^2,$$
$$T^*M = T^*\mathbb{R}^2 = \mathbb{R}^{2*} \times \mathbb{R}^2 = \left\{ \lambda = (\xi, x) \mid x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \xi = (\xi_1, \xi_2) \right\}.$$

The control-dependent Hamiltonian function of PMP is

$$h_u(\xi, x) = (\xi_1, \xi_2) \begin{pmatrix} x_2 \\ u \end{pmatrix} = \xi_1 x_2 + \xi_2 u,$$

and the corresponding Hamiltonian system has the form

$$\begin{cases} \dot{x} = \dfrac{\partial h_u}{\partial \xi}, \\ \dot{\xi} = -\dfrac{\partial h_u}{\partial x}. \end{cases}$$

In coordinates this system splits into two independent subsystems:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u, \end{cases} \qquad \begin{cases} \dot{\xi}_1 = 0, \\ \dot{\xi}_2 = -\xi_1. \end{cases} \tag{12.1}$$

By PMP, if a control $\tilde{u}(\cdot)$ is time-optimal, then the Hamiltonian system has a nontrivial solution $(\xi(t), x(t))$, $\xi(t) \not\equiv 0$, such that

$$h_{\tilde{u}(t)}(\xi(t), x(t)) = \max_{|u| \leq 1} h_u(\xi(t), x(t)) \geq 0.$$

From this maximality condition, if $\xi_2(t) \neq 0$, then $\tilde{u}(t) = \operatorname{sgn} \xi_2(t)$. Notice that the maximized Hamiltonian

$$\max_{|u| \leq 1} h_u(\xi, x) = \xi_1 x_2 + |\xi_2|$$

is not smooth. So we cannot apply Proposition 11.1, but we can obtain description of optimal controls directly from Pontryagin Maximum Principle, without preliminary maximization of Hamiltonian.

Since

$$\ddot{\xi}_2 = 0,$$

then ξ_2 is linear:

$$\xi_2(t) = \alpha + \beta t, \qquad \alpha,\ \beta = \text{const},$$

hence the optimal control has the form

$$\tilde{u}(t) = \operatorname{sgn}(\alpha + \beta t).$$

So $\tilde{u}(t)$ is piecewise constant, takes only the extremal values ± 1, and has not more than one switching (discontinuity point).

New we find all trajectories $x(t)$ that correspond to such controls and come to the origin. For controls $u = \pm 1$, the first of subsystems (12.1) reads

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \pm 1. \end{cases}$$

Trajectories of this system satisfy the equation

$$\frac{d\,x_1}{d\,x_2} = \pm x_2,$$

thus are parabolas of the form

$$x_1 = \pm \frac{x_2^2}{2} + C, \qquad C = \text{const}\,.$$

First we find trajectories from this family that come to the origin without switchings: these are two semiparabolas

$$x_1 = \frac{x_2^2}{2}, \qquad x_2 < 0, \quad \dot{x}_2 > 0, \tag{12.2}$$

and

$$x_1 = -\frac{x_2^2}{2}, \qquad x_2 > 0, \quad \dot{x}_2 < 0, \tag{12.3}$$

for $u = +1$ and -1 respectively.

Now we find all extremal trajectories with one switching. Let $(x_{1s}, x_{2s}) \in \mathbb{R}^2$ be a switching point for anyone of curves (12.2), (12.3). Then extremal trajectories with one switching coming to the origin have the form

$$x_1 = \begin{cases} -x_2^2/2 + x_{2s}^2/2 + x_{1s}, & x_2 > x_{2s}, \quad \dot{x}_2 < 0, \\ x_2^2/2 & 0 > x_2 > x_{2s}, \quad \dot{x}_2 > 0, \end{cases} \tag{12.4}$$

and

$$x_1 = \begin{cases} x_2^2/2 - x_{2s}^2/2 + x_{1s}, & x_2 < x_{2s}, \quad \dot{x}_2 > 0, \\ -x_2^2/2 & 0 < x_2 < x_{2s}, \quad \dot{x}_2 < 0. \end{cases} \tag{12.5}$$

It is easy to see that through any point (x_1, x_2) of the plane passes exactly one curve of the forms (12.2)–(12.5). So for any point of the plane there exists exactly one extremal trajectory steering this point to the origin. Since optimal trajectories exist, then the solutions found are optimal.

12.2 Control of a linear oscillator

Consider a linear oscillator whose motion can be controlled by force bounded in absolute value. The corresponding control system (after appropriate rescaling) is

$$\ddot{x}_1 + x_1 = u, \qquad |u| \le 1, \quad x_1 \in \mathbb{R},$$

or, in the canonical form:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + u, \end{cases} \qquad |u| \le 1, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2.$$

We consider the time-optimal problem for this system.

By Filippov's theorem, optimal control exists. Similarly to the previous problem, we apply Pontryagin Maximum Principle: the Hamiltonian function is

$$h_u(\xi, x) = \xi_1 x_2 - \xi_2 x_1 + \xi_2 u, \qquad (\xi, x) \in T^*\mathbb{R}^2 = \mathbb{R}^{2*} \times \mathbb{R}^2,$$

and the Hamiltonian system reads

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + u, \end{cases} \qquad \begin{cases} \dot{\xi}_1 = \xi_2, \\ \dot{\xi}_2 = -\xi_1. \end{cases}$$

The maximality condition of PMP yields

$$\xi_2(t)\tilde{u}(t) = \max_{|u| \le 1} \xi_2(t) u,$$

thus optimal controls satisfy the condition

$$\tilde{u}(t) = \operatorname{sgn} \xi_2(t) \qquad \text{if } \xi_2(t) \ne 0.$$

For the variable ξ_2 we have the ODE

$$\ddot{\xi}_2 = -\xi_2,$$

hence

$$\xi_2 = \alpha \sin(t + \beta), \qquad \alpha,\ \beta = \text{const}.$$

Notice that $\alpha \ne 0$: indeed, if $\xi_2 \equiv 0$, then $\xi_1 = -\dot{\xi}_2(t) \equiv 0$, thus $\xi(t) = (\xi_1(t), \xi_2(t)) \equiv 0$, which is impossible by PMP. Consequently,

$$\tilde{u}(t) = \operatorname{sgn}(\alpha \sin(t + \beta)).$$

This equality yields a complete description of possible structure of optimal control. The interval between successive switching points of $\tilde{u}(t)$ has the length π. Let $\tau \in [0, \pi)$ be the first switching point of $\tilde{u}(t)$. Then

$$\tilde{u}(t) = \begin{cases} \operatorname{sgn} \tilde{u}(0), & t \in [0, \tau) \cup [\tau + \pi, \tau + 2\pi) \cup [\tau + 3\pi, \tau + 4\pi) \cup \dots \\ -\operatorname{sgn} \tilde{u}(0), & t \in [\tau, \tau + \pi) \cup [\tau + 2\pi, \tau + 3\pi) \cup \dots \end{cases}$$

That is, $\tilde{u}(t)$ is parametrized by two numbers: the first switching time $\tau \in [0, \pi)$ and the initial sign $\operatorname{sgn} \tilde{u}(0) \in \{\pm 1\}$.

Optimal control $\tilde{u}(t)$ takes only the extremal values ± 1. Thus optimal trajectories $(x_1(t), x_2(t))$ consist of pieces that satisfy the system

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 \pm 1, \end{cases} \tag{12.6}$$

i.e., arcs of the circles

$$(x_1 \pm 1)^2 + x_2^2 = C, \qquad C = \text{const},$$

passed clockwise.

Now we describe all optimal trajectories coming to the origin. Let γ be any such trajectory. If γ has no switchings, then it is an arc belonging to one of the semicircles

$$(x_1 - 1)^2 + x_2^2 = 1, \qquad x_2 \le 0, \tag{12.7}$$

$$(x_1 + 1)^2 + x_2^2 = 1, \qquad x_2 \ge 0 \tag{12.8}$$

and containing the origin. If γ has switchings, then the last switching can occur at any point of these semicircles except the origin. Assume that γ has the last switching on semicircle (12.7). Then the part of γ before the last switching and after the next to last switching is a semicircle of the circle $(x_1 + 1)^2 + x_2^2 = C$ passing through the last switching point. The next to last switching of γ occurs on the curve obtained by rotation of semicircle (12.7) around the point $(-1, 0)$ in the plane (x_1, x_2) by the angle π, i.e., on the semicircle

$$(x_1 + 3)^2 + x_2^2 = 1, \qquad x_2 \ge 0. \tag{12.9}$$

To obtain the geometric locus of the previous switching of γ, we have to rotate semicircle (12.9) around the point $(1, 0)$ by the angle π; we come to the semicircle

$$(x_1 - 5)^2 + x_2^2 = 1, \qquad x_2 \le 0.$$

The previous switching of γ takes place on the semicircle

$$(x_1 + 7)^2 + x_2^2 = 1, \qquad x_2 \ge 0,$$

and so on.

The case when the last switching of γ occurs on semicircle (12.8) is obtained from the case just considered by the central symmetry of the plane (x_1, x_2)

w.r.t. the origin: $(x_1, x_2) \mapsto (-x_1, -x_2)$. Then the successive switchings of γ (in the reverse order starting from the end) occur on the semicircles

$$\begin{aligned}
(x_1 + 1)^2 + x_2^2 &= 1, \qquad x_2 \geq 0,\\
(x_1 - 3)^2 + x_2^2 &= 1, \qquad x_2 \leq 0,\\
(x_1 + 5)^2 + x_2^2 &= 1, \qquad x_2 \geq 0,\\
(x_1 - 7)^2 + x_2^2 &= 1, \qquad x_2 \leq 0,
\end{aligned}$$

etc. We obtained the switching curve in the plane (x_1, x_2):

$$\begin{aligned}
(x_1 - (2k-1))^2 + x_2^2 = 1, \qquad x_2 \leq 0, \quad k \in \mathbb{N},\\
(x_1 + (2k-1))^2 + x_2^2 = 1, \qquad x_2 \geq 0, \quad k \in \mathbb{N}.
\end{aligned} \tag{12.10}$$

This switching curve divides the plane (x_1, x_2) into two parts. Any extremal trajectory $(x_1(t), x_2(t))$ in the upper part of the plane is a solution of ODE (12.6) with -1 in the second equation, and in the lower part it is a solution of (12.6) with $+1$. For any point of the plane (x_1, x_2) there exists exactly one curve of this family of extremal trajectories that comes to the origin (it has the form of a "spiral" with a finite number of switchings). Since optimal trajectories exist, the constructed extremal trajectories are optimal.

The time-optimal control problem is solved: in the part of the plane (x_1, x_2) over the switching curve (12.10) the optimal control is $\tilde{u} = -1$, and below this curve $\tilde{u} = +1$. Through any point of the plane passes one optimal trajectory which corresponds to this optimal control rule. After finite number of switchings, any optimal trajectory comes to the origin.

Now we consider optimal control problems with the same dynamics as in the previous two sections, but with another cost functional.

12.3 The cheapest stop of a train

As in Section 12.1, we control motion of a train. Now the goal is to stop the train at a fixed instant of time with a minimum expenditure of energy, which is assumed proportional to the integral of squared acceleration.

So the optimal control problem is as follows:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = u, \end{cases} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2, \quad u \in \mathbb{R},$$

$$x(0) = x^0, \quad x(t_1) = 0, \quad t_1 \text{ fixed},$$

$$\frac{1}{2} \int_0^{t_1} u^2 \, dt \to \min .$$

Filippov's theorem cannot be applied directly since the right-hand side of the control system is not compact. Although, one can choose a new time $t \mapsto$

$\frac{1}{2}\int_0^t u^2(\tau)\,d\tau + C$ and obtain a bounded right-hand side, then compactify it and apply Filippov's theorem. In such a way existence of optimal control can be proved. See also the general theory of linear quadratic problems below in Chapter 14.

To find optimal control, we apply PMP. The Hamiltonian function is

$$h_u^\nu(\xi, x) = \xi_1 x_2 + \xi_2 u + \frac{\nu}{2}u^2, \qquad (\xi, x) \in \mathbb{R}^{2*} \times \mathbb{R}^2.$$

Along optimal trajectories

$$\nu \leq 0, \quad \nu = \text{const}\,.$$

From the Hamiltonian system of PMP, we have

$$\begin{cases} \dot{\xi}_1 = 0, \\ \dot{\xi}_2 = -\xi_1. \end{cases} \tag{12.11}$$

Consider first the case of abnormal extremals:

$$\nu = 0.$$

The triple (ξ_1, ξ_2, ν) must be nonzero, thus

$$\xi_2(t) \not\equiv 0.$$

But the maximality condition of PMP yields

$$\tilde{u}(t)\xi_2(t) = \max_{u \in \mathbb{R}} u\, \xi_2(t). \tag{12.12}$$

Since $\xi_2(t) \neq 0$, the maximum above does not exist. Consequently, there are no abnormal extremals.

Consider the normal case: $\nu \neq 0$, we can take $\nu = -1$. The normal Hamiltonian function is

$$h_u(\xi, x) = h_u^{-1}(\xi, x) = \xi_1 x_2 + \xi_2 u - \frac{1}{2}u^2.$$

Maximality condition of PMP is equivalent to $\frac{\partial h_u}{\partial u} = 0$, thus

$$\tilde{u}(t) = \xi_2(t)$$

along optimal trajectories. Taking into account system (12.11), we conclude that optimal control is linear:

$$\tilde{u}(t) = \alpha t + \beta, \qquad \alpha,\ \beta = \text{const}\,.$$

The maximized Hamiltonian function

$$H(\xi, x) = \max_u h_u(\xi, x) = \xi_1 x_2 + \frac{1}{2}\xi_2^2$$

is smooth. That is why optimal trajectories satisfy the Hamiltonian system

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \xi_2, \\ \dot{\xi}_1 = 0, \\ \dot{\xi}_2 = -\xi_1. \end{cases}$$

For the variable x_1 we obtain the boundary value problem

$$\begin{aligned} & x_1^{(4)} = 0, \\ & x_1(0) = x_1^0, \quad \dot{x}_1(0) = x_2^0, \quad x_1(t_1) = 0, \quad \dot{x}_1(t_1) = 0. \end{aligned} \tag{12.13}$$

For any (x_1^0, x_2^0), there exists exactly one solution $x_1(t)$ of this problem — a cubic spline. The function $x_2(t)$ is found from the equation $x_2 = \dot{x}_1$.

So through any initial point $x^0 \in \mathbb{R}^2$ passes a unique extremal trajectory arriving at the origin. It is a curve $(x_1(t), x_2(t))$, $t \in [0, t_1]$, where $x_1(t)$ is a cubic polynomial that satisfies the boundary conditions (12.13), and $x_2(t) = \dot{x}_1(t)$. In view of existence, this is an optimal trajectory.

12.4 Control of a linear oscillator with cost

We control a linear oscillator, say a pendulum with a small amplitude, by an unbounded force u, but take into account expenditure of energy measured by the integral $\frac{1}{2}\int_0^{t_1} u^2(t)\,dt$. The optimal control problem reads

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + u, \end{cases} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2, \quad u \in \mathbb{R},$$

$$x(0) = x^0, \quad x(t_1) = 0, \quad t_1 \text{ fixed},$$

$$\frac{1}{2}\int_0^{t_1} u^2\,dt \to \min.$$

Existence of optimal control can be proved by the same argument as in the previous section.

The Hamiltonian function of PMP is

$$h_u^\nu(\xi, x) = \xi_1 x_2 - \xi_2 x_1 + \xi_2 u + \frac{\nu}{2} u^2.$$

The corresponding Hamiltonian system yields

$$\begin{cases} \dot{\xi}_1 = \xi_2, \\ \dot{\xi}_2 = -\xi_1. \end{cases}$$

In the same way as in the previous problem, we show that there are no abnormal extremals, thus we can assume $\nu = -1$. Then the maximality condition yields

$$\tilde{u}(t) = \xi_2(t).$$

In particular, optimal control is a harmonic:

$$\tilde{u}(t) = \alpha \sin(t + \beta), \qquad \alpha,\ \beta = \text{const}\,.$$

The system of ODEs for extremal trajectories

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + \alpha \sin(t + \beta) \end{cases}$$

is solved explicitly:

$$\begin{aligned} x_1(t) &= -\frac{\alpha}{2} t \cos(t + \beta) + a \sin(t + b), \\ x_2(t) &= \frac{\alpha}{2} t \sin(t + \beta) - \frac{\alpha}{2} \cos(t + \beta) + a \cos(t + b), \quad a,\ b \in \mathbb{R}. \end{aligned} \tag{12.14}$$

Exercise 12.1. Show that exactly one extremal trajectory of the form (12.14) satisfies the boundary conditions.

In view of existence, these extremal trajectories are optimal.

12.5 Dubins car

Consider a car moving in the plane. The car can move forward with a fixed linear velocity and simultaneously rotate with a bounded angular velocity. Given initial and terminal position and orientation of the car in the plane, the problem is to drive the car from the initial configuration to the terminal one for a minimal time.

Admissible paths of the car are curves with bounded curvature. Suppose that curves are parametrized by length, then our problem can be stated geometrically. Given two points in the plane and two unit velocity vectors attached respectively at these points, one has to find a curve in the plane that starts at the first point with the first velocity vector and comes to the second point with the second velocity vector, has curvature bounded by a given constant, and has the minimal length among all such curves.

Remark. If curvature is unbounded, then the problem, in general, has no solutions. Indeed, the infimum of lengths of all curves that satisfy the boundary conditions without bound on curvature is the distance between the initial and terminal points: the segment of the straight line through these points can be approximated by smooth curves with the required boundary conditions. But this infimum is not attained when the boundary velocity vectors do not lie on the line through the boundary points and are not collinear one to another.

After rescaling, we obtain a time-optimal problem for a nonlinear system:

$$\begin{cases} \dot{x}_1 = \cos\theta, \\ \dot{x}_2 = \sin\theta, \\ \dot{\theta} = u, \end{cases} \tag{12.15}$$

$$x = (x_1, x_2) \in \mathbb{R}^2, \quad \theta \in S^1, \quad |u| \leq 1,$$
$$x(0),\ \theta(0),\ x(t_1),\ \theta(t_1) \text{ fixed},$$
$$t_1 \to \min.$$

Existence of solutions is guaranteed by Filippov's Theorem. We apply Pontryagin Maximum Principle.

We have $(x_1, x_2, \theta) \in M = \mathbb{R}^2_x \times S^1_\theta$, let (ξ_1, ξ_2, μ) be the corresponding coordinates of the adjoint vector. Then

$$\lambda = (x, \theta, \xi, \mu) \in T^*M,$$

and the control-dependent Hamiltonian is

$$h_u(\lambda) = \xi_1 \cos\theta + \xi_2 \sin\theta + \mu u.$$

The Hamiltonian system of PMP yields

$$\dot{\xi} = 0, \tag{12.16}$$
$$\dot{\mu} = \xi_1 \sin\theta - \xi_2 \cos\theta, \tag{12.17}$$

and the maximality condition reads

$$\mu(t)u(t) = \max_{|u|\leq 1} \mu(t)u. \tag{12.18}$$

Equation (12.16) means that ξ is constant along optimal trajectories, thus the right-hand side of (12.17) can be rewritten as

$$\xi_1 \sin\theta - \xi_2 \cos\theta = \alpha \sin(\theta + \beta), \qquad \alpha,\ \beta = \text{const}, \quad \alpha = \sqrt{\xi_1^2 + \xi_2^2} \geq 0. \tag{12.19}$$

So the Hamiltonian system of PMP (12.15)–(12.17) yields the following system:

$$\begin{cases} \dot{\mu} = \alpha \sin(\theta + \beta), \\ \dot{\theta} = u. \end{cases}$$

Maximality condition (12.18) implies that

$$u(t) = \operatorname{sgn} \mu(t) \quad \text{if } \mu(t) \neq 0. \tag{12.20}$$

If $\alpha = 0$, then $(\xi_1, \xi_2) \equiv 0$ and $\mu = \text{const} \neq 0$, thus $u = \text{const} = \pm 1$. So the curve $x(t)$ is an arc of a circle of radius 1.

Let $\alpha \neq 0$, then in view of (12.19), we have $\alpha > 0$. Conditions (12.16), (12.17), (12.18) are preserved if the adjoint vector (ξ, μ) is multiplied by any positive constant. Thus we can choose (ξ, μ) such that $\alpha = \sqrt{\xi_1^2 + \xi_2^2} = 1$. That is why we suppose in the sequel that

$$\alpha = 1.$$

Condition (12.20) means that behavior of sign of the function $\mu(t)$ is crucial for the structure of optimal control. We consider several possibilities for $\mu(t)$.

(0) If the function $\mu(t)$ does not vanish on the segment $[0, t_1]$, then the optimal control is constant:

$$u(t) = \text{const} = \pm 1, \qquad t \in [0, t_1], \tag{12.21}$$

and the optimal trajectory $x(t)$, $t \in [0, t_1]$, is an arc of a circle. Notice that an optimal trajectory cannot contain a full circle: a circle can be eliminated so that the resulting trajectory satisfy the same boundary conditions and is shorter. Thus controls (12.21) can be optimal only if $t_1 < 2\pi$.

In the sequel we can assume that the set

$$N = \{\tau \in [0, t_1] \mid \mu(\tau) = 0\}$$

is nonempty. Since N is open, it is a union of open intervals in $[0, t_1]$, plus, may be, semiopen intervals of the form $[0, \tau_1)$, $(\tau_2, t_1]$.

(1) Suppose that the set N contains an interval of the form

$$(\tau_1, \tau_2) \subset [0, t_1], \qquad \tau_1 < \tau_2. \tag{12.22}$$

We can assume that the interval (τ_1, τ_2) is maximal w.r.t. inclusion:

$$\mu(\tau_1) = \mu(\tau_2) = 0, \qquad \mu|_{(\tau_1, \tau_2)} \neq 0.$$

From PMP we have the inequality

$$h_{u(t)}(\lambda(t)) = \cos(\theta(t) + \beta) + \mu(t)u(t) \geq 0.$$

Thus

$$\cos(\theta(\tau_1) + \beta) \geq 0.$$

This inequality means that the angle

$$\widehat{\theta} = \theta(\tau_1) + \beta$$

satisfies the inclusion

$$\widehat{\theta} \in \left[0, \frac{\pi}{2}\right] \cup \left[\frac{3\pi}{2}, 2\pi\right).$$

Consider first the case

$$\widehat{\theta} \in \left(0, \frac{\pi}{2}\right].$$

Then $\dot{\mu}(\tau_1) = \sin \widehat{\theta} > 0$, thus at τ_1 control switches from -1 to $+1$, so

$$\dot{\theta}(t) = u(t) \equiv 1, \qquad t \in (\tau_1, \tau_2).$$

We evaluate the distance $\tau_2 - \tau_1$. Since

$$\mu(\tau_2) = \int_{\tau_1}^{\tau_2} \sin(\widehat{\theta} + \tau - \tau_1)\, d\tau = 0,$$

then $\tau_2 - \tau_1 = 2(\pi - \widehat{\theta})$, thus

$$\tau_2 - \tau_1 \in [\pi, 2\pi). \tag{12.23}$$

In the case

$$\widehat{\theta} \in \left[\frac{3\pi}{2}, 2\pi\right)$$

inclusion (12.23) is proved similarly, and in the case $\widehat{\theta} = 0$ we obtain no optimal controls (the curve $x(t)$ contains a full circle, which can be eliminated).

Inclusion (12.23) means that successive roots τ_1, τ_2 of the function $\mu(t)$ cannot be arbitrarily close one to another. Moreover, the previous argument shows that at such instants τ_i optimal control switches from one extremal value to another, and along any optimal trajectory the distance between any successive switchings τ_i, τ_{i+1} is the same.

So in case (1) an optimal control can only have the form

$$u(t) = \begin{cases} \varepsilon, & t \in (\tau_{2k-1}, \tau_{2k}), \\ -\varepsilon, & t \in (\tau_{2k}, \tau_{2k+1}), \end{cases} \tag{12.24}$$

$$\varepsilon = \pm 1,$$

$$\tau_{i+1} - \tau_i = \text{const} \in [\pi, 2\pi), \qquad i = 1, \dots, N-1, \tag{12.25}$$

$$\tau_1 \in (0, 2\pi),$$

here we do not indicate values of u in the intervals before the first switching, $t \in (0, \tau_1)$, and after the last switching, $t \in (\tau_N, t_1)$. For such trajectories, control takes only extremal values ± 1 and the number of switchings is finite on any compact time segment. Such a control is called *bang-bang*.

Controls $u(t)$ given by (12.24), (12.25) satisfy PMP for arbitrarily large t, but they are not optimal if the number of switchings is $N > 3$. Indeed, suppose that such a control has at least 4 switchings. Then the piece of trajectory $x(t)$, $t \in [\tau_1, \tau_4]$, is a concatenation of three arcs of circles corresponding to the segments of time $[\tau_1, \tau_2]$, $[\tau_2, \tau_3]$, $[\tau_3, \tau_4]$ with

$$\tau_4 - \tau_3 = \tau_3 - \tau_2 = \tau_2 - \tau_1 \in [\pi, 2\pi).$$

Draw the segment of line

$$\tilde{x}(t), \qquad t \in [(\tau_1 + \tau_2)/2, (\tau_3 + \tau_4)/2]\,, \qquad \left|\frac{d\,\tilde{x}}{d\,t}\right| \equiv 1,$$

the common tangent to the first and third circles through the points $x\left((\tau_1+\tau_2)/2\right)$ and $x\left((\tau_3+\tau_4)/2\right)$. Then the curve

$$y(t)=\begin{cases} x(t), & t\notin\left[(\tau_1+\tau_2)/2,(\tau_3+\tau_4)/2\right], \\ \tilde{x}(t), & t\in\left[(\tau_1+\tau_2)/2,(\tau_3+\tau_4)/2\right], \end{cases}$$

is an admissible trajectory and shorter than $x(t)$. We proved that optimal bang-bang control can have not more than 3 switchings.

(2) It remains to consider the case where the set N does not contain intervals of the form (12.22). Then N consists of at most two semiopen intervals:

$$N=[0,\tau_1)\cup(\tau_2,t_1], \qquad \tau_1\le\tau_2,$$

where one or both intervals may be absent. If $\tau_1=\tau_2$, then the function $\mu(t)$ has a unique root on the segment $[0,t_1]$, and the corresponding optimal control is determined by condition (12.20). Otherwise

$$\tau_1<\tau_2,$$

and

$$\mu|_{[0,\tau_1)}\neq 0, \quad \mu|_{[\tau_1,\tau_2]}\equiv 0, \quad \mu|_{(\tau_2,t_1]}\neq 0. \tag{12.26}$$

In this case the maximality condition of PMP (12.20) does not determine optimal control $u(t)$ uniquely since the maximum is attained for more than one value of control parameter u. Such a control is called *singular*. Nevertheless, singular controls in this problem can be determined from PMP. Indeed, the following identities hold on the interval (τ_1,τ_2):

$$\dot{\mu}=\sin(\theta+\beta)=0 \quad\Rightarrow\quad \theta+\beta=\pi k \quad\Rightarrow\quad \theta=\text{const} \quad\Rightarrow\quad u=0.$$

Consequently, if an optimal trajectory $x(t)$ has a singular piece, which is a line, then τ_1 and τ_2 are the only switching times of the optimal control. Then

$$u|_{(0,\tau_1)}=\text{const}=\pm1, \quad u|_{(\tau_2,t_1)}=\text{const}=\pm1,$$

and the whole trajectory $x(t)$, $t\in[0,t_1]$, is a concatenation of an arc of a circle of radius 1

$$x(t), \quad u(t)=\pm1, \qquad t\in[0,\tau_1],$$

a line

$$x(t), \quad u(t)=0, \qquad t\in[\tau_1,\tau_2],$$

and one more arc of a circle of radius 1

$$x(t), \quad u(t)=\pm1, \qquad t\in[\tau_2,t_1].$$

So optimal trajectories in the problem have one of the following two types:

(1) concatenation of a bang-bang piece (arc of a circle, $u=\pm1$), a singular piece (segment of a line, $u=0$), and a bang-bang piece, or

(2) concatenation of bang-bang pieces with not more than 3 switchings, the arcs of circles between switchings having the same central angle $\in [\pi, 2\pi)$.

If boundary points $x(0)$, $x(t_1)$ are sufficiently far one from another, then they can be connected only by trajectories containing singular piece. For such boundary points, we obtain a simple algorithm for construction of an optimal trajectory. Through each of the points $x(0)$ and $x(t_1)$, construct a pair of circles of radius 1 tangent respectively to the velocity vectors $\dot{x}(0) = (\cos\theta(0), \sin\theta(0))$ and $\dot{x}(t_1) = (\cos\theta(t_1), \sin\theta(t_1))$. Then draw common tangents to the circles at $x(0)$ and $x(t_1)$ respectively, so that direction of motion along these tangents was compatible with direction of rotation along the circles determined by the boundary velocity vectors $\dot{x}(0)$ and $\dot{x}(t_1)$. Finally, choose the shortest curve among the candidates obtained. This curve is the optimal trajectory.

Chapter 13

Linear time-optimal problem

13.1 Problem statement

In this chapter we study the following optimal control problem:

$$\begin{aligned}
&\dot{x} = Ax + Bu, \qquad x \in \mathbb{R}^n, \quad u \in U \subset \mathbb{R}^m, \\
&x(0) = x_0, \quad x(t_1) = x_1, \quad x_0,\ x_1 \in \mathbb{R}^n \text{ fixed}, \\
&t_1 \to \min,
\end{aligned} \tag{13.1}$$

where U is a compact convex polytope in $\mathbb{R}^m$, and A and B are constant matrices of order $n \times n$ and $n \times m$ respectively. Such problem is called *linear time-optimal problem*.

The polytope U is the convex hull of a finite number of points $a_1, \dots, a_k$ in $\mathbb{R}^m$:

$$U = \operatorname{conv}\{a_1, \dots, a_k\}.$$

We assume that the points a_i do not belong to the convex hull of all the rest points a_j, $j \neq i$, so that each a_i is a vertex of the polytope U.

In the sequel we assume the following *General Position Condition*:

For any edge $[a_i, a_j]$ of U, the vector $e_{ij} = a_j - a_i$ satisfies the equality

$$\operatorname{span}(Be_{ij}, ABe_{ij}, \dots, A^{n-1}Be_{ij}) = \mathbb{R}^n. \tag{13.2}$$

This condition means that no vector Be_{ij} belongs to a proper invariant subspace of the matrix A. By Theorem 3.1, this is equivalent to controllability of the linear system $\dot{x} = Ax + Bu$ with the set of control parameters $u \in \mathbb{R}e_{ij}$. Condition (13.2) can be achieved by a small perturbation of matrices A, B.

We already considered examples of linear time-optimal problems in Sections 12.1, 12.2. Here we study the structure of optimal control, prove its uniqueness, evaluate the number of switchings.

Existence of optimal control for any points x_0, x_1 such that $x_1 \in \mathcal{A}(x_0)$ is guaranteed by Filippov's theorem. Notice that for the analogous problem with an unbounded set of control parameters, optimal control may not exist: it is easy to show this using linearity of the system.

Before proceeding with the study of linear time-optimal problems, we recall some basic facts on polytopes.

13.2 Geometry of polytopes

The convex hull of a finite number of points $a_1, \dots, a_k \in \mathbb{R}^m$ is the set

$$U = \operatorname{conv}\{a_1, \dots, a_k\} \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^k \alpha_i a_i \mid \alpha_i \geq 0,\ \sum_{i=1}^k \alpha_i = 1 \right\}.$$

An affine hyperplane in $\mathbb{R}^m$ is a set of the form

$$\Pi = \{u \in \mathbb{R}^m \mid \langle \xi, u \rangle = c\}, \qquad \xi \in \mathbb{R}^{m*} \setminus \{0\}, \quad c \in \mathbb{R}.$$

A supporting hyperplane to a polytope U is a hyperplane Π such that

$$\langle \xi, u \rangle \leq c \quad \forall u \in U$$

for the covector ξ and number c that define Π, and this inequality turns into equality at some point $u \in \partial U$, i.e., $\Pi \cap U \neq \emptyset$.

A polytope $U = \operatorname{conv}\{a_1, \dots, a_k\}$ intersects with any its supporting hyperplane $\Pi = \{u \mid \langle \xi, u \rangle = c\}$ by another polytope:

$$\begin{aligned} &U \cap \Pi = \operatorname{conv}\{a_{i1}, \dots, a_{il}\}, \\ &\langle \xi, a_{i1} \rangle = \dots = \langle \xi, a_{il} \rangle = c, \\ &\langle \xi, a_j \rangle < c, \quad j \notin \{i_1, \dots, i_l\}. \end{aligned}$$

Such polytopes $U \cap \Pi$ are called faces of the polytope U. Zero-dimensional and one-dimensional faces are called respectively vertices and edges. A polytope has a finite number of faces, each of which is the convex hull of a finite number of vertices. A face of a face is a face of the initial polytope. Boundary of a polytope is a union of all its faces. This is a straightforward corollary of the separation theorem for convex sets (or the Hahn-Banach Theorem).

13.3 Bang-bang theorem

Optimal control in the linear time-optimal problem is bang-bang, i.e., it is piecewise constant and takes values in vertices of the polytope U.

Theorem 13.1. *Let $u(t)$, $0 \leq t \leq t_1$, be an optimal control in the linear time-optimal control problem* (13.1)*. Then there exists a finite subset*

$$\mathcal{T} \subset [0, t_1], \qquad \#\mathcal{T} < \infty,$$

such that

$$u(t) \in \{a_1, \dots, a_k\}, \qquad t \in [0, t_1] \setminus \mathcal{T}, \tag{13.3}$$

and restriction $u(t)|_{t \in [0,t_1] \setminus \mathcal{T}}$ *is locally constant.*

Proof. Apply Pontryagin Maximum Principle to the linear time-optimal problem (13.1). State vector and adjoint vector are

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \qquad \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^{n*},$$

and a point in the cotangent bundle is

$$\lambda = (\xi, x) \in \mathbb{R}^{n*} \times \mathbb{R}^n = T^*\mathbb{R}^n.$$

The control-dependent Hamiltonian is

$$h_u(\xi, x) = \xi A x + \xi B u$$

(we multiply raws by columns). The Hamiltonian system and maximality condition of PMP take the form:

$$\begin{cases} \dot{x} = Ax + Bu, \\ \dot{\xi} = -\xi A, \end{cases}$$
$$\xi(t) \neq 0,$$
$$\xi(t) B u(t) = \max_{u \in U} \xi(t) B u. \tag{13.4}$$

The Hamiltonian system implies that adjoint vector

$$\xi(t) = \xi(0) e^{-tA}, \qquad \xi(0) \neq 0, \tag{13.5}$$

is analytic along the optimal trajectory.

Consider the set of indices corresponding to vertices where maximum (13.4) is attained:

$$J(t) = \left\{ 1 \leq j \leq k \mid \xi(t) B a_j = \max_{u \in U} \xi(t) B u = \max\{\xi(t) B a_i \mid i = 1, \dots, k\} \right\}.$$

At each instant t the linear function $\xi(t)B$ attains maximum at vertices of the polytope U. We show that this maximum is attained at one vertex always except a finite number of moments.

Define the set

$$\mathcal{T} = \{ t \in [0, t_1] \mid \# J(t) > 1 \}.$$

By contradiction, suppose that $\mathcal{T}$ is infinite: there exists a sequence of distinct moments

$$\{\tau_1, \dots, \tau_n, \dots\} \subset \mathcal{T}.$$

Since there is a finite number of choices for the subset $J(\tau_n) \subset \{1, \dots, k\}$, we can assume, without loss of generality, that

$$J(\tau_1) = J(\tau_2) = \cdots = J(\tau_n) = \cdots .$$

Denote $J = J(\tau_i)$.

Further, since the convex hull

$$\operatorname{conv}\{a_j \mid j \in J\}$$

is a face of U, then there exist indices $j_1, j_2 \in J$ such that the segment $[a_{j_1}, a_{j_2}]$ is an edge of U. We have

$$\xi(\tau_i) B a_{j_1} = \xi(\tau_i) B a_{j_2}, \qquad i = 1, 2, \dots .$$

For the vector $e = a_{j_2} - a_{j_1}$ we obtain

$$\xi(\tau_i) B e = 0, \qquad i = 1, 2, \dots .$$

But $\xi(\tau_i) = \xi(0) e^{-\tau_i A}$ by (13.5), so the analytic function

$$t \mapsto \xi(0) e^{-tA} B e$$

has an infinite number of zeros on the segment $[0, t_1]$, thus it is identically zero:

$$\xi(0) e^{-tA} B e \equiv 0.$$

We differentiate this identity successively at $t = 0$ and obtain

$$\xi(0) B e = 0, \quad \xi(0) A B e = 0, \quad \dots, \quad \xi(0) A^{n-1} B e = 0.$$

By General Position Condition (13.2), we have $\xi(0) = 0$, a contradiction to (13.5). So the set $\mathcal{T}$ is finite.

Out of the set $\mathcal{T}$, the function $\xi(t)B$ attains maximum on U at one vertex $a_{j(t)}$, $\{j(t)\} = J(t)$, thus the optimal control $u(t)$ takes value in the vertex $a_{j(t)}$. Condition (13.3) follows. Further,

$$\xi(t) B a_{j(t)} > \xi(t) B a_i, \qquad i \neq j(t).$$

But all functions $t \mapsto \xi(t) B a_i$ are continuous, so the preceding inequality preserves for instants close to t. The function $t \mapsto j(t)$ is locally constant on $[0, t_1] \setminus \mathcal{T}$, thus the optimal control $u(t)$ is also locally constant on $[0, t_1] \setminus \mathcal{T}$. □

In the sequel we will need the following statement proved in the preceding argument.

Corollary 13.1. *Let $\xi(t)$, $t \in [0, t_1]$, be a nonzero solution of the adjoint equation $\dot{\xi} = -\xi A$. Then everywhere in the segment $[0, t_1]$, except a finite number of points, there exists a unique control $u(t) \in U$ such that $\xi(t) B u(t) = \max_{u \in U} \xi(t) B u$.*

13.4 Uniqueness of optimal controls and extremals

Theorem 13.2. *Let the terminal point x_1 be reachable from the initial point x_0:*

$$x_1 \in \mathcal{A}(x_0).$$

Then linear time-optimal control problem (13.1) *has a unique solution.*

Proof. As we already noticed, existence of an optimal control follows from Filippov's Theorem.

Suppose that there exist two optimal controls: $u_1(t)$, $u_2(t)$, $t \in [0, t_1]$. By Cauchy's formula:

$$x(t_1) = e^{t_1 A} \left(x_0 + \int_0^{t_1} e^{-tA} B u(t)\, dt \right),$$

we obtain

$$e^{t_1 A} \left(x_0 + \int_0^{t_1} e^{-tA} B u_1(t)\, dt \right) = e^{t_1 A} \left(x_0 + \int_0^{t_1} e^{-tA} B u_2(t)\, dt \right),$$

thus

$$\int_0^{t_1} e^{-tA} B u_1(t)\, dt = \int_0^{t_1} e^{-tA} B u_2(t)\, dt. \tag{13.6}$$

Let $\xi_1(t) = \xi_1(0) e^{-tA}$ be the adjoint vector corresponding by PMP to the control $u_1(t)$. Then equality (13.6) can be written in the form

$$\int_0^{t_1} \xi_1(t) B u_1(t)\, dt = \int_0^{t_1} \xi_1(t) B u_2(t)\, dt. \tag{13.7}$$

By the maximality condition of PMP

$$\xi_1(t) B u_1(t) = \max_{u \in U} \xi_1(t) B u,$$

thus

$$\xi_1(t) B u_1(t) \geq \xi_1(t) B u_2(t).$$

But this inequality together with equality (13.7) implies that almost everywhere on $[0, t_1]$

$$\xi_1(t) B u_1(t) = \xi_1(t) B u_2(t).$$

By Corollary 13.1,

$$u_1(t) \equiv u_2(t)$$

almost everywhere on $[0, t_1]$. □

So for linear time-optimal problem, optimal control is unique. The standard procedure to find the optimal control for a given pair of boundary points x_0, x_1 is to find all extremals $(\xi(t), x(t))$ steering x_0 to x_1 and then to seek for the best among them. In the examples considered in Sections 12.1, 12.2, there was one extremal for each pair x_0, x_1 with $x_1 = 0$. We prove now that this is a general property of linear time-optimal problems.

Theorem 13.3. *Let $x_1 = 0 \in \mathcal{A}(x_0)$ and $0 \in U \setminus \{a_1, \dots, a_k\}$. Then there exists a unique control $u(t)$ that steers x_0 to 0 and satisfies Pontryagin Maximum Principle.*

Proof. Assume that there exist two controls

$$u_1(t), \quad t \in [0, t_1], \qquad \text{and} \qquad u_2(t), \quad t \in [0, t_2],$$

that steer x_0 to 0 and satisfy PMP.

If $t_1 = t_2$, then the argument of the proof of preceding theorem shows that $u_1(t) \equiv u_2(t)$ a.e., so we can assume that

$$t_1 > t_2.$$

Cauchy's formula gives

$$e^{t_1 A}\left(x_0 + \int_0^{t_1} e^{-tA} B u_1(t)\, dt\right) = 0,$$
$$e^{t_2 A}\left(x_0 + \int_0^{t_2} e^{-tA} B u_2(t)\, dt\right) = 0,$$

thus

$$\int_0^{t_1} e^{-tA} B u_1(t)\, dt = \int_0^{t_2} e^{-tA} B u_2(t)\, dt. \tag{13.8}$$

According to PMP, there exists an adjoint vector $\xi_1(t)$, $t \in [0, t_1]$, such that

$$\xi_1(t) = \xi_1(0) e^{-tA}, \qquad \xi_1(0) \neq 0, \tag{13.9}$$
$$\xi_1(t) B u_1(t) = \max_{u \in U} \xi_1(t) B u. \tag{13.10}$$

Since $0 \in U$, then

$$\xi_1(t) B u_1(t) \geq 0, \qquad t \in [0, t_1]. \tag{13.11}$$

Equality (13.8) can be rewritten as

$$\int_0^{t_1} \xi_1(t) B u_1(t)\, dt = \int_0^{t_2} \xi_1(t) B u_2(t)\, dt. \tag{13.12}$$

Taking into account inequality (13.11), we obtain

$$\int_0^{t_2} \xi_1(t) B u_1(t)\, dt \leq \int_0^{t_2} \xi_1(t) B u_2(t)\, dt. \tag{13.13}$$

But maximality condition (13.10) implies that

$$\xi_1(t)Bu_1(t) \geq \xi_1(t)Bu_2(t), \qquad t \in [0, t_2]. \tag{13.14}$$

Now inequalities (13.13) and (13.14) are compatible only if

$$\xi_1(t)Bu_1(t) = \xi_1(t)Bu_2(t), \qquad t \in [0, t_2],$$

thus inequality (13.13) should turn into equality. In view of (13.12), we have

$$\int_{t_1}^{t_2} \xi_1(t)Bu_1(t)\, dt = 0.$$

Since the integrand is nonnegative, see (13.11), then it vanishes identically:

$$\xi_1(t)Bu_1(t) \equiv 0, \qquad t \in [t_1, t_2].$$

By the argument of Theorem 13.1, the control $u_1(t)$ is bang-bang, so there exists an interval $I \subset [t_1, t_2]$ such that

$$u_1(t)|_I \equiv a_j \neq 0.$$

Thus

$$\xi_1(t)Ba_j \equiv 0, \qquad t \in I.$$

But $\xi_1(t)0 \equiv 0$, this is a contradiction with uniqueness of the control for which maximum in PMP is obtained, see Corollary 13.1. □

13.5 Switchings of optimal control

Now we evaluate the number of switchings of optimal control in linear time-optimal problems. In the examples of Sections 12.1, 12.2 we had respectively one switching and an arbitrarily large number of switchings, although finite on any segment. It turns out that in general there are two cases: non-oscillating and oscillating, depending on whether the matrix A of the control system has real spectrum or not. Recall that in the example with one switching, Section 12.1, we had

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \operatorname{Sp}(A) = \{0\} \subset \mathbb{R},$$

and in the example with arbitrarily large number of switchings, Section 12.2,

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \operatorname{Sp}(A) = \{\pm i\} \not\subset \mathbb{R}.$$

We consider systems with scalar control:

$$\dot{x} = Ax + ub, \qquad u \in U = [\alpha, \beta] \subset \mathbb{R}, \quad x \in \mathbb{R}^n,$$

under the General Position Condition

$$\operatorname{span}(b, Ab, \dots, A^{n-1}b) = \mathbb{R}^n.$$

Then attainable set of the system is full-dimensional for arbitrarily small times. We can evaluate the minimal number of switchings necessary to fill a full-dimensional domain. Optimal control is piecewise constant with values in $\{\alpha, \beta\}$. Assume that we start from the initial point x_0 with the control α. Without switchings we fill a piece of a 1-dimensional curve $e^{(Ax+\alpha b)t}x_0$, with 1 switching we fill a piece of a 2-dimensional surface $e^{(Ax+\beta b)t_2} \circ e^{(Ax+\alpha b)t_1}x_0$, with 2 switchings we can attain points in a 3-dimensional surface, etc. So the minimal number of switchings required to reach an n-dimensional domain is $n-1$.

We prove now that in the non-oscillating case we never need more than $n-1$ switchings of optimal control.

Theorem 13.4. *Assume that the matrix A has only real eigenvalues:*

$$\operatorname{Sp}(A) \subset \mathbb{R}.$$

Then any optimal control in linear time-optimal problem (13.1) *has no more than $n-1$ switchings.*

Proof. Let $u(t)$ be an optimal control and $\xi(t) = \xi(0)e^{-tA}$ the corresponding solution of the adjoint equation $\dot{\xi} = -\xi A$. The maximality condition of PMP reads

$$\xi(t)bu(t) = \max_{u \in [\alpha, \beta]} \xi(t)bu,$$

thus

$$u(t) = \begin{cases} \beta & \text{if } \xi(t)b > 0, \\ \alpha & \text{if } \xi(t)b < 0. \end{cases}$$

So the number of switchings of the control $u(t)$, $t \in [0, t_1]$, is equal to the number of changes of sign of the function

$$y(t) = \xi(t)b, \qquad t \in [0, t_1].$$

We show that $y(t)$ has not more than $n-1$ real roots.

Derivatives of the adjoint vector have the form

$$\xi^{(k)}(t) = \xi(0)e^{-tA}(-A)^k.$$

By Cayley Theorem, the matrix A satisfies its characteristic equation:

$$A^n + c_1A^{n-1} + \cdots + c_n \operatorname{Id} = 0,$$

where

$$\det(t \operatorname{Id} - A) = t^n + c_1t^{n-1} + \cdots + c_n,$$

thus

$$(-A)^n - c_1(-A)^{n-1} + \cdots + (-1)^n c_n \operatorname{Id} = 0.$$

Then the function $y(t)$ satisfies an n-th order ODE:

$$y^{(n)}(t) - c_1 y^{(n-1)}(t) + \cdots + (-1)^n c_n y(t) = 0. \tag{13.15}$$

It is well known (see e.g. [4]) that any solution of this equation is a quasipolynomial:

$$\begin{aligned} & y(t) = \sum_{i=1}^{k} e^{-\lambda_i t} P_i(t), \\ & P_i(t) \text{ a polynomial}, \\ & \lambda_i \neq \lambda_j \text{ for } i \neq j, \end{aligned}$$

where λ_i are eigenvalues of the matrix A and degree of each polynomial P_i is less than multiplicity of the corresponding eigenvalue λ_i, thus

$$\sum_{i=1}^{k} \deg P_i \leq n - k.$$

Now the statement of this theorem follows from the next general lemma. □

Lemma 13.1. *A quasipolynomial*

$$\begin{aligned} & y(t) = \sum_{i=1}^{k} e^{\lambda_i t} P_i(t), \qquad \sum_{i=1}^{k} \deg P_i \leq n - k, \\ & \lambda_i \neq \lambda_j \textit{ for } i \neq j, \end{aligned} \tag{13.16}$$

has no more than $n-1$ *real roots.*

Proof. Apply induction on k.

If $k = 1$, then a quasipolynomial

$$y(t) = e^{\lambda t} P(t), \qquad \deg P \leq n - 1,$$

has no more than $n - 1$ roots.

We prove the induction step for $k > 1$. Denote

$$n_i = \deg P_i, \qquad i = 1, \ldots, k.$$

Suppose that the quasipolynomial $y(t)$ has n real roots. Rewrite the equation

$$y(t) = \sum_{i=1}^{k-1} e^{\lambda_i t} P_i(t) + e^{\lambda_k t} P_k(t) = 0$$

as follows:

$$\sum_{i=1}^{k-1} e^{(\lambda_i-\lambda_k)t} P_i(t) + P_k(t) = 0. \tag{13.17}$$

The quasipolynomial in the left-hand side has n roots. We differentiate this quasipolynomial successively (n_k+1) times so that the polynomial $P_k(t)$ disappear. After (n_k+1) differentiations we obtain a quasipolynomial

$$\sum_{i=1}^{k-1} e^{(\lambda_i-\lambda_k)t} Q_i(t), \qquad \deg Q_i \le \deg P_i,$$

which has $(n-n_k-1)$ real roots by Rolle's Theorem. But by induction assumption the maximal possible number of real roots of this quasipolynomial is

$$\sum_{i=1}^{k-1} n_i + k - 2 < n - n_k - 1.$$

The contradiction finishes the proof of the lemma. □

So we completed the proof of Theorem 13.4: in the non-oscillating case an optimal control has no more than $n-1$ switchings on the whole domain (recall that $n-1$ switchings are always necessary even on short time segments since the attainable sets $\mathcal{A}_{q_0}(t)$ are full-dimensional for all $t>0$).

For an arbitrary matrix A, one can obtain the upper bound of $(n-1)$ switchings for sufficiently short intervals of time.

Theorem 13.5. *Consider the characteristic polynomial of the matrix A:*

$$\det(t\,\mathrm{Id} - A) = t^n + c_1 t^{n-1} + \cdots + c_n,$$

and let

$$c = \max_{1\le i\le n} |c_i|.$$

Then for any time-optimal control $u(t)$ and any $\bar{t} \in \mathbb{R}$, the real segment

$$\left[\bar{t}, \bar{t} + \ln\left(1 + \frac{1}{c}\right)\right]$$

contains not more than $(n-1)$ switchings of an optimal control $u(t)$.

In the proof of this theorem we will require the following general proposition, which we learned from S. Yakovenko.

Lemma 13.2. *Consider an ODE*

$$y^{(n)} + c_1(t) y^{(n-1)} + \cdots + c_n(t) y = 0$$

with measurable and bounded coefficients:

$$c_i = \max_{t\in[\bar t,\bar t+\delta]} |c_i(t)|.$$

If

$$\sum_{k=1}^{n} c_k \frac{\delta^k}{k!} < 1, \tag{13.18}$$

then any nonzero solution $y(t)$ of the ODE has not more than $n-1$ roots on the segment $t \in [\bar t, \bar t+\delta]$.

Proof. By contradiction, suppose that the function $y(t)$ has at least n roots on the segment $t \in [\bar t, \bar t+\delta]$. By Rolle's Theorem, derivative $\dot y(t)$ has not less than $n-1$ roots, etc. Then $y^{(n-1)}$ has a root $t_{n-1} \in [\bar t, \bar t+\delta]$. Thus

$$y^{(n-1)}(t) = \int_{t_{n-1}}^{t} y^{(n)}(\tau)\, d\tau.$$

Let $t_{n-2} \in [\bar t, \bar t+\delta]$ be a root of $y^{(n-2)}(t)$, then

$$y^{(n-2)}(t) = \int_{t_{n-2}}^{t} d\tau_1 \int_{t_{n-1}}^{\tau_1} y^{(n)}(\tau_2)\, d\tau_2.$$

We continue this procedure by integrating $y^{(n-i+1)}(t)$ from a root $t_{n-i} \in [\bar t, \bar t+\delta]$ of $y^{(n-i)}(t)$ and obtain

$$y^{(n-i)}(t) = \int_{t_{n-i}}^{t} d\tau_1 \int_{t_{n-i+1}}^{\tau_1} d\tau_2 \cdots \int_{t_{n-1}}^{\tau_{i-1}} y^{(n)}(\tau_i)\, d\tau_i, \qquad i = 1, \dots, n.$$

There holds a bound:

$$\begin{aligned}\left|y^{(n-i)}(t)\right| &\le \int_{t_{n-i}}^{t} d\tau_1 \int_{t_{n-i+1}}^{\tau_1} d\tau_2 \cdots \int_{t_n-1}^{\tau_{i-1}} \left|y^{(n)}(\tau_i)\right| d\tau_i \\ &\le \int_{\bar t}^{\bar t+\delta} d\tau_1 \int_{\bar t}^{\tau_1} d\tau_2 \cdots \int_{\bar t}^{\tau_{i-1}} \left|y^{(n)}(\tau_i)\right| d\tau_i \le \frac{\delta^k}{k!} \sup_{t\in[\bar t,\bar t+\delta]} \left|y^{(n)}(t)\right|.\end{aligned}$$

Then

$$\left|\sum_{i=1}^{n} c_i(t) y^{(n-i)}(t)\right| \le \sum_{i=1}^{n} |c_i(t)| \left|y^{(n-i)}(t)\right| \le \sum_{i=1}^{n} c_i \frac{\delta^k}{k!} \sup_{t\in[\bar t,\bar t+\delta]} \left|y^{(n)}(t)\right|,$$

i.e.,

$$\left|y^{(n)}(t)\right| \le \sum_{i=1}^{n} c_i \frac{\delta^k}{k!} \sup_{t\in[\bar t,\bar t+\delta]} \left|y^{(n)}(t)\right|,$$

a contradiction with (13.18). The lemma is proved. □

Now we prove Theorem 13.5.

Proof. As we showed in the proof of Theorem 13.4, the number of switchings of $u(t)$ is not more than the number of roots of the function $y(t) = \xi(t)b$, which satisfies ODE (13.15).

We have

$$\sum_{k=1}^{n} |c_k| \frac{\delta^k}{k!} < c(e^\delta - 1) \quad \forall \delta > 0.$$

By Lemma 13.2, if

$$c(e^\delta - 1) \leq 1, \tag{13.19}$$

then the function $y(t)$ has not more than $n-1$ real roots on any interval of length δ. But inequality (13.19) is equivalent to the following one:

$$\delta \leq \ln\left(1 + \frac{1}{c}\right),$$

so $y(t)$ has not more than $n-1$ roots on any interval of the length $\ln\left(1+\frac{1}{c}\right)$. $\square$

Chapter 14

Linear-quadratic problem

14.1 Problem statement and assumptions

In this chapter we study a class of optimal control problems very popular in applications, *linear-quadratic problems*. That is, we consider linear systems with quadratic cost functional:

$$
\begin{aligned}
&\dot{x} = Ax + Bu, \qquad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^m, \\
&x(0) = x_0, \quad x(t_1) = x_1, \qquad x_0, x_1, t_1 \text{ fixed}, \\
&J(u) = \frac{1}{2}\int_0^{t_1} \langle Ru(t), u(t)\rangle + \langle Px(t), u(t)\rangle + \langle Qx(t), x(t)\rangle \, dt \to \min .
\end{aligned}
\tag{14.1}
$$

Here A, B, R, P, Q are constant matrices of appropriate dimensions, R, Q are symmetric:

$$R^* = R, \qquad Q^* = Q,$$

and angle brackets $\langle \cdot, \cdot \rangle$ denote the standard inner product in $\mathbb{R}^m$ and $\mathbb{R}^n$.

One can show that the condition $R \geq 0$ is necessary for existence of optimal control. We do not touch here the case of degenerate R and assume that $R > 0$. The substitution of variables $u \mapsto v = R^{1/2}u$ transforms the functional $J(u)$ to a similar functional with the identity matrix instead of R. That is why we assume in the sequel that $R = \mathrm{Id}$. Another change of variables kills the matrix P (exercise: find this change of variables). So we can write the cost functional as follows:

$$J(u) = \frac{1}{2}\int_0^{t_1} |u(t)|^2 + \langle Qx(t), x(t)\rangle \, dt.$$

For dynamics of the problem, we assume that the linear system is controllable:

$$\operatorname{rank}(B, AB, \dots, A^{n-1}B) = n. \tag{14.2}$$

14.2 Existence of optimal control

Since the set of control parameters $U = \mathbb{R}^m$ is noncompact, Filippov's Theorem does not apply, and existence of optimal controls in linear-quadratic problems is a nontrivial problem.

In this chapter we assume that admissible controls are square-integrable:

$$u \in L_2^m[0, t_1]$$

and use the L_2^m norm for controls:

$$\|u\| = \left(\int_0^{t_1} |u(t)|^2 \, dt\right)^{1/2} = \left(\int_0^{t_1} u_1^2(t) + \cdots + u_m^2(t) \, dt\right)^{1/2}.$$

Consider the set of all admissible controls that steer the initial point to the terminal one:

$$U(x_0, x_1) = \{u \in L_2^m[0, t_1] \mid x(t_1, u, x_0) = x_1\}.$$

We denote by $x(t, u, x_0)$ the trajectory of system (14.1) corresponding to an admissible control $u \in L_2^m$ starting at a point $x_0 \in \mathbb{R}^n$. By Cauchy's formula, the endpoint mapping

$$u \mapsto x(t_1, u, x_0) = e^{t_1 A} x_0 + \int_0^{t_1} e^{(t_1 - \tau)A} B u(\tau) \, d\tau$$

is an affine mapping from $L_2^m[0, t_1]$ to $\mathbb{R}^n$. Controllability of the linear system (14.1) means that for any $x_0 \in \mathbb{R}^n$, $t_1 > 0$, the image of the endpoint mapping is the whole $\mathbb{R}^n$. Thus

$$U(x_0, x_1) \subset L_2^m[0, t_1]$$

is an affine subspace,

$$U(0, 0) \subset L_2^m[0, t_1]$$

is a linear subspace, and

$$U(x_0, x_1) = u + U(0, 0) \quad \text{for any } u \in U(x_0, x_1).$$

Thus it is natural that existence of optimal controls is closely related to behavior of the cost functional $J(u)$ on the linear subspace $U(0, 0)$.

Proposition 14.1. (1) *If there exist points $x_0, x_1 \in \mathbb{R}^n$ such that*

$$\inf_{u \in U(x_0, x_1)} J(u) > -\infty, \tag{14.3}$$

then

$$J(u) \geq 0 \quad \forall u \in U(0, 0).$$

(2) *Conversely, if*

$$J(u) > 0 \quad \forall u \in U(0,0) \setminus 0,$$

then the minimum is attained:

$$\exists \min_{u \in U(x_0,x_1)} J(u) \quad \forall x_0,\ x_1 \in \mathbb{R}^n.$$

Remark. That is, the inequality

$$J|_{U(0,0)} \geq 0$$

is necessary for existence of optimal controls, at least for one pair (x_0, x_1), and the strict inequality

$$J|_{U(0,0)\setminus 0} > 0$$

is sufficient for existence of optimal controls for all pairs (x_0, x_1).

In the proof of Proposition 14.1, we will need the following auxiliary proposition.

Lemma 14.1. *If $J(v) > 0$ for all $v \in U(0,0) \setminus 0$, then*

$$J(v) \geq \alpha \|v\|^2 \qquad \text{for some } \alpha > 0 \text{ and all } v \in U(0,0),$$

or, which is equivalent,

$$\inf\{J(v) \mid \|v\| = 1,\ v \in U(0,0)\} > 0.$$

Proof. Let v_n be a minimizing sequence of the functional $J(v)$ on the sphere $\{\|v\| = 1\} \cap U(0,0)$. Closed balls in Hilbert spaces are weakly compact, thus we can find a subsequence weakly converging in the unit ball and preserve the notation v_n for its terms, so that

$$\begin{aligned} &v_n \to \widehat{v} \text{ weakly as } n \to \infty, \qquad \|\widehat{v}\| \leq 1, \quad \widehat{v} \in U(0,0), \\ &J(v_n) \to \inf\{J(v) \mid \|v\| = 1,\ v \in U(0,0)\}, \qquad n \to \infty. \end{aligned} \tag{14.4}$$

We have

$$J(v_n) = \frac{1}{2} + \frac{1}{2} \int_0^{t_1} \langle Q x_n(\tau), x_n(\tau) \rangle \, d\tau.$$

Since the controls converge weakly, then the corresponding trajectories converge strongly:

$$x_n(\cdot) \to x_{\widehat{v}}(\cdot), \qquad n \to \infty,$$

thus

$$J(v_n) \to \frac{1}{2} + \frac{1}{2} \int_0^{t_1} \langle Q x_{\widehat{v}}(\tau), x_{\widehat{v}}(\tau) \rangle \, d\tau, \qquad n \to \infty.$$

In view of (14.4), the infimum in question is equal to

$$\frac{1}{2} + \frac{1}{2} \int_0^{t_1} \langle Q x_{\widehat{v}}(\tau), x_{\widehat{v}}(\tau) \rangle \, d\tau = \frac{1}{2} \left(1 - \|\widehat{v}\|^2\right) + J(\widehat{v}) > 0.$$

□

Now we prove Proposition 14.1

Proof. (1) By contradiction, suppose that there exists $v \in U(0,0)$ such that $J(v) < 0$. Take any $u \in U(x_0, x_1)$, then $u + sv \in U(x_0, x_1)$ for any $s \in \mathbb{R}$.

Let $y(t)$, $t \in [0, t_1]$, be the solution to the Cauchy problem

$$\dot{y} = Ay + Bv, \qquad y(0) = 0,$$

and

$$J(u, v) = \frac{1}{2} \int_0^{t_1} \langle u(\tau), v(\tau) \rangle + \langle Qx(\tau), y(\tau) \rangle \, d\tau.$$

Then the quadratic functional J on the family of controls $u + sv$, $s \in \mathbb{R}$, is computed as follows:

$$J(u + sv) = J(u) + 2sJ(u, v) + s^2 J(v).$$

Since $J(v) < 0$, then $J(u + sv) \to -\infty$ as $s \to \infty$. The contradiction with hypothesis (14.3) finishes the proof of item (1) of this proposition.

(2) We have

$$J(u) = \frac{1}{2} \|u\|^2 + \frac{1}{2} \int_0^{t_1} \langle Qx(\tau), x(\tau) \rangle \, d\tau.$$

The norm $\|u\|$ is lower semicontinuous in the weak topology on L_2^m, and the functional $\int_0^{t_1} \langle Qx(\tau), x(\tau) \rangle \, d\tau$ is weakly continuous on L_2^m. Thus $J(u)$ is weakly lower semicontinuous on L_2^m. Since balls are weakly compact in L_2^m and the affine subspace $U(x_0, x_1)$ is weakly compact, it is enough to prove that $J(u) \to \infty$ when $u \to \infty$, $u \in U(x_0, x_1)$.

Take any control $u \in U(x_0, x_1)$. Then for any $v \in U(0,0) \setminus 0$, the control $u + v$ belongs to $U(x_0, x_1)$ and

$$J(u + v) = J(u) + 2\|v\| J\left(u, \frac{v}{\|v\|}\right) + J(v).$$

Denote $J(u) = C_0$. Further, $\left| J\left(u, \frac{v}{\|v\|}\right) \right| \leq C_1 = \text{const}$ for all $v \in U(0,0) \setminus 0$. Finally, by Lemma 14.1, $J(v) \geq \alpha \|v\|^2$, $\alpha > 0$, for all $v \in U(0,0) \setminus 0$. Consequently,

$$J(u + v) \geq C_0 - 2\|v\| C_1 + \alpha \|v\|^2 \to \infty, \qquad v \to \infty, \ v \in U(0,0).$$

Item (2) of this proposition follows. □

So we reduced the question of existence of optimal controls in linear-quadratic problems to the study of the restriction $J|_{U(0,0)}$. We will consider this restriction in detail later.

14.3 Extremals

Now we write PMP for linear-quadratic problems. The control-dependent Hamiltonian is

$$h_u(\xi, x) = \xi A x + \xi B u - \frac{\nu}{2}(\|u\|^2 + \langle Qx, x\rangle), \qquad x \in \mathbb{R}^n, \ \xi \in \mathbb{R}^{n*}.$$

Consider first the abnormal case:

$$\nu = 0.$$

By PMP, adjoint vector along an extremal satisfies the ODE $\dot{\xi} = -\xi A$, thus $\xi(t) = \xi(0)e^{-tA}$. The maximality condition

$$\xi(t)Bu(t) = \max_{u \in \mathbb{R}^n} \xi(t)Bu \tag{14.5}$$

implies that

$$0 \equiv \xi(t)B = \xi(0)e^{-tA}B.$$

We differentiate this identity $n-1$ times, take into account the controllability condition (14.2) and obtain $\xi(0) = 0$. This contradicts PMP, thus there are no abnormal extremals.

In the sequel we consider the normal case: $\nu \neq 0$, thus we can assume

$$\nu = 1.$$

Then the control-dependent Hamiltonian takes the form

$$h_u(\xi, x) = \xi A x + \xi B u - \frac{1}{2}(\|u\|^2 + \langle Qx, x\rangle), \qquad x \in \mathbb{R}^n, \ \xi \in \mathbb{R}^{n*}.$$

The term $\xi B u - \frac{1}{2}\|u\|^2$ depending on u has a unique maximum in $u \in \mathbb{R}^m$ at the point where

$$\frac{\partial h_u}{\partial u} = \xi B - u^* = 0,$$

thus

$$u = B^* \xi^*.$$

So the maximized Hamiltonian is

$$\begin{aligned} H(\xi, x) &= \max_{u \in \mathbb{R}^m} h_u(\xi, x) = \xi A x - \frac{1}{2}\langle Qx, x\rangle + \frac{1}{2}|B^*\xi^*|^2 \\ &= \xi A x - \frac{1}{2}\langle Qx, x\rangle + \frac{1}{2}|B\xi|^2. \end{aligned}$$

The Hamiltonian function $H(\xi, x)$ is smooth, thus extremals are solutions of the corresponding Hamiltonian system

$$\begin{cases} \dot{x} = Ax + BB^*\xi^*, \\ \dot{\xi} = x^*Q - \xi A. \end{cases}$$

14.4 Conjugate points

Now we study conditions of existence and uniqueness of optimal controls depending upon the terminal time. So we write the cost functional to be minimized as follows:

$$J_t(u) = \frac{1}{2}\int_0^t |u(\tau)|^2 + \langle Qx(\tau), x(\tau)\rangle \, d\tau.$$

Denote

$$U_t(0,0) = \{u \in L_2^m[0,t] \mid x(t,u,x_0) = x_1\},$$
$$\mu(t) \stackrel{\text{def}}{=} \inf\{J_t(u) \mid u \in U_t(0,0), \ \|u\| = 1\}. \tag{14.6}$$

We showed in Proposition 14.1 that if $\mu(t) > 0$ then the problem has solution for any boundary conditions, and if $\mu(t) < 0$ then there are no solutions for any boundary conditions. The case $\mu(t) = 0$ is doubtful. Now we study properties of the function $\mu(t)$ in detail.

Proposition 14.2. (1) *The function $t \mapsto \mu(t)$ is monotone nonincreasing and continuous.*

(2)

$$1 \geq 2\mu(t) \geq 1 - \frac{t^2}{2} e^{2t\|A\|}\|B\|^2\|Q\|. \tag{14.7}$$

(3) *If $1 > 2\mu(t)$, then the infimum in* (14.6) *is attained, i.e., it is minimum.*

Proof. (3) Denote

$$I_t(u) = \frac{1}{2}\int_0^t \langle Qx(\tau), x(\tau)\rangle \, d\tau,$$

the functional $I_t(u)$ is weakly continuous on L_2^m. Notice that

$$J_t(u) = \frac{1}{2} + I_t(u) \quad \text{on the sphere } \|u\| = 1.$$

Take a minimizing sequence of the functional $I_t(u)$ on the sphere $\{\|u\| = 1\} \cap U_t(0,0)$. Since the ball $\{\|u\| \leq 1\}$ is weakly compact, we can find a weakly converging subsequence:

$$u_n \to \widehat{u} \text{ weakly as } n \to \infty, \qquad \|\widehat{u}\| \leq 1, \quad \widehat{u} \in U_t(0,0),$$
$$I_t(u_n) \to I_t(\widehat{u}) = \inf\{I_t(u) \mid \|u\| = 1, \ u \in U_t(0,0)\}, \qquad n \to \infty.$$

If $\widehat{u} = 0$, then $I_t(\widehat{u}) = 0$, thus $\mu(t) = \frac{1}{2}$, which contradicts hypothesis of item (3).

So $\widehat{u} \neq 0$, $I_t(\widehat{u}) < 0$, and $I_t\left(\frac{\widehat{u}}{\|\widehat{u}\|}\right) \leq I_t(\widehat{u})$. Thus $\|\widehat{u}\| = 1$, and $J_t(u)$ attains minimum on the sphere $\{\|u\| = 1\} \cap U_t(0,0)$ at the point $\widehat{u}$.

(2) Let $\|u\| = 1$ and $x_0 = 0$. By Cauchy's formula,

$$x(t) = \int_0^t e^{(t-\tau)A} Bu(\tau)\, d\tau,$$

thus

$$|x(t)| \leq \int_0^t e^{(t-\tau)\|A\|} \|B\| \cdot |u(\tau)| \, d\tau$$

by Cauchy-Schwartz inequality

$$\begin{aligned} &\leq \|u\| \left(\int_0^t e^{(t-\tau)2\|A\|} \|B\|^2 \, d\tau \right)^{1/2} \\ &= \left(\int_0^t e^{(t-\tau)2\|A\|} \|B\|^2 \, d\tau \right)^{1/2}. \end{aligned}$$

We substitute this estimate of $x(t)$ into J_t and obtain the second inequality in (14.7).

The first inequality in (14.7) is obtained by considering a weakly converging sequence $u_n \to 0$, $n \to \infty$, in the sphere $\|u_n\| = 1$, $u_n \in U_t(0,0)$.

(1) Monotonicity of $\mu(t)$. Take any $\hat{t} > t$. Then the space $U_t(0,0)$ is isometrically embedded into $U_{\hat{t}}(0,0)$ by extending controls $u \in U_t(0,0)$ by zero:

$$\begin{aligned} &u \in U_t(0,0) \quad \Rightarrow \quad \widehat{u} \in U_{\hat{t}}(0,0), \\ &\widehat{u}(\tau) = \begin{cases} u(\tau), & \tau \leq t, \\ 0, & \tau > t. \end{cases} \end{aligned}$$

Moreover,

$$J_{\hat{t}}(\widehat{u}) = J_t(u).$$

Thus

$$\begin{aligned} \mu(t) &= \inf\{J_t(u) \mid u \in U_t(0,0), \|u\| = 1\} \\ &\geq \inf\{J_{\hat{t}}(u) \mid u \in U_{\hat{t}}(0,0), \|u\| = 1\} = \mu(\hat{t}). \end{aligned}$$

Continuity of $\mu(t)$: we show separately continuity from the right and from the left.

Continuity from the right. Let $t_n \searrow t$. We can assume that $\mu(t_n) < \frac{1}{2}$ (otherwise $\mu(t_n) = \mu(t) = \frac{1}{2}$), thus minimum in (14.6) is attained:

$$\mu(t_n) = \frac{1}{2} + I_{t_n}(u_n), \qquad u_n \in U_{t_n}(0,0), \quad \|u_n\| = 1.$$

Extend the functions $u_n \in L_2^m[0, t_n]$ to the segment $[0, t]$ by zero. Choosing a weakly converging subsequence in the unit ball, we can assume that

$$u_n \to u \text{ weakly as } n \to \infty, \qquad u \in U_t(0,0), \quad \|u_n\| \leq 1,$$

thus

$$I_{t_n}(u_n) \to I_t(u) \geq \inf\{I_t(v) \mid v \in U_t(0,0), \quad \|v\| = 1\}, \quad t_n \searrow t.$$

Then

$$\mu(t) \le \frac{1}{2} + \lim_{t_n \searrow t} I_{t_n}(u_n) = \lim_{t_n \searrow t} \mu(t_n).$$

By monotonicity of μ,

$$\mu(t) = \lim_{t_n \searrow t} \mu(t_n),$$

i.e., continuity from the right is proved.

Continuity from the left. We can assume that $\mu(t) < \frac{1}{2}$ (otherwise $\mu(\tau) = \mu(t) = \frac{1}{2}$ for $\tau < t$). Thus minimum in (14.6) is attained:

$$\mu(t) = \frac{1}{2} + I_t(\widehat{u}), \qquad \widehat{u} \in U_t(0,0), \quad \|\widehat{u}\| = 1.$$

For the trajectory

$$\widehat{x}(\tau) = x(\tau, \widehat{u}, 0),$$

we have

$$\widehat{x}(\tau) = \int_0^\tau e^{(\tau-\theta)A} B\widehat{u}(\theta)\, d\theta.$$

Denote

$$\alpha(\varepsilon) = \| \, \widehat{u}|_{[0,\varepsilon]} \, \|$$

and notice that

$$\alpha(\varepsilon) \to 0, \qquad \varepsilon \to 0.$$

Denote the ball

$$B_\delta = \{u \in L_2^m \mid \|u\| \le \delta, \ u \in U(0,0)\}.$$

Obviously,

$$x(\varepsilon, B_{\alpha(\varepsilon)}, 0) \ni \widehat{x}(\varepsilon).$$

The mapping $u \mapsto x(\varepsilon, u(\cdot), 0)$ from L_2^m to $\mathbb{R}^n$ is linear, and the system $\dot{x} = Ax + Bu$ is controllable, thus $x(\varepsilon, B_{\alpha(\varepsilon)}, 0)$ is a convex full-dimensional set in $\mathbb{R}^n$ such that the positive cone generated by this set is the whole $\mathbb{R}^n$. That is why

$$x(\varepsilon, 2B_{\alpha(\varepsilon)}, 0) = 2x(\varepsilon, B_{\alpha(\varepsilon)}, 0) \supset O_{x(\varepsilon, B_{\alpha(\varepsilon)}, 0)}$$

for some neighborhood $O_{x(\varepsilon, B_{\alpha(\varepsilon)}, 0)}$ of the set $x(\varepsilon, B_{\alpha(\varepsilon)}, 0)$. Further, there exists an instant $t_\varepsilon > \varepsilon$ such that

$$\widehat{x}(t_\varepsilon) \in x(\varepsilon, 2B_{\alpha(\varepsilon)}, 0),$$

consequently,

$$\widehat{x}(t_\varepsilon) = x(\varepsilon, v_\varepsilon, 0), \qquad \|v_\varepsilon\| \le 2\alpha(\varepsilon).$$

Consider the following family of controls that approximate $\widehat{u}$:

$$u_\varepsilon(\tau) = \begin{cases} v_\varepsilon(\tau), & 0 \le \tau \le t_\varepsilon, \\ \widehat{u}(\tau + t_\varepsilon - \varepsilon), & t_\varepsilon < \tau \le t + \varepsilon - t_\varepsilon. \end{cases}$$

We have

$$u_\varepsilon \in U_{t+\varepsilon-t_\varepsilon}(0,0),$$
$$\|\widehat{u} - u_\varepsilon\| \to 0, \qquad \varepsilon \to 0.$$

But $t+\varepsilon-t_\varepsilon < t$ and μ is nonincreasing, thus it is continuous from the left.

Continuity from the right was already proved, hence μ is continuous. □

Now we prove that the function μ can have not more than one root.

Proposition 14.3. *If $\mu(t) = 0$ for some $t > 0$, then $\mu(\tau) < 0$ for all $\tau > t$.*

Proof. Let $\mu(t) = 0$, $t > 0$. By Proposition 14.2, infimum in (14.6) is attained at some control $\widehat{u} \in U_t(0,0)$, $\|\widehat{u}\| = 1$:

$$\begin{aligned} \mu(t) &= \min\{J_t(u) \mid u \in U_t(0,0),\ \|u\| = 1\} \\ &= J_t(\widehat{u}) = 0. \end{aligned}$$

Then

$$J_t(u) \geq J_t(\widehat{u}) = 0 \quad \forall u \in U_t(0,0),$$

i.e., the control $\widehat{u}$ is optimal, thus it satisfies PMP. There exists a solution $(\xi(\tau), x(\tau))$, $\tau \in [0,t]$, of the Hamiltonian system

$$\begin{cases} \dot{\xi} = x^* Q - \xi A, \\ \dot{x} = Ax + BB^*\xi, \end{cases}$$

with the boundary conditions

$$x(0) = x(t) = 0,$$

and

$$u(\tau) = B^*\xi^*(\tau), \qquad \tau \in [0,t].$$

We proved that for any root t of the function μ, any control $u \in U_t(0,0)$, $\|u\| = 1$, with $J_t(u) = 0$ satisfies PMP.

Now we prove that $\mu(\tau) < 0$ for all $\tau > t$. By contradiction, suppose that the function μ vanishes at some instant $t' > t$. Since μ is monotone, then

$$\mu|_{[t,t']} \equiv 0.$$

Consequently, the control

$$u'(\tau) = \begin{cases} \widehat{u}(\tau), & \tau \leq t, \\ 0, & \tau \in [t,t'], \end{cases}$$

satisfies the conditions:

$$u' \in U_{t'}(0,0), \quad \|u'\| = 1,$$
$$J_{t'}(u') = 0.$$

Thus u' satisfies PMP, i.e.,

$$u'(\tau)B^*{\xi^*}'(\tau), \qquad \tau \in [0,t'],$$

is an analytic function. But $u'|_{[t,t']} \equiv 0$, thus $u' \equiv 0$, a contradiction with $\|u'\| = 1$. □

It would be nice to have a way to solve the equation $\mu(t) = 0$ without performing the minimization procedure in (14.6). This can be done in terms of the following notion.

Definition 14.1. A point $t > 0$ is *conjugate* to 0 for the linear-quadratic problem in question if there exists a nontrivial solution $(\xi(\tau), x(\tau))$ of the Hamiltonian system

$$\begin{cases} \dot{\xi} = x^*Q - \xi A, \\ \dot{x} = Ax + BB^*\xi \end{cases}$$

such that $x(0) = x(t) = 0$.

Proposition 14.4. *The function μ vanishes at a point $t > 0$ if and only if t is the closest to 0 conjugate point.*

Proof. Let $\mu(t) = 0$, $t > 0$. First of all, t is conjugate to 0, we showed this in the proof of Proposition 14.3.

Suppose that $t' > 0$ is conjugate to 0. Compute the functional $J_{t'}$ on the corresponding control $u(\tau) = B^*\xi^*(\tau)$, $\tau \in [0,t']$:

$$\begin{aligned} J_{t'}(u) &= \frac{1}{2}\int_0^{t'} \langle B^*\xi^*(\tau), B^*\xi^*(\tau)\rangle + \langle Qx(\tau), x(\tau)\rangle \, d\tau \\ &= \frac{1}{2}\int_0^{t'} \langle BB^*\xi^*(\tau), \xi^*(\tau)\rangle + \langle Qx(\tau), x(\tau)\rangle \, d\tau \\ &= \frac{1}{2}\int_0^{t'} \xi(\tau)(\dot{x}(\tau) - Ax(\tau)) + x^*(\tau)Qx(\tau) \, d\tau \\ &= \frac{1}{2}\int_0^{t'} (\xi\dot{x} + \dot{\xi}x) \, d\tau \\ &= \frac{1}{2}(\xi(t')x(t') - \xi(0)x(0)) = 0. \end{aligned}$$

Thus $\mu(t') \le J_{t'}\left(\frac{u}{\|u\|}\right) = 0$. Now the result follows since μ is nonincreasing. □

The first (closest to zero) conjugate point determines existence and uniqueness properties of optimal control in linear-quadratic problems.

Before the first conjugate point, optimal control exists and is unique for any boundary conditions (if there are two optimal controls, then their difference gives rise to a conjugate point).

At the first conjugate point, there is existence and nonuniqueness for some boundary conditions, and nonexistence for other boundary conditions.

And after the first conjugate point, the problem has no optimal solutions for any boundary conditions.

Chapter 15

Sufficient optimality conditions, Hamilton-Jacobi equation, and Dynamic Programming

15.1 Sufficient optimality conditions

Pontryagin Maximum Principle is a universal and powerful necessary optimality condition, but the theory of sufficient optimality conditions is not so complete. In this section we consider an approach to sufficient optimality conditions that generalizes fields of extremals of the Classical Calculus of Variations.

Consider the following optimal control problem:

$$\dot{q} = f_u(q), \qquad q \in M, \ u \in U, \tag{15.1}$$

$$q(0) = q_0, \ q(t_1) = q_1, \quad q_0, q_1, t_1 \text{ fixed}, \tag{15.2}$$

$$\int_0^{t_1} \varphi(q(t), u(t))\, dt \to \min . \tag{15.3}$$

The control-dependent Hamiltonian of PMP corresponding to the normal case is

$$h_u(\lambda) = \langle \lambda, f_u(q) \rangle - \varphi(q, u), \qquad \lambda \in T^*M, \ q = \pi(\lambda) \in M, \ u \in U.$$

Assume that the maximized Hamiltonian

$$H(\lambda) = \max_{u \in U} h_u(\lambda) \tag{15.4}$$

is defined and smooth on T^*M. We can assume smoothness of H on an open domain $O \subset T^*M$ and modify respectively the subsequent results. But for simplicity of exposition we prefer to take $O = T^*M$. Then trajectories of the Hamiltonian system

$$\dot{\lambda} = \vec{H}(\lambda)$$

are extremals of problem (15.1)–(15.3). We assume that the Hamiltonian vector field $\vec{H}$ is complete.

Fix an arbitrary smooth function

$$a \in C^\infty(M).$$

Then the graph of differential da is a smooth submanifold in T^*M:

$$\mathcal{L}_0 = \{d_q a \mid q \in M\} \subset T^*M,$$
$$\dim \mathcal{L}_0 = \dim M = n.$$

Translations of $\mathcal{L}_0$ by the flow of the Hamiltonian vector field

$$\mathcal{L}_t = e^{t\vec{H}}(\mathcal{L}_0)$$

are smooth n-dimensional submanifolds in T^*M, and the graph of the mapping $t \mapsto \mathcal{L}_t$,

$$\mathcal{L} = \{(\lambda, t) \mid \lambda \in \mathcal{L}_t,\ 0 \le t \le t_1\} \subset T^*M \times \mathbb{R}$$

is a smooth $(n+1)$-dimensional submanifold in $T^*M \times \mathbb{R}$.

Consider the 1-form

$$s - H\,dt \in \Lambda^1(T^*M \times \mathbb{R}).$$

Recall that s is the tautological 1-form on T^*M, $s_\lambda = \lambda \circ \pi_*$, and its differential is the canonical symplectic structure on T^*M, $ds = \sigma$. In mechanics, the form $s - H\,dt = p\,dq - H\,dt$ is called the *integral invariant of Poincare-Cartan* on the extended phase space $T^*M \times \mathbb{R}$.

Proposition 15.1. *The form* $(s - H\,dt)|_{\mathcal{L}}$ *is exact.*

Proof. First we prove that the form is closed:

$$0 = d(s - H\,dt)|_{\mathcal{L}} = (\sigma - dH \wedge dt)|_{\mathcal{L}}. \qquad (15.5)$$

(1) Fix $\mathcal{L}_t = \mathcal{L} \cap \{t = \text{const}\}$ and consider restriction of the form $\sigma - dH \wedge dt$ to $\mathcal{L}_t$. We have

$$(\sigma - dH \wedge dt)|_{\mathcal{L}_t} = \sigma|_{\mathcal{L}_t}$$

since $dt|_{\mathcal{L}_t} = 0$. Recall that $\left(e^{t\vec{H}}\right)^* \sigma = \sigma$, thus

$$\sigma|_{\mathcal{L}_t} = \left(\left(e^{t\vec{H}}\right)^* \sigma\right)\Big|_{\mathcal{L}_0} = \sigma|_{\mathcal{L}_0} = ds|_{\mathcal{L}_0}.$$

But $s|_{\mathcal{L}_0} = d(a \circ \pi)|_{\mathcal{L}_0}$, hence

$$ds|_{\mathcal{L}_0} = d \circ d(a \circ \pi)|_{\mathcal{L}_0} = 0.$$

We proved that $(\sigma - dH \wedge dt)_{\mathcal{L}_t} = 0$.

(2) The manifold $\mathcal{L}$ is the image of the smooth mapping

$$(\lambda, t) \mapsto \left(e^{t\vec{H}}\lambda, t\right),$$

thus the tangent vector to $\mathcal{L}$ transversal to $\mathcal{L}_t$ is

$$\vec{H}(\lambda) + \frac{\partial}{\partial t} \in T_{(\lambda,t)}\mathcal{L}.$$

So

$$T_{(\lambda,t)}\mathcal{L} = T_{(\lambda,t)}\mathcal{L}_t \oplus \mathbb{R}\left(\vec{H}(\lambda) + \frac{\partial}{\partial t}\right).$$

To complete the proof, we insert the vector $\vec{H}(\lambda) + \frac{\partial}{\partial t}$ as the first argument to $\sigma - dH \wedge dt$ and show that the result is equal to zero. We have:

$$i_{\vec{H}}\sigma = -dH, \quad i_{\frac{\partial}{\partial t}}\sigma = 0,$$
$$i_{\vec{H}}(dH \wedge dt) = \underbrace{\left(i_{\vec{H}}dH\right)}_{=0} \wedge dt - dH \wedge \underbrace{\left(i_{\vec{H}}dt\right)}_{=0} = 0,$$
$$i_{\frac{\partial}{\partial t}}(dH \wedge dt) = \underbrace{\left(i_{\frac{\partial}{\partial t}}dH\right)}_{=0} \wedge dt - dH \wedge \underbrace{\left(i_{\frac{\partial}{\partial t}}dt\right)}_{=1} = -dH,$$

consequently,

$$i_{\vec{H}+\frac{\partial}{\partial t}}(\sigma - dH \wedge dt) = -dH + dH = 0.$$

We proved that the form $(s - H\,dt)|_{\mathcal{L}}$ is closed.

(3) Now we show that it is exact, i.e.,

$$\int_\gamma s - H\,dt = 0 \tag{15.6}$$

for any closed curve

$$\gamma \;:\; s \mapsto (\lambda(s), t(s)) \in \mathcal{L}, \qquad s \in [0,1].$$

The curve γ is homotopic to the curve

$$\gamma_0 \;:\; s \mapsto (\lambda(s), 0) \in \mathcal{L}_0, \qquad s \in [0,1].$$

Since the form $(s - H\,dt)|_{\mathcal{L}}$ is closed, Stokes' theorem yields that

$$\int_\gamma s - H\,dt = \int_{\gamma_0} s - H\,dt.$$

But the integral over the closed curve $\gamma_0 \subset \mathcal{L}_0$ is easily computed:

$$\int_{\gamma_0} s - H\,dt = \int_{\gamma_0} s = \int_{\gamma_0} da = 0.$$

Equality (15.5) follows, i.e., the form $(s - H\,dt)|_{\mathcal{L}}$ is exact. □

Now we prove a statement that gives sufficient conditions for optimality in problem (15.1)–(15.3).

Theorem 15.1. *Assume that the restriction of projection $\pi|_{\mathcal{L}_t}$ is a diffeomorphism for any $t \in [0, t_1]$. Then for any $\lambda_0 \in \mathcal{L}_0$, the normal extremal trajectory*

$$\widehat{q}(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad 0 \le t \le t_1,$$

realizes a strict minimum of the cost functional $\int_0^{t_1} \varphi(q(t), u(t))\, dt$ among all admissible trajectories $q(t)$, $0 \le t \le t_1$, of system (15.1) *with the same boundary conditions:*

$$q(0) = \widehat{q}(0), \quad q(t_1) = \widehat{q}(t_1). \tag{15.7}$$

Remarks. (1) Under the hypotheses of this theorem, no check of existence of optimal control is required.

(2) If all assumptions (smoothness of H, extendibility of trajectories of $\vec{H}$ to the time segment $[0, t_1]$, diffeomorphic property of $\pi|_{\mathcal{L}_t}$) hold in a proper open domain $O \subset T^*M$, then the statement can be modified to give local optimality of $\widehat{q}(\cdot)$ in $\pi(O)$. These modifications are left to the reader.

Now we prove Theorem 15.1.

Proof. The curve $\widehat{q}(t)$ is projection of the normal extremal

$$\widehat{\lambda}_t = e^{t\vec{H}}(\lambda_0).$$

Let $\widehat{u}(t)$ be an admissible control that maximizes the Hamiltonian along this extremal:

$$H(\widehat{\lambda}_t) = h_{\widehat{u}(t)}(\widehat{\lambda}_t).$$

On the other hand, let $q(t)$ be an admissible trajectory of system (15.1) generated by a control $u(t)$ and satisfying the boundary conditions (15.7). We compare costs of the pairs $(\widehat{q}, \widehat{u})$ and (q, u).

Since $\pi : \mathcal{L}_t \to M$ is a diffeomorphism, the trajectory $\{q(t) \mid 0 \le t \le t_1\} \subset M$ can be lifted to a smooth curve $\{\lambda(t) \mid 0 \le t \le t_1\} \subset T^*M$:

$$\forall t \in [0, t_1] \quad \exists!\ \lambda(t) \in \mathcal{L}_t \quad \text{such that} \quad \pi(\lambda(t)) = q(t).$$

Then

$$\begin{aligned} \int_0^{t_1} \varphi(q(t), u(t))\, dt &= \int_0^{t_1} \langle \lambda(t), f_{u(t)}(q(t)) \rangle - h_{u(t)}(\lambda(t))\, dt \\ &\ge \int_0^{t_1} \langle \lambda(t), \dot{q}(t) \rangle - H(\lambda(t))\, dt \\ &= \int_0^{t_1} \langle s_{\lambda(t)}, \dot{\lambda}(t) \rangle - H(\lambda(t))\, dt \\ &= \int_\gamma s - H\, dt, \end{aligned} \tag{15.8}$$

where

$$\gamma \,:\, t \mapsto (\lambda(t), t) \in \mathcal{L}, \qquad t \in [0, t_1].$$

By Proposition 15.1, the form $(s - H\,dt)|_{\mathcal{L}}$ is exact. Then integral of $(s - H\,dt)|_{\mathcal{L}}$ along a curve depends only upon endpoints of the curve. The curves γ and

$$\widehat{\gamma} \,:\, t \mapsto (\widehat{\lambda}_t, t) \in \mathcal{L}, \quad \widehat{\lambda}_t = e^{t\vec{H}}(\lambda_0), \qquad t \in [0, t_1],$$

have the same endpoints, thus

$$\begin{aligned}\int_\gamma s - H\,dt &= \int_{\widehat{\gamma}} s - H\,dt = \int_0^{t_1} \langle \widehat{\lambda}_t, \dot{\widehat{q}}(t)\rangle - H(\widehat{\lambda}_t)\,dt \\ &= \int_0^{t_1} \langle \widehat{\lambda}_t, f_{\widehat{u}(t)}(\widehat{q}(t))\rangle - h_{\widehat{u}(t)}(\widehat{\lambda}_t)\,dt \\ &= \int_0^{t_1} \varphi(\widehat{q}(t), \widehat{u}(t))\,dt.\end{aligned}$$

So

$$\int_0^{t_1} \varphi(q(t), u(t))\,dt \geq \int_0^{t_1} \varphi(\widehat{q}(t), \widehat{u}(t))\,dt, \tag{15.9}$$

i.e., the trajectory $\widehat{q}(t)$ is optimal.

It remains to prove that the minimum of the pair $(\widehat{q}(t), \widehat{u}(t))$ is strict, i.e, that inequality (15.9) is strict.

For a fixed point $q \in M$, write cotangent vectors as $\lambda = (p, q)$, where p are coordinates of a covector λ in T_q^*M. The control-dependent Hamiltonians $h_u(p, q)$ are affine w.r.t. p, thus their maximum $H(p, q)$ is convex w.r.t. p. Any vector $\xi \in T_qM$ such that

$$\langle p, \xi\rangle = \max_{u \in U} \langle p, f_u(q)\rangle$$

defines a hyperplane of support to the epigraph of the mapping $p \mapsto H(p, q)$. Since H is smooth in p, such a hyperplane of support is unique and maximum in (15.4) is attained at a unique velocity vector:

$$H(p, q) = h_u(p, q) \text{ at unique vector } \dot{\widehat{q}} = f_u(q).$$

If $q(t) \not\equiv \widehat{q}(t)$, then inequality (15.8) becomes strict, as well as inequality (15.9) (for details on convex functions, see e.g. [13]). The theorem is proved. □

Sufficient optimality condition of Theorem 15.1 is given in terms of the manifolds $\mathcal{L}_t$, which are in their turn defined by a function a and the Hamiltonian flow of $\vec{H}$. One can prove optimality of a normal extremal trajectory $\widehat{q}(t)$, $t \in [0, t_1]$, if one succeeds to find an appropriate function $a \in C^\infty(M)$ for which the projections $\pi \,:\, \mathcal{L}_t \to M$, $t \in [0, t_1]$, are diffeomorphisms.

For $t = 0$ the projection $\pi \,:\, \mathcal{L}_0 \to M$ is a diffeomorphism. So for small $t > 0$ any function $a \in C^\infty(M)$ provides manifolds $\mathcal{L}_t$ well projected to M, at least if we restrict ourselves by a compact $K \Subset M$. Thus the sufficient optimality condition for small pieces of extremal trajectories follows.

Corollary 15.1. *For any compact $K \Subset M$ that contains a normal extremal trajectory*

$$\widehat{q}(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad 0 \le t \le t_1,$$

there exists $t_1' \in (0, t_1]$ such that the trajectory

$$\widehat{q}(t), \qquad 0 \le t \le t_1',$$

is optimal w.r.t. all trajectories contained in K.

In many problems, one can choose a sufficiently large compact $K \supset \widehat{q}$ such that the functional J is separated from below from zero on all trajectories leaving K (this is the case, e.g., if $\varphi(q,u) > 0$). Then small pieces of $\widehat{q}$ are globally optimal.

15.2 Hamilton-Jacobi equation

Suppose that conditions of Theorem 15.1 are satisfied. As we showed in the proof of this theorem, the form $(s - H\,dt)|_{\mathcal{L}}$ is exact, thus it coincides with differential of some function:

$$(s - H\,dt)|_{\mathcal{L}} = dg, \qquad g : \mathcal{L} \to \mathbb{R}. \tag{15.10}$$

Since the projection $\pi : \mathcal{L}_t \to M$ is one-to-one, we can identify $(\lambda, t) \in \mathcal{L}_t \times \mathbb{R} \subset \mathcal{L}$ with $(q,t) \in M \times \mathbb{R}$ and define g as a function on $M \times \mathbb{R}$:

$$g = g(q,t).$$

In order to understand the meaning of the function g for our optimal control problem, consider an extremal

$$\widehat{\lambda}_t = e^{t\vec{H}}\lambda_0$$

and the curve

$$\widehat{\gamma} \subset \mathcal{L}, \quad \widehat{\gamma} : t \mapsto (\widehat{\lambda}_t, t),$$

as in the proof of Theorem 15.1. Then

$$\int_{\widehat{\gamma}} s - H\,dt = \int_0^{t_1} \varphi(\widehat{q}(\tau), \widehat{u}(\tau))\,d\tau, \tag{15.11}$$

where $\widehat{q}(t) = \pi(\widehat{\lambda}_t)$ is an extremal trajectory and $\widehat{u}(t)$ is the control that maximizes the Hamiltonian $h_u(\lambda)$ along $\widehat{\lambda}_t$. Equalities (15.10) and (15.11) mean that

$$g(\widehat{q}(t), t) = g(q_0, 0) + \int_0^t \varphi(\widehat{q}(\tau), \widehat{u}(\tau))\,d\tau,$$

i.e., $g(q,t) - g(q_0, 0)$ is the optimal cost of motion between points q_0 and q for the time t. Initial value for g can be chosen of the form

$$g(q_0, 0) = a(q_0), \qquad q_0 \in M. \tag{15.12}$$

Indeed, at $t = 0$ definition (15.11) of the function g reads

$$dg|_{t=0} = (s - H\,dt)|_{\mathcal{L}_0} = s|_{\mathcal{L}_0} = da,$$

which is compatible with (15.12).

We can rewrite equation (15.10) as a partial differential equation on g. In local coordinates on M and T^*M, we have

$$q = x \in M, \quad \lambda = (\xi, x) \in T^*M, \quad g = g(x,t).$$

Then equation (15.10) reads

$$(\xi\,dx - H(\xi,x)\,dt)|_{\mathcal{L}} = dg(x,t),$$

i.e.,

$$\begin{cases} \frac{\partial g}{\partial x} = \xi, \\ \frac{\partial g}{\partial t} = -H(\xi,x). \end{cases}$$

This system can be rewritten as a single first order nonlinear PDE:

$$\frac{\partial g}{\partial t} + H\left(\frac{\partial g}{\partial x}, x\right) = 0, \tag{15.13}$$

which is called *Hamilton-Jacobi equation*. We showed that the optimal cost $g(x,t)$ satisfies Hamilton-Jacobi equation (15.13) with initial condition (15.12).

Characteristic equations of PDE (15.13) have the form

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial \xi}, \\ \dot{\xi} = -\frac{\partial H}{\partial x}, \\ \dfrac{d}{dt} g(x(t),t) = \xi\dot{x} - H. \end{cases}$$

The first two equations form the Hamiltonian system $\dot{\lambda} = \vec{H}(\lambda)$ for normal extremals. Thus solving our optimal control problem (15.1)–(15.3) leads to the method of characteristics for the Hamilton-Jacobi equation for optimal cost.

15.3 Dynamic programming

One can derive the Hamilton-Jacobi equation for optimal cost directly, without Pontryagin Maximum Principle, due to an idea going back to Huygens and constituting a basis for Bellman's method of *Dynamic Programming*, see [6]. For this, it is necessary to assume that the optimal cost $g(q,t)$ exists and is C^1-smooth.

Let an optimal trajectory steer a point q_0 to a point q for a time t. Apply a constant control u on a time segment $[t, t+\delta t]$ and denote the trajectory starting at the point q by $q_u(\tau)$, $\tau \in [t, t+\delta t]$. Since $q_u(t+\delta t)$ is the endpoint of an admissible trajectory starting at q_0, the following inequality for optimal cost holds:

$$g(q_u(t+\delta t), t+\delta t) \leq g(q,t) + \int_t^{t+\delta t} \varphi(q_u(\tau), u)\, d\tau.$$

Divide by δt:

$$\frac{1}{\delta t}(g(q_u(t+\delta t), t+\delta t) - g(q,t)) \leq \frac{1}{\delta t}\int_t^{t+\delta t} \varphi(q_u(\tau), u)\, d\tau$$

and pass to the limit as $\delta t \to 0$:

$$\left\langle \frac{\partial g}{\partial q}, f_u(q) \right\rangle + \frac{\partial g}{\partial t} \leq \varphi(q,u).$$

So we obtain the inequality

$$\frac{\partial g}{\partial t} + h_u\left(\frac{\partial g}{\partial q}, q\right) \leq 0, \qquad u \in U. \tag{15.14}$$

Now let $(\tilde{q}(t), \tilde{u}(t))$ be an optimal pair. Let $t > 0$ be a Lebesgue point of the control $\tilde{u}$. Take any $\delta t \in (0,t)$. A piece of an optimal trajectory is optimal, thus $\tilde{q}(t-\delta t)$ is the endpoint of an optimal trajectory, as well as $\tilde{q}(t)$. So the optimal cost g satisfies the equality:

$$g(\tilde{q}(t), t) = g(\tilde{q}(t-\delta t), t-\delta t) + \int_{t-\delta t}^{t} \varphi(\tilde{q}(\tau), \tilde{u}(\tau))\, d\tau.$$

We repeat the above argument:

$$\frac{1}{\delta t}(g(\tilde{q}(t), t) - g(\tilde{q}(t-\delta t), t-\delta t)) = \frac{1}{\delta t}\int_{t-\delta t}^{t} \varphi(\tilde{q}(\tau), \tilde{u}(\tau))\, d\tau,$$

take the limit $\delta t \to 0$:

$$\frac{\partial g}{\partial t} + h_{\tilde{u}(t)}\left(\frac{\partial g}{\partial q}, q\right) = 0. \tag{15.15}$$

This equality together with inequality (15.14) means that

$$h_{\tilde{u}(t)}\left(\frac{\partial g}{\partial q}, q\right) = \max_{u \in U} h_u\left(\frac{\partial g}{\partial q}, q\right).$$

We denote

$$H(\xi, q) = \max_{u \in U} h_u(\xi, q)$$

and write (15.15) as Hamilton-Jacobi equation:

$$\frac{\partial g}{\partial t} + H\left(\frac{\partial g}{\partial q}, q\right) = 0.$$

Thus derivative of the optimal cost $\dfrac{\partial g}{\partial q}$ is equal to the impulse ξ along the optimal trajectory $\tilde{q}(t)$.

We do not touch here a huge theory on nonsmooth generalized solutions of Hamilton-Jacobi equation for smooth and nonsmooth Hamiltonians.

Appendix A

Lemma A.1. *On any smooth manifold M, there exists a function $a \in C^\infty(M)$ such that for any $N > 0$ exists a compact $K \Subset M$ for which*

$$a(q) > N \qquad \forall\, q \in M \setminus K.$$

Proof. Let e_k, $k \in \mathbb{N}$, be a partition of unity on M: the functions $e_k \in C^\infty(M)$ have compact supports $\operatorname{supp} e_k \Subset M$, which form a locally finite covering of M, and $\sum_{k=1}^\infty e_k \equiv 1$. Then the function $\sum_{k=1}^\infty k e_k$ can be taken as a. □

Now we recall and prove Proposition 2.1.

Proposition 2.1. *Let $\varphi : C^\infty(M) \to \mathbb{R}$ be a nontrivial homomorphism of algebras. Then there exists a point $q \in M$ such that $\varphi = \widehat{q}$.*

Proof. For the homomorphism $\varphi : C^\infty(M) \to \mathbb{R}$, the set

$$\operatorname{Ker} \varphi = \{ f \in C^\infty(M) \mid \varphi f = 0 \}$$

is a maximal ideal in $C^\infty(M)$. Further, for any point $q \in M$, the set of functions

$$I_q = \{ f \in C^\infty(M) \mid f(q) = 0 \}$$

is an ideal in $C^\infty(M)$. To prove the proposition, we show that

$$\operatorname{Ker} \varphi \subset I_q \tag{A.1}$$

for some $q \in M$. Then it follows that $\operatorname{Ker} \varphi = I_q$ and $\varphi = \widehat{q}$.

By contradiction, suppose that $\operatorname{Ker} \varphi \not\subset I_q$ for any $q \in M$. This means that

$$\forall\, q \in M \quad \exists\, b_q \in \operatorname{Ker} \varphi \quad \text{s. t.} \quad b_q(q) \neq 0.$$

Changing if necessary the sign of b_q, we obtain that

$$\forall\, q \in M \quad \exists\, b_q \in \operatorname{Ker} \varphi,\ O_q \subset M \quad \text{s. t.} \quad b_q|_{O_q} > 0. \tag{A.2}$$

Fix a function a given by Lemma A.1. Denote $\varphi(a) = \alpha$, then $\varphi(a - \alpha) = 0$, i.e.,

$$(a - \alpha) \in \operatorname{Ker} \varphi.$$

Moreover,

$$\exists\, K \Subset M \quad \text{s. t.} \quad a(q) - \alpha > 0 \quad \forall\, q \in M \setminus K.$$

Take a finite covering of the compact K by the neighborhoods O_q as in $(A.2)$:

$$K \subset \bigcup_{i=1}^{n} O_{q_i},$$

and let $e_0, e_1, \dots, e_n \in C^\infty(M)$ be a partition of unity subordinated to the covering of M:

$$M \setminus K, O_{q_1}, \dots, O_{q_n}.$$

Then we have a globally defined function on M:

$$c = e_0(a - \alpha) + \sum_{i=1}^{n} e_i b_{q_i} > 0.$$

Since

$$1 = \varphi\left(c \cdot \frac{1}{c}\right) = \varphi(c) \cdot \varphi\left(\frac{1}{c}\right),$$

then

$$\varphi(c) \neq 0.$$

But $c \in \operatorname{Ker} \varphi$, a contradiction. Inclusion $(A.1)$ is proved, and the proposition follows. □

Now we formulate and prove the theorem on regularity properties of composition of operators in $C^\infty(M)$, in particular, for nonautonomous vector fields or flows on M.

Proposition A.1. *Let A_t and B_t be continuous w.r.t. t families of linear continuous operators in $C^\infty(M)$. Then the composition $A_t \circ B_t$ is also continuous w.r.t. t. If in addition the families A_t and B_t are differentiable at $t = t_0$, then the family $A_t \circ B_t$ is also differentiable at $t = t_0$, and its derivative is given by the Leibniz rule:*

$$\left.\frac{d}{dt}\right|_{t_0} (A_t \circ B_t) = \left(\left.\frac{d}{dt}\right|_{t_0} A_t\right) \circ B_{t_0} + A_{t_0} \circ \left(\left.\frac{d}{dt}\right|_{t_0} B_t\right).$$

Proof. To prove the continuity, we have to show that for any $a \in C^\infty(M)$, the following expression tends to zero as $\varepsilon \to 0$:

$$(A_{t+\varepsilon} \circ B_{t+\varepsilon} - A_t \circ B_t)\, a = A_{t+\varepsilon} \circ (B_{t+\varepsilon} - B_t)\, a + (A_{t+\varepsilon} - A_t) \circ B_t a.$$

By continuity of the family A_t, the second term $(A_{t+\varepsilon} - A_t) \circ B_t a \to 0$ as $\varepsilon \to 0$. Since the family B_t is continuous, the set of functions $(B_{t+\varepsilon} - B_t)\, a$ lies in any preassigned neighborhood of zero in $C^\infty(M)$ for sufficiently small ε. For any $\varepsilon_0 > 0$, the family $A_{t+\varepsilon}$, $|\varepsilon| < \varepsilon_0$, is locally bounded, thus equicontinuous by the

Banach-Steinhaus theorem. Consequently, $A_{t+\varepsilon} \circ (B_{t+\varepsilon} - B_t)\, a \to 0$ as $\varepsilon \to 0$. Continuity of the family $A_t \circ B_t$ follows.

The differentiability and Leibniz rule follow similarly from the decomposition

$$\frac{1}{\varepsilon}\,(A_{t+\varepsilon} \circ B_{t+\varepsilon} - A_t \circ B_t)\, a = A_{t+\varepsilon} \circ \frac{1}{\varepsilon}\,(B_{t+\varepsilon} - B_t)\, a + \frac{1}{\varepsilon}\,(A_{t+\varepsilon} - A_t) \circ B_t a.$$

□

Bibliography

[1] A. A. Agrachev, R. V. Gamkrelidze, The exponential representation of flows and the chronological calculus. (Russian) *Mat. sb.*, **107** (1978), 467–532. (English translation: Math. USSR Sb. **35** (1979), 727–785.)

[2] A. A. Agrachev, R. V. Gamkrelidze, and A. V. Sarychev, Local invariants of smooth control systems. *Acta Appl. Math.*, **14** (1989), 191–237.

[3] V.I. Arnold, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics **60**, Springer-Verlag, 1989.

[4] V.I. Arnold, *Ordinary differential equations*, Springer-Verlag, 1992.

[5] V.I. Arnold, A. B. Givental, *Symplectic geometry*, Itogi Nauki, Sovr. probl. matem., Fundamentalnye napravleniya 4 (1985), VINITI, Moscow, 5–139; English translation in Encyclopaedia of Math. Scien., vol. 4, Springer.

[6] R. Bellman, *Introduction to the mathematical theory of control processes*, vol. 1, 2. New York : Academic Press, 1967.

[7] J. Hale, *Ordinary differential equations*, Robert E. Krieger Publishing Company, Inc., 1980.

[8] J.-P. Gauthier, I. A. K. Kupka, *Deterministic observation theory and applications*, Cambridge University Press, 2001.

[9] A. Isidori, *Nonlinear control systems: an introduction*, Springer-Verlag, 1985.

[10] V. Jurdjevic, *Geometric control theory*, Cambridge University Press, 1997.

[11] H. Nijmeijer, A. van der Schaft, *Nonlinear dynamical control systems*, Springer-Verlag, 1990.

[12] L. S. Pontryagin, V. G. Boltyanskij, R. V. Gamkrelidze, E. F. Mishchenko, *The mathematical theory of optimal processes*, Oxford, Pergamon Press, 1964.

[13] R. T. Rockafellar, *Convex analysis*, Princeton, Princeton University Press, 1972.

[14] W. Rudin, *Functional analysis*, McGraw-Hill, Inc., 1973.

[15] I. E. Segal, R. A. Kunze, *Integrals and operators*, Springer, 1978.

[16] E. D. Sontag, *Mathematical control theory : deterministic finite dimensional systems*, Springer-Verlag, 1990.

[17] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, Glenview, Ill. : Scott, Foresman, 1971.