Part I

General Theory

1

Introductory Chapter

1.1 Differentiable Manifolds

A *manifold* of dimension n is essentially a space that locally resembles the Euclidean space $\mathbb{R}^n$. Every point of the manifold has a neighborhood homeomorphic to an open set of $\mathbb{R}^n$, called a *chart*. The coordinates of the point are the coordinates induced by the chart. Since a point can be covered by several charts, these changes of the coordinates have to be correlated when changing from one chart to another. More precisely, we have the following definitions.

Definition 1.1.1. *Let M be a Hausdorff separated topological space. Then the pair (V, ψ) is called a chart (coordinate system) if $\psi : V \to \psi(V) \subset \mathbb{R}^n$ is a homeomorphism of the open set V in M onto an open set $\psi(V)$ of $\mathbb{R}^n$. The coordinate functions on V are defined as $x^j : V \to \mathbb{R}^n$ and $\psi(p) = (x^1(p), \dots, x^n(p))$; namely, $x^j = u^j \circ \psi$, where $u^j : \mathbb{R}^n \to \mathbb{R}$, $u^j(a_1, \dots, a_n) = a_j$, is the jth projection.*

Definition 1.1.2. *The space M is called a differentiable manifold if there is a collection of charts $\{(V_\alpha, \psi_\alpha)\}_\alpha$ such that*

(1) $V_\alpha \subset M$, $\bigcup_\alpha V_\alpha = M$ *(V_α covers M)*
(2) *if $V_\alpha \cap V_\beta \neq \emptyset$, the map*

$$\Phi_{\alpha\beta} = \psi_\alpha \circ \psi_\beta^{-1} : \psi_\beta(V_\alpha \cap V_\beta) \to \psi_\alpha(V_\alpha \cup V_\beta)$$

is smooth; i.e., the systems of coordinates overlap smoothly.

Since most of the computations in this book have a local character, we may consider that $M = \mathbb{R}^n$. The results are sometimes proved for this particular case; the extension to a general manifold case is left as an exercise for the reader.

1.2 Submanifolds

A *submanifold* is a subset of a manifold that behaves as a manifold. More precisely, we have the following definition.

Definition 1.2.1. *Consider a differentiable manifold M and let N be a subset of M. Let $f : N \to M$ be a smooth function such that*

(1) *f is one-to-one*
(2) *f is an immersion (f_* is one-to-one).*

The pair (N, f) is called a submanifold of M. A map with the properties (1) *and* (2) *is called an imbedding.*

Example 1.2.2. *Any inclusion is an imbedding. For instance, if we consider the standard inclusion $i : \mathbb{S}^2 \to \mathbb{R}^3$, then $\mathbb{S}^2$ becomes a submanifold of $\mathbb{R}^3$.*

Remark 1.2.3. *It is possible for f to be one-to-one without being an imbedding. For instance, $f : (-1, 1) \to \mathbb{R}$, $f(t) = t^3$ is one-to-one but does not have $f'(t)$ one-to-one.*

The fact that $(f_*)_p$ is one-to-one for all $p \in N$ makes possible the identification of the tangent spaces T_pN and $(f_*)_p(T_pN) \subset T_{f(p)}M$. Hence we can consider the tangent space T_pN as a subspace of the tangent space $T_{f(p)}M$.

A classical result dealing with imbedding was proved by Whitney (see [40]).

Theorem 1.2.4 (Whitney's Imbedding Theorem, 1937)**.** *Every n-dimensional manifold imbeds in $\mathbb{R}^{2n+1}$.*

1.3 Distributions

A *distribution* $\mathcal{D}$ of rank k on a manifold M assigns to each point p of M a k-dimensional subspace $\mathcal{D}_p$ of T_pM.

The distribution $\mathcal{D}$ is called *differentiable* if every point p has a neighborhood $\mathcal{V}$ and k differentiable vector fields on $\mathcal{V}$ denoted by $X_1, X_2, \dots, X_k$, which form a basis of $\mathcal{D}_q$ for all $q \in \mathcal{V}$. We shall write $\mathcal{D} = span\{X_1, \dots, X_k\}$ on $\mathcal{V}$. In future, by a distribution we will mean a differentiable distribution. k is called the *rank* of the distribution.

The distribution $\mathcal{D}$ is called *involutive* if $[X, Y] \in \mathcal{D}$ for any X, Y in $\mathcal{D}$. An *integral manifold* of the distribution $\mathcal{D}$ is a connected submanifold N of M such that

$$f_*(T_pN) = \mathcal{D}_p, \quad \forall p \in N,$$

where $f : N \to M$ is the imbedding map.

N is called the *maximal integral manifold* of $\mathcal{D}$ if there is no other integral manifold of $\mathcal{D}$ that contains N. A distribution $\mathcal{D}$ that admits a unique maximal manifold through each point is called *integrable*. The classical theorem of

Frobenius states the relationship between the aforementioned two concepts (see, for instance, [52]).

Theorem 1.3.1 (Frobenius). *A distribution $\mathcal{D}$ is involutive if and only if it is integrable.*

In sub-Riemannian geometry the negation statement is used more often: *$\mathcal{D}$ is not involutive if and only if $\mathcal{D}$ is nonintegrable.*

1.4 Integral Curves of a Vector Field

A vector field X on a manifold M can be considered as a particular case of a distribution of rank 1. Since $[X, X] = 0$, the distribution is involutive and hence integrable. The integral manifold has dimension 1 and is called the *integral curve* of X. If t is the parameter along the integral curve $c(t)$, then for any parameter value t_0 the vector $X_{c(t_0)}$ is tangent to the curve $c(t)$ at $c(t_0)$. The existence of integral curves holds locally; i.e., for any $p_0 \in M$, there is $\epsilon > 0$ such that the integral curve $c(t)$ is defined on $(-\epsilon, \epsilon)$ and $c(0) = t_0$. This assertion can be shown as in the following: If $(x^1, \ldots, x^m)$ is a local system of coordinates on M in a neighborhood $\mathcal{U}$ of p_0, then the integral curve $c(t)$ is a solution of the following ODE (ordinary differential equation) system:

$$\frac{dc^j(t)}{dt} = X^j\big(c(t)\big), \quad j = 1, \ldots, m,$$

where $X = \sum_{i=1}^m X^i \frac{\partial}{\partial x_i}$ on $\mathcal{U}$ and $c^j(t) = x^j \circ c(t)$. The fundamental theorem of local existence and uniqueness of solutions of ODEs provides the proof of our assertion.

Let X be a vector field and define $\varphi_t(p) = c(t)$, where $c(t)$ is the integral curve of X passing through p at $t = 0$. The diffeomorphisms $\varphi_t : M \to M$ form a local one-parameter group of transformations of M; i.e.,

$$\varphi_{t+s}(p) = \varphi_t\big(\varphi_s(p)\big) = \varphi_s\big(\varphi_t(p)\big), \quad \forall t, s, t+s \in (-\epsilon, \epsilon).$$

One may show that the converse is also true; i.e., any local one-parameter group of diffeomorphisms generates locally a vector field. Sometimes φ_s is regarded as the flow in the direction of the vector field X.

We shall denote by $\Gamma(\mathcal{D})$ the set of vector fields tangent to the distribution $\mathcal{D}$. This notation agrees with the notation used for the sections of a subbundle.

Consider a noninvolutive distribution $\mathcal{D}$ and two vector fields X and Y tangent to the distribution. In the following we shall show how the one-parameter group of diffeomorphisms generated by $[X, Y]$ can be written in terms of the one-parameter group of diffeomorphisms associated with the vector fields X and Y. We shall start with an example.

Let $X = \partial_{x_1} + 2x_2\partial_{x_3}$ and $Y = \partial_{x_2} - 2x_1\partial_{x_3}$ be two vector fields on $\mathbb{R}^3$. Consider the ODE system satisfied by the integral curve $c(s) = \big(x_1(s), x_2(s), x_3(s)\big)$ of X:

$$\begin{aligned}\dot{x}_1(s) &= 1\\ \dot{x}_2(s) &= 0\\ \dot{x}_3(s) &= 2x_2(s)\end{aligned}$$

with the solution

$$x(s) = x(0) + s\big(1, 0, 2x_2(0)\big),$$

where $c(0) = \big(x_1(0), x_2(0), x_3(0)\big)$ is the initial point. Then the one-parameter group of diffeomorphisms associated with X is

$$\varphi_s(x) = x + s(1, 0, 2x_2).$$

In a similar way, the flow associated with the vector field Y is

$$\psi_s(x) = x + s(0, 1, -2x_1).$$

Since $[X, Y] = -4\partial_{x_3} \neq 0$, the flows φ_s and ψ_s do not commute. We shall compute next the difference $\varphi_s \circ \psi_s - \psi_s \circ \varphi_s$.

$$\begin{aligned}(\varphi_s \circ \psi_s)(x) &= \varphi_s(x_1, x_2 + s, x_3 - 2sx_1)\\ &= (x_1, x_2 + s, x_3 - 2sx_1) + s\big(1, 0, 2(x_2 + s)\big)\\ &= \big(x_1 + s, x_2 + s, x_3 + 2s(x_2 - x_1) + 2s^2\big),\end{aligned}$$

and

$$\begin{aligned}(\psi_s \circ \varphi_s)(x) &= \psi_s(x_1 + s, x_2, x_3 + 2sx_2)\\ &= (x_1 + s, x_2, x_3 + 2sx_2) + s\big(0, 1, -2(x_1 + s)\big)\\ &= \big(x_1 + s, x_2 + s, x_3 + 2s(x_2 - x_1) - 2s^2\big).\end{aligned}$$

We note that

$$\psi_s \circ \varphi_s(x) - \varphi_s \circ \psi_s(x) = s^2(0, 0, -4) = s^2[X, Y](x).$$

Let τ_s be the flow associated with the vector field $[X_1, X_2] = -4\partial_{x_3}$. We have

$$\tau_s(x) = x + (0, 0, -4s).$$

Then the preceding relation can also be written as

$$\psi_s \circ \varphi_s(x) - \varphi_s \circ \psi_s(x) = \tau_{s^2}(x) - x.$$

In particular, when $x = 0$ we get

$$\psi_s \circ \varphi_s(0) - \varphi_s \circ \psi_s(0) = \tau_{s^2}(0).$$

We shall prove these identities in a more general case.

Proposition 1.4.1. *Let $(\psi_s)_s$ and $(\varphi)_s$ be the one-parameter groups of diffeomorphisms associated with the vector fields X and Y on a manifold M. Then for any smooth function $f \in \mathcal{F}(M)$ we have*

$$f\big(\psi_t \circ \varphi_s(x)\big) - f\big(\varphi_s \circ \psi_t(x)\big) = ts[X, Y](f)(x) + o(s^2 + t^2).$$

Proof. Let $f \in \mathcal{F}(M)$ and consider the smooth function of two variables

$$u(t, s) = f\big(\psi_t \circ \varphi_s(x)\big) - f\big(\varphi_s \circ \psi_s(x)\big).$$

The Taylor expansion of u about $(0, 0)$ is

$$\begin{aligned} u(t, s) &= \sum_{n,m\geq 0} \partial_t^n \partial_s^m u(t, s)_{|_{t=s=0}} t^n s^m \\ &= u(0, 0) + \partial_t u(t, 0)_{|_{t=0}} t + \partial_s u(0, s)_{|_{s=0}} s + \partial_t^2 u(t, 0)_{|_{t=0}} t^2 \\ &\quad + \partial_s^2 u(0, s)_{|_{s=0}} s^2 + \partial_t \partial_s u(t, s)_{|_{t=s=0}} ts + o(s^2 + t^2). \end{aligned}$$

Since $\varphi_0(x) = x$ and $\psi_0(x) = x$, we have

$$\begin{aligned} u(0, 0) &= f\big(\psi_0 \circ \varphi_0(x)\big) - f\big(\varphi_0 \circ \psi_0(x)\big) = 0 \\ u(t, 0) &= f\big(\psi_t \circ \varphi_0(x)\big) - f\big(\varphi_0 \circ \psi_s(x)\big) = f\big(\psi_t(x)\big) - f\big(\psi_t(x)\big) = 0 \\ u(0, s) &= f\big(\psi_0 \circ \varphi_s(x)\big) - f\big(\varphi_s \circ \psi_0(x)\big) = f\big(\varphi_t(x)\big) - f\big(\varphi_t(x)\big) = 0, \end{aligned}$$

and then

$$\partial_t u(t, 0)_{|_{t=0}} = 0, \qquad \partial_s u(0, s)_{|_{s=0}} = 0, \qquad \partial_t^2 u(t, 0)_{|_{t=0}} = 0, \qquad \partial_s^2 u(0, s)_{|_{s=0}} = 0.$$

It follows that

$$u(t, s) = \partial_t \partial_s u(t, s)_{|_{t=s=0}} ts + o(s^2 + t^2). \tag{1.4.1}$$

It suffices to compute the mixed derivative at $t = s = 0$. Using the definition of a vector at a point we have

$$\partial_s f\big(\psi_s \circ \varphi_t(x)\big)_{|_{s=0}} = \partial_s f\big(\psi_s(\varphi_t(x))\big)_{|_{s=0}} = (Xf)(\varphi_t(x)) = g\big(\varphi_t(x)\big),$$

where $g = Xf$. Then

$$\partial_t \partial_s f\big(\psi_s(\varphi_t(x))\big)_{|_{t=s=0}} = \partial_t g\big(\varphi_t(x)\big)_{|_{t=0}} = (Yg)(x) = YX(f)(x).$$

Similarly we obtain

$$\partial_s \partial_t f\big(\varphi_t \circ \psi_s(x)\big)_{|_{t=s=0}} = XY(f)(x).$$

Using (1.4.1) yields

$$u(t, s) = ts[Y, X](f)(x) + o(s^2 + t^2).$$ ■

When $s = t$ we obtain the following consequence.

Corollary 1.4.2. *In the hypothesis of Proposition* 1.4.1 *we have*

$$f\big(\psi_s \circ \varphi_s(x)\big) - f\big(\varphi_s \circ \psi_s(x)\big) = s^2[Y, X](f)(x) + o(s^2).$$

Lemma 1.4.3. *If $(\tau_s)_s$ is the one-parameter group of diffeomorphisms associated with the vector field Z, then*

$$\tau_s(x) = x + sZ(x) + o(s^2).$$

Proof. It follows from

$$\lim_{s\to 0} \frac{\tau_s(x) - \tau_0(x)}{s - 0} = Z(x)$$

and $\tau_0(x) = x$. ∎

In the following we shall consider $M = \mathbb{R}^m$ and choose $f = x^i$ to be the ith coordinate function. Then Corollary 1.4.2 becomes

$$\big(\psi_s \circ \varphi_s(x)\big)^i - \big(\varphi_s \circ \psi_s(x)\big)^i = s^2[Y, X]^i(x) + o(s^2), \quad i = 1, \dots, m.$$

In vectorial notation we have

$$\psi_s \circ \varphi_s(x) - \varphi_s \circ \psi_s(x) = s^2[Y, X](x) + o(s^2). \tag{1.4.2}$$

Using Lemma 1.4.3 yields

$$\psi_s \circ \varphi_s(x) - \varphi_s \circ \psi_s(x) = \tau_{s^2}(x) - x + o(s^2),$$

where τ_s is the one-parameter group of diffeomorphisms of $[Y, X]$.

Denote $q = \varphi_s \circ \psi_s(x)$. Then

$$\psi_s \circ \varphi_s(x) = \psi_s \circ \varphi_s \circ \psi_s^{-1} \circ \varphi_s^{-1}(q)$$

and (1.4.2) becomes

$$\psi_s \circ \varphi_s \circ \psi_s^{-1} \circ \varphi_s^{-1}(q) - q = s^2[Y, X] + o(s^2).$$

We arrive at the following result.

Proposition 1.4.4. *Let $[\psi_s, \varphi_s] := \psi_s \circ \varphi_s \circ \psi_s^{-1} \circ \varphi_s^{-1}$. Then*

$$\begin{aligned}[\psi_s, \varphi_s](q) &= q + s^2[Y, X](q) + o(s^2)\\ &= \tau_{s^2}(q) + o(s^2).\end{aligned}$$

If $X, Y \in \Gamma(\mathcal{D})$ and $[X, Y] \notin \Gamma(\mathcal{D})$, then we can move in the $[X, Y]$ direction by just going along the integral curves of X and Y. This is the main idea of the proof of Chow's theorem of connectivity by horizontal curves. In other words, if a creature lives in a universe where it is constrained to move only along a noninvolutive distribution, then it can move in any direction just by taking tangent paths to the distribution (see Fig. 1.1).

The commutator in local coordinates. Given two tangent vector fields U and V to the differentiable manifold M, their commutator vector field is defined by

$$[U, V] = UV - VU = \nabla_U V - \nabla_V U.$$

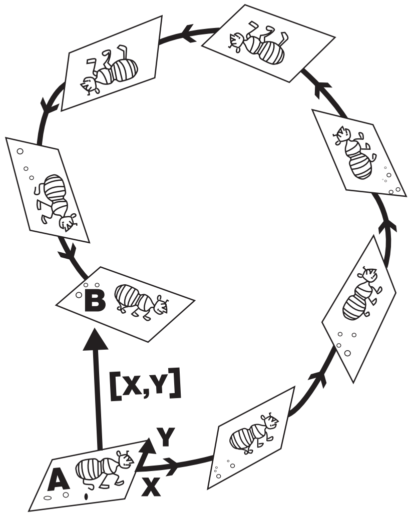

Figure 1.1. The ant can go in the $[X, Y]$ direction by just walking along the integral curves of the noncommuting vector fields X and Y

If $U = \sum_i U^i \partial_{x_i}$ and $V = \sum_i V^i \partial_{x_i}$ are the representations in a local chart $(x_1, \ldots, x_n)$, then the commutator in local coordinates becomes

$$[U, V] = UV - VU = \Big(U^i\, \partial_{x_i}(V^j) - V^i\, \partial_{x_i}(U^j)\Big)\partial_{x_j},$$

with summation in the repeated indices. The reader can verify the following properties of the commutator:

(1) The commutator is skew-symmetric: $[U, V] = -[V, U]$.
(2) Jacobi's identity is satisfied:

$$[U, [V, W]] + [V, [W, U]] + [W, [U, V]] = 0.$$

(3) For any smooth functions f and g on M we have

$$[fU, hV] = fh[U, V] + f(Uh)V - h(Vf)U.$$

Geometrical interpretation of a vanishing commutator. Let φ_t and ϕ_s be the one parameter groups of diffeomorphisms associated with the vector fields U and V. Then $[U, V] = 0$ if and only if $\varphi_t(\phi_s(p)) = \phi_s(\varphi_t(p))$; i.e., if starting at any point p and going in an arc s along the integral curve of V and then an arc t along the integral curve of U, we end up at the same point as if we performed the procedure in the reverse way.

1.5 Independent One-Forms

We review here a few basic notions regarding one-forms, which will be useful in the future presentation. One reason for studying them is that a distribution can also be defined in terms of one-forms.

Let M be a differentiable manifold. A one-form on M is a section through the cotangent bundle T^*M, i.e., a smooth assignment $M \ni p \to T_p^*M$. If $(x_1, \dots, x_n)$ are the coordinates on an open domain $U \subset M$, a one-form ω can be written in local coordinates as $\omega = \sum_{i=1}^n \omega_i(x)dx_i$, where $\omega_i(x)$ are smooth functions of x. Since all the computations in this section have a local character, we may assume $M = \mathbb{R}^n$.

Consider two one-forms

$$\omega_1 = \sum_{j=1}^n \omega_1^i \, dx_i, \qquad \omega_2 = \sum_{j=1}^n \omega_2^i \, dx_i$$

on $\mathbb{R}^n$ and let

$$S_i = ker\, \omega_{i\,|p} = \{X \in T_pM;\ \omega_{i\,|p}(X) = 0\}, \quad i \in \{1, 2\}$$

be the $(n-1)$-dimensional vectorial subspaces of T_pM defined by the preceding one-forms at p.

Definition 1.5.1. *The spaces S_1 and S_2 are called transversal if they are not parallel. We shall write in this case $S_1 \not\parallel S_2$.*

Let $\langle\ ,\ \rangle$ be the natural inner product of $\mathbb{R}^n$. If $X = \sum_{k=1}^n X^k \partial_{x_k} \in ker\, \omega_i$, then we can write

$$\begin{aligned} 0 = \omega_i(X) &= \sum_{j=1}^n \omega_i^j \, dx_j(X) \\ &= \sum_{j=1}^n \omega_i^j X^j = \langle \nu_i, X \rangle, \end{aligned}$$

and hence $\nu_i = \sum_{i=1}^n \omega_i^j \partial_{x_j}$ is a normal vector field to the space $S_i = ker\, \omega_i$.

Definition 1.5.2. *Two one-forms ω_1 and ω_2 are called functionally independent if*

$$rank \begin{pmatrix} \omega_1^i \\ \omega_2^j \end{pmatrix}_{1 \le i,j \le n} = 2.$$

k one-forms $\omega_1, \dots, \omega_k$ are called functionally independent if

$$rank \begin{pmatrix} \omega_1^{i_1} \\ \vdots \\ \omega_k^{i_n} \end{pmatrix}_{1 \le i_1, \dots, i_n \le n} = k;$$

i.e., the coefficients matrix has maximum rank.

Remark 1.5.3. *Definition* 1.5.2 *does not depend on the choice of the basis of one-forms. If $\omega = \sum \omega^i dx_i = \sum \widetilde{\omega}^j d\widetilde{x}_j$ is the representation of the one-form in two local systems of coordinates, then $\omega^i = \widetilde{\omega}^j d\widetilde{x}_j(\partial_{x_i}) = \widetilde{\omega}^j (\partial \widetilde{x}_j / \partial x_i)$, and hence $rank\, \widetilde{\omega}_p^j = rank\, \omega_p^j$.*

Proposition 1.5.4. (1) *The spaces* $ker\ \omega_1$ *and* $ker\ \omega_2$ *are transversal at* p *if and only if* ω_1 *and* ω_2 *are linearly independent at* p.

(2) $\bigcap_{j=1}^{k} ker\ \omega_j \neq \emptyset$ *if and only if the one-forms* $\omega_1, \ldots, \omega_k$ *are functionally independent.*

Proof.

(1) Let $S_i = ker\ \omega_i$. Then the spaces $S_1 \not\parallel S_2$ if and only if the normal vectors are not parallel, i.e., $\nu_1 \not\parallel \nu_2$, or $(\omega_1^1, \ldots, \omega_1^n)$ and $(\omega_2^1, \ldots, \omega_2^n)$ are not proportional. This means that there is a 2×2 nondegenerate minor matrix

$$\det \begin{pmatrix} \omega_1^{i_1} & \omega_1^{i_2} \\ \omega_2^{j_1} & \omega_2^{j_2} \end{pmatrix} \neq 0,$$

and therefore the rank of the coefficients matrix is 2. Hence ω_1 and ω_2 are functionally independent.

(2) We leave this as an exercise for the reader. ∎

Example 1.5.1. *The following two one-forms on* $\mathbb{R}^4_{(x_1,x_2,y_1,y_2)}$

$$\omega_1 = dy_1 - x_1 dx_2$$
$$\omega_2 = dy_2 - \frac{1}{2} x_1^2 dx_2$$

are functionally independent.

Example 1.5.2. *The following three one-forms on* $\mathbb{R}^4_{(x_1,x_2,y_1,y_2)}$

$$\begin{aligned} \omega_1 &= dx_1 + x_1 dx_2 + y_2 dy_1 + y_1 dy_2 \\ \omega_2 &= dx_2 + x_2^2 dy_1 + y_2 dy_2 \\ \omega_3 &= dy_1 + y_1^2 dy_2 \end{aligned}$$

are functionally independent.

1.6 Distributions Defined by One-Forms

Codimension 1. The simplest case is when the distribution is defined by only one one-form ω as $\mathcal{D} = ker\ \omega$. We note that for any $f \neq 0$ the distribution is still given by $\mathcal{D} = ker\ f\omega$, and therefore the one-form is unique up to a multiplicative nonvanishing function.

Codimension 2. Consider the case of a distribution defined by two functionally independent one-forms ω_1 and ω_2. Then we define the distribution by

$$\mathcal{D}_{(\omega_1,\omega_2)} = ker\ \omega_1 \cap ker\ \omega_2.$$

The following result shows the invariance of the distribution under some algebraic operations with one-forms.