GEOMETRICAL MECHANICS

Part I

Lectures by Saunders MacLane

Department of Mathematics
The University of Chicago

Winter Quarter 1968

These Lecture Notes were prepared with assistance by a grant from the Office of Naval Research

Notes by Raphael Zahler

Assisted by G. T. Angwin
W. M. Corvette
P. H. Palmquist
M. K. Smith

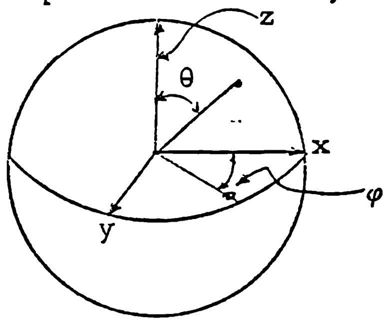
TABLE OF CONTENTS

Introduction

"Kinetic energy is a Riemann Metric on Configuration space."

We examine this statement in detail in order to illustrate the method and purpose of this course. First, we define kinetic energy as $T = \frac{1}{2} mv^2$; in detail, the kinetic energy T of a particle with mass m, moving along an arc $s = s(t)$ at velocity $v = \frac{ds}{dt}$ is $T = \frac{1}{2} mv^2$.. In 3-space, with coordinates x, y, and z, $ds^2 = dx^2 + dy^2 + dz^2$.

If the particle is moving on the surface of a sphere of radius r, its position may be given by spherical coordinates:



$$x = r \sin\theta \cos\varphi$$
$$y = r \sin\theta \sin\varphi$$
$$z = r \cos\theta$$

where θ depends on the "latitude" and φ on the "meridian." Since r is fixed,

$$dx = r \cos\theta \cos\varphi \, d\theta - r \sin\theta \sin\varphi \, d\varphi$$
$$dy = r \cos\theta \sin\varphi \, d\theta + r \sin\theta \cos\varphi \, d\varphi$$
$$dz = -r \sin d\theta$$

An elementary calculation gives

$$ds^2 = dx^2 + dy^2 + dz^2 = r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2 .$$

The equation

$$ds^2 = r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2$$

is an example of a Riemann metric. It is a symmetric (in fact, diagonal) quadratic form in the differentials $d\theta$ and $d\varphi$. This metric on the $\varphi - \theta$ rectangle $(0; 2\pi) \times (0; \pi)$ gives arc length on the sphere.

However, this single $(\varphi - \theta)$ chart is not enough. (A chart on the sphere is a smooth 1-1 correspondence between an open set in the plane and part of the sphere. In this chart, the angles (φ, θ) are mapped to the point they determine $(r \sin\theta \cos\varphi, r \sin\theta \sin\varphi, r \cos\theta)$.) This chart cannot describe neighborhoods of the north or south pole smoothly. So more charts are needed.

In fact, it would be better to start over and use the two charts based on stereographic projections from the two poles. The first chart would be the mapping of the whole x-y plane onto the sphere minus the north pole. This is done by placing the sphere's south pole tangent to the origin of the x-y plane and mapping each point (x, y) in the plane onto the point of the sphere where a line (segment) from (x, y) to the north pole intersects the sphere. The other chart is made similarly by placing the x-y plane tangent to the north pole.

The sphere together with these charts is an example of a Differentiable Manifold. We will frequently use differentiable manifolds (e.g. configuration space will be defined as a differentiable manifold ...)

Two possible references are:

S. Sternberg, Lectures on Differential Geometry, Prentice Hall 1964
(It contains references to mechanics)

and

Noel J. Hicks, Notes on Differential Geometry, van Nostrand

The texts are:

Ralph Abraham, Foundations of Mechanics, Benjamin 1967
and
Mac Lane & Birkhoff, Algebra, Macmillan 1967 (contains references on vectors, quadratic forms, modules ...)

Let M be a "configuration space" with coordinates $q^1, \ldots, q^n$. We pose the problem: given n particles, each moving in one dimension, with masses $m_1, \ldots, m_n$, can we formulate the kinetic energy of this system as that of one particle of mass m moving in n-space. (i.e., $T = \frac{1}{2} m(\frac{ds}{dt})^2$ where s denotes an element of arc in n-space)?

The total energy T is the sum of the energies T_i, where T_i is the energy of the i^{th} particle. Thus

$$T = \sum_{i=1}^{n} \frac{1}{2} m_i (\frac{dq^i}{dt})^2 .$$

We need only define

$$ds^2 = \sum_{i=1}^{n} (\frac{m_i}{m})(dq^i)^2 \tag{1}$$

to obtain $T = \frac{1}{2} m(\frac{ds}{dt})^2$. Moreover, (1) is a Riemann metric on configuration space !

In general, a Riemann metric is of the form

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}\, dq^i dq^j$$ where (g_{ij}) is a symmetric and positive definite matrix. Each g_{ij} is a constant or, more generally, a smooth function.

§1 Modules (including vector spaces).

Let K be a commutative ring. That is, K is a set of elements (scalars) which is an abelian group under the binary operation + (addition), with $0 \in K$ as the neutral element: that is,

for all $k, k' \in K$, $k + k' \in K$; $0 + k = k$; there exists $-k \in K$ such that

$$k + (-k) = 0$$

for all $k_1, k_2, k_3 \in K$, $(k_1 + k_2) + k_3 = k_1 + (k_2 + k_3)$; $k_1 + k_2 = k_2 + k_1$.

Also there is a binary operation $\cdot$ (multiplication), with $1 \in K$ as the neutral element, satisfying : for all $k_1, k_2, k_3 \in K$,

$1 \cdot k_1 = k_1$, $k_1 \cdot k_2 \in K$, $(k_1 \cdot k_2) \cdot k_3 = k_1 \cdot (k_2 \cdot k_3)$, $k_1 \cdot k_2 = k_2 \cdot k_1$.

Moreover, the distributive laws hold, viz., $k_1 \cdot (k_2 + k_3) = k_1 \cdot k_2 + k_1 \cdot k_3$.

Examples are $\mathbb{Z}$ the ring of integers, $\mathbb{Q}$ the ring of rational numbers, and $\mathbb{R}$ the ring of real numbers. Moreover, $\mathbb{Q}$ and $\mathbb{R}$ are fields (a commutative ring K, is a field if for each $k \in K$, $k \neq 0$ there is a $k^{-1} \in K$ such that $k \cdot k^{-1} = 1$).

Definition. A K-module A is an abelian group A with right (module) action by K

$$A \times K \longrightarrow A$$

defined by $(a, k) \rightsquigarrow ak$ and satisfying the laws

1. $a(k + k') = ak + ak'$
2. $(a + a')k = ak + a'k$
3. $a1 = a$
4. $(ak)k' = a(kk')$.

If K is a field, A is a vector space over K.

(Note: We employ the following "arrow" notation; for sets X and Y the straight arrow $X \longrightarrow Y$ denotes a function from X into Y; the wavy arrow $x \rightsquigarrow y$ shows the value y the function takes at $x \in X$. If we want to label the function f, we write $\begin{array}{c} f: X \longrightarrow Y \\ x \rightsquigarrow f(x) \end{array}$ or $\begin{array}{c} X \xrightarrow{f} Y \\ x \rightsquigarrow f(x) \end{array}$. X is called the domain of f, Y is called the codomain (or range) of f, and if they are clear, we may write $f: x \rightsquigarrow f(x)$.)

<u>Definition.</u> $f: A \longrightarrow B$ (with $a \rightsquigarrow f(a)$) is a <u>homomorphism of K-modules,</u> if f is a homomorphism of abelian groups (i.e., $f(a+a') = f(a)+f(a')$) and $f(ak) = (fa)k$ for all $a \in A$, $k \in K$.

If K is a field, f is usually called a <u>linear transformation.</u>

<u>Definition.</u> $\mathrm{Hom}_K(A, B) = \{f: A \longrightarrow B \mid f$ is a homomorphism of K-modules$\}$

= the set of all K-module homomorphisms of A into B.

The set $\mathrm{Hom}_K(A, B)$ is itself a K-module under the following definitions

1° $(f+g)(a) = fa + ga$

2° $(fk)a = (fa)k$.

The reader unfamiliar with modules is invited to check that 1° gives an abelian group and that 2° satisfies the module laws.

Note: K itself is a K-module (the right action is just multiplication $K \times K \xrightarrow{\cdot} K$)

$A^* \stackrel{\text{def.}}{=\!=} \mathrm{Hom}_K(A, K)$ is called the <u>dual</u> (or conjugate) of A. For example, if $K = \mathbb{R}$, and $A = V$ a vector space, then $V^* = \mathrm{Hom}\ (V, \mathbb{R})$ is the dual space. If V is finite dimensional with basis $e_1, \dots, e_n$, then V^* has the dual basis $e^1, \ldots, e^n$ where $e^i\colon V \longrightarrow \mathbb{R}$ is defined by

$$e^i(e_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

There is an alternate development in terms of coordinates: If $v = \sum_{i=1}^{n} e_i x^i \in V$, define $e^i\colon V \longrightarrow \mathbb{R}$ $(v \rightsquigarrow x_i)$, that is, $e^i(\sum_{i=1}^{n} e_j x^j) = x^i$.

<u>Further examples of rings and modules</u>:

$\mathbb{R}[x]$ = the ring of all polynomials in x with real coefficients

$$= \{a_o + a_1 x + \ldots + a_k x^k \mid k \geq 0,\ a_i \in \mathbb{R},\ 0 \leq i \leq k\}.$$

We illustrate how modules differ from vector spaces. Let $K = \mathbb{Z}$.

A $\mathbb{Z}$-module is just an abelian group

$$\begin{aligned} a \cdot n &= \underbrace{a + \ldots a}_{n \text{ summands}} \quad \text{if } n \geq 0 \\ &= \underbrace{(-a) + \ldots + (-a)}_{(-n) \text{ summands}} \quad \text{if } n < 0 \end{aligned}$$

The $\begin{cases}\text{abelian group} \\ \mathbb{Z}\text{-module}\end{cases}$ $\mathbb{Z}_3 = \{0, 1, 2\}$ has addition modulo 3. For example,

$2 + 2 \equiv 2 + 2 - 3 = 1$. The dual module $(\mathbb{Z}_3)^* = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_3, \mathbb{Z}) = \{f\colon \mathbb{Z}_3 \longrightarrow \mathbb{Z} \mid f$ is a homomorphism of $\mathbb{Z}$-modules$\}$, is 0, because $f(1) + f(1) + f(1) = f(1+1+1) = f(0) = 0 \in \mathbb{Z}$. Therefore $f(1) = 0$, and $f = 0$, and $(\mathbb{Z}_3)^* = 0$. However, it is well known that the dual V^* of any n-dimensional vector space V also has dimension n. Therefore, if we choose $n > 0$, then $V^* \cong V \neq 0$.

We develop further notation for $V^* = \{f\colon V \longrightarrow K \mid f$ is a K-linear transformation$\}$, where V is again a finite dimensional vector space over a field K. Write

$(f,v) \overset{\text{def.}}{=\!=} f(v) \in K$, for $f \in V^*$ and $v \in V$. The equations

$$(f, v_1 + v_2) = (f, v_1) + (f, v_2)$$
$$(f, vk) = (f, v)k$$

show that f is a linear transformation. The definitions

$$(f_1 + f_2, v) = (f_1, v) + (f_2, v)$$
$$(fk, v) = (f, v)k$$

show <u>how</u> V^* is a vector space.

Define, for all $v \in V$, $\bar{v} : V^* \longrightarrow K$ $(f \rightsquigarrow f(v))$; that is, $\bar{v}$ is the function $(-, v) : f \rightsquigarrow (f, v)$. Now $\bar{v}$ is in $V^{**} = (V^*)^*$ and $v \rightsquigarrow \bar{v}$ defines a linear transformation $V \xrightarrow{\theta} V^{**}$. (The proof is straightforward.) θ is one-to-one (i. e. $\bar{v} = 0$ implies $v = 0 \in V$) and V and V^{**} have the same dimension, therefore θ is an isomorphism between V and its "double dual" V^{**}. The isomorphism is natural (see <u>Algebra</u>, Ch. 15, § 5) and we will identify

$$\begin{aligned} V &\equiv\!\!> V^{**} \\ v &\rightsquigarrow \bar{v} \end{aligned}$$

by this isomorphism.

We review dual bases in the (,)-notation. If V is n-dimensional with basis $e_1, \ldots, e_n$, then V^* is n-dimensional with basis $e^1, \ldots, e^n$ where

$$(e^i, e_j) = \delta^i_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

<u>Proposition.</u> (e^i, v) is the i^{th} coordinate of v in the basis $e_1, \ldots, e_n$.

<u>Proof.</u> Let $v = e_1 x^1 + \ldots + e_n x^n$, then $(e^i, v) = (e^i, e_1 x^1 + \ldots + e_n x^n)$ $= (e^i, e_1)x^1 + \ldots + (e^i, e_n)x^n = 0 + \ldots + (e^i, e_i)x^i + \ldots + 0 = x^i$.

§2 Euclidean Vector Spaces

A Euclidean vector space is a finite dimensional vector space W over $\mathbb{R}$ with an inner product $W \times W \xrightarrow{\cdot} \mathbb{R}$ $((v,w) \leadsto v \cdot w)$ satisfying

$$\left.\begin{array}{ll} 1.\ \text{linear} & (v_1k_1 + v_2k_2)\cdot w = (v_1 \cdot w)k_1 + (v_2 \cdot w)k_2 \\ 2.\ \text{symmetric} & v \cdot w = w \cdot v \end{array}\right\} \text{therefore bilinear}$$

3. positive definite $\quad v \neq 0 \Longrightarrow v \cdot v > 0$

For example, let V be all n-tuples $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ with inner product $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i$. The length of v is $\sqrt{v \cdot v}$. Since we wrote the elements of V as column vectors, it is suggestive to write the elements of V^* as row vectors $(a_1, \ldots, a_n)$. Then $(a,x) = ((a_1, \ldots, a_n), \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix})$ equals $\sum_{i=1}^{n} a_i x^i$ which is the matrix product $(a_1, \ldots, a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

If V has an inner product, then we have a natural isomorphism $V \cong V^*$ $(v \leadsto \tilde{v})$, where $\tilde{v}(w) = v \cdot w$. $\tilde{v}$ is linear because $\tilde{v} = v \cdot -$, so indeed $\tilde{v} \in V^*$. $V \longrightarrow V^*$ is linear, because

$$(\widetilde{v+v'})(w) = (v+v')\cdot w = (v \cdot w) + (v' \cdot w) = \tilde{v}(w) + \tilde{v}'(w) = (\tilde{v} + \tilde{v}')(w),$$

and

$$\widetilde{vk}(w) = (vk)\cdot w = \tilde{v}(w)k = (\tilde{v}k)(w).$$

$V \longrightarrow V^*$ is one-to-one, since $\tilde{v} = 0 \Rightarrow v \cdot v = 0 \Rightarrow v = 0$. V and V^* have the same dimension, so $V \to V^*$ is onto. We identify $V = V^*$ by this isomorphism.

Let V have an inner product $v \cdot w$, Take any basis $e_1, \ldots, e_n$. Then let $g_{ij} = e_i \cdot e_j \in \mathbb{R}$. $g = (g_{ij})$ is an $n \times n$ matrix, symmetric and positive definite. Moreover, the matrix g determines the inner product!

$$\left[\left(\sum_{i=1}^{n} e_i x^i,\ \sum_{j=1}^{n} e_j y^j\right) = \sum_{i,j=1}^{n} (e_i, e_j) x^i x^j = \sum_{i,j=1}^{n} g_{ij} x^i x^j\right]$$

But $V = V^*$, so the dual basis $e^1, \ldots, e^n$ is also a basis of V, while the equation $g^{ij} = e^i \cdot e^j$ defines a different symmetric, positive definite $n \times n$ matrix of real numbers! However,

$$(e^i, e_j) = e^i \cdot e_j = \delta^i_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

<u>Proposition.</u> $e^i = \sum_{j=1}^{n} g^{ij} e_j$ and $e_i = \sum_{j=1}^{n} g_{ij} e^j$.

<u>Proof.</u> Because of the duality it suffices to prove only the first equality. It is enough to test the equality by application to each basis vector e^k (since for all k, $v \cdot e^k = v' \cdot e^k \Rightarrow$ for all k, $(v - v') \cdot e^k = 0 \Rightarrow v - v' = 0 \Longrightarrow v = v'$).

We test

$$\left(\sum_{j=1}^{n} g^{ij} e_j\right) \cdot e^k = \sum_{j=1}^{n} g^{ij} (e_j, e^k) = \sum_{j=1}^{n} g^{ij} \delta^k_j = g^{ik} \stackrel{\text{def}}{=} e^i \cdot e^k .$$

We summarize: the g^{ij} change upper indices to lower ones, and the g_{ij} change lower indices to upper ones. Moreover, g^{ij} is the inverse matrix of g_{ij} .

<u>Definition.</u> A Riemann metric on $\mathbb{R}^n$ (with coordinates $q^1, \ldots, q^n$) is a function $G: \mathbb{R}^n \longrightarrow n \times n$ matrices over $\mathbb{R}$ where $(q^1, \ldots, q^n) \rightsquigarrow (g_{ij} = g_{ij}(q^1, \ldots, q^n))$ and (g_{ij}) is a positive definite, symmetric matrix, and for each i and j the functions $g_{ij}(q^1, \ldots, q^n)$ has continuous derivatives of all orders $(g_{ij}(q^1, \ldots, q^n) \in C^\infty)$. We let

$ds^2 = \sum_{i,j=1}^{n} g_{ij}(q^1, \ldots, q^n) dq^i dq^j$ so that arc length is given by the integral

$$s = \int_{t=0}^{1} \sqrt{\sum_{i,j=1}^{n} q_{ij}(q^1(t), \ldots, q^n(t)) \frac{dq^i}{dt} \frac{dq^j}{dt}}\, dt \quad ,$$

Again, to keep things simple, we consider a system of n particles, each moving one-dimensionally. Our configuration space is $\mathbb{R}^n$, where the coordinates $q^1, \ldots, q^n$ correspond to the position of the n particles. If the i^{th} particle has mass m_i, its kinetic energy T is

$$T = \frac{1}{2} m_i \left(\frac{dq^i}{dt}\right)^2 = \frac{1}{2} m_i (V^i)^2 \quad .$$

The second law of motion tells us that, if F_i is the force on the i^{th} particle then $m_i \frac{d^2 q^i}{dt^2} = F_i$ for $i = 1, \ldots, n$. We also assume that the system is <u>conservative</u>, that is, there exists a suitable potential energy function $V: \mathbb{R}^n \longrightarrow \mathbb{R}$ (i.e., a real-valued function on the configuration space) such that the forces are $F_i = -\frac{\partial V}{\partial q^i}$.

The above second-order system of differential equations is difficult to work with, but by the standard trick of doubling the number of variables we get the equivalent first-order system

$$m_i \frac{dV^i}{dt} = F_i \quad , \quad \frac{dq^i}{dt} = V^i \, , \qquad i = 1, \ldots, n$$

in a 2n-dimensional space with coordinates $q^1, \ldots, q^n; v^1, \ldots, v^n$. For reasons that we hope to make clear later, we again shift coordinates by transforming to momentum, p_i :

(1) $$p_i = m_i V^i = \frac{dT}{dV^i} \quad , \qquad i = 1, \ldots, n.$$

The Riemann metric, which you may recall we identified with the kinetic-energy form, is a matrix g_{ij} whose only entries in this case are the numbers m_i on the diagonal; the transformation given by this metric is exactly the transformation of equation (1). In our new coordinates $q^1, \ldots, q^n, p_1, \ldots, p_n$, we define the Hamiltonian H = T + V, and we get

$$T = \frac{1}{2} \sum \frac{p_i^2}{m_i}, \quad V_i = \frac{dT}{dp_i} = \frac{\partial H}{\partial p_i}$$

since V does not depend on momentum; and by the second law

$$\frac{dp_i}{dt} = - \frac{dV}{dq_i} = - \frac{\partial H}{\partial q^i}$$

or

$$\left.\begin{aligned} \frac{dp_i}{dt} &= - \frac{\partial H}{\partial p_i} \\ \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} \end{aligned}\right\} \quad i = 1, \ldots, n .$$

Typical conservative mechanical systems can be described by equations in this so-called Hamiltonian form. The system of equations refers to the coordinates $q^1, \ldots, q^n, p_1, \ldots, p_n$ of a point in phase space; in the most general case, the first n of these coordinates will not necessarily describe a point of a vector space, as they do in our simple-minded example, but a point of a more general mathematical object. The last n coordinates, though, will often refer to a vector space. The whole business will be described, mathematically, as the cotangent bundle of a differentiable manifold, which is just a method of expressing the properties of the usual phase spaces of mechanics in a systematic and presumably more comprehensive way.

Chapter I. Local Mechanics

§1. Functions

First, some preliminaries about notation. Following mathematical usage, we refer to "the function sin," not "the function sin(x)," reserving the expression sin(x) for the value of the function sin at the point x. "The function e^x" we write as "e^-", and "the function x^2" comes out as "$(-)^2$". The value of the function f at x is f(x), or sometimes fx.

Next, we review some basic definitions from calculus and show how they may be understood intuitively in terms of easy topological notions. Recall

Definition. If f is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, f is continuous at a if, given $\epsilon > 0$, there exists a number $\delta > 0$, such that $|x-a| < \delta$ implies $|fx - fa| < \epsilon$. If f is a function mapping $\mathbb{R}^n$ to $\mathbb{R}^n$ (in symbols $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$), we replace $|x-a|$ by $\sqrt{\sum (x_i - a_i)^2}$.

Now some topological definitions. In $\mathbb{R}^n$, given a point $a = (a_1, \ldots, a_n)$, the open ball of radius δ with center a is $\{(x^1, \ldots, x^n) \mid \sqrt{\sum (x_i - a_i)^2} < \delta\}$. If $n = 1$, an open ball is just an open interval (open = not including end points), while in dimension two an open ball becomes just a disk, (open = not including the points on the circumference). We generalize this property of "a set which contains none of its boundary points" in the next definition.

Definition. An open set U in $\mathbb{R}^n$ is any union of open balls.

Be aware that there may be infinitely many open balls in the union making up U, and that they may overlap; thus an open set may be a very complicated object, with ragged edges, holes, disconnected pieces, and other peculiarities you can visualize. But we can state the following (the proof is easy): U is an open set if and only if, given a in U, there is a number $\delta > 0$, such that the ball of radius δ with center a lies in U. In fact, in terms of open sets the definition of continuity assumes the following new and interesting form.

Theorem. Let U be an open set in $\mathbb{R}^n$, and $f: U \longrightarrow \mathbb{R}^m$. Then f is continuous at every point of U if and only if, for every open set V in $\mathbb{R}^m$, $f^{-1}V$ is open in $\mathbb{R}^n$, where $f^{-1}V = \{x \in U \mid f(x) \in V\}$.

Here $f^{-1}V$, called the inverse image of V, is merely the set of all points of U which f maps into V.

We quickly sketch the proof of the theorem: if f is continuous by our first definition, pick a point of $f^{-1}V$, and take a ball of radius ϵ around its image in V; then by the first definition there will be a ball of radius δ surrounding our original point and lying in $f^{-1}V$, which shows that $f^{-1}V$ is open. The other implication is proved similarly. For details on this, as well as more facts about general topology, refer to any book on general topology.

The i^{th} coordinate of a point in $\mathbb{R}^m$ may be viewed as a real-valued function $q^i: \mathbb{R}^m \longrightarrow \mathbb{R}$ defined by $q^i(a) = q^i(a_1, \ldots, a_n) = a_i$, $i = 1, \ldots, m$. Thus $\varphi: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ yields m coordinate functions φ^i, defined by

$\varphi^i(x) = q^i(\varphi(x))$. By the partial derivatives of the function φ we mean the usual partial derivatives $\dfrac{\partial\varphi^i}{\partial q^j}$.

Definition. φ is C^1 if all the first-order partial derivatives exist and are continuous; it is C^k if for each φ^i, all possible partial derivatives of order $\leq k$ exist and are continuous; it is C^∞ if it is C^k for every $k > 0$. A smooth function will usually mean a C^∞ function.

Thus, the set $\mathcal{F}$ of all smooth functions $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ forms a ring, since if f and g are smooth, so is the sum $f+g$, where $(f+g)(x) = f(x) + g(x)$, and so is the product fg, where $(fg)(x) = f(x)g(x)$; and since the ring axioms hold for this sum and product of functions.

§2 Paths, Functions, and Tangent Spaces.

Now we come to the fundamental duality involved in describing the action of physical systems, a type of duality which we will see again and again in this course. We have already met the smooth function $f: U \longrightarrow \mathbb{R}$ for U open in $\mathbb{R}^n$. Its counterpart is the path, a smooth function $c: I \longrightarrow \mathbb{R}^n$, where I is an interval of $\mathbb{R}$. (Think of a point $t \in I$ as a "time".) Notice that for us a path is not just a string of points but a function, which specifies for each $t \in I$, the point $c(t)$ reached at time t. At any point x of $\mathbb{R}^n$, consider all paths passing through x. Each has its associated tangent vector at x, which is shorter or longer depending on the speed with which the path is traversed, that is, on the parametrization of the function c. We are about to describe how these vectors form a tangent space.

First, we define an operation relating the dual objects f and c.

Definition. $<f, c>_a = [\frac{d}{dt}(f \circ c)]_{t=0}$, if $c(0) = a$.

Notice that we have the property

$$<f_1 k_1 + f_2 k_2, c>_a = <f_1, c>_a k_1 + <f_2, c>_a k_2,$$

Our object is to use the operation $<\ ,\ >$ to establish a dual vector space relation between the tangent space and the set of differentials, which will form the cotangent space. To see why this is possible, let us for the moment put coordinates on $\mathbb{R}^n$. The path c maps a subset I of $\mathbb{R}$ into $U \subset \mathbb{R}^n$, and f maps U into $\mathbb{R}$ again; thus the composite function $f \circ c$, defined by the rule $(f \circ c)(t) = f(c(t))$, is just a real-valued function defined on the subset I of the real line. The chain rule for functions of several variables gives us

$$\frac{d(f \circ c)}{dt} = \frac{df(c^1(t), \ldots, c^n(t))}{dt} = \sum_{i=1}^{n} (\frac{\partial f}{\partial q^i})_{q=a} (\frac{dc^i}{dt})_{t=0} .$$

Thus, it seems possible to represent the tangent vector to c by the vector $(\frac{dc^1}{dt}, \ldots, \frac{dc^n}{dt})_0$, and the differential of f by the vector $(\frac{\partial f}{\partial q^1}, \ldots, \frac{\partial f}{\partial q^n})_a$; then $<f, c>_a$ is exactly the ordinary euclidean inner product of these two vectors.

We choose, however, to develop these ideas by the more intuitive, coordinate-free approach. To do this we need the notion of equivalence relation, a generalization of the idea of equality. We say $\equiv$ is an equivalence relation on the set S if for all elements a, b, and c of S, we have $a \equiv a$; $a \equiv b$ implies $b \equiv a$; and $a \equiv b$ and $b \equiv c$ implies $a \equiv c$. In words, $\equiv$ is reflexive, symmetric, and transitive.

It is an easy theorem that $\equiv$ divides S into equivalence classes; that is, subsets of S consisting of mutually equivalent elements, such that if any two elements of S are equivalent then they are in the same subset. The crucial idea here is that we can now view each equivalence class as itself an element in a new set W; we decide that, since all the elements of an equivalence class are equivalent, we might as well consider them as the same object. As an example, let S be the set of real numbers, and let s be equivalent to t if s - t is an integral multiple of 2π; the set of equivalence classes can be identified with the unit circle in the complex plane. (See Mac Lane and Birkhoff, Algebra, Chapter 1, §7.)

We now use the idea of equivalence class to identify all those functions on U which have the same "cotangent" vector at the point $a \in U$: We define $f \equiv_a g$ iff $\langle f, c\rangle_a = \langle g, c\rangle_a$ for all c. In coordinate notation, this means the same thing as saying

$$\left(\frac{\partial f}{\partial q^i}\right)_a = \left(\frac{\partial g}{\partial q^i}\right)_a \qquad \text{for all } i.$$

It is not hard to see that the equivalence classes of functions now form a vector space, which we call T^aU, the cotangent space to U at a, or the space of differentials. We just have to check the vector space axioms, first noticing that

$$f_1 \equiv_a g_1\,,\ f_2 \equiv_a g_2 \ \text{ implies } \ f_1+f_2 \equiv_a g_1+g_2$$

$$f \equiv_a g\,,\ k \text{ a scalar implies } fk \equiv_a gk$$

since

$$\langle f_1+f_2, c\rangle_a = \langle f_1, c\rangle_a + \langle f_2, c\rangle_a$$

and

$$\langle fk, c\rangle_a = \langle f, c\rangle_a k\,;$$

and then, defining $d_a f$ to be the equivalence class of the function f, we see that we may write $d_a(f+g) = d_a f + d_a g$, and $d_a(fk) = (d_a f)k$, and the vector space axioms are satisfied.

Now let $b, c : I \longrightarrow U$ be two paths with $b(0) = c(0) = a$, and write c^i for $q^i c: I \longrightarrow \mathbb{R}$. In a similar way we define $b \equiv_a c$ to mean that, for all f, $<f, b>_a = <f, c>_a$ (in coordinate notation, $(\frac{db^i}{dt})_{t=0} = (\frac{dc^i}{dt})_{t=0}$ for all i). Intuitively speaking, b and c are equivalent curves if and only if they kiss at a; that is, if they have the same tangent vector there (same length and same direction). If we write $T_a U$ for the set of all congruence classes $\tau_a c$: of paths, we find that we can no longer duplicate the vector space construction above, since the latter depended on being able to add two functions f and g; whereas there is no direct and natural way of defining the sum of two paths. We rely instead on the operation $< , >$ to transfer to vector space structure on $T^a U$ to the set $T_a U$. Here $T^a U$ is the space of all differentials $d_a f$, while $T_a U$ is the set of all tangents $\tau_a c$ to paths c through a.

Since the value of $<f, c>_a$ depends only on the equivalence classes of f and c, we can define $<d_a f, \tau_a c>_a$ to be the number $<f, c>_a$. Now the function which takes $\tau_a c$ to the linear functional $<-, \tau_a c> : T^a(U) \longrightarrow \mathbb{R}$ is a map from $T_a U$ into the dual space of $T^a U$. We have

$$\begin{aligned} \tau_a c = \tau_a b \iff c \equiv b \iff & <f, c> = <f, b> \quad \text{all } f \\ \iff & <-, \tau_a c> = <-, \tau_a b> ; \end{aligned}$$

thus, the map $T_a U \longrightarrow (T^a U)^*$ is one-to-one. If we can prove that this map is also onto, we will be able to transfer the vector space structure on the range space $(T^aU)^*$ to the domain space T_aU in such a way that the map ρ becomes a vector-space isomorphism. For this proof we must refer to the coordinates q^i of $\mathbb{R}^n$. Recall that $d_a f = d_a g$ if and only if $\left(\frac{\partial f}{\partial q^i}\right)_a = \left(\frac{\partial g}{\partial q^i}\right)_a$ for all i. The function q^i, defined by $q^i(a_1, \ldots, a_n) = a_i$, corresponds to the cotangent vector $d_a q^i$ which is written in coordinates as $(0, 0, \ldots, 1, \ldots, 0)$, where the 1 is in the i^{th} place. Hence $d_a q^1, \ldots, d_a q^n$ is a basis of T_aU. In a similar way, we have n paths running along the coordinate axes: $x^i(t) = (a^1, a^2, \ldots, a^i + t, \ldots, a^n)$. Thus $\tau_a x^1, \ldots, \tau_a x^n$ belong to T_aU. (Note that $x^i(t)$ is defined for t sufficiently small, so $x^i : I \longrightarrow U$ is a path.)

<u>Theorem</u>. $\tau_a x^1, \ldots, \tau_a x^n$ form a basis of T_aU, a vector space ;

$d_a q^1, \ldots, d_a q^n$ form a basis of T^aU ;

and

$T_aU \cong (T^aU)^*$ under the map ρ.

<u>Proof</u>. From what we have already done, it is enough to show that, using the addition and scalar multiplication operations "pulled back" from $(T^aU)^*$ by the above map ρ, $\tau_a x^1, \ldots, \tau_a x^n$ form a vector-space basis of T_aU, and that it is the dual basis of $d_a q^1, \ldots, d_a q^n$. The $\tau_a x^i$ span T_aU, since if c is any path,

$$\left\langle f, c - \sum_1^n \left(\frac{dc^i}{dt}\right)_{t=0} \tau_a x^i \right\rangle = 0,$$

hence $c \equiv_a \sum_1^n \left(\frac{dc^i}{dt}\right)_{t=0} x^i$. So we can express any $\tau_a c$ as $\sum \left(\frac{dc^i}{dt}\right)_{t=0} \tau_a x^i$.

The dual basis condition is expressed by the fact that

$$< d_a q^i, \tau_a x^j >_a = \left(\frac{d(q^i \circ x^j)}{dt}\right)_{t=0} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases} \qquad \text{Q.E.D.}$$

(Remember that this is not an inner product but a function on two different spaces, and thus we have not a single orthonormal basis but two bases which are dual, in the sense that $e^i e_j = < e^i, e_j > = \delta_{ij}$.)

Now at last we can justify our use of the term "differentials" for elements of the cotangent space, for we compute

$$< f - \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q^i}\right)_a q^i, c > = 0,$$

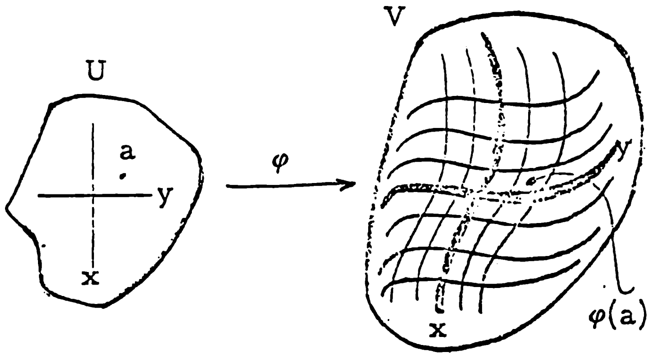
and hence

$$d_a f = \sum \left(\frac{\partial f}{\partial q^i}\right)_a d_a q^i$$

which is just the usual formula for the differential of a function of several variables. Since $< f, x^i >_a = \left(\frac{\partial f}{\partial q^i}\right)_a$, we will also write $\tau_a x^i = \left(\frac{\partial}{\partial q^i}\right)_a$.

Our coordinate-free set-up becomes really useful when we consider the effect of a smooth function φ mapping U to another open set V. If c is any path in U passing through a, φ carries that path into a path in V passing through $\varphi(a)$, namely the path represented by the composite function $\varphi \circ c$. It is easy to check that this gives us a map from $T^a U$ to $T^a V$, defined by $\tau_a c \rightsquigarrow \tau_{\varphi(a)}(\varphi \circ c)$. This map will be called $\tau_a \varphi$, or φ_* . The dual situation is completely symmetric, only everything is reversed. Namely, if h is a smooth function on V, then $h \circ \varphi$ is a smooth function on U, and thus $d_{\varphi(a)} h \rightsquigarrow d_a(h \circ \varphi)$ maps $T^{\varphi(a)} V \longrightarrow T^a U$. (Note that this map is backwards, from V to U.)

If we view intuitively a tangent vector as a small geometric vector (an arrow) lying in our open set U, then the mapping φ_* "carries" each such arrow in U to a corresponding arrow in V. Dually, we may view a function f from U to $\mathbb{R}$ as a "collapsing" of U onto some line, analogous to the projections of a two-dimensional set in the plane onto the x- and y-axes. The effect of the mapping φ^* or $d_a\varphi$ which φ induces on the cotangent space is to take a collapsing of V and from it get a collapsing of U by first mapping U to V and then collapsing V. We summarize the situation as follows



$$\begin{tikzcd}
\mathbb{R} \supset W \arrow[r, "c"] \arrow[dr, "\varphi\circ c"'] & \mathbb{R}^n \arrow[d, "\varphi"] \arrow[dr, "h\circ\varphi"] & \\
 & \mathbb{R}^n \arrow[r, "h"] & \mathbb{R}
\end{tikzcd}$$

$$\begin{tikzcd}
d_a(\tilde{f}\circ\varphi) & T^aU & T_aU \arrow[d, "T_a\varphi = \varphi_*"] & \tau_a c \arrow[d, rightsquigarrow] \\
d_{\varphi(a)}\tilde{f} \arrow[u, rightsquigarrow] & T^{\varphi(a)}V \arrow[u, "T^a\varphi = \varphi"] & T_{\varphi(a)}V & \tau_{\varphi(a)}\varphi(c)
\end{tikzcd}$$

Moreover, we have the following "self-adjointness" rule,

$$< d_a\tilde{f}\varphi, \tau_a c >_a \; = \; < d_{\varphi(a)}\tilde{f}, \tau_{\varphi(a)}\varphi c >_{\varphi(a)} \; = \frac{d}{dt}(\tilde{f}\circ\varphi\circ c)\Big|_{t=0}$$

called that because it can be abbreviated $<\tilde{f}\varphi, c> = <\tilde{f}, \varphi c>$.

When we introduce coordinates q^i on U and $\tilde{q}^i$ on V, the chain rule gives us the following formula for transforming the base $\partial/\partial\tilde{q}^i$ of $T_{\varphi(a)}V$ to that of T_aU:

$$\partial/\partial q^i = \sum \frac{\partial(\tilde{q}^j \circ \varphi)}{\partial q^i} \frac{\partial}{\partial \tilde{q}^j} .$$

Thus we may write

$$\varphi_* \frac{\partial}{\partial q^i} = \sum a_i^j \frac{\partial}{\partial \tilde{q}^j} ,$$

where a_i^j is the $m \times n$ Jacobian matrix

$$a_i^j = \left.\frac{\partial(\tilde{q}^j \circ \varphi)}{\partial q^i}\right|_a = [\varphi_* (\frac{\partial}{\partial q^i})](\tilde{q}^j) .$$

Similarly

$$\varphi^* (d_{\varphi(a)}\tilde{q}^j) = d_a(\tilde{q}^j \varphi) = \sum \left(\frac{\partial(\tilde{q}^j \circ \varphi)}{\partial q^i}\right)_a dq^i = \sum a_i^j dq^i .$$

If now $U_1 \xrightarrow{\varphi} U_2 \xrightarrow{\psi} U_3$ we get induced linear transformations $T_a(U_1) \xrightarrow{\varphi^*} T_{\varphi(a)}(U_2) \xrightarrow{\psi^*} T_{\psi(\varphi(a))}(U_3)$; by considering the effect of the composite map $\psi \circ \varphi$ on a typical curve in U_1, we see that $(\psi\varphi)_* = \psi_*\varphi_*$. In the dual case, there is a reversal of order, called <u>contravariance</u> in general:

$$(\psi\varphi)^* df = d(f\psi\varphi) = \varphi^* d(f\psi) = \varphi^*\psi^* df.$$

If 1 denotes the identity map taking every point to itself, then it is clear that we have $1_* = 1$, $1^* = 1$; and thus, in particular, if $\psi = \varphi^{-1}$ then $\psi_* = (\varphi_*)^{-1}$.

§3 Tangent Bundles

Now imagine our open set U of n-space as if it were a 2-dimensional sheet in space; over each point of U sits the tangent space at that point, an n-dimensional vector space. If we now imagine the tangent spaces as stalks which are tightly bound together by the structure of U we have envisioned the tangent bundle of U, written $T.(U)$. The points of $T.(U)$ are the pairs (a,v), where a is a point of U and v is a point of $T_a(U)$. Since a and v are both points of n-space, we imagine the pair (a,v) as lying in $\mathbb{R}^{2n}$. This concept should be clear enough from physics: the point a is the position, and the value of v is the (directed) velocity at a, which taken together form a point of phase space; since in general, to describe the future motion of a particle we need to know only its position and velocity, it seems likely that the tangent bundle will be a natural setting for the study of mechanics. Even more useful is the cotangent bundle, $T^{\cdot}(U)$, which is defined to be the set of pairs (a,w), where this time $w \in T^a(U)$; that is, $w = d_a f$ for some function f. We have natural maps, projections, in both cases:

$$\pi_{\cdot} : T_{\cdot}(U) \longrightarrow U; \qquad \pi_{\cdot}(a,v) = a$$

$$\pi^{\cdot} : T^{\cdot}(U) \longrightarrow U; \qquad \pi^{\cdot}(a,w) = a$$

Notice that if a is a point of U, the complete inverse image $(\pi^{\cdot})^{-1}(a)$ is always a vector space.

By a vector field X on U we will mean an assignment of a vector Xa of $T_a U$ to every point a of U, such that the correspondence $a \rightsquigarrow Xa$ is "smooth"; to be precise about this, we regard X as a map of U into $T.(U)$. Since $Xa \in T_a(U)$, we can write X as $a \rightsquigarrow (a, Xa)$.

Definition. A vector field is a smooth function $X: U \longrightarrow T.(U)$ such that the composite $\pi \circ X$ is the identity function: $\pi \circ X = 1_U$.

The last requirement means simply that

$$X: u \rightsquigarrow (u, \text{some tangent vector at } u).$$

Recall that the vector space T^aU has the basis $d_aq^1, \ldots, d_aq^n$, where the q^i are coordinates in U. It seems natural to put coordinates on $T^\cdot(U)$ so that the point $(a, w) \in T^\cdot(U)$ is viewed as $(q^1, \ldots, q^n, h^1 d_a q^1, \ldots, h^n d_q q^n)$, where the q^i and the h^i are the coordinates of a and w, respectively. Thus $w = h^1 d_a q^1 + \ldots + h^n d_a q^n$. In a similar way, we put $2n$ coordinates on the tangent bundle. We have seen that, if f is a function on U, then

$$d_a f = \sum \left(\frac{\partial f}{\partial q^i}\right)_a d_a q^i$$

is a point of T^aU. Hence we can write

$$(*) \qquad df = \sum \left(\frac{\partial f}{\partial q^i}\right) dq^i$$

for the function which assigns to every point a of U the cotangent vector $d_a f$; it is thus the cotangent-space analogue of a vector field. Expressions like the above constantly pop up in physics; for example, if the three components of force in space are represented by the functions F_i of x, y, and z, the infinitesimal work is usually defined by

$$dW = F_1 dx + F_2 dy + F_3 dz .$$

When we interpret dW, dx, and so forth, not as infinitesimals, but as cotangent vector fields, this equation resembles equation (*). This will actually be the way we put the notion of "infinitesimal" on a sound mathematical basis.

§4 Vector Bundles

Take a long narrow strip of paper, draw a line down the middle lengthwise, and paste the ends together by twisting once to form a figure called the möbius strip. The line down the middle now becomes a circle, and the surface can be thought of as composed of vectors ("fibers") perpendicular to the circle and radiating from it. We are about to see how the mobius strip can be viewed as part of a vector bundle over the circle.

First we define the simpler notion of pre-bundle. Given two sets U and V, their cartesian product $U \times V$ is defined to be the set of all pairs (u, v) with $u \in U$, $v \in V$. Given sets G, E, U, and maps σ and π forming the following diagram

$$\begin{tikzcd} & E \arrow[d, "\pi"] \\ G \arrow[r, "\sigma"] & U \end{tikzcd}$$

we define the pullback $G \times_U E$ as the universal object making the square

$$\begin{tikzcd} G \times_U E \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\ G \arrow[r] & U \end{tikzcd}$$

commutative; (universal objects and commutative squares will be defined in due course). Specifically, the pullback can be defined as the following set:

$$G \times_U E = \{(e, g) \mid e \in E,\ g \in G,\ \pi e = \sigma g\}.$$

A pre-bundle, usually written $\begin{matrix} B \\ \downarrow \pi \\ U \end{matrix}$, is a pair of sets B, U and a map between them such that B and U are open in some Euclidean space, π is a smooth map, and on each fiber $\pi^{-1}(u) = \{b \mid b \in B, \pi b = u\}$ there is a smooth addition and a scalar multiplication defined so that each fiber forms a vector space. By this we mean that

(a) if $b_1, b_2 \in B$ with $\pi(b_1) = \pi(b_2)$ (i.e., b_1 and b_2 are in the same fiber), and if k is a real number, then there are points $b_1 + b_2$ and $b_1 k$ in B with $\pi(b_1 + b_2) = \pi(b_1 k) = \pi(b_1)$,

(b) the resulting operations on each fiber satisy the vector space axioms, so that each fiber by itself becomes a real vector space,

(c) the maps $+: B \times_U B \longrightarrow B$ and $\cdot : B \times \mathbb{R} \longrightarrow B$ are smooth.

To state property (c) we must make the additional postulate that $B \times_U B$, the pullback in the diagram

$$\begin{tikzcd} B \times_U B \arrow[r] \arrow[d] & B \arrow[d, "\pi"] \\ B \arrow[r, "\pi"] & U \end{tikzcd}$$

may be embedded in some Euclidean space as an open set. U is usually called the base space, B the total space, and π the projection of B on U. Notice that while U is an open subset of some vector space $\mathbb{R}^k$, U is not itself a vector space. Neither is B; instead, it is made up of a lot of vector spaces glued together like a bundle of sticks.

The easiest example of a pre-bundle is the product space $U \times V$, where U is an open set in $\mathbb{R}^n$ and V is a real vector space, $V = \mathbb{R}^m$ for some m. When we define $\pi(u, v) = u$, we see that the fiber $\pi^{-1}(a) = \{(a, v) \mid v \in V\}$ is just a copy of V, and so inherits the vector-space structure of V. One can check that the other requirements are satisfied. Such a pre-bundle is called a <u>local vector bundle</u>.

Suprisingly enough, the tangent bundle to an open set U is a pre-bundle, (indeed, a local vector bundle) with U the base space, $T.(U)$ the total space, and $\pi: (a, v) \rightsquigarrow v$ the projection. Moreover, there is an isomorphism $T.(U) \xrightarrow{\cong} U \times \mathbb{R}^n$ obtained by mapping $(a, v) \rightsquigarrow (a, (v^1, \ldots, v^n))$, where v has coordinates v^i ; that is,

$$v = v^1\left(\frac{\partial}{\partial q^1}\right)_a + \ldots + v^n\left(\frac{\partial}{\partial q^n}\right)_a .$$

This isomorphism is a bundle map, in the sense that points lying in the fiber over a point a of U are mapped into points in the same fiber. More generally, the pair of smooth maps (H, h) in the diagram

$$\begin{array}{ccc} B & \xrightarrow{H} & \tilde{B} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\tilde{\pi}} \\ U & \xrightarrow{h} & \tilde{U} \end{array}$$

is called the <u>bundle map</u> if $\tilde{\pi}H = h\pi$ (that is, if the diagram commutes), and if the map H is linear on each fiber. The first condition means that the fiber over a is mapped into the fiber over $h(a)$; the second can be stated

$$\pi b_1 = \pi b_2 \Longrightarrow H(b_1 + b_2) = H(b_1) + H(b_2), \quad H(bk) = H(b)k.$$

A <u>cross-section</u> of a pre-bundle is a smooth map $\chi : U \longrightarrow B$ such that $\pi\chi = 1_U$; in other words, $\chi(u)$ lies in the fiber over u . For example, we always have the zero section: $\chi(u) = 0$, all $u \in B$. In the case of a local vector bundle, for any constant vector v_o of V the map $u \rightsquigarrow (u, v_o)$ is a cross-section. In general, the set of all cross-sections of a bundle π, denoted by $\Gamma(\pi)$, has an additive structure $[(\chi_1 + \chi_2)(u) = \chi_1(u) + \chi_2(u)]$ as well as right multiplication not just by ordinary scalars but by real-valued smooth functions: if $f \in \mathcal{F}$, $(\chi f)(u) = \chi(u)\cdot f(u)$. Since χf and $\chi_1 + \chi_2$ are also smooth, we find that $\Gamma(\pi)$ forms a module over $\mathcal{F}$.

Now that we know that $T.(U)$ is a pre-bundle, we can redefine a vector field to be a cross-section of $T.(U)$. The inevitable dual object, a cross-section of $T^{\cdot}(U)$, will be called a <u>differential one-form</u>, or a <u>co-vector field.</u> Before considering these creatures, we observe that an ordinary vector field may act on functions much like a directional derivative; that is, given a function and a vector field, we define a new function whose value at every point is the derivative of the original function in the direction given by the vector field at that point. Formally, given $g \in \mathcal{F}$ and $X \in \Gamma(\pi)$, define

$$(Xg)(a) = \langle d_a g, Xa \rangle .$$

It is important not to confuse Xg, which is a function, with the vector field Xg, the product of the cross-section X of $T.(U)$ with the function g. The context should usually make clear which one is meant.

Since $T^{\cdot}(U)$ is a pre-bundle in exactly the same way that $T.(U)$ was, it makes sense to speak of a cross-section of $T^{\cdot}(U)$, and such a function ω will be called a <u>differential one-form</u> on U. For example, if f is a smooth function taking U into $\mathbb{R}$, the function df defined by $df(a) = (a, d_a f) \in T^{\cdot}(U)$

is a differential one-form. From the i^{th}-coordinate function q^i, which maps a point of U to its i^{th}-coordinate, we get a form dq^i, and we can express df for a general f in terms of this basis:

$$(1) \qquad df(a) = \left(q^1(a), \ldots, q^n(a), \left(\frac{\partial f}{\partial q^1}\right)_a d_a q^1, \ldots, \left(\frac{\partial f}{\partial q^n}\right)_a d_a q^n\right).$$

The differential one-forms are a module over $\mathcal{F}$, just as the vector fields were; given a form ω and a function $g \in \mathcal{F}$, we define $g\omega$ (a new form) by the equation $g\omega(a) = (a, g(a)\omega_a)$ if $\omega(a) = (a, \omega_a)$. Then formula (1) shows that

$$df = \sum \left(\frac{\partial f}{\partial q^i}\right) dq^i ,$$

which is now a rigorously based statement of the familiar law of partial differentiation.

It is easy to see that any differential one-form on U can be written as $\omega = \sum_{i=1}^{n} h^i dq^i$, where the h^i are n smooth functions on U. But it is an important fact that in general not every form ω is of the form df for some smooth f. This problem is related to the well-known condition for exactness of a first-order differential equation (cf. Apostol, vol. II, p. 239). Thus we can define a one-form ω to be <u>exact</u> if there is a smooth function f on U such that $\omega = df$. In a simply-connected two-dimensional region, like the interior of a circle, a necessary and sufficient condition for the one-form $\omega = h\,dx + k\,dy$ to be exact is that $\partial h/\partial y = \partial k/\partial x$. In general, however, this is only a necessary condition (cf. Mackey, p. 18, footnote), so we define a new expression, the differential of ω, as

$$d\omega = \sum_{\substack{i,j=1 \\ i<j}}^{n} \left(\frac{\partial h^i}{\partial q^j} - \frac{\partial h^j}{\partial q^i} \right) dq^i \, dq^j \qquad \text{if} \quad \omega = \sum h^i dq^i$$

and we say that ω is closed if $d\omega = 0$. Thus any exact form is closed but not vice versa. The proverbial "alert reader" may have noticed that the above conclusion depends on the definition of exactly what $dq^i dq^j$ means; for the moment, however, we merely point out the analogy between the form of $d\omega$ and that of the "curl" of a vector field in three-space, which allows us to compare and contrast our condition for a form to be exact with the commonplace of physics that a field has a potential function if and only if its curl is zero. (It is worth pointing out that in n-space there are $\frac{n(n-1)}{2}$ different $dq^i dq^j$ for i and j ranging between 1 and n, $i<j$; when $n = 3$ it happens that $\frac{1}{2}n(n-1)$ is also 3. So while in the three-dimensional case the curl of a field is a vector of the same dimension as the field, this need not be true in general.)

We previously considered the effect of a mapping $\varphi: U \longrightarrow \tilde{U}$ on cotangent vectors; for each point a of U, φ induces a linear map $\varphi^*: T^{\varphi(a)}\tilde{U} \longrightarrow T^a U$. It is easy to generalize this to the case of differential forms. Let $\tilde{\omega}$ be a one-form on $\tilde{U}$; we desire to construct a form $\varphi^*\tilde{\omega}$ on U. Given $a \in U$, we have $\varphi(a) \in \tilde{U}$. If $\tilde{\omega}(\varphi(a)) = (\varphi(a), \omega)$, $\omega \in T^{\varphi(a)}(\tilde{U})$, then define $(\varphi^*\tilde{\omega})(a) = (a, \varphi^*(\omega))$ where $\varphi^*(\omega) \in T^a(U)$, so $(a, \varphi^*(\omega)) \in T\cdot(U)$. In a typical physical application, $\tilde{\omega}$ would represent the work for some displacement, and $\varphi^*\omega$ would be the same work function expressed in terms of a different generalized coordinate system, given by the change φ of coordinates from U to $\tilde{U}$.

In the case where $\omega = d\tilde{f}$, where $\tilde{f}$ is a real-valued smooth function on $\tilde{U}$, we can derive

$$\varphi^*(d\tilde{f}) = d(\tilde{f}\circ\varphi) = \sum_{j=1}^{n} \frac{\partial(\tilde{f}\circ\varphi)}{\partial q^j}\, dq^j .$$

In particular,

$$\varphi^*(d\tilde{q}^i) = \sum_{j=1}^{m} \frac{\partial(\tilde{q}^i\circ\varphi)}{\partial q^j}\, dq^j .$$

Here the coefficients of dq^j form the familiar Jacobian matrix $\dfrac{\partial(\tilde{q}^i\circ\varphi)}{\partial q^j}$.

Moreover, $\varphi^*(\tilde{f}\,\omega) = (\varphi^*\tilde{f})(\varphi^*\omega) = (\tilde{f}\circ\varphi)(\varphi^*\omega)$, where $\varphi^*\tilde{f} = \tilde{f}\circ\varphi$; and when we have defined the operator d on forms it will turn out that $d(\varphi^*\omega) = \varphi^*(d\omega)$. These equations can be interpreted as properties of the map $\varphi^*: \Gamma(T^{\bullet}\tilde{U}) \longrightarrow \Gamma(T^{\bullet}U)$, where $\Gamma(T^{\bullet}(U))$ is the $\mathcal{F}$-module of all differential one-forms on U. . Finally, if we have the diagram of maps $U_1 \xrightarrow{\varphi} U_2 \xrightarrow{\psi} U_3$ then $\Gamma(T^{\bullet}U_3) \xrightarrow{\psi^*} \Gamma(T^{\bullet}U_2) \xrightarrow{\varphi^*} \Gamma(T^{\bullet}U_1)$ and we have $(\psi\varphi)^*\omega = \varphi^*(\psi^*(\omega))$. We say that $(\;)^*$ is a "contravariant functor".

We can pull-back not only differential forms but also whole pre-bundles along the map φ. Specifically, given the situation

$$\begin{array}{ccc} & & \tilde{B} \\ & & \downarrow{\scriptstyle\pi} \\ U & \xrightarrow{\varphi} & \tilde{U} \end{array}$$

where $\tilde{B}$ is a pre-bundle, construct a new set

$$B = \{(a,\tilde{b}) \mid a \in U,\ \tilde{b} \in \tilde{B},\ \text{and}\ \varphi a = \pi\tilde{b}\} .$$

(This set B is exactly the "pullback" $U \times_{\tilde{U}} \tilde{B}$ discussed earlier.) If we define $\pi(a,\tilde{b}) = a \in U$, then it can be shown that B is a pre-bundle. Intuitively speaking, given any point a of U we have found the fiber over the image of a under φ and made that fiber the fiber over a in our new bundle. Notice that if $\widetilde{B}$ is a tangent bundle then its pullback need not be; for example, if $\widetilde{U}$ is some open set in the plane, and U is a line lying in $\widetilde{U}$, the pullback B will have a two-dimensional vector space over each point of U, whereas the tangent bundle to a line is always one-dimensional.

§5. The Lagrange Equations

(As a reference for the following discussion of the Lagrange equations, cf. Goldstein, pp. 10-18, and Whittaker, pp. 30-35.)

We consider a system of N particles with coordinates $x^1, \ldots, x^{3N}$, moving subject to contraints, in such a way that the system can be described by n generalized coordinates $q^1, \ldots, q^n$. For example, the position of a sphere rolling on a plane is completely determined by x and y, the coordinates of the center of the sphere, and the three Eulerian angles θ, ψ, φ which determine how much the sphere rotates. We make the assumption that the constraints are holonomic; that is, we can make small displacements in each of the q^j independently. Given applied forces $F_i^{(a)}$ and constraint forces $F_i^{(c)}$ in each of the euclidean directions x^i, $i = 1, \ldots, 3N$, and under the assumption that the forces of constraint do no work, we will show that the system obeys the Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^j}\right) - \frac{\partial \mathcal{J}}{\partial q^j} = Q_j \quad , \quad j = 1, \ldots, n$$

where $\mathcal{J} = \mathcal{J}(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n)$ is the kinetic-energy function, and Q_j is called the "generalized force in direction j" as defined below.

Newton's laws tell us exactly how the system behaves when viewed in the framework of euclidean space, $\mathbb{R}^{3N}$; our problem is to describe it in the configuration space U whose coordinates are the q^j. Since we have an evident map $\varphi: U \longrightarrow \mathbb{R}^{3N}$ which merely changes the q^j-description of a state of the system to the x^i-coordinates of $\mathbb{R}^{3N}$, a natural step is to try to "pull back" the laws of motion in $\mathbb{R}^{3N}$ along φ. For example, if

$$W = \sum_i^{3N} (F_i^{(a)} + F_i^{(c)})\, dx^i$$

is the differential form corresponding to work in euclidean space, then $\varphi^* W$ gives the work in terms of the q^j. This is so because the change of the state of the system in time is described by a path in $\mathbb{R}^{3N}$; but since we consider only motion subject to the constraints, each such path is the image under φ of a path in U. Now the "self-adjointness" rule given above for pullbacks makes it clear that $W^U = \varphi^* W$ has the same effect on a path of U as does W on the corresponding path in $\mathbb{R}^{3N}$. Computing,

$$\begin{aligned} W^U &= \sum_{i=1}^{3N} \varphi^*(F_i^{(a)} + F_i^{(c)}) \sum_{j=1}^{n} \frac{\partial x^i}{\partial q^j}\, dq^j \\ &= \sum_{j=1}^{n} \Big(\sum_i F_i^{(a)} \frac{\partial x^i}{\partial q^j} \Big) dq^j + \sum_{j=1}^{n} \Big(\sum_i F_i^{(c)} \frac{\partial x^i}{\partial q^j} \Big) dq^j , \end{aligned}$$

where we have identified F and $\varphi^* F = F \circ \varphi$ (that is, the force F as a function of the x^i and F viewed as a function of the q^i) as is the custom in physics. Since we have assumed that the $F_i^{(c)}$ do no work, we get

$$W^u = \sum_{j=1}^{n} Q_j\, dq^j \qquad \text{where} \quad Q_j = \sum_{i=1}^{3N} F_i^{(a)} \frac{\partial x^i}{\partial q^j}$$

are the generalized forces; "generalized" since they need not have the dimensions of force, although the product $Q_j q^j$ always has the dimensions of work.

The situation is now as follows

$$\begin{tikzcd}
T.(U) \arrow[r, "\Phi"] \arrow[d] & T.(\mathbb{R}^{3N}) \arrow[rr, "\mathcal{J}"] \arrow[d] \arrow[dr, shift left] \arrow[dr, shift right] & & \mathbb{R} \\
U \arrow[r, "\varphi"] \arrow[dr, shift left, "q^n"] \arrow[dr, shift right, "q^1"'] & \mathbb{R}^{3N} \arrow[dr, shift left, "x^{3N}"] \arrow[dr, shift right, "x^1"'] & (x^1, \dots, x^{3N}, \dot{x}^1, \dots \dot{x}^{3N}) & \\
& {} & {} &
\end{tikzcd}$$

where the map Φ is defined by $\Phi(a, v) = (\varphi a, \varphi^* v)$. As above, we will follow the convention of writing $\frac{\partial \mathcal{J}}{\partial q^i}$ when we really mean $\frac{\partial(\mathcal{J} \circ \Phi)}{\partial q^i}$.

Then we have the following straightforward derivation of the Lagrange equations, starting from Newton's second law:

$$m_i \frac{d^2 x^i}{dt^2} = F_i^{(a)} + F_i^{(c)} \quad , \quad i = 1, \dots, 3N\,;$$

therefore

$$(1) \quad \sum_{j=1}^{n} \Big(\sum_{i=1}^{3N} m_i \frac{d^2 x^i}{dt^2} \frac{\partial x^i}{\partial q^j} \Big) dq^j = \sum_{i=1}^{3N} (F_i^{(a)} + F_i^{(c)}) \sum_{j=1}^{n} \frac{\partial x^i}{\partial q^j} dq^j = \sum_{j=1}^{n} Q_j dq^j .$$

Now $\dot{x} = \frac{dx}{dt} = \sum_{j=1}^{n} \frac{\partial x}{\partial q^j} \dot{q}^j$. Therefore $\frac{\partial \dot{x}}{\partial \dot{q}^j} = \frac{\partial x}{\partial q^j}$. Hence

$$\frac{d^2 x}{dt^2} \frac{\partial x}{\partial q^j} = \frac{d}{dt}(\dot{x}) \frac{\partial x}{\partial q^j} = \frac{d}{dt}\Big(\dot{x} \frac{\partial x}{\partial q^j}\Big) - \dot{x} \frac{d}{dt}\Big(\frac{\partial x}{\partial q^j}\Big) = \frac{d}{dt}\Big(\dot{x} \frac{\partial \dot{x}}{\partial \dot{q}^j}\Big) - \dot{x} \frac{\partial}{\partial q^j}(\dot{x})$$

$$= \frac{d}{dt}\Big(\frac{\partial}{\partial \dot{q}^j} \cdot \frac{\dot{x}^2}{2}\Big) - \frac{\partial}{\partial q^j}\Big(\frac{\dot{x}^2}{2}\Big) .$$

By (1), $$\sum_{j=1}^{n} Q_j dq^j = \sum_{j=1}^{n} \left[\frac{d}{dt}\frac{\partial}{\partial \dot{q}^i}\sum_i \frac{m_i \dot{x}_i^2}{2} - \frac{\partial}{\partial q^j}\sum_i \frac{m_i \dot{x}_i^2}{2}\right]dq^j$$
$$= \sum_{j=1}^{n}\left(\frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^j}\right) - \frac{\partial \mathcal{J}}{\partial q^j}\right)dq^j .$$

Since the dq^j are independent, we can equate the individual coefficients and get the desired Lagrange equations.

What is really going on here? A first clue is that our formula $\frac{dx}{dt} = \sum \frac{\partial x}{\partial q^j}\dot{q}^j$ is just a special case of the formula in coordinates for $T_a(\varphi)$, the map induced by φ on the tangent bundle. We have really "lifted" the path $c: I \longrightarrow U$, which describes the change of the system in time, to a path $\tilde{c}: I \longrightarrow T.(U)$, where $\tilde{c}(t) = (c(t), \dot{c}(t))$, and examined its form under the map Φ. More pertinent is the reason why we have organized our equations of motion in this form to begin with. Examining the derivation above shows that we expressed Newton's second law, $m_i \frac{d^2x^i}{dt^2} = F^i$, in terms of $\mathcal{J} = \sum \frac{1}{2} m_i (\dot{x}^i)^2$, getting $\frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{x}^i}\right) = \frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{x}^i}\right) - \frac{\partial \mathcal{J}}{\partial x^i} = F_i$, $i = 1, \ldots, 3N$, the Lagrange equations in euclidean space; we then pulled back by φ to get a system of equations <u>in the same form</u> with respect to the q^i. What matters here is that the Lagrange equations in fact remain invariant under pullback by <u>any</u> 1-1 map φ. This will be done in the next section.

§6. <u>Lifting Paths</u>

Let W be a differentiable manifold, and $c: I \longrightarrow W$ be a path in W. We define the <u>lifted path</u> $\tilde{c}: I \longrightarrow T.(W)$ in the tangent bundle $T.(W)$ by

$\tilde{c}(t) = (c(t), \tau_{c(t)}c)$. Let $\varphi: W \longrightarrow U$. Then $(T.\varphi)(\tilde{c}) = (\widetilde{\varphi c})$, since by definition $\widetilde{\varphi c}(t) = (\varphi c(t), \tau_{\varphi ct}(\varphi c))$ and $(T.\varphi)(\tilde{c}(t)) = (\varphi c(t), T_{ct}\varphi(\tau_{ct}c))$ by definition of $T.\varphi$. We indicate the relationships of the above functions by a commutative diagram

$$\begin{tikzcd}
& T.W \arrow[r, "T.\varphi"] \arrow[d, "\pi"] & T.U \arrow[d, "\pi"] \\
I \arrow[ur, "\tilde{c}"] \arrow[r, "c"'] \arrow[urr, bend left, "\widetilde{\varphi c}"] & W \arrow[r, "\varphi"'] & U
\end{tikzcd}$$

If we choose coordinates $r^1, \ldots, r^m: W \longrightarrow \mathbb{R}$ for W and $q^1, \ldots, q^n: U \longrightarrow \mathbb{R}$ for U, we have coordinates $\dot{r}^1, \ldots, \dot{r}^m; r^1\pi, \ldots, r^m\pi: T.W \longrightarrow \mathbb{R}$ for the tangent bundle of W and $\dot{q}^1, \ldots, \dot{q}^n; q^1\pi, \ldots, q^n\pi: T.U \longrightarrow \mathbb{R}$ as coordinates for $T.U$.

The Jacobian $J = J(\varphi)$ of φ is the $n \times m$ matrix defined by $J^i_j = \dfrac{\partial q^i\varphi}{\partial r^j}$ $\left(= \dfrac{\partial q^i}{\partial r^j}\right.$ for short), where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The Jacobian of $T.\varphi$ is defined similarly. It is a $2n \times 2m$ matrix of the form shown below

$$\begin{array}{c|cc|}
 & r^1 \ldots r^m & \dot{r}^1 \ldots \dot{r}^m \\
\hline
q^1 \\ \vdots & J(\varphi) & 0 \\ q^n \\
\hline
\dot{q}^1 \\ \vdots & ? & J(\varphi) \\ \dot{q}^n \\
\hline
\end{array}$$

The lower right block can be calculated as follows

$$\frac{dq^i}{dt} = \sum_{j=1}^{m} \frac{\partial q^i}{\partial r^j} \frac{dr^j}{dt} \; ; \quad \dot{q}^i = \sum_{j=1}^{m} \frac{\partial q^i}{\partial r^j} \dot{r}^j \; .$$

Therefore

$$\frac{\partial \dot{q}^i}{\partial \dot{r}^j} = \frac{\partial q^i}{\partial r^j} \quad \text{for all } i \text{ and } j \, .$$

Thus the equations for the transformation $T.\varphi$ in terms of coordinates are

$$\dot{q}^i = \sum_{j=1}^{m} \frac{\partial q^i}{\partial r^j} \dot{r}^j \quad , \quad q^i = \varphi(r^1, \ldots, r^m).$$

The invariant description is $T.\varphi(b, \tau_b c) = (\varphi b, \tau_{\varphi b}(\varphi c))$.

Given U, ω a 1-form on U, and $\mathcal{J}: T.U \longrightarrow \mathbb{R}$ a smooth function, we define: a path c in U <u>satisfies Lagrange's equation</u> (with respect to $\mathcal{J}$ and ω) in the coordinates $q^1, \ldots, q^n$ of U where

$$\omega = \sum_{j=1}^{n} Q_j dq^j \quad , \quad \mathcal{J} = \mathcal{J}(q^1, \ldots, q^n; \dot{q}^1, \ldots, \dot{q}^n)$$

if

$$(1) \qquad \frac{d}{dt}\left[\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^i}\right)(\tilde{c})\right] - \left(\frac{\partial \mathcal{J}}{\partial q^i}\right)(\tilde{c}) = Q_i c \quad , \quad i = 1, \ldots, n .$$

The functions are indicated in the diagram:

$$\begin{tikzcd} & T.U \arrow[r, "\mathcal{J}"] \arrow[d] & \mathbb{R} \\ I \arrow[ur, "\tilde{c}"] \arrow[r, "c"] & U \arrow[r, "Q_i"] & \mathbb{R} \end{tikzcd}$$

As a special case, suppose the forces are <u>conservative</u>. By definition, the 1-form ω (the work) is conservative if and only if there exists a smooth function $\mathcal{V}: U \longrightarrow \mathbb{R}$ with $\omega = -d\mathcal{V}$, in coordinates

$$\sum_{i=1}^{n} Q_i dq^i = -\sum_{i=1}^{n} \frac{\partial \mathcal{V}}{\partial q^i} dq^i \quad \text{so} \quad Q_i = \frac{-\partial \mathcal{V}}{\partial q^i} .$$ Thus (1) becomes

$$\frac{d}{dt}\left[\left(\frac{\partial L}{\partial \dot{q}^i}\right)\tilde{c}\right] - \left(\frac{\partial L}{\partial q^i}\right)\tilde{c} = 0,$$ where L is defined by

$$L = (\mathcal{J} - \mathcal{V} \circ \pi): T.U \longrightarrow \mathbb{R}$$

We now state the theorem asserting that Lagrange's equation can be "pulled back" along a smooth map φ.

Theorem. Let $\varphi: W \longrightarrow U$ be a smooth map between open sets in euclidean spaces, while ω is a 1-form on U and $\mathcal{J}$ a smooth function on $T.U$, as above. If $c: I \longrightarrow W$ is a path in W such that the composite path φc satisfies Lagrange's equation (with respect to $\mathcal{J}$ and ω), then the path c in W satisfies Lagrange's equation (with respect to $\mathcal{J}(T.\varphi)$ and $\varphi^* \omega$).

Proof. If $\omega = \sum Q_i dq^i$, for Q_i smooth functions on U, the form $\varphi^* \omega$ is given in the coordinates r^j of W as

$$\varphi^* \omega = \sum_j \sum_{i=1}^{n} Q_i \frac{\partial q^i}{\partial r^j} dr^j .$$

This can be written as $\sum R_j dr^j$, where each

$$R_j = \sum_{i=1}^{n} Q_i \frac{\partial q^i}{\partial r^j} \quad , \quad j = 1, \dots, m .$$

We are given, on the path $\varphi \circ c$, the equations

$$\frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^i}\right) - \frac{\partial \mathcal{J}}{\partial q^i} = Q_i \quad , \quad i = 1, \dots, n,$$

(here $\mathcal{J}$ is short for $T \circ \varphi \circ c$, Q_i for $Q_i \circ \varphi \circ c$). Multiply the i^{th} equation by $\frac{\partial q^i}{\partial r^j}$ and add over i to get

$$\sum_i \frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^i}\right) \frac{\partial q^i}{\partial r^j} - \sum_i \frac{\partial \mathcal{J}}{\partial q^i} \frac{\partial q^i}{\partial r^j} = \sum_i Q_i \frac{\partial q^i}{\partial r^j} = R_j$$

(here again everything is, through c, a function of t) By the rule for differentiating the product $\frac{\partial \mathcal{J}}{\partial \dot{q}^i} \frac{\partial q^i}{\partial r^j}$, this can be rewritten as

$$\sum_i \frac{d}{dt}\left[\frac{\partial \mathcal{J}}{\partial \dot{q}^i}\ \frac{\partial q^i}{\partial r^j}\right] - \sum_i \frac{\partial \mathcal{J}}{\partial \dot{q}^i}\ \frac{d}{dt}\left(\frac{\partial q^i}{\partial r^j}\right) - \sum_i \frac{\partial \mathcal{J}}{\partial q^i}\ \frac{\partial q^i}{\partial r^j} = R_j .$$

The second and third terms combine to give $\frac{\partial \mathcal{J}\varphi}{\partial r^j}$; as for the first term,

$$\frac{\partial \mathcal{J}\varphi}{\partial \dot{r}^j} = \sum_i \frac{\partial \mathcal{J}}{\partial \dot{q}^i}\ \frac{\partial \dot{q}^i}{\partial \dot{r}^j} + \sum_i \frac{\partial \mathcal{J}}{\partial q^i}\ \frac{\partial q^i}{\partial \dot{r}^j} ,$$

(here $\frac{\partial q^i}{\partial \dot{r}^j} = 0$ and $\frac{\partial \dot{q}^i}{\partial \dot{r}^j} = \frac{\partial q^i}{\partial r^j}$, as noted above). Hence the whole equation (really with the path c substituted in) becomes

$$\frac{d}{dt}\left[\frac{\partial \mathcal{J}\varphi}{\partial \dot{r}^j}\right] - \frac{\partial \mathcal{J}\varphi}{\partial r^j} = R_j$$

and this is Lagrange's equation on W. q.e.d.

As a corollary we can prove the invariance of Lagrange's equation under a change of coordinates (i.e., under a smooth map φ with a two-sided inverse).

<u>Corollary.</u> If $\varphi: W \longrightarrow U$ is a smooth map with a smooth inverse $\varphi^{-1}: U \longrightarrow W$, while $c: I \longrightarrow W$ is a path, then $\varphi \circ c$ satisfies Lagrange's equation with respect to a function L and a form ω if and only if c satisfies Lagrange's equation with respect to $(T.\varphi)^* L$ and $\varphi^* \omega$.

This invariance under any change of coordinates is the advantage of Lagrange's equation over the original Newtonian equations. In the next section we give another explanation of this invariance.

The <u>conservative case</u> is that in which the 1-form ω on U which represents the work is the differential

$$\omega = d\mathcal{V}$$

of a smooth function $\mathcal{V} : U \longrightarrow \mathbb{R}$ called the <u>potential.</u> The function $\mathcal{V}$ can be lifted to the tangent bundle as $\mathcal{V} \circ \pi : T.U \longrightarrow \mathbb{R}$, and we often write $\mathcal{V}$ for $\mathcal{V} \circ \pi$. Then the Lagrangian function L is by definition

$$L = \mathcal{T} - \mathcal{V} : T.U \longrightarrow \mathbb{R}$$

and Lagrange's equations clearly take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0 \ , \ i = 1, \ldots, n.$$

Here L is of course short for $L \circ \tilde{c}$, c a path. Expanding the derivative, these equations are

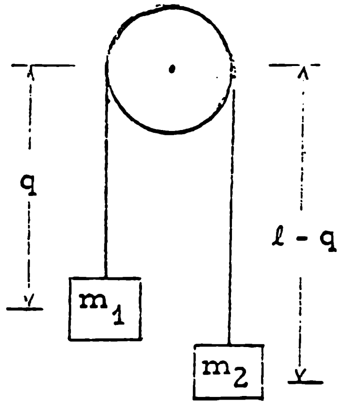
$$\sum_{j=1}^{n} \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \ddot{q}^j + \sum_{j=1}^{n} \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} \dot{q}^j - \frac{\partial L}{\partial q^i} = 0,$$

a system of n second order differential equations.

We give an application of Lagrange's equation. Recall $L = (\mathcal{T} - \mathcal{V}) : T.U \longrightarrow \mathbb{R}$ and Lagrange's equation is

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}^i}\right] = \frac{\partial L}{\partial q^i} \qquad i = 1, \ldots, n .$$

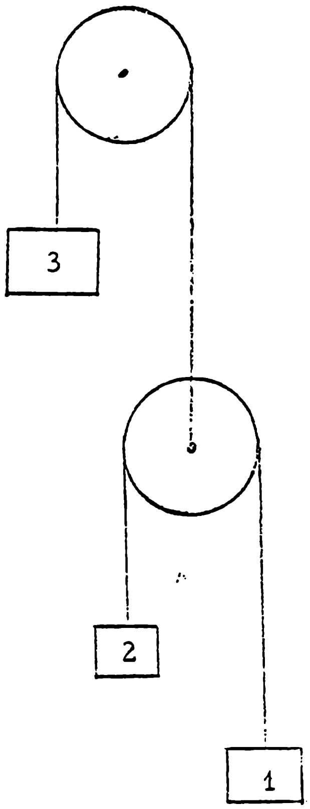
We will use this to derive the equation of motion (acceleration) for Atwood's machine:



a (massless) pulley supporting a (massless) chain with a weight m_1 a distance q below the center of the pulley and a weight m_2 a distance $(\ell - q)$ below the center of the pulley, where ℓ is fixed. As indicated, there is just one generalized coordinate q.

We calculate $\mathcal{T} = \frac{1}{2} m_1\dot{q}^2 + \frac{1}{2} m_2\dot{q}^2$ and $\mathcal{V} = -m_1gq - m_2g(\ell - q) =$ $(m_2 - m_1)gq +$ constant, therefore $L = \frac{1}{2}(m_1 + m_2)\dot{q}^2 - (m_2 - m_1)gq +$ constant, We differentiate $\frac{\partial L}{\partial \dot{q}} = (m_1 + m_2)\dot{q}$, $\frac{d}{dt}[\frac{\partial L}{\partial \dot{q}}] = (m_1 + m_2)\ddot{q}$ and $\frac{\partial L}{\partial q} = -(m_2 - m_1)g$. We equate $(m_1 + m_2)\ddot{q} = (m_1 - m_2)g$ and solve to get $\ddot{q} = ((m_1 - m_2)/(m_1 + m_2))g$, a second order differential equation which can readily be integrated.

We give a problem for mathematicians: Find the accelerations of the system shown below. The mass of the weights is indicated by the numbers labeling them.



§7 Hamilton's Principle.

The fact that a path $c: I \longrightarrow U$ satisfies Lagrange's equation is, we have seen, independent of the choice of coordinates in the configuration space U. This fact can also be explained by Hamilton's principle, which asserts that the solutions of Lagrange's equations are exactly the curves which "minimize" a certain integral formed from the Lagrangian function L. To cover the most general case, we will assume that L depends not only on position and velocity but also on time; that is, that L is a smooth function $L: T.U \times I \longrightarrow \mathbb{R}$. In coordinates, this means that $L(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n, t) \in \mathbb{R}$.

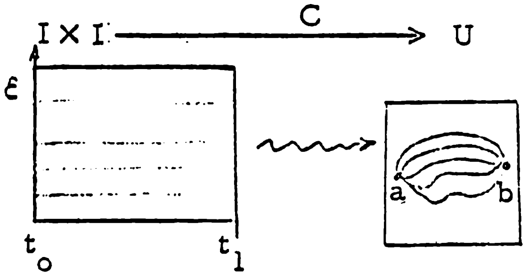
Given fixed points a and b in U, we consider paths c_o from a to b; that is, $c_o: I \longrightarrow U$ with $t_o, t_1 \in I$ and $c_o(t_o) = a$, $c_o(t_1) = b$. We can lift c_o to a path $\tilde{c}_o: I \longrightarrow T.U \times I$, the identity on I. We want to compare c_o with other smooth paths with the same endpoints a and b at the same time t_o and t_1. For any path c, let $J(c) = \int_{t_o}^{t_1} (L\tilde{c})dt$. We want to prove

Hamilton's Principle: If $c_o: I \longrightarrow U$ is a path from a to b in U, then c_o satisfies Lagranges's equations for L if and only if the corresponding integral $J(c)$ is "stationary" for $c = c_o$.

It remains to explain "stationary" for functions of paths like J. In the simpler case of a function f of a real number ε, we say that f is stationary at ε_o if $\frac{df}{d\varepsilon} = 0$ for $\varepsilon = \varepsilon_o$. Thus a function is stationary at a maximum, at a minimum, and at other points (horizontal inflections). Similarly, $J(c_o)$ will be stationary for c_o if c_o is a minimum $(J(c_o) \leq J(c))$ or a maximum or ... . More exactly, J is stationary if the corresponding function of ε is stationary at c_o whenever c_o is embedded in a one-parameter family of paths. Such a family is described by a smooth function

$$C: I \times I \longrightarrow U,$$
$$(\varepsilon, t) \rightsquigarrow C(\varepsilon, t), \qquad \varepsilon \in I, \ t \in I$$

where $C(\varepsilon, t_o) = a$, $C(\varepsilon, t_1) = b$, and $C(0,t) = c_o(t)$. In other words, for each ε, $C(\varepsilon, -)$ is a path from a to b, while for $\varepsilon = 0$, $C(0,-)$ is the given path c_o. The situation is that of the following figure



By lifting each path, we get $\tilde{C}: I \times I \longrightarrow T.U \times I$. Then $J(\tilde{C}) = \int_{t_o}^{t_1} (L\tilde{C})dt$, is a function of ε. Calculate

$$\frac{dJ(\tilde{C})}{d} = \int_{t_o}^{t_1} \frac{\partial}{\partial\varepsilon}(L\tilde{C})\, dt = \int_{t_o}^{t_1} \left(\sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}^i} \frac{\partial \dot{q}^i}{\partial \varepsilon} + \sum_{i=1}^{n} \frac{\partial L}{\partial q^i} \frac{\partial q^i}{\partial \varepsilon}\right) dt.$$

Note that $\frac{\partial \dot{q}^i}{\partial \varepsilon} = \frac{d}{dt}\left(\frac{\partial q^i}{\partial \varepsilon}\right)$ and the integration-by-parts formula

$$\int_{t_o}^{t_1} U \frac{dV}{dt}\, dt = -\int_{t_o}^{t_1} \frac{dU}{dt} V\, dt + UV \Big|_{t_o}^{t_1} .$$

Therefore, $\frac{\partial J(\tilde{c})}{\partial \varepsilon} = \int \left[\sum [-\frac{d}{dt} (\frac{\partial L}{\partial \dot{q}^i})](\frac{dq^i}{d\varepsilon}) + \frac{\partial L}{\partial q^i} \frac{\partial q^i}{\partial \varepsilon} \right] dt + \underbrace{\sum \frac{dL}{\partial \dot{q}^i} \frac{\partial q^i}{\partial \varepsilon} \Big|_{t_o}^{t_1}}_{=0}$

(the last part equals zero because the paths have the same endpoints)

$$= \sum_{i=1}^{n} \int [-\frac{d}{dt} (\frac{\partial L}{\partial \dot{q}^i}) + \frac{\partial L}{\partial q^i}](\frac{\partial q^i}{\partial \varepsilon}) \, dt \; .$$

Therefore, if the Lagrange equations hold along c_o, then $\frac{dJ(\tilde{c})}{d\varepsilon} = 0$.

Conversely, we show for a path c_o that if $J(\tilde{C})$ is stationary for every one parameter family C containing c_o as above, then c_o must satisfy Lagrange's equations. The proof will use the "variation" of the path given by a smooth $\eta : I \longrightarrow \mathbb{R}^n$ with $\eta(t_o) = \eta(t_1) = 0$. The "varied" path is the one parameter family C with

$$C(\varepsilon, t) = c_o(t) + \varepsilon \eta(t)$$

(assume that $\eta(t)$ is small, so that $C(\varepsilon, t)$ still lies in the open set U of $\mathbb{R}^n$).

Then for the coordinates we have

$$q^i C(\varepsilon, t) = q^i \circ c_o + \varepsilon q^i \circ \eta \quad , \quad \frac{\partial q^i}{\partial \varepsilon} = q^i \eta .$$

For this family the calculation above shows that

$$\frac{dJ(\tilde{C})}{d\varepsilon} \Big|_{\varepsilon = 0} = \sum_{i=1}^{n} \int_{t_o}^{t_1} [-\frac{d}{dt} (\frac{\partial L}{\partial \dot{q}^i} \tilde{c}_o) \quad \frac{\partial L}{\partial q^i} \tilde{c}_o] \, q^i \eta \, dt = 0.$$

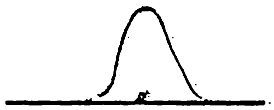
This holds for all η; we wish to conclude that the expression in brackets is zero for each i, for this will give us Lagrange's equations. If we call this expression M, this will follow from the following lemma.

<u>Lemma.</u> If $M: I \longrightarrow \mathbb{R}$ is a smooth function and if

$$\int_{t_o}^{t_1} M\eta\, dt = 0$$

for all smooth functions $\eta: I \longrightarrow \mathbb{R}$ with $\eta(t_o) = \eta(t_1) = 0$, then $M(t) = 0$ for all t with $t_o \leq t \leq t_1$.

<u>Proof.</u> Suppose instead that $M(t) \neq 0$ for some $t = t_2$, say that $M(t_2) > 0$. Then $M(t) > 0$ on some small interval about t_2 and we can choose a "bump" function $b: I \longrightarrow \mathbb{R}$ which is smooth, zero outside the interval, positive inside this interval, and 1 at t_2:



t_2

Then choosing $\eta = bM$ as variation, the hypothesis gives

$$\int_{t_o}^{t_1} M\eta\, dt = \int_{t_o}^{t_1} bM^2\, dt > 0.$$

This contradiction gives $M = 0$, as desired.

The methods used here are those of the Calculus of Variations. The result can be formulated more generally, as follows

Given $h: T.U \times I \longrightarrow \mathbb{R}$, consider paths $c_o: I \longrightarrow U$ which make $\int_{t_o}^{t_1} h\tilde{c}\, dt$ stationary in comparison with other paths c, $c(t_o) = c_o(t_o)$, $c(t_1) = c_o(t_1)$. A necessary condition for this is Euler's equation:

$$\frac{d}{dt}\left(\frac{\partial h}{\partial \dot{q}^i}\right) = \frac{\partial h}{\partial q^i}, \quad i = 1, \ldots, n.$$

In the special case when h is the Lagrangian function L_1, Euler's equations are Lagrange's equations. In more general treatments, the smooth paths used above can be replaced by "piecewise" smooth paths.

§ 8 Bilinear and Quadratic Forms

The kinetic energy $\mathcal{J}$ is usually a quadratic function of the velocities; that is, $\mathcal{J} : T.U \longrightarrow \mathbb{R}$ restricted to the fiber (tangent space) over a point of U is a quadratic function on that tangent space. We now study certain properties of such quadratic functions.

Let V be a finite dimensional vector space. Consider a function $B: V \times V \longrightarrow \mathbb{R}$ $((v, w) \rightsquigarrow B(v, w))$. We define B to be bilinear if $B(v, w)$ is linear in v (with w fixed) and linear in w (with v fixed). We define $Q: V \longrightarrow \mathbb{R}$ to be quadratic when,

1° $Q(-v) = Q(v)$

2° $Q(u+v) - Q(u) - Q(v) \overset{\text{def.}}{=} 2Q^\flat(u, v)$ is bilinear in u and v.

That is, Q determines a symmetric bilinear function $Q^\flat$.

As a consequence of bilinearity, we have

$$Q(u+v+w) - Q(u) - Q(v+w) = Q(u+v) - Q(u) - Q(v) + Q(u+w) - Q(u) - Q(w).$$

Letting $u = v = -w$, $Q(u) - Q(u) = Q(2u) - Q(u) - Q(u) + 0 - Q(u) - Q(u)$ and thus $Q(2u) = 4Q(u)$. (We must have $Q(0) = 0$ since $Q^\flat(0, 0) = 0$.)

The assignment $Q \underset{\text{quadratic}}{\rightsquigarrow} Q^\flat \overset{\text{def.}}{=} \frac{1}{2}[Q(u+v) - Q(u) - Q(v)]$ symmetric and bilinear, has an inverse $B \rightsquigarrow B^\#$. Define $B^\#(u) = B(u, u)$, for B symmetric and bilinear. Clearly, $(Q^\flat)^\# = Q$, since

$$(Q^b)^\#(u) = Q^b(u,u) = \frac{1}{2}[Q(u+u) - Q(u) - Q(u)] = \frac{1}{2}[4Q(u) - 2Q(u)] = Q(u).$$

Conversely, $B^\#$ is quadratic:

1° $B^\#(-u) = B(-u,-u) = B(u,u)$

and

2° $B^\#(u+v) - B^\#(u) - B^\#(v)$ $\left(\overset{\text{def}}{=\!=}\ 2(B^\#)^b(u,v)\right) = B(u+v,u+v) - B(u,u) - B(v,v)$

$$= B(u,v) + B(v,u) = 2B(u,v).$$

And from this calculation, clearly $(B^\#)^b = B$.

Given $Q: V \longrightarrow \mathbb{R}$ quadratic and $W \xrightarrow{\varphi} V$ a linear transformation, then $Q\varphi: W \longrightarrow \mathbb{R}$ is quadratic. For the proof, note

$$Q = V \xrightarrow{\Delta} V \times V \xrightarrow{B} \mathbb{R}$$
$$v \rightsquigarrow (v,v) \rightsquigarrow B(v,v) = Q(v)$$

where $B = Q^b$ is symmetric and bilinear. Check that $W \times W \xrightarrow{\varphi \times \varphi} V \times V \xrightarrow{B} \mathbb{R}$ is symmetric and bilinear. The rest of the proof is indicated by the commutative diagram:

$$\begin{tikzcd} W \arrow[r, "\varphi"] \arrow[d, "\Delta"'] & V \arrow[d, "\Delta"'] \arrow[dr, "Q"] & \\ W \times W \arrow[r, "\varphi \times \varphi"] & V \times V \arrow[r, "B"] & \mathbb{R} \end{tikzcd}$$

Choosing a basis $e_1, \ldots, e_n$ for V, and letting $v = \sum_{i=1}^{n} q^i e_i$, we have $B(\sum_i q^i e_i, \sum q^j e_j) = \sum_{i,j} q^i q^j B(e_i, e_j)$. Defining $g_{ij} = B(e_i, e_j)$, we have $Q(v) = \sum_{i,j=1}^{n} g_{ij} q^i q^j$; we may call $\|g_{ij}\|$ the matrix of Q for the basis e_i.

Consider the functions indicated in the following diagram

$$\begin{tikzcd} \mathbb{R} & T.U \arrow[l, "\dot q^1"', shift right] \arrow[l, "\dot q^n", shift left] \arrow[r, "J", shift left] \arrow[r, "L"', shift right] \arrow[d, "\pi"] & \mathbb{R} \\ \mathbb{R} & U \arrow[l, "q^1"', shift right] \arrow[l, "q^n", shift left] \arrow[r, "\mathcal{V}"] & \mathbb{R} \end{tikzcd}$$

The $\dot{q}^i$ and q^i are the usual coordinates and $\mathcal{J}$ is quadratic and positive definite on each fiber, thus $\mathcal{J}$ is a Riemann metric on U. $L \overset{\text{def}}{=\!=} \mathcal{J} - \mathcal{V}\pi$ is a quadratic plus a constant on each fiber. Let $\mathcal{J} = \sum_{i,j=1}^{n} g_{ij}\dot{q}^i\dot{q}^j$ where $g_{ij}: U \longrightarrow \mathbb{R}$, that is for $u \in U$, $(g_{ij}(u))$ is the positive definite symmetric matrix associated with the quadratic form $\mathcal{J}$ induces on T_uU, the fiber over U. But we have already noted that such a Riemann metric gives an isomorphism

$$T_uU \longrightarrow (T_uU)^* = T^uU$$

of the tangent space to its dual, the cotangent space. This isomorphism carries a point with coordinates $(\dot{q}^1, \ldots, \dot{q}^n)$ to the point with coordinate p_i, where

$$p_i = \sum g_{ij}\dot{q}^j = \frac{\partial \mathcal{J}}{\partial \dot{q}^i} \quad .$$

If we apply this isomorphism to each fiber of the tangent bundle, we get a smooth map $\mathcal{L}: T.U \longrightarrow T^{\cdot}U$ called the Legendre transformation, as in the diagram

$$\begin{tikzcd}
\mathbb{R} & T.U \arrow[l, shift right=1ex, "\dot{q}^1"'] \arrow[l, shift left=1ex, "\dot{q}^n"] \arrow[rr, "\mathcal{L}"] \arrow[dr, "\pi."'] & & T^{\cdot}U \arrow[r, shift left=1ex, "p^1"] \arrow[r, shift right=1ex, "p^n"'] \arrow[dl, "\pi^{\cdot}"] & \mathbb{R} \\
& & U \arrow[r, shift left=1ex, "q^1"] \arrow[r, shift right=1ex, "q^n"'] & \mathbb{R} &
\end{tikzcd}$$

The coordinates for $T.U$ are $q^1\pi., \ldots, q^n\pi., \dot{q}_1, \ldots, \dot{q}_n$ and for $T^{\cdot}U$ they are $q^1\pi^{\cdot}, \ldots, q^n\pi^{\cdot}, p_1, \ldots, p_n$. The map $\mathcal{L}$ is then defined by

$$q^i\pi^{\cdot}\mathcal{L} = q^i\pi. \quad , \quad p_i\mathcal{L} = \frac{\partial \mathcal{J}}{\partial \dot{q}^i} \quad .$$

We now consider the properties of this Legendre transformation $\mathcal{L}$

§9 The Legendre Transformation

Suppose U is an open set in $\mathbb{R}^n$ with coordinates $q^1, \ldots, q^n$ and we are given a smooth kinetic energy function $\mathcal{T}$ from $T.U$ to $\mathbb{R}$. We shall assume $\mathcal{T}$ quadratic; i.e., $\mathcal{T} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial \dot{q}^i} \dot{q}^i$. Such a function determines a smooth function $\mathcal{L}$ defined by $p_i \mathcal{L} = \frac{\partial \mathcal{T}}{\partial \dot{q}^i}$, $q^i \pi^{\cdot} \mathcal{L} = q^i \pi_{\cdot}$. We set $\hat{\mathcal{T}} = \mathcal{T}\mathcal{L}^{-1}$. From the quadratic assumption on $\mathcal{T}$ we have

$$\begin{tikzcd} \mathbb{R} & T.U \arrow[l, "\mathcal{T}"'] \arrow[r, "\mathcal{L}"] \arrow[dr, "\pi_{\cdot}"'] & T^{\cdot}U \arrow[ll, bend right, "\hat{\mathcal{T}}"'] \arrow[d, "\pi^{\cdot}"] \\ & & U \end{tikzcd}$$

$$2d\mathcal{T} = \sum^{n} d\left(\frac{\partial \mathcal{T}}{\partial \dot{q}^i}\right)\dot{q}^i + \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial \dot{q}^i} d\dot{q}^i .$$

Subtract $\quad d\mathcal{T} = \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial q^i \pi_{\cdot}} dq^i \pi_{\cdot} + \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial \dot{q}^i} d\dot{q}^i \quad$ to get

$d\mathcal{T} = \sum_{i=1}^{n} d\left(\frac{\partial \mathcal{T}}{\partial \dot{q}^i}\right)\dot{q}^i - \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial q^i \pi_{\cdot}} dq^i \pi_{\cdot} \qquad$ Then

$$\begin{aligned} d\hat{\mathcal{T}} &= d(\mathcal{T}\mathcal{L}^{-1}) = (\mathcal{L}^{-1})^* \, d\mathcal{T} \\ &= \sum_{i=1}^{n} d\left(\frac{\partial \mathcal{T}}{\partial \dot{q}^i}\mathcal{L}^{-1}\right)\dot{q}^i \mathcal{L}^{-1} - \sum_{i=1}^{n} \frac{\partial \mathcal{T}}{\partial q^i \pi_{\cdot}} \mathcal{L}^{-1} d(q^i \pi_{\cdot} \mathcal{L}^{-1}) . \end{aligned}$$

(Recall that if $X \xrightarrow{\emptyset} Y \xrightarrow{f} \mathbb{R}$ are smooth functions, then f pulls back via $\emptyset$ to the function $f\emptyset$ and we have $d(f\emptyset) = \emptyset^* df$, $\emptyset^*(gdf) = (g\emptyset)\emptyset^* df$.)

Finally, using the defining equations for $\mathcal{L}$,

$$d\hat{\mathcal{T}} = \sum_{i=1}^{n} (dp_i)\dot{q}^i \mathcal{L}^{-1} - \sum_{i=1}^{n} \left(\frac{\partial \mathcal{T}}{\partial q^i \pi_{\cdot}} \mathcal{L}^{-1}\right) dq^i \pi^{\cdot} .$$

But also $d\hat{\mathcal{J}} = \sum_{i=1}^{n} \frac{\partial \hat{\mathcal{J}}}{\partial p_i} dp_i + \sum_{i=1}^{n} \frac{\partial \hat{\mathcal{J}}}{\partial q^i \pi^\cdot} dq^i \pi^\cdot$, and the coefficients of the differentials are unique. Thus

$$\dot{q}^i \mathcal{L}^{-1} = \frac{\partial \hat{\mathcal{J}}}{\partial p_i} \quad \text{and} \quad - \frac{\partial \mathcal{J}}{\partial q^i \pi_\cdot} \mathcal{L}^{-1} = \frac{\partial \hat{\mathcal{J}}}{\partial q^i \pi^\cdot} ,$$

which are the equations of the transferred kinetic energy $\hat{\mathcal{J}}$ in terms of the given one $\mathcal{J}$.

Now if we have a mechanical system with Lagrangian $L = \mathcal{J} - \mathcal{V}$ in which a path c satisfies Lagrange's equations, then these last equations yield Hamilton's equations for c. Let's follow the convention in mechanics of not writing in the maps $\pi_\cdot , \pi^\cdot$ and $\mathcal{L}$, so the last equations are

$$\frac{dq^i}{dt} = \frac{\partial \hat{\mathcal{J}}}{\partial p_i} \quad \text{and} \quad - \frac{\partial \mathcal{J}}{\partial q^i} = \frac{\partial \hat{\mathcal{J}}}{\partial q^i} .$$

By definition of $\mathcal{L}$

$$\frac{dp_i}{dt} = \frac{d}{dt}\left(\frac{\partial \mathcal{J}}{\partial \dot{q}^i}\right) = \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i}\right)$$

$$= \frac{\partial \mathcal{L}}{\partial q^i} = \frac{\partial \mathcal{J}}{\partial q^i} - \frac{\partial \mathcal{V}}{\partial q^i} = - \left(\frac{\partial \hat{\mathcal{J}}}{\partial q^i} + \frac{\partial \mathcal{V}}{\partial q^i} \right).$$

So by setting $\mathcal{H} = \hat{\mathcal{J}} + \mathcal{V}\mathcal{L}^{-1}$, we have Hamilton's equations

$$\frac{dq^i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} \quad \text{and} \quad \frac{dp_i}{dt} = - \frac{\partial \mathcal{H}}{\partial q^i} .$$

<u>Exercise:</u> Let $n = 1$ and $\mathcal{J} = g\dot{q}^2$, where $U \xrightarrow{g} \mathbb{R}$. Calculate $\hat{\mathcal{J}}$.

We want to understand better how this scheme produced these equations. Notice first that χ maps $T_a U$ into $T^a U$ -- which is to say $(\chi, 1_U): \pi_{\bullet} \longrightarrow \pi^{\bullet}$ is a morphism of prebundles. Thus we may as well look at χ on each fibre of $T.U$ and paste the fibres together where we must. So consider a finite dimensional vector space V (think of this as $T_a U$ for some $a \in U$):

a) At each point $v \in V$, $T_v V \cong V$ $(\tau_v c \rightsquigarrow w)$, where $c(t) = v + tw$ is a curve $I \longrightarrow V$. (Identify V with $T_v V$ by this isomorphism.)

b) At each $v \in V$, $T^v V \cong V^*$ $(d_v f \rightsquigarrow \overline{d_v f}\ )$ where

$\overline{d_v f}\,(w) = \frac{d}{dt}\,(f(v + tw))\Big|_{t=0}$ and $V \xrightarrow{f} \mathbb{R}$. (Again identify V^* and $T^v V$.)

c) $T^{\bullet} V$ can be identified with $V \times V^*$ via

$$(v, d_v f) \rightsquigarrow (v, \overline{d_v f})$$

$$\begin{tikzcd} T^{\bullet}V \arrow[r] \arrow[d, "\pi^{\bullet}"'] & V \times V^* \arrow[d, "\pi"] \\ V \arrow[r, "="] & V \end{tikzcd}$$

where $\pi(v, \alpha) = v$.

We have another projection $V \times V^* \longrightarrow V^*$ $(v, \alpha) \rightsquigarrow \alpha$.

d) A 1-form ω on V is a smooth function $\omega: T$— $T^{\bullet}V$ such that $\pi \circ \omega = 1_V$. Composition of σ with a 1-form ω produces a smooth function

$$\chi_{\omega} = \sigma \circ \omega .$$

$$\begin{tikzcd} V \times V^* \arrow[r, "\sigma"] \arrow[d, shift left, "\pi"] & V \\ V \arrow[u, shift left, "\omega"] \arrow[ur, "\chi_{\omega}"'] & \end{tikzcd}$$

e) In particular, suppose $L: V \longrightarrow \mathbb{R}$ is smooth (think of L as a Lagrangian function restricted to one fiber of $T.U$). Then dL is a 1-form and so determines $V \xrightarrow{\mathcal{L}_{dL}} V^*$. To compare this with the example at the beginning of this section, let the potential energy be zero so $L = \mathcal{J}$. The first defining equation for $\mathcal{L}$ says $\mathcal{L}_{d\mathcal{J}} = \sigma \circ d\mathcal{J}$ when written in coordinates. Returning to the general case, the explicit formula for $\mathcal{L}_{dL}$ gives each value $\mathcal{L}_{dL}V$ as a function of w:

$$\mathcal{L}_{dL}V(w) = \frac{d}{dt} L(v+tw)\Big|_{t=0} .$$

f) Let $e_1, \ldots, e_n$ be a basis for V with coordinates $e^1, \ldots, e^n$; then $e^1, \ldots, e^n$ are a dual basis for V^* with coordinates $e_1, \ldots, e_n$. The formula for $\mathcal{L}_{dL} = \mathcal{L}$ in these bases is

$$\mathcal{L}\,v(w) = \sum_{j=1}^{n} \left(\frac{\partial L}{\partial e^j} \Big|_v \right) e^j w .$$

Apply e_i to both sides and use $e_i e^j = \delta^i_j$ to get $e_i \mathcal{L}\, v = \frac{\partial L}{\partial e^i}\Big|_v$ or, as functions

$$e_i \mathcal{L} = \frac{\partial L}{\partial e^i} \quad , \quad i = 1, \ldots, n$$

In the mechanical rotation $e^i = \dot{q}^i$ and $e_i = p_i$, since $V = T_a U$ and $V^* = T^a U$ for $a \in U$, U open in $\mathbb{R}^n$, and the result reads

$$p_i \mathcal{L} = \frac{\partial L}{\partial \dot{q}^i}$$

as before.

The Hamiltonian function arises from asking the question: when is $\mathcal{L}$ invertible? (This is probably not the way Hamilton found it.)

By the Inverse Function Theorem (see Abraham p. 10), $\mathcal{L}^{-1}$ exists if and only if the matrix $\left(\dfrac{\partial^2 L}{\partial e^i \partial e^j}\Big|_v \right)$ is non-singular at every $v \in V$.

g) Suppose $\mathcal{L}$ invertible. Does $\mathcal{L}^{-1}$ come from a smooth function on V^* in the same manner $\mathcal{L}$ came from a smooth L on V? That is, is there $V^* \xrightarrow{H} \mathbb{R}$ such that $\mathcal{L}^{-1} = \pi \circ dH$?

$$\begin{tikzcd} V \times V^* \arrow[r, "\pi"] \arrow[d, shift left, "\sigma"] & V \arrow[dl, "\mathcal{L}^{-1}"] \\ V^* \arrow[u, shift left, "dH"] & \end{tikzcd}$$

From part f) we have

$$dL = \sum_{i=1}^{n} (e_i \mathcal{L}) de^i .$$

Let's try the formula dual to this one:

We want H so that $dH = \sum_{i=1}^{n} (e^i \mathcal{L}^{-1}) de_i$. Use the derivation property of d:

$$\sum_{i=1}^{n} (e^i \mathcal{L}^{-1}) de_i = d\Big[\sum_{i=1}^{n} (e^i \mathcal{L}^{-1}) e_i\Big] - \sum_{i=1}^{n} e_i d(e^i \mathcal{L}^{-1}).$$

The second term on the right is

$$-\sum_{i=1}^{n} e_i \mathcal{L} \mathcal{L}^{-1} d(e^i \mathcal{L}^{-1}) = -(\mathcal{L}^{-1})^* \sum_{i=1}^{n} (e_i \mathcal{L}) de^i$$

$$= -(\mathcal{L}^{-1})^* dL = -d(L \mathcal{L}^{-1}).$$

Substitute this in our conjectured dH to get

$$dH = d\Big[\sum_{i=1}^{n} (e^i \mathcal{L}^{-1}) e_i - L \mathcal{L}^{-1}\Big],$$

so H should be

$$H = \sum_{i=1}^{n} (e^i \mathcal{L}^{-1}) e_i - L \mathcal{L}^{-1} .$$

The steps reverse so this is indeed the right formula. On elements,

$$H_y = \mathcal{L}^{-1}_y \cdot y - L \mathcal{L}^{-1}_y \qquad \text{for} \quad y \in V^*$$

usual inner product

and in mechanical notation

$$H = \sum_{i=1}^{n} \dot{q}^i p_i - L.$$

(Notice that we are leaving out $\mathcal{L}$ and $\mathcal{L}^{-1}$ as is customary in mechanics.) If, in particular, $L = \mathcal{J} - \mathcal{V}$ where $\mathcal{V}$ is a function of only the q^i's and $\mathcal{J}$ is quadratic as in our example, then

$$H = \sum_{i=1}^{n} \dot{q}^i \frac{\partial \mathcal{J}}{\partial \dot{q}^i} - L = 2\mathcal{J} - L = \mathcal{J} + \mathcal{V}$$

Given the Lagrangian L, define the <u>Action</u> $A: V \longrightarrow \mathbb{R}$, and <u>Energy</u> $E: V \longrightarrow \mathbb{R}$ by $Av = \langle \mathcal{L}_{dL} v, v \rangle$, $Ev = \langle \mathcal{L}_{dL} v, v \rangle - Lv$. Both are smooth functions.

The last lecture proved most of the following theorem:

<u>Theorem</u> (Legendre): If V is a finite dimensional real vector space of dimension n, then

1) for each $v \in V$, there is a natural isomorphism $T^v V \cong V^*$ (which we consider equality below)

2) for each smooth $L: V \longrightarrow \mathbb{R}$, there is a smooth

$$V \xrightarrow{\mathcal{L}_{dL}} V^* \qquad (v \rightsquigarrow d_v L)$$

In coordinates

$$dL = \sum_{i=1}^{n} (e_i \mathcal{L}_{dL}) de^i$$

with $e_1, \ldots, e_n$ a basis for V. $\mathcal{L}_{dL}$ is called the <u>Legendre transformation for L</u>.

3) the function $\mathcal{L} = \mathcal{L}_{dL}$ is invertible if and only if the matrix $\left(\frac{\partial^2 L}{\partial e^i \partial e^j}\Big|_v\right)$ is non-singular at each $v \in V$. If this is so $\mathcal{L}^{-1}: V^* \longrightarrow V$ is the Legendre transformation for $H: V^* \longrightarrow \mathbb{R}$ defined by

$$Hy = \langle y, \mathcal{L}^{-1} y \rangle - L\mathcal{L}^{-1} y \quad = E\mathcal{L}^{-1} y .$$

4) in particular, if L is quadratic, then $\mathcal{L}$ is the isomorphism of V with its dual given by the inner product induced by L and $E = L$ in this case.

[See Sternberg pp. 150-153, Goldstein pp. 215-   , Abraham §17 .]

Corollary 1. If U open in $\mathbb{R}^n$, I open interval $\subset \mathbb{R}$ and $L: T.U \times I \longrightarrow \mathbb{R}$ is smooth, then L on each fibre determines

$$\begin{tikzcd}
T.U \times I \arrow[rr, "\mathcal{L}"] \arrow[dr] & & T^{\cdot}U \times I \arrow[dl] \\
& U \times I &
\end{tikzcd}$$

and all parts of the theorem hold for this $\mathcal{L}$. (Abraham calls $\mathcal{L}$ the fibre derivative of L.)

Corollary 2. Let c be a path in U, $\tilde{c}$ lifted path in $T.U$. If $\tilde{c}$ satisfies Lagrange's equations for L, then $\mathcal{L}\tilde{c}$ satisfies the canonical differential equations for H (Hamilton's equations):

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i} .$$

Exercise Prove Corollary 2.

Forget the projection, so $T^{\cdot}U$ is an open set in $\mathbb{R}^{2n}$ and on it sits a first order differential equation -- Hamilton's equation of Corollary 2. We shall consider changes of coordinates in this 2n-dimensional space which leave this differential equation invariant. To do this we need to consider 2-forms, and in general k-forms.

§10. Vector Fields

Let U be open in $\mathbb{R}^n$ with coordinates $q^1, \ldots, q^n$. A vector field X on U is a cross-section of the tangent bundle of U; i.e., $\pi_\cdot X(a) = a$ for all $a \in U$.

$$\begin{tikzcd} T_\cdot U \arrow[d, bend left, "\pi"] \\ U \arrow[u, bend left, "X"] \end{tikzcd}$$

For example, for given coordinates q^i, we can define a vector field D "along the axis q^i" by $D^i(a) = (a, \overline{D}^i a)$, where $\overline{D}^i(a) = \tau_a$ (path along i^{th} coordinate axis) = unit vector in i^{th} direction in $T_a U$, for $i = 1, \ldots, n$. This clearly defines a cross-section D^i of the tangent bundle. The set of vector fields on U is an $\mathcal{F}$-module, where $\mathcal{F}$ is the ring of smooth functions $U \longrightarrow \mathbb{R}$. The $\mathcal{F}$-module structure is given by the equations

$$(X_1 + X_2)a = X_1 a + X_2 a$$
$$fX_1(a) = f(a) \cdot X_1 a ,$$

for $a \in U$, $f \in \mathcal{F}$ and X_1, X_2 vector fields. The vectors $D^i(a)$, $i = 1, \ldots, n$ form a basis of $T_a U$, so

$$Xa = \sum_{i=1}^{n} (X_i a) D^i(a);$$

thus $X = \sum_{i=1}^{n} X_i D^i$, where the functions $X_i \colon U \longrightarrow \mathbb{R}$ are smooth and unique. This says that the vector fields $D^1, \ldots, D^n$ are a basis for the set of vector fields on U as a real vector space.

Each vector field X produces a function called the Lie derivative $\mathcal{F} \xrightarrow{L_X} \mathcal{F}$ with

$$L_X f(a) = \langle d_a f, Xa \rangle = \text{derivative of } f \text{ along } X.$$

Here we need "smooth" to mean C^∞, since otherwise $L_X f$ has one lower order of differentiability than f. The function L_X has the properties

(1) L_X is $\mathbb{R}$-linear,

(2) $L_X(f \cdot g) = f \cdot L_X g + g \cdot L_X f$.

Property (1) is a consequence of the linearity of d_a and $\langle\ , Xa\rangle$. For (2)

$$\begin{aligned}
L_X(f\cdot g)a &= \langle d_a f\cdot g, Xa\rangle \\
&= \langle f(a)\cdot d_a g + g(a)\cdot d_a f, Xa\rangle \\
&= f(a)\langle d_a g, Xa\rangle + g(a)\langle d_a f, Xa\rangle \\
&= (f\cdot L_X g + g\cdot L_X f)a .
\end{aligned}$$

In coordinates, $X = \sum_{i=1}^{n} X_i D^i$ so

$$\begin{aligned}
L_X f &= \left\langle \sum_{j=1}^{n} \frac{\partial f}{\partial q^i}\, dq^j ,\ \sum_{i=1}^{n} X_i D^i \right\rangle \\
&= \sum_{i=1}^{n} X_i \sum_{j=1}^{n} \frac{\partial f}{\partial q^i} \langle dq^j, D^i\rangle \\
&= \sum_{i=1}^{n} X_i \frac{\partial f}{\partial q^i} , \quad \text{since } \langle dq^j, D^i\rangle = \delta_i^j .
\end{aligned}$$

In particular, if $X = D^i$, then

$$L_{D^i} f = \frac{\partial f}{\partial q^i} ,$$

so L_{D^i} is sometimes written $\partial/\partial q^i$.

__Definition.__ A derivation on the ring $\mathfrak{F}$ is an $\mathbb{R}$-linear function $\theta\colon \mathfrak{F} \longrightarrow \mathfrak{F}$ such that

$$\theta(f\cdot g) = f\cdot \theta g + g\cdot \theta f.$$

Each L_X is a derivation on $\mathfrak{F}$; in fact, these are all the derivations on $\mathfrak{F}$.

__Theorem.__ For every derivation θ on $\mathfrak{F}$, there is a unique vector field X such that $\theta = L_X$.

__Proof.__ Take $a \in U$. Since translations are invertible functions which preserve all differentiable structures, we may as well assume $a = 0$. Since U is open, it contains a ball with center at the origin: for any u in that ball define a path c in U by

$$c(t) = tu.$$

The Fundamental Theorem of Calculus gives us the equation

$$\begin{aligned} fc(1) - fc(0) &= \int_0^1 \frac{dfc}{dt}\, dt \\ &= \int_0^1 \sum_{i=1}^{n} \frac{\partial fc}{\partial q^i}\, q^i(u)\, dt. \end{aligned}$$

Set $h_i(u) = \int_0^1 \frac{\partial fc}{\partial q^i}\, dt$ and notice that $q^i(u)$ is independent of t:

$$f(u) = f(0) + \sum_{i=1}^{n} h_i(u) q^i(u).$$

Then, by the defining property of θ

$$\begin{aligned} \theta f(0) &= \sum_{i=1}^{n} h_i(0)\theta q^i(0) + q^i(0)\theta h_i(0) \\ &= \sum_{i=1}^{n} \theta q^i(0) \left.\frac{\partial f}{\partial q^i}\right|_0 , \end{aligned}$$

since $q^i(0) = 0$ and $h_i(0) = \left.\dfrac{\partial f}{\partial q^i}\right|_0$. For any $f \in \mathcal{F}$ then we have

$$\theta f = \sum_{i=1}^{n} \theta q^i \cdot \frac{\partial f}{\partial q^i} ,$$

which is exactly $L_X f$ when $X_i = \theta q^i$. The X_i uniquely determine X, so the theorem is proved.

§11. The Tensor Product

This section begins with the material in MacLane and Birkhoff Algebra, Chapter VI, §§4 and 5, and Chapter IX §§7 and 8.

A tensor is sometimes described by symbols with many indices, upper and lower. To really understand tensors, we must understand their relation to the basic vector space V under discussion. Tensors are in fact elements of new vector spaces built up out of V and its dual space by the operation of tensor product.

Given vector spaces V and W, a tensor product of V and W is a vector space, which we will write $V \otimes W$, together with a bilinear function $\otimes : V \times W \longrightarrow V \otimes W$, which have the following property: if $B: V \times W \longrightarrow U$ is any bilinear function, then there is a unique linear map $F: V \otimes W \longrightarrow U$ such that the diagram below commutes:

$$\begin{tikzcd} V \times W \arrow[r] \arrow[dr, "B"'] & V \otimes W \arrow[d, "F"] \\ & U \end{tikzcd}$$

Briefly, we say that $\otimes$ is "universal" among bilinear functions on $V \times W$. If we write the image of the pair $(v, w) \in V \times W$ under the map $\otimes$ as

$v \otimes w$, then the commutativity of the diagram is expressed by the equation $B(v, w) = F(V \otimes W)$. Similarly, the bilinearity of $\otimes$ is equivalent to the equations

(i) $V \otimes (w_1 \times w_2) = v \otimes w_1 \times v \otimes w_2$, (iii) $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$,

(ii) $v \otimes kw = k(v \otimes w)$, (iv) $kv \otimes w = k(v \otimes w)$

for all $v, v_1, v_2 \in V$; $w, w_1, w_2 \in W$, and $k \in \mathbb{R}$.

The universality of $\otimes$ means that the elements $v \otimes w$ generate $V \otimes W$ as a vector space. Thus

$$V \otimes W = \left\{ \sum_{i=1}^{n} (v_i \otimes w_i) k_i \;\middle|\; v_i \in V,\ w_i \in W,\ k_i \text{ scalars and satisfying the relations (i) - (iv) above} \right\}$$

describes $V \otimes W$ in terms of elements but without bases. This decomposition may be used to prove the existence of the tensor product (<u>Algebra</u>, Ch. IX) What in the world is this space $V \otimes W$? Let $V = W = \mathbb{R}^3$, and let e_1, e_2, e_3 be a basis for $\mathbb{R}^3$. A bilinear function $B: \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow U$ satisfies

$$B\Big(\sum_{i=1}^{3} x_i e_i \, , \sum_{j=1}^{3} y_j e_j\Big) = \sum_{i,j=1}^{3} x_i y_i \, B(e_i, e_j),$$

so B is determined by the 3×3 matrix $(B(e_i, e_j))$. Let U be the artificial space on the basis $\{ e_{ij} \mid i, j = 1, 2, 3 \}$ and set $B(e_i, e_j) = e_{ij}$. Then by the universal property of tensor products we have a unique linear h such that $e_{ij} = B(e_i, e_j) = h(e_i \otimes e_j)$

$$\begin{tikzcd} \mathbb{R}^3 \times \mathbb{R}^3 \arrow[rr, "\otimes"] \arrow[dr, "B"'] & & \mathbb{R}^3 \otimes \mathbb{R}^3 \arrow[dl, dashed, bend left, "h"] \\ & U \arrow[ur, "g"] & \end{tikzcd}$$

On the other hand, define a linear transformation $g: U \longrightarrow \mathbb{R}^3 \otimes \mathbb{R}^3$ by $g(e_{ij}) = e_i \otimes e_j$. Then U is isomorphic to $\mathbb{R}^3 \otimes \mathbb{R}^3$, since

$$gh(e_i \otimes e_j) = e_i \otimes e_j$$

and

$$hg(e_{ij}) = e_{ij} \ .$$

This analysis works for any pair of finite dimensional vector spaces V and W but not in more general situations. If dim V = n and dim W = n, it proves that $V \otimes W$ is a finite dimensional vector space of dimension mn.

In forming the tensor product of given spaces, we must say "<u>a</u> tensor product of V and W" instead of "<u>the</u> tensor product," because for all we know there may be many non-isomorphic spaces $V \otimes W$ and maps $\otimes$ enjoying the above properties. These doubts are removed by the following theorem.

<u>Theorem</u>. Any two tensor products of V and W are isomorphic.

Proof. We can show somewhat more. For let $V \otimes W$ and $V \square W$ be two tensor products of V and W, with associated maps $\otimes$ and $\square$ respectively: that is, $\otimes : (v, w) \rightsquigarrow v \otimes w$ and $\square : (v, w) \rightsquigarrow v \square w \in V \square W$. Then we will show that there is a unique isomorphism $\ell : V \otimes W \longrightarrow V \square W$ such that $\ell(v \otimes w) = v \square w$.

First, since $V \otimes W$ is a tensor product of V and W, we may replace U and B in the definition above by $V \square W$ and $\square$, getting a map f making the following diagram commute

$$\begin{tikzcd} V \times W \arrow[rr, "\otimes"] \arrow[dr, "\square"'] & & V \otimes W \arrow[dl, "B"] \\ & V \square W & \end{tikzcd}$$

But $\square$ is also a tensor product of V and W, so we may reverse the roles of $\otimes$ and $\square$ to get the map g in the following commutative diagram

$$\begin{tikzcd} V \times W \arrow[r, "\otimes"] \arrow[dr, "\square"'] & V \otimes W \arrow[d, shift right, "f"'] \\ & V \square W \arrow[u, shift right, "g"'] \end{tikzcd}$$

Thus commutativity means that for all $(v, w) \in V \times W$, we have

$$f(v \otimes w) = v \square w, \qquad g(v \square w) = v \otimes w \, .$$

In particular, $gf(v \otimes w) = v \otimes w$, $fg(v \square w) = v \square w$. Now use again the fact that $V \otimes W$ is a tensor product, replacing U and B in the definition this time by $V \otimes W$ and $\otimes$. We have just shown that gf and $1_{V \otimes W}$ both make the diagram commute; hence, by theuniqueness assertion in the definition, $1_{V \otimes W} = gf$. Similarly, we show that $fg = 1_{V \square W}$. This means exactly that f is an isomorphism of $V \otimes W \longrightarrow V \square W$ with inverse g.

$$\begin{tikzcd} V \times W \arrow[r, "\otimes"] \arrow[dr, "\otimes"'] & V \otimes W \arrow[d, bend right, "gf"'] \arrow[d, bend left, "1_{V \otimes W}"] \\ & V \otimes W \end{tikzcd}$$

We have shown that if tensor products exist, they are unique, "up to isomorphism". But there might not be any tensor products of V and W at all.

<u>Theorem</u> If V and W are vector spaces then they have a tensor product.

Proof. Even though V and W may have infinite dimension, we can still find bases $\{e_i\}$ for V, $i \in I$, and $\{d_j\}$ for W, $j \in J$. The basis elements are indexed not by integers but by members of the (possibly uncountable) sets I and J; that is, to each element i of I there exists a basis element e_i of V. Then every element v of V will be uniquely expressible as a linear combination of some finite subset of the $\{e_i\}$.

We are likewise free to form a new set of symbols b_{ij}, one for each element (i, j) of the cartesian product $I \times J$. The set of all possible finite expressions $\sum_{k=1}^{m} r_k b_{(k)}$, where each $b_{(k)}$ is a b_{ij}-symbol for some (i, j), and each $r_k \in \mathbb{R}$, forms a perfectly good abstract vector space when we define addition and scalar multiplication in the obvious way. We claim that this new space L is a tensor product of V and W.

Define $\square : V \times W \longrightarrow L$ by $(\sum x^i e_i, \sum y^j d_j) \rightsquigarrow \sum x^i y^i b_{ij}$.

Clearly, $\square$ is bilinear. If now $B: V \times W \longrightarrow U$ is bilinear, we have $B(\sum x^i e_i, \sum y^i d_j) = \sum_i x^i B(e_i, \sum y^j d_j) = \sum_{i,j} x^i y^j B(e_i, d_j)$. Hence if we define $f: L \longrightarrow U$ on the basis elements $\{b_{ij}\}$ of L by the formula $f(b_{ij}) = B(e_i, d_j)$, computation shows that the tensor product diagram commutes, and that this is the only f which will make the diagram commutative. Hence L is a tensor product of V and W.

Denoting the elements b_{ij} by the symbols $e_i \otimes d_j$, we derive as a corollary that if $\{e_i\}$ is a basis of the vector space V and $\{d_j\}$ is a basis of W, then $\{e_i \otimes d_j\}$ is a basis of $V \otimes W$. Notice that not every vector of $V \otimes W$ is of the form $v \otimes w$ for some v in V and w in W. There will usually be sums $\sum_i v_i \otimes w_i$ which cannot be reduced to single terms of the form $v \otimes w$; e.g. $e_1 \otimes d_1 + e_2 \otimes d_2$.

We now define the space of 2-tensors on V as $T_2(V) = V \otimes V$. Since $V \otimes V$ has basis $\{e_i \otimes e_j\}$, we can write any element t of $T_2(V)$ as $t = \sum_{i,j} x^{ij} e_i \otimes e_j$, where the x^{ij} are real numbers. The traditional viewpoint is that the tensor is the array x^{ij}. Of course, the matrix elements depend on which basis of V we pick; we can derive the rule for transforming to the new basis e'_i, where $e^i = \sum a^j_i e'_j$, as follows

$$e_i \otimes e_k = \sum a^j_i e'_j \otimes \sum a^\ell_k e'_\ell = \sum_{j,k} a^j_i a^\ell_k e'_j \otimes e'_\ell$$

Therefore,

$$t = \sum x^{ik} e_i \otimes e_k = \sum_{i,j,k,\ell} x^{ik} a^j_i a^\ell_k e'_j \otimes e'_\ell .$$

So
$$x'^{j\ell} = \sum_{i,k} x^{ik} a^j_i a^\ell_k \tag{1}$$

Thus we have replaced the usual opaque definition that a tensor is an array of coordinates relative to a basis which transforms according to equation (1).

Our new definition helps us see how tensors behave under linear transformations. In the most general case, we have linear maps $g: V \to V'$ and $h: W \to W'$ giving us a map $g \times h: V \times W \longrightarrow V' \times W'$ defined by sending (v, w) to (gv, hw). The composite $\otimes(g \times h)$ is a bilinear map on $V \times W$, hence by the definition of tensor product, factors uniquely through

$V \otimes W$ as below, giving us a new map $g \otimes h: V \otimes W \longrightarrow V' \otimes W'$. Thus $g(v \otimes w) = gv \otimes hw$. It is also easy to see that if $g': V' \longrightarrow V''$ and $h': W' \longrightarrow W''$, then $g'g \otimes h'h = (g' \otimes h')(g \otimes h)$ (use the uniqueness property in the definition of tensor product).

$$\begin{tikzcd}
V \times W \arrow[rr, "\otimes"] \arrow[dr, "g \times h"'] & & V \otimes W \arrow[dd, dashed, "g \otimes h"] \\
& V' \times W' \arrow[dr, "\otimes"'] & \\
& & V' \otimes W'
\end{tikzcd}$$

A historical note: we have defined elements of the tangent space to be "contravariant" vectors. Actually, modern usuage considers tangent vectors to be covariant in nature, since given $\varphi: V \longrightarrow U$, the induced map φ_* on the tangent spaces maps $T_{\cdot}(V) \longrightarrow T_{\cdot}(U)$; if the induced map reversed the arrow, taking $T_{\cdot}(U) \longrightarrow T_{\cdot}(V)$, it would be not "co" but "contra." The reason for the traditional terminology (which we will stick to) is that the coordinate transformations under change of basis, given by (1), do interchange the position of primed and unprimed letters.

Similar arguments to the ones we have been using prove that either $V \otimes (W \otimes U)$ or $(V \otimes W) \otimes U$ is a universal object for trilinear maps from $V \times W \times U$. (cf. MacLane and Birkhoff, Ch. 16). Thus we can unambiguously define the tensor product $V \otimes W \otimes U$ to be $V \otimes (W \otimes U)$ and similarly the tensor product of any finite number of vector spaces. The elements of $T_n(V) = V \otimes V \otimes \ldots \otimes V$ (n times) are called n-contravariant tensors, or contravariant tensors of rank n.

Theorem. Let V and W be finite-dimensional vector spaces, $V \otimes W$ a tensor product, and $V^* \square W^*$ a tensor product of the dual spaces V^* and W^*. Then there is an isomorphism $V^* \square W^* \xrightarrow{\cong} (V \otimes W)^*$.

Proof. Recall that V^* was the set of linear maps $f: V \longrightarrow \mathbb{R}$. If now $(f, g) \in V^* \times W^*$, $f \otimes g: V \otimes W \longrightarrow \mathbb{R} \otimes \mathbb{R} \cong \mathbb{R}$; therefore $f \otimes g \in (V \otimes W)^*$. (cf. the corollary above). Furthermore, the map $(f, g) \rightsquigarrow f \otimes g \in (V \otimes W)^*$ is bilinear. By the universality of $\square$, there is a map $h: V^* \square W^* \longrightarrow (V \otimes W)^*$ such that $h(f \square g) = f \otimes g$. We wish to show that h is an isomorphism. It suffices to prove that h maps the basis $\{e^i \square d^j\}$ of $V^* \square W^*$ to a basis of $(V \otimes W)^*$. Thus it will be enough to show that $\{e^i \otimes d^j\}$ is the dual basis of $\{e_i \otimes d_j\}$. But by the definition of the tensor product of two maps,

$$(e^i \otimes d^j)(e_k \otimes d_\ell) = (e^i e_k \otimes d^i d_\ell) = (e^i e_k)(d^j d_\ell) \quad \text{since } k \otimes \ell = k\ell \text{ for } k, \ell \in$$

$$= \delta^i_k \, \delta^j_\ell = \begin{cases} 1 & \text{if } i = k, j = \ell \\ 0 & \text{otherwise} \end{cases}.$$

Thus $e^i \otimes d^j$ is the dual basis. We can now state the conclusion of the theorem as $(V \otimes W)^* = V^* \otimes W^*$; that is, we identity $(V \otimes W)^*$ with $V^* \otimes W^*$ by the isomorphism here established.

§12. Tensor Algebras and Graded Algebras.

Let us now consider all the vector spaces $T_n(V) = V \otimes \ldots \otimes V$ (n times), for $n = 1, 2, \ldots$. For $n = 0$ we will define $T_0(V) = \mathbb{R}$. Suppose that $a = v_1 \otimes v_2 \otimes \ldots \otimes v_r$ is in $T_r(V)$ and that $b = w_1 \otimes w_2 \ldots \otimes w_s$

is in $T_s(V)$. Since all of the vectors v_i and w_j belong to V, we can form $v_1 \otimes \ldots \otimes v_r \otimes w_1 \otimes \ldots \otimes w_s \in T_{r+s}(V)$, which we will call the product of a and b. Thus if we refer to the whole collection of spaces $\{T_n(V)\}$ as $T_*(V)$, extending the above multiplication by the distributive law gives a natural multiplicative structure on this set, which has the unusual property that the product of an element of $T_r(V)$ with one of $T_s(V)$ lies in $T_{r+s}(V)$, so that the product is usually in a space different from those of the factors.

We generalize this situation as follows. An algebra A is a vector space which in addition possesses a multiplication; that is, not only can we multiply elements of the space by scalars, but any two elements of the space itself have a product. This multiplication is required to be bilinear and associative, and to have a unit element. Thus, we have the rules

$$(k_1a_1 + k_2a_2)b = k_1(a_1b) + k_2(a_2b) \ , \ b(k_1a_1 + k_2a_2) = k_1(ba_1) + k_2(ba_2)$$

$$a(bc) = (ab)c \ , \ 1a = a1 = a$$

$$(ka)b = k(ab) \ , \ a(kb) = k(ab)$$

for $a, b, a_1, a_2 \in A$ and k, k_1, and k_2 scalars. A graded algebra is a string G of vector spaces, $G_0, G_1, \ldots$, with an additional product structure such that if $a \in G_n$ and $b \in G_m$, then $ab \in G_{n+m}$. For each m and n, this product must be a bilinear function from $G_n \times G_m$ to G_{n+m}, must be associative, and must have a unit element $1 \in G_0$. Notice in particular that a graded algebra G is not an algebra, since the sum of any two elements is defined only if they both have the same degree ; that is, if they both lie in the same G_i.* Now we can see that $T_*(V)$ is a graded algebra, with unit element $1 \in T_0(V) = \mathbb{R}$. Call $T_*(V)$ the tensor algebra of V. There are other examples of graded algebras: consider the set G_n of all homogeneous polynomials of degree n in the two letters X and Y. A typical element of G_n is $\sum_{i=0}^{n} a_i X^i Y^{n-i}$, where $a_i \in \mathbb{R}$. G_n is a vector space under the usual addition and scalar multiplication of polynomials, and since the product of a homogeneous polynomial of degree n and one of degree m is a homogeneous polynomial of degree n+m, it is easy to see that the set of all polynomials in S and Y contains the graded algebra G.

§13. Exterior Algebra.

An exterior algebra E is a graded algebra with the property that the square of any element of degree one is zero. Now if a and b are of degree one, so is a+b, hence $0 = (a+b)^2 = 0 + ab + ba + 0$. Thus in any exterior algebra, $ab = -ba$ for any two elements of degree one. What does the rest of an exterior algebra look like? We can get some idea of the answer by considering the simple example of an exterior algebra Λ for which Λ_1 is a two-dimensional vector space. If $\{e^1, e^2\}$ is a basis for Λ_1, then $e^1e^2 = -e^2e^1$ will lie in Λ_2. Thus Λ_2 is forced to be at least one-dimensional. (Note: We assume that no other relations besides

---

*If $a \in G_i$ then we say a has degree i.

($a^2 = 0$ for deg $a = 1$) are satisfied; in particular, $e^1e^2 \neq 0$.) It's also easy to see that any product of elements of Λ_1 which lies in Λ_n (for $n > 2$) must be zero. So we are tempted to form Λ by letting Λ_n be the zero vector space for $n > 2$, with Λ_2 a one-dimensional vector space whose basis is $\{e^{12}\}$, where by definition $e^{12} = e^1e^2$. This does give us an exterior algebra: for example, if $x_1e^1 + x_2e^2$ is any element of Λ_1, the formula

$$(x_1e^1 + x_2e^2)(y_1e^1 + y_2e^2) = (x_1y_2 - x_2y_1)e^1e^2$$

tells us that $(x_1e^1 + x_2e^2)^2 = 0$.

Move now to the case where Λ_1 is three-dimensional, with basis $\{e^1, e^2, e^3\}$. Then e^1e^2, e^2e^3, e^1e^3 all must lie in Λ_2; denote them by e^{12}, e^{23}, e^{13} respectively. Then we can let Λ_2 be three-dimensional, with basis $\{e^{12}, e^{23}, e^{13}\}$ with Λ_3 a one-dimensional space generated by $e^{123} = e^1e^2e^3$, and $\Lambda_n = 0$ for $n > 3$. Multiplication is now always possible since for example $e^2e^{12} = e^2e^1e^2 = -e^2e^2e^1 = 0$ and $e^3e^2e^1 = -e^3e^1e^2 = e^1e^3e^2 = -e^1e^2e^3 = -e^{123}$. Calculation gives the rule

$$\left(\sum_{i=1}^{3} x_ie^i\right) \wedge \left(\sum_{j=1}^{3} x_je^j\right) \wedge \left(\sum_{k=1}^{3} z_ke^k\right) = \det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} e^{123}$$

where we have used the traditional "wedge" symbol to denote the anti-symmetric multiplication of an exterior algebra.

It is now apparent how to extend the above construction to the n-dimensional case.

<u>Theorem</u>. Given a vector space W of dimension m, there exists a (free) exterior algebra $\Lambda = \Lambda(W)$ with $\Lambda_1 = W$.

<u>Proof</u>. Let $\{e^1, \ldots, e^m\}$ be a basis of W. Let $\Lambda_k(W)$ be a vector space whose basis is the set of symbols $\{e^{i_1 i_2 \ldots i_k}\}$, where $1 \le i_1 < i_2 < \ldots < i_k \le m$. Thus $\Lambda_k(W)$ has dimension $\binom{m}{k} = \frac{m!}{k!(m-k)!}$. To write out the rules of multiplication in a simple form, we will use the language of permutations, considering a permutation σ of order n as a one-to-one function from the set of the first n integers to that set. The sign of σ, sometimes written $(-1)^\sigma$, is defined to be $+1$ if σ is the product of an even number of transpositions, and -1 otherwise. Then if $i_1 < \ldots < i_n$, while σ is a permutation, we define $e^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}} = (-1)^\sigma e^{i_1 \ldots i_k}$; it is not hard to see that this agrees with our two- and three-dimensional examples. Now define exterior multiplication by

$$e^{i_1 \ldots i_k} \wedge e^{j_1 \ldots j_\ell} = \begin{cases} 0 & \text{if some } i_s \text{ equals some } j_t \\ e^{i_1 i_2 \ldots i_k j_1 \ldots j_\ell} & \text{otherwise} \end{cases}$$

Finally, let $1 \cdot e^{j_1 j_2 \cdots j_r} = e^{j_1 j_2 \cdots j_r}$, and extend this multiplication linearly to all of $\Lambda(W)$. Now all the properties in the definition of exterior algebra follow: for example

$$(e^{i_1 \ldots i_r} \wedge e^{j_1 \ldots j_s}) \wedge e^{k_1 \ldots k_t} = e^{i_1 \ldots i_r j_1 \ldots j_s k_1 \ldots k_t}$$
$$= e^{i_1 \ldots i_r} \wedge (e^{j_1 \ldots j_s} \wedge e^{k_1 \ldots k_t})$$

is the associative law. Moreover, if $a_i = \sum_{i=1}^{r} c_{ij} e^j$, then the formula

$$a_1 \wedge \dots \wedge a_r = \det(\cancel{x}_{ij})\cancel{e}^{\cancel{i_1 \dots i_r}} .$$

$\det(c_{ij})\, e^{1 \dots r}$

holds. If $a = \sum x_i e^i$ is any element of degree one,

$$a^2 = \sum x_i x_j e^i \wedge e^j = \sum_{i=j} + \sum_{i<j} + \sum_{i>j} = 0 + \sum_{i<j} x_i x_j e^i \wedge e^j + \sum_{i>j} x_i x_j e^j \wedge e^i = 0,$$

so Λ is an exterior algebra.

This construction is in fact the most general way of obtaining an n-dimensional free exterior algebra; it does not depend on the choice of basis. To prove this, we note first that if E is any exterior algebra, and f is a linear transformation mapping W to E_1, then there are unique linear transformations $f_i : \Lambda_i W \longrightarrow E_i$ such that the collection $\{f_i\}$ forms a morphism of graded algebras; that is, $f_o(1) = 1$, and $f_i(a)f_j(b) = f_{i+j}(ab)$ if a and b are of degree i and j respectively. This is true since we must have $f_k(e^{i_1 \dots i_k}) = f_k(e^{i_1} \wedge e^{i_2} \wedge \dots \wedge e^{i_k}) = f(e^{i_1}) \wedge f(e^{i_2}) \wedge \dots \wedge f(e^{i_k})$; since f_k is now defined on a basis, it extends uniquely to all of Λ_k.

Our proof that the construction of an exterior algebra on V does not depend on the choice of basis of V will be based on the following very general principles, which we have been using implicitly for some time. Let us define a functor on algebras to vector spaces to be a function $\mathcal{F}$ which to every algebra A assigns a vector space $\mathcal{F}(A)$, and to every map $A \xrightarrow{f} A'$ of algebras assigns a map $\mathcal{F}(f): \mathcal{F}(A) \longrightarrow \mathcal{F}(A')$ of vector spaces (a linear transformation) with the rules that $\mathcal{F}(1_A) = 1_{\mathcal{F}(A)}$, where 1_A is the identity map on A, and $\mathcal{F}(gf) = \mathcal{F}(g)\,\mathcal{F}(f)$ if $A \xrightarrow{f} A' \xrightarrow{g} A''$.

Given $\mathfrak{A}$ and a vector space V, a <u>universal construction</u> for $\mathfrak{A}$ and V is an <u>algebra</u> R and an arrow $u: V \longrightarrow \mathfrak{A}(R)$ (that is, a linear transformation) with the following property: given any arrow $t: V \longrightarrow \mathfrak{A}(A)$, there is a unique map of algebras $f: R \longrightarrow A$ such that $t = \mathfrak{A}(f)u$. That is, $\mathfrak{A}(f)$ makes the following diagram commute and f is the only map which has this property.

$$\begin{tikzcd} V \arrow[r, "u"] \arrow[dr, "t"'] & \mathfrak{A}(R) \arrow[d, "\mathfrak{A}(f)"] \\ & \mathfrak{A}(A) \end{tikzcd} .$$

We saw this situation in the construction of the tensor product (§11 above). In that case we showed that any two universal arrows, i.e., any two tensor products, were isomorphic. Exactly the same proof works in the general situation:

<u>Theorem.</u> If (R, u) and (R', u') are both universal arrows for $\mathfrak{A}$ and V, then there is a unique isomorphism $h: R \longrightarrow R'$ such that $\mathfrak{A}(h)u = u'$.

To prove this, simply repeat the proof of the above-mentioned theorem about tensor products; since nothing in the argument there depended on whether the objects in question were algebras or vector spaces, this <u>universal property</u> is all we need in the proof.

Now consider the functor $\mathfrak{G}$ from graded algebras to vector spaces, defined by $\mathfrak{G}(G) = G_1$.

Theorem 1. The tensor algebra T_*V is a universal construction for V and $\mathcal{B}$; that is, given a graded algebra G and a map $V \xrightarrow{t} G_1$, there is a unique map $T_*(V) \xrightarrow{r} G$ of algebras making the diagram below commute.

$$\begin{tikzcd} V \arrow[r, "1_V"] \arrow[dr] & \mathcal{B}(T_*(V)) = V \arrow[d, "\mathcal{B}(r) = r_1"] \\ & \mathcal{B}(G) = G_1 \end{tikzcd}$$

Proof. It is easy to see that we are forced to take $r_o = 1_{\mathbb{R}}$, $r_1 = t$; and to define $r_n(v_1 \otimes \ldots \otimes v_n) = (tv_1)(tv_2)\ldots(tv_n)$.

Theorem 2. Define $\mathcal{B}$ as before, but mapping graded exterior algebras to vector spaces. Then if W is a finite-dimensional vector space, $\Lambda(W)$ is a universal object for W and $\mathcal{B}$.

Proof. In the same situation as that of Theorem 1, defining $r_n(a_1 \wedge \ldots \wedge a_n) = (ta_1) \wedge (ta_2) \wedge \ldots \wedge (ta_n)$ does the trick.

Now we can prove our assertion about the construction of the exterior algebra on a given finite dimensional vector space V. Theorem 2 shows that the algebra $\Lambda(W)$ constructed from a particular basis $e_1, \ldots, e_n$ is a universal object; likewise, the exterior algebra coming from a different basis $d_1, \ldots, d_n$ is universal. Since any two universal objects for a functor are isomorphic, the two exterior algebras we have constructed are really the same. Incidentally, it is possible to define $\Lambda(W)$ directly as an invariant object; for details, see the last chapter of MacLane and Birkhoff.

§14. Alternating Tensors

We now go on to derive a new and very useful way of looking at $\Lambda(W)$ as a special subspace of $T_*(W)$. In §1~~2~~3 we studied elements of S_k, the set of permutations of k letters. Now each $\sigma \in S_k$ can be interpreted as a linear transformation on $T_k(W)$: define $\sigma(w_1 \otimes \ldots \otimes w_k) = w_{\sigma(1)} \otimes \ldots \otimes w_{\sigma(k)}$ and check that this can be extended to a well-defined linear map (this amounts to checking that $\bar{\sigma}$ on $W \times \ldots \times W$ defined by $\bar{\sigma}(w_1 \times \ldots \times w_k) = w_{\sigma(1)} \otimes \ldots \otimes w_{\sigma(k)}$ is a k-linear map).

<u>Definition</u>. A tensor $t \in T_k(W)$ is called <u>alternating</u> if $\sigma(t) = (-1)^{\sigma} t$ for all $\sigma \in S_k$.

<u>Theorem</u>. The set of alternating tensors in $T_*(W)$ can, in a natural way, be made into a graded exterior algebra isomorphic to $\Lambda(W)$.

<u>Proof</u>. The set of alternating k-tensors clearly forms a subspace of $T_k(W)$; furthermore, any tensor t can be symmetrized to yield an alternating tensor. Specifically, let $\overline{A}(t) = \sum_{\sigma \in S_k} (-1)^{\sigma} \sigma t$ if t is <u>any</u> k-tensor, not necessarily alternating. Then

$$\tau \overline{A}(t) = \sum_{\sigma \in S_k} (-1)^{\sigma} \tau\sigma(t) = (-1)^{\tau}(-1)^{\tau} \sum_{\sigma \in S^k} (-1)^{\sigma} \tau\sigma(t) = (-1)^{\tau} \sum_{\sigma \in S^k} (-1)^{\sigma\tau} \tau\sigma(t),$$

since in general $(-1)^{\sigma}(-1)^{\tau} = (-1)^{\sigma\tau}$. But since S_k is a group, as σ runs over all the elements of S_k, $\sigma\tau$ also runs over all the elements of S_k (perhaps in a different order, but each element is counted once and only once). Hence $\tau\overline{A}(t) = (-1)^{\tau}\overline{A}(t)$, for $t \in T_k(W)$, so $\overline{A}(t)$ is alternating, whether or not t is. If t does happen to be alternating, $\overline{A}(t) = (k!)t$, since there are $k!$ elements in S_k; so the mapping A defined by $A(t) = (1/k!)\overline{A}(t)$ has the property that At is always alternating, and that if t is already alternating, then $At = t$.

Now if s and t are any two tensors in $T_*(W)$, define $s \wedge t$ to be $A(s \otimes t)$, where $(s,t) \rightsquigarrow s \otimes t$ is just the usual product in the tensor algebra $T_*(W)$. We assert that the alternating tensors form a graded exterior algebra under $\wedge$. First, given t of degree one, we have $t \wedge t = 0$, since S_2 consists of only two permutations, one of each sign, and hence $A(t \times t) = t \otimes t - t \otimes t = 0$. Next, we check the associative law. The proof of this is divided into steps.

1. $A(\tau t) = (-1)^\tau A(t)$ for $t \in T_k(W)$, $\tau \in S_k$.

This follows easily from the definition of A.

2. Given $\sigma \in S_k$, define $\tilde{\sigma} \in S_{m+k}$ by

$$\tilde{\sigma}(i) = i \quad \text{for } 1 \leq \pm m$$

$$\tilde{\sigma}(i+m) = \sigma(i) + m \quad \text{for } m+1 \leq m+i \leq m+k .$$

Then $(-1)^\sigma = (-1)^{\tilde{\sigma}}$.

3. If s is an m-tensor and t is a k-tensor, then $s \wedge t = s \wedge (At)$.

<u>Proof.</u>

$$s \wedge At = A(s \otimes At) = \frac{1}{k!} A(\sum_{\sigma \in S_k} s \otimes (-1)^\sigma \sigma t) = \frac{1}{k!} A(\sum_{\sigma \in S_k} (-1)^\sigma \tilde{\sigma}(s \otimes t))$$

$$= \frac{1}{k!} \sum_{\sigma \in S_k} (-1)^\sigma A(\tilde{\sigma}(s \otimes t)) = (\text{by } 1) \frac{1}{k!} \sum_{\sigma \in S_k} (-1)^\sigma (-1)^{\tilde{\sigma}} A(s \otimes t)$$

$$= A(\frac{1}{k!} \sum_{\sigma} s \otimes t) = A(s \otimes t) = s \wedge t.$$

4. $r \wedge (s \wedge t) = r \wedge A(s \otimes t) = r \wedge (s \otimes t)$ by 3

$$= A(r \otimes (s \otimes t)) = A((r \otimes s) \otimes t) = (r \otimes s) \wedge t$$

$$= A(r \otimes s) \wedge t = (r \wedge s) \wedge t.$$

Similarly, we check that the alternating tensors have the other properties of an exterior algebra.

To finish the proof we must establish theisomorphism between the set of alternating tensors and $T_*(W)$. Let $e_1, \ldots, e_n$ be a basis of W; then a linear map on $\Lambda(W)$ is determined by its values on all basis elements $e_{i_1} \wedge \ldots \wedge e_{i_k}$. Then mapping $e_{i_1} \wedge \ldots \wedge e_{i_k}$ to the alternating tensor $A(e_{i_1} \otimes \ldots \otimes e_{i_k})$ is an isomorphism: it is one-to-one since

$$A\Big(\sum_i a_i (e_{i_1} \otimes \ldots \otimes e_{i_k})\Big) = \sum_i a_i \Big(\frac{1}{k!}\Big) \sum_{\sigma \in S_k} (-1)^\sigma e_{\sigma(i_1)} \otimes \ldots \otimes e_{\sigma(i_k)}$$

is a sum of distinct basis elements of $T_k(W)$ with non-zero coefficients, hence non-zero; it is onto since given any alternating tensor $t \in T_k(W)$, write $t = \sum_i a_i e_{i_1} \otimes \ldots \otimes e_{i_k}$; then the element $\sum a_i e_{i_1} \wedge \ldots \wedge e_{i_k}$ of $\Lambda(W)$ is mapped to $A(\sum a_i e_{i_1} \otimes \ldots \otimes e_{i_k}) = At = t$. This completes the proof.

§15. Local Manifolds: Invariant Description.

Eventually we will describe mechanical systems (Lagrange's equations, Hamiltonians, etc.), on smooth manifolds obtained by piecing together open sets of Euclidean spaces $\mathbb{R}^n$. Hitherto, all our treatment has been local, so has been formulated for <u>open sets</u> U in $\mathbb{R}^n$. Each such set $U \subset \mathbb{R}^n$ comes equipped with a natural set of coordinates -- those of $\mathbb{R}^n$. However, the basic constructions such as the tangent bundle $T.U$ or the Lagrange equations for a smooth function $L: T.U \longrightarrow \mathbb{R}$ have in fact been independent of the choice of coordinates. Indeed, all of our previous discussion of such open sets U can be made more clearly invariant if U is replaced by a "local manifold" M in the sense of the following definition. The sense is this: Given coordinates q^i on U, a function $f: U \longrightarrow \mathbb{R}$ is smooth if all the higher partial derivatives $\dfrac{\partial f}{\partial q^i}$ are continuous. The set $\mathcal{F}$ of <u>all</u> smooth functions is then the same for <u>any</u> allowable choice of coordinates; hence we can give an invariant description in terms of $\mathcal{F}$.

<u>Definition.</u> A <u>local manifold</u> is a set M together with a set $\mathcal{F}$ of functions $f: M \longrightarrow \mathbb{R}$ such that

1) there exist $q^1, \ldots, q^n \in \mathcal{F}$ such that the map

$$m \overset{\varphi}{\rightsquigarrow} (q^1 m, \ldots, q^n m)$$
$$M \xrightarrow{\varphi} \mathbb{R}^n$$

is one-to-one onto an open set $U \subset \mathbb{R}^n$,

2) $f \in \mathcal{F}$ if and only if $f\varphi^{-1}$ is smooth on U.

We leave the reader to carry out the replacement.

§16. The Exterior Bundle

We have seen that given any finite-dimensional vector space W, there exists (uniquely up to isomorphism) the exterior algebra $\Lambda = \Lambda(W)$ of W, i.e., a graded algebra universal for the properties

1. W = elements of degree 1,

2. $w \wedge w = 0$.

We have also seen that

$$\Lambda_k(W) \simeq \text{alternating tensors on } W$$
$$= \{t \in T_k(W) \mid \sigma t = (-1)^\sigma t \quad \text{for all } \sigma \in S_k\}.$$

As we have done in the case of a two-fold tensor product, we can show that

$$T_k(W) \simeq \text{Mult}\,(\overbrace{W^*, \ldots, W^*}^{\text{k-factors}} : \mathbb{R})$$
$$= \text{all multilinear maps } f: \overbrace{W^* \times \ldots \times W^*}^{\text{k factors}} \longrightarrow \mathbb{R}.$$

Then an alternating tensor corresponds to an alternating multilinear map

$$a: \underbrace{W^* \times \ldots \times W^*}_{\text{k factors}} \longrightarrow \mathbb{R}$$

such that

$$a(v_1, \ldots, v_k) = (-1)^\sigma a(v_{\sigma(1)}, \ldots, v_{\sigma(k)}), \quad \text{for all } \sigma \in X_k, \; v_i \in W^*.$$

Suppose U is an open set in $\mathbb{R}^n$ (or better, a local manifold as defined above). We define the <u>k^{th} exterior bundle over U</u> to be

$$\begin{array}{ccccc} \Lambda^k U & = & \Lambda_k(T^\cdot U) & = & \{(a,d) \mid a \in U,\ d \in \Lambda_k(T^a U)\}, \\ \big\downarrow & & & & \big\downarrow \\ U & & & & a \in U \end{array}$$

i.e., the k^{th} exterior bundle is the bundle whose fibers are the k^{th} exterior algebra of the cotangent spaces of U. If $k = 0$, each $\Lambda_0(T^aU)$ is just $\mathbb{R}$, so $\Lambda^0(U) = U \times \mathbb{R}$ (all fibers isomorphic to the 1-dimensional vector space $\mathbb{R}$.

$\Lambda^k U$ is a local manifold -- for if U has coordinates $q^1, \ldots, q^n$ and T^aU has basis $e^1, \ldots, e^n$, then $\Lambda^k U$ has coordinates $q^1, \ldots, q^n$ and $p_{i_1 \ldots i_k}$, where the $p_{i_1 \ldots i_k}$ are the dual basis to the basis $\{e^{i_1} \wedge \ldots \wedge e^{i_k}\}$ for $\Lambda_k(T^aU)$; we can define a function on $\Lambda^k U$ to be smooth if it is smooth in terms of these coordinates.

For any integers k and m, we may form the "pullback" of $\Lambda^k U$ and $\Lambda^m U$ over U

$$\Lambda^k U \underset{U}{\times} \Lambda^m U = \{(x, y) \in \Lambda^k U \times \Lambda^m U \mid \pi(x) = \pi(y)\}$$

$$\begin{tikzcd}
 & \Lambda^k U \underset{U}{\times} \Lambda^m U \arrow[dl, "\rho_k"'] \arrow[dr, "\rho_m"] & \\
\Lambda^k U \arrow[dr, "\pi"'] \arrow[rr, phantom, "\text{pullback}"] & & \Lambda^m U \arrow[dl, "\pi"] \\
 & U &
\end{tikzcd}$$

This pullback is itself a bundle over U. Now we can define a multiplication on $\Lambda^k U \underset{U}{\times} \Lambda^m U$ to $\Lambda^{k+m} U$ by means of the multiplication in $\Lambda(T^aU)$ for each $a \in U$:

$$\Lambda^k U \underset{U}{\times} \Lambda^m U \xrightarrow{\ \wedge\ } \Lambda^{k+m} U$$

$$((a, d), (a, d')) \overset{\wedge}{\rightsquigarrow} (a, d \wedge d') \in \Lambda^{k+m} U .$$

<u>Definition.</u> An exterior k-form ω on U $(k = 0, 1, \ldots, n)$ is a smooth cross-section of
$$\begin{array}{c} \Lambda^k U \\ \big\downarrow \pi \\ U \end{array}.$$

Thus a k-exterior form ω on U is a smooth map $U \xrightarrow{\omega} \Lambda^k U$ of the form $a \rightsquigarrow (a, \tilde{\omega}a)$ where $\tilde{\omega}a \in \Lambda_k(T^a U)$.

Note that a 1-form by this definition coincides with our previous definition of a 1-form, and that a 0-form is just a smooth map from U to $\mathbb{R}$.

<u>Definition.</u> $\Omega^k(U)$ = the set of all k-forms ω on U.

We remark that

1) $\Omega^k(U)$ is a vector space over $\mathbb{R}$, with vector space opeations as follows: if $\omega \in \Omega^k(U)$, $\alpha \in \mathbb{R}$ and ω: $a \rightsquigarrow (a, \tilde{\omega}a)$, then

$$\alpha\omega: a \rightsquigarrow (a, \alpha(\tilde{\omega}a)) \; ; \; \omega_1 + \omega_2 : a \rightsquigarrow (a. \tilde{\omega}_1 a + \tilde{\omega}_2 a).$$

2) $\Omega^k(U)$ is an $\mathcal{F}$-module, where $\mathcal{F}$ = the ring of smooth functions $f: U \longrightarrow \mathbb{R}$. The module action is given by $f\omega: a \rightsquigarrow (a, f(a)\tilde{\omega}a)$.

3) For each k and m, there is an exterior product

$$\Omega^k(U) \times \Omega^m(U) \longrightarrow \Omega^{k+m}(U)$$
$$(\omega, \eta) \rightsquigarrow \omega \wedge \eta$$

defined by

$$\omega \wedge \eta : a \rightsquigarrow (a, \tilde{\omega}(a) \wedge \tilde{\eta}(a)) .$$

This is well-defined since $\tilde{\omega}(a)$ and $\tilde{\eta}(a)$ are both elements of $\Lambda(T^a U)$. The exterior product is associative and bilinear. Thus the $\Omega^k(U)$ form a graded algebra $\Omega^*(U)$ over $\mathcal{F}$.

Since $\Lambda^k(U)$ has as basis all k-fold wedge products of the dq^i, where $q^1, \ldots, q^n$ are coordinates in U, any k-form ω is of the form

$$\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} f^{i_1 \ldots i_k} \wedge dq^{i_1} \wedge \ldots \wedge dq^{i_k} ,$$

where $f^{i_1 \ldots i_k} : U \longrightarrow \mathbb{R}$. If the $p^{i_1 \ldots i_k}$ are the coordinates corresponding to the $dq^{i_1} \wedge \ldots \wedge dq^{i_k}$, then

$$f^{i_1 \ldots i_k} = p^{i_1 \ldots i_k} \circ \omega .$$

Given a smooth map $U \xrightarrow{\varphi} U'$, we have an induced map $T^a U \xleftarrow{\varphi^*} T^a U'$, which in turn determines, via the properties of the exterior algebra, a map $\Omega^k U \xleftarrow{\varphi^*} \Omega^k U'$ such that φ^* is linear and

$$\varphi^*(\omega \wedge \eta) = (\varphi^* \omega) \wedge (\varphi^* \eta).$$

Then if $q^1, \ldots, q^n$ are coordinates in U and $r^1, \ldots, r^n$ in U', it will follow that

$$\varphi^*(dq^1 \wedge \ldots \wedge dq^n) = \det\left(\frac{\partial q^i \varphi}{\partial r^j}\right) dr^1 \wedge \ldots \wedge dr^n ,$$

which is just the usual change-of-coordinates rule.

Now we define the basic operations of exterior differentiation of forms.

<u>Theorem</u>. There exists a unique linear map

$$d: \Omega^k(U) \longrightarrow \Omega^{k+1}(U) \qquad \text{for each } k$$

such that

1) for $f \in \Omega^0$, df = usual differential of f

2) $d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^k \omega \wedge d\eta$, where k = degree of ω [i.e.,

every time d is moved past something of degree 1, the sign changes]

3) $dd\omega = 0$ for all ω

(d is called the exterior derivative)

Proof. If d exists, then by linearity and (1)-(3),

$$d(\sum f^{i_1 \cdots i_k} \wedge dq^{i_1} \wedge \ldots \wedge dq^{i_k})$$
$$= \sum d(f^{i_1 \cdots i_k} \wedge dq^{i_1} \wedge \ldots \wedge dq^{i_k})$$
$$= \sum df^{i_1 \cdots i_k} \wedge dq^{i_1} \wedge \ldots \wedge dq^{i_k} \quad ,$$

since all other summands have a factor of the form ddq^i and hence are zero by (3).

Since every k-form ω can be written as

$$\omega = \sum f^{i_1 \cdots i_k} dq^{i_1} \ldots dq^{i_k}$$

for smooth functions $f^{i_1 \cdots i_k}$, the above shows that d, if it exists, must be unique, since it is determined by its effect on 0-forms, which is specified by (1). But the above expression also defines a function $d: \Omega^k(U) \longrightarrow \Omega^{k+1}(U)$ for each k. It is easily verified that this d is linear and satisfies (1)-(3).

Chapter III. HAMILTONIAN MECHANICS

17. Calculus of Variations

Suppose that M is a local manifold with coordinates $y^1, \ldots, y^m$, and that

$$K: T.M \times I \longrightarrow \mathbb{R}$$

is a smooth function. We have already seen that the integral

$$\int_{t_o}^{t_1} K\, dt$$

is stationary along the path $c: I \longrightarrow M$ if and only if c satisfies Euler's Equations for K:

$$\frac{d}{dt} \frac{\partial K}{\partial \dot{y}^i} - \frac{\partial K}{\partial y^i} = 0 \quad , \quad i = 1, \ldots, m.$$

In particular, using $K = L$, we obtained Lagrange's equations in this way.

Theorem. (Generalized Hamilton's Principle): Given U with coordinates $q^1, \ldots, q^n$, additional coordinates $p_1, \ldots, p_n$ in T^*U, and a smooth function $H: T^*U \longrightarrow \mathbb{R}$, then Hamilton's equations for H are just the Euler equations for the function

$$K = \sum_{i=1}^{n} p_i \dot{q}^i - H,$$

where $M = T^*U$.

Proof. T^*U has coordinates $q^1, \ldots, q^n$ and $p_1, \ldots, p_n$, so Euler's equations correspondingly take two forms:

1) for the q^i's

$$\frac{d}{dt} \frac{\partial K}{\partial \dot{q}^i} - \frac{\partial K}{\partial q^i} = 0$$

or

$$\frac{d}{dt} p_i = - \frac{\partial H}{\partial q^i} \quad ,$$

2) for the p_i's

$$\frac{d}{dt}\frac{\partial K}{\partial \dot{p}_i} - \frac{\partial K}{\partial p_i} = 0$$

or

$$-\dot{q}^i + \frac{\partial H}{\partial p_i} = 0\ , \text{ since } K \text{ is independent of } \dot{p}_i\ .$$

In the above, we would like to express

$$\int \sum p_i \dot{q}^i \, dt$$

in terms of a differential form. Suppose

$$c : I \longrightarrow M$$

is a curve in M, lifted to

$$\tilde{c} : I \longrightarrow T.M.$$

Let ω be the 1-form

$$\omega = \sum_{i=1}^{n} p_i dq^i \quad \text{on } M.$$

Pulling ω back to I via c, we get

$$c^*\omega = \sum_{i=1}^{n} p_i c(c^* dq^i) = \sum_{i=1}^{n} p_i c \frac{dq^i c}{dt} \, dt\ .$$

Thus what we mean by

$$\int_{t_o}^{t_1} \sum_{i=1}^{n} p_i \dot{q}^i \, dt$$

is just

$$\int_{t_o}^{t_1} (c^*\omega) \, dt\ .$$

In what follows, we'll find it convenient to use the 2-form

$$\Omega = d\omega = \sum_{i=1}^{n} dp_i \wedge dq^i$$

rather than ω itself.

$M = T.U$ is called the <u>phase space</u> of the system, and U is called the <u>configuration space</u>.

<u>Definition.</u> Let M' and M be phase spaces with coordinates $Q^1, \ldots, Q^n; P_1, \ldots, P_n$ and $q^1, \ldots, q^n; p_1, \ldots, p_n$ respectively. A smooth, one-to-one mapping $M' \longrightarrow M$ is called a <u>canonical transformation,</u> or <u>contact transformation</u>, with respect to the given coordinates, if

$$\varphi^*\Big(\sum_{i=1}^{n} dp_i \wedge dq^i\Big) = \sum_{i=1}^{n} dP_i \wedge dQ^i .$$

The importance of these transformations is that they preserve Hamilton's equations:

<u>Theorem</u>. Given a canonical transformation $\varphi\colon M' \longrightarrow M$, a smooth map $H\colon M \longrightarrow \mathbb{R}$, and a path c in M' such that the image of c is contained in some open, simply connected subset of M'. If φc satisfies Hamilton's canonical equations in M for H, then c satisfies them in M' for $\varphi^* H$

$$\begin{tikzcd}
\mathbb{R} \arrow[d, "c"'] \arrow[dr, dashed, "\varphi c"] & \\
M' \arrow[r, "\varphi"] \arrow[dr, dashed, "\varphi^* H"'] & M \arrow[d, "H"] \\
& \mathbb{R}
\end{tikzcd}$$

<u>Proof.</u> We want to show that

$$\int_{t_0}^{t_1} \Big(\sum_{i=1}^{n} P_i dQ^i - \varphi^* H\Big)\, dt$$

is stationary over the path c with respect to nearby smooth paths with the same end-points.

Since the transformation φ is canonical

$$d\varphi\Big(\sum p_i dq^i\Big) = \varphi^*\Big(\sum dp_i \wedge dq^i\Big)$$
$$= \sum dP_i \wedge dQ^i = d\Big(\sum P_i dQ^i\Big),$$

so the 1-form

$$\rho = \varphi^*(\sum p_i dq^i) - \sum P_i dQ^i$$

is closed, i.e., $d\rho = 0$. Thus (by Poincaré's lemma, to be proved below) there exists a 0-form F on the open, simply connected subset of M containing the image of c, such that

$$\rho = dF$$

on that subset. Then over any curve γ in this subset,

$$\begin{aligned}
\int_{t_o}^{t_1} &\left(\sum_{i=1}^{n} P_i dQ^i - \varphi^* H \right) dt \\
&= \int_{t_o}^{t_1} (\varphi^*(\sum p_i dq^i) - \rho - \varphi^* H)\, dt \\
&= \int_{t_o}^{t_1} [\varphi^*(\sum p_i dq^i - H) - dF]\, dt \\
&= \int_{t_o}^{t_1} \varphi^*(\sum p_i dq^i - H)\, dt - F(\gamma(t_1)) + F(\gamma(t_o)) \\
&= \int_{t_o}^{t_1} (\sum p_i dq^i - H)\, dt - F(\gamma(t_1)) + F(\gamma(t_o)),
\end{aligned}$$

where the last integration is over $\varphi\gamma$. But we are comparing c with nearby curves γ with the same endpoints. Thus $F(\gamma(t_1)) - F(\gamma(t_o))$ is a constant. Then since

$$\int_{t_o}^{t_1} (\sum p_i dq^i - H)\, dt$$

is stationary for the path φc, the above relation says that

$$\int_{t_o}^{t_1} (\sum P_i dQ^i - \varphi^* H)\, dt$$

is stationary over c.

Examples:

1) Let $M \xrightarrow{\varphi} M$ be the transformation defined by

$$p_i\varphi = -Q^i$$
$$q^i\varphi = P_i .$$

Then
$$\begin{aligned}\varphi^*\Big(\sum_{i=1}^{n} dp_i \wedge dq^i\Big) &= \sum_{i=1}^{n} d(p_i\varphi)\wedge d(q^i\varphi)\\ &= -\sum_{i=1}^{n} dQ^i \wedge dP_i\\ &= \sum_{i=1}^{n} dP_i \wedge dQ^i .\end{aligned}$$

Thus φ is a canonical transformation.

2) We will see later that if $V \xrightarrow{\varphi} U$ is a smooth, invertible map of configuration spaces, then the induced map $T^{\cdot}V \xleftarrow{\Phi} T^{\cdot}U$ of the cotangent spaces is a canonical transformation. Such a transformation is called a (canonical) point transformation.

§18. Application: The Harmonic Oscillator

a) The linear oscillator (cf. Goldstein, p. 24):

This concerns the following phase space:

U is one-dimensional with coordinate q,

$$T = \frac{1}{2} m \dot{q}^2 = \frac{1}{2m} p^2 ,$$
$$V = \frac{1}{2} k q^2 ,$$
$$H = T + V = \frac{1}{2}\Big[\frac{p^2}{m} + kq^2\Big].$$

We would like to find a canonical transformation of $T^{\cdot}U$ so that H is of a simpler form.

Let $\quad p = \sqrt{km}\ F \cos Q,$

$q = F \sin Q$, where F is a function of P.

Then

$$H = \frac{1}{2m}[p^2 + kmq^2] = \frac{k}{2}F^2 ,$$

$$dp\wedge dq = \sqrt{km}\,(F' \cos Q\, dP - F \sin Q\, dQ)\wedge(F' \sin Q\, dP + F \cos Q\, dQ)$$

$$= \sqrt{km}\ FF'\, dP\wedge dQ .$$

So our transformation will be canonical if

$$dP\wedge dQ = \sqrt{km}\ FF'\, dP\wedge dQ ,$$

i.e., if

$$\sqrt{km}\ FF' = 1.$$

We can integrate this to get

$$F = \frac{\sqrt{2p}}{(km)^{1/4}} .$$

Let $\omega = (km)^{1/2}$. Then

$$H = \frac{k}{2}F^2 = \omega P,$$

so Hamilton's equations become

$$\frac{dQ}{dt} = \omega \quad , \quad \frac{dP}{dt} = 0 .$$

These may be immediately integrated as

$$Q = \omega t + \alpha \quad , \qquad P = \text{constant},$$

and

$$p = m\omega\,(2p/m\omega)^{1/2} \cos\,(\omega t + \alpha),$$

$$q = (2p/m\omega)^{1/2} \sin(\omega t + \alpha).$$

These are of course the familiar equations for the (one-dimensional) harmonic oscillator.

b) As a further example of the technique of canonical transformations, we consider the case of any oscillation around an equilibrium position. A point of equilibrium is characterized, in configuration space, by the equations $\partial V/\partial q^i = 0$, all i. For $n = 2$, imagine the potential function V to be represented by a surface in space; at an equilibrium point this surface has a critical point, which in the stable case is a local minimum. The state of the system behaves like a marble rolling on the potential surface; it oscillates back and forth in the potential well. To see this, we expand T and V by Taylor series about the origin, neglecting all except the quadratic terms:

$$T = \sum a_{ij}\dot{q}^i\dot{q}^j \quad , \quad V = \sum b_{ij}q^i q^j \ ,$$

where the a_{ij} and b_{ij} are constants (here we've used the fact that the point at which $q^1 = q^2 = \ldots = q^n = 0$ is a critical point of V to eliminate the first-order terms in the Taylor expansion of V). We can find a linear transformation to new coordinates $\{r^i\}$ in which T is diagonal: $T = \sum (\dot{r}^i)^2$. The well-known principal-axis theorem now allows us to change coordinates again so that V also assumes a diagonal form, $V = \sum k_i(r^i)^2$; since these changes may be made by an orthogonal transformation (which preserves the inner product), T remains diagonal. But now that we have diagonalized both T and V, we see that each of the i coordinates r satisfies the equations $T = \dot{r}^2$, $V = kr^2$, which we have shown lead to simple harmonic motion. We say that small perturbations around a point of stable equilibrium produce simple harmonic motion in each suitably chosen coordinate.

§19. Canonical Transformations.

We will now prove that every point transformation (one that is given by a smooth, 1-1, onto map of configuration space) is a canonical transformation; at the same time we will be able to get a more natural invariant description of the basic form Ω, which we have been writing as $\Omega = \sum dp_i \wedge dq^i$. Recall that to each map $\varphi: U \to U'$ we associated a linear map $\varphi^*: T^{\varphi(a)}(U') \to T^a(U)$. In particular, we can regard $T^\cdot(U)$ as a local manifold, and the canonical projection π onto U as a smooth map of local manifolds. Then given $w \in T^a(U)$, π maps (a, w) to a and so π^* maps the cotangent space to U at a to the cotangent space at (a, w) to $T^\cdot(U)$:

$$\pi^*: T^a U \longrightarrow T^{(a,w)}(T^\cdot U).$$

Define the one-form ω on $T^\cdot U$ by $\omega(c) = (c, \pi^*(w))$ where $c = (a, w)$ is a point of $T^\cdot U$ (that is, $a \in U$ and $w \in T^a U$), and $\omega(c)$ is a point of $T^\cdot(T^\cdot U)$, since $c \in T^\cdot U$ and $\pi^* w \in T^c(T^\cdot U)$. Let us find what this invariant description becomes in terms of coordinates $\{q^i\}$ in U, and $\{q^i \circ \pi, p_i\}$ in $T^\cdot U$. We can always write w as $d_a f$ for some smooth function f; then

$$w = d_a f = \sum \frac{\partial f}{\partial q^i}\bigg|_a d_a q^i = \sum p_i(c) d_a q^i .$$

Hence

$$\pi^* w = \sum p_i(c)\, d_{(a,w)}(q^i \circ \pi) = \sum p_i(c)\, d_c(q^i \circ \pi) ,$$

so

$$\omega(c) = (c, \sum p_i(c) d_c(q^i \circ \pi)).$$

Thus by abuse of notation $\omega = \sum p_i d(q^i \circ \pi) = \sum p_i dq^i$, which is the same form we have been working with all along. In particular, $\Omega = d\omega$, having been described invariantly, is independent of the particular coordinates we use. But our smooth bijective point transformation φ may be interpreted as nothing but a change of coordinates: if $\{q^i\}$ is a coordinate system on U, then so is $\{q^i \circ \varphi\}$. To say that Ω remains invariant under φ is to say that Ω is the same whether expressed in terms of $\{q^i \circ \varphi\}$ or $\{q^i\}$. Hence φ is a canonical transformation.

For those who like to get their hands dirty, here is a direct proof that φ is canonical: let $\{P_i, Q^i\}$ be the new coordinates on T^*U induced by φ; then

$$dq^i = \sum_u \frac{\partial q^i}{\partial Q^j} dQ^j = \sum_j a^i_j \, dQ^j ,$$

$$p_i = \frac{\partial}{\partial q^i} = \sum_j \frac{\partial}{\partial Q^j} \frac{\partial Q^j}{\partial q^i} = \sum_j p_j b_i^j ,$$

where b_i^j and a_j^i are in fact inverse matrices. Then

$$\omega = \sum_i p_i \wedge dq^i = \sum_i \Big(\sum_j b_i^j P_j\Big) \wedge \Big(\sum_k a_k^i \, dQ^k\Big)$$

$$= \sum_{j,k} \Big(\sum_i \underbrace{b_i^j a_k^i}_{\delta_j^k}\Big) P_j \wedge dQ^k = \sum_j P_j \wedge dQ^k .$$

so φ is indeed canonical.

<u>Definition.</u> A smooth family of maps φ_t is called an <u>infinitessimal canonical transformation</u> if the induced coordinates $\{P_i(t), Q^i(t)\}$ in

phase space satisfy

$$\frac{d}{dt}\left[\sum_{i=1}^{n} dP_i \wedge dQ^i\right]_{t=0} = 0.$$

<u>Theorem.</u> Every motion in phase space satisfying Hamilton's equation is an infinitessimal canonical transformation.

<u>Proof.</u> We will prove more; in fact we will show that the map which takes every point of phase space at $t = 0$ to the point representing the corresponding state of the system at time t is a canonical transformation. For this we calculate:

$$\frac{d}{dt}\left(\sum_i dP_i \wedge dQ^i\right) = \sum_i d\left(\frac{dP_i}{dt}\right)\wedge dQ^i + \sum_i dP_i \wedge d\left(\frac{dQ^i}{dt}\right)$$

(since if f is a function $\frac{\partial}{\partial t}(df) = d(\frac{\partial f}{\partial t})$ is easy to derive)

$$= \sum_i -d\left(\frac{\partial H}{\partial Q^i}\right)\wedge dQ^i + \sum_i dP_i \wedge d\left(\frac{\partial H}{\partial P_i}\right)$$

$$= \sum_i -\left(\sum_j \frac{\partial^2 H}{\partial Q^i \partial Q^j} dQ^j \wedge dQ^i + \sum_j \frac{\partial^2 H}{\partial Q^i \partial P_j} dP_j \wedge dQ^i\right)$$

$$+ \sum_i \left(\sum_j \frac{\partial^2 H}{\partial P_i \partial Q^j} dP_i \wedge dQ^j + \sum_j \frac{\partial^2 H}{\partial P_i \partial P_j} dP^i \wedge dQ^j\right).$$

Since mixed partial derivatives of smooth functions are equal, the second and third terms cancel. But the first and fourth terms are both zero, since $dQ^i \wedge dQ^j = -dQ^j \wedge dQ^i$. Hence $\frac{d}{dt}\left(\sum_i dP_i \wedge dQ_i\right) = \frac{d}{dt}\Omega$ is zero for all times t, so the motion of the system to time t is a canonical transformation for any t.

We sketch another proof of the preceding theorem. Let φ be a smooth map of an open set in $\mathbb{R}^2$ into M, where M is now viewed as any 2n-dimensional manifold. Let the coordinates on $\mathbb{R}^2$ be u and v, and the coordinates on M the usual $\{q^i, p_i\}$. It is easy to see that any 2-form ω on M is determined by the set $\{\varphi^*\omega\}$ for all possible such φ. But

$$\varphi^*(\Sigma\, dp_i \wedge dq^i) = \sum_i d(p_i\varphi)\wedge d(q^i\varphi) = \sum_i \left(\frac{\partial p_i\varphi}{\partial u}du + \frac{\partial p_i\varphi}{\partial v}dv\right) \wedge \left(\frac{\partial q^i\varphi}{\partial u}du + \frac{\partial(q^i\varphi)}{\partial v}dv\right)$$

$$= \sum_i \left(\frac{\partial(p_i\varphi)}{\partial u}\frac{\partial(q^i\varphi)}{\partial v} - \frac{\partial(q^i\varphi)}{\partial u}\frac{\partial(p_i\varphi)}{\partial v}\right) du \wedge dv .$$

The coefficient of $du \wedge dv$ in this formula is called $[u, v]$, the <u>Lagrange bracket</u> of u and v, to prove the theorem, we must show that $d/dt[u, v] = 0$ at zero for all φ. We do the calculation only in the case where M is 2-dimensional, with coordinates p and q:

$$\frac{d}{dt}\left(\frac{\partial p}{\partial u}\frac{\partial q}{\partial v}\right) = \frac{\partial}{\partial u}\left(-\frac{\partial H}{\partial q}\right)\frac{\partial q}{\partial v} + \frac{\partial p}{\partial u}\frac{\partial}{\partial v}\left(\frac{\partial H}{\partial p}\right) \quad \text{(interchanging differential operators)}$$

$$= \frac{\partial^2 H}{\partial p \partial q}\frac{\partial p}{\partial u}\frac{\partial q}{\partial v} - \frac{\partial^2 H}{\partial q\partial q}\frac{\partial q}{\partial u}\frac{\partial q}{\partial v} + \frac{\partial^2 H}{\partial p\partial q}\frac{\partial p}{\partial u}\frac{\partial q}{\partial v} + \frac{\partial^2 H}{\partial p\partial p}\frac{\partial p}{\partial u}\frac{\partial p}{\partial v} .$$

The first and third terms above cancel; and when we subtract $d/dt\left(\frac{\partial p}{\partial v}\frac{\partial q}{\partial u}\right)$ corresponding terms also cancel each other, so the result is zero.

§20. Symplectic Spaces

We now turn to the problem of finding a standard way of writing 2-forms on M. It will turn out that, under suitable conditions, <u>any</u> closed and non-degenerate 2-form can be expressed as $\Sigma\, dp_i \wedge dq^i$ for some set of coordinates $\{p_i, q^i\}$. We look first at the situation on a vector space (i.e., on a single fiber of phase space).

<u>Theorem</u>. Let V be a finite-dimensional vector space, $\omega \in \Lambda_2(V^*)$. Then there is a basis $\{e_1, \ldots, e_m\}$ of V and an integer r such that

$$\omega = e^1 \wedge e^{n+1} + \ldots + e^r \wedge e^{2r} .$$

<u>Proof</u>. Regard ω as an alternating bilinear form on V. We may assume $\omega \neq 0$; then we can find linearly independent vectors e_1, e_2 with $\omega(e_1, e_2) \neq 0$. By scalar multiplication we can adjust e_1 and e_2 so that $\omega(e_1, e_2) = 1$. Now let S be the subspace of all $v \in V$ satisfying $\omega(e_1, v) = \omega(e_2, v) = 0$. Calculation shows that no linear combination of e_1 and e_2 lies in S. Furthermore, V is spanned by S, e_1, and e_2. For let z be any vector of V; we wish to find numbers x and y such that $v = z - xe_1 - ye_2$ lies in S. To force $\omega(e_1, v) = 0$ we must have $\omega(e_1, z - xe_1 - ye_2) = \omega(e_1, z) - y = 0$. Thus $y = \omega(e_1, z)$; similarly, we can take $x = -\omega(e_2, z)$. This accomplished, we now apply the same technique to the form ω restricted to S. We find $e_3, e_4 \in S$ and a subspace $S' \subsetneq S$ such that no linear combination of e_3 and e_4 lies in S', but e_3, e_4 and S' span S. Continuing in this fashion, we eventually find $S^{(k)}$ on which ω is identically zero. Then, choosing any basis $\{e_{2k+3}, \ldots, e_m\}$

for $S^{(k)}$, we get a basis $\{e_1, \ldots, e_m\}$ for V with the property that $\omega(e_1, e_2) = \omega(e_3, e_4) = \ldots = \omega(e_{2k+1}, e_{2k+2}) = 1$, and all other $\omega(e_i, e_j) = 0$ (except for reversals of the above, e.g., $\omega(e_2, e_1) = -1$). This means that $\omega = e^1 \wedge e^2 + e^3 \wedge e^4 + \ldots + e^{2k+1} \wedge e^{2k+2}$. Renumbering the e's gives us the desired formula, as in the theorem.

Each ω determines a linear map $\omega^\flat : V \longrightarrow V^*$ given by $[\omega^\flat(v)]v' = \omega(v, v')$. We say ω is <u>non-degenerate</u> if $\omega^\flat$ is an isomorphism. Since V and V^* are finite-dimensional, this is equivalent to saying that $\omega^\flat$ has zero null-space; in other words, $\omega(v, v') = 0$ for all v' implies $v = 0$. Using now the canonical form given in the theorem, we derive

<u>Corollary 1</u>. If ω is non-degenerate and in the form given by the theorem, then V is $2r$-dimensinal.

For if $m > 2r$, $\omega(e_{2r+1}, v) = 0$ for all $v \in V$.

More computation with the canonical form establishes

<u>Corollary 2.</u> If V is $2n$-dimensional, ω is non-degenerate if and only if $\omega \wedge \omega \wedge \ldots \wedge \omega$ (n times) $= 0$.

<u>Corollary 3.</u> The integer r in the theorem is determined by

$$\omega^r \neq 0 \ , \ \omega^{r+1} = 0 \ .$$

The number $2r$ is called the <u>rank</u> of ω.

<u>Definition.</u> A <u>symplectic vector space</u> is a finite-dimensional vector space V with a non-degenerate form $\omega \in \Lambda_2(V^*)$.

We have proved that every symplectic vector space has dimension $2n$ for some integer n and possesses a symplectic basis $\{e_i\}$ for which $\omega = \sum_{i=1}^{n} e^i \wedge e^{n+i}$. In this basis, the matrix of the bilinear form ω is

$$\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Furthermore,

$$(\omega^\flat e_i)(e_j) = \omega(e_i, e_j)$$
$$= \sum_{1 \le k \le n} \frac{1}{2}(e^k \otimes e^{n+k} - e^{n+k} \otimes e^k)(e_i, e_j) .$$

Hence $\omega^\flat(e_i) = \frac{1}{2} e^{n+i}$, $1 \le i \le n$. Similarly, $\omega^\flat(e_{n+i}) = -\frac{1}{2} e^i$. If ω is non-degenerate, $\omega^\flat$ is an isomorphism, and so is $2\omega^\flat$.

This map $2\omega^\flat : V \longrightarrow V^*$ is given in this basis by the convenient formulas

$$(2\omega^\flat)(e_i) = e^{n+i} ,$$
$$(2\omega^\flat)(e_{n+i}) = -e^i , \quad i = 1, \ldots, n.$$

We let $\omega^\# = (2\omega^\flat)^{-1}$ be the inverse map.

§21 Hamilton's equations

Now apply this machinery to $T.(M)$. Here $M = T^{\cdot}U$ is our 2n-dimensional phase space, and we are given a non-degenerate closed 2-form ω on M. This form ω yields maps $2\omega^\flat$ and $\omega^\#$ on each tangent space of M, and hence gives a bundle map $2\omega^\flat : T.(M) \longrightarrow T^{\cdot}(M)$. Also, the Hamiltonian H is a function on M, and thus gives rise to a one-form dH on M. Here is the diagram:

$$\begin{tikzcd}
T_{\cdot}(M) \arrow[rr, shift right, "2\omega^{\flat}"'] \arrow[dr] & & T^{\cdot}(M) \arrow[ll, shift right, "\omega^{\#}"'] \arrow[dl, shift right] \\
& M \arrow[ur, shift right, "dH"'] &
\end{tikzcd}$$

.

Explicitly, if c is a point of M, we have the map $2\omega^{\flat}: T_c M \to T^c M$, where we have identified V above with $T_c M$, V^* with $T^c M$. This allows us to pass from vector fields to one-forms and vice versa, since $\omega^{\flat}$ is an isomorphism. Specifically, $\omega^{\#}(dH)$ is a vector field on M. Identifying the $\{dq^i\}$ with $e^1, \ldots, e^n$ and the $\{dp_i\}$ with $e^{n+1}, \ldots, e^n$, we compute

$$dH = \sum \frac{\partial H}{\partial q^i} dq^i + \sum \frac{\partial H}{\partial p_i} dp_i \;;$$

that is,

$$dH = \left(\frac{\partial H}{\partial q^i}, \ldots, \frac{\partial H}{\partial q^n}, \frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n} \right).$$

Therefore

$$\begin{aligned} \omega^{\#}(dH) &= \left(\frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q^1}, \ldots, -\frac{\partial H}{\partial q^n} \right) \\ &= \sum \frac{\partial H}{\partial q^i} \frac{\partial}{\partial q^i} - \sum \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p_i} . \end{aligned}$$

The vector field $X = \omega^{\#}(dH)$ determines a system of differential equations on M, namely $dc/dt = X(c)$, where c is a path in M. This equation means that the curve c threads its way through the tangent vector field X on M in such a way that the tangent vector to c at any point is exactly the same as the value of the field X at that point. But in coordinates, the paths which satisfy the equation $dc/dt = X(c)$ are those which satisfy Hamilton's equations:

$$\frac{\partial q^i}{\partial t} = \frac{\partial H}{\partial p_i}, \quad \frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q^i}$$

if $X = \omega^{\#}(dH) = (\frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q^i}, \ldots, -\frac{\partial H}{\partial q^n})$. In other words, we have reached an invariant way of stating Hamilton's equations on a local manifold.

§22 Symplectic manifolds

Generalizing the results of our last lecture, we define a symplectic manifold to be a manifold possessing a closed non-degenerate two-form ω. We have seen that $\omega \in \Lambda_2(V^*)$ may be viewed as an alternating bilinear function on $V \times V$ to $\mathbb{R}$; thus if v and v' are vectors of V, $\omega(v, v')$ is a real number. We will say that a linear transformation $f: V \to V$ is symplectic if it is one-to-one onto and preserves the form ω; that is, $f^*\omega = \omega$, or $\omega(fv, fv') = \omega(v, v')$ for all $v, v' \in V$. The set of all such f is called the symplectic group; it is in fact a group in the mathematical sense, since the operation of composition is associative, has inverses, and has a unit element. In particular though, a symplectic transformation is a linear transformation on a finite-dimensional vector space, and thus may have eigenvalues.

Theorem. If λ is an eigenvalue of a symplectic transformation, then so are $\bar{\lambda}$ (the complex conjugate of λ), $1/\lambda$ and $1/\bar{\lambda}$; in particular $\lambda \neq 0$.

Proof. Over the complex numbers, eigenvalues always exist; so our first task is to create a complex vector space out of the real vector space V; that is, to find a way of multiplying by complex as well as real scalars. It is not hard to see that taking the tensor product $\mathbb{C} \otimes V = \tilde{V}$ results in a complex vector space $\tilde{V}$, where the product of a vector $\sum c_i \otimes v_i$ by a complex scalar c is $c(\sum c_i \otimes v_i) = \sum cc_i \otimes v_i$. Also, the real dimension of $\tilde{V}$ is twice that of V: if $\{e_1, \ldots, e_m\}$ is a basis of V, then since $\{1, i\}$ is a basis of $\mathbb{C}$ considered as a real vector space, a basis of $\tilde{V}$ as a real vector space is $1 \otimes e_1, \ldots, 1 \otimes e_m, i \otimes e_1, \ldots, i \otimes e_m$; there are 2m of these vectors. But the m vectors $1 \otimes e_1, \ldots, 1 \otimes e_m$ form a basis of $\tilde{V}$ over $\mathbb{C}$ since scalar multiplication by i converts $1 \otimes e_j$ into $i \otimes e_j$. We now extend ω to a form $\tilde{\omega}$ on $V \otimes \mathbb{C}$, defining

$$\tilde{\omega}(c \otimes v, c' \otimes v') :: cc'\omega(v, v')$$

for any two complex numbers c and c' and any two vectors $v, v' \in V$, and extending this definition by linearity. Similarly, each linear $f: V \longrightarrow V$ goes over to $\tilde{f} = 1 \otimes f: \mathbb{C} \otimes V \longrightarrow \mathbb{C} \otimes V$. Moreover, the formula $\tilde{\omega}(\tilde{f}v, \tilde{f}v') = \tilde{\omega}(v, v')$ still holds; that is, if f is symplectic, $\tilde{f}$ is symplectic with respect to $\tilde{\omega}$.

Now suppose u is an eigenvector corresponding to λ : $u \neq 0$, $f(u) = \lambda u$. If λ were zero, u would be in the null-space of f, contradicting the assumption that f was one-to- one. But then if u' is any other vector, $\tilde{\omega}(u, u') = \tilde{\omega}(\tilde{f}u, \tilde{f}u') = \tilde{\omega}(\lambda u, \tilde{f}u') = \tilde{\omega}(u, \lambda \tilde{f}u')$, so $\tilde{\omega}(u, u' \quad \lambda\tilde{f}(u')) = 0$. Now if the map taking $u' \rightsquigarrow u' - \lambda\tilde{f}(u')$ were onto, every vector w of $\tilde{V}$ could be written in the form $w = v - \lambda\tilde{f}(v)$. Then we'd have $\omega(u, w) = 0$ for every $w \in \tilde{V}$, which would mean that $u = 0$ since ω is non-singular. This is a contradiction. Hence the map mentioned above is not onto; since it's a linear transformation between vector spaces of the same dimension, it's also not one-to-one. Thus there is a u' with $u' - \lambda\tilde{f}(u') = 0$; this says that $\tilde{f}(u') = \frac{u'}{\lambda}$, so $\frac{1}{\lambda}$ is an eigenvalue of $\tilde{f}$, and hence an eigenvalue of f. Also $\tilde{f}(\bar{v}) = \overline{\tilde{f}(v)}$ by the definition of f, so $\tilde{f}(u) = \lambda u$ implies $\overline{\tilde{f}(u)} = \overline{\lambda u}$ which implies $\tilde{f}(\bar{u}) = \bar{\lambda}\bar{u}$. Hence $\bar{\lambda}$ is an eigenvalue, so by what we've already proved, $\frac{1}{\bar{\lambda}}$ is an eigenvalue. This completes the proof of the theorem.

§23 <u>The Poincaré Lemma</u>.

We return to the problem of expressing our closed form ω in some system of coordinates as $\omega = \sum dp_i \wedge dq^i$ in some <u>region</u>, not just at a point. We first prove the Poincaré lemma:

<u>Theorem</u> Let U be an open ball in $\mathbb{R}^n$. Let ω be a closed k-form on U; that is, $d\omega = 0$. Then there is a (k-1)-form η on U such that $d\eta = \omega$. (succinctly: closed forms are exact on U).

<u>Proof.</u> We will first derive a new formula for the differential d, which makes (k+1)-forms out of k-forms. We will then use the assumption on U to define a new map s which makes (k-1)-forms out of k-forms

for each k, such that $ds(\omega) + sd(\omega) = \omega$ for every form ω. If then we have an ω with $d\omega = 0$, it will follow that $\omega = d(s\omega)$, showing that ω is exact. We will also let V denote the tangent space (at any point) of U

A k-form ω may be regarded as a smooth map from U to $\Lambda_k(V^*)$, the space of alternating k-tensors on V. Thus for each $u \in U$, ω_u is an alternating k-tensor: $v_1, \ldots, v_k \in V$ implies that $\omega_u(v_1, \ldots, v_k) \in \mathbb{R}$. Write $\omega(u, v_1, \ldots, v_k) = \omega_u(v_1, \ldots, v_k)$; then ω is a function smooth in the first argument, and linear and alternating in the last k arguments.

Suppose f is a smooth real-valued function on U. We define a new function $Df: U \times V \longrightarrow \mathbb{R}$ by letting $Df(u, v) = \langle d_u f, v \rangle$; that is, $Df(u, v) = \left.\frac{d(f \circ \tilde{v})}{dt}\right|_{t=0}$ where v is the path defined by $\tilde{v}(t) = u + tv$. Hence Df is nothing more than the directional derivative of f in the direction v at the point u. Now if f happens to be a function of other variables as well, we can still form Df by ignoring those other variables as we take the derivative, and then putting them back: thus if $f = f(u, w_1, \ldots, w_r)$,

$$Df(u, v, w_1, \ldots, w_r) = (d/dt) f(u+tv, w_1, \ldots, w_r)\Big|_{t=0}$$

Notice that Df is a linear function of v; if also f happens to be a linear function (in u), $Df(u, v) = f(v)$.

If ω is a k-form, redefine the (k+1)-form $d\omega$ by

$$(d\omega)(u, v_o, \dots, v_k) = \sum_{\ell=0}^{k} (-1)^\ell (D\omega)(u, v_\ell, v_o, v_1, \dots, \hat{v}_\ell, \dots, v_k) .$$

(Here the $\wedge$ over v_ℓ means that v_ℓ is omitted.) We claim this $d\omega$ is the same as the $d\omega$ defined previously. This is checked by showing that this $d\omega$ is linear and alternating in the $v_o, \dots, v_k$, and has the same values on the basis elements of $V \times V \times \dots \times V$ as the old $d\omega$. The linearity is clear, given our comments regarding the operator D; $d\omega$ is alternating since computation shows that it vanishes when any two successive arguments are equal. Suppose now ω is a one-form ; $\omega = \sum w_i dq^i$, where $\{q^i\}$ are coordinates on M and $\{e_i\}$ are the corresponding basis elements of $V \cong T_u(M)$. Then $w_i(u) = \omega(u, e_i)$. By our old definition

$$d\omega = \sum_{i<j} \left(\frac{\partial w_j}{\partial q^i} - \frac{\partial w_i}{\partial q^j}\right) dq^i \wedge dq^j = \sum_{i<j} dw(u, e_i, e_j) dq^i \wedge dq^j .$$

To prove that the two definitions coincide for one-forms it will thus suffice to show that $d\omega(u, e_i, e_j)$ is the same as in the new definition. But in the new definition

$$d\omega(u, e_i, e_j) = D\omega(u, e_i, e_j) - D\omega(u, e_j, e_i),$$

and

$$Df(u, e_i) = \partial f / \partial q_i .$$

Hence

$$d\omega(u, e_i, e_j) = \frac{\partial\omega(u, e_j)}{\partial q^i} - \frac{\partial\omega(u, e_j)}{\partial q^j} = \frac{\partial\omega_j}{\partial q^i} - \frac{\partial\omega_i}{\partial q^j} ,$$

which is what we were trying to prove. Similar techniques show that the two definitions are the same for general k-forms.

We are now ready to define the map s which makes a (p-1)-form out of every p-form. If ω is a k-form, let

$$(s\omega)(u;v_1,\dots,v_{k-1}) = \int_0^1 t^{k-1}\omega(tu;u,v_1,\dots,v_{k-1})\,dt.$$

Here we consider the open set U as part of the vector space $V = \mathbb{R}^n$, which has also been identified with $T_p(U)$. Thus on the right-hand side of the equation, the second argument, $u \in U$, is viewed as a vector of V. But since U is an open ball, tu, the first argument, is in U for all $t \neq 1$. It is now easy to check that $s\omega$ is a (k-1)-form -- linear, alternating, and smooth as a function of u.

We now take a k-form ω and show, at last, that $ds(\omega) + sd(\omega) = \omega$. First,

$$D(s\omega)(u,v;v_1,\dots,v_{k-1}) = \int_0^1 D[t^{k-1}\omega(tu,v,u;v_1,\dots,v_{k-1})]dt$$

(since all functions involved are smooth and bounded)

$$= \int_0^1 t^k D\omega(tu,v,u,v_1,\dots,v_{k-1})\,dt + \int_0^1 t^{k-1}\omega(tu,v;v_1,\dots,v_{k-1})\,dt\,.$$

The latter term appears as it does since ω is linear in the third variable, and it was proved that if f is linear, $Df(u,v) = f(v)$. Now

$$d(s\omega)(u, v_1, \ldots, v_k) = \sum_{\ell=1}^{k} (-1)^{\ell-1} D(s\omega)(u, v_\ell, v_1, \ldots, \hat{v}_\ell, \ldots, v_k)$$

$$= \sum_{\ell=1}^{k} (-1)^{\ell-1} [\int_0^1 t^k D\omega(tu, v_\ell, u, v_1, \ldots, \hat{v}_\ell, \ldots, v_k)\, dt$$

$$+ \int_0^1 t^{k-1} \omega(tu, v_\ell, v_1, \ldots \hat{v}_\ell, \ldots, v_k)\, dt] ,$$

and

$$s(d\omega)(u, v_1, \ldots, v_k) = \int_0^1 t^k d\omega(tu, u, v_1, \ldots, v_k)\, dt$$

$$= \int_0^1 \sum_{\ell=1}^{k} (-1)^\ell t^k D\omega(tu, v_\ell, u, v_1, \ldots, \hat{v}_\ell, \ldots, v_k)\, dt$$

$$+ \int_0^1 t^k D\omega(tu, u, v_1, \ldots, v_k)\, dt .$$

When we add $d(s\omega)$ and $s(d\omega)$, the first terms of each expression cancel; also,

$$\sum_{\ell=1}^{k} (-1)^{\ell-1} \int_0^1 t^{k-1} \omega(tu, v_\ell, v_1, \ldots, \hat{v}_\ell, \ldots, v_k)\, dt$$

$$= \sum_{k=1}^{\ell} (-1)^{\ell-1} \int_0^1 (-1)^{\ell-1}\, t^{k-1} \omega(tu, v_1, \ldots, v_k)\, dt \qquad \text{since } \omega \text{ is alternating}$$

$$= k \int_0^1 t^{k-1} \omega(tu, v_1, \ldots, v_k) dt .$$

Hence

$$(sd\omega + ds\,\omega)(u, v_1, \ldots, v_k) = \int_0^1 [t^k D\omega(tu, u, v_1, \ldots, v_k) + kt^{k-1} \omega(tu, v_1, \ldots, v_k)] dt$$

$$= \int_0^1 \frac{d}{dt} [t^k \omega(tu, v_1, \ldots, v_k)]\, dt$$

(since $D\omega(tu, u, \ldots)$ is just the directional derivative in direction u of $\omega(tu, v_1, \ldots)$);

$$= 1^k \omega(u, v_1, \ldots, v_k) - 0 = \omega(u, v_1, \ldots, v_k). \quad \text{Q. E. D.}$$

§ 24. The Lie Derivative

Let X be a vector field on U. There is an operation on the ring $\mathcal{F}$ of smooth functions on U defined by

$$L_X(f) = Df(u, X(u)) = (X_u)(f).$$

This operaton L_X is a derivation, since each $X_u \in T_u(U)$ is a derivation. L_X is called the Lie derivative. We have shown that a vector field is determined by the way it acts on the functions of $\mathcal{F}$; this means that knowing the operator L_X determines X.

Now it is easy to check that if θ and ψ are derivations of $\mathcal{F}$, then so is $\theta\psi - \psi\theta$. Call this new derivation $[\theta, \psi]$. Then $[L_X, L_Y]$ is a derivation, so to it there is associated a unique vector field. This vector field is called the Lie bracket of X and Y and is written $[X, Y]$. In coordinates $\{q^1, \ldots, q^n\}$, let $X = \sum x^i \frac{\partial}{\partial q^i}$, $Y = \sum y^i \frac{\partial}{\partial q^i}$, $x^i, y^i \in \mathcal{F}$. Then $L_X(f) = \sum x^i \frac{\partial f}{\partial q^i}$ for $f \in \mathcal{F}$, so

$$L_{[X, Y]} = \sum_i \Big(\sum_j x^j \frac{\partial y^i}{\partial q^j} - y^j \frac{\partial x^i}{\partial q^j} \Big) \frac{\partial}{\partial q^i}.$$

In fact, it is easy to check without using coordinates that $[\ ,\ ]$ is linear in each argument and satisfies the relations

$$[X, X] = 0 \quad , \quad [X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0 \quad \text{(Jacobi identity).}$$

We will now show that the Lie bracket provides a natural way of extending the Lie-derivative operation to apply to all tensor fields. There is an identity tensor $\delta \in \mathcal{J}^1_1(V)$; δ corresponds to the identity map on V^* under the series of identifications

$$\mathcal{J}^1_1(V) = V \otimes V^* \cong \mathrm{Hom}(V^* \otimes V, \mathbb{R}) \cong \mathrm{Hom}(V^*, \mathrm{Hom}(V, \mathbb{R})) \cong \mathrm{Hom}(V^*, V^*).$$

In coordinates $\{e_j\}$ we find $\delta = \sum e_i \otimes e^i$.

Theorem. Given any vector field X on U, there is a unique linear map L_X, where $L_X: \mathcal{J}^r_s(U) \longrightarrow \mathcal{J}^r_s(U)$ for all non-negative integers r and s, with the properties

1. $L_X f = \langle df, X \rangle$ for any smooth function f on U,
2. $L_X Y = [X, Y]$ for any vector field Y,
3. $L_X \delta = 0$,
4. L_X is a derivation; that is,

$$L_X(\tau \otimes \tau') = (L_X \tau) \otimes \tau' + \tau \otimes (L_X \tau'),$$

for any tensor fields τ and τ'.

Proof. We put coordinates $q^1, \ldots, q^n$ on U and show that any operator L_X satisfying the given conditions must satisfy

$$L_X(dq^k) = \sum_i \frac{\partial x^k}{\partial q^i} dq^i \qquad \text{for all } k. \tag{5}$$

Write $X = \sum X^i \frac{\partial}{\partial q^i}$; then we have shown that

$$L_X Y = \sum_{i,j} \left(X^j \frac{\partial Y^i}{\partial q^j} - Y^j \frac{\partial X^i}{\partial q^j}\right) \frac{\partial}{\partial q^i}, \quad \text{if } Y = \sum Y^i \frac{\partial}{\partial q^i}.$$

In particular, $L_X\left(\frac{\partial}{\partial q^k}\right) = -\sum_i \frac{\partial X^i}{\partial q^k} \frac{\partial}{\partial q^i}$.

Now $\delta = \sum_k \frac{\partial}{\partial q^k} \otimes dq^k$. So

$$0 = L_X\delta = \sum_k L_X\left(\frac{\partial}{\partial q^k}\right)\otimes dq^k + \frac{\partial}{\partial q^k} \otimes L_X dq^k$$

$$= -\sum_k \sum_i \frac{\partial X^i}{\partial q^k}\left(\frac{\partial}{\partial q^i} \otimes dq^k\right) + \sum_k \frac{\partial}{\partial q^k} \otimes L_X dq^k .$$

If we write $L_X(dq^k) = \sum c_k^i \, dq^i$, it is clear that we are forced to take $c_k^i = \partial X^k/\partial q^i$. But it is clear from (4) that once L_X is defined on functions, vector fields, and co-vector fields, it extends uniquely to all tensor fields. So we merely define L_X by (1), (2), and (5), and check that (3) and (4) are satisfied. Notice that this is really an invariant proof, since we have shown that any extension of the Lie derivative satisfying 1-4 must, when expressed in coordinates, agree with the operator we've defined.

Corollary 1. $L_X(df) = d(L_X f)$.

For example,

$$d(L_X q^k) = d\left(\sum X^i \frac{\partial q^k}{\partial q^i}\right) = dX^k = \sum_i \frac{\partial X^k}{\partial q^i} dq^i = L_X(dq^k).$$

Corollary 2. If V is an open subset of U, and $|_V$ denotes the restriction to V, then

$$(L_X\tau)|_V = L_{(X|_V)}(\tau|_V).$$

Corollary 3. L_X maps $\Omega_k(U)$ into $\Omega_k(U)$.

Proof. Think of an exterior form as an alternating tensor; recall that a tensor τ is alternating if and only if $A\tau = \tau$. Hence, we must show that if τ is alternating, $A(L_X\tau) = L_X\tau$. In fact, $A(L_X\tau) = L_X(A\tau)$, since A is a sum of permutation operators, and it is easy to see that L_X commutes with permutations.

Corollary 4. $L_X(\omega\wedge\eta) = L_X\omega\wedge\eta + \omega\wedge L_X\eta$, if ω and η are exterior forms.

Proof.
$$\begin{aligned} L_X(\omega\wedge\eta) &= L_X(A(\omega\otimes\eta)) = AL_X(\omega\otimes\eta) \\ &= A(L_X\omega\otimes\eta + \omega\otimes L_X\eta) \\ &= L_X\omega\wedge\eta + \omega\wedge L_X\eta . \end{aligned}$$

Corollary 5. $d(L_X\omega) = L_X(d\omega)$ if ω is a k-form.

Proof. Write $\omega = \sum f dq^1 \wedge dq^2 \wedge \ldots \wedge dq^k$. Then $d\omega = \sum df \wedge dq^1 \wedge \ldots \wedge dq^k$, while
$$L_X\omega = \sum L_X f\, dq^1 \wedge \ldots\ dq^k + \sum\sum_{i=1}^{k} f\, dq^1 \wedge \ldots \wedge L_X dq^i \wedge \ldots \wedge dq^k.$$
Computation now proves the equality, with the aid of the preceding corollaries.

§25 Transportation along Trajectories

This purely formal proof of the properties of the Lie derivative does not really shed much light on the geometrical meaning of this operator. Actually, as we are about to show, the Lie derivative can be interpreted in a way very much like the ordinary definition of a derivative: the limit of a difference quotient.

Recall that every smooth vector field X on a local manifold M has integral curves c passing through every point of M. An integral curve is one whose tangent at every point p is the same as the tangent vector which is the value of the vector field X at p; symbolically, if $p = c(t_0)$,

$$dc/dt(t_0) = X_{c(t_0)} , \quad \text{or} \quad dc/dt = X \circ c.$$

The standard existence and uniqueness theorem for ordinary differential equations, when applied to this equation expressed in coordinates, guarantees the existence of at least one integral curve through each point of M (although each curve may be defined only on a small interval on the real line); furthermore, two integral curves passing through the same point must agree wherever they are both defined.

Now change the point of view slightly, and consider the motion of M which takes each point p to the point t units along the integral curve passing through p. For a fixed t for which all the integral curves are defined, this would describe a map from M to M. In these terms, the existence and uniqueness theorem may be stated:

> If X is a smooth vector field on a local manifold M, there is a unique trajectory of X through each point, and for each point p of M there is a neighborhood U of p and an interval $I \subset \mathbb{R}$, together with a function $F: I \times U \longrightarrow M$, such that $F(0, u) = u$ (initial conditions) and for fixed u, the map taking t to $F(t, u)$ is a trajectory of X.

Write F_t for the map of M to M defined by $F_t(u) = F(t, u)$. In the case where M and U happen to be all of $\mathbb{R}^n$ and $I = \mathbb{R}$; we must have $F_t F_{t'}(u) = F_{t+t'}(u)$ for all t and t'. Indeed. $t \rightsquigarrow F_t(F_{t'}(u))$ is a trajectory starting at $F_{t'}(u)$, but so is $t \rightsquigarrow F_{t+t'}(u)$; by uniqueness, they must be equal. In particular,

$$\left.\begin{aligned} F_t(F_{-t}(u)) &= u \\ F_{-t}(F_t(u)) &= u \end{aligned}\right\} \text{ for each } u \in M ,$$

so each F_t has a smooth inverse map. In this case we say that F_t is a diffeomorphism. Then $t \rightsquigarrow F_t$ is a map of the additive group of real numbers into the group of diffeomorphisms of $\mathbb{R}^n$. In general, of course, the map F will be defined only on subsets I and U of $\mathbb{R}$ and $\mathbb{R}^n$; it is not hard to see, however, that restricting further to an interval $I' \epsilon I$ and an open set $U' \epsilon U$ gives us a map $F' = F|_{I' \times U'}$ for which F'_t has the smooth inverse F'_{-t} for each $t \epsilon I'$. (Cf. Abraham pp. 39-40). Such a system F' will be called a <u>flow box</u> of X.

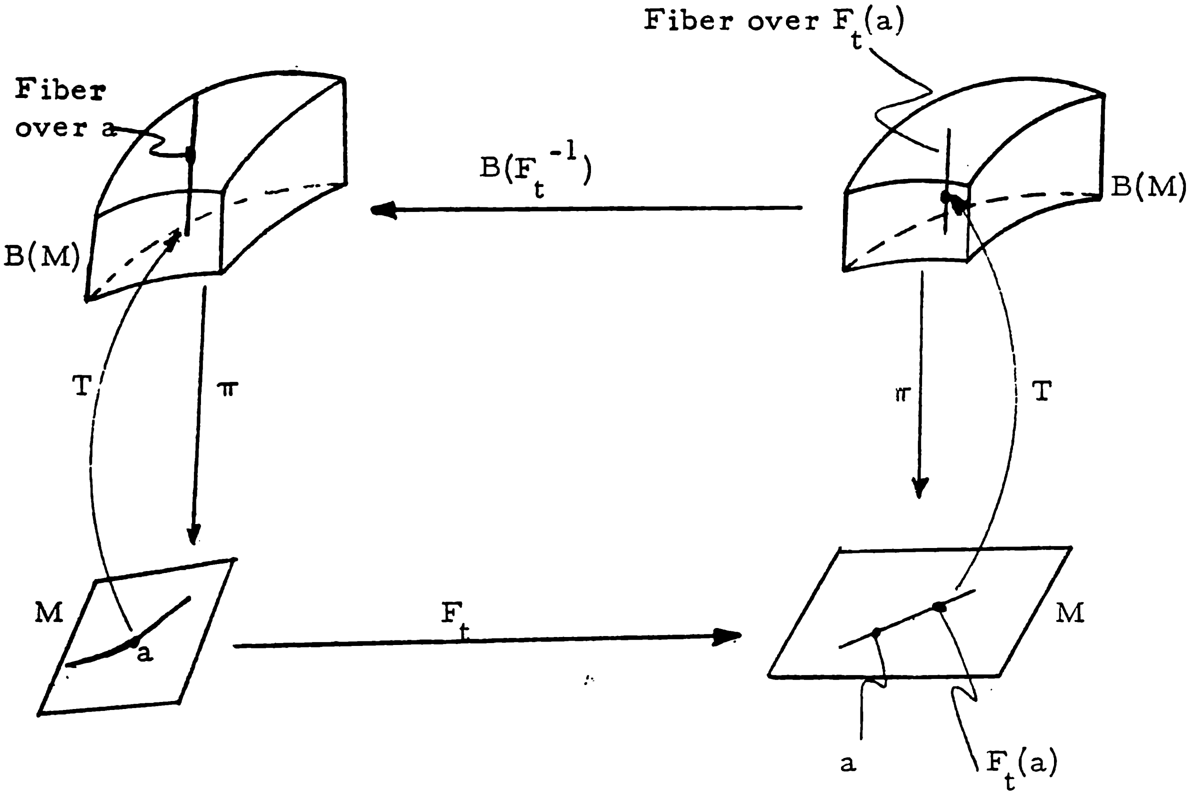
Suppose now that B is a functor from vector spaces to vector spaces: for example, $B(V) = V^*$, or $B(V) = V \times V$, or $B(V) = V \wedge \ldots \wedge V$. Associated to such a B is a bundle $B(M)$ over M, whose fiber over p is $B(T_p M)$. Thus if $B(V) = V$, $B(M) = T.M$; if $B(V) = V \wedge V$, $B(M)$ is the bundle of all 2-forms on M; and so on. Furthermore, if $f: M \longrightarrow N$, is a smooth map, then there is a map

$B(f)\colon B(M) \longrightarrow B(N)$ if B is covariant;

$B(f)\colon B(N) \longrightarrow B(M)$ if B is contravariant.

For instance, if $B(V) = V^*$, $B(f) = f^*\colon T^{\cdot}(N) \longrightarrow T^{\cdot}(M)$.

To define the Lie derivative of a general field T with respect to X we wish to do the following: given a point p, move along the trajectory of X through p for a time t; at this point, find the value of T, and now move T back along the trajectory to get a tensor at p. This pulled-back tensor will not generally be the same as the value of the tensor field at p; but we can form the difference quotient.



Definition. If X is a smooth vector field on M and T is a co- or contra-variant field on M, define the operator K_X by

$$K_X(T) = \begin{cases} d/dt(B(F_t^{-1}) \circ T \circ F_t)\big|_{t=0} & \text{if } T \text{ is covariant} \\ d/dt(B(F_t) \circ T \circ F_t)\big|_{t=0} & \text{if } T \text{ is contravariant}, \end{cases}$$

where F is a flow box for X.

Notice that this definition does not depend on the choice of F, by the uniqueness of flow boxes.

Theorem. K_X is the same as the Lie derivative L_X.

Proof. By the Theorem of § 24 characterizing L_X, it will suffice to show that K_X is a derivation, $K_X\delta = 0$, and that K_X agrees with L_X on functions and vector fields. In fact, examination of the proof of that theorem makes it clear that we can show $K_X = L_X$ on covector fields instead of on vector fields, and the result will still follow.

First, $K_X\delta = 0$ since δ may be expressed as an identity matrix invariant under F_t and $B(F_t)$; its derivative is zero. Showing that K_X is a derivation will involve working with functors B, B', and B'', and a covariant, bilinear, natural transformation $\square : B \times B' \longrightarrow B''$. Then if $T: M \longrightarrow B(M)$, $T': M \longrightarrow B'(M)$, we can define $T \times T'$, mapping M into the pullback bundle $B(M) \times_M B'(M)$; $T \square T'$ is the composite map

$$M \xrightarrow{T \times T'} B(M) \times_M B'(M) \xrightarrow{\square_M} B''(M)$$

To show that $K_X(T \square T') = K_X T \square T' + T \square K_X T'$, we compute

$$B''(F_t^{-1})\circ T \,\Box\, T'\circ F_t = [B(F_t^{-1})T \,\Box\, B'(F_t^{-1})T']\circ F_t$$
$$= [B(F_t^{-1})T]F_t \,\Box\, [B'(F_t^{-1})T']F_t \, .$$

Now in general, suppose $\sigma(t), \sigma'(t)$ are maps of an interval I to V and V', respectively; $\sigma(t)\,\Box\,\sigma'(t) \in V \,\Box\, V' = V''$. Since $\Box$ is bilinear, we can write

$$e''_i(\sigma(t)\,\Box\,\sigma'(t)) = \sum_{j,k} c^i_{jk}\, e_j(\sigma(t))e'_k(\sigma'(t))$$

for some constants c^i_{jk}; where $\{e_i\}$ are coordinates on V, and so on. Now the ordinary Leibnitz rule for the derivative of a product of real-valued functions applies, and it follows that K_X is a derivation.

If f is a real-valued function on M, the fiber over every point is $\mathbb{R}$, and in this case $B(F_t^{-1})$ is always the identity map. Thus

$$K_X(f) = \frac{d}{dt}(f(F_t(a))) = \sum_i \frac{\partial f}{\partial q^i}\frac{\partial F_t^i}{dt} = \sum x^i \frac{\partial f}{\partial q^i} = L_X f \; ,$$

where $q^i F_t = F_t^i$; and $X = \sum_i x^i \dfrac{\partial}{\partial q^i}$

Finally, to show that K_X agrees with L_X on covector fields, we can use the properties already proved to find $K_X(\sum f_i dq^i)$, once we know $K_X(dg)$ for any function g; hence it will be enough to prove $K_X(dg) = L_X(dg)$. Now in this case $B(V) = V^*$, so $B(F_t) = F_t^*$. In general, however, we defined the pullback of a form ω by a map φ as the form $\varphi^*\omega$ given by

$$(\varphi^*\omega)(a) = \varphi^*(\omega_{\varphi(a)}) = (\varphi^*\circ\omega\circ\varphi)(a) \; .$$

Thus

$$
\begin{aligned}
K_X(dg) = \frac{d}{dt}(B(F_t)\circ dg\circ F_t) &= \frac{d}{dt}(F_t^{*}\circ dg\circ F_t) \\
&= \frac{d}{dt}(F_t^{*}(dg)) = \frac{d}{dt}(d(g\circ F_t)) \\
&= d(\frac{d}{dt}(g\circ F_t)) = d(L_X g) = L_X(dg)
\end{aligned}
$$

by previous part of the theorem.

This completes the proof of the theorem.

<u>Note</u>. In the above proof we considered a k-form as a cross-section of a suitable (exterior) bundle. This means in particular that we should consider a 0-form (= a smooth function) as a cross-section; we will show that it is a cross-section of the trivial bundle. Explictly, take that functor B which sends each vector space V to the one-dimensional space $\mathbb{R}$ and each linear transformation $f: V \longrightarrow V'$ to $1: \mathbb{R} \longrightarrow \mathbb{R}$. (Thus B is a "constant" functor). If B is used to construct a fiber bundle over M it gives the bundle

$$
\begin{array}{ccc}
M \times \mathbb{R} & & (m, k) \\
\downarrow & & \rotatebox[origin=c]{-90}{$\rightsquigarrow$} \\
M & & m
\end{array}
$$

. A cross-section here is clearly just a smooth function $M \longrightarrow \mathbb{R}$.

§26. <u>Canonical Transformations described by generating functions</u>

A function $F(q^1, \ldots, q^n, P_1, \ldots, P_n)$ of $2n$ variables will yield a canonical transformation. We first describe informally how this arises. Suppose that the quantities $q^1, \ldots, q^n, P_1, \ldots, P_n$ are coordinates on some local manifold U and that the matrix

$$\left\| \frac{\partial^2 F}{\partial q^i \partial P_j} \right\|$$

is everywhere non-singular. Define 2n more quantities p_i and Q^i (= smooth functions) on U by the equations

$$q^i = q^i \; ; \quad i = 1, \ldots, n \qquad Q^i = \frac{\partial F}{\partial P_i}$$
$$p_i = \frac{\partial F}{\partial q^i} \, , \; i = 1, \ldots, n \qquad P_i = P_i \; , \quad i = 1, \ldots, n \, .$$

The assumption on the matrix above, plus the standard implicit function theorem, tells us that the p_i, q^i or the P_i, Q^i may also serve as coordinates on U. In particular, there is then a transformation from the $p_i q^i$ to the $P_i Q^i$ coordinates. <u>This</u> is the transformation "generated" by the given function F. To show that it is indeed a canonical transformation we calculate the differential

$$d\left(\sum P_i Q^i - F\right) = \sum dP_i Q^i + \sum P_i dQ^i - \sum \frac{\partial F}{\partial q^i} dq^i - \sum \frac{\partial F}{\partial P_i} dP^i \, .$$

Inserting the values chosen for p_i and Q^i above gives

$$d\left(\sum P_i Q^i - F\right) = \sum P_i dQ^i - \sum p_i dq^i \, .$$

Taking the differential once more gives

$$\sum dP_i \wedge dQ^i = \sum dp_i \wedge dq^i \, .$$

so the indicated transformation is indeed canonical.

Similar transformations may be generated from functions G of other sets of variables, say $G(Q^1, \ldots, Q^n, p_1, \ldots, p_n)$. The formalism may be found in Goldstein; we turn now to a more conceptual explanation.

<u>Theorem</u>. Let M be a 2n-dimensional local manifold with coordinates $\{\begin{smallmatrix} P_i \\ q^i \end{smallmatrix}\}$. $M \xrightarrow{F} \mathbb{R}$ a smooth function and $\det \left\| \frac{\partial^2 F}{\partial P_i \partial q^i} \right\| \neq 0$ everywhere. Then $\omega = -\sum_{i,j} \frac{\partial^2 F}{\partial q^i \partial P_j} dq^i \wedge dP_j$ is a closed 2-form with $\underbrace{\omega \wedge \ldots \wedge \omega}_{n} \neq 0$. Thus M is symplectic.

<u>Proof</u>. We must first show that $d\omega = 0$. But

$$d\omega = -\sum \frac{\partial^3 F}{\partial q^k \partial q^i \partial P_j} \underbrace{dq^k \wedge dq^i \wedge dP_j}_{\text{alternating}} - \sum \frac{\partial^3 F}{\partial q^i \partial P_k \partial P_j} dP_k \wedge dq^i \wedge dP_j = 0$$

Next we must show the n-fold exterior product $\omega \wedge \ldots \wedge \omega \neq 0$. Write

$$\omega = \sum a_{ij} dq^i \wedge dP_j \ .$$

In the n-fold product many terms (iterated factors) drop out; there remain the following terms, for all permutations σ and τ of the symmetric group on n letters:

$$\sum \pm \left(\prod a_{\sigma_i \tau_j} \right) dq^i \wedge \ldots \wedge dq^n \wedge dP_j \wedge \ldots \wedge dP_n$$

so one gets the determinant $n!$ times and

$$\omega \wedge \ldots \wedge \omega = n! \det \| a_{ij} \| \neq 0 \ .$$

Thus (M, ω) is a symplectic manifold, as required.

Now the definitions

$$\left. \begin{aligned} p_i &= \frac{\partial F}{\partial q^i} \\ q^i &= q^i \end{aligned} \right\} \quad i = 1, \ldots, n$$

give 2n coordinates; since $\omega = \sum dp_i \wedge dq^i$ in these coordinates, they are canonical coordinates. Indeed,

$$\sum dp_i \wedge dq^i = \sum d\left(\frac{\partial F}{\partial q^i}\right) \wedge dq^i = \underbrace{\sum \frac{\partial^2 F}{\partial q^i \partial q^i} \wedge dq^i}_{0} + \sum \frac{\partial^2 F}{\partial q^i \partial p_j} \wedge dp^j \wedge dq^i .$$

This is exactly the 2-form ω defined above.

Similarly, the definitions

$$\left.\begin{aligned} P_i &= P_i \\ Q_i &= \frac{\partial F}{\partial P_i} \end{aligned}\right\} \quad i = 1, \ldots, n$$

give $2n$ coordinates which are canonical coordinates, since the 2-form may be written

$$\omega = \sum dP_i \wedge dQ^i .$$

Note the advantage of using differential forms. Specifying any closed non-degenerate 2-form makes U symplectic -- no matter what the coordinates. Here there are three possible coordinate systems

$$\begin{array}{ll} q^1, \ldots, q^n; p_1, \ldots, p_n & \omega = \sum dp_i \wedge dq^i \\ q^1, \ldots, q^n, P_1, \ldots, P_n & \omega = - \sum \dfrac{\partial^2 F}{\partial q^i \partial P_j} dq^i \wedge dP_j \\ Q^1, \ldots, Q^n, P_1, \ldots, P_n & \omega = - \sum dP_i \wedge dQ^i . \end{array}$$

The first and third systems are symplectic, but the transformation is generated by going through the intermediate non-symplectic coordinate system.

Such transformations may now be used to simplify a given Hamiltonian function H (a smooth function on U, given in terms of the coordinates q^i, p_i). If we relabel the generating function F as W, then $p_i = \partial W/\partial q^i$, and

$$H(q^1, \ldots, q^n, p_1, \ldots, p_n) = H(q^1, \ldots, q^n, \frac{\partial W}{\partial q^1}, \ldots, \frac{\partial W}{\partial q^n}) .$$

We propose to choose new canonical coordinates Q^i, P_i so that H will become simply P_1, the first P-coordinate. Writing α_1 for P_1, and considering α_1 as a constant (a parameter) this yields the equation

$$H(q^1, \ldots, q^n, \frac{\partial W}{\partial q^1}, \ldots, \frac{\partial W}{\partial q^n}) = \alpha_1 .$$

This is a first order partial differential equation for the function W. It is called the <u>Hamilton-Jacobi</u> partial differential equation. If we find a solution W depending on $n-1$ additional "constants of integration"

$$W = W(q^1, \ldots, q^n; \alpha_1, \ldots, \alpha_n)$$

satisfying the condition

$$\det \left\| \frac{\partial^2 W}{\partial q^i \partial \alpha_j} \right\| \neq 0 , \quad i, j = 2, \ldots, n$$

we can then prove that

$$\det \left\| \frac{\partial^2 W}{\partial q^i \partial \alpha_j} \right\| \neq 0 , \quad i, j = 1, \ldots, n .$$

Then W, with the α's replaced by P's, may serve as the generating function of a canonical transformation. In the new canonical coordinates the Hamiltonian function is just P_1. Therefore Hamilton's equations become

$$\dot{Q}_i = \delta_{i1} ,$$

$$\dot{P}_i = 0 , \quad i = 1, \ldots, n$$

and may be immediately integrated as

$$Q_i = t\delta_{i1} + \beta_i \ , \quad P_i = \alpha_i \ , \qquad i = 1, \ldots, n \ .$$

From the transformation equations

$$p_i = \frac{\partial W}{\partial q^i} \ , \quad Q^j = \frac{\partial W}{\partial \alpha_j} = t\delta_{j1} + \beta_j$$

one may then solve for the original coordinates p_i and q_i as functions of t.

§27. The Top

As an example, we consider the rigid motion of a heavy top, using the "Euler angles" θ, ψ, and $\emptyset$ as the parametrization of the rotation group in 3-space, with the axis of the top along the z-axis (see figure). Let ω_x, ω_y, and ω_z be the angular velocities about the x, y, and z axes. The kinetic energy of the top is then

$$T = \frac{1}{2}[I_1 \omega_x{}^2 + I_2 \omega_y{}^2 + I_3 \omega_z{}^2] \ ,$$

where I_1, I_2, I_3 are the diagonal terms of the moment of inertia tensor. (The axes are chosen so that the non-diagonal elements are zero.) Choose the top symmetric with respect to the x- and y-axes so $I_1 = I_2$. One may calculate from the definition of the Euler angles that

$$\omega_z = \dot{\psi} + \dot{\emptyset} \cos\theta \ ,$$

$$\omega_y = \dot{\emptyset} \sin\theta \sin\psi + \dot{\theta} \cos\psi \ ,$$

$$\omega_x = \dot{\emptyset} \sin\theta \cos\psi - \dot{\theta} \sin\psi \ .$$

Therefore

$$T = \frac{1}{2}[I_1(\dot\theta^2 + \dot\phi^2 \sin^2\theta) + I_3(\dot\psi + \dot\phi \cos\theta)^2] .$$

Since the momentum coordinates p_i are defined as $p_i = \frac{\partial T}{\partial \dot q^i}$, we have

$$p_1 = I_1\dot\theta ,$$

$$p_2 = I_3(\dot\psi + \dot\phi \cos\theta),$$

$$p_3 = I_3(\dot\psi \cos\theta) + (I_1 \sin^2\theta + I_3 \cos^2\phi)\dot\phi$$

from which we obtain

$$H = \frac{1}{2}\left[\frac{p_1^2}{I_1} + \frac{p_2^2}{I_3} + \frac{1}{I_1}\left(\frac{p_2 \cos q^1 - p_3}{\sin q^1}\right)^2\right] + Mg\ell \cos q^1$$

where ℓ is the distance along z from the origin to the center of mass (see figure).

Substituting $p_i = \frac{\partial W}{\partial q^i}$ gives the Hamilton-Jacobi partial differential equation for W. Since H also is independent of q^2 and q^3, we know that $p_2 = \alpha_2$ and $p_3 = \alpha_3$, α_2 and α_3 constants are solutions of Hamilton's equations. Hence we "separate the variables" in W, writing

$$W = W_1(q^1) + W_2(q^2) + W_3(q^3) ,$$

where we may take

$$W_2(q^2) = \alpha_2 q_2 + \text{constant},$$

$$W_3(q^3) = \alpha_3 q^3 + \text{constant},$$

since $\qquad \frac{\partial W}{\partial q^i} = p_i = \alpha_i \quad$ for $i = 2, 3$.

The remaining equation for $\frac{dW}{dq^1}$ is

$$\left(\frac{dW}{dq^1}\right)^2 = I_1\left[2\alpha_1 - 2Mg\ell \cos q^1 - \frac{\alpha_2^2}{I_3} - \frac{1}{I_1}\left(\frac{\alpha_2 \cos q^1 - \alpha_3}{\sin q^1}\right)^2\right].$$

Set $u = \cos q^1$; the equation becomes

$$(1 - u^2)\left(\frac{dW}{dq^1}\right)^2 = I_1\left[\left(2\alpha_1 - \frac{\alpha_2^2}{I_3} - 2Mg\ell u\right)(1 - u^2) - \frac{1}{I_1}(\alpha_2 u - \alpha_3)^2\right]$$

$$= F(u, \alpha_1) = (1 - u^2)(h - ku) - \left(\frac{\alpha_2}{I_1} u - \frac{\alpha_3}{I_1}\right)^2,$$

where h and k are defined accordingly so

$$W = \int \sqrt{\frac{F(u, \alpha_1)}{1 - u^2}}\, dq^1 .$$

But we want $Q^1 = \frac{\partial W}{\partial \alpha_1}$ so $\dot{Q}^1 = \frac{\partial H}{\partial p_1} = 1$ and $Q_1 = t + \beta$, so

$$Q^1 = \frac{\partial W}{\partial \alpha_1} = \int \frac{dq_1(1 - u^2)}{\sqrt{1 - u^2}\sqrt{F(u, \alpha_1)}} = \int \frac{du}{\sqrt{F(u, \alpha_1)}};$$

so

$$\beta_1 + t = \int \frac{du}{\sqrt{F(u, \alpha_1)}}, \quad \text{where } u = \cos q^1 = \cos \theta.$$

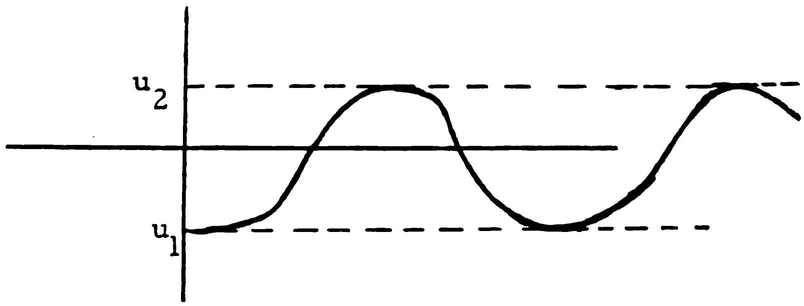
Since F is a cubic polynomial in u, this is an elliptic integral; the detailed explicit solution is given in Klein and Sommerfeld (4-volumes) on the gyroscope.

We consider now the qualitative properties of the solution. Since $F(u) = ku^3 + \ldots$, with $k > 0$, we have $F(\infty) = +\infty$, $F(-\infty) = -\infty$, and $F(u) = 0$ has three roots. The roots u_1 and u_2 between -1 and +1 (see graphs) are the only ones of physical interest since $-1 \leq u \leq 1$ is the only part which is physically possible. Now

$$\frac{dt}{du} = \frac{1}{\sqrt{F(u)}} \quad \text{implies} \quad \left(\frac{du}{dt}\right)^2 = F(u)$$

so the zeroes of $F(u)$ are where $\frac{du}{dt}$ is zero.

If we take the positive square root of F, we get a solution for u increasing from u_1 to u_2 as time goes from 0 to, say, A. This may be extended by reflection (negative root of F) to give a solution decreasing from u_2 to u_1. Continuing, as in the figure, we have a solution for all t

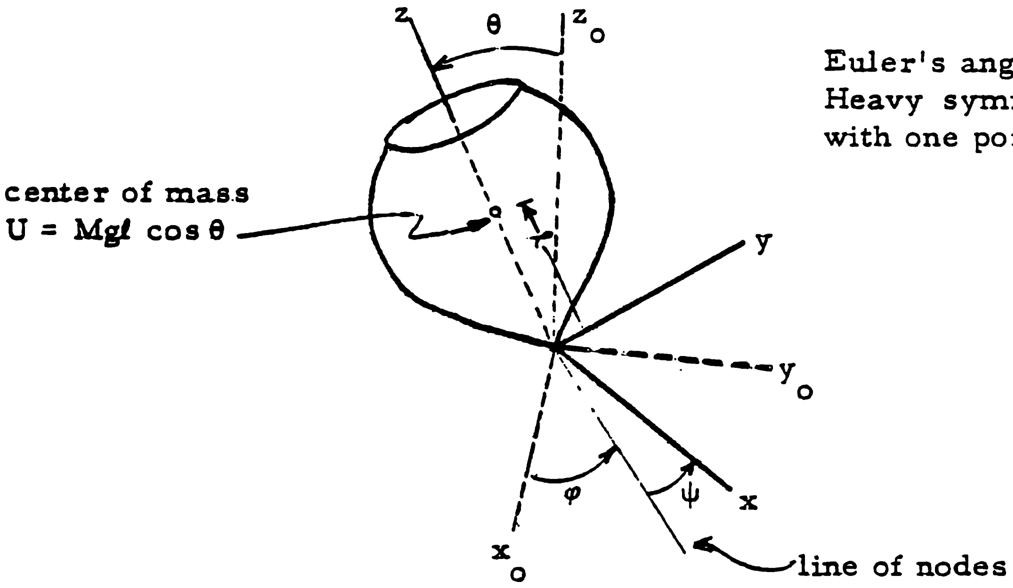


of the general form indicated, so that $u(t)$ is periodic with period 2A.

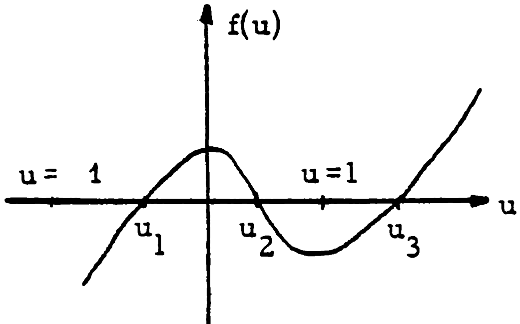
This solution may be pictured in terms of the angles (θ, φ) which give the point where the axis of the top pierces the unit sphere. For example, $p_2 = \alpha_2$ and $p_3 = \alpha_3$ allows us to solve the equations above for $\dot{\varphi}$ as

$$\dot{\varphi} = \frac{\alpha_3 - \alpha_2 \cos\theta}{I_1 \sin^2\theta} = \frac{\alpha_3 - \alpha_2 u}{I_1(1 - u^2)} .$$

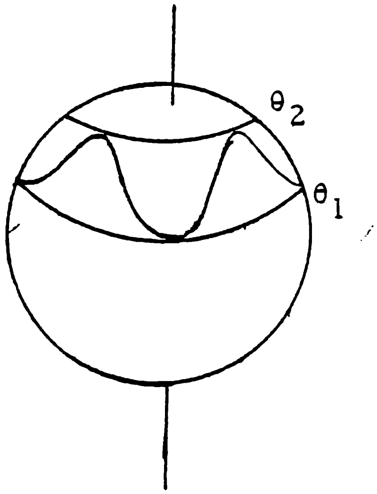
Here $\dot{\varphi}$ is the speed of <u>precession</u> of the axis of the top around the vertical, while $\dot{\theta}$ is the speed of rotation. For further qualitative discussion, see Goldstein, <u>Classical Mechanics</u>, pp. 164-175 or W.F. Osgood, <u>Mechanics</u>.



Euler's angles and the Heavy symmetric top with one point fixed



Location of turning angles



One possible Locus of the figure axis on the unit sphere

§ 28. Darboux's Theorem.

The typical example of a symplectic manifold is a cotangent bundle of configuration space. In such a bundle we have the usual position and momentum coordinates q^i and p_i, and the basic 2-form of the manifold is given in terms of these coordinates as

$$\omega = \sum dp^i \wedge dq^i \ .$$

The definition of a symplectic manifold (M, ω) was apparently much more general: Any 2n-dimensional manifold with a 2-form ω which is closed $(d\omega = 0)$ and non-degenerate ($\omega \wedge \ldots \wedge \omega$, to n factors, nowhere 0).

The added generality is illusory: Darboux's theorem asserts that at each point a of such a manifold there is always an open set U containing a and coordinates q^i and p_i -- good in this neighborhood U -- for which ω has the special form above.

One proof of this theorem is done by systematically exploiting the correspondence between vector fields and forms which is given by the basic form ω. This is the proof given in Abraham, pp. 92-94. Another proof (see Sternberg p. 137) depends on a more general theorem of Frobenius on the integration of differential systems. We refer to these texts for details.

January 1977

Corrections to Geometrical Mechanics, Part I, Saunders Mac Lane

(1-4 means 4 lines from foot of the page)

p. 10, line 2 (display) gij after $\sum$ should be g_{ij}

p. 10, line 7, 1-5, 1-1: All V^1 should be 1.e. v^1.

p. 11, line 6 (display): $V_1 \longrightarrow F_1$

p. 11, line 10 (display): (see below)

p. 18: Proof of theorem incomplete because addition of tangent vectors is not explicitly defined. Definition should be by map $\rho: T_a U \longrightarrow (T^a U)^*$ given in line 1 via

$$\rho(\tau_a c + \tau_a c') = \rho\tau_a c + \rho t_a c' \quad (\text{addition in } T^a U)^*$$

This requires proof that ρ is onto. Use local coordinates $q^1, \ldots, q^n$

$$\theta: T_a U \longrightarrow \mathbb{R}^n \text{ by } \theta\tau_a c = \left(\frac{\partial c}{\partial q^1}, \ldots, \frac{\partial c}{\partial q^n}\right)$$

$$\varphi: T^a U \longrightarrow (\mathbb{R}^n)^* \text{ by } \varphi df = \left(\frac{\partial f}{\partial q^1}, \ldots, \frac{\partial f}{\partial q^n}\right)$$

$$\varphi^*: \mathbb{R}^n \longrightarrow (T^a U)^*$$

and thus show by calculation that the diagram

$$\begin{tikzcd} T_a U \arrow[rr, "\rho"] \arrow[dr, "\theta"'] & & (T^a U)^* \\ & \mathbb{R}^n \arrow[ur, "\varphi^*"'] & \end{tikzcd}$$

with θ, φ^* both isomorphisms, commute ($\varphi^*\theta = \rho$)

p. 20, 1-1,-2 $\hat{f}$ should be $k: V \longrightarrow \mathbb{R}$

p. 11, line 10 (display): ∂p_1 should be ∂q^1 (1st equation)

GEOMETRICAL MECHANICS

Part II

Lectures by Saunders MacLane

Department of Mathematics
The University of Chicago

Spring Quarter 1968

These Lecture Notes were prepared with assistance by a grant from the Office of Naval Research

With a guest lecture by René Thom, a chapter by David Golber and supplementary mater ial by Raphael Zahler

Notes by: George Angwin
David Golber
Reinhard Schultz
Martha Smith
Stanley Weiss
Raphael Zahler

TABLE OF CONTENTS

Introduction to Part II

This sequel to Part I completes the notes of my two-quarter course on Geometrical Mechanics, except for the final section of the course which discussed Relativity Theory, the Schwarzschild metric, and the relativistic explanation of the advance in the perihelion of Mercury. (These lectures have not been reduced to written form.)

These notes have many of the imperfections of a first course on a new subject. Here the new subject is the use of modern geometrical ideas in the long-stagnant treatment of classical mechanics. The initiative of George Mackey has been vital for this subject, and the books by Ralph Abraham and Schlomo Sternberg are excellent guides. A few of the topics covered here are apparently not to be found in this form in the literature: The treatment of the Legendre transformation (§9 of Chapter I), the conceptual treatment of the generating functions for canonical transformations (§26 of Chapter III and §44 of Chapter VI), the description of manifolds by means of germs (Chapter IV, §30) and the geometric description of the characteristics of first order partial differential equations (Chapter VI, §46). This, with the material on contact transformations, may suggest how much of classical Mathematics stands in need of modern conceptual formulation.

I am much indebted to the students whose notes have improved and codified my lectures, and to René Thom for permission to include the material of his guest lectures.

The University of Chicago

Saunders MacLane
October 1968

29 Topological Spaces

To define manifolds, we first review the basic properties of topological spaces.

Definition. A topological space is a pair (X, t) where X is a set and t is a collection of subsets of X such that:

1° $\emptyset \in t$, $X \in t$;

2° $U \cap V \in t$ whenever $U \in t$ and $V \in t$;

3° If $\{U_\alpha\}_{\alpha \in A}$ is a collection of subsets of X such that $U_\alpha \in t$ for each α, then $\bigcup_{\alpha \in A} U_\alpha \in t$.

Here t is called the topology of X. The sets in t are called open sets. A subset F of X is called closed if $X - F \in t$, where $X - F = \{x \in X \mid x \notin F\}$ is the complement of F in X. We will often use just X to refer to the topological space (X, t) when it is clear what topology on X is intended.

Example. $\mathbb{R}^n$ together with the subsets which we have previously called open is a topological space.

If (X, t) and (X', t') are topological spaces, a function $f: X \longrightarrow X'$ is continuous if $f^{-1}(V) \in t$ whenever $V \in t'$, where the set $f^{-1}(V) = \{x \in X \mid f(x) \in V\}$ is the inverse image of V under f.

The function $f: X \longrightarrow X'$ is a homeomorphism if it is a bijection (one-to-one onto) and both f and f^{-1} are continuous.

A <u>neighborhood</u> of a point $x \in X$ is any open set in X containing x.

A function $f: X \longrightarrow X'$ is <u>continuous at x</u>, where $x \in X$, if for every neighborhood V of $f(x)$ in X', there exists a neighborhood U of x in X with

$$f(U) \subset V \qquad (\text{i.e., } U \subset f^{-1}(V))$$

It is easy to show that $f: X \longrightarrow X'$ is continuous if and only if f is continuous at every point $x \in X$.

<u>Examples:</u>

1^{o} If (X, t) is a topological space and S is any subset of X, let

$$t' = \{U \cap \rho \mid U \in t\}.$$

Then t' is a topology for S, called the <u>relative topology</u>.

2^{o} If (x, t) is a topological space and the function $X \xrightarrow{p} S$ maps X onto the set S, let

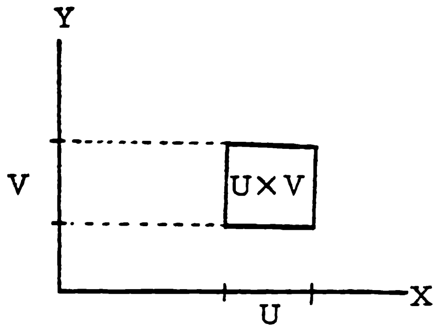
$$t' = \{V \subset S \mid p^{-1}(V) \in t\}.$$

Then t' is a topology, called the <u>quotient topology</u>, for S, and $(X, t) \xrightarrow{p} (S, t')$ is a continuous map.

3^{o} If (X, t) and (Y, t') are topological spaces, let $X \times Y$ denote the ordinary cartesian product of the sets X and Y. Let

$\bar{t}_o = \{U \times V \mid U \in t, \ V \in t'\}$,

$\bar{t}$ = all subsets of $X \times Y$ which are unions of sets in $\bar{t}_o$.



Then $(X \times Y, \bar{t})$ is a topological space, and $\bar{t}$ is called the <u>product topology</u>.

We have the two projection mappings onto X and Y:

$$p: X \times Y \longrightarrow X \qquad q: X \times Y \longrightarrow Y$$
$$(x, y) \longmapsto x \qquad (x, y) \longmapsto y$$

These are both continuous when $X \times Y$ is given the product topology.

<u>Definition:</u> If (X, t) is a topological space, a <u>basis</u> for (X, t) (or a "basis for the open sets of (X, t)") is a collection t_o of open subsets of X such that every member of t is a union of members of t_o.

Thus in Example 3^o, $\bar{t}_o$ is a basis for $(X \times Y, \bar{t})$.

A topological space is <u>separable</u> if it has a countable basis.

A <u>sub-basis</u> t_1 for (X, t) is a collection of subsets of X such that the set t_o of all finite intersections of members of t_1 forms a basis for X. Given any collection t_1 of subsets of a set X such that $X = \bigcup_{t_1} U$, there exists a unique topology t for X having t_1 as a sub-basis -- namely, t consists of all unions of finite intersections of members of t_1.

If $\{(X_\alpha, t_\alpha)\}_{\alpha \in A}$ is a family of topological spaces, let $X = \prod_\alpha X_\alpha$ be the cartesian product of the sets X_α and let $P_\alpha : X \longrightarrow X_\alpha$ be the projection onto the α^{th} coordinate space. Let

$$t_1 = \bigcup_{\alpha \in A} \{P_\alpha^{-1}(V) \mid V \in t_\alpha\} .$$

The topology t for X having t_1 as sub-basis is called the product topology. Then each $P_\alpha : (X, t) \longrightarrow (X_\alpha, t_\alpha)$ is continuous. If (X', t') is another space, and if for each $\alpha \in A$ we are given a continuous map $f_\alpha : (X', t') \longrightarrow (X_\alpha, f_\alpha)$, there exists a unique function $f : X' \longrightarrow X$ such that $P_\alpha f = f_\alpha$ for each α, since the set X is the set product of the X_α. Then if $V \in t_1$ -- say $V = P_\alpha^{-1}(U_\alpha)$ where U_α is open in X_α, it follows that $f^{-1}V = f^{-1}P_\alpha^{-1}U_\alpha = (P_\alpha f)^{-1}U_\alpha = f_\alpha^{-1}U_\alpha$ is open in X'. Since t_1 is a sub-basis for (X, t), it follows that f is continuous (This means that (X, t) is the product of the (X_α, t_α) in the category of all topological spaces.)

Suppose the X_α as above are disjoint (if not, take disjoint homeomorphic copies). Then we can topologize their disjoint union $Y = \coprod_\alpha X_\alpha$ as follows:

$$U \subset Y \text{ is open in } Y \Longleftrightarrow U \cap X_\alpha \text{ is open in } X_\alpha \text{ for each } \alpha.$$

In a fashion similar to that above, if (X', t') is another topological space and $g_\alpha : X_\alpha \longrightarrow X'$ is a continuous map for each α, then there exists a unique continuous map $g : X \longrightarrow X'$ such that $g_\alpha = g q_\alpha$, where $q_\alpha : X_\alpha \longrightarrow X$ is the injection of X_α into the disjoint union. This means that Y is the "coproduct" of the (X_α, t_α) in the category of all topological spaces.

Suppose we are given a set X and subsets X_α each with a topology t_α such that

1° $X = \bigcup_\alpha X_\alpha$

2° $Y_{\alpha\beta} = X_\alpha \cap X_\beta$ is open in both X_α and X_β, and the relative topologies on $Y_{\alpha\beta}$ induced from (X_α, t_α) and (X_β, t_β) coincide.

Then X has a topology

$$t = \{U \subset X \mid U \cap X_\alpha \in t_\alpha \text{ for all } \alpha\}.$$

This situation can be expressed by the statement that X is the coequalizer, in the category of topological spaces and continuous maps, of the maps

$$\coprod_{\alpha,\beta} Y_{\alpha\beta} \rightrightarrows \coprod_\alpha X_\alpha ,$$

where one map injects $Y_{\alpha\beta}$ into X_α, the other into X_β.

An <u>open covering</u> for a topological space (X, t) is a collection of open sets of X whose union is X. If $\{U_\alpha\}$ is an open covering for (X, t), it is easily checked that a function $f: (X, t) \longrightarrow (Y, t')$ is continuous if and only if $f|U_\alpha: U_\alpha \longrightarrow Y$ is continuous in the relative topology of U_α for each α.

A topological space X is <u>Hausdorff</u> if for every pair of points $x, y \in X$ there exist open sets $U, V \subset X$ with $x \in U$, $y \in V$ such that $U \cap V = \emptyset$.

30. Manifolds.

Let X be a topological space, $x \in X$. Every function $f: U \longrightarrow \mathbb{R}$ such that $x \in U$ and U is an open subset of X determines the <u>germ</u> f_x of f at x, where $f_x = g_x$ if $g: V \longrightarrow \mathbb{R}$ and there exists $W \subset U \cap V$ such that $x \in W$, W is open in X and $f|W = g|W$.

Let C_x denote the set of germs of all continuous functions to $\mathbb{R}$ defined on some neighborhood of x. C_x is an algebra.

<u>Definition.</u> A <u>loaded space</u> is a triple (X, t, G) where (X, t) is a topological space and G assigns to each point $x \in X$ a set G_x of germs at x (germs of the "good" functions).

Unless otherwise specified, we will assume that $G_x \subset C_x$. Often we will require that G_x be an algebra.

<u>Examples:</u>

1° $X = U_o$ open in $\mathbb{R}^n$ (e.g., $X = \mathbb{R}$) and G = germs of all C^∞ functions at x. Call this loaded space (U_o, C^∞).

2° If (X, G) is a loaded space and V is open in X, then $(V, G|V)$ is a loaded space.

3° Let (X, t) be a topological space and $\mathcal{F}$ any set of continuous functions $f: X \longrightarrow \mathbb{R}$. Set $\mathcal{G}_x = \{f_x | f \in \mathcal{F}\}$. Then $(X, t, \mathcal{G})$ is a loaded space.

If $(X, \mathcal{G})$ is a loaded space and U is open in X, define

$$\mathcal{G}(U) = \{f | f: U \longrightarrow \mathbb{R} \text{ continuous and } f_x \in \mathcal{G}_x, \text{ for all } x \in U\}.$$

$\mathcal{G}(U)$ has the <u>sheaf property</u>: if $U = \bigcup_\alpha V_\alpha$ where the V_α are open, then $f \in \mathcal{G}(U)$ if and only if for each α, $f|V_\alpha \in \mathcal{G}(V_\alpha)$.

If $(X, \mathcal{G})$ and $(Y, \mathcal{H})$ are loaded spaces, a <u>loaded map</u> $(X, \mathcal{G}) \xrightarrow{\varphi} (Y, \mathcal{H})$ is a map $X \xrightarrow{\varphi} Y$ such that

1° φ is continuous,

2° $x \in X$, $h_{\varphi(x)} \in \mathcal{H}_{\varphi(x)}$ implies $(h\varphi)_x \in \mathcal{G}_x$.

Notice the similarity of this definition to that of a continuous map.

The following facts follow easily from the last definition:

1° The composite of loaded maps is loaded.

2° V open in Y implies $\varphi^* \mathcal{H}(V) \subset \mathcal{G}(\varphi V)$.

3° φ is loaded if and only if at each $x \in X$, φ is continuous and carries "good" germs at $\varphi(x)$ to "good" germs at x.

A <u>loaded isomorphism</u> is a loaded map $(X, \mathcal{G}) \xrightarrow{\varphi} (Y, \mathcal{H})$ such that

1° $X \xrightarrow{\varphi} Y$ is a topological isomorphism (i.e., a <u>homeomorphism</u>)

2° for each $x \in X$, the correspondence $\mathcal{H}_{\varphi(x)} \longrightarrow \mathcal{G}_x$ induced by φ is one-to-one and onto.

<u>Definition.</u> A C^∞ n-chart on $(X, \mathcal{G})$ consists of

1° an open set U of X, called the <u>domain</u> of the chart,

2° a loaded isomorphism $(U, \mathcal{G}|U) \simeq (U_o, C^\infty)$, where U_o is open in $\mathbb{R}^n$.

A C^∞ <u>n-manifold</u> is a loaded space $(M, \mathcal{G})$ such that M is Hausdorff and the domains of all C^∞ n-charts on $(M, \mathcal{G})$ cover M. We will usually also require that M be separable.

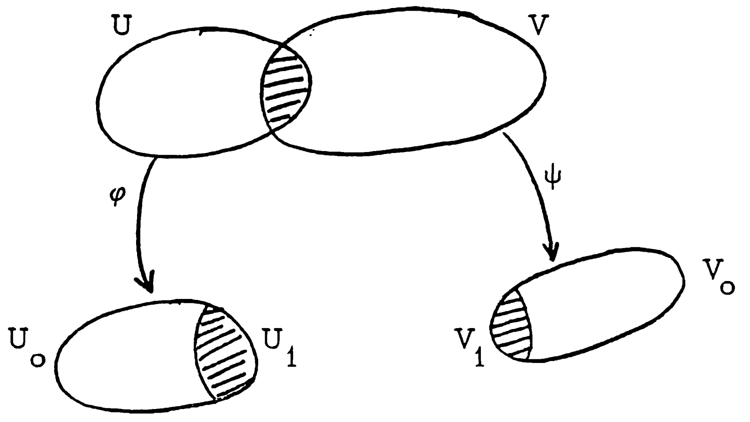
<u>Example:</u> Any open set in $\mathbb{R}^n$ is a C^∞ n-manifold.

An <u>atlas</u> of a C^∞ n-manifold $(M, \mathcal{G})$ is a set of n-charts whose domains cover M. The same manifold can have many atlases; the only "invariant" one is the maximal atlas (<u>all</u> charts).

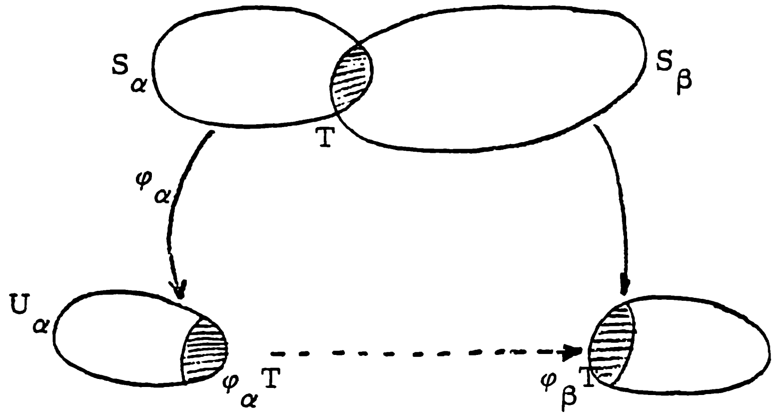
Uf $U \overset{\varphi}{\simeq} U_o$ and $V \overset{\psi}{\simeq} V_o$ are charts of the C^∞ manifold $(M, \mathcal{G})$, then the induced map

$$\begin{tikzcd} U_1 \arrow[r, phantom, "\underset{\varphi}{\cong}"] \arrow[rr, bend left, "\theta"] & U \cap V \arrow[r, phantom, "\underset{\psi^{-1}}{\simeq}"] & V_1 \end{tikzcd}$$

is C^∞, for if $x^1, \ldots, x^n$ are smooth coordinates on V_1, then it follows that each $x^i \circ \theta$ is smooth.



Suppose S is a set. A <u>chart</u> on S is a one-to-one function $S_\alpha \xrightarrow{\varphi_\alpha} U_\alpha$ where $S_\alpha \subset S$, U_α is open in $\mathbb{R}^n$, and φ_α maps S_α onto U_α. Two charts φ_α and φ_β are <u>compatible</u> if, for $T = S_\alpha \cap S_\beta$, $\varphi_\alpha T$ is open in U_α, $\varphi_\beta T$ is open in U_β, and $(\varphi_\beta | T) \circ (\varphi_\alpha^{-1} | \varphi_\alpha T)$ is a C^∞ map from $\varphi_\alpha T$ onto $\varphi_\beta T$.



If $\{(S_\alpha, \varphi_\alpha)\}$ is a family of pairwise compatible charts on S such that the S_α cover S, then the φ_α collectively determine a topology on S. If this topology is Hausdorff, then S becomes a manifold. More generally, we have the following theorem which is often used to construct a manifold from overlapping pieces M_α (especially with each M_α an open set in $\mathbb{R}^n$):

<u>Theorem.</u> If X is a set and $X = \bigcup_\alpha M_\alpha$ where each $M_\alpha = (M_\alpha, t_\alpha, \mathcal{G}^{(\alpha)})$ is a C^∞ n-manifold such that for each α and β

1° $T_{\alpha\beta} = M_\alpha \cap M_\beta$ is open in both M_α and M_β;

2° if $x \in T_{\alpha\beta}$ and g is a real-valued function defined near x, then

$$g_x \in \mathcal{G}^{(\alpha)} \Longleftrightarrow g_x \in \mathcal{G}^{(\beta)} ;$$

3° if $U \subset T_{\alpha\beta}$, then $U \in t_\alpha \Longleftrightarrow U \in t_\beta$;

then

a) X has a topology, namely W is open in X if and only if for every α, $W \cap M_\alpha$ is open in M_α. (We've seen this part before. In particular, each M_α is open in X.)

b) X is a loaded space, where for $x \in M$, $\mathcal{G}_x = \mathcal{G}_x^{(\alpha)}$ with $x \in M_\alpha$.

c) If X is Hausdorff with the topology in (a), then X is a C^∞ n-manifold. If $U \xrightarrow{\varphi} U_o$ is a chart in M_α for some α, then it is also a chart in X.

If $U \xrightarrow{\varphi} U_o$ is a chart on the n-manifold M and $x^1, \ldots, x^n$ are coordinates on $U_o \subset \mathbb{R}^n$, then $q^i = x^i \circ \varphi$ are called <u>coordinates on U.</u>

We will say that a function $(X, \mathcal{G}) \xrightarrow{\varphi} (Y, \mathcal{H})$ between two loaded spaces is <u>loaded at</u> $x \in X$, or <u>smooth at</u> x if it is continuous at x and satisfies condition (2) of the definition of loaded map for x.

<u>Lemma.</u> If $(X, \mathcal{G}) \xrightarrow{h} (Y, \mathcal{H})$ is any function between C^∞ manifolds, $x \in X$, and $q^1, \ldots, q^n$ are coordinates on the domain U of a chart $U \xrightarrow{\varphi} U_o \subset \mathbb{R}^n$ such that $h(x) \in U$, then h is loaded at x if and only if each $(q^i \circ h)_x \in \mathcal{G}_x$.

<u>Proof.</u> h is loaded at x if and only if φh is loaded at x. So if h is loaded at x, then for every C^∞ function $U_o \xrightarrow{k} \mathbb{R}$, $(k\varphi h)_x \in \mathcal{G}_x$. In particular, $(q_i h)_x = (x^i \varphi h)_x \in \mathcal{G}_x$.

Conversely, if each $(q^i h)_x \in \mathcal{G}_x$, let $V \xrightarrow{\psi} V_o$ be a chart of X such that $x \in V$. Then $q^i h \psi^{-1}$ must be C^∞, so for any k as above, $k\varphi h\psi^{-1} = k(q^1 h\psi^{-1}, \ldots, q^n h\psi^{-1})$ is the composite of C^∞ functions and hence C^∞. Then since ψ is loaded, $k\varphi h \in \mathcal{G}_x$. This holds for all C^∞ functions k, so φh and hence h are loaded.

A smooth map (a C^∞-map) $h: M \to N$ between C^∞ manifolds is now defined to be a continuous map which is loaded at each point $x \in M$. In other words, a function h is smooth if it is continuous and if it carries good germs at each point h(x) of N back into good germs at x. It follows that the composite of smooth maps is smooth.

Example. The sphere S^n is an n-manifold. The usual manifold structure is a generalization to higher dimensions of the charts obtained by stereographic projection of S^2. However, for $n \geq 7$ there exist other manifold structures on S^n, giving the so-called exotic spheres. In other words, there exist two manifold structures $\mathcal{G}$ and $\mathcal{H}$ on S^n such that the identity function $(S^n, \mathcal{G}) \to (S^n, \mathcal{H})$ is not smooth.

We have described a manifold as a topological space with a function $\mathcal{G}$ assigning good germs. This function may be replaced by the function $U \mapsto \mathcal{G}(U)$ described above and called a "sheaf" (more exactly, the sheaf of germs of C^∞ functions. This sheaf-theoretic definition of a manifold is equivalent to a different definition by atlases (A manifold is a topological space equipped with a suitable "maximal" atlas).

<u>The product:</u> If M and N are C^∞ manifolds, let $M \times N$ be their product as topological spaces. If $\{U_\alpha \overset{\varphi_\alpha}{\simeq} U_\alpha^o \subseteq \mathbb{R}^m\}$ and $\{V_\beta \overset{\psi_\beta}{\simeq} V_\beta^o \subseteq \mathbb{R}^n\}$ are collections of charts covering M and N respectively, then the $U_\alpha \times U_\beta$ cover M. Furthermore, $U_\alpha \times V_\beta \overset{\varphi_\alpha \times \psi_\beta}{\simeq} U_\alpha^o \times U_\beta^o$, which is open in $\mathbb{R}^m \times \mathbb{R}^n$, so each $U_\alpha \times V_\beta$ is a manifold -- if $U_\alpha \times V_\beta \supset W \xrightarrow{g} \mathbb{R}$ and $x \in W$, then $g_x \in \mathcal{G}_x$ if and only if $g(\varphi_\alpha \times \psi_\beta)^{-1}$ is C^∞ at $(\varphi_\alpha \times \psi_\beta)_x$. The manifold structures coincide on the overlaps, so by the theorem $M \times N$ is a manifold -- indeed, the topology given in the theorem is the product topology. The projections

$$\begin{array}{ccccc} M & \xleftarrow{p} & M \times N & \xrightarrow{p'} & N \\ m & \leftarrow\!\!\sim\!\sim & (m, n) & \sim\!\sim\!\!\rightarrow & n \end{array}$$

are smooth maps. If K is a manifold and f, f' are smooth maps

$$\begin{tikzcd} & K \arrow[dl, "f"'] \arrow[dr, "f'"] & \\ M & & N \end{tikzcd},$$

they are in particular continuous, so since $M \times N$ is a topological product, there exists a unique continuous map $K \xrightarrow{h} M \times N$ such that $f = ph$ and $f' = p'h$. By selecting suitable charts and coordinates we can easily show that h is smooth. Hence $M \times N$ is the "categorial" product.

31. The Tangent Bundle.

The tangent bundles defined in Part I, § 3 for open sets of $\mathbb{R}^n$, can now be defined for manifolds.

If M is a C^∞ n-manifold, the tangent bundle $\begin{array}{c} T.M \\ \downarrow \pi. \\ M \end{array}$ consists of

all points $\begin{array}{c} (a, \tau_a c) \\ \downarrow \pi. \\ a \end{array}$ where $a \in M$ and $\tau_a c$ is a tangent vector at a.

(More precisely, $\tau_a c$ is a tangent vector at φ_a where $a \in U \xrightarrow{\varphi} U_o \subseteq \mathbb{R}^n$ is some fixed chart of M.)

For each chart $U \xrightarrow{\varphi} U^o \subseteq \mathbb{R}^n$ of M,

$$U^o \times \mathbb{R}^n \simeq T.U \subset T.M$$

via coordinates $q^1, \ldots, q^n; \dfrac{\partial}{\partial q^1}, \ldots, \dfrac{\partial}{\partial q^n}$ for $T.U$. This defines a chart on $T.M$. Any two such charts are compatible since the Jacobian of a change of coordinates in M is non-zero. Thus $T.M$ is a manifold. (Apply the theorem of the previous paragraph constructing a manifold from the overlapping pieces $T.U$.)

A pre-bundle is $\begin{array}{c} B \\ \downarrow \pi \\ M \end{array}$ where B and M are C^∞ manifolds, π is

smooth, and each $\pi^{-1}(m)$ is a vector space.

Example. If U_o is open in $\mathbb{R}^n$, then
$$\begin{array}{ccc} U_o \times \mathbb{R}^\ell & & (u_o, v) \\ \big\downarrow \pi_o & & \big\downarrow \\ U_o & & u_o \end{array}$$
with the standard vector space structure on each $\{u_o\} \times \mathbb{R}^\ell \cong \mathbb{R}^\ell$ is a pre-bundle called a special pre-bundle.

A chart of a pre-bundle $\begin{array}{c} B \\ \big\downarrow \pi \\ M \end{array}$ consists of an open set $U \quad M$ and a pre-bundle isomorphism of $\begin{array}{c} \pi^{-1}(U) \\ \big\downarrow \pi | \pi^{-1}(U) \\ U \end{array}$ with a special pre-bundle.

A vector bundle is a pre-bundle covered by charts, (i. e., the U's of all possible charts cover M and the $\pi^{-1}(U)$'s cover B.)

Note that it is not necessary to require that addition be smooth in a vector bundle, since a vector bundle is locally like a special bundle, in which addition is automatically smooth.

$T.M$ is a vector bundle with the charts described above. Similarly, we define the cotangent bundle $T^{\cdot}M$ and the bundles constructed from the various mixed tensors. We can consider each of these as a functor which takes smooth functions between manifolds into smooth functions between vector bundles.

32. Bump Functions and the Extension of Germs.

We could have defined manifold with "analytic", "piecewise linear" or "continuous" replacing C^∞ throughout. If we did this for "continuous", we would get a topological manifold -- a topological space which locally looks like Euclidean space.

Recall that in Part I we defined a local manifold as a set M with a set $\mathcal{F}$ of "smooth" functions such that $M \simeq U_o \subsetneq \mathbb{R}^n$ and $\mathcal{F}$ corresponds one-to-one to the C^∞ functions on U_o. We could have defined a manifold in a similar manner: as a set M with a set $\mathcal{F}$ of "smooth" functions which would determine both the topology of M and the "good" germs on M as the germs of the functions in $\mathcal{F}$. We used germs of functions defined only on open sets of M, but the following theorem shows that it suffices to consider only germs of functions defined "in the large" (on all of M).

Theorem. If M is a C^∞ manifold and g_x is a smooth germ at $x \in M$, then there exists a C^∞ function $f: M \longrightarrow \mathbb{R}$ such that $f_x = g_x$.

Before proving the theorem we will need some preliminary results.

Definition. A topological space X is compact if every open covering of X has a finite subcover.

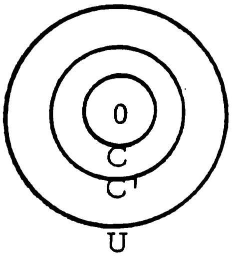
Theorem. A compact subspace of a Hausdorff topological space is closed.

(This is a standard result and can be found in any text on point-set topology.)

Lemma (Existence of "bump" functions): If M is a manifold, $U \subset M$ is open, and $x \in U$, then there exist compact subsets C and C' of U and a C^∞ function $h: U \to \mathbb{R}$ such that $x \in C \subset C' \subset U$ and $h = 1$ on C, $h = 0$ outside C'.

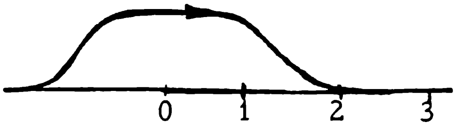
Proof of Lemma. It suffices to consider a chart containing x. We may take a chart in Euclidean space containing a disc about 0 of radius 3. (If not, blow up the chart by a large enough factor.)

Take C = closed disc about 0 of radius 1, C' = closed disc about 0 of radius 2.



If $n = 1$, define $h_o = 0$ for $x > 2$, $h_o = 1$ for $0 \leq x \leq 1$, and to be any suitable C^∞ function which is 1 at 1 and 0 at 2, for $1 \leq x \leq 2$. (Problem: give an explicit formula.) Define $h_o(x) = h_o(-x)$ for $x < 0$.

For general n, let



$$h(x) = h(|x|^2),$$

where $|x|^2 = (\Sigma x_i^2)$ for $x = (x_1, \ldots, x_n)$.

Proof of Theorem. Let U be the domain of a chart containing x. Construct a smooth "bump" function b on U by the above lemma. Let

$$f = \begin{cases} gb & \text{on } U \\ 0 & \text{outside of } U \end{cases} .$$

Then f is continuous -- for $V \subset \mathbb{R}$ open, $f^{-1}(V) = (gb)^{-1}(V)$ is open if $0 \notin V$, and if $0 \in V$, $f^{-1}(V) = (gb)^{-1}(V) \cup (M - C')$ is open, since C' is closed by the above remark. Moreover f is smooth, since g and b are both smooth.

The theorem shows that we could have defined a manifold in terms of functions defined on the entire manifold. However, such a definition would make it more difficult to show that certain manifolds (such as tangent bundles) can be constructed by piecing together other manifolds.

Note that the Theorem would also hold for topological manifolds, but does not hold for analytic manifolds, because the bump function cannot be made analytic.

33. Volumes on Symplectic and Contact Manifolds.

Let us now review the standard set-up we use for discussing mechanics on a general differentiable manifold. Configureation space, C, is an n-dimensional manifold whose points correspond, roughly speaking, to "configurations" or "positions" of the mechanical system. Phase space is the cotangent bundle, T^*C, with the canonical 2-form ω; in local coordinates, $\omega = \Sigma\, dp_i \wedge dq^i$. We define event space to be the topological product $C \times I$, where I is an interval of time, that is, an interval on the real line with t as coordinate. A point (c, t) of event space represents the state c at the time t. Finally, state space is defined to be the product manifold $T^*C \times I$, endowed with the canonical one-form given in local coordinates as $\theta = -\Sigma\, p_i \wedge dq^i + dt$.

Now recall from Part I, § 22, that a symplectic manifold (M, ω) is a manifold M of even dimension $2n$ together with a closed 2-form ω such that $\omega \wedge \ldots \wedge \omega$ (n factors) is nowhere zero. Each phase space T^*C is a symplectic manifold. Similarly, a contact manifold (M, θ) is a manifold of dimension $2n + 1$, where n is an integer, with a one-form θ such that the $(2n+1)$-form $\theta \wedge d\theta \wedge \ldots \wedge d\theta$ (n factors $d\theta$) is non-zero everywhere. State space is an example of a contact manifold. (Note: These contact manifolds are called "exact contact manifolds" in Abraham, loc. cit.)

Both a symplectic manifold and a contact manifold have a non-zero form of highest dimension; that is an "element of volume". For example, in euclidean three-space an element of volume is usually written $dx\,dy\,dz = dx \wedge dy \wedge dz$ with respect to rectangular coordinates; $r^2 \sin\theta\, dr\, d\theta\, d\varphi$ with respect to spherical coordinates, and so on. In general a volume element on an n-dimensional vector space W is a non-zero element $b \in \Lambda_n(W)$. Since the n-th exterior power $\Lambda_n(W)$ is a one-dimensional vector space, any two volume elements b and b' on W are proportional: $b' = rb$, where r is a non-zero number. Now we often speak of "right-handed" and "left-handed" coordinate systems on

Euclidean three-space; similarly, there may be two types of volume elements. To see this, say that b and b' are <u>equivalent</u> if the proportionality constant r is positive. This divides the volume elements up into equivalence classes: $dx \wedge dy \wedge dz = -dx \wedge dz \wedge dy = dz \wedge dx \wedge dy$, so the elements $dx \wedge dy \wedge dz$ and $dz \wedge dx \wedge dy$ are equivalent.

A <u>volume</u> on an n-dimensional manifold M is a form Ω on $\Omega_n(M)$ which is non-zero everywhere on M. Any two volumes Ω and Ω' are related by the formula $\Omega = f\Omega'$, where f is a smooth non-zero real-valued function on M. If f is positive everywhere, call Ω and Ω' equivalent. Then an orientation of M is defined to be an equivalence class of volumes. Since M may not have a volume in the first place, M may not be <u>orientable</u>; however, we have seen that symplectic manifolds and exact contact manifolds are orientable. A möbius strip is an example of a non-orientable manifold.

Let Ω be a volume on the n-dimensional manifold M. If X is a vector field on M, the Lie derivative $L_X\Omega$ is another n-form. But any two n-forms at a point are proportional. Thus there is a smooth function f such that $L_X(\Omega) = f\Omega$. We write $f = \text{div } X$; notice that div X depends on the choice of a volume element. Does this agree with the usual notion of the divergence of a vector field? In the situation $M = \mathbb{R}^n$, with coordinates $x^1, \ldots, x^n$, we can write $\Omega = dx^1 \wedge \ldots \wedge dx^n$, and $X = \Sigma X^i \dfrac{\partial}{\partial x^i}$. Then

$$L_X\Omega = \sum_i dx^1 \wedge \ldots \wedge L_X dx^i \wedge \ldots \wedge dx^n .$$

But

$$L_X dx^i = \sum_j (X^j \frac{\partial}{\partial x^j})\, dx^i = \frac{\partial X^i}{\partial x^i}\, dx^i .$$

Hence

$$L_X(\Omega) = [\sum_{i=1}^{n} (\frac{\partial X^i}{\partial x^i})]\Omega$$

so $\operatorname{div} X = \Sigma \frac{\partial X^i}{\partial x^i}$, as expected.

Moreover, our generalized definition of divergence proves a suitable extension of the idea of divergence as the infinitesimal change of volume at a point. For (Part I, 24) the derivative $L_X\Omega$ describes the rate of change of the volume along the trajectories of X.

34. Poisson Brackets.

Let (M, ω) be a symplectic manifold, with symplectic coordinates $\{p_i, q^i\}$; if f and g are two real-valued functions on M, the poisson bracket of f and g with respect to the coordinates p_i, q^i is the smooth function defined by

$$\{f, g\} = \sum_{i=1}^{n} (\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i}) .$$

It can be shown that the value of the poisson bracket $\{f, g\}$ of f and g does not depend on the choice of coordinates; however, we seek an invariant description of this function, since it will help us find a formulation of the laws of mechanics leading naturally to quantum mechanics.

We now develop the general algebraic machinery needed for this invariant definition of the Poisson brackets. We previously defined the exterior derivative d, which takes k-forms into $(k+1)$-forms for every non-negative integer k. Given a vector field X, there is likewise an operation i_X mapping $(k+1)$-forms ω to k-forms $i_X\omega$:

$$i_X\omega(X_1,\ldots,X_k) = (k+1)\,\omega(X,X_1,\ldots,X_k)$$
$$= \sum_i (-1)^i\,\omega(X_1,\ldots,X_{i-1},X,X_i,\ldots,X_k).$$

Finally, the Lie derivative L_X takes k-forms to k-forms. We can now state three identities:

(1) $i_X d + d\,i_X = L_X$ ("homotopy identity");

(2) $L_X(\eta(X_1,\ldots,X_k)) = (L_X\eta)(X_1,\ldots,X_k) + \sum_{i=1}^{n} \eta(X_1,\ldots,L_X X_i,\ldots,X_k)$,

where η is a k-form, so that $\eta(X_1,\ldots,X_k)$ is a function on M;

(3) $2\,d\omega(X,Y) = L_X\omega(Y) - L_Y\omega(X) - \omega([X,Y])$,

where ω is a one-form.

<u>Proof</u>. For (1), we first notice that i_X is an antiderivation: that is if α is a k-form,

$$i_X(\alpha\wedge\beta) = (i_X\alpha)\beta + (-1)^k\alpha\wedge(i_X\beta).$$

This is an easy computation from the definition of i_X.

Now we prove $i_X d\alpha + d i_X\alpha = L_X\alpha$ by induction on k: for a function f (a 0-form), $i_X f$ is defined to be zero, and we have $i_X df = \langle df, X\rangle = L_X f$. Assuming the result true for k-forms, write a general $(k+1)$-form, α, as $\Sigma\, df_i\wedge\omega_i$; by linearity it will suffice to prove the result for each summand.

But

$$L_X(df\wedge\omega) = (L_X df)\wedge\omega + df\wedge(L_X\omega),$$

while

$$\begin{aligned} &i_X d(df\wedge\omega) + di_X(df\wedge\omega) \\ &\quad = -i_X(df\wedge d\omega) + d(i_X df\wedge\omega - df\wedge i_X\omega) \\ &\quad = -(i_X df)\wedge d\omega + df\wedge(i_X d\omega) \\ &\qquad\quad + (di_X df)\wedge\omega + (i_X df)\wedge d\omega + df\wedge(di_X\omega). \end{aligned}$$

Here the first and fourth terms cancel, giving

$$\begin{aligned} &df\wedge(i_X d\omega) + (di_X df)\wedge\omega + df\wedge(di_X\omega) \\ &\quad = df\wedge L_X\omega + (di_X df\wedge\omega) \quad \text{(by inductive assumption)} \\ &\quad = df\wedge L_X\omega + (d(L_X f)\wedge\omega) = df\wedge L_X\omega + (L_X df)\wedge\omega. \end{aligned}$$

This proves (1).

For part (2), recall that L_X commutes with contractions, while evaluation of η at $(X_1,\ldots,X_k)$ is nothing but the contraction $\delta(\eta\otimes X_1\otimes\ldots\otimes X_k)$. With this observation (2) follows from the fact that L_X is a derivation. To derive (3), we verify the formula for $\omega = g\,dq$, since we can then extend to a general one-form ω by linearity. But since $d\omega = dg\wedge df$, it is easy to show that both sides of (3) reduce to $(L_X g)(L_Y f) - (L_X f)(L_Y g)$. A similar argument shows that for ω any k-form,

$$(4)\quad (k+1)d\omega(X_o,\ldots,X_k) = \sum_{i=0}^{k}(-1)^i L_{X_i}(\omega(X_o,\ldots,\hat{X}_i,\ldots,X_k)) + \sum_{0\le i<j\le k}(-1)^{i+j}\omega([X_i,X_j],X_o,\ldots,\hat{X}_i,\ldots,\hat{X}_j,\ldots,X_k).$$

(Here the $\wedge$ over X_i means "omit X_i".)

Now let (M, ω) be a symplectic manifold. ω induces linear mappings (§21) $X \longrightarrow X^\flat$ and $\alpha \longrightarrow \alpha^\#$ taking vector fields into co-vector fields (= 1-forms), and vice versa. We may now define the poisson bracket of two one-forms, α and β, by

<u>Definition.</u> $\{\alpha, \beta\} = -[\alpha^\#, \beta^\#]^\flat$.

In other words, we turn the forms temporarily into vector fields, take the Lie bracket, and return to the space of forms. The minus sign is chosen for convenience in proving such formulas as

<u>Proposition.</u> $\{\alpha, \beta\} = -L_{\alpha^\#}\beta + L_{\beta^\#}\alpha + d(i_{\alpha^\#}L_{\beta^\#}\omega)$.

<u>Proof.</u> ω is closed, hence by (4) above

$$0 = 3d\omega(X, Y, Z) = L_X(\omega(Y, Z)) + L_Y(\omega(Z, X)) + L_Z(\omega(X, Y)) - \omega([X, Y], Z) + \omega([X, Z], Y) - \omega([Y, Z], X).$$

If we let $X = \alpha^\#$, $Y = \beta^\#$, and recall that, by the definition of $\#$, we have $\omega(\alpha^\#, Y) = \frac{1}{2}\alpha(Y)$, the above equation becomes

$$0 = L_{\alpha^\#}(\tfrac{1}{2}\beta(Z)) - L_{\beta^\#}(\tfrac{1}{2}\alpha(Z)) + L_Z(\omega(\alpha^\#, \beta^\#)) + \omega(\{\alpha, \beta\}^\#, Z) + \omega([\alpha^\#, Z], \beta^\#) - \omega([\beta^\#, Z], \alpha^\#).$$

Therefore

$$\begin{aligned} -L_{\alpha^\#}(\tfrac{1}{2}\beta(Z)) + L_{\beta^\#}(\tfrac{1}{2}\alpha(Z)) - L_Z(\omega(\alpha^\#, \beta^\#)) &= \tfrac{1}{2}(\{\alpha, \beta\})Z - \tfrac{1}{2}\beta[\alpha^\#, Z] + \tfrac{1}{2}\alpha[\beta^\#, Z] \\ &= \tfrac{1}{2}(\{\alpha, \beta\})Z - \tfrac{1}{2}\beta L_{\alpha^\#}Z + \tfrac{1}{2}\alpha L_{\beta^\#}Z \ . \end{aligned}$$

Now the proposition follows from the three identities

$$L_{\alpha^\#}(\tfrac{1}{2}\beta(Z)) = (\tfrac{1}{2}L_{\alpha^\#}\beta)Z + \tfrac{1}{2}\beta(L_{\alpha^\#}Z)$$

$$-L_{\beta^\#}(\tfrac{1}{2}\alpha(Z)) = -\tfrac{1}{2}(L_{\beta^\#}\alpha)Z - \tfrac{1}{2}\alpha(L_{\beta^\#}Z)$$

$$-2L_Z(\omega[\alpha^\#, \beta^\#]) = d(i_{\alpha^\#}i_{\beta^\#}\omega)Z,$$

of which the first two are merely (2), above, and the third is a consequence of the equation $i_X\omega(Y) = 2\omega(X, Y)$.

Corollary 1. If β is closed then $\{\alpha, \beta\} = L_{\beta^\#}\alpha$.

<u>Proof</u>. By the homotopy identity,

$$\begin{aligned} L_{\alpha^\#}\beta &= i_{\alpha^\#}d\beta + d i_{\alpha^\#}\beta \\ &= 0 + 2d(\tfrac{1}{2}\beta(\alpha^\#)) = 2d(\omega(\beta^\#, \alpha^\#)) = d(i_{\alpha^\#}i_{\beta^\#}\omega) \end{aligned}$$

Now use the proposition.

<u>Corollary 2.</u> If α and β are closed, $\{\alpha, \beta\} = L_{\beta^\#}\alpha = -L_{\alpha^\#}\beta = 2d(\omega(\beta^\#, \alpha^\#))$.

<u>Corollary 3.</u> If α and β are closed, $\{\alpha, \beta\}$ is exact.

For $\{\alpha, \beta\} = d(2\omega(\beta^\#, \alpha^\#))$.

Now by using #, we see that each function f on M determines a vector field $X_f = (df)^\#$.

<u>Corollary 4.</u> If f and g are smooth functions on M, then

$$\begin{aligned} \{df, dg\} &= L_{X_g}(df) = d(L_{X_g}f) \\ &= -L_{X_f}(dg) = -d(L_{X_f}g) \\ &= 2d(\omega(X_g, X_f)) . \end{aligned}$$

Definition. The Poisson bracket of the functions f and g is

$$\{f, g\} = L_{X_g} f \qquad (X_g = (dg)^{\#}). \text{ (Hence } d\{f, g\} = \{df, dg\}).$$

Proposition. $L_{X_g} f = -L_{X_f} g = -2\omega(X_g, X_f)$.

Proof.

$$\begin{aligned} L_{X_g} f &= \langle df, X_g\rangle = \langle df, dg^{\#}\rangle = 2\omega(df^{\#}, dg^{\#}) \\ &= 2\omega(X_f, X_g) = -2\omega(X_g, X_f) = -2\omega(dg^{\#}, df^{\#}) \\ &= -\langle dg, df^{\#}\rangle \\ &= -L_{X_f} g. \end{aligned}$$

In particular, $L_{X_g} f = 0$ if and only if $L_{X_f} g = 0$. Thus f is constant on the trajectories of g if and only if g is constant on the trajectories of f. (By the trajectories of f we mean those of the vector field X_f.)

Of course, we must check that this definition agrees with our coordinate-wise notion of poisson bracket. Let the symplectic coordinates be $\{p_i, q^i\}$. This means that $\omega = \Sigma\, dp_i \wedge dq^i$. Any 1-form α can be written

$$\alpha = \Sigma\, h_i dq^i + \Sigma\, k^j dp_j ,$$

while any vector field X can be written

$$X = \Sigma\, X^i \frac{\partial}{\partial q^i} + \Sigma\, T^j \frac{\partial}{\partial p_j} .$$

Then we have seen that

$$\begin{aligned} X^{\flat} &= \Sigma\, T^i dq^i - \Sigma\, X^i dp_i \\ \alpha^{\#} &= -\Sigma\, k^i \frac{\partial}{\partial q^i} + \Sigma\, h_i \frac{\partial}{\partial p_i} . \end{aligned}$$

Thus $\{f, g\} = -L_{X_f} g = \Sigma\left(\frac{\partial f}{\partial p_j}\frac{\partial g}{\partial q^j} - \frac{\partial f}{\partial q^j}\frac{\partial g}{\partial p_j}\right)$ as expected.

Since our poisson bracket $\{\ ,\ \}$ was defined invariantly from the 2-form ω, the formula holds for any symplectic (= canonical) coordinates. In particular, this formula gives the poisson bracket of any two coordinate functions. We can deduce hence that a set of $2n$ functions $Q^1, \ldots, Q^n, P_1, \ldots, P_n$ on a symplectic manifold are symplectic coordinates if and only if they satisfy the relations

$$\{P_i, Q_j\} = \delta_{ij}$$
$$\{P_i, P_j\} = \{Q_i, Q_j\} = 0$$

for all i and j.

One can also prove that a transformation $\varphi:(M, \omega) \longrightarrow (M, \omega)$ is symplectic if and only if it preserves all poisson brackets of functions.

<u>Proposition.</u> For any three smooth functions on a symplectic manifold

$$\{f, \{g, h\}\} + \{g\{h, f\}\} + \{h, \{f, g\}\} = 0.$$

This asserts that the set of smooth functions is a Lie algebra under the poisson bracket $\{\ ,\ \}$.

<u>Proof.</u>

$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\}$

$$= L_{X_f} L_{X_g} h - L_{X_g} L_{X_f} h + L_{X_{\{f, g\}}} h$$

and this is zero, because

$$X_{\{f, g\}} = (d\{f, g\})^{\#} = \{df, dg\}^{\#} = -\{df^{\#}, dg^{\#}\} = -[X_f, X_g].$$

Here is an application of the antisymmetry of the poisson bracket. Consider k particles moving in three dimensional space. Their position is then specified by $3k$ coordinates

$$(x_1, y_1, z_1; x_2, y_2, z_2; \ldots, x_k, y_k, z_k),$$

so that the configuration space C is $\mathbb{R}^{3k}$. In the corresponding phase space $M = T^*C$ we can write down the Hamiltonian function H in terms of the potential energy V and the usual kinetic energy of the 3k particles. If we assume that V depends only on the distances between particles, then the Hamiltonian H is left fixed by the transformations of M induced by rigid motions, like translations and rotations, in $\mathbb{R}^3$. Let X_g be the vector field corresponding to such a translation; then X_g leaves H invariant. By anti-symmetry, X_H must leave g invariant; that is, since the system moves along the trajectories of H, g is a constant of the motion. For translations, g turns out to be the linear momentum, while for rotations g is angular momentum. We have just derived the familiar conservation-of-momentum laws. In general, any function f with $\{f, H\} = 0$ is a constant of the motion.

35. Submanifolds and Immersions.

We will study "energy surfaces" (submanifolds of constant energy); for this we need some facts about submanifolds. In a number of places in these lectures we have used (and will be using) the theorem below and its corollaries. (Here $Df(m)$ is the map induced by f on the tangent space at the point m.)

Theorem. (Inverse Function Theorem): Let $M \xrightarrow{f} N$ be a smooth function. If $Df(m)$ is an isomorphism, then f is a local diffeomorphism at m; i.e., there are neighborhoods U of m and V of fm such that $f(U) = V$ and $f|U: U \longrightarrow V$ has a smooth inverse.

Corollary 1. (Implicit Function Theorem): Let $M \xrightarrow{f} N$ be a smooth function. If $Df(m)$ is a surjection, then f is locally a projection; i.e., there are charts (U, ϕ) at m and (V, ψ) at fm such that

$\phi U = U' \times V'$

$\psi V = V'$

and $\psi \circ f \circ \phi^{-1}$ is the projection of $U' \times V'$ onto V'.

$$\begin{tikzcd}
M \arrow[r, "f"] & N \\
\bigcup & \bigcup \\
m \in U \arrow[d, "\phi", "\cong"'] & fm \in V \arrow[d, "\psi", "\cong"'] \\
U' \times V' \arrow[r, "\psi \circ f \circ \phi^{-1}"] & V'
\end{tikzcd}$$

Corollary 2. Let $M \xrightarrow{f} N$ be a smooth function. If $Df(m)$ is an injection, then f is locally an injection; i.e., there are charts (U, ϕ) at m and (V, ψ) at fm such that $\phi U = U'$, $\psi V = U' \times V'$, and $\psi \circ f \circ \phi^{-1}$ is the injection of U' into $U' \times V'$ as $U' \times 0$.

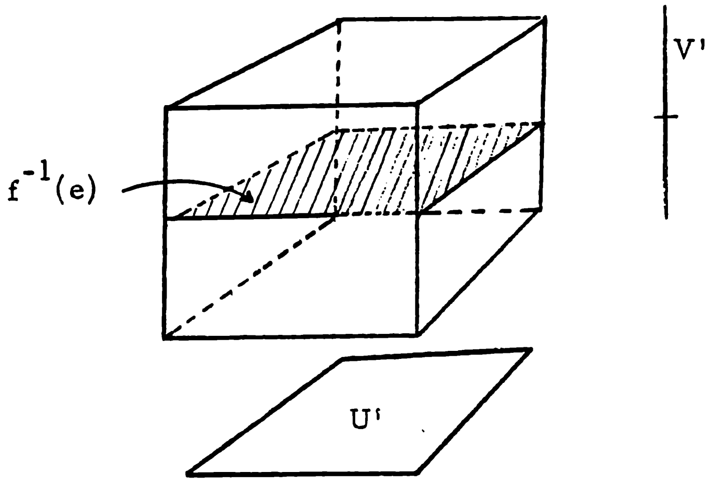
In what follows, M and N are manifolds and $a \in M$. An <u>immersion</u> of M in N is a smooth function $M \xrightarrow{f} N$ with the property: for each $f(a) \in N$, there is a chart at $f(a)$ with coordinates $q^1, \ldots, q^n$ such that $q^1 f, \ldots, q^d f$ are coordinates for a chart at a, for some $d \leq n$.

By Corollary 2 above, a smooth function $M \xrightarrow{f} N$ is an immersion if and only if $Df(a)$ is an injection for every $a \in M$. An <u>embedding</u> is an immersion which is a homeomorphism onto its image endowed with the subspace topology. A weaker notion of embedding which sometimes is used is an immersion that is an injection (1-1 function); but the stronger sense seems to be what we want for mechanics. If $M \subset N$ and the inclusion is an embedding, then M is a <u>submanifold</u> of N. One last definition: the point $e \in N$ is a <u>regular value</u> of f if and only if the Jacobian $Df(a)$ has maximum rank for every a such that $f(a) = e$. Since $Df(a)$ is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ it will have maximum rank when it is surjective if $m \geq n$ and when it is injective if $m \leq n$.

<u>Theorem</u>. Let N, P be manifolds and $N \xrightarrow{f} P$ a smooth function. If $e \in P$ is a regular value of f, then $f^{-1}(e)$ is a submanifold of N.

<u>Proof</u>. Since the inclusion $f^{-1}(e) \subset \overset{N}{M}$ is clearly a homeomorphism onto its image we need only show the immersion property. Take $a \in N$ so that $f(a) = e$.

Case (i). $Df(a)$ is surjective: Then Corollary 1 of the Inverse Function Theorem gives a chart $(U, \emptyset)$ at a such that $\emptyset(f^{-1}(e) \cap U) = U'$ and $\emptyset U = U' \times V'$. Locally the picture shows what the manifold M looks like near a.



Case (ii). $Df(a)$ is injective: Then f is an injection near a, so $f^{-1}(e)$ is a set of isolated points each of which is a submanifold. Thus the union of these is a submanifold, so the theorem is proved.

If $P = \mathbb{R}$ in the theorem, it is useful to have explicit local coordinates for a point a such that $f(a) = e$. Let $q^1, \ldots, q^n$ be any coordinates around a; then $Df(a) = \left(\frac{\partial f}{\partial q^1}, \ldots, \frac{\partial f}{\partial q^n}\right)_a$ where e is a regular value means one of the entries is non-zero, say $\frac{\partial f}{\partial q^n}$. The change of coordinates $\begin{array}{l} q^i \longrightarrow q^i \quad 1 \leq i \leq n-1 \\ q^n \longrightarrow f \end{array}$ has Jacobian matrix

$$\begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & \frac{\partial f}{\partial q^n} \end{pmatrix}$$

so $q^1, \ldots, q^{n-1}, f$ are coordinates around a in N such that $q^1, \ldots, q^{n-1}$ are coordinates around a in $f^{-1}(e)$.

36. Invariants on a Symplectic Manifold.

We first study quantities invariant under a vector field on any manifold.

Definition. Let M be a manifold and X a vector field on M; then a k-form α is invariant under X if and only if $L_X\alpha = 0$.

We have the equivalences

$$\begin{aligned} L_X\alpha = 0 \ &\text{iff } F_t^*\alpha = \alpha, \text{ where } F \text{ is a flow of } X, \\ &\text{iff } \alpha \text{ is constant on integral curves of } X. \end{aligned}$$

The following properties are easy to prove.

(i) α invariant under X implies $i_X\alpha$ and $d\alpha$ are invariant under X.

(ii) α and β invariant under X implies $\alpha\wedge\beta$ is invariant under X.

(iii) α invariant under X and $L_XY = 0$ implies $i_Y\alpha$ is invariant under X,

where X, Y are vector fields and α, β forms.

Proof of (i) for $i_X\alpha$: By the "homotopy identity" (1) of 33, $L_X = i_Xd + di_X$, thus

$$L_X(i_X\alpha) = i_Xdi_X\alpha + di_Xi_X\alpha = -i_Xi_Xd\alpha + di_Xi_X\alpha.$$

But the definition of i_X shows that $i_X i_X = 0$, since we are dealing with alternating tensors, so $L_X(i_X\alpha) = 0$.

Proof of (ii): L_X is a derivation.

Proof of (iii):

$$\begin{aligned} L_X(i_Y\alpha) &= L_X(\mathcal{C}(\alpha \otimes Y)) \\ &= \mathcal{C}(L_X(\alpha \otimes Y)) \\ &= \mathcal{C}(L_X\alpha \otimes Y + \alpha \otimes L_X Y) = 0 \end{aligned}$$

where $\mathcal{C}$ is the contraction operator which commutes with Lie derivatives.

A submanifold V of M is an <u>invariant submanifold of X</u> if for each $a \in V$, $X_a \in T_a V \subset T_a M$.

<u>Theorem</u>. Suppose M a manifold and X a vector field on M. If the function $M \longrightarrow \mathbb{R}$, as a 0-form, is invariant under X and if e is a regular value of k, then $k^{-1}(e)$ is an invariant submanifold of X.

<u>Proof</u>. By the previous theorem, $k^{-1}(e)$ is a submanifold of M and for each $a \in M$ such that $k(a) = e$ there are coordinates $q^1, \ldots, q^m$ around a in M with $q^1, \ldots, q^{m-1}$ coordinates for a in $k^{-1}(e)$ and $\frac{\partial k}{\partial q^m} \neq 0$. By definition of L_X for $X = \Sigma X^i \frac{\partial}{\partial q^i}$,

$$0 = L_X k = \langle dk, X \rangle = \sum_{i=1}^{m-1} \frac{\partial k}{\partial q^i} X^i + \frac{\partial k}{\partial q^m} X^m = \frac{\partial k}{\partial q^m} X^m$$

at a because k is constant on $k^{-1}(e)$. Thus $X^m a = 0$; but

$T_a(k^{-1}(e)) = \{w \in T_aM \mid \text{the last component } w_m = 0\}$. So $X_a \in T_a(k^{-1}(e))$, which was to be proved.

<u>Proposition.</u> If the function $M \xrightarrow{k} \mathbb{R}$ is an invariant of the vector field X on the manifold M and if e is a regular value of K, then a trajectory of X which meets a connected component Σ_e of $k^{-1}(e)$ lies entirely in Σ_e.

(Here <u>connected</u> means <u>path connected</u>; i.e., any two points of Σ_e can be connected by a path lying entirely in Σ_e.)

<u>Proof.</u> Let γ be a trajectory for X starting at the point $a \in \Sigma_e$. Because Σ_e is an invariant submanifold of X, $X|\Sigma_e$ is a vector field on Σ_e. The existence and uniqueness theorems for differential equations say there is a unique trajectory γ' in Σ_e starting at a which satisfies the differential equation for $X|\Sigma_e$. But γ is such a trajectory, thus $\gamma = \gamma'$ and γ lies in Σ_e.

Now consider invariants for a symplectic manifold (M, ω) of dimension $2n$.

<u>Definition.</u> A vector field X on (M, ω) is <u>locally Hamiltonian</u> if and only if ω is invariant under X; i.e., $L_X\omega = 0$. Equivalent conditions are

(a) $di_X\omega = 0$,

(b) $X^\flat$ is closed,

(c) $X^\flat = dH$ locally, for some function H.

The vector field X is globally Hamiltonian if and only if there is a smooth $M \xrightarrow{H} \mathbb{R}$ such that $X = (dH)^{\#}$, or equivalently, $X^{\flat}$ is exact.

Recall that a volume on M is a nowhere zero $2n$-form. For example $\Omega = \frac{-1^{n/2}}{n!} \omega \wedge \ldots \wedge \omega$ (n times) is a volume. If X is a locally Hamiltonian vector field, then $L_X \omega = 0$, so $L_X \Omega = 0$. The volume element thus is constant under motion along X. This statement becomes Liouville's Theorem when translated into the language of statistical mechanics. In detail, in statistical mechanics a system of n particles is replaced by a single particle in $3n$-dimensional configuration space, and hence by a point moving in $6n$-dimensional phase space. An ensemble of systems thus corresponds to an ensemble of points in phase space. Liouville's Theorem states that the density of this ensemble is constant along the trajectories.

37. Submanifolds of Constant Energy.

The last proposition shows that for the trajectories defined by the globally Hamiltonian vector field $(dH)^{\#}$ it is appropriate to restrict consideration to the submanifolds where H is constant. We now examine the structure of such submanifolds for any suitable function K.

<u>Theorem.</u> (Hamilton-Jacobi): Let X be a vector field on an m-dimensional manifold M, $M \xrightarrow{K} \mathbb{R}$ an invariant of X, e a regular value of K, and V a connected component of $K^{-1}(e)$; then

1° V is an embedded submanifold of dimension $m-1$.

2° If M is oriented, so is V.

3° If M has an X-invariant volume Ω, then V has an $X|V$-invariant volume Ω_V.

<u>Proof.</u> 1° has already been proven.

2°. We will need a "normal" vector field N for V; that is, a smooth function $V \longrightarrow T.M$ with the property that for each $a \in V$

$$T_{\nu a}M = \nu_*(T_aV) \oplus \mathbb{R}\cdot Na$$

where ν is the inclusion of V in M. To get such an N let g be any Riemannian metric on M (see the end of the proof for a different method), and set $N = g_\# \circ dK \circ \nu$. Thus Na for each point a is the unique vector such that $g(Na, -) = d_{\nu a}K$ is the differential of K at νa.

$$\begin{tikzcd}
 & T.M \arrow[dr, "\pi."] & & T^*M \arrow[ll, "g_\#"'] \\
V \arrow[ur, dashed, "N"] \arrow[rr, "\nu"'] & & M \arrow[ur, "dK"'] &
\end{tikzcd}$$

To prove that N is a "normal" vector field it suffices to show that $N_a \notin \nu_*(T_aV)$. Now

$$\nu_*(T_aV) = \{Y \in T_{\nu a}M \mid \langle d_{\nu a}K, Y\rangle = 0\},$$

but $\langle d_{\nu a}K, Na\rangle = g(Na, Na)$ which is not zero unless $Na = 0$. Since

e is a regular value $d_{\nu a}K = g(Na, -) \neq 0$, so Na cannot be zero.

Suppose Ω is a non-zero volume on M, then $\nu^*(i_N\Omega)$ is an $(m-1)$-form on V and

$$\begin{aligned}\nu^*(i_N\Omega)(w_1, \ldots, w_{m-1}) &= i_N\Omega(\nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= m\Omega(N, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &\neq 0.\end{aligned}$$

Therefore, $\nu^*(i_N\Omega)$ is a volume on V. Consequently, V is orientable. This applies in particular when M is symplectic, and hence orientable.

3° First we need a lemma:

Lemma. If γ is an $(m-1)$-form on M, $a \in V$, and $w_1, \ldots, w_{m-1} \in T_aV$, then

$$\begin{aligned}&(dK \wedge \gamma)_a(Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &\qquad = c \langle dK, N \rangle_a \gamma a(\nu_* w_1, \ldots, \nu_* w_{m-1})\end{aligned}$$

with c a non-zero constant.

<u>Proof of Lemma.</u> We use the definition of a form as an alternating tensor. Indeed

$$(dK \otimes \gamma)_a(Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) = \langle dK, N \rangle_a \gamma_a(\nu_* w_1, \ldots, \nu_* w_{m-1}),$$

so $(dK \wedge \gamma)_a(Na, \nu_* w_1, \ldots, \nu_* w_{m-1})$

$$\begin{aligned}&= \frac{1}{m!} \sum_{\sigma \in S_m} (-1)^{\sigma} (dK \otimes \gamma)_a(Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= \frac{1}{m} \langle dK, N \rangle_a \gamma_a(\nu_* w_1, \ldots, \nu_* w_{m-1}),\end{aligned}$$

because $\langle dK, w_i \rangle_a = 0$ and γ is a form. Putting $c = \frac{1}{m}$ proves the lemma.

Now assume Ω is an X-invariant volume on M and let $\theta = i_N\Omega$. Then $dK \wedge \theta$ is an m-form, and hence is a multiple of Ω:

$$dK \wedge \theta = h \cdot \Omega \qquad \text{where } h: M \longrightarrow \mathbb{R} \text{ is smooth.}$$

In order to make $\nu^*\theta$ invariant, we want $h = 1$ -- if $h \neq 0$, we can multiply by h^{-1}. But

$$\begin{aligned} ha\Omega_a&(Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= (dK \wedge \theta)_a (Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= c \langle dK, N\rangle_a \theta_a (\nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= c' \langle dK, N\rangle_a \Omega_a (Na, \nu_* w_1, \ldots, \nu_* w_{m-1}). \end{aligned}$$

For $w_1, \ldots, w_{m-1}$ linearly independent in T_aV, all the terms, except ha, at both ends of this equation are non-zero; therefore ha is also non-zero. Thus set $\bar{\theta} = h^{-1}\theta$; then $dK \wedge \bar{\theta} = \Omega$ and $\nu^*\bar{\theta}$ is a volume (as in 2^o). We must finally show $\nu^*\bar{\theta}$ invariant under $XN = \nu^*X$. First

$$0 = L_X\Omega = L_X dK \wedge \bar{\theta} + dK \wedge L_X\bar{\theta}$$

so $dK \wedge L_X\bar{\theta} = 0$.

Using the lemma again

$$\begin{aligned} 0 &= (dK \wedge L_X\bar{\theta})_a (Na, \nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= c \langle dK, N\rangle_a (L_X\bar{\theta})_a (\nu_* w_1, \ldots, \nu_* w_{m-1}) \\ &= c \langle dK, N\rangle_a \; \nu^*(L_X\bar{\theta})_a (w_1, \ldots, w_{m-1}) \\ &= c \langle dK, N\rangle_a (L_{X|V} \nu^*\bar{\theta})_a (w_1, \ldots, w_{m-1}). \end{aligned}$$

Since the first two factors c and $\langle dK, N\rangle_a$ on the right are non-zero, we get $L_{X|V}(\nu^*\bar{\theta}) = 0$, which proves 3^o.

This proof of the Hamilton-Jacobi theorem is a presentation of that given in Abraham, loc. cit., in numbers 11.11, 11.15, 15.13, 16.27.

This proof made use of the fact that there exists a Riemannian metric on any manifold. The use of this result is efficient and suggestive, but there is no unique or canonical such metric. After the lecture, Alan Waterman suggested the following (standard) way of avoiding the choice of a metric.

Lemma. For M, K, and V as in the theorem, let Ω be any m-form on M. Then there is a 1-form θ on V such that for any point a of V there is a coordinate neighborhood of a in M and an $(m-1)$-form β on the neighborhood with

$$\theta = \nu^{*}\beta \quad , \quad \Omega = \beta \wedge dK.$$

Proof. Let $\nu: V \to M$ be the inclusion. Since $dK \neq 0$ at e, we can take K as one of the coordinates at a in M. The form Ω of maximal dimension can then be written locally as $\Omega = \beta \wedge dK$, where β is some $(m-1)$-form on M. This form β is not unique, but if also $\Omega = \beta' \wedge dK$ in the same coordinate neighborhood, then a representation with coordinates shows that $\nu^{*}\beta = \nu^{*}\beta'$. Therefore $\theta = \nu^{*}\beta = \nu^{*}\beta'$ is an $(m-1)$-form well-defined everywhere on V, and the lemma holds.

The theorem itself is now readily proved from this lemma; in particular, since Ω and K are both invariants for X, it follows that θ is an $X|V$-invariant volume.

Chapter V. QUALITATIVE PROPERTIES OF VECTOR FIELDS

38. Orbits.

This chapter is devoted to the study of the qualitative properties of the trajectories (integral curves) of vector fields on manifolds. Except for the first section, the chapter consists of notes of lectures given by Prof. René Thom of the Institute des Hautes Etudes Scientifiques (France).

Hence suppose M is a manifold and X a vector field on M. Consider trajectories of X: that is, the curves $c: I \longrightarrow M$ such that the tangent vector to c at each point is the value of X at that point. We shall assume that c and the given interval I are chosen so that $0 \in I$.

Let

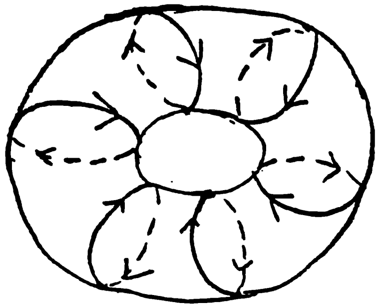
$$\mathcal{D}_X = \{(a,t) \mid a \in M, t \in \mathbb{R} \text{ and there exists a trajectory } c: I \longrightarrow M \text{ of } X \text{ with } c(0) = a \text{ and } t \in I\}.$$

By the uniqueness theorem for differential equations, there exists a smooth map (a "flow")

$$F: \mathcal{D}_X \longrightarrow M$$

such that for each a, $F(a, -)$ is a trajectory for X -- in fact is the maximal trajectory through a. ($\mathcal{D}_X$ is an open subset of $M \times \mathbb{R}$, so it is meaningful to require that F be smooth.)

An <u>orbit</u> of X is the image in M of a maximal trajectory. A <u>closed orbit</u> of X is an orbit which is compact. For example, the vector field going around the torus as indicated has (infinitely) many closed orbits.



Finally, the support of X is the closure of the set $\{a \in M | X(a) \neq 0\}$.

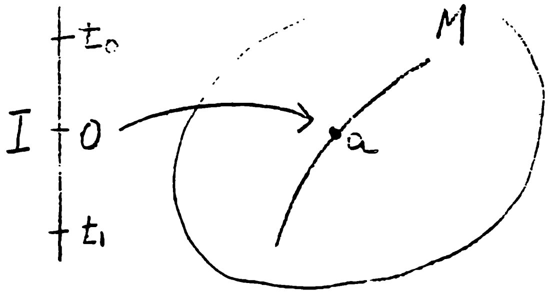
A vector field X is complete when $\mathcal{V}_X = M \times \mathbb{R}$. It is equivalent to say that every integral curve of X can be extended to one whose domain of definition is the whole real line. This cannot always happen. For example, let U be the first quadrant of $\mathbb{R}^2$, with the usual coordinates x, y. Then the vector field $X(x, y) = (\frac{1}{x}, 1)$ is not complete. (Verification is left to the reader.) If however, the closure of the set of points p where $X(p) \neq 0$ is compact , then X is complete.

The most important thing about a complete vector field is that it yields a one-parameter group of diffeomorphisms. For each $s \in R$, there is a diffeomorphism Φ_s of the manifold such that $\Phi_o = 1$ and $\Phi_s \Phi_t = \Phi_{s+t}$. Explicitly, $\Phi_s(p)$ is the value of the integral curve of X with initial conditions p at time s. In contrast a non-complete vector field yields only local diffeomorphisms rather than global ones.

Theorem. Let X be a vector field on M. If M is a compact manifold or if X has compact support, then X is complete.

We sketch the proof in the case that M is compact. We shall use the fact that in a compact space every infinite sequence of points has a convergent subsequence.

We want to show that we can prolong any trajectory $c: I \longrightarrow M$ where $I = (t_o, t_1)$ for $t_o < 0 < t_1$. Suppose c were maximal and $t_1 < \infty$. Let $\{t_i\}$ be a sequence of points in I convergent to t_1. Then the sequence $\{c(t_i)\}$ in M has a limit point m (i.e., some subsequence of $\{c(t_i)\}$ converges to m). Apply the existence theorem of differential equations to get a flow box $F: U \times I_1 \longrightarrow M$ at m. Thus $U \subset M$ is an open set containing m. But m must be in the closure of the image of c. Therefore, U contains some point $c(\bar{t})$ in the image of c. $F(c(\bar{t}), -)$ is a trajectory through $c(\bar{t})$ which must extend c.



A similar argument applies to the case when M is not compact but X has compact support.

Suppose next that K is a compact orbit of X. Then we may as well assume that K is not a point, and consequently that X is never zero on K. The main result is that compact orbits are periodic.

Theorem. Suppose $\varphi: \mathbb{R} \longrightarrow M$ is a non-constant integral curve of X with compact image K. Then there is a $\tau > 0$ such that

$\varphi(t+\tau) = \varphi(t)$, all t.

The least such t is called the period of the orbit.

Proof. We first notice that we can assume the vector field X to be complete. For if we take a neighborhood W of K with compact closure and a C^∞ function α which is 1 near K and 0 off W, then αX is a vector field with compact support which agrees with X near K; also φ is an integral curve of αX.

It suffices to show that there are points $t \neq t'$ such that $\varphi(t) = \varphi(t')$. For if $\{\Phi_s : M \longrightarrow M\}$ is the corresponding 1-parameter group of diffeomorphisms of M, it then follows that

$$\varphi(s+t-t') = \Phi_{s-t'}\varphi(t) = \Phi_{s-t'}\varphi(t') = \varphi(s)$$

for all $s \in R$. If $t > t'$, this is the conclusion $(\tau = t - t')$ of the theorem; since t and t' are symmetric, we may suppose this is so.

We thus want to prove the following

Lemma. There are points $t \neq t'$ such that $\varphi(t) = \varphi(t')$.

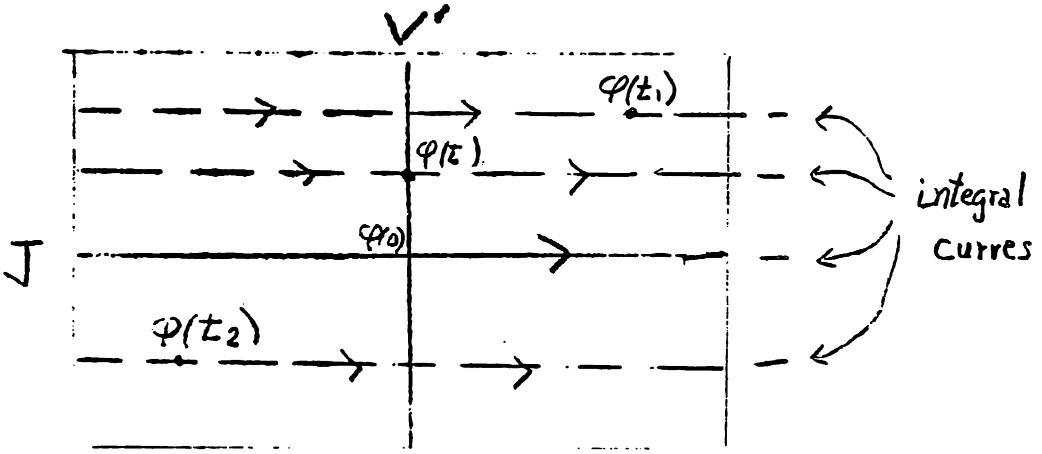
Proof As noted earlier, X is non-zero on a neighborhood of K. Therefore, by the inverse function theorem, there exists an open interval $J \subset \mathbb{R}$ containing 0 and an open disk V' of radius ε such that $J \times V'$ is a coordinate neighborhood at $X(0)$ with coordinates t and $v^1, \ldots, v^{n-1}$ and such that locally

$$\varphi(t) = (t, 0).$$

By continuity we may assume that if

$$X = \sum w_i \frac{\partial}{\partial v^i} + y \frac{\partial}{\partial t}$$

then y is bounded away from zero (and in fact is positive on $J \times V'$).



Suppose $\varphi(t) \in J \times V'$. By our hypothesis that y is positive, a point can travel backward or forward along the trajectory of φ until it meets the disk $0 \times V'$. (In the above picture, the point travels backwards from $\varphi(t_1)$ and forwards from $\varphi(t_2)$.)

Again by our hypotheses on X, if $\varphi(t) = (0, v')$, then no other nearby point can also get mapped into $0 \times V'$. This is because the component y is bounded away from zero, and hence the integral curves must have at least some fixed positive velocity in the $\frac{\partial}{\partial t}$ direction. In other words, the set of points t such that $\varphi(t) \in 0 \times V'$ is an isolated subset of R. But it is easy to show that any such subset is countable; that is, it is in 1-1 correspondence with a subset of the positive integers.

Since $K \cap (0 \times V')$ is countable and $[0, 1)$ is uncountable, there exists an $\varepsilon' < \varepsilon$ such that no point of K has distance exactly ε' from the origin. Let V be the closed disk of radius ε'; then $K \cap (0 \times V)$ is compact and countable, and is contained in the interior of V. As a subset of a metric space, it has a notion of distance.

It is left to the reader to prove the next result, which is a trivial consequence of Baire's theorem (see e.g., Kelley, General Topology)

Sublemma. Let Y be a compact metric space which is countable. Then Y has an isolated point.

Therefore, there is a t_o and a neighborhood U of $\varphi(t_o)$ in V such that $K \cap U = \emptyset$. Again, it follows by continuity that there is a $U' \subseteq U$ neighborhood of $\varphi(t_o)$ and a $J_o \subseteq J$ neighborhood of 0 such that $K \cap (J_o \times U') = J_o \times \varphi(t_o)$

Composing the result with the diffeomorphism Φ_s for $s = t_i - t_o$, we see that for any point $\varphi(t_i) \in K$, there exists a coordinate neighborhood $J \times U$ of $\varphi(t_i)$ such that the intersection of K with the neighborhood is $J \times 0$.

Since K is compact, there is a finite collection of $(J_i \times U_i)$ covering it. Since the inverse images of the $J_i \times 0$ then cover $\mathbb{R}$, it follows that some point t' outside some interval $t_k + J_k$ gets mapped into $J_k \times 0$. But we know some $t \in t_k + J_k$ gets mapped onto any point of $J_k \times 0$, and for these choices $\varphi(t') = \varphi(t)$.

We now want to ask what the critical elements of a vector field X are. These are of two types:

1) $a \in M$ such that $X(a) = 0$ (for example, the south pole on the sphere with a vector field which everywhere points downward). Since

$X(a) = 0$, the Jacobian

$$\left\| \frac{\partial x^i}{\partial q^j} \right\|$$

provides a linear approximation to X near a. For example, in dimension 1, a Taylor expansion gives

$$\frac{dq}{dt} = \lambda q + \text{higher order terms},$$

where λ is the eigenvalue of the Jacobian. Thus

$$\frac{dq}{dt} = \lambda q$$

gives a first approximation to solutions near a. In higher dimensions it is generally possible to choose coordinates $q^1, \ldots, q^n$ so that

$$\frac{dq^i}{dt} = \lambda^i q^i$$

give first approximations to the solutions near a, where the λ^i are the eigenvalues of the Jacobian matrix near a.

2) Closed orbits. Nearby trajectories may be studied by taking a normal cross-section to the given closed orbit. Again, suitable eigenvalues determine the behavior; they are obtained by mapping the cross-section on itself by following along trajectories going "once around" the orbit.

Structural Stability -- René Thom*

The purpose of mechanics is to describe the motion of physical bodies. Recently the theories developed for this aim have also been used to study chemical and even biological phenomena.

Two separate theories of mechanics have evolved. Time reversible mechanics is based on the assumption that the time parameter can be reversed without changing the qualitative aspects of the phenomenon being studied. Vibration without damping is an example of such a phenomenon. Time reversible mechanics has been dominated by Hamiltonian theory and is centered on the concept of Invariance of Energy. Time reversible mechanics suffers from the defect that it is in most cases an idealization of nature. Time-irreversible mechanics is more true to nature but has been studied less than time-reversible mechanics. It is dominated by the study of gradient-like systems and centered on the concept of Increase of Entropy. More explicitly, if X is a vector field on a phase space M, then Increase of Entropy is satisfied if there exists a function $S: M \longrightarrow \mathbb{R}$ (the entropy function) such that $S(m_t)$ is monotone increasing, where $\frac{dm_t}{dt} = X$. Otherwise put, X is transversal in an increasing direction to the level varieties of S.

(*) Prof. Thom wishes to thank Prof. S. MacLane for having been given this opportunity to expose some favorite ideas in the field of Goemetrical Mechanics.

39. Gradient Vector Fields.

Let M be a manifold with a Riemannian metric $<,>$. Then there is a correspondence between vector fields X on M and 1-forms α_X on M given by

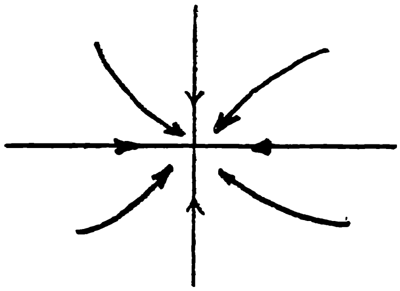
$$<X,u> = \alpha_X(u) \quad \text{for } u \in T_m M,$$

at each $m \in M$. (This is just the correspondence induced by the isomorphism of tangent and cotangent bundles given by $<,>$.) If $\alpha_X = dU$, then we set

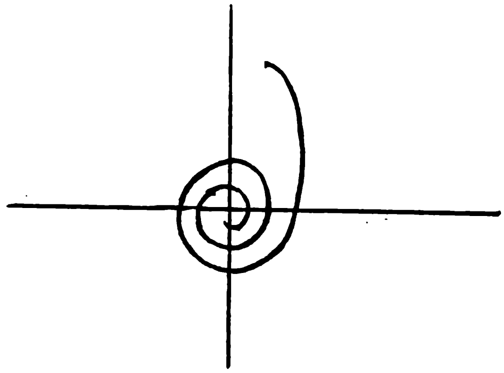
$$X = \text{grad } U.$$

The situation may be described by saying that X is orthogonal to the level surfaces of U. This is just like an entropy situation, particularly since we are free to choose a convenient Riemann metric.

At regular points (that is, points where $X \neq 0$) there is little difficulty in determining the nature of the trajectories. At a singular point the situation becomes more complicated. The behavior of trajectories near a singular point of a gradient field mightbe as in the picture below.



In a more general situation a focus, as in the diagram below, might occur; this cannot happen with a gradient field, however. (This is because of the symmetry of the second order derivatives $\frac{\partial^2 V}{\partial x_i \partial x_j}$



40. Qualitative Dynamics

The classical approach to dynamics was to try to solve the Hamilton-Jacobi equation

$$\frac{dm}{dt} = X(m)$$

explicitly. This appraoch was beset with difficulties. Frequently, X is not known exactly -- say, if not all the forces acting on a system are explicitly known. In this case, empirical formulae are used to approximate the desired information. Given a known vector field X, there are not always adequate means of integration available, as for the three-body problem of Newtonian mechanics. Then the solution must be approximated. To approximate reasonably, it is necessary to know how much a slight perturbation of X affects the global solution.

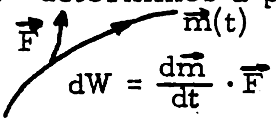
In the 1880's, Poincare introduced the study of Qualitative Dynamics, which aims to describe solutions rather than find them explicitly. Once the geometric picture formed by the trajectories is found, one can pose the <u>structural stability problem:</u> to determine whether this picture is invariant up to homeomorphism under small perturbations of X. We whall examine these problems for the case of gradient fields. We will deal with a potential function

$$M \xrightarrow{V} \mathbb{R}$$

and the corresponding vector field

$$X = -\operatorname{grad} V.$$

First, a digression concerning the nature of potential functions. If the vector field X corresponds as above to the one-form α_X, then $\alpha_X(z)$ is called the <u>work</u> of the field of force z. If α is closed -- $d\alpha = 0$ -- then we say "V determines a potential." Either of two cases may occur:



1) If α_X is homologous to zero in the one dimensional cohomology group then $H^1(M, \mathbb{R})$, then $\alpha = d(-V)$.

2) If α_X is not homologous to zero in $H^1(M, \mathbb{R})$, then the potential is "multi-valued"; i.e., it is defined up to multiples of periods. We shall consider only the first case.

<u>Definition.</u> $p \in M$ is a <u>critical point</u> of X if $X(p) = 0$; i.e., if p is a singular point of X.

If $X = -\operatorname{grad} V$, p will be a singular point of X if and only if $dV_p = 0$; i.e., if and only if p is a critical point of V in the usual sense.

Suppose M is compact. Let

$$h_t : M \xrightarrow{\text{onto}} M$$

be a one-parameter family of diffeomorphisms obtained by integrating X, such that $h_t(m_o) = m_t$. (That is, h_t is a flowbox for X.) Consider the h_t's applied to one point $m \in M$; since M is compact, the resulting set $\{h_t(m)\}$ has limit points.

Claim: Any limit point of a trajectory m_t for X is a critical point of V.

Proof. Suppose q is a limit point of m_t. Then the trajectory m_t keeps coming back near q. By definition, $V(m_t)$ is decreasing. Since M is compact, $\{V(m_t)\}$ has a lower bound. Thus we must have $V(m_t) \downarrow V(q)$ as $t \to \infty$. If q is not critical, then V has a non-zero gradient at q, so there would be nearby points with values less than V(q), contrary to $V(m_t) \downarrow V(q)$.

Thus we have shown that "all trajectories go to critical sets." The same reasoning shows that "all trajectories start from critical sets."

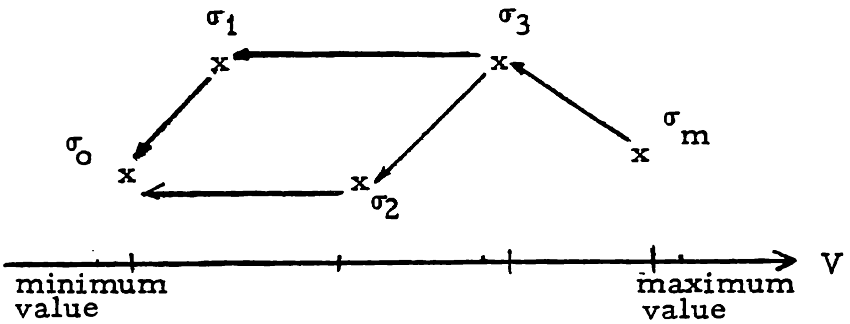
Definition: The critical set of X is $(dV)^{-1}(0)$. Thus, if V is C^1, then the critical set is closed.

Is V constant on any connected component of the critical set? If the components are all differentiable-arc-connected (i.e., if any two points in the same component can be connected by a differentiable arc lying in that component), then the answer is "yes".

Theorem (A.P. Morse)[1]: If V is of class C^m where $m \geq n = \dim M$, then V is constant on any component of the critical set.

There is a counterexample, due to Whitney, showing that V can be non-constant on a component if its class is too low.

When V is constant on each component of the critical set, the components can be ordered via the values which V assumes on them. The structure of the gradient field may then be described by the following procedure. Let σ_i be the components of the critical set. To each σ_i associate a point above the value in $\mathbb{R}$ which V assumes on σ_i. Draw an arrow from σ_i to σ_j whenever there is a trajectory for X in M which starts in a neighborhood of σ_i and ends in a neighborhood of σ_j. Such a graph might look like the following:



In what follows, assume V is C^∞. Let $V(x_i)$ denote the potential as a function of the coordinates x_i. Suppose the origin is a critical point -- i.e.,

$$\frac{\partial V}{\partial x_1}(0) = \dots = \frac{\partial V}{\partial x_n}(0) = 0.$$

---

1) See for instance G. de Rham, Variétés Differentiables, Hermann Paris, Th. 9, p. 10; or S. Sternberg, Lectures on Differential Geometry, Prentice Hall -- Sard's Theorem --Theorem 3.1, p. 47.

The origin is a non-degenerate critical point if the Hessian $\left\| \frac{\partial^2 V}{\partial x_i \partial x_j} \right\|$ is non-zero there. Equivalently, if

$$V(x_i) = V(0) + \varphi_2(x_i) + \ldots$$

is the Taylor expansion for V, then 0 is a non-degenerate critical point if and only if the quadratic form $\varphi_2(x_i)$ is non-degenerate. In this case, φ_2 can be reduced to the sum of squares of linearly independent forms:

$$\varphi_2(x_i) = \sum \pm x_i^2 = -x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_n^2 .$$

Then k is the index of the quadratic form and, along with the fact that 0 is a non-degenerate critical point, it is invariant under changes of coordinates.

Suppose $V(x_i)$ is a potential function which admits 0 as a non-degenerate critical point. Perturb V slightly to

$$V(x_i) + \delta V(x_i),$$

where $\delta V(x_i)$ is small in a suitable norm. Then the new function has near 0 a non-degenerate critical point of the same index.

To show this we introduce the "auxiliary map"

$$\mathbb{R}^n \xrightarrow{G} \mathbb{R}^n$$

given by $x_i \rightsquigarrow u_i = \frac{\partial V}{\partial x_i}(x_1, \ldots, x_n)$. Then the critical set is just $G^{-1}(0)$, and 0 is non-degenerate if and only if G has rank n at 0. Perturb V slightly in the C_2 norm; this perturbs G to a new map G' which is a C_1 approximation of G. Therefore G' has rank n in some neighborhood of 0, so $(G')^{-1}(0)$ must contain a point in this neighborhood. The invariance of index follows from the continuity of the second derivatives.

Thus we have that "non-degenerate critical points are invariant under small perturbations". This theorem is due to M. Morse. We next state without proof the following result of M. Morse:

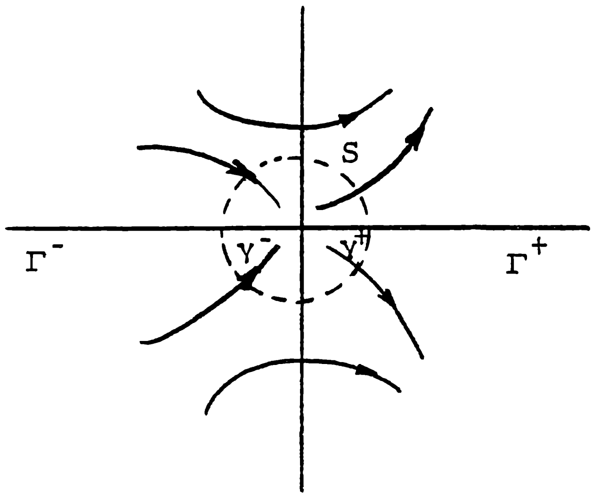
Lemma. (M. Morse): If $V(x_i) = V(0) + \varphi_2(x_i) + \varphi_3(x_i) + \dots$, then there is a change of coordinates $x_i \to x_i'$ near 0 such that

$$V(x_i') = V(0) + \sum \pm \; x_i'^2 \; ,$$

i.e., V can always be reduced to a quadratic form (plus a constant).

If a critical point p is non-degenerate, we can get a good description of the gradient field near p. Before attempting this, we state a related conjecture:

Suppose f is an analytic real-valued function on $\mathbb{R}^n$ with 0 as an isolated singular point; let (f_{x_i}) be the ideal (in the ring of analytic functions at 0) generated by the first partial derivatives of f. Then (f_{x_i}) contains a power of the ideal generated by the coordinate functions; in other words, any monomial of sufficiently high degree is a sum of multiples of the first partials of f. Consider the set of trajectories of grad f. There will generally be a set Γ^- of trajectories tending toward the origin, and another, Γ^+, of trajectories emanating from the origin. The problem is to show that Γ^- and Γ^+ have a nice topological structure; in particular, that they may be triangulated, preferably by a triangulation which can be extended to the whole space. It has so far been shown that if Γ^+ and Γ^- are cut by a suffi-



ciently small sphere S, yielding sets γ^+ and γ^-, then γ^+ is a deformation retract of $S-\gamma^-$, and vice-versa.

Now suppose 0 is a non-degenerate critical point of $F(x_i)$

$$F_{x_i}(0) = 0$$

and

$$F_{x_i x_j}(0) \neq 0.$$

What can be said about the qualitiative structure of the gradient field near 0?

By M. Morse's theorem, we can choose coordinates $x_1, \ldots, x_k$, $y_1, \ldots, y_{n-k}$ around the critical point 0 so that

$$F = \sum_{i=1}^{k} x_i^2 - \sum_{j=1}^{n-k} y_j^2 .$$

<u>Special case:</u> Take as the Riemann metric about 0

$$ds^2 = dx^2 + dy^2 = \sum dx_i^2 + \sum dy_j^2 .$$

Now

$$\frac{1}{2} F_{x_i} = x_i ,$$

$$\frac{1}{2} F_{y_j} = -y_j ,$$

and

$$\frac{1}{2} \operatorname{grad} F = \sum F_{x_i} \frac{\partial}{\partial x_i} - \sum F_{y_j} \frac{\partial}{\partial y_j} ,$$

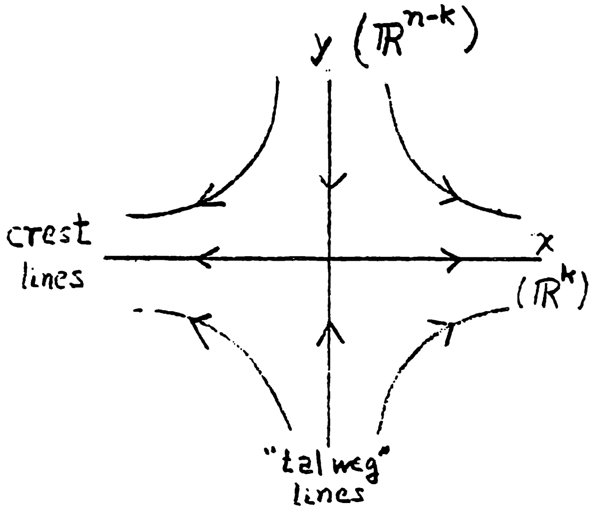
so integrating the system amounts to solving

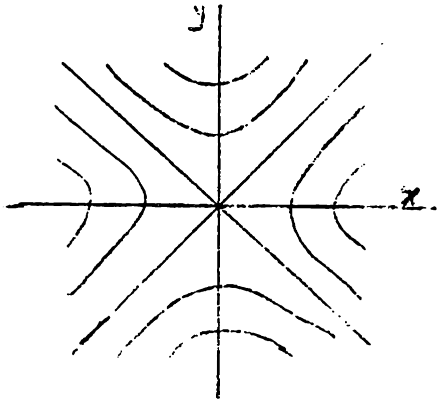
$$\frac{dx_i}{dt} = x_i \quad , \quad \frac{dy_j}{dt} = -y_j$$

which we can readily do to get

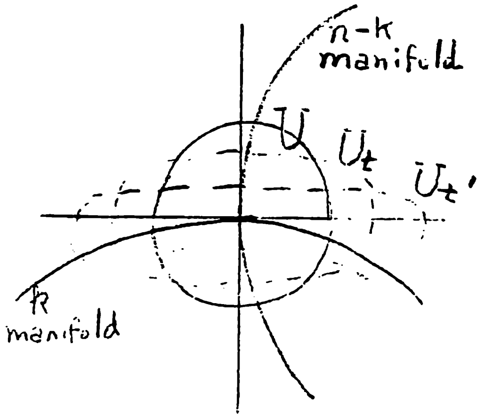
$$x_i = a_i e^t \quad , \quad y_j = b_j e^{-t} .$$

The orbits of trajectories tending to or from the origin fill up the "axes" R^{n-k} and $\mathbb{R}^k$, where $x = 0$ and $y = 0$, respectively. The trajectory through any point not in one of these sets does not pass through the origin. In general (if $k \neq 0, n$), there will be a saddle point -- as is pictured in the two-dimensional case for $F = x^2 - y^2$. (The level curves for F itself in the case $n = 2$ are pictured in the middle diagram)

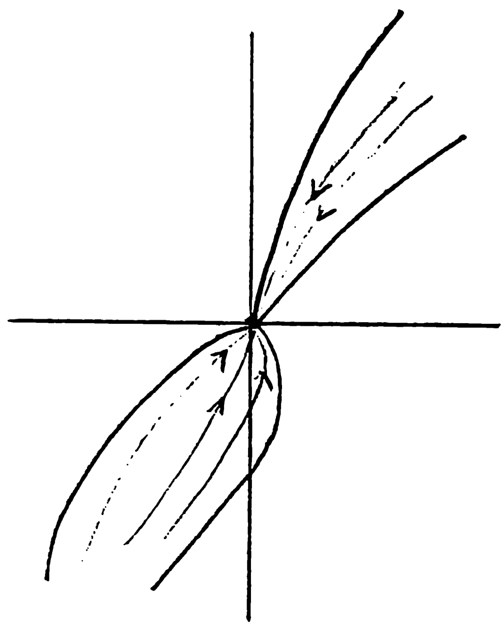




If instead we take an arbitrary Riemann metric in a neighborhood of the critical point, an analysis more delicate than that above will yield the same results. There will be a k-manifold passing through 0 formed by trajectories leaving 0, and an n-k manifold orthogonal to it formed by trajectories tending to 0, where n-k is the index of the critical point. The general method is to take a small neighborhood U of 0 and consider its images $U_t = h_t(U)$,



where the h_t form as above a flow box for X. The point 0 is a fixed point of h_t, so each U_t is a neighborhood of 0. As $t \to \infty$, these neighborhoods stretch along the x-axis, tending toward a limit which is a k-dimensional manifold of trajectories leaving the origin. As $t \to -\infty$, the U_t tend toward the n-k dimensional manifold of trajectories approaching the origin. P. Hartman has done the analysis, that of "unstable critical sets", involved.[3)]



The above results hold only for a C^2 function F. For example, if $F(x, y)$ is a C^1 but not C^2 function of two variables, the trajectories arriving at the origin might form a full sector.

It is known that the homotopy types of the sets of trajectories entering and leaving a critical point are fairly well determined. However, other topological properties may vary widely.

---

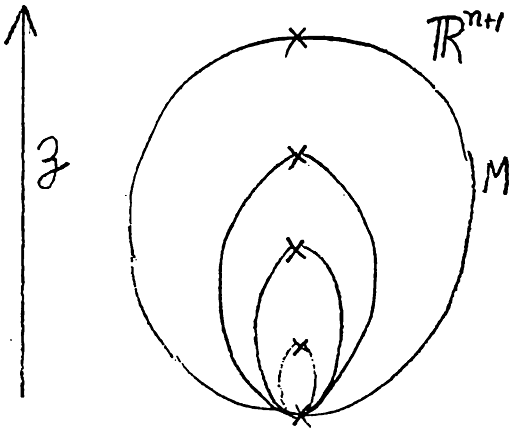
3) P. Hartman -- On the local linearization of differential equations, Proceedings Amer. Math. Soc., 1963, 14, pp. 568-73.

41. Morse Theory.

Theorem (M. Morse): Any real-valued function of class C^m on a compact manifold M^n, where $m > n$, can be approximated in the C^n topology by a function which admits only non-degenerate critical points.

Consider $\mathcal{F}^m(M^n)$ = all real-valued C^m functions on M^n, with the C^m topology; it is a Banach space. Morse's theorem says that the set of functions in $\mathcal{F}^m(M^n)$ which admit only non-degenerate critical points is open and dense in $\mathcal{F}^m$.

Now suppose $\mathfrak{z}$ is a function defined on a compact n-manifold M^n imbedded in $\mathbb{R}^{n+1}$ in such a way that $\mathfrak{z}$ is a coordinate in $\mathbb{R}^{n+1}$. By Morse's theorem, we may approximate $\mathfrak{z}$ by a function with only non-degenerate critical points. So without loss of generality, we may assume $\mathfrak{z}$ has only finitely many critical points, each of which is non-degenerate. Associate to each of these the set of "descending trajectories" from that point (i. e., the set of trajectories of -grad $\mathfrak{z}$). These form a k-dimensional cell $\mathfrak{z}_k$, where n-k is the index of the critical point. The cells $\mathfrak{z}_k$ form a partition of M -- for any point of M lies on a gradient trajectory with $\mathfrak{z}$ increasing, and this trajectory must tend toward a critical point. Thus we have the basic objective of Morse theory a representation of M as a union of k-cells $\mathfrak{z}_k$.



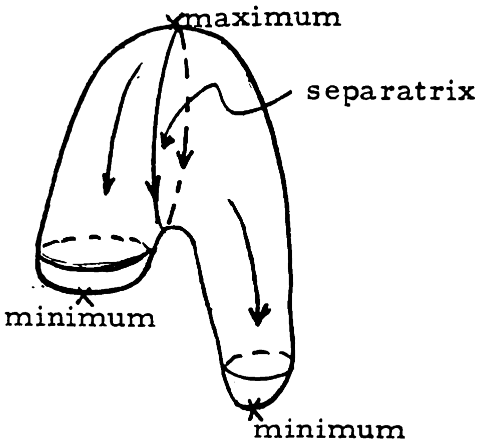
This partition is in fact a C-W complex. (For details, see Milnor's book on Morse theory.)

If i_k = the number of critical points of index k, then it can be shown that

$$i_k \geq b_k(M^n) \quad , \quad \text{(the Morse inequalities);}$$

i.e., i_k is greater than any k-dimensional betti-number of M^n. Consequently, the number of critical points has a lower bound determined by the topological structure of the manifold. This fact has interesting applications in mechanics, since a non-degenerate critical point corresponds to an equilibrium position. However, not all the non-degenerate critical points correspond to stable equilibrium positions -- in fact, it is possible to leave the critical point along an edge of the corresponding cell unless the point is a minimum.

In dimension 3, there may be several minima. Each determines a "basin"; the manifold minus the separatrices between adjacent basins is globally partitioned into basins. In general, we can't expect the separatrices to behave nicely, as will be shown shortly.



Suppose that 0 is a minimum with value 0 and suppose further that this is a quadratic minimum, i.e., if

$$z = \varphi_2(x, y) + \varphi_3(x, y) + \ldots$$

is the Taylor expansion near 0, then there exists an orthogonal change of coordinates such that

$$\varphi_2(x, y) = Ax^2 + By^2,$$

where $A > B > 0$. Consider

$$-\text{grad } z = X \frac{\partial}{\partial x} + Y \frac{\partial}{\partial y},$$

where $\varphi = \varphi_3 + \ldots$ and

$$X = -Ax - \frac{\partial \varphi}{\partial x}(x, y) \ldots$$

$$Y = -By - \frac{\partial \varphi}{\partial y}(x, y)\ldots .$$

<u>Claim:</u> The trajectories h_t for $-\text{grad } z$ are <u>contracting</u> near 0, i.e., $|h_t| \to 0$ as $t \to \infty$, for trajectories h_t near 0.

<u>Proof.</u> Let $\rho^2 = x^2 + y^2 = |h_t|^2$

$$\frac{d}{dt}(\rho^2) = 2x(-Ax - \frac{\partial \varphi}{\partial x}) + 2y(-By - \frac{\partial \varphi}{\partial y})$$

$$= -2(Ax^2 + By^2) - 2(x \frac{\partial \varphi}{\partial x} + y \frac{\partial \varphi}{\partial y}).$$

Now $|\frac{\partial \varphi}{\partial x}| < M\rho^2$, $|\frac{\partial \varphi}{\partial y}| < M\rho^2$, so the second term in the expression for $\frac{d}{dt}(\rho^2)$ is dominated by ρ^3. But the first term dominates ρ^2. Hence for ρ small,

$$Ax^2 + By^2 > | x \frac{\partial \varphi}{\partial y} + y \frac{\partial \varphi}{\partial y} | ;$$

so

$$\frac{d}{dt}(\rho^2) < 0.$$

An alternative way of looking at the problem would be to consider the inner product between the vector field at a point $h_t = (x, y)$ and the vector from 0 to (x, y). This is just

$$xX + yY = -(Ax^2 + By^2) - (x \frac{\partial \varphi}{\partial x} + y \frac{\partial \varphi}{\partial y}).$$

The above reasoning shows that this inner product is negative for $\rho^2 = x^2 + y^2$ small; hence the angle between the two vectors is between $\pi/2$ and $3\pi/2$, so the vector field always enters any circle of sufficiently small radius ρ.

<u>Claim:</u> All trajectories except the one along the x-axis arrive toward the origin tangential to the y-axis.

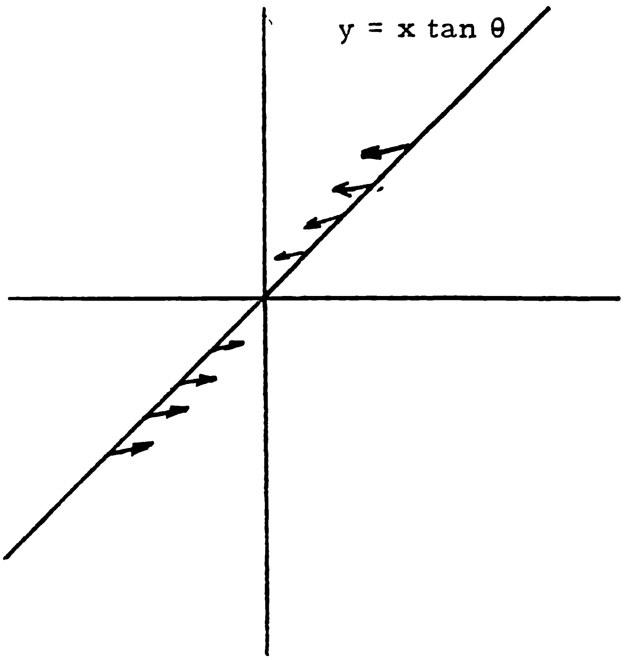
<u>Proof.</u> Consider how the vector field acts along a line $y = x\tan\theta$ through 0:

$$\frac{Y}{X} = \frac{-By - \frac{\partial\varphi}{\partial y}}{-Ax - \frac{\partial\varphi}{\partial x}}$$

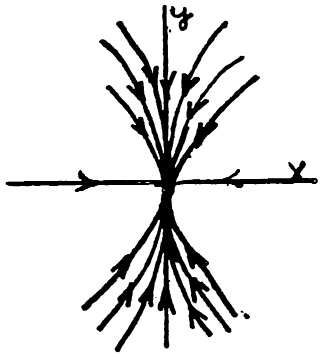
$$= \frac{-B\rho\sin\theta - \frac{\partial\varphi}{\partial y}(x,y)}{-A\rho\sin\theta - \frac{\partial\varphi}{\partial x}(x,y)}$$

$$= \frac{-B\sin\theta + \varepsilon(\rho)}{-A\cos\theta}$$

for ρ small, where $\varepsilon(\rho) \to 0$ as $\rho \to 0$. For ρ small enough the vector field is pointed at an angle of approximately $\arctan(\frac{B}{A}\tan\theta)$ with the x-axis. Since $B < A$, this angle is smaller than θ. Thus for small ρ, the vector field enters the angle formed by this line and the y-axis. Since B/A is a constant, every trajectory near 0 which is not along the



x-axis must enter every angle containing the y-axis, hence arrives tangential to it.

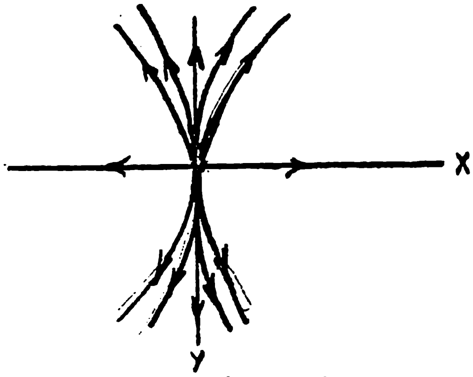


This last result says that the trajectories near a quadratic minimum look as pictured.

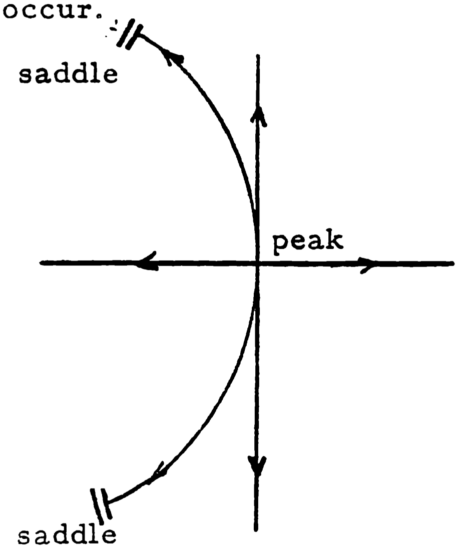
The same analysis may be applied to a quadratic maximum, i.e., a critical point where

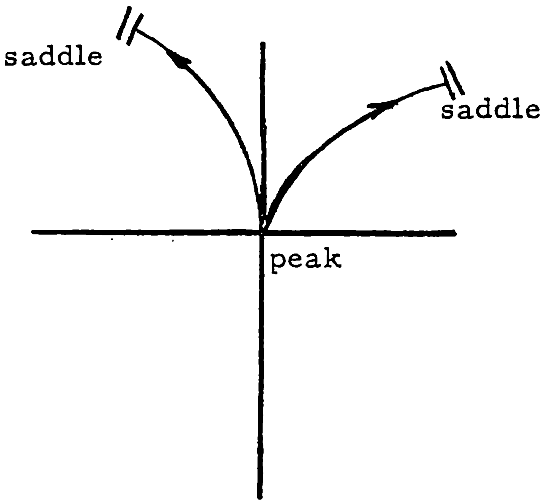
$$\varphi^2 = Ax^2 + By^2$$

and $A, B < 0$. If $A < B$, then almost all trajectories leave the maximum tangent to the y-axis.



In particular, if there are two separatrices going from the same quadratic maximum to saddle points, the two situations pictured might occur.





In the first case, the two separatrices will together form a smooth curve. In the second case, a cusp is formed, so the separatrices do not form a submanifold.

From our previous analysis, we see that the question of whether there is a trajectory from a given maximum to a given minimum is decidable by a finite procedure -- just look at the components corresponding to each of the critical points. This is not true in a system more complicated than a gradient system. In certain cases, the decidability of this question depends on arithmetic properties of the coordinates; in some cases, the situation is, practically speaking, indeterminate.

42. Critical points in the degenerate situation.

So far a few results are known which apply to degenerate as well as non-degenerate singular points, but knowledge of the degenerate case remains sketchy compared to the non-degenerate case.

The theory of the Lusternik-Schnirelmann category states that there is a lower bound for the number $i = i(M)$ of all critical points of a manifold:[4)]

$$c(M^n) < i(M^n).$$

It has also been shown that on a manifold M^n of dimension n there is always a function with only $n+1$ critical points.[5)] This is

---

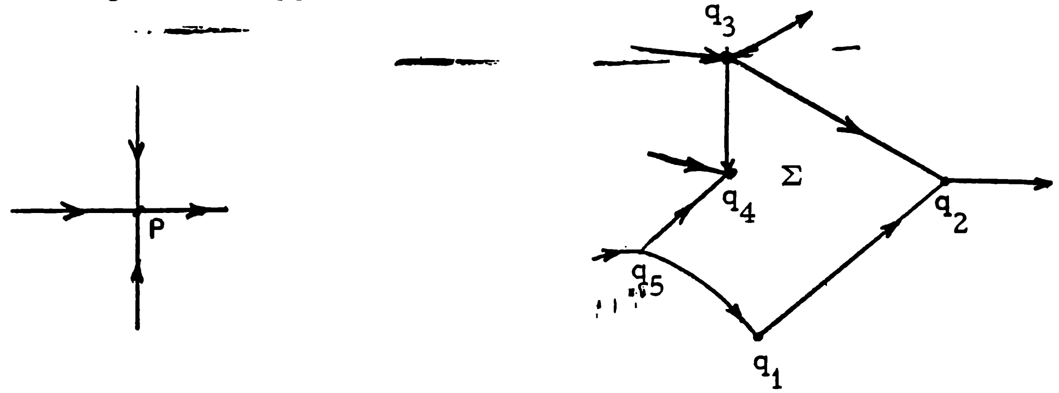
4) Liusternik - The topology of calculus of variations in the large, Translation of math. monographs, Amer. Math. Soc., 1966.

5) Èl'sgol'c, L. È. - Estimation of the number of critical points, Uspeki Matem. Nauk 5 n. 6(40), p. 52-87, 1950 (Russian) Math. Review p. 721, vol. 12, 1951.

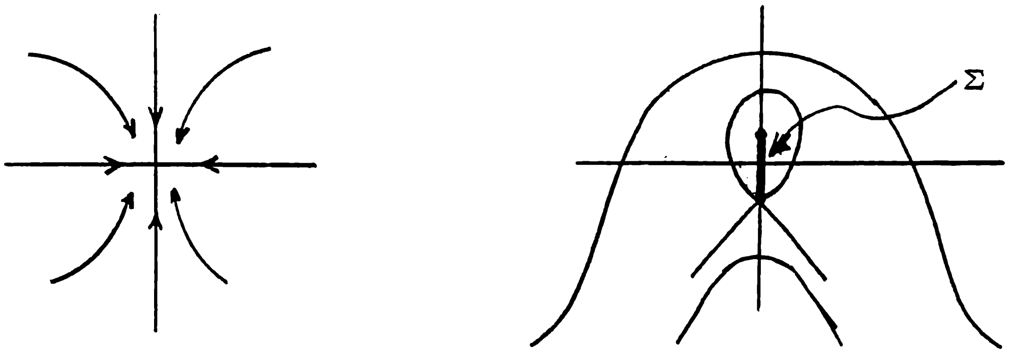
similar to the result that M^n can be covered by n+1 charts! For example, in the case of a Riemannian manifold, we can do this in terms of the cut locus of a point $0 \in M$: For each $m \in M$, consider all geodesics joining 0 to m. There is one or more of shortest length. Let K_o be the set of points of M which have at least two such geodesics of shortest length. K_o is the cut locus of 0. $M-K_o$ is an open cell which is everywhere dense. If $x_o, \ldots, x_n$ are n+1 points of M in "general position", then $\bigcap K_{x_i} = \emptyset$, so the n+1 open cells $M-K_{x_i}$ cover M.

In studying a degenerate critical point p, we may use the following method. Perturb F to get a new function $F + \delta F$. By the Morse theorem, we may assume the new function has only non-degenerate critical points.. Thus we have a finite number of non-degenerate critical points of the new function, corresponding to the original degenerate critical point, to study.

We can now state a conjecture as to the nature of degenerate critical points. Suppose the perturbation is also such that the gradient trajectories to and from the critical points always intersect transversally. The trajectories between two of the relevant critical points form a set Σ. The conjecture asserts that Σ is contractible and that the degenerate case arises by collapsing Σ, (i.e., that the sets Γ_p^- and Γ_p^+ of trajectories entering and leaving p are determined by the sets $\Gamma^-_{q_i}$ and $\Gamma^+_{q_i}$ for the q_i, perhaps by taking the disjoint union of the $\Gamma^{\pm}_{q_i}$ and identifying the points q_i). Then degenerate points could be characterized by the number of points and their indices in the non-degenerate approximation.



For example, let $f = x^3 + y^3$ in dimension 2. Perturb f to get $x^3 + y^3 - \lambda y$. For $\lambda > 0$, there will be two critical points. The new function gives the diagram on the right, which collapses to give the diagram for f on the left.



CHAPTER VI. FIRST ORDER PARTIAL DIFFERENTIAL EQUATIONS

43. The Hamilton-Jacobi Equations.

In the first part, in §26 (page 118) we had occasion to consider the Hamilton-Jacobi Partial Differential Equation for a Hamiltonian function $H(q^1, \ldots, q^n, p_1, \ldots, p_n)$, and to show how suitable solutions of this equation determined the trajectories of the corresponding mechanical system. We now return to this topic, for the case when H also depends on the time (time-dependent Hamiltonian).

Let C be a configuration space, $M = T^* C$ the corresponding phase space. In the metric spaces $\mathbb{R} \times C$ and $\mathbb{R} \times M$ the projection onto $\mathbb{R}$ will be written as t, so that $t: \mathbb{R} \times M \to \mathbb{R}$ is the "time coordinate". Consider functions

$$H: \mathbb{R} \times M \to \mathbb{R} \quad , \quad S: \mathbb{R} \times C \to \mathbb{R}$$

(H is the hamiltonian, S the "entropy"). The function $S_t = S(t, -): C \to \mathbb{R}$ for each t determines a 1-form

$$dS_t: C \to M$$

(a cross section of the cotangent bundle). The corresponding equation

$$\frac{\partial S}{\partial t} + H(t, dS_t) = 0$$

is the (time-dependent) Hamilton-Jacobi equation.

Given the function S, each curve $b: I \to C$ lifts to a curve $c(t) = dS_t \circ b(t)$ in $T^* C$. The equations (for $q^1, \ldots, q^n$ coordinates in C)

$$\frac{dq^i b}{dt} = \frac{\partial H}{\partial p_i}(t, dS_t b) \ , \quad i = 1, \ldots, n$$

will be called the first Hamilton equations, as they are the first half of the 2n Hamilton equations for C. The role of the Hamilton-Jacobi equation may be expressed by the following theorem (formulated by George Mackey):

Theorem. Let S be a solution of the Hamilton-Jacobi partial differential equation. Then if the path b in C satisfies the first Hamilton equation, the lifted path c satisfies Hamilton's equation. Conversely, let $S: \mathbb{R} \times C \to \mathbb{R}$ be a smooth function such that every b satisfying the first Hamilton equations yields a c satisfying (all of) Hamilton's equations. Then there is a smooth function $\psi: \mathbb{R} \to \mathbb{R}$ such that $S - \psi \circ t$ satisfies the Hamilton-Jacobi partial differential equation.

Proof. First we calculate

$$\begin{aligned}\frac{dp_i c}{dt} &= -\frac{d}{dt}\frac{\partial S}{\partial q^i}(t, q^1 b, \ldots, q^n b) \\ &= \frac{\partial^2 S}{\partial t\, \partial q^i} + \sum_j \frac{\partial^2 S}{\partial q^i \partial q^j}\frac{dq^j b}{dt} .\end{aligned}$$

By assumption, the first Hamilton equation holds. Hence

$$(1) \qquad \frac{dp_i c}{dt} = \frac{\partial^2 S}{\partial t\, \partial q^i} + \sum_j \frac{\partial^2 S}{\partial q^i \partial q^j}\frac{\partial H}{\partial p_j} .$$

(Here each second partial of S has the evident arguments.)

On the other hand, applying $\frac{\partial}{\partial q^i}$ to Hamilton-Jacobi gives the following "derived H-J equation:

$$(2) \qquad 0 = \frac{\partial^2 S}{\partial t\, \partial q^i} + \frac{\partial H}{\partial q^i}(t, dS_t) + \sum \frac{\partial H}{\partial p_j} \frac{\partial^2 S}{\partial q^j \partial q^i} \quad .$$

Now $\frac{\partial^2 S}{\partial q^j \partial q^i} = \frac{\partial^2 S}{\partial q^i \partial q^j}$. Hence subtracting the last two equations gives (with suitable arguments)

$$(3) \qquad \frac{\partial p_i c}{dt} = \frac{\partial H}{\partial q^i} \quad .$$

This is the second half of Hamilton's equations.

In the last equation everything is to be regarded as a function of t. The "suitable arguments" required to make this the case are indicated without ambiguity by the diagram of the functions involved.

$$\begin{tikzcd}
 & I \times T^*C \arrow[r, "H"] & \mathbb{R} \\
I \arrow[ur, "(1,c)"] \arrow[r, "(1,b)"'] & I \times C \arrow[u, "dS"'] &
\end{tikzcd}$$

It should be possible to make a systematic use of such mapping diagrams to indicate which (composite) arguments we intended in equations (such as the Hamilton and Hamilton-Jacobi equations).

Now consider the converse part of the theorem. Take a solution b of the first Hamilton equations; then equations (1) above hold. By hypothesis (1) implies (3); subtracting, (2) holds along b. But by the existence theorem for ordinary differential equations, there is a solution b through each point of $I \times C$; hence (2) holds at any point $(t, q^1, \ldots, q^n)$. But (2) states that

$$\frac{\partial}{\partial q^i}(HJ(S)) = 0 \quad , \quad i = 1, \ldots, n,$$

where $HJ(S)$ denotes the left-hand side of the Hamilton-Jacobi equations.
depends only on t, so
Therefore, there is a smooth function $\theta : \mathbb{R} \to \mathbb{R}$ with $HJ(S) = \theta \circ t : I \times C \to \mathbb{R}$. Take a function ψ such that $\frac{d\psi}{dt} = \theta$. Then it is not hard to see that

$$HJ(S - \psi) = 0.$$

This gives the conclusion of the theorem.

44. Transformation to Equilibrium.

Now let Y be an n-dimensional manifold, and suppose $S : \mathbb{R} \times C \times Y \to \mathbb{R}$ is a function such that everywhere

$$\det \left| \frac{\partial^2 S}{\partial q^i \partial y^j} \right| \neq 0$$

where q^i are coordinates for C and the y^j are coordinates for Y. Let $\Theta : R \times C \times Y \to R \times T^* C$ be given by

$$p_i \Theta = \frac{\partial S}{\partial q^i}$$

$$t\,\Theta = t$$

$$q^i \Theta = q^i \quad , \quad i = 1, \ldots, n.$$

Thus the assumption on S is equivalent to saying that Θ is regular everywhere. Consider also the mapping $\Psi : \mathbb{R} \times C \times Y \to \mathbb{R} \times T^* Y$ given by

$$t\Psi = t$$

$$y^i\Psi = y^i$$

$$x_i\Psi = \frac{\partial S}{\partial y_i} \quad , \quad i = 1, \ldots, n,$$

where the y^i are coordinates on Y and the x_i the corresponding (momentum) coordinates on T^*Y. This map is also regular everywhere. Hence we have the diagram

$$\begin{tikzcd} R \times T^*C & & \mathbb{R} \times T^*Y \arrow[ll, dotted, "\chi"'] \\ & \mathbb{R} \times C \times Y = N \arrow[ul, "\Theta"] \arrow[ur, "\Psi"'] & \end{tikzcd} ,$$

and locally at least there is a map χ from one time dependent phase space to another (compare §26).

Theorem. Take $H: \mathbb{R} \times T^*C \to \mathbb{R}$. For each point $a \in Y$, suppose that $S: N \to \mathbb{R}$ satisfies the H-J partial differential equation for the function H. Let $c: \mathbb{R} \to \mathbb{R} \times T^*Y$ be a curve of the form $c(t) = (t, \text{const.})$. Then the curve χc satisfies the Hamilton equations for H on $\mathbb{R} \times T^*C$.

Proof. It will be more convenient to look at everything in N rather than in $\mathbb{R} \times T^*C$. To do this we pull everything back locally by Θ^{-1}. Thus we are interested in the curve $\Psi^{-1}c$ and the function $H\Theta$.

Let x_i denote the coordinates as above. Then $x_i = \dfrac{\partial S}{\partial y_i}$ on N. Taking $\dfrac{d}{dt}$ of this, we get

$$0 = \frac{\partial^2 S}{\partial y^i \partial t} + \sum \frac{\partial^2 S}{\partial y^i \partial q^j} \frac{dq^j}{dt} + \sum \frac{\partial^2 S}{\partial y^i \partial y^j} \frac{\partial y^i}{\partial t} .$$

The third sum vanishes because $\frac{\partial y^i}{\partial t} = 0$. Since we are assuming

$$\frac{\partial S}{\partial t} + H = 0$$

on N (more precisely, we should write $H\Theta$ instead of H), applying $\frac{\partial}{\partial y^i}$ yields

$$\frac{\partial^2 S}{\partial t \partial y^i} + \sum \frac{\partial H}{\partial p^j} \frac{\partial^2 S}{\partial q^i \partial y^j} = 0,$$

which holds on the curve $\Psi^{-1}c$. Hence, again on the curve,

$$\sum \frac{\partial^2 S}{\partial q^i \partial y^j} \left(\frac{dq^j}{dt} - \frac{\partial H}{\partial p_j}\right) = 0.$$

Since the determinant of $\left(\frac{\partial^2 S}{\partial q^i \partial y^j}\right)$ was assumed to be non-zero, this means that

$$\frac{dq^j}{dt} = \frac{\partial H}{\partial p_j}$$

holds for all j. This is the first Hamilton equation. By the previous theorem, we get the remaining half of the Hamilton equations.

For fixed t_o, the submanifolds of N of the form $t_o \times C \times Y$ have a symplectic form given by

$$\sum \frac{\partial^2 S}{\partial q^i \partial y^j} dq^i \wedge dy^j .$$

By the theorem proved in § 26 of Part I, the functions Θ and Ψ are symplectic mappings, whence the usual symplectic structures are taken on T^*C and T^*Y.

In the theorem just proved, the trajectories c in $\mathbb{R} \times T^*Y$ are constant in T^*Y. Hence one says that the map χ^{-1} of the theorem transforms the Hamiltonian H "to equilibrium".

45. Characteristics.

The previous results indicate a close relation between the Hamilton-Jacobi equation, a partial differential equation, and Hamilton's equations, a system of ordinary first order differential equations. This is a special case of the theory which relates a first order partial differential equation to its characteristics, which are solutions of a corresponding system of ordinary first order differential equations.

Sources: a) Courant and Hilbert, Methods of Mathematical Physics, II

b) Caratheodory, Calculus of Variations and PDE's of 1^{st} order, Part I.

The case of the arbitrary first order equation will be reached in stages. We first consider the linear case, involving the following functions on a configuration space C:

$$\mathbb{R} \xrightarrow{c} C \xrightarrow[q^1, \ldots, q^n]{u} \mathbb{R},$$

$$(1) \qquad \sum_{i=1}^{n} a_i \frac{\partial u}{\partial q^i} = bu + d \quad , \qquad \begin{cases} a_i: C \to \mathbb{R}, \\ b: C \to \mathbb{R}, \\ d: C \to \mathbb{R} \ . \end{cases}$$

Equation (1) for the linear case has a_i, b, d functions of position in C. The a_i determine a vector field $X = \sum_{i=1} \frac{\partial}{\partial q^i}$ on C which appears in the following coordinate independent form of (1): $L_X u = bu + d$. Call c a characteristic curve of the PDE (1) when

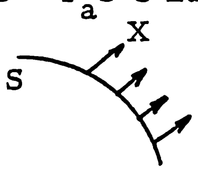
$$(2) \qquad \frac{d(q^i c)}{dt} = a^i \quad , \quad i = 1, \ldots, n.$$

Thus the characteristics are the trajectories of the vector field X. In view of the definition of X, this equation can be written as

$$\frac{duc}{dt} = bu + d. \tag{2'}$$

First suppose that $b = d = 0$. Then if the function u satisfies the PDE (1), it is constant along the characteristics of (1). More generally, for any b and d, the equation states that the values of u along a characteristic are determined by the value at any one (initial) point there. This suggests that we can obtain a solution u by taking initial values along a suitable set S, and then prolong these values by solving (2').

More explicitly, find a submanifold S of dimension n-1 transverse to X (i.e., with $T_aC = T_aS \oplus \mathbb{R}X(a)$ at each point a of S). According



to the basic theorem on the integration of (smooth) vector fields, the trajectories of X through S cover some neighborhood of S, determining on some neighborhood of any $\overset{o}{S}$ a unique function u which agrees on $\overset{o}{S}$ with some chosen $u_o : \overset{o}{S} \to \mathbb{R}$, and which satisfies (2) along characteristics (trajectories of X) on the neighborhood. Here $\overset{o}{S}$ is an open submanifold of S with compact closure in S. In local coordinates it is immediate that, for smooth u_o, the function u is smooth and satisfies (1). So we have found a local solutuon.

Next we consider a first order P.D.E. in an unknown u, of the form

$$(1) \qquad \sum_{i=1}^{n} a_i(u, q^1, \ldots, q^n) \frac{\partial u}{\partial q^i} = b(u, q^1, \ldots, q^n) \quad , \quad u = u(q^1, \ldots, q^n).$$

This is linear in all the partial derivatives of u, but not in u itself, hence is said to be quasilinear. We interpret the $q^1, \ldots, q^n$ as coordinates in an n-dimensional configuration space C, so that $u: C \to \mathbb{R}$. The equation thus has the form

$$(1') \qquad \sum_{i=1} a_i \frac{\partial u}{\partial q^i} = b, \qquad \begin{tikzcd} \mathbb{R} \times C \arrow[r, bend left, "a_1"] \arrow[r, "a_n"] \arrow[r, bend right, "b"'] & \mathbb{R}, \end{tikzcd}$$

for given coefficient functions a_i and b.

We plan to reduce this to the previous case for a linear P.D.E. in an unknown $v: \mathbb{R} \times C \longrightarrow \mathbb{R}$ in one more variable, constructing the function u via its graph $\hat{u}: C \xrightarrow{\hat{u}} \mathbb{R} \times C$. Let $r: \mathbb{R} \times C \to \mathbb{R}$ be the projection on the first coordinate. We introduce a function $v: \mathbb{R} \times C \xrightarrow{v} \mathbb{R}$ defined by $v = u - r$. Now $\frac{\partial v}{\partial r} = -1$ and $\frac{\partial v}{\partial q^i} = \frac{\partial u}{\partial q^i}$ (while on hypersurfaces $v = 0$ we will have $a_i(u, q^i) = a_i(r, q)$ and $b(u, q) = b(r, q)$) so that (1) becomes

$$0 = \sum_{i=1}^{n} a_i \frac{\partial u}{\partial q^i} - b = \sum_{i=1}^{n} a_i \frac{\partial v}{\partial q_i} + b \frac{\partial v}{\partial r} = \Big(\sum_{i=1}^{n} a_i \frac{\partial}{\partial q_i} + b \frac{\partial}{\partial r} \Big) v.$$

This is a homogeneous linear P.D.E. in $v: \mathbb{R} \times C \to \mathbb{R}$. Its characteristics are thus given by a suitable vector field $\hat{X}$. Indeed we now define the vector field $\hat{X}$ on $\mathbb{R} \times C$ by $\hat{X} = \sum_{i=1}^{n} a_i \frac{\partial}{\partial q^i} + b \frac{\partial}{\partial r}$, then (on the hypersurface $v = 0$) the equation (1) becomes:

$$0 = \left(\sum a_i \frac{\partial}{\partial q^i} + b\frac{\partial}{\partial r}\right)v = \langle dv, \hat{X}\rangle .$$

So $\hat{X}$ at each point is in the tangent plane to the hypersurface at that point, and the trajectories of $\hat{X}$ remain in the hypersurface. Moreover:

<u>Proposition.</u> For $x_o \in \mathbb{R}\times C$, let $v: \mathbb{R}\times C \to \mathbb{R}$ with

$$\begin{cases} \langle dv, \hat{X}\rangle = 0 \\ \left.\dfrac{\partial v}{\partial r}\right|_{x_o} \neq 0, \ v(x_o) = 0. \end{cases}$$

Then the function $u: N_{x_o} \to \mathbb{R}$ such that $v(q^1, \ldots, q^n, u) = 0$, constructed for a suitable neighborhood $N_{x_o} \subset C$ via the implicit function theorem, is a solution of the P.D.E. (1).

<u>Proof.</u> By the construction of u,

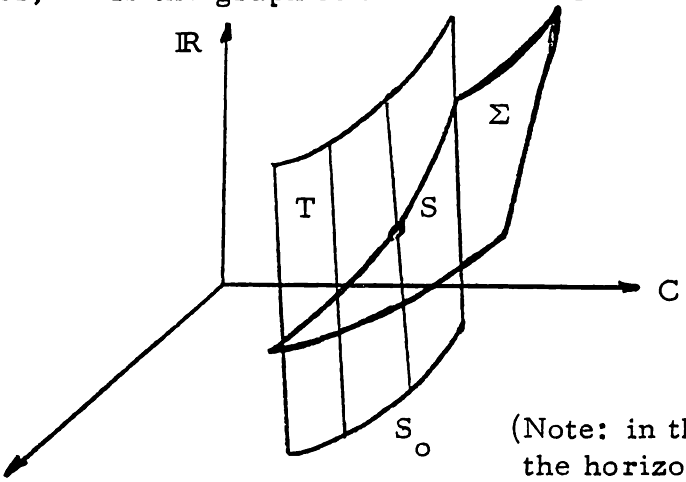
$$0 = \frac{\partial}{\partial q_i}(v \circ \hat{u}) \qquad\qquad C \xrightarrow{\hat{u}} \mathbb{R}\times C \xrightarrow{v} \mathbb{R},$$

$$= \frac{\partial v}{\partial q^i} + \frac{\partial v}{\partial r}\frac{\partial u}{\partial q^i} \quad \text{for each } i\text{, so that we have}$$

$$0 = \langle dv, \hat{X}\rangle = \sum_{i=1}^{n} a_i \frac{\partial v}{\partial q_i} + b\frac{\partial v}{\partial r} = \sum_{i=1}^{n} -a_i \frac{\partial v}{\partial r}\frac{\partial u}{\partial q_i} + b\frac{\partial v}{\partial r} .$$

Therefore for $\frac{\partial v}{\partial r} \neq 0$, u will be a solution of (1), q.e.d.

Let S_o be a submanifold of C of dimension $n-1$, and let $u_o: S_o \to \mathbb{R}$ be a smooth function. Through S_o in $\mathbb{R}\times C$ pass the vertical hypersurface $T = \{(r,x) \mid r \in \mathbb{R},\ x \in S_o\}$. Define $v_o: T \to \mathbb{R}$ by $v_o(r,x) = u_o(x) - r$. Suppose that the characteristic field $\hat{X}$ is transverse to T at a point x_o of the $n-1$ dimensional submanifold S on which $v_o = 0$. Then, it is immediate in local coordinates (see the figure below) that the trajectories of $\hat{X}$ through some neighborhood in S of x_o determine an n-dimensional submanifold Σ of $\mathbb{R} \times C$ (locally unique). Moreover, Σ is the graph of the desired solution u. For the



(Note: in the figure, C is the horizontal plane and the $\mathbb{R}$ axis is the vertical.)

fact that $\hat{X}$ is transversal to T at x_o implies that the function $v(x, t) = v_o(x)$ (where $x \in T$ and $\hat{X} = \frac{\partial}{\partial t}$) is well-defined on a neighborhood of x_o in $\mathbb{R} \times C$ and satisfies the conditions of the previous proposition $\left(\frac{\partial v}{\partial r} \Big|_{x_o} = -1 \neq 0 \right)$.

Any point of $\mathbb{R} \times C$ at which $\hat{X}$ is non-vertical lies on the graph of such solutions. Explicitly, the hypersurface $S_o \subset C$ may be described as the locus where some smooth function $f: C \to \mathbb{R}$ is constant (i.e., as a level hypersurface of f). Then the vertical hypersurface T is

$$T = \{(r, y) \mid f(y) = f(y_o)\}$$

for y_o a fixed and y any point of C.

We have proved

Theorem. For smooth functions $a_1, \ldots, a_n, b: \mathbb{R} \times C \to \mathbb{R}$ let $S_o \subset C$ be defined by a point $y_o \in C$ and a smooth function $f: C \to \mathbb{R}$ as

$$S_o = \{y \mid y \in C \text{ and } f(y) = f(y_o)\}.$$

If $u_o: S_o \to \mathbb{R}$ is a smooth function satisfying the "transversality" condition

$$\sum a_i(u_o(y_o), y_o) \frac{\partial f}{\partial q^i} \neq 0,$$

then in some neighborhood of y_o there is a unique solution u of the P.D.E. $\sum a_i \frac{\partial u}{\partial q^i} = b$ with values u_o on S_o.

46. The General First Order P.D.E.

Consider an equation

$$E(u, q^1, \ldots, q^n, \frac{\partial u}{\partial q^1}, \ldots, \frac{\partial u}{\partial q^n}) = 0 \tag{1"}$$

in an unknown function $u: C \to \mathbb{R}$, where $q^1, \ldots, q^n$ are local coordinates in the configuration space C. We can regard the "equation" E as a given (smooth) function $E: \mathbb{R} \times T^*C \to \mathbb{R}$. The differential du is a function $C \to T^*C$; we also have $d'u: C \to \mathbb{R} \times T^*C$ given locally as

$$(q^1, \ldots, q^n) \to (u, q^1, \ldots, q^n, \frac{\partial u}{\partial q^1}, \ldots, \frac{\partial u}{\partial q^n}).$$

Thus the equation (1") becomes $E \circ d'u = 0$. If $r: \mathbb{R} \times T^*C \to \mathbb{R}$ is the projection on the first factor, then $\frac{\partial}{\partial q^i}$ applied to (1") yields the i^{th} derived P.D.E.

$$\frac{\partial E}{\partial r} \frac{\partial u}{\partial q^i} + \frac{\partial E}{\partial q^i} + \sum_{j=1}^{n} \frac{\partial E}{\partial p_j} \frac{\partial}{\partial q^i}\left(\frac{\partial u}{\partial q^j}\right) = 0, \quad i = 1, \ldots, n.$$

Interchanging the order of partial derivatives, this is

$$(2) \qquad \frac{\partial E}{\partial r}\frac{\partial u}{\partial q^i} + \frac{\partial E}{\partial q^i} + \sum_{j=1}^{n} \frac{\partial E}{\partial p_j}\frac{\partial}{\partial q^j}\left(\frac{\partial u}{\partial q^i}\right) = 0.$$

By way of motivation, observe that the i^{th} equation of (2) may now be regarded as a quasilinear P.D.E. in the unknown $p_i = \frac{\partial u}{\partial q^i}$. The characteristics of this quasilinear equation are then given by the (n+1)-dimensional vector field (see above):

$$\hat{X} = \sum_j \frac{\partial E}{\partial p_j}\frac{\partial}{\partial q^j} + \left(-\frac{\partial E}{\partial q^i} - \frac{\partial E}{\partial r} p_i\right)\frac{\partial}{\partial r}.$$

The differential equations of these characteristics are then the n+1 equations

$$\frac{dq^j}{dt} = \frac{\partial E}{\partial p_j}, \qquad j = 1,\ldots,n,$$

$$\frac{dp_i}{dt} = -\frac{\partial E}{\partial q^i} - \frac{\partial E}{\partial r} p_i.$$

As i varies, the first n equations are the same. Note also that this reduces to Hamilton's equations when $\frac{\partial E}{\partial r} = 0$.

Our actual interpretation of (2) will be slightly different, as an equation on $\mathbb{R} \times T^*C$ itself, with characteristics in $\mathbb{R} \times T^*C$ which are solution curves of:

This [illegible] not [illegible]

$$(3) \qquad \begin{aligned} &\frac{dq^j c}{dt} = \frac{\partial E}{\partial p_j}, \quad \frac{dp_i c}{dt} = -\frac{\partial E}{\partial q_i} - \frac{\partial E}{\partial r} p_i, \\ &\frac{dr}{dt} = \sum_j p_j \frac{\partial E}{\partial p_j}. \end{aligned}$$

The third set of equations is included since $E = E(r,q,p)$ is constant along trajectories of the vector field

$$(3') \quad X_E = \sum \frac{\partial E}{\partial p_j} \frac{\partial}{\partial \cancel{p}^j} + \sum_j \left(- \frac{\partial E}{\partial q^j} - p_j \frac{\partial E}{\partial r} \right) \frac{\partial}{\partial p_j} + \sum p_j \frac{\partial E}{\partial p_j} \frac{\partial}{\partial r}$$

on $T^{*}C \times \mathbb{R}$. For

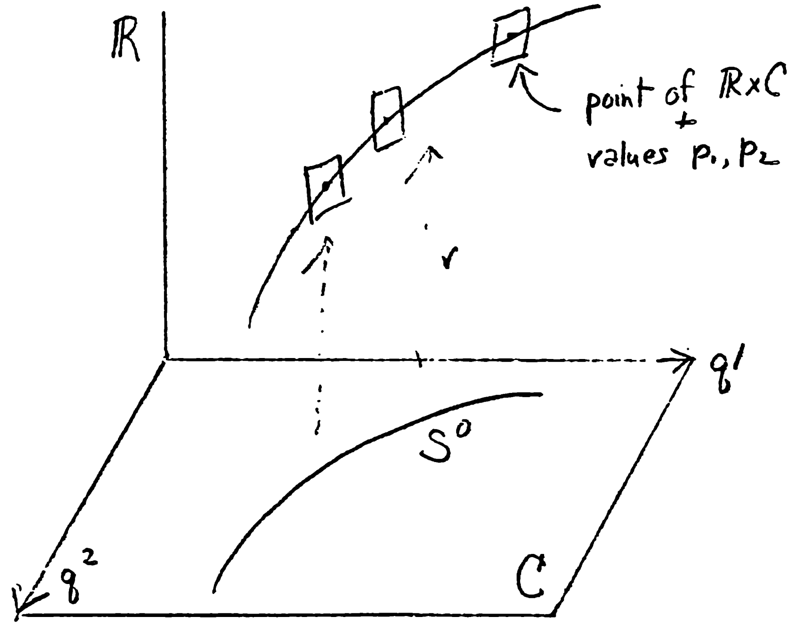
$$L_{X_E} E = \sum_j \frac{\partial E}{\partial p_j} \frac{\partial E}{\partial \cancel{p}^j} + \left(\sum_j - \frac{\partial E}{\partial q^j} \frac{\partial E}{\partial p_j} - \sum p_j \frac{\partial E}{\partial r} \frac{\partial E}{\partial p_j} \right) + \sum p_j \frac{\partial E}{\partial p_j} \frac{\partial E}{\partial r} = 0 .$$

For our previous cases, the last equation disappears, while the middle equations all collapse to the equation for r (relabeled $p_i c$) from the quasilinear case, giving trajectories "parallel" to those of the earlier cases. We state our existence theorem in the form:

<u>Theorem</u>: Given in C a compact submanifold S_o of dimension $n-1$ and initial values u_o of u on S_o such that a certain determinant (which appears as (5) below) does not vanish, then there exists an open set $U \supset S_o$ and a smooth function $u: U \to \mathbb{R}$ which satisfies E on U and agrees on S_o with u_o.

It will be clear from the proof that the conditions on the initial surface could be taken as before, and that the determinant condition corresponds to our previous transversality condition, with no loss of applicability.

<u>Proof</u>. We operate in $\mathbb{R} \times T^{*}C$, where we already have defined in (3) the characteristic curves. In the submanifold T of dimension $2n$ above S_o, we distinguish a surface S which will correspond to $\hat{U} \cap T$. This submanifold (diffeomorphic to S_o) with local coordinates $x^1, \ldots, x^{n-1}$ embedded by $v: S_o \xrightarrow{v} S \subsetneq T$, should have



Proof. In the configuration space C we have local coordinates $q^1, \ldots, q^n$, an initial manifold $S_o \subset C$ of dimension 1 and initial values $u_o : S_o \to \mathbb{R}$. On $\mathbb{R} \times T^*C$ we have $2n+1$ local coordinates $r, q^1, \ldots, q^n, p_1, \ldots, p_n$. We can define a map $v: S_o \to \mathbb{R} \times T^*C$; this amounts to choosing "initial" values of r, q^i and p_j along S_o. Specifically we make $r \circ v = u_o, q^i \circ v = q^i$ and then we choose $p_1, \ldots, p_n$ so that

$$E = 0 \quad , \quad dr - \sum p_i dq^i = 0$$

both along S_o. The last condition on dr may be written in terms of local parameters $x^1, \ldots, x^{n-1}$ on the $(n-1)$-manifold S_o as

(4) $$0 = du_o - \sum_{i=1}^{n} p_i dq^i = \sum_k \left(\frac{\partial u}{\partial x^k} - \sum_i p_i \frac{\partial q^i}{\partial x^k} \right) dx^k .$$

Hence $p_1, \ldots, p_n$ are determined uniquely along S_o if

$$
(5) \qquad \begin{vmatrix} \frac{\partial E}{\partial p_1} & \frac{\partial E}{\partial p_2} & \cdots & \frac{\partial E}{\partial p_n} \\ \frac{\partial q^1}{\partial x^1} & & \cdots & \frac{\partial q^n}{\partial x^1} \\ \vdots & & & \\ \frac{\partial q^1}{\partial x^{n-1}} & & \cdots & \frac{\partial q^n}{\partial x^{n-1}} \end{vmatrix} \neq 0 .
$$

This condition may be readily satisfied, since we can assume that the first row is nowhere zero. (This amounts to assuming that the given partial differential equation effectively involves at least one of the partial derivatives $p_i = \frac{\partial u}{\partial q^i}$). Given such a first row, the submanifold S_o can be chosen to make (5) hold; for example, if $\frac{\partial E}{\partial p_n} \neq 0$ we can choose the submanifold S_o given locally by the equation $q^n = 0$, with local coordinates $x^1 = q^1, \ldots, x^{n-1} = q^{n-1}$; then the determinant (5) is simply $(-1)^n \frac{\partial E}{\partial p_n}$. Indeed, the condition (5) is then exactly the condition that S_o be transversal to the projection of X_E.

We now have $v: S_o \to \mathbb{R} \times T^*C$, with image an (n-1)-manifold S in $R \times T^*C$; moreover one can show S transversal to the characteristic vector field X_E. Therefore the trajectories of X_E through points of S fill up locally a manifold T of dimension n. Now $E = 0$ holds along S, so by the properties earlier established for characteristics it holds along T. In other words, T gives the graph of functions $u, q^1, \ldots, q^n, p_1, \ldots, p_n$ on C (or on a neighborhood of S_o in C) which satisfy $E(u, q^1, \ldots, q^n, p_1, \ldots, p_n) = 0$.

What remains to be verified is that $p_i = \frac{\partial u}{\partial q_i}$ for $i = 1, \ldots, n$ on this manifold T. If $\nu : T \to \mathbb{R} \times T^*C$ is the inclusion map, this amounts to showing that the induced 1-form $\theta = \nu^*(dr - \sum p_i dq^i)$ is zero on T. We may calculate θ in local coordinates $x^1, \ldots, x^{n-1}$ (on S_o) and t (the parameter along the trajectories of X_E) as

$$\theta = \sum_{k=1}^{n-1} \left(\frac{\partial r}{\partial x^k} - \sum p_i \frac{\partial q^i}{\partial x^k} \right) dx^k + \left(\frac{\partial r}{\partial t} - \sum p_i \frac{\partial q^i}{dt} \right) dt .$$

The last term is zero by the equation (3) for the characteristics. It thus remains to show that

$$D_k = \frac{\partial r}{\partial x^k} - \sum p_i \frac{\partial q^i}{\partial x^k} , \quad k = 1, \ldots, n-1$$

is zero. But $D_k = 0$ on S_o by the choice of the initial values of p_i, while a calculation with the equations (3) shows

$$\frac{\partial D_k}{\partial t} = \frac{\partial E}{\partial x^k} + \frac{\partial E}{\partial r} D_k = \frac{\partial E}{\partial r} D_k .$$

This is a linear first order differential equation for D_k as a function of the parameter t, with initial values zero on S_o. Hence (by the uniqueness of the solutions of such equations) $D_k = 0$, q.e.d..

47. <u>Contact Manifolds</u>. The use of the characteristic vector field X_E for the partial differential equation E raises the following question. For a symplectic manifold any two smooth functions f and g have a Poisson bracket given by

$$\{f, g\} = L_{X_f} g = -L_{X_g} f .$$

On the manifold $N = \mathbb{R} \times T^{*}C$, each smooth function $F: N \to \mathbb{R}$ determines (by characteristics, as in (3) above) a vector field X_F, and hence two such functions F and G have a "superbracket" defined as $X_F(G)$. We wish to examine the geometric structure producing this operation; it will turn out that this structure depends essentially on the 1-form $dr - \sum p_i dq^i$ used in the calculations to the last theorem.

Another approach is in terms of "elements". An "element" of the space $\mathbb{R} \times C$ is a *point* of this space plus a (non-vertical) *hyperplane* through this point; for example, if C has dimension 2 an element is just this: $\boxed{\cdot}$. In coordinates $r, q^1, \ldots, q^n$ any hyperplane through the origin has an equation $a_o r + a_1 q^1, \ldots, + a_n q^n = 0$ for suitable constants a_i; it is non-vertical precisely when $a_o \neq 0$, and in this case we may take $a_o = 1$ and write $r + a_1 q^1 + \ldots a_n q^n = 0$. Thus the hyperplane is determined by $a_1, \ldots, a_n$, which we now write as $p_1, \ldots, p_n$ (in case $n = 3$ they are the direction cosines of the normal to the hyperplane). Thus an *element* is given by coordinates $r, q^1, \ldots, q^n, p_1, \ldots, p_n$, and so is exactly a point in $\mathbb{R} \times T^{*}C$.

Take a curve $c: \mathbb{R} \to \mathbb{R} \times T^{*}C$; it consists of *points* of $\mathbb{R} \times T^{*}C$ and so may instead be regarded as a curve in $\mathbb{R} \times C$ consisting of *elements* there. In the classical treatments, such a curve of elements is called a *characteristic strip* when the 1-form $\theta = dr - \sum p_i dq_i$ is zero along the strip.

These elements are also used in geometrical optics in the space $\mathbb{R} \times C$. Huyghens principle for the propagation of a wave front W gives the new wave front W_t after time t as the envelope of the spherical waves centered at all the points of W. The transformation from W to W_t then does not carry points of the space (say the space $\mathbb{R} \times C$) into points, but elements into elements, so is really a transformation of $\mathbb{R} \times T^{*}C$ into itself. Since such a transformation of elements carries tangent wave fronts V and W into tangent wave fronts, it is called a "contact transformation".

This pictorial representation of a contact transformation is again connected with first order partial differential equations. One finds (for example, see Luneburg, Mathematical theory of optics) that Maxwell's equations yield wave fronts of the form

$\psi(x,y,z) - ct = 0$ where ψ satisfies $(\frac{\partial\psi}{\partial x})^2 + (\frac{\partial\psi}{\partial y})^2 (\frac{\partial\psi}{\partial z})^2 - n^2 = 0$

in the medium whose "index of refraction" is n.

A contact manifold is a manifold N of dimension $2n+1$ with a distinguished one-form θ such that $\theta \wedge (d\theta)^n \neq 0$ everywhere. We also consider submanifolds $\nu : S \to N$ such that $\nu^{*}\theta = 0$; one dimensional such are called strips. A transformation $(N, \theta) \xrightarrow{h} (N', \theta)$ is called a contact transformation when there is a map $\rho: N \to \mathbb{R}$ for which $h^{*}\theta' = \rho\theta$.

Theorem. A mapping h is a contact transformation if and only if it takes strips into strips.

Proof. $\Longrightarrow$ Trivial since $R \xrightarrow{c} (N,\theta) \xrightarrow{h} (N',\theta)$.

$\Longleftarrow$ Left to the reader.

Clearly, this notion of a contact transformation is preserved under composition; however, we also needed the characteristics before. Given smooth mappings E and F of N into R, there were X_E defined by (3') and $X_E(F) = [E,F]$, analogous to the Poisson bracket though not quite satisfying the Jacobi identity.

Suppose we are given a mapping $\Phi: T^*(N) \longrightarrow T_*(N)$, with

$$\Phi(dq^i) = -\frac{\partial}{\partial p_i}$$

$$\Phi(dp^i) = \frac{\partial}{\partial q^i} + p_i \frac{\partial}{\partial r}$$

$$\Phi(dr) = -\sum p_i \frac{\partial}{\partial p_i}$$

$$\begin{tikzcd} T_*N & & T^*N \arrow[ll, "\Phi"'] \arrow[dl] \\ & N \arrow[from=ul] & \end{tikzcd}$$

In particular (by construction) $\Phi(dE) = X_E$. The matrix of Φ will be of the form:

$$\Phi^{\alpha\beta} = \begin{array}{c c} & \begin{array}{ccc} q & p & r \end{array} \\ \begin{array}{c} q \\ p \\ r \end{array} & \begin{pmatrix} 0 & -I & p \\ I & 0 & p \\ 0 & -p & 0 \end{pmatrix} \end{array}$$

where α and β are indices ranging from 1 to 2n+1. It will be of rank 2n. Since $\Phi(dr - \sum p_i dq^i) = -\sum p_i \frac{\partial}{\partial p_i} + \sum p_i \frac{\partial}{\partial p_i} = 0$, the form θ will be determined up to a scalar factor as the kernel of Φ. At any point

Φ is a twice contravariant tensor field on N. We may define such a pair (N, Φ) to be a bracket manifold, where Φ is determined up to multiplication by a smooth $\rho \neq 0$. Given a configuration space C, we may construct $N = T^{*}C \times \mathbb{R}$ and define $\Phi^{\alpha\beta}$ as above, showing that this matrix as defined in local coordinates transforms properly under coordinate changes. In terms of Φ, we have $[E, F] = \langle dF, \Phi(dE) \rangle$. Thus the tensor Φ on N is indeed sufficient to define the bracket operation. For a mapping h to be a bracket transformation of (N, Φ) into (N', Φ'), we require that $h^{*}[E, F] = \rho[h^{*}E, h^{*}F]$ for some $\rho: N \rightarrow N'$ with $\rho \neq 0$ on N. We could instead require that h^{*} multiply by a common $\rho \neq 0$ the relations between canonical coordinates:

$$[q^i, q^j] = 0 = [p_i, p_j]$$
$$[q^i, p_j] = \delta^i_j$$
$$[q^i, r] = 0 \qquad [p_i, r] = p_i$$

We suggest that all contact manifolds are bracket manifolds. Note that $d\theta(X_E, X_F) = [E, F]$.

We now develop the suggestion of the previous paragraph. References are Cartan's Lecons sur les integrals invariants (1922) or the article by John Gray in the "Annals of Mathematics" 69(1959), pp. 421-450. Let (N, θ) be a given contact manifold. We will define in terms of the basis form θ a vector field Y_θ and a bracket $[\ \ ]_\theta$: Since the matrix of the 2-form $d\theta$ is of rank 2n, we may define a vector field Y by:

$$L_Y(d\theta) = 0 \quad \text{and} \quad \langle \theta, Y \rangle = 1.$$

For it suffices to do this locally, where we may take θ in the form $\theta = dr - \sum p_i dq^i$ by the Darboux theorem, in which case we have

$$d\theta = \sum dq^i \wedge dp_i \quad , \quad Y = \sum Y_i \frac{\partial}{\partial q^i} + \sum \overline{Y}_i \frac{\partial}{\partial p_i} + Z \frac{\partial}{\partial r}$$

and $L_Y d\theta = 0$ if and only if $Y = Z \frac{\partial}{\partial r}$ with $< \theta, Z \frac{\partial}{\partial r} > = Z$.

Equivalent would be an appeal to the fact that the $(2n+1)$ forms on N are spanned at each point by $\theta \wedge (d\theta)^n$, so that any such form, and in particular the form $dE \wedge (d\theta)^n$ may be represented uniquely as $h(r, p, q)\theta \wedge (d\theta)^n$ for some smooth function h. The quantity $h \overset{\text{def}}{\equiv} Y_\theta(E)$ is easily shown to be a derivation, and therefore determines a vector field Y_θ. That this is the vector field of the previous paragraph is verified by evaluation of $< Y_\theta, \theta >$ and $i_{Y_\theta} d\theta$.

We claim that θ also defines $[E, F]$ by:

$$dE \wedge dF \wedge \theta \wedge (d\theta)^{n-1} = [E, F]\theta \wedge d\theta^n \ .$$

The verification that the function so defined is $[E, F]$ is by direct calculation in canonical coordinates:

$$\theta \wedge (d\theta)^{n-1} = (dr - \sum_i p_i dq^i) \wedge (\sum_i dp_i \wedge dq^i)^{n-1}$$

$$= (dr - \sum_i p_i dq^i) \wedge (n-1)! \sum_i (-1)^{(n-1)(n-2)/2} dp_{\hat{j}} \wedge dq_{\hat{j}}$$

$$= (-1)^{(n-1)(n-2)/2}(n-1)! [\sum dp_{\hat{j}} \wedge dq_{\hat{j}} \wedge dr - \sum (-1)^{n+j-2} p_j dp_{\hat{j}} \wedge dq,$$

where $dp_{\hat{j}}$ is short for the product of all the dp_i with only dp_j omitted from the product, while $dq = dq^1 \wedge \ldots \wedge dq^n$ with no terms omitted.

We have $dE \wedge dF$ computed as:

$$\left(\sum_i \frac{\partial E}{\partial q^i} dq^i + \sum_i \frac{\partial E}{\partial p_i} dp_i + \frac{\partial E}{\partial r} dr\right)\wedge\left(\sum_i\left(\frac{\partial F}{\partial q^i} dq^i + \frac{\partial F}{\partial p_j} dp_j\right) + \frac{\partial F}{\partial r} dr\right)$$

and by direct computation the nonzero terms of $dE\wedge dF\wedge(\theta\wedge(d\theta)^{n-1})$, the only one without products $\theta\wedge\theta$, are:

$$(-1)^{n+\{(n-1)(n-2)/2\}}(n-1)! \sum_i\left(\frac{\partial E}{\partial q^i}\frac{\partial F}{\partial p_i} - \frac{\partial E}{\partial p_i}\frac{\partial F}{\partial q^i} - p_i \frac{\partial E}{\partial p_i}\frac{\partial F}{\partial r} + p_i \frac{\partial E}{\partial r}\frac{\partial F}{\partial p_i}\right)dp\wedge dq\wedge dr$$

$= (-1)^{n(n-1)/2}(n-1)!\,[E,F]dp\wedge dq\wedge dr$ according to our former definition. Since $\theta\wedge(d\theta)^n$ is $(-1)^{n(n-1)/2}\, n!\, dp\wedge dq\wedge dr$, our two definitions of $[E,F]$ have been shown to agree (except for $n!$ versus $(n-1)!$).

CHAPTER VII. COVARIANT DIFFERENTIATION

By David Golber

The following material summarizes, in outline form, lectures on covariant differentiation.

I (§48). Riemannian and Pseudo-Riemannian Metrics

A. Definition. A Riemannian metric g on a manifold M assigns in a C^∞ fashion to each point x of M an inner product g_x on the tangent vector space T_xM. A pseudo-Riemannian metric g assigns, in a C^∞ fashion, to each point $x \in M$ a non degenerate symmetric bilinear form g_x on T_xM.

Certain important results, especially III B. 3 below hold for pseudo-Riemannian as well as Riemannian metrics.

B. Local expression: In a coordinate patch on M, we have coordinates $x_1, \dots, x_n$, and vector fields $\partial/\partial x_1, \dots, \partial/\partial x_n$. Let us use the abbreviation $\partial/\partial x_i = \partial_i$ for these fields. Suppose M has a Riemannian or pseudo-Riemannian metric g. Then we can define C^∞ functions on the coordinate patch by

$$g_{ij}(x) = g_x(\partial_i(x), \partial_j(x)).$$

If g is Riemannian, then, for each x, the matrix $(g_{ij}(x))$ is positive definite and symmetric. If g is pseudo-Riemannian, then $(g_{ij}(x))$ is non-singular and symmetric.

C. How to get Riemannian or pseudo-Riemannian metrics.

(1) On one coordinate patch or on a Euclidean space, we have a system of vector fields $\partial_1, \ldots, \partial_n$ valid everywhere. It suffices to define the functions $g_{ij}(x)$. We could, for example, set $g_{ij}(x) = \delta_{ij}$, making the vector fields $\{\partial_i\}$ orthonormal at each point.

(2) On a manifold, we can use (1) to construct metrics on coordinate patches. Then we can use "partitions of unity" to combine these metrics into a metric on the whole manifold. This is the method usually used to show that any manifold has a metric.

(3) The usual way in which Riemannian metrics arise in practice is as follows: Suppose N is a space (often a Euclidean space) which already has a Riemannian metric h. Suppose we have a manifold M and a C^∞ function $f: M \to N$ which is an immersion (that is, the Jacobian matrix of f is non-singular at every point of M). Then we define a metric $f^*(h) = g$ on M by letting $g_x(X, Y) = h_{f(x)}(f_*(X), f_*(Y))$ for $X, Y \in T_\bullet M$

In local coordinates, this goes as follows: Let $x_1, \ldots, x_m$ be local coordinates on M, and $y_1, \ldots, y_n$ be local coordinates on N. Then f is given by $y_i = f_i(x_1, \ldots, x_m)$ $(i = 1, \ldots, n)$. Let h be given by $h_{ij}(y)$. Then $g = f^*(h)$ is given by

$$
\begin{aligned}
g_{ij}(x) &= \sum_{k,\ell=1}^{n} h_{k,\ell}(f(x)) \frac{\partial f_k}{\partial x_i} \cdot \frac{\partial f_\ell}{\partial x_j} \\
&= \sum_{k,\ell=1}^{n} h_{k,\ell}(f(x)) \frac{\partial y_k}{\partial x_i} \cdot \frac{\partial y_\ell}{\partial x_j} \quad .
\end{aligned}
$$

$f^*(h)$ is called the <u>pullback of h by f.</u>

II (§49). Covariant Differentiation

A. Motivation. We want some sort of directional derivative on a manifold. Will L_X do? No! Why?

(1) A directional derivative in the direction X should depend only on the value of X at the point in question. But $L_{f\cdot X}Y = f\cdot L_X Y - (Y\cdot f)X$, showing that $L_X Y$ depends on how X is changing at the point in question (note the term $(Y\cdot f)X$).

(2) We will be interested in Newton's laws, and therefore in acceleration as we move along a curve; i. e., the derivative of the velocity vector in the direction of the velocity vector. But the Lie derivative $L_X X$ is always zero. Thus we cannot use L_X to discuss acceleration.

B. The abstract covariant derivative.

(1) Definition: An (affine) connection ∇ on M is a rule which assigns to two smooth vector fields X and Y on M another smooth vector field $\nabla_X Y$ on M, called the covariant derivative of Y in the direction X (with respect to ∇), obeying

(a) $\nabla_{X_1+X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y$ and $\nabla_{fX} Y = f\cdot\nabla_X Y$,

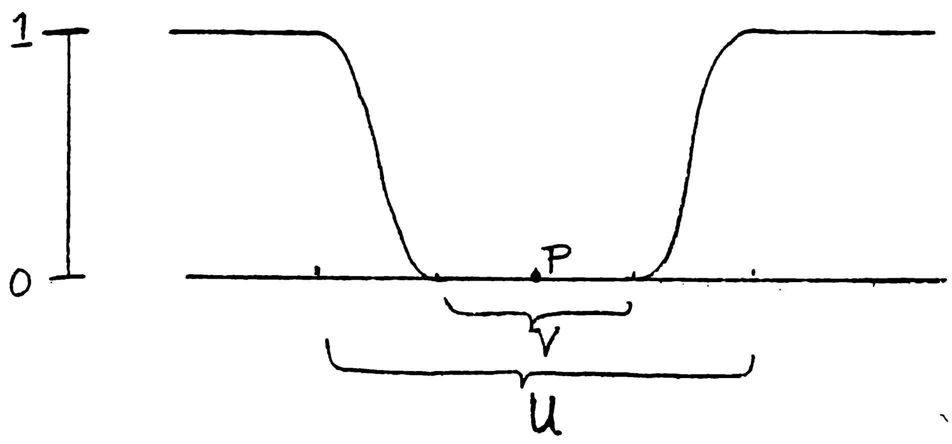
(b) $\nabla_X(Y_1+Y_2) = \nabla_X Y_1 + \nabla_X Y_2$ and $\nabla_X(fY) = (X\cdot f)Y + f\cdot\nabla_X Y$

for $f \in \mathcal{F}(M)$, X_i, Y_i vector fields on M.

(2) If $\nabla_X Y$ is to fulfill our expectations of what a directional derivative ought to be, then the following proposition should hold:

Proposition. For any point p on M, $(\nabla_X Y)_p$ depends only on X_p and on the behavior of Y in a neighborhood of p (actually on the "germ" of Y at p).

Proof. If $Y = Y'$ in a neighborhood U of p, then we take a "bump" function f which is one outside of U and zero on some neighborhood $V \subsetneq U$ of p.



Then $Y - Y' = f\cdot(Y - Y')$, so that

$$\begin{aligned}(\nabla_X(Y - Y'))_p &= [\nabla_X(f\cdot(Y - Y'))]_p \\ &= (X\cdot f)_p(Y - Y')_p + f(p)\nabla_X(Y - Y')_p \\ &= 0\cdot(Y - Y')_p + 0\cdot\nabla_X(Y - Y')_p = 0,\end{aligned}$$

so $(\nabla_X Y)_p = (\nabla_X Y')_p$.

If $X_p = X'_p$, then we can write $X - X' = \sum f_i P_i$, where $f \in \mathcal{F}(M)$ and P_i are vector fields on M, with $f_i(p) = 0$. (Details left to the reader.) Then

$$(\nabla_X Y)_p - (\nabla_{X'} Y)_p = (\nabla_{X-X'} Y)_p = (\nabla_{\sum f_i P_i} Y)_p$$

$$= \sum f_i(p)(\nabla_{P_i} Y)_p = 0 , \quad \text{Q.E.D.}$$

(3) By the proposition above, $\nabla_X Y$ is well-defined at a point, even if X and Y are defined only in a neighborhood of that point (rather than on the whole manifold). Thus the following definition makes sense:

For $\{x_1, \ldots, x_n\}$ a local coordinate system on M, $\partial_i = \partial/\partial x_i$ as before, we define n^3 smooth functions $\Gamma_{ij}^k(x)$ $(i, j, k = 1, \ldots, n)$ on the coordinate patch by

$$\Delta_{\partial_i}(\partial_j)_x = \sum_k \Gamma_{ij}^k(x)\, \partial_k(x).$$

The functions Γ_{ij}^k are called the <u>Christoffel symbols</u> of the connection We can calculate that for $X = \sum a_i(x)\partial_i$, $Y = \sum b_j(x)\partial_j$,

$$\nabla_X Y = \sum_i a_i \left[\sum_j \frac{\partial b_j}{\partial x_i} \partial_j + \sum_{j,k} b_j \Gamma_{ij}^k \partial_k \right].$$

(4) Examples:

(a) In $\mathbb{R}^n$, with the usual coordinate system, let Γ_{ij}^k be identically zero. Then we get the usual directional derivative of vector fields.

(b) If M is embedded in N (particularly $\mathbb{R}^n$), and if N has a connection ∇^N and a Riemannian metric, then we can use these to define a connection on M as follows: For $p \in M$ and X, Y vector fields defined on M in a neighborhood of p, extend X and Y to vector fields defined on N in a neighborhood of p. Define ∇^M by setting $(\nabla^M_X Y)_p = \mathrm{proj}_M((\nabla^N_X Y)_p)$, where proj_M denotes the perpendicular projection of T_pN onto T_pM. It is easy to verify that ∇^M satisfies the definition of a connection. $(\nabla^M_X Y)_p$ is independent of the extensions of X and Y to N, and ...)

C. Covariant derivative of a vector field along a curve.

(1) Define a vector field X along a curve $c: I \to M$ to be a map X such that

$$\begin{tikzcd} & T.M \arrow[d] \\ I \arrow[ur, "X"] \arrow[r, "c"] & M \end{tikzcd}$$

commutes.

Note difficulties involved in extending X to M when c crosses itself, has cusps, stationary points, etc. An example of a vector field along a curve c is the velocity $\dot{c}$

(2) For X a vector field along c, define the <u>covariant derivative of X along</u> c, $\nabla_{\dot c} X$, by

(a) where $\dot c(t) \neq 0$, extend X to a neighborhood of $c(t)$ in M, and let the covariant derivative along the curve just be the ordinary covariant derivative in M, $\nabla_{\dot c}(X)$. We show that the result is independent of the extension by showing that

$$\nabla_{\dot c}(Y) = \sum_i (\dot c \cdot y_i)\partial_i + \sum_{i,j,k} \dot c_i y_j \Gamma^k_{ij} \, \partial_k ;$$

where $Y = \sum y_i \partial_i$

(b) Where $\dot c = 0$, let $\nabla_{\dot c}(Y) = 0$.

D. Parallel translation

(1) For M a manifold with connection ∇; c a smooth curve in M and X a vector field along c, we say that X is parallel along c if $\nabla_{\dot{c}} X = 0$ holds everywhere on c.

In local coordinates $x_1, \ldots, x_n$: let c be given by $c_1(t), \ldots, c_n(t)$; let X be given by $X(t) = \sum X_i(t)\partial_i(c(t))$. Then the equation $\nabla_{\dot{c}} X = 0$ is equivalent to

$$\frac{dX_i}{dt}(t) + \sum_{k,\ell} \Gamma^i_{k\ell}(c(t)) \frac{dc_k}{dt} \cdot X_\ell(t) = 0, \qquad (i = 1, \ldots, n).$$

This is a system of n linear differential equations in n variables. For an initial value t_o and an arbitrarily chosen vector $X(t_o)$ in $T_{c(t_o)}M$, there is a unique vector field $X(t)$ along c which coincides with $X(t_o)$ at $c(t_o)$. The value of this vector field at $c(t_1)$ is said to be the parallel translation of $X(t_o)$ along c to $c(t_1)$.

(2) Note that the parallel translation along c from $c(a)$ to $c(b)$ gives an invertible linear map of $T_{c(a)}M$ to $T_{c(b)}M$. This linear map depends very heavily on c (unless the "curvature" of the connection is zero).

(3) Relation of parallel translation and ∇.

Proposition. Let $X \in T_pM$, Y be a vector field defined in some neighborhood of p. Take any curve c such that $\dot{c}(0) = X$. Then

$$(\nabla_X Y)_p = \lim_{t \to \infty} \frac{(\|^o_{c,t} \; Y(c(t))) - Y(p)}{t}$$

(where $\|^o_{c,t}$ denotes the parallel translation along c from $c(t)$ to $c(0) = p$.)

Proof. Let $\{Z_1, \ldots, Z_n\}$ be a basis of T_pM. Extend Z_i by parallel translation to a vector field along c. Thus, for each t, $\{Z_1(t), \ldots, Z_n(t)\}$ is a basis for $T_{c(t)}M$.

Write $Y(c(t)) = \sum y_i(t) Z_i(t)$. As parallel translation is linear and the Z_i's are parallel along c, we get that

$$\|^{o}_{c,t}\ Y(c(t)) = \sum y_i(t) Z_i(0).$$

Taking the difference and the limit, we find that the right hand side of our conclusion becomes

$$\sum_i \left(\lim_{t\to 0} \frac{y_i(t) - y_i(0)}{t}\right) Z_i(0) = \sum_i (\dot{c} \cdot y_i)_0 Z_i(0) .$$

But, as $\nabla_{\dot{c}} Z_i = 0$, this equals the left hand side of our conclusion:

$$\begin{aligned} \nabla_X Y &= \nabla_{\dot{c}}\left(\sum y_i(t) Z_i(t)\right) \\ &= \sum (\dot{c} \cdot y_i) \cdot Z_i(t) + \sum y_i(\nabla_{\dot{c}} Z_i) \\ &= \sum (\dot{c} \cdot y_i) \cdot Z_i(t) + 0 \quad , \qquad \text{Q.E.D.} \end{aligned}$$

(4) Note the similarity of the above proposition to the proposition giving the Lie derivative L_X in terms of the flow of X. As a parallel to Willmore's theorem, we have:

<u>Theorem</u>. We can extend ∇_X to a unique linear map of the various tensor bundles

$$\nabla_X : T_s^{\,r}(M) \to T_s^{\,r}(M)$$

such that

(1) $\nabla_X f = X \cdot f$ for $f \in \mathcal{F}(M)$,

(2) For Y a vector field on M, $\nabla_X Y$ is the given covariant derivative.

(3) $\nabla_X \delta = 0$, where $\delta = \sum_i e^i \otimes e_i$.

(4) ∇_X is a derivation of the tensor algebra:

$$\nabla_X(\tau \otimes \tau') = (\nabla_X \tau) \otimes \tau' + \tau \otimes (\nabla_X \tau')$$

Further, we can also extend the notion of parallel translation along c to

$$\|_{c,a}^{b} : T_s^{\,r}(M)_{c(a)} \longrightarrow T_s^{\,r}(M)_{c(b)}$$

and, for any tensor field τ, $\nabla_X \tau$ is given by a limit, as in the previous proposition.

<u>Example</u>: Using (3) and (4), we can find

$$\nabla_{\partial/\partial x_i}(dx^j) = -\sum_k \Gamma^j_{ik}\, dx^k .$$

III (§50) <u>Nice Covariant Derivatives</u>.

A. Torsion

(1) Symmetry. Suppose $S: \mathbb{R}^2 \to M$ is a smooth map. (Call S a "parametrized surface".) Then we get two vector fields on S,

$$S_x = \frac{\partial S}{\partial x} \quad \text{and} \quad S_y = \frac{\partial S}{\partial y} ,$$

$$\begin{tikzcd} & TM \arrow[d] \\ \mathbb{R}^2 \arrow[ur, "S_x, S_y"] \arrow[r, "S"] & M \end{tikzcd}$$

We can form $\nabla_{S_x} S_y$ and $\nabla_{S_y} S_x$. (These correspond to covariant derivatives along the curves $S(t, y_o)$ and $S(x_o, t)$ respectively.) In general, it is not true that

$$\nabla_{S_x} S_y = \nabla_{S_y} S_x .$$

It this condition is satisfied for all parametrized surfaces S, then we say that the connection ∇ is <u>torison free</u> or <u>symmetric</u>. (Note: this condition does <u>not</u> correspond to the property $\partial^2/\partial x \partial y = \partial^2/\partial y \partial x$ in ordinary Euclidean space. <u>That</u> property corresponds to the "curvature" of the connection being zero.)

(2) The torsion tensor.

For vector fields X, Y on M, define

$$\mathrm{Tor}(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

Show

(a) Tor is $\mathcal{F}(M)$-linear in X and Y,

(b) Follows from (a) that $\mathrm{Tor}(X, Y)_p$ depends only on X_p and Y_p, and bilinearly on these.

(c) (b) means that Tor is a tensor, the <u>torsion tensor</u> of the connection ∇. (Actually, more properly speaking, the torsion tensor is the tensor τ of type $\binom{2}{1}$ given by

$$\tau(X, Y, \omega) \equiv \omega(\mathrm{Tor}(X, Y))$$

for X, Y vector fields and ω a 1-form.)

(d) Calculate local form: If we let

$$\mathrm{Tor}(\partial_i, \partial_j) = \sum_k \mathrm{Tor}_{ij}^{\ \ k} \ \partial_k$$

then we find

$$\mathrm{Tor}_{ij}^{\ \ k} = \Gamma_{ij}^{\ \ k} - \Gamma_{ji}^{\ \ k} \ .$$

(3) Relation of Tor and symmetry.

<u>Theorem.</u> $\mathrm{Tor} \equiv 0$ if and only if $\nabla_{S_x} S_y = \nabla_{S_y} S_x$ for all parametrized surfaces.

<u>Proof.</u> ($\Longleftarrow$). For any two vectors X and Y at p, we can choose S so that $(S_x)_p = X$ and $(S_y)_p = Y$. It is easy to calculate that, because S_x and S_y both come from S, $[S_x, S_y] = 0$. (The calculation reduces to $\dfrac{\partial^2}{\partial y \partial x} = \dfrac{\partial^2}{\partial y \partial x}$ on $\mathbb{R}^2$.) Then

$$\nabla_{S_x} S_y = \nabla_{S_y} S_y \text{ implies } \nabla_X Y - \nabla_Y X = 0 \text{ and } [X, Y] = 0,$$

so $\mathrm{Tor}(X, Y) = 0$ for all X, Y.

($\Longrightarrow$). $\mathrm{Tor}(S_x, S_y) = 0$. But, again $[S_x, S_y] = 0$. Q.E.D.

B. Invariance of g under parallel translation.

(1) <u>Definition.</u> If M has both a connection ∇ and a Riemannian or pseudo-Riemannian metric $g(X, Y) = (X, Y)$, then it will be nice if <u>parallel translation preserves inner products</u>; i.e., whenever $X(t)$ and $Y(t)$ are parallel along c, then $(X(t), Y(t))$ is independent of t.

(2). Proposition. (1) above holds if and only if the following condition holds: if A and B are vector fields on M, then

$$X\cdot(A,B) = (\nabla_X A, B) + (A, \nabla_X B).$$

Proof. ($\Longrightarrow$) Take c a curve with $\dot{c}(0) = X$. Take an orthonormal basis $Y_1, \ldots, Y_n$ at $c(0)$. Extend these by parallel translation along c. By our hypothesis (1) above, the vectors $Y_1(t), \ldots, Y_n(t)$ form an orthonormal basis in $T_{c(t)}M$ for each t.

We can write

$$A(c(t)) = \sum f_i(t) Y_i(t),$$

$$B(c(t)) = \sum g_i(t) Y_i(t).$$

Then

$$(A(c(t)), B(c(t))) = \sum f_i(t)\cdot g_i(t),$$

and

$$\begin{aligned} X\cdot(A,B) &= \frac{d}{dt}(A(c(t)), B(c(t))) = \frac{d}{dt}\left(\sum_i f_i(t)\cdot g_i(t)\right) \\ &= \sum_i [(X\cdot f_i) g_i + f_i (X\cdot g_i)] \\ &= (\nabla_X A, B) + (A, \nabla_X B). \end{aligned}$$

($\Longleftarrow$) is even easier. If A and B are parallel along c, then $\nabla_{\dot{c}} A = \nabla_c B = 0$, so the derivative of (A, B) along c is

$$\begin{aligned} \dot{c}\cdot(A,B) &= (\nabla_{\dot{c}} A, B) + (A, \nabla_{\dot{c}} B). \\ &= 0 + 0 = 0. \end{aligned}$$

Therefore (A, B) is constant along c. Q.E.D.

Note: If we regard (,) as a tensor $g \in T_0^2(M)$, then the condition that parallel translation preserve inner products is equivalent to $\nabla_X g = 0$ for all vector fields X on M. Here, ∇_X is as described in the theorem at the top of page 96.

Note: The theorem above also holds for pseudo-Riemannian metrics. The modification of the proof is left to the reader.

(3) Main theorem (Holds for pseudo-Riemannian metrics).

Theorem. Given M with a pseudo-Riemannian metric (,), there is a unique connection ∇ on M satisfying

(1) $\mathrm{Tor} = 0$

(2) parallel translation preserves inner products.

Proof. Uniqueness: we have from (2) that

$$X\cdot(Y, Z) = (\nabla_X Y, Z) + (Y, \nabla_X Z).$$

Using (1), this becomes

$$X\cdot(Y, Z) = (\nabla_X Y, Z) + (Y, \nabla_Z X) + (Y, [X, Z]).$$

Cyclically permuting X, Y and Z, we get two other equations. Solving for $(\nabla_X Y, Z)$ and eliminating the terms involving $\nabla_Y Z$ and $\nabla_Z X$ (using the symmetry of (,)) we get

$$2(\nabla_X Y, Z) = X\cdot(Y, Z) + Y\cdot(Z,X) - Z\cdot(X, Y) - (Y, [X, Z])$$
$$- (Z, [Y, X]) + (X, [Z, Y]).$$

As (,) is nonsingular, this shows that $\nabla_X Y$ is determined.

Conversely, if we define $\nabla_X Y$ by using this formula, then we find that condition (1) and condition (2) of the theorem are satisfied. Q.E.D.

(4) Local form of the above result

In local coordinates, using the fact that $[\partial_i, \partial_j] = 0$ we get

$$2\Gamma_{ij}^{\;k} = \sum_{\ell} [\partial_i(g_{j\ell}) + \partial_j(g_{i\ell}) - \partial_\ell(g_{ij})] g^{\ell k} ,$$

where $\partial_i = \dfrac{\partial}{\partial x_i}$, and $(g^{\ell k})$ is the inverse matrix of (g_{ij}).

C. <u>Example.</u> Suppose N (especially $\mathbb{R}^n$) is a manifold with a metric g and the unique corresponding covariant derivative $\overset{N}{\nabla}$. Let M be embedded in N. M inherits a metric h (see I. C. 3) and a connection $\overset{M}{\nabla}$ (see II. B. 4). <u>Claim</u> that $\overset{M}{\nabla}$ is the unique connection on M corresponding to the metric h.

<u>Proof.</u> (1) Tor is zero: if S is a surface in M, then it is a surface in N. Then $\overset{N}{\nabla}_{S_x} S_y = \overset{N}{\nabla}_{S_y} S_x$, so their projections $\overset{M}{\nabla}_{S_x} S_y$ and $\overset{M}{\nabla}_{S_y} S_x$ into M are equal.

(2) Show

$$X\cdot(Y, Z) = (\overset{M}{\nabla}_X Y, Z) + (Y, \overset{M}{\nabla}_X Z) \quad \text{for } X, Y, Z \text{ tangent fields}$$

to M. But the left hand side is independent of whether we look in M or in N. The equation holds in N. $\overset{N}{\nabla}_X Y$ differs from $\overset{M}{\nabla}_X Y$ by a vector perpendicular to M, so $(\overset{M}{\nabla}_X Y, Z) = (\overset{N}{\nabla}_X Y, Z)$, and so on. Q.E.D.

IV. (§51) Lagrange's Equations.

Suppose N particles move in space, subject to certain constraints. For each allowed configuration of the N particles, we get a point in 3N-space. We assume that the set of allowed configurations is a submanifold M of $\mathbb{R}^{3N}$ of dimension n and that arbitrary motions on the submanifolds are possible. This is what it means for the constraints to be "holomonic".

Put the metric on $\mathbb{R}^{3N}$ given by

$$h = \sum_{i=1}^{N} m_i(dx_i^2 + dy_i^2 + dz_i^2),$$

where m_i is the mass and (x_i, y_i, z_i) the coordinates of the i^{th} particle.

Let ω denote the 1-form on $\mathbb{R}^{3N}$ given by

$$\omega = \sum_{i=1}^{N} F_{ix}dx^i + F_{iy}dy^i + F_{iz}\,dz^i ,$$

where F_{ix} is the force on the i^{th} particle in the x-direction, etc.

Now, we have the inclusion $M \overset{f}{\hookrightarrow} \mathbb{R}^{3N}$. Let $g = f^*(\omega)$, and let ∇ be the unique nice connection on M associated with g. Now, g produces an isomorphism of T_*M with T^*M. Let X_ω be the vector field corresponding to ω_Q under this isomorphism. Then the equations of motion may be expressed for a path c in M as

$$\nabla_{\dot{c}}\,\dot{c} = X_\omega$$

That is, given an initial position $c(t_o)$ and an initial velocity $\dot{c}(t_o)$ the system follows the unique path c(t) satisfying this equation for these initial conditions.

Proof. First look at unconstrained motion in $\mathbb{R}^{3N}$. It is easy to see that the metric h gives us the usual connection $\nabla^R (\Gamma^k_{ij} \equiv 0)$ on $\mathbb{R}^{3N}$. Now specify the path c by giving coordinates c_{jx}, c_{jy}, c_{jz} for the j^{th} particle, $j = 1, \ldots, n$. The definition of the covariant derivative along c then gives

$$\nabla^R_{\dot{c}}\dot{c} = \sum_j \left(\frac{d^2 c_{jx}}{dt^2} \frac{\partial}{\partial x_j} + \frac{d^2 c_{jy}}{dt^2} \frac{\partial}{\partial y_j} + \frac{d^2 c_{jz}}{dt^2} \frac{\partial}{\partial z_j} \right)$$

Further, if X^R_ω is the vector field which corresponds to the form ω by way of h then

$$X^R_\omega = \sum_j \left(\frac{F_j x}{m_j} \frac{\partial}{\partial x_j} + \frac{F_j y}{\partial y_j} + \frac{F_j z}{m_j} \frac{\partial}{\partial z_j} \right).$$

Therefore the statement

$$\nabla^R_{\dot{c}} \dot{c} = X^R_\omega$$

is equivalent to Newton's equations in $\mathbb{R}^N$.

Let us decompose this vector identity into components parallel an perpendicular to $M \subsetneq R^N$.

$$(\nabla^R_{\dot{c}}\dot{c})_{||} + (\nabla^R_{\dot{c}}\dot{c})_{\perp} = (X^R_\omega)_{||} + (X^R_\omega)_{\perp}.$$

We know (III. C.) that $(\nabla^R_{\dot{c}}\dot{c})_{||} = \nabla_{\dot{c}}\dot{c}$ in M. Therefore Newton's equations are equivalent to:

$$\nabla_{\dot{c}} \dot{c} = (X^R_\omega)_{||}$$

$$(\nabla^R_{\dot{c}}\dot{c})_{\perp} = (X^R_\omega)_{\perp}.$$

The statement that the motion is constrained to M says that the second equation must balance. Therefore, the first equation is our equation of motion. We will have proven our result if we show that $(X_\omega^R)_{||} = X_\omega$, as defined at the start. As these are both tangent fields to M, it is enough to show that

$$g((X_\omega^R)_{||}, Y) = g(X_\omega, Y) \qquad \text{for } Y \in T_*M.$$

But

$$\begin{aligned} g((X_\omega^R)_{||}, Y) &= h(X_\omega^R, Y) \qquad \text{as } Y \in T.M \\ &= \omega(Y) \qquad \text{by definition of } X_\omega^R \\ &= \langle \omega, f_* Y\rangle \\ &= \langle f^*\omega, Y\rangle \\ &= g(X_\omega, Y) \quad \text{by definition of } X_\omega. \qquad \text{Q.E.D.} \end{aligned}$$

We make several comments on the material above.

(1) The equation $\nabla_{\dot{c}}\dot{c} = X_\omega$ is the "same" as Lagrange's equations. To see that, take $g(-, \frac{\partial}{\partial x_j})$ of both sides. Then

$$g(X_\omega, \frac{\partial}{\partial x_j}) = \omega(\frac{\partial}{\partial x_j}) = Q_j ,$$

the generalized force in the j^{th} direction. To analyze the term $g(\nabla_{\dot{c}}\dot{c}, \frac{\partial}{\partial x_j})$, let Z be a vector field extending $\dot{c}$.

$$g(\nabla_Z Z, \frac{\partial}{\partial x_j}) = Z\cdot g(\frac{\partial}{\partial q^j}, Z) + g(Z, [\frac{\partial}{\partial q^j}, Z]) - \frac{1}{2}\frac{\partial}{\partial q^j}\cdot g(Z, Z),$$

(using the relation of the Theorem III. B. 3).

Claim that this expression is $\frac{d}{dt}(\frac{\partial T}{\partial \dot{q}^j}) - \frac{\partial T}{\partial q^j}$. But, for $v_o \in M_*$

$T(v_o) = \frac{1}{2} g(v_o, v_o)$. Thus $\frac{\partial T}{\partial \dot{q}^j}\Big|_{v_o} = g(v_o, \frac{\partial}{\partial q^j})$, and, as $Z = \dot{c}$,

$\frac{d}{dt}(\frac{\partial T}{\partial \dot{q}^j}) = Z \cdot g(Z, \frac{\partial}{\partial q^j})$. To show that

$$g(Z, [\frac{\partial}{\partial q^j}, Z]) - \frac{1}{2}\frac{\partial}{\partial q^j} \cdot g(Z, Z) = -\frac{\partial T}{\partial q^j} : \quad \text{Let } Z = \sum_\ell z^\ell \frac{\partial}{\partial x^\ell}.$$

Then

$$[\frac{\partial}{\partial q^j}, Z] = \sum_m \frac{\partial z^m}{\partial q^j} \cdot \frac{\partial}{\partial q^m}.$$

Let g be given by g_{ij}. Then

$$g(Z, [\frac{\partial}{\partial q^j}, Z]) = \sum_{k,\ell} g_{k\ell}\, z^k \frac{\partial z^\ell}{\partial q^j}$$

and

$$\frac{1}{2}\frac{\partial}{\partial q^j}(g(Z, Z)) = \frac{1}{2}\frac{\partial}{\partial q^j}\Big(\sum_{k,\ell} g_{k\ell}\, z^k z^\ell\Big)$$

$$= \frac{1}{2}\sum_{k,\ell}\frac{\partial g_{k\ell}}{\partial q^j} z^k z^\ell + \sum_{k,\ell} g_{k\ell}\, z^k \frac{\partial z^\ell}{\partial q^j},$$

and therefore

$$g(Z, [\frac{\partial}{\partial q^j}, Z]) - \frac{1}{2}\frac{\partial}{\partial q^j} g(Z, Z) = -\frac{1}{2}\sum_{k,\ell}\frac{\partial g_{k\ell}}{\partial q^j} z^k z^\ell$$

$$= (-\frac{\partial T}{\partial q^j})(Z),$$

(where $T(q^1, \ldots, q^n; \dot{q}^1, \ldots, \dot{q}^n) = \frac{1}{2}\sum_{k,\ell} g_{k\ell}(q^1, \ldots, q^n) \cdot \dot{q}^k \cdot \dot{q}^\ell$).

Thus we have shown that we do indeed have Lagrange's equations here.

(By being more sophisticated, we could probably have done this with less involvement in the coordinates. But note that our final result involves the coordinates, so we cannot avoid them entirely.)

(2) Note: X_ω does not depend on forces which are perpendicular (relative to h) to M. These "forces of constraint" may therefore be ignored in

setting up the equations $\nabla_{\dot{c}}\dot{c} = X_\omega$ for solution. This is the whole advantage of the method.

(3) If the forces parallel to M are zero, $\nabla_{\dot{c}}\dot{c} = 0$ says that the path is a "geodesic". For a single particle moving on a surface (e. g. a marble moving (without gravity) on a cone) this is particularly reasonable, as it just says that the acceleration is perpendicular to the surface. (Recall III. C.)

(4) Note finally that if we take the inner product of $\nabla_{\dot{c}}\dot{c} = X_\omega$ with $\dot{c}$ and apply III. B. 3 again, we get

$$\dot{c}\cdot[\tfrac{1}{2}g(\dot{c}, \dot{c})] = \omega(\dot{c})$$

which simply says that the rate of change of the kinetic energy T $(=\frac{1}{2}g(\dot{c}, \dot{c}))$ is given by the work-form applied to $\dot{c}$, as it should be.

SUPPLEMENT - EULER'S EQUATIONS

By Raphael Zahler

Frequently in classical mechanics it happens that the configuration space of the dynamical system in question has the structure of a Lie group. This means that it is a differentiable manifold with an additional "multiplication" operation related to the structure of the manifold. The points of the manifold are then thought of as motions of the system; the product xy stands for the motion resulting from the combined effect of the motion x followed by the motion y. For example, the group of all rotations of an asymmetrical three-dimensional body which leaves a particular point fixed is the familiar Lie group $SO(3)$. The reader may consult Helgason, Differential Geometry and Symmetric Spaces, (Academic Press) for a full treatment of the mathematical theory of Lie groups; here we will briefly outline some important facts. For any fixed element g of the Lie group G, multiplication on the left by g gives a map L_g of G into itself called "left translation by g". The induced map on tangent spaces, L_{g*}, maps $T_e(G)$ to $T_g(G)$, where e is the identity element of the group G. In this way the structure of the vector space $T_e(G)$ is closely related to the overall structure of G. $T_e(G)$ is called the Lie algebra of G. There is a function, called the exponential map, which takes vectors of $T_e(G)$ to points of G; if $\exp X$ is the point corresponding to the vector X, then $\exp(t_1+t_2)X = (\exp t_1X)(\exp t_2X)$; in particular, $\exp 0 = e$.

Suppose now that the kinetic-energy metric on our Lie group is left-invariant: that is,

$$(X,Y)_g = (L_{h*}X, L_{h*}Y)_{hg}$$

for all $g, h \in G$, $X, Y \in T_g(G)$. (The subscript "g" in $(X,Y)_g$ denotes that the metric is being applied to tangent vectors at the point g.) It is then a fact that the geodesics of the metric, which represent the motion of the system in time, can be described near e by $c(t) = \exp X(t)$, where $X(t) \in T_e(G)$, all t. We will investigate the behavior of these trajectories close to e.

Consider a geodesic $\exp \gamma(t)$, and let its tangent vector at time t be $\dot{\gamma}(t)$. Then the kinetic energy as a function of time is

$$T(t) = \frac{1}{2}(\dot{\gamma}, \dot{\gamma})_{\exp \gamma} = \frac{1}{2}(L_{\exp(-\gamma)*}\dot{\gamma}, L_{\exp(-\gamma)*}\dot{\gamma})_e ,$$

or, if we write $\xi(t) = L_{\exp(-\gamma)*}\dot{\gamma}$,

$$T(t) = \frac{1}{2}(\xi(t), \xi)t)).$$

Let us assume that there is no potential energy. Then we will be able to show that ξ satisfies Euler's equation: $\dot{\xi} = -B(\xi, \xi)$, where the function $B: T_e(G) \times T_e(G) \to T_e(G)$ is defined uniquely by

$$([X, Y], Z) = (B(Z, X), Y) \quad , \quad \text{all } Y.$$

First of all, since everything involved is invariant under left translation, it will suffice to consider the case $\gamma(0) = 0$; any other geodesic will be a left translation of one of these. Next, it may be proved using a "Taylor expansion" technique that

$$L_{\exp X*} Y = Y - \frac{1}{2}[X, Y] + O(|X|^2),$$

where the symbols X and Y on the right-hand side are understood in terms of a special "canonical" coordinate system $\{x_1, \ldots, x_n\}$ in a neighborhood of e by which we identify the vectors of the various different tangent spaces near e. Let us plug this into our formula for the Lagrangian:

$$L = T = \frac{1}{2}(\xi,\xi)_e = \frac{1}{2}(L_{\exp(-\gamma)*}\dot{\gamma}, L_{\exp(-\gamma)*}\dot{\gamma})_e$$

$$= \frac{1}{2}(\dot{\gamma} - \frac{1}{2}[-\gamma,\dot{\gamma}] + O(|\gamma|^2), \dot{\gamma} - \frac{1}{2}[\gamma,\dot{\gamma}] + O(|\gamma|^2))_e$$

$$= \frac{1}{2}((\dot{\gamma},\dot{\gamma})_e - 2(\dot{\gamma}, \frac{1}{2}[-\gamma,\dot{\gamma}])_e + O(|\gamma|^2))$$

$$= \frac{1}{2}(\dot{\gamma},\dot{\gamma})_e - \frac{1}{2}(\dot{\gamma},[\dot{\gamma},\gamma])_e + O(|\gamma|^2)$$

$$= \frac{1}{2}(\dot{\gamma},\dot{\gamma})_e - \frac{1}{2}(B(\dot{\gamma},\dot{\gamma}),\gamma)_e + (|\gamma|^2).$$

We now invoke the Euler-Lagrange equations, which must be satisfied by any geodesic: in terms of the canonical coordinates,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_r} = \frac{\partial L}{\partial x_r} \qquad r = 1,\ldots,n.$$

Writing $L = L(t,\gamma,\xi) = L(t, x_1,\ldots,x_n, \dot{x}_1,\ldots,\dot{x}_n)$, we get $L = \frac{1}{2}(\xi,\xi)_e = \frac{1}{2}\sum g_{ii}\dot{x}_i^2$, where we assume that the matrix of constants $\{g_{ij}\}$ representing the metric at e has been diagonalized. Hence

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_r} = g_{rr}\frac{d\dot{x}_r}{dt} \qquad \text{(neglecting the "0" term)}$$

Next : $\dfrac{\partial L}{\partial x_r} = \dfrac{\partial}{\partial x_r}\left(\frac{1}{2}(\dot{\gamma},\dot{\gamma}) - \frac{1}{2}(B(\dot{\gamma},\dot{\gamma}),\gamma)\right).$

If we write $B(\dot{\gamma},\dot{\gamma}) = \sum b_i \dfrac{\partial}{\partial x_i}$ then $\dfrac{\partial \dot{x}_r}{dt} = -b_r$, and the Euler-Lagrange equations now imply that Euler's equation is satisfied near the origin by the trajectories of the dynamical system, when the configuration space happens to form a Lie group with left-invariant metric. (Note: this derivation is due to Arnold (Comptes Rendues, v. 260 (May 31, 1965), p. 5668); a derivation independent of Lie-group theory is found in Loomis and Sternberg, Advanced Calculus (Addison-Wesley), p. 541 ff.)

Let us apply this to the case of rigid-body motion. Considering only rotations leaving a point fixed, a moment's reflection shows that the rotation group SO(3) is actually the configuration space of our system; just take a fixed reference position of the body and consider any other position as a rotation from the reference position. The tangent space to the Lie group SO(3) at the origin can be identified with the vector space of all skew-symmetric three-by-three matrices; these are usually referred to in physics texts as "infinitesimal rotations", and this is the Lie algebra we must work with.

Let $F(t)$ be a curve in configuration space; then a particle at a point p of Euclidean space moves in the trajectory $(F(t))(p)$. A physically sensible Riemannian metric in this case is the inertia tensor $(A, B) = \int m(A_p, B_p)dp$ where A and B are skew-symmetric matrices. In general, this is a left- but not right-invariant metric. In terms of a basis $\{e_i\}$ of $\mathbb{R}^3$, we have

$$\begin{aligned}(A, B) &= \int m(A(\sum r_i e_i), B(\sum r_j e_j))\, dp \\ &= \sum_{i,j=1}^{3} (Ae_i, Be_j) \int m r_i r_j \, dp \\ &= \sum_{i,j=1}^{3} I_{ij}(Ae_i, Be_j).\end{aligned}$$

I, the coordinatized version of the inertia tensor, may be diagonalized (principal axis theorem); picking an obvious basis E_{12}, E_{13}, E_{23} of the Lie algebra gives

$$(E_{ij}, E_{ij}) = \begin{cases} I_i + I_j & i \neq j \\ 0 & \text{otherwise} \end{cases}.$$

It is now possible to substitute in Euler's equation as we derived it above, to obtain

$$\frac{da_{12}}{dt}(I_1 + I_2) - (I_2 - I_1)a_{13}a_{23} = 0,$$

and two similar equations obtained by cyclic permutations of the indices. This is the form of Euler's equations without potential usually found in physics texts; it may be used to solve problems like those involving the spinning top.

A similar situation occurs in the physics of fluid flow. If we have a domain filled with a uniform incompressible ideal fluid, the group of volume-preserving diffeomorphisms of this domain forms configuration space, and, in certain conditions, is a Lie group. Euler's equation , in the form in which we have derived it, now yields

$$\frac{d}{dt}(\mathrm{curl}\ \vec{\xi}) = \mathrm{curl}(\vec{\xi} \times \mathrm{curl}\ \vec{\xi})$$

where ξ is now interpreted as the velocity vector field of the fluid. This is known as Euler's equation for fluid mechanics.