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

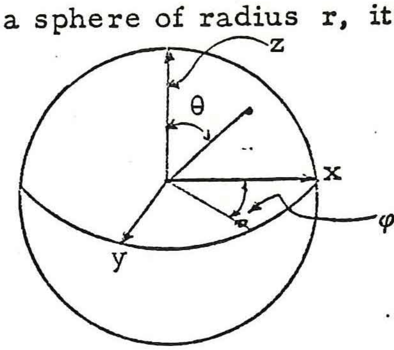
GEOMETRICAL MECHANICS

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 φ - θ 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> $\mathrm{Hom}_K(A,B) = \{f: A \longrightarrow B \mid f \text{ is a homomorphism of } K\text{-modules}\}$

= 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 $K \times K \xrightarrow{\cdot} K$)

$A^* \overset{\text{def.}}{=\!=} \mathrm{Hom}_K(A,K)$ 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 $e^i : V \longrightarrow \mathbb{R}$ 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 $e^i : V \longrightarrow \mathbb{R}$ $(v \rightsquigarrow x_i)$, 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
$= \{a_0 + a_1 x + \ldots + a_k x^k \mid k \geq 0, \ a_i \in \mathbb{R}, \ 0 \leq i \leq k\}$.

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

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

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

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 $(\mathbb{Z}_3)^* = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_3, \mathbb{Z}) = \{f : \mathbb{Z}_3 \longrightarrow \mathbb{Z} \mid f$ 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 $V^* = \{f : V \longrightarrow K \mid f$ is a K-linear transformation$\}$, where V is again a finite dimensional vector space over a field K. Write

$(f,v) \stackrel{\text{def.}}{=\!=} f(v) \in K$, 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 how 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 Algebra, Ch. 15, § 5) and we will identify

$$V \xrightarrow{\;\equiv\;} V^{**}$$
$$v \rightsquigarrow \bar{v}$$

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

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

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

Proof. Let $v = e_1x^1 + \ldots + e_nx^n$, then $(e^i, v) = (e^i, e_1x^1 + \ldots + e_nx^n)$ $= (e^i, e_1)x^1 + \ldots + (e^i, e_n)x^n = 0 + \ldots + (e^i, e_i)x^i + \ldots + 0 = x^i$.

§2 Euclidean Vector Spaces

A Euclidean vector space is a finite dimensional vector space W over $\mathbb{R}$ with an inner product $W \times W \xrightarrow{\cdot} \mathbb{R}$ $((v,w) \rightsquigarrow v\cdot w)$ 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 $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\cdot\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_iy_i$. 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 $\sum_{i=1}^{n} a_ix^i$ 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 $V \cong V^*$ $(v \rightsquigarrow \tilde{v})$, 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 $V \longrightarrow V^*$ 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 co-ordinates $q^1, \ldots, q^n$ correspond to the position of the n particles. If the i^{th} particle has mass m_i, its kinetic energy T is

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

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

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

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

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

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

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 $f:\mathbb{R}^n \to \mathbb{R}^n$), we replace $|x-a|$ by $\sqrt{\sum (x_i - a_i)^2}$.

Now some topological definitions. In $\mathbb{R}^n$, given a point $a = (a_1, \dots, a_n)$, the open ball of radius δ with center a is $\{(x^1, \dots, x^n) \mid \sqrt{\sum (x_i - a_i)^2} < \delta\}$. 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.

<u>Definition.</u> 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.

<u>Theorem.</u> 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 <u>inverse image</u> 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 $T^a U$, 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 $\quad \langle f_1 + f_2, c \rangle_a = \langle f_1, c \rangle_a + \langle f_2, c \rangle_a$

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

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

$$\tau_a c = \tau_a b \Longleftrightarrow c \equiv b \Longleftrightarrow <f, c> = <f, b> \quad \text{all } f$$
$$\Longleftrightarrow <-, \tau_a c> = <-, \tau_a b> ;$$

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 $d_af = d_ag$ 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_aq^i which is written in coordinates as $(0, 0, \ldots, 1, \ldots, 0)$, where the 1 is in the i^{th} place. Hence $d_aq^1, \ldots, d_aq^n$ 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 $\tau_ax^1, \ldots, \tau_ax^n$ 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> $\tau_ax^1, \ldots, \tau_ax^n$ form a basis of T_aU, a vector space ;
$d_aq^1, \ldots, d_aq^n$ 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, $\tau_ax^1, \ldots, \tau_ax^n$ form a vector-space basis of T_aU, and that it is the dual basis of $d_aq^1, \ldots, d_aq^n$. The τ_ax^i span T_aU, since if c is any path,

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

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

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

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

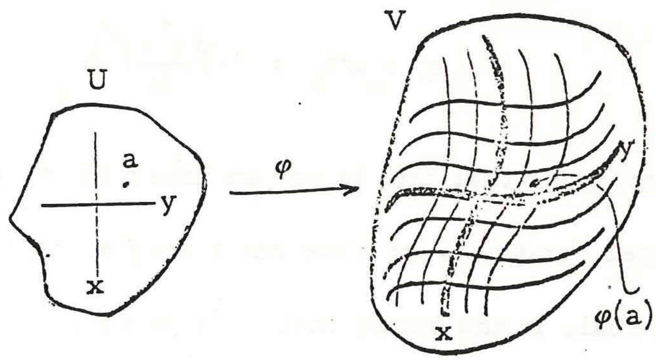
and hence

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

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



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

$$\begin{tikzcd}
d_a(\tilde{f} \circ \varphi) & T^a U & T_a U \arrow[d, "T_a\varphi = \varphi_*"] & \tau_a c \arrow[d, squiggly] \\
d_{\varphi(a)} \tilde{f} \arrow[u, squiggly] & T^{\varphi(a)} V \arrow[u, "T^a\varphi = \varphi"'] & T_{\varphi(a)} V & \tau_{\varphi(a)} \varphi(c)
\end{tikzcd}$$

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

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

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_aU to every point a of U, such that the correspondence $a \rightsquigarrow Xa$ is