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

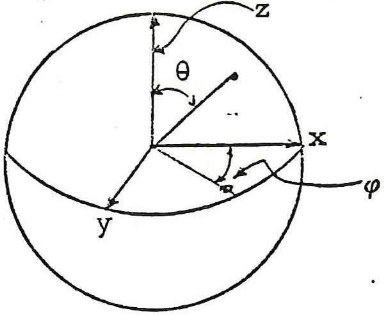
GEOMETRICAL MECHANICS

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 $\varphi - \theta$ 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 $f: \begin{array}{c} X \longrightarrow Y \\ x \rightsquigarrow f(x) \end{array}$ 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> $\operatorname{Hom}_K(A, B) = \{f: A \longrightarrow B \mid f$ is a homomorphism of K-modules$\}$

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

The set $\operatorname{Hom}_K(A, B)$ 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 $K \times K \xrightarrow{\cdot} K$)

$A^* \overset{\text{def.}}{=} \operatorname{Hom}_K(A, K)$ is called the <u>dual</u> (or conjugate) of A. For example, if $K = \mathbb{R}$, and $A = V$ a vector space, then $V^* = \operatorname{Hom}\ (V, \mathbb{R})$ is the dual space. If V is finite dimensional with basis $e_1, \ldots, e_n$, then V^* has the dual basis $e^1, \dots, e^n$ 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, $e^i(\sum_{i=1}^{n} e_j x^j) = x^i$.

Further examples of rings and modules:

$\mathbb{R}[x]$ = the ring of all polynomials in x with real coefficients
$= \{a_0 + a_1 x + \dots + a_k x^k \mid k \geq 0,\ a_i \in \mathbb{R},\ 0 \leq i \leq k\}$.

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

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

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

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 $(\mathbb{Z}_3)^* = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_3, \mathbb{Z}) = \{f: \mathbb{Z}_3 \longrightarrow \mathbb{Z} \mid f$ 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

$(f,v) \overset{\text{def.}}{=\!=} f(v) \in K$, 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 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

$$\begin{tikzcd} V \arrow[r, equal] & V^{**} \\ v \arrow[r, squiggly] & \bar{v} \end{tikzcd}$$

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 $v = e_1 x^1 + \ldots + e_n x^n$, 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 $W \times W \xrightarrow{\cdot} \mathbb{R}$ $((v, w) \rightsquigarrow v \cdot w)$ 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 $v \neq 0 \Longrightarrow v \cdot v > 0$

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 $(a, x) = \left((a_1, \ldots, a_n), \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$ 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 $V \cong V^*$ $(v \rightsquigarrow \tilde{v})$, 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,

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

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

Proof. 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

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

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} .

Definition. 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

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

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:

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

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 $f: \mathbb{R}^n \to \mathbb{R}^n$), 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.