VARIATIONAL AND TENSOR CALCULUS[1]

U. H. Gerlach

August 21, 2025

Beta Edition

[1]gerlach.1@osu.edu

Contents

[1]In this subsection vectors are identified by Latin letters with or without *sub*scripts, while their components also by Latin letters but with *super*scripts, as, for example, in

$$a = a^i e_i \equiv a^1 e_1 + \cdots + a^n e_n \ .$$

By contrast covectors are identified by Greek letters with or without *super*scripts, while their components also by Greek letters but with *sub*scripts, as, for example, in

$$\alpha = \alpha_j \omega^j \equiv \alpha_1 \omega^1 + \cdots + \alpha_n \omega^n \ .$$

Chapter 1

FUNDAMENTAL IDEAS

Lecture 1

1.1 Multivariable Calculus as a Prelude to the Calculus of Variations.

Calculus of several variables deals with the behaviour of (multiply) differentiable functions whose domain is spanned by a finite number of coordinates.

The utility of this calculus stems from the fact that it provides, among others, methods for finding the "critical" points of a given function. For, say $f(x^1, x^2, \cdots, x^n)$, this function has points where it has an extremum, which is characterized by

$$\frac{\partial f}{\partial x^i}(x_0^1, x_0^2, \ldots, x_0^n) = 0$$

The critical point $\vec{x}_0 = (x_0^1, x_0^2, \ldots, x_0^n)$ is a point where the function has a maximum, a minimum or an inflection behaviour. Which type of these extremal behaviours the function has in the neighborhood of its critical point can, in general, be inferred from the properties of the (symmetric) second derivative matrix,

$$\left[\frac{\partial^2 f(\vec{x}_0)}{\partial x^i \partial x^j}\right] \quad \left.\begin{matrix} i \\ j \end{matrix}\right\} 1, \ldots n$$

at the critical point $\vec{x}_0$.

The *calculus of variations* extends these methods from a finite to an infinite dimensional setting.

The distinguishing feature is that each point in this infinite dimensional setting is a *function*. Furthermore, the task of finding the extremum of a function becomes the task of finding the extremum of a "functional".

A "functional" is a map whose domain is the infinite dimensional space, each of whose points is a function. A "critical" point in this space is an "optimal" function which maximizes, minimizes, or in general, extremizes the given functional.

1.2 Some Typical Problems in the Calculus of Variations.

Nearly all problems in mathematical engineering, physics and geometry are closely related in one form or another to the calculus of variations. It turns out that these problems can be expressed in terms of some kind of optimization principle which says that some functional must be maximized, minimized or, in general, extremized. Let us consider some typical examples of such problems.

Problem 1. (The Suspended Cable Problem)

Consider a pearl necklace of fixed length ℓ and $n+1$ pearls. This necklace is a discrete approximation to a continuous cable.

Question: What is the shape of this necklace when suspended freely between two fixed points?

Answer: Its shape is such as to minimize the total potential energy (P.E.) of the system pearls. As depicted in in Figure 1, this shape is specified by the locations

$$y_i(x_i) \text{ or } x_i(y_i); \quad i = 0, 1, \cdots, n$$

of its $n+1$ pearls.

Figure 1.1: FIGURE GOES HERE. Necklace of fixed length suspended between fixed points $(x_0, y_0) = (a, b)$ and $(x_n, y_n) = (c, d)$

.

The total potential energy of the freely suspended necklace is

$$P.E.(x_0, x_1, \cdots, x_n) = -mg \sum_{i=0}^{n} x_i. \tag{1.1}$$

The length of the necklace is

$$\ell = \sum_{i=0}^{n} d_i = \sum_{i=0}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{1.2}$$

The fixed ends are at

$$(x_0, y_0) = (a, b) \text{ and } (x_n, y_n) = (c, d) \tag{1.3}$$

The task is: FIND

$$y_i(x_i) \text{ or } x_i(y_i) for i = 1, \cdots, n-1$$

such that

$$P.E. = \text{minimum},$$

but subject to the constraints

$$\sum_{i=0}^{n} d_i = \ell \text{ fixed}$$

and

$$(x_0, y_0) = (a, b) \text{ and } (x_n, y_n) = (c, d).$$

Comment:
This problem is a multivariable calculus problem, easy to set up and easy to solve for $n = 2$ (three pearls), but becomes challenging for $n = 4, 5, \cdots$. However, as we shall see, it becomes easily solvable when $n \to \infty$ and the necklace becomes a continuous cable suspended between two fixed points.

Problem 2. (The Brachistochrone Problem)

Given: A particle moves along a curve $y(x)$ without friction. The particle starts with zero velocity from point $(x_1, y_1) = (0, 0)$ and advances, because of a vertical constant gravitational field, to the point (x_2, y_2).

Question: For what curve is the travel time a minimum?

Let us formulate the problem more precisely. The travel time is given by the integral

$$\text{time} = \int_{t_1}^{t_2} dt = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{d(\text{path length})}{\text{speed}} \equiv \int_{1}^{2} \frac{ds}{v} \ .$$

For a typical curve $y(x)$ with slope $\dfrac{dy}{dx}$, the speed v and the path length differentials are obtained from the following considerations:

(a) *time independent* gravitational field *implies* that the *total energy* of a system *is conserved*, i.e. independent of time. Hence,

$$\text{T.E.} \equiv \text{K.E.}+\text{P.E.} = \text{constant} \ ,$$

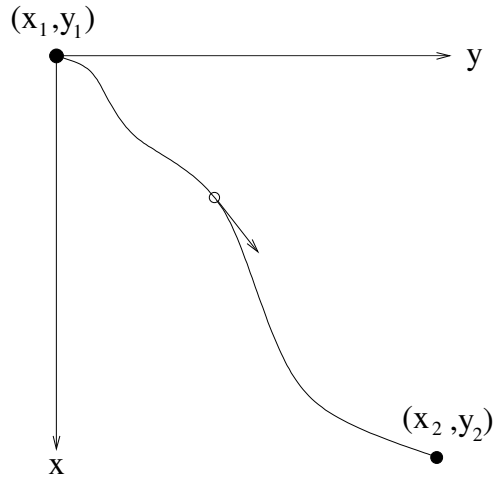


Figure 1.2: Particle sliding down a bent wire.

or

$$\frac{1}{2}mv^2 - mgx = 0 \ .$$

Here we have chosen the P.E. energy to be zero at $x = 0$:

$$\text{P.E.}(x = 0) = 0$$

Thus

$$v = \sqrt{2gx} \ .$$

Comment: The fact that the total energy is independent of time,

$$\text{T.E.=constant} \ ,$$

is an illustration of a general principle, the principle of energy conservation. It applies to all dynamical systems evolving in a time-invariant environment, e.g. a freely swinging pendulum or vibrating body. It does not apply to dynamical systems subjected to externally applied driving forces which depend on time.

(b) The element of path length is

$$ds = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + \left(\frac{dy}{dx}\right)^2} \ .$$

The total travel time is therefore

$$t_{12} = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}{\sqrt{2gx}} \, dx \ ,$$

which is a function of the function $y(x)$, i.e. t_{12} is a "functional".

Thus, to solve the brachistochrone problem requires that we find a function $y(x)$ such that the "functional" t_{12}=minimum!

Problem 2. (Minimum Surface of Revolution)

Given: Some curve between two fixed points (x_1, y_1) and (x_2, y_2) in the plane $z = 0$. A surface of revolution is formed by revolving the curve around the y-axis.

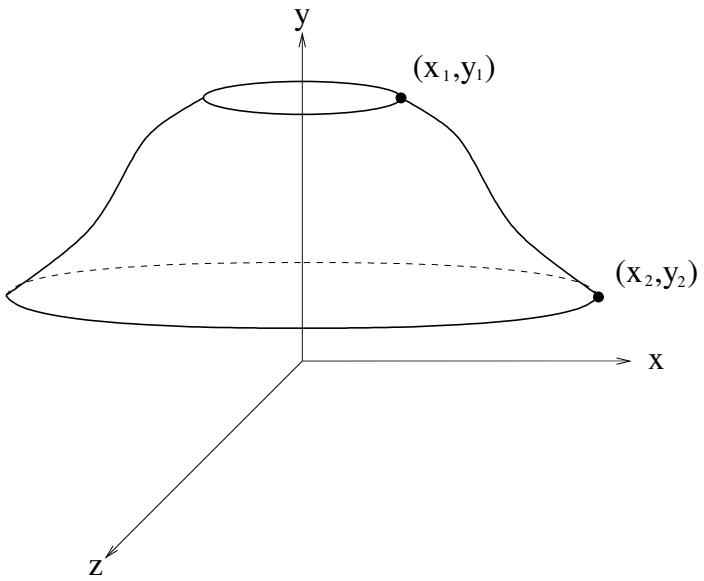


Figure 1.3: Surface area of revolution generated by a curve in the $x - y$ plane.

Question: For what curve $y(x)$ is the surface area a minimum?

One must find a function $y(x)$ such that the functional

$$\begin{aligned} \text{Area}[y] &= \int_1^2 2\pi x ds = 2\pi \int_{(x_1,y_1)}^{(x_2,y_2)} x\sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx \\ &= \text{minimum!} \end{aligned}$$

Problem 3. (Fermat's Principle of Least Time)

Given: An inhomogeneous medium through which light (or sound) can propagate with a velocity that depends on position,

$$v(x, y) = \frac{c}{n(x, y)} \ .$$

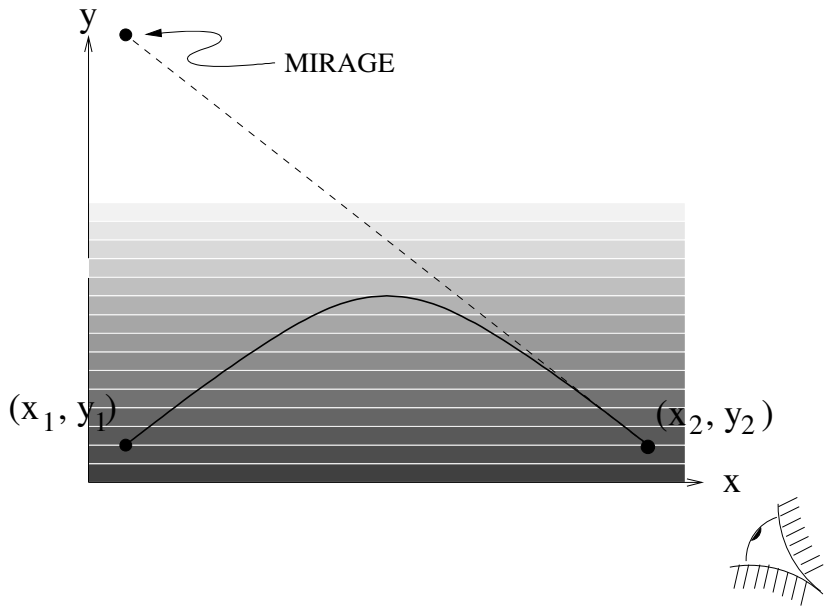


Figure 1.4: Optical path as determined by Fermat's principle of least time. In the lighter regions the propagation speed of light is faster. Thus the optimal (least time) path includes such regions in its domain of passage.

Here $n(x, y)$ is the position dependent refractive index, which is assumed to be given.

Statement of Fermat's Principle:
A light (or a radio, or a sound) beam takes that path $y(x)$ through two points (x_1, y_1) and (x_2, y_2) for which the travel time is a minimum.

This principle subsumes many natural phenomena, including 1.) mirages [1], 2.) ionospheric bending of radio beacons, 3.) twinkling of stars[2], 4.) sonar beams refracted by underwater temperature and salinity changes, 5.) light beam traversing an optical interface, etc.

The application of Fermat's principle consists of finding $y(x)$ such that

$$t_{12} = \int_1^2 \frac{ds}{v(x,y)} = \int_{x_1}^{x_2} n(x,y) \frac{\sqrt{1+(y')^2}}{c}\, dx = \text{minimum!}$$

[1] Due to a refractive index increasing with altitude.
[2] Due to light rays from a star propagating through atmospheric plumes of warm and cold air.

1.3 Methods for Solving Problems in Calculus of Variations.

There are three known ways of finding an optimal function which minimizes (or maximizes) the given integral expression for the functional.

1. the method of finite differences,
2. the method of variations,
3. the direct method via minimizing sequences.

We shall briefly discuss the first, extensively develop the second, and postpone discussion of the third until later.

1.3.1 Method of Finite Differences.

In order to appreciate the basic meaning of problems in the calculus of variations and methods for solving them, it is important to see how they are related to the calculus of n variables. We are given a functional of the form

$$J[y] = \int_a^b F(x, y, y')dx \quad \text{with } y(a) = A \ \text{ and } \ y(b) = B.$$

This is a map which assigns a number to each curve $y(x)$ passing through the two points (a, A) and (b, B). One may find the curve which minimizes the integral J by the following approximation scheme, which is based on the construction of a Riemann sum with the derivative function approximated by the partition-induced set of differential quotients:

1. Subdivide the interval $[a, b]$ into $n + 1$ equal parts by using the points

$$x_0 = a, \ x_1, \ldots, x_n, \ x_{n+1} = b$$

2. Replace the curve $y(x)$ with the polygonal line whose vertices are

$$(x_0, A), (x_1, y(x_1)), \ldots, (x_n, y(x_n)), (x_{n+1}, B)$$

3. Approximate the integral $J[y]$ by a finite Riemann sum,

$$J(y_1, \ldots, y_n) = \sum_{i=1}^{n+1} F\left(x_i, y_i, \frac{y_i - y_{i-1}}{\Delta x}\right) \ \Delta x \ ,$$

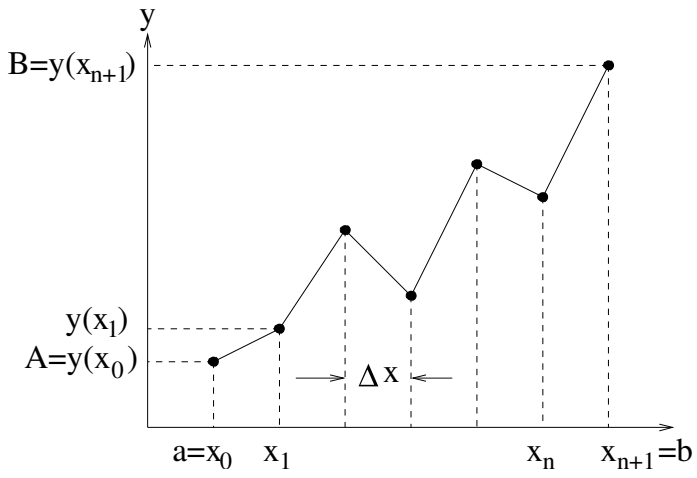


Figure 1.5: Curve approximated by a polygonal line.

where

$$y_i = y(x_i) \text{ and } \Delta x = x_i - x_{i-1}$$

Thus, each polygonal line is uniquely determined by the ordinates $y_1, y_2, \ldots, y_n$ of the vertices because the end points are assumed to be *fixed*. Consequently, the *functional* $J[y]$ may be considered as *being approximated* by the function $J(y_1, \ldots, y_n)$, and the problem of finding the optimal curve $y(x)$ which minimizes $J[y]$ may be approximated by the problem of finding the critical point $(y_1^*, \ldots, y_n^*)$ which minimizes the finite Riemann sum $J(y_1, \ldots, y_n)$.

This minimization process consists of solving the following n simultaneous equations

$$\frac{\partial J}{\partial y_j} = 0 \qquad j = 1, \ldots, n$$

Using the expression for J, namely

$$\begin{aligned} J(y_1, \ldots, y_n) &= F\left(x_1, y_1, \frac{y_1 - y_0}{\Delta x}\right)\Delta x + F\left(x_2, y_2, \frac{y_2 - y_1}{\Delta x}\right)\Delta x \\ \ldots &+ F\left(x_j, y_j, \frac{y_j - y_{j-1}}{\Delta x}\right)\Delta x + F\left(x_{j+1}, y_{j+1}, \frac{y_{j+1} - y_j}{\Delta x}\right)\Delta x \\ + \ldots &+ F\left(b, B, \frac{B - y_n}{\Delta x}\right)\Delta x, \end{aligned}$$

let us calculate the partial derivative w.r.t y_j:

$$\begin{aligned}\frac{\partial J}{\partial y_j} = \frac{\partial}{\partial y_j}\ \{F(x_j, y_j, \underbrace{\frac{y_j - y_{j-1}}{\Delta x}}_{y'_{j-1}})\ \Delta x \quad &+ \quad F(x_{j+1}, y_{j+1} \underbrace{\frac{y_{j+1} - y_j}{\Delta x}}_{y'_j})\ \Delta x\} \\ = F_{y_j}(x_j, y_j, \frac{y_j - y_{j-1}}{\Delta x})\ \Delta x \quad &+ \quad F_{y'_j}(x_j, y_j, y'_{j-1}) \frac{1}{\Delta x} \Delta x \\ &+ \quad F_{y'_j}(x_{j+1}, y_{j+1}, y'_j)\ \frac{-1}{\Delta x}\ \Delta x\end{aligned}$$

As $\Delta x \to 0$; x_j, $x_{j-1} \to x$; $y_j \to y$ the number of subintervals increases without limit and the expression approaches zero. But if we divide by Δx, then we have

$$\lim_{\Delta x \to 0}\ \frac{1}{\Delta x}\ \frac{\partial J}{\partial y_j} \equiv \frac{\delta J}{\delta y}$$

Remark: The expression $\Delta x \partial y_j$ has a direct geometrical significance as the area between the given curve and the varied (dashed) curve. The limit of the differential

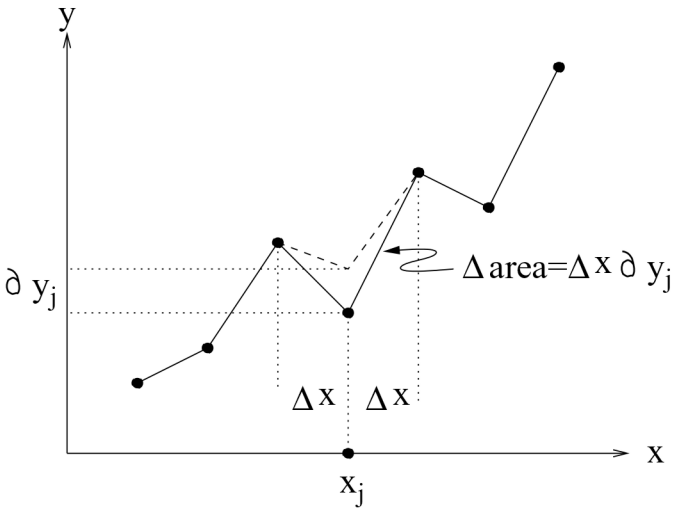


Figure 1.6: The area between the solid and the dashed curve is $\Delta x \partial y_j$. In the continuum limit, Figure 1.12, the solid (given) curve becomes a "trial function" and the dashed (varied) curve becomes its "variant".

quotient is the variational derivative of J,

$$\frac{\delta J}{\delta y} = \frac{\partial F(x, y, y')}{\partial y} - \frac{d}{dx}\ \frac{\partial F(x, y, y')}{\partial y'}$$

The process of minimizing the functional $J[y]$ by first approximating it as $J(y_1, \ldots, y_n)$, a function of the n independent variables $y_1, \ldots, y_n$, is called the *method of finite differences*.

Leonard Euler used it to solve problems in the calculus of variations. By replacing smooth curves with polygonal lines he solved such problems as multivariable problems in n-dimensions. By letting $n \to \infty$, he obtained the mathematically exact[3] solution.

Thus the functional $J[y]$ may be considered a function of infinitely many variables, and the *calculus of variations may be considered an extension* of multivariable calculus *into a space of infinite dimensions*.

Lecture 2

1.4 The Method of Variations.

The method of finite differences first approximates an infinite dimensional problem by one of finite dimensions. Solving the finite dimensional problem consists of finding the critical point of a scalar function having a finite number independent variables. The coordinates of this critical point, say $\{y_i = y(x_i) : i = 1, \ldots, n\}$, are the vertices of a polygonal curve, the finite approximation to the curve that optimizes the scalar functional. One obtains the exact solution, the optimal curve, to the infinite dimensional problem by letting n, the dimension of the approximation space, become infinite. There exist numerical optimization techniques for finding the critical point scalar function on a finite dimensional approximation space. One obtains the exact solution to the infinite problem by letting n, the dimension of the approximation space, become infinite. But in that case the numerical optimization technique becomes very computationally expensive.

However, one can dispense with the intermediate approximation process entirely by dealing with the infinite dimensional problem directly. In this strategy one does not obtain the optimal curve directly. Instead, one obtains a differential equation which this curve must satisfy. However, finding solutions to a differential equation is a highly developed science. Hence, dispensing with the intermediate approximation process is an enormous advantage if, as is usually the case, one can solve the differential equation.

It is difficult to overstate the importance of the method which allows us to do this. It is called the method of variations, which we shall now describe.

[3]Mathematical exactness is the limit of *physical exactness*. The latter presupposes a specific context of the measurements underlying the results. This context is limited both by the *range* and the *precision* of the measured data it subsumes. Thus, given such a context, a result is said to be physically exact in relation to such a specified context. The process of going to the above limit of physical exactness is made mathematically precise by means of the familiar $\delta - \epsilon$ process.

1.4.1 Variants and Variations

The basic problem of the calculus of variation is this:

Given: (a) $F(x; q, p)$ a function of three variables and (b) a pair of points (x_1, y_1) and (x_2, y_2).

Determine: That "optimal" function $y(x)$ which passes through (x_1, y_1) and (x_2, y_2) such that

$$J[y] = \int_{x_1}^{x_2} F(x; y(x), y'(x))dx \tag{1.4}$$

has the largest (or smallest or, in general, extremal) value for $y(x)$ as compared to any *variant* of $y(x)$ also passing through (x_1, y_1) and (x_2, y_2).
Remark 1: The dictionary definition of "variant" is as follows:

variant – something that differs in form only slightly from something else. Remark

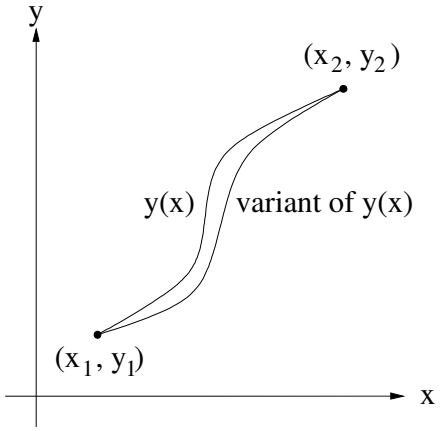


Figure 1.7: A function $y(x)$ and one of it variants

2: In order to determine whether $y(x)$, which passes through (x_1, y_1) and (x_2, y_2), is an "optimal" curve, one must compare the functional J evaluated for all variants of $y(x)$, including $y(x)$ itself.
Remark 3: We shall *assume* that an optimal curve exists, and then deduce as a consequence the conditions that $y(x)$ must satisfy. In other words, we shall determine the *necessary conditions* for $y(x)$ to exist. We are *not* developing *sufficient conditions*[4] for the existence of an optimal curve.

[4]Sufficient conditions for the functional $J[y]$ to have an extremum are discussed and given in Chapters 5 and 6 in *Calculus of Variations* by I.M. Gelfand and S.V. Fomin; Dover Publications, Inc., Mineola, N.Y., 2000.

To test whether a curve is optimal by comparing values of J, pick a curve suspected of being an optimal one and deform it slightly. This yields a *variant*, a *deformed curve*, of the original curve, namely

$$\begin{aligned} y &= f(x, \alpha_i) \\ &\equiv f(x, 0) + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots \\ &\equiv f(x, 0) + \delta f(x) \end{aligned}$$

Here

$$y = f(x, 0) \equiv f(x)$$

is the *optimal curve* under consideration, while

$$\delta f(x) = \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots$$

is a *variation* in the function f, a variation which is a generic sum of linearly independent functions $\{h_k(x)\}$.

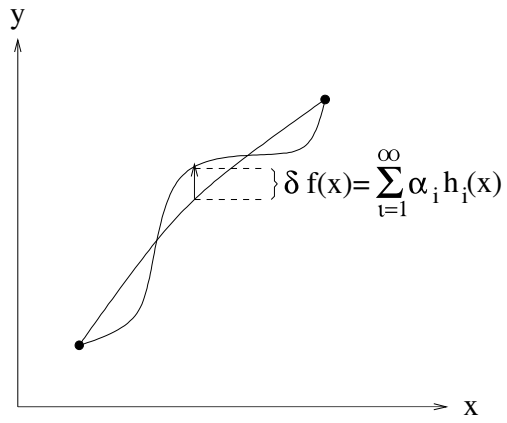


Figure 1.8: Arbitrary variation of f as a superposition of basis functions.

Definition. The fact that all variants of f pass through the same pair of endpoints we express by saying that they are *admissible* functions; in other words admissible functions have the property that their variations vanish at the endpoints,

$$\boxed{\delta f(x_1) = \delta f(x_2) = 0}.$$

Because of the linear independence of the h_i's, this is equivalent to

$$\boxed{h_i(x_1) = h_i(x_2) = 0 \quad i = 1, 2 \dots}$$

Relative to the set of basis functions $\{h_i\}$ the variants

$$\begin{aligned} A: &\quad f(x,0) - .2h_1(x) \\ B: &\quad f(x,0) - .1h_1(x) \\ C: &\quad f(x,0) \\ D: &\quad f(x,0) + .1h_1(x) \\ E: &\quad f(x,0) + .2h_1(x) \\ F: &\quad f(x,0) + .2h_1(x) + .1h_2(x) \\ G: &\quad f(x,0) + .2h_1(x) + .2h_2(x) \end{aligned}$$

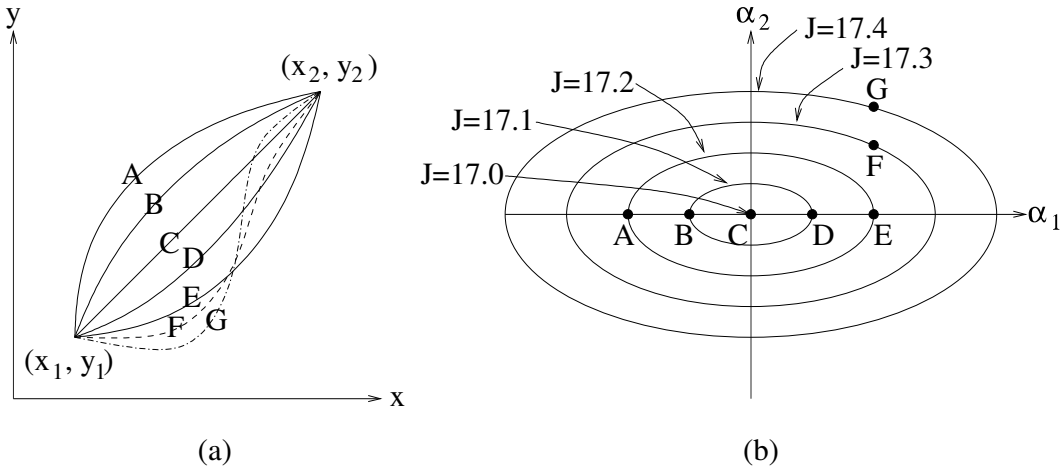


Figure 1.9: (a) Optimal curve (C) and some of its variants in x-y space. (b) Isograms of J on the function space of variants.

are represented as points

$$\begin{aligned} A: &\quad (-.2, 0, \dots) \\ B: &\quad (-.1, 0, \dots) \\ C: &\quad (0, 0, \dots) \\ D: &\quad (.1, 0, \dots) \\ E: &\quad (.2, 0, \dots) \\ F: &\quad (.2, .1, \dots) \\ G: &\quad (.2, .2, \dots) \end{aligned}$$

in the function space*coordinatized* by α_1, α_2, etc. as depicted in Figure 1.9.

Lecture 3

1.4.2 The Euler-Lagrange Equation

The problem of variational calculus is the task of optimizing Eq.(1.4). In particular, the mathematization of the calculus of variations consists of setting up the problem, defining the framework, its key concepts, and the causal relations between them.

The solution to the problem consists of identifying the optimization method in mathematical terms.

I. The Problem

The achievement of these goals requires three steps.
(i) Specify the space of trial functions on which that integral is defined,
(ii) recognize that this integral J is a mapping from these trial function into the reals, and
(iii) find the critical point(s) in the domain space for which the mapping has extremal values in this space, i.e. find the optimal function(s).

The space of trial functions is a non-linear submanifold of the linear space of functions which have the *norm*

$$\|y(x)\| = \max_{x_1 \le x \le x_2} |y(x)| + \max_{x_1 \le x \le x_2} |y'(x)| < \infty \ . \tag{1.5}$$

These functions form a vector space. Because of its norm structure it is called a "Banach" space.

The space of trial functions does *not* consist of *all* Banach space elements, only of those functions $y(x)$ whose graphs run through the fixed ends points (x_1, y_1) and (x_2, y_2) as in Figures 1.7, 1.8, 1.9(a):

$$y(x_1) = y_1 \text{ and } y(x_2) = y_2 \ . \tag{1.6}$$

The set of functions that satisfy Eqs.(1.5) and (1.6) (integrability and the end point conditions) forms the set of the already mentioned *trial functions*. For emphasis one also calls them *admissible* trial functions. Taking into account

1. that they make up the above-mentioned non-linear submanifold, S,

$$S = \{y : y(x) \text{ satisfies } Eqs.(1.5) \text{ and } (1.6)\}, \tag{1.7}$$

   the set of admissible functions, and

2. that the variational integral

$$J[y] = \int_{x_1}^{x_2} F\left(x; y, \frac{dy}{dx}\right) dx. \tag{1.8}$$

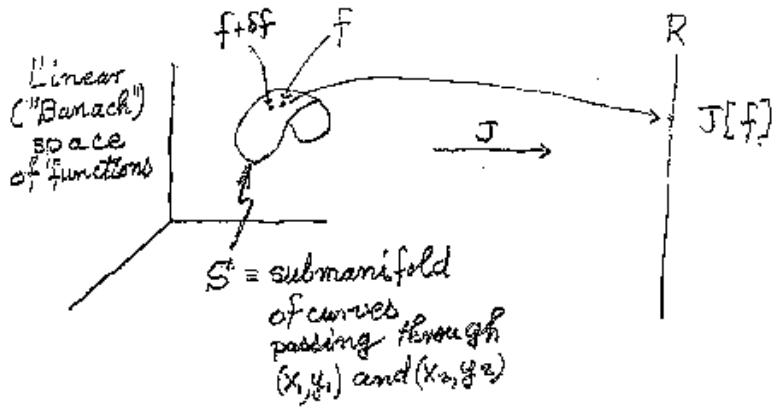


Figure 1.10: The functional J maps its domain space, the submanifold S, into the target space R, the reals

being real-valued, one arrives at the conclusion that, as depicted in Figure 1.10, J is a real-valued functional that maps the submanifold into the reals

$$J: \quad S \rightarrow R = \text{reals} \tag{1.9}$$

$$y \rightsquigarrow J[y] = \int_{x_1}^{x_2} F\left(x; y, \frac{dy}{dx}\right) dx \tag{1.9'}$$

3. Among the set S of admissible functions find the function f which extremizes $J[y]$.

II. The Solution

Based on observations and experiments (ultimately via one's five senses) we have already illustrated the existence of optimal solutions to some typical problems in the calculus of variations. Denoting an optimal solution by f, one mathematize the method for finding it by calculating

$$J[f + \delta f] - J[f] \equiv \Delta J$$

and setting its principal linear part to zero. The calculation is very compact and proceeds as follows:

$$\begin{aligned}\Delta J &= \int_{x_1}^{x_2} \left\{ F(x;y,y')|_{y=f(x)+\delta f(x)} - F(x;y,y')|_{y=f(x)} \right\} dx \\ &= \int_{x_1}^{x_2} \Bigg\{ \cancel{F(x;f,f')} + \frac{\partial F}{\partial y}(x;f,f')\delta f(x) + \frac{\partial F}{\partial y'}(x;f,f') \overbrace{\frac{d(\delta f(x))}{dx}}^{\delta(f'(x))} \\ &\qquad + \begin{pmatrix} \text{higher} \\ \text{order} \\ \text{non-linear} \\ \text{terms} \end{pmatrix} - \cancel{F(x;f,f')} \Bigg\} dx\end{aligned}$$

The Taylor series expansion theorem was used to obtain this result. It depends on the variation δf and its derivative. However, we would like to have the result depend only on δf, and not on its derivative.This desire can be realized by integrating the second term by parts. This means, we recall, that one use $\frac{\partial F}{\partial y'}\frac{d(\delta f)}{dx} = \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\delta f\right) - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\delta f$. One obtains

$$\Delta J = \underbrace{\int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} \right) \delta f(x) dx + \frac{\partial F}{\partial y'}\delta f(x)\Big|_{x_1}^{x_2}}_{\text{“Principal Linear Part” of } \Delta J} + \begin{pmatrix} \text{higher} \\ \text{order} \\ \text{non-linear} \\ \text{terms} \end{pmatrix} \tag{1.10}$$

The fact that $f + \delta f \in S$ implies that

$$\delta f(x_1) = \delta f(x_2) = 0,$$

which we specified already on page 18. Consequently, the boundary contribution to ΔJ vanishes. We now consider the *Principal Linear Part* of ΔJ, which is linear in δf. We designate it by δJ, and find

$$\delta J = \int_{x_1}^{x_2} \left[F_y - \frac{d(F_{y'})}{dx} \right] \delta f(x) dx. \tag{1.11}$$

This is the *1st variation* of the functional J at $y = f(x, 0)$. It is a linear functional which is also called the *differential* of J.

1.4.3 Variational Derivative

The fact that an arbitrary variant of $f(x,0)$ is given by

$$f(x,\alpha_i) = f(x,0) + \sum_{i=1}^{\infty} \alpha_i h_i(x)$$

implies that the variations of f,

$$\delta f(x) = \sum_{i=1}^{\infty} \alpha_i h_i(x) \tag{1.12}$$

form a linear vector space which is spanned by the basis vectors $\{h_i(x) :\ i = 1, 2, \dots\}$. Taking a cue from multivariable calculus, we can use the differential δJ of J, and identify the *directional derivatives* of J along the respective directions h_k in the space of admissible functions defined on page 18.

These directional derivatives, we recall, are given with the help of

$$\Delta J = \overbrace{J(\alpha, \cdots, \alpha_k, \cdots)}^{J(x_i)} - J(0, \cdots, 0, \cdots),$$

and of Eqs.(1.11) and (1.12) by

$$\begin{aligned} \left.\frac{\partial J(\alpha_i)}{\partial \alpha_k}\right|_{\alpha_i=0} &= \lim_{\alpha_k \to 0} \frac{\Delta J(0, \dots, \alpha_k, \dots)}{\alpha_k} \\ &= \int_{x_1}^{x_2} \left[F_y - \frac{d}{dx}(F_{y'})\right] h_k(x)dx, \quad k = 1, 2, \dots \end{aligned} \tag{1.13}$$

$$\tag{1.14}$$

In order to establish the parallel with multivariable calculus, let us designate

$$F_y - \frac{d}{dx}\ F_{y'} \equiv \frac{\delta J}{\delta y(x)} \qquad x_1 < x < x_2 \tag{1.15}$$

as the (*first*) *variational derivative of* $J[y]$.

It is clear that this derivative depends on x.

Reminder:

Recall that in multivariable calculus the gradient $\vec{\nabla} f$ of a function is related to its *directional derivative* $D_{\vec{e}_k} f$ along the same basis vector e_k by the equation

$$D_{e_k} f = (\vec{\nabla} f) \cdot \vec{e}_k$$

One may look at the right hand side of the directional derivative, Eq. (1.13), in the same way. It follows that the variational derivative of the functional J,

$$\frac{\delta J}{\delta y(x)} = F_y - \frac{d}{dx}\, F_{y'} \qquad x_1 < x < x_2,$$

corresponds to what in multivariable calculus is the gradient of a function. The obvious difference between the two is that the number of components of a gradient in multivariable calculus is finite and they are labeled by integers. By contrast the "gradient" of a function J, i.e. its variational derivative, has a continuous infinity of components which are labeled by the values of x ($x_1 < x < x_2$).

1.4.4 Euler's Differential Equation

The usefulness of the first variational arises when one considers an "optimal" curve, say

$$y = f(x) \tag{1.16}$$

for which J has a maximal (or minimum, or extremal) value. In such a circumstance all the directional derivates, Eq. (1.13), vanish, i.e.

$$0 = \frac{\partial J(\alpha_i)}{\partial \alpha_k} = \int_{x_1}^{x_2} \frac{\delta J}{\delta y(x)} h_k(x) dx \qquad k = 1, 2, \ldots$$

It follows that for $x_1 < x < x_2$

$$\frac{\delta J}{\delta y(x)} \equiv F_y(x; y(x), y'(x)) - \frac{d}{dx} F_{y'}(x; y(x), y'(x)) = 0. \tag{1.17}$$

This is *Euler's differential equation.* This is the equation which the curve, Eq. (1.16), must satisfy if it extremizes the functional J.

Why must one have $\dfrac{\delta J}{\delta y(x)} = 0$ for $x_1 < x < x_2$? Suppose $\dfrac{\delta J}{\delta y(x)} \neq 0$ in the neighborhood of some point $x = x'$. Then from the set

$$\{h_k(x) : k = 1, 2, \ldots\}$$

we choose or construct a localized "blip" function[5] (see Figure 1.11)

$$\delta f(x) = \sum_{k=1}^{\infty} \beta_k h_k(x) \ .$$

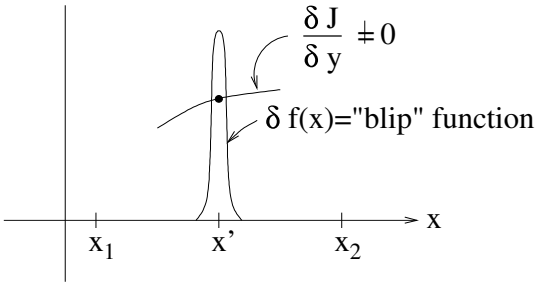


Figure 1.11: A "blip" function is a variation which is differentiable but is nonzero only in a neighborhood around a point x'.

A "blip" function is a type of variation $\delta f(x)$, which is the difference between a trial function $f(x)$ and its variant, say, $f(x)+\delta f(x)$, as in Figure 1.12. Its use would yield

$$0 \neq \int_{x_1}^{x_2} \frac{\delta J}{\delta y(x)} \ \delta f(x) dx = \sum_{k=1}^{\infty} \frac{\partial J}{\partial \alpha_k} \beta_k$$

This contradicts the fact that

$$0 = \frac{\partial J}{\partial \alpha_k} = 0 \qquad \forall \ k.$$

The conclusion is therefore that Euler's differential Eq.(1.17) is indeed the equation which an optimal curve must satisfy. Explicitly, with $F = F(x; y(x), y'(x))$, this equation is

$$\boxed{\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0} \ ,$$

[5] An example of a smooth (i.e. infinitely differentiable) blip function which vanishes for

$$r \equiv |x - x'| \geq \epsilon$$

is

$$\delta f(x) = \begin{cases} \exp\left(-\dfrac{\epsilon^2}{\epsilon^2 - r^2}\right) & \text{for } r < \epsilon \\ 0 & \text{for } r \geq \epsilon. \end{cases}$$

.

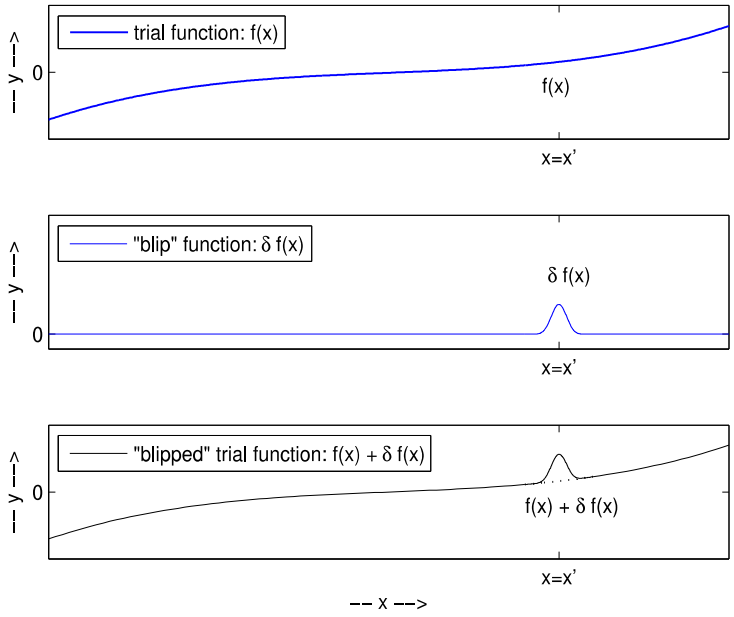


Figure 1.12: The area under any "blip" function equals the area between the trial function $f(x)$ and its variant $f(x) + \delta f(x)$ as depicted in the bottom panel. This area is what in Euler's discrete formulation is the area between a polygonal curve and its variant (the dashed one) in Figure 1.6.

or

$$y'' F_{y'y'} + y' F_{y'y} + F_{y'x} - F_y = 0 \ .$$

This Euler's differential equation is one of second order. It constitutes a necessary condition which $y(x)$ must satisfy if it is to be an optimal curve for the functional $J[y]$.

The Euler equations for the three extremum problems in Section 2 are as follows:

1. Brachistochrone:

$$F = \sqrt{\frac{1 + y'^2}{x}} \qquad \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + y'^2}}\ \frac{1}{\sqrt{x}}\right) = 0$$

2. Least area:

$$F = x\sqrt{1 + y'^2} \qquad xy'' + y'(1 + y'^2) = 0$$

3. Fermat:

$$F = n(x, y)\sqrt{1 + y'^2} \qquad ny'' = (n_y - n_x y')(1 + y'^2)$$

1.5 Solved Example

Let us solve the Brachistochrone problem for which $F = \sqrt{\frac{1+y'^2}{x}}$. The fact that $\frac{\partial F}{\partial y} = 0$ implies that the Euler equation is

$$0 = \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) \ .$$

Consequently

$$\frac{y'}{\sqrt{1+y'^2}}\ \frac{1}{\sqrt{x}} = \text{const} \equiv \left(\frac{1}{2a}\right)^{1/2} \ .$$

Square the result to obtain

$$\frac{1}{1+y'^2}\ \frac{y'^2}{x} = \frac{1}{2a}$$

or

$$y'^2 2a = x + xy'^2 \ .$$

Thus

$$\left(\frac{dy}{dx}\right)^2 = \frac{x}{2a-x} = \frac{x^2}{2ax-x^2} \ . \tag{1.18}$$

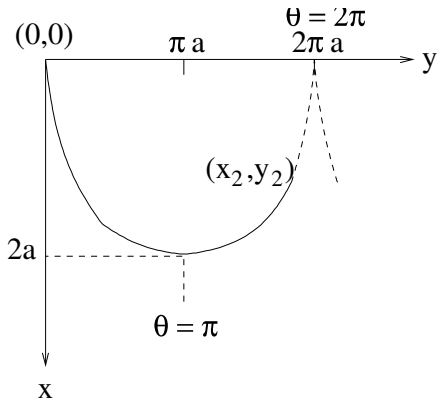


Figure 1.13: The solution curve for the brachistochrone problem is a cycloid.

Equation (1.18) can be integrated. One obtains

$$\begin{aligned} y(x) &= \int_{x_1=0}^{x} \frac{x dx}{(a^2 - a^2 + 2ax - x^2)^{1/2}} \\ &= \int_{x_1=0}^{x} \frac{x dx}{\sqrt{a^2 - (a-x)^2}} \end{aligned}$$

We make the substitution $a - x = a\cos\theta$, or

$$x = a(1 - \cos\theta)$$

to obtain

$$y = \int_{\theta_1}^{\theta} \frac{a(1-\cos\theta)a\sin\theta}{a\sin\theta} d\theta \ .$$

Thus we have

$$y = a(\theta - \sin\theta) + const. \ \leftarrow \text{zero to make } y_1 = 0 \text{ the starting height.}$$
$$x = a(1 - \cos\theta) \ .$$

This is a θ-parametrized cycloid whose amplitude a must be adjusted so that the curve goes through (x_2, y_2).

Lecture 4

1.6 Integration of Euler's Differential Equation.

I. The Brachistochrome problem where

$$F = \sqrt{\frac{1 + y'^2}{x}} \tag{1.19}$$

illustrates the simplification which occurs if the variational integrand F is *independent* of y. In that case

$$\frac{\partial F}{\partial y'} = \text{const.} \tag{1.20}$$

along the curve, and one has to deal only with a 1st order differential equation as was done on page 27. Equation (1.20) is called an *integral of motion.*

II. If F is *independent* of x, as, for example, in Fermat's problem for a stratified atmosphere where

$$F = n(y)\sqrt{1 + y'^2},$$

then an analogous simplification is possible. This is seen best from the "Second Form" of Euler's Equation. Euler's motivational method towards this equation was his desire to obtain an alternate differential equation which contains *no second derivatives of* y whenever his original equation

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0 \tag{1.21}$$

was satisfied. To this end he noticed that

$$\frac{dF(x,y,y')}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\ y' + \frac{\partial F}{\partial y'}\ y''$$

and

$$\frac{d}{dx}\left(y'\frac{\partial F}{\partial y'}\right) = y''\frac{\partial F}{\partial y'} + y'\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\ .$$

To obtain an expression with no derivatives higher than the first, he subtracted to eliminate the y'' term and obtained

$$\frac{d}{dx}\left(F - y'\ \frac{\partial F}{\partial y'}\right) = \frac{\partial F}{\partial x} + \left(\frac{\partial F}{\partial y} - \frac{d}{dx}\ F_{y'}\right) y'$$

By Euler's equation, Eq.(1.21), the last term vanishes and we are left with the "second form" of Euler's equation

$$\boxed{\frac{d}{dx}\left(F - y'\ F_{y'}\right) = \frac{\partial F}{\partial x}} \tag{1.22}$$

Remark:

1. If the variational integrand F is independent of x, i.e. if $\frac{\partial F}{\partial x} = 0$, then one obtains an *integral of motion*,

   $$F - y'F_{y'} = \text{constant}. \tag{1.23}$$

   This expression involves only the first derivative of the to-be-determined function $y(x)$. It is evident that we have *reduced* the problem from solving a second order differential equation to solving one of first order.

2. By switching independent variables from x to y:

   $$F\left(x, y,\ \frac{dy}{dx}\right)\frac{dx}{dy}\ dy = G\left(y, x, \frac{dx}{dy}\right)\ dy$$

   the Euler equation becomes

   $$0 = \frac{d}{dy}\left(\frac{\partial G}{\partial\left(\frac{dx}{dy}\right)}\right) - \frac{\partial G}{\partial x}\ ,$$

   which implies that

   $$\frac{\partial G}{\partial\left(\frac{\partial x}{\partial y}\right)} = const.$$

   is an integral of motion. Thus there is an alternate way of arriving at the integral of motion (1.23).

Example 1 Fermat's Principle for a Stratified Atmosphere.

The variational integrand for least time propagation in an x-translation invariant but y-stratified medium is

$$F = n(y)\sqrt{1 + y'^2}.$$

The corresponding E-L equation is a differential equation of second order, and because it *is* of second order, the method of solving it is a two-stage process.

Stage 1.

In light of the second form of Euler's equation, the first stage reduces to a single step: by inspection obtain a first integral of motion,namely

$$F - y'F_{y'} = n(y)\left[\sqrt{1 + y'^2} - \frac{y'^2}{\sqrt{1 + y'^2}}\right] = \frac{n(y)}{\sqrt{1 + y'^2}} = constant \equiv c$$

The result is a c-parametrized family of first order differential equations

$$\frac{dy}{dx} = \sqrt{\frac{n^2}{c^2} - 1}\ .$$

From its right hand side one draws a key conclusion,

$$\boxed{\text{No real solution exists whenever } n^2(y) < c^2.}$$

This mathematizes the fact that, for an integral of motion satisfying this inequality, no propagation is posible.

To concretize this feature, and to exhibit a type of stratified Fermat problem which is explicitly soluble in mathematically closed form, consider propagation in an exponential atmosphere. There the refractive index has the form

$$n(y) = n_0 e^{-y}.$$

Focus on propagation which starts at the origin $(x = 0, y = 0)$. There the initial data is

$$\left.\frac{dy}{dx}\right|_{x=0} = y'_0 \equiv \tan\alpha \text{ (“initial slope”)}, \tag{1.24}$$

$$n(y = 0) = n_0 \qquad \text{(“refractive index at ground level”)}. \tag{1.25}$$

This initial data determines the integral of motion:

$$c = \frac{n_0}{\sqrt{1 + y'^2_0}} = n_o \cos\alpha, \tag{1.26}$$

and the equation governing the corresponding unique motion is

$$\left(\frac{dy}{dx}\right) = \pm\sqrt{\frac{n^2(y)}{c^2} - 1} = \pm\sqrt{\frac{e^{-2y}}{\cos^2\alpha} - 1}. \tag{1.27}$$

The upper sign mathematizes the upward progress of the trajectory $y(x)$, the lower sign its downward progress. These two branches of $y(x)$ meet at the turning point where their slopes vanish.
Thus the first stage of the solution process yields, among others, the result the maximum height y_{max} is achieved when $\frac{dy}{dx} = 0$:

$$e^{-y_{max}} = \cos\alpha \tag{1.28}$$

or

$$y_{max} = -\log\cos\alpha. \tag{1.29}$$

This maximum height depends on the launching angle α:

$$\alpha = 0 \Rightarrow \ y_{max} = 0 \tag{1.30}$$

$$\alpha \to \frac{\pi}{2} \Rightarrow \ y_{max} \to +\infty \tag{1.31}$$

Stage 2.

That being is the case, what is the trajectory's return distance d? How does it depend on the initial angle α in Figure 1.14? The answer to these questions lies in the second integral obtained by integrating the differential Eq.(1.27). Upon solving this differential equation by separating variables, using the properties of elementary functions, and introducing the appropriate integration limits, one finds that

$$d = 2\alpha, \quad \text{where } \cos\alpha = e^{-y_{max}}. \tag{1.32}$$

Example 2. Geodesic on a Sphere.

A geodesic is a line which represents the shortest path between two given points when the path is restricted to lie on some surface. On a sphere of radius ρ the square element of length is

$$(ds)^2 = \rho^2 d\theta^2 + \rho^2 \sin^2\theta d\varphi^2$$

Hence the length of a path $\theta(\varphi)$ between two points, say 1 and 2, is

$$s = \int_1^2 ds = \rho \int_1^2 \sqrt{\left(\frac{d\theta}{d\varphi}\right)^2 + \sin^2\theta(\varphi)}\, d\varphi$$

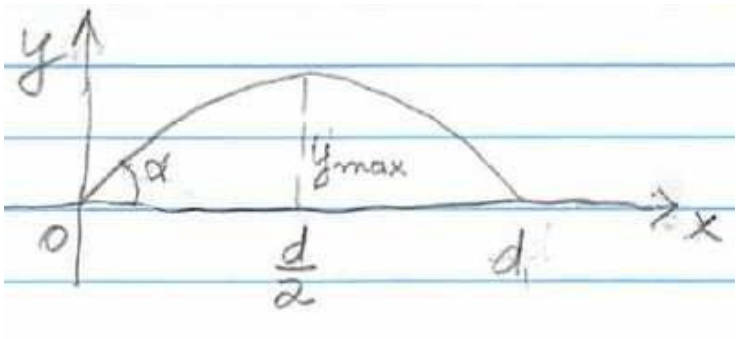


Figure 1.14: Trajectory of light ray in an exponentially stratified medium

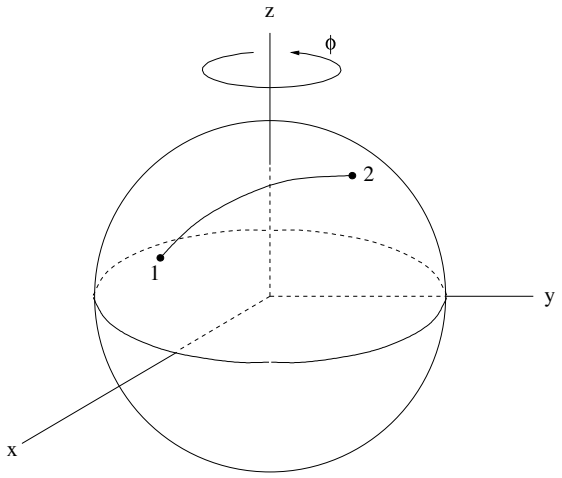


Figure 1.15: Curve on a two-sphere of radius ρ.

Here the variational integrand is

$$F = \sqrt{\theta_\varphi^2 + \sin^2\theta}$$

Note that F is independent of φ. Consequently, one can apply the "second form" of Euler equation. One finds that

$$F - \theta_\varphi \frac{\partial F}{\partial \theta_\varphi} = \text{constant} \equiv a.$$

Explicitly one has

$$\sqrt{\theta_\varphi^2 + \sin^2\theta} - \theta_\varphi \frac{\theta_\varphi}{\sqrt{\theta_\varphi^2 + \sin^2\theta}} = a$$

This is a differential equation which one readily solves in the usual way: Multiply by $\sqrt{\quad}$ and cancel θ_φ^2 to obtain

$$\sin^2\theta = a\sqrt{\left(\frac{d\theta}{d\varphi}\right)^2 + \sin^2\theta} \tag{1.33}$$

Isolate the derivative by solving for it. One obtains

$$\begin{aligned}\frac{d\varphi}{d\theta} &= \left(\frac{\sin^4\theta}{a^2} - \sin^2\theta\right)^{-1/2} \\ &= \frac{\csc^2\theta}{\sqrt{\frac{1}{a^2} - \csc^2\theta}} = \frac{\csc^2\theta}{\sqrt{\beta^2 - \cot^2\theta}}.\end{aligned}$$

Here $\beta^2 = \frac{1}{a^2} - 1$. The antiderivative is

$$\varphi(\theta) = -\sin^{-1}\frac{\cot\theta}{\beta} + \alpha\ .$$

Consequently, one obtains

$$-\cot\theta = \beta\sin(\varphi - \alpha)$$

This is the equation of a plane through the origin: Multiply by $\rho\sin\theta$ and expand

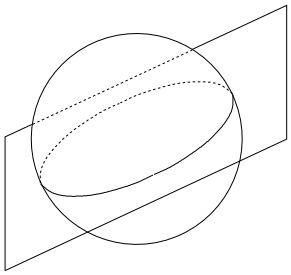


Figure 1.16: A great circle is the intersection of a sphere with a plane through the origin.

the sine function,

$$-\underbrace{\rho\cos\theta}_{z} = \beta\underbrace{\rho\sin\theta\sin\varphi}_{y}\cos\alpha - \beta\underbrace{\rho\sin\theta\cos\varphi}_{x}\sin\alpha$$

$$\boxed{z + y\,\beta\cos\alpha - x\,\beta\sin\alpha = 0}$$

Thus, the geodesic is the locus of the sphere intersected by a plane. In other words, the geodesic is a segment of a great circle.

Remark: The variational integral is extremized by *many* solutions to the Euler equation. The shortest geodesic minimizes s. Other solutions are geodesics which consist of great circles that wind around the sphere. They constitute *saddle points* of s in the space of functions!

Chapter 2

GENERALIZATIONS

Lecture 5

2.1 Functional with Several Unknown Functions

The world is not two-dimensional but three-dimensional, and if you take time into account, it is four-dimensional. Suppose we need to find the optimal curve between two points in three dimensions. All curves between these points are specified by one of the chosen coordinates, say x, as the curve parameter. Thus every point on this curve has its three coordinate values

$$\vec{c}(x) = (x, y_1(x), y_2(x))$$

More generally, consider a variational integral on an $n+1$-dimensional domain. The integral is a functional of n functions with the value of the functions fixed at the endpoints. The extremization of such an integral, namely

$$J[y_i] = \int_{x_1}^{x_2} F(x; y_1, y_1', y_2, y_2', \ldots, y_n, y_n')dx,$$

where $y_i(x)$ $i = 1, \ldots, n$ satisfy the prescribed boundary conditions at $x = x_1$ and $x = x_2$, is achieved by a method analogous to the one described before. Thus, we consider a system of variants

$$y_j = y_j(x, 0) + \sum_{i=1} \alpha_i^j h_i(x) = f_j(x, 0) + \sum_{i=1} \alpha_i^j h_i(x) j = 1, \ldots, n$$

of the system of functions

$$y_j = y_j(x, 0) = f_j(x, 0) \quad j = 1, \ldots, n$$

which we assume extremizes the functional $J[y_i]$.

The fact that these functions and their variants have prescribed values at the endpoints x_1 and x_2 can be stated by saying that each of the α-parametrized curves,

$$\{y_j = f_j(x, \alpha_i^j), j = 1, \ldots n\} \ ,$$

passes through the same two end points. This means that the system of variations

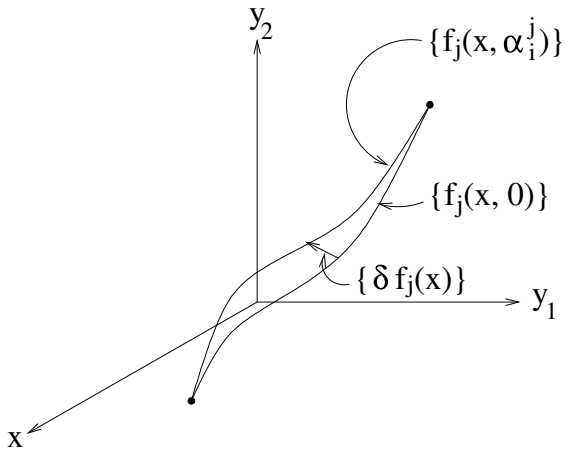


Figure 2.1: A variant and its system of variations $\delta\{f_j(x)\}$

$$\delta f_j(x) \quad j = 1, \ldots, n \ ,$$

which can be pictured as an n-component vector, vanishes at the endpoints:

$$\delta f_j(x_1) = \delta f_j(x_2) = 0 \quad j = 1, \ldots, n \ .$$

The procedure for obtaining the necessary condition for J to be an extremum follows the same pattern as before. The first variation of the functional J is

$$\delta J = \sum_{j=1}^{n} \int_{x_1}^{x_2} \left\{ \frac{\partial F}{\partial y_j} \, \delta f_j + \frac{\partial F}{\partial y_j'} \, \delta \left(\frac{df_j}{dx} \right) \right\} dx$$

Using the fact that

$$\begin{aligned} \delta \left(\frac{df_j}{dx} \right) &= \frac{df_j(x, \alpha_i^j)}{dx} - \frac{df_j(x, 0)}{dx} \\ &= \frac{d}{dx} \, \delta f_j(x) \quad j = 1, \ldots, n \end{aligned}$$

and integrating by parts yields

$$\delta J = \sum_{j=1}^{n} \int_{x_1}^{x_2} \left\{ \frac{\partial F}{\partial y_j} - \frac{d}{dx}\left(\frac{\partial F}{\partial y_j'}\right) \right\} \delta f_j(x) dx + \frac{\partial F}{\partial y_j'}\, \delta f_j \Big|_{x_1}^{x_2}.$$

The boundary term vanishes. The fact that for J to be an extremum for arbitrary variations δf_j $j = 1, \ldots, n$, implies that

$$\frac{\delta J}{\delta y_j(x)} \equiv F_{y_j} - \frac{d}{dx}(F_{y_j'}) = 0 \quad j = 1, \ldots, n \tag{2.1}$$

the Euler equations for "a system with n degrees of freedom". This set of equations constitutes the necessary condition for the variational integral $J[y_1, \ldots, y_n]$ to be an extremum. It is evident that this condition generalizes Euler equation from a single degree of freedom, Eq.(1.17) on page 24, to many degrees of freedom.

There is, however, one qualitatively new feature. It may happen that the vanishing of the first variation of J is due to the fact that J is an *maximum* with respect to variations in, say, $y_1(x)$ but is *minimum* with respect to variations in, say, $y_2(x)$.

Whatever the case may be, we shall call a curve $\{y_1(x), \ldots, y_n(x)\}$ which solves the set of Euler equations (2.1) an "*extremal curve*", even though often an "*optimal curve*" may be more descriptive.

Example (Fermat's Principle in three dimensions).

The propagation of a beam of light or sound whose local propagation speed is

$$v = \frac{c}{n(x, y, z)}$$

is governed by the extremum principle

$$J[y, z] = c \int_{x_1}^{x_2} \frac{\sqrt{1 + y'^2 + z'^2}}{n(x, y, z)} dx = \text{Minimum}.$$

This principle assumes that the local propagation speed depends only on the position (x, y, z) of the propagating beam. In that case the propagation medium in said to be *inhomogeneous* but *isotropic*. If, however, the propagation also depends on the direction of the beam, i.e.

$$v = \frac{c}{n(x, y, z, y', z')}$$

then the medium is said to be *inhomogeneous* and *anisotropic*.

In that case the propagation law is given by

$$J[y, z] = c \int_{x_1}^{x_2} \frac{\sqrt{1 + y'^2 + z'^2}}{n(x, y, z, y', z')}\, dx = \text{Minimum}.$$

Lecture 6

Example (Geodesic on a Sphere)

Consider the collection of all paths $(\theta(\lambda),\ \varphi(\lambda))$ joining two fixed points on the two-sphere. The parameter λ may be any variable which expresses monotonic progress along the path. For example, λ = time t of a moving ship on the earth, azimuthal angle φ, arclength s, or any other curve parameter monotonically related to the others.

The task is to find the λ-parametrized path $(\theta(\lambda),\ \varphi(\lambda))$ on the two sphere which has the property that its length

$$J = \int \sqrt{\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\theta \left(\frac{d\varphi}{d\lambda}\right)^2}\, d\lambda = \text{Extremum}\ .$$

For this problem one must find two functions that satisfy two differential equations:
(i) $\quad \dfrac{d}{d\lambda}\left(\dfrac{\partial F}{\partial \theta'}\right) = \dfrac{\partial F}{\partial \theta}$
or

$$\frac{d}{d\lambda}\ \frac{\frac{d\theta}{d\lambda}}{\sqrt{\theta'^2 + \sin^2\theta\varphi'^2}} = \frac{\sin\theta\cos\theta}{\sqrt{\theta'^2 + \sin^2\varphi'^2}}\ \left(\frac{d\varphi}{d\lambda}\right)^2 \tag{2.2}$$

(ii) The second equation is

$$\frac{d}{d\lambda}\left(\frac{\partial F}{\partial \varphi'}\right) = \frac{\partial F}{\partial \varphi}$$

The fact that the *variational integrand* F is independent of φ (one says that F is "cyclic" in φ) means that

$$\frac{\partial F}{\partial \varphi} = 0$$

Consequently, we have immediately an integral of motion, namely

$$\frac{\partial F}{\partial \varphi'} = \frac{\varphi' \sin^2\theta}{\sqrt{\theta'^2 + \sin^2\theta\varphi'^2}} = const. \equiv P_\varphi\ . \tag{2.3}$$

Equation (2.2) and Eq. (2.3) constitute the system of ordinary differential equations which we must solve. If we let $\lambda = \varphi$ be the curve parameter, then Eq.(2.3) becomes Eq.(1.33) (on page 33) and we can solve the problem. However, if the variational integrand had been

$$F \ = \ \left[g_{11}\left(\frac{dy_1}{d\lambda}\right)^2 + g_{22}\left(\frac{dy_2}{d\lambda}\right)^2 + 2g_{12}\frac{dy_1}{d\lambda}\ \frac{dy_2}{d\lambda}\right]^{1/2},$$

where the g_{ij}'s are functions which depend on the coordinates y_1, and y_2, then we could not have solved the problem. Nevertheless, a crisp and simple conclusion is possible.

What one does is to introduce the arclength as the curve parameter:

$$s = \int_0^\lambda \sqrt{\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\theta \left(\frac{d\varphi}{d\lambda}\right)^2}\, d\lambda$$

or

$$ds = \sqrt{\theta'^2 + \sin^2\theta\varphi'^2}\, d\lambda \ .$$

This brings about a tremendous simplification. The arclength parameter is introduced as follows

$$\frac{d}{ds} = \frac{d\lambda}{ds}\frac{d}{d\lambda} = \frac{1}{\sqrt{\theta'^2 + \sin^2\theta\varphi'^2}}\frac{d}{d\lambda} \ . \tag{2.4}$$

One applies this differential operator to θ and obtains

$$\frac{1}{\sqrt{\ }}\frac{d\theta}{d\lambda} = \frac{d\theta}{ds} \ .$$

Indicating differentiation with respect to s by a dot, $\frac{d\theta}{ds} = \dot{\theta}$, one obtains for Eq.(2.2)

$$\frac{1}{\sqrt{\ }}\frac{d\dot{\theta}}{d\lambda} = \sin\theta\cos\theta\left(\frac{1}{\sqrt{\ }}\frac{d\varphi}{d\lambda}\right)^2 \ .$$

Applying the above differential operator to $\dot{\theta}$ and to φ twice, one obtains

$$\ddot{\theta} = \dot{\varphi}^2 \sin\theta\cos\theta \ .$$

Similarly, the constant of motion, Eq.(2.3) becomes

$$P_\varphi = \dot{\varphi}\sin^2\theta \ .$$

If s is proportional to time, then in mechanics this is the "angular momentum" around the z-axis.

The introduction of the arclength s as the curve parameter always implies that

$$\left(\frac{d\theta}{ds}\right)^2 + \sin^2\left(\frac{d\varphi}{ds}\right)^2 = constant$$

along a curve. This expression is an integral of motion even for curves which are not extremals. The availability of this integral is guaranteed without having to integrate the Euler equations for the variational problem. Indeed, one always has

$$1 = \frac{ds}{ds} = \frac{\sqrt{d\theta^2 + \sin^2 d\varphi^2}}{ds} = \sqrt{\left(\frac{d\theta}{ds}\right)^2 + \sin^2\left(\frac{d\varphi}{ds}\right)^2} \ . \tag{2.5}$$

The constancy of this quantity will be discussed and used again on page 58, but this time in the context of a *constrained* variational problem.
Remarks:

1. The arclength parameter *always* simplifies the parametric representation in this way.

2. Equation (2.5) constitutes an integral of motion:

$$1 = \left(\frac{d\theta}{ds}\right)^2 + \sin^2\theta\left(\frac{d\varphi}{ds}\right)^2 .$$

   It is a constant of motion along *any curve*, geodesic or non-geodesic on S^2, provided s is the arclength parameter. In other words,

$$\frac{d}{ds}\left[\sqrt{\left(\frac{d\theta}{ds}\right)^2 + \sin^2\theta\left(\frac{d\varphi}{ds}\right)^2}\right] = 0$$

   along *any* curve. Thus there always is an "energy type" integral

$$\frac{1}{2}\left[\left(\frac{d\theta}{ds}\right)^2 + \sin^2\theta\left(\frac{d\varphi}{ds}\right)^2\right] = constant.$$

   The same conclusion holds in the general context as expressed by the variational integrand given by Eq. (2.4).

   The conclusion is this: For a geodesic we have

$$\begin{aligned} &\ddot{\theta} = \dot{\varphi}^2 \sin\theta\cos\theta \\ &\dot{\varphi}\sin^2\theta = const. && \text{"angular momentum"} \\ &\frac{1}{2}(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2) = \frac{1}{2} && \text{"energy"} \end{aligned}$$

3. These three equations in two unknowns are not independent. Instead, any *two* equations imply the *third*. For example, $\frac{d}{ds}(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2) = 0$ because of the two Euler-Lagrange equations. The benefit is that one can grasp the trajectories on a two-sphere in terms of the two simple integrable equations,

$$\frac{1}{2}\left(\dot{\theta}^2 + \frac{(const.)^2}{\sin^2\theta}\right) = constant$$

   and

$$\dot{\varphi}\sin^2\theta = const. ,$$

   whose solutions are arclength-parametrized curves.

2.2 Extremum Problem with Side Conditions.

Lecture 7

The fundamental problem of variational calculus is to maximize, minimize, or extremize a scalar functional on the unrestricted set of admissible variants. But variational calculus also accommodates the extremization on submanifolds, on subsets of such variants, namely those which are subject to one or more constraints.

The most famous such problem is the isoperimetric problem already known to the ancient Greeks. It consists of finding a curve such that

$$\text{Area} = \iint dxdy = \oint ydx = J[y] = \text{extremum}$$

subject to the constraint that the bounding curve $y(x)$ has fixed length

$$\ell = \oint \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \text{constant}.$$

An interesting generalization of this problem was the one which confronted Dido[1],

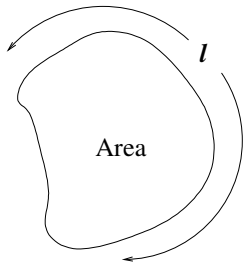


Figure 2.2: Area having perimeter of length ℓ.

the founder of ancient Carthage.

[1]According to legend, the mythical Queen Dido founded the ancient city of Carthage using this problem as a trick. Dido had to flee her home to escape from her murderous brother, and eventually ended up in the realm of King Hiarbas. She asked Hiarbas to be given as much land as could be enclosed by a single ox hide. "Sounds fair enough," the king must have thought, and granted her wish. But Dido was clever: rather than taking the ox hide as it was, she cut it into thin strips, attached them at their ends, and then laid the long strip out in a circle. This gave her an area large enough to found a kingdom on.

The fact that the circle is the shape that gives you the largest area becomes almost obvious. Yet, it took mathematicians a very long time to prove it. While Dido is supposed to have lived nearly 3000 years ago, a complete proof of her problem didn't arrive until 1879.

An important generalization of her problem was to circumscribe by means of a curve of *fixed length* an area which yields the largest amount of fertility. In other words, for a given function $\rho(x, y)$ (say, yield per unit area), extremize the weighted area integral

$$\iint \rho(x, y)dxdy$$

subject to the condition that the bounding curve $y(x)$ has fixed length,

$$\int \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \equiv \ell = \text{fixed}.$$

More generally, the problem is to find in the space of functions that function $y(x)$ for which $J[y] \equiv \int_{x_1}^{x_2} F(x; y, y')dx = \text{extremum}$, but which at the same time satisfies the constraints

$$\begin{aligned} K_1[y] &\equiv \int_{x_1}^{x_2} G_1(x; y, y')dx = \text{fixed} \equiv \kappa_1 \\ &\vdots \\ K_k[y] &\equiv \int_{x_1}^{x_2} G_k(x; y, y')dx = \text{fixed} \equiv \kappa_k \end{aligned} \tag{2.6}$$

2.2.1 Heuristic Solution

Lagrange's method of variations lends itself to solving this constrained optimization problem. One sets up the necessary conditions the desired function $y(x)$ must satisfy. As usual, consider the variants,

$$y(x) + \delta y(x)$$

of $y(x)$. Each variant is uniquely characterized by an infinitesimal displacement vector $\delta y(x)$, the variation of $y(x)$ in the s*space of admissible functions*,

$$S = \left\{ y(x) : \ y(x_1) = y_1, y(x_2) = y_2; \int_{x_1}^{x_2} F(x; y(x), y'(x))dx \text{ is well-defined} \right\} \tag{2.7}$$

The extremum problem under consideration is this:
Among $y \in S$ FIND an element $y^* \in S$ which extremizes

$$J[y] = \int_{x_1}^{x_2} F(x; y(x), y'(x))dx \tag{2.8}$$

subject to

$$\left.\begin{array}{l} K_1[y] = \int_{x_1}^{x_2} G_1(x; y(x), y'(x))dx = \kappa_1 \leftarrow \\ \vdots \\ K_k[y] = \int_{x_1}^{x_2} G_k(x; y(x), y'(x))dx = \kappa_k \leftarrow \end{array}\right\} \text{fixed and given} \tag{2.9}$$

Solution:
Consider the isograms of J and those of $K_1, \cdots, K_k$ on their common domain S of admissible functions. The function $y^*(x)$ which extremizes the scalar functional J

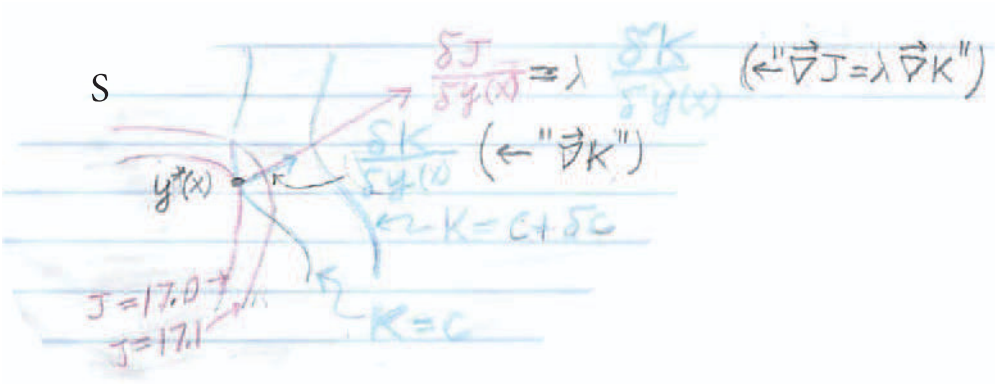


Figure 2.3: Isogram $J = 17.0$ osculates the isogram $K = \kappa$ at $y^*(x) \in S$.

while satisfying each of the constraints

$$\begin{array}{c} K_1[y] = \kappa_1 \\ \vdots \\ K_k[y] = \kappa_k \end{array}$$

is that critical point $y^*(x) \in S$ where an isogram of J simultaneously osculates all k isograms $K_1 = \kappa_1, \cdots, K_k = \kappa_k$. This means that there is a plane which is tangent to that isogram of $J[y]$ at the point y^* which isimultaneously tangent to the isograms of $K_1[y] = \kappa_1, \cdots, K_k[y] = \kappa_k$ at the same point in S.

This osculating condition, which is depicted in Figure 2.3 above, implies that the gradient $\frac{\delta J}{\delta y(x)}$ (= "$\vec{\nabla} J$") is a linear combination of the gradients $\frac{\delta K_1}{\delta y(x)}, \cdots, \frac{\delta K_k}{\delta y(x)}$, i.e. there exist constants $\lambda_1, \cdots, \lambda_k$ such that

$$\frac{\delta J}{\delta y(x)} = \lambda_1 \frac{\delta K_1}{\delta y(x)}, \cdots, \lambda_k \frac{\delta K_k}{\delta y(x)}. \tag{2.10}$$

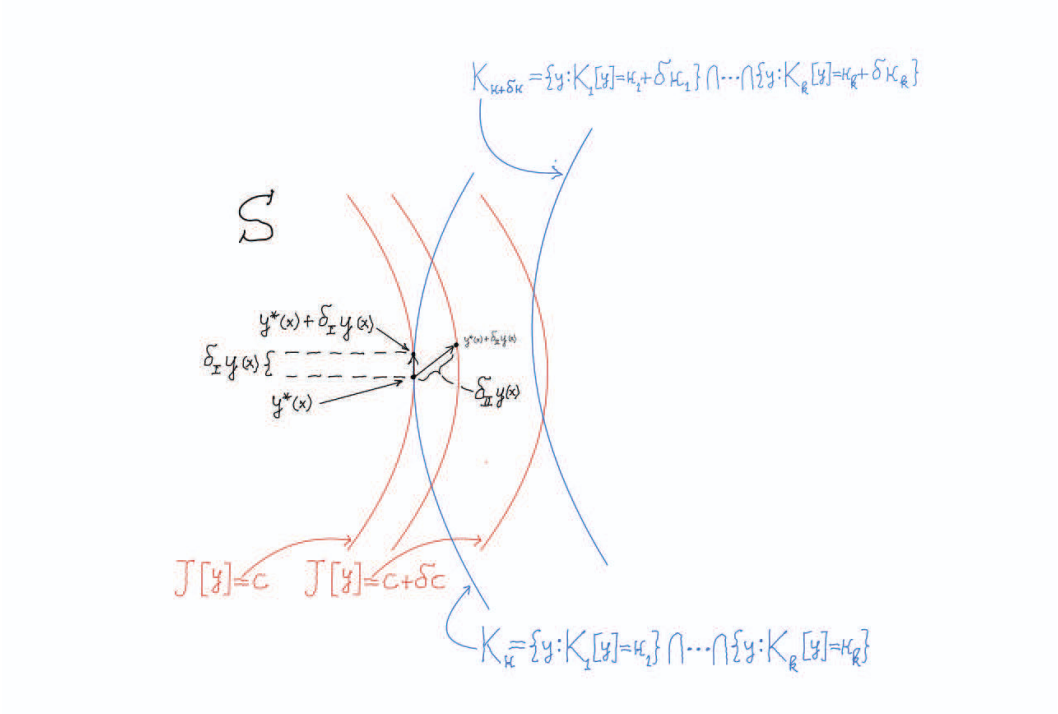


Figure 2.4: Two types of displacement vectors $\delta_I y(x)$ and $\delta_{II} y(x)$ at the point $y^*(x)$ in S, the space of admissible functions.

The line of reasoning leading to this conclusion is depicted in Figure 2.4 above and is based on linear algebra in four steps.

STEP 1.

Consider a generic variation $\delta y(x)$ of $y^*(x) \in S$. This variation is an infinitesimal displacement vector away from $y^*(x) \in S$. The corresponding variations in J and K_i

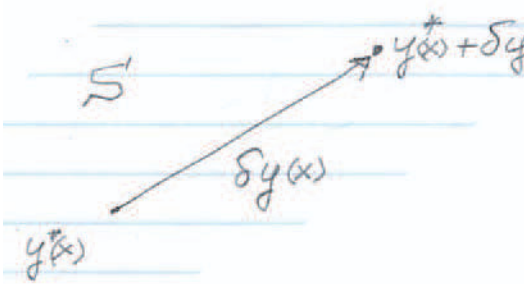


Figure 2.5: A variation $\delta y(x)$ viewed as a displacement vector in S.

are

$$\delta J = \int_{x_1}^{x_2} \underbrace{\frac{\delta J}{\delta y(x)}}_{F_y - \frac{d}{dx}F_{y'}} \delta y(x)\, dx$$

and

$$\delta K_i = \int_{x_1}^{x_2} \underbrace{\frac{\delta K_i}{\delta y(x)}}_{G_{iy} - \frac{d}{dx}G_{iy'}} \delta y(x)\, dx \quad i = 1, \cdots, k$$

STEP 2.
As depicted in Figure 2.4 on page 44, there are two types of variations $\delta y(x)$ away from $y^*(x)$:

(I)

$$\delta_I y(x) : \ \delta K_i = \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)} \delta y(x)\, dx = 0 \quad \text{for } i = 1, \cdots, k$$

These are the "YES" variations. Their associated variants $y^*(x) + \delta_I y(x)$ satisfy all the constraints

$$K_i[y^*(x) + \delta_I y(x)] = K_i[y^*(x)] \ (= \kappa_i^*) \ \ i = 1, \cdots, k.$$

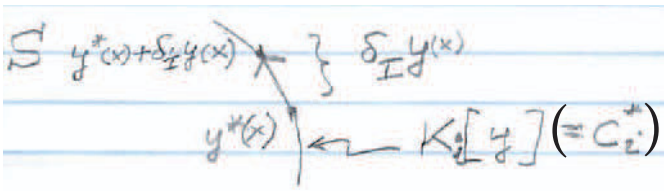


Figure 2.6: "YES" variation $\delta_I y(x)$ viewed as a vector tangent to the intersection of those isograms of K_i that contain the the point $y^*(x) \in S$.

(II)

$$\delta_{II} y(x) : \ \delta K_i = \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)} \delta y(x)\, dx \neq 0 \quad \text{for } i = 1, \cdots, k$$

These are the "NO" variations. Their associated variants $y^*(x) + \delta_{II}y(x)$ do *not* satisfy all the constraints

$$\underbrace{K_i[y^*(x) + \delta_{II}y(x)]}_{\kappa_i^* + \delta\kappa_i} \neq \underbrace{K_i[y^*(x)]}_{\kappa_i^*} \quad i = 1, \cdots, k.$$

Comments:

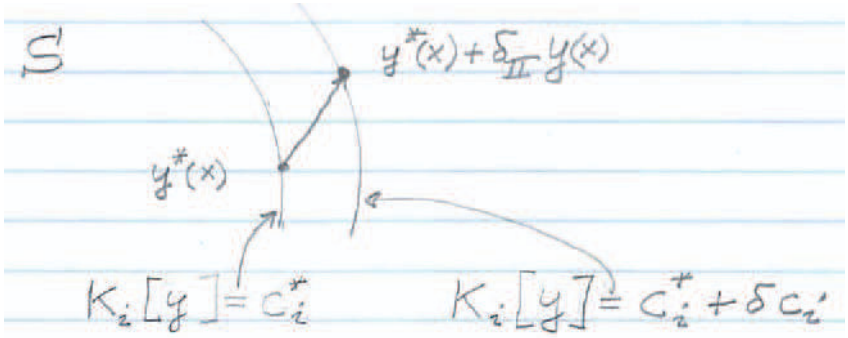


Figure 2.7: "NO" variation $\delta_{II}y(x)$ displacement vector which connects two different isograms of K_i in S.

(a) The sets of "YES" variations and "NO" variations at $y^*(x)$ are *jointly exhaustive* and *mutually exclusive* sets:

$$\{\delta_I y\} \cup \{\delta_{II} y\} = \{\delta y\} \quad \text{(vector space of displacements)}$$
$$\{\delta_I y\} \cap \{\delta_{II} y\} = \{\emptyset\} \quad \text{(empty set)}$$

(b) Thus there are precisely two classes of variations
$\delta_I y$ is a class I variation vector.
$\delta_{II} y$ is a class II variation vector.

(c) $\{\delta_I y\}$ is a subspace of $\{\delta y\}$, the vector space at $y^*(x)$.

STEP 3.
Let $y^*(x)$ be an optimal function.

(a) Let $\delta_I y$ be a class I variation vector. Then

$$\delta J = \int_{x_1}^{x_2} \frac{\delta J}{\delta y(x)} \delta_I y(x)\, dx = 0$$

and

$$\delta K_i = \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)} \delta_I y(x)\, dx = 0 \quad i = 1, \cdots, k$$

(b) Let $\delta_{II} y$ be a class II variation vector. Then

$$\delta K_i = \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)} \delta_{II} y(x)\, dx \neq 0$$

for at least some $i = 1, \cdots, k$, say

$$\delta K_1 = \int_{x_1}^{x_2} \frac{\delta K_1}{\delta y(x)} \delta_1 y(x)\, dx \neq 0$$

but

$$\delta K_j = \int_{x_1}^{x_2} \frac{\delta K_j}{\delta y(x)} \delta_1 y(x)\, dx = 0 \quad \text{for } j \neq 1.$$

Observe that

$$V_1 = \{\delta y = \delta_I y\}$$

is a 1-d subspace at $y^*(x)$.

(c) Calculate

$$\delta J = \int_{x_1}^{x_2} \frac{\delta J}{\delta y(x)} \delta_{II} y(x)\, dx$$

those which do not

satisfy the given constraints.

I. Those which do satisfy the contraints have variations which are perpendicular to the gradient of the constraint surfaces $K_i = c_i (i = 1, \ldots, k)$, i.e. those for which

$$\begin{aligned} \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)}\, \delta y(x) dx &= 0 \quad \forall i = 1, \ldots, k \\ K_i[y] &= \kappa_i \quad i = 1, \ldots, k_1 \end{aligned} \tag{2.11}$$

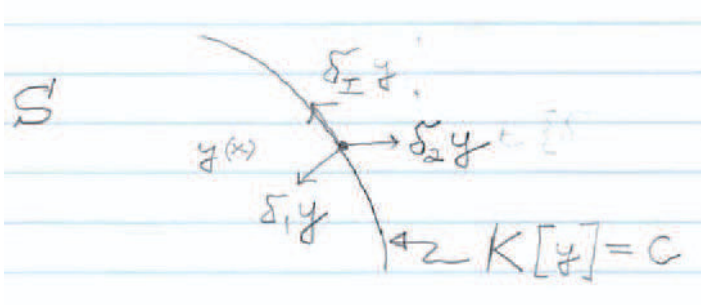


Figure 2.8: "NO" variation $\delta_{II}y(x)$ displacement vector which connects two different isograms of K_i in S.

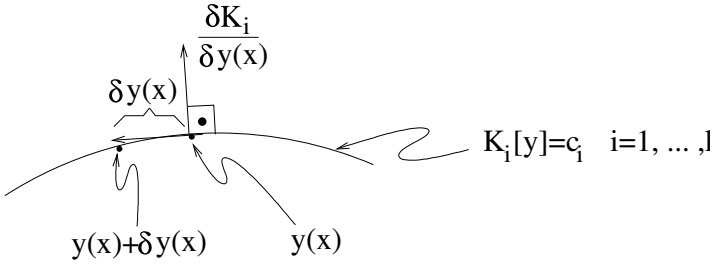


Figure 2.9: ith constraint manifold: its gradient $\frac{\delta K_i}{\delta y(x)}$ which is perpendicular to any class I ("yes") vector $\delta y(x)$ which is tangent to that manifold, $K_i[y] = \kappa_i$.

We shall call such vectors the "yes" vectors. They are tangent to all the constraint manifolds.

II. Those variants which do *not* satisfy the constraints have variations which are *not* perpendicular to the constraint surfaces $K_i = c_i$, i.e. those for which

$$\int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)}\, \delta y(x)dx \neq 0 \quad \text{for some } i = 1, \ldots, k$$

We shall call such vectors the "no" vectors.

Thus[2]a variation vector $\delta y(x)$ always belongs either to class I ("yes") or to class II ("no"): If δy belongs to class I, then

$$\delta J = \int_{x_1}^{x_2} \frac{\delta J}{\delta y(x)}\, \delta y(x)dx = 0$$

[2]Following Aristotle's law of the excluded middle.

because $y(x)$ was assumed to be an optimal function which satisfies the constraint, i.e. for which

$$\delta K_i = \int_{x_1}^{x_2} \frac{\delta K_i}{\delta y(x)}\, \delta y(x) dx = 0 \quad i = 1, \ldots, k \text{ (for class I variants).}$$

If, on the other hand, δy belongs to class II, i.e. such that

$$\delta K_i \equiv \int \frac{\delta K_i}{\delta y(x)}\, \delta y(x) dx \neq 0 \quad \text{for some } i = 1, \ldots, k,$$

then we can, and we will, choose constants $\lambda_1, \ldots, \lambda_k$ such that

$$\delta J = \lambda_1 \delta K_1 + \ldots \lambda_k \delta K_k.$$

The conclusion is therefore this: *all* variation vectors $\delta y(x)$ satisfy the equation

$$\int_{x_1}^{x_2} \left[\frac{\delta J}{\delta y(x)} - \lambda_1 \frac{\delta K_1}{\delta y(x)} - \ldots \lambda_k \frac{\delta K_k}{\delta y(x)} \right] \delta y(x) dx = 0$$

whenever

$$J - \lambda_1 K_1 - \cdots - \lambda_k K_k = \text{extremum}$$

where the constants $\lambda_1, \ldots, \lambda_k$ *must be chosen* appropriately. They are determined by the values $c_1, \ldots, c_k$ of the constraints.

Summary:

To solve the optimization problem with constraints, solve the following $k + 1$ equations

$$\begin{aligned} \frac{\delta}{\delta y(x)} \left(J[y] - \lambda_1 K_1[y] - \ldots \lambda_k K_k[y] \right) &= 0 \\ K_1[y] &= \kappa_1 \\ &\vdots \\ K_k[y] &= \kappa_k \end{aligned}$$

for $y(x)$ and $\lambda_1, \ldots, \lambda_k$. The constants $\lambda_1, \ldots, \lambda_k$ are called “Lagrange multipliers".

2.2.2 Solution via Constraint Manifold

The “yes” and “no” vectors play a basic role in establishing necessary conditions for the existence of an extremum subject to constraints. It is therefore appropriate to highlight how these vectors tie into well-established theory. We shall do this by splitting the space of tangent vectors into two orthogonal vector spaces, whose only common vector is the zero vector.

Review: Finite-dimensional Space

Our task is to extremize a given function $f(\vec{x})$ on the finite-dimensional space R^n subject to a finite number of constraints, say, $g_1(x) = c_1, \cdots, g_k(x) = c_k$, with $k < n$. More precisely, we wish to find the equations for those critical points $\vec{x}_0$ of $f(\vec{x})$ which lie on the constraint manifold, which is the intersection of the constraint hypersurfaces given by the finite number of constraints.

We shall solve this problem in two steps. First, we shall develop the chain of reasoning when there are no constraints present. Then we shall extend this reasoning to the case where k constraints are present.

Extremum Without Constraints

Consider the trajectory $\vec{x} = \vec{r}(t)$ of a particle in the conservative force field ∇f of a potential $f(\vec{x})$. Then

$$\frac{df(\vec{r}(t))}{dt} = \sum_i \frac{\partial f}{\partial x^i}\frac{dr^i}{dt} \equiv \nabla f \cdot \frac{d\vec{r}}{dt} = \text{“(-)force} \times \text{velocity”} \tag{2.12}$$

is the rate at which the force field does work on the particle.

Suppose there exists a point $\vec{x}_0$ such that

$$\nabla f(\vec{x}_0) \cdot \frac{d\vec{r}}{dt} = 0$$

for the tangent vectors of *all* curves $\vec{x} = \vec{r}(t)$ passing through this point, or, more generally, for all vectors $\vec{v}$ such that

$$\nabla f(\vec{x}_0) \cdot \vec{v} = 0.$$

These vectors form the *tangent space* at $\vec{x}_0$. Observe that

$$\nabla f(\vec{x}_0) \cdot \vec{v} = 0$$

implies that

$$\nabla f(\vec{x}_0) \in \text{“orthogonal complement” of the tangent space at } \vec{x}_0 \ .$$

Being the orthogonal complement of the space of *all* tangent vectors, it consists of the zero vector only; it is zero dimensional. Thus

$$\nabla f(\vec{x}_0) = \vec{0} \ .$$

This is the necessary condition for $\vec{x}_0$ to be the location of an unconstrained extremum of f.

Extremum with Constraints

Next, suppose there exists a point $\vec{x}_0$ such that

$$\nabla f(\vec{x}_0) \cdot \frac{d\vec{r}}{dt} = 0 \tag{2.13}$$

for the tangent vectors $\frac{d\vec{r}}{dt}$ of only those curves $\vec{x} = \vec{r}(t)$ passing through $\vec{x}_0$ which lie in the *constraint manifold*

$$\{\vec{x} : \ g_1(\vec{x}) = c_1, \cdots, g_k(\vec{x}) = c_k\} \ , \tag{2.14}$$

and hence satisfy

$$\nabla g_1(\vec{x}_0) \cdot \frac{d\vec{r}}{dt} = 0, \cdots, \nabla g_k(\vec{x}_0) \cdot \frac{d\vec{r}}{dt} = 0 \ . \tag{2.15}$$

The collection of such vectors $\frac{d\vec{r}}{dt}$ forms a *subspace* of the set of all vectors at x_0: any linear combination of the subspace vectors also satisfies these k constraints. This subspace,

$$W = \{\vec{v} : \nabla g_1(\vec{x}_0) \cdot \vec{v} = 0, \cdots, \nabla g_k(\vec{x}_0) \cdot \vec{v} = 0\} \tag{2.16}$$

is called the *constraint subspace* at $\vec{x}_0$. These subspace vectors are the "yes" vectors for the infinite dimensional domain in Eq.(2.11) on page 47.

One now makes the fundamental and final observation that $\nabla f(\vec{x}_0)$ belongs to the the orthogonal complement of the constraint subspace W, Eq.(2.16). This orthogonal complement consists precisely of those vectors which have the form $\lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k$. In particular,

$$\nabla f(\vec{x}_0) = \lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k \tag{2.17}$$

for some set of coefficients $\lambda_1, \cdots, \lambda_k$.

These constants *must* exist. This is because our focus is on a non-trivial solution $\frac{d\vec{r}}{dt}$ to Eqs.(2.13) and (2.15).

They are called the *Lagrange multipliers* of the constrained extremum problem. It is evident that their values and the location of the critical point of f on the constraint manifold are given by the following "Lagrange Multiplier" theorem

Theorem 1.

Given: *Let f be a scalar function which has an extremum at $\vec{x}_0$ on the intersection of the k constraint hypersurfaces ("the constraint manifold"), Eq.(2.14).*

Conclusion: *The location of the constrained critical point is determined by*

$$\boxed{\nabla f(\vec{x}_0) - \lambda_1 \nabla g_1(\vec{x}_0) - \cdots - \lambda_k \nabla g_k(\vec{x}_0) = 0\}} \tag{2.18}$$

where the coefficients $\lambda_1, \cdots, \lambda_k$ are obtained in conjunction with

$$\begin{aligned} g_1(\vec{x}_0) &= c_1 \\ &\vdots \\ g_k(\vec{x}_0) &= c_k \end{aligned}$$

Remark 1: In mechanics where $f(\vec{x})$ is a scalar potential and $\nabla f(\vec{x})$ its (conservative) force field, the force

$$\nabla f(\vec{x}_0) = \lambda_1 \nabla g_1(\vec{x}_0) + \cdots \lambda_k \nabla g_k(\vec{x}_0)$$

is the constraining force which guarantees that the point particle stays in equilibrium ("no net forces") on the constraint manifold.
Remark 2: This theorem is easily visualized in three dimensions, $R^n = R^3$, with two constraints, say $g_1(\vec{x}) = c_1$ and $g_2(\vec{x}) = c_2$. There the constraint manifold is one-dimensional, a single curve $\vec{r}(t)$, the intersection of the two surfaces $g_1(\vec{x}) = c_1$ and $g_2(\vec{x}) = c_2$. Consequently,

$$\begin{aligned} \vec{\nabla} g_1 \cdot \frac{d\vec{r}}{dt} &= 0 \\ \vec{\nabla} g_2 \cdot \frac{d\vec{r}}{dt} &= 0 \end{aligned}$$

The fact that f has an extremum on this curve implies

$$\vec{\nabla} f \cdot \frac{d\vec{r}}{dt} = 0$$

This is a system of three homogeneous equations in the three unknowns of $\frac{d\vec{r}}{dt}$. Such a system has a non-trivial solution if and only if the set of vectors $\{\vec{\nabla} f, \vec{\nabla} g_1, \vec{\nabla} g_2\}$ form a linearly dependent set. Consequently, $\exists\ \lambda_1$ and λ_2 such that

$$\nabla f = \lambda_1 \vec{\nabla} g_1 + \lambda_2 \vec{\nabla} g_2.$$

Example (Function with two finite constraints):
In multivariable calculus the problem

$$f(x, y, z) = \text{extremum}$$

subject to

$$\begin{aligned} g(x, y, z) &= c_1 \\ h(x, y, z) &= c_2 \end{aligned}$$

is solved by considering

$$f^* = f - \lambda_1 g - \lambda_2 h$$

and solving

$$\begin{aligned} \vec{\nabla}(f - \lambda_1 g - \lambda_2 h) &= 0 \\ g(x,y,z) &= c_1 \\ h(x,y,z) &= c_2 \end{aligned}$$

for x_0, y_0, z_0, λ_1 and λ_2. These are 3+2 equations in 3+2 unknowns.

Infinite-dimensional Space

The Lagrange Multiplier theorem readily extends to a functional which is defined on curves in function space, which is infinite-dimensional. The development runs parallel to the one preceeding Theorem 1.

First of all recall that a "curve in function space" is a 1-parameter set of functions

$$y(x,t) \qquad a \le t \le b \ . \tag{2.19}$$

Second, the rate of change of the functional $J[y]$ along this 1-parameter set is

$$\begin{aligned} \frac{d}{dt}J[y] &= \frac{d}{dt}\int_{x_1}^{x_2} F(x; y(x,t), y'(x,t))dx \\ &= \int_{x_1}^{x_2} \left[\left.\frac{\partial F}{\partial y}\right|_{y(x,t)} \frac{\partial y(x,t)}{\partial t} + \left.\frac{\partial F}{\partial y'}\right|_{y(x,t)} \frac{\partial y'(x,t)}{\partial t} \right] dx \\ &= \int_{x_1}^{x_2} \left.\frac{\delta J}{\delta y(x)}\right|_{y(x,t)} \frac{\partial y(x,t)}{\partial t} dx \ , \end{aligned}$$

where $\delta J/\delta y(x)$ is the variational derivative of J as defined by Eq.(1.15) on page 23.

Third, if $J[y]$ has an extremal value on the parametrized set of functions $y(x,t)$ at $t = t_0$, which is to say that its rate of change $dJ[y]/dt$ vanishes at $t = t_0$, then

$$0 = \left.\frac{d}{dt}J[y]\right|_{t=t_0} = \int_{x_1}^{x_2} \left.\frac{\delta J}{\delta y(x)}\right|_{y(x,t_0)} \left.\frac{\partial y(x,t)}{\partial t}\right|_{t_0} dx \ . \tag{2.20}$$

replaces Eq.(2.13) as the extremum condition.

Fourth, the fact that the one-parameter family $y(x,t)$ lies in the intersection of the constraints, Eqs.(2.6) on page 42, implies that

$$\begin{aligned} \frac{dK_1[y]}{dt} &= \int_{x_1}^{x_2} \left.\frac{\delta K_1}{\delta y(x)}\right|_{y(x,t)} \frac{\partial y(x,t)}{\partial t} dx = 0 \\ &\vdots \\ \frac{dK_k[y]}{dt} &= \int_{x_1}^{x_2} \left.\frac{\delta K_k}{\delta y(x)}\right|_{y(x,t)} \frac{\partial y(x,t)}{\partial t} dx = 0 \ . \end{aligned} \tag{2.21}$$

At $t = t_0$ these equations replace Eqs.(2.15).

Finally, Eq.(2.17) gets replaced by

$$\frac{\delta J}{\delta y(x)} = \lambda_1 \frac{\delta K_1}{\delta y(x)} + \cdots + \lambda_k \frac{\delta K_k}{\delta y(x)} \ .$$

Nota bene: A class I ("yes") or a class II ("no") vector $\delta y(x)$ on page 48 is an infinitesimal vector ("a variation")

$$\delta y(x) = \frac{\partial y(x,t)}{\partial t} \delta t$$

tangent to one of the t-parametrized sets of functions ("curve in function space"), Eq.(2.19), on page 53. A curve which satisfies the integral constraints has tangent vectors which obey Eq.(2.21). These tangents are proportional to the "yes" vectors. The orthogonal complement of all such vectors make up the vector space of "no" vectors. Each of them is some linear combination of $\delta K_1/\delta y(x), \cdots, \delta K_k/\delta y(x)$.

In spite of the fact that the domain of the functional $J[y]$ is infinite-dimensional, the line of reasoning that leads to the infinite-dimensional version of the Lagrange Multiplier Theorem 1 remains unchanged. In fact, the variational (i.e. "Euler") derivatives of $J[y]$ and of the $K_i[y]s$ correspond one for one to the familar gradient of $f(\vec{x}_0)$ and of the $g_i(\vec{x}_0)$s

$$\begin{aligned} \frac{\delta J[y]}{\delta y(x)} &\leftrightarrow \nabla f(\vec{x}_0) \\ \frac{\delta J K_1[y]}{\delta y(x)} &\leftrightarrow \nabla g_1(\vec{x}_0) \\ \vdots &\leftrightarrow \vdots \\ \frac{\delta J K_k[y]}{\delta y(x)} &\leftrightarrow \nabla g_k(\vec{x}_0) \end{aligned} \tag{2.22}$$

The only difference between the first and the third colums is that the variational derivatives are on an infinite dimensional background, while the gradients are on its finite dimensional background.

The constrained optimization Theorem 2 below is the extension to an infinite dimensional background of the constrained optimization Theorem 1 for a finite dimensional background

Theorem 2.

Given: *Let*

$$J[y] = \int_{x_1}^{x_2} F(x, y, y')\, dx$$

be a scalar functional which has an optimal function $y(x)$ satisfying the k constraint conditions (2.6) on page 42.

Conclusion: *This optimal function satisfies the differeential equation*

$$F_y - \frac{d}{dx}F_{y'} - \lambda_1 \left(G_{1y} - \frac{d}{dx}G_{1y'} \right) - \cdots - \lambda_k \left(G_{ky} - \frac{d}{dx}G_{ky'} \right) = 0 \ , \quad (2.23)$$

where the coefficients $\lambda_1, \cdots, \lambda_k$ are obtained with the help of

$$\int_{x_1}^{x_2} G_1(x, y, y')\, dx = c_1$$

$$\vdots$$

$$\int_{x_1}^{x_2} G_k(x, y, y')\, dx = c_k \ .$$

Example (Functional with two integral constraints):

Apply Lagrange's method to solve the variational problem

$$J[y] \equiv \int_{x_1}^{x_2} F(x, y, y')\, dx = \text{extremum}$$
$$K[y] \equiv \int_{x_1}^{x_2} G(x, y, y')\, dx = c_1 = \text{fixed}$$
$$L[y] \equiv \int_{x_1}^{x_2} H(x, y, y')\, dx = c_2 = \text{fixed}$$

This problem is a generalization to infinite dimensions of the example on page 52. One proceeds by first constructing

$$J^* = J - \lambda_1 K - \lambda_2 L$$

and then considering

$$\delta J^* = \delta(J - \lambda_1 K - \lambda_2 L) \ .$$

Thus

$$\delta J^* = \int \left(\frac{\partial F}{\partial y} - \frac{d}{dx} \frac{\partial F}{\partial y'} \right) \delta f(x) dx - \lambda_1 \int \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \frac{\partial G}{\partial y'} \right) \delta f(x) dx$$
$$- \lambda_2 \int \left(\frac{\partial H}{\partial y} - \frac{d}{dx} \frac{\partial H}{\partial y'} \right) \delta f(x) dx \ .$$

The problem is solved by setting

$$\delta J^* = \delta J - \lambda_1 \delta K - \lambda_2 \delta L = 0 \ .$$

Thus one has the following set of equations

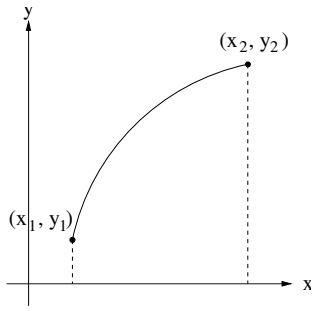
$$\begin{aligned} \frac{\partial}{\partial y}(F - \lambda_1 G - \lambda_2 H) - \frac{d}{dx} \frac{\partial}{\partial y'}(F - \lambda_1 G - \lambda_2 H) &= 0 \\ \int_{x_1}^{x_2} G(x; y, y') dx &= c_1 \\ \int_{x_1}^{x_2} H(x; y, y') dx &= c_2 \end{aligned}$$

The solutions to these equations yield 4 constants:
(1) 2 integration constants
(2) 2 Lagrange multipliers
The 4 constants are determined by
(1) 2 boundary conditions on $y(x)$
(2) 2 constraints $K[y] = c_1$ and $L[y] = c_2$.

Lecture 8

Example. (Isoperimetric Problem):
Among all the curves $(x(\lambda), y(\lambda))$ of fixed length ℓ,

$$\int_{\lambda_1}^{\lambda_2} \sqrt{\left(\frac{dx}{d\lambda} \right)^2 + \left(\frac{dy}{d\lambda} \right)^2} \, d\lambda = \text{constant} = \ell,$$



through the points (x_1, y_1) and (x_2, y_2) find that curve for which

$$\int_{\lambda_1}^{\lambda_2} y \, \frac{dx}{d\lambda} \, d\lambda = \text{extremum}$$

i.e. extremize the area!

Remark. Other formulations of this problem are

(i) $\int y \, dx$=extremum
subject to

$$\int \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \, dx = \text{fixed}$$

(ii) $\int y \, \frac{dx}{ds} \, ds$=extremum
subject to

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1$$

Solution: Consider the variational integrand

$$F^* = yx' - \mu\sqrt{x'^2 + y'^2}$$

where μ is a Lagrange multiplier. The Euler equations are

$$\frac{d}{d\lambda}\frac{\partial F^*}{\partial x'} = \frac{\partial F^*}{\partial x}$$

$$\frac{d}{d\lambda}\frac{\partial F^*}{\partial y'} = \frac{\partial F^*}{\partial y}$$

The fact that x is a cyclic coordinate, i.e. that $\frac{\partial F^*}{\partial x} = 0$ implies that $\frac{\partial F^*}{\partial x'}$ is constant:

$$y - \mu\frac{x'}{\sqrt{x'^2 + y'^2}} = \text{constant} \equiv C$$

The second equation gives us

$$\frac{d}{d\lambda}\left((-)\mu\frac{y'}{\sqrt{x'^2+y'^2}}\right)-x'=0$$

The form of these equations suggests that one introduce arclength,

$$ds=\sqrt{x'^2+y'^2}d\lambda,$$

as a new curve parameter. In terms of this parameter we have

$$\frac{1}{\sqrt{\left(\frac{dx}{d\lambda}\right)^2+\left(\frac{dy}{d\lambda}\right)^2}}\,\frac{d}{d\lambda}=\frac{d}{ds}=\text{“ . ”}$$

We also notice that for *any* curve extremal or nonextremal, the use of s as a curve parameter implies

$$\sqrt{\left(\frac{dx}{ds}\right)^2+\left(\frac{dy}{ds}\right)^2}=1 \tag{2.24}$$

In terms of the derivative with respect to arclength the Euler equations become

$$y-\mu\dot{x}=c \tag{2.25}$$

and

$$\mu\ddot{y}+\dot{x}=0 \tag{2.26}$$

Comments about consistency: Any two of the equations (2.24), (2.25), and (2.26) imply the third. Thus, notice that the last two equations are consistent with the requirement that

$$\boxed{\dot{x}^2+\dot{y}^2=\text{const}=1.} \tag{2.27}$$

This requirement was already highlighted on page 39, but we also validate it by direct differentiation: Differentiate Eq(2.25)

$$\mu\ddot{x}-\dot{y} = 0 \tag{2.28}$$

$$\mu\ddot{y}+\dot{x} = 0 \tag{2.29}$$

These are the unintegrated Euler equations. Insert them into the derivative of $\dot{x}^2+\dot{y}^2$,

$$\frac{d}{ds}(\dot{x}^2+\dot{y}^2)=2\ddot{x}\dot{x}+2\ddot{y}\dot{y}=2\ddot{x}(-)\lambda\ddot{y}+2\ddot{y}\lambda\ddot{x}=0 \tag{2.30}$$

Thus $\dot{x}^2+\dot{y}^2$=constant along the curve. This is consistent with Eq.(2.24). This constancy is a general property and we shall return to it when we discuss the symmetries of the functional J.

Now return to the task of solving Eqs (2.25) and (2.26). Insert Eq.(2.25) into (2.26) and obtain

$$\ddot{y} + \frac{1}{\mu^2}\, y = \frac{c}{\mu^2}$$

The solution is

$$\boxed{y = A\cos\left(\frac{s}{\mu} + \delta\right) + c} \tag{2.31}$$

To obtain $x(s)$, use Eq.(2.25),

$$\dot{x} = \frac{y - c}{\mu} = \frac{A}{\mu}\cos\left(\frac{s}{\mu} + \delta\right) \ ,$$

and obtain

$$\boxed{x = A\sin\left(\frac{s}{\mu} + \delta\right) + x_0} \tag{2.32}$$

Equations (2.31) and (2.32) are parametrized equations for a *circular arc* because $(y - c)^2 + (x - x_0)^2 = A^2$.

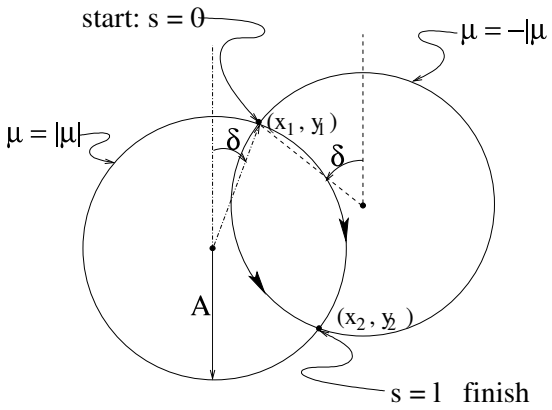


Figure 2.10: Two intersecting circles as solutions to the isoperimetric problem.

We have five constants A, δ, μ, c, and x_0 whose values are determined by five conditions, namely,

$$\text{At } s = 0: \quad x(0) = x_1 \tag{2.33}$$

$$y(0) = y_1 \tag{2.34}$$

$$\text{At } s = \ell: \quad x(\ell) = x_2 \tag{2.35}$$

$$y(\ell) = y_2 \tag{2.36}$$

$$\int_0^\ell \sqrt{\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2}\, ds = \ell \ . \tag{2.37}$$

This integral constraint together with Eq.(2.30) imply that that

$$\dot{x}^2 + \dot{y}^2 = 1,$$

which is to say that the squared length of the tangent to the curve has unit length. Q: What does this equation tell us about the Lagrange multiplier μ? A: With the help of the solutions $x(t)$ and $y(t)$ it implies that

$$\frac{A^2}{\mu^2} = 1 \ . \tag{2.38}$$

It follws that $\mu = \pm A$. Thus there are *two* solutions, one corresponding to $\mu = A$, which gives an arc whose length increases with angle in a clockwise fashion; the other corresponds to $\mu = -A$, which gives an arc whose length increases with angle in a counter-clockwise fashion as the curve proceeds from point 1 to point 2.
The other implication is also geometrical: for both circles the Langrange multiplier $|\mu|$ is the radius of their circular arcs.

Economy and Transparency

Mathematical economy and geometrical transparency are always cognitive virtues. In the above isoperimetric problem they are achieved by the method of complex arithmetic: Consolidate the two real degrees of freedom into a single complex degree of freedom,

$$z(s) = y(s) + ix(s),$$

and its complex conjugate,

$$\overline{z}(s) = y(s) - ix(s),$$

This collapses the two Euler Eqs.(2.28)-(2.29) into a single equation

$$\mu\ddot{z} - i\dot{z} = 0$$

whose solution is

$$\begin{aligned} z(s) &= Ce^{i\frac{s}{\mu}} \\ &= Ae^{i(\frac{s}{\mu}+\delta)} \qquad (2.39) \\ &= \underbrace{A\cos\left(\frac{s}{\mu}+\delta\right)}_{y(s)} + i\underbrace{A\sin\left(\frac{s}{\mu}+\delta\right)}_{x(s)} \qquad (2.40) \end{aligned}$$

The Lagrange multiplier μ is real. Consequently, the complex conjugate equation $\mu\ddot{\bar{z}} + i\dot{\bar{z}} = 0$ yields merely the complex conjugate of Eq.(2.40). Furthermore, the constraint Eq.(2.27) when expressed in terms of the complex solution, Eq.(2.39) and its complex conjugate, namely

$$\dot{z}\dot{\bar{z}} = 1,$$

leads to the real result Eq.(2.38) on page 60, as it must.

2.2.3 Variational Problems with Finite Constraints

Scientiofic knowledge consists of understanding observations and facts by asking and answering

- Why do they occur?
- What are the primciples that imply them?
- Where do they come from?

Variational calculus is the means of mathematizing many questions of this type.

Aside from isoperimetric variation problems, i.e. those having a set of integral constraints as subsidiary conditions, another important class is exemplified by the following:
Find the curve $\{y(x), z(x)\}$ through points $(x_1, y_1, z_1), (x_2, y_2, z_2)$ such that

$$J = \int F(x, y, z, y', z')dx = \text{extremum}$$

subject to the "finite" constraint

$$G(x, y, z) = 0 \ .$$

Given that this problem has a solution, we consider the first variation of J due to deviations $\delta y(x), \delta z(x)$ of the variants $y + \delta y$ and $z + \delta z$ away from the functions $y(x)$ and $z(x)$:

$$\begin{aligned} \delta J \ = \ & \int_{x_1}^{x_2} \left[\left(F_y - \frac{d}{dx} F_{y'} \right) \delta y(x) + \left(F_z - \frac{d}{dx} F_{z'} \right) \delta z(x) \right] dx \\ & + F_{y'} \delta y \Big|_{x_1}^{x_2} + F_{z'} \delta z \Big|_{x_1}^{x_2} \end{aligned}$$

The fact that the variant $\{y(x) + \delta y(x),\ z(x) + \delta z(x)\}$ is required to lie on the constraint manifold $G = 0$, i.e., it must satisfy

$$G(x, y + \delta y, z + \delta z) = 0 \ ,$$

implies that the variations δy and δz are not independent. They must satisfy

$$\frac{\partial G}{\partial y}\delta y(x) + \frac{\partial G}{\partial z}\delta z(x) = 0$$

so that

$$\delta z = -\frac{G_y}{G_z}\delta y.$$

Insert this into the expression for the first variation and obtain

$$\delta J = \int_{x_1}^{x_2} \left\{ (F_y - \frac{d}{dx}F_{y'}) - (F_z - \frac{d}{dx}F_{z'})\frac{G_y}{G_z} \right\} \delta y dx + F_{y'}\delta y|_{x_1}^{x_2} - F_{z'}\frac{G_y}{G_z}\delta y \Big|_{x_1}^{x_2}$$

The endpoint terms are zero because all variants pass through the same pair of end points (x_1, y_1, z_1) and (x_2, y_2, z_2) so that $\delta y(x_1) = \delta y(x_2) = 0$.
The variations

$$\delta y = \sum \alpha_i h_i(x)$$

are arbitrary. Consequently, $\delta J = 0$ implies that for each x one has

$$\frac{F_y - \frac{d}{dx}F_{y'}}{G_y} = \frac{F_z - \frac{d}{dx}F_{z'}}{G_z}$$

Let us designate this common function by $\lambda(x)$. Thus one obtains two Euler equations,

$$\boxed{F_y - \frac{d}{dx}F_{y'} - \lambda(x)G_y = 0 \text{ and } F_z - \frac{d}{dx}F_{z'} - \lambda(x)G_z = 0}\ .$$

These equations together with

$$\boxed{G(x, y(x), z(x)) = 0}$$

determine the three functions

$$y(x), z(x), \text{ and } \lambda(x).$$

Even though $\lambda = \lambda(x)$ not a constant, it still is referred to a “Lagrange multiplier."

2.3 Variable End Point Problem

It is not necessary that the endpoints of a variational problem be fixed. Consider as an example the brachistochrone problem (Figure 1.2 on page 10) in which the second

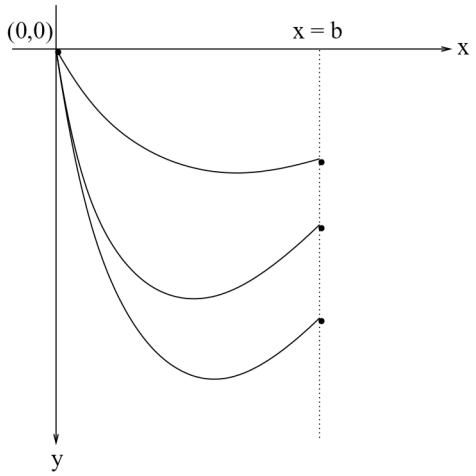


Figure 2.11: Brachistochrone trial curves with different end points.

end point, instead of being fixed, is required to lie somewhere on the vertical line. Which of the curves will allow a sliding bead to reach its destination in the least amount of time?

The general problem is this: Find that curve $y(x)$, which starts at $x = a$ and ends at $x = b$, for which

$$J = \int_a^b F(x, y, y')dx + \psi(y(b)) - \varphi(y(a)) = \text{ extremum}, \tag{2.41}$$

where F, ψ, and φ are given.

The variational principles we have considered so far can be said to have had *Dirichelet* boundary conditions. By contrast, the present variational principle can be said to have just the opposite type, namely "natural" or "free" boundary conditions.

Having intergrated by parts, one finds that the first variation of the functional J is

$$\delta J = \int_a^b [F_y - \frac{d}{dx}F_{y'}]\delta f(x)dx + \left(\frac{\partial F}{\partial y'} + \frac{\partial \psi}{\partial y}\right)\delta f(x)\Bigg|^{x=b} - \left(\frac{\partial F}{\partial y'} + \frac{\partial \varphi}{\partial y}\right)\delta f(x)\Bigg|^{x=a}$$

The variations δf fall into two main classes:

class 1: $\delta f : \delta f(a) = 0 \;\; \delta f(b) = 0$

class 2: $\delta f : \delta f(a) \neq 0 \;\; \delta f(b) \neq 0$

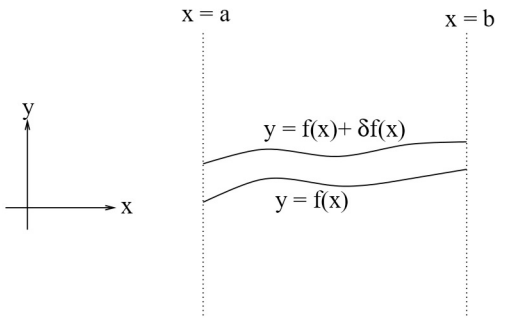


Figure 2.12: Trial curves with variable starting points and end points.

We know that $J =$ extremum $\Rightarrow \delta J = 0$ for all variations δf .
First we let δf be of class 1. In that case there are no endpoint terms in δJ. Consequently, $\delta J = 0 \Rightarrow F_y - \frac{d}{dx}(F_{y'}) = 0$.
Thus we are left with

$$\delta J = \left(F_{y'} + \frac{\partial \psi}{\partial y}\right)\delta f(x)\Big|^{x=b} - \left(F_{y'} + \frac{\partial \varphi}{\partial y}\right)\delta f(x)\Big|^{x=a}.$$

Next we let δf be of class 2. Then

$$\delta J = 0 \Rightarrow \left(F_{y'} + \frac{\partial \psi}{\partial y}\right)\Big|^{x=b} = 0 \text{ and } \left(F_{y'} + \frac{\partial \varphi}{\partial y}\right)\Big|^{x=a} = 0.$$

These are the optimal boundary conditions associated for the given variational principle, Eq.(2.41), with variable endpoints as expressed by the boundary terms ψ and ϕ.

2.3.1 Extremum Principle at a Moment of Time Symmetry

The extremum principle with side conditions and with variable end points is illustrated by the following example.

Consider a vibrating string imbedded in an elastic medium and with a spring attached to each end. The tension of the string, the coefficient of restitution of the elastic medium, and the spring constants are properties which determine the temporal evolution, i.e. the vibrations, of the system. These properties we assume to be independent of time. Consequently, the system is said to be time-invariant. One can show that for such a time-invariant system its total energy,

$$\text{T.E.} = \text{Kinetic energy} + \text{Potential energy}$$

is independent of time, i.e. the total energy of a time invariant system is conserved.

In general the vibratory motion of the system is such that its kinetic energy is never zero. Moreover, there exist certain types of motion, namely those in which the vibrations are characterized by a single frequency. Such states of motions are called *normal modes*. They have the property that at some instant of time the system is momentarily stationary. At that moment the system's velocity is zero, its kinetic energy is zero and all of its energy is in the form of potential energy,

$$\text{T.E.} = 0 + \text{P.E. (at a moment of time symmetry)}$$

We inquire as to the amplitude profile of the system at this moment of time symmetry. The amplitude profile which nature allows is governed by a extremum principle.

Consider the total energy of the system at a moment of time symmetry ($\frac{\partial y}{\partial t} = 0$)

$$\begin{aligned} \text{T.E.} = J[y] &= \frac{1}{2}\int_a^b \left[T(x)\left(\frac{dy}{dx}\right)^2 + \kappa(x)y^2\right] dx \\ &+ \frac{1}{2}k_1\left(y(a) - y_1\right)^2 + \frac{1}{2}k_2\left(y(b) - y_2\right)^2. \end{aligned}$$

This is the potential energy of the system. It consists of three parts:

1. The "stretch" energy

$$\int_a^b T\sqrt{(dx)^2 + (dy)^2} - \int_a^b T dx \approx \frac{1}{2}\int_a^b T\left(\frac{dy}{dx}\right)^2 dx$$

   Here $T = T(x)$ is the (position dependent) string tension.

2. The displacement energy due to the string being displaced in an elastic medium with a local restoration ("Hook's") coefficient

$$\kappa(x)dx$$

   for each incremental interval dx.

3. The potential energy stored in the two springs at $x = a$ and $x = b$. Their spring constants are k_1 and k_2.

The extremum principle governing the amplitude profile at a moment of time symmetry is given by

$$J[y] = \text{extremum} \tag{2.42}$$

subject to

$$K[y] = \frac{1}{2}\int_a^b \rho(x)y^2 dx = 1. \tag{2.43}$$

Here $\rho(x)$ is the linear mass density of the string.
With the help of the Lagrange multiplier λ, this principle reads

$$J[y] - \lambda K[y] = \text{extremum.}$$

Nota bene. It turns out that this principle is equivalent to

$$\frac{J[y]}{K[y]} = \text{extremum.}$$

This is Rayleigh's variational principle and we shall consider it later.

The Euler equation for either variational principle yield the same differential equation, namely

$$-\frac{d}{dx}T(x)\frac{dy}{dx} + (K(x) - \lambda\rho(x))y = 0 \tag{2.44}$$

This is the result of what we called "class 1" variations on page 63.

By contrast,"class 2" variations, which do not vanish at $x = a$ yield

$$T(x)y'(x)\Big|^a + k_1(y(b) - y_1) = 0$$

or

$$\frac{T(a)}{k_1}y'(a) + y(a) = y_1 \tag{2.45}$$

Similarly at $x = b$ one has

$$\frac{T(b)}{k_2}y'(b) + y(b) = y_2 \tag{2.46}$$

Equations 2.45 and 2.46 are known as the inhomogeneous mixed Dirichelet Neumann boundary conditions for the vibrating string. If we let the springs become infinitely stiff, i.e. let $k_1 \to \infty$ and $k_2 \to \infty$, then the problem reduces to a Dirichelet boundary value problem where the end point amplitude has the fixed values

$$y(a) = y_1$$
$$y(b) = y_2$$

We conclude that the Sturm-Liouville boundary value problem, Equations 2.44, 2.45 and 2.46 is a consequence of the variational principle,

$$\text{P.E.} - \lambda \int_a^b \frac{1}{2}\rho(x)y^2\,dx = \text{extremum.} \tag{2.47}$$

Remark 1. In vibrational problems the Lagrange multiplier λ, the eigenvalue of the S-L problem, refers to the squared vibrational angular frequency,

$$\lambda \equiv \omega^2 = (2\pi \text{ frequency})^2,$$

of the system, and the kinetic energy (K.E.) of the system vibrating at this frequency is

$$\frac{1}{2}\omega^2 \int_a^b \rho(x) y^2 dx.$$

Consequently, the amplitude profile of such a normal mode is governed by the variational principle

$$\text{P.E.} - \text{K.E.} = \text{extremum}.$$

We shall see later that the "least action principle"

$$\int_{t_1}^{t_2} (\text{P.E. - K.E.})\, dt = \text{extremum}$$

yields not only the amplitude profile at any instant of time, but also the evolution of this profile with time. Evidently, equation (2.47), the least energy problem with its amplitude constraint is a specialization of the least action principle.
Remark 2. This principle can be extended to two and three dimensions where the vibration can be those of a drum, solid, liquid sphere (vibrating sphere, vibrating neutron star, vibrating uranium nucleus), vibrating black hole, sound in cavity, vibrating bridge, etc.

Lecture 9

2.4 Generic Variable Endpoint Problem

There are variable end point problems in which the endpoint lies on a curve or a surface. For example,

1. what is the optimal curve from a point to a surface, or for that matter from one surface to another?

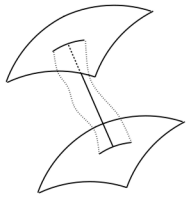


Figure 2.13: Optimal curve and its variants between two surfaces.

2. How does an extremal meet a terminating surface?

3. Given that a functional extends over two different media,

$$J = \int_a^c F(x, y, y')\, dx \; + \; \int_c^b F^*(x, y, y')\, dx \; ,$$

how must an extremal in $x < c$ be related to to its continuation in $c < x$, i.e. what are the junction conditions that prevail at the interface. For example, what path does light take as it passes from one medium to another?

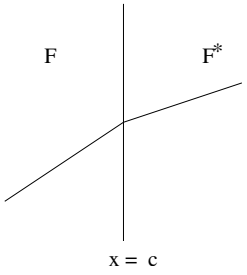


Figure 2.14: Curve crossing the boundary between two media characterized by F and F^*.

These questions and others can be answered within the framework of the following generic variable end point problem:
Find the curve $\{y_1(x), \ldots, y_n(x)\}$ in an n+1 dimension space for which the variational integral

$$J[y_1, \ldots, y_n] = \int_{x_0}^{x_1} F(x, y, \ldots, y_n, y_n', \ldots, y_n')dx = \text{extremum}$$

subject to the constraint that the starting and end points $\{y_i(x_0)\}$ and $\{y_i(x_1)\}$ lie on two prescribed curves (or surfaces).
This variational principle asks us to consider changes in J not only due to variation in the curves,

$$y_i \rightarrow y_i + \delta y_i,$$

but also due to variations in the points

$$\begin{aligned} x_0 &\rightarrow x_0 + \delta x_0 \\ x_1 &\rightarrow x_1 + \delta x_0, \end{aligned}$$

where the curves start and finish. The determination of these changes in J is a task which one must do before the endpoint constraints are imposed.
To simplify discussion we consider the simplest case

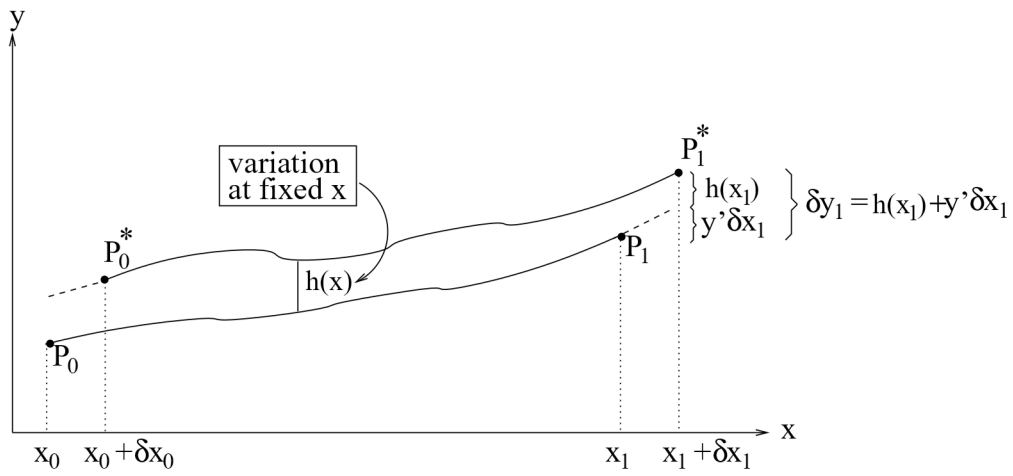


Figure 2.15: A curve and its generic variant.The total vertical variation δy_1 includes only the principal linear part. With that stipulation one has $h(x_1) = h(x_1 + \delta x_1)$.

$$J = \int_{x_0}^{x_1} F(x, y, y')dx = \text{extremum}$$

and calculate the first variation δJ allowing the endpoints to vary. Consider the curves

		starts	ends
curve:	$y(x)$	$P_0 = (x_0, y_0)$	$P_1 = (x_1, y_1)$
its variant:	$y^*(x)$	$P_0^* = (x_0 + \delta x_0, y_0 + \delta y_0)$	$P_1^* = (x_1 + \delta x_1, y_1 + \delta y_1)$

$$\boxed{h(x) = y^*(x) - y(x) = \text{difference at fixed } x}$$

2.4.1 General Variations in the Functional

The total variation in the functional to be extremized is

$$\begin{aligned}
\Delta J &= \int_{x_0+\delta x_0}^{x_1+\delta x_1} F(x, y+h, y'+h')dx - \int_{x_0}^{x_1} F(x, y, y')dx \\
&= \int_{x_0}^{x_1} [F(x, y+h, y'+h') - F(x, y, y')]dx \\
&+ \int_{x_0+\delta x_0}^{x_0} F(x, y+h, y'+h')dx + \int_{x_1}^{x_1+\delta x_1} F(x, y+h, y'+h')dx
\end{aligned}$$

whose principal linear part is

$$\delta J = \int_{x_0}^{x_1} [F_y - \frac{d}{dx} F_{y'}]h(x)dx + F_{y'}h(x)\Big|_{x_0}^{x_1} + F(x, y, y')\delta x\Big|_{x_0}^{x_1}.$$

The last term is new. It is proportional to the shift in x :

$$\delta x\Big|^{x_0} = \delta x_0 \quad \text{and} \quad \delta x\Big|^{x_1} = \delta x_1.$$

Referring to Figure 2.15, we express any type of end point variation in terms of the independent variations δy_i and δx_i :

$$\begin{aligned} h(x_0) = &\quad \delta y_0 - y'(x_0)\delta x_0 \\ h(x_1) = &\quad \delta y_1 - y'(x_1)\delta x_1 \end{aligned}$$

Thus the first variation is

$$\begin{aligned} \delta J \;\; = \;\; & \int_{x_0}^{x_1} \frac{\delta J}{\delta y(x)} h(x)dx + F_{y'}\delta y\Big|_{x=x_0}^{x=x_1} \\ & +(F - y'\frac{\partial F}{\partial y'})\delta x\Big|_{x=x_0}^{x=x_1} \end{aligned} \tag{2.48}$$

We have two endpoint terms because in general both y and x are allowed to change at the boundary.
This formula opens the door to a new mathematical perspective in optimizing the integral $J[y]$. It is the means

(a) to mathematize boundary conditions

(b) to mathematize junction conditions

(c) of entering into the “canonical" formulation of mechanics

(d) of entering into the Hamilton-Jacobi formulation of mechanics.

As we shall see in Chapter 3.5, the conceptual economy of this Hamilton-Jacobi formulation is breath-taking. This is because a system of several ordinary differential equations, the E-L Eqs.(2.1) on page 37, gets replaced by a *single* partial differential equation, the H-J equation.

In order to get already now a general idea of the relevant line of reasoning, consider the context of two close-by optimal curves $y(x)$ and $y^*(x)$, both satisfying the same E-L equation

$$0 = \frac{\delta J}{\delta y(x)} \equiv F_y - \frac{d}{dx}F_{y'} \;, \tag{2.49}$$

and each of them passing through its pairs of points

$$P_0 = (x_0, y_0) \qquad \& \qquad P_1 = (x_1, y_1) \tag{2.50}$$

and through

$$P_0^* = (x_0 + \delta x_0, y_0 + \delta y_0) \qquad \& \qquad P_1^* = (x_1 + \delta x_1, y_1 + \delta y_1) \ . \tag{2.51}$$

In light of Eq.(2.49) the expression for ΔJ, Eq.(2.48), becomes

$$\Delta J = F_{y'}\delta y\Big|_{x=x_0}^{x=x_1} + (F - y'\frac{\partial F}{\partial y'})\delta x\Big|_{x=x_0}^{x=x_1} \ .$$

It implies that J is a well-defined function of its end points,

$$J = J[y](x_0, y_0; x_1, y_1),$$

whenever $y(x)$ is an optimal curve. Because of this one has the following two partial derivatives

$$\frac{\partial J}{\partial x}\Big)_y \equiv \lim_{\delta x_1 \to 0} \frac{\Delta J}{\delta x_1}\Big)_y = F(x; y, y') - y'\frac{\partial F}{\partial y'}(x; y, y') \tag{2.52}$$

$$\frac{\partial J}{\partial y}\Big)_x \equiv \lim_{\delta y_1 \to 0} \frac{\Delta J}{\delta y_1}\Big)_x = F_{y'}(x; y, y') \tag{2.53}$$

Solving Eq.(2.53) for y' in terms of $\frac{\partial J}{\partial y}$ and substituting it for y' in Eq.(2.52), one obtains the partial differential equation

$$\frac{\partial J}{\partial x} + H\left(x; y, \frac{\partial J}{\partial y}\right) = 0 \ .$$

Its solution is the scalar function

$$J = J(x, y) \ .$$

Its determination replaces the task of solving the Euler-Lagrange differential Eq.(2.49), and hence the task of exhibiting the optimal extremals $y(x)$ in closed form. The physical significance, the mathematical nature of this scalar as well as the method by which it yields the extremal curves comprize the heart of the new mathematical perspective. We shall develop these in Chapter 3.5 after the Lagrangian and Hamiltonian mathematization of dynamics.

2.4.2 Transversality Conditions

Continuing the mathematization of optimized junction and boundary conditions, we now subject the first variation to prescribed endpoint constraints. We require that the paths start and terminate on the two respective given curves

$$y = \varphi(x) \tag{2.54}$$

and

$$y = \psi(x) \tag{2.55}$$

or equivalently,

$$x = \Phi(y) \tag{2.56}$$

and

$$x = \Psi(y) \tag{2.57}$$

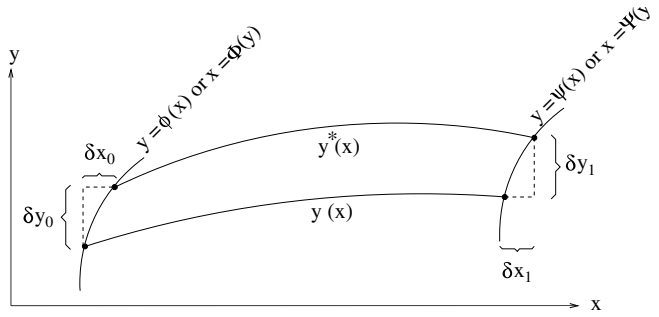


Figure 2.16: A curve and its variant with end point constraints.

In that case the vertical and horizontal variations in the end point are related by

$$\delta y_0 = \varphi'(x_0)\delta x_0 \quad \text{and} \quad \delta y_1 = \psi'(x_1)\delta x_1 \tag{2.58}$$

Consequently, the first variation in J is

$$\begin{aligned} \Delta J &= \int_{x_0}^{x_1} \frac{\delta J}{\delta y} h(x)dx + [F_{y'}\psi'(x) + (F - y'F_{y'})]_{x=x_1}\delta x_1 \\ &\quad -[F_{y'}\varphi'(x) + (F - y'F_{y'})]_{x=x_0}\delta x_0 \end{aligned}$$

If $y(x)$ is an optimal curve, then $\Delta J = 0$ for arbitrary variations $h(x)$ and arbitrary variations δx_i at the end points. The former implies that $y(x)$ satisfies the Euler equation. The latter implies

$$\begin{aligned} \Big[F + (\psi' - y')F_{y'}\Big]_{x=x_1} &= 0 \\ \Big[F + (\varphi' - y')F_{y'}\Big]_{x=x_0} &= 0, \end{aligned}$$

which are known as the *delta x transversality conditions*. By contrast, the *delta y transverality conditions* are

$$\left[F_{y'} + (F - y'F_{y'})\,\Psi'(y)\right]_{(x,y)=(x_1,y_1)} = 0$$
$$\left[F_{y'} + (F - y'F_{y'})\,\Phi'(y)\right]_{(x,y)=(x_0,y_0)} = 0$$

One must use this kind of transversality condition if there are many degrees of freedom, $\{\, y_1(x), y_2(x), \cdots, y_n(x)\,\}$, as is the case in Section 2.5 on page 75.

2.4.3 Junction Conditions

Lecture 10

Suppose one wished to find an optimal curve between two points in adjoining regions characterized by different variational integrands. Thus we consider an optimal curve which leaves one region characterized by F and enters another characterized by F^* so that

$$J = \int_a^c F(x, y, y')dx + \int_c^b F^*(x, y, y')dx = \text{extremum}.$$

The boundary between the two adjoining regions is $x = c.$, and it is depicted in Figure 2.14 on page 68. The extremum principle establishes a relationship between the properties of the curve in the two regions. To determine this relation we, as usual, consider the total first variation

$$\begin{aligned}\Delta J \;=\; & \int_a^c \frac{\delta J}{\delta y}\delta y(x)dx + \int_c^b \frac{\delta J}{\delta y}\delta y(x)dx \\ & + F_{y'}\delta y\Big|_a^{c^-} + F^*_{y'}\delta y\Big|_{c^+}^b \\ & + (F - y'\frac{\partial F}{\partial y'})\delta x\Big|_a^{c^-} + (F^* - y'\frac{\partial F^*}{\partial y'})\delta x\Big|_{c^+}^b \;.\end{aligned}$$

The fact that the curve is optimal implies

$$\frac{\delta J}{\delta y} = 0.$$

The fact that the boundary between the two regions is $x = c$, implies that $\delta x(c) = 0$, i.e. it is fixed and given.
Consequently the extremal condition implies

$$0 = \frac{\Delta J}{\delta y(c)} = F_{y'}\Big|_{c^+}^{c^-}$$

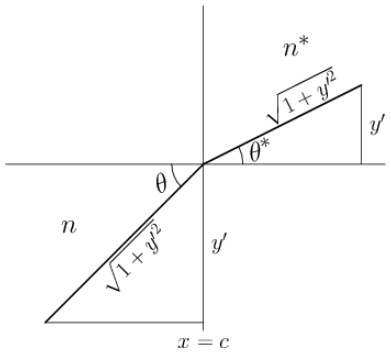


Figure 2.17: Ray refraction at the interface between two media having constant refractive indeces n and n^*.

$$\boxed{F_{y'}\big|_{c^+} = F^*_{y'}\big|_{c^-}} \text{ The Junction Condition}$$

Example 1. *(Snell's Law).*

Consider a light beam propagating from one medium where the refractive index is $n(x, y)$ to another where it is $n^*(x, y)$. The variational principle is

$$J = \int_a^c n(x,y)\,\sqrt{1+y'^2}dx + \int_c^b n^*(x,y)\,\sqrt{1+y'^2}dx = \text{extremum!}$$

The junction condition is

$$n(x,y)\frac{y'}{\sqrt{1+y'^2}}\bigg|_{x=c^-} = n^*(x,y)\frac{y'}{\sqrt{1+y'^2}}\bigg|_{x=c^+},$$

which according to Figure 2.17 is geometrized by the statement

$$n\sin\theta = n^*\sin\theta^*.$$

This is Snell's law of refraction.

Example 2. *(Segmented string)*

Consider a string with discontinuous tension. Such a discontinuity would prevail at the junction of two sting segments $0 \le x < c$ and $c < x \le \ell$ which are joined at $x = c$ by a frictionless ring. This ring (having neglegible mass) is constrained to slide without friction along a rod transverse to the segmented string. However, at the same time this rod pulls with non-zero force along the longitudinal direction of

that string. Consequently, there is a corresponding non-zero jump in the tension of the string segments along their junction. In fact, the difference, $T_2(c^+) - T_1(c^-)$, in the tension

$$T(x) = \begin{cases} T_1(x) & x < c \\ T_2(x) & c < x \end{cases}$$

on the two sides of the junction $x = c$ equals the longitudinal force provided by the frictionless rod.
The equation for the amplitude profile of a normal mode of any string is

$$\frac{d}{dx}T(x)\frac{dy}{dx} + [q(x) - \lambda\rho(x)]\, y = 0 \tag{2.59}$$

Q What are the junction conditions across the frictionless support rod?

A The variational principle for the amplitude profile is

$$\text{P.E.} = \frac{1}{2}\int_0^c \left[T_1\, y'^2 + q\, y^2\right] dx + \frac{1}{2}\int_c^\ell \left[T_2\, y'^2 + q\, y^2\right] dx = \text{extremum} \tag{2.60}$$

subject to

$$\int_0^\ell \rho\, y^2 dx = 1. \tag{2.61}$$

The Euler equationfor this "isoperimetric" problem is given by Eq.(2.59), where λ is the Lagrange multiplier for the given constraint, Eq.(2.61). The junction condition

$$F_{y'}|_{x=c^+} = F_{y'}|_{x=c^-}$$

yields

$$T_2(x)\left.\frac{dy}{dx}\right|_{c^+} = T_1(x)\left.\frac{dy}{dx}\right|_{c^-}.$$

In other words, the *transverse* force on the string parallel to the boundary $x = c$ is continuous across the boundary.

2.5 Many Degrees of Freedom

One would like to generalize the variable end point problem from extremal paths in the plane to extremal path in a higher dimensional space. In that case the two curves (2.56) and (2.57) generalize to the two surfaces

$$x = \Phi(y_1, \ldots, y_n) \tag{2.62}$$

and

$$x = \Psi(y_1, \ldots, y_n) \tag{2.63}$$

for the loci of points where the extremal path starts and finishes.

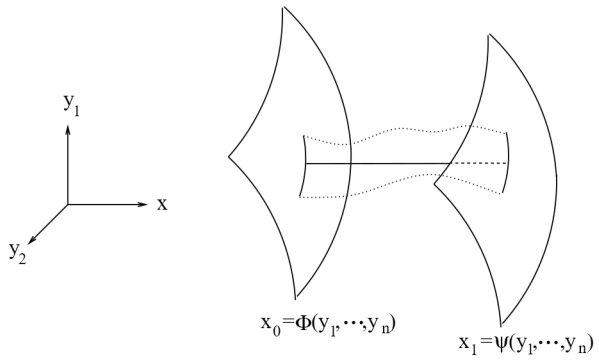


Figure 2.18: A curve and its variant between surfaces.

The corresponding generalization for the total first variation consists of replacing the equation (2.48) with

$$\begin{aligned} \Delta J &= \int_{x_0}^{x_1} \sum_{i=1}^{n} \frac{\delta J}{\delta y_i(x)} \delta y_i(x) dx + \sum_{i=1}^{n} F_{y_i'} \delta y_i \Big|_{x_0}^{x_1} \\ &\quad + \left(F - \sum_{j=1}^{n} y_j' \frac{\partial F}{y_j'} \right) \delta x \Big|_{x_0}^{x_1} \end{aligned}$$

Conceptual unit economy demands that it is best not to single out one of the y_i's over all the others. Consequently, we express the variation in δx at in terms of the variations of all the δy_i's in the surfaces (2.62) and (2.63),

$$\begin{aligned} \delta x_0 &= \sum_{i=1}^{n} \frac{\partial \Phi}{\partial y_i} \, \delta y_i \\ \delta x_1 &= \sum_{i=1}^{n} \frac{\partial \Psi}{\partial y_i} \, \delta y_i. \end{aligned}$$

With the help of these expressions the transversality conditions at these starting and terminating surface are given by

$$\left[F_{y_i'} + \frac{\partial \Phi}{\partial y_i} \left(F - \sum_{j=1}^{n} y_j' \frac{\partial F}{\partial y_j'} \right) \right]_{x=x_0} = 0, \qquad i = 1, \ldots, n$$

and analogous condition at $x = x_1$, where the gradient of the surface is $\left\{\frac{\partial \Psi}{\partial y_i}\right\}_{i=1}^{n}$.

Example 3. *(Fermat's Principle of Least Time).*
Consider Fermat's principle of least time for a light beam propagating through a medium whose refractive index is $n(x, y, z)$. Let the light beam start at (x_0, y_0, z_0) and finish at some point x_1, y_1, z_1 on the surface

$$x = \Psi(y, z)$$

such that the path minimizes the time of travel. The quantity to be extemized is

$$\begin{aligned} J[y, z] &= \frac{1}{c}\int_0^1 n(x, y, z)\sqrt{dx^2 + dy^2 + dz^2} \\ &= \frac{1}{c}\int_{x_0}^{x_1} n(x, y, z)\sqrt{1 + y'^2 + z'^2}dx \end{aligned}$$

One applies the transversality conditions to the variational integrand. We have

$$\begin{aligned} F &= n(x, y, z)\sqrt{1 + y'^2 + z'^2} \\ \frac{\partial F}{\partial y'} &= \frac{ny'}{\sqrt{1 + y'^2 + z'^2}}; \qquad \frac{\partial F}{\partial z'} = \frac{nz'}{\sqrt{1 + y'^2 + z'^2}} \\ F - y'F_{y'} - z'F_{z'} &= n\sqrt{1 + y'^2 + z'^2} - \frac{y'^2 + z'^2}{\sqrt{1 + y'^2 + z'^2}} \\ &= \frac{n}{\sqrt{1 + y'^2 + z'^2}} \end{aligned}$$

There are two transversality conditions at $x = x_1$. They are

$$\frac{ny'}{\sqrt{1 + y'^2 + z'^2}} + \frac{\partial \Psi}{\partial y}\frac{n}{\sqrt{1 + y'^2 + z'^2}} = 0$$

and

$$\frac{nz'}{\sqrt{1 + y'^2 + z'^2}} + \frac{\partial \Psi}{\partial z}\frac{n}{\sqrt{1 + y'^2 + z'^2}} = 0 \ .$$

Thus after cancellation we obtain

$$(y', z') = (-\frac{\partial \Psi}{\partial y}, -\frac{\partial \Psi}{\partial z})$$

or

$$(x', y', z') = (1, -\frac{\partial \Psi}{\partial y}, -\frac{\partial \Psi}{\partial z}) = \overrightarrow{\nabla} g$$

The left hand side is $\frac{d\vec{x}}{dx}$, the tangent to the light beam. the right hand side is the gradient of the function $g(x, y, z) = x - \Psi(y, z)$. This gradient is $\perp$ to the surface. Thus the transversality condition says that *the optimal light beam strikes the surface perpendicularly.*

2.6 Parametrization Invariant Problem

A mathematical problem is said to be geometrical if it is a statement whose validity does not depend on the coordinate system used in its enunciation.

Thus, the shortest trajectory between two point is a straight line, or the Pythagorean theorem, are geometrical statements.

In geometrical problems where one is dealing with the determination of curves which are optimal, it is more suitable not to single out any one coordinate over the others as an independent variable. Instead, represent these curves parametrically. Thus let

$$x = x(\lambda)$$
$$y = y(\lambda)$$

The choice of curve parameterization is optional, i.e. different parametrizations represent the same curve provided the functions x(λ) and y(λ) do *not* satisfy

$$\frac{dx}{d\lambda} = 0 \qquad \text{and} \qquad \frac{dy}{d\lambda} = 0$$

simultaneously. This resttriction guarantees that λ expresses non-zero progress along the curve.

2.6.1 Parametrization Invariance via Homogeneous Function

Consider the simplest variational problem. Its variational integral has the form

$$J = \int_{x_0}^{x_1} F\left(x, y, \frac{dy}{dx}\right) dx = \int_{\lambda_0}^{\lambda_1} G\left(x, y, \frac{dx}{d\lambda}, \frac{dy}{d\lambda}\right) d\lambda$$

where

$$G = \frac{dx}{d\lambda} F\left(x, y, \frac{\frac{dy}{d\lambda}}{\frac{dx}{d\lambda}}\right)$$

This integrand is a function which is "homogeneous" in the variables $\frac{dx}{d\lambda}$ and $\frac{dy}{d\lambda}$ and its "degree of homogeneity" is 1. It satisfies the homogeneity relation

$$G\left(x, y, k\frac{dx}{d\lambda}, k\frac{dy}{d\lambda}\right) = kG\left(x, y, \frac{dx}{d\lambda}, \frac{dy}{d\lambda}\right) \tag{2.64}$$

Following Euler we differentiate this expression with respect to k, and then settig $k = 1$, we obtain

$$x'G_{x'} + y'G_{y'} = G, \text{ where} x' = \frac{dx}{d\lambda} \text{ and} y' = \frac{dy}{d\lambda}. \tag{2.65}$$

Conversely, suppose that G is some homogeneous function of degree one in x' and y', i.e G satisfies Eq.(2.64), then the variational principle

$$\int_{\lambda_0}^{\lambda_1} G d\lambda = extremum$$

determines a parametrization invariant path. This is so becuase the parameter change $\lambda = \lambda(\bar{\lambda})$ yields

$$\begin{aligned}\int_{\bar{\lambda}_0}^{\bar{\lambda}_1} G\left(x, y, \frac{dx}{d\bar{\lambda}}, \frac{dy}{d\bar{\lambda}}\right) d\bar{\lambda} &= \int_{\bar{\lambda}_0}^{\bar{\lambda}_1} G\left(x, y, x'\frac{dx}{d\bar{\lambda}}, y'\frac{dy}{d\bar{\lambda}}\right) d\bar{\lambda} \\ &= \int_{\bar{\lambda}_0}^{\bar{\lambda}_1} G\left(x, y, x', y'\right) \frac{d\lambda}{d\bar{\lambda}} d\bar{\lambda} \\ &= \int_{\lambda_0}^{\lambda_1} G(x, y, x', y') d\lambda\end{aligned}$$

We conclude that *the variational problem is invariant under a change in parametrization*, and that the extremal path is independent of one's choice of parametrization used to represent it. The extremal path satisfies the Euler equations

$$G_x - \frac{d}{d\lambda} G_{x'} = 0 \qquad \text{and} \qquad G_y - \frac{d}{d\lambda} G_{y'} = 0. \tag{2.66}$$

These equations must be equivalent to the single equation

$$F_y - \frac{d}{dx} \frac{\partial F}{\partial \left(\frac{dy}{dx}\right)} = 0$$

This means that somehow the two equations (2.66) cannot be independent equations. There must exist an identity between them. In fact, one can show with the help of Eq. (2.65) that

$$x'(G_x - \frac{d}{d\lambda} G_{x'}) + y'(G_y - \frac{d}{d\lambda} G_{y'}) = 0$$

is an identity. It holds for an any path, extremal or non-extreamal.

We shall now apply these ideas to the geometrical problem of finding the geodesics in a space coordinatized by curvilinear coordinates.

2.7 Variational Principle for a Geodesic

Lecture 11

Consider flat Euclidean space coordinatized by (x,y,z) $= \vec{x}$. Introduce curvilinear coordinates (x^1, x^2, x^3) as an alternative means of labelling the points

$$\left.\begin{array}{l} x(x^1,x^2,x^3) \\ y(x^1,x^2,x^3) \\ z(x^1,x^2,x^3) \end{array}\right\} \vec{r}(x^i).$$

We would like to reexpress distance

$$(ds)^2 = dx^2 + dy^2 + dz^2$$

in terms of these curvilinear coordinates. Thus we consider

$$\begin{aligned} dx &= \sum_{i=1}^{3} \frac{\partial x}{\partial x^i} dx^i \\ dy &= \sum_{i=1}^{3} \frac{\partial y}{\partial x^i} dx^i \\ dz &= \sum_{i=1}^{3} \frac{\partial z}{\partial x^i} dx^i \end{aligned}$$

The distance formula becomes

$$\begin{aligned} (ds)^2 &= \sum_i \sum_j dx^i \frac{\partial x}{\partial x^i}\frac{\partial x}{\partial x^j} dx^j + dx^i \frac{\partial y}{\partial x^i}\frac{\partial y}{\partial x^j} dx^j + dx^i \frac{\partial z}{\partial x^i}\frac{\partial z}{\partial x^j} dx^j \\ &= \sum_i \sum_j dx^i \frac{\partial \vec{r}}{\partial x^i} \cdot \frac{\partial \vec{r}}{\partial x^j} dx^j \\ &= g_{ij}(x^k) dx^i dx^j \end{aligned}$$

Remark. The last line intoduces the *Einstein summation convention* in which a *pair of repeated indeces* implies a summation over the relevent coordinates. Also note that the coefficients

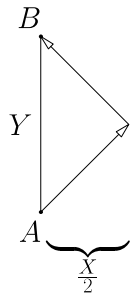
$$g_{ij} = \frac{\partial \vec{r}}{\partial x^i} \cdot \frac{\partial \vec{r}}{\partial x^j} = \vec{e}_i \cdot \vec{e}_j \tag{2.67}$$

are the inner products between ith and jth coordinate tangent vectors. They would not be mutually orthonormal in an oblique curvilinear coordinate system.

Q: How does one determine the optimal path between two given points A and B?

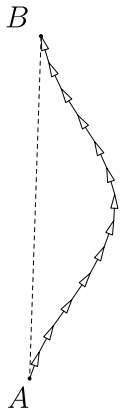
A: The optimal path is the one which extremizes the path length between A and B.

The simplest case is a broken curve consisting of two parts:



$$\begin{aligned} distance &= 2\sqrt{\left(\frac{Y}{2}\right)^2+\left(\frac{X}{2}\right)^2}=\sqrt{Y^2+X^2} \\ &= \text{extremum for the direct } (X{=}0) \text{ path} \end{aligned}$$

For broken curves consisting of several pieces, we have



$$\sum_i \sqrt{\Delta x_i^2+\Delta y_i^2}=extremum.$$

Among arbitrary curves beween A and B, the path length between A and B

$$\int_A^B \left[g_{ij}dx^i dx^j\right]^{1/2} = \left(\begin{array}{c}\text{minimum for a straightline}\\ \text{as compared to}\\ \text{any of its variants}\end{array}\right) ,$$

i.e. *extremal length* is an indicator of straightness.

Remark. The introduction of curvilinear coordinates in Euclidean space is optional. In such a space one can always find a global coordinate system x,y,z such that

$$(ds)^2 = dx^2 + dy^2 + dz^2.$$

In other spaces, e.g. a curved surface or more generally a Rimannian space, there is no option. The non-Euclidean nature of the space makes the introduction of curvilinear coordiantes mandatory. Relative to such coordinates one necessarily has

$$(ds)^2 = g_{11}\left(dx^1\right)^2 + 2g_{12}dx^1dx^2 + g_{22}\left(dx^2\right)^2 + \cdots$$

A particularly exotic example is the interior of a star. In terms of spherical coordinates (r,θ,ϕ), the distance formula is not the expected expression $(ds)^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$, but instead is given by

$$(ds)^2 = \frac{dr^2}{1 - \frac{8\pi}{3}\frac{G}{c^2}\rho r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

where the constants ρ, G, and c are

$$\begin{aligned} \rho &= \frac{\text{mass}}{\text{volume}} \\ G &= \text{Newton's gravitational constant } \left(= \frac{1}{15,\,000,\,\,000} \text{ in cgs units}\right) \\ c &= \text{speed of light } (= 3 \times 10^{10} \text{ in cgs units}) \end{aligned}$$

This expression is a consequence of Einstein's law of graviation.

Introducing the spatial scale parameter

$$a = \left(\frac{8\pi}{3}\frac{G}{c^2}\rho\right)^{-1/2}, \tag{2.68}$$

the distance formula has the form

$$(ds)^2 = \frac{dr^2}{1 - \frac{r^2}{a^2}} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right).$$

This suggests that one let

$$r = a\sin\chi$$

for the purpose of simplifying the distance formula. Doing so results in

$$(ds)^2 = a^2\left(d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\chi d\phi^2\right)\right).$$

One can show that this is the element of distance on part of a three-sphere S^3

$$x^2 + y^2 + z^2 + w^2 = a^2$$

of radius a, and which is coordinatized by

$$\begin{aligned} x &= a\ \sin\chi \sin\theta \cos\phi \\ y &= a\ \sin\chi \sin\theta \sin\phi \\ z &= a\ \sin\chi \cos\theta \\ w &= a\ \cos\chi \quad . \end{aligned}$$

The remarkable feature of this three-sphere is that its radius a is determined by the stars mass density[3] in Eq.(2.68).

2.8 Equation of Geodesic Motion

To determine the curve having extremal distance between A and B, we pick a curve with the suspected property and deform it slightly, but arbitrarily

$$\begin{aligned} \text{original curve: } x^i &= a^i(\lambda) \\ \text{variant = deformed curve: } x^i &= a^i(\lambda) + \delta a^i(\lambda)\,. \end{aligned}$$

1. Along either curve the distance is

$$s = \int_A^B ds = \int_0^1 \left[g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right]^{1/2} d\lambda\,. \tag{2.69}$$

2. At *fixed* λ we find that

   (a) the metric coefficient $g_{ij}\left(x^b(\lambda)\right)$ differs from one cuve to the other by

$$\delta g_{ij} \equiv g_{ij}\left(a^k(\lambda) + \delta a^k(\lambda)\right) - g_{ij}\left(a^k(\lambda)\right) = \frac{\partial g_{ij}}{\partial x^k} \delta a^k(\lambda)\,,$$

   (b) the components $\frac{dx^i}{d\lambda}$ of the tangent vector differ by

$$\delta\left(\frac{dx^i}{d\lambda}\right) \equiv \frac{d(a^i + \delta a^i)}{d\lambda} - \frac{da^i}{d\lambda} = \frac{d}{d\lambda}(\delta a^i)\,,$$

[3]For example, with a density of 1 gram/cc, that "radius" would be ∼60 solar radii.

(c) the application of these changes in g_{ij} and $\frac{dx^i}{d\lambda}$ to the integrand of Eq.(2.69) results in the following change in the integrated distance, Eq.(2.69),

$$\delta s = \int_0^1 \frac{\frac{1}{2} g_{ij} \frac{d}{d\lambda}(\delta a^i) \frac{da^j}{d\lambda} + \frac{1}{2} g_{ij} \frac{da^i}{d\lambda} \frac{d}{d\lambda}(\delta a^j) + \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \delta a^k \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{1/2}} \, d\lambda \,,$$

The first two term are equal. Integrate them by parts and omit the boundary term because both paths must pass through A and B:

$$\delta a^i(0) = \delta a^i(1) = 0 \,.$$

Thus obtain

$$\delta s = \int_0^1 f_k(\lambda) \delta a^k \left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{1/2} \, d\lambda \,,$$

where

$$f_k(\lambda) = \frac{-1}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{1/2}} \frac{d}{d\lambda} \left(\frac{g_{kj} \frac{da^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{1/2}} \right) + \frac{\frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]} \,. \quad (2.70)$$

An extremum is achieved, and the first order change δs vanishes for every first order deformation δa^k of the path $a^k(\lambda)$, when the three quantities that multiply the δa^k all vanish. Thus one obtains the n conditions

$$f_k(\lambda) = 0 \qquad k = 1, 2, 3, \cdots, n$$

for the determination of an optimal path.

In 3-dimensional Euclidean or Riemannian space these are three equations. In 4-dimensional Lorentz space-time these are four equations, even though in this latter case the space-time interval is given by

$$(d\tau)^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \,,$$

provided gravitation is absent.

Lecture 12

2.9 Geodesics: Their Parametrization.

The equations $f_k(\lambda) = 0$ constitute overkill! Why?

2.9.1 Parametrization Invariance.

It is because their number is more than enough to express the extrmal nature of

$$s = \int_0^1 [g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}]^{\frac{1}{2}} d\lambda$$

This follows from the parametrization independence of this integral. The reparametrization

$$\begin{aligned} \lambda \to \bar{\lambda} &= \lambda + h(\lambda) \\ \bar{\lambda}(0) &= 0 \\ \bar{\lambda}(1) &= 1 \Rightarrow h(0) = h(1) = 0 \end{aligned}$$

does not change the integral. It corresponds to a mere "repositioning of beads along a string" (= reparametrization)

$$s = \int_0^1 [g_{ij} \frac{dx^i}{d\bar{\lambda}} \frac{dx^j}{d\bar{\lambda}}]^{\frac{1}{2}} d\bar{\lambda}$$

The change in $a^i(\lambda)$ brought about by such a reparametrization is

$$a^i(\lambda) \to a^i(\lambda + h(\lambda)) = a^i(\lambda) + \delta a^i(\lambda)$$

where

$$\delta a^i(\lambda) = \frac{da^i}{d\lambda} h(\lambda)$$

The fact that such variations cannot change the variational integral for arbitrary reparametrization magnitudes $h(\lambda)$ implies that

$$\boxed{f_k(\lambda) \frac{da^k)}{d\lambda} = 0} \tag{2.71}$$

The remarkable feature of this equation is that it holds for *any* path $a^k(\lambda)$, even if it is not optimal.

An equation which holds true whether or not the quantities obey any differential equation is called an *identity*.

Thus instead of needing n independent equations, one needs only $n-1$ equations to specify an extremal path (in n dimensions).

2.9.2 Parametrization in Terms of Curve Length

This reparametrization freedom can be exploited to introduce a physically fundamental parameter, the length parameter itself,

$$ds = \left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{\frac{1}{2}} d\lambda$$

Let $x^i = a^i(\lambda)$ be the optimal path. With s as the new path parameter call this extrmal path $x^i(s)$. Making this replacement in the differential equation

$$f_k(\lambda) = 0$$

one obtains

$$0 = g_{kj}\frac{d^2x^j}{ds^2} + \frac{\partial g_{kj}}{\partial x^m}\frac{dx^m}{ds}\frac{dx^j}{ds} - \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^i}{ds}\frac{dx^j}{ds}$$

The middle term can be rewritten as

$$\frac{1}{2}\frac{\partial g_{kj}}{\partial x^m}\dot{x}^m\dot{x}^j + \frac{1}{2}\frac{\partial g_kj}{\partial x^m}\dot{x}^j\dot{x}^m$$

Let $m = \bar{j}$, $j = \bar{m}$ and then drop the bar. The result of this "index gymnastics" is

$$0 = g_{kj}\frac{d^2x^j}{ds^2} + \frac{1}{2}\left(\frac{\partial g_{kj}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{mj}}{\partial x^k}\right)\frac{dx^m}{ds}\frac{dx^j}{ds} \equiv f_k \tag{2.72}$$

Introduce the inverse g^{lk} of the matrix g_{kj}:

$$g^{lk}g_{kj} = \delta^l_j$$

The equation of the geodesic becomes

$$\boxed{\frac{d^2x^l}{ds^2} + \Gamma^l_{\ mj}\frac{dx^m}{ds}\frac{dx^j}{ds} = 0} \tag{2.73}$$

where

$$\Gamma^l_{\ mj} = \frac{1}{2}g^{lk}\left(\frac{\partial g_{kj}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{mj}}{\partial x^k}\right)$$

is the so called "Christoffel symbol of the 2^{nd} kind".

Remark. By contrast, without the $g^{\ell k}$, the coefficients

$$\Gamma_{k\ mj} \equiv \frac{1}{2}\left(\frac{\partial g_{kj}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{mj}}{\partial x^k}\right)$$

are the so called the "Christoffel symbols of the 1^{nd} kind". As we shall see later, both kinds characterize the differential geometric law of parallel transport. Both kinds have at first sight a daunting appearance, but they are computationally very powerful in obtaining this law from one's knowledge of the inner products $g_{ij} = \vec{e}_i \cdot \vec{e}_j$ at each point.

Comment. The identitiy Eq.(2.71) on page 85, or equivalently,

$$f_k \frac{dx^k}{ds} = 0 \tag{2.74}$$

should not come as a surprise. Recall that the Pythagorean theorem, $(\Delta s)^2 = g_{ij}\Delta x^i \Delta x^j$, implies

$$g_{kj} \frac{dx^k}{ds} \frac{dx^j}{ds} = 1, \tag{2.75}$$

even for nongeodesic curves; and, because of the arclength parametrization, the tangent $\left\{\frac{dx^k}{ds}\right\}$ has been normalized to unit length. It therefore is an integral of motion of the system of differential Eqs.(2.72) or (2.73). Indeed, differentiating the l.h.s. of Eq.(2.75) with respect to s yields

$$\frac{d}{ds}\left(g_{kj} \frac{dx^k}{ds} \frac{dx^j}{ds}\right) = f_k \frac{dx^k}{ds}, \tag{2.76}$$

where f_k is given by the r.h.s. of Eq.(2.72). The result is therefore this:
If $f_k = 0$ or if Eq.(2.73) holds, then $g_{kj}\frac{dx^k}{ds}\frac{dx^j}{ds}$ is a constant, an integral of motion. Conversely, given *any* s-parametrized curve along which $g_{kj}\frac{dx^k}{ds}\frac{dx^j}{ds}$ is constant, then Eq.(2.74) must be satisfied, *even* if $f_k \neq 0$.

2.10 Physical Significance of the Equation for a Geodesic

Lecture 13

Q What is the physical significance and purpose of the Γ-coefficients in the equation for a geodesic?

A These coefficients express the acceleration of a free particles relative to a given frame of reference. There are three archetypical reference frames: (i) free float coordinate frames, e.g. an orbiting space capsule not rotating relative to the fixed stars, (ii) rotating coordinate frames, and (iii) linearly accelerated coordinate frames.

2.10.1 Free float frame

Also called "inertial" frames of reference, such frames are defined by Newton's 1st Law of Motion:
This means that relative to a free float frame free particle move uniformly along

straight lines, i.e. these particles obey Newton's 1st law of motion. Mathematically we have

$$\text{mass} \times \text{acceleration}=0 \;\Rightarrow \frac{dx^i}{ds} = (const)^i \Leftrightarrow (\text{straight line motion})$$

where the travelled distance is

$$s = (constant) \times \text{time} \;\Leftrightarrow (\text{uniform motion})$$

2.10.2 Rotating Frame

Consider a frame of reference which rotates with angular velocity $\vec{\omega}$ relative to the fixed stars. First, consider a vector, say $\vec{G}$, which is rigidly attached to this rotating frame. The vector will rotate relative to the static inertial frame, which is static relative to the stars. In fact, during a time interval dt this rotating vector $\vec{G}$ will have changed by an amount which is given by

$$(d\vec{G})_{static} = dt\,\vec{\omega} \times \vec{G}$$

This expresses an infinitesimal rotation around the axis $\vec{\omega}$. Second, consider the circumstance in which the vector $\vec{G}$ is not rigidly attached to the rotating frame. Instead, let it change by the amount $(d\vec{G})_{rot}$ during the time interval dt, then the total change in this vector will be

$$(d\vec{G})_{static} = (d\vec{G})_{rot} + dt\,\vec{\omega} \times \vec{G}$$

relative to the inertial frame static with respect to the fixed stars. Thus

$$\frac{d}{dt}\Big)_{static} = \frac{d}{dt}\Big)_{rot} + \vec{\omega}\times$$

Finally, apply this transformation law to the postion vector $\vec{R}$,

$$\frac{d\vec{R}}{dt}\Big)_{static} = \frac{d\vec{R}}{dt}\Big)_{rot} + \vec{\omega} \times \vec{R}$$

and then to the velocity vector $\frac{d\vec{R}}{dt}\Big)_{static} \equiv \vec{v}_s$. Assuming that $\vec{\omega}$ is constant, one obtains that the equation of motion for a free particle is

$$0 \;=\; m\frac{d^2\vec{R}}{dt^2}\Big)_{static} = m\left[\vec{a}_{rot} + 2\vec{\omega} \times \vec{v}_{rot} + \vec{\omega} \times (\vec{\omega} \times \vec{R})\right]$$

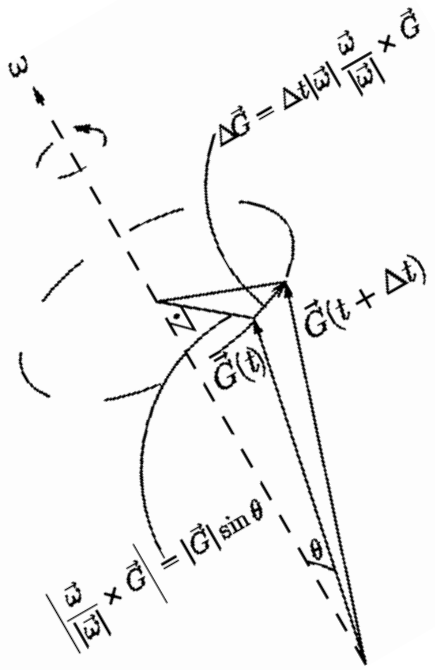


Figure 2.19: Rigid frame rotating relative to the fixed stars with angular velocity $\vec{\omega}$. Its direction as well as its magnitude $|\vec{\omega}|$ are taken to be fixed relative to the fixed stars. Being fixed relative to the rotating frame, the vector $\vec{G}(t)$ rotates relative to the fixed stars with angular velocity $\vec{\omega}$. The vector $\frac{\vec{\omega}}{|\vec{\omega}|}\times\vec{G}$ is $\perp$ to $\vec{\omega}$ and $\vec{G}$, but its length $\left|\frac{\vec{\omega}}{|\vec{\omega}|}\times\vec{G}\right|$ is equal to that of the radius vector $\perp$ to $\vec{\omega}$.

with

$$\vec{v}_{rot} = \left.\frac{d\vec{R}}{dt}\right)_{rot}$$

and

$$\vec{a}_{rot} = \frac{d\vec{v}_{rot}}{dt}\Big)_{rot}$$

In terms of components one has

$$m\frac{d^2x^i}{dt^2}\Big)_{rot} = \underbrace{-2m[\vec{\omega}\times\vec{v}_{rot}]^i}_{Coriolis\ force}\underbrace{-m[\vec{\omega}\times(\vec{\omega}\times\vec{R})]^i}_{centrifugal\ force} \qquad (8.8) \qquad (2.77)$$

relative to the *rotating* coordinate frame. The two terms on the right hand side of this equation are the Coriolis force" and the "centrifugal force" in these frames. Suppose we try to compare these Newtonian equations of motion with those of a geodesic,

$$\frac{d^2x^i}{ds^2} = -\Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds}$$

This is a good thing to do because both sets of equations have seond derivatives on the left hand side. The obstacles are the independent variables, time t for Newton's equations, and geodesic length s for the geodesic equations. But this difference is no problem because for each free particle path, length and time are linearly related:

$$\begin{aligned} s &= (constant)\times \text{time} \\ &\equiv (constant)\, t \end{aligned} \qquad (2.78)$$

Consequently, the geodesic equations are equivalent to

$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{jk}\frac{dx^i}{dt}\frac{dx^k}{dt} \qquad (8.10) \qquad (2.79)$$

$$\frac{d^2t}{dt^2} = 0\,. \qquad (8.11) \qquad (2.80)$$

Suppose we introduce time

$$x^0 = t\,.$$

as another coordinate so that

$$\frac{dx^0}{dt} = 1$$

and

$$\frac{d^2x^0}{dt^2} = 0\,. \qquad (2.81)$$

We now consolidate the two equations (2.79) and (2.81) into the single four-component equation

$$\frac{d^2x^\mu}{dt^2} = -\sum_{\alpha=0}^{3}\sum_{\beta=0}^{3}\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} \qquad \mu = 0,1,2,3 \tag{2.82}$$

$$= -\Gamma^\mu_{00}\frac{dx^0}{dt}\frac{dx^0}{dt} - \sum_{k=1}^{3}(\Gamma^\mu_{0k}+\Gamma^\mu_{k0})\frac{dx^0}{dt}\frac{dx^k}{dt} - \sum_{j=1}^{3}\sum_{k=1}^{3}\Gamma^\mu_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} \tag{2.83}$$

The fact that $\frac{dx^0}{dt} = 1$ results in

$$\frac{d^2x^\mu}{dt^2} = -\Gamma^\mu_{00} - \sum_{k=1}^{3}(\Gamma^\mu_{0k}+\Gamma^\mu_{k0})\frac{dx^k}{dt} - \sum_{j=1}^{3}\sum_{k=1}^{3}\Gamma^\mu_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt}$$

What are these Γ-coefficients? They are determined by doing a component-by-component comparison between this four-component equation with the three-plus-one set of Eqs.(2.77) and (2.81) and requiring that these two sets are equivalent for all particles. For example, the first component ($i = 1$) of Eq.(2.77) is

$$\frac{d^2x^1}{dt^2} = -2(\omega^2\frac{dx^3}{dt} - \omega^2\frac{dx^2}{dt}) - (\vec{\omega}\cdot\vec{\omega}x^1 - \vec{\omega}\cdot\vec{R}\omega^2)$$

For $\mu = 0$ equivalence implies

$$\Gamma^0_{\alpha\beta} = 0 \qquad \alpha,\beta = 0,1,2,3\ ,$$

while for $\mu = i\ (= 1,2,3)$ one finds that in the rotating frame

$$\Gamma^i_{0k}\frac{dx^k}{dt} = \left(\vec{\omega}\times\frac{d\vec{R}}{dt}\right)^i \qquad i = 1,2,3 \quad \text{(Coriolis acceleration)}$$

so that[4]

$$\begin{matrix} i=1\rightarrow \\ i=2\rightarrow \\ i=3\rightarrow \end{matrix}\begin{bmatrix} \\ \Gamma^i_{0k} \\ \\ \end{bmatrix}\begin{bmatrix} dx^1/dt \\ dx^2/dt \\ dx^3/dt \end{bmatrix} = \begin{bmatrix} 0 & \omega^3 & -\omega^2 \\ -\omega^3 & 0 & \omega^1 \\ \omega^2 & -\omega^1 & 0 \end{bmatrix}\begin{bmatrix} dx^1/dt \\ dx^2/dt \\ dx^3/dt \end{bmatrix}.$$

Furthermore,

$$\Gamma^i_{00} = \left[\vec{\omega}\times(\vec{\omega}\times\vec{R})\right]^i \qquad i = 1,2,3 \quad \text{(centrifugal acceleration)}\,.$$

[4]The coefficients Γ^μ_{0k} and of Γ^μ_{k0} of the quadratic form in the geodesic equation occur only in the symmetric combination $(\Gamma^\mu_{0k}+\Gamma^\mu_{k0})$. Consequently, one may assume without loss of generality that $\Gamma^\mu_{0k} = \Gamma^\mu_{k0}$.

In other words,

$$\begin{aligned} \Gamma^i_{00} &\neq 0 \Leftrightarrow \text{centrifugal force} \neq 0 \\ \Gamma^i_{0k} &\neq 0 \Leftrightarrow \text{Coriolis force} \neq 0 \end{aligned}$$

This says that the "Christoffel symbols" of the equations for a geodesic are an expression of centrifugal and Coriolis accelerations, which are associated with a rotating coordinate frame. Let us now extend these same considerations to an accelerated frame of reference.

2.10.3 Uniformly Accelerated Frame

Recall that relative to a free float ("inertial") frame the equation for the free particle moving along the x-direction is

$$\frac{d^2x}{dt^2} = 0\,. \tag{2.84}$$

Consider the motion of such a particle relative to an accelerated frame.

Q What is the equation of motion of this particle relative to a frame accelerating into the positive x-direction with acceleration g?

A To find out, consider a meter rod parallel to the x-direction, but attached to the accelerated frame.

Let ξ be the coaccelerating coordinate displacement along this meter rod. Thus a coaccelerating point on this meter stick is measured to be ξ units from the floor.

However, relative to the nonaccelerated inertial frame the x-coordinate of this point on the meter stick is

$$x = \xi + \frac{1}{2}g\tau^2$$

with

$$t = \tau$$

These two equations comprise the coordinate transformation $(\tau, \xi) \rightarrow (t, x)$ between the two frames. Let

$$\xi = \xi(\tau)$$

be the particle trajectory as measured in the accelerated frame. Appling the coordinate transformation to Newton's Eq.(2.84) for the free particle yields

$$0 = \frac{d^2}{d\tau^2}[\xi(\tau) + \frac{1}{2}g\tau^2] = \frac{d^2\xi}{d\tau^2} + g\,.$$

Furthermore,

$$0 = \frac{d^2t}{d\tau^2}\,.$$

Let $x^0 = t$ and $x^1 = \xi$. The equations assume the form

$$\begin{aligned}\frac{d^2x^0}{d\tau^2} &= 0\\ \frac{d^2x^1}{d\tau^2} &= -g\end{aligned}$$

Compare them to the equation for a geodesic in two dimensions:

$$\begin{aligned}\frac{d^2x^0}{d\tau^2} &= -\Gamma^0_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}\\ \frac{d^2x^1}{d\tau^2} &= -\Gamma^1_{00}\left(\frac{dx^0}{d\tau}\right)^2 - 2(\Gamma^1_{01}+\Gamma^1_{10})\frac{dx^0}{d\tau}\frac{dx^1}{d\tau} - \Gamma^1_{11}\left(\frac{dx^1}{d\tau}\right)^2\end{aligned}$$

Assume they apply to all possible particle trajectories, we find

$$\Gamma^1_{00} = g = \text{"inertial acceleration"} \qquad (8.15) \qquad (2.85)$$

All the other Γ's are zero. One concludes therefore that relative to a uniformly accelerated frame

$$\Gamma^1_{00} \neq 0 \Leftrightarrow \text{"non-inertial force"} \neq 0\,.$$

2.11 The Equivalence Principle and "Gravitation"="Geometry"

Einstein directed attention to the fact that there is no way that one can tell the difference between (i) the motion of free particles in a uniformly accelerated frame, and (ii) the motion of particle falling freely in a uniform gravitational field.

This holds true regardless of the particle's composition. In other words, the particle trajectories are identical, no matter whether a particle is made of gold, aluminum, or snakewood. The measurements were first done by Lorand von *Eötvös*. Thus the inertial force is equvalent to, i.e. physically 'indistinguishable from, the gravitational force.

This equivalence is called *the equivalence principle*. It implies that

$$\text{"inertial force"} = \text{"gravitational force"}\,, \qquad (2.86)$$

or

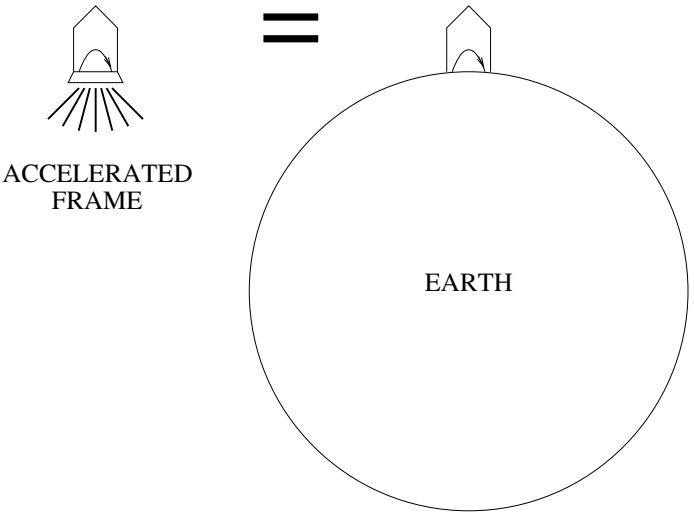


Figure 2.20: Equivalence Principle: the trajectories of particles in an accelerated frame and in a uniform gravitational field are indistinguishable regardless of their composition.

$$\text{mass}_{inertial} \times \text{"inertial acceleration"} = \text{mass}_{gravitational} \times \text{gravitational field}$$

This equation holds regarless of whether the inertial mass and the gravitational mass refer to particles made of gold, aluminum, etc. Thus

$$\frac{m_{in}}{m_{acc}} = 1\,,$$

which expresses the fact that this ratio is independent of the composition of the paricles. Doing the cancellation and appying Eq.(2.85) to Eq. (2.86), we find that the Christoffel symbol Γ^1_{00} is to be identified with the graviational field, i.e.

$$\Gamma^i_{00} = (\text{"gravitational field"})^i \quad i = 1, 2, 3 \qquad (8.18a) \tag{2.87}$$

if we had extended our considerations to motion in three dimenssions.

Apply this result to a static gravitational field. In that case

$$\Gamma^i_{00} = \sum_{\alpha=0}^{3} \frac{1}{2} g^{i\alpha}(g_{\alpha 0,0} + g_{\alpha 0,0} - g_{00,\alpha})$$

The fact that the gravitatioal field is independent of time means that all time-derivatives must vanish. Consequently, the expression reduces to

$$\Gamma^i_{00} = \sum_{j=1}^{3} \frac{1}{2} g^{ij}(-) g_{00,i} \tag{2.88}$$

If the gravitational field is weak, then we can say that distance measurements in three dimenssional space are governed by Euclidean geometry as expressed by the theorem of Pythagoras,

$$(ds)^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \,.$$

This implies that the matrix, Eq. (2.67), on page 80 has the simple form

$$[g_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Consequently, its inverse $[g^{ij}]$is also the identity matrix,

$$[g^{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \,.$$

Applying this to the simplified Christoffell symbol, Eq.(2.88), give the result

$$\Gamma^i_{00} = -\frac{1}{2}\frac{\partial g_{00}}{\partial x^i} \qquad i = 1, 2, 3$$

Introduce this simplified result into Eq. (2.87) and obtain

$$-\frac{1}{2}\frac{\partial g_{00}}{\partial x^i} = (\text{gravitational field})^i \qquad i = 1, 2, 3 \,.$$

Recall that the gravitational field is the negative gradient of Newton's gravitational potential $\Phi(x^1, x^2, x^3)$,

$$(\text{gravitational field})^i = -\frac{\partial}{\partial x^i}\Phi \,.$$

Consequently,

$$\Phi(x^1, x^2, x^3) = \frac{1}{2} g_{00}(x^1, x^2, x^3) \,.$$

This is a new result. It says that

$$\text{"gavitational potential"} \;\; = \text{"metric coefficient"} \,.$$

In other words, by examing the motion of particles we find that gravitation manifests itself by altering the properties of space (and time) as expressed by the coefficients in the "metric"

$$(ds)^2 = g_{\alpha\beta} dx^\alpha dx^\beta$$

and by the concomitant Christoffel symbols

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2} g^{\mu\gamma}(g_{\gamma\alpha,\beta} + g_{\gamma\beta,\alpha} - g_{\alpha\beta,\gamma}) \qquad \mu, \alpha, \beta = 0, 1, 2, 3$$

One summarizes this result by saying that Einstein's *equivalence principle* leads to the conclusion that

$$\text{“gravitation"} = \text{“geometry"}\,.$$

Remark. The reasoning process that leads from (a) the equivalence between (i) trajectories of free particles in an accelerated frame and (ii) the trajectories of particles in a gravitaqtional field to (b) the identification of gravitation as a geometrical property is an example of the scientific method, induction based on experiments and/or observation.

Chapter 3

Variational Formulation of Mechanics

3.1 Hamilton's Principle

Lecture 14

Recall Newton's second law of motion applied to a force derived from a potential $U(x, y, z, t)$ is

$$\frac{d}{dt}\left(m\dot{\vec{x}}\right) = -\vec{\nabla}U(x, y, z, t)\,.$$

Like all differential equations, this law expresses – in mathematical form – a causal relation in the world. Here the change in momentum is caused by a force. This law governs the change in the particle's momentum along every point of the particle's trajectory. This law is therefore a local one, it is a differential statement about the motion of the particle. By integrating this differential equation one obtains the global trajectory, an integral curve. This is another causal relation, this time between the particle's initial position and velocity on one hand and its final position and velocity on the other. If one knows the initial velocity at point event A, then the integration of Newton's equation yields a unique curve which terminates at, say, point event B.

One now asks: given that there exists a unique curve through A and B, does there exist a principle which singles out this curve over any other – a principle with a mathematical starting point different from Newton's local equations of motion?

The answer is given by the principle of least (more generally, stationary) action, also called *Hamilton's Principle*:

> Of all possible paths along which a dynamical system may (be imagined to) move from one point to another within a specified time interval (consistent with any constraints), the actual path followed is that which minimizes the time integral of the difference between the kinetic and potential energies.

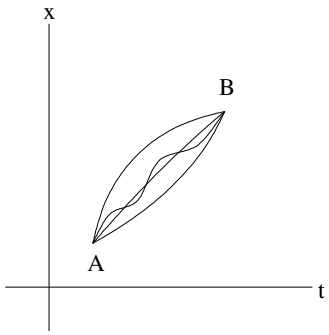


Figure 3.1: Optimal and non-optimal paths in spacetime

3.1.1 Prologue: Why $\int(\boldsymbol{K.E.} - \boldsymbol{P.E.})\boldsymbol{dt} = \boldsymbol{minimum}$?

Example. In its simplest manifestations, Hamilton's principle is a mathematical expression of two fundamental physical aspects of nature: Newton's First Law of Motion and the Equivalence Principle. This can be seen from the following considerations:
A. Launch a particle vertically from x_1 at time t_1, watch it reach its maximum height, and then catch it at time t_2 at the instant it is located at x_2.

From Galileo we learned that in its travel from $(t_1,\, x_1)$ to $(t_2,\, x_2)$ the particle traces a space-time trajectory which is given by a parabola. Why so? Answer:

1. Newton's 1st Law: Every body continues in its state of rest, or of uniform motion in a straight line, unless it iscompelled to change that state by forces impressed upon it.

2. The principle of equivalence.

A. Simpler case: Free Particle.
Consider the motion of a particle moving freely in a free float ("inertial") frame. This particle moves with constant velocity, i.e. its space-time trajectory is a straight line.

The implication of this fact is that for such a curve the integral

$$\frac{1}{(t_2 - t_1)} \int_{t_i}^{t_2} \left(\frac{dx(t)}{dt} \right)^2 dt \equiv \langle v^2 \rangle = min\,!$$

as compared to other curves having the same starting and termination points.
Q: Why?

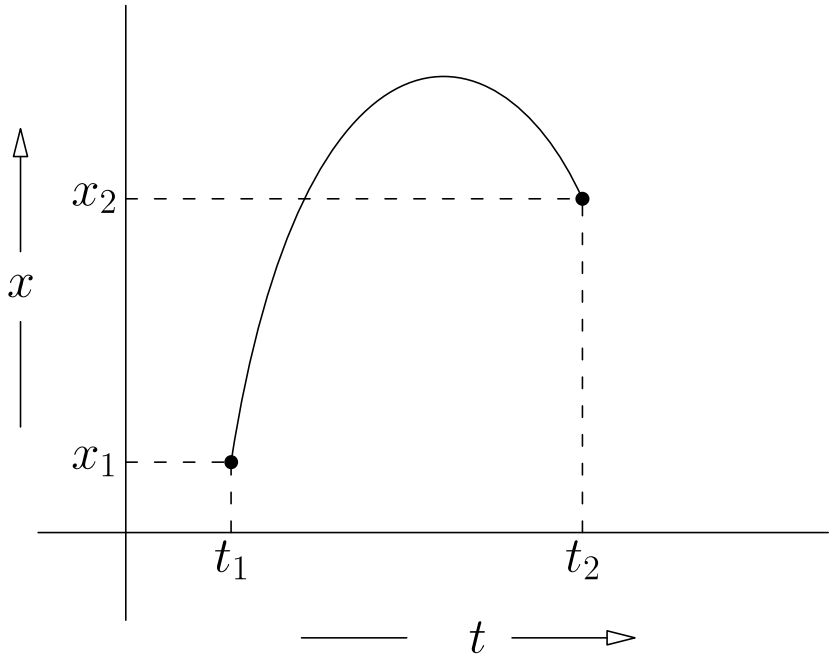


Figure 3.2: Spacetime trajectory of a particle thrown into the air.

A: All such curves have the same average velocity,

$$\langle \bar{v} \rangle = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \frac{d\bar{x}}{dt} dt = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \frac{dx}{dt} dt = \langle v \rangle = v,$$

which means that the area under the curves $\bar{v}(t)$and $v(t) = v$ are the same.

Applying this fact to the positivity of the averaged squared deviation (away from the average),

$$0 \leqslant \langle (\bar{v} - \langle \bar{v} \rangle)^2 \rangle = \langle \bar{v}^2 \rangle - (\langle \bar{v} \rangle)^2 = \langle \bar{v}^2 \rangle - (\langle v \rangle)^2 ,$$

one has

$$\langle \bar{v}^2 \rangle \geqslant (\langle v \rangle)^2 = v^2,$$

or

$$\int_{t_i}^{t_2} \left(\frac{d\bar{x}(t)}{dt} \right)^2 dt \geqslant \int_{t_i}^{t_2} \left(\frac{dx(t)}{dt} \right)^2 dt.$$

This says that a free particle moves so that the integral of its kinetic energy is a minimum:

$$\int_{t_i}^{t_2} K.E.\, dt \equiv \int_{t_i}^{t_2} \frac{1}{2} m \left(\frac{dx(t)}{dt} \right)^2 dt = min\,!$$

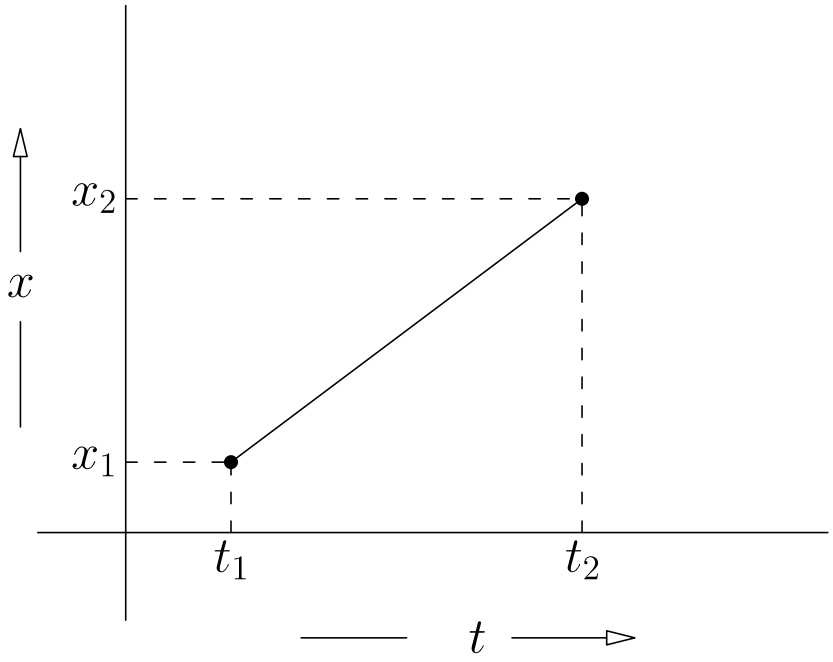
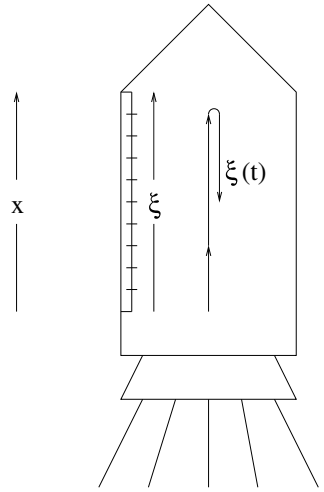


Figure 3.3: Spacetime trajectory of a free particle is a straight line.

B. Free particle in an accelerated frame.

Consider the motion of the same particle moving freely in a frame accelerated uniformly with acceleration g.



A point ξ fixed in the accelerated frame will move relative to the free float frame

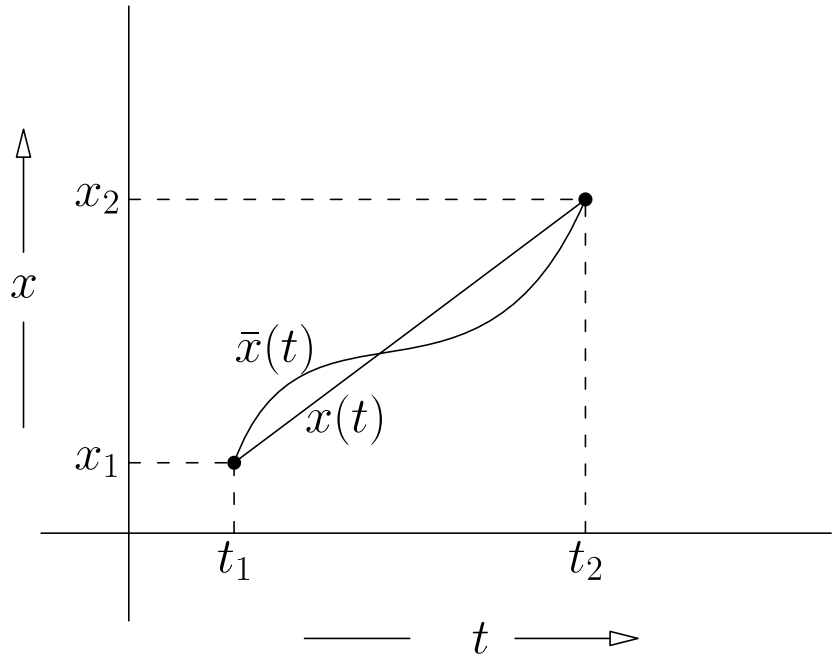


Figure 3.4: Straight line $x(t)$ and its variant $\bar{x}(t)$ have the same average velocity: $\langle \bar{v} \rangle = v \, (= const.)$.

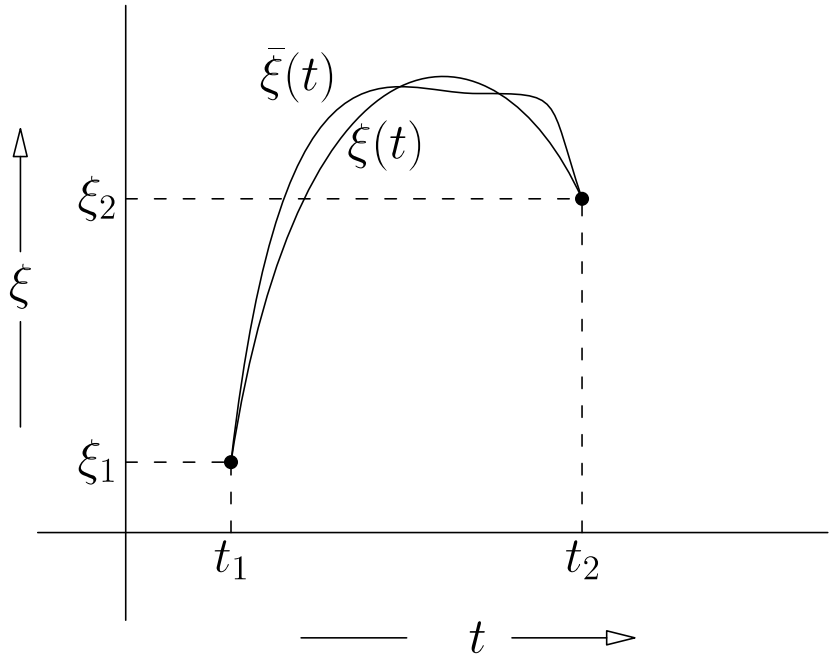


Figure 3.5: Minimizing trajectory $\xi(t)$ and one of its variants $\bar{\xi}(t)$.

according to

$$x = \xi + \frac{1}{2}gt^2.$$

It follows that, relative to the accelerated frame, the spacetime trajectory of the particle, $\xi(t)$, is given by

$$x(t) = \xi(t) + \frac{1}{2}gt^2. \tag{3.1}$$

Here $x(t)$ is the linear trajectory in Figure 3.3.

The to-be-minimized integral takes the form

$$\begin{aligned} min = \int_{t_i}^{t_2} \left(\frac{dx(t)}{dt}\right)^2 dt &= \int_{t_i}^{t_2} \left(\frac{d\xi}{dt} + gt\right)^2 dt \\ &= \int_{t_i}^{t_2} \left\{ \left(\frac{d\xi}{dt}\right)^2 + 2gt\frac{d\xi}{dt} + g^2t^2 \right\} dt \\ &= \int_{t_i}^{t_2} \left\{ \left(\frac{d\xi}{dt}\right)^2 - 2g\xi \right\} dt + 2gt\xi|_{t_1}^{t_2} + \frac{1}{3}gt^2|_{t_1}^{t_2} \end{aligned}$$

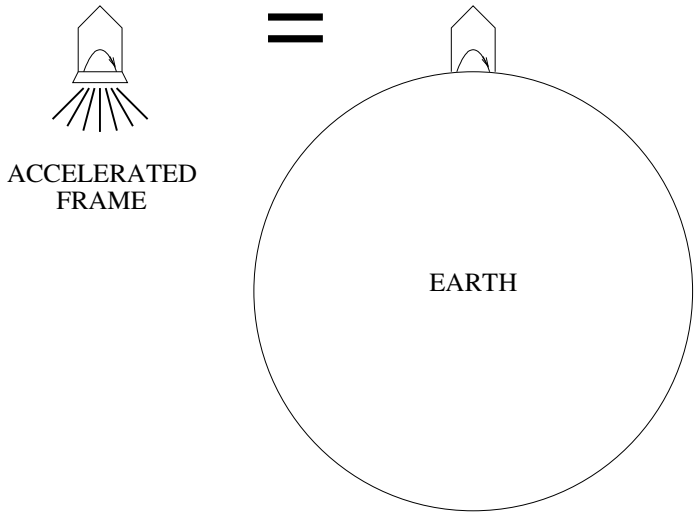


Figure 3.6: Trajectories in an accelerated frame are indistinguishable from those in a gravitational field. In particular the motion of particles of different composition (gold, aluminum, snakewood, etc.) is independent of their composition.

The last line is the result of an integration by parts. The last two terms are the same for all trajectories passing through the given points $(t_1,\, x_1)$ and $(t_2,\, x_2)$. Consequently,

$$\int_{t_i}^{t_2} \frac{1}{2} m \left(\frac{dx(t)}{dt} \right)^2 dt = min \iff \int_{t_i}^{t_2} \left\{ \frac{m}{2} \left(\frac{d\xi}{dt} \right)^2 - mg\xi \right\} dt = min$$

C. Free particle in an *equivalent* gravitational field.

The *equivalence principle* is an observation of the fact that in an accelerated frame the laws of moving bodies are the same as those in a homogeneous gravitational field.

Recall that in a gravitational field

$$mg\xi = P.E.$$

represents the potential energy of a mass m at a height ξ. Consequently, the trajectory of a particle in a gravitational field is determined by

$$\int_{t_i}^{t_2} (\, K.E. \, - P.E. \,)\, dt \equiv \int_{t_i}^{t_2} L(\dot{x},\, x,\, t)\, dt = min.$$

In fact, the trajectory of a particle which satisfies this minimum condition satisfies the Euler-Lagrange

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x},$$

which is Newton's second law of motion

$$ma = F$$

for the one-dimensional motion of a particle.
Nota bene:

1. The same minimum priciple hold even if g, and hence the potential energy $P.E.$, depends explicitly on time.

2. This principle is a special case of what is known as Hamilton's principle of least action. The difference is that the latter also accomodates motion which are subject to constraints.

D. Extension to multi dimensions and generic potentials.

The Lagrangian formulation opens new vistas on the notion of bodies. It can be fruitfully implemented for more general motions and potentials. These generalizations are alternate but equivalent formulations of Newtonian mechanics. They are simply expressed by the statement that

$$\int_{t_i}^{t_2} (\, K.E. \, - P.E. \,)\, dt = min$$

with

$$K.E. = \frac{1}{2} \sum_{i=1}^{n} m_i \dot{\vec{x}}_i \cdot \dot{\vec{x}}_i$$

$$P.E. = U(t, \vec{x_i})$$

on the class of all system trajectories having fixed endpoints.

The advantage of Lagrangian Mechanics becomes evident in the process of setting up Newton's equations of motion. In Newtonian Mechanics one must do this for each force component separately, a task which becomes non-trivial relative to curvilinear coordinate frames (spherical, cylindrical, etc.). By contrast, in the Lagrangian approach one merely identifies the two scalars $K.E$ and $P.E.$ relative to the coordinate frame of one's choice. The remaining task of setting up the differential equations of motion is done automatically by merely writing down the Euler-Lagrange equations.

In terms of the calculus of variations, the principle of stationary action is simply

$$\int_{B}^{A} (T - U) dt = \text{extremum!}$$

Here $T(\dot{x}^i)$ is the kinetic energy (*K.E.*) and $U(x^i, t)$ is the potential energy (*P.E.*). The integrand $T(\dot{x}^i) - U(x^i, t) \equiv L(\dot{x}^i, x^i, t)$ is the *Lagrangian* of the system. Thus the stationary action principle can be stated as

$$\delta \int_A^B L(x^i, \dot{x}^i, t) dt = 0 \, .$$

3.1.2 Hamilton's Principle: Its Conceptual Economy in Physics and Mathematics

Lecture 15

Hamilton's principle lends itself to a treatment within the frame work of the calculus of variations if one identifies the action with the variational integral

$$J[y_i] = \int_{x_0}^{x_1} F(x, y_i, y_i') dx$$

By making the identificaiton

$$\begin{aligned} x &\rightarrow t \\ y_i(x) &\rightarrow x^i(t) \\ y_i'(x) &\rightarrow \dot{x}^i(t) \\ F(x, y_i, y_i^{'}) &\rightarrow L(x^i, \dot{x}^i, t) \end{aligned}$$

one obtains the Euler-Lagrange equations,

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} = 0 \, .$$

These equations of motion reproduce the Newtonian equations of motion. This fact is illustrated by means of the following example. For n particles, each with mass m_i one has relative to cartesian coordinates (x,y,z)

$$T = \frac{1}{2} \sum_{i=1}^{n} m_i \left(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2 \right)$$

and

$$U = U(t, x_1, y_1, z_1, \cdots, x_n, y_n, z_n) \, .$$

The Euler-Lagrange equations are

$$\begin{aligned} -\frac{\partial U}{\partial x_i} &= m_i \ddot{x}_i \\ -\frac{\partial U}{\partial y_i} &= m_i \ddot{y}_i \\ -\frac{\partial U}{\partial z_i} &= m_i \ddot{z}_i \qquad i = 1, 2, 3 \end{aligned}$$

Thus one has $3n$ differential equations for $3n$ unknown functions. These equations are Newton's equations of motion for a system of n particles because

$$\left(-\frac{\partial U}{\partial \dot{x}_i}, -\frac{\partial U}{\partial \dot{y}_i}, -\frac{\partial U}{\partial \dot{z}_i}\right) = \overrightarrow{Force} \text{ on the ith particle .}$$

The advantage of the variational formulation of mechanics via Hamilton's principle over Newton's formulation is twofold:

1. From the viewpoint of physics it opens up new vistas. In particular, Hamilton's principle is a *bridge to quantum mechanics*. In fact, it is quantum mechanics which sanctions and justifies Hamilton's principle of least action. We shall return to that point later.

2. From the viewpoint of mathematics this formulation of mechanics constitutes an enormous streamlining, both in setting up and in solving the equations of motion for complex Newtonian dynamical systems.

   (a) Setting up the equations has been reduced to constructing a single scalar function, the Lagrangian, and its derivatives.

   (b) Solving the resulting system of differential equations is greatly facilitated by the flexibility inherent in a scalar function. The function, and hence the system of equations, can be exhibited relative to any system of coordinates one chooses, including those relative to which the equations are so simple that they can be readily analysed, if not solved.

The conceptual unit economy, and hence the technical advantage and the mathematical power of Hamilton's principle, arises from the fact that the Lagrangian is a single scalar

$$L = K.E. - P.E.\,,$$

and from the fact that it may be expressed in terms of virtually any type of coordinates. These "generalized coordinates"

$$q^1(t), q^2(t), \cdots, q^s(t)$$

and their time derivatives

$$\dot{q}^1(t), \dot{q}^2(t), \cdots, \dot{q}^s(t)$$

characterize the state of motion of the mechanical system. Quite often they are (but need not be) related to, say, the cartesian coordinates, $\vec{x} = (x, y, z)$ by

$$\vec{x}_i = \vec{x}_i(q^1, \cdots, q^s, t) = \vec{x}_i(q^j, t)$$

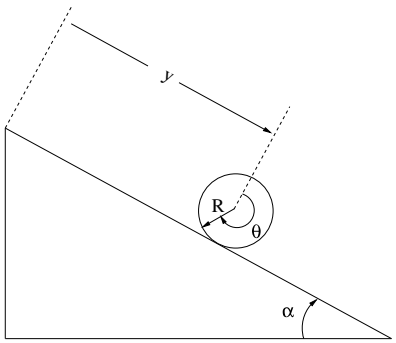


Figure 3.7: Disk rolling down an inclined plane

and their time derivative

$$\dot{\vec{x}}_i = \dot{\vec{x}}_i(\dot{q}^j, q^k, t)\,.$$

In any case, in terms of generalized coordinates Hamilton's principle becomes

$$\delta \int_{t1}^{t_2} L(q^j, \dot{q}^j, t) = 0\,.$$

The associated Euler-Lagrange equations of motion are

$$\frac{\partial L}{\partial q^j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j}\,. \tag{3.2}$$

The ease with which the equations of motion can be set up is illustrated by the following simple

Example. Consider a disk rolling down a plane of lenth ℓ inclined at an angle α relative to the horizontal . The task of setting up the equations for the system is archetypical:

1. The total kinetic energy breaks up into its translational and its rotational part (around the center of mass)

$$\begin{aligned} T &= \frac{1}{2}M\dot{y}^2 + \frac{1}{2}I\dot{\theta}^2 \\ &= \frac{1}{2}M\dot{y}^2 + \frac{1}{4}MR^2\dot{\theta}^2 \end{aligned}$$

   Here $I = \frac{1}{2}MR^2$ is the moment of inertia of the disc whose radius is R and whose mass is M.

2. The potential energy is

$$U = Mg\,(\ell - y)\,\sin(\alpha)$$

so that U=0 at the bottom ($y = \ell$) of the inclined plane.

3. The Lagrangian is

$$L = T - U = \frac{1}{2}M\dot{y}^2 + \frac{1}{4}MR^2\dot{\theta}^2 + Mg(y-\ell)\sin(\alpha)\,.$$

4. The equation of constraint between tranlation and rotation is

$$G(y,\theta) = y - R\theta = 0\,.$$

5. Finally, the Euler-Lagrange equations are

$$\begin{aligned}\frac{d}{dt}\frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} &= \lambda(t)\frac{\partial G}{\partial y}\\ \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} &= \lambda(t)\frac{\partial G}{\partial \theta}\end{aligned}\quad .$$

Explicitly, they are

$$M\ddot{y} - Mg\,\sin\alpha = \lambda \tag{3.3}$$

$$\frac{1}{2}MR^2\ddot{\theta} = -\lambda R\,. \tag{3.4}$$

The constraint is

$$y = R\theta\,. \tag{3.5}$$

By differentiating the last equation, using the result to rewrite the antecedent equation, and introducing the resultant expression for λ, these three equations yield

$$\begin{aligned}\ddot{\theta} &= \frac{\ddot{y}}{R}\\ \lambda &= -\frac{1}{2}M\ddot{y}\\ \ddot{y} &= \frac{2g\,\sin\alpha}{3}\,.\end{aligned}$$

Consequently,

$$\begin{aligned}\lambda &= \frac{-Mg\,\sin\alpha}{3}\\ \ddot{\theta} &= \frac{2g\,\sin\alpha}{3R}\,.\end{aligned}$$

Remark. The introduction of the Lagrange multiplier λ into the variational principle reveals physical aspects of the mechanical system which would have stayed hidden without that multiplier. In this example the right hand side of Eq.(3.3),

$$\lambda = \frac{-Mg\,\sin\alpha}{3},$$

is a *constraint force*. On the other-hand, the right hand side of Eq.(3.4),

$$-\lambda R = \frac{Mg\,\sin\alpha}{3} R,$$

is a constraint torque. These constraint-induced quantities (i.e. λ and -λR) are non-zero in order to guarantee that the constraint, namely, *no slippage*, stays enforced. In particular, note that λ is a constraint force along the y-direction, and that λR is a positive torque. This guarantees that the disk spins up in the positive θ direction at the expense of the disk not gaining velocity as rapidly as it would without the constraint.

The fundamental challenge in one's acquisition of knowledge is to be able to take a concept, no matter how abstract, and be able to trace it back to its origin in reality, i.e. to *concretize* it.

In the context of mathematizing a system subject to constraints, mathematizing their existence in terms of corresponding Lagrange multipliers gives one's knowledge that extra vividness and extra relevance, which would have been lost had one merely used the constraints to decrease the complexity of the system.

3.2 Hamilton-Jacobi Theory

Lecture 16

Newton created mechanics and formulated it in terms of his differential laws of motion. Using Hamilton's Principle, Lagrange streamlined Newton's formulations by giving it an innovative mathematical formulation which is based on a single scalar, the Lagrangian. As a consequence, there is an infinite flexibility in describing the mechanical system; in other words, the equations of motion can be exhibitied relative to any chosen set of generalized coordinates.

Hamilton-Jacobi theory, which also springs forth from Hamilton's principle, constitutes an additional step forward in physics as well as in mathematics. The advance in mathematics consists of the fact that H-J theory constitutes a powerful and systematic way of integrating the equations of motion of mechanics. The advance in physics consists of the fact that H-J theory provides a bridge between classical mechanics and quantum mechanics, as well as between ray optics and electromagnetic wave theory (wave optics).

Pedigree of classical mechanics and the times its authors flourished			
	Newton	1672-1727	17^{th} century
Systematic method of setting up Newton's equations of motion	↓ Euler Lagrange	1707-1783 1736-1813	18^{th} century
Streamline their mathematical formulation	↓ Hamilton	1805-1865	19^{th} century
Systematize the process of solving Hamilton's equations	↓ Jacobi	1804-1851	19^{th} century
Global and geometrical formulation ("KAM theory")	↓ Kolmogorov Arnold Moser	1903-1987 1937-2010 1928-1999	20^{th} century

It is difficult to point to a mathematical theory which brings into our grasp a larger and more diverse number of aspects of the physical world than H-J theory. No wonder many physicists express their admiration and enthusaism for H-J theory in a way which borders on the poetic. For example, Cornelius Lanczos in his book "The Variational Principles of Mechanics" (1949), writes: "We now are in the rarified atmosphere of theories of excessive beauty and we are nearing a high plateau on which geometry, optics, mechanics, and wave mechanics meet on common ground. Only concentrated thinking, and a considerable amount of re-creation, will reveal the full beauty of our subject in which the last words have not yet been spoken. We start with the integration theory of Jacobi and continue with Hamilton's own investigations in the realm of goemetrical optics and mechanics. The combination of these two approaches leads to de Broglie's and Schroedinger's great discoveries, and we come to the end our journey."

3.3 The Dynamical Phase

The equations of motion of a dynamical system are obtained from Hamilton's principle, namely the require ment that the action functional

$$I = \int_{x',t'}^{x,t} L\left(x, \frac{dx}{dt}, t\right) dt \tag{3.6}$$

be an extremum for the actual path along which the system travels in space and in time. This actual path we call an *extremal* path. Suppose we consider only extremal

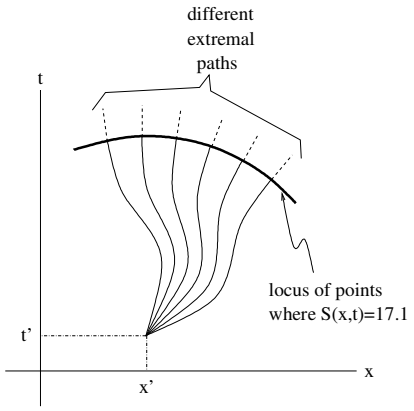


Figure 3.8: Isogram of the dynamial phase S(x,t)

paths, in particular those that have the same common starting point (x', t') in space and time. The termination point of each extemal path is different, and we designate it by the variable point (x, t). For of these extremal path the action integral has a (unique) extemal value, namely

$$\int_{x',t'}^{x,t} L\left(x, \frac{dx}{dt}, t\right) dt = I_{extremum}\,.$$

For each endpoint (x, t) of an extemal path there is number $I_{extremum}(x, t)$. Thus we have a real valued function which maps (x, t) into the reals:

$$S(x, t) \equiv I_{extemum}(x, t)\,.$$

The level surfaces (isograms) of this scalar function are easy to construct. Suppose we wish to construct the isogram

$$S(x, t) = 17.1\,.$$

We consider the set of extemal paths emanating from the common point (x', t'). At that point $S(x', t') = 0$. We now move along a particular extemal path and evaluate the action integral, Eq.(3.6), along this path. When the value of this integral is $I = 17.1$, we stop. This stopping point is a point on the isogram $S(x, t) = 17.1$.

The function defined by this construction is called the "dynamical phase", or the "action" of the system at (x, t). It also is called the "eikonal" in optics. This function is easy to remember because in wave mechanics it is the phase of the system. In fact, we assert that its relation to the wave function of the system is give by

$$\Psi(x, t) = \mathcal{A}e^{iS(x,t)/\hbar}.$$

Here $\hbar$ is a constant which guarantees that the exponent is dimensionless. For a mechanical system this constant is Planck's constant (6.1×10^{-27}erg sec) divided by 2π. The wave function Ψ satisfies a wave equation. Whether this wave equation is the Schroedinger equation of the system, the scalar wave equation for a light ray, or some other linear wave equation is of great importance, but of no concern to us at present. For us, our present interest lies only in the dynamical phase $S(x,t)$,

1. in the equation for $S(x,t)$ (Section 3.5) and
2. in the reconstruction of the extremal paths (Sections 3.6 and 3.12) from $S(x,t)$.

The physical perspective on our reconstruction of the extremal paths is very different from that of Newton. Whereas his focus is on the particle aspects of bodies, ours is on their wave aspects. In spite of that difference our wave perspective started with Newton. It grew by a well-defined process of inductive reasoning consisting of Newton →Euler-Lagrange →Hamilton →Jacobi + experimental observations about the dynamical behavior of matter – observations which were performed during the first half of the twentieth century. In spite of the great conceptual distance between the particle and the wave perspectives, the former is a correspondence (asymptotic, shortwave length) limit of the latter. This fact is developed in Sections 3.6-3.12 and summarized in Figure 3.14 and 3.15 on pages 129 and 133 respectively.

3.4 Momentum and the Hamiltonian

We now proceed with the process of deriving the partial differential equation which the scalar function $S(x,t)$ satisfies. In this process we shall repeatedly encounter two quantities. They are (i) the *generalized momentum* and (ii) the *Hamiltonian* of the system. Specifically we let

$$\frac{\partial L}{\partial \dot{q}^i} \equiv p_i \qquad i = 1, ...s \tag{3.7}$$

be the ith generalized momentum. Its importance follows from the equation of motion. If the Lagrangian is independent of the ith generalized coordinate q^i, i.e.

$$\frac{\partial L}{\partial q^i} = 0$$

then the Euler-Lagrange equations of motion demand that

$$\frac{dp_i}{dt} = 0\,.$$

In other words, the ith generalized momentum of the system does not change; this *momentum is "conserved".*

Example. Consider a system of n particles in a potential. The Lagrangian is

$$L = \frac{1}{2}\sum_{j=1}^{n} m_j \left(\dot{x}_j^2 + \dot{y}_j^2 + \dot{z}_j^2\right) - U\left(x_j, y_j, z_j, t\right) .$$

The x-component of the jth particle *momentum* is

$$\frac{\partial L}{\partial \dot{x}_j} = m\dot{x}_j \equiv p_{xj} .$$

The other repeatedly occuring quantity is

$$\sum_{i=1}^{s} \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} - L \equiv H , \tag{3.8}$$

the *Hamiltonian* of the system. This quantity is conserved if the Lagrangian of the system has no explicit time dependence. Indeed, differentiating each term of that expression for the Hamiltonian and keeping in mind that L depends on q^i, $\dot{q}^i$, and t, one obtains

$$\frac{dH}{dt} = \sum_{i=1}^{s}\left[\ddot{q}^i\frac{\partial L}{\partial \dot{q}^i} + \dot{q}^i\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i\right] - \frac{\partial L}{\partial t} .$$

Here the two middle terms cancel because of the E-L equations of motion. Consequently, we are left with

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} . \tag{3.9}$$

We conclude that if $\frac{\partial L}{\partial t} = 0$, then

$$H = constant$$

along the extremal path of the system.

Remark. If the system has only a single degree of freedom ($s = 1$) then $H = \dot{q}\frac{\partial L}{\partial \dot{q}} - L$, and Eq.(3.9) is simply the "2nd form of the Euler equation", which was considered at the end of Chapter I.

The generalized momenta p_j , $j = 1, \cdots, s$ as well as the Hamiltonian H will play the key roles in establishing the partial diffential equation which the dynamical phase $S(q^1, q^2, \cdots, q^s, t)$ must satisfy. In fact, in order to obtain this equation, we shall find that one must reexpress the Hamiltonian, Eq.(3.8), in terms of the generalized momenta defined by Eq.(3.7) in such a way that the Hamiltonian has no functional dependence on the generalized velocities $\dot{q}^j$, j= *1,*$\cdots$*,s*. This goal is achieved by solving the defining equations (3.7) for the generalized velocities

$$\dot{q}^i = \dot{q}^i(q^k, p_j)$$

in terms of the momenta, and then introducing these expressions into the Hamiltonian:

$$\sum_{i=1}^{s} \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} - L = H(q^k, p_j, t)\,. \tag{3.10}$$

To verify that H does indeed depend only on q^k and p_j, but not on $\dot{q}^i$, we observe that[1]

$$\begin{aligned} dH &= d\left(\dot{q}^i p_i - L\right) \\ &= d\dot{q}^i p_i + \dot{q}^i dp_i - \frac{\partial L}{\partial q^i} dq^i - \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i - \frac{\partial L}{\partial t} dt\,. \end{aligned}$$

The first and the fourth term cancel with the help of Eq.(3.7). What remains is

$$dH = \dot{q}^i dp_i - \frac{\partial L}{\partial q^i} dq^i - \frac{\partial L}{\partial t} dt\,.$$

There are two benefits. First, this shows that, with the help of the defining relation (3.7), the Hamiltonian is functionally independent of $\dot{q}^j$, and depends only on q^k and p_j instead. Second, by comparing this expression with the differential of Eq.(3.10),

$$dH = \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial t} dt\,,$$

by using the E-L equations of motion (3.2),

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt} p_i\,,$$

and by taking advantage of the linear independence of the differentials $\{dp_i, dq^i, dt\}$, one obtains the three equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}, \quad \text{and} \quad \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}\,. \tag{3.11}$$

These are *Hamilton's equations of motion.* They govern the dynamics of the mechanical system. Its motion is represented by the trajectory of a moving point $\{q^i(t), p_i(t)\}_1^s$ in the $2s$-dimensional space spanned by $\{q^i, p_i\}$. The rate of change of

[1] In the interest of conceptual unit economy we are introducing the Einstein summation convention according to which a summation is implied whenever a pair of repeated (summation!) indices occur in a product expression. For example, $\sum_{i=1}^{s} \dot{q}^i p_i \equiv \dot{q}^i p_i$.

the Hamiltonian function $H(q^i, p_i, t)$ along the trajectory is

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial H}{\partial q^i}\dot{q}^i + \frac{\partial H}{\partial p_i}\dot{p}_i + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q^i} + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial t}\,.\end{aligned}$$

Thus, if t is a cyclic coordinate, i.e. the Hamiltonian is manifestly independent of time ($\partial H/\partial t = 0$), then H is evidently a constant of motion.

3.4.1 The Legendre Transformation

In Lagrangian mechanics the evolution of a system proceeds in the form its trajectory, a moving point in the domain spanned by $\{q^i, \dot{q}^j\}$. In Hamiltonian mechanics that evolution proceeds along its trajectory in the dual domain[2], which is spanned by $\{q^i, p_j\}$. These two domains are mathematically equivalent because their respective mathematizations are related by means of *Legendre's transformation*,

$$\{q^i, \dot{q}^j\} \leftrightarrow \{q^i, p_j\},$$

which is one-to-one and onto. To concretize this feature,consider the graph of the convex function

$$L(v) = \frac{1}{2}mv^2 - U(x)$$

and its tangent $\dfrac{\partial L}{\partial v} \equiv p$ as well as the graph of line parallel through the origin as shown in Figure 3.9. For fixed x the relation between the v coordinate and the tilt of the tangent,

$$\frac{\partial L(v)}{\partial v} \equiv p,$$

to the curve $L(v)$ is one-to-one. Thus one can always find the inverse relation

$$p \rightsquigarrow v = v(p)$$

[2]This duality is in the second argument:

$$\{\dot{q}^j\} \leftrightarrow \{p_j\}.$$

Here $\{\dot{q}^j\}$ are the components of a vector in the vectorspace $V_{\{q^i\}}$ attached to the base point $\{q^i\}$, while $\{p_j\}$ are the components of a covector in the dual space $V^*_{\{q^i\}}$, i.e. the space of linear functions on $V_{\{\dot{q}^i\}}$. Thus $\{\dot{q}^j\}$ refers to an element in the tangent space at $\{q^j\}$, while $\{p_j\}$ refers to an element in the cotangent space at $\{q^j\}$.

Introducing it into the difference $pv - L(v)$, one obtains its unique value, the value of the transformed function $H(p)$

$$pv(p) - L(v(p)) \equiv H(p)$$

at p. The convexity of $L(v)$ guarantees that this can be done for every v. Consequently, as constructed in Figure 3.9, for every function $L(v)$ on $\{v\}$ there exists a uniquely defined function $H(p)$ on $\{p\}$:

$$\left.\begin{array}{l} L : \text{`}\{v\} \to R \\ \qquad v \rightsquigarrow L(v) \end{array}\right\} \leftrightarrow \left\{\begin{array}{l} H : \ \{p\} \to R \\ \qquad\quad p \rightsquigarrow H(p). \end{array}\right.$$

Here

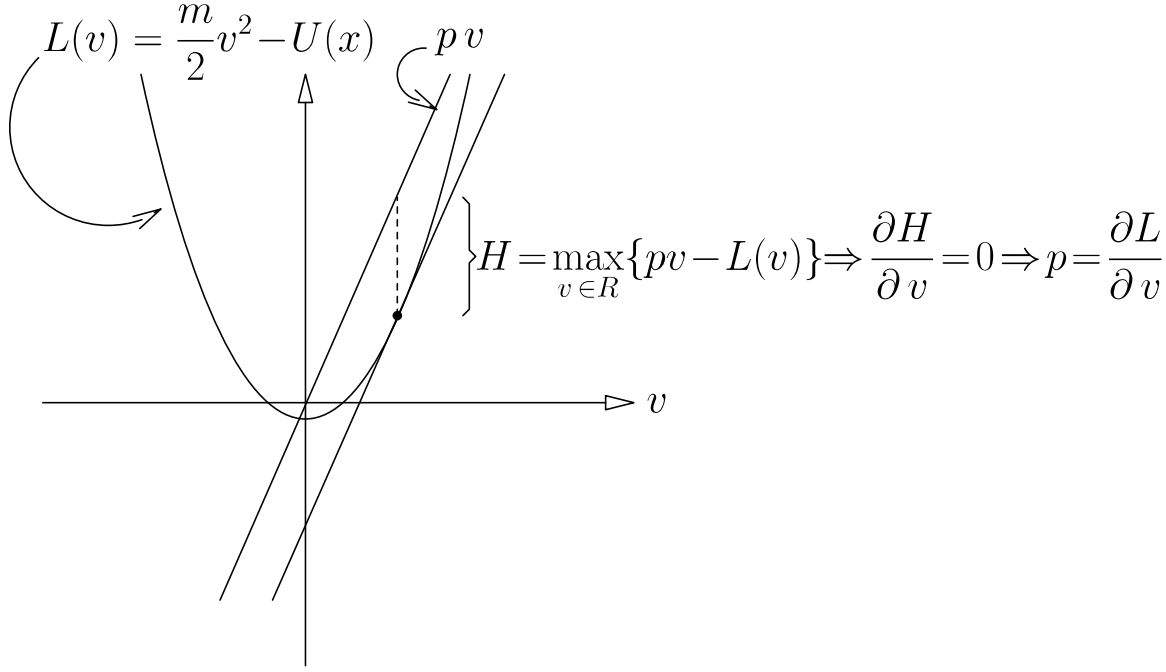


Figure 3.9: The Legendre transformation process is a constructive process which maximizes the difference between (i) the linear function whose isogram through nthe origin is parallel to the tangent to the given function at a point and (ii) the value of that function at that point of tangency.

The Hamiltonian equations of motion are equivalent to those of Euler-Lagrange. The important mathematical difference is that the former is a system of *2s first* order equations, while the latter is one of *s second* order equations.

Example. (*Particles in the Field of a Potential)*
The Lagrangian is

$$L = \sum_{i=1}^{n} \frac{1}{2} m_i \vec{\dot{x}}_i \vec{\dot{x}}_i - U(\vec{x}_k, t) \, .$$

For each of the particles the momentum is

$$\vec{p}_i = (p_{xi}, \, p_{yi}, \, p_{zi}) = m_i \vec{\dot{x}}_i \, . \tag{3.12}$$

The Hamiltonian is

$$\begin{aligned} H &= \sum_j \dot{q}^j \frac{\partial L}{\partial \dot{q}_j} - L \\ &= \sum_{i=1}^{n} (\dot{x}_i m_i \dot{x}_i + \dot{y}_i m_i \dot{y}_i + \dot{z}_i m_i \dot{z}_i) - \frac{1}{2} \sum_{i=1}^{n} (\dot{x}_i m_i \dot{x}_i + \dot{y}_i m_i \dot{y}_i + \dot{z}_i m_i \dot{z}_i) + U(\vec{x}_k, t) \\ &= \frac{1}{2} \sum_{i=1}^{n} m_i \vec{\dot{x}}_i \cdot \vec{\dot{x}}_i + U(\vec{x}_k, t) \\ &= K.E. + P.E. = \text{Total Energy} \, . \end{aligned}$$

With the momentum defined by Eq. (3.12), the Hamilitonian assumes the form

$$H(\vec{p}_i, \vec{x}_i, t) = \sum_{i=1}^{n} \frac{\vec{p}_i \cdot \vec{p}_i}{2m_i} + U(\vec{x}_k, t) \, ,$$

and the two sets of Hamiltonian equations of motion are

$$\begin{aligned} \frac{d\vec{x}_i}{dt} &= \frac{\vec{p}_i}{m} \\ \frac{d\vec{p}_i}{dt} &= -\vec{\nabla}_i U(\vec{x}_k, t) \, . \end{aligned}$$

If the potential energy is time-independent and hence depends only on the position $\vec{x}_i$ of the particles, then H is cyclic in t:

$$0 = \frac{\partial H}{\partial t} \, .$$

Consequently,

$$H = T.E. = \text{constant} \, .$$

Thus the total energy of a system of particles moving in a time independent potential is a constant of motion, a conclusion one also infer directly from the fact that the Lagrangian does not contain any explicit dependence on time t.

Remark. That the Hamiltonian is the total energy of the system is true in a non-rotating coordinate frame. However, in a rotating coordinate frame the Hamiltonian is not equal to the system's total energy. However, if the Lagrangian has no explicit time dependence ("time invariant system") then the Hamiltonian is conserved nevertheless. To illustrate the issue, consider the above example of particles-in-a-potential problem from the perspective of a rotating frame, relative to which one achieves substantial simplifcation in the analysis of charged particles circulating in a magnetic field or of satellites in an earth-centered reference frame.

Problem. (*Particle motion in a rotating frame)*
Given a frame (x', y', z') which rotates with constant angular frequency ω relative to a given inertial frame (x,y,z),

$$\begin{aligned} x &= x' \cos \omega t - y' \sin \omega t \\ y &= x' \sin \omega t + y' \sin \omega t \\ z &= z' \end{aligned}$$

determine for a single particle in a rotationally symmetric potential $U\left(x^2+y^2,\, z\right)$ (i) its Lagrangian, (ii) its Hamiltonian relative the rotating frame, and then (iii) compare the two Hamiltonians in the two frames.
Solution. The Lagrangian of a particle moving in the potential U is

$$L = \frac{m}{2}\left(\dot{x}^2+\dot{y}^2+\dot{z}^2\right) - U\left(x^2+y^2,\, z\right) .$$

Relative to the rotating frame this Lagrangian,

$$L = \frac{m}{2}\left(\dot{x}'^2+\dot{y}'^2+\dot{z}'^2\right) - U\left(x'^2+y'^2,\, z'\right) + \frac{m}{2}\omega^2\left(x'^2+y'^2\right) + \omega\, m\left(x'\dot{y}' - y'\dot{x}'\right),$$

has acquired an additional centrifugal potential energy, and a velocity dependent potential energy. In the rotating frame the Hamiltonian is

$$\begin{aligned} H' &= \dot{x}'\frac{\partial L}{\partial \dot{x}'} + \dot{y}'\frac{\partial L}{\partial \dot{y}'} + \dot{z}'\frac{\partial L}{\partial \dot{z}'} - L \\ &= \frac{m}{2}\left(\dot{x}'^2+\dot{y}'^2+\dot{z}'^2\right) + U\left(x'^2+y'^2,\, z'\right) - \frac{m}{2}\omega^2\left(x'^2+y'^2\right) . \end{aligned}$$

Introducing the defining relations, Eq.(3.7) on page 112

$$\begin{aligned} p'_x &= m\dot{x}' - m\omega y' \quad &\rightarrow \dot{x}' = \frac{p'_x}{m} + \omega y' \\ p'_y &= m\dot{y}' + m\omega x' \quad &\rightarrow \dot{y}' = \frac{p'_y}{m} - \omega x' \\ p'_z &= m\dot{z}' \quad &\rightarrow \dot{z}' = \frac{p'_z}{m} , \end{aligned}$$

one finds that this Hamiltonian is

$$H' = \frac{(p'_x)^2}{2m} + \frac{(p'_y)^2}{2m} + \frac{(p'_z)^2}{2m} + U\left(x'^2 + y'^2,\, z'\right) - \omega\left(x'p'_y - y'p'_x\right).$$

The particle's inertial Hamiltonian

$$\begin{aligned} H &= \frac{(p_x)^2}{2m} + \frac{(p_y)^2}{2m} + \frac{(p_z)^2}{2m} + U\left(x^2 + y^2,\, z\right) \\ &= \text{Kinetic Energy} + \text{Potential Energy} \end{aligned}$$

expresses its total energy. It is an integral of motion. By contrast the particle's rotational Hamiltonian,

$$H' = (\text{Kinetic Energy})' + (\text{Potential Energy})' - \omega L_z\,,$$

even though it is also an integral of motion, does not express a mere total energy, kinetic plus potential in the rotating frame. There is an additional energy-like quantity $-\omega L_z$ which expresses a kind of interaction between (a) whatever aspect (proportional to ω) is responsible for the curved meniscus in Newton's bucket experiment[3] and (b) the particle's moment of momentum (i.e. angular momentum),

$$\begin{aligned} L_z &= \left(x'p'_y - y'p'_x\right) \\ &= \left(x\,p_y - y\,p_x\right) \end{aligned}$$

It is evident that this quantity is the same in the rotating and the inertial frame. It therefore refers to a property independent of any rotational aspect of the frame. Furthermore, it is an integral of motion, a fact which is implied by Hamilton's equations of motion. Thus there are three integrals of motion: (i) H', (ii) (K.E.)′+(P.E.)′, (iii) L_z and any two imply the third.

3.5 The Hamilton-Jacobi Equation

We now return to the task of determining the equation that governs the dynamical phase for a system with a single degree of freedom. The extension to several degrees takes little effort by comparison.

[3]Newton considered a water-filled bucket rotating around its axis, as in Figure 2.19 on page 89 – rotating relative to a free float (i.e. inertial) frame – a frame relative to which all free bodies obey Newton's first law of motion, a frame which today we recognize as having the additional feature of *not* rotating relative to the fixed stars. The rotating frame associated with this bucket is characterized by a single distinguishing feature, its angular velocity ω. In this frame the surface of unmoving water curves upward with increasing radius. This curvature is absent in the inertial frame, which is not rotating relative to the fixed stars.

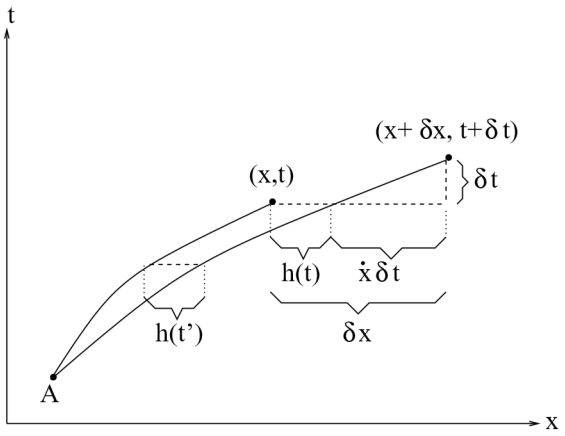


Figure 3.10: Two nearby extremal paths.

3.5.1 Single Degree of Freedom

Consider a set of extremal paths emanating from event A, i.e. they all start at the same time t' at the same point x' in space. These paths being extremal, they all satisfy the same Euler-Lagrange equation. They also pass through the same starting event, but they have different initial velocities. We evaluate the "action" along each extremal path,

$$I_{extremum}(t,t') = \int_{t'}^{t} L\left(x(\lambda), \dot{x}(\lambda), \lambda\right) d\lambda \,.$$

Consider the locus of points where, on different extremals, the integral $I_{extreme}$ has the same value, say

$$I_{extremum} = I_{extremum.}(t,t') \,.$$

This locus of points forms the level surfaces of a function *S(x,t)* where

$$S(x,t) = \text{extremum value of} \int_{x',t'}^{x,t} L(x,\dot{x},t)dt \,.$$

We shall now see that the partial derivatives of this scalar function are the momenta and the negative Hamiltonian (usually, but not always, the energy), i.e.

$$\delta S = p\delta x - H\delta t$$

We consider two *extemal* paths. They have the same starting point A. But their end points are respectively (x,t) and (x+δx,t+δt). The difference in the dynamical phase

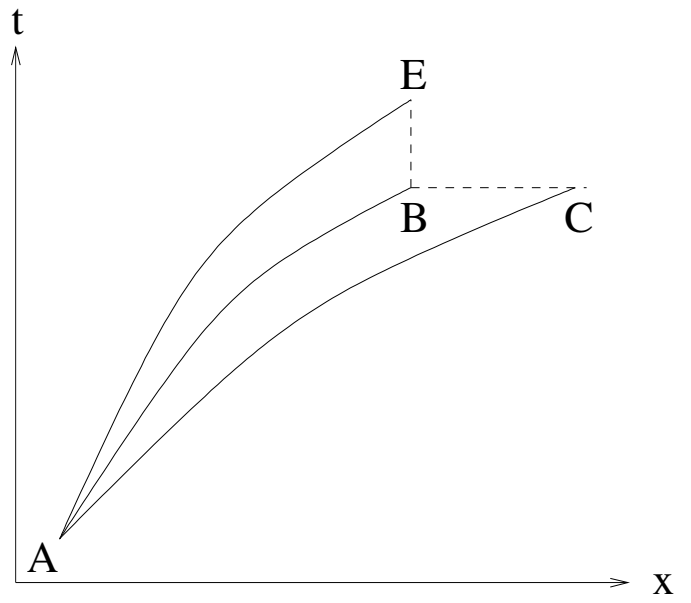


Figure 3.11: Three extremal curves

(action) of these endpoints is

$$\begin{aligned}
\delta S &= \delta I_{extremum} \\
&= \int_A^{t+\delta t} L(x+h, \dot{x}+\dot{h}, t)dt - \int_A^t L(x, \dot{x}, t)dt \\
&= L\delta t + \int_A^t \left(\frac{\partial L}{\partial x} h + \frac{\partial L}{\partial \dot{x}} \dot{h} \right) dt \\
&= L\delta t + \frac{\partial L}{\partial \dot{x}} h + \int_A^t \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) h \, dt \, .
\end{aligned}$$

We only consider extremal curves. Thus $\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0$, and the integral vanishes. The variations δx and δt refer to the coordinate differences of arbitrarily located endpoint of the two extremals. But the variation $h(t)$ connects pairs of points having the same curve parameter t. Consequently, as one can see in Fig. 3.10,

$$h(t) = \delta x - \dot{x}\delta t \, .$$

Thus

$$\begin{aligned}
\delta S &= \frac{\partial L}{\partial \dot{x}} \delta x - \left(\frac{\partial L}{\partial \dot{x}} \dot{x} - L \right) \delta t \\
&\equiv p\delta x - H\delta t
\end{aligned}$$

An inquiry as to how S(x,t) changes into the x-direction, keeping t fixed, yields

$$\left(\begin{array}{c}\text{rate of change of}\\ \text{dynamical phase}\\ \text{with position}\end{array}\right) = (\text{momentum}) \tag{3.13}$$

$$= \lim_{\substack{C\to B\\ t=fixed}} \frac{\delta S}{\delta x} = \frac{\partial S}{\partial x} \tag{3.14}$$

$$= \frac{\partial L(x,\dot{x},t)}{\partial \dot{x}} = p\,. \tag{3.15}$$

Similarly,

$$-\left(\begin{array}{c}\text{rate of change of}\\ \text{dynamical phase}\\ \text{with time}\end{array}\right) = (\text{energy}) \tag{3.16}$$

$$= -\lim_{\substack{E\to B\\ x=fixed}} \frac{\delta S}{\delta t} = -\frac{\partial S}{\partial t} \tag{3.17}$$

$$= \frac{\partial L}{\partial \dot{x}}\dot{x} - L = H\,. \tag{3.18}$$

These two equations, yield the "dispersion relation" at (x,t). Indeed, in Eq. (3.18) replace the $\dot{x}$-dependence with a dependence on p by solving Eq. (3.15) for $\dot{x}$. The result is

$$(\text{energy}) = H(x,p,t)\,.$$

Finally introduce into this equation the partial derivatives calculated in Eqs. (3.14) and (3.17):

$$-\frac{\partial S}{\partial t} = H\left(x, \frac{\partial S}{\partial x}, t\right)\,.$$

This is the Hamilton-Jacobi equation, a single first order differential equation for the dynamical phase $S(x,t)$. This equation is a dispersion relation because it expresses how the time rate of change of the phase depends on its rate of change with position. For a system with several degrees of freedom the H-J equation is

$$H\left(q^i, \frac{\partial S}{\partial q^j}, t\right) + \frac{\partial S}{\partial t} = 0.$$

Lecture 17

Problem. (*Particle in a potential*)

Set up and solve the Hamilton-Jacoby equation for a particle in a one dimensional potential $U(x)$.

Solution. Setting up the H-J equation is a three step process.

1. Exhibit the Lagrangian:
$$L = \frac{1}{2}m\dot{x}^2 - U(x)\,.$$

2. Determine the momentum and the Hamiltonian:
$$\begin{aligned} p &= \frac{\partial L}{\partial \dot{x}} \\ &= m\dot{x}\,; \\ H &= \dot{x}\frac{\partial L}{\partial \dot{x}} - L \\ &= \frac{1}{2}m\dot{x}^2 + U(x)\,. \end{aligned}$$

3. Express the Hamiltonian in terms of the momentum:
$$H = \frac{p^2}{2m} + U(x)\,.$$

4. Write down the H-J equation $-\frac{\partial S}{\partial t} = H\left(x, \frac{\partial S}{\partial x}\right)$:
$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + U(x)\,.$$

This a first order non-linear partial differential equation that needs to be solved for the scalar function $S(x,t)$.
This p.d.e. lends itself to being solved by the method of separation of variables according to which one finds solutions of the form
$$S(x,t) = T(t) + X(x)\,. \tag{3.19}$$
Introducing this form into the H-J equation, one finds
$$-\frac{dT(t)}{dt} = \frac{1}{2m}\left(\frac{dX(x)}{dx}\right)^2 + U(x)\,.$$
This equation says that the left hand side is independent of x, while the right hand side is independent of t. Being equal, the l.h.s. is also independent of x. Being independent of both t and x, it is a constant. Letting this "separation" constant be equal to E, one obtains two equations
$$\begin{aligned} -\frac{dT(t)}{dt} &= E \\ \frac{1}{2m}\left(\frac{dX(x)}{dx}\right)^2 + U(x) &= E\,. \end{aligned}$$

These are two ordinary equations for T and X. Inserting these equations into Eq. (3.19), one obtains the sought after solution to the H-J equation,

$$S(x,t) = -Et + \int^x \sqrt{2m\,(E - U(x'))}dx' + \delta(E)\,.$$

Here the "integration constant" $\delta(E)$ is an arbitrary function of E. Furthermore, observe that S depends on E also. This means that one has an E-parametrized family of solutions. Thus, properly speaking, separation of variables yields many solutions to the H-J equation, in fact, a one-parameter family of them

$$S(x,t) = S_E(x,t)\,.$$

3.5.2 Several Degrees of Freedom

We shall see in a subsequent section that whenever the H-J for a system with several degrees of freedom, say $\{q^i\}$, lends itself to being solved by the method of the separation of variables, i.e.

$$S(q^i,t) = T(t) + \sum_{i=1}^{s} Q_i(q^i)\,,$$

the solution has the form

$$S = -\int^t Edt + \sum_{i=1}^{s}\int^{q^i} p_i(x^i; E, \alpha_1, \cdots, \alpha_{s-1})dq^i + \delta\,(E, \alpha_1, \cdots, \alpha_{s-1})$$

Here δ is an arbitrary function of E and the other separation constants that arise in the process of solving the H-J equation. We see that for each choice of $(E, \alpha_1, \cdots, \alpha_{s-1})$ we have a different solution S. Thus, properly speaking, we have $S_{E,\alpha_1,\cdots,\alpha_{s-1}}$, a multi-parametrized family of solutions to the H-J equation.

We shall now continue our development and show that Hamilton-Jacobi Theory is

a) A new and rapid way of integrating the E-L equations

b) The bridge to wave (also "quantum") mechanics.

The virtue of Hamilton's principle is that once the kinetic and potential energy of the system are known, the equations of motion can be set up with little effort. These Euler-Lagrange equations are Newton's equations of motion for the system. Although setting up the equations of motion for a system is a routine process, solving them can be a considerable challenge. This task can be facilitated considerably by using an entirely different approach. Instead of setting up and solving the set of coupled Newtonian ordinary differential equations, one sets up and solves a single partial differential equation for a single scalar function. Once one has this scalar function,

one knows everything there is to know about the dynamical system. In particular, we shall see that by differentiating this scalar function (the dynamical phase, the Hamilton-Jacobi function, the eikonal) one readily deduces all possible dynamical evolutions of the system.

3.6 Hamilton-Jacobi Description of Motion

Hamilton-Jacobi theory is an example of the *principle of unit economy*[4], according to which one condenses a vast amount of knowledge into a smaller and smaller number of principles. Indeed, H-J theory condenses all of classical mechanics and all of wave mechanics (in the asymptotic high-frequency/short-wavelength (a.k.a. W.K.B.) approximation) into two conceptual units,, (i) the H-J equation and (ii) the principle of constructive interference. These two units are a mathematical expression of the fact that classical mechanics is an asymptotic limit of wave mechanics.

Hamilton thinking started with his observations of numerous known analogies between "particle world lines" of mechanics and "light rays" of geometric optics. These observations were the driving force of his theory. With it he developed classical mechanics as an asymptotic limit in the same way that ray optics is the asymptotic limit of wave optics. Ray optics is a mathematically precise asymptotic limit of wave optics. Hamilton applied this mathematical formulation to classical mechanics. He obtained what nowadays is called the Hamilton-Jacobi formulation of mechanics. Even though H-J theory is a mathematical limit of wave mechanics, in Hamilton's time there was no logical justification for attributing any wave properties to material particles. (That justification did not come until experimental evidence to that effect was received in the beginning of the 20th century.) The most he was able to claim was that H-J theory is a mathematical method with more unit economy than any other formulation of mechanics. The justification for associating a wave function with a mechanical system did not come until observational evidence to that effect was received in the beginning of the 20th century.

We shall take advantage of this observation (in particular by Davidson and Germer, 1925) implied association by assigning to a mechanical system a wave function. For our development of the H-J theory it is irrelevant whether it satisfies the

[4]The *principle of unit economy*, also known informally as the "crow epistemology", is the principle that stipulates the formation of a new concept

1. when the description of a set of elements of knowledge becomes too complex,
2. when the elements comprising the knowledge are used repeatedly, or
3. when the elements of that set require further study.

Pushing back the frontier of knowledge and successful navigation of the world demands the formation of a new concept under any one of these three circumstances.

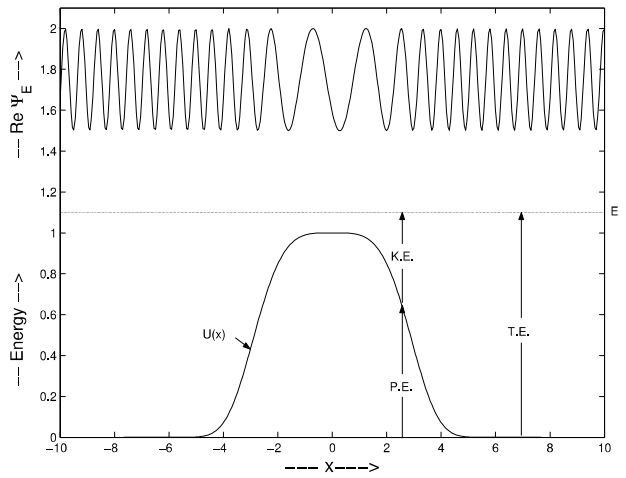


Figure 3.12: The spatial oscillation rate of the wave function $Re\,\Psi_E$ at $t = const.$ is proportional to its x-momentum, whose square is proportional to the kinetic energy (K.E.=T.E.-P.E.).

Schroedinger, the Klein-Gordon, or some other quantum mechanical wave equation. Furthermore, whatever the form of the wave equation governing this wave function, our focus is only on those circumstances where the wave function has the form

$$\Psi_E(x,t) = \underbrace{\mathcal{A}(x,t)}_{\textit{slowly}\text{ varying function of } x \text{ and } t} \times \underbrace{\exp\left(\frac{i}{\hbar}S_E(x,t)\right)}_{\textit{rapidly}\text{ varying function of } x \text{ and } t} \tag{3.20}$$

This circumstance is called the "high frequency" limit or the "semi-classical" approximation. It can be achieved by making the energy E of the system large enough. In that case

$$1 << \frac{S_E(x,t)}{\hbar}$$

with the consequence that the phase factor oscillates as a function of x and t rapidly indeed. The existence of such a wave function raises a non-trivial problem:
If the wave and its dynamical phase, and hence the wave intensity, is defined over all of space-time, how is it possible that a particle traces out a sharp and well defined path in space-time when we are left with three delemas?

1. The large magnitude ($S \gg \hbar$=1.05×10^{-27}[erg sec]) of the action for a classical particle is certainly of no help.

2. Neither is the simplicity of the H-J equation

$$\frac{\partial S}{\partial t} + H(x, \frac{\partial S}{\partial x}, t) = 0$$

which governs the dynamical phase in

$$\Psi = \mathcal{A} \exp\left(i \frac{S}{\hbar}\right) ,$$

3. Nor is the simplicity of the solution S for a particle of energy E,

$$S(x,t) = -Et + \int_{x_o}^{x} \sqrt{2m(E - U(x))} dx + \delta(E)$$

of any help in identifying a localized trajectory ("world line") of the particle in space-time coordinatized by x and t.

What *is* of help is the basic implication of associating a wave function with a moving particle, namely, it is a linear superposition of monochromatic waves, Eq. (3.20), which gives rise to a travelling wave packet – a localized moving wave packet whose history is the particle's world line. To validate this claim we shall give two heuristic arguements (i-ii), one application (iii), a more precise argument (iv) and an observation (v).

(i) The most elementary superposition monochomatic waves is given by the sum wave trains with different wavelengths

$$\Psi(x,t) = \Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots .$$

(ii) In space-time one has the following system of level surfaces for $S_E(x,t)$ and $S_{E+\Delta E}(x,t)$

Destructive interference between different waves comprising Ψ(x,t) occures everywhere except where the phase of the waves agree:

$$S_E(x,t) = S_{E+\Delta E}(x,t)$$

At the locus of events satisfying this condition, the waves interfere constructively and wave packet has non-zero amplitude. The wave mechanical priciple says that this condition of constructive interference

$$0 = \lim_{\Delta E \to 0} \frac{S_{E+\Delta E}(x,t) - S_E(x,t)}{\Delta E} = \frac{\partial S_E(x,t)}{\partial E}$$

yields a Newtonian worldline, i.e. an extremal paths.

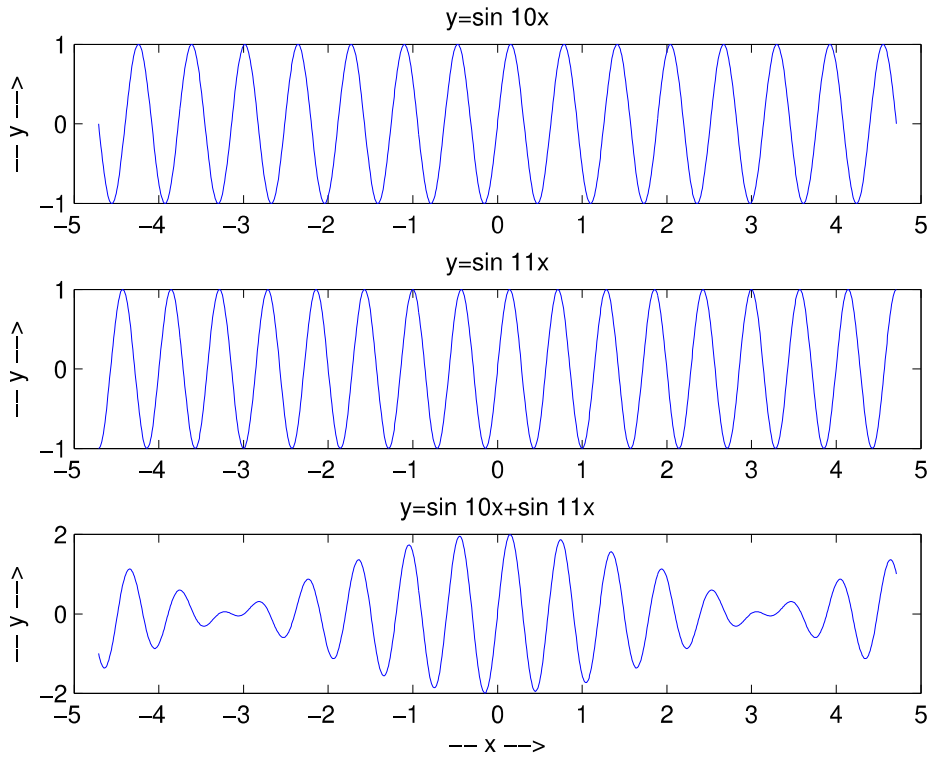


Figure 3.13: Photographic snapshot in space of two interfering wave trains and their resulting wave packet.

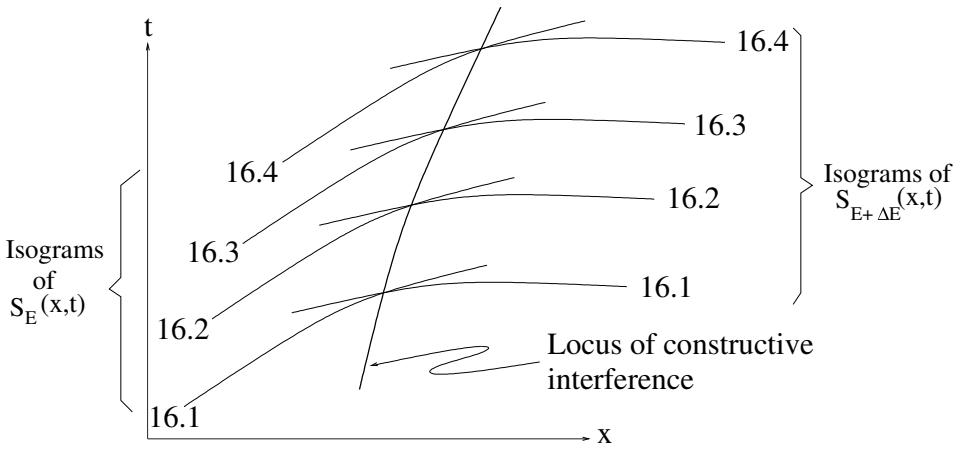


Figure 3.14: Constructive interference represented in space-time. The intersection of the respective isograms of $S_E(x,t)$ and $S_{E+\Delta E}(x,t)$ locates the events (x,t) which make up the trajectory of the particle in x-t space – the locus of constructive interference.

(iii) Apply this condition to the action S(x,t) of a single particle. One obtains the time the particle requires to travel to point x,

$$0 = -t + \int_{x_0}^{x} \sqrt{\frac{m}{2}} \left(\frac{1}{E - U(x)} \right)^{\frac{1}{2}} dx + t_0$$

with

$$t_0 \equiv \frac{\partial \delta(E)}{\partial E} \, .$$

This condition yields the Newtonian worldline indeed. The precise argument is on page 139 in Section 3.12

Lecture 18

3.7 Constructive Interference

Our formulation of constructive interference is based on a picture in which at each time t a superposition of wave trains

$$\Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots \equiv \Psi(x,t)$$

yields a wave packet at time t. The principle of constructive interference itself,

$$\frac{\partial S_E(x,t)}{\partial E} = 0$$

is a condition which at each time t locates where the maximum amplitude of the wave packet is.
It is possible to bring into much sharper focus the picture of superposed wave trains and thereby not only identify the location of the resultant wave packet maximum, but also that packet's width and how it changes with time.

3.8 Spacetime History of a Wave Packet

The sharpened formulation of this picture consists of replacing a *sum* of superposed wave amplitudes with an *integral* of wave amplitudes

$$\begin{aligned} \Psi(x,t) &= \text{"}\Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots\text{"} \\ &= \int_{-\infty}^{\infty} f(E)\, e^{\frac{i}{\hbar}S_E(x,t)} dE\,. \end{aligned} \tag{3.21}$$

A very instructive example is that of a superpostion of monochromatic ("single energy") wavetrains, each one weighted by the amplitude $f(E)$ of a Gaussian window in the Fourier ("energy") domain,

$$f(E) = Ae^{-(E-E_0)^2/\epsilon^2} \tag{3.22}$$

The dominant contribution to this integral comes from within the window, which is centered around the location of E_0 of the Gaussian maximum and has width 2ϵ, which is small for physical reasons. Consequently, it suffices to represent the phase function as a Taylor series around that central point E_0, namely

$$S_E(x,t) = S_{E_0}(x,t) + \left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0}(E-E_0) + \frac{1}{2}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0}(E-E_o)^2 + \begin{matrix} higher \\ order \\ terms \end{matrix}, \tag{3.23}$$

and neglect the higher order terms. Keeping only the first three terms and ignoring the remainder allows an exact evaluation of the Gaussian superposition integral. This evaluation is based on the following formula

$$\int_{-\infty}^{\infty} e^{\alpha z^2+\beta z} dz = \sqrt{\frac{\pi}{-\alpha}} e^{-\frac{\beta^2}{4\alpha}}\,. \tag{3.24}$$

Applying it to the superpostion integral, Eq. (3.21) together with Eqs. (3.22) and (3.23), we make the following identification

$$\begin{aligned} z =& E - E_0 \,; \quad dz = dE \,, \\ \alpha =& -\frac{1}{\epsilon^2} + \frac{i}{\hbar}\frac{1}{2}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} \equiv -\frac{1}{\epsilon^2}(1 - i\sigma) \,, \\ -\frac{1}{\alpha} =& \frac{\epsilon^2}{1 - i\sigma} = \epsilon^2 \frac{1 + i\sigma}{1 + \sigma^2} \,, \\ \sigma =& \frac{1}{2}\frac{1}{\hbar}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} \epsilon^2 \,, \\ \beta =& \frac{i}{\hbar}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} . \end{aligned} \tag{3.25}$$

Inserting these expressions into the righthand side of the formula (3.24), one obtains

$$\begin{aligned} \Psi(x,t) &= A\sqrt{\pi}\epsilon\sqrt{\frac{1+i\sigma}{1+\sigma^2}} \exp\left\{ -\frac{1}{4}\left(\frac{\frac{\partial S(x,t)}{\partial E_o}}{\hbar} \right)^2 \epsilon^2 \left(\frac{1+i\sigma}{1+\sigma^2} \right) \right\} e^{i\frac{S_{E_o}(x,t)}{\hbar}} \\ &\equiv \underbrace{\mathcal{A}(x,t)}_{\substack{slowly \\ varying}} \underbrace{e^{i\frac{S_{E_o}(x,t)}{\hbar}}}_{\substack{rapidly \\ varying}} . \end{aligned}$$

This is the *rapidly oscillating* function

$$e^{i\, S_{E_0}(x,t)/\hbar}$$

modulated by a *slowly varying* amplitude $\mathcal{A}(x,t)$. For each time t this product represents a wave packet. The location of the maximum of this wave packet is given implicitly by

$$\left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0} = 0 \,. \tag{3.26}$$

As t changes, the x-location of the maximum changes. Thus we have curve in x-t space of the locus of those events where the slowly varying amplitude $\mathcal{A}$ has a maximum. In other words, this wave packet maximum condition locates those events (= points in spacetime) where constructive interference takes place.

A wave packet has finite extent in space and in time. This extent is governed by its squared modulus, i.e. the squared magnitude of its slowly varying amplitude,

$$|\Psi(x,t)|^2 = |\mathcal{A}|^2 = A^2\pi\epsilon^2 \frac{1}{\sqrt{1+\sigma^2}} \exp \underbrace{\left\{ -\frac{\epsilon^2}{2}\frac{1}{\sqrt{1+\sigma^2}} \frac{\left(\left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0} \right)^2}{\hbar^2} \right\}}_{\mathbf{E}(x,t)} \tag{3.27}$$

We see that this squared amplitude has nonzero value even if the condition for constructive interference, Eq.(3.26), is violated. This violation is responsible for the finite width of the wave packet. More precisely, its shape is controlled by the exponent $\mathbf{E}(x,t)$,

$$\mathbf{E}(x,t) \equiv \left\{ -\frac{\epsilon^2}{2} \frac{1}{\sqrt{1+\left(\frac{\epsilon^2}{2\hbar}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0}\right)^2}} \frac{\left(\left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0}\right)^2}{\hbar^2} \right\} \neq 0\,.$$

The spacetime evolution of this shape is exhibited in Figure 3.15. Thus the world-line of the particle is not a sharp one, but instead has a slight spread in space and in time. How large is this spread?

The magnitude of the wave mechanical ("non-classical") spread in the world line is the width of the Gaussian wave packet. This spread is Δx, the amount by which one has to move away from the maximum in order that the amplitude profile change by the factor $e^{\frac{1}{2}}$ from the maximum value. Let us calculate this spread under the circumstance where the effect due to dispersion is a minimum, i.e. when σ is neglegibly small. In that case the condition that $\mathbf{E}(x+\Delta x,t) = -1$ becomes

$$\left| \frac{\epsilon}{\hbar} \left.\frac{\partial S_E(x+\Delta x,t)}{\partial E}\right|_{E_0} \right| = 1\,.$$

Expand the left hand side to first order, make use of the fact that (x,t) is a point in spacetime where the wavepacket profile has a maximum, i.e. satisfies Eq.(3.26). One obtains

$$\left| \epsilon \frac{\partial^2 S}{\partial E \partial x} \Delta x \right| = \hbar$$

or, in light of $\partial S_E(x,t)/\partial x \equiv p(x,t;E)$,

$$\left| \epsilon \frac{\partial p}{\partial E} \Delta x \right| \;\; = \;\; \hbar\,,$$

and hence

$$\boxed{\Delta p \Delta x = \hbar}$$

Similarly the temporal extent Δt, the amount by which one has to wait (at fixed x) for the wave amplitude profile to decrease by the factor $e^{-1/2}$ from its maximum value, satisfies the condition

$$\left| \frac{\epsilon}{\hbar} \left.\frac{\partial S_E(x,t+\Delta t)}{\partial E}\right|_{E_0} \right| = 1$$

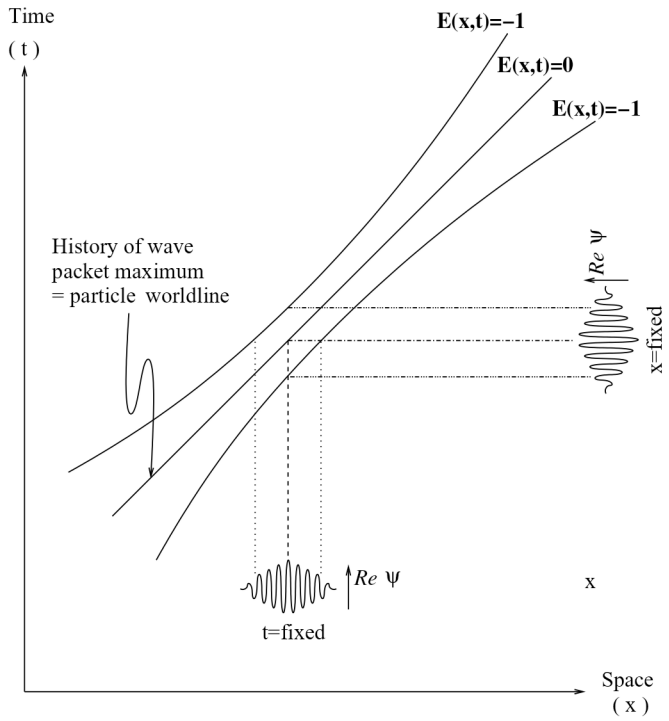


Figure 3.15: Spacetime particle trajectory ("the $\mathbf{E}(x,t) = 0$ isogram") and the dispersive wave packet amplitude histories surrounding it. The two mutually diverging ones (both characterized by $\mathbf{E}(x,t) = -1$) in this figure refer to the front and the back end of the wave packet at each instant $t = fixed$, or to the beginning and the end of the wave disturbance passing by a fixed location $x = fixed$. The particle and the wave packet maximum are moving with a velocity given by the slope of the $\mathbf{E}(x,t) = 0 = \left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0}$ isogram, which is the locus of constructive interference exhibited in Figure 3.14

On the other hand, the convergence and subsequent divergence ("dispersion") of the wave packet is controlled (and expressed mathematically) by the behavior of the second derivative, $\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0}$ of the dynamical phase $S_E(x,t)$. Whereas the behavior of its first derivative characterizes the difference in the motion of particles launched with difference initial conditions, its second derivative characterizes the intrinsically wave mechanical aspects of each of these particles.

which become

$$\left| \epsilon \frac{\partial^2 S_E}{\partial E \partial t}\bigg|_{E_0} \Delta t \right| = \hbar$$
$$\left| \epsilon (-) \frac{\partial E}{\partial E}\bigg|_{E_0} \Delta t \right| = \hbar$$

or

$$\boxed{\Delta E \Delta t = \hbar}\,.$$

The two boxed equation are called the Heisenberg indeterminacy relation. Even though we started with the dynamical phase S (see page 112) with $\Psi \sim e^{\frac{iS}{\hbar}}$ to arrive at the extremal path in spacetime, the constant $\hbar$ ("quantum of action") never appeared in the final result for the spacetime trajectory. The reason is that in the limit

$$\frac{S}{\hbar} \to \infty$$

the location of the wave packet reduces to the location of the wave crest. Once one knows the dynamical phase $S(x,t)$ of the system, the condition of constructive interference gives *without approximation* the location of the sharply defined Newtonian world line, the history of this wave crest, an extremal path through spacetime.

3.9 Hamilton's Equations of Motion

Lecture 19

To validate the claim that constructive interference leads to the extremal paths determined by the E-L equations, one must first recast them into a form that involves only q^i and p_j instead of $q^i$and $\dot{q}^j$. Holding off that validation until Section 3.12, we achieve the transition from $(q^i, \dot{q}^j)$ to (q^i, p_j) as follows:
The Lagrangian is a function of $q^i$and $\dot{q}^i$. Consequently,

$$dL = \sum_i \frac{\partial L}{\partial q^i} dq^i + \sum_i \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i$$

which may be rewritten as

$$dL = \sum_i \dot{p}_i dq^i + \sum_i p_i d\dot{q}^i\,,$$

where

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

and

$$\dot{p}_i = \frac{\partial L}{\partial q^i}$$

by the E-L equations. Using the fact that

$$p_i dq^i = d(p_i \dot{q}^i) - \dot{q}^i dp_i \,,$$

one obtains, after a sign reversal, an expression which depends only on q^i and p_i:

$$d \underbrace{\left(\sum_i p_i \dot{q}^i - L \right)}_{H} = -\sum_i \dot{p}_i dq^i + \sum_i \dot{q}^i dp_i \,. \tag{3.28}$$

Introduce the Hamiltonian of the system

$$H(p,q,t) \equiv \sum p_i \dot{q}^i - L \,.$$

Compare its differential

$$dH = \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial p_i} dp_i + zero \times d\dot{q}^i$$

with the one given above by Eq.(3.28). Recall that two differentials are equal if and only if the coefficients of the (arbitrary) coordinate differences (i.e. $dq^1, \cdots, dq^s$, $dp_1, \cdots, dp_s$) are equal. Consequently, one has

$$\boxed{\dot{q}^i = \frac{\partial H}{\partial p_i} \,, \qquad p_i = -\frac{\partial H}{\partial q_i} \,.}$$

These are the *Hamilton's* or the *canonical equations of motion.* They are equivalent to those of E-L. Comment 1: The fact that H does not depend on $\dot{q}^i$ follows directly from

$$\frac{\partial H}{\partial \dot{q}^i} = p_i - \frac{\partial L}{\partial \dot{q}^i} = 0$$

Comment 2: a) It follows from Hamilton's equations of motion that

$$\begin{aligned} \frac{dH}{dt} &= \frac{\partial H}{\partial q^i}\dot{q}^i + \frac{\partial H}{\partial p_i}\dot{p}_i + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial t} \end{aligned}$$

In other words, if H is independent of any *explicit* time dependence, i.e. time then H i*s a constant* along the trajectory of the evolving system.
Comment 2: b.) If H is independent of the *generalized coordinate* q^k, then

$$\frac{dp^k}{dt} = 0$$

i.e. p_k *is a constant* for the evolving system.

3.10 Lagrangian vs. Hamiltonian Mechanics: An Overview

There are two closely related ways of mathematizing a system with n degrees of freedom.

In the Lagrangian approach the state of the system at any instant of time is specified by its instantaneous generalized position $(q^1(t), \cdots, q^n(t))$ and its corresponding instantaneous velocity $(v^1, \cdots, v^n) = (\dot{q}^1(t), \cdots, \dot{q}^n(t))$. Thus a particular evolution of the system is mathematized in terms of a particular curve, a moving point,

$$\left(q^1(t), \cdots, q^n(t), \dot{q}^1(t), \cdots, \dot{q}^n(t)\right) \tag{3.29}$$

in the $2n$-dimensional continuum. This continuum is coordinatized by the base space coordinates $\{q^i\}_{i=1}^n$ and the velocity coordinates $\{v^j\}_{j=1}^n$ in the tangent space attached to each base point. The form of the dynamically allowed curves, Eq.(3.29), is determined by the system's Lagrangian

$$L(q^1, \dot{q}^1, \cdots, q^n, \dot{q}^n).$$

In the Hamiltonian approach the dynamics of the system is controlled by the Hamiltonian

$$H(q^1, \cdots, q^n, p_1, \cdots, p_n)$$

whose domain is spanned by the $2n$ coordinates

$$\left(q^1, \cdots, q^n, p_1, \cdots, p_n\right) \ . \tag{3.30}$$

As noted on page 115 in Section 3.4.1, the Legendre tranformation is a one-to-one onto mapping which transforms the Lagrangian and its domain coordinatized by Eq.(3.29) in a one-to-one manner onto the Hamiltonian and its domain coordinatized by Eq.(3.30):

$\{q^i, \dot{q}^i\}_{i=1}^n$		$\{q^i, p_i\}_{i=1}^n$
$L(q^i, \dot{q}^i)$		$H(q^i, p_i)$
$\dfrac{d}{dt}\left(\dfrac{\partial L}{\partial \dot{q}^i}\right) = \dfrac{\partial L}{\partial q^i}$ $i = 1, \cdots, n$		$\dfrac{dq^i}{dt} = \dfrac{\partial H}{\partial p_i}; \ \dfrac{dp_i}{dt} = -\dfrac{\partial H}{\partial q^i}$ $i = 1, \cdots n$
defined on the disjoint union of the tangent spaces of the base space $\{q^i\}$: $TM = \{q^i\} \times \{v^j \frac{\partial}{\partial q^j}\}$	$\leftrightarrow$	defined on the disjoint union of the cotangent spaces of the base space $\{q^i\}$: $T^*M = \{q^i\} \times \{p_j dq^j\}$

3.11 The Phase Space of a Hamiltonian System

The 2n-dimentional space spanned by the coordinates

$$\{q^1, \cdots, q^n, p_1, \cdots, p_n\}$$

is called the *phase space* of the system. In this space, the curve

$$\{q^i(t), p_i(t)\}$$

is an integral curve of the Hamiltonian vector field

$$\left(\frac{\partial H(q,p)}{\partial p_i}, -\frac{\partial H(q,p)}{\partial q^i}\right)$$

This means that the tangents to this curve are given by

$$(\dot{q}^i, \dot{p}_i) = \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right)$$

Example: For the simple harmonic oscilator the Lagarangian is:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

and the Hamiltonian is:

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 = E$$

a) The phase space of this system is spanned by x and p.
The Hamiltonian vector field is

$$\left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right) = \left(\frac{p}{m}, -kx\right)$$

b) The area of one of the phase-path ellipses is

$$\text{area} = \int p dx \qquad \text{and it has the dimension of "action"}$$

According to quantum mechanics the action of a periodic system must obey the Bohr quantization condition

$$\int p dx = \left(n + \frac{1}{2}\right) h\,, \qquad n = 1, 2, \cdots \tag{3.31}$$

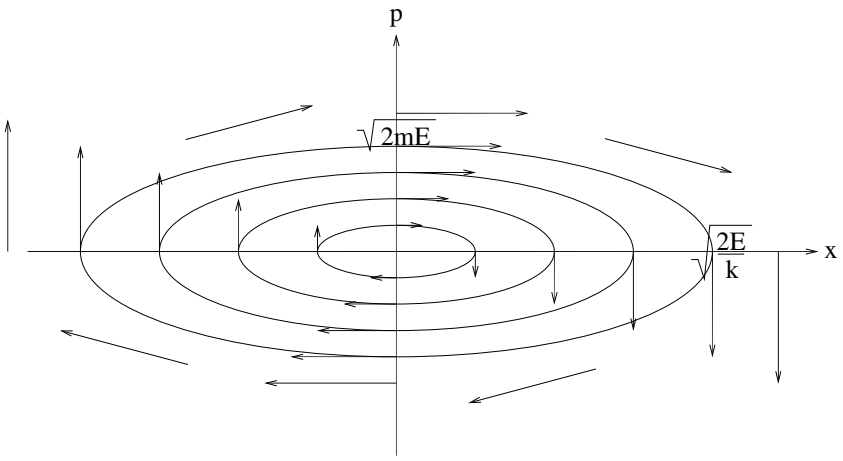


Figure 3.16: Hamiltonian vector field of a simple harmonic oscillator (s.h.o.) of mass m and spring constant k. The ellipses are integral curves whose tangents at each point are the vectors of that field. The major axis, $\sqrt{\frac{2E}{k}}$, and the minor axis, $\sqrt{2mE}$, of each ellipse are determined by the energy E of the s.h.o. The area of any particular ellipse is $2\pi\, E\, \frac{m}{k}$.

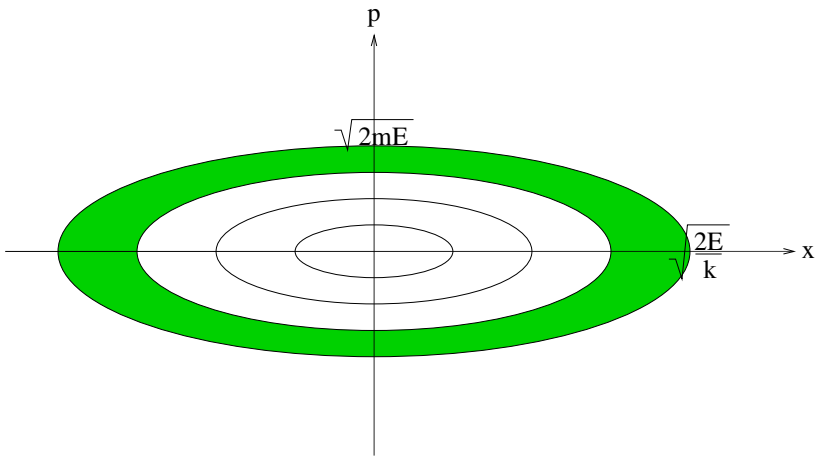


Figure 3.17: The shaded difference between the areas of adjacent phase space ellipses, Eq.(3.31), is precisely $h = 6.27\times10^{-27}$erg sec, which is one quantum of action.

Thus, as depicted in Figure 3.17, the quantum mechanically allowed phase space ellipses differ in area from one another by precisely $h = 6.27\times10^{-27}$erg sec, which is one quantum of action.

For the simple harmonic oscilator the area of one of these ellipses is $\int pdx = \pi\sqrt{2mE}\sqrt{\frac{2E}{k}} = 2\pi E\sqrt{\frac{m}{k}} = 2\pi\frac{E}{\omega}$. Thus the Bohr quantization condition yields

$$2\pi\frac{E}{\omega} = \left(n + \frac{1}{2}\right)h$$

or with $\frac{\omega}{2\pi} = frequency$

$$E = (n + \frac{1}{2})h \times frequency$$

3.12 Constructive interference ⇒ Hamilton's Equations

The principle of constructive interference provides the bridge between particle and wave mechanics. This fact is validatede by the following theorem.

Theorem. *Constructive interference conditions imply the Hamilton's equantions of motion and hence determine the existence of an extremal path.*

Proof: Step 1.) Consider a *complete integral* of the H-J equation

$$S = S(t, q^1, \cdots, q^s, \alpha_1, \cdots, \alpha_s)$$

i.e. a solution which has as many independent arbitrary constants as there are independent coordinates[5]. The constructive interference conditions are

$$\frac{\partial S}{\partial \alpha_k} = 0 \qquad k = 1, \cdots, s$$

[5]Such independence is expressed mathematically by the fact that

$$det\left|\frac{\partial^2 S}{\partial q^i \partial \alpha_j}\right| \neq 0\,.$$

This condition would be violated if the dependence on two constants were of the form $S(t, q^i, f(\alpha_1, \alpha_2), \alpha_3, \cdots, \alpha_s)$. Under such a violation S would depend only on $s-1$ independent parameters instead of s of them.

They determine implicitly a trajectory $q^i = q^i(t),\ i = 1, \cdots, s$.
Step 2.) Take the total derivative and obtain

$$\begin{aligned} 0 &= \frac{d}{dt}\frac{\partial S}{\partial \alpha_k} = \frac{\partial^2 S}{\partial t \partial \alpha_k} + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= -\frac{\partial}{\partial \alpha_k} H\left(q, \frac{\partial S(t,q,\alpha)}{\partial x^i}, t\right) + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= -\frac{\partial H}{\partial p_i}\frac{\partial^2 S}{\partial \alpha_k \partial q^i} + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= \frac{\partial^2 S}{\partial \alpha_k \partial q^i}\left(\frac{dq^i}{dt} - \frac{\partial H}{\partial p_i}\right), \end{aligned}$$

which implies the 1st half of Hamilton's equations,

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i},$$

provided $\frac{\partial^2 S}{\partial \alpha_k \partial q^i}$ is non-singular.
Step 3.) Differentiate both sides of the H-J equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial q^i}\left[\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right)\right] \\ &= \frac{\partial}{\partial t}\left(\frac{\partial S}{\partial q^i}\right) + \left(\frac{\partial H}{\partial p_k}\right)_q \frac{\partial^2 S}{\partial q^k \partial q^i} + \left(\frac{\partial H}{\partial q^i}\right)_p \\ &= \frac{\partial}{\partial t}\left(\frac{\partial S}{\partial q^i}\right) + \frac{dq^k}{dt}\frac{\partial}{\partial q^k}\left(\frac{\partial S}{\partial q^i}\right) + \left(\frac{\partial H}{\partial q^i}\right)_p \\ &= \frac{d}{dt}p_i + \left(\frac{\partial H}{\partial q^i}\right)_p \end{aligned}$$

which is the 2nd half of Hamilton's equation's,

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}.$$

QED. Thus the two Hamilton's equations of motion are implied by the principle of constructive interference indeed.

Lecture 20

3.13 Applications

Two of the most important applications of Hamilton-Jacobi theory are found in

1. the motion of bodies on the astronomical scale, for example, space craft, comets, or planets moving in a gravitational field, and in

2. the motion of bodies on the atomic scale, for example, a charged particle (electron) moving in the potential of an atomic nucleus or in th electromagnetic field of a pulsed laser.

The mathematical procedure for these and numerous other examples is routine and always the same:

(i) Construct the Hamiltonian for the system

(ii) Write down and solve the H-J equation

(iii) Apply the conditions of constructive interference to obtain the trajectories of the body.

Let us describe how this three step procedure is done in practice.

3.13.1 H-J Equation Relative to Curvilinear Coordinates

In constructing the Hamiltonian one must choose some specific set of coordinates. For a single particle it is is difficult to find an easier way of writing down the H-J equation than the way whose starting point is the element of arclength

$$\begin{aligned}(ds)^2 &= dx^2 + dy^2 + dz^2 \quad \text{(Cartesian coordinates)} \\ &= g_{ij}dx^i dx^j \qquad \text{(curvilinear coordinates)}\end{aligned} \tag{3.32}$$

This element of arclength is the best starting point because it is related so closely to the Lagrangian of the system. Indeed, one has

$$\begin{aligned}L &= \frac{1}{2}m\vec{x}\cdot\vec{x} - U \\ &= \frac{1}{2}m\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right] - U \\ &= \frac{1}{2}m\, g_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} - U\,.\end{aligned}$$

In other words, the Lagrangian is constructed relative to curvilinear coordinates by inspection. The steps leading to the H-J equation are now quite routine.

The momenta are

$$p_j = \frac{\partial L}{\partial \dot{x}^j} = m\, g_{ij}\dot{x}^i \,.$$

Let g^{kj}be the inverse of g_{ji}: $g^{kj}g_{ji} = \delta^k_i$ so that

$$\dot{x}^i = \frac{1}{m}\, g^{ij}p_j$$

and

$$H = p_j\dot{x}^j - L = \frac{1}{2m}\, g^{ij}p_ip_j + U$$

Thus the Hamilton-Jacobi equation is

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\, g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} + U$$

in terms of the inverse metric.

3.13.2 Separation of Variables

The most important way of solving the H-J equation is by the method of sep[aration of variables. To illustrate this, consider the following

Example (Particle in a dipole potential) Consider the motion of a particle in the potential of a combined dipole and monopole field. Relative to spherical coordinates the metric is

$$(ds)^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2$$

and that potential has the form

$$U(r,\theta) = \mu\frac{\cos\theta}{r^2} - \frac{k}{r}$$

Its equipotential surfaces are rotationally symmetric around the z-axis. The Lagrangian is

$$\begin{aligned} L &= \text{Kinetic Energy } - \text{Potential Energy} \\ &= \frac{1}{2}m\, g_{ij}\dot{q}^i\dot{q}^j - U(q^1,q^2,q^3) \\ &= \frac{1}{2}m\,(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2) - \mu\frac{\cos\theta}{r^2} + \frac{k}{r}\,. \end{aligned}$$

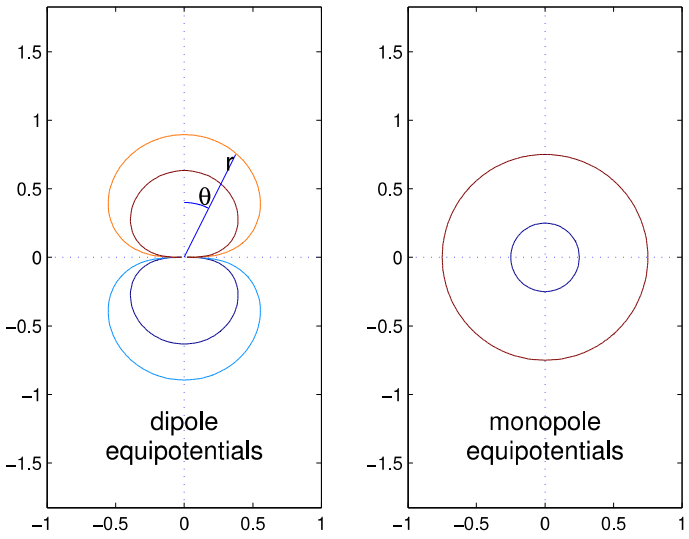


Figure 3.18: Rotationally symmetric potential as the sum a dipole potential $\left(\mu\frac{\cos\theta}{r^2}\right)$ plus a monopole potential $\left(-\frac{k}{r}\right)$.

The corresponding Hamilton-Jacobi equation is

$$\begin{aligned} 0 &= \frac{\partial S}{\partial t} + H \\ &= \frac{\partial S}{\partial t} + \frac{1}{2m} g^{ij} \frac{\partial S}{\partial q^i} \frac{\partial S}{\partial q^j} + U(q^1, q^2, q^3) \\ &= \frac{\partial S}{\partial t} + \frac{1}{2m}\left[\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{1}{r^2\sin^2\theta}\left(\frac{\partial S}{\partial \phi}\right)^2\right] + \mu\frac{\cos\theta}{r^2} - \frac{k}{r}\,. \end{aligned}$$

This equation can be solved by the method of separation of variables. This method is condensed into the following three definitions and propositions:

1. *Definition (Separable variables).* The variables $q^1, q^2, \cdots, q^s$ in the H-J equation

$$0 = \frac{\partial S}{\partial t} + H\left(t, q^i, \frac{\partial S}{\partial q^j}\right) \equiv \mathcal{H}\left(t, \frac{\partial S}{\partial t}, q^i, \frac{\partial S}{\partial q^j}\right)$$

are said to be *separable* if it has a "complete" solution of the form

$$S = S_0(t, \alpha_0) + S_1(q^1, \alpha_0, \alpha_1) + S_2(q^2, \alpha_0, \alpha_1, \alpha_2) + S_3(q^3, \alpha_0, \alpha_1, \alpha_2, \alpha_3) \tag{3.33}$$

where each S_i depends only on t and q^i respectively.

2. *Definition (Complete solution).* A solution is said to be *complete* if

$$\det \left| \frac{\partial^2 S}{\partial \alpha_i \partial q^j} \right| \neq 0$$

Remark 1: We saw (in Section 3.12 on page 139) in the context of reconstructing the classical trajectories of a Hamiltonian system from the principle of constructive interference it was essential that the matrix $[\partial^2 S/\partial \alpha_i \partial q^j]$ be non-singular.
Remark 2: The solution, Eq.(3.33), is complete indeed, because

$$\det \left| \frac{\partial^2 S}{\partial \alpha_i \partial q^j} \right| = \det \begin{array}{|ccc|c|c} 1 & 2 & 3 & \leftarrow & j \\ * & 0 & 0 & & 0 \\ * & * & 0 & \neq 0 & 1 \\ * & * & * & & 2 \\ & & & & \uparrow \\ & & & & i \end{array}$$

and its diagonal elements are not zero.

3. *Definition (Separability condition).* The Hamilton-Jacobi equation is said to satisfy the separability criterion if its Hamiltonian is of the form

$$\mathcal{H}\left(t, \frac{\partial S}{\partial t}, q^i, \frac{\partial S}{\partial q^j}\right) = f_3\left(f_2\left(f_1\left(f_0\left(t, \frac{\partial S}{\partial t}\right), q_1, \frac{\partial S}{\partial q^1}\right), q^2, \frac{\partial S}{\partial q^2}\right), q^3, \frac{\partial S}{\partial q^3}\right)$$

(for $s = 3$ degrees of freedom). This functional form is said satisfy the condition of separability because the solution to this first order p.d.e. has the separated form, Eq.(3.33). In other words, the claim is

Proposition: Definition 3 implies Definition 1.
Proof: Step 1.) According to Definition 3 the H-J equation is

$$f_3\left(f_2\left(f_1\left(f_0\left(t, \frac{\partial S}{\partial t}\right), q_1, \frac{\partial S}{\partial q^1}\right), q^2, \frac{\partial S}{\partial q^2}\right), q^3, \frac{\partial S}{\partial q^3}\right) = 0\,. \tag{3.34}$$

The method of solution via separation of variables starts by solving for f_0. One finds

$$f_0\left(t, \frac{\partial S}{\partial t}\right) = \text{an expression involving } q^1, q^2, q^3, \frac{\partial S}{\partial q^1}, \frac{\partial S}{\partial q^2}, \text{ and } \frac{\partial S}{\partial q^3} \text{ only .}$$

It follows that the solution must have the separated form

$$S = T(t) + S'(q^1, q^2, q^3)\,. \tag{3.35}$$

This is because the common value of

$$\underbrace{f_0\left(t, \frac{dT(t)}{dt}\right)}_{\substack{\text{independent} \\ \text{of } q^1, q^2, q^3}} = \underbrace{\text{an expression that depends only on } q^1, q^2, q^3}_{\text{independent of t}}$$

is independent of all variables. This common independence implies that f_0 is a constant, say, α_0:

$$f_0\left(t, \frac{dT(t)}{dt}\right) = \alpha_0\,. \tag{3.36}$$

Solving for $T(t)$, one obtains

$$T(t) = S_0(t, \alpha_0)\,. \tag{3.37}$$

and, in light of Eq(3.35),

$$\frac{dT}{dt} = \frac{\partial S}{\partial t} \tag{3.38}$$

so that

$$f_0\left(t, \frac{\partial S}{\partial t}\right) = \alpha_0\,. \tag{3.39}$$

Step 2.) Introduce this function value into Eq.(3.34). This remove its t-dependence. The simplified functional equation is

$$f_3\left(f_2\left(f_1\left(\alpha_0, q^1, \frac{\partial S'}{\partial q^1}\right), q^2, \frac{\partial S'}{\partial q^2}\right), q^3, \frac{\partial S'}{\partial q^3}\right) = 0\,.$$

Now solve for f_1 and find that

$$f_1\left(\alpha_0, q^1, \frac{\partial S'}{\partial q^1}\right) = \text{an expression involving } \; q^2, q^3, \frac{\partial S'}{\partial q^2}, \text{and} \frac{\partial S'}{\partial q^3} \; \text{ only.}$$

It follows that the solution S' has the separated form

$$S' = Q_1(q^1) + S''(q^2, q^3)\,. \tag{3.40}$$

This is because

$$\underbrace{f_1\left(\alpha_0, q^1, \frac{dQ_1(q^1)}{dq^1}\right)}_{\substack{\text{independent} \\ \text{of } q^2, q^3}} = \underbrace{\text{an expression that depends only on } \text{ on } q^2, q^3}_{\text{independent of q}^1}.$$

This common independence implies that f_1 is a constant, say, α_1:

$$f_1\left(\alpha_0, q^1, \frac{dQ_1(q^1)}{dq^1}\right) = \alpha_1\,.$$

Solving for $Q_1(q^1)$, one obtains

$$Q_1(q^1) = S_1(q^1, \alpha_1, \alpha_0)\,. \tag{3.41}$$

and, in light of Eq.(3.40),

$$\frac{dQ_1}{dq^1} = \frac{\partial S'}{\partial q^1} \tag{3.42}$$

At this stage one has brought the solution, Eq.(3.35), to the H-J equation has the into the separated form

$$S = T(t) + Q_1(q^1) + S''(q^2, q^3)\,.$$

Step 3.) Repeat Step 2.) two more times to obtain

$$f_2\left(\alpha_0, \alpha_1, q^2, \frac{dQ_2(q^2)}{dq^2}\right) = \alpha_2$$
$$f_3\left(\alpha_0, \alpha_1, \alpha_2, q^3, \frac{dQ_3(q^3)}{dq^{23}}\right) = \alpha_3\,.$$

Notice, however, that the H-J Eq.(3.34) implies that $\alpha_3 = 0$, always. Consequently, there are only *three* independent separation constants, $(\alpha_0, \alpha_1, \alpha_2)$, while the number of independent variables, (t, q^1, q^2, q^3) is four. It follows that

$$S = T(t) + Q_1(q^1) + Q_2(q^2) + Q_3(q^3)$$

and hence with Eqs.(3.37), (3.41), $\cdots$etc. The to-be-determined solution S appears in the H-J equation only in the form of its derivatives. Consequently, S necessarily contains an additive constant, which in general depends on the three separation constants $\alpha_0, \alpha_1, \alpha_3$. Thus the dynamical phase S has indeed the separated form whenever its H-J equation has the form (3.34):

$$\begin{aligned} S = {} & S_0(t, \alpha_0) + S_1(t, q^1, \alpha_0, \alpha_1) \\ & + S_2(t, q^2, \alpha_0, \alpha_1, \alpha_2) + S_3(t, q^3, \alpha_0, \alpha_1, \alpha_2, \alpha_3 = 0) + \delta(\alpha_0, \alpha_1, \alpha_2)\,. \end{aligned}$$

The purpose and the specific values of $\alpha_0, \alpha_1, \alpha_3$ and the form of δ mathematize the specific nature of the particular system under consideration.

Motion in the Field of Two Unequal Bodies: Prolate Spheroidal Coordinates

As an application of Hamilton-Jacobi theory, consider the problem of determining the motion of a particle under the influence of forces toward two fixed centers.

The force due to each center is directed radially and proportional to the inverse squared distance.

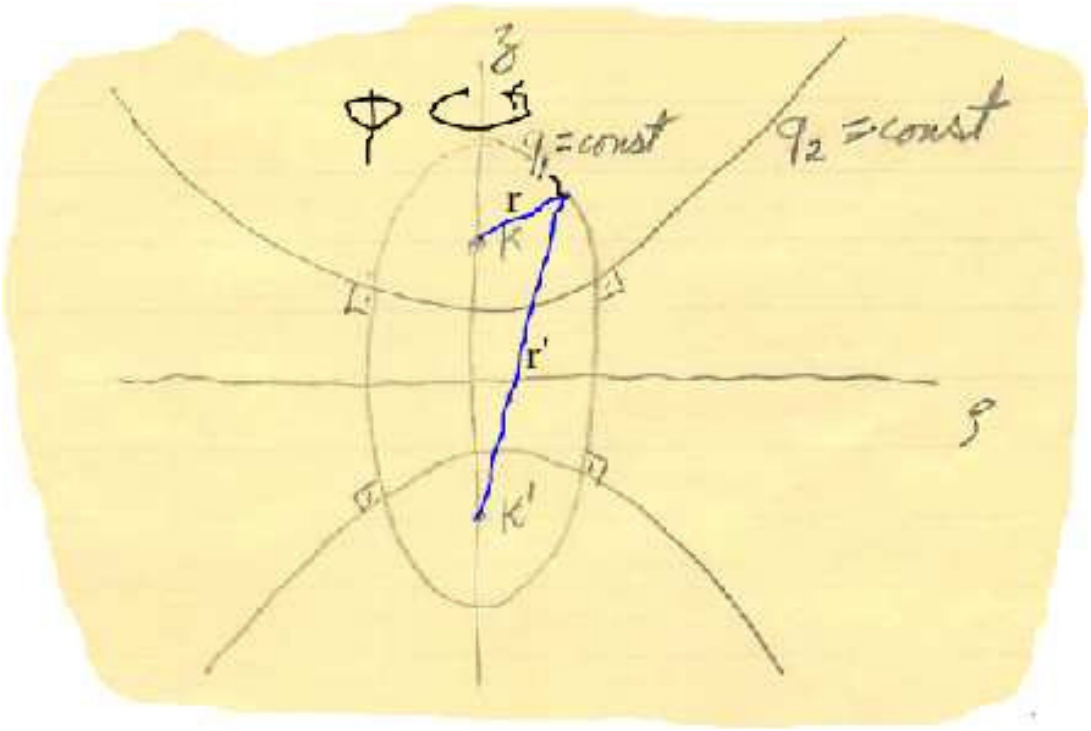


Figure 3.19: Prolate coordinates q_1, q_2 and the azimuthal angle ϕ coordinatize the geometry of two static attractive centers of strengths κ and κ' located at $z = c$ and $z = -c$ respectively. These locations are the foci of the confocal ellipsoids q_1 and confocal hyperboloids q_2. The ellipsoids, $q_1 =$ fixed are the loci of points where the sum $r' + r =$ constant; the hyperboloids $q_2 =$ fixed are the loci where the difference $r' - r =$ constant.

In light of Figure 3.19 the two body potential is therefore

$$U = \frac{\kappa}{r} + \frac{\kappa'}{r'} = \frac{\kappa}{\sqrt{\rho^2 + (z-c)^2}} + \frac{\kappa'}{\sqrt{\rho^2 + (z+c)^2}}. \tag{3.43}$$

Here $2c$ is the distance between the two centers and κ_1 and κ_2 are their respective

sources of the force fields. The problem of an otherwise free particle can be solved completely for such a potential.

This potential has various limiting forms, for which the dynamics of the particle can therefore also be solved exactly. They are:

1. Dipole field The sources (i) have very large equal and opposite signs,

$$\kappa = -\kappa' \quad \longrightarrow \infty,$$

(ii) are very closely spaced

$$2c \longrightarrow 0,$$

but (iii) the product of source strength and source separation,

$$\kappa \cdot 2c = \text{fixed} \equiv \mu \quad (\text{“dipole moment”}),$$

remains fixed. In that case

$$U \to \mu \frac{\cos\theta}{r^2} \quad (\text{“dipole potential”})$$

2. Uniform field For this case we arrange the attractive centers so that first of all the attractive centers are located at $\bar{z} = -2c$ and $\bar{z} = 0$ and then let $-2c \to -\infty$.

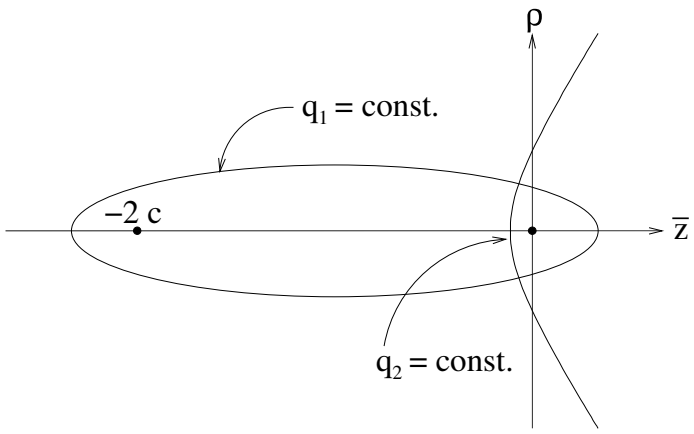


Figure 3.20: Shifted spheroidal coordinate system.

The result is a set of confocal parabolas whose focus is the origin, as in Figure 3.21.

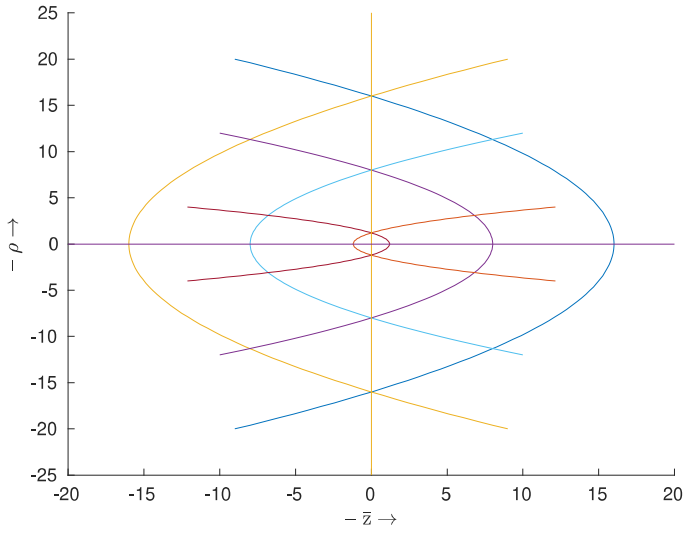


Figure 3.21: Confocal parabolas with common focus at the origin as the asymtotic limit of prolate spheroidal coordinates, Figure 3.19.

A. Particle in the *field of a dipole*:

$$\begin{array}{c}\kappa = -\kappa' \\ \kappa \to \infty \\ c \to 0\end{array} \quad \text{but} \quad c\kappa = \text{const} = \frac{\mu}{2}$$

[FIGURE]

B. Particle in the field of an isolated body immersed in a uniform field. For example, the hydrogen atom in a uniform electric field.

[FIGURE]

For this case we arrange the attractive centers so that

[FIGURE]

$$\begin{array}{c}c \to \infty \\ \kappa' \to \infty \\ c \to 0\end{array} \quad \text{but} \quad \frac{\kappa'}{c^2} = \text{const} = 4F$$

In that case

$$U \to \frac{\kappa}{\sqrt{\rho^2 + \bar{z}^2}} - F\bar{z}$$

The first task is to establish appropriate coordinates. The rotational symmetry around the z-axis suggests that we start out with cylindrical coordinates relative to the axis containing the sources, in which the squared element of arclength ("metric") has the form

$$ds^2 = g_{ij}dx^i dx^j = d\rho^2 + dz^2 + \rho^2 d\phi^2.$$

It is sufficient to focus on a plane $\phi = \text{const}$.

Instead of letting the coordinate surfaces be $\rho = \text{const}$ and $z = \text{const}$, we introduce elliptical and hyperbolic surfaces of revolution, with the force centers as their common foci.

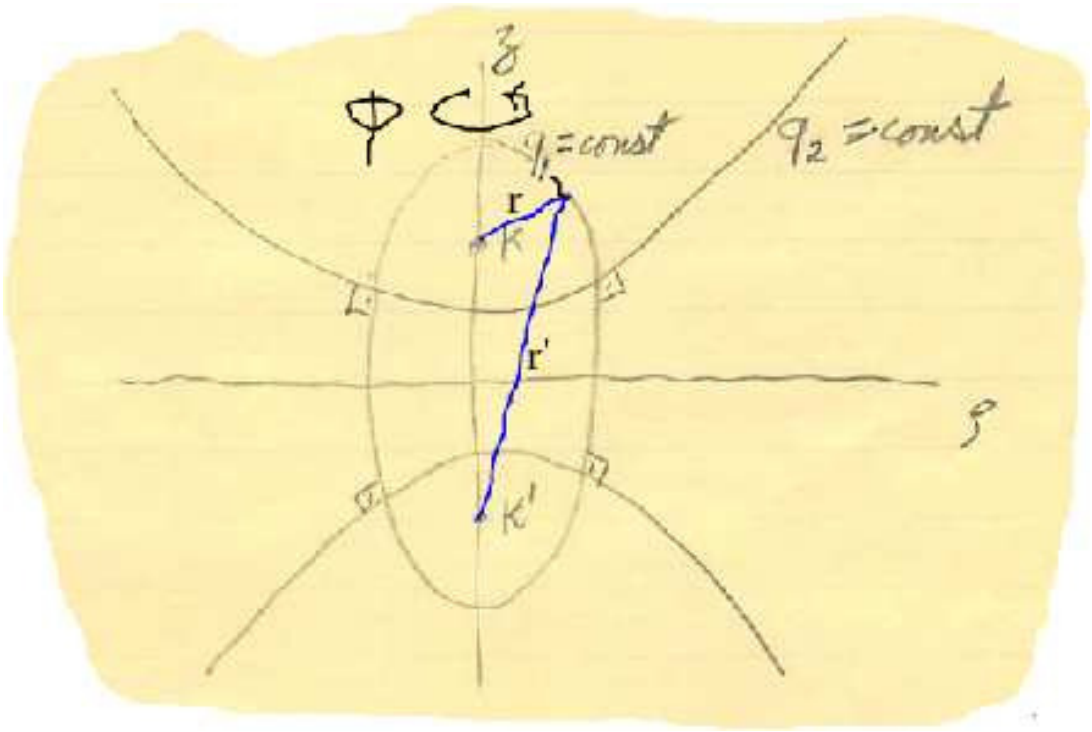


Figure 3.22: Prolate spherioidal coordinates.

Such a set of orthogonal coordinate surfaces is most easily introduced by means of the conformal transformation

$$z + i\rho \equiv w = f(q) = c\cosh q = c\cosh(q_1 + iq_2)$$

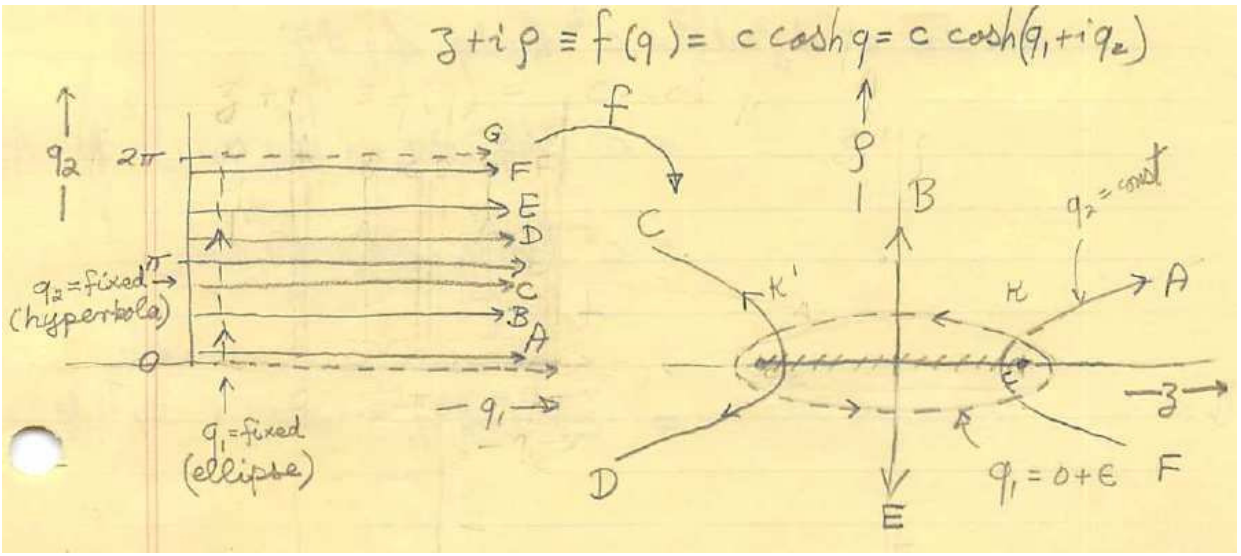


Figure 3.23: complex $q = q_1 + iq_2$ plane $\xrightarrow{f}$ complex w plane

The real and imaginary parts of this transformation are

$$\begin{aligned} w = z + i\rho &= c\cosh(q_1 + iq_2) \\ &= c\,(\cosh q_1 \cos q_2 + i \sinh q_1 \sin q_2) \\ z &= c\cosh q_1 \cos q_2 \qquad (3.44) \\ \rho &= c \sinh q_1 \sin q_2 \qquad (3.45) \end{aligned}$$

so that the locus of constant q_1 and q_2 yields

$$\begin{array}{ll} \text{ellipses} & \dfrac{z^2}{c^2\cosh^2 q_1} + \dfrac{\rho^2}{c^2 \sinh^2 q_1} = 1 \qquad 0 \le q_1 < \infty \\ \text{and} & \\ \text{hyperbolas.} & \dfrac{z^2}{c^2\cos^2 q_2} - \dfrac{\rho^2}{c^2 \sin^2 q_2} = 1 \qquad 0 \le q_2 < 2\pi \end{array}$$

The coordinates q_1 and q_2 together with ϕ are called prolate spheroidal coordinates.

The squared element of arclength (i.e. the *metric*) relative to these prolate ellipsoidal coordinates is

$$\begin{aligned} ds^2 &= dz^2 + d\rho^2 + \rho^2 d\phi^2 \\ &= dw\, d\bar{w} + \rho^2 d\phi^2 \\ {\scriptstyle w=f(q)} \quad &= \underbrace{f'(q)\overline{f'(q)}}_{c^2|\sinh q|^2} \underbrace{(dq_1^2 + dq_2^2)}_{dq\, d\bar{q}} + \rho^2 d\phi^2. \end{aligned}$$

Here

$$|f'(q)|^2 = c^2|\sinh q|^2 = c^2\big[\underbrace{\sinh^2 q_1}_{\cosh^2 q_1 - 1}\cos^2 q_2 + \cosh^2 q_1 \underbrace{\sin^2 q_2}_{1-\cos^2 q_2}\big]$$
$$= c^2\left[\cosh^2 q_1 - \cos^2 q_2\right] \equiv c^2\left[\sinh^2 q_1 + \sin^2 q_2\right] \tag{3.46}$$

Thus the metric is

$$(ds)^2 = c^2\left(\sinh^2 q_1 + \sin^2 q_2\right)\left(dq_1^2 + dq_2^2\right) + c^2\sinh^2 q_1 \sin^2 q_2\, d\phi^2$$

where $\rho^2 = c^2\sinh^2 q_1 \sin^2 q_2$. The Lagrangian is therefore

$$L = \frac{m}{2}c^2\left[\left(\sinh^2 q_1 + \sin^2 q_2\right)\left(\dot{q}_1^2 + \dot{q}_1^2\right) + \sinh^2 q_1 \sin^2 q_2\, \dot{\phi}^2\right] - U.$$

The potential $U = \frac{\kappa}{\sqrt{\rho^2+(z-c)^2}} + \frac{\kappa'}{\sqrt{\rho^2+(z+c)^2}}$ couples the two cylindrical degrees of freedom, z and ρ, in an inextricable fashion. This is because of the square root dependencies. However, these offending square roots disappear relative to the oblate spheroidal coordinates q_1 and q_2 defined by Eqs.(3.44)-(3.45). They disappear because relative to these coordinates

$$\begin{aligned}
\rho^2 + (z \pm c)^2 &= c^2\sin^2 q_2 \sinh^2 q_1 + c^2\left(\cosh q_1 \cos q_2 \pm 1\right)^2 \\
&= c^2\left[\sin^2 q_2 \sinh^2 q_1 + \cosh^2 q_1 \cos^2 q_2 \pm 2\cosh q_1 \cos q_2 + 1\right] \\
&= c^2\left[\left(1 - \cos^2 q_2\right)\left(\cosh^2 q_1 - 1\right) + \cosh^2 q_1 \cos^2 q_2 \pm 2\cosh q_1 \cos q_2 + 1\right] \\
&= c^2\left[\cosh^2 q_1 + \cos^2 q_2 \pm 2\cosh q_1 \cos q_2\right] \\
&= c^2\left(\cosh q_1 \pm \cos q_2\right)^2 .
\end{aligned}$$

This perfect square liberates the U from those offending square roots:

$$\begin{aligned}
U(q_1, q_2) &= \frac{\kappa}{c}\frac{1}{\cosh q_1 - \cos q_2} + \frac{\kappa'}{c}\frac{1}{\cosh q_1 + \cos q_2} \\
&= \frac{\left(\kappa + \kappa'\right)\cosh q_1 + \left(\kappa - \kappa'\right)\cos q_2}{\underbrace{\cosh^2 q_1 - \cos^2 q_2}_{\sinh^2 q_1 + \sin^2 q_2}}
\end{aligned}$$

From the Lagrangian one finds that the generalized momenta are

$$\begin{aligned}
p_i &= mc^2\left(\cosh^2 q_1 - \cos^2 q_2\right)\dot{q}_i \qquad i = 1, 2 \\
&= mc^2\left(\sinh^2 q_1 + \sin^2 q_2\right)\dot{q}_i
\end{aligned}$$

and

$$p_\phi = mc^2\left(\sinh^2 q_1 \sin^2 q_2\right)\dot{\phi}.$$

The Hamiltonian is

$$H = \frac{1}{2mc^2}\left\{\frac{p_1^2+p_2^2}{\sinh^2 q_1+\sin^2 q_2}+\left[\frac{1}{\sinh^2 q_1 \sin^2 q_2}\right]p_\phi^2\right\}+U\left(q_1,q_2\right).$$

Note: t and ϕ are cyclic coordinates. Thus H is a constant, say E. Therefore the Hamilton-Jacobi equation is, with the help of Eq.(3.46),

$$\begin{aligned}E = &\frac{1}{2mc^2\left(\sinh^2 q_1+\sin^2 q_2\right)}\times\\ &\left\{\left(\frac{\partial S_0}{\partial q_1}\right)^2+\left(\frac{\partial S_0}{\partial q_2}\right)^2+\left[\frac{\sinh^2 q_1+\sin^2 q_2}{\sinh^2 q_1\sin^2 q_2}\right]\left(\frac{\partial S_0}{\partial \phi}\right)^2\right\}\\ &+\frac{1}{c}\frac{(\kappa+\kappa')\cosh q_1+(\kappa-\kappa')\cos q_2}{\sinh^2 q_1+\sin^2 q_2}.\end{aligned}$$

The separability of this equation becomes more evident once one multiplies it by $c^2\left(\sinh^2 q_1+\sin^2 q_2\right)$:

$$\begin{aligned}0 =&\frac{1}{2m}\left\{\left(\frac{\partial S_0}{\partial q_1}\right)^2+\left(\frac{\partial S_0}{\partial q_2}\right)^2+\left[\frac{1}{\sin^2 q_2}+\frac{1}{\sinh^2 q_1}\right]\left(\frac{\partial S_0}{\partial \phi}\right)^2\right\}\\ &+c(\kappa+\kappa')\cosh q_1+c(\kappa-\kappa')\cos q_2\\ &-2mEc^2\sinh^2 q_1-2mEc^2\sin^2 q_2\ .\end{aligned}$$

One infers that the general solution to this equation is

$$S=-Et+S_1(q_1)+S_2(q_2)+p_\phi\phi$$

where S_1 and S_2 satisfy the ordinary differential equations

$$\left(\frac{dS_1}{dq_1}\right)^2-2mc^2E\sinh^2 q_1+\frac{p_\phi^2}{\sinh^2 q_1}+2mc(\kappa+\kappa')\cosh q_1=\alpha_1$$

and

$$\left(\frac{dS_2}{dq_2}\right)^2-2mc^2E\sin^2 q_2+\frac{p_\phi^2}{\sin^2 q_2}+2mc(\kappa-\kappa')\cos q_2=-\alpha_1\ .$$

Here α_1 is the common separation constant. It follows that the solution to the H-J equation is

$$\begin{aligned}S=&-Et+p_\phi\phi\\ &+\int^{q_1}dq_1\left[\alpha_1-2mc^2E\sin^2 q_1-\frac{p_\phi^2}{\sinh^2 q_1}-2mc(\kappa+\kappa')\cosh q_1\right]^{1/2}\\ &+\int^{q_2}dq_2\left[-\alpha_1-2mc^2E\sin^2 q_2-\frac{p_\phi^2}{\sin^2 q_2}-2mc(\kappa-\kappa')\cos q_2\right]^{1/2}\\ &+\ \delta(E,p_\phi,\alpha_1).\end{aligned}$$

The orbits are determined by the constructive interference conditions

$$\frac{\partial S}{\partial E} = 0, \qquad \frac{\partial S}{\partial p_\phi} = 0, \qquad \frac{\partial S}{\partial \alpha_1} = 0.$$

The intersection of these three surfaces in the 4-d space spanned by (t, ϕ, q_1, q_2) yields the $E, p_\phi, \alpha_1)$-parametrized family of globally defined trajectories in that 4-d spacetime. Their velocity components are

$$\frac{dq_1}{dt} = \frac{\partial H}{\partial p_1} = \frac{p_1}{mc^2(\sinh^2 q_1 + \sin^2 q^2} \tag{3.47}$$

$$\frac{dq_2}{dt} = \frac{\partial H}{\partial p_2} = \frac{p_2}{mc^2(\sinh^2 q_1 + \sin^2 q^2} \tag{3.48}$$

$$\frac{d\phi}{dt} = \frac{\partial H}{\partial \phi} = \frac{p_\phi}{mc^2\ \sinh^2 q_1 \sin^2 q^2} \tag{3.49}$$

$$\tag{3.50}$$

The method of separating variables is one that gives the momentum components

$$\begin{aligned}
\frac{\partial S}{\partial t} &= -E\ (= const.) \\
\frac{\partial S}{\partial \phi} &= p_\phi\ (= const.) \\
\frac{\partial S}{\partial q_1} &= p_1(q_1) \\
\frac{\partial S}{\partial q_2} &= p_2(q_2),
\end{aligned}$$

and the solution to the H-J equation in explicit form. Apply them to the velocity expressions. Thereby obtain the instantaneous velocities (tangents) at each point of a trajectory:

$$\frac{dq_1}{dt} = \frac{1}{mc^2(\sinh^2 q_1 + \sin^2 q^2} \left[\alpha_1 - 2mc^2 E \sin^2 q_1 - \frac{p_\phi^2}{\sinh^2 q_1} - 2mc(\kappa + \kappa') \cosh q_1\right]^{1/2}$$

$$\frac{dq_2}{dt} = \frac{1}{mc^2(\sinh^2 q_1 + \sin^2 q^2} \left[-\alpha_1 - 2mc^2 E \sin^2 q_2 - \frac{p_\phi^2}{\sin^2 q_2} - 2mc(\kappa - \kappa') \cos q_2\right]^{1/2}$$

$$\frac{d\phi}{dt} = \frac{1}{mc^2\ \sinh^2 q_1 \sin^2 q^2}\ \ p_\phi$$

These expressions furnish the boundaries between those phase space domains accessible to a particle where $\frac{dq_1}{dt}$ and $\frac{dq_2}{dt}$ are real, and those where they are imaginary

and hence inaccessible and strictly forbidden (at least "classically"). The boundary between these phase space domains is where the velocities $\frac{dq_1}{dt}$ and $\frac{dq_2}{dt}$ vanish, i.e. where

$$\left[\alpha_1 - 2mc^2 E \sin^2 q_1 - \frac{p_\phi^2}{\sinh^2 q_1} - 2mc(\kappa+\kappa')\cosh q_1\right]^{1/2} = 0$$
$$\left[-\alpha_1 - 2mc^2 E \sin^2 q_2 - \frac{p_\phi^2}{\sin^2 q_2} - 2mc(\kappa-\kappa')\cos q_2\right]^{1/2} = 0$$

How Spheroidal Coordinates can be Deformed into Spherical Coordinates
Consider the potential due to two attractive centers.

[FIGURE]

Consider the case $\kappa = -\kappa'$. Now let $c \to 0$, $\kappa \to \infty$ so that

$$2c\kappa = \text{const} = \mu.$$

The two prolate spheroidal coordinates q_1 and q_2, which are determined by

$$\rho = c \sinh q_1 \sin q_2$$
$$z = c \cosh q_1 \cos q_2$$

become the spherical coordinates θ and r. This is so because for fixed ρ and z ("same point in space") $c \sinh q_1$ and $c \cosh q_1$ remain fixed. Thus as $c \to 0$, $q_1 \to \infty$. Consequently

$$\left.\begin{matrix} c \sinh q_1 \\ c \cosh q_2 \end{matrix}\right\} \to \text{same constant} \equiv r$$

Thus, setting $q_1 \equiv \theta$, one obtains

$$\rho = r \sin\theta$$
$$z = r \cos\theta$$

[FIGURE]

In other words, the prolate spheroidal coordinates become the spherical coordinates.

You can convince yourself that

$$\lim_{\substack{c\to 0 \\ 2\kappa c=\mu}} \underbrace{\frac{\kappa}{\sqrt{\rho^2+(z-c)^2}} - \frac{\kappa}{\sqrt{\rho^2+(z+c)^2}}}_{U} = \frac{2\kappa c z}{\sqrt{\rho^2+z^2}} = \mu \frac{\cos\theta}{r^2}.$$

In other words, the potential becomes

$$\boxed{U = \mu \frac{\cos\theta}{r^2}},$$

that of a dipole of strength μ. The H-J equation separates for this potential, and the trajectories can therefore be determined.

The limit of the potential U is

$$\begin{aligned} U &= \frac{\kappa}{\sqrt{\rho^2 + (z-c)^2}} - \frac{\kappa}{\sqrt{\rho^2 + (z+c)^2}} \\ &= \frac{\kappa\left[\sqrt{\rho^2 + z^2 + 2zc + c^2} - \sqrt{\rho^2 + z^2 - 2zc + c^2}\right]}{\sqrt{\left(\rho^2 + z^2 + c^2\right)^2 - (2cz)^2}}. \end{aligned}$$

As $c \to 0$ the difference in the brackets goes to 0. To preserve accuracy, we must expand around $p^2 + z^2$. We obtain

$$U = \kappa\left(\rho^2 + z^2\right)^{1/2} \frac{\left[1 + \frac{zc}{\rho^2+r^2} + \cdots - \left(1 - \frac{zc}{\rho^2+r^2} + \cdots\right)\right]}{(\rho^2 + z^2)\left\{1 - \frac{2c^2z^2}{(\rho^2+z^2)^2} + \cdots\right\}}$$

$$\therefore \quad \lim_{\substack{c\to 0 \\ 2\kappa c = \mu}} U = \frac{2\kappa c z}{\left(\rho^2 + z^2\right)^{1/2}} = \mu \frac{\cos\theta}{r^2}$$

This is the potential of a dipole whose dipole strength is μ.

[FIGURE]

Figure 3.24: Potential of a field whose strength is $\mu = 2\kappa c$.

One can solve exactly the problem of a particle moving in such a field. The Lagrangian is

$$L = \frac{1}{2} m\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2\right) - \mu\frac{\cos\theta}{r^2}.$$

The Hamilton-Jacobi equation is

$$0 = \frac{\partial S}{\partial t} + \frac{1}{2m}\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{1}{r^2\sin^2\theta}\left(\frac{\partial S}{\partial \phi}\right)^2 - \frac{\mu}{r^2}\cos\theta.$$

This equation can be solved using the separation of variables technique:

$$S = -Et + R(r, E, \ell, p_\phi) + \Theta(\theta, \ell, p_\phi) + p_\phi\phi + \delta(E, \ell, p_\phi).$$

Here

$$\alpha_1 = E \qquad \alpha_2 = \ell \qquad \alpha_3 = p_\phi$$

are the separation constants. The orbits are given by the constructive interference condition

$$\frac{\partial S}{\partial E} = 0, \qquad \frac{\partial S}{\partial \ell} = 0, \qquad \frac{\partial S}{\partial p_\phi} = 0.$$

3.14 Hamilton's Principle for the Mechanics of a Continuum

Lecture 21

If a principle is to be gauged by the *scope* of the phenomena and processes it brings into our grasp and by the *brevity* of its annunciation, then the extremum principles of physics and engineering do not have many rivals. Variational calculus (and more generally, mathematics) highlights the potency of the unaided human mind. It gives a person the methods for graspin g the physical world from the nooks and crannies here on earth, on the submicroscopic through the biological scale, all the way the farthest reaches of the universe.

Most prominent among these principles is Hamilton's principle, which is based on the action integral

$$I = \int_{t_0}^{t_1} (\text{K.E.} - \text{P.E.}) dt.$$

(K.E.=Kinetic Energy, P.E.=Potential Energy)

We would like to extend its range of applicability from systems having a finite number of degrees of freedom to systems having an infinite number.

Hamilton's principle accommodates such an extension, provided (i) we can identify the P.E. and the K.E. of such systems and then (ii) give a mathematical formulation for optimizing their performance as ascertained by their action integral.

Examples. (i) The P.E. of a string is the finite sum of the potential energy of its atoms. However, we shall approximate this discrete and finite system by a continuum whose state is specified by the transverse displacement $\psi(x)$. If the string is under tension T [force], then its potential energy is the energy necessary to bring it from its unstretched to the stretched state associated with its transverse displacement $\psi(x)$. The stretch energy is

$$\begin{aligned} \text{P.E.} &= \int T \left[\sqrt{1 + \left(\frac{d\psi}{dx} \right)^2} - 1 \right] dx \\ &\cong \frac{1}{2} \int T \left(\frac{d\psi}{dx} \right)^2 dx \quad \text{(small amplitude approximation)} \end{aligned}$$

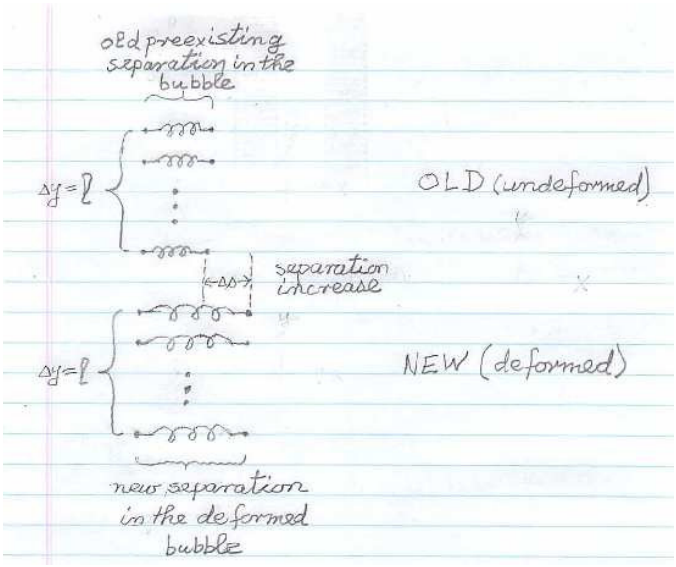


Figure 3.25: Molecular structural change in a bubble due to its deformation. The associated change in an element of area is $\ell \Delta s$.

(ii) If the string is embedded in an elastic medium whose elastic constant is κ and if the string is subjected to a given transverse force density $F(x)$ $[\frac{\text{force}}{\text{length}}]$, then the potential energy of the string is

$$\text{P.E.} = \int \left[\frac{1}{2} T \left(\frac{d\psi}{dx} \right)^2 + \frac{1}{2} \kappa \psi^2 - F\psi \right] dx$$

(iii) Suppose we have a soap bubble under surface tension T $[\frac{\text{force}}{\text{length}} = \frac{\text{energy}}{\text{area}}]$. This means that the energy necessary to increase the separation between adjacent molecules in the bubble Δs. This separation increase refers to the change in the separation between two columns (both of length $\Delta y = \ell$) of adjacent molecules, a change from the the old preexisting (undeformed) separation to the new separation. This change is depicted in Figure 3.25.

As a result, the area subtended by the two adjacent columns in the bubble increases by

$$\Delta s \, \ell = \Delta x \left[\sqrt{1 + \left(\frac{d\psi}{dx} \right)^2} - 1 \right] \Delta y \ .$$

This is the area increase when the stretch is strictly into the x-direction. The rota-

tionally invariant generalization of this is

$$\Delta s\ell = \Delta x \Delta y \left[\sqrt{1 + \left(\frac{d\psi}{dx}\right)^2 + \left(\frac{d\psi}{dy}\right)^2} - 1 \right] .$$

It follows that the associated increase in potential energy is

$$\begin{aligned} \text{P.E.} &= \iint T \left[\sqrt{1 + \left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2} - 1 \right] dx\, dy \\ &\cong \frac{1}{2} \iint T \left[\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2 \right] dx\, dy \end{aligned}$$

(iv) Suppose we have a rod or a column of cross sectional area A and whose Young's modulus is Y $[\frac{\text{force}}{\text{area}}]$. The potential energy when the rod is in a compressed (or stretched) state is

$$\begin{aligned} \text{P.E.} &= A\tfrac{1}{2} \int Y \left(\frac{\partial\psi}{\partial x}\right)^2 dx \\ &= \frac{1}{2} \iiint Y \left(\frac{\partial\psi}{\partial x}\right)^2 dx\, dy\, dz \end{aligned}$$

(v) The kinetic energy in all these examples has the form

$$\begin{aligned} &\text{K.E.} = \tfrac{1}{2} \int \rho \left(\tfrac{\partial\psi}{\partial t}\right)^2 dx && \rho = \left[\tfrac{\text{mass}}{\text{length}}\right] \\ &\quad \text{or} \\ &\text{K.E.} = \tfrac{1}{2} \iint \rho \left(\tfrac{\partial\psi}{\partial t}\right)^2 dx\, dy && \rho = \left[\tfrac{\text{mass}}{\text{area}}\right] \\ &\quad \text{or} \\ &\text{K.E.} = \tfrac{1}{2} \iiint \rho \left(\tfrac{\partial\psi}{\partial t}\right)^2 dx\, dy\, dz && \rho = \left[\tfrac{\text{mass}}{\text{volume}}\right] \end{aligned}$$

(vi) The typical Lagrangian for a three-dimensional medium undergoing only small deformations has the form

$$L = \text{K.E.} - \text{P.E.} = \iiint \mathcal{L}\ dx\, dy\, dz$$

where

$$\mathcal{L} = \frac{1}{2}\rho \left(\frac{\partial\psi}{\partial t}\right)^2 - \frac{1}{2}k \left[\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2 + \left(\frac{\partial\psi}{\partial z}\right)^2 \right] - \frac{\kappa}{2}\psi^2 + F\psi$$

is the *Lagrangian density* of the system.

3.14.1 Variational Principle

Consider a general three-dimensional system whose state is characterized by the function $\psi(x, y, z, t)$. The dynamics of this system is governed by the extremum principle

$$I[\psi] \equiv \int_{t_0}^{t_1}\iiint_{\mathcal{V}} \mathcal{L}\left(t, x^i, \frac{\partial\psi}{\partial t}, \frac{\partial\psi}{\partial x^i}\right) dx^1\, dx^2\, dx^3\, dt$$
$$= \text{extremum!}$$

Here $\mathcal{L}$, the Lagrangian density of the system, may depend on time as well as the space coordinates $x^i, i = 1, 2, 3$. The integration is over the spatial volume $\mathcal{V}$ which is spanned by the coordinates x^1, x^2, and x^3.

Nature dictates that the actual performance (i.e. the evolution in time) of the system is that one which is optimal in the sense that it extremizes the variational integral $I[\psi]$.

This means that if $\psi(t, x^i)$ is such an optimal evolution, then the variational integral will be unchanged to first order if one evaluates this integral for a variant evolution, say $\psi(t, x^i) + \delta\psi(t, x^i)$. In other words

$$I[\psi + \delta\psi] - I[\psi] \equiv \delta I[\psi] = O[\delta\psi]$$

where $O[\delta\psi]$ is a quantity of second or higher order in $\delta\psi$. It is, however, understood that the optimal evolution and its variant agree initially at t_0 and finally at t_i

$$\psi(t_0, x^i) = \psi(t_0, x^i) + \delta\psi(t_0, x^i)$$
$$\psi(t_1, x^i) = \psi(t_1, x^i) + \delta\psi(t_1, x^i).$$

In other words the variation $\delta\psi(t, x^i)$ vanishes at the ends of the time interval,

$$\left.\begin{array}{l}\delta\psi(t_0, x^i) = 0\\ \delta\psi(t_1, x^i) = 0\end{array}\right\} \quad \forall\ (x^1, x^2, x^3) \in \mathcal{V}. \tag{3.51}$$

Furthermore it is also understood that the variants $\psi + \delta\psi = \psi$ on the boundary for all times $t_0 \le t \le t_1$. In other words, the variation $\delta\psi(t, x^i)$ satisfies the homogeneous boundary condition that for all times

$$\delta\psi(t, x^i) = 0 \quad \text{whenever} \quad (x^1, x^2, x^3) \in \partial\mathcal{V}. \tag{3.52}$$

Here $\partial\mathcal{V} \equiv$ boundary of $\mathcal{V}$.

Equations (3.51) and (3.52) can be restated by observing that $\delta\psi(t, x^i)$ is in general non-zero in the interior of the four-dimensional spacetime region $[t_0, t_1] \times \mathcal{V}$, but

$$\boxed{\delta\psi(t, x^i) = 0 \quad (t, x^1, x^2, x^3) \in \partial\big([t_0, t_1] \times \mathcal{V}\big)}$$

i.e. $\delta\psi = 0$ whenever (t, x^1, x^2, x^3) lies on the *boundary* of the four-dimensional spacetime cube $[t_0, t_1] \times \mathcal{V}$.

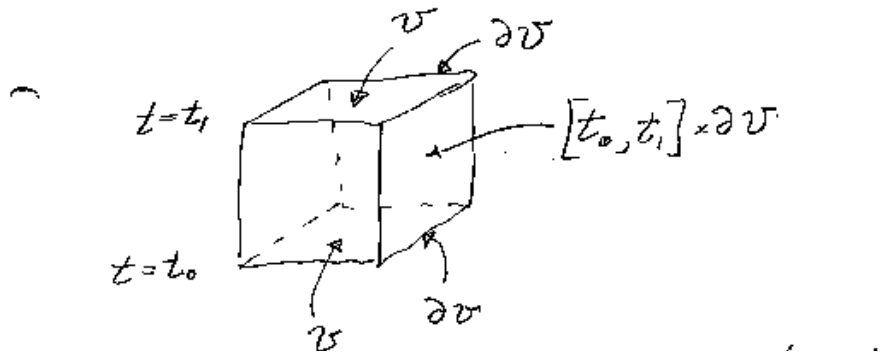


Figure 3.26: Four-dimensional spacetime cube and its boundary.

3.14.2 Euler-Lagrange Equation

Subject to these *boundary conditions* on the variants, the optimal evolution satisfies then the necessary condition

$$\delta I[\psi] = 0.$$

We now resort in our customary fashion to the familiar calculus of variations technique to obtain the Euler-Lagrange equation implied by this necessary condition.

The variation in the Lagrangian density due to the variation $\delta\psi$ is

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\psi}\delta\psi + \frac{\partial\mathcal{L}}{\partial\dot{\psi}}\delta\left(\tfrac{\partial\psi}{\partial t}\right) + \sum_{i=1}^{3}\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^i}\right)}\delta\left(\tfrac{\partial\psi}{\partial x^i}\right).$$

Hamilton's principle therefore becomes

$$0 = \delta I = \int_1^2\iiint_{\mathcal{V}}\Big\{\frac{\partial\mathcal{L}}{\partial\psi}\delta\psi + \underbrace{\frac{\partial\mathcal{L}}{\partial\dot{\psi}}\delta\left(\tfrac{\partial\psi}{\partial t}\right)}_{\boxed{1}} + \sum_{i=1}^{3}\underbrace{\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^i}\right)}\delta\left(\tfrac{\partial\psi}{\partial x^i}\right)}_{\boxed{2}}\Big\}d^3x\,dt.$$

Make use of the same integration by parts occurring in the derivation of the ordinary Euler-Lagrange equations:

$$\boxed{1} = \frac{\partial\mathcal{L}}{\partial\dot{\psi}}\underbrace{\left[\frac{\partial(\psi+\delta\psi)}{\partial t} - \frac{\partial\psi}{\partial t}\right]}_{\frac{\partial}{\partial t}(\delta\psi)} = -\int\frac{\partial}{\partial t}\left(\frac{\partial\mathcal{L}}{\partial\dot{\psi}}\right)\delta\psi\,dt.$$

The integral involving the spatial derivatives can be manipulated in a similar manner:

$$\boxed{2} = \int\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)}\left[\frac{\partial(\psi+\delta\psi)}{\partial x^k} - \frac{\partial\psi}{\partial x^k}\right]dx^k = \int\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)}\frac{\partial}{\partial x^k}(\delta\psi)\,dx^k$$

An integration by parts yields:

$$\left.\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)}\delta\psi\right|_{\partial\mathcal{R}} - \int \frac{\partial}{\partial x^k}\left[\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)}\delta\psi\right] dx^k. \qquad \underset{\text{no sum}}{k=1,2,3}$$

The integrated term vanishes because it is evaluated at the two end points of the x^k interval, and they lie on the boundary $\partial\mathcal{R}$.

The variation in I becomes therefore

$$\delta I = \int_1^2 \iiint_{\mathcal{R}} \left\{ \frac{\partial \mathcal{L}}{\partial\psi} - \frac{\partial}{\partial t}\left(\frac{\partial \mathcal{L}}{\partial\dot{\psi}}\right) - \sum_{k=1}^{3} \frac{\partial}{\partial x^k}\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)} \right\} \delta\psi \; dt\, d^3x.$$

The optimal nature of ψ demands that $\delta I = 0$ for arbitrary ψ, in particular those variations $\delta\psi$ which have a "blip" function behaviour around $(t, x^i) \in [t_0, t_1] \times \mathcal{V}$ as in Figure 1.11. Consequently the variational derivative of I,

$$\boxed{\frac{\delta I}{\delta\psi(t,x^i)} \equiv \frac{\partial \mathcal{L}}{\partial\psi} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial\dot{\psi}} - \sum_{k=1}^{3} \frac{\partial}{\partial x^k}\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^k}\right)} = 0},$$

must vanish.

Note: If we let

$$\left.\begin{array}{ll} x^0 & \equiv t \\ x^1 & = x^1 \\ x^2 & = x^2 \\ x^3 & = x^3 \end{array}\right\} x^\mu \qquad \mu = 0,1,2,3$$

then the Euler-Lagrange equation can be written in the more symmetric form

$$\boxed{\frac{\partial \mathcal{L}}{\partial\psi} - \sum_{\mu=0}^{3} \frac{\partial}{\partial x^\mu}\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial\psi}{\partial x^\mu}\right)} = 0}.$$

Comment.
(i) These Euler-Lagrange equations have a structure that puts the time on an equal status with the space coordinates.

Furthermore, the action integral

$$I[\psi] = \iiint\!\!\int \mathcal{L}\, d^4x$$

has been extremized on the set of all functions $\psi(x^\mu)$ which have fixed values on the 3-dimensional boundary of a given 4-dimensional region such as depicted in Figure 3.26

on page 161. It turns out that the general shape can be, within certain limits, fairly general.
(ii) If the action is a functional of several functions, so that

$$\mathcal{L} = \mathcal{L}\Big(x^\mu, \psi^1, \cdots, \psi^s, \frac{\partial \psi^1}{\partial x^\mu}, \cdots, \frac{\partial \psi^s}{\partial x^\mu}\Big),$$

then the Euler equations for the functional are

$$\boxed{\frac{\partial \mathcal{L}}{\partial \psi^n} - \sum_{\mu=0}^{3} \frac{\partial}{\partial x^k} \frac{\partial \mathcal{L}}{\partial\big(\frac{\partial \psi^n}{\partial x^k}\big)} = 0} \qquad n = 1, \ldots, s.$$

3.14.3 Examples and Applications

Example 1. *The Wave Equation.*

Consider the action for the wave equation,

$$I = \frac{1}{2} \iiiint \left[\left(\frac{\partial \psi}{\partial t}\right)^2 - \left(\frac{\partial \psi}{\partial x}\right)^2 - \left(\frac{\partial \psi}{\partial y}\right)^2 - \left(\frac{\partial \psi}{\partial z}\right)^2 \right] dt\, dx\, dy\, dz.$$

The E-L equation yields

$$\boxed{-\frac{\partial^2 \psi}{\partial t^2} + \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = 0},$$

the familiar wave equation.

Example 2. *The (Time Independent) Helmholtz equation.*

Consider the variational problem

$$\iiint \mathcal{F} dx\, dy\, dz = \iiint \left[\left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2 + \left(\frac{\partial \psi}{\partial z}\right)^2 \right] dx\, dy\, dz = \text{extremum}$$

for all those functions ψ which are constrained by the condition

$$\iiint \psi^2 dx\, dy\, dz = 1$$

This problem is solved by the method of Lagrange multipliers. Let the variational density be

$$\mathcal{F}' = \mathcal{F} - \lambda \psi^2.$$

The Euler equation for this integrand is

$$\frac{\partial \mathcal{F}'}{\partial \psi} - \sum_{i=1}^{3} \frac{d}{dx^i} \frac{\partial \mathcal{F}'}{\partial\left(\frac{\partial \psi}{\partial x^i}\right)} = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \lambda \psi = \nabla^2 \psi + \lambda \psi = 0,$$

which is the Helmholtz equation.

Example 2. *Helmholtz equation relative to arbitrary coordinates.*

Suppose we wish to rewrite this equation in terms of the new curvilinear coordinates

$$\begin{aligned} \bar{x}^1 &= \bar{x}^1(x^1, x^2, x^3) \\ \bar{x}^2 &= \bar{x}^2(x^1, x^2, x^3) \\ \bar{x}^3 &= \bar{x}^3(x^1, x^2, x^3) \end{aligned}$$

(a) In the intererst of conceptual unit-economy (a.k.a. informally as "crow epistemology"), instead of applying this transformation to the differential equation, apply it to the variational principle, and then write down the Euler equation relative to these new coordinates:

$$\begin{aligned} \mathcal{F} = \nabla\psi \cdot \nabla\psi &= \left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2 + \left(\frac{\partial \psi}{\partial z}\right)^2 \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} \Bigg(\underbrace{\frac{\partial \bar{x}^i}{\partial x^1} \frac{\partial \bar{x}^j}{\partial x^1} + \frac{\partial \bar{x}^i}{\partial x^2} \frac{\partial \bar{x}^j}{\partial x^2} + \frac{\partial \bar{x}^i}{\partial x^3} \frac{\partial \bar{x}^j}{\partial x^3}}_{\equiv g^{ij} \text{ where } g^{ij} g_{jk} = \delta^i_k} \Bigg) \frac{\partial \psi}{\partial \bar{x}^i} \frac{\partial \psi}{\partial \bar{x}^j}. \end{aligned}$$

It is easy to show that

$$\left[\frac{\partial \bar{x}^i}{\partial x^1} \frac{\partial \bar{x}^j}{\partial x^1} + \frac{\partial \bar{x}^i}{\partial x^2} \frac{\partial \bar{x}^j}{\partial x^2} + \frac{\partial \bar{x}^i}{\partial x^3} \frac{\partial \bar{x}^j}{\partial x^3} \right] \equiv \left[g^{ij} \right]$$

is the matrix inverse to $[g_{ij}]$, which defined the metric

$$\begin{aligned} \left(ds\right)^2 &= \left(dx^1\right)^2 + \left(dx^2\right)^2 + \left(dx^3\right)^2 \\ &= \delta_{ij}\, dx^i\, dx^j \\ &= \delta_{ij} \frac{\partial x^i}{\partial \bar{x}^k} \frac{\partial x^j}{\partial \bar{x}^\ell}\, d\bar{x}^k\, d\bar{x}^\ell \\ &= g_{k\ell}\, d\bar{x}^k\, d\bar{x}^\ell. \end{aligned}$$

In other words,

$$\boxed{\left[g^{ij} \right] \left[g_{jk} \right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}.$$

Thus relative to these new coordinates $\bar{x}^i$ one has

$$\boxed{\nabla\psi \cdot \nabla\psi = g^{ij}\frac{\partial\psi}{\partial\bar{x}^i}\frac{\partial\psi}{\partial\bar{x}^j}}$$

which is easy to write down once g_{ij} is known.

[FIGURE]

(b) The volume element relative to the new coordinates is expressed in terms of the Jacobian determinant J:

$$\begin{vmatrix} \frac{\partial x^1}{\partial\bar{x}^1}\,d\bar{x}^1 & \frac{\partial x^2}{\partial\bar{x}^1}\,d\bar{x}^1 & \frac{\partial x^3}{\partial\bar{x}^1}\,d\bar{x}^1 \\ \frac{\partial x^1}{\partial\bar{x}^2}\,d\bar{x}^2 & \frac{\partial x^2}{\partial\bar{x}^2}\,d\bar{x}^2 & \frac{\partial x^3}{\partial\bar{x}^2}\,d\bar{x}^2 \\ \frac{\partial x^1}{\partial\bar{x}^3}\,d\bar{x}^3 & \frac{\partial x^2}{\partial\bar{x}^3}\,d\bar{x}^3 & \frac{\partial x^3}{\partial\bar{x}^3}\,d\bar{x}^3 \end{vmatrix} = \det J\; d\bar{x}^1\,d\bar{x}^2\,d\bar{x}^3, \quad \text{where} \quad J = \left[\frac{\partial x^i}{\partial\bar{x}^j}\right].$$

This determinant is readily available from the transformed metric coefficients:

$$\begin{aligned} & & g_{k\ell} &= \tfrac{\partial x^i}{\partial\bar{x}^k} \quad \delta_{ij} \quad \tfrac{\partial x^k}{\partial\bar{x}^\ell} \\ & \text{or} & & \\ & & G &= J^T \quad I \quad J \\ & \text{so that} & & \\ & & \det G &= (\det J)^2 \\ & \text{or} & & \\ & & \det J &= \sqrt{\det G} \;\equiv\; \sqrt{g}\ . \end{aligned}$$

(c) The transformed variational integral is

$$\iiint g^{k\ell}\frac{\partial\psi}{\partial\bar{x}^k}\frac{\partial\psi}{\partial\bar{x}^\ell}\sqrt{g}\;d\bar{x}^1\,d\bar{x}^2\,d\bar{x}^3 = \text{extremum}$$

and the normalization condition is

$$\iiint \psi^2\sqrt{g}\;d\bar{x}^1\,d\bar{x}^2\,d\bar{x}^3 = 1.$$

(d) The Euler-Lagrange equations are

$$\frac{\partial}{\partial x^k}\left[\sqrt{g}\,g^{k\ell}\frac{\partial\psi}{\partial x^\ell}\right] + \lambda\sqrt{g}\,\psi = 0$$

or

$$\boxed{\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^k}\left[\sqrt{g}\,g^{k\ell}\frac{\partial\psi}{\partial x^\ell}\right] + \lambda\psi = 0}.$$

Comparing this equation with the one in example 2, we see that

$$\boxed{\nabla^2\psi = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^k}\left[\sqrt{g}\, g^{k\ell}\frac{\partial\psi}{\partial x^\ell}\right]}$$

is the expression for the Laplacian relative to curvilinear coordinates.

Example: *Find the Laplacian relative to spherical coordinates.*

Step 1: Write down the metric relative to these coordinates.

$$(ds)^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2$$

$$\sqrt{g} = \sqrt{\det g_{ij}} = \sqrt{r^4\sin^2\theta} = r^2\sin\theta$$

Step 2:

$$g^{ij} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{1}{r^2\sin^2\theta} \end{bmatrix}.$$

Thus

$$\nabla^2\psi = \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}r^2\sin\theta\frac{\partial\psi}{\partial r} + \frac{\partial}{\partial\theta}\frac{r^2\sin\theta}{r^2}\frac{\partial\psi}{\partial\theta} + \frac{\partial}{\partial\phi}\frac{r^2\sin\theta}{r^2\sin^2\theta}\frac{\partial\psi}{\partial\phi}\right]$$

$$\boxed{\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial\psi}{\partial r} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial\psi}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}}.$$

Chapter 4

DIRECT METHODS IN THE CALCULUS OF VARIATIONS

Lecture 22

We know how to minimize the functional $J[\psi] = \iiint \mathcal{L}(x^i, \psi, \frac{\partial \psi}{\partial x^i})\, d^3x$ indirectly. We write down the Euler equation, a differential equation for $\psi(x^i)$. Then we solve that equation by finding that function $\psi(x^i)$ which satisfies the conditions at the boundary. This can be a daunting task because what one needs to do is solve the differential equation in the *whole* region enclosed by the boundary, even if one wishes to find only an approximate solution.

The ultimate object of interest is a function which *extremizes* $J[\psi]$ and *not* necessarily a function which *satisfies the Euler equation.* Consequently, it is appropriate to ask:

> "Can one not find this optimizing function (if not exactly, then at least approximately) without having to bother with the intermediate expense of having to integrate the Euler equation over a finite multidimensional region?"

The answer to this question leads us from the *indirect variational* methods based on solving partial differential equations to *direct variational* methods which dispense with these equations altogether.

The direct methods have two additional attractive features:

(i) they can be used to solve differential equations and

(ii) they can be used to calculate a solution to the differential equation with arbitrary accuracy.

Thus suppose that one can show by some means that the functional $J[\psi]$ has an extremum for a sufficiently smooth admissible function ψ. Then this very fact proves the existence of a solution to the corresponding Euler differential equation, and thereby gives us a rough idea about the nature of the solution.

4.1 Minimizing Sequence.

The general idea behind the direct method is as follows.

Consider the problem of finding the minimum of a functional $J[y]$ on the space $\mathcal{M}$ of allowed ("admissible") functions. (This usually means the functions must be sufficiently differentiable.) We assume (a) for each of these functions the functional J is *finite*, i.e.

$$J[y] < \infty$$

and (b) $J[y]$ has a *minimum*, i.e. $\exists \mu$ such that

$$\inf_{y \in \mathcal{M}} J[y] = \mu > -\infty.$$

Here $\inf_{y \in \mathcal{M}} J[y]$ is the *greatest lower bound* ("infimum"=GLB) for the set $\{J[y] : y \in \mathcal{M}\}$; in other words

$$\inf_{y \in \mathcal{M}} J[y] = \max\{m : J[y] \geq m; y \in \mathcal{M}\}.$$

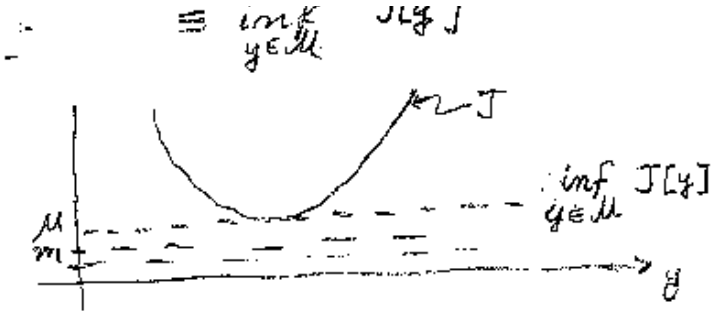


Figure 4.1: The greatest lower bound ("infimum") of the functional $J[y]$.

The existence of such a greatest lower bound implies that $\exists$ a sequence of functions in $\mathcal{M}$ (a "minimizing sequence")

$$y_1, y_2, \ldots$$

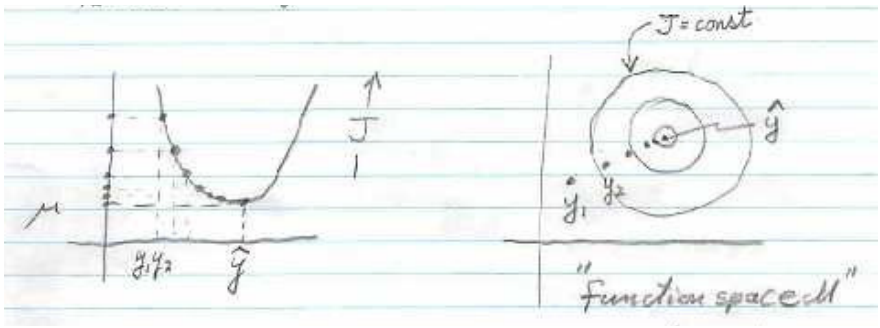


Figure 4.2: The greatest lower bound of J and a convergent sequence $\{y_n\}$ implied by it.

such that

$$\lim_{n\to\infty} J[y_n] = \mu.$$

If the sequence $y_1, y_2, \cdots$ has a limit, say $\hat{y}$:

$$\lim_{n\to\infty} y_n = \hat{y},$$

then one can write

$$\lim_{n\to\infty} J[y_n] = J[\hat{y}]$$
$$\lim_{n\to\infty} J[y_n] = J\Big[\lim_{n\to\infty} y_n\Big]$$

or $J[\hat{y}] = \mu$.

In that case $\hat{y}$ is the solution to the problem and the elements of the minimizing sequence y_n can be considered as approximations to the problem.

4.2 Implementation via Finite-Dimensional Approximation

An implementation of this idea is as follows: Consider the following sequence

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_n \subset \cdots \subset \mathcal{M}$$

of spaces of functions

$$\begin{array}{rcll}
\mathcal{M}_1 & = & \{ & y_1 = y_1(x^i, \alpha_1) \quad -\infty < \alpha_1 < \infty \ \} \\
\mathcal{M}_2 & = & \{ & y_2 = y_2(x^i, \alpha_1, \alpha_2) \quad -\infty < \alpha_1, \alpha_2 < \infty \\
 & & & y_2(x^i, \alpha_1, 0) = y_1(x^i, \alpha_1) \ \} \\
\mathcal{M}_3 & = & \{ & y_3 = y_3(x^i, \alpha_1, \alpha_2, \alpha_3) \quad -\infty < \alpha_1, \alpha_2, \alpha_3 < \infty \\
 & & & y_3(x^i, \alpha_1, \alpha_2, 0) = y_1(x^i, \alpha_1, \alpha_2) \ \} \\
 & \vdots & &
\end{array}$$

Thus each $\mathcal{M}_n$ is coordinatized by its n coordinates $(\alpha_1, \cdots, \alpha_n)$.

The functional J can be minimized on each subspace $\mathcal{M}_n$. Thus

(1) on the 1-dimensional space $\mathcal{M}_1$ consider $\min_{y \in \mathcal{M}_1} J[y] = \mu_1$. This is obtained by letting

$$J(\alpha_1) = J[y_1(\alpha_1)]$$

requiring

$$\frac{\partial J(\alpha_1)}{\partial \alpha_1} = 0,$$

and then letting $\mu_1 = J(\alpha_1)$.

(2) on the 2-dimensional space $\mathcal{M}_2$ consider $\min_{y \in \mathcal{M}_2} J[y] = \mu_2$. This is obtained by letting

$$J(\alpha_1, \alpha_2) = J[y_2(\alpha_1, \alpha_2)]$$

requiring

$$\frac{\partial J(\alpha_1, \alpha_2)}{\partial \alpha_i} = 0 \qquad i = 1, 2,$$

and then letting $\mu_2 = J(\alpha_1, \alpha_2)$.

$\vdots$

(n) on the n-dimensional space $\mathcal{M}_n$ consider $\min_{y \in \mathcal{M}_n} J[y] = \mu_n$. This is obtained by letting

$$J(\alpha_1, \ldots, \alpha_n) = J[y_n(\alpha_1, \ldots, \alpha_n)]$$

requiring

$$\frac{\partial J(\alpha_1, \ldots, \alpha_n)}{\partial \alpha_i} = 0 \qquad i = 1, \ldots, n,$$

and then letting $\mu_n = J(\alpha_1, \ldots, \alpha_n)$.

$\vdots$

This results in a non-increasing sequence of minima

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq \cdots \geq \inf_{y \in \mathcal{M}} J[y] = \hat{\mu}$$

because

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_n \subset \cdots \subset \mathcal{M}.$$

In other words, the minimum can never increase as we increase the size of the function space.

4.3 Rayleigh's Variational Principle

Let us apply the direct method to one of the most important variational problems, namely the one governing a linear vibrating system, say, a vibrating compressible medium enclosed in a three-dimensional volume:

$$\iiint (\nabla\psi) \cdot (\nabla\psi)\, dV = \text{extremum}$$

subject to the constraint

$$\iiint \psi^2\, dV = 1; \quad \psi = 0 \quad \text{on the boundary of the volume.}$$

This variational problem extremizes the potential energy of the system at a moment of time symmetry (see page 65 in Section 2.3.1). The Euler equation

$$\nabla^2\psi + \lambda\psi = 0 \tag{4.1a}$$

with the boundary condition

$$\psi = 0 \quad \text{on the boundary of the volume,} \tag{4.1b}$$

is what in one dimension corresponds to the Sturm-Liouville boundary value problem. Here we have its three-dimensional version.

The interesting feature about this variational problem is that it has many (discrete) solutions. In other words, we are asked to find not only a function which *minimizes* the variational integral, but also those functions which correspond to *saddle points* in the function space on which the integral is defined.

4.3.1 The Rayleigh Quotient

In order to help us see the relationship between these different solution and help us develop methods for determining them at least approximately, we shall use a powerful device which is called the Rayleigh Quotient after the person who invented it and made extensive use of it in analyzing vibrational problems.

The Rayleigh Quotient for our variational problem consists of

$$\frac{\iiint(\vec{\nabla}\psi)\cdot(\vec{\nabla}\psi)\,dV}{\iiint\psi^2\,dV}\equiv\frac{N}{D}.$$

The utility of this quotient lie in the fact that the variational problem ("Rayleigh's variational principle")

$$\frac{N}{D}=\text{extremum}$$

subject to

$$\psi=0\quad\text{on the boundary,}$$

is equivalent to the original variational problem, Eqs. (4.1). This is a big advantage because Rayleigh's variational problem permits us to consider variants of ψ which are arbitrary instead of being restricted by the constraint condition

$$\iiint\psi^2\,dV=1.$$

The *first* significant property of the Rayleigh quotient lies in the fact that if ψ satisfies the eigenvalue ("Helmholtz") equation

$$\nabla^2\psi=-k^2\psi,$$

then

$$\boxed{\frac{N}{D}=k^2}.$$

Indeed, using Gauß's divergence theorem and the boundary condition that $\psi=0$ on the boundary we find

$$\begin{aligned}\frac{N}{D}&=\frac{\iiint(\nabla\psi)\cdot(\nabla\psi)\,dV}{\iiint\psi^2\,dV}\\&=\frac{\iiint[\nabla\cdot(\psi\vec{\nabla}\psi)-\psi\nabla\cdot\vec{\nabla}\psi]\,dV}{\iiint\psi^2\,dV}\\&=\frac{-\iiint\psi\nabla^2\psi\,dV+\iint\psi\vec{\nabla}\psi\cdot\vec{n}\,dA}{\iiint\psi^2\,dV}\\&=\frac{-\iiint\psi(-k^2\psi)\,dV+0}{\iiint\psi^2\,dV}\\&=k^2.\end{aligned}$$

Remark. If ψ satisfies the homogeneous boundary condition

$$\vec{n} \cdot \vec{\nabla}\psi + f\psi = 0,$$

then the Rayleigh quotient is

$$\frac{N}{D} = \frac{\iiint (\nabla\psi)^2 \, dV + \iint f\psi^2 \, dA}{\iiint \psi^2 \, dV} \tag{4.2}$$

instead. But we still have the basic fact,

$$\left.\begin{array}{rl} \nabla^2\psi + k^2\psi & = 0 \\ \vec{n} \cdot \vec{\nabla}\psi + \psi\big|_{\text{boundary}} & = 0 \end{array}\right\} \quad \text{implies} \quad \frac{N}{D} = k^2.$$

The *second* significant property of the Rayleigh quotient is that extremizing it is equivalent to the original extremum principle; in other words, if ψ is an admissible function (i.e. it satisfies, among others, the boundary conditions) which extremizes the Rayleigh quotient, then this function ψ satisfies the Helmholtz equation. More briefly, we have

$$\frac{N}{D} = \text{extremum} \qquad \psi\big|_{\text{boundary}}$$

implies that

$$\nabla^2\psi + k^2\psi = 0$$

where

$$k^2 = \frac{N}{D}.$$

The variational calculation leading to this conclusion is based on the Rayleigh variational principle

$$\delta\left(\frac{N}{D}\right) = 0. \qquad \text{("Rayleigh variational principle")}$$

The details are easy to perform:

$$\delta\left(\frac{N}{D}\right) = \frac{\delta N}{D} - \frac{N}{D}\frac{\delta D}{D}.$$

Using the fact that $\delta\psi = 0$ on the boundary, one has

$$\begin{aligned}
\delta N &= \delta \iiint \nabla\psi \cdot \nabla\psi \, dV \\
&= 2 \iiint \nabla\psi \cdot \nabla\delta\psi \, dV \\
&= 2 \iiint \left[\nabla \cdot (\delta\psi \, \vec{\nabla}\psi) - \nabla \cdot \vec{\nabla}\psi \, \delta\psi \right] dV \\
&= -2 \iiint (\nabla^2\psi) \, \delta\psi \, dV + 2 \iint \vec{n} \cdot \nabla\psi \, \delta\psi \, dA \\
&= -2 \iiint (\nabla^2\psi) \, \delta\psi \, dV + 0 \\
\delta D &= 2 \iiint \psi \, \delta\psi \, dV
\end{aligned}$$

Thus

$$0 = \delta\left(\frac{N}{D}\right) = -\frac{2}{D} \iiint \left[\nabla^2\psi + \frac{N}{D}\psi \right] \delta\psi \, dV.$$

Inside the integration volume, the variation $\delta\psi$ is arbitrary. Consequently,

$$\nabla^2\psi + \frac{N}{D}\psi = 0$$

or

$$\nabla^2\psi + k^2\psi = 0$$

if we set

$$k^2 = \frac{N}{D}.$$

Remark. If the trial functions ψ are required to satisfy the homogeneous boundary condition

$$\vec{n} \cdot \nabla\psi + f\psi\big|_{\text{boundary}} = 0$$

on the boundary, then Rayleigh's variational principle still holds provided one uses the corresponding Rayleigh quotient, Eq. (4.2).

Remark. $\delta\left(\frac{N}{D}\right) = 0$ implies the original variational principle, Eqs. (4.1): $\delta(N - k^2 D) = 0$. The *third* significant property of the Rayleigh quotient is that the variational principle based on it reveals the relationship between the various solutions which extremize it.

1. We note that when

$$k^2 = \frac{\iiint (\nabla\psi)^2 \, dV}{\iiint \psi^2 \, dV}$$

is evaluated for all possible trial functions ψ, there will exist an *absolute minimum*, k_0^2. Let ψ_0 be that optimal trial function corresponding to this minimum:

$$\psi_0 \longleftrightarrow k_0^2$$

2. It is not difficult to show that: If k_i^2 and k_j^2 are two eigenvalues of $(\nabla^2 + k^2)\psi = 0$, then

$$k_i^2 \neq k_j^2 \Rightarrow \iiint \psi_i \, \psi_j \, dV = 0$$

2b. Thus one may obtain a variational principle for k_1^2, where $k_1^2 > k_0^2$ but $k_n^2 > k_1^2 \quad (n > 1)$. One simply imposes the additional constraint

$$\iiint \psi \, \psi_0 \, dV = 0$$

on the set of trial functions. Then

$$\frac{\iiint (\nabla\psi)^2 \, dV}{\iiint \psi^2 \, dV} = \text{minimum}$$

subject to

$$\iiint \psi \, \psi_0 \, dV = 0$$

yields

$$k_1^2 \qquad \text{and} \qquad \psi_1.$$

This process may be continued to obtain k_2^2 and ψ_2, and other eigenfunctions and eigenvalues:

$$k_0^2 \leq k_1^2 \leq k_2^2 \leq \cdots$$

$$\psi_0, \quad \psi_1, \quad \psi_2, \quad \cdots$$

It is therefore evident that the Rayleigh quotient reveals:

(a) the *existence* of an *infinite* sequence of eigenvalues and eigenfunctions of the boundary value problem

$$\nabla^2\psi + k^2\psi = 0$$

(b) the sequence of eigenvalues is *non-decreasing*

(c) the sequence of functions which extremize the variational integral are *orthogonal*.

4.3.2 Rayleigh-Ritz Principle

The Rayleigh variational principle based on the Rayleigh quotient furnishes us with a framework for an *approximate determination* of the eigenvalues and their eigenfunctions. This framework is a variational method which consists of letting the trial function ("variant")

$$\phi = \phi(x, \alpha_1, \ldots, \alpha_s)$$

depend on a finite number of parameters $\alpha_1, \ldots, \alpha_s$. The dependence of ϕ on α_i may be a non-linear one or a linear one.

The Rayleigh-Ritz Method is a special method. It consists of letting the trial functions depend *linearly* on these parameters,

$$\phi = \sum_{i=1}^{s} \alpha_i \phi_i(x).$$

Let us introduce the following matrix elements:

$$\int \phi_i \, \phi_j \, dV \equiv B_{ij}$$

and

$$\int \vec{\nabla}\phi_i \cdot \vec{\nabla}\phi_j \, dV = -\int \phi_i \nabla^2 \phi_j \, dV \equiv A_{ij}.$$

We now take advantage of the fact that Rayleigh's variational principle

$$\delta\left(\frac{N}{D}\right) = 0$$

implies

$$\delta(N - k^2 D) = 0.$$

This is the condition for extremizing the expression

$$N - k^2 D = \sum_{i=1}^{s}\sum_{j=1}^{s}(A_{ij} - k^2 B_{ij})\alpha_i \alpha_j \equiv J(\alpha_1, \ldots, \alpha_s).$$

Thus we have

$$\frac{\partial J}{\partial x^i} = 0 \qquad i = 1, \ldots, s.$$

This results in

$$2\sum_{j=1}^{s}(A_{ij} - k^2 B_{ij})\alpha_j = 0 \qquad i = 1, \ldots, s$$

or

$$[A]\,\vec{\alpha} = k^2\,[B]\,\vec{\alpha}.$$

These equations have a solution if

$$\det\left|A_{ij} - k^2 B_{ij}\right| = 0.$$

This is the familiar secular equation for the symmetric matrices $[A_{ij}]$ and $[B_{ij}]$. The solutions are eigenvalues which approximate those determined by the problem

$$(\nabla^2 + k^2)\psi = 0,$$

plus homogeneous Dirichlet or Neumann boundary conditions.

The solution consists of an approximation to the first s eigenvalues:

$$k_0^2 \le k_1^2 \le \cdots \le k_s^2$$

and their corresponding approximate solutions

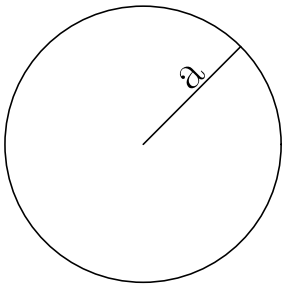
$$\phi^{(0)} = \sum \alpha_i^{(0)}\phi_i(x), \quad \ldots, \quad \phi^{(s))} = \sum \alpha_i^{(s)}\phi_i(x)$$

where $\alpha_i^{(m)}$ is an eigenvalue for the problem

$$[A]\,\vec{\alpha}^{(m)} = (k_m)^2\,[B]\,\vec{\alpha}^{(m)}.$$

4.3.3 Vibration of a Circular Membrane

Consider a clamped circular membrane of radius a.



The variational principle for this problem is

$$[k^2] = \frac{\iiint (\nabla\psi)^2\,dV}{\iiint \psi^2\,dV} \equiv \frac{N}{D}.$$

The amplitude ψ satisfies

$$\psi(a) = 0$$

with ψ and $\nabla\psi$ *continuous* inside the disc. For algebraic simplicity, consider *only circularly symmetric* modes. Thus

$$\psi = \psi(r)$$

is a function of r only, independent of angle.

The variational principle becomes

$$k^2 = \frac{\int_0^a (\frac{\partial\psi}{\partial r})^2 \, r \, dr}{\int_0^a \psi^2 \, r \, dr} \equiv \frac{N}{D}.$$

Comment: The differential equation corresponding to this problem is

$$\frac{1}{r}\frac{d}{dr} r \frac{d\psi}{dr} + k^2\psi = 0$$

whose eigenvalues k^2 are determined by $J_0(ka) = 0$. It is advantageous to introduce the dimensionless independent variable

$$x \equiv \frac{r}{a}.$$

Consequently

$$(ka)^2 = \frac{\int_0^1 (\frac{\partial\psi}{\partial x})^2 \, x \, dx}{\int_0^1 \psi^2 \, x \, dx} \equiv \frac{N}{D}.$$

(A) *Lowest* eigenvalue λ_0 and its eigenfunction $\phi_1^{(0)}$: Consider the possible trial functions. The simplest one which vanishes at $x = 1$ and which has *continuous gradient at* $x = 0$ is

$$\phi_1^{(0)} = 1 - x^2 \quad \text{for the lowest eigenvalue in the } 1^{st} \text{ approximation.}$$

Comment: $1 - x$ won't do because its slope at $x = 0$ is -1, giving rise to a cusp at the center.

Insert $\phi_1^{(0)}$ into the above variational principle and obtain λ_0, the lowest eigenvalue in the 1^{st} approximation:

$$\lambda_0 = (ka)^2 = \frac{\int_0^1 (-2x)^2 \, x \, dx}{\int_0^1 (1-x^2)^2 \, x \, dx} = \frac{\int_0^1 4x^3 \, dx}{-\frac{(1-x^2)^2}{6}\Big|_0^1} = \frac{x^4\big|_0^1}{\frac{1}{6}} = 6.$$

The *exact* value of $(ka)^2$ is determined by the first zero of J_0:

$$J_0(ka) = 0 \Rightarrow \begin{array}{cc} \text{Lowest root} & \text{Next lowest root} \\ k_0 a = 2.40483; & k_1 a = 5.52008 \\ (k_0 a)^2 = 5.78319; & (k_1 a)^2 = 30.4713 \end{array}$$

Thus the error in the crude variational estimate for the lowest eigenvalue is

$$\frac{6 - 5.78319}{6} = \frac{0.21681}{6} \approx 0.036, \qquad \text{i.e.} \qquad 3.6\%$$

(B) The next approximation uses an improved trial function[1]

$\phi_2 = \alpha_1 \left(1 - x^2\right) + \alpha_2 \left(1 - x^2\right)^2$ for the lowest eigenvalue in the 2^{nd} approximation.

This choice is based on the fact that any function which (a) is zero at $x = 1$ and (b) has zero slope at $x = 0$ can be expanded in a *power series in* $(1 - x^2)$. *It is observations like these about the (geometrical) nature of trial functions which makes the variational principle practical.*

Insert this trial function into the expression for $(ka)^2$ and obtain

$$\frac{N}{D} = (ka)^2 \equiv \lambda = \frac{\vec{\alpha}^T A \, \vec{\alpha}}{\vec{\alpha}^T B \, \vec{\alpha}}$$

or

$$A_{11}\alpha_1^2 + 2A_{12}\alpha_1\alpha_2 + A_{22}\alpha_2^2 = \lambda \left(B_{11}\alpha_1^2 + 2B_{12}\alpha_1\alpha_2 + B_{22}\alpha_2^2\right).$$

(*Nota bene:* The matrices A and B are only two-dimensional because they mathematize the 2^{nd} approximation. Higher order approximations would be mathematized by higher dimensional matrices.)

[1] In this context the convention is that the superscript "(0)" refers to the eigenvalue, and the subscript "2" refers to the approximation , i.e. the dimension of the approximating subspace, which now also accommodates the eigenfinction $\phi_2^{(1)}$ of the next eigenvalue λ_1 in the 2^{nd} approximation.

Here

$$B_{11} = \int_0^1 \left(1-x^2\right)^2 x\,dx = \frac{1}{6}$$
$$B_{12} = \int_0^1 \left(1-x^2\right)^3 x\,dx = \frac{1}{8}$$
$$B_{22} = \int_0^1 \left(1-x^2\right)^4 x\,dx = \frac{1}{10}$$
$$A_{11} = \int_0^1 \left[\frac{d}{dx}(1-x^2)\right]^2 x\,dx = \frac{1}{6}$$
$$A_{12} = \int_0^1 \frac{d}{dx}\left(1-x^2\right)\frac{d}{dx}\left(1-x^2\right)^2 x\,dx = \frac{2}{3}$$
$$A_{22} = \int_0^1 \left[\frac{d}{dx}\left(1-x^2\right)^2\right]^2 x\,dx = \frac{2}{3}.$$

Differentiate with respect to α_1 and α_2 and set

$$\frac{\partial}{\partial\alpha_1}\left(\frac{N}{D}\right) = \frac{\partial(ka)^2}{\partial\alpha_1} = 0$$
$$\frac{\partial}{\partial\alpha_2}\left(\frac{N}{D}\right) = \frac{\partial(ka)^2}{\partial\alpha_2} = 0$$

to obtain

$$(A_{11}-\lambda B_{11})\alpha_1 + (A_{12}-\lambda B_{12})\alpha_2 = 0$$
$$(A_{21}-\lambda B_{21})\alpha_1 + (A_{22}-\lambda B_{22})\alpha_2 = 0\ .$$

This is a pair of linear equations for α_1 and α_2. It has a solution only if $\det|\lambda B - A| = 0$, i.e. if

$$\det|\lambda B - A| = 0 \Rightarrow \begin{vmatrix} \frac{1}{6}\lambda - 1 & \frac{1}{8}\lambda - \frac{2}{3} \\ \frac{1}{8}\lambda - \frac{2}{3} & \frac{1}{10}\lambda - \frac{2}{3} \end{vmatrix} = 0.$$

This is a quadratic equation for $\lambda = (ka)^2$. There are two roots. The *smaller* root is

$$\lambda_0 = (ka)^2 = 5.78413 \quad \text{(lowest eigenvalue in the } 2^{st} \text{ approximation)},$$

which is very close to the exact value

$$\lambda_{\text{exact}} = (ka)^2_{\text{exact}} = 5.78319.$$

The ratio $\frac{\alpha_2}{\alpha_1}$ is, from the second equation,

$$\frac{\alpha_2}{\alpha_1} = -\frac{\frac{2}{3} - \frac{1}{8}\lambda}{\frac{2}{3} - \frac{1}{10}\lambda} = 0.639.$$

Thus the wave function is

$\phi_2^{(0)} \propto (1-x^2)+0.639\left(1-x^2\right)^2$ for the lowest eigenvalue in the 2^{st} approximation.

One should compare the first trial function

$$\phi_1^{(0)} = 1 - x^2$$

and the second trial function

$\phi_2^{(0)} = \dfrac{1}{1.639}\left[\left(1-x^2\right)+0.639\left(1-x^2\right)^2\right]$ for the lowest eigenvalue in the 2^{st} approximation

with the exact wave function

$$J_0(2.40483x). \leftarrow \text{exact}$$

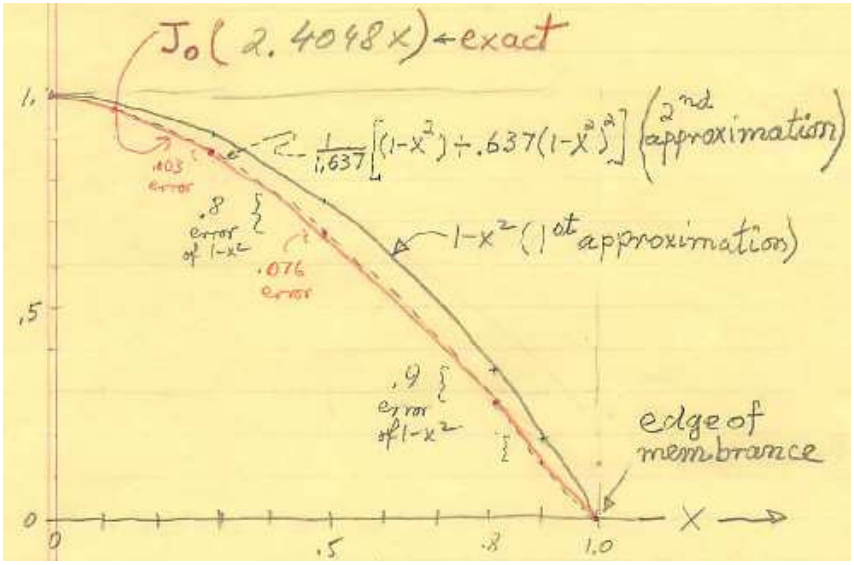


Figure 4.3: Exact vs. 1^{st} and 2^{nd} approximations to the lowest radial vibrational normal mode amplitudes.

Comment:

1. The larger second root of the secular quadratic equation is $\lambda_1 = (k_1 a)^2 = 36.883$, to be compared with the exact result $(k_1 a)^2_{\text{exact}} = 30.471$.

2. The approximate wave corresponding to this eigenvalue is orthogonal to the wave function determined by the first root.

$$\boxed{\lambda_1 = 36.883} \Rightarrow \frac{\alpha_2}{\alpha_1} = -\frac{\frac{2}{3} - \frac{\lambda}{8}}{\frac{2}{3} - \frac{\lambda}{10}}$$
$$= -\frac{\frac{2}{3} - 4.6103}{\frac{2}{3} - 3.6883} = -\frac{3.9437}{3.0216}$$
$$= -1.30516$$

The wave function is therefore

$$\phi_2^{(1)} = (1 - x^2) - 1.30516\left(1 - x^2\right)^2.$$

This wave function is orthogonal to $\phi_2^{(0)}$:

$$\int_0^1 \phi_2^{(0)}\, \phi_2^{(1)}\, x\, dx = 0$$

in the 2-d approximation subspace.

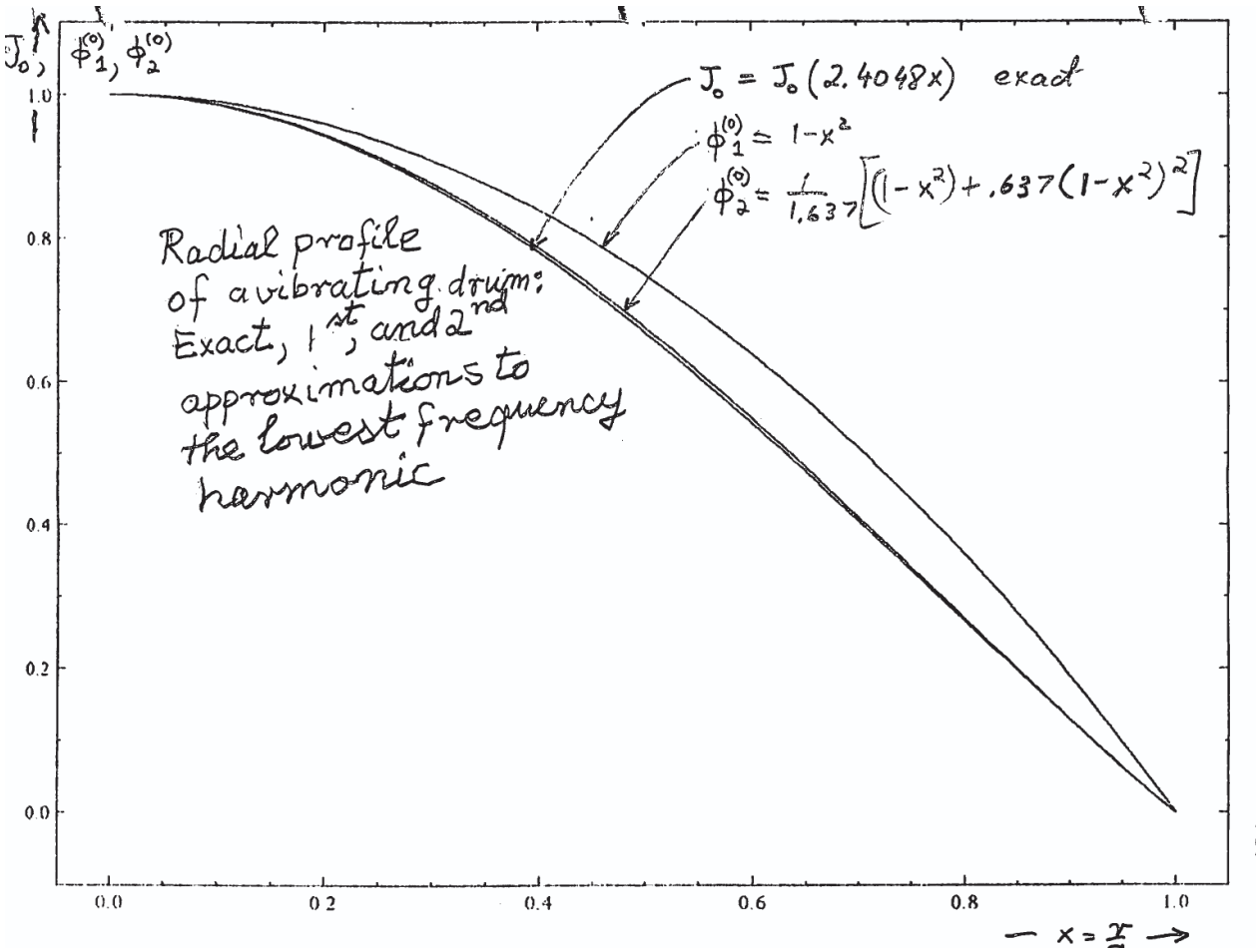


Figure 4.4: Graphs of a vibrating drum's lowest frequency radial amplitude profiles: in its mathematically exact form, in its first and its second Rayleigh-Ritz approximation.

Chapter 5

TENSOR CALCULUS

Tensor calculus unifies linear geometrical structures on a vector space with multivariable differential and integral calculus. It mathematizes the unification of the concept of linearity with that of change.

A geometrical structure on a vector space is a real-valued multilinear map (or function) which exists in numerical form relative to some standard basis but may exist relative to any basis of the vector space. In other words, the multilinear map is independent of the particular basis, which serves as a (contextually defined) standard by which vectors are specified quantitatively.

The unification is not a mere juxtaposition of two items; it is an integration, a mathematical blending of the two into a new mental entity, a *tensor field* on an underlying manifold.

5.1 The Dual of a Vector Space

A multilinear map is formed from two types of fundamental building blocks, elements of the vector space and covectors, the elements of its dual space. For finite dimensions these two spaces always exist in pairs. The existence of one always implies the other. It is not difficult to find a driving force conceptualizing "the dual of a given vector space".

Lecture 22

What facts of reality give rise to the concept "the dual of a given vector space"? Why do we need such a concept?

I. Consider the familiar circumstance, the set of fruit inventories in a supermarket. Designate a typical inventory of α^a apples, α^b bananas, α^c coconuts, $\cdots$ by

$$\vec{x} = \alpha^a \overrightarrow{apples} + \alpha^b \overrightarrow{bananas} + \alpha^c \overrightarrow{coconuts} + \cdots \tag{5.1}$$

or, more succinctly, by

$$\vec{x} = \alpha^a \vec{e_a} + \alpha^b \vec{e_b} + \alpha^c \vec{e_c} + \cdots \tag{5.2}$$

Fruit inventories like these form a vector space, V, closed under addition and scalar multiplication[1].
Nota bene: In this vector space there is no Pythagorean theorem, no distance function, and no angle between any pair of inventories. This is because in the present context apples, bananas, and coconuts are in-commensurable, i.e. "one does not mix apples and bananas".

II. Next consider a particular purchase price function, $\$_f$. Its values yield the cost of any fruit inventory $\vec{x}$:

$$\$_f(\vec{x}) = \$_f(\alpha^a \vec{e_a} + \alpha^b \vec{e_b} + \alpha^c \vec{e_c} + \cdots) \tag{5.3}$$

$$= \alpha^a \underbrace{\$_f(\vec{e_a})}_{\substack{\text{purchase} \\ \text{price/apple}}} + \alpha^b \underbrace{\$_f(\vec{e_b})}_{\substack{\text{purchase} \\ \text{price/banana}}} + \alpha^c \underbrace{\$_f(\vec{e_c})}_{\substack{\text{purchase} \\ \text{price/coconut}}} + \cdots \tag{5.4}$$

Also consider the cost of fruit inventory $\vec{y}$:

$$\$_f(\vec{y}) = \beta^a \$_f(\vec{e_a}) + \beta^b \$_f(\vec{e_b}) + \beta^c \$_f(\vec{e_c}) + \cdots \ . \tag{5.5}$$

We see that

$$\$_f(\vec{x} + c\vec{y}) = \$_f(\vec{x}) + c\$_f(\vec{y}) \tag{5.6}$$

Thus

$$\begin{aligned} \$_f : \ & V \to R \\ & \vec{x} \rightsquigarrow \$_f(\vec{x}) \end{aligned}$$

is a *linear* function on the vector space V of fruit inventories.

III. The set of purchase price functions forms a vector space. Indeed, consider the purchase price functions $\$_f$, $\$_g$, $\$_h$, $\cdots$ of different fruit wholesalers, and introduce the *combined purchase price function* $\$_f + \$_g$ by the requirement that

$$(\$_f + \$_g)(\vec{x}) = \$_f(\vec{x}) + \$_g(\vec{x}) \ \text{ for all } \vec{x} \text{ in } V \tag{5.7}$$

and $c\$_f$, the c-multiple of $\$_f$, by

$$(c\$_f)(\vec{x}) = c\,(\$_f(\vec{x})) \ .$$

We infer that the set of purchase price functions forms a vector space, V^*, the space *dual* to V.

[1]We are taking it for granted that the supermarket also carries inventories of (dried) cut-up apples, (dried) fractional bananas, and shredded amounts of whole coconuts.

5.2 Linear Functions

The space of duals is variously referred to as
the space of linear functions
the space of linear functionals
the space of covectors
the space of duals.

The concept of a dual is a new concept. It combines two concepts into one. It is a marriage between the concept of a function and the concept of a vector. Such a marriage is possible only for linear functions, a fact formalized by Theorem 1 on page 189.

In the hierarchy of concepts a dual is a derived concept, it depends on the existence and knowledge of the entities that make up a vector space. A dual is an abstraction which conceptualizes a measurable property of these entities. For example, if one introduces a basis for the vector space, then each of the associates coordinate functions is a dual. This fact is depicted in Figure 5.2 on page 192. Each one is a measurable property of a vector, with the relevant basis vector serving as the relevant measurement standard. Properties such as these, and others, are mathematized by means of linear functions which are identified by the following

Definition 1. *(Linear Function)*
Let V be a vector space. Consider a scalar-valued linear function *f defined on V as follows:*

$$\begin{aligned} f: \quad & V \to R \\ & x \rightsquigarrow f(x) \end{aligned}$$

such that

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \text{ where } x, y \in V;\ \alpha, \beta \in \{scalars\}\ .$$

Example 1: (Linear function as a row vector)
Let $V = R^n = \{(x_1, x_2, \cdots, x_n)\}$ then

$$\begin{array}{rccll} f: & R^n & \to & R & \\ & \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} & \rightsquigarrow & f(x_1, \cdots, x_n) & \equiv [\xi_1, \cdots, \xi_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ & & & & = \xi_1 x_1 + \cdots + \xi_n x_n \end{array}$$

is a linear function whose domain is $V = R^n$, and whose form is determined by a given row vector $[\xi_1, \cdots, \xi_n]$. See Figure 5.1.

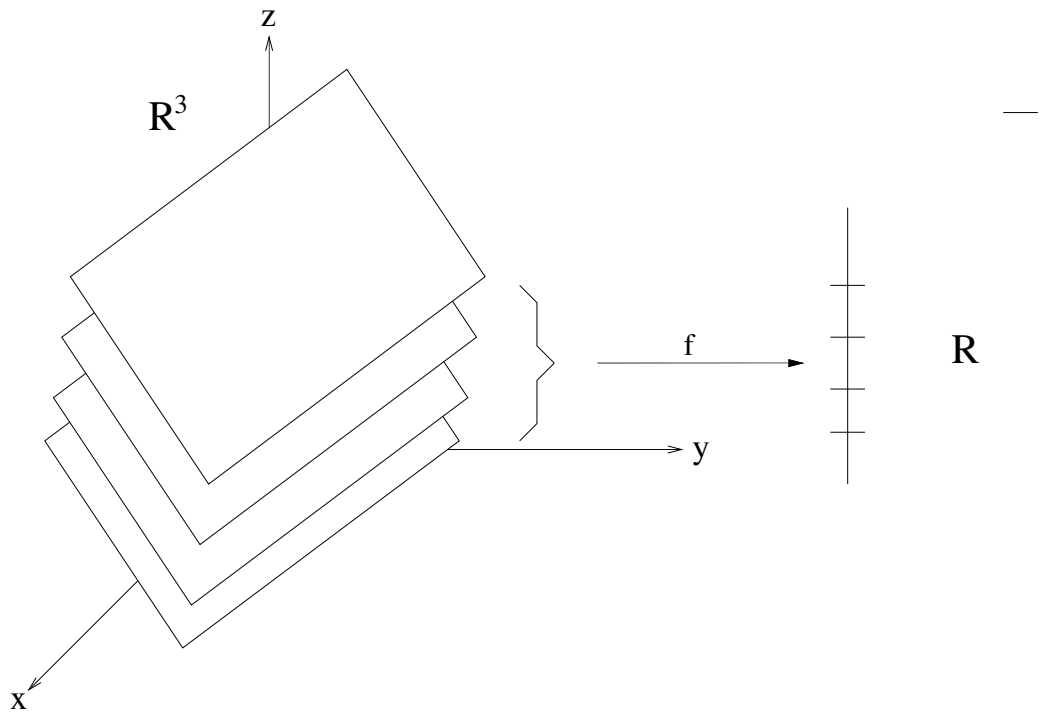


Figure 5.1: A linear map f. Its *isograms* (= loci of point where f has constant value) are parallel planes in its domain, R^3, with the isogram defined by $f = 0$ passing through the origin. The image of these planes are points in R

Example 2: (Weighted Sum of Samples) Consider $V = C[a, b]$, the vector space of functions continuous on the closed interval $[a, b]$:

$$V = \{\psi : \psi(s) \text{is continuous on } [a, b]\} \equiv C[a, b] \ .$$

Consider the following linear scalar-valued functions on V:

1. For any point $s_1 \in [a, b]$ the "s_1-evaluation map" (also known as the "s_1-sampling function") f

$$\begin{array}{rll} V = C[a, b] & \xrightarrow{f} R & \text{(reals)} \\ \psi & \rightsquigarrow f(\psi) = \psi(s_1) & \end{array}$$

   The function f is *linear* because

$$f(c_1\psi + c_2\phi) = c_1 f(\psi) + c_2 f(\phi) \tag{5.8}$$

2. Let $\{s_1, s_2, \cdots, s_n\} \subset [a,b]$ be a specified collection of points in $[a,b]$, and let $\{k_1, k_2, \cdots, k_n\} \equiv \{k_j\}_{j=1}^n$ be a set of scalars. Then the function g defined by

$$\begin{array}{rl} C[a,b] & \xrightarrow{g} R \\ \psi & \rightsquigarrow g(\psi) = \sum_{j=1}^n k_j \psi(s_j) \end{array}$$

is a linear function on $V = C[a,b,]$

3. Similarly, the map h defined by

$$h(\psi) = \int_a^b \psi(s)\, ds$$

is also a linear map on $v = C[a,b]$.

Example 3:
Consider the vector space of infinitely differentiable functions,

$$V = \{\psi : \psi \text{ is } C^\infty \text{ on } (a,b)\} \equiv C^\infty(a,b)$$

on (a,b). Furthermore, let

$$d^j\psi(s) = \left.\frac{d^j\psi}{ds^j}\right|_{x=s}$$

be the j^{th} derivative of ψ at $x = s$. Then for any *fixed* $s \in (a,b)$,the map h defined by

$$\begin{array}{rl} V = C[a,b] & \xrightarrow{h} R \\ \psi & \rightsquigarrow h(\psi) = \sum_{j=1}^n a_j d^j\psi(s) \end{array}$$

is a linear function on $V = C^\infty(a,b,)$

5.3 The Vector Space V^* Dual to V

Given a vector space V, the consideration of all possible linear functions defined on V gives rise to

$$V^* = \text{ set of all linear functions on V .}$$

These linear functions form a vector space in its own right, the *dual space* of V. Indeed, we have the following

Theorem 3.
The set V^* is a vector space.

Comment and proof:

1. In physics the elements of the vector space V^* are called *covectors*.

2. That V^* does indeed form a vector space is verified by observing that the collection of linear functions satisfies the familiar ten properties of a vector space.

   Thus, if f, g, h are linear functions and $\alpha, \beta \in R$, then

   (a) $f + g$ is also a linear function defined by the formula

   $$(f + g)(\vec{x}) = f(\vec{x}) + g(\vec{x}) \quad \forall \vec{x} \in V \ .$$

   Consequently, the following concomitant properties are satisfied automatically:

   (b) $f + g = g + f$

   (c) $f + (g + h) = (f + g) + h$

   (d) the zero element $\underline{0}$ (="additive identity") is the constant zero function

   (e) the additive inverse of f is $-f$.

   Furthermore,

   (i) αf is also a linear function defined by the formula

   $$(\alpha f)(\vec{x}) = \alpha f(\vec{x}) \quad \forall \vec{x} \in V \text{ and } \alpha \in R$$

   In light of this formula the following properties are also satisfied automatically:

   (ii) $\alpha(\beta f) = (\alpha\beta) f$

   (iii) $1f = f$

   (iv) $\alpha(f + g) = \alpha f + \alpha g$

   (v) $(\alpha + \beta) f = \alpha f + \beta f$

5.4 Dirac's Bracket Notation

To emphasize the duality between the two vector spaces, one takes advantage of Dirac's *bra-ket notation*, which he originally introduced into quantum mechanics.

If f is a linear function on V and $f(x)$ is its value at $x \in V$, then one also writes

$$f(x) \equiv \langle f|x\rangle \equiv \langle \underline{f}|\vec{x}\rangle \tag{5.9}$$

Thus the underscore under f is a reminder that $\underline{f} \in V^*$, while x, or better $\vec{x}$ is an element of V. We say that f *operates* on the vector x and produces

$$\langle f|x\rangle$$

To emphasize that f is a linear "machine", we write

$$f = \langle \underline{f}| \quad (\in V^*) \tag{5.10}$$

for the covector (which Dirac called a bra) and

$$x = |\vec{x}\rangle \quad (\in V) \tag{5.11}$$

for the vector (which Dirac called a ket). They combine to form

$$\langle \underline{f}|\vec{x}\rangle \quad (\in R)$$

5.5 The Duality Principle

Lecture 23

As we shall see, mathematically it is the existence and uniqueness of a vector's scalar coefficients relative to a chosen (or given) basis that makes the concept of duality so important. Indeed, such a basis makes the introduction of V^* inevitable. This is because a basis determines unique scalar values for each vector, which is to say that it determines scalar functions on V.

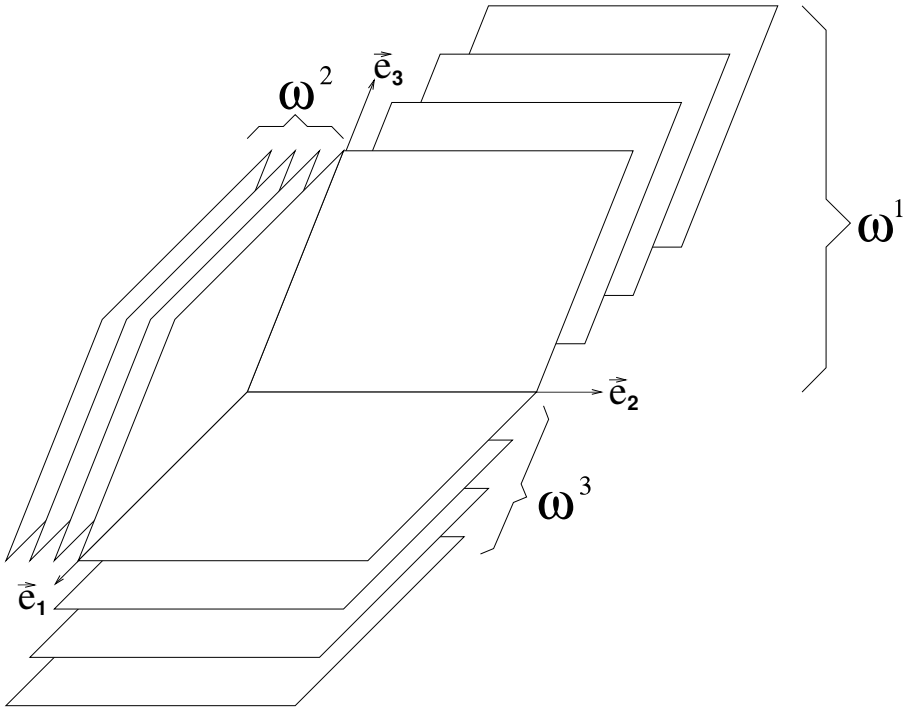


Figure 5.2: Level surfaces (="isograms") of the coordinate functions, which (i) are linear on V and (ii) are determined by the basis $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n$. The tip of a vector (not shown here), say, $x = \alpha^1\vec{e}_1 + \alpha^2\vec{e}_2 + \alpha^3\vec{e}_3$, is located at the intersection of three isograms. This is the point where $\omega^1(x) = \alpha^1$, $\omega^2(x) = \alpha^2$, and $\omega^3(x) = \alpha^3$. *Nota bene:* The upper indeces 1, 2, and 3 are not powers; they are superscripts.

The problem, therefore, is: What are these scalar functions? Are they elements of V^*? If so, do they form a linearly independent set? Do they span V^*?

The answer to these questions gives rise to

The Duality Principle:
For each ordered basis

$$\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$$

of a finite dimensional vector space V, there exists a corresponding basis

$$\{\underline{\omega}^1, \underline{\omega}^2, \cdots, \underline{\omega}^n, \}$$

for V^*, and vice versa, such that

$$\langle \underline{\omega}^i | \vec{e}_j \rangle = \delta^i_j \ . \tag{5.12}$$

Warning. The evaluation, Eq.(5.12), is not to be confused with an inner product. The existence of the duality between V and V^* by itself does not at all imply the existence of an inner product. We shall see that the existence of an inner product on a vector space establishes a unique basis-independent (= "natural") isomorphic correspondence between V and V^*. In the absence of an inner product such a correspondence does not exist.

The validation of the duality principle consists of the actual three-step construction of the basis dual to the given basis, which we denote by

$$B = \{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\} \subset V \quad \text{(basis for } V.\text{)}$$

Step I.
For all vectors x and y one has the following unique expansions:

$$\begin{aligned} x &= \alpha^1 \vec{e}_1 + \cdots + \alpha^n \vec{e}_n \\ y &= \beta^1 \vec{e}_1 + \cdots + \beta^n \vec{e}_n \\ x + y &= (\alpha^1 + \beta^1)\vec{e}_1 + \cdots + (\alpha^n + \beta^n)\vec{e}_n \\ cx &= c\alpha^1 \vec{e}_1 + \cdots + c\alpha^n \vec{e}_n \qquad (c \text{ is a scalar}) \end{aligned} \tag{5.13}$$

Note that

α^1	is	uniquely	determined	by	x
β^1	$''$	$''$	$''$	$''$	y
$\alpha^1 + \beta^1$	$''$	$''$	$''$	$''$	$x + y$
$c\alpha^1$	$''$	$''$	$''$	$''$	cx

Step II.
These four relations determine a linear function, call it ω^1,

$$\omega^1 : \ V \to \mathbb{R}.$$

Its defining properties are

$$\begin{aligned} \omega^1(x) &= \alpha^1 \\ \omega^1(y) &= \beta^1 \\ \omega^1(x + y) &= \alpha^1 + \beta^1 \\ \omega^1(cx) &= c\alpha^1 \end{aligned}$$

which imply

$$\omega^1(x+y) = \omega^1(x) + \omega^1(y)$$
$$\omega^1(cx) = c\omega^1(x)$$

In particular, using Eq.(5.13) on page 193, one has

$$\omega^1(\vec{e}_1) = 1$$
$$\omega^1(\vec{e}_2) = 0$$
$$\vdots$$
$$\omega^1(\vec{e}_n) = 0.$$

We conclude that ω^1 is a linear function, indeed. The function ω^1 is called the first coordinate function.

Step III.

Similarly the j^{th} coordinate function, a.k.a. "the j^{th} parallel projection", is defined by

$$\omega^j(x) = \alpha^j \quad \text{for } j = 2, 3, \cdots, n \ .$$

By applying ω^j to the ith basis vector $\vec{e}_i$, and using Eq.(5.13) on page 193 one obtains

$$\omega^j(\vec{e}_i) \equiv \langle \omega^j | \vec{e}_i \rangle = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

or in terms of the Kronecker delta,

$$\boxed{\langle \omega^j | \vec{e}_i \rangle = \delta^j{}_i} \ .$$

This is called a duality relation or duality principle. The choice of a different vector basis would have resulted in a correspondingly different set of coordinate functions, but would have again resulted in a duality relation.

Being elements in V^*, do these coordinate functions form a basis for V^*? The answer to this important question is answered in the affirmative by the following

Theorem 4. *(Dual Basis)*

Given: A basis $B = \{\vec{e}_1, \cdots, \vec{e}_n\}$ for V.

Conclusion: The set of linear functions $B^* = \{\omega^j\}_{j=1}^n$ which satisfies the duality relation

$$\langle \omega^j | \vec{e}_i \rangle = \delta^j{}_i \tag{5.14}$$

is a basis for V^*.

The proof of the spanning property of B^* hinges on the spanning property of B as follows: Let $f \in V^*$ be some linear function on V. Evaluate $f(x)$ and use $x = \sum_i \alpha^i \vec{e}_i$. Thus

$$\begin{aligned} f(x) &= f\left(\sum_i \alpha^i \vec{e}_i\right) \\ &= \sum_i f(\vec{e}_i)\,\alpha^i \qquad \alpha^i \text{ is the i}^{th} \text{ coord. of } x, \text{ i.e. } \alpha^i = \omega^i(x); \\ &= \sum_i f(\vec{e}_i)\,\omega^i(x) \qquad \forall x \in V \end{aligned}$$

This holds for all $x \in V$. Consequently,

$$f = \sum_i f(\vec{e}_i)\,\omega^i, \tag{5.15}$$

which is an expansion of f in terms of the elements of B^*, which means that B^* is a spanning set for V^* indeed.
To show that B^* has the linear independence property, we consider the equation

$$c_1\omega_1 + c_2\omega_2 + \cdots + c_n\omega_n = \underline{0}$$

where $\underline{0}$ is the function with constant value zero on V. By evaluating both sides on the ith basis vector $\vec{e}_i$ and using Eq.(5.14) one obtains

$$c_i = 0 \qquad \text{for } i = 1, 2, \cdots, n$$

Consequently, B^* does have the linear independence property. Together with its spanning property, this validates the claim made in the Theorem that B^* is a basis for V^*.
Example 1 (Column space*=Row space)
GIVEN:
Let

$$B = \left\{ \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \vec{e}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}; \vec{e}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

be a basis for the column space $V = R^3$.
a) IDENTIFY V^*, the space dual to V.
b) FIND the basis $B^* = \{\omega_1; \omega_2; \omega_3\}$ dual to B, i.e. exhibit elements ω^j which satisfy Eq.(5.14).
Solution
a) The space dual to V consists of the *row space*

$$V^* = \{\sigma = [a\,b\,c] : \ a, b, c \in R\}.$$

Indeed, for any $x = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in R^3$

$$\langle \sigma | x \rangle \equiv \sigma \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right)$$
$$= [a\ b\ c] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax + by + cz$$

Question:
What line of reasoning led to the fact that the answer to a) is the space of *row* vectors?
Answer:
A prerequisite for the course "Linear Mathematics in Finite Dimensions" is a knowledge of matrix theory (as, for example, in chapter 1 of Johnson, Riess, and Arnold). One of the concepts in this chapter is that of a *row vector*. One of the constitutive properties of a row vector is that it can be multiplied by a column vector and thereby produces a scalar number. This property was stored in our subconscious (the "hard disk" of our consciousness) where it has been ever since.
Now part a) of the above problem, asks for linear functions on the space of column vectors. That is the standing order, to search our subconscious for a concept with this requisite property. The success of this search was not immediate. In fact, we had to "sleep on it". However, the already-known concept "row vector" is the *green light* to inferring generalizations from particular instances. In particular, a row vector, when applied to column vectors, produces scalar numbers. The generalization is "row vectors give rise to (i.e. imply) linear functions". This inference is mandatory for two reasons: (i) a function is precisely the process of assigning scalar values to elements, here column vectors in the function's domain and (ii) the process is a linear one.

This generalization is a causal relation between row vectors and linear functions. Like all generalizations it is new knowledge. We arrived at it not deductively ("All men are mortal; Socrates is a man; hence Socrates is mortal"), but by the process of induction. This process is much more difficult and requires much more effort because it involved *all* our relevant knowledge, namely matrix theory.

To state it negatively and more generally: generalizations are not obtained by "intuition", "inspiration", "revelation", or by some other kind of pseudo explanation. Instead, the road to success is paved by hard work together with by not letting one's subconscious "goof off", but giving it a standing order(s) consisting of valid concepts.

b) Each ω^j is a row vector. They must satisfy

$$\left.\begin{array}{l} \langle\omega^1|\vec{e}_1\rangle = [a\ b\ c]\begin{bmatrix}1\\0\\0\end{bmatrix} = 1 \\ \langle\omega^1|\vec{e}_2\rangle = [a\ b\ c]\begin{bmatrix}1\\1\\0\end{bmatrix} = 0 \\ \langle\omega^1|\vec{e}_3\rangle = [a\ b\ c]\begin{bmatrix}1\\1\\1\end{bmatrix} = 0 \end{array}\right\} \Rightarrow \omega^1 = [1\ {-1}\ 0] \tag{5.16}$$

$$\left.\begin{array}{l} \langle\omega^2|\vec{e}_1\rangle = [d\ e\ f]\begin{bmatrix}1\\0\\0\end{bmatrix} = 0 \\ \langle\omega^2|\vec{e}_2\rangle = [d\ e\ f]\begin{bmatrix}1\\1\\0\end{bmatrix} = 1 \\ \langle\omega^2|\vec{e}_3\rangle = [d\ e\ f]\begin{bmatrix}1\\1\\1\end{bmatrix} = 0 \end{array}\right\} \Rightarrow \omega^2 = [0\ 1\ {-1}] \tag{5.17}$$

$$\left.\begin{array}{l} \langle\omega^3|\vec{e}_1\rangle = [u\ v\ w]\begin{bmatrix}1\\0\\0\end{bmatrix} = 0 \\ \langle\omega^3|\vec{e}_2\rangle = [u\ v\ w]\begin{bmatrix}1\\1\\0\end{bmatrix} = 0 \\ \langle\omega^3|\vec{e}_3\rangle = [u\ v\ w]\begin{bmatrix}1\\1\\1\end{bmatrix} = 1 \end{array}\right\} \Rightarrow \omega^3 = [0\ 0\ 1] \tag{5.18}$$

Thus the basis of duals for V^*, the space dual to $V = R^3$ is

$$B^* = \{\omega^j\}_{j=1}^3 = \{\ [1\ {-1}\ 0],\ [0\ 1\ {-1}],\ [0\ 0\ 1]\ \}$$

Example 2
Same as Example 1 on page 195, except that

$$B = \left\{\vec{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}; \vec{e}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}; \vec{e}_3 = \begin{bmatrix}1\\1\\2\end{bmatrix}\right\}$$

Answer:

$$B^* = \{\omega^j\}_{j=1}^3 = \{\ [1\ {-1}\ 0],\ [0\ \ 1\ -\frac{1}{2}],\ [0\ \ 0\ \frac{1}{2}]\ \}$$

Comment.
Note that changing only one element of B, say, $\vec{e}_j \to \vec{e}'_j$ changes several elements of B^*, as in Figure 5.3 on page 199. This implies that there is as-yet no basis independent correspondence between V^* and V. A basis independent correspondence would have required that changing only one of the basis vectors in Vwould have produced a corresponding change in only one basis vector in V^*.
Summary.
There does exist a unique correspondence between ordered basis sets in V and V^*

$$\{\vec{e}_i\}_{i=1}^n \leftrightarrow \{\omega^j\}_{j=1}^n,$$

but not between individual vectors in V and V^*:

$$\{\text{coordinate vectors}\} \nleftrightarrow \{\text{coordinate surfaces}\}$$

More succinctly, one says that there exists no natural (i.e., basis independent) isomorphism between V and V^*.

However, we shall see in the next lecture that if V is endowed with an *inner product*, then there *does* exist a natural (and unique!) isomorphism between V and V^*.

5.5.1 Multivariable Calculus and Linear Algebra: Differentials as Dual Elements

Consider the Taylor series expansion of the multivariable function $f(x^1, \cdots, x^n) \equiv f(x)$ in a neighborhood of a point $(a^1, \cdots, a^n) \equiv a$ in an n-dimensional space,

$$f(x) - f(a) = \underbrace{\frac{\partial f}{\partial x^i}(a)\left[x^i - a^i\right]}_{\substack{\text{“Principal} \\ \underline{\text{Linear Part}}\text{”}}} + \underbrace{\frac{1}{2!}\frac{\partial^2 f}{\partial x^j \partial x^i}(a)\left[x^j - a^j\right]\left[x^i - a^i\right] + \cdots}_{\substack{\text{neglegible} \\ \text{higher order} \\ \text{terms}}} \tag{5.19}$$

(Einstein summation convention for pairs of repeated indeces.)

Observe that $\{x^i - a^i\}_{i=1}^n$ are the components of the column vector

$$\begin{bmatrix} x^1 - a^1 \\ \vdots \\ x^n - a^n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{e}_1}(x^1 - a^1) + \cdots + \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}}_{\vec{e}_n}(x^n - a^n)$$

$$\equiv \vec{e}_i\,(x^i - a^i) \tag{5.20}$$

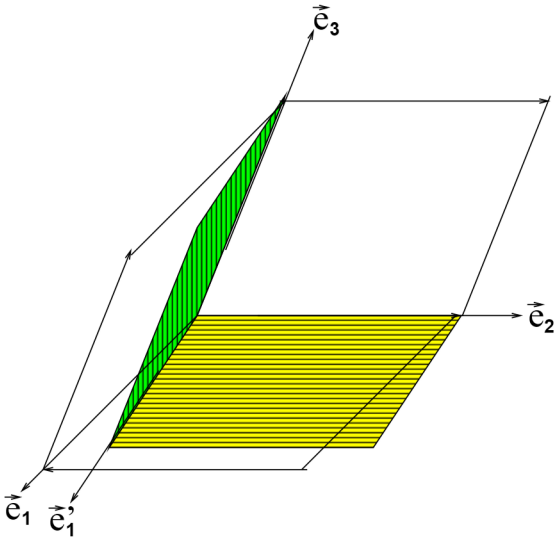


Figure 5.3: Changing only one vector of a basis, for example $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\} \longrightarrow \{\vec{e}_1{}', \vec{e}_2, \vec{e}_3\}$, changes more than one dual element as evidenced by the tilting of their coordinate surfaces in the vector space.

and that $\{\frac{\partial f}{\partial x^i}\}_{i=1}^n$ are the components of the row vector

$$\begin{aligned}\left[\frac{\partial f}{\partial x^1}, \cdots, \frac{\partial f}{\partial x^n}\right] &= \frac{\partial f}{\partial x^1}\underbrace{[1, \cdots, 0]}_{\omega^1} + \cdots + \frac{\partial f}{\partial x^n}\underbrace{[0, \cdots, 1]}_{\omega^n} \\ &\equiv \frac{\partial f}{\partial x^j}\,\omega^j \end{aligned} \tag{5.21}$$

The $\vec{e}_i$'s and the ω^j's form the bases for the column space V and its dual, the row space V^*:

$$\omega^j(\vec{e}_i) \equiv \langle \omega^j | \vec{e}_i \rangle = \delta^j{}_i \ .$$

It follows that the multivariable Taylor series expansion of $f(x)$ is

$$\begin{aligned} f(x) - f(a) &= \frac{\partial f}{\partial x^j}(a)\,\langle \omega^j | \vec{e}_i \rangle\,(x^i - a^i) \quad + \text{higher order terms} \\ &= \underbrace{\langle\ \frac{\partial f}{\partial x^j}(a)\omega^j}_{\in V^*}\ |\ \underbrace{\vec{e}_i(x^i - a^i)\ \rangle}_{\in V} + \cdots \end{aligned} \tag{5.22}$$

Thus the principal linear part of f's Taylor series consists of the covector $\frac{\partial f}{\partial x^j}(a)\omega^j \in V^*$ being evaluated on the vector $\vec{e}_i(x^i - a^i) \in V$.

These observations are condensed into the following

Definition 2. *The linear function* $\dfrac{\partial f}{\partial x^j}(a)\,\omega^j$*, also written* [2] *as* $\dfrac{\partial f}{\partial x^j}\,dx^j$*, or more briefly as*

$$df(a;\,\cdot\,):\quad V \to R$$
$$\vec{e}_i v^i \rightsquigarrow df(a;\vec{e}_i v^i) = \frac{\partial f}{\partial x^j}(a)\,v^i\ ,$$

is an element in the vector space V^**, which is attached to the point* $a = (a^1, \cdots, a^n)$*.*

It is a function of two variables: it depends (in general) non-linearly on the first argument a, but linearly on the second variable.

These observations compel us, at the pain of contradiction, to attach conceptually the vector space V and its dual V^* to the point $a = (a^1, \cdots, a^n)$ in an n-dimensional space. The resulting picture is one where the base space, which is parametrized by a, has assigned to each of its points copies of V and V^*.

5.6 Metric as a Bilinear Function on a Vector Space

Lecture 24

The vector space arenas developed so far are in skeleton form but fundamental to all of mathematics. In physics and engineering terminology their linearity is captured by means of the *superposition principle*. In mathematics, by means of *closure under linear combination*.

The bare bones attributes introduced so far are the *linear (in)dependence* and the *spanning property* of a set of vectors. These properties are sufficient for characterizing a vector space in terms of coordinate systems introduced via any chosen (or given) basis. As a result every vector space V accommodates its dual space V^*, the space of linear functions[3]. This space is a vector space in its own right, and any basis for

[2] There is a marked difference between the expression $\dfrac{\partial f}{\partial x^j}\,dx^j$, which refers to a linear combination of linear functions, namely, $\omega^1 \equiv dx^1, \cdots,$ and $\omega^n \equiv dx^n$ as compared $\dfrac{\partial f}{\partial x^i}\,\Delta x^i$, which is the value of $\dfrac{\partial f}{\partial x^j}\,dx^j$ when it is evaluated on the vector $\vec{e}_i\,\Delta x^i$.

Failure to recognize the distinction between a function and its possible values is a case of the fallacy of "package dealing", failing to discriminate crucial differences, namely, to treat together, as a conceptual whole or "package", elements that differ essentially in nature, truth-status, importance or value.

[3] In engineering, physics, and mathematics, the space of such functions includes the space of Fourier transforms, the space of (generalized) Fourier series, the space of X-ray crystal diffraction patterns, the space of wavelet transforms of L^2 functions, and others.

V determines a unique corresponding basis for V^*. Indeed, the dimensions of V and V^* are the same, a fact which is a consequence of the duality principle

$$\langle \underline{\omega}^j | \vec{e}_j \rangle = \delta^j{}_i$$

In spite of this duality, there is no natural (i.e. basis-independent) correspondence between V and its dual space of covectors, V^*.

This deficiency, as we shall see, disappears once one has identified an inner product on the given vector space.

5.6.1 Bilinear Functional; the Metric

There is no natural isomorphism between V and V^*. However, if the vector space has an inner product defined on it, then such an isomorphism is determined.

Definition 3. *(Bilinear Form)*
Given: a vector space U and a vector space V.
A bilinear functional *(or* "form"*) on $U \times V$ (= pairs of elements, one from U and one from V) is a function w,*

$$\begin{array}{rl} w: & U \times V \stackrel{w}{\to} \mathbb{R} \\ & (x,y) \rightsquigarrow \quad w(x,y) \end{array}$$

with the properties

$$\begin{aligned} w(\alpha^1 x_1 + \alpha^2 x_2, y) &= \alpha^1 w(x_1, y) + \alpha^2 w(x_2, y) \\ w(x, \beta^1 y_1 + \beta^2 y_2) &= \beta^1 w(x, y_1) + \beta^2 w(x, y_2); \end{aligned}$$

in other words, w is linear in each argument.

In this definition U and V can be vector spaces of different dimensions. The concept of a metric arises if the two vector spaces are one and the same and is given by the following definition:

Definition 4. *(Metric) A* metric *(or* inner product*) is a bilinear functional g on $V \times V$ (pairs of elements in V)*

$$g: \ (x,y) \rightsquigarrow \quad g(x,y)$$

with the property

$$g(x,y) = g(y,x) \ .$$

In other words, a real-valued metric is symmetric.
Comment.
If the metric were complex-valued, then the symmetry condition get replaced by

$$g(x,y) = \overline{g(y,x)}$$

The metric $g(\ ,\ )$ is said to be an *inner product* whenever g is positive definite, i.e. $g(x,x) > 0\ \forall x \neq 0$.

Example (Basis Expansion of the Metric) Let

$$x = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n$$

be a representation of a vector x relative to a basis $\{\vec{e}_1, \cdots, \vec{e}_n\}$ for V. Let g be a metric on V. Then

$$\begin{aligned} g(x,y) &= g(x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n, y^1\vec{e}_1 + y^2\vec{e}_2 + \cdots + y^n\vec{e}_n) \\ &= x^1y^1g(\vec{e}_1,\vec{e}_1) + (x^1y^2 + x^2y^1g(\vec{e}_1,\vec{e}_2) + x^2y^2g(\vec{e}_2,\vec{e}_2) + \cdots \\ &= x^1y^1\vec{e}_1\cdot\vec{e}_1 + (x^1y^2 + x^2y^1)\vec{e}_1\cdot\vec{e}_2 + x^2y^2\vec{e}_2\cdot\vec{e}_2 + \cdots \\ &= x^1y^1g_{11} + (x^1y^2 + x^2y^1)g_{12} + x^2y^2g_{22} + \cdots \\ &= x^iy^jg_{ij} \quad \text{(Einstein summation convention for pairs of repeated indeces.)} \end{aligned}$$

The coefficients $g_{ij} \equiv \vec{e}_i\cdot\vec{e}_j \equiv g(\vec{e}_i,\vec{e}_j)$ are the *components* of the metric g *relative* to the given basis. They are the inner products (i.e., the "dot" products) of all pairs of basis vectors.
Nota bene: In the Einstein summation convention the placement of super and subscripts (one "up" the other "down") provides an error correction code. That placement insures that a particular sum of products is always <u>invariant</u> under <u>any</u> change of bases. Thus each of the sums

$$\begin{aligned} x^1\vec{e}_1 + \cdots + x^n\vec{e}_n \equiv x^i\vec{e}_i = \vec{x} = x'^1\vec{e'}_1 + \cdots + x'^n\vec{e'}_n \equiv x'^i\vec{e'}_i \\ x_1\omega^1 + \cdots + x_n\omega^n \equiv x_i\omega^i = \underline{x} = x'_1\omega'^{\,1} + \cdots + x'_n\omega'^{\,n} \equiv x'_i\omega'^{\,i} \\ x_1y^1 + \cdots + x_ny^n \equiv x_iy^i = x'_1y'^{\,1} + \cdots + x'_ny'^{\,n} \equiv x'_iy'^{\,i} \end{aligned}$$

does not change upon transitioning from the unprimed to the primed basis, even though the corresponding terms are not equal.

The Einstein summation convention is the means of mathematizing basis-independent aspects of the world.

5.6.2 Metric as an Isomorphism between Vector Space and its Space of Duals

The scalar product

$$\begin{aligned} g: \quad V \times V &\rightarrow R \\ (x,y) &\rightsquigarrow g(x,y) \end{aligned} \tag{5.23}$$

is a bilinear function. Consequently, it can be evaluated on only one of its arguments, $g(x,\ )$.The result is a linear function. More precisely, a metric establishes a natural (i.e. basis independent) isomorphism between the vector space V and its space of duals, V^*. In order to conserve notation we shall use the same symbol g to designate this correspondence. Its defining property is

$$\begin{aligned} g: \quad V &\rightarrow V^* \\ x &\rightsquigarrow g(x,\ ) \equiv \underline{x} \ = \ x\cdot \end{aligned} \tag{5.24}$$

Here $\underline{x}$ is that linear function which, when evaluated on $y \in V$, yields $g(x,y)$:

$$\begin{aligned} \underline{x} = \ x\cdot: \quad V &\stackrel{\underline{x}}{\rightarrow} R \\ y &\rightsquigarrow \langle \underline{x}|y\rangle = x\cdot y \equiv g(x,y) \end{aligned}$$

If g maps $\vec{x}$ to its image $\underline{x}$, what is the image of the set of components of $\vec{x}$? The answer is given by the following proposition,

Proposition

Given the vector $\vec{x} = x^k\vec{e}_k$, the numerical coefficients x_j of the corresponding covector $\underline{x} = x_j\underline{\omega}^j$ are given explicitly by the following computation:

$$\begin{aligned} \underline{x} &= g(\vec{x},\ \ ) \\ &= x^k g(\vec{e}_k,\ ) \end{aligned}$$

Taking advantage of the spanning property of $\{\omega^j\}$, Eq.(5.15) on page 195), we find that the to-be-determined x_j satisfy

$$x^k g(\vec{e}_k,\ ) = x_j\omega^j$$

Evaluating both side on each of the basis vectors e_i, and using the duality relation Eq.(5.14), we obtain

$$\begin{aligned} x_i &= x^k g(\vec{e}_k, \vec{e}_i) \\ &= x^k g_{ki} \\ &\equiv \vec{x}\cdot\vec{e}_i, \end{aligned} \tag{5.25}$$

the components of $x_j\omega^j$, which is the image of $\vec{x} = x^k\vec{e}_k$ under g.

Although the two mappings, Eq.(5.23) and 5.24) are entirely different functions (one bilinear, the other linear) they can be represented by a single explicit formula,

$$\boxed{g = g_{ij}\,\omega^i \otimes \omega^j}\ .$$

The (ij)th term in this double sum is the metric coefficient g_{ij} multiplied by the "tensor product" of ω^i and ω^j, namely $\omega^i \otimes \omega^j$. The virtue of these "tensor products" is twofold: (i) each one unites the bilinear and the linear mapping property into a single concept, and (ii) this linearity and bilinearity is preserved under linear combinations. The meaning of this tensor product arises from the meaning $g(\vec{x}, \vec{y})$. The bilinearity g implies that

$$\begin{aligned} g(\vec{x}, \vec{y}) &= \vec{x} \cdot \vec{y} \\ &= \vec{e}_i \cdot \vec{e}_j\, x^i y^j \\ &= g_{ij}\,\omega^i(\vec{x})\,\omega^j(\vec{y}) \\ &\equiv g_{ij}\,\omega^i \otimes \omega^j(\vec{x}, \vec{y}) \qquad \forall\, \vec{x}, \vec{y} \in V. \end{aligned} \tag{5.26}$$

Consequently,

$$g = g_{ij}\,\omega^i \otimes \omega^j\ .$$

Here the tensor product symbol $\otimes$ establishes an ordered juxtaposition of two *linear* function(al)s, thereby yielding a *bi*linear function(al).
On the other hand, the linearity of $g = g_{ij}\,\omega^i \otimes \omega^j$ in each argument implies that for a given $\vec{x}$ one has

$$\begin{aligned} g_{ij}\,\omega^i \otimes \omega^j(\vec{x}, \vec{y}) &= g_{ij}\omega^i(\vec{x})\,\omega^j(\vec{y}) \\ &= g_{ij}\,x^i\omega^j(\vec{y}) \\ &= x_j\omega^j(\vec{y}) \qquad \text{(as defined by Eq.(5.25) on page 203)} \\ &\equiv \underline{x}\,(\vec{y}) \qquad \forall y \in V \\ &= \langle \underline{x} | \vec{y} \rangle \qquad \text{(Dirac notation)} \end{aligned}$$

Consequently,

$$\begin{aligned} \vec{x} \overset{g}{\rightsquigarrow} g_{ij}\,\omega^i \otimes \omega^j(\vec{x},\ ) &\equiv g_{ij}\omega^i(\vec{x})\,\omega^j \\ &= \underline{x} \\ &= \langle \underline{x}| \quad \text{(Dirac notation)} \end{aligned}$$

which is to say that $g = g_{ij}\,\omega^i \otimes \omega^j$ is a *linear* function(al) which assigns $\vec{x} \in V$ to $\underline{x} \in V^*$.

Implicit in the metric-induced isomorphism

$$\begin{aligned} V &\xrightarrow{\ g\ } V^* \\ \vec{x} &\rightsquigarrow \underline{x} \end{aligned}$$

is the existence of a vector normal to $\underline{x}$'s isograms[4] in V. In fact, this normal is precisely the preimage of $\underline{x}$, namely $\vec{x}$ itself. Moreover, each set of isograms of each coordinate function ω^i has such a normal. This circumstance gives rise to the basis *reciprocal* to the originally given basis. This reciprocal basis mathematizes the obliqueness of the given coordinate surfaces. It is developed in the next subsection.

5.7 Mathematizing an Oblique Coordinate System

Lecture 25

Historically the "contravariant" components of a vector $\vec{x}$ are its components relative to a chosen/given basis. Indeed, they are "the" components relative to such a basis, regardless of what metric the vector space V is endowed with.

If V is endowed with a specific metric then there exists a second well-defined basis, "reciprocal" to the first one. It turns out that the components of the vector $\vec{x}$ relative to this second (metric-induced) reciprocal basis are precisely the x_i's, Eq.(5.25), and it is these components that historically have been called the "covariant" components of a vector. As we shall see, the reciprocal basis brings the geometry of the dual basis into very sharp focus.

Warning:
The appellations "contravariant vector" and "covariant vector" are invalid concepts. They are oxymorons, examples of mixing incommensurable categories, an attempt to blend mutually exclusive ideas into a single unit. Indeed, a vector is a *basis independent* concept, while "contravariant" or "covariant" are attributes of the components of a vector and thus are *relative to some basis.*

The reciprocal basis and its properties arise as follows:

Start with an oblique basis $\{\vec{e}_1, \vec{e}_2\}$ and its coordinate system as in Fig. 5.4. Next introduce a vector, $\vec{e}_1^*$, which is perpendicular to $\vec{e}_2$ and is "normalized" by being reciprocal to $\vec{e}_1$ as in Figure 5.5 on page 208

$$\vec{e}_1^* \cdot \vec{e}_2 = 0$$
$$\vec{e}_1^* \cdot \vec{e}_1 = 1$$

In a similar way introduce $\vec{e}_2^*$ which is perpendicular to $\vec{e}_1$ and reciprocal to $\vec{e}_2$:

$$\vec{e}_2^* \cdot \vec{e}_1 = 0$$
$$\vec{e}_2^* \cdot \vec{e}_2 = 1$$

[4]Recall (from page 192) that an *isogram* (a.k.a. a level surface) of a function is the locus of points where the function has the same constant value.

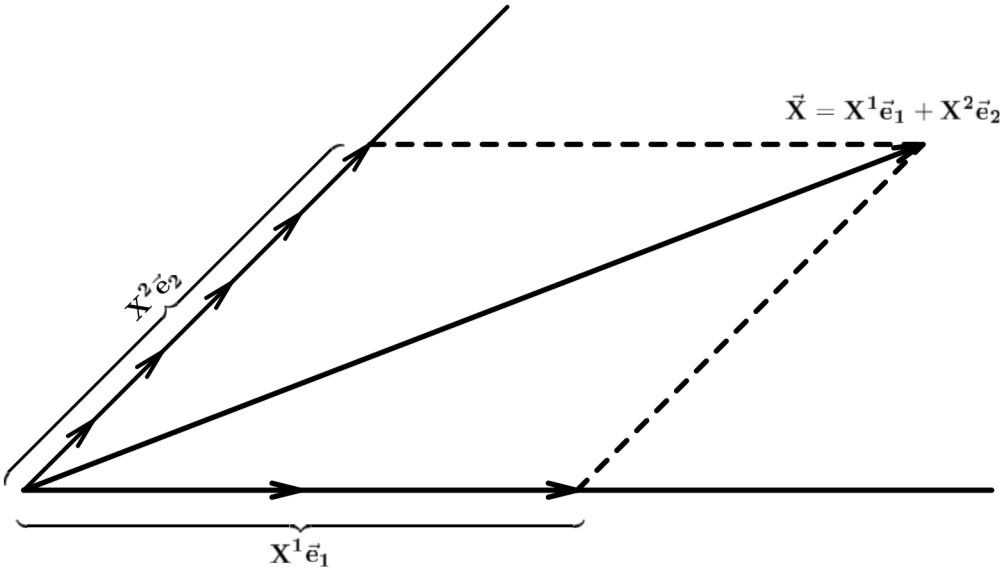


Figure 5.4: Oblique basis and its oblique coordinate system.

The basis $\{\vec{e}_1^*, \vec{e}_2^*\} \equiv R^*$ constructed in this manner is *reciprocal* to the given basis $B = \{\vec{e}_1, \vec{e}_2\}$ because it satisfies

$$\vec{e}_i^* \cdot \vec{e}_j = \delta_{ij}.$$

More generally, we have the following

Definition 5. *(Reciprocal Basis)*

Given (i) the metric $g ='' \cdot''$*, a metric on* v

(ii) a basis $\{\vec{e}_1, \cdots, \vec{e}_n\}$ *for* V*,*

then the set of vectors

$$\{\vec{e}_1^*, \vec{e}_2^*\},$$

where

$$\vec{e}_k^* \cdot \vec{e}_j = \delta_{kj}, \tag{5.27}$$

is the basis reciprocal *to* $\{\vec{e}_j\}_{j=1}^n$.

This definition says that $\{\vec{e}_k\}$ is a vector perpendicular to the plane containing the vectors $\{\vec{e}_1, \cdots, \vec{e}_{k-1}, \vec{e}_{k+1}, \cdots, \vec{e}_n\}$, i.e.

$$\boxed{\vec{e}^*_k \cdot \vec{e}_j = 0 \quad j \neq k.}$$

Furthermore, $\vec{e^*}_k$ is scaled such that

$$\boxed{\vec{e}^*_k \cdot \vec{e}_k = 1 \quad (\text{No sum over } k)}$$

It is also clear that if the basis $\{\vec{e}^*_k\}_{k=1}^n$ is reciprocal to $\{\vec{e}_j\}_{j=1}^n$, then $\{\vec{e}_j\}_{j=1}^n$ is reciprocal to $\{\vec{e}^*_k\}_{k=1}^n$.

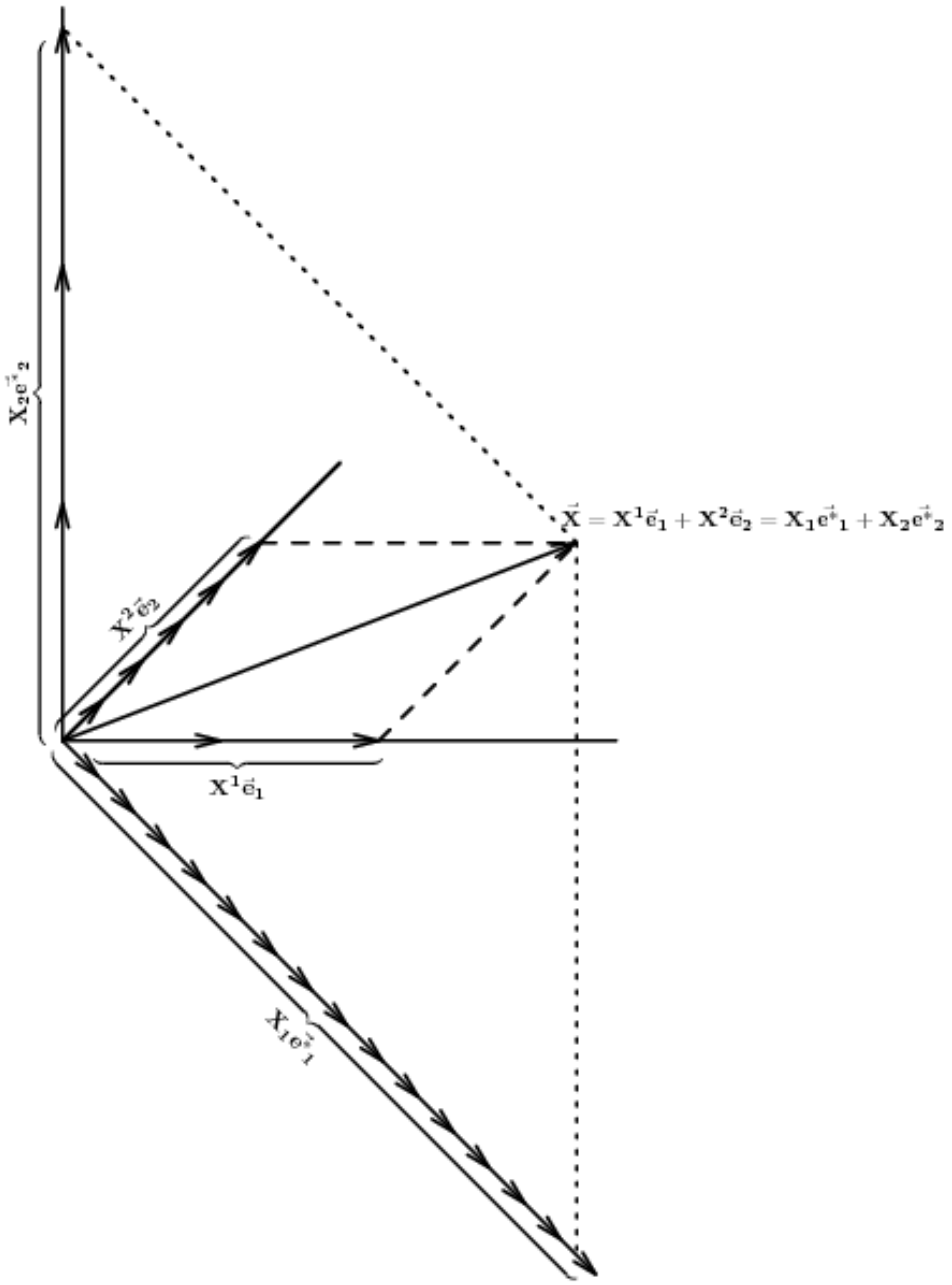


Figure 5.5: Oblique basis $\{\vec{e}_1, \vec{e}_2\}$ and its oblique reciprocal basis $\{\vec{e}^*_1, \vec{e}^*_2\}$

5.7.1 The Geometric Relation Between a Vector Space and its Dual[5]

First of all, recall that from one's knowledge of a given inner product on the vector space V one identifies by inductive reasoning a unique one-to-one mapping from V to V^*. This mapping which does not depend on one's chosen basis $\{e_i\}$ for V. Thus, if one is given or chooses a basis, then one has

$$V \overset{g}{\to} V^* \tag{5.28}$$

$$a \rightsquigarrow \underline{\alpha} = g(a, \cdot) \tag{5.29}$$

$$\left.\begin{array}{l} = g_{ij}\omega^i \otimes \omega^j(a, \cdot) \\ = g_{ij}\langle\omega^i|a\rangle\langle\omega^j| \end{array}\right\} : \begin{array}{l} V \overset{g}{\to} R \\ x \rightsquigarrow g(a,x) = \left[g_{i1}\omega^i(a)\omega^1 + \cdots + g_{in}\omega^i(a)\omega^n\right](x) \\ \qquad\qquad = \left[\quad a_1 \quad \omega^1 + \cdots + \quad a_n \quad \omega^n\right](x) \end{array} \tag{5.30}$$

Secondly recall that the elements of the reciprocal basis $\{e_k\}$ are determined by the condition

$$e_k^* \cdot e_\ell = \delta_{k\ell}.$$

This condition implies that

$$e_\ell = \sum_{k=1}^{n} g_{\ell k}\, e_k^* \ , \quad \ell = 1, \cdots, n \tag{5.31}$$

and

$$e_k^* = \sum_{j=1}^{n} g^{kj} e_j \ , \quad k = 1, \cdots, n \tag{5.32}$$

where $[g^{kj}]$ is the inverse of $[g_{\ell k}]$:

$$g_{\ell k}\, g^{kj} = \delta_\ell{}^j \ . \tag{5.33}$$

[5]In this subsection vectors are identified by Latin letters with or without *sub*scripts, while their components also by Latin letters but with *super*scripts, as, for example, in

$$a = a^i e_i \equiv a^1 e_1 + \cdots + a^n e_n \ .$$

By contrast covectors are identified by Greek letters with or without *super*scripts, while their components also by Greek letters but with *sub*scripts, as, for example, in

$$\alpha = \alpha_j \omega^j \equiv \alpha_1 \omega^1 + \cdots + \alpha_n \omega^n \ .$$

With either the given or the reciprocal basis at one's disposal, a vector x has the two corresponding expansions

$$x = x^1 e_1 + \cdots + x^n e_n = \sigma_1 e_1^* + \cdots + \sigma_n e_n^* \ .$$

The matrix of metric coefficients $g_{i\ell} = e_i \cdot e_\ell$ provides the relation between the two. Indeed, taking the inner product with e_ℓ leads to

$$x \cdot e_\ell = x^i \underbrace{e_i \cdot e_\ell}_{g_{i\ell}} = \sum_{k=1}^{n} \sigma_k \underbrace{e_k^* \cdot e_\ell}_{\delta_{k\ell}},$$

so that

$$\sigma_\ell = x^i g_{i\ell}$$

and hence that the $\{e_k^*\}$-expansion is

$$x = \sum_{k=1}^{n} x^i g_{ik} e_k^* \ .$$

On the other hand, taking the inner product with e_ℓ^*, one finds the $\{e_\ell\}$-expansion coefficients $x^\ell = \omega^\ell(x)$, namely

$$x^\ell = x \cdot e_\ell^* = g(e_\ell^*, x) \ . \tag{5.34}$$

These observations lead to the following
Reciprocal Basis Principle:
In the context of an inner product space, whenever the given basis is oblique[6], the reciprocal basis is $\{e_\ell^*\}$ is – as shown in Eq.(5.34) – the mandatory means for finding the components

$$x^\ell = \omega^\ell(x) \ . \tag{5.35}$$

Compare Eq.(5.34) with (5.35). Leaving x unspecified, one arrives at the coordinate basis elements in V^*,

$$\omega^\ell = g(e_\ell^*, \ ) \ .$$

More generally one has

$$\alpha = g(a, \ ).$$

Going into the reverse direction , the converse question is:
GIVEN:

$$\alpha \in V^* : \ x \rightsquigarrow \alpha(x) = \left[a_1\omega^1 + \cdots + a_n\omega^n\right](x) \in R, \tag{5.36}$$

[6] i.e. whenever $[e_i \cdot e_j]$ is non-orthogonal; $\{e_i\}$ is not orthonormal.

WHAT vector $a \in V$ does α correspond to geometrically?

Note thet α determines the following $(n-1)$-dimensional subspace V,

$$V_\alpha = \left\{x: \ \alpha(x) = \left[a_1\omega^1 + \cdots + a_n\omega^n\right](x) = 0\right\} \subset V$$

Geometrically this is an $n-1)$-dimensional plane in V passing through its origin.

Q: What are $\underbrace{\text{“the”}}$vectors that span this plane?

“using the given basis $\{e_i\}$”

A: First use the reciprocal basis $\{e_1^*, \cdots, e_n^*\}$ to construct the vector

$$a_1e_1^* + \cdots + a_ne_n^* \equiv a \in V \ .$$

Then, with $x = x^ie_i$, one has

$$\begin{aligned} a \cdot x &= [a_1e_1^* + \cdots + a_ne_n^*] \cdot x \\ &= a_1x^1 + \cdot + a_nx^n \\ &= a_i\omega^i(x) \\ &= \alpha(x) \end{aligned}$$

Applying this result to the subspace definition

$$V_\alpha = \{x: \ 0 = \alpha(x) = a \cdot x\}$$

leads to the

Conclusion: The plane V_α consists of all vectors x perpendicular to a; this plane is the zero isogram of the given covector α, Eq.(5.36), and the unique vector, $a = a_1e_1^* + \cdots + a_ne_n^*$ constructed from the expansion coefficients of α, is perpendicular to the isograms of α

SUMMARY: The reciprocal basis, Eq.(5.32), induces a mapping which assigns to $\alpha \in V^*$ that vector $a \in V$,

$$\alpha \overset{\{e_1^*, \cdots, e_n^*\}}{\rightsquigarrow} a \ ,$$

which is perpendicular to α's isograms in V. Together with the additional fact that

$$a \overset{g}{\rightsquigarrow} \alpha \ ,$$

one has the result that the inner product g establishes a unique isomorphism between V and V^*,

$$V \overset{g}{\leftrightarrow} V^* \ .$$

5.7.2 Mathematization of the Law of X-ray Diffraction by a Crystal

A beam of X-rays striking the periodic lattice of a crystal gets refracted into discrete directions according to Bragg's Law. For a three-dimensional crystal this law is mathematized as follows.

The crystal consists of atoms arranged periodically into a lattice. They are located in a 3-d vector space. Its basis vectors $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ are the displacements into three different directions from the origin, the location of an arbitrarily chosen reference atom in the crystal. These displacements are determined by the three neighboring atoms closest to the atom at the origin.

All atoms of the crystal are located at some integral multiple linear combination of the basis vectors. Thus a typical atom is located at

$$\vec{x} = h\vec{e}_1 + k\vec{e}_2 + \ell\vec{e}_3 \qquad (h, k, \ell \in \{\text{integers}\})$$

As is the case for an isoclinic crystal, the basis

$$B = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\} \tag{5.37}$$

is oblique in general: the basis vectors need not be orthogonal nor of unit length.

Consider a beam of electromagnetic radiation (X-rays). For a plane wave this disturbance is characterized by its amplitude profile

$$\psi(\vec{x}) = Ae^{i\underline{\phi}(\vec{x})}; \quad (\nabla^2 + k^2)\psi = 0.$$

Here

$$\underline{\phi}(\vec{x}) = k_1x^1 + k_2x^2 + k_3x^3.$$

is the value of the phase $\underline{\phi}$ at location

$$\vec{x} = x^1\vec{e}_1 + x^2\vec{e}_2 + x^3\vec{e}_3.$$

It follows that relative to the dual basis

$$B^* = \{\omega^1, \omega^2, \omega^3 : \ \langle\omega^j|\vec{e}_i\rangle = \delta^j{}_i\},$$

the plane wave phase function is

$$\underline{\phi} = k_1\omega^1 + k_2\omega^2 + k_3\omega^3.$$

When such a plane wave enters a crystal, it is observed[7] that emerging from this crystal there are discrete plane wave beams. Their directions relative to the incident beam is determined entirely by the atomic crystal basis, Eq.(5.37), more precisely, by the set of parallel crystal planes. Before expressing this deterministic relation in mathematical terms, one must first mathematize these crystal planes.

[7]Observed and explained by father and son W.H. Bragg and W.L. Bragg in 1913.

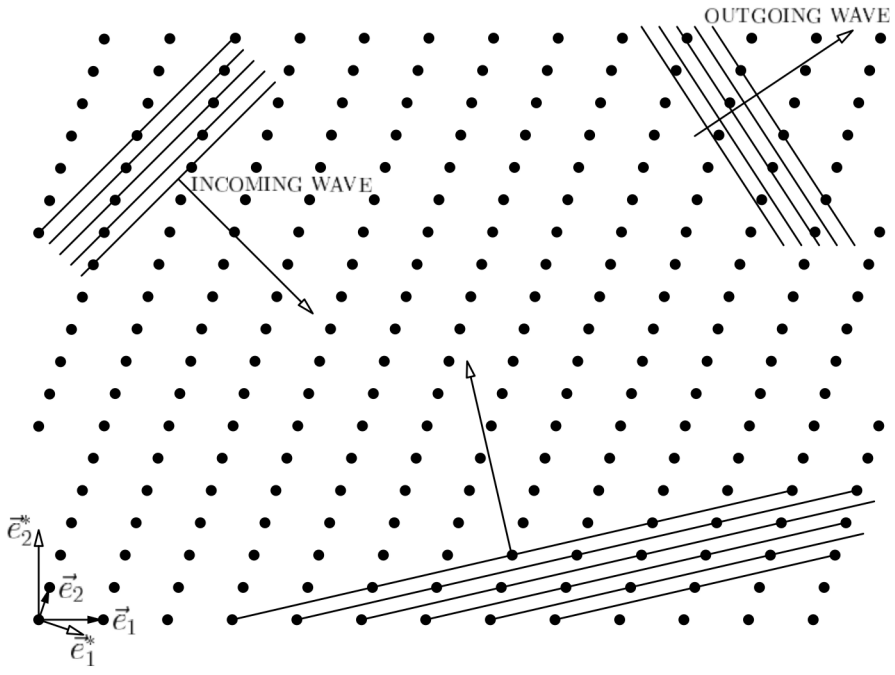


Figure 5.6: Crystal lattice with incoming and outgoing phase fronts. The outgoing wave is the result of a diffraction process. The difference between the in- and outgoing wave front normals is the vector perpendicular to the crystal planes indicated near the bottom. The basis vectors $\vec{e}_1$ and $\vec{e}_2$ are generated by directed pairs of nearest neighbor atoms (black circles).

Mathematized Crystal Planes

This is done by first observing that each one is one of the isograms of

$$\underline{f} = h\omega^1 + k\omega^2 + \ell\omega^3 \qquad (h, k, \ell \in \{\text{integers}\}) \tag{5.38}$$

The integers (h, k, l), which are understood to be relative prime (i.e. have no common integral divisor) are the *Miller indices* of a given set of parallel crystal planes. Such a linear combination of dual basis elements with relative prime integral coefficients we shall call a *Miller covector*. It is an element of the dual space and each one of its isograms passing through an integral linear combination of basis vectors has an integral value.

There is a on-to-one correspondence between a set of parallel crystal planes and the Miller covector corresponding to this set. The following problem illustrates this fact.

Problem
Find the Miller covector for the set of parallel crystal planes one of which contains the set of linearly independent vectors $S = \{\vec{e}_1 + \vec{e}_2, 2\vec{e}_2, 3\vec{e}_3\}$.
Solution
This is a two-step process.

Step 1: Let

$$\underline{g} = a\omega^1 + b\omega^2 + c\omega^3,$$

and find a, b, c so that $\underline{g}$ has an isogram , say $\underline{g} = 1$, passing through $\vec{e}_1 + \vec{e}_2, 2\vec{e}_2$, and $3\vec{e}_3\}$:

$$\begin{aligned} \underline{g}(\vec{e}_1 + \vec{e}_2) &= a + b + 0 = 1 \\ \underline{g}(2\vec{e}_2) &= 0 + 2b + 0 = 1 \\ \underline{g}(3\vec{e}_2) &= 0 + 0 + 3c = 1 \end{aligned}$$

Thus

$$\underline{g} = \frac{1}{2}\omega^1 + \frac{1}{2}\omega^2 + \frac{1}{3}\omega^3$$

Step 2: The Miller indices are mutually prime integers. Thus multiplying $\underline{g}$ by the least common denominator yield the Miller covector,

$$\underline{f} = 3\omega^1 + 3\omega^2 + 2\omega^3 \tag{5.39}$$

The Miller indices of $\underline{f}$ are therefore

$$(h, k, \ell) = (3, 3, 2).$$

Mathematized Diffraction Law

Focus on two X-ray beams and their respective phase functions $\underline{\phi}$,

$$\begin{aligned} \psi(\vec{x})^{incident}: \quad \underline{\phi}^{inc} &= k_1^{inc}\ \omega^1 + k_2^{inc}\ \omega^2 + k_3^{inc}\ \omega^3 \\ \psi(\vec{x})^{diffracted}: \quad \underline{\phi}^{diff} &= k_1^{diff}\omega^1 + k_2^{diff}\omega^2 + k_3^{diff}\omega^3. \end{aligned} \tag{5.40}$$

To be diffracted by the set of crystal planes whose Miller covector is Eq.(5.38), the phase function of the diffracted beam must satisfy

$$\begin{aligned} \underline{\phi}^{inc} - \underline{\phi}^{diff} &= \underline{f} \\ \Delta k_1\omega^1 + \Delta k_2\omega^2 + \Delta k_3\omega^3 &= h\omega^1 + k\omega^2 + \ell\omega^3 \end{aligned} \tag{5.41}$$

where

$$\Delta k_1 \equiv (k_1^{inc} - k_1^{diff}) = h \tag{5.42}$$

$$\Delta k_2 \equiv (k_2^{inc} - k_2^{diff}) = k \tag{5.43}$$

$$\Delta k_3 \equiv (k_3^{inc} - k_3^{diff}) = \ell \tag{5.44}$$

are the components of the difference between the incident and the diffracted phase functions (in physics, a.k.a. "propagation covectors") $\underline{\phi}^{inc}$ and $\underline{\phi}^{diff}$. They also equal the $\{\vec{e}_1^*, \vec{e}_2^*, \vec{e}_3^*\}$-basis components of the normals to the set of crystal planes, Eq.(5.41). The atoms in these planes are responsible for the particular phase fronts, Eq.(5.40). The three conditions, Eq.(5.42)-(5.44), for Bragg diffraction are known as Laue's equations.

5.7.3 Sampling Theorem as a Corollary to the Duality Principle

All observations and measurements processed by our mind into concepts and knowledge are finite. Concepts such as "infinity", "limit", "continuity", "derivative", etc., are not metaphysical[8] attributes of the world, but instead are mathematical methods. They are objective in that their composite nature reflects their nature of the world and the nature of our mind in grasping it.

One of the most ubiquitous concepts in the hierarchical network of mathematical methods is that of functions continuous on, say, the interval $[0, 2\pi]$. They form an infinite-dimensional vector space, which subsumes an unlimited number of different kinds of finite-dimensional vector spaces. Among them are those subspaces which are spanned by bases that reflect the particular manner of observation or measurement, specifically those those in which a function is sampled at equal intervals, say,

$$x_k = \frac{2\pi}{2N+1}\, k \qquad k = 0, 1, \cdots, 2N.$$

Recall that a chosen basis for a given vector space induces a unique set of linear functions. They are the coordinate functions on this vector space. These functions are also vectors. In fact, they form a basis, but for the dual vector space, which is entirely distinct. Its dimension is the same as that of the given vector space. The duality relation, Eq.(5.14) on page 194 mathematizes the duality principle. The sampling theorem is an application of the dual space concept.

Example 1 (Sampling a Band-Limited Function)
GIVEN:

[8] in the Greek sense, pertaining to the nature of reality.

a) The vector space of band-limited functions of period 2π,

$$V = \{f : \ f(x) = \sum_{m=0}^{N} a_m \cos mx + \sum_{m=1}^{N} \sin mx\} \equiv \mathcal{B}_N.$$

This is a $(2N+1)$-dimensional space with its standard trigonometric basis

$$B_{trig} = \{1, \cos mx, \sin mx : \ m = 1, \cdots, N\}, \tag{5.45}$$

or its exponential basis

$$B_{exp} = \{e^{imx}\}_{m=-N}^{N}. \tag{5.46}$$

b) The values $f(x_k)$ of f at the sampling points $\{x_k = \frac{2\pi}{2N+1}k$ with $k = 0, 1, \cdots, 2N\}$

FIND:

The function $f(x)$ for *all* x in terms of its known sampled values $\{f(x_k)\}_{k=0}^{2N}$.

SOLUTION:
The task at hand consists of answering the following question: Can one reconstruct f over the whole x-domain from one's knowledge of the f-values at the $2N+1$ sample points x_k only? If, yes, HOW?
COMMENT:

This question cannot be answered without specifying a particular $(2N+1)$-dimensional subspace of $C[0, 2\pi]$, the infinite-dimensional subspace of functions continuous on $[0, 2\pi]$.

There are many such subspaces, and $V = \mathcal{B}_N$, the above space of band limited 2π-periodic functions of the present Example 1, is only one of them. Another one, considered in Example 2, below, on page 220, is $V = CPL[0, 2\pi]$, the $(2N+1)$-dimensional subspace of continuous functions piecewise linear on the closed interval $[0, 2\pi]$.

In both subspaces a vector is specified by the same $2N+1$ values of the sampled function. However, inbetween its sampling points, the function is interpolated in entirely different ways. The two subspaces are entirely different, but their dimensions are the same.

The answer to the posed question is that for sampling purposes the bases (5.45) or (5.46) on page 216 do not give good representations of elements in V. Instead, we

construct x_k-localized functions by means of the following linear superpositions

$$\vec{e}_k(x) = \frac{1}{2N+1}\sum_{m=-N}^{N} e^{im(x-x_k)}, \text{ where } x_k = \frac{2\pi}{2N+1}k \text{ with } k = 0, 1, \cdots, 2N; \tag{5.47}$$

$$= 1 + \frac{1}{2N+1}\sum_{m=1}^{N} \cos m(x - x_k) \tag{5.48}$$

These are also band-limited functions, vectors in V. In fact, being mere geometrical series, their summed values are

$$\vec{e}_k(x) = \frac{1}{2N+1}\frac{\sin\left(N+\frac{1}{2}\right)(x-x_k)}{\sin\left(\frac{x-x_k}{2}\right)}. \tag{5.49}$$

These functions are x_ℓ-localized. They satisfy

$$\vec{e}_k(x_\ell) = \delta_{k\ell} \equiv \begin{cases} 0 & \ell \neq k \\ 1 & \ell = k \end{cases}.$$

Their graphs are exhibited in Figure 5.7 and they form a basis for V,

$$C = \{\vec{e}_0, \vec{e}_1, \cdots, \vec{e}_{2N}\}.$$

The reason for introducing this basis is that (i) sampling a function at a particular point

$$x_\ell = \frac{2\pi}{2N+1}\,\ell \quad \ell = 0, 1, \cdots, 2N+1$$

constitutes a linear map on the space of functions $f \in V = \mathcal{B}_N$:

$$\omega^\ell(f) = f(x_\ell),$$

and that (ii) these linear maps, which comprise the set

$$\{\omega^\ell\}_{\ell=0}^{2N},$$

have the property that

$$\omega^\ell(e_k) = e_k(x_\ell) = \delta_{\ell k}.$$

This is the duality relation. Thus the set of sampling maps

$$\{\omega^0, \omega^1, \omega^2, \cdots, \omega^{2N}\}$$

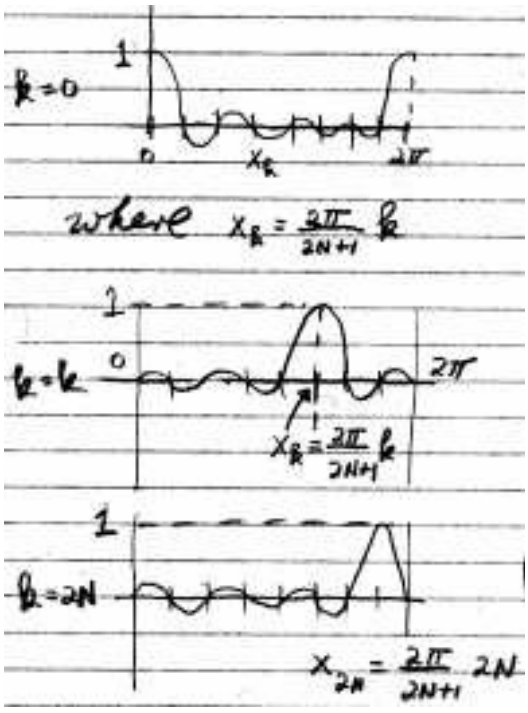


Figure 5.7: Graphs of x_k-localized functions $\vec{e}_k(x)$. The set $\{\vec{e}_k\}_{k=0}^{2N} \equiv B$ forms an alternative linearly independent spanning set (basis) for V.

are precisely the basis elements dual to the constructed x_k-localized basis

$$\{e_0, e_1, e_2, \cdots, e_{2N}\}$$

as given by Eq.(5.49).

The reason for introducing this particular basis comes from our goal to characterize an arbitrary band-limited $f \in V = \mathcal{B}_N$ in terms of its sampled values at $x = x_k,\ k = 0, 1, \cdots, 2N$:

$$f \overset{\mathcal{B}_N}{\rightsquigarrow} \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_{2N}) \end{bmatrix}_{\mathcal{B}_N}$$

There are $2N + 1$ sampled values for each and every $f \in V$. This circumstance is

mathematized by means of $2N+1$ sample-valued maps $\omega^0, \omega^1, \cdots, \omega^{2N}$ on V, namely

$$\begin{aligned}
\omega^0(f) &= f(x_0) \\
\omega^1(f) &= f(x_1) \\
&\vdots \\
\omega^\ell(f) &= f(x_\ell) \\
&\vdots \\
\omega^{2N}(f) &= f(x_{2N}).
\end{aligned}$$

These linear maps are precisely the coordinate functionals $\{\omega^\ell\}_{\ell=0}^{2N}$ induced by the vector basis $C = \{\vec{e}_k\}_{k=0}^{2N}$. This claim is validated by the fact that, according to Eq.(5.49) on page 217, the sampled values of the $\vec{e}_k$'s are

$$\vec{e}_k(x_\ell) = \delta_{k\ell} \equiv \begin{cases} 0 & \ell \neq k \\ 1 & \ell = k \end{cases},$$

in other words,

$$\omega^\ell(\vec{e}_k) \equiv \langle \omega^\ell | \vec{e}_k \rangle = \delta_{\ell k}$$

This is the duality relation: the basis $C = \{\vec{e}_k\}_{k=0}^{2N}$ has as its dual the set of basis (linear) functionals $C^* = \{\omega^\ell\}_{\ell=0}^{2N} \subset V^*$.

1. This set has two distinguishing properties:

   (a) On one hand each ω^ℓ *samples* any $f(x) \in V$ at $x = x_\ell$, and thereby yields $\omega^\ell(f) = f(x_\ell)$, the ℓ^{th} coordinate of $f \in V$ relative to B;

   (b) on the other hand, and at the same time, each ω^ℓ *is a covector* which, together with the other elements in B^*, forms that basis for V^* which is dual to B.

2. The success of the sampling theorem hinges on the existence of the above two features:

   (a) The $\vec{e}_k$'s must form a basis for V. Consequently, one has

   $$f(x) = \sum_{k=0}^{2N} \alpha_k \vec{e}_k,$$

   and

(b) each $\vec{e}_k$ is a function with zero values at all equally spaced points x_ℓ, except one, where its value does not vanish. Consequently,

$$\underbrace{\omega^\ell(f)}_{f(x_\ell)} = \sum_{k=0}^{2N} \alpha_k \underbrace{\omega^\ell(\vec{e}_k)}_{\delta_k^\ell} = \sum_{k=0}^{2N} \alpha_k \underbrace{\vec{e}_k(x_\ell)}_{\delta_k^\ell}$$

or

$$f(x_\ell) = \alpha_\ell$$

Consequently, $f(x)$ is given by

$$f(x) = \sum_{k=0}^{2N} f(x_k)\, \vec{e}_k,$$

a mathematically 100% accurate reconstruction of $f(x)$ in terms of its sampled values. This is the sampling theorem for band-limited functions $\mathcal{B}_N$.

Example 2 (Piecewise Linear Function via a Sampling Sequence)
GIVEN:

1. The closed interval $[x_0, x_n]$ which is partitioned by $x_0 < x_1 < \cdots < x_n$ into n *equally spaced* subintervals.

2. The values $y_0, y_1, \cdots, y_n$ of a function $f \in C[x_0, x_n]$ sampled at the above equally spaced points:

$$\begin{array}{rl} y_0 = & f(x_0) \\ y_0 = & f(x_1) \\ & \vdots \\ y_k = & f(x_k) \\ & \vdots \\ y_n = & f(x_n) \end{array} \tag{5.50}$$

3. The set

$$CPL(\{x_0, x_1, \cdots, x_n\}) = \{\psi : \ \psi \in C[x_0, x_n] \text{ and } \psi \text{ is linear on each subinterval } [x_{k-1}, x_k]\},$$

which, being closed under addition and multiplication by scalars, is a subspace of $C[x_0, x_n]$.

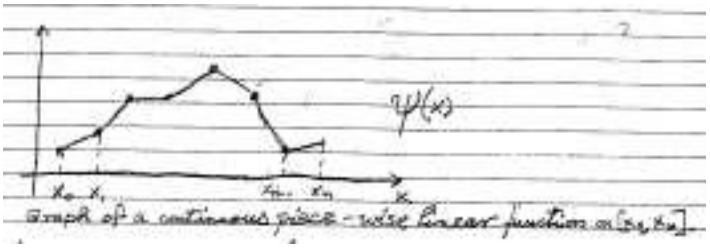


Figure 5.8: Graph of a CPL-function, an element of the vector space $CPL \subseteq C[x_0, x_n]$

EXHIBIT:

1. A basis for CPL whose elements $\psi_k(x)$ (like those of Eq.(5.49)) are x_ℓ-localized:

$$\psi_k(x_\ell) = \delta_{k\ell} \equiv \begin{cases} 0 & \ell \neq k \\ 1 & \ell = k \end{cases}.$$

2. The dual basis $\{\omega^j\}_{j=1}^n$ for CPL^*

3. For the given sampling sequence, Eq.(5.50) of the function $f \in C[x_0, x_n]$, the function $\psi(x) \in CPL$ such that

$$\begin{aligned} \psi(x_0) &= y_0 \\ &\vdots \\ \psi(x_n) &= y_n \end{aligned}$$

SOLUTION:
1-2:

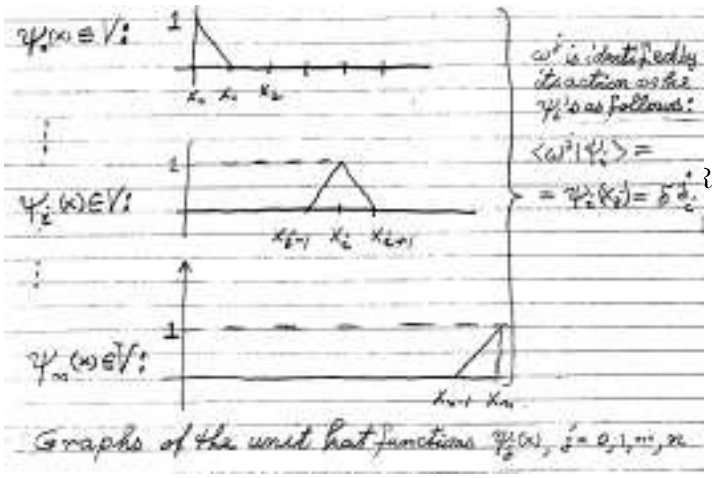


Figure 5.9: Graphs of the x_k-localized unit roof functions. They comprise a basis for CPL.

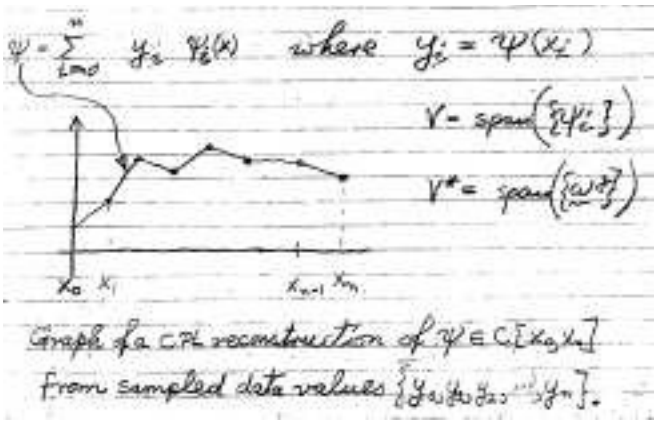


Figure 5.10: Graph of CPL-interpolation of sampled data $y_0, y_1, \cdots, y_n$.

5.8 PROBLEMS

1. (DUAL BASIS AS A SET OF MULTIVARIABLE FUNCTIONS)
   Let $B = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be a basis for R^3 defined by

$$\vec{v}_1 = (1, 0, 1)^t$$
$$\vec{v}_2 = (1, 1, 1)^t$$
$$\vec{v}_3 = (2, 2, 0)^t$$

   a) FIND the basis $\{f, g, h\}$ of linear functions (i.e. row vectors) dual to B.
   b) EXHIBIT

$$f(\vec{x}) = f(x, y, z)$$
$$g(\vec{x}) = g(x, y, z)$$
$$h(\vec{x}) = h(x, y, z)$$

2. (VECTOR BASES AND THEIR DUAL BASES)
Consider 3-dimensional vector space spanned by

$$\vec{e}_1 = \vec{i} + \vec{j} + \vec{k}$$
$$\vec{e}_2 = -\vec{i} + \vec{j} + \vec{k}$$
$$\vec{e}_3 = -\vec{i} - \vec{j} + \vec{k}$$

where $\vec{i}, \vec{j}, \vec{k}$ are the usual orthogonal basis vectors.

(a) If $\{\underline{\tau}^1, \underline{\tau}^2, \underline{\tau}^3\}$ is the basis *dual* to $\{i, j, k\}$, i.e.

$$\langle \underline{\tau}^1 | \vec{i} \rangle = 1 \qquad \langle \underline{\tau}^1 | \vec{j} \rangle = 0 \qquad \langle \underline{\tau}^1 | \vec{k} \rangle = 0$$
$$\langle \underline{\tau}^2 | \vec{i} \rangle = 0 \qquad \langle \underline{\tau}^2 | \vec{j} \rangle = 1 \qquad \langle \underline{\tau}^2 | \vec{k} \rangle = 0$$
$$\langle \underline{\tau}^3 | \vec{i} \rangle = 0 \qquad \langle \underline{\tau}^3 | \vec{j} \rangle = 0 \qquad \langle \underline{\tau}^3 | \vec{k} \rangle = 1$$

FIND the basis $\{\underline{\omega}^1, \underline{\omega}^2, \underline{\omega}^3\}$ *dual* to $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$.

(b) Let h, k, ℓ be three scalars.
FIND that linear function, call it $\underline{f}$, which has the value 1 at each of the three points $\frac{\vec{e}_1}{h}, \frac{\vec{e}_2}{k}, \frac{\vec{e}_3}{\ell}$.
Thus *write down* this function in terms of $\{\underline{\tau}^1, \underline{\tau}^2, \underline{\tau}^3\}$ and in terms of $\{\underline{\omega}^1, \underline{\omega}^2, \underline{\omega}^3\}$.

(c) FIND the set of reciprocal basis vectors (e_1^*, e_2^*, e_3^*), which satisfy

$$\vec{e}_i \cdot \vec{e}_j^* = \delta_{ij} \, .$$

(Here "$\cdot$" is the familiar inner product obtained from $\vec{i} \cdot \vec{i} = 1$, $\vec{i} \cdot \vec{j} = 0$, $\vec{i} \cdot \vec{k} = 0$, etc.)

(d) What relation, if any, does there exist between these basis vectors $\vec{e}_1^*$, $\vec{e}_2^*$, and $\vec{e}_3^*$ and the level surfaces of $\underline{\omega}^1$, $\underline{\omega}^2$, and $\underline{\omega}^3$?

(e) FIND the unit vector perpendicular to the level surface $\underline{f} = 1$.

(f) Write down the distance from the origin to $\underline{f} = 1$.
REMINDER: If you get bogged down in detailed computation, you are not making optimal use of the nature of the dual basis!

Comment: The components of $\underline{f}$ found in (b) relative to $\{\underline{\omega}^i\}$ are the *Miller indices* of a set of parallel planes in a crystal whose primitive translation vectors are $\vec{e}_1, \vec{e}_2, \vec{e}_3$.
SEE C. KITTEL, Introduction to SOLID STATE PHYSICS.

3. (DEFINITE INTEGRALS DUALS ON THE SPACE OF POLYNOMIALS) Let $V = \mathcal{P}_2$ be the vector space of all polynomial functions p from R into R which have degree 2 or less:

$$p(x) = a_0 + a_1 x + a_2 x^2 \,.$$

Define the following linear functionals on V by

$$\begin{aligned} \langle f_1 \mid p \rangle \equiv f_1(p) &= \int_0^1 p(x)dx \\ \langle f_2 \mid p \rangle \equiv f_2(p) &= \int_0^2 p(x)dx \\ \langle f_3 \mid p \rangle \equiv f_3(p) &= \int_0^{-1} p(x)dx \,. \end{aligned}$$

SHOW that $\{f_1, f_2, f_3\}$ is a basis for V^* by exhibiting the basis for V of which it is the dual.

5.9 Tensor Algebra

Lecture 26

By generalizing linear and bilinear maps one arrives at the idea of a tensor. In fact, a tensor is a multilinear map. The idea of multilinearity is illustrated by the following
Example (*The determinant*)
Consider the determinant each of whose rows is the set of components of a vector:

$$\left.\begin{array}{llll} \vec{A}_1: & A_1^1, & A_1^2, \cdots, & A_1^n \\ \vec{A}_2: & A_2^1, & A_2^2, \cdots, & A_2^n \\ \vdots & & & \\ \vec{A}_n: & A_n^1, & A_n^2, \cdots, & A_n^n \end{array}\right\} \overset{\det}{\rightsquigarrow} \det(\vec{A}_1, \cdots, \vec{A}_n) \,. \tag{5.51}$$

This determinant

$$\det(\vec{A}_1, \cdots, \vec{A}_n) = \begin{vmatrix} A_1^1, \cdots, & A_1^n \\ \vdots & \\ A_n^1, \cdots, & A_n^n \end{vmatrix} \in R$$

is a multilinear map,

$$V \times \cdots \times V \overset{\det}{\to} R$$

because

$$\det(\cdots, \alpha\vec{A}_i + \beta\vec{B}_i, \cdots) = \alpha \det(\cdots, \vec{A}_i, \cdots) + \beta \det(\cdots, \vec{B}_i, \cdots) \ .$$

The most encompassing definition of a multilinear map does not stipulate the dimensionality of each vector, nor does it demand a choice of basis.

5.9.1 Tensor as a Multilinear Map

By generalizing bilinear maps to multilinear maps one arrives at the idea of a tensor. In fact, a tensor is a multilinear map. We start with the concept of multilinearity.
Definition (*Multilinearity*)
Let $V_1, V_2, \cdots, V_q$ be vector spaces. Then the map

$$\begin{aligned} \mathbf{H}: \ & V_1 \times V_2 \times \cdots \times V_q \to R \\ & (v_1, v_2, \cdots, v_q) \rightsquigarrow \mathbf{H}(v_1, v_2, \cdots, v_q) \end{aligned}$$

is said to be *multilinear* if it is linear in each of its arguments:

$$\mathbf{H}(v_1, \cdots, \alpha v_i + \beta w_i, \cdots, v_q) = \alpha \mathbf{H}(v_1, \cdots, \alpha v_i, \cdots, v_q) + \beta \mathbf{H}(v_1, \cdots, \beta w_i, \cdots, v_q) \quad \forall\, 1 \le i \le q$$

Definition (*Tensor*)
Let

$$\begin{aligned} V_1 = \ & \cdots \ = V_n \ & = V* \\ V_{n+1} = \ & \cdots \ = V_{n+m} \ & = V \end{aligned}$$

then the multilinear map

$$\mathbf{H}: \ \underbrace{V^* \times V^* \times \cdots \times V^*}_{\equiv (V^*)^n \ n \text{ copies}} \times \underbrace{V \times V \times \cdots \times V}_{\equiv V^m \ m \text{ copies}} \overset{\mathbf{H}}{\to} R \tag{5.52}$$

$$(\underbrace{\underline{\sigma}, \underline{\lambda}, \cdots, \underline{\beta}}_{n \text{ copies}}, \underbrace{\mathbf{u}, \mathbf{v}, \cdots, \mathbf{w}}_{m\text{vectors}}) \overset{\mathbf{H}}{\rightsquigarrow} \mathbf{H}(\underline{\sigma}, \underline{\lambda}, \cdots, \underline{\beta}, \mathbf{u}, \mathbf{v}, \cdots, \mathbf{w}) \tag{5.52$'$}$$

is a tensor of rank $\binom{n}{m}$. Here n and m are called the "contravariant rank" and the "covariant rank" of $\mathbf{H}$.

Comment One can add tensors of rank $\binom{n}{m}$ as well or mutiply them by a scalar. The result is still a tensor of the same rank. Thus $\binom{n}{m}$-rank tensors form a vector space, the *tensor space* of $\binom{n}{m}$-rank tensors. From Eqs.(5.52) one finds that its dimension is $(\dim V)^{n+m}$.

Examples of Tensors

Name	Symbol	Mapping	Rank
covector	$\underline{\omega}$	$V \to R$ $\mathbf{v} \rightsquigarrow \underline{\omega}(\mathbf{v}) \equiv \langle \underline{\omega} \vert \mathbf{v} \rangle$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
metric	$\mathbf{g}$	$V \times V \to R$ $(\mathbf{u}, \mathbf{v}) \rightsquigarrow \mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$	$\begin{pmatrix} 0 \\ 2 \end{pmatrix}$
vector	w	$V^* \to R$ $\underline{\sigma} \rightsquigarrow \mathbf{w}(\underline{\sigma}) = \underline{\sigma}(\mathbf{w}) \equiv \langle \underline{\sigma} \vert \mathbf{w} \rangle$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
inverse metric	$\mathbf{g}^{-1}$	$V^* \times V^* \to R$ $(\underline{f}, \underline{h}) \rightsquigarrow \mathbf{g}^{-1}(\underline{f}, \underline{h})$	$\begin{pmatrix} 2 \\ 0 \end{pmatrix}$

5.9.2 Coordinate Components of a Tensor

A vector has components relative to a given basis. So does a covector. This concept "components relative to a basis" can be extended to tensors. The basis-dependent coordinate components of a tensor are obtained by projecting them out with the basis elements as follows:

Definition *(Tensor Components Relative to a Given Basis)*
Let $\{e_i\}$ be a basis for V.
Let $\{\omega^j\}$ be its dual basis for V^*.
Then the numbers

$$\mathbf{H}(\omega^{j_1}, \omega^{j_2}, \cdots, \omega^{j_n}, e_{i_1}, e_{i_2}, \cdots, e_{i_m}) \equiv H^{j_1 \cdots j_n}{}_{i_1 \cdots i_m} \tag{5.53}$$

are the components of $\mathbf{H}$ relative to the given basis.
Examples

1. Let $\boldsymbol{\alpha}$ be a tensor of rank $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then its coordinate components are

$$\begin{aligned} \boldsymbol{\alpha}(e_k) &= \alpha_j \underbrace{\omega^j(e_k)}_{\delta^j{}_k} = \alpha_k \\ &= k^{th} \text{ component relative to } \{\omega^j\}. \end{aligned}$$

2. Let $\mathbf{u}$ be a tensor of rank $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then its coordinate components are

$$\begin{aligned} \mathbf{u}(\omega^k) &= u^i e_i(\omega^k) \\ &\equiv u^i \underbrace{\omega^k(e_i)}_{\delta_i{}^k} = u^k \\ &= k^{th} \text{ component relative to } \{e_i\}. \end{aligned}$$

3. Let $\mathbf{g}$ be the metric, a tensor of rank $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Then its coordinate components are

$$\begin{aligned} \mathbf{g}(e_k, e_\ell) &= g_{k\ell} \\ &= (k\ell)^{th} \text{ component relative to } \{e_i\} \ . \end{aligned}$$

Comment 1: It is understood, but worthwhile reemphasizing, that the tensor $\mathbf{H}$ is a geometrical object, by which one means that it is a concept which unites all possible basis representations under one roof, namely the multilinear map $\mathbf{H}$ of rank $\begin{pmatrix} n \\ m \end{pmatrix}$.
Comment 2: The result of evaluating $\mathbf{H}$ on some arbitrary $(n+m)$-tuple $(\underline{\sigma}, \cdots, \underline{\beta}, \mathbf{u}, \cdots, \mathbf{w})$ is

$$\mathbf{H}(\sigma_{j_1}\omega^{j_1}, \cdots, \beta_{j_n}\omega^{j_n}, u^{i_1}e_{i_1}, \cdots, u^{i_m}e_{i_m}) = H^{j_1\cdots j_n}{}_{i_1\cdots i_m}\sigma_{j_1}\cdots\beta_{j_n}u^{i_1}\cdots u^{i_m} \ ,$$

which is an $(n+m)$-fold sum.
Comment 3: It is worthwhile to remind ourselves that

(a) the distinction of upper vs. lower indices is to be observed with rigid rigor. This is because such positioning is a defense against potential logical disasters.

(b) The summation (a.k.a. "dummy") indices appear in distinct pairs. This prevents defective thinking in regard to multiple sums.

Lecture 27

5.9.3 The Tensor Product

In 3-dimensional Euclidean space consider a vector $\vec{v}$ rotating with angular velocity[9] $\vec{\omega}$ around a given axis. The vectorial change $\Delta\vec{v}$ of this $\vec{v}$ during time interval Δt is

$$\Delta\vec{v} = \Delta t \quad \vec{\omega} \times \vec{v} \ .$$

[9]The components of $\vec{\omega}$, namely the coefficients $\omega^1, \omega^2, \omega^3$ in Eq.(5.54) are not to be confused with the basis elements dual to the given basis.

In terms of orthonormal basis vectors this cross product has the form

$$= \Delta t \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ \omega^1 & \omega^2 & \omega^3 \\ v^1 & v^2 & v^3 \end{vmatrix} . \tag{5.54}$$

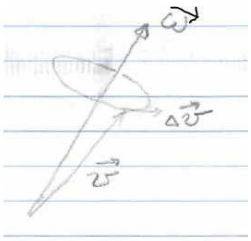


Figure 5.11: Vectorial change $\Delta\vec{v}$ in $\vec{v}$ due to its rotation around $\vec{\omega}$. The tip of $\vec{v}$ moves in a plane perpendicular to $\vec{\omega}$ and the tip's angle of rotation during time interval Δt is $|\vec{\omega}|\Delta t$.

Expand Eq.(5.54) in terms of the orthonormal basis vectors and find

$$\Delta\vec{v} = \Delta t \left[-\omega^1(\vec{e}_2 v^3 - \vec{e}_3 v^2) + \omega^2(\vec{e}_1 v^3 - \vec{e}_3 v^1) - \omega^3(\vec{e}_1 v^2 - \vec{e}_2 v^1) \right] .$$

Expressing the components of $\vec{v}$ in terms of inner products yields

$$= \Delta t \left[-\omega^1(\vec{e}_2 \otimes \vec{e}_3 - \vec{e}_3 \otimes \vec{e}_2) + \omega^2(\vec{e}_1 \otimes \vec{e}_3 - \vec{e}_3 \otimes \vec{e}_1) - \omega^3(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \right] \cdot \vec{v} \tag{5.55}$$

The bivector $[\cdots]$ is a linear combination of "tensor products". There are three of them. The difference $(\vec{e}_2 \otimes \vec{e}_3 - \vec{e}_3 \otimes \vec{e}_2)$ generates rotation in the plane spanned by $\vec{e}_2$ and $\vec{e}_3$. The coeffient $-\Delta t\, \omega^1$ is the angular amount of that rotation, and similarly for the other pairs of spanning vectors. The sum total in $[\cdots]$ of Eq.(5.55) is the rotation in the plane perpendicular to the rotation axis.

The mathematical generalization of those "tensor products" is given by the following

Definition (*Tensor Product*[10])
Let $\vec{a}, \vec{b}, \cdots, \vec{c} \in V$
and $\underline{\alpha}, \underline{\beta}, \cdots, \underline{\gamma} \in V^*$

[10]not to be confused with the Catesian Product.

The multilinear map

$$\underbrace{\vec{a}\otimes\vec{b}\otimes\cdots\otimes\vec{c}}_{n}\otimes\underbrace{\underline{\alpha}\otimes\underline{\beta}\otimes\cdots\otimes\underline{\gamma}}_{m}:$$

$$\underbrace{V^*\times V^*\cdots\times V^*}_{n}\times\underbrace{V\times V\times\cdots\times V}_{m}\to R$$

$$(\ \underbrace{\underline{\sigma},\underline{\rho},\cdots,\underline{\tau}}_{n}\ ,\ \underbrace{\vec{u},\vec{v},\cdots,\vec{w}}_{m}\ )\rightsquigarrow\langle\underline{\sigma}|\vec{a}\rangle\langle\underline{\rho}|\vec{b}\rangle\cdots\langle\underline{\tau}|\vec{c}\rangle\langle\underline{\alpha}|\vec{u}\rangle\langle\underline{\beta}|\vec{v}\rangle\cdots\langle\underline{\gamma}|\vec{w}\rangle$$

is the tensor product $\vec{\alpha},\cdots,\gamma$. It is a tensor of rank $\begin{pmatrix} n \\ m \end{pmatrix}$.

5.9.4 Tensors: Their Basis Representation

Having generated $\begin{pmatrix} n \\ m \end{pmatrix}$-rank tensors by taking tensor products, one can take linear combinations of products of basis elements and thereby decompose any tensor into a unique linear combination of such basis products. In fact, a linear combination of tensor products of basis elements yields a (chosen/given) basis representation of an $\begin{pmatrix} n \\ m \end{pmatrix}$-rank tensor. This idea is condensed into the following

Proposition: (*Basis Representation of a Tensor*)
GIVEN:

(i) a basis $\{e_i\}$ of V, and hence also its dual basis ω^j for V^*.

(ii) the tensor $\mathbf{H}$ of rank $\begin{pmatrix} n \\ m \end{pmatrix}$.

CONCLUSION:

$$\boxed{\mathbf{H}=H^{j_1\cdots j_n}{}_{i_1\cdots i_n}\vec{e}_{j_1}\otimes\cdots\otimes\vec{e}_{j_n}\otimes\underline{\omega}^{i_1}\otimes\cdots\otimes\underline{\omega}^{i_m}} \tag{5.56}$$

This is the representation of $\mathbf{H}$ in terms of the basis and its dual. The validity of this boxed equation depends on showing that the value of the linear map on the l.h.s. equals the value of the linear map on the r.h.s. for all $(n+m)$-tuples of covectors and vectors.

To concretize this line of reasoning apply it to the archetypical case of a tensor of rank $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$,

$$\mathbf{H}\ :V^*\times V\to R$$

for which one must show that

$$\boxed{\mathbf{H} = H^{j}{}_{i} e_j \otimes \underline{\omega}^{i}} \tag{5.57}$$

The validation consists of showing that

$$\begin{array}{rcl} \mathbf{H}(\underline{\sigma}, \vec{v}) & = & H^{j}{}_{i} e_j \otimes \underline{\omega}^{i}(\underline{\sigma}, \vec{v}) \\ & & \uparrow \\ & & \forall\, (\underline{\sigma}, \vec{v}) \in V^* \times V \ . \end{array} \tag{5.58}$$

Because $\mathbf{H}$ is linear, it suffices to show this for $(\underline{\omega}^{j'}, \vec{e}_{i'}) \in V^* \times V$. This is a two-step process.
(i) One observes immediately that

$$\mathbf{H}(\underline{\omega}^{j'}, \vec{e}_{i'}) = H^{j'}{}_{i'} \ , \tag{5.59}$$

which, we recall from Eq.(5.53) on page 226, is the $(j', i')^{th}$ component of $\mathbf{H}$.
(ii) On the other hand, one has

$$\begin{array}{rcll} \mathbf{H}(\underline{\omega}^{j'}, \vec{e}_{i'}) & = & H^{j}{}_{i} \ \ \vec{e}_j \otimes \underline{\omega}^{i}(\underline{\omega}^{j'}, \vec{e}_{i'}) & \\ & = & H^{j}{}_{i} \ \ \vec{e}_j(\underline{\omega}^{j'}) \ \ \underline{\omega}^{i}(\vec{e}_{i'}) & \text{(definition of a tensor product)} \\ & \equiv & H^{j}{}_{i} \ \ \underbrace{\langle \underline{\omega}^{j'} | \vec{e}_j \rangle}_{\delta^{j'}{}_{j}} \ \ \underbrace{\langle \underline{\omega}^{i} | \vec{e}_{i'} \rangle}_{\delta^{i}{}_{i'}} & \text{(linear transformation)} \\ & & & \text{(duality)} \\ & = & H^{j'}{}_{i'} \ . & \end{array} \tag{5.60}$$

Equations (5.59) and (5.60) hold for all basis elements, and as already mentioned, because of the linearity of $\mathbf{H}$, it holds for all pairs $(\underline{\sigma}, \vec{v}) = (\sigma_{j'} \underline{\omega}^{j'}, v^{i'} \vec{e}_{i'})$ in Eq.(5.58). Thus Eq.(5.57) is a valid tensorial basis expansion indeed.

5.9.5 Examples of Tensorial Basis Expansions

Metric Tensor

$$\mathbf{g} = g_{ij} \underline{\omega}^{i} \otimes \underline{\omega}^{j}$$

Inverse Metric Tensor

$$\mathbf{g}^{-1} = g^{ij} \vec{e}_i \otimes e_j$$

Cartan's Unit Tensor

$$\begin{array}{rcl} d\mathcal{P} & = & \delta^{i}{}_{j} \underline{\omega}^{j} \otimes \vec{e}_i \\ & = & \underline{\omega}^{i} \otimes \vec{e}_i \end{array}$$

The Totally Antisymmetric (Levi-Civita) Tensor (a.k.a. the Volume Tensor in n-Dimensions

$$\boldsymbol{\epsilon} = \epsilon_{i_1\cdots i_n}\underline{\omega}^{i_1}\otimes\cdots\otimes\underline{\omega}^{j_1}\ .$$

Here $\epsilon_{i_1\cdots i_n}$ is the totally antisymmetric (Levi-Civita) symbol,

$$\epsilon_{i_1\cdots i_n} = \begin{cases} 0 & \text{if any pair of indices are the same} \\ +\epsilon_{1\cdots n} & \text{if } i_1\cdots i_n \text{ is an even permutation of } 1\cdots n \\ -\epsilon_{1\cdots n} & \text{if } i_1\cdots i_n \text{ is an odd permutation of } 1\cdots n \end{cases} \tag{5.61}$$

An equivalent but basis ("frame") independent definition is

$$\boldsymbol{\epsilon}(\vec{A}_1,\cdots,\vec{A}_n) = \det\begin{vmatrix} \underline{\omega}^1(\vec{A}_1) & \underline{\omega}^2(\vec{A}_1) & \cdots & \underline{\omega}^n(\vec{A}_1) \\ \underline{\omega}^1(\vec{A}_2) & \underline{\omega}^2(\vec{A}_2) & \cdots & \underline{\omega}^n(\vec{A}_2) \\ \vdots & & & \\ \underline{\omega}^1(\vec{A}_n) & \underline{\omega}^2(\vec{A}_n) & \cdots & \underline{\omega}^n(\vec{A}_n) \end{vmatrix} = \begin{array}{l}\text{volume of a} \\ \text{parallelopiped} \\ \text{subtended by} \\ \{\vec{A}_i\}_{i=1}^n \text{ in } R^n\end{array}$$

5.9.6 Tensor Space

Tensors of rank $\begin{pmatrix} n \\ m \end{pmatrix}$ can be added and multiplied by scalars. This feature is already implied when the tensorial basis expasions were introduced in subsections 5.9.4 and 5.9.5. More formally one has the following

Proposition (*Tensor Space*)

Tensors of rank $\begin{pmatrix} n \\ m \end{pmatrix}$ form a vector space. This space of tensors is denoted by the tensor space $\underbrace{V\otimes\cdots\otimes V}_{n \text{ factors}}\otimes\underbrace{V^*\otimes\cdots\otimes V^*}_{m \text{ factors}}$.

5.9.7 Unit-Economy via New Tensors

The complexity of mathematics is a reflection of the causal relations that exist in the universe. Linearity, which includes the algebra of tensors, is one of the forms of grasping and conquering this complexity. The power of tensor algebra lies in its conceptual nature: new tensors can be formed by addition and by (scalar and tensorial) multiplication,as in subsections 5.9.3 and 5.9.4. However, there are other ways of forming tensors of different rank.

Raising and Lowering Tensor Indices

Given a metric

$$\mathbf{g}:\ V\to V^*$$

one has

$$\vec{u} = u^i \vec{e}_i \rightsquigarrow \underline{u} = u_j \underline{\omega}^j$$
$$\{u^i\} \rightsquigarrow \{u_j\}$$

where

$$u_j = g_{ji} u^i \ .$$

This causal relation[11] between $\vec{u}$ and $\underline{u}$, as well as between the u^i's and the u_i's, is called lowering the indices. This correspondence is generalized to tensors as follows:
Proposition (*Lowering of Indices*)
g lowers the indices of a tensor:

$$\mathbf{g}: \underbrace{V \otimes \cdots \otimes V}_{n \text{ factors}} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{m \text{ factors}} \to \underbrace{V \otimes \cdots \otimes V}_{(n-1) \text{ factors}} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{(m+1) \text{ factors}}$$
$$\begin{pmatrix} n \\ m \end{pmatrix} \text{tensors} \to \begin{pmatrix} n-1 \\ m+1 \end{pmatrix} \text{tensors} \ .$$

Explicitly one has

$$\mathbf{H} = H^{j_1 \cdots j_{n-1} j_n}{}_{i_1 \cdots i_n} \vec{e}_{j_!} \otimes \cdots \otimes \vec{e}_{j-1} \otimes \boxed{\vec{e}_{j_n}} \otimes \underline{\omega}^{i_1} \otimes \cdots \otimes \underline{\omega}^{i_m} \rightsquigarrow$$
$$H^{j_1 \cdots j_{n-1}}{}_{j_n i_1 \cdots i_n} \vec{e}_{j_!} \otimes \cdots \otimes \vec{e}_{j-1} \otimes \boxed{\underline{\omega}^{j_n}} \otimes \underline{\omega}^{i_1} \otimes \cdots \otimes \underline{\omega}^{i_m} \ ,$$

or in terms of the coordinate components with their n^{th} superscript (j_n) lowered,

$$\{H^{j_1 \cdots j_{n-1} j_n}{}_{i_1 \cdots i_n}\} \rightsquigarrow \{H^{j_1 \cdots j_{n-1} k}{}_{i_1 \cdots i_m} g_{k j_n}\} \equiv \{H^{j_1 \cdots j_{n-1}}{}_{j_n i_1 \cdots i_m}\}$$

Contraction of a Tensor

Regardless of what the metric on V is, one can lower the rank of a tensor by the contraction map as follows:
Definition (*Contraction of a Tensor*)
The contraction map C is an operation which equates one of the superscripts to one of the subscripts and then sums over the dimension of the vector space:

$$C: \begin{pmatrix} n \\ m \end{pmatrix} \text{tensors} \to \begin{pmatrix} n-1 \\ m-1 \end{pmatrix} \text{tensors} \ .$$
$$\{H^{j_1 \cdots j_{n-1} j_n}{}_{i_1 \cdots i_n}\} \rightsquigarrow \{H^{j_1 \cdots j_{n-1} k}{}_{k\, i_2 \cdots i_m}\} \ .$$

Lecture 28

[11] Quite generally, in the relation $y = f(x)$ in which expresses an action whose cause is x and whose effect is y.

5.10 Tensor Calculus

Tensor calculus integrates the algebra of multilinear maps with multivariable differential and integral calculus.

Following Aristotle's observation that all concepts have their ultimate basis in sense perceptions, A. Einstein in 1907 revolutionized theoretical physics by mathematizing an accelerated frame of reference in terms of a one-parameter family of instantaneous inertial frames as defined by Newton's 1^{st} law of motion of freely floating bodies.

This motivated E. Cartan, a one-sided penfriend (Cartan $\rightarrow$ Einstein) of Einstein, to introduce a perspective which revolutionized 20^{th} century mathematics. He did this by assigning to each point (event) of a domain (spacetime) the vector space of the tangents to the curves (trajectories of particles) running through this point. The concept implied by this assignment is that of the tangent space at a point. When distributed over the domain, it generalizes to a "tangent bundle", to a "vector bundle", or most generally to a "fiber bundle" in local and global modern (post WW II) differential geometry[12].

We shall develop the distribution of vectors and their covector counterparts from the post-WW II perspective. In doing so we shall take cognizance of the hierarchical nature of the concepts to be mathematized.

First focus on entities – bodies/particles in motion – and thus on the tangents to their trajectories in Section 5.10.1 below. Then, via differentials in Section 5.10.2, on the change in the ambient properties (e.g. pressure, temperature, electric field, etc.) permeating the domain that accommodates these moving particles.

5.10.1 Vector as a Derivation

Consider a scalar function $f(x)$ and a curve $c(t)$ in the neighborhood of the reference point $x = a$, which we shall take as the starting point for the curve,

$$c(t) : \ c(0) = a \ .$$

Relative to a given/chosen coordinate system, f and c are represented by

$$f(x) = f(x^1, \cdots, x^n)$$
$$c(t) = \{c^j(t)\}_{j=1}^n \ .$$

[12]The fundamental novelty of this perspective was not recognized initially because Cartan invariably presented it by implication instead of explicitly. However, after WW II C.Chevalley and S.S. Chern made it explicit to the mathematicians, and C.W.Misner in 1963 to the physicists.

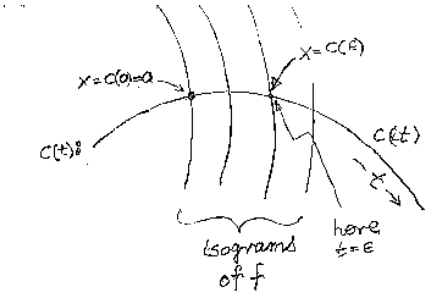


Figure 5.12: Curve $c(t)$ passing through the isograms of f.

The two Taylor series for these functions are

$$f(x) - f(a) = \left.\frac{\partial f}{\partial x^i}\right|_a (x^i - a^i) + \left.\frac{\partial^2 f}{\partial x^i \partial x^j}\right|_a \frac{(x^i - a^i)(x^j - a^j)}{2!} + \cdots \tag{5.62}$$

$$c^i(\epsilon) = c^i(0) + \left.\frac{dc^i}{dt}\right|_0 \epsilon + \left.\frac{d^2c^i}{dt^2}\right|_0 \frac{\epsilon^2}{2!} + \cdots ; \quad i = 1, \cdots, n \tag{5.63}$$

The most important part of these series is their *Principal Linear Part* (PLP). It is a mathematical "green light" on the road to the unification of linear algebra and calculus. The avievement of this goal is mediated by exhibiting two vector spaces:
(i) the vector space of tangents of preexisting curves through a point and
(ii) the dual space of differentials of preexisting functions on the neighborhood of this point.

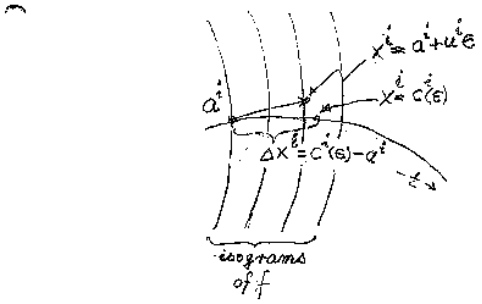


Figure 5.13: Value of f at $x^i = c^i(\epsilon)$ based on the Taylor series expansion of f vs value of f at $x^i = a^i + \vec{u}^i\epsilon$ based on the *Principal Linear Part* (PLP) of that series. The value difference at the two points is due to the non-linear part of the Taylor series.

To do this highlight the PLP of the Taylor series Eq.(5.63) of $c(t)$ in a neighbor-

hood around $c(t=0)=a$:

$$\underbrace{\left[c^i(\epsilon)-a^i\right]}_{\Delta x^i} = \underbrace{\left.\frac{dc^i}{dt}\right|_{t=0}\times\epsilon}_{\equiv\ u^i\epsilon} + \left.\frac{d^2c^i}{dt^2}\right|_{t=0}\times\frac{\epsilon^2}{2!}+\cdots; \quad i=1,\cdots,n\ .$$

$\equiv u^i\epsilon$ = PLP of the Taylor series expansion of $c(t)$

Introduce it into the series expansion, Eq.(5.62), including its PLP:

$$f\left(c(\epsilon)\right)-f(a) = \underbrace{\left.\frac{\partial f}{\partial x^i}\right|_a \underbrace{\left[c^i(\epsilon)-a^i\right]}_{\Delta x^i}}_{\equiv\ \text{PLP of the Taylor series expansion of } f} + \left.\frac{\partial^2 f}{\partial x^i\partial x^j}\right|_a \frac{\Delta x^i\Delta x^i}{2!}+\cdots \tag{5.64}$$

or

$$f\left(c^i(0)+u^i\epsilon+\cdots\right)-f\left(c^i(0)\right) = \left.\frac{\partial f}{\partial x^i}\right|_a\left[u^i\epsilon+\ddot{c}^i(0)\frac{\epsilon^2}{2!}+\cdots\right] + \left.\frac{\partial^2 f}{\partial x^i\partial x^j}\right|_a\frac{\Delta x^i\Delta x^j}{2!}+\cdots$$

Divide by ϵ and go to the limit $\epsilon\to 0$. Only the PLP of the r.h.s. gives a nonzero contribution to the derivative $\frac{df(c(t))}{dt}$, the rate of change of f into the direction $\vec{u}$ of the curve $c(t)$ when $t=0$:

$$\left.\frac{df\left(c(t)\right)}{dt}\right|_{t=0} = \lim_{\epsilon\to 0}\frac{f\left(c(\epsilon)\right)-f\left(c(t)\right)}{\epsilon} = u^i\big|_{t=0}\left.\frac{\partial f}{\partial x^i}\right|_{a=c(0)}\ . \tag{5.65}$$

One therefore concludes that the existence of a Taylor series for smooth functions guarantees that their respective rates of change at a point is given by Eq.(5.65). Based on this observation one now forms a new concept by means of the following process:

(i) In one's mind isolate two or more functions by means of one attribute common to them, namely the existence of their respective rates of change, Eq.(5.65), at point a (which is their common distinguishing feature),

(ii) retain their existence while

(iii) omitting the particular rate of change values on the principle that these values must have some value but may have any value[13].

This line of easoning yields the following
Definition (*Tangent to s Curve*)
Given: A curve $c(t) = (c^1(t), \cdots, c^n(t))$ with $c(0) = a$ and $\dot{c}^i(0) = u^i$. The mapping

$$\vec{u} = u^i \frac{\partial}{\partial x^i} : \quad C^\infty \to R \tag{5.66}$$

$$f \rightsquigarrow u^i \left.\frac{\partial f}{\partial x^i}\right|_a \equiv \underbrace{\vec{u}(f)}_{\substack{\text{unit-economy} \\ \text{notation}}} \tag{5.67}$$

is called the *tangent* to the curve $c(t)$ at $c(0) = a$. Its defining properties are those of a derivation:

$$1.\ \vec{u}(f+g) = \vec{u}(f) + \vec{u}(g); \quad f, g \in C^\infty \tag{5.68}$$

$$2.\ u(\lambda f) = \lambda \vec{u}(f); \quad \lambda \in R \tag{5.69}$$

$$3.\ \vec{u}(fg) = \vec{u}(f)g + f\vec{u}(g); \quad \text{product (a.k.a. "Leibnitz") rule} \tag{5.70}$$

These derivations form a vector space, the space of tangents at point $x = a$. Indeed, consider the set of curves

$$c_1,\ c_2,\ \cdots$$

through their common point

$$c_1(0),\ c_2(0),\ \cdots,\ a$$

and their tangents at this point

$$\dot{c}_1(0) = \vec{u}_1,\ \dot{c}_2(0) = \vec{u}_2,\ \cdots$$

This context leads to the following
Proposition (*Tangent Space at a Point*)
The set of derivations at a form a vector space, V_a at a.

[13]Steps (ii) and)iii) comprise a mental integration process which blends functions that obey Eq.(5.65) into a constellation of (interchangeable) units.

The validity of this proposition follows from the fact that derivations are closed under addition and scalar multiplication. Indeed, let

$$\vec{u} = u^i \frac{\partial}{\partial x^i} \quad \text{at } a \tag{5.71}$$

$$\vec{v} = v^i \frac{\partial}{\partial x^i} \quad \text{at } a \tag{5.72}$$

be derivations at a. Then

a)

$$\begin{array}{lllll} \vec{u}+\vec{v}: & C^\infty & \to & R & \\ & f & \rightsquigarrow & (\vec{u}+\vec{v})(f)\big|_a & \equiv \vec{u}(f)\big|_a + \vec{v}(f)\big|_a \\ & & & & = u^i \frac{\partial f}{\partial x^i}\Big|_a + v^i \frac{\partial f}{\partial x^i}\Big|_a \\ & & & & = \quad (u^i + v^i)\frac{\partial f}{\partial x^i}\Big|_a \end{array} \tag{5.73}$$

is also a derivation. Thus we have closure under addition. On the other hand,

b)

$$\begin{array}{lllll} \alpha\,\vec{u}: & C^\infty & \to & R & \\ & f & \rightsquigarrow & (\alpha\,\vec{u})(f)\big|_a & \equiv \alpha\ \vec{u}(f)\big|_a \\ & & & & = \alpha\left(u^i \frac{\partial f}{\partial x^i}\Big|_a\right) \\ & & & & = (\alpha u^i)\frac{\partial f}{\partial x^i}\Big|_a \end{array} \tag{5.74}$$

is also a derivation. Thus we have closure under scalar multiplication.

Conclusion: the set of derivations at $x = a$, $u^i \frac{\partial f}{\partial x^i}$, $v^i \frac{\partial f}{\partial x^i}$, $\cdots$ form a vector space, the *tangent space* V_a at $x = a$.

Lecture 29

5.10.2 The Differential of a Function

Every vector space has a dual space.
Question: What is the space V^* dual to V_a? What are its elements? The answers to these questions lie in the PLP,

$$u^i \frac{\partial f}{\partial x^i}\bigg|_{a=c(0)} = \dot{c}^i(0) \frac{\partial f}{\partial x^i}\bigg|_{a=c(0)} \tag{5.75}$$

$$= \frac{df(c(t))}{dt}\bigg|_{t=0} , \tag{5.76}$$

of the Taylor series expansion, Eq.(5.64) on page 235. For the vector

$$\begin{aligned} \vec{u} = u^i \frac{\partial}{\partial x^i} \quad &\in V_a \\ = \dot{c}^i(0) \frac{\partial}{\partial x^i} & \end{aligned}$$

tangent to $c(t)$ at $c(0) = a$, apply that PLP to the following set of n+1 C^∞-functions

$$\begin{aligned} f^1(x) &= x^1 \\ f^2(x) &= x^2 \\ &\vdots \\ f^n(x) &= x^n \\ f(x) &= f\left(x^, \cdots, x^n\right) \quad \text{(an arbitrarily chosen function)} \end{aligned}$$

These functions yield linear functions $\omega^1, \cdots, \omega^n$ and h on V_a with their own distinguished properties:

$$\begin{aligned} f^1: \quad & u^i \frac{\partial x^1}{\partial x^i}\bigg|_a \equiv \omega^1(a;\vec{u}) = u^1 \quad (= D_{\vec{u}} x^1) \\ f^2: \quad & u^i \frac{\partial x^2}{\partial x^i}\bigg|_a \equiv \omega^2(a;\vec{u}) = u^2 \quad (= D_{\vec{u}} x^2) \\ & \vdots \\ f^n: \quad & u^i \frac{\partial x^n}{\partial x^i}\bigg|_a \equiv \omega^n(a;\vec{u}) = u^n \quad (= D_{\vec{u}} x^n) \\ f: \quad & u^i \frac{\partial f}{\partial x^i}\bigg|_a \equiv \quad h(a;\vec{u}) \qquad\quad (= D_{\vec{u}} f\ ) \end{aligned}$$

From these linear functions one infers two new concepts:

1. The set $\{\omega^1, \cdots, \omega^n\}$ forms the coordinate-induced basis for V_a^*, the space dual to V_a.
   This inference one obtains from the distinguishing properties of the ω^js in

two steps: (i) Evaluate each of the linear functions $\omega^1, \cdots, \omega^n$ on each of the elements of the coordinate-induced basis for V_a,

$$B = \left\{ \vec{e}_1 = \frac{\partial}{\partial x^1}, \cdots, \vec{e}_n = \frac{\partial}{\partial x^n} \right\} .$$

The result is

$$\omega^j(a, \vec{e}_i) = \frac{\partial x^j}{\partial x^i} = \delta^j{}_i .$$

(ii) It follows that

$$B^* = \{\omega^1, \cdots, \omega^n\}$$

is the sought-after basis for V_a^* dual to V_a at point $x = a$.

2. The linear function

$$h(a; \vec{u}) = u^i \frac{\partial f}{\partial x^i} \tag{5.77}$$

is the differential of f, an element of in V_a^*. Indeed, insert the distinguishing properties of the ω^js,

$$\omega^j(a, \vec{u}) = u^j \quad j = 1, \cdots, n$$

into Eq.(5.77). The result

$$h(a; \vec{u}) = \frac{\partial f}{\partial x^i} \omega^i(a, \vec{u})$$

holds for all $\vec{u} \in V_a$.

Omitting reference to the value $h(a, \vec{u})$ for any particular value $\vec{u}$ on the principle that $h(a, \vec{u})$ must have been evaluated for some $\vec{u}$, but may have been evaluated for any $\vec{u}$, one obtains the concept of the function

$$h(a, \cdot) = \frac{\partial f}{\partial x^i} \omega^i(a, \cdot) \in V_a^* : \quad \begin{array}{ccc} V_a & \to & R \\ \vec{u} & \rightsquigarrow & \frac{\partial f}{\partial x^i} \omega^i(a, \vec{u}) \end{array} \quad . \tag{5.78}$$

This function on V_a is called the differential of $f(x^1, \cdots, x^n)$ at the point $x = a$. The standard mathematical notation it is

$$\boxed{df = \frac{\partial f}{\partial x^i} dx^i \quad (\in V_a^*)} \tag{5.79}$$

In terms of Dirac bracket notation this is

$$\langle df| = \frac{\partial f}{\partial x^i} \langle dx^i| \ .$$

In this notation its value on the vector $\vec{u} = u^j \frac{\partial}{\partial x^j}$ is

$$\begin{aligned} df(\vec{u}) \equiv \langle df|\vec{u}\rangle &= \langle \frac{\partial f}{\partial x^i}\langle dx^i|u^j\frac{\partial}{\partial x^j}\rangle \\ &= \frac{\partial f}{\partial x^i}\, u^j \underbrace{\frac{\partial x^i}{\partial x^j}}_{\delta^i{}_j} = \frac{\partial f}{\partial x^i}u^i \\ &= D_{\vec{u}}f \end{aligned} \tag{5.80}$$

Summary
The geometrization of calculus is characterized by the following battery of concepts at each point a:

1. $\left\{\vec{e}_1 = \frac{\partial}{\partial x^1}, \cdots, \vec{e}_n = \frac{\partial}{\partial x^n}\right\}$ is a vector basis for V_a.
2. $\{\omega^1 = dx^1, \cdots, \omega^n = dx^n\}$ is a covector basis for V_a^*.
3. These two bases are dual to one another., This is mathematized by the statement

$$\langle dx^i|\frac{\partial}{\partial x^j}\rangle = \frac{\partial x^i}{\partial x^j} = \delta^*{}_j. \tag{5.81}$$

4. Given $f \in C^\infty$, one infers from thr PLP of the Taylor series expansion, Eq.(5.62), the existence of the covector

$$\frac{\partial f}{\partial x^i}dx^i\Big|_a = \begin{pmatrix} \text{“ the differential} \\ \text{of } f \text{ at } x = a \text{ "} \end{pmatrix} \in V_a^* \ .$$

When evaluated on the vector $\vec{u} \in V_a$, this covector yields the directional derivative of f, Eq.(5.80).

5.10.3 Duality's Coordinate Invariance

Any element of a (finite dimensional) vector space can be represented in terms its componets relative to a chosen or given basis. The purpose of these representation components is that they form primitive data from which a reasoning mind forms higher level concepts, the building blocks of scientific knowledge, principles that govern the lawful nature of the universe.

The mathematical principle of duality applied to V_x consists of having the parallel isograms of a linear function on V_x, the space of derivations, become a single point

in V_x^*, the space of differentials. This single point is df, and we have mathematized this assignment in Section 5.10.2 by the statement

$$\langle df|\vec{u}\rangle = \frac{\partial f}{\partial x^i}u^i \tag{5.82}$$

for all $\vec{u} \in V_x$. We have done this concretely[14] relative to some chosen basis representation, namely,

$$\vec{u} = u^j \frac{\partial}{\partial x^j} \tag{5.83}$$

and

$$df = \frac{\partial f}{\partial x^i}dx^i \ . \tag{5.84}$$

This seems to make this line of reasoning vulnerable to the charge of being subjective, i.e. dependent on the chosen basis. To show that this is not the case, we change coordinates. This alters the basis for the vector space V_x as well as that for the covector space V_x^* at each point x. Then we ask and answer the following question:
Q: Will it also alter the dualty relations, Eq.(5.81) and (5.82), between them?
A: No.

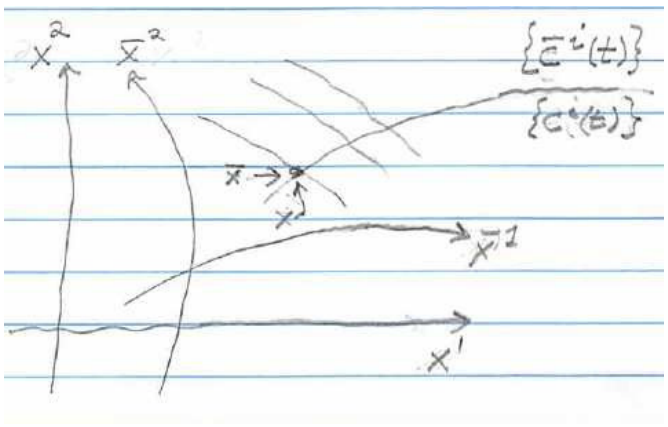


Figure 5.14: Curve piercing the isograms of the function f and depicted relative to the x^1-x^2 and the $\overline{x}^1$-$\overline{x}^2$ coordinate systems.

[14]To prevent "duality" from becoming a floating abstraction, i.e. from not having its basis in reality.

According to the following
Proposition
Duality is invariant under coordinate change.
The validity of this proposition follows from the effect of an explicit coordinate change,

$$x^i = x^i(\overline{x}^k) \tag{5.85}$$
$$\overline{x}^k = \overline{x}^k(x^i) \tag{5.86}$$

Based on the calculus chain rule, the vector $\vec{u} = u^I \frac{\partial}{\partial x^i}$ has the alternate barred coordinate basis expansion

$$\vec{u} = u^i \frac{\partial \overline{x}^k}{\partial x^i} \frac{\partial}{\partial \overline{x}^k}$$

Introduce this alternative into the eveluation of $\langle df|\vec{u}\rangle$ and obtain the following sequence of equalities,

$$\langle df|\vec{u}\rangle| \overset{0}{=} \frac{\partial f(x)}{\partial x^i} \overbrace{\frac{dc^i}{dt}}^{u^i} \tag{5.87}$$
$$\overset{1}{=} \frac{\partial f(x)}{\partial x^i} \underbrace{\frac{\partial x^i}{\partial \overline{x}^k}}_{\delta^i{}_k} \frac{d\overline{c}^k}{dt}$$
$$\overset{2}{=} \frac{\partial f(x(\overline{x}))}{\partial \overline{x}^k} \frac{d\overline{c}^k}{dt}$$
$$\overset{3}{=} \frac{\partial \overline{f}(\overline{x})}{\partial \overline{x}^k} \underbrace{\frac{d\overline{c}^k}{dt}}_{\overline{u}^k} \tag{5.88}$$

Equality 0 evaluates $\langle df|\vec{u}\rangle$ relative to the unbarred coordinate frame.
1 uses the above-mentioned chain rule.
2 uses the chain rule applied to the derivative of the composite function $f(x(\overline{x}))$.
3 introduces the coordinate representative $\overline{f}$ of f relative to the barred cordinate system.
Thus $\langle df|\vec{u}\rangle$ is invariant under a change of coordinate-induced basis relative to which $\langle df|\vec{u}\rangle$ is evaluated[15].

Lecture 30

[15]This evaluation takes place at the *same* point as before, but it now has been coordinatized by $\overline{x}^k$.

5.10.4 The Graph of a Vector Field

Distributed Vector Spaces

By repeating the vector space construction at every point of the domain (a.k.a. the *base space*), one arrives at a *system of vector spaces* $V_{a_1}, V_{a_1}, \cdots$ distributed over the given domain and having their origins located at the points of the domain/base space. Each vector space has its coordinate induced basis

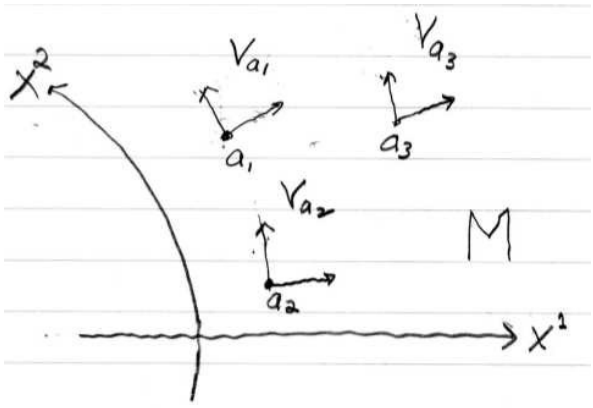


Figure 5.15: Base manifold M populated with vector spaces V_x attached to each point $a_1, a_2, a_3, \cdots$.

$$B_{a_1} = \left\{ \frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n} \right\} \text{ for } V_{a_1} \tag{5.89}$$

$$B_{a_2} = \left\{ \frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n} \right\} \text{ for } V_{a_2} \tag{5.90}$$

$$\vdots \tag{5.91}$$

If

$$M = \{(x^1, \cdots, x^n)\} \equiv \{x\}$$

is the coordinatized base space domain, then the system of distributed tangent vector spaces

$$\bigcup_{x \in M} V_x \equiv TM$$

is called the *tangent bundle* of M and

$$\bigcup_{x \in M} V_x^* \equiv T^*M$$

is called the *cotangent bundle* of M

Vector Field

The tangent bundle TM accommodates the assignment of a vector

$$\vec{u}(x) = u^i(x^1, \cdots, x^n)\frac{\partial}{\partial x^i} \ \in V_x \tag{5.92}$$

to each vector space. We thererfore have the
Definition (*Vector Field*)
The mapping

$$u : \left\{ \begin{array}{lcll} M & \to & TM & \\ x & \rightsquigarrow & u(x) = \{x, \vec{u}(x)\} & = \left\{x^1, \cdots, x^n; u^i, \cdots, u^n)\frac{\partial}{\partial x^i}\right\} \\ & & & \equiv \underbrace{\{(x^1, \cdots, x^n); (u^1. \cdots, u^n)\}}_{\begin{array}{c}\text{coordinate}\\ \text{representative}\\ \text{of } u(x) \in TM\end{array}} \end{array} \right. \tag{5.93}$$

is a vector field on M. It is the *graph* of u in the tangent bundle space TM.

We shall consider only *smooth* vector fields. The vector field $\vec{u}(x)$ is said to be *smooth* whenever

$$u^i(x^1, \cdots, x^n)\frac{\partial f(x^1, \cdots, x^n)}{\partial x^i}$$

is smooth for all $f \in C^\infty$.

Flow Field: Integral Curves of a Vector Field

A smooth vector field $\vec{u}(x)$ determines a set of curves whose tangents equal the values of the vector field along these curves.

I. Integral Curves

GIVEN: A continuous vector field, an assignment of a vector

$$\vec{u} = u^i(x^1, \cdots, x^n)\frac{\partial}{\partial x^i} \quad \in V_x; \quad x \in M$$

to each point $x = (x^1, \cdots, x^n)$ of the base M.
QUESTION: Within the framework of geometry: Do there exist curves $x(t)$ whose vector $\frac{dx^i}{dt}\frac{\partial}{\partial x^i}$ tangents to $x(t)$ at each point equals the value of $\vec{u}$ at that point ?
Within the framework of the mechanics of calculus: Can one find curves $x(t)$ such that they obey the system of o.d.e.'s equations,

$$\frac{dx^i}{dt}\frac{\partial}{\partial x^i} = u^i(x^1, \cdots, x^n)\frac{\partial}{\partial x^i} \tag{5.94}$$

or

$$\frac{dx^i}{dt} = u^i(x^1, \cdots, x^n) \quad i = 1, \cdots, n\,? \tag{5.95}$$

ANSWER: Yes, one can. The theory of o.d.e.'s informs us that one can solve this system of n equations in n unknowns: For every chosen point $x_0 \in M$ there exist n parameters $t_0, \alpha^1, \cdots, \alpha^{n-1}$ and curves $c(t)$

a) which depend smoothly on these parameters,

b) each of which curves satisfies

$$\frac{dc^i}{dt} = u^i(c^1, \cdots, c^n) \text{ and the initial condition } c(t = t_0) = x_0,$$

c) and which together form an n-parameter family of curves which maps each $(t_9, \alpha^1, \cdots, \alpha^{n-1})$ homeomorphically into M

$$c(t): \ (t - t_0, \alpha^1, \cdots, \alpha^{n-1}) \overset{c}{\rightsquigarrow} \ x^i = c^i(t - t_0, \alpha^1, \cdots, \alpha^{n-1})$$

Therefore, in the space R^n spanned by t and $\alpha^1, \cdots, \alpha^{n-1}$ (comoving coordinates) this family straightens out the vector field $\vec{u}(x)$

Flow Field

The vector field $\vec{u}(x) = u^i(x^j)\frac{\partial}{\partial x^i}$, and hence its integral curves $c(t - t_0.\alpha^i)$, induce a *flow* on the lab (Eulerian) coordinatized manifold.

Arrive at this flow by shifting attention from each curve indivdually to a set of neighboring curves. In Figure 5.16 they are depicted as $c(t), c_0(t), c_1(t)$, and $c_2(t)$. Their distinguishing characteristic is that each curve has its own starting point when $t = t_0$. The location of this point is mathematized by the physical attributes of the curve. They are expressed by the constants of integration $t_0, \alpha^1, \cdots, \alpha^{n-1}$ depicted in the left hand panel of Figure 5.16.

Observe that by individuating each curve in terms of its starting point at $t = t_0$,

$$x = c(t_0): \ \{x^i\} = \{c^i(0, \alpha^1, \cdots, \alpha^{n-1})\}_{i=1}^n,$$

one finds that neighboring curves have the same distinguishing characteristic, the mutual proximity of their starting points. This proximity is measured by their $(\alpha^1, \cdots, \alpha^{n\ 1}$w-values which fall into the same $n-1)$-dimensional neighborhood of $(\alpha^1, \cdots, \alpha^{n\ 1})$-values. This neighborhood is depicted in the lefthand panel of Figure 5.16.

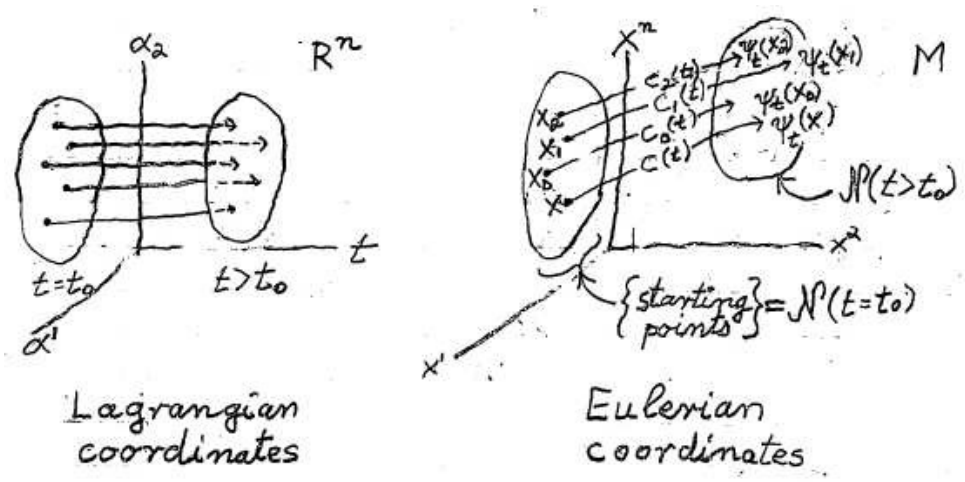


Figure 5.16: Integral curves of vector field $\vec{u}(x)$ relative to comoving (Lagrangian) and lab (Eulerian) coordinates. The observed starting points of curves are mathematized by the measured multi-parameter $(\alpha^1, \cdots, \alpha^{n-1})$ when $t = t_0$ in the left hand panel. The right hand panel depicts the action of the flow field transformation $\psi_t : \mathcal{N}(t = t_0) \to \mathcal{N}(t > t_0)$.

Within the $c(t)$-induced context of the set of physically determined integration constants $(t_0, \alpha^1, \cdots, \alpha^{n\ 1}$ one arrives at the new concept, the *flow-field transformation*[16] ψ_t.

$$\psi_t : \begin{cases} \mathcal{N}(t = t_0) & \to \mathcal{N}(t > t_0) \\ x \equiv \{c^i(t - t_0 = 0; \alpha^1, \cdots, \alpha^{n-1})\} & \rightsquigarrow \psi_t(x) \equiv \{c^i(t - t_0 > 0; \alpha^1, \cdots, \alpha^{n-1})\} \end{cases} \tag{5.96}$$

The properties of this flow transformation are

[16]The mental process for producing it is in essence a two-step process of "measurement omission": (i) Mentally isolate two or more instances of the distinguishing characteristic. (There are four of them in the left hand panel of Figure 5.16.They are instances of measured multi-dimensional parameter $(t_0, \alpha^1, \cdots, \alpha^{n\ 1})$ values that fall within the $t = t_0$ neighborhood as well as within the $t > t_0$ neighborhood.)
(ii) Retain their characteristic while omitting their particular measurements (i.e. the values of thje multi-dimensional parameter $(\alpha^1, \cdots, \alpha^{n\ 1})$ – on the principle that these exist in some (multi-dimensional) quantity, but may in any (such) quantity.
In light of the above-mentioned context, the product of this two-step process is the transformation ψ_t in the text.

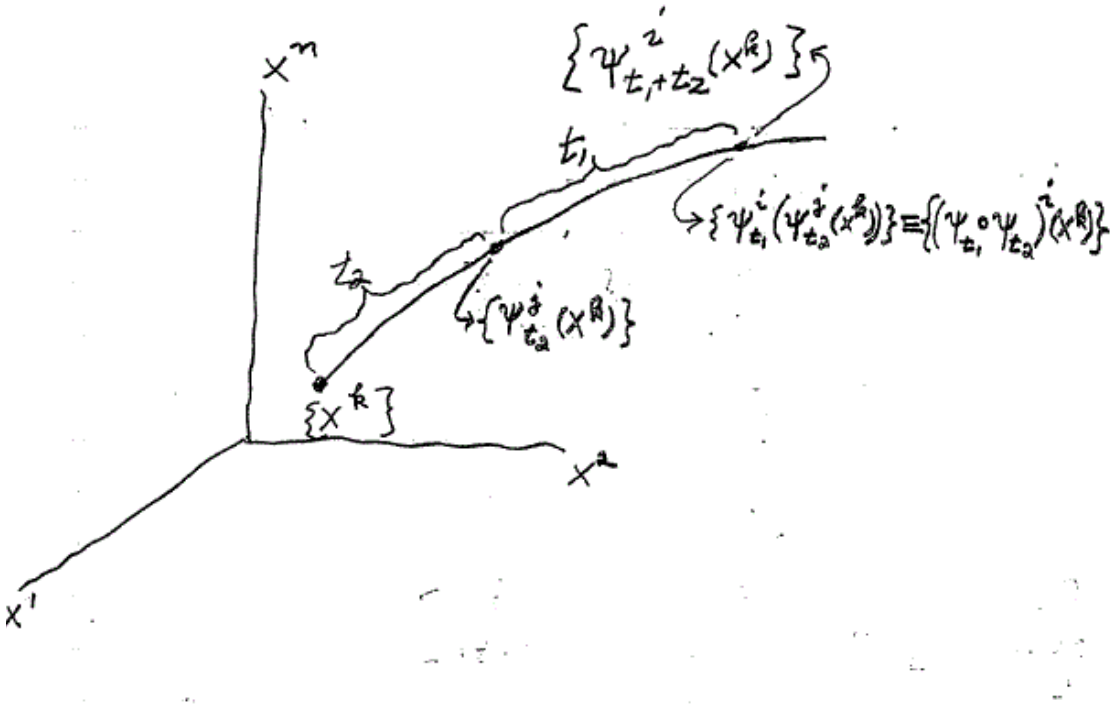


Figure 5.17: Composite of flow transformations ψ_{t_1} with ψ_{t_2}. The result, $\psi_{t_1} \circ \psi_{t_2}$, equals $\psi_{t_1+t_2}$. .

a)

$$\psi_t \equiv \{\psi_t^i\} : \begin{cases} M & \to M \\ x \equiv \{x^j\} & \rightsquigarrow \psi_t(x) \equiv \{\psi_t^i(x^j)\} \end{cases} \tag{5.97}$$

where

$$\psi_0^i(x^j) = x^i \quad i = 1, \cdots, n$$

is the *identity* traransformation,

$$\boxed{\psi_0 = I}$$

b) The *composite* of two transformations is

$$\boxed{\psi_{t_1} \circ \psi_{t_2} = \psi_{t_1+t_2}}$$

This is the mathematization of the fact that the solution to the system of o.d.e.'s is *unique* This uniqueness is depicted in Figure 5.17.
b) The inverse of the transformation ψ_t is

$$\boxed{(\psi_t)^{-1} = \psi_{-t}}$$

This simplicity follows from

$$(i)\ \psi_t \circ \psi_{-t} = \psi_{t-t} = \psi_0 = I = \begin{pmatrix} \text{the identity} \\ \text{transformation} \end{pmatrix}$$
$$(ii)\ \psi_t \circ (\psi_t)^{-1}$$
$$\therefore\ \ (\psi_t)^{-1} = \psi_{-t}\ .$$

Conclusion The above three boxed equations are summarized by the statement that the solutions to a system of autonomous[17] o.d.e.'s induce flow field transformations which form an "commutative group" under functional composition.

Lecture 31

5.10.5 The Commutator of Two Vector Fields

Geometrized calculus is bare bones differential geometry. Its domain consists of the vector space of tangents and its covector space of differentials attached smoothly to each point of the base manifold. The resulting domains are the tangent bundle TM and the cotangent bundle T^*M of the base space manifold. The bundle perspective, introduced by Einstein on physical grounds in 1907, and mathematized subsequently by E. Cartan, revolutionized 20^{th} mathematics.

Geometrized calculus is multivariable calculus on steroids. This is because its initial formulation starts relative to a chosen set of coordinates, but once finished, the resulting propositions are independent of one's chosen coordinates. The means by which this is achieved is that geometric calculus is an integration of vector space theory with calculus. This integration glues together distinct vector spaces by the mere requirement of continuity and differentiability. But it does so without assuming that these vector spaces have any particular metric structure (length, inner product, etc.), any orientation structure or even any structure of parallelism between vector spaces adjacent to one another. There is no assumed geometrical relation between them except that of continuity and differentiability.

The integral curves of a given vector field, Eq.(5.92), are 1-dimensional continua parametrized by their curve parameters. For two or more vector fields this result generalizes to continua of two or more dimensions. Under this and other circumstances[18] the *commutator* of two vector fields plays a fundamental role.

[17] i.e. the coefficient in the system are independent of the independendent variable.

[18] In differential geometry and differential topology via Frobenius's theorem; in mathematical engineering via non-linear control theory of dynamical systems.

Consider two vector fields

$$\vec{u}(x) = u^i(x)\frac{\partial}{\partial x^i}$$
$$\vec{v}(x) = v^i(x)\frac{\partial}{\partial x^i}$$

and their integral curves $c_{\vec{u}}(t)$ and $c_{\vec{v}}(s)$
Q: Can these curves serve as a coordinate mesh with coordinate lines t and s ?
A: The existence of such a coordinate mesh depends on whether the quadrilateral formed by pairs of equal curve segments closes or not. If the points $\mathcal{P}_1$ and $\mathcal{P}_2$

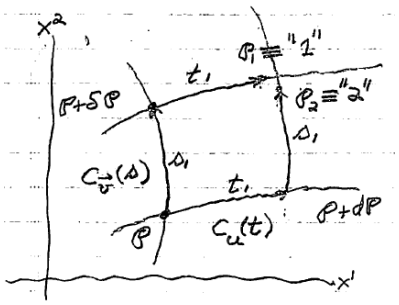


Figure 5.18: Open quadrilateral formed from two pairs of equal curve segments. The above quadrilateral is not closed because the broken paths $\mathcal{P} \to \mathcal{P} + \delta\mathcal{P} \to \mathcal{P}_1(\equiv$ point "1") and $\mathcal{P} \to \mathcal{P} + \mathcal{P} \to \mathcal{P}_2(\equiv$ point "2") do not have the same termination pont.

coincide,

$$\mathcal{P}_1 = \mathcal{P}_2 \ ,$$

then their common coordinate label is

$$(t_1, s_1) \ .$$

The criterion as to whether these two points are the same is mathematized by the
Definition

$$\mathcal{P}_1 = \mathcal{P}_2 \ \Leftrightarrow \ f(1) = f(2) \quad \forall f \in C^\infty \ ,$$

i.e. for all physical properties that identify points "1" and "2".

Use Taylor series to calculate $f(1)$ and $f(2)$ from f's values and its derivatives at $\mathcal{P}$.

a) Computing $f(1)$ as a Taylor series along the upper broken path $1 \to \mathcal{P}+\delta\mathcal{P} \to \mathcal{P}$, one finds to second order accuracy

$$\begin{aligned} f(1) &= f(\mathcal{P}+\delta\mathcal{P}) + t_1\vec{u}(f)|_{\mathcal{P}+\delta\mathcal{P}} + \frac{t_1^2}{2!}\vec{u}\,(\vec{u}(f))\Big|_{\mathcal{P}+\delta\mathcal{P}} + \cdots \\ &= f(\mathcal{P}) + s_1\vec{v}(f)|_{\mathcal{P}} + \frac{s_1^2}{2!}\vec{v}\,(\vec{v}(f))\Big|_{\mathcal{P}} + \cdots \\ &\qquad + t_1\vec{u}(f)|_{\mathcal{P}} + t_1 s_1\ \vec{v}(\vec{u}(f))|_{\mathcal{P}} + \cdots + \frac{t_1^2}{2!}\vec{u}\,(\vec{u}(f))\Big|_{\mathcal{P}} + \cdots \end{aligned}$$

b) Similarly along the lower broken path $2 \to \mathcal{P}+d\mathcal{P} \to \mathcal{P}$ one finds

$$\begin{aligned} f(1) &= f(\mathcal{P}) + t_1\vec{u}(f)|_{\mathcal{P}} + \frac{t_1^2}{2!}\vec{u}\,(\vec{u}(f))\Big|_{\mathcal{P}} + \cdots \\ &\qquad + s_1\vec{v}(f)|_{\mathcal{P}} + s_1 t_1\ \vec{u}(\vec{v}(f))|_{\mathcal{P}} + \cdots + \frac{s_1^2}{2!}\vec{v}\,(\vec{v}(f))\Big|_{\mathcal{P}} + \cdots \end{aligned}$$

c) Note that $f(2)$ and $f(1)$ are equal except for the order of the mixed directional derivatives $\vec{u}(\vec{v}(f)) \equiv D_{\vec{u}}D_{\vec{v}}f$ and $\vec{v}(\vec{u}(f)) \equiv D_{\vec{v}}D_{\vec{u}}f$. The difference between $f(2)$ and $f(1)$ reduces therefore to

$$\begin{aligned} \frac{f(2)-f(1)}{s_1 t_1} &= \Big(\vec{u}(\vec{v}(f)) - \vec{v}(\vec{u}(f))\Big)\Big|_{\mathcal{P}} + \mathcal{O}(s_1 t_1) \\ &= [\vec{u}\vec{v} - \vec{v}\vec{u}](f)\Big|_{\mathcal{P}} + \mathcal{O}(s_1 t_1) \\ &\equiv [\vec{u},\vec{v}]\Big|_{\mathcal{P}} + \mathcal{O}(s_1 t_1) \end{aligned} \tag{5.98}$$

Definition The difference quotient Eq.(5.98),

$$\vec{u}\vec{v} - \vec{v}\vec{u} \equiv [\vec{u},\vec{v}],$$

is the *commutator* of the vector fields $\vec{v}$ and $\vec{u}$.

Proposition

The commutator $[\vec{u},\vec{v}]\Big|_{\mathcal{P}}$ is a derivation at point $\mathcal{P}$ and hence an element of $V_{\mathcal{P}}$.

Exercise 1: Infer the validity of this proposition by verifying that the commutator satisfies the three properties, Eqs.(5.68)-(5.70), on page 236. **Exercise 2:** GIVEN: Two vector fields

$$\vec{u} = u^i\frac{\partial}{\partial x^i}$$
$$\vec{v} = v^j\frac{\partial}{\partial x^j}$$

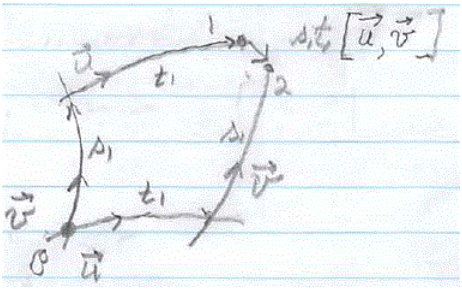


Figure 5.19: The commutator of two vector fields.

FIND: The components of their commutator

$$[\vec{u},\vec{v}] = [\vec{u},\vec{v}]^k \frac{\partial}{\partial x^k} = [?]^k \frac{\partial}{\partial x^k}$$

relative to the coordinate basis $\frac{\partial}{\partial x^k}$

Problem

What is the commutator of

$$(i)\ \vec{e}_\theta = \frac{\partial}{\partial\theta} \text{ and } \vec{e}_r = \frac{\partial}{\partial r}$$
$$(ii)\ \vec{e}_{\hat{\theta}} = \frac{1}{r}\frac{\partial}{\partial\theta} \text{ and } \vec{e}_{\hat{r}} = \frac{\partial}{\partial r}$$
$$(iii)\ \text{Draw a picture of each.}$$

Solution:

$$[\frac{\partial}{\partial r},\frac{\partial}{\partial\theta}] = \frac{\partial}{\partial r}\frac{\partial}{\partial\theta} - \frac{\partial}{\partial\theta}\frac{\partial}{\partial r} = 0$$
$$[\frac{\partial}{\partial\hat{r}},\frac{\partial}{\partial\hat{\theta}}] = \frac{\partial}{\partial r}\frac{1}{r}\frac{\partial}{\partial\theta} - \frac{1}{r}\frac{\partial}{\partial\theta}\frac{\partial}{\partial r}$$
$$= -\frac{1}{r^2}\frac{\partial}{\partial\theta} + \frac{1}{r}\frac{\partial^2}{\partial r\partial\theta} - \frac{1}{r}\frac{\partial^2}{\partial\theta\partial r}$$
$$= -\frac{1}{r^2}\frac{\partial}{\partial\theta} = -\frac{1}{r} = \vec{e}_{hat\theta}$$

These two commutators are depicted in the figure below.

Lecture 32

5.11 Geometrical Structures

The distinguishing feature of Geometrical Calculus is its bare bones structure. Recall that at each point $\mathcal{P}$ of a manifold M one observes the existence of $V_\mathcal{P}$, the vector

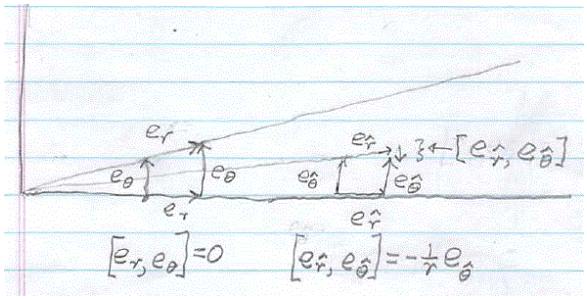


Figure 5.20: The value of the commutator of two different vector fields (unhatted and hatted) compared and contrasted.

space of tangents to the curves $c(t)$ running through $\mathcal{P}$: $c(t=0) = \mathcal{P}$. Each point $\mathcal{P}$ has its own tangent space $V_\mathcal{P}$ In this context one's curiosity leads one to ask wether

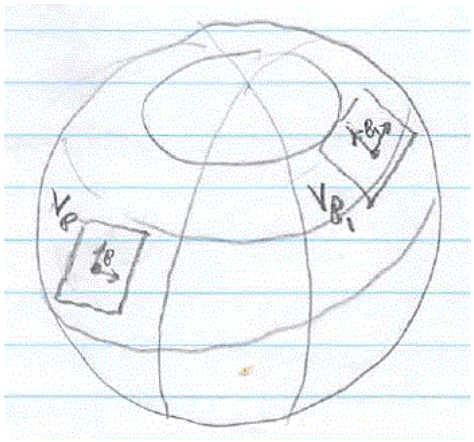


Figure 5.21: Attached to each point $\mathcal{P}$ of a manifold is its tangent space $V_\mathcal{P}$.

there exists a natural (i.e. a basis-independent) relation between them. Without further observational input the answer is "NO". However, suppose we are confronted with a manifold in which one observes curves which are straight, meaning that betwee a give pair of points there exists an extremal, say a shortest path.

5.11.1 Parallelism Via Schild's Ladder

Consider a generic curve $c(t)$ on the manifold, say the sphere S^2, and a displacement vector $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$ at the the initial point $\mathcal{P}_0$ of the curve. This vector is an infinitesimal straight line segment which connects the two points $\mathcal{P}_0$ and $\mathcal{A}_0$. From some other

nearby point, say $\mathcal{P}_1$ also on $c(t)$, draw the straight line segment $\overrightarrow{\mathcal{P}_1\mathcal{A}_0}$ and mark off its center $\mathcal{C}$ as in Figure 5.22. Next draw the line sgment $\overrightarrow{\mathcal{P}_0\mathcal{C}}$ and extend it by bthe same amount beyond $\mathcal{C}$ with its termination point $\mathcal{A}_1$ twice as far from $\mathcal{P}_0$. The point $\mathcal{A}_1$ together with $\mathcal{P}_1$ form the straight line segment $\overrightarrow{\mathcal{P}_1\mathcal{A}_1}$ which is parallell to $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$.

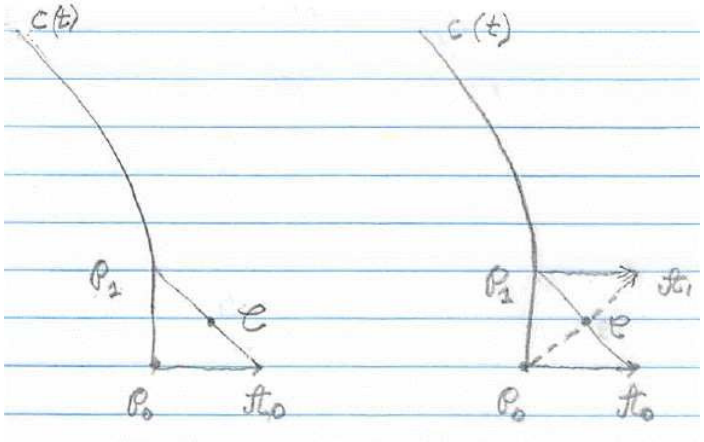


Figure 5.22: Schild ladder construction of vector $\overrightarrow{\mathcal{P}_1\mathcal{A}_1}$ at $\mathcal{P}_1$ parallel to the initial vector $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$ at $\mathcal{P}_0$.

As shown in Figure 5.23 below, this "Schild Ladder" construction can be repeated with the result that one obtains a sequence of parallelograms with a field of parallel vectors $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$, $\overrightarrow{\mathcal{P}_1\mathcal{A}_1}$, $\overrightarrow{\mathcal{P}_2\mathcal{A}_2}, \cdots$ along the given curve $c(t)$.
Remark 1:
The fact that the diagonals of a parallelogram besect each other, if not already known to the Greeks, was undoubtedly known to Newton. Here the Schild ladder construction is a converse of this fact: if two line sgments bisect each other, then its vertices form a parallelogram.
Remark 2:
a) The initial vector $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$ can be an arbitrary element of $V_{\mathcal{P}_0}$, and the Schild ladder construction yields a unique element $\overrightarrow{\mathcal{P}_1\mathcal{A}_1} \in V_{\mathcal{P}_1}$ which is said to be parallel to $\overrightarrow{\mathcal{P}_0\mathcal{A}_0} \in V_{\mathcal{P}_0}$:

$$\boxed{\forall\, \overrightarrow{\mathcal{P}_0\mathcal{A}_0} \in V_{\mathcal{P}_0}\ \exists \text{ a unique } \overrightarrow{\mathcal{P}_1\mathcal{A}_1} \in V_{\mathcal{P}_1}}$$

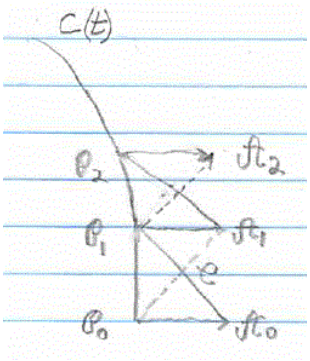


Figure 5.23: Sequential construction of $\overrightarrow{\mathcal{P}_1\mathcal{A}_1}$ at $\mathcal{P}_1$ parallel to $\overrightarrow{\mathcal{P}_0\mathcal{A}_0}$, $\overrightarrow{\mathcal{P}_2\mathcal{A}_2}$ at $\mathcal{P}_2$ parallel to $\overrightarrow{\mathcal{P}_1\mathcal{A}_1}$, etc.

b) Thus the Schild ladder construction establishes an isomorphism[19]

$$V_{\mathcal{P}_0} \leftrightarrow V_{\mathcal{P}_1}$$

between adjacent vector spaces separated by the curve segment

$$\overrightarrow{\mathcal{P}_0\mathcal{P}_1} = \Delta t \left.\frac{dc^i}{dt}\right|_{t=0} \frac{\partial}{\partial x^i} = \Delta\ \vec{u}|_{\mathcal{P}_0}$$

of curve $c(t)$ at $t = 0$.
Remark 3:
The construction is natural in that it is independent of any preexisting bases for vector spaces $V_{\mathcal{P}_0}$ and $V_{\mathcal{P}_1}$.
Remark 4:
a) This isomorphism between adjacent vector spaces is the mathematical formulation of parallel transport. Relative to a basis chosen (or given) for each vector space $V_{\mathcal{P}}$ this transformation is expressed by means of a matrix. Thus, given such a matrix, one has a law of parallel treansport, and given a law of parallel transport (and a coice of basis) one has a matrix.
b) The next task is to exhibit this matrix representation explicitly.

5.11.2 Mathematization of Schild's Ladder

Elie Cartan introduced a new way of conceptualizing a vector field. He viewed it as a graph. The set of vector fields is multifarious. But there is one type which stands out from all the others, namely, vector fields which are *parallel*. Faced with the bewildering diversity of vector fields follow Thales of Miletus by asking,

[19] A one-to-one linear transformation which is onto.

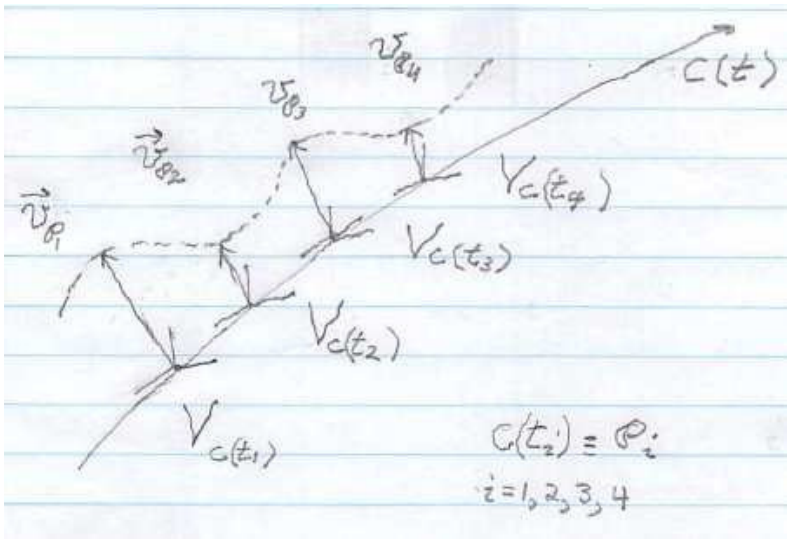


Figure 5.24: Graph of vector field $\vec{v} = v^k(\mathcal{P})\frac{\partial}{\partial x^k}$. The coordinate representation of this graph is in terms of the 2n-tuples $(\mathcal{P}, \vec{v}(\mathcal{P})) \equiv (x^1, \cdots, x^n; v^1(x^1 \cdots, x^n), \cdots, v^n(x^1 \cdots, x^n)$. They refer to points in the tangent bundle $TM = \bigcup_{\mathcal{P}\in M} V_{\mathcal{P}}$. Here they are depicted along the curve $\mathcal{P}(t) = c(t)$ in the base manifold M.

1. What is the one in the many among such vector fields?

2. Where does their commonality come from? And following Euclid and E. Cartan ask:

3. How does one mathematize it?

Answer: Apply Schild's ladders to Cartan's moving frames. From the application of this ladder construction one infers the non-parallelness of the basis elements as in Figure 5.25 or as in Figure 5.26 of adjacent generic vectors. Parallelism via Schild ladder is unique in that it yields adjacent parallelograms whose edges form closed figures. However, in the physics of continuaous media there exists crystaline matter with an imbedded distribution of dislocations or of intrinsic spins. Such media give rise to non-closed parallelograms. Such a parallelism is characterized by Cartan's tortion tensor and its associated Burgers vector. They will be defined in the next subsection. Thus, in reality there are many different types of parallel transport, but they all lend themselves to be mathematized of appropriate isomorphisms between adjacent vector spaces such as $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$. Figure 5.27 depicts such an isomorphism

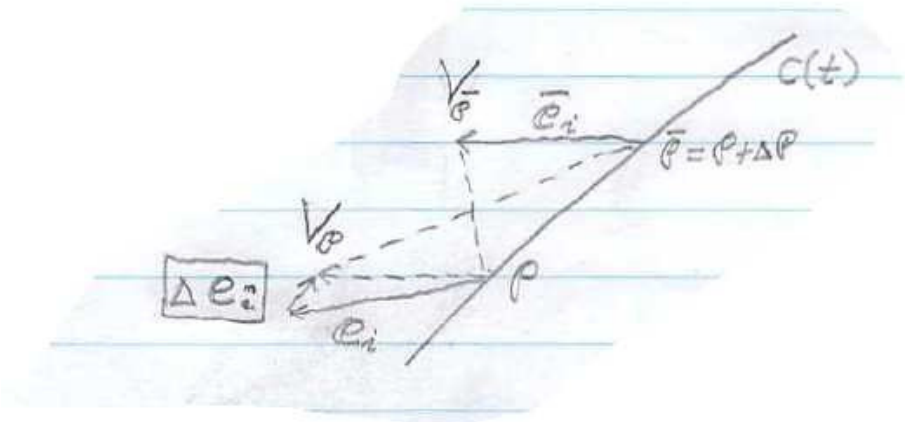


Figure 5.25: Schild ladder construction reveals the non-parallelness $\boxed{\Delta\overline{\mathbf{e}}_i}$ between the bases $\{\overline{\mathbf{e}}_i\}$ and $\{\mathbf{e}_i\}$ for the vector spaces $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$.

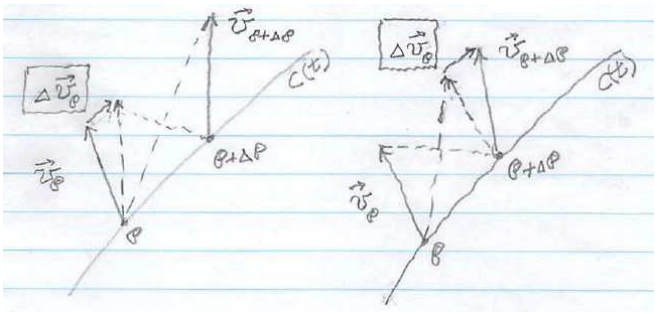


Figure 5.26: Schild ladder construction reveals the non-parallelness $\boxed{\Delta\vec{v}_{\mathcal{P}}}$ between the bases $\{\overline{\mathbf{e}}_i\}$ and $\{\mathbf{e}_i\}$ for the vector spaces $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$.

relative to a particular given/chosen basis. Its matrix representation is

$$
\begin{aligned}
& V_{\overline{\mathcal{P}}} \to V_{\mathcal{P}} \\
& \overline{\mathbf{e}}_i \rightsquigarrow \mathbf{e}_i + \Delta\mathbf{e}_i \equiv \mathbf{e}_j \left[\delta^j{}_i + \omega^j{}_i(\Delta)\right] \\
(0,\cdots,1,\cdots,0) \rightsquigarrow & \Big(\omega^1{}_i(\Delta),\cdots,\omega^{j-1}{}_i(\Delta), 1 + \underbrace{\omega^i{}_i(\Delta)}_{\substack{\uparrow \\ \text{no sum!}}},\omega^{j+1}{}_i(\Delta),\cdots,\omega^n{}_i(\Delta)\Big)
\end{aligned}
\tag{5.99}
$$

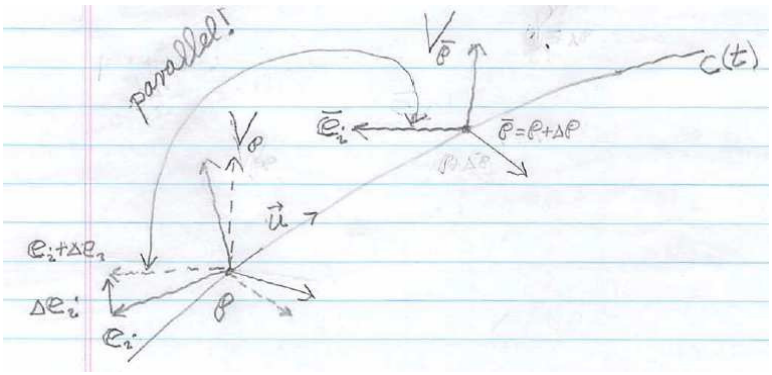


Figure 5.27: Parallelism-induced isomorphism between adjacent vector spacesc $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$.

The matrix for this linear transfortmation

$$[\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n] \begin{bmatrix} 1+\omega^1{}_1(\Delta) & \omega^1{}_2(\Delta) & \cdots & \omega^1{}_n(\Delta) \\ \omega^2{}_1(\Delta) & 1+\omega^2{}_2(\Delta) & \cdots & \omega^2{}_n(\Delta) \\ \vdots & \vdots & \vdots & \vdots \\ \omega^n{}_1(\Delta) & 1+\omega^n{}_2(\Delta) & \cdots & 1+\omega^n{}_n(\Delta) \end{bmatrix} = [\overline{\mathbf{e}}_1, \overline{\mathbf{e}}_2, \cdots, \overline{\mathbf{e}}_n]$$

is determined entirely by the coefficients

$$\omega^j{}_i(\Delta)$$

These numerical coefficients mathematize the parallelism between the vectors of $V_{\overline{\mathcal{P}}}$ and those of $V_{\mathcal{P}}$.
Remark 1:
These coefficients depend on the basis chosen in $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$. Altering the bases for these adjacent vector spaces changes the matrix representation of the isomorphism. This basis dependency is a seemingly fatal flaw in mathematizing parallelism in a basis-independent way. On the contrary! As we shall see, this dependency serves as a portal for an inductive line of reasoning that leads to a basis independent mathematization.
Remark 2:
The matrix representation depends continuously on the separation $\Delta\mathcal{P} = \Delta x^i \frac{\partial}{\partial x^i}$ between the two vector spaces. In moving from $\mathcal{P}$ to $\overline{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$ dominant changes are mathematized by their Principal Linear Parts (P.L.P.). They are the changes along any curve, say $c(t)$, passing smoothly through these points.

In fact, in compliance with the observed nature of things, assume that for small separations the matrix elements depend linearly on the separation between the points $\mathcal{P}$ and $\overline{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$ on a curve $c(t)$ passing through them.

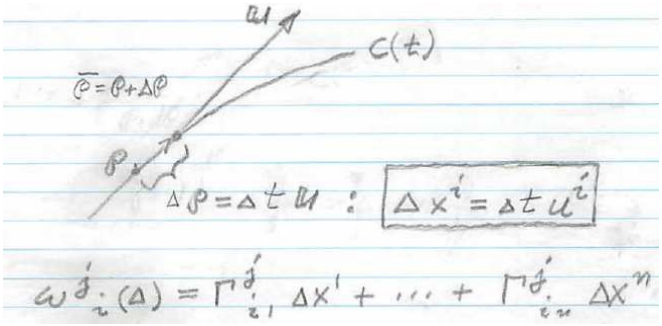


Figure 5.28: Principal Linear Part $[\omega^j{}_i(\Delta)] = [\Gamma^j{}_{ik}\Delta x^k]$, the matrix which generates the parallelism transformation between adjacent vector spaces $V_{\overline{\mathcal{P}}}$ and $V_{\mathcal{P}}$.

Remark 3:
The P.L.P. of the vectorial change, designated by

$$\Delta \mathbf{e}_i = \mathbf{e}_j \omega^j{}_i(\Delta), \text{ with } \omega^j{}_i(\Delta) = \Gamma^j{}_{ik}\Delta x^k$$

is the linear approximation of the difference between the preexisting vector and the one parallel transported by the above-discussed isomorphism. The dominant P.L.P. ignores the non-linear contributions due to the motion between $\mathcal{P}$ and $\overline{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$. Consequently, one has the following equivalent depictions of $\Delta\mathbf{e}_i$:

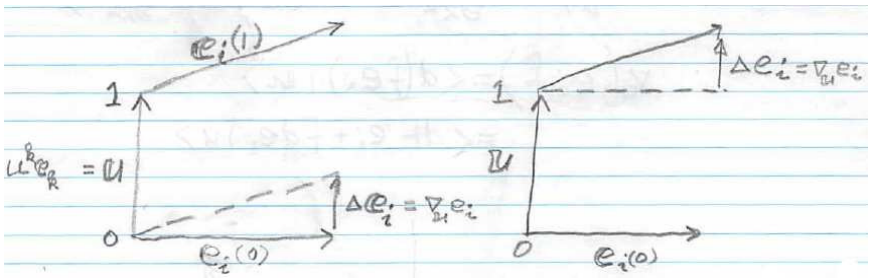


Figure 5.29: Principal Linear Part of vectorial changes

Lecture 33

5.11.3 Vectorial Change: Differential vs. Directional Derivative

Parallel transport is mathematized in two different but related ways:

I. First way: Via the differential of a vector:
Every point of a manifold has a tangent space. Compare the elements in one space, say $V_{\mathcal{P}}$, with those in its neighbor $V_{\overline{\mathcal{P}}}$ at $\mathcal{P}$ and at $\overline{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$

These two vector spaces are separated by the displacement vector

$$\Delta\mathcal{P} = \Delta x^i \frac{\partial}{\partial x^i} = \Delta u^i \frac{\partial}{\partial x^i} \equiv \Delta t\, \mathbf{u}$$

along the curve whose tangent is $\mathbf{u}$.

There are as many forms of parallelism between them as there are isomorphism between

$$V_{\mathcal{P}} = \text{span}\left(\{\mathbf{e}_i\}\right) \text{ and } V_{\overline{\mathcal{P}}} = \text{span}\left(\{\overline{\mathbf{e}}_i\}\right) \ .$$

Recall that parallel transport is a relation which carries a vector $\overline{\mathbf{e}}_i|_{\overline{\mathcal{P}}}$ at $\overline{\mathcal{P}}$ to its parallel image $(\mathbf{e}_i + \Delta\mathbf{e}_i)|_{\mathcal{P}}$ at $\mathcal{P}$:

$$\{\overline{\mathbf{e}}_i\}_{\overline{\mathcal{P}}} \rightsquigarrow \{\mathbf{e}_i + \Delta\mathbf{e}_i\}_{\mathcal{P}},$$

Given the preexisting bases at $\overline{\mathcal{P}} = \mathcal{P} + \Delta\mathcal{P}$ and at $\mathcal{P}$, this isomorphism is mathematized by the Principal Linear Part (P.L.P.) of the change

$$\{\Delta\mathbf{e}_i = \mathbf{e}_j \omega^j{}_i(\Delta t\, \mathbf{u}) : \ i = 1, \cdots, n\}$$

away from the preexisting basis $\{\mathbf{e}_i\}$. This P.L.P. is linear in the separation vector $\Delta t\, \mathbf{u}$.

By leaving it unspecified one has formed a new concept, the *vectorial differential*

$$\boxed{d\mathbf{e}_i \equiv \mathbf{e}_j \otimes \omega^j{}_i \,, \ i = 1, \cdots, n \ .} \tag{5.100}$$

More explicitly, expand $\omega^j{}_i$ in terms of the coordinate basis covectors dx^k with the $\Gamma^i{}_{ik}$'s, the "Christoffel symbols of the second kind", as the expansion coefficients. In these terms the vectorial differential is

$$\boxed{d\mathbf{e}_i = \mathbf{e}_j \otimes \Gamma^i{}_{ik} dx^k \,, \ i = 1, \cdots, n \ .}$$

This is *the vectorial rate of change of* $\mathbf{e}_i$ *due to motion into an as-yet-unspecified direction.* As depicted in Figure 5.29, this change is vector-valued, and it is relative to a physically defined standard of parallelism between adjacent vector spaces. The Schild ladder construction, Figure 5.26, is a particular example of this parallelism.

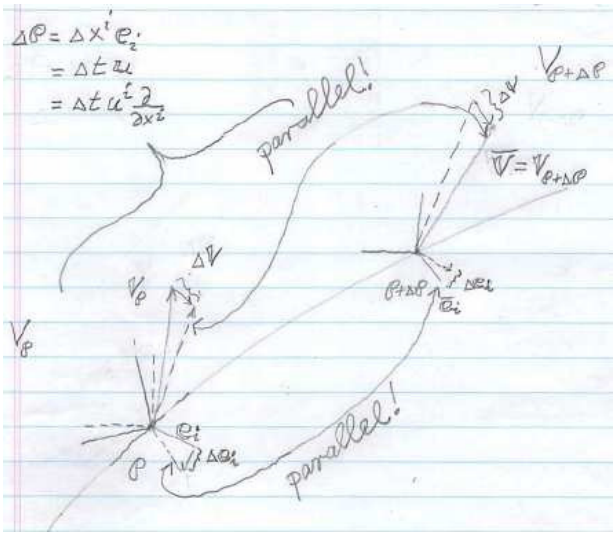


Figure 5.30: Principal Linear Part of vectorial changes $\Delta\mathbf{e}_i$ and $\Delta\mathbf{v}$ relative to the observed, isomorphically defined standard of being parallel.

The vector field $\mathbf{v} = v^i\mathbf{e}_i$ is a linear combination of basis vectors with scalar-function coefficients. The differential of this vector field is a matter of identifying the P.L.P. of the change in each of the products comprising that sum. The P.L.P. follows the product rule. Consequently,

$$\begin{aligned} d\mathbf{v} &= d(\mathbf{e}_i v^i) \\ &= \mathbf{e}_i \otimes dv^i + v^i d\mathbf{e}_i \\ &= \mathbf{e}_j \otimes \frac{\partial v^i}{\partial x^k} dx^k + v^i \mathbf{e}_j \otimes \omega^j{}_i \\ &= \mathbf{e}_j \otimes \left[\frac{\partial v^i}{\partial x^k} + v^i \Gamma^i{}_{ik}\right] dx^k \qquad (5.101) \\ &\equiv \mathbf{e}_j \otimes v^j{}_{;k} dx^k \qquad (5.101') \end{aligned}$$

II. Second way: Via the directional derivative of a vector:
Mathematize the concept of parallelism by introducing the directional derivative ∇. Do this by actually evaluating $d\mathbf{v}$, Eq.(5.101), the "rate of change – away from being parallel – of $\mathbf{v}$ due to motion into an as-yet-unspecified direction", on the the direction $\mathbf{u}$. This evaluation process is governed by $d\mathbf{v}$, the vector-valued linear map

$$d\mathbf{v} : \; V_{\mathcal{P}} \to V_{\mathcal{P}} \qquad (5.102)$$

$$\mathbf{u} \rightsquigarrow \langle d\mathbf{v}|\mathbf{u}\rangle \equiv \nabla_{\mathbf{u}}\mathbf{v} \; , \qquad (5.103)$$

It has the following four fundamental properties

$$\boxed{\begin{array}{lll} (1) & \nabla_{\mathbf{u}}(\mathbf{v}_1+\mathbf{v}_2) & = \nabla_{\mathbf{u}}\mathbf{v}_1+\nabla_{\mathbf{u}}\mathbf{v}_2 \\ (2) & \nabla_{\mathbf{u}_1+\mathbf{u}_2}(\mathbf{v}) & = \nabla_{\mathbf{u}_1}\mathbf{v}+\nabla_{\mathbf{u}_2}\mathbf{v} \\ (3) & \nabla_{f\mathbf{u}}\mathbf{v} & = f\nabla_{\mathbf{u}}\mathbf{v} \\ (4) & \nabla_{\mathbf{u}}(f\mathbf{v}) & = f\nabla_{\mathbf{u}}\mathbf{v}+\mathbf{v}\nabla_{\mathbf{u}}f \end{array}} \tag{5.104}$$

Remark 1:
(1) and (2) say that $\nabla_{\mathbf{u}}\mathbf{v}$ say that is pont-wise linear in $\mathbf{u}$.
(4) is the "Leibnitz product rule" for differentiation.
Remark 2:
The directional derivative $\nabla_{\mathbf{u}}f$ of a scalar function f satisfies the same four properties.
Definition *Covariant Derivative*
The mapping

$$\nabla\mathbf{v}\,(=d\mathbf{v}):\ V_{\mathcal{P}}\to V_{\mathcal{P}} \tag{5.105}$$

$$\mathbf{u}\rightsquigarrow\nabla_{\mathbf{u}}\mathbf{v} \tag{5.106}$$

which satisfies the above boxed four properties is called the *covariant derivative* of the vector field $\mathbf{v}$.
Remark:
The covariant derivative is related to the covariant differential by the fact that the two mappings are one and the same

$$\nabla_{\mathbf{u}}\mathbf{v}=\langle d\mathbf{v}|\mathbf{u}\rangle\quad\forall\mathbf{u}\in V_{\mathcal{P}}$$

so that

$$\boxed{\nabla\mathbf{v}=d\mathbf{v}}$$

5.11.4 Covariant Differentiation

Vector

Covector

Tensor

5.11.5 Tensor Map as Pointwise Linear: Covariant Derivative vs. Commutator

5.12 Geometrical Structures

A geometrical structure is the assignment of a geometrical object to each point of the manifold.

5.12.1 Torsion as a Point-wise Displacement Trait to Parallel Transport

5.12.2 Curvature as a Point-wise Rotation Trait to Parallel Transport

5.12.3 Torsion and Curvature as Point-wise Linear Maps

5.12.4 Equation of Geodesic Deviation

5.12.5 Metric Tensor Field

5.12.6 Elastic Media

Strain Vector

Strain Tensor

5.12.7 Metric Induced Parallel Transport

Metric Compatibility

Compatibility Via Extremal Paths

LECTURE 34

The Covariant Differential

The Covariant Derivative

Reminder

Parallelism, which is illustrated by the Schild ladder construction, is mathematized by the statement that between any pair of curve-connected vector spaces V_ρ and $V_{\rho+\Delta\rho}$ there is an isomorphism which informs as to which vector in one space is parallel the which vector in the other space:

$$V_\rho \longleftrightarrow V_{\rho+\Delta\rho}$$

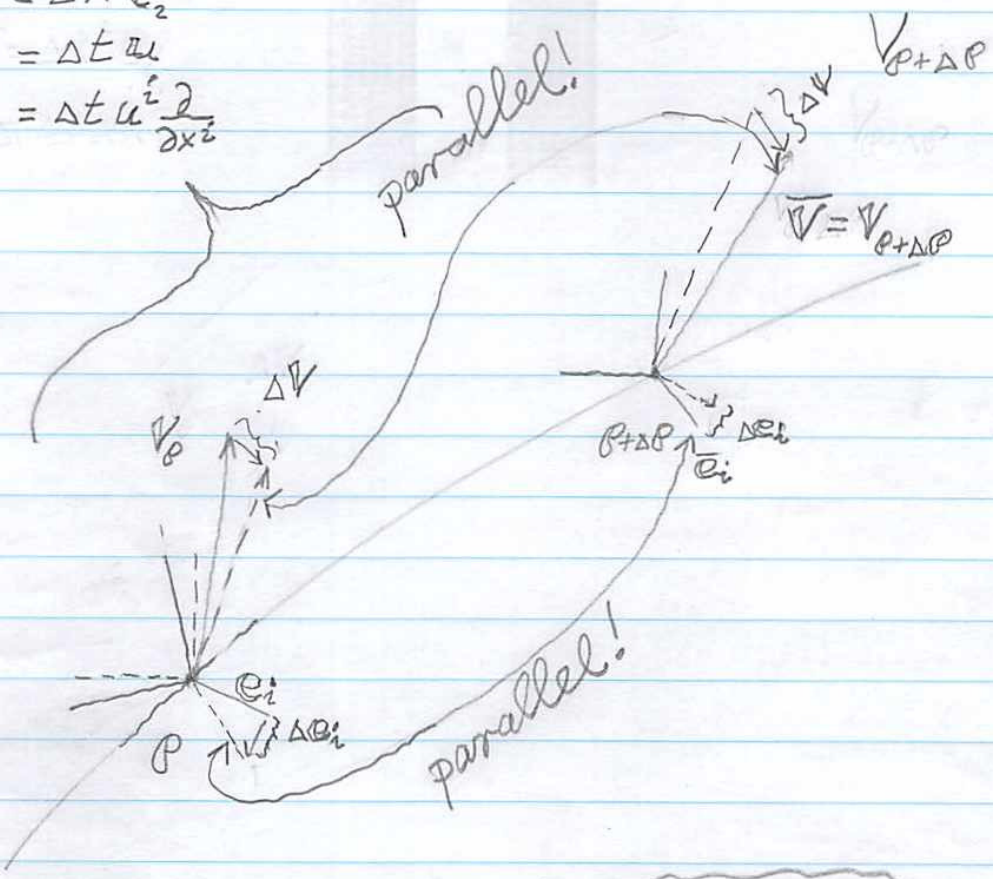
The existence of this parallelism, when applied to a preexisting vector field v, allows us to mathematize the concept of vectorial change relative to this observed and isomorphically defined parallelism

| Parallelism is the condensation of a vast body of observations

$\Delta P = \Delta x^i e_i$

$= \Delta t\, u$

$= \Delta t\, u^i \frac{\partial}{\partial x^i}$



of Δe_i and ΔV

$\Delta e_i, \Delta V$ = P, L, P, of ⇒ vectorial changes relative to the observed, and isomorphically defined standard of being parallel.

A,) The Covariant Differential

The isomorphism which mathematizes the concept of parallelism (or parallelness) is

$$\bar{e}_i \rightsquigarrow e_i + \Delta e_i = e_j \left[\delta^j_i + \omega^j_i(\Delta)\right]$$

The basis vectors in V_P are not parallel to those in $V_{P+\Delta P}$.

The change in the i^{th} basis vector e_i relative to the observed and isomorphically defined parallelism is

$$\Delta e_i = e_j\, \omega^j_i(\Delta P)$$

$$= e_j\, \omega^j_i(\underbrace{\Delta x^k e_k}_{\Delta t u = \Delta t\, u^k \frac{\partial}{\partial x^k}})$$

$$= PLP \text{ of } C_u(t)$$

$$= e_j \left(\Gamma^j_{i1}\Delta x^1 + \cdots + \Gamma^j_{in}\Delta x^n\right) = PLP \text{ of } \Delta e_i$$

$$= e_j\, \Gamma^j_{ik}\, \Delta x^k -$$

$$\Delta e_i = e_j \Gamma^j_{ik} \langle dx^k | \Delta x^\ell \frac{\partial}{\partial x^\ell} \rangle ; \quad \left(\text{where } \langle dx^k | \frac{\partial}{\partial x^\ell} \rangle = \delta^k_\ell \text{ expresses duality} \right)$$

$$= e_j \Gamma^j_{ik} \langle dx^k | \Delta \tau u \rangle$$

Thus we have the following

Definitions

$$de_i = e_j \otimes \Gamma^j_{ik} dx^k \text{ , a tensor of rank } \binom{1}{1},$$

$$= e_j \otimes \omega^j_i$$

a) is the covariant diff'l of the basis vector field e_i, $i = 1, \ldots, n$;

b) $\omega^j_i = \Gamma^j_{ik} dx^k \in \binom{0}{1}$

are the connection one-forms.

c) and

$$\Gamma^j_{ik} \in R$$

are the Christoffel symbols of the 2nd kind.

Change in V relative to the observed isomorphically defined parallelism is

$$\Delta V = \Delta(v^i e_i)$$

$$= (v^i + \Delta v^i)(e_i + \Delta e_i) - v^i e_i$$

$$= \Delta v^i e_i + v^i \Delta e_i \quad \text{P.L.P. of } \Delta V$$

$$= e_i \frac{\partial v^i}{\partial x^k} \Delta x^k + e_j v^i \Gamma^j_{ik} \Delta x^k$$

$$= e_j \left(\frac{\partial v^j}{\partial x^k} + v^i \Gamma^j_{ik} \right) \langle dx^k, \underbrace{\Delta t\, u^p \frac{\partial}{\partial x^k}}_{\Delta e} \rangle$$

$$dV = e_j \left(v^j_{,k} + \Gamma^j_{ik} v^i \right) dx^k \in \binom{1}{1}$$

is the <u>covariant differential</u> of the vector $V_p \in V_p$

$$dV \equiv e_j \; v^j_{,k} \, dx^k$$

Here $v^j_{,k} \equiv \frac{\partial v^j}{\partial x^k} + \Gamma^j_{ik} v^i$, the $(j,k)^{th}$ component of dV has two parts

(i) the change of V relative to the chosen basis and

(ii) the change in V to the non-parallelism of the chosen basis.

B) The Covariant Derivative

Note that

$\underset{\sim}{d}V$ = rate of change of V (relative to the observed isomorphically defined parallelism) due to motion by an as-yet-unspecified displacement

Let the displacement be

$$\Delta P = \Delta t\, u,$$

Then

$$\langle dV \mid \Delta t u \rangle = e_j \left(\frac{\partial v^j}{\partial x^k} + \Gamma^j_{ik} v^i \right) \langle dx^k \mid \Delta t \vec{u} \rangle$$

$$= e_j \left(\quad " \quad \right) u^k \Delta t$$

$$= \underbrace{e_j \left(\frac{\partial v^j}{\partial x^k} u^k + \Gamma^j_{ik} v^i u^k \right)}_{(\nabla_u V)} \Delta t$$

$\equiv$ Covariant derivative of V into the direction of u

LECTURE 35 (a superset of Lec. 34)

Covariant derivative of

a) vectors
b) covectors

Reviews Leading to

the Covariant Derivative

1. Summary of the differential of

   (i) a function
   (ii) a vector field

2. Rules for taking Covariant Derivatives

3. Parallel Transport in the Euclidean Plane

4. Covariant Derivatove of a

   (i) vector field
   (ii) covector field
   (iii) tensor field

Reminder:

1. The differential of a scalar field $f(x^i) \equiv f(p)$ is obtained from the Principal Linear Part (PLP) of its Taylor series around a point

$$f(x^k + \Delta x^k) - f(x^k) = \underbrace{\frac{\partial f}{\partial x} \Delta x^k}_{PLP} + \dots, \qquad \Delta x = \Delta t\, u$$

$$= \frac{\partial f}{\partial x^k} \Delta t\, u^k + \dots,$$

$$= \left\langle \frac{\partial f}{\partial x^k}\Big|_p \middle| dx^k \right| \Delta t\, u\Big|_p \rangle + \dots.$$

Thus

$$df = \frac{\partial f}{\partial x^k}\Big|_p dx^k \qquad V_p \to R$$

$$u \rightsquigarrow \boxed{\langle df | u \rangle = \frac{\partial f}{\partial x^k} u^k \Big|_p = \nabla_u f} \tag{1}$$

is a non-linear function of p, and a linear fn on V_p.

Omitting particular reference to vectors in V_p on the principle that df is to be evaluated on some vectors but may be evaluated on any vector. Thus we have

df = rate of deviation away from $f(x^k)$ due to motion into an as-yet-unspecified direction.

2. The differential of a vector field V

$$V(x^i) \equiv V(P) = e_i(P)\, v^i(P)$$

is

$$dV = e_i\, dv^i + de_i\, v^i$$

$$de_i = e_j \underbrace{\Gamma^j{}_{ik}\, dx^k}_{\omega^j{}_i}$$

$$= e_j \left(\frac{\partial v^j}{\partial x^k} + \Gamma^j{}_{ik} \right) dx^k$$

$$dV(\Delta t U) = \langle dV \mid \Delta t U \rangle$$

$$= e_j \left(\frac{\partial v^j}{\partial x^k} + \Gamma^j{}_{ik} \right) \Delta t\, U^k \equiv e_j\, v^j{}_{,k}\, \Delta t\, U^k \quad \text{(notation)}$$

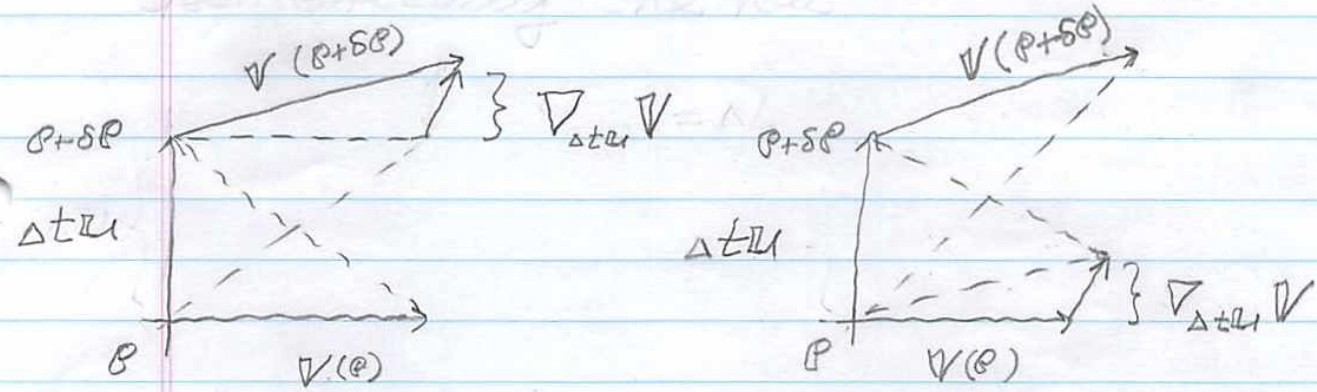
(2) $$\boxed{\langle dV \mid \Delta t\, U \rangle \equiv \nabla_{\Delta t U} V}$$

$\langle dV|u\rangle = \nabla_u V$ = rate of deviation away from

= parallelism
(i.e. rate of change from parallel value of V_p)

due to motion into direction
u

Geometrically one has



From Eqs. (1) & (2) on pages 34.1 & 34.2 one observes that the mapping

$dV: \quad V_p \longrightarrow V_p$

$u \rightsquigarrow \langle dV|u\rangle \equiv \nabla_u V$

has the following four properties

(1) $\nabla_u(V_1+V_2) = \nabla_u V_1 + \nabla_u V_2$

(2) $\nabla_{u_1+u_2}(V) = \nabla_{u_1} V + \nabla_{u_2} V$ } Pointwise linearity in u

(3) $\nabla_{fu} V = f \nabla_u V$ }

(4) $\nabla_u(fV) = f\nabla_u V + V \nabla_u f$ } the Product Rule

Note: The direction derivative $\nabla_u \psi$ of a scalar fn ψ satisfies the same four properties.

The observation on 34.1 - 34.3 are condensed into the following

Definition (Covariant Derivative)

The mapping

$$\nabla V \ (= dV): V_0 \longrightarrow V_0$$

$$u \rightsquigarrow \nabla_u V$$

which satisfies the above boxed four properties is called the covariant derivative of the vector field V.

Remark: The concept of a covariant derivative

(a) is an alternate way of mathematizing parallelism

(b) is the concept of a directional derivative applied to a vector field.

(c) unifies, i.e. integrates, the methods of derivatives of scalars and that of vectors into a single method.

(d) is related to the covariant differential by the fact that the two mapping are one and the same

$$\nabla_u V = \langle dV | u \rangle \quad \forall u \in T_p$$

so that

$$\boxed{\nabla V = dV}$$

Note! The covariant derivatives $\nabla_{e_k} V$ $(k = 1, \ldots, n)$ are given by

a) $\nabla_{e_k} V = \nabla_{e_k} v^l e_l$

$= (\nabla_{e_k} v^l) e_l + v^l \nabla_{e_k} e_l$

$= \frac{\partial v^l}{\partial x^k} e_l + v^l \langle de_l | e_k \rangle$

$= \quad " \quad + v^l \langle e_j \underbrace{\Gamma^j_{li} dx^i}_{\Gamma^j_{lk}} | e_k \rangle$

$$\nabla_{e_k} V = e_j \left(\frac{\partial v^j}{\partial x^k} + v^l \Gamma^j_{lk} \right)$$

b) $\nabla_{e_k} e_l = e_j \Gamma^j_{lk}$

Q: What are the covariant derivatives of the dual basis elements $\{\omega^i\}$?

$$\langle \omega^i | e_j \rangle = \delta^i_j .$$

A: They are determined by the requirement that duality is

invariant under // transport

$$\nabla_{e_k}\left(\langle \omega^i | e_j \rangle = \delta^i_j\right) = 0.$$

A direct calculation applied to the PLP yields

$$0 = \nabla_{e_k}(\delta^i_j) = \nabla_k \langle \underset{\sim}{\omega}^i | e_j \rangle$$

$$= \langle \nabla_k \underset{\sim}{\omega}^i | e_j \rangle + \langle \omega^i | \nabla_k e_j \rangle$$

$$= \underbrace{\langle \nabla_k \underset{\sim}{\omega}^i | e_j \rangle}_{} + \Gamma^i_{jk}$$

$$\langle \left(\nabla_k \underset{\sim}{\omega}^i\right)_m \omega^m | e_j \rangle = \left(\nabla_k \underset{\sim}{\omega}^i\right)_j$$

$$\therefore \left(\nabla_k \underset{\sim}{\omega}^i\right)_j = -\Gamma^i_{jk}$$

or $$\boxed{\nabla_k \underset{m}{\omega}{}^i = -\Gamma^i_{jk}\, \underset{m}{\omega}{}^j}$$

More generally, using the product rule one has

$$\nabla_k \underset{\sim}{\sigma} = \nabla_k (\sigma_i \underset{\sim}{\omega}^i)$$

$$= \sigma_{i,k}\, \underset{\sim}{\omega}^i + \sigma_i \nabla_k \underset{\sim}{\omega}^i$$

$$= \left(\sigma_{j,k} - \sigma_i \Gamma^i_{jk}\right) \underset{\sim}{\omega}^j$$

$$\nabla_{e_k} \underset{\sim}{\sigma} = \underbrace{\sigma_{j;k}}_{(jk)^{th}} \omega^j$$

component of the covariant derivative of $\underset{\sim}{\sigma} = \sigma_i \underset{\sim}{\omega}^i$

Different but also good

Covariant Derivative of a Vector Field

Let $V = e_i v^i(x)$ be an arbitrary vector field. The derivative of V at a point into the direction of the vector $U = e_k u^k(x)$ is obtained by the following computation

$\nabla_U V = \nabla_U (v^i e_i)$ $\neq 0$ whenever e_i is not parallel.

$= \nabla_U (v^i) e_i + v^i \nabla_U e_i$

$= u^k \frac{\partial v^i}{\partial x^k} e_i + v^i \langle de_i | U \rangle$

$= u^k \frac{\partial v^j}{\partial x^k} e_j + v^i \langle e_j \otimes \Gamma^j_{ik} dx^k | U \rangle$ $\quad (\nabla_{e_k} e_i = \Gamma^j_{ik} e_j)$

$= u^k \frac{\partial v^j}{\partial x^k} e_j + v^i (e_j \Gamma^j_{ik}) u^k$

$\equiv e_j \{ v^j_{,k} + v^i \Gamma^j_{ik} \} u^k$

$\equiv e_j \; v^j_{;k} \; u^k$.

The elements $v^j_{;k}$ are referred to as the components of the covariant derivative of V.

Covariant Derivative of a Covector Field.

Parallel transport of $\underset{\sim}{\omega}^i$ is determined by the preservation of duality.

Reminder: $\nabla_k e_j = e_\ell \Gamma^\ell_{jk}$

$$0 \equiv \nabla_{e_k} \delta^i_j = \nabla_{e_k} \langle \underset{\sim}{\omega}^i | e_j \rangle = \langle \nabla_k \underset{\sim}{\omega}^i | e_j \rangle + \langle \omega^i | \nabla_k e_j \rangle$$

$\nabla_k \omega^i$: j^{th} coordinate of $\nabla_k \omega^i$; Γ^i_{jk} : j^{th} coordinate of $\nabla_k e_j$

Thus
$$\langle \nabla_k \omega^i | e_j \rangle = -\Gamma^i_{jk}$$

Reminder: $\langle \underset{\sim}{\sigma} | e_j \rangle = \langle \sigma_m \omega^m | e_j \rangle = \sigma_j = j^{th}$ coord of $\underset{\sim}{\sigma}$; δ^m_j

or
$$\nabla_k \underset{\sim}{\omega}^i = \underset{\sim}{\omega}^j \langle \nabla_k \omega^i | e_j \rangle = -\underset{\sim}{\omega}^j \Gamma^i_{jk}$$

For a general 1-form $\sigma_i \omega^i$ one uses the product rule of differentiation and obtains

$$\nabla_k (\sigma_i \underset{\sim}{\omega}^i) = (\sigma_{j;k} - \Gamma^i_{jk} \sigma_i) \underset{\sim}{\omega}^j$$

Thus
$$\boxed{\nabla_k \underset{\sim}{\sigma} \equiv \sigma_{j;k} \underset{\sim}{\omega}^j}$$

$$\nabla_{u^k e_k} \underset{\sim}{\sigma} = \nabla_{u^k e_k} \underset{\sim}{\sigma} = u^k \nabla_{e_k} \underset{\sim}{\sigma}$$
$$= \sigma_{j;k} u^k \underset{\sim}{\omega}^j$$

LECTURE 36

1. Covariant Derivative of a Tensor
2. What is a Tensor Map?
3. The Torsion Tensor

I) Covariant Derivative of a (1,1) Tensor via the covariant derivative's product rule

$$\nabla_u T = \nabla_u \left(T^i{}_j \, e_i \otimes \omega^j \right)$$

$$= \nabla_u (T^i{}_j) \, e_i \otimes \omega^j + T^i{}_j (\nabla_u e_i) \otimes \omega^j + T^i{}_j \, e_i \otimes \nabla_u \omega^j$$

$$= \frac{\partial T^i{}_j}{\partial x^k} u^k e_i \otimes \omega^j + T^i{}_j \, e_\ell \Gamma^\ell{}_{ik} u^k \otimes \omega^j + T^i{}_j \, e_i \otimes (-) \Gamma^j{}_{\ell k} \omega^\ell u^k$$

Use index gymnastics in order to isolate the $(i\,j\,k)^{th}$ component of $\nabla_u T$:

$$\nabla_u T = u^k \left[\frac{\partial T^i{}_j}{\partial x^k} e_i \otimes \omega^j + T^\ell{}_j \, \Gamma^i{}_{\ell k} \, e_i \otimes \omega^j - T^i{}_\ell \, \Gamma^\ell{}_{jk} \, e_i \otimes \omega^j \right]$$

$(\ell \leftrightarrow i)$ $(\ell \leftrightarrow j)$

$$= \underbrace{\left[T^i{}_{j,k} + T^\ell{}_j \, \Gamma^i{}_{\ell k} - T^i{}_\ell \, \Gamma^\ell{}_{jk} \right]}_{\| \; T^i{}_{j;k}} u^k e_i \otimes \omega^j$$

observe that with each index "i" and "j" there is a the respective correction term $T^{\ell}_{ij}\,\Gamma^{i}_{\ell k}$ and $-T^{i}_{\ell j}\Gamma^{\ell}_{jk}$.

II. Tensor map vs non-tensor map as illustrated by

2.7 Covariant Derivative vs Lie Derivative

Given two vector fields, say $\mathbb{U}$ and $\mathbb{V}$, the covariant derivative $\nabla_{\mathbb{U}}\mathbb{V}$ and the commutator $[\mathbb{V},\mathbb{U}] = \mathcal{L}_{\mathbb{V}}\mathbb{U}$, also known at the "Lie derivarite with respect to $\mathbb{V}$" are two ways of generating new vectors at each point of the nonlinear manifold. One is based on a parallel transport structure, the other on the differential structure. One can, however, point to a more specific distinction between the two: the covariant derivative is a <u>tensor map</u> while the Lie bracket is not.

Question: In what sense is $\nabla_{\mathbb{U}}\mathbb{V}$ a tensor map?

Answer: Given a vector

$$\nabla\mathbb{V} : V_p \to V_p$$
$$\mathbb{U} \mapsto \nabla_{\mathbb{U}}\mathbb{V}$$

is a tensor map because ∇V is pointwise linear, i.e.,

$$f u_1 + g u_2 \rightsquigarrow f \nabla_{u_1} V + g \nabla_{u_2} V .$$

Thus ∇V is a tensor (map) at each point p.

By contrast, the Lie derivative (or "Lie bracket")

$$\mathcal{L}_V : V_p \longrightarrow V_p$$
$$u \rightsquigarrow \mathcal{L}_V u \equiv [V, u]$$

is not a tensor map because

$$f u \rightsquigarrow [V, fu] = f[V, u] + V(f)\, u$$
$$\left(= f \mathcal{L}_V u + u \mathcal{L}_V(f)\right)$$

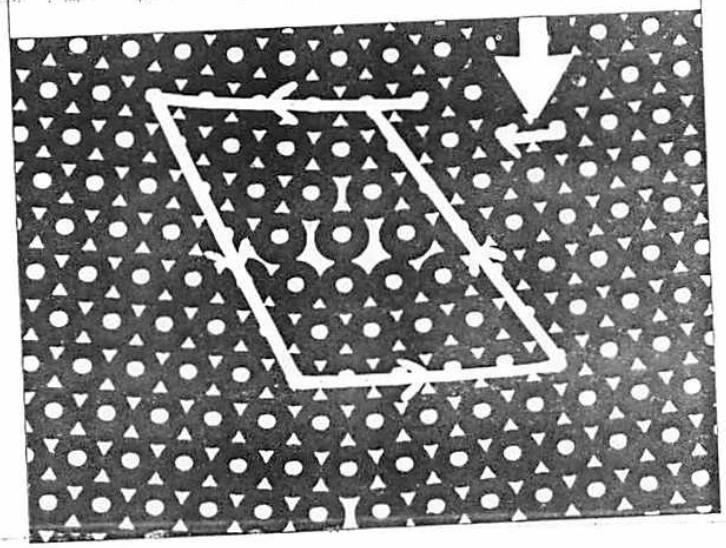
is not pointwise linear.

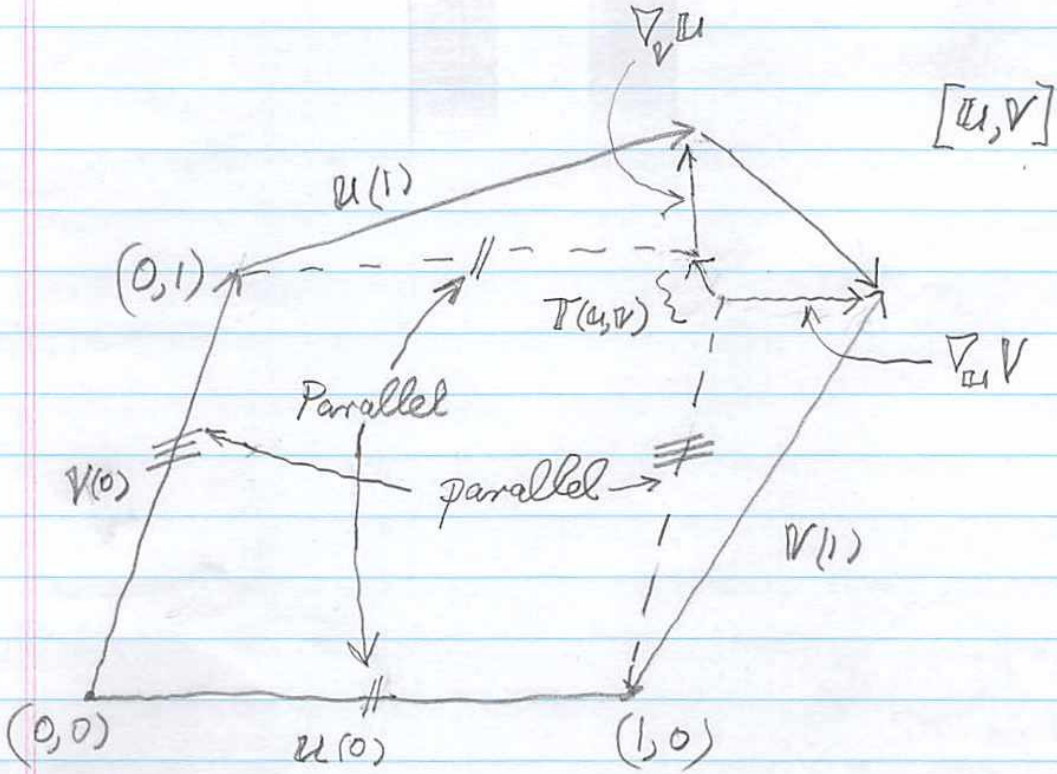
III. Torsion as a Trait of Parallel Transport.

There exist other types of parallel transport besides the one induced by the Schild ladder construction highlighted near the end of Lecture 32. There the parallelograms are closed closed. However, in condensed matter physics there is a parallelism where the parallelogram are not closed. This is the case with crystaline material permeated by a density of dislocations (point, edge, screw, etc). The figure below depicts a single point dislocation inside a parallelogram whose opposite and parallel sides consist of arrays

having an equal number of atoms. The non-closure of the parallelogram is depicted by a vector of one atomic unit pointing into the negative horizontal unit. This non-closure vector is the Burgers vector. It is the value of the "Cartan torsion tensor" which mathematizes the dislocations per parallelogram area,

Torsion $\neq 0$





Nonzero edge or a skrew dislocations prevent the parallelograms from being closed. The resulting <u>Burger vector</u>

$$\nabla_U V - \nabla_V U - [U,V] \equiv T(U,V),$$

which mathematizes the distribution of dislocations in the u-v quadrilateral,

is non zero.

T is called the Cartan torsion. It is the dislocation density inside the u-v quadrilateral.

Remark

a) Cartan torsion is a tensor map. This is because a calculation based on the properties of the covariant derivative and the commutator shows that

$$T(fu, v) = \nabla_{fu} v - \nabla_v (fu) - [fu, v]$$
$$= fT(u, v)$$

and

$$T(u, gv) = gT(u, v).$$

This pointwise linearity shows that

$$\Pi : V_p \times V_p \longrightarrow V_p$$

$$(u, v) \rightsquigarrow \Pi(u, v)$$

is a <u>tensor</u> indeed.

b) $T_p \in V_p^* \otimes V_p^* \otimes V_p$

Q: What is the rank $\binom{n}{m}$ of this tensor?

A: $\binom{n}{m} = \binom{2}{1}$,

Q: a) Given Γ^i_{jk}, what are the <u>components</u> of Π relative the coordinate-induced basis $\{e_i = \frac{\partial}{\partial x^i}\}$?

b) What is Π expressed in terms of these basis elements and its duals?

LECTURE 37

T E

Curvature

1. a) Rotation of a Vector (Parallel)

   b) Transported Around a Loop

2. Riemann Curvature Tensor

not developed in class:

3. Equation of Geodesic Deviation (but in LECTURE 38, 39)

4. Rotation = Relative Acceleration

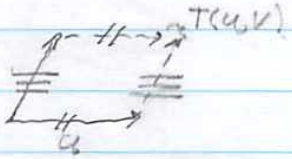
Wheeler's First Moral Principle

Never make a calculation until you know the answer. Make an estimate before every calculation, try a simple physical argument (symmetry! invariance! conservation!) before every derivation, guess the answer to every puzzle.

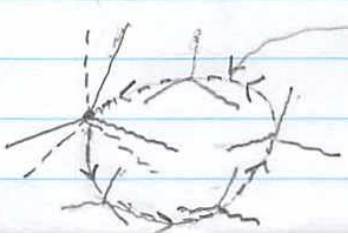
(Taken from Space-Time Physics, 2nd Edition, Taylor & Wheeler 1993)

Parallel transport mathematized by the covariant derivative ∇ has two constitutive traits:

a) Torsion, which is a tensor that mathematizes a non-zero gap (the Burgers vector) between the parallel transported edges of a parallelogram



b) Curvature, which is a tensor that mathematizes a non-zero rotation resulting from parallel transporting a basis of vector around a closed loop



In a curved 3-d space with an inner product in each tangent space V_p, this rotation is expressed in term of the angular velocity

$$\vec{\omega} = (\omega_x, \omega_y, \omega_z)$$

by means of the orthogonal matrix $R_{\vec{\omega}}(t)$:

$$R_{\vec{\omega}}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} t + \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}^2 \frac{t^2}{2!} + \cdots$$

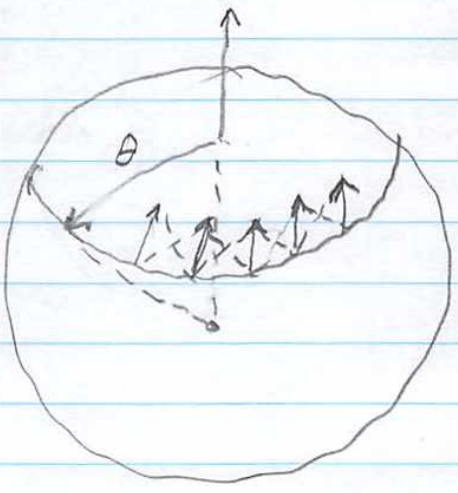
$$R_{\vec{\omega}}(t) \simeq e^{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} t}$$

curvature induced "rotation" due to motion around a closed loop, in time t

"Principal Linear Part" of $R_{\vec{\omega}}(t)$.

Curvature-induced rotation is concretized by parallel transporting each basis vector individually around a loop permeated by curvature.

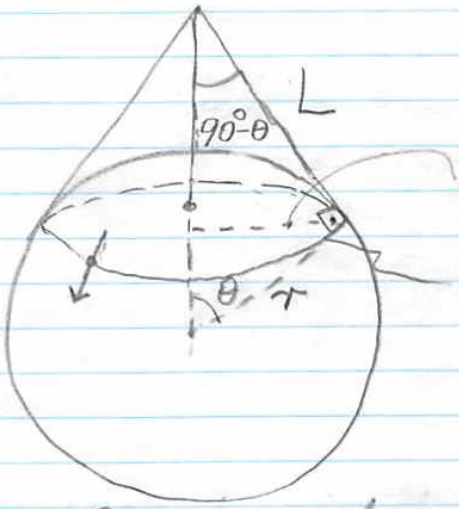
Apply the Schild ladder parallel transport to a vector on the θ-latitude of S^2



Q: What is the result of this transport around the closed θ-laditude circle

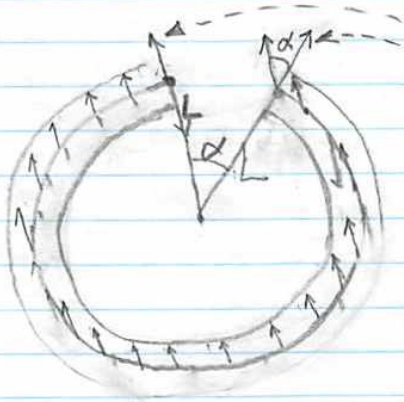
A: The result of this $\parallel$ transport is exhibited by

cutting out an annular strip surrounding the circle of constant latitude, θ = constant. This annular strip is tangent to the bottom of a cone whose apex angle is $2 \times (90° - \theta)$.



θ-latitude circle
circumference =
$= 2\pi L \sin(90° - \theta)$

Cut this cone along an edge, and flatten the cone out so that it becomes a disk with a missing sector of angle α.



same vector on S^2.

radius = L

Total circumference = $2\pi L$
of the completed
circle

The parallel vector field is along the arc perimeter of the disk. It is evident that after a full circuit of parallel transport, a vector gets rotated by the angle α.

More precisely, we have

α = Rotation angle of a vector after parallel transport around a meridean circle of latitude θ away from the North pole

$$= \frac{\text{arclength of completed circle} - \text{arclength of the } \theta\text{-latitude circle}}{L}$$

$$= \frac{2\pi L - 2\pi L \sin(90^\circ - \theta)}{L}$$

$$= 2\pi(1 - \cos\theta)$$

This angle leads us to the following tentative definition:

The curvature permeating an enclosed area yields the angular rotation suffered by a vector parallel transported around the perimeter of that area:

$$\boxed{\alpha \equiv (\text{"curvature"}) \times \text{enclosed area}}$$

intrinsic to S^2

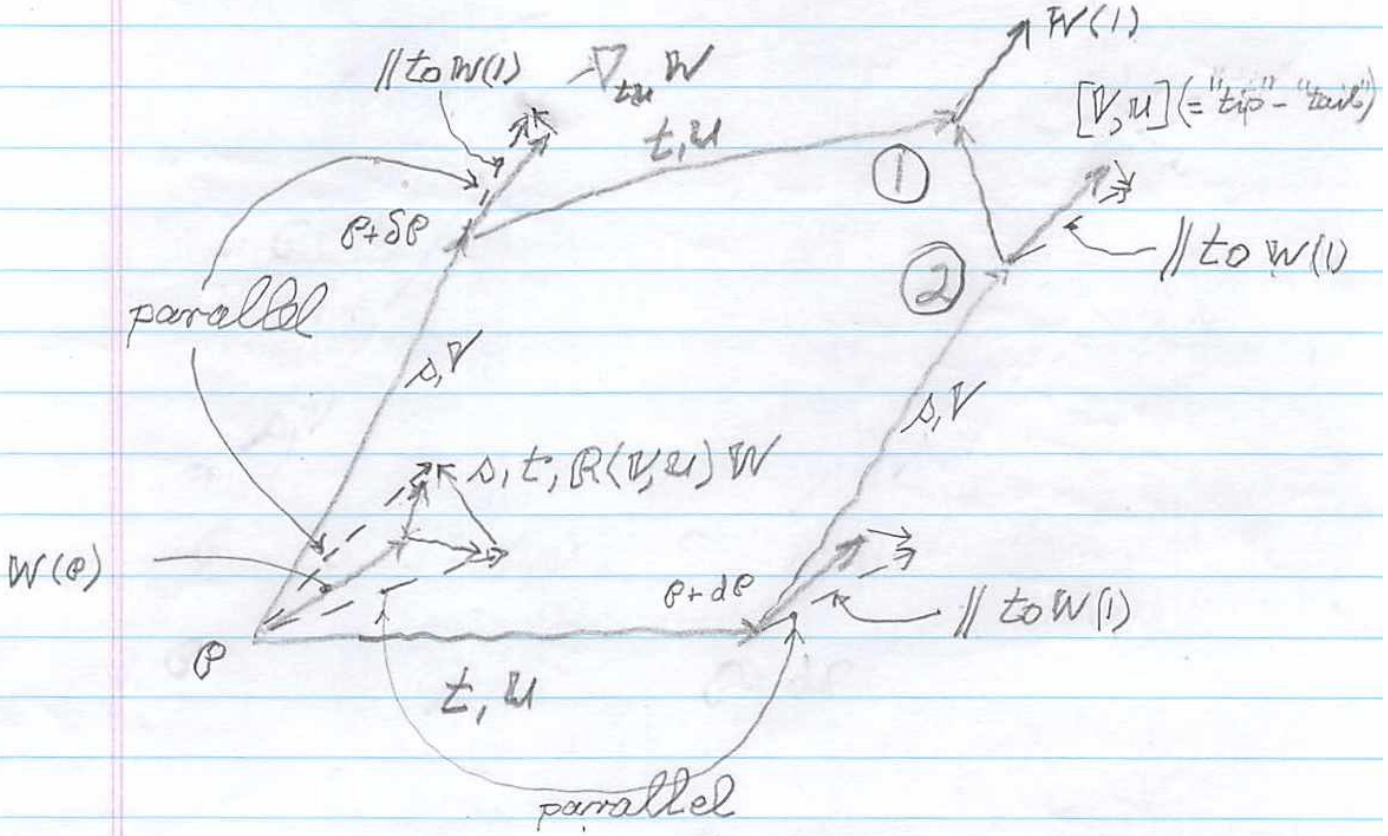
or

$$\text{curvature} \equiv \frac{\alpha}{\underbrace{\int_0^{2\pi} d\varphi \int_0^{\theta} r^2 \sin\theta \, d\theta}_{\text{enclosed area}}} = \frac{2\pi(1-\cos\theta)}{2\pi r^2(1-\cos\theta)} = \frac{1}{r^2} \equiv \text{curvature} \left(\begin{array}{c}\text{independent of} \\ \text{enclosed} \\ \text{area}\end{array}\right)$$

The radius r is called the radius of curvature of the enclosed area.

Curvature Mathematized via Covariant Derivative

a) Consider a closed loop bounded by generic vector fields u, v and $[v,u]$ and a generic vector fied w on this loop



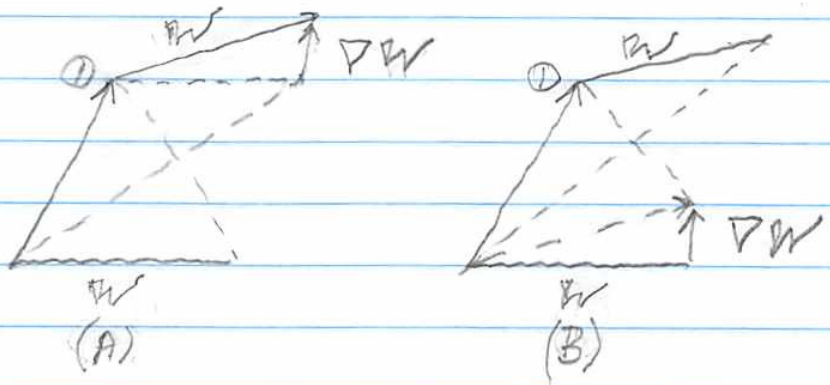
There are two paths along which to parallel transport (via the Schild ladder construction) from point ① to point P

(i) ① → P+δP → P

and

(ii) ① → ② → P+dP → P.

Of the two Schild ladder constructions



use the pullback construction (B) in order to construct the parallel translate of W in term in terms of the pre-existing W plus the vectorial change ΔW.

LECTURE 37

1. Curvature

a) Rotation of a vector parallel transported around a loop.

b) Rieman curvature tensor
   (i) Definition
   (ii) Its pointwise linearity.

Wheeler's First Moral Principle

Never make a calculation until you know the answer. Make an estimate before every calculation, try a simple physical argument (symmetry! invariance! conservation!), before every derivation guess the answer to every puzzle.

(Taken from Spacetime Physics, 2nd Edition by Taylor and Wheeler

The parallel translates of W at D, which are depicted as dashed line segments in the Figure on page 37.7, form vector fields along broken curves (i) and (ii). However, the values of these vector fields at their common reference point P are not the same. The difference, s.t. $R(V,U)W$, is a vector which is non-zero. This vector mathematizes the amount of curvature-induced rotation suffered by a vector W as it is parallel translated around the closed curve

$$P \to P+dP \to (2) \to (1) \to P+\delta P \to P$$

This closed curve decomposes into two different paths

$$① \to P+\delta P \to P$$

and

$$① \to ② \to P+d P \to P$$

They accommodate two different vectorial Taylor series expansions relative to their common expansion point P.

b) The vectorial Taylor series expansion along $① \to P+\delta P \to P$ (the "high road") relative to P is done as follows:

Parallel transport (go to next page)

yields the // translate first to $P+\delta P$,

$$(i)\quad W(P+\delta P) + t_1 \nabla_U W\Big|_{P+\delta P} + \frac{t_1^2}{2!} \nabla_U \nabla_U W\Big|_{P+\delta P} + \cdots .$$

= dotted vector at $P+\delta P$ which is // to $W(1)$

and then to P

$$(ii)\quad W(P) + \delta_1 \nabla_V W\Big|_P + \frac{\delta_1^2}{2!} \nabla_V \nabla_V W\Big|_P + \cdots$$

$$+ t_1 \left[\nabla_U W\Big|_P + \delta_1 \nabla_V \nabla_U W\Big|_P + \cdots \right]$$

$$+ \frac{t_1^2}{2!} \left[\nabla_U \nabla_U W\Big|_P + \cdots \right]$$

= vector at P // to $W(1)$ via $1 \to P+\delta P \to P$,

c) Parallel transport along

① → ② → P+dP → P (the "low road")

yields the following vector

at point ②

$= W(2) + s, t, \nabla_{[U,V]} W + \cdots$

At point P+dP

$= W(P+dP) + s, \nabla_V W\Big|_{P+dP} + \frac{s^2}{2} \nabla_V \nabla_U W\Big|_{P+dP} + \cdots$

At point P

$+ s, t, \nabla_{[U,V]} W\Big|_{P+dP} + \cdots$

$= W(P) + t, \nabla_U W\Big|_P + \frac{t^2}{2!} \nabla_U \nabla_U W + \cdots$

$s, \left[\nabla_V W\Big|_P + t, \nabla_U \nabla_U W\Big|_P + \cdots \right]$

$+ \frac{s^2}{2} \left[\nabla_V \nabla_U W\Big|_P + \cdots \right] + s, t, \nabla_{[U,V]} W$

This is the same as for $1 \to P+\delta P \to P$ except that $t, U \leftrightarrow s, V$ plus the additional term

$s, t, \nabla_{[U,V]} W$

d) The difference at $\mathcal{P}$ is

(Parallel translate of $W(1)$ to point $\mathcal{P}$ around 1st ("upper") broken path $① \to \mathcal{P}+\delta\mathcal{P} \to \mathcal{P}$) −

(parallel translate of $W(1)$ to point $\mathcal{P}$ around 2nd ("lower") broken path $① \to ② \to \mathcal{P}+d\mathcal{P} \to \mathcal{P}$)

$$= \delta, t, \underbrace{\left[\nabla_V \nabla_U - \nabla_U \nabla_V - \nabla_{[V,U]}\right]}_{\text{the "curvature" operator}} W$$

$$\equiv \delta, t, \overbrace{R(V,U)}\, W$$

$$\equiv \delta, \tau, \text{"Riemann}(\cdots, W, V, U) = \begin{pmatrix} \text{a vector} \\ \text{at } \mathcal{P} \end{pmatrix}$$

= vectorial amount of rotation that W @ $\mathcal{P}$ undergoes by parallel transporting it around the closed quadrilateral.

The curvature operator

$$R(V,U) = \nabla_V \nabla_U - \nabla_U \nabla_V - \nabla_{[V,U]}$$

is a 2nd order operator when it is applied to the vector field W.

Nevertheless

$$R(V,U)W = \{\nabla_V \nabla_U - \nabla_U \nabla_V - \nabla_{[V,U]}\} W$$

is pointwise linear, i.e. upon

replacing V with fV
U with gU
W with hW

all 1st and 2nd derivatives of f, g, and h cancel out, and one is left with

$$R(fV,U)W = f\,R(V,U)$$
$$R(V,gU)W = g\,R(V,U)$$
$$R(V,U)hW = h\,R(V,U).$$

whenever f, g, and are smooth scalar functions.

Conclusion

R is a tensor

Let $V = e_i v^i$

$U = e_j u^j$

$W = e_k w^k$

Then

$$R(\cdots, W, V, U) = \left[\nabla_V \nabla_U - \nabla_U \nabla_V - \nabla_{[V,U]}\right] W$$

$$= e_\ell R^\ell{}_{k i j}\, w^k v^i u^j$$

$$= e_\ell \otimes \omega^k R^\ell{}_{k i j}\, \omega^i \wedge \omega^j / 2!\,(W, V, U)$$

Q: (Problem 2b on HW set 7)

What are the components of R relative to a coordinate basis?

LECTURE 38

I. Torsion and Curvature:

Their basis representation

Their components relative a a given/chosen basis and its dual.

II. The Equation of Geodesic Deviation.

Quote of the Day:

"In the long ago, no man was a man who could wield a sword. Today nobody can be anybody who cannot express himself with clarity and force"

— Ilsion —

I. Tensor calculus mathematizes change of geometrical objects (vectors, covectors, tensors on a continuum) relative to a specified law of parallel transport in terms of the covariant derivative ∇. There exist many different types of parallel transport. They are distinguished by their respective tensor maps,

Torsion: $T(u,v) = \nabla_u v - \nabla_v u - [u,v]$

Curvature: $R(u,v)w = \nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]}\} w \equiv R(\cdot,w,u,v)$

Their point wise linearity allows them being expanded relative to

any chosen/given basis

$u = e_i u^i$

$v = e_j v^j$

$w = e_k w^k$

Problem 1

a) Compute and exhibit relative to a coordinate basis the components of the torsion T in terms of the Christoffel symbols Γ^i_{jk}.

b) Same as a) except that the components are those of the curvature R.

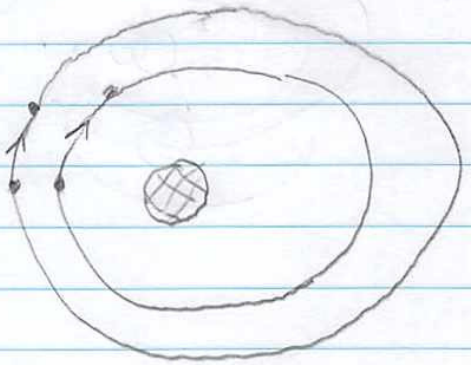
Problem 2

a) Exhibit the torsion tensor as a tensor product of the basis elements of $\{e_i\}$ and and the dual basis $\{\omega^i\}$.

b) Same as a) except for the curvature tensor R

II. Curvature as a mathematizer of tidal acceleration: The equation of geodesic deviation.

a) Consider the trajectories of a pair of bodies (e.g. a space station and a supply ship) orbiting around the earth or around a black hole.

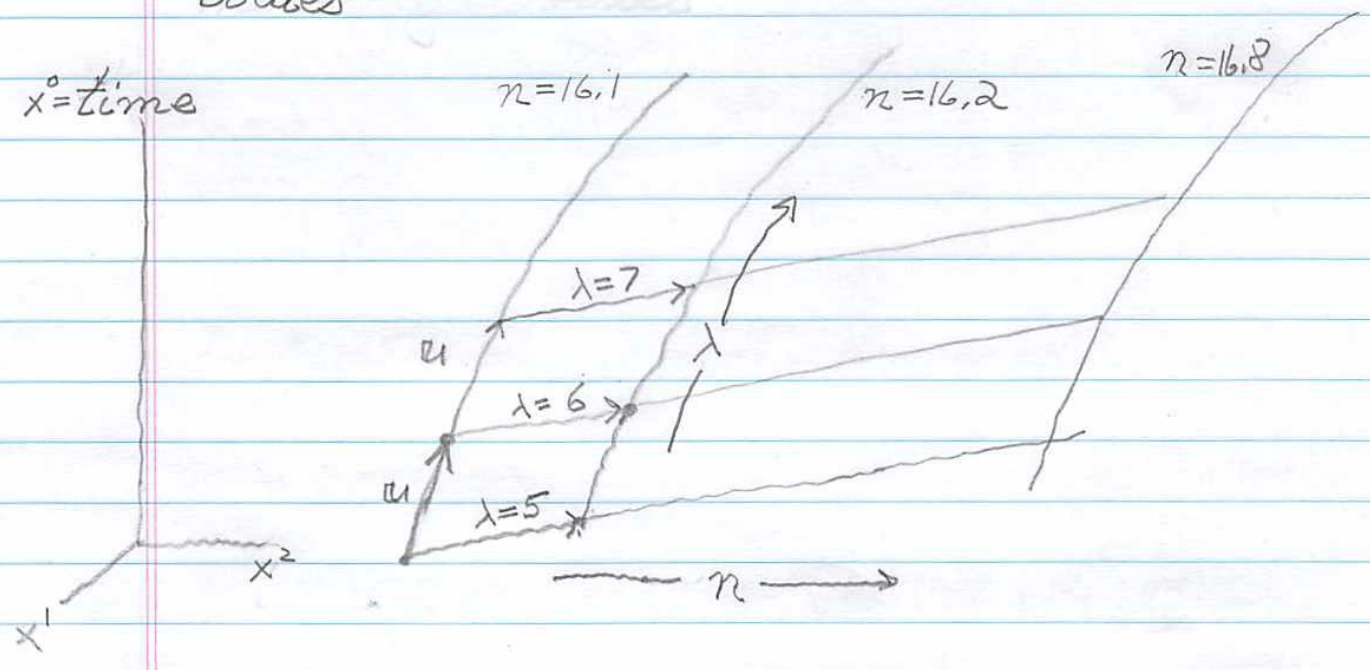


In Newtonian gravity the separation between them is mathematized in terms of their tidal acceleration, which is due

to the inhomogeneous gravitational field between them.

Within the geometrical framework of Einstein's mathematization of gravity this acceleration is due to the curvature between two geodesics

b) Consider the spacetime worldlines of adjacent freely (i.e. geodesically) moving bodies



1. In order to mathematize the relative motion of these bodies identify their motions as a 1-parameter family of geodesics

$$P(\lambda, n),$$

where n, which labels the bodies, connects equal values of λ common to their geodesics.

The curves corresponding to fixed parameter n are geodesics whose tangent are

$$u = \frac{\partial}{\partial \lambda}\Big)_n$$

They satisfy the geodesic equation

$\nabla_u u = 0$

The tangent vector u is the 4-velocity of a particular body.

$\left(e_{i'} \left(\frac{\partial u^{i'}}{\partial x^k} + u^{j'} \Gamma^{i'}_{jk} \right) u^k = 0 ; \text{ the tangents are auto-parallel} \right)$

2. Let

$$\mathbb{R} = \frac{\partial}{\partial n}\Big)_\lambda$$

be the displacement vector which connect adjecent worldlines common time parameter λ. The vector mathematizes their coordinate differences.

$$\mathbb{R}(x^{i'}) = \frac{\partial x^{i'}}{\partial n}\Big)_\lambda$$

or the change in any other scalar fn:

$$\mathbb{R}(f) = \frac{\partial f}{\partial n}\Big)_\lambda$$

The vector $\mathbb{R}$ is called the deviation vector

From the definition of the vectors $u = \frac{\partial}{\partial \lambda}\Big)_n$ and $n = \frac{\partial}{\partial n}\Big)_\lambda$ as world line tangents and connecting vectors between adjacent bodies, it follows they commute

$$[n, u] = 0$$

3. a) Consider the curvature

$$R(n, u)\,u = \left[\nabla_n \nabla_u - \nabla_u \nabla_n - \nabla_{[n,u]}\right] u$$

in the plane spanned by λ and n.

In light of $\nabla_u u = 0$ and $[n, u] = 0$

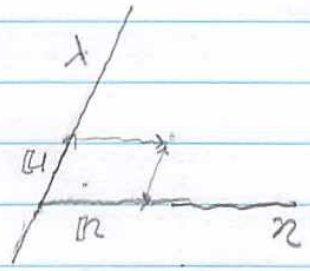
$$\nabla_u \nabla_n u + R(n, u)\,u = 0$$

b) Physical observations tell us that there is no torsion:

$$T(n, u) \equiv \nabla_n u - \nabla_u n = 0$$

Thus one find that the separation vector satisfies the equation of geodesic deviation

$$\boxed{\nabla_u \nabla_u \eta + R(\eta, u)u = 0}$$



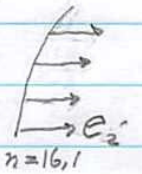
where $\rho(\lambda, \eta) = \{ x^i(\lambda, \eta)$

next extreme

Let $\eta = \eta^i e_i$ where $\eta^i = \frac{\partial x^i}{\partial \eta}$

where $\{e_1, \ldots, e_n\}$ is a basis parallel along $\eta = 16,1$

$\nabla_u e_i = 0$;



Then $\nabla_u \eta = \nabla_u (\eta^i) e_i = \frac{\partial x^k}{\partial \lambda} \frac{\partial \eta^i}{\partial x^k} e_i = \frac{d\eta^i}{d\lambda} e_i$

and

$$\frac{d^2 \eta^i}{d\lambda^2} e_i + R(\eta^j e_j, u) \overbrace{u^k e_k}^{u} = 0$$

$$\overbrace{u^l e_l}$$

$$e_i \frac{d^2 \eta^i}{d\lambda^2} + e_i R^i{}_{kjl} u^l u^k \eta^j = 0$$

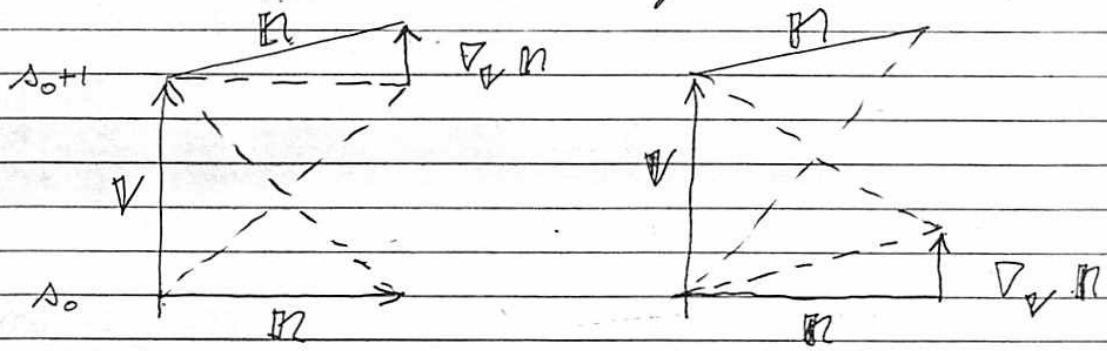
4.4 Rotation = Tidal Acceleration = Geodesic Deviation } not done in class 38.9

(i.e. tidal forces)

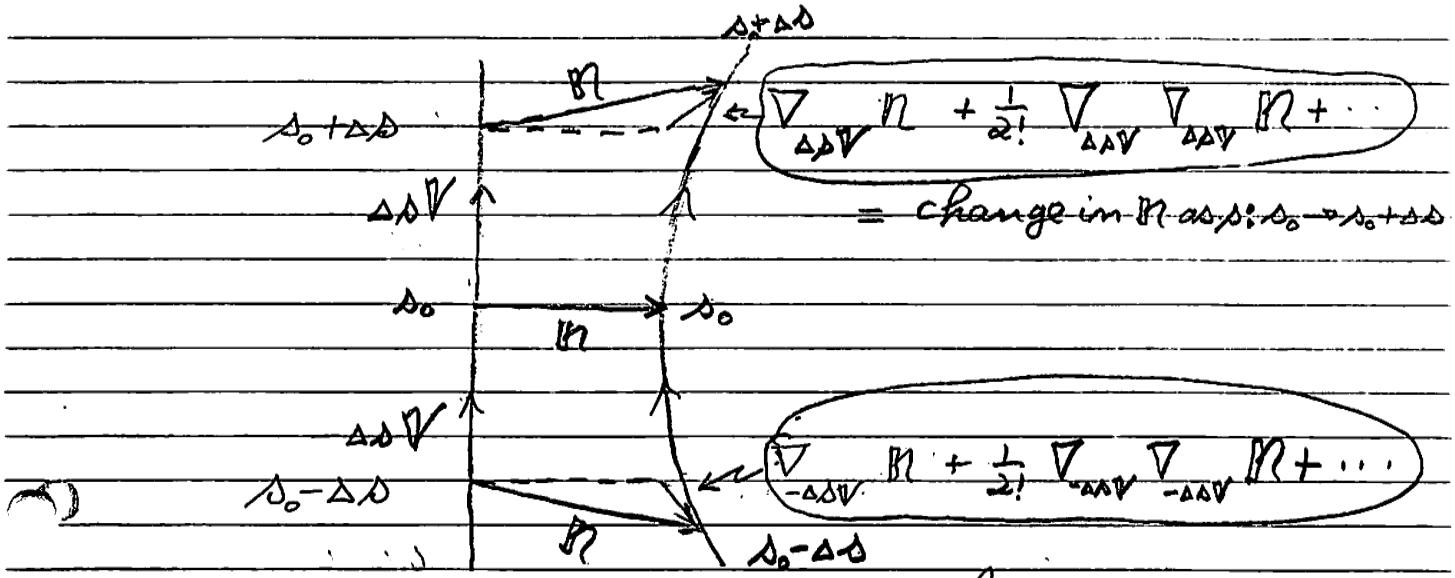
Curvature manifests itself in two seemingly different ways: (i) Rotation of a vector upon being carried around a closed loop and (ii) relative acceleration between a pair of nearby geodesics.

We shall show that these two are in fact different aspects of curvature.

Everything hinges directly on the idea of paralell transport which is expressed by the two mnemonic diagrams



Consider two close-by geodesics and the evolution of the connecting vector $\vec{n}$ between them



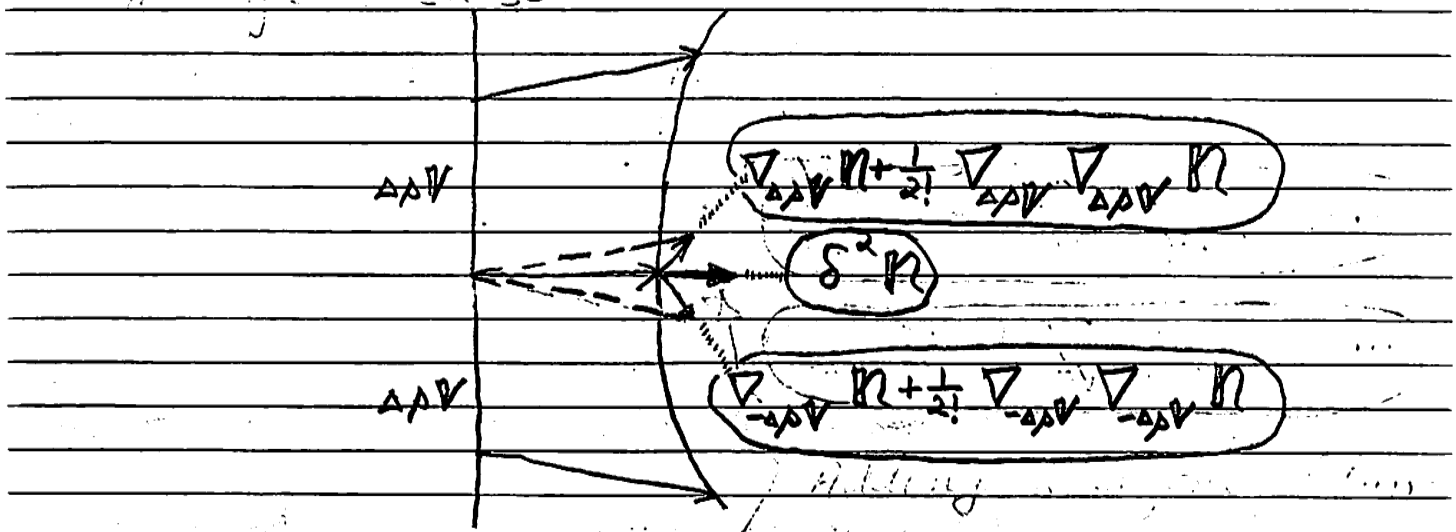
$$\nabla_{\Delta s \vec{V}} \vec{n} + \frac{1}{2!}\nabla_{\Delta s \vec{V}}\nabla_{\Delta s \vec{V}}\vec{n} + \cdots$$

= change in $\vec{n}$ as $s: s_0 \to s_0 + \Delta s$

$$\nabla_{-\Delta s \vec{V}} \vec{n} + \frac{1}{2!}\nabla_{-\Delta s \vec{V}}\nabla_{-\Delta s \vec{V}}\vec{n} + \cdots$$

= change in $\vec{n}$ as $s: s_0 \to s_0 - \Delta s$

or equivalently

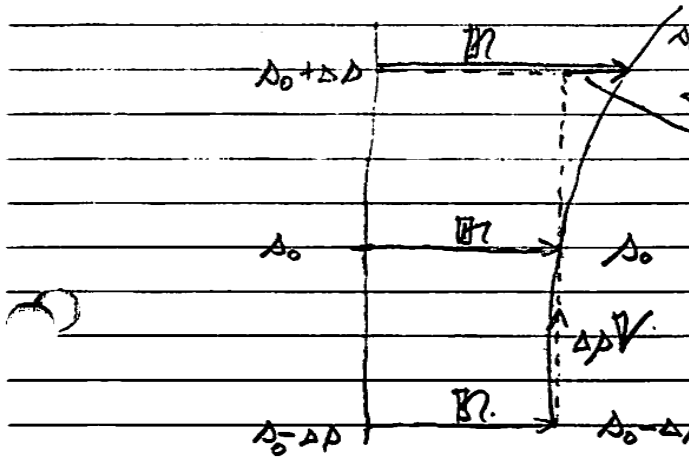


Adding the two, one obtains the total change in $\vec{n}$

as $s: s_0 - \Delta s \to s_0 + \Delta s$

$$\boxed{\delta^2\vec{n} = \frac{2}{2!}(\Delta s)^2 \nabla_{\vec{V}}\nabla_{\vec{V}}\vec{n}}$$

Without changing the relative acceleration between the two geodesics, we can tilt the right hand geodesic clockwise around s_0 until the two geodesics are parallel at $s_0 - \Delta s$. In that case the picture at the top of PG becomes



$$\nabla_{\Delta_p V}\nabla_{\Delta_p V}\eta = \delta^2\eta$$

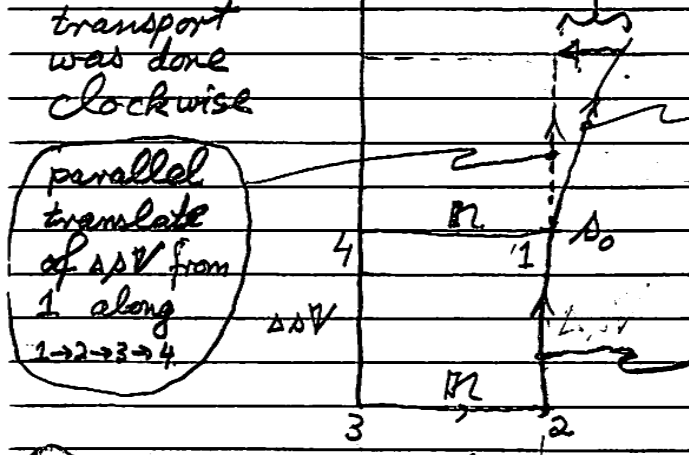
or

$$-(\Delta s)^2\nabla_V\nabla_V\eta = -\delta^2\eta$$

EQUAL

$$-\delta^2\eta \left[= -R(\Delta_s V, \eta)\Delta_p V\right]$$

minus sign because transport was done clockwise



parallel translate of $\Delta_p V$ from 1 along $1\to2\to3\to4$

$\Delta_p V$ = tangent at s_0

PARALLEL because curve is a geodesic: $\nabla_V V = 0$

$\Delta_p V$ = tangent at $s_0 - \Delta s$

Comparing the two pictures on this page we obtain

$$\frac{\delta^2\eta}{(\Delta s)^2} = -R(V,\eta)V$$

or

$$\boxed{\frac{\delta^2\eta}{(\Delta s)^2} + R(\eta, V)V = 0}$$

as it should be

LECTURE 39 (continuation of Lec. 38)

I. Understanding the eq'n of geodesic deviation

II. Metric Tensor Field

1. 2-d surface in E^3

consigned to Lecture 40:

2. strain tensor as a distortion in the metric tensor

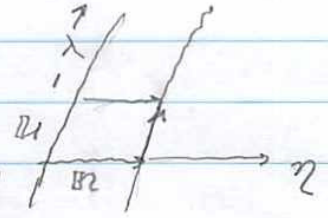
Thus one finds that the separation vector satisfies the eq'n of geodesic deviation

$$\nabla_u \nabla_u \eta + R(\eta, u) u = 0$$

$$\nabla_u \nabla_u \eta + R(\dots, u, \eta, u) = 0$$

$$\rho(\lambda, \eta) = \{ x^i(\lambda, \eta) \}; \quad u = \frac{\partial \rho}{\partial \lambda} = \frac{\partial x^k}{\partial \lambda} \frac{\partial}{\partial x^k} \equiv u^k e_k$$

$$\eta = \frac{\partial \rho}{\partial \eta} = \frac{\partial x^i}{\partial \eta} \frac{\partial}{\partial x^i} \equiv \eta^i e_i$$



The best way of understanding the content of the eq'n of geodesic deviation is to choose an appropriate vector basis at each point

consigned to LECTURE 39

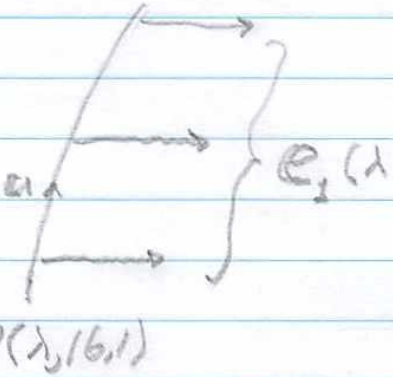
At one point point of the geodesic, say $\eta = 16.1$, choose any basis

$$\{ e_1, e_2, \dots, e_n \}$$

which may contain u as one of the basis vectors.

Extend this basis to all points of the geodesic by using these basis vectors as initial condition for the solutions to the eq'ns of parallel transport

$$\nabla_u e_i = 0$$



Using these basis of vectors, which are parallel (constant) along the curve write

$$\mathcal{N} = n^i e_i$$

zero by construction

Thus

$$\nabla_u \mathcal{N} = \nabla_u(n^i) e_i + n^i \nabla_u e_i$$

$$= \frac{dn^i}{d\lambda} e_i$$

$\mathcal{N}$ is not a parallel vector field in general.

and the equation of geodesic deviation becomes

$$\frac{d^2 n^i}{d\lambda^2} e_i + R(n^j e_j, u) u = 0$$

$e_i R^i{}_{kjl} u^k n^j u^l$

or

$$e_i \left[\frac{d^2 n^i}{d\lambda^2} + R^i{}_{uju} n^j = 0 \right]$$

relative to a basis field of $\parallel$ vectors

In the absence curvature we see that the deviation vector depends linearly on λ,

$$n^i = a^i \lambda + b^i$$

Evidently the curvature term in the equation measures precisely the extent to which this linear deviation law for

$R(u,v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u,v]} w$

$R(e_\rho e_\sigma) e_\nu = (e_\rho e_\sigma - e_\sigma e_\rho) e_\nu$

$R^\mu{}_{\nu\rho\sigma} e_\mu = (\Delta_\rho \Delta_\sigma - \Delta_\sigma \Delta_\rho) e_\nu$

space-time in a falling elevator

time

space

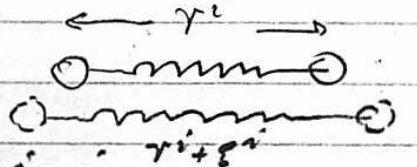
time

space

straight lines fails.
This term evidently measures the curvature of nearby (small n-16.1) geodesics relative to the one taken as a base line (n=16.1.

In Newtonian theory $\lambda \propto$ time and the equation yield the relative acceleration between adjacent freely orbiting particle.
One therefore infers from this fact that curvature is an expression of gravitational field gradient. mathematizes the

Consider



$$\eta^i = r^i + \xi^i$$

with $\frac{dr^i}{d\lambda} = 0 \quad \forall \lambda$.

Let $f^i = -k^i_j \xi^j$ be the restoring force

Then

$$\frac{d^2\xi^i}{d\lambda^2} + R^i_{uju}(r^j \xi^j) = -\frac{1}{m} k^i_j \xi^j$$

so that

$$\frac{d^2\xi^i}{d\lambda^2} + \left(\frac{1}{m} k^i_j + R^i_{uju}\right)\xi^j = -R^i_{ujv} r^j$$

is the eq'n governing the displacement of ξ^i

5. Metric Tensor Field.

(i) Parallel transport establishes a correspondence between vectors in the vector space at one point and those in the vector space at another point. A pair of corresponding vectors are decreed to be parallel by this law of parallel transport. $\nabla_u V = 0$ $\vec{V}(t=\Delta t) \dashrightarrow \vec{V}(t=0)$

Parallelograms, but not angles or lengths

(ii) Parallel transport makes no statements about angles between vectors or their length. Such statements are in the purview of a different geometrical structure, namely that of a metric (tensor) on a vector space, or more generally, a metric tensor field, which is an assignment of a metric tensor, say

$$g \equiv \text{"}\cdot\text{"} \equiv ds^2 = g_{ij}\, \omega^i \otimes \omega^j$$

to each point P of the manifold. Here

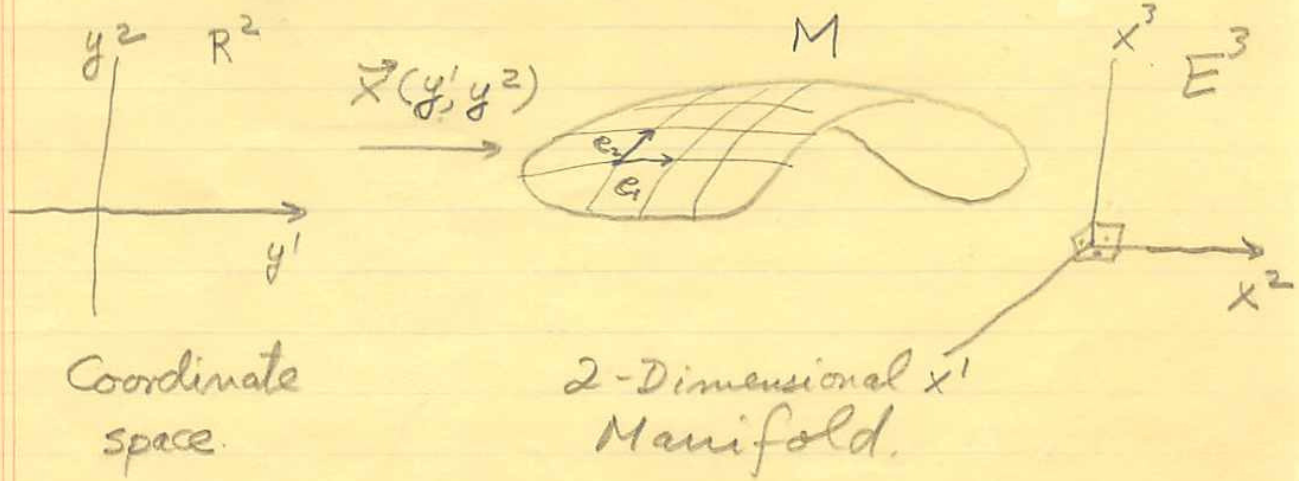
$$g_{ij} = e_i \cdot e_j$$

are the inner products of the basis vectors for V_P.

Let us consider two examples, one from geometry, the other from continuum mechanics.

First Example.
(Metric induced on a 2-dimensional surface by the geometry of the ambient Euclidean space)

In Euclidean 3 dimensional space consider a surface parametrized by the coordinates y^1 and y^2:



Coordinate space.

2-Dimensional Manifold.

~~28.3~~
39,3

The tangent space V_p at each point of M is spanned by the coordinate basis

$$\left\{ e_1 = \frac{\partial}{\partial y^1} \, , \, e_2 = \frac{\partial}{\partial y^2} \right\}$$

$$= \left\{ \frac{\partial x^1}{\partial y^1}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial y^1}\frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial y^1}\frac{\partial}{\partial x^3} \, , \, \frac{\partial x^1}{\partial y^2}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial y^2}\frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial y^2}\frac{\partial}{\partial x^3} \right\}$$

e_1 : $\rightarrow$ The partial derivatives

e_i : $\left\{ \frac{\partial x^1}{\partial y^i} \, , \, \frac{\partial x^2}{\partial y^i} \, , \, \frac{\partial x^3}{\partial y^i} \right\}_{i=1}^{i=2} \longleftrightarrow \frac{\partial \vec{x}}{\partial y^i} \qquad i = 1,2$

are the components of the vectors e_1 and e_2 relative to the orthonormal coordinates of the ambient three dimensional Euclidean space. This Euclidean space quite naturally induces the following inner product on the tangent space V_p at the point P of M. One has

Metric induced covariant derivative (parallel transport) ~~28.4~~ 39.4

$$e_i \cdot e_j = g_{ij} = \left(= \frac{\partial}{\partial y^i} \cdot \frac{\partial}{\partial y^j} \right) = \sum_{m=1}^{3} \frac{\partial x^m}{\partial y^i} \frac{\partial x^m}{\partial y^j} \left(= \frac{\partial \vec{x}}{\partial y^i} \cdot \frac{\partial \vec{x}}{\partial y^j} \right)$$

Thus the metric tensor on this imbedded manifold is

$$g = \frac{\partial \vec{x}}{\partial y^i} \cdot \frac{\partial \vec{x}}{\partial y^j} \; dy^i \otimes dy^j = g_{ij} \, dy^i \otimes dy^j$$

Thus the inner product on E^3 induces a natural metric tensor field on the imbedded submanifold M indeed.

end of Lecture 39.

LECTURE 40

Metric mathematization of elastic media.

I. Strain displacement

II. The strained state

III. Strain tensor = deformed metric.

An elastic medium is a continuum generalization of a system of springs.

Consider a bowl of jello with some raisens and grapes imbedded in it.

The purpose is to mathematize the medium in a state of deformation away from equilibrium. Do this in three steps.

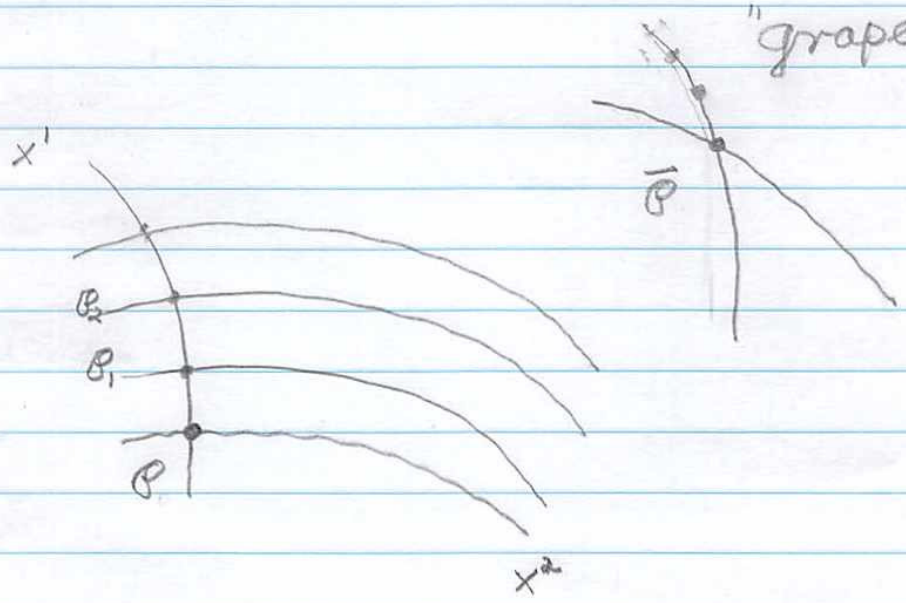
I. Conceptualize the strain displacement vector

II. Mathematize the strained state.

III. Geometrize the altered state in terms of the strain tensor, a deformation of the metric.

I. The Strain Displacement Vector Field.

First consider the elastic medium in its unconstrained state, and coordinatize the undisplaced raisins and grapes in terms of the lab coordinates, i.e. those of the bowl.

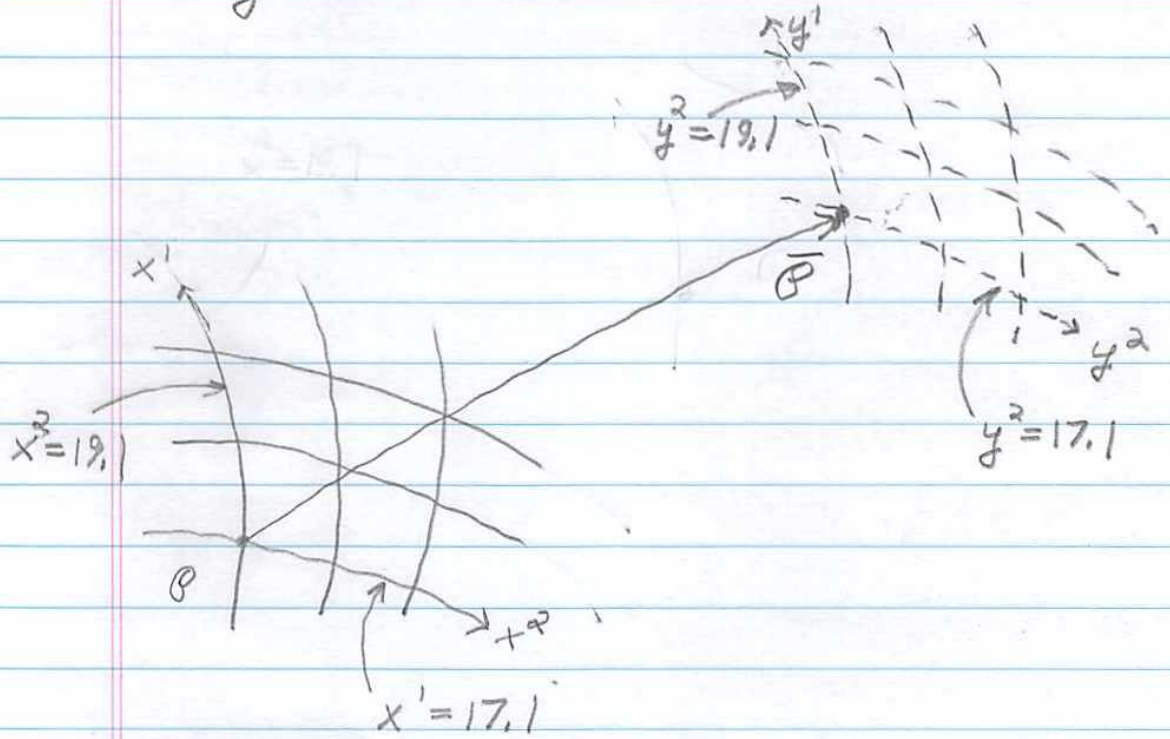


Undisplaced raisins and grapes.

The lab coordinates assigned to a typical raisin are

$$\{x^1(P), x^2(P), x^3(P)\}$$

The strain displacement vector field moves the raisin at P and its neighbors to the point $\bar{P}$ and its neighborhood. The grapes get moved away from their original location



(a) (b)

(a) Unconstrained coord. system before deformation
(b) Strained coord. system after deformation

The undisplaced raisins are connected to the displaced raisins by means of the displacement

$$\delta \mathbf{p} = \bar{\mathbf{p}} - \mathbf{p}.$$

In the small strain (i.e. linear) approximation, this displacement is a vector:

$$\boxed{\delta \mathbf{p} = \text{"displacement vector"}}$$

Q: What are the coordinate components of this displacement vector?

Remark:

The effect of the deformation is three-fold:

(relative to the lab)

(i) It changes the location of each atom/raisin in the elastic medium and

(ii) It changes the curvilinear coordinate system used to map out the atoms/raisins.

(iii) It changes the distance between neighboring atoms/raisins.

The deformation, say F, is a map.

From the medium onto itself

$$F: \quad M \longrightarrow M$$
$$P \rightsquigarrow \bar{P} = F(P)$$

Consider the curvilinear lab coordinate system $x^i(P)$ determined by the location of the atoms/raisons of the medium <u>before</u>, as compared to the <u>new</u> $y^i(\bar{P})$ obtained from the old coordinates

P	$\bar{P}$ ← displaced point object
$x^i(P)$	$y^i(\bar{P}) = x^i(P)$ ← its coordinates
given	defines $y^i(P)$
old (lab) coord. system	new (lab) coord. system

Definition

Assign old coordinate at the old point to the new coordinate at the new point

$$\boxed{y^{i'}(\bar{P}) = x^{i'}(P)}$$

II. The Strain Vector

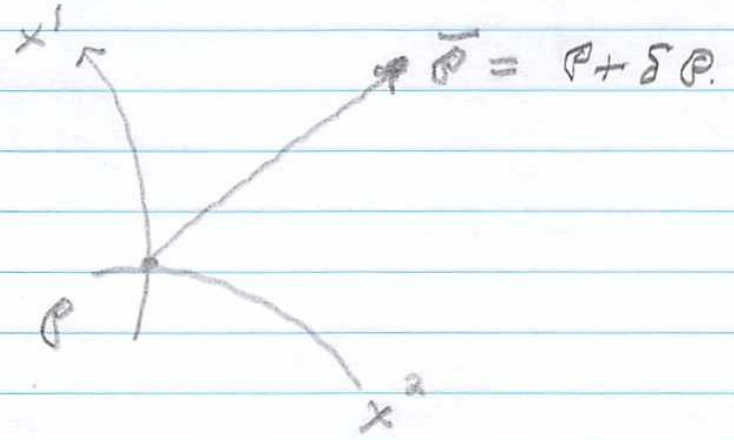
The coord. values of the old coord fn $x^{i'}$ are related to those of the new coord fn $y^{i'}$ by means of the infinitesimal (= Principal Linear Part = P.L.P.) strain displacement, a vector function:

Using

$$x^i(P) = y^i(\bar{P}),$$

one has

$$\xi^i = x^i(\bar{P}) - x^i(P) = x^i(\bar{P}) - y^i(\bar{P}) \quad (*)$$



$$x^{i'}(\bar{P}) = x^{i'}(P+\delta P)$$

$$= x^{i'}(P) + \xi^{i'}(P) + \text{higher order terms}$$

Drop the bar and obtain from Eq. (*)

$$\boxed{y^{i'}(P) = x^{i'}(P) - \xi^{i'}(P)}$$

Remark 1:

The boxed eq'n is an infinitesimal coordinate transformation

Remark 2:

$$\xi^{i'}(\bar{P}) = \xi^{i'}(P+\delta P) = \xi^{i'}(P) + \text{higher order terms}$$

i.e

$$\xi^{i'}(y^{i'}) = \xi^{i'}(x^{i'}) + \text{neglected terms}$$

Once the strain displacement $\{\xi^{i'}(x)\}$ is given, the new coord. system is determined.

STEP III: Strain tensor as the distortion in the metric tensor

The deformation $\mathcal{P} \xrightarrow{F} \bar{\mathcal{P}}$ establishes the map $V_{\mathcal{P}} \longrightarrow V_{\bar{\mathcal{P}}}$. It does this by mapping the old basis vectors at $\mathcal{P}$ to the new distorted basis at $\bar{\mathcal{P}}$. This distortion can be a shear, a volume change, or any combination of such changes in the basis vectors. The change in the inner product between any pair of distorted basis vectors expresses such a distortion. In fact,

$g_{old} = g_{ij}(x(\mathcal{P}))\, dx^i dx^j$ is the old inner product for $V_{\mathcal{P}}$ while the new, distorted inner product at $V_{\mathcal{P}}$ is

$$g_{ij}(y(\bar{\mathcal{P}}))\, dy^i \otimes dy^j = g_{ij}(x^k - \xi^k)\, d(x^i - \xi^i) \otimes d(x^j - \xi^j)$$

keep only the principal linear part:

$$= g_{ij}(x)\, dx^i dx^j - g_{ij,k}\, \xi^k\, dx^i dx^j$$

$$- g_{ij}\, \xi^i{}_{,k}\, dx^k dx^j - g_{ij}\, \xi^j{}_{,k}\, dx^i dx^k$$

$$\boxed{g_{NEW} = g_{ij}\, dx^i \otimes dx^j - \left(\frac{\partial g_{ij}}{\partial x^k} \xi^k + g_{kj} \frac{\partial \xi^k}{\partial x^i} + g_{ik} \frac{\partial \xi^k}{\partial x^j} \right) dx^i \otimes dx^j}$$

$$+ \text{terms of higher order in } \xi$$

$S_{ij}\, dx^i \otimes dx^j$ = strain tensor,

One sees that the distortion, (i.e perturbation,) in the metric, the strain tensor

$$S_{ij}\, dx^i \otimes dx^j = \left(g_{ij,k}\, \xi^k + g_{kj}\, \xi^k{}_{,i} + g_{ik}\, \xi^k{}_{,j} \right) dx^i \otimes dx^j$$

(a) is symmetric,

and $$S_{ij} = S_{ji},$$

(b) relative to a linear (as compared to a curvilinear) coordinate system, for which $g_{ij} = \text{const.}$ so that $g_{ij,k} = 0$, one has

$$\boxed{S_{ij} = \xi_{j,i} + \xi_{i,j}}$$

where we "lowered" the index of ξ^k by means of

$$g_{kj}\, \xi^k{}_{,i} = (g_{kj}\, \xi^k)_{,i} \equiv \xi_{j,i}$$

STEP IV. Strain tensor relative to curvilinear coordinates

~~28.13~~ 40.12

Relative to curvilinear coordinates, in which case one has $\Gamma^k_{ei} \neq 0$, the strain tensor components are

$$\boxed{S_{ij} = \xi_{j;i} + \xi_{i;j}}$$

Indeed one has

$$S_{ij} = g_{ij,k}\,\xi^k + g_{jk}\,\xi^k_{,i} + g_{ik}\,\xi^k_{,j}$$

Add and subtract

$$g_{jk}\left(\Gamma^k_{ei} - \Gamma^k_{ei}\right)\xi^e \quad + \quad (j \leftrightarrow i)$$

Thus one obtains

$$S_{ij} = g_{ij,k}\,\xi^k + g_{jk}\,\xi^k_{;i} + g_{ik}\,\xi^k_{;j}$$

$$- \underbrace{g_{jk}\,\tfrac{1}{2}\,g^{km}}_{\tfrac{1}{2}\,\delta_j^{\,m}}\left(g_{me,i} + g_{mi,e} - g_{ei,m}\right)\xi^e$$

$$- \underbrace{g_{ik}\,\tfrac{1}{2}\,g^{km}}_{\tfrac{1}{2}\,\delta_i^{\,m}}\left(g_{me,j} + g_{mj,e} - g_{ej,m}\right)\xi^e$$

Performing the sums over "m" one obtains.

~~28.14~~
40.13

$$S_{ij} = g_{ij,k}\xi^k + g_{jk}\xi^k{}_{;i} + g_{ik}\xi^k{}_{;j}$$

$$-\frac{1}{2}\left(g_{je,i} + g_{ji,e} - g_{ei,j}\right)\xi^e$$

$$-\frac{1}{2}\left(g_{ie,j} + g_{ij,e} - g_{ej,i}\right)\xi^e$$

cancel, cancel, cancel

Thus we are left with

$$\boxed{S_{ij} = g_{jk}\xi^k{}_{;i} + g_{ik}\xi^k{}_{;j}}$$

Using the fact that

$$\left(g_{jk}\xi^k\right)_{;i} = g_{jk,i}\xi^k + g_{jk}\xi^k{}_{;i}$$

$\xi_{j;i}$ ← product rule; $g_{jk,i}$ = 0 "metric compatibility"

One find that the component of the strain tensor are

$$\boxed{S_{ij} = \xi_{j;i} + \xi_{i;j}}$$

"Strain tensor components".

LECTURE 41

Metric-induced parallel transport

1. The Metric compatibility condition
2. Metric compatibility via extremal paths

Metric-induced Parallel transport

A metric tensor field assigns an inner product

$$g(\vec{u},\vec{v}) = g_{ij}\,\omega^i \otimes \omega^j(\vec{u},\vec{v}) \doteq g_{ij}\,u^i v^j = \vec{u}\cdot\vec{v}$$

to each vector space V_p of a manifold.

The metric tensor field induces a law of parallel transport of vectors along a given curve: at an initial point on the curve start with a vector that has a specific length and makes a specific angle with the tangent to the curve. Moving At an adjacent point on the curve draw a vector having the same length

and making the same angle with the curve tangent at that point. Applying this process to every element of the vector space at the starting establishes an isometric isomorphism between the two adjacent vector spaces.

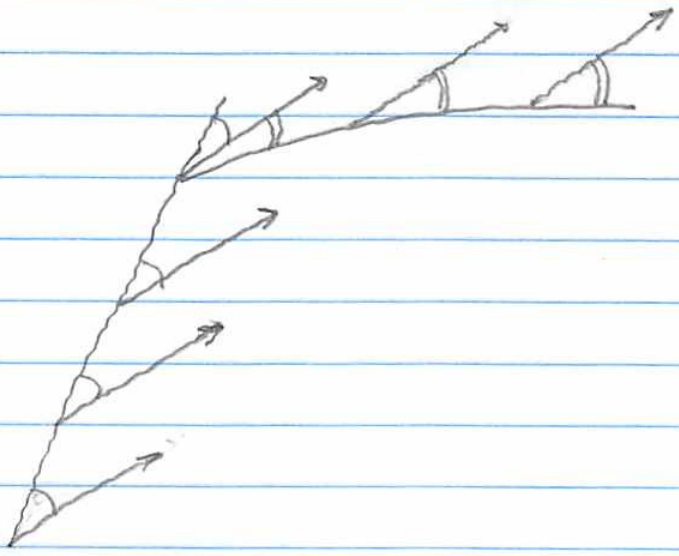


Figure 1: Vectors parallel in compliance with the metrically induced law of parallelism.

Proposition:

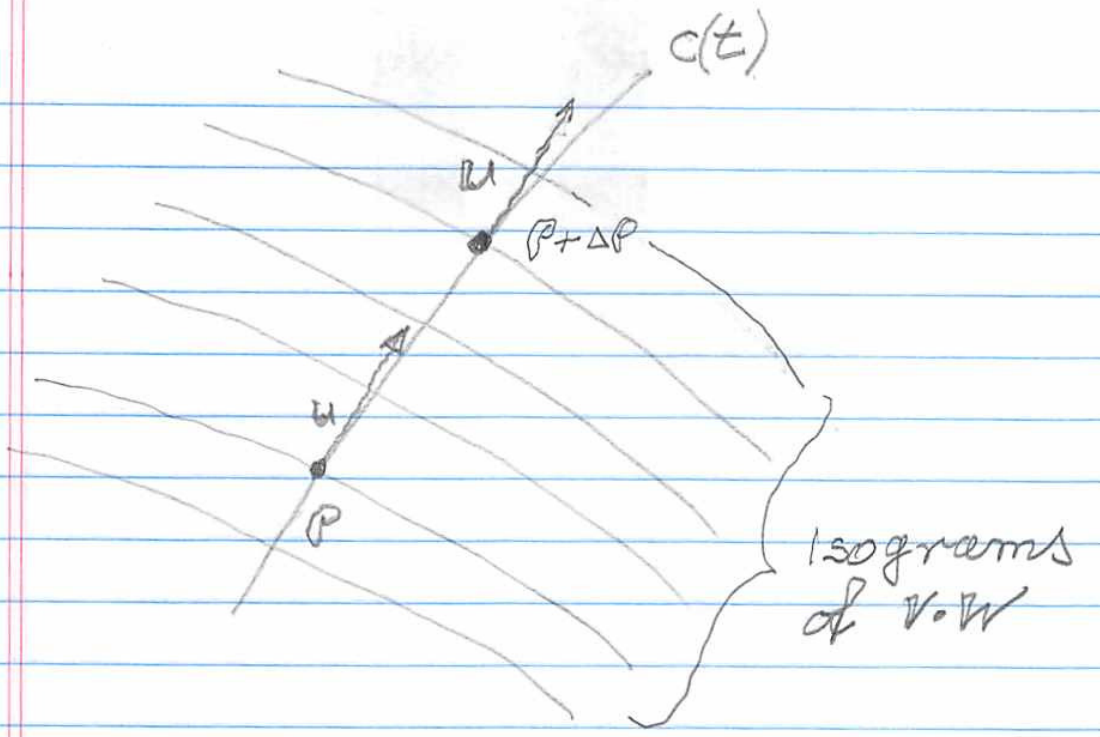
A metric tensor field determines a unique torsionless law of parallel transport.

The validity of this proposition is ascertained in 3 steps.

Step 1: Motion-induced change in the Inner Product.

Consider two arbitrary differentiable vector fields V and W as well as a curve whose tangents belong to a chosen vector field u

The inner product $g(V, W) = V \cdot W$ in each tangent space is a scaler field whose isograms intersect the chosen curve



igure 2: Isograms of the scalar field V·W.

In moving from point P to point P+ΔP on the curve, the change in this scalar is

$$V\cdot W\Big|_{P+\Delta P} - V\cdot W\Big|_{P} = \Delta(V\cdot W) \qquad (1)$$

Focusing on two adjacent points P and P+ΔP on the curve c(t), note

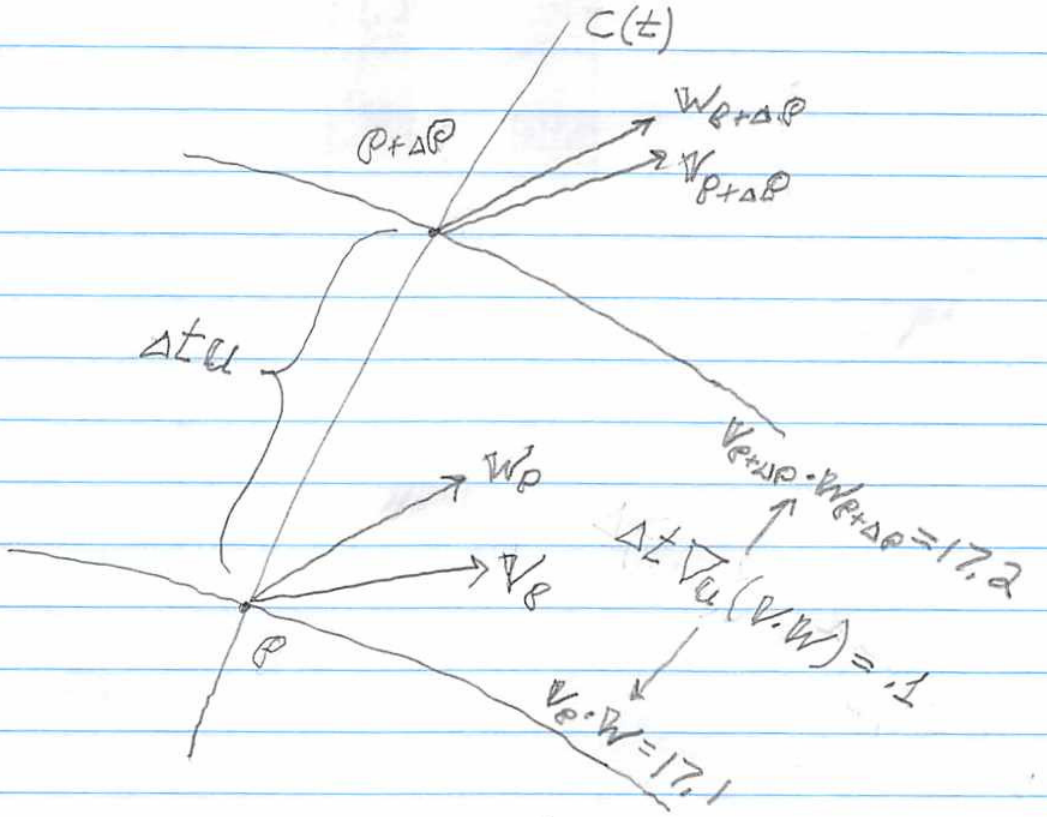


Figure 3: Difference between isogram values at adjacent points on the chosen curve.

that this difference is

$$\Delta(V \cdot W) = \Delta t U(V \cdot W) + \begin{matrix}\text{term of} \\ \text{higher} \\ \text{order in } \Delta t\end{matrix}$$

Retain the Principal Liner Part of this difference, and express in term the familia scalar version of the directional derivative, ∇_U

$$\Delta(V \cdot W) = \Delta t U(V \cdot W) = \Delta t \nabla_U (V \cdot W).$$

Step 2: The Metric Compatibility Condition

Deduce the to-be-formulated law of parallelism from the condition of compatibility

$$\boxed{\nabla_u(V\cdot W) = (\nabla_u V)\cdot W + V\cdot(\nabla_u W)}$$

between the inner product structure expressed by the given metric tensor $g = g_{ij}\,\omega^i\otimes\omega^j$ and parallelism expressed by the covariant derivatives $\nabla_u V$ and $\nabla_u W$. The above equation for metric compatible parallelism is a product of inductive reasoning. Consign it to Step 3.

The deduction leading to $\nabla_U V$ and $\nabla_U W$ consists of applying the compatibility condition to the basis vector fields $V = e_i$, $W = e_j$, and $U = e_k$. The result is

$$\nabla_{e_k}(e_i \cdot e_j) = (\nabla_{e_k} e_i) \cdot e_j + e_i \cdot (\nabla_{e_k} e_j)$$

or $$\frac{\partial g_{ij}}{\partial x^k} = e_\ell \Gamma^\ell_{ik} \cdot e_j + e_i \cdot e_\ell \Gamma^\ell_{jk}$$

For typographic efficiency introduce

$$g_{i\ell} \Gamma^\ell_{ik} = \Gamma_{\ell\, jk} \quad \left(\text{"christoffel symbol of the first kind"}\right)$$

Using these, obtain

$$g_{i\ell,k} = \Gamma_{j\, ik} + \Gamma_{i\, jk}. \qquad \left.\begin{matrix} i \\ j \\ k \end{matrix}\right\} = 1, \ldots, n$$

Solve these equations by adding unto the negative of this equation the other two with their indeces cyclically permuted:

$$-\left(g_{ij,k} = \Gamma^{1}_{ji,k} + \Gamma^{2}_{ij,k}\right) \qquad (1)$$

$$+\left(g_{jk,i} = \Gamma^{3}_{kj,i} + \Gamma^{4}_{jk,i}\right) \qquad (2)$$

$$+\left(g_{ki,j} = \Gamma^{5}_{ik,j} + \Gamma^{6}_{ki,j}\right) \qquad (3)$$

This system does not have a unique solution for the Γ's. However consider the type of parallel transport whose <u>torsion is zero</u>!

$$\nabla_{e_k} e_j - \nabla_{e_j} e_k - [e_k, e_j] = 0$$

Relative to a coordinate induced basis

$$[e_k, e_j] = 0.$$

Consequently,

$$e_\ell \Gamma^{\ell}_{jk} = e_\ell \Gamma^{\ell}_{kj}$$

which implies

$$\Gamma_{i,jk} = \Gamma_{i,kj},$$

i.e the Christoffel Γ's are symmetric in the last two indeces.

In light of this symmetry the sum of the r.h.s of the three Eqs (1)-(3) reduces to twice a single Γ,

$$g_{kl,ij} + g_{kj,il} - g_{ij,kl} = 2\,\Gamma_{k,ij}.$$

Thus
$$\boxed{\Gamma^{l}_{ij} = \tfrac{1}{2}\, g^{lk}\left(g_{ki,j} + g_{kj,i} - g_{ij,k}\right)} \qquad (4)$$

These are the unique metric determined Christoffel symbols.

The vector field $Z(c(t)) = e_j(c(t))\, z^j(t) \equiv z(t)$ to be parallel along the given curve $\{c^k(t)\}$, whose tangents are $u(t) = \dot{c}^k(t)\, e_k(c(t))$, when $Z(c(t))$ satisfies the system of linear diff.'l eq'ns

$$0 = \nabla_u Z = e_j\left(\left.\frac{\partial Z^j(x)}{\partial x^k}\right|_{x=c(t)} + z^i(t)\,\Gamma^j_{ik}(c(t))\right)\frac{dc^k}{dt}$$

$$= e_j\left[\frac{d\,z^j(t)}{dt} + z^i(t)\,\Gamma^j_{ik}(c(t))\frac{dc^k}{dt}\right]$$

41.10

Step 3: Parallelism and the Covariant Derivative

The Directional Derivative

The familiar directional derivative is the mathematization of change as it applies to scalar functions on a manifold. It starts with the value of the scalar field at a reference point of interest. The focus then shifts to the value of the scalar at an adjacent point separated by the displacement vector connecting the reference (1^{st}) point to the adjacent (2^{nd}) point. The difference between the values of the scalar field at these two vector-separated points is the quantitative change that goes into the process of defining the directional derivative at the chosen reference point.

The Covariant Derivative

The covariant derivative follows the same process. Start with the value of the vector field at a reference point of interest. That value is a vector in the tangent space at that point. Then shift attention to the value of the vector field in the tangent space at the adjacent point, the one separated from the reference point by the displacement vector between the two.

This leads to the question: What is the change in the values of the vector field in moving from the first tangent space to the second adjacent? This is a mathematically illegitimate question because the difference between vectors in different vector spaces is undefined.

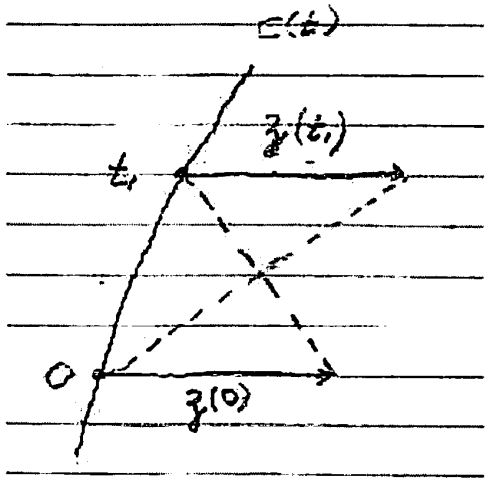


Figure 1: Schild-Ladder-induced isomorphic alignment of vector $\mathbf{z}(t_1) \in V_{c(t_1)}$ with $\mathbf{z}(o) \in V_{c(0)}$.

SKIP

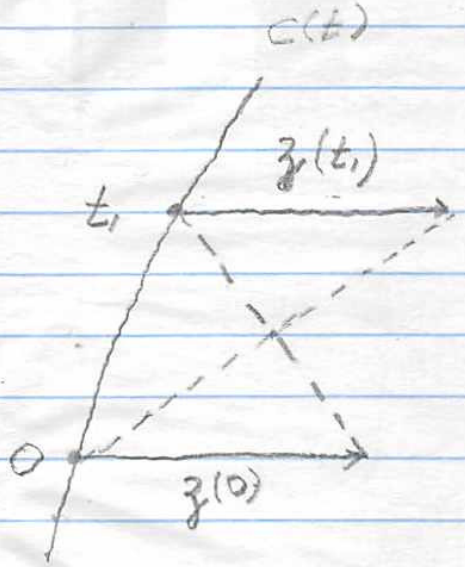


Figure 1: Isomorphic alignment via the Schild ladder construction. Vector $z(t_1) \in V_{c(t_1)}$ is isomorphically aligned with vector $z(0) \in V_{c(0)}$.

41, 11

However, it *is* well defined when there exists a ***parallelism*** between tangent spaces. Parallelism is a joint property of two tangent spaces which are separated by a vectorial displacement: every vector in one tangent space is observed to be ***isomorphically aligned*** with a unique vector in the adjacent tangent space.

Because of this parallelism, a vector in the tangent space at the reference point can be moved to its adjacent neighbor, and can therefore be compared by subtraction with the preexisting value of the vector field. The resulting difference is the quantitative change that goes into the process of defining the directional covariant derivative at the chosen reference point.

This covariant derivative extends the directional derivative applied to scalar field to that applied to vector fields. It mathematizes the extent by which a vector field is not constant.

Parallel Vector Field

An everywhere unchanging, i.e. parallel, vector field, say $\mathbf{Z}(x)$ satisfies the diffential equations

$$0 = \nabla_{\mathbf{u}}\mathbf{Z} = \left\{ \frac{\partial Z^i(x)}{\partial x^k} u^k(x) + Z^\ell(x)\Gamma^i{}_{\ell k}(x)u^k(x) \right\} \frac{\partial}{\partial x^i} \ . \tag{1}$$

Except in a Euclidean space, or a manifold with a similar parallelism structure, there does not exist a non-zero solution to this system of partial differential equations.

However, there *does* exist a solution to this system if its domain is restristed to a curve $\{c(t) : \{c^k(t)\}_{k=1}^n\}$. Its tangents

$$\frac{dc^k(t)}{dt}\frac{\partial}{\partial x^k} \equiv \dot{c}(t)$$

form a vector field on this one-dimensional domain, namely

$$\mathbf{u}(c(t)) = u^k(c(t))\frac{\partial}{\partial x^k} \equiv \dot{c}(t) \ .$$

The existence of a vector field parallel along this curve is guaranteed by the existence of a solution to Eq.(1) with its domain restricted to $c(t)$, namely

$$\begin{aligned} 0 = \nabla_{\mathbf{u}}\mathbf{Z}|_{c(t)} &= \left\{ \left.\frac{\partial Z^i(x)}{\partial x^k}\right|_{c(t)} \frac{dc^k(t)}{dt} + Z^\ell(c(t))\Gamma^i{}_{\ell k}(c(t))\frac{dc^k(t)}{dt} \right\} \frac{\partial}{\partial x^i} \\ &= \left\{ \frac{d\mathfrak{z}^i(t)}{dt} + \mathfrak{z}^\ell(t)\Gamma^i{}_{\ell k}(c(t))\dot{c}^k(t) \right\} \frac{\partial}{\partial x^i} \end{aligned} \tag{2}$$

Given an initial vector

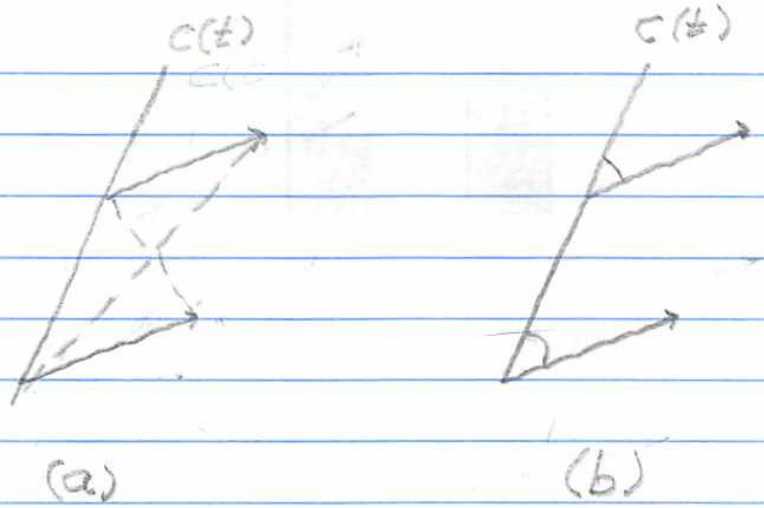
$$\mathbf{Z}|_{c(0)} \equiv \mathfrak{z}(0) = \mathfrak{z}^i(t=0)\frac{\partial}{\partial x^i}$$

on $c(t)$, this linear system of n o.d.e.'s has a unique solution for the parallel vector field

$$\mathbf{Z}|_{c(t)} \equiv \mathfrak{z}(t) = \mathfrak{z}^i(t)\frac{\partial}{\partial x^i} \tag{3}$$

Metric Compatibility

The fact that there are different types of parallelism is a reflection of the fact that there are different geometrical structures that a vector space can accommodate. Figure 41.2 exhibits this aspect of the nature of things where parallelism is a consequence of equal norms (bisecting diagonals for the Schild ladder parallelism) vs a consequence of equal inner products (equal angles and lengths for metric induced parallelism).



Parallelism via Schild Ladder (bisecting diagonals)

Parallelism via Metric (equal angles & lengths)

Figure 41.2: Different parallelisms via different geometric structures.

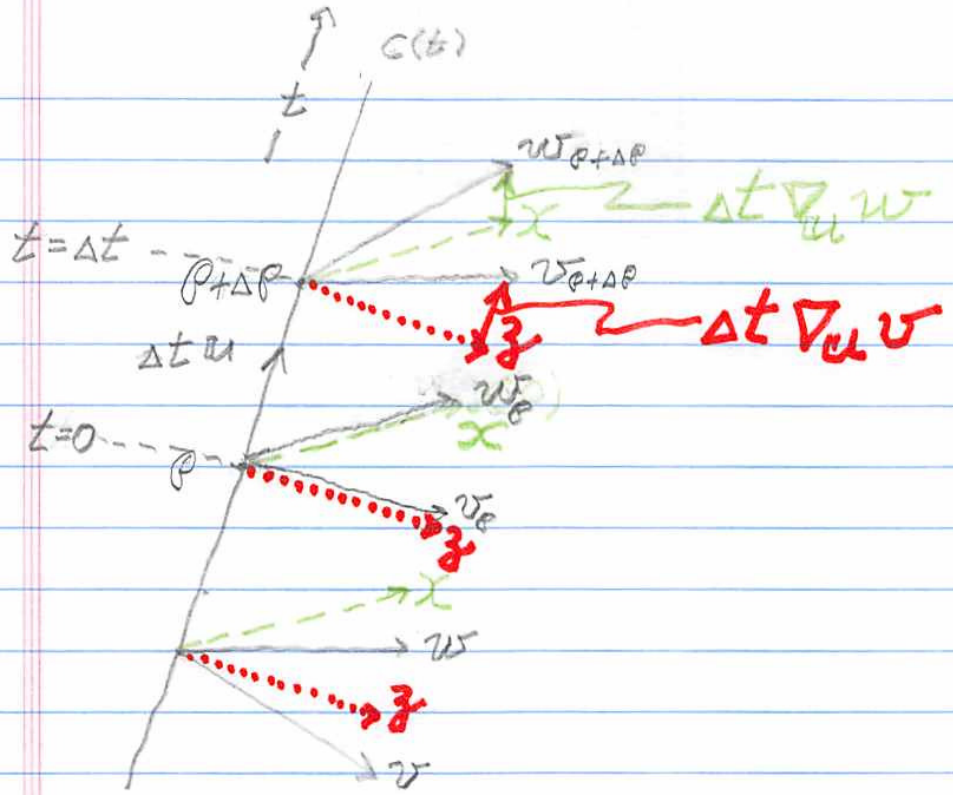


FIGURE 41.3

Along the curve $c(t)$ whose tangent is $u(t)$ there are (a) two preexisting (in general non-parallel) vector fields v and w,

(b) Two vector fields x and z parallel along the curve $c(t)$;

$$\nabla_u x = 0 \qquad x(0) = w_p$$

$$\nabla_u z = 0 \qquad z(0) = v_p$$

The metric compatibility condition is that the inner product of $x(t)$ and $z(t)$ be constant along the curve

$$\boxed{x(t)\cdot z(t) = x(0)\cdot z(0) = v_p \cdot w_p} \quad (2)$$

for all t.

Looking at Figure 41,3 on page 41.14, note that:

(1) The given preexting vectors at point p are v_p and w_p

(2) At point $p+\Delta p$ the preexisting vectors $v_{p+\Delta p}$ and $w_{p+\Delta p}$ differ from the parallel translates $x(\Delta t)$ and $z(\Delta t)$ by $\Delta t\, \nabla_u v$ and $\Delta t\, \nabla_u w$

$$v_{p+\Delta p} = x(\Delta t) + \Delta t\, \nabla_u v$$
$$w_{p+\Delta p} = z(\Delta t) + \Delta t\, \nabla_u w$$

The difference in their inner product, Eq.(1) on page 41.4 is therefore

(3) $\left(x(\Delta t)+\Delta t\,\nabla_u v\right)\cdot\left(z(\Delta t)+\Delta t\,\nabla_u w\right)-v_\rho\cdot w_\rho=\Delta(v\cdot w)$

(3) (i) Inserting the compatibility condxn Eq(2) on page 41.15, namely

$$x(\Delta t)\cdot z(\Delta t)=v_\rho\cdot w_\rho$$

into the Principal Linear Part of Eq.(3) at the top of this page 41.16,

(ii) noting that

$$\Delta(v\cdot w)=\Delta t\,\nabla_u(v\cdot w),$$

(iii) noting that

$$x(\Delta t)=v_{\rho+\Delta\rho}-\Delta t\,\nabla_u v$$

$$z(\Delta t)=w_{\rho+\Delta\rho}-\Delta t\,\nabla_u w$$

(iv) neglecting non-linear terms

one obtains

$$\Delta t(\nabla_u v)\cdot w + \Delta t\, v\cdot \nabla_u w + \begin{pmatrix}\text{negligible}\\ \text{higher}\\ \text{order}\\ \text{terms}\end{pmatrix} = \Delta t\, \nabla_u (v\cdot w)$$

Consequently

$$\boxed{\nabla_u (v\cdot w) = (\nabla_u v)\cdot w + v\cdot(\nabla_u w)}\,.$$

This is the <u>metric compatibility condxn.</u> as expressed by the covariant derivative in STEP 2 on page 41.6.

Metric Compatibility via Extremal Paths

An alternative and direct way of obtaining the metric compatible law of torsionless parallel transport is via the extremal paths, the geodesics, determined by the metric. Recall the extremal paths are obtained from the variational principle

$$\int_{\lambda_1}^{\lambda_2} \sqrt{g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}}\, d\lambda = \text{extr.}$$

The E-L eq'ns for the variational integral read

$$\frac{d^2x^i}{ds^2} + \frac{dx^j}{ds} \Gamma^i_{jk} \frac{dx^k}{ds} = 0 \qquad (5)$$

These Christoffel symbols are furnished directly by the variational principle.

The fact that

$$\Gamma^{\ell}_{ij} = \tfrac{1}{2} g^{\ell k}\left(g_{kj,i} + g_{ki,j} - g_{ij,k}\right)$$

is guaranteed by the E-L eq'ns

Indeed those Γ's which are non-zero are furnished directly by the variational principle. The ones which are zero don't even appear and thus do not have to be computed explicitly. This was the gist of the "geodesic Lagrangian method" (Homework 3)

The geometric meaning of Eq.(5) on page 41.11 is quite perspicuous: Let the tangent to each geodesic be

$$\frac{dx^k}{ds}\frac{\partial}{\partial x^k} = u^k \frac{\partial}{\partial x^k} = U,$$

Then in terms of this tangent the eq'n reads (with the help of the chain rule of calculus)

$$e_\ell \left(\frac{\partial u^\ell}{\partial x^k} + u^j \Gamma^\ell_{jk} \right) u^k = 0$$

or $$\boxed{\nabla_u u = 0}$$

This expresses the fact that the tangents of the extremal curves are autoparallel.

Furthermore, using the metric compatibility condition

$$0 = (\nabla_u u) \cdot u + u \cdot \nabla_u u = \nabla_u (u \cdot u).$$

This means that the magnitude of $u \cdot u$ of the auto-parallel tangent is a constant along the geodesic.