

Todo list

Prove the theorem	25
Study notes on Riemannian Geometry	45
Prove the Properties of Open Subsets of a Metric Space (Lee, p.397)	47
Prove the Properties of Closed Subsets of a Metric Space (Lee, p.397)	47
Finish the proof	47
Figure: Draw rectangular box	71



Contents

I Special Relativity

9

Chapter 1

Einstein's Original Paper on "Special Relativity"

1.0.1 Reading Notes...	9
§1. Definition of Simultaneity	10
§2. Something Stops Working...	11
§3. Lorentz Transformation - Quantifying Non-Simultaneity	13

II A First Course on General Relativity

21

Chapter 2

Special Relativity

2.1 Fundamental principles of special relativity theory (SR)	21
2.1.1 On "Principle of relativity (Galileo)"	21
Galilean invariance	21
2.2 Construction of the coordinates used by another observer	21
2.2.1 Why would the tangent of the angle is the speed in Fig. 1.2?	21
2.3 Invariance of the interval	22

III Mathematics

29

Chapter 3

Calculus

3.1 Calculus	29
3.1.1 The Definition of the Derivative	29
3.1.2 Continuous Function	29

31

	Chapter 4	
	Calculus of Variation	
	4.1 Calculus of Variation 31	
	4.1.1 Origin of Calculus of Variations 31	
	The Brachistochrone Challenge	31
	The Famous Solutions	31
	Formalization by Euler and Lagrange	32
33	Chapter 5	
	Topology	
	5.1 Metric Spaces 33	
	5.2 Indefinite Inner Products 37	
39	Chapter 6	
	Linear Algebra	
	6.1 Linear Algebra 39	
	6.1.1 Eigenvalues and Eigenvectors 40	
	Matrix Diagonalization	40
45	Chapter 7	
	Riemannian Geometry	
	7.1 What are Manifolds? 45	
	7.1.1 Intuitive Meaning 45	
	7.1.2 Formal Definition of Manifolds 46	
	IV Electromagnetism	
53	Chapter 8	
	Mathematics	
	8.1 Differential Calculus 53	
	8.1.1 Gradient 55	
	8.1.2 Divergence 56	
	The Meaning of Divergence	56
	8.1.3 Curl 57	
	The Cross Product of Two Vectors	57
	Curl	60
	8.1.4 Line Integrals 61	
	The Fundamental Theorem of Calculus	62
	Independence of Path & Conservative Fields	68
	8.1.5 Green's Theorem 71	
73	Chapter 9	
	Maxwell's Equations	
	9.1 Gauss's Law for Electric Fields 73	
	9.1.1 The Electric Field 73	
75	Chapter 10	
	Euclid's Elements	

V Quantum Mechanics

79Chapter 11
Introduction

11.1	The Schrödinger Equation	79
11.1.1	Statistical Interpretation	79
11.2	Probability	79
11.2.1	Normalization	80
11.2.2	Momentum	82
11.3	Formalism	82

Part I

Special Relativity

Chapter 1

Einstein's Original Paper on "Special Relativity"



The source of this chapter is the [original German version of Einstein's paper on Special Relativity](#)



It is not recommended to read English translations because many translators, thinking they were smart enough, unethically tempered the original version dramatically, and thus disguised the readers^a

^asource

Reading the original paper requires the prerequisites of



Michelson-Morley Experiment

- [An excellent experiment intro](#)
- What I care most about this experiment is the **way we handle "unsolvable" problems**. Michelson-Morley experiment had led to extensive followups trying to explain what was seen in the experiment. All the *mediocre* conclusion simply said: "Dude, we don't know." Albert Einstein innovated a new era of Physics out of this conflict. **When a problem seems to lead to a dead end, it's time to innovate; it's time to take on the risk and bring the human into a new world of new opportunities!**



[Maxwell's Electrodynamics](#)

1.0.1 Reading Notes...

...of [the Paper](#)

§1. Definition of Simultaneity

Definition 1: Simultaneity

If an event occurs at (t, x, y, z) , then all observers would see this event at (t, x', y', z') , where $x \neq x'$, $y \neq y'$, and $z \neq z'$

For example, If I say that “a train arrives here at 7 o’clock,” that means when a clock on that train ticks to 7, my hand watch also points at 7 sharp. There is no “delay” caused by information propagation between the train clock and my hand watch; they just happen to “synchronize” that way. We say the 7 on the hand watch and the 7 on the train are **simultaneous events**

This definition is ideal when the clock and the attached event are located at the same place because we can read the clock precisely at that location. But what if there are series of events at other locations and we need to link them temporally? Can we simply tell which of the two events happened first simply by comparing the time we see the incident light of the two events? No, because, as Einstein said, “such an assignment has the drawback that it is not independent of the position of the observer”. For example, suppose you are reading this PDF document at 9 am in the morning; 20 minutes later your special apparatus on your desk detected the light of a collision of two stars in universe that happened 2 billion light years away. In this case, *can we say you saw this PDF file earlier than the star collision?*

What is missing here is the ordering of multiple events. In other words, if we could somehow “synchronize” the clock located at us, the PDF file, and the star collision, we will be able to say, by looking at all 3 clocks synchronized, the star collision happened first.

Einstein offered an approach to synchronize the clocks in the following way. He got inspired by Differential Geometry by realizing that the clock C_1 of the event located at the *vicinity* of another clock C_2 is infinitesimally synchronized with clock C_2 . If there is a clock at point A of space, then an observer located at A can evaluate the time of the events in the immediate vicinity of A by simply look at their hand watch. If there is also a clock at point B , then the time of the events in the immediate vicinity of B can likewise be evaluated by an observer located at B . But it is not possible to compare the time of an event at A with one at B without a further stipulation; thus far we have only defined an “ A -time” and a “ B -time” but not a “time” common to A and B . The latter can now be determined by establishing by definition that the “time” needed for the light to travel from A to B is equal to the “time” it needs to travel from B to A . Thus we have the following definition

Definition 2: Synchronism (Location-Independent)

Suppose a ray of light leaves from A toward B at “ A -time” at t_A , is reflected from B toward A at “ B -time” t_B , and arrives back at A at “ A -time” t'_A . The two clocks are synchronous by definition if

$$t_B - t_A = t'_A - t_B \quad (1.1)$$

It follows naturally that

1. if the clock in B is synchronous with the clock in A , then the clock in A is synchronous with the clock in B , and
2. if the clock in A is synchronous with the clock in B as well as with the clock in C , then the clocks in B and C are also synchronous relative to each other

This is a very powerful assertion, because it states that for 2 clocks to be “synchronous”, not only do they need to be ticking at the same rate, but also they must always be pointing at the exact same absolute instance of time, i.e. both must be both ticking at 9:00:00 am sharp, not one at 9:00:00 and the other at 9:00:03. Another example is that two clocks at Chicago and New York City are never synchronous.

Applying this Def. 2, we are able to solve the PDF-collision problem. Since the PDF and us are in pretty much vicinity of each other, we can approximate the two to be synchronous with each other. To synchronize with the star collision, our watch will need to “move backwards” in time by little more than 2 billion years. According to the transitivity law of Synchronism, the clocks of PDF and the collision are now synchronized. The ordering of the events should be pretty clear now.

The **speed of light as a universal constant in empty space is thus:**

$$V = \frac{2\overline{AB}}{t'_A - t_A} \quad (1.2)$$

! and there is a BIG assumption: all clocks are at *rest* in a system at *rest*

With that, we have a pretty good mechanism to talk about series of events in a system happening at different time, because we know how to synchronize them 

§2. Something Stops Working...

Principle of Relativity

The laws governing the changes of the state of any physical system do not depend on which one of two coordinate systems in uniform translational motion relative to each other these changes of the state are referred to

Principle of the Constancy of the Velocity of Light

Each ray of light moves in the coordinate system “at rest” with the definite velocity V independent of whether this ray of light is emitted by a body at rest or a body in motion

2 principles above along with the Def.2 shall present us an surprising result that a moving rod with a stationary length r_{AB} will be measured to have a different length measured by an moving observer

Suppose the rod is moving in the x-direction at a constant speed v . Let the length of the moving rod, measured in the system at rest, be denoted as r_{AB} , where A and B are the two ends of the rod. In addition, we imagine that the two ends (A and B) of the rod are equipped with clocks (C_1 and C_2) that are *synchronous* with the clocks of the system at rest. Hence, C_1 and C_2 are **synchronous for the observer in the system at rest** whose readings always correspond to the "time of the system at rest" at the

Classically

An observer co-moving with the rod measures this stationary rod to have a length of $t_B - t_A = t'_A - t_B = \frac{r_{AB}}{V}$, where V is the speed of light and t_B, t_A, t'_A, t'_B are all drawn from Def.2.

Another stationary observer, who sees the rod moving, measures the length of the (moving) rod seen from a system at rest. The light travels, due to **Classical Intertia**, at a speed of $v+V$, relative to the observer. The observer see the light reaches the other end of the rod at $\Delta t = t_B - t_A$. The light ends up travelling a distance of $(v+V)\Delta t$; this distance must also equal to the sum of the *stationary rod length* plus the *extra distance that the rod travels*:

$$(v+V)\Delta t = r_{AB} + v\Delta t \quad (1.3)$$

Thus

$$t_B - t_A = \frac{r_{AB}}{V} \quad (1.4)$$

When the light reflects back:

$$(V-v)\Delta t = v\Delta t - r_{AB} \quad (1.5)$$

$$t'_A - t_B = \frac{r_{AB}}{V} \quad (1.6)$$

Therefore

$$t_B - t_A = t'_A - t_B \quad (1.7)$$

Relativistically

An observer co-moving with the rod measures this stationary rod to have a length of $t_B - t_A = t'_A - t_B = \frac{r_{AB}}{V}$, where V is the speed of light and t_B, t_A, t'_A, t'_B are all drawn from Def.2.

Another stationary observer, who sees the rod moving, measures the length of the (moving) rod seen from a system at rest. The light travels, by the **Principle of the Constant Speed of Light**, at a speed of V , relative to the observer. The observer see the light reaches the other end of the rod at $\Delta t = t_B - t_A$. The light ends up travelling a distance of $V\Delta t$; this distance must also equal to the sum of the *stationary rod length* plus the *extra distance that the rod travels*:

$$V\Delta t = r_{AB} + v\Delta t \quad (1.8)$$

Thus

$$t_B - t_A = \frac{r_{AB}}{V-v} \quad (1.9)$$

When the light reflects back:

$$V\Delta t = v\Delta t - r_{AB} \quad (1.10)$$

$$t'_A - t_B = \frac{r_{AB}}{V+v} \quad (1.11)$$

Therefore

$$t_B - t_A \neq t'_A - t_B \quad (1.12)$$

Note that neither Eq.1.9 nor Eq.1.11 is related to the time = $\frac{\text{distance}}{\text{speed}}$. The numerators of both equations are NOT the distance traveled and the denominator is NOT the speed of the travel

Now, let's further imagine that each clock of C_1 and C_2 has an observer co-moving with it, and that these observers apply to the two clocks the criterion for synchronism formulated in Def 2. Classically, both observer co-moving with the moving rod and the observer in the system at rest declare C_1 and C_2 to be synchronous

$$t_B - t_A = t'_A - t'_B \quad (1.13)$$

Relativistically, however, observer co-moving with the moving rod, By Def. 2, finds C_1 and C_2 do not run synchronously

$$t_B - t_A \neq t'_A - t'_B \quad (1.14)$$

while the observer in the system at rest declare them synchronous, in this case

we have a contradiction

How do we resolve the contradiction?

There is No *Absolute* Simultaneity

Instead, two events that are simultaneous when observed from some particular coordinate system can no longer be considered simultaneous when observed from a system that is moving relative to that system

The fact that is no absolute conformance in simultaneity indicates that *there must be some form of transformation where two simultaneous events are transformed to be simultaneous in another system*. The next section explore this transformation which is famously called the "Lorentz transformation"

§3. Lorentz Transformation - Quantifying Non-Simultaneity

Let's imagine there are 2 coordinate systems, K and k , both "at rest" in their own right. Now let system k start moving in the increasing x -direction relative to K with a speed of v . Let's further image that K and k are contained within a "larger" coordinate system or space called S , where $S \neq K \neq k$. We now imagine the space S to be measured both from



the system K at rest by means of the measuring rod at rest; the measurement is obtained as (x, y, z) , and



the moving system k by means of the measuring rod moving along with it; the measurement is obtained as (ξ, η, ζ)

In the setup above, we assign a measuring rod to both K and k . Further, all clocks in K are **synchronized** to have a time t and all clocks in k are also **synchronized** to have a time τ . For every event (x, y, z, t) measured in the system K at rest, there corresponds to a "fixed" or transformed event (ξ, η, ζ, τ) measured in the system k . Our target is to derive the relation between $K(x, y, z, t)$ and $k(\xi, \eta, \zeta, \tau)$

In this setup, let's talk about how observer k measures things in K . Suppose there is a point at rest in $K : (x, y, z)$. Let's denote this point measured from within system k as $k : (x', y', z')$.

We will assume that the origins of k and K coincide initially^a. That is at $t_0 = 0, x = x'$

^aThe Collected Papers of Albert Einstein, Vol.2, page 149

After some time t , since K is effectively moving to in the negative x -direction relative to k , observer in k will now see this event further to the left, i.e. the coordinate of this event now becomes

$$x' = x + c - vt = x - vt \quad (1.15)$$

$c = 0$ because of the [assumption](#) we have made. Eq. 1.15 allow us to link systems k and K by a common attribute - x' . So let's start from here.

Looking at k only

Suppose that at time τ_0 a light ray is sent from the origin of the system k along the x -axis to x' and is reflected from there at time τ_1 toward the origin, where it arrives at time τ_2 ; we then must have

$$\begin{aligned} \tau_2 - \tau_1 &= \tau_1 - \tau_0 \\ \frac{1}{2}(\tau_0 + \tau_2) &= \tau_1 \end{aligned} \quad (1.16)$$

Since it is also true that $\tau_2 = \tau_0 + \underbrace{\frac{x'}{V-v}}_{\text{Eq. 1.9}} + \underbrace{\frac{x'}{V+v}}_{\text{Eq. 1.11}}$, by writing out the parameters of τ in Eq. 1.16, we have the following general equation of

$$\frac{1}{2} \left[\tau(0, 0, 0, t) + \tau \left(0, 0, 0, \tau_0 + \frac{x'}{V-v} + \frac{x'}{V+v} \right) \right] = \tau \left(x', 0, 0, t + \frac{x'}{V-v} \right) \quad (1.17)$$

Taking the derivative of τ with respect to x' and applying the Chain Rule of

$$\frac{\partial \tau}{\partial x'} = \frac{\partial \tau}{\partial t} \frac{\partial t}{\partial x'}$$

give us

$$\frac{1}{2} \frac{\partial \tau}{\partial t} \left[\frac{1}{V-v} + \frac{1}{V+v} \right] = \frac{\partial \tau}{\partial x'} + \left(\frac{1}{V-v} \right) \frac{\partial \tau}{\partial t} \quad (1.18)$$

which simplifies down to

$$\frac{\partial \tau}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial \tau}{\partial t} = 0 \quad (1.19)$$

It should be noted that $\mathcal{L} = \frac{\partial}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial}{\partial t}$ is a Linear operator ^a because

$$\mathcal{L}(\tau_1 + \tau_2) = \frac{\partial(\tau_1 + \tau_2)}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial(\tau_1 + \tau_2)}{\partial t} \quad (1.20)$$

$$= \frac{\partial \tau_1}{\partial x'} + \frac{\partial \tau_2}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial \tau_1}{\partial t} + \frac{v}{V^2 - v^2} \frac{\partial \tau_2}{\partial t} \quad (1.21)$$

$$= \mathcal{L}(\tau_1) + \mathcal{L}(\tau_2) \quad (1.22)$$

and

$$\mathcal{L}(c\tau) = \frac{\partial c\tau}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial c\tau}{\partial t} = c \left(\frac{\partial \tau}{\partial x'} + \frac{v}{V^2 - v^2} \frac{\partial \tau}{\partial t} \right) = c\mathcal{L}(\tau) \quad (1.23)$$

Therefore Eq. 1.19 has a solution of the form ^b

$$\tau = f\left(\frac{v}{V^2 - v^2}x - t\right) \quad (1.24)$$

We shall declare that

$$\tau = \underbrace{f\left(\frac{v}{V^2 - v^2}x' - t\right)}_{\text{Mathematics}} = \underbrace{\varphi(v)\left(t - \frac{v}{V^2 - v^2}x'\right)}_{\text{Physics}} \quad (1.25)$$

which assumes that transformation between τ and t must be *linear* which can be inferred by the following fundamental postulate^c:

Space and time are homogeneous^a

^aWhat proof do we have that the universe is homogenous? In physics we can't prove something like this. It must be a postulate - something we take as a fundamental assumption on which to base our theories. If the assumption is wrong then eventually we will find experimental evidence of this. What we can say is that currently there is no evidence for any lack of homogeneity or isotropy in the universe.

With Eq.1.25, we shall deduct the first 3 positional transformations in (ξ, η, ζ, τ) . For a light ray emitted at time $\tau = 0$ in the direction of increasing ξ , we have

$$\xi = V\tau = V\varphi(v)\left(t - \frac{v}{V^2 - v^2}x'\right) \quad (1.26)$$

which, combined with Eq.1.9, gives

$$\xi = \varphi(v)\left(\frac{V^2}{V^2 - v^2}\right)x' \quad (1.27)$$

Analogously, by considering light rays **moving along the two other moving axes while k is still moving in the x-direction**:

$$\begin{cases} \eta = V\varphi(v)\left(t - \frac{v}{V^2 - v^2}x'\right) \\ y^2 + (vt)^2 = (Vt)^2 \\ x' = 0 \end{cases} \Rightarrow \eta = \varphi(v)\frac{V}{\sqrt{V^2 - v^2}}y \quad (1.28)$$

$$\begin{cases} \zeta = V\varphi(v)\left(t - \frac{v}{V^2 - v^2}x'\right) \\ z^2 + (vt)^2 = (Vt)^2 \\ x' = 0 \end{cases} \Rightarrow \zeta = \varphi(v)\frac{V}{\sqrt{V^2 - v^2}}z \quad (1.29)$$

Note that Eq.1.27, 1.28, and 1.29 are based on the assumption that k starts out at the same point of K 's origin

^aSee [Partial Differential Equations](#), Strauss, page 2 for definition of *Linearity*

^bSee [Partial Differential Equations](#), Strauss, page 6

^cIn a [1964 paper](#), Erik Christopher Zeeman showed that the Lorentz transformation can be proven by applying the homogeneity principle to equations that represent the corresponding quantities at rest. This shows that the Lorentz transformation must be linear in both space and time coordinates

Now let's bridge k and K with Eq. 1.15 by plugging it into Eq. 1.25, 1.27, 1.28, 1.29:

$$\tau = \varphi(v) \left[t - \frac{v}{V^2 - v^2} (x - vt) \right] \quad (1.30)$$

$$= \varphi(v) \left(\frac{V^2 t - v^2 t - vx + v^2 t}{V^2 - v^2} \right) \quad (1.31)$$

$$= \varphi(v) \left(\frac{V^2 t - vx}{V^2 - v^2} \right) \quad (1.32)$$

$$= \varphi(v) \frac{V^2}{V^2 - v^2} \left(t - \frac{v}{V^2} x \right) \quad (1.33)$$

$$\xi = \varphi(v) \left(\frac{V^2}{V^2 - v^2} \right) (x - vt) \quad (1.34)$$

$$\eta = \varphi(v) \frac{V}{\sqrt{V^2 - v^2}} y \quad (1.35)$$

$$\zeta = \varphi(v) \frac{V}{\sqrt{V^2 - v^2}} z \quad (1.36)$$

$$(1.37)$$

It does not lose generality to say if k is moving relative to K at speed v then K is moving relative to k at speed $-v$. Directly measuring coordinates in K is effectively the same thing as transforming (x, y, z, t) in K to (ξ, η, ζ, τ) and then back to (x, y, z, t) so we have along the x-axis:

$$x = \varphi(-v) \left(\frac{V^2}{V^2 - v^2} \right) \varphi(v) \left(\frac{V^2}{V^2 - v^2} \right) x = \varphi(v) \varphi(-v) \left(\frac{V^2}{V^2 - v^2} \right)^2 x \quad (1.38)$$

where v is the speed of K relative to K , which is clearly 0. Therefore

$$x = \varphi(v) \varphi(-v) \left(\frac{V^2}{V^2 - 0^2} \right)^2 x = \varphi(v) \varphi(-v) x \quad (1.39)$$

which leads to the identity transformation of

$$\varphi(v) \varphi(-v) = 1 \quad (1.40)$$

Let's place a rod of length l at the origin of k perpendicular to ξ -axis. If k is moving to the x-direction relative to K at speed v , the measurement of the rod length, which we denote y_1 , in K satisfy

$$l = \varphi(v) \frac{V}{\sqrt{V^2 - v^2}} y_1 \quad (1.41)$$

By symmetry

$$l = \varphi(-v) \frac{V}{\sqrt{V^2 - v^2}} y_1 \quad (1.42)$$

$$\varphi(v) = \varphi(-v) \quad (1.43)$$

Combining 1.43 and 1.40 leads to

$$\varphi(v) = \pm 1 \quad (1.44)$$

By definition we know $l > 0$, $y_1 > 0$, and $V > 0$, therefore,

$$\varphi(v) = 1 \quad (1.45)$$

Since we focus on the x -direction only, it is our intuition that nothing shall change in y or z direction, i.e.

$$\eta = y \quad (1.46)$$

$$\zeta = z \quad (1.47)$$

Given that Eq. 1.25, 1.27, 1.28, 1.29 becomes

$$\tau = t - \frac{v}{V^2 - v^2} x' \quad (1.48)$$

$$\xi = \left(\frac{V^2}{V^2 - v^2} \right) x' \quad (1.49)$$

Part II

A First Course on General Relativity

Chapter 2

Special Relativity

2.1 Fundamental principles of special relativity theory (SR)

2.1.1 On “Principle of relativity (Galileo)”

Galilean invariance

[Newton’s laws of motion](#) hold in all frames related to one another by a [Galilean transformation](#). In other words, all frames related to one another by such a transformation are inertial (meaning, Newton’s equation of motion is valid in these frames).¹ The proof has been given by the book on page 2.

2.2 Construction of the coordinates used by another observer

2.2.1 Why would the tangent of the angle is the speed in Fig. 1.2?

Suppose $\mathcal{O}$ and $\bar{\mathcal{O}}$ both start out at the same position where $\bar{\mathcal{O}}$ moves along the x at some speed. After t_1 , observer $\mathcal{O}$ sees $\bar{\mathcal{O}}$ at position x_1 :

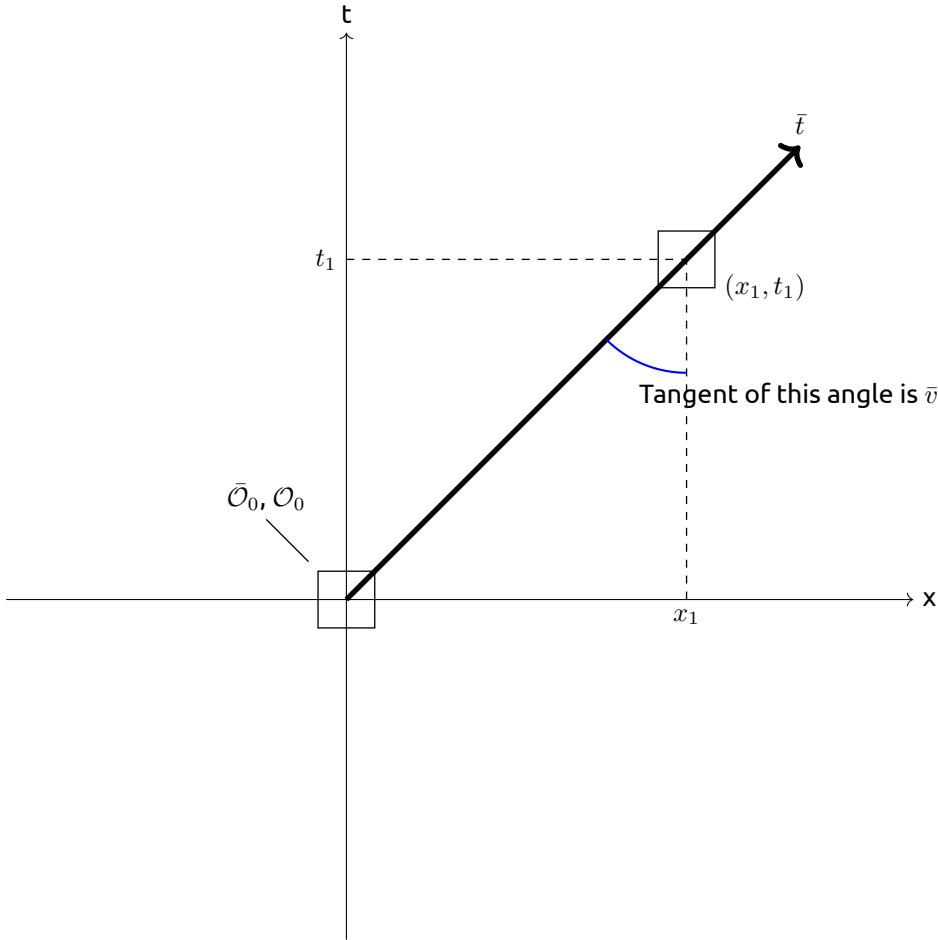
$$\bar{\mathcal{O}}_1 = (x_1, t_1)$$

Observer $\bar{\mathcal{O}}$, however, still sees themselves at $x = 0$:

$$\bar{\mathcal{O}}_1 = (0, t_1)$$

By definition where “ $\bar{t}$ ” is the locus of events at constant $\bar{x} = 0$ ”, $\bar{t}$ is the straight line that passes the origin and the (x_1, t_1) :

¹ [Galilean invariance](#)



2.3 Invariance of the interval

Why $(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (\Delta t)^2 = 0$ for two events in the same light beam?

Let's say, in a simplified 1D case, event $\mathcal{E} = (x_0, t_0)$ and $\mathcal{P} = (x_1, t_1)$.

$$(\Delta x)^2 - (\Delta t)^2 = (x_1 - x_0)^2 - (t_1 - t_0)^2$$

Since the speed of light is 1,

$$(x_1 - x_0)^2 - (t_1 - t_0)^2 = (x_1 - x_0)^2 - (t_1 \times 1 - t_0 \times 1)^2 = (x_1 - x_0)^2 - (x_1 - x_0)^2 = 0$$

Why does the equation contains only $M_{\alpha\beta} + M_{\beta\alpha}$ terms when $\alpha \neq \beta$, which guarantees $M_{\alpha\beta} = M_{\beta\alpha}$?

$$\Delta \bar{s}^2 = \sum_{\alpha=0}^3 \sum_{\beta=0}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta)$$

Before spending too much time on expanding the equation, we can pick up a pair of indices of $(\alpha, \beta) = (\alpha^*, \beta^*)$ where $\alpha^* \neq \beta^*$. Then we would definitely have the following 2 terms in the expansion:

$$M_{\alpha^* \beta^*} (\Delta x^{\alpha^*}) (\Delta x^{\beta^*})$$

$$M_{\beta^* \alpha^*} (\Delta x^{\beta^*}) (\Delta x^{\alpha^*})$$

Since

$$(\Delta x^{\alpha^*}) (\Delta x^{\beta^*}) = (\Delta x^{\beta^*}) (\Delta x^{\alpha^*})$$

We can then group these 2 terms and factor out the product, leaving

$$(\Delta x^{\alpha^*}) (\Delta x^{\beta^*}) (M_{\alpha^* \beta^*} + M_{\beta^* \alpha^*})$$

The terms of expanded $\Delta \bar{s}^2$ can be expressed in a matrix of

$$\begin{bmatrix} M_{00} \Delta x^0 \Delta x^0 & M_{01} \Delta x^0 \Delta x^1 & M_{02} \Delta x^0 \Delta x^2 & M_{03} \Delta x^0 \Delta x^3 \\ M_{10} \Delta x^1 \Delta x^0 & M_{11} \Delta x^1 \Delta x^1 & M_{12} \Delta x^1 \Delta x^2 & M_{13} \Delta x^1 \Delta x^3 \\ M_{20} \Delta x^2 \Delta x^0 & M_{21} \Delta x^2 \Delta x^1 & M_{22} \Delta x^2 \Delta x^2 & M_{23} \Delta x^2 \Delta x^3 \\ M_{30} \Delta x^3 \Delta x^0 & M_{31} \Delta x^3 \Delta x^1 & M_{32} \Delta x^3 \Delta x^2 & M_{33} \Delta x^3 \Delta x^3 \end{bmatrix}$$

Because the off-diagonal terms always appear in pairs above, we could effectively replace them with their mean value:

$$M_{\alpha^* \beta^*} = M_{\beta^* \alpha^*} = \frac{(M_{\alpha^* \beta^*} + M_{\beta^* \alpha^*})}{2}$$

where $\alpha^* \neq \beta^*$. And since $M_{\alpha\beta} = M_{\beta\alpha}$ if $\alpha = \beta$, we conclude that

$$M_{\alpha\beta} = M_{\beta\alpha} \text{ for all } \alpha \text{ and } \beta$$

Why do we have the 2nd term in equation

$$\Delta \bar{s}^2 = M_{00} (\Delta r)^2 + 2 \left(\sum_{i=1}^3 M_{0i} \Delta x^i \right) \Delta r + \sum_{i=1}^3 \sum_{j=1}^3 M_{ij} \Delta x^i \Delta x^j$$

$$\Delta \bar{s}^2 = \sum_{\alpha=0}^3 \sum_{\beta=0}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.1)$$

$$= \sum_{\alpha=0}^0 \sum_{\beta=0}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) + \sum_{\alpha=0}^3 \sum_{\beta=0}^0 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.2)$$

$$= \sum_{\beta=0}^3 M_{0\beta} \Delta t (\Delta x^\beta) + \sum_{\alpha=0}^3 M_{\alpha 0} (\Delta x^\alpha) \Delta t + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.3)$$

$$= M_{00} (\Delta t)^2 + \sum_{\beta=1}^3 M_{0\beta} \Delta t (\Delta x^\beta) + \sum_{\alpha=1}^3 M_{\alpha 0} (\Delta x^\alpha) \Delta t + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.4)$$

$$= M_{00} (\Delta t)^2 + \left[2 \sum_{i=1}^3 M_{0i} \Delta t (\Delta x^i) \right] + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.5)$$

Why would $M_{0i} = 0$ for $i = 1, 2, 3$ and $M_{ij} = -M_{00}\delta_{ij}$ in Equation 2.5?

The answer is: **not necessarily**. We are probably looking at a wrong problem.

The solution to exercise 1.8 takes $\Delta x_1 = -\Delta x_2$ to simplify the equation 2.10. This is not sufficient, because what if $\Delta x_1 \neq -\Delta x_2$? This box takes a general approach where we **do not assume any relationship between Δx_1 and Δx_2**

Note that this statement is based on the aforementioned assumption that $\Delta \bar{s}^2 = \Delta s^2 = 0$, which has been proved [here](#). Therefore, by 2.5, we have

$$\Delta \bar{s}^2(\Delta t, \Delta x_1) - \Delta \bar{s}^2(\Delta t, \Delta x_2) \quad (2.6)$$

$$= M_{00} (\Delta t)^2 + 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x^i) \right] + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.7)$$

$$= 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x_1^i) \right] + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x_1^\alpha) (\Delta x_1^\beta) - 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x_2^i) \right] - \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x_2^\alpha) (\Delta x_2^\beta) \quad (2.8)$$

$$= \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x_1^\alpha) (\Delta x_1^\beta) - \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x_2^\alpha) (\Delta x_2^\beta) + 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x_1^i) \right] - 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x_2^i) \right] \quad (2.9)$$

$$= \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} \left[(\Delta x_1^\alpha) (\Delta x_1^\beta) - (\Delta x_2^\alpha) (\Delta x_2^\beta) \right] + 2 \left[\sum_{i=1}^3 M_{0i} \Delta t (\Delta x_1^i - \Delta x_2^i) \right] = 0 \quad (2.10)$$

We won't be able to go further unless with some assumed relationships between Δx_1^i and Δx_2^i .



But since we do not assume any relations between them...

...let's step back and re-think about this problem then and forget about Δx_1^i and Δx_2^i .

We go through all these for the proof of invariance of the interval. This is to work out a relation between Δs^2 and $\Delta \bar{s}^2$. The detail is about Δx_1^i and Δx_2^i but the goal is to derive some form of

$$\Delta \bar{s}^2 = f(\Delta s^2) = \sum_{\alpha=0}^3 \sum_{\beta=0}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta)$$

where

$$\Delta s^2 = -(\Delta t)^2 + \sum_{i=1}^3 (\Delta x^i)^2$$

Let's work on $f(\Delta s^2)$ directly toward that goal then

Assuming $\Delta \bar{s}^2 = \Delta s^2 = 0$, we have $\Delta t = \pm \Delta x$; plugging it into Eq. 2.5 gives us

$$\Delta \bar{s}^2 = M_{00} \sum_{i=1}^3 (\Delta x^i)^2 + 2 \left[\sum_{i=1}^3 M_{0i} (\Delta x^i)^2 \right] + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.11)$$

$$= (M_{00} + 2M_{0i}) \sum_{i=1}^3 (\Delta x^i)^2 + \sum_{\alpha=1}^3 \sum_{\beta=1}^3 M_{\alpha\beta} (\Delta x^\alpha) (\Delta x^\beta) \quad (2.12)$$

Eq.2.12 seems to suggest a linear relationship between Δs^2 and $\Delta \bar{s}^2$. How do we go about proving it? We now start the formal proof of Invariance of Interval

Theorem 1 (Proportionality of ds^2 and ds'^2)

Let $n, p \geq 1$ be integers, $d := n + p$ and V a vector space over $\mathbb{R}$ of dimension d . Let h be an indefinite-inner product on V with signature type (n, p) . Suppose g is a symmetric bilinear form on V such that the null set of the associated quadratic form of h is contained in that of g (i.e. suppose that for every $v \in V$, if $h(v, v) = 0$ then $g(v, v) = 0$). Then, there exists a constant $C \in \mathbb{R}$ such that $g = Ch$. Furthermore, if we assume $n \neq p$ and that g also has signature type (n, p) , then we have $C > 0$

Prove the theorem

The assumptions above on h means that $h : V \times V \rightarrow \mathbb{R}$ is a bilinear form which is symmetric and non-degenerate such that there exists an ordered basis $\{v_1, \dots, v_n, v_{n+1}, \dots, v_d\}$ of V for which

$$h(v_a, v_b) = \begin{cases} -1 & a = b, \text{ where } a, b \in \{1, \dots, n\} \\ 1 & a = b, \text{ where } a, b \in \{n+1, \dots, d\} \\ 0 & \text{otherwise} \end{cases} \quad (2.13)$$

An equivalent way of saying this is that h has the matrix representation $\begin{bmatrix} -I_n & 0 \\ 0 & I_p \end{bmatrix}$ relative to the ordered basis $\{v_1, \dots, v_d\}$ of V

If we are considering the special case where $n = 1, p = 3$ then we are talking about the situation of a matrix signature in 4-dimensions

Part III

Mathematics

Chapter 3

Calculus

3.1 Calculus

Calculus is usually divided into two main parts

1. **Differential Calculus**, whose basic problem is the *problem of tangents*: Calculating the slope of the tangent line to the graph at a given point P
2. **Integral Calculus**, whose basic problem is the *problem of areas*: Calculating the area under the graph between the points $x = a$ and $x = b$

Much of the basic calculus focus on these 2 problems, on the ideas and techniques that have been developed for solving them, and on the applications that arise from them

3.1.1 The Definition of the Derivative

Definition 3: Derivative

Given any function $f(x)$, its derivative $f'(x)$ is the new function whose value at a point x is defined by

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = \frac{d}{dx} f(x) \quad (3.1)$$

If the limit exists for $x = a$, then the function is said to be *differentiable at a* . A *differentiable function* is one that is differentiable at each point of its domain.

The requirement that a limit "exists" in calculus inherently means that it evaluates to a **single, unique, finite real number**.

Definition 4: Differentiation

The process of actually forming the derivative $f'(x)$ is called the *differentiation of the given function $f(x)$*

3.1.2 Continuous Function

A function is continuous at the point a if

$$\lim_{\Delta x \rightarrow 0} f(x) = f(a) \quad (3.2)$$

A function is called a continuous function if it is continuous at every point in its domain.

With a slight change of notation, we can express the continuity of our function $f(x)$ at a point x (instead of a) in either of the equivalent forms

$$\lim_{\Delta x \rightarrow 0} f(x + \Delta x) = f(x) \quad \text{or} \quad \lim_{\Delta x \rightarrow 0} [f(x + \Delta x) - f(x)] = 0$$

and if we write $\Delta y = f(x + \Delta x) - f(x)$, then this condition becomes

$$\lim_{\Delta x \rightarrow 0} \Delta y = 0 \quad (3.3)$$

A function which is differentiable at a point is continuous at that point, because

$$\lim_{\Delta x \rightarrow 0} \Delta y = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \Delta x = \left[\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \right] \left[\lim_{\Delta x \rightarrow 0} \Delta x \right] = \frac{dy}{dx} 0 = 0$$

"The converse of this statement, however, is not true"

A function can be perfectly continuous at a point but fail to be differentiable if the curve has a sharp turn or a vertical slope.

A classic example is the absolute value function $f(x) = |x|$ at $x = 0$. The graph is entirely connected (continuous), but the limit of the secant slopes from the left is -1 while the limit from the right is $+1$. Because the one-sided limits of the difference quotient do not agree, the derivative does not exist at the origin.

Chapter 4

Calculus of Variation

4.1 Calculus of Variation

4.1.1 Origin of Calculus of Variations

The origin of the **calculus of variations** is a classic example of how a single, elegant challenge can birth an entire field of mathematics. While mathematicians had toyed with optimization for centuries, the formal discipline emerged from the **Brachistochrone problem** in 1696.

The Brachistochrone Challenge

In June 1696, **Johann Bernoulli** issued a public challenge to the greatest mathematicians of the era. He posed a deceptively simple question:

“Given two points A and B in a vertical plane, what is the path a particle must take to travel from A to B under the influence of gravity in the shortest possible time?”

— Johann Bernoulli

Common sense might suggest a straight line (the shortest distance), but gravity accelerates the particle differently depending on the slope. Johann knew the answer was a cycloid - the curve traced by a point on the rim of a rolling wheel - but he wanted to see who else could prove it.

The Famous Solutions

5 mathematicians successfully solved the challenge: Johann Bernoulli himself, his brother Jacob Bernoulli, Gottfried Leibniz, Guillaume de l'Hôpital, and an anonymous submission from Isaac Newton (who reportedly solved it in a single evening after a long day at the Mint).

While Johann's solution used an ingenious analogy with the refraction of light (Fermat's Principle), his brother **Jacob Bernoulli** provided a more rigorous mathematical foundation. Jacob's approach focused on making "infinitesimal" changes to the path itself to see how it affected the total time, which is the core logic of the calculus of variations.

Formalization by Euler and Lagrange

The field moved from a collection of isolated problems to a formal branch of analysis in the mid-1700s.

Leonhard Euler began generalizing these problems, moving beyond "shortest time" to "optimal anything." In 1744, he published the first systematic treatment of the subject titled *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes* (Latin for "A method for finding curved lines enjoying properties of maximum or minimum"). It is widely considered the **first textbook on the calculus of variations** and transformed the field from a set of clever tricks for specific problems (like the Brachistochrone) into a systematic branch of mathematical analysis.

In this text, Euler derived the first version of the fundamental differential equation of the calculus of variations. While he initially used a more geometric and "finite element" style of proof, the resulting equation is what we still use today.

The book also famously includes **The Principle of Least Action** applied to real-world physics by formulating a rigorous version of this principle for a single particle, suggesting that nature always acts to minimize a specific quantity (the integral of momentum over distance).

Interestingly, Euler's 1744 method was quite different from the modern "variational" approach. He essentially broke the curve into a series of straight-line segments (similar to how we think of discrete steps in modern computing) and used standard calculus to optimize the position of a single point.

It wasn't until **1755**, when the 19-year-old **Joseph-Louis Lagrange** sent Euler a letter in 1755 that introduced a purely algebraic method using the symbol δ to represent variations. Euler was so impressed by this "variational" approach that he dropped his own terminology and adopted Lagrange's, coining the term **"Calculus of Variations."**

Their collaboration resulted in the **Euler-Lagrange equation**, the bedrock of the field. It allows us to find a function $y(x)$ that minimizes or maximizes an integral (a functional) of the form:

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

The solution must satisfy:

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

Today, this is the language of modern physics, forming the basis of **Lagrangian mechanics** and the **Principle of Least Action**, which dictates how everything from planets to subatomic particles behaves.

Chapter 5

Topology

The name of this division ("Topology") is in honor of, from my sincere respect, the great book by George F. Simmons, [Introduction to Topology and Functional Analysis](#)



5.1 Metric Spaces

Definition 5: Metric Space

Let X be a non-empty set. A **metric**^a on X is a real function d of ordered pairs of elements of X which satisfies the following conditions



$$d(x, y) \geq 0, \text{ and } d(x, y) = 0 \iff x = y$$



$$d(x, y) = d(y, x) \text{ (symmetry)}$$



$$d(x, y) \leq d(x, z) + d(z, y) \text{ (the triangle inequality)}$$

^aIntroduction to Topology and Functional Analysis by George F. Simmons, p.51

$d(x, y)$ is called the *distance* between x and y . Thus a metric space consists of 2 objects:

1. a non-empty set X , and
2. a metric d on X

¹ Now let X be a metric space with metric d , and let

$$\{x_n\} = \{x_1, x_2, \dots, x_n, \cdot\} \tag{5.1}$$

be a sequence of points in X . We say that $\{x_n\}$ is **convergent** if there exists a point x in X such that either

1. for each $\epsilon > 0$, there exists a positive integer n_0 such that $n \geq n_0 \Rightarrow d(x_n, x) < \epsilon$, or, equivalently,

¹ Introduction to Topology and Functional Analysis by George F. Simmons, p.70

2. for each open sphere $S_\epsilon(x)$ centered on x , there exists a positive integer n_0 such that x_n is in $S_\epsilon(x)$ for all $n \geq n_0$

We say in this case that x_n converges to x and x is called the *limit* of the sequence $\{x_n\}$ and we sometimes write $x_n \rightarrow x$ in the form of

$$\lim x_n = x$$

Now let's consider two cases of convergence:



for each $\epsilon > 0$, there exists a positive integer n_0 such that $n \geq n_0 \Rightarrow d(x_n, x) < \epsilon$ (the same definition above)



for each $\epsilon > 0$, there exists a positive integer n_0 such that $m \geq n_0 \Rightarrow d(x_m, x_n) < \epsilon$

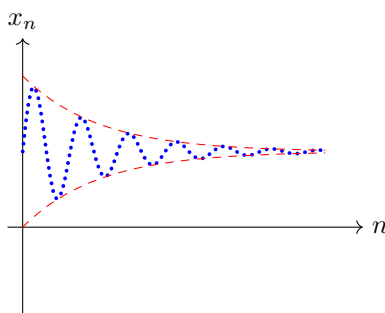
By the *symmetry* and *triangle inequality* of the Def. 5:

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) = d(x_m, x) + d(x_n, x) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad (5.2)$$

for all $m, n \geq n_0$. Therefore, every convergent sequence $\{x_n\}$ has the following property:

For each $\epsilon > 0$, there exists a positive integer n_0 such that $m, n \geq n_0 \Rightarrow d(x_m, x_n) < \epsilon$

A sequence with this property is called a **Cauchy sequence**. Intuitively, a Cauchy sequence is a sequence whose elements become arbitrarily close to each other as the sequence progresses:



In addition, we have also shown that every convergent sequence is a Cauchy sequence. *The converse of this, however, is not necessarily true.* That is, a Cauchy sequence is not necessarily convergent. As an example, consider the subspace $X = (0, 1]$ of the real line where $x_n \in X$. The sequence defined by $x_n = \frac{1}{n}$ is easily seen to be a Cauchy sequence in this space, but it is not convergent, because the point $0 \notin X$ is not a point of the space

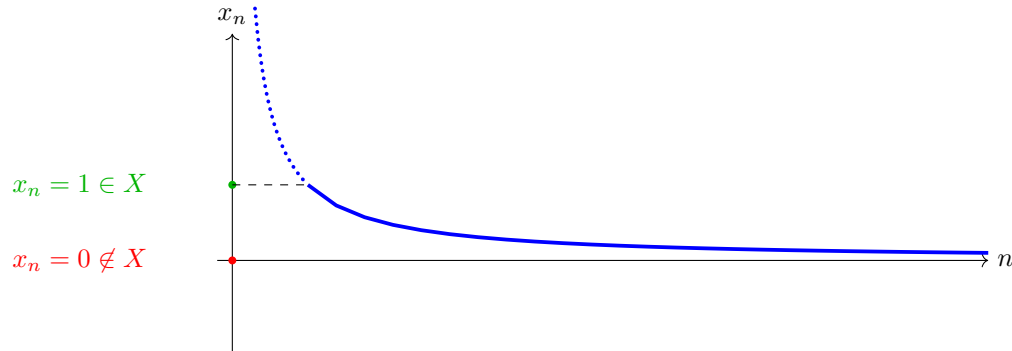


Figure 5.1: Here, the real numbers in the range of $X = (0, 1]$ takes on the values “y-axis”. We use a solid line to denote the range of n values under the subspace of X . It really should be a dotted line because n can only be integers. We do that in order to distinguish it from the range we are not concerned with, i.e. $X = (1, \infty)$

The notion of a convergent sequence is not intrinsic to the sequence itself, but also depends on the structure of the space in which it lies. A convergent sequence is not convergent “on its own”; it must converge to some point in the space...

...Sounds wordy?  Let's put it this way: when the space has a hole in it. Like if you remove 0 from $\mathbb{R}$ like in the case above, then $\frac{1}{n}$ is a Cauchy sequence that doesn't converge.

We are now ready to introduce the concept of *complete metric space*, which we define as follows:

Definition 6: Complete Metric Space

A complete metric space is a metric space in which every Cauchy sequence is convergent

Definition 7: Linear Space

Let L be a non-empty set, and assume that each pair of elements x and y in L can be combined by a process called **addition** to yield an element z in L denoted by $z = x + y$. Assume also that this operation of addition satisfies the following conditions:



$$x + y = y + x$$



$$x + (y + z) = (x + y) + z$$



There exists in L a unique element, denoted by 0 and called the **zero element**, or the origin, such that $x + 0 = x$ for every x



To each element x in L there corresponds a unique element in L , denoted by $-x$ and called the **negative of x** , such that $x + (-x) = 0$.

We adopt the device of referring to the system of real numbers or to the system of complex numbers as the **scalars**. We now assume that each scalar α and each element x in L can be combined by a process called **scalar multiplication** to yield an element y in L denoted by $y = \alpha x$ in such a way that



$$\alpha(x + y) = \alpha x + \alpha y$$



$$(\alpha + \beta)x = \alpha x + \beta x$$



$$(\alpha\beta)x = \alpha(\beta x)$$



$$1 \cdot x = x$$

The algebraic system L defined by these operations and axioms is called a **linear space**^a

^aIntroduction to Topology and Functional Analysis by George F. Simmons, p.81

Depending on the numbers admitted as scalars (only the real numbers, or all the complex numbers), we distinguish when necessary between *real linear spaces* and *complex linear spaces*. A linear space is often called a **vector space** and its elements are spoken of as *vectors*.

Definition 8: Normed Linear Space

A **Normed Linear Space**^a is a linear space on which there is a **norm** defined, i.e. a function which assigns to each element x in the space a real number $\|x\|$ in such a manner that



$$\|x\| \geq 0, \text{ and } \|x\| = 0 \iff x = 0$$



$$\|x + y\| \leq \|x\| + \|y\|$$



$$\|\alpha x\| = |\alpha| \|x\|$$

^aIntroduction to Topology and Functional Analysis by George F. Simmons, p.81

Intuitively, a normed linear space is simply a linear space in which a notion of the distance from an arbitrary element to origin is defined.

Definition 9: Banach space

A **Banach space**^a is a normed linear space which is complete as a metric space.

^aIntroduction to Topology and Functional Analysis by George F. Simmons, p.81

One of the principal applications of the theory of Banach algebras developed from Def. 5 to Def 9 is to the study of operators in Hilbert spaces.²

Definition 10: Hilbert Space

A **Hilbert space**^a is a complex Banach space whose norm arises from an inner product in which there is defined a complex function (v, w) of vectors v with the following properties:



$$(\alpha v + \beta w, u) = \alpha(v, u) + \beta(w, u)$$

²Introduction to Topology and Functional Analysis by George F. Simmons, p.243



$$\overline{(v, w)} = (w, v)$$



$$(v, v) = \|x\|^2$$

^aIntroduction to Topology and Functional Analysis by Goerge F. Simmons, p.245

A complex function^a is one whose range consists of complex numbers

If $z = a + ib$ is a complex number, then its conjugate $\bar{z}$ is defined by $\bar{z} = a + i(-b)$ ^b

^aIntroduction to Topology and Functional Analysis by Goerge F. Simmons, p.17

^bIntroduction to Topology and Functional Analysis by Goerge F. Simmons, p.53

5.2 Indefinite Inner Products

Definition 11: Indefinite Inner Product

Let $\mathbb{C}^n$ be the n -dimensional complex Hilbert space consisting of all column vectors x with complex coordinate $x^{(j)}$, $j = 1, 2, \dots, n$. The typical column vector x will be written in the form $x = \langle x^{(1)}, x^{(1)}, \dots, x^{(n)} \rangle$. The standard inner product in $\mathbb{C}^n$ is denoted by $(\cdot, \cdot)$. Thus,

$$(x, y) = \sum_{j=1}^n x^{(j)} \overline{y^{(j)}} \quad (5.3)$$

where $x = \langle x^{(1)}, x^{(1)}, \dots, x^{(n)} \rangle$, $y = \langle y^{(1)}, y^{(1)}, \dots, y^{(n)} \rangle$

A function $(\cdot, \cdot)$ from $\mathbb{C}^n \times \mathbb{C}^n$ to $\mathbb{C}$ is called an **indefinite inner product** in $\mathbb{C}^n$ ^a if the following axioms are satisfied:



Linearity in the first argument:

$$[\alpha x_1 + \beta x_2, y] = \alpha[x_1, y] + \beta[x_2, y] \quad (5.4)$$

for all $x_1, x_2, y \in \mathbb{C}^n$ and all complex numbers α, β



Antisymmetry:

$$[x, y] = \overline{[y, x]} \quad (5.5)$$

for all $x, y \in \mathbb{C}^n$



Nondegeneracy: if $[x, y] = 0$ for all $y \in \mathbb{C}^n$, then $x = 0$

^aIndefinite Linear Algebra and Applications, L. Rodman, Peter Lancaster, Israel Gohberg. 2005, p.7

Chapter 6

Linear Algebra

6.1 Linear Algebra

Definition 12: Vector Space

The space $\mathbb{R}^n$ consists of all column vectors v with n components^a

^aIntroduction to Linear Algebra, Strang, 4th Edition, 2009, p. 120

The components of v are real numbers, which is the reason for the letter $\mathbb{R}$. A vector whose n components are complex numbers lies in the space $\mathbb{C}^n$

Definition 13: Subspace

A **subspace**^a of a **vector space** is a set of vectors (including 0) that satisfies 2 requirements: If v and w are vectors in the subspace and c any scalar, then



$v + w$ is in the subspace



cv is in the subspace

In other words, the set of vectors is “**closed**” under addition and multiplication - **all linear combinations stay in the subspace**

^aIntroduction to Linear Algebra, Strang, 4th Edition, 2009, p. 122

Intuitively, we can visualize a subspace in the 3-dimensional space $\mathbb{R}^3$. Choose a plane through the origin $(0, 0, 0)$. That plane is a vector space in its own right. If we add two vectors in the plane, their sum is in the plane; if we multiply an in-plane vector by 2 or -5 , it is still in the plane. This plane is a vector space **inside** $\mathbb{R}^3$ or is a subspace of the full vector space $\mathbb{R}^3$

Definition 14: Column Space

The **column space**, $C(A)$, consists of all linear combinations of the columns, i.e. the combinations of all possible vectors Ax

The subspace $C(A)$ is the “**span**” of matrix A

Definition 15: Span

A set of vectors spans a space if their linear combinations fill the space

Definition 16: Basis for a Vector Space

A **basis**^a for a vector space is a sequence of vectors with 2 properties



The basis vectors are linearly independent, and



they **span** the space

^aIntroduction to Linear Algebra, Strang, 4th Edition, 2009, p. 172

6.1.1 Eigenvalues and Eigenvectors

All matrices in this section are square

Almost all vectors change direction, when they are multiplied by a matrix A . Certain exceptional vectors x don't and are in the same direction as Ax . Those are the **eigenvectors**¹. By definition, the result of this multiplication shall be a number times the original x , i.e.

$$Ax = \lambda x \quad (6.1)$$

We call λ an **eigenvalue** of A .

The eigenvalue λ tells whether the special vector x is stretched (e.g. 2) or shrunk (e.g. $\frac{1}{2}$) or reversed (-1) or left unchanged (1) when multiplied by A . When $\lambda = 0$, Then $Ax = 0x$ means this eigenvector x is in nullspace²

Matrix Diagonalization**Why do we study diagonalizing a matrix?**

When x is an eigenvector, multiplication by A is just multiplication by a number λ : $Ax = \lambda x$. All the difficulties of matrices are swept away, because the matrix A turns into a diagonal matrix Λ when we use the eigenvectors properly.

Suppose a n by n matrix A has n linearly independent eigenvectors $x_1, \dots, x_n$. Put them into the columns of an **eigenvector matrix** S . Multiplying A with S ³:

¹Introduction to Linear Algebra, Strang, 4th Edition, 2009, p. 283

²

Definition 17: Nullspace

The nullspace of A consists of all solutions to $Ax = 0$. These vectors x are in $\mathbb{R}^n$. The nullspace containing all solutions of $Ax = 0$ is denoted by $N(A)$ (Introduction to Linear Algebra, Strang, 4th Edition, 2009, p. 132)

³Introduction to Linear Algebra, Strang, 4th Edition, 2009, p. 298

$$AS = A \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \quad (6.2)$$

$$= \begin{bmatrix} \lambda_1 x_1 & \dots & \lambda_n x_n \end{bmatrix} \quad (6.3)$$

$$= \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad (6.4)$$

$$= S \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad (6.5)$$

Let

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad (6.6)$$

Then

$$AS = S\Lambda \quad \text{or} \quad S^{-1}AS = \Lambda \quad \text{or} \quad A = S\Lambda S^{-1} \quad (6.7)$$

It is important to keep in mind that S has an inverse because it's columns are linearly independent



We call Λ the **Eigenvalue Matrix**

Definition 18: Diagonalization

Suppose a n by n matrix A has n linearly independent eigenvectors $x_1, \dots, x_n$. Then $S^{-1}AS$ is the **eigenvalue matrix** Λ

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad (6.8)$$

Definition 19: Bilinear Form

Let F be a field and V a vector space over F . A bilinear form on V is a function $B : V \times V \rightarrow F$ that is linear in each variable when the other one is fixed. That is

$$B(v + v', w) = B(v, w) + B(v', w) \quad (6.9)$$

$$B(cv, w) = cB(v, w) \quad (6.10)$$

for all $v, v', w \in V$ and $c \in F$, and

$$B(v, w + w') = B(v, w) + B(v, w') \quad (6.11)$$

$$B(v, cw) = cB(v, w) \quad (6.12)$$

for all $v, w, w' \in V$ and $c \in F$

We call B symmetric when

$$B(v, w) = B(w, v) \quad (6.13)$$

for all $v, w \in V^a$

Definition 20: Linear Transformation

The transformation is **linear** if it meets these requirements for all v and w^a :

$$T(cv + dw) = cT(v) + dT(w) \quad (6.14)$$

for all c and d

^aIntroduction to Linear Algebra, Strang, 4th Edition, 2009, p. 375

Definition 21: Field

A field is a set of elements in which a pair of operations called multiplication and addition is defined analogous to the operations of multiplication and addition in the real number system (which is itself an example of a field)^a

^aA. Fields, Chapter I: Linear Algebra, Galois Theory: Lectures Delivered at the University of Notre Dame, Project Euclid

Definition 22: Group

A group^a is a non-empty set G together with a rule that assigns to each pair g and h of elements of G and element $g * h$ such that



$g * h \in G$, which we say G is **closed** under $*$



$g * (h * k) = (g * h) * k$ for all $g, h, k \in G$, which we call $*$ being **associative**



There exists an **identity** element $e \in G$ such that $e * g = g * e$ for all $g \in G$



Every element $g \in G$ has an **inverse** g^{-1} such that $g * g^{-1} = g^{-1} * g = e$

^aIntroduction to Group Theory, Michael Wemyss

Definition 23: Vector Space (Field Theory)

If V is an additive abelian group with elements $A, B, \dots$, F a field with elements $a, b, \dots$, and if for each $a \in F$ and $A \in V$ the product aA denotes element of V , then V is called a **(left) vector space over F** if the following assumptions hold:

$$a(A + B) = aA + aB \quad (6.15)$$

$$(a + b)A = aA + bA \quad (6.16)$$

$$a(bA) = (ab)A \quad (6.17)$$

$$1A = A \quad (6.18)$$

$$(6.19)$$

^aKeith Conrad

Chapter 7

Riemannian Geometry

This chapter was raised by the discussion of [Green's Theorem](#)

There are lots of texts around the topics of Riemannian Geometry. Here is a list of threads in which great people recommended good books on it



[Riemannian and Pseudo-Riemannian Geometry](#)



[Beginner's book for Riemannian geometry](#)



[Need help finding a good book on Riemann Geometry](#)



[Introductory text on Riemannian geometry](#)

One will need quite a solid background in topology, and especially differential geometry. I would recommend the book series by [Lee: Introduction to topological manifolds](#) / introduction to smooth manifolds / introduction to riemannian manifolds.¹

Study notes on Riemannian Geometry

7.1 What are Manifolds?

7.1.1 Intuitive Meaning

A manifold is a space that "looks like" regular old Euclidean space (like a line, a plane, 3d space, and so on) if we zoom in enough. More formally, every point in a manifold has a neighborhood that is homeomorphic to ("the same as") a neighborhood in some Euclidean space.

Easy examples include circles and spheres. A circle is curved in a global sense, but if you're reallllllly close to a circle, it looks like a line. Similarly, a sphere is curved in a global sense, but if you're really close, it looks like a plane (which is why the Earth looks flat when we live on it). If we cut off a little chunk of circle, it's essentially just a line (that is, it's the same as 1D Euclidean space). If you cut off a little chunk of sphere, it's essentially just a plane (that is, it's the same as 2D Euclidean space)².

¹ [Book recommendations](#)

² [r/explainlikeimfive](#)

This sounds like a manifold is always *1 dimension higher* than Euclidean space. In this sense, an n -dimensional Euclidean space $\mathbb{R}^n$ is the **prototype** of an n -dimensional manifold.

It should be noted that the number “ n ” in manifold is not the same thing as the number of coordinates we use to locate a position in Euclidean space.

For instance, manifolds of dimension 1 are lines and curves. It is easy for us to see that a real line is an example of such case. The space curves, which are often described parametrically by equations such as $(x, y, z) = (f(t), g(t), h(t))$, are also 1-dimensional manifolds. See Fig 7.1 below:

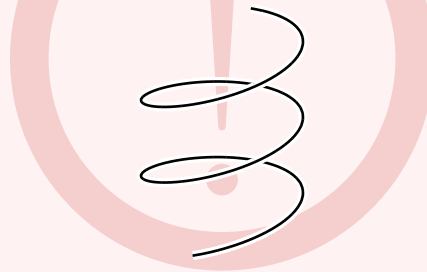


Figure 7.1: Intuitively, if we keep zooming in, we get a straight line on the spiral

In essence, an n -dimensional manifold “looks like” $\mathbb{R}^n$ locally.

7.1.2 Formal Definition of Manifolds

³The chief problem with the [intuitive introduction of manifolds](#) is that it depends on having an “ambient Euclidean space” in which our n -manifold lives. This introduces a great deal of extraneous structure that is irrelevant to our purpose. Instead, we would like to view a manifold as a mathematical object in its own right, not as a subset of some larger space. The key concept that makes this possible is that of a *topological space*.

We begin by defining topological spaces, motivated by the open subset criterion from continuity metric spaces.

Definition 24: Continuous Mapping

If (M_1, d_1) and (M_2, d_2) are metric spaces and x is a point in M_1 , a map $f : M_1 \rightarrow M_2$ is said to be **continuous at x** if for any $\epsilon > 0$ there exists $\delta > 0$ such that $d_1(x, y) < \delta$ implies $d_2(f(x), f(y)) < \epsilon$ for all $y \in M_1$; and f is **continuous** if it is continuous at every point of M_1 ^a

^aIntroduction to Topological Manifolds, John M. Lee, 2nd, P.398

Definition 25: Image

Let $f : X \rightarrow Y$ be a function. If $S \subset X$, the **image of S under f** , denoted by $f(S)$, is the subset of Y defined by^a

$$f(S) = \{y \in Y : y = f(x) \text{ for some } x \in S\} \quad (7.1)$$

^aIntroduction to Topological Manifolds, John M. Lee, 2nd, P.387

³Introduction to Topological Manifolds, John M. Lee, 2nd, P.19

Definition 26: Preimage

If T is a subset of Y , the **preimage of T under f** (also called the **inverse image**) is the subset $f^{-1}(T) \subseteq X$ defined by^a

$$f^{-1}(T) = \{x \in X : f(x) \in T\} \quad (7.2)$$

^aIntroduction to Topological Manifolds, John M. Lee, 2nd, P.388

Definition 27: Open Subset

Let M be a metric space^a



For any $x \in M$ and $r > 0$, the **(open) ball of radius r around x** is the set

$$B_r(x) = \{y \in M : d(y, x) < r\} \quad (7.3)$$

and the **closed ball of radius r around x** is

$$\overline{B}_r(x) = \{y \in M : d(y, x) \leq r\} \quad (7.4)$$



A subset $A \subseteq M$ is said to be an **open subset of M** if it contains an open ball around each of its points



A subset $A \subseteq M$ is said to be a **closed subset of M** if $M \setminus A$ (set subtraction) is open

^aIntroduction to Topological Manifolds, John M. Lee, 2nd, P.397

Prove the Properties of Open Subsets of a Metric Space (Lee, p.397)

Prove the Properties of Closed Subsets of a Metric Space (Lee, p.397)

Theorem 2 (Open Subset Criterion for Continuity)

A map $f : M_1 \rightarrow M_2$ between metric spaces is **continuous** if and only if the **preimage** of every open subset of M_2 is open; that is, whenever U is an open subset of M_2 , its preimage $f^{-1}(U)$ is open in M_1

Proof First assume f is continuous, and let $U \subseteq M_2$ be an open set.

Let $x \in M_1$ be an element of the **preimage** of U ; that is, x is any point in

$f^{-1}(U) = \{x \in M_1 : f(x) \in U\}$. Since U is open, there exists some $r = \epsilon > 0$ such that $B_\epsilon(\text{element in } U) = B_\epsilon(f^{-1}(U)) \subseteq U$

Finish the proof

Part IV

Electromagnetism

Why study Electromagnetism?

This question arose because I noticed Einstein dedicated a huge portion of his famous [Special Relativity paper](#) on its application in Electromagnetism.

In fact, Einstein was trying ridiculously hard to unify his General Theory of Relativity with Electromagnetism^a. Why?

Relativity describes our Universe pretty good but works horribly in the microscopic world like sub-atomic particles, whereas Quantum Physics works from the opposite. Relativity is, in some sense, *not complete*

Those who are familiar with Maxwell's Equations know by heart that Electrodynamics is a beautifully *complete* and successful theory. It has become a king of paradigm for physicists: an ideal model that other theories strive to emulate.

Studying Electromagnetism is same thing as studying the standard model of Physics which shall guide my study of General Theory of Relativity.

^a...[Albert Einstein, who attempted to unify his general theory of relativity with electromagnetism ...](#), Wikipedia

Chapter 8

Mathematics

8.1 Differential Calculus

Theorem 3 (Fundamental Lemma^a)

^aCalculus with Analytic Geometry by George F. Simmons, 2nd, p. 680

Suppose that a function $z = f(x, y)$ and its partial derivatives f_x and f_y are defined at a point (x_0, y_0) , and also through some neighborhood of this point. Suppose further that f_x and f_y are continuous at (x_0, y_0) . Then the increment Δz can be expressed in the form of

$$\Delta z = f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \epsilon_1\Delta x + \epsilon_2\Delta y \quad (8.1)$$

where ϵ_1 and $\epsilon_2 \rightarrow 0$ as Δx and $\Delta y \rightarrow 0$

To prove this statement¹, we analyze the change Δz in 2 steps as shown in Fig. 8.1:

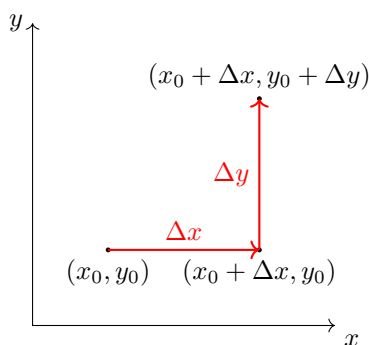


Figure 8.1: We assume $\Delta z = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$ and $\Delta z = \Delta_1 z + \Delta_2 z$

1. changing x alone and moving from (x_0, y_0) to $(x_0 + \Delta x, y_0)$, and then
2. changing y alone and moving from $(x_0 + \Delta x, y_0)$ to $(x_0 + \Delta x, y_0 + \Delta y)$

We denote the first change in z by $\Delta_1 z$, so that

¹Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 841

$$\Delta_1 z = f(x_0 + \Delta x, y_0) - f(x_0, y_0) \quad (8.2)$$

By The Mean Value Theorem², we can write this as

$$\Delta_1 z = \Delta x f_x(x_1, y_0) \quad (8.3)$$

where x_1 is between x_0 and $x_0 + \Delta x$. Smilary, if we denote the second part of the change in z by $\Delta_2 z$, so that

$$\Delta_2 z = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) \quad (8.4)$$

then

$$\Delta_2 z = \Delta y f_y(x_0 + \Delta x, y_1) \quad (8.5)$$

where y_1 is between y_0 and $y_0 + \Delta y$.

Now as Δx and $\Delta y \rightarrow 0$, $x_1 \rightarrow x_0$ and $y_1 \rightarrow y_0$. By the assumed continuity of f_x and f_y at (x_0, y_0) , we can write

$$f_x(x_1, y_0) = f_x(x_0, y_0) + \epsilon_1 \quad (8.6)$$

$$f_y(x_0 + \Delta x, y_1) = f_y(x_0, y_0) + \epsilon_2 \quad (8.7)$$

where ϵ_1 and $\epsilon_2 \rightarrow 0$ as Δx and $\Delta y \rightarrow 0$. Plugging Eq.8.6 into Eq.8.3 gives us

$$\Delta_1 z = \Delta x [f_x(x_0, y_0) + \epsilon_1] = \Delta x f_x(x_0, y_0) + \Delta x \epsilon_1 \quad (8.8)$$

and similarly Eq.8.7 into Eq.8.5

$$\Delta_2 z = \Delta y [f_y(x_0, y_0) + \epsilon_2] = \Delta y f_y(x_0, y_0) + \Delta y \epsilon_2 \quad (8.9)$$

2

Theorem 4 (The Mean Value Theorem^a)

^aCalculus with Analytic Geometry by George F. Simmons, 2nd, p. 76

Let $y = f(x)$ be a function with the following two properties:

1. $f(x)$ is continuous on the closed interval $[a, b]$; and
2. $f(x)$ is differentiable on the open interval (a, b)

Then there exists at least one point c in the open interval (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

or equivalently,

$$f(b) - f(a) = f'(c)(b - a)$$

Since we have assumed $\Delta z = \Delta_1 z + \Delta_2 z$

$$\Delta z = \Delta x f_x(x_0, y_0) + \Delta x \epsilon_1 + \Delta y f_y(x_0, y_0) + \Delta y \epsilon_2 = f_x(x_0, y_0) \Delta x + f_y(x_0, y_0) \Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y \quad (8.10)$$



³ Now Let $f(x, y, z)$ be a function of 3 variables defined throughout some region of three-dimensional space, and let P be a point in this region. At what rate does f change as we move away from P in a specified direction? In the directions of the positive x , y , and z -axes, we know that the rates of change off are given by the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$. But how do we calculate the rate of change of f if we move away from P in a direction that is not a coordinate direction?

Let $P = (x, y, z)$ and $\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ being the position vector of P . If we move away from P to a nearby point $Q = (x + \Delta x, y + \Delta y, z + \Delta z)$, then the function will change by an amount Δf . Let Δs denote the distance between P and Q , then we have

$$\frac{df}{ds} = \lim_{\Delta s \rightarrow 0} \frac{\Delta f}{\Delta s} \quad (8.11)$$

We further assume that $f(x, y, z)$ has continuous partial derivatives with respect to x , y , and z .

! Unless explicitly stated otherwise, all functions we deal with are always continuous in all of our discussions

The **Fundamental Lemma** enables us to write Δf in the form of

$$\Delta f = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y + \frac{\partial f}{\partial z} \Delta z + \epsilon_1 \Delta x + \epsilon_2 \Delta y + \epsilon_3 \Delta z \quad (8.12)$$

As $\Delta s \rightarrow 0$, i.e. as $\Delta x \rightarrow 0$, $\Delta y \rightarrow 0$, and $\Delta z \rightarrow 0$, $\epsilon_1, \epsilon_2, \epsilon_3 \rightarrow 0$. Dividing Eq.8.12 by Δs gives

$$\lim_{\Delta s \rightarrow 0} \frac{\Delta f}{\Delta s} = \frac{\partial f}{\partial x} \frac{dx}{ds} + \frac{\partial f}{\partial y} \frac{dy}{ds} + \frac{\partial f}{\partial z} \frac{dz}{ds} \quad (8.13)$$

Combing Eq.8.13 and 8.11 results in

$$\frac{df}{ds} = \frac{\partial f}{\partial x} \frac{dx}{ds} + \frac{\partial f}{\partial y} \frac{dy}{ds} + \frac{\partial f}{\partial z} \frac{dz}{ds} \quad (8.14)$$

8.1.1 Gradient

⁴ The theorem on **partial derivaves states** that

$$dT = \left(\frac{\partial T}{\partial x} \right) dx + \left(\frac{\partial T}{\partial y} \right) dy + \left(\frac{\partial T}{\partial z} \right) dz \quad (8.15)$$

³ Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 681

⁴ Introduction to Electrodynamics by Griffiths, 3rd, p. 13

Writing it in the dot product form:

$$dT = \left(\frac{\partial T}{\partial x} \hat{x} + \frac{\partial T}{\partial y} \hat{y} + \frac{\partial T}{\partial z} \hat{z} \right) \cdot (dx \hat{x} + dy \hat{y} + dz \hat{z}) \quad (8.16)$$

$$= \nabla T \cdot d\mathbf{l} \quad (8.17)$$

where

$$\nabla T \equiv \frac{\partial T}{\partial x} \hat{x} + \frac{\partial T}{\partial y} \hat{y} + \frac{\partial T}{\partial z} \hat{z} \quad (8.18)$$

is the **gradient** of T . We also call ∇ as the **vector operator** that *acts upon* T

The Geometrical Interpretation of the Gradient

The dot product 8.17 can be written as

$$dT = \nabla T \cdot d\mathbf{l} = |\nabla T| |d\mathbf{l}| \cos \theta \quad (8.19)$$

We soon realize that the *maximum* change of T occurs when $\theta = 0$, therefore

The gradient ∇T points in the direction of the maximum increase of T , and its magnitude $|\nabla T|$ gives the slope (rate of increase) along this maximal direction

Now there are 3 ways the operator ∇ can act:

1. On a scalar function T : ∇T (the gradient, which we've discussed so far)
2. On a vector function via the dot product: $\nabla \cdot \mathbf{v}$ (the **divergence**)
3. On a vector function via the cross product: $\nabla \times \mathbf{v}$ (the **curl**)

8.1.2 Divergence

From the definition of ∇ , we construct the divergence

$$\nabla \cdot \mathbf{v} = \left(\frac{\partial T}{\partial x} \hat{x} + \frac{\partial T}{\partial y} \hat{y} + \frac{\partial T}{\partial z} \hat{z} \right) \cdot (v_x \hat{x} + v_y \hat{y} + v_z \hat{z}) \quad (8.20)$$

$$= \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \quad (8.21)$$

The Meaning of Divergence

The Geometrical Interpretation of the Divergence

The divergence is a measure of how much the vector $\mathbf{v}$ spreads out (diverges) from the point in question

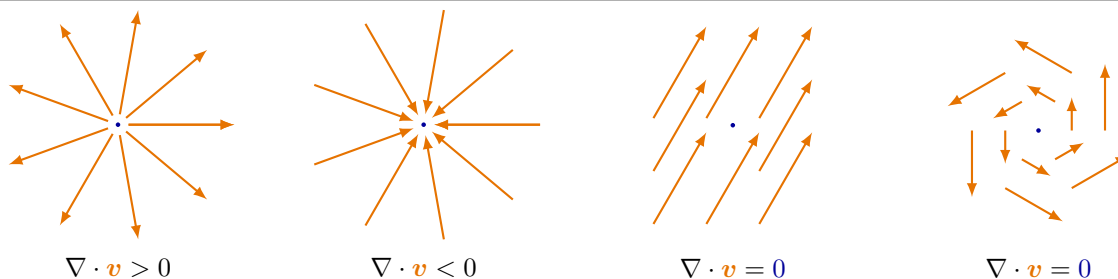


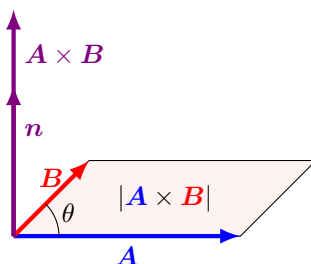
Figure 8.2: Divergence of vector fields

8.1.3 Curl

The Cross Product of Two Vectors

⁵ Many problems in geometry require us to find a vector that is perpendicular to each of two given vectors A and B . A routine way of doing this is provided by the *cross product* (or *vector product*) of A and B , denoted by $A \times B$. This cross product is very different from the *dot product* $A \cdot B$, because $A \times B$ is a vector while $A \cdot B$ is a scalar.

Consider two nonzero vectors A and B . Suppose their tails coincide and let θ be the angle from A to B (not from B to A) with $0 \leq \theta \leq \pi$.

Figure 8.3: The plane defined by A and B

These 2 vectors determine a plane, as shown in Fig. 8.3. We now choose the unit vector n which is normal (perpendicular) to this plane and whose direction is determined by the *right-hand thumb rule*⁶. This gives the direction of $A \times B$.

Vectors A and B also defines a parallelogram in this plane of area $|A||B| \sin \theta$, which defines the *magnitude* of $A \times B$.

Definition 28: Cross Product of A and B

$$A \times B = |A||B| \sin \theta \quad (8.22)$$

Our next objective is to develop a convenient formula for calculating $A \times B$ where

$$A = a_1 \hat{i} + a_2 \hat{j} + a_3 \hat{k} \quad \text{and} \quad B = b_1 \hat{i} + b_2 \hat{j} + b_3 \hat{k} \quad (8.23)$$

We need to know that the cross product possesses the following algebraic properties

⁵Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 640

⁶This means that if the right hand is placed so that the thumb is perpendicular to the plane of A and B and the fingers curl from A to B in the direction of angle θ , then n points in the same direction as the thumb of this hand

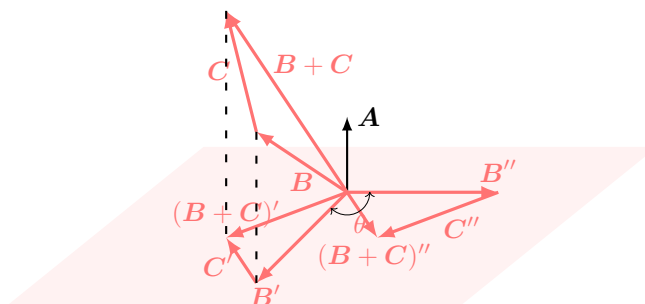
$$(cA) \times B = c(A \times B) = A \times (cB), \quad (8.24)$$

$$A \times (B + C) = A \times B + A \times C, \quad (8.25)$$

$$(A + B) \times C = A \times C + B \times C \quad (8.26)$$

Property 8.24, also called *homogeneous* in each argument⁷, is easily established directly from Definition 28.

The proof of Eq. 8.25 starts with a unit vector $\hat{n}$ and 3 arbitrary vectors B , C , and $B + C$. $\hat{n} \times (B + C)$ can be constructed by performing the following two operations:



1. Project B , C , and $(B + C)$ onto the plane perpendicular to $\hat{n}$ to obtain a vector B' , C' , and $(B + C)'$. By the nature of projection, the head and tails of B' , C' , and $(B + C)'$ still coincide. Then,
2. rotate the triangle formed by B' , C' , and $(B + C)'$ by 90 degrees counterclockwise with respect to the tail of $\hat{n}$ to obtain B'' , C'' , and $(B + C)''$, which still form a triangle.

Therefore, we have

$$(B + C)'' = B'' + C''$$

What this means is, geometrically, the operation of $\hat{n} \times (B + C)$ and $\hat{n} \times B + \hat{n} \times C$ produces the same result, i.e. the vector $(B + C)''$. Therefore

$$\hat{n} \times (B + C) = \hat{n} \times B + \hat{n} \times C$$

Now let

$$A = c\hat{n} \quad (8.27)$$

We will then have

$$\frac{1}{c}A \times (B + C) = \frac{1}{c}A \times B + \frac{1}{c}A \times C$$

ending up with the original formula of

$$A \times (B + C) = A \times B + A \times C$$

⁷Cross product, Wikiversity



Eq. 8.26 follows from Eq 8.25 combined with the corollary of

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \quad (8.28)$$

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) \times \mathbf{C} &= -[\mathbf{C} \times (\mathbf{A} + \mathbf{B})] \\ &= -(\mathbf{C} \times \mathbf{A} + \mathbf{C} \times \mathbf{B}) \\ &= -\mathbf{C} \times \mathbf{A} - \mathbf{C} \times \mathbf{B} \\ &= \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C} \end{aligned}$$



We continue with our task of multiplying out the [cross product of the vectors](#) using Eq. 8.24, 8.25, and 8.26. By substituting $b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$ with $\mathbf{B}$:

$$\mathbf{A} \times \mathbf{B} = (a_1\hat{i} + a_2\hat{j} + a_3\hat{k}) \times (b_1\hat{i} + b_2\hat{j} + b_3\hat{k}) \quad (8.29)$$

$$= (a_1\hat{i} + a_2\hat{j} + a_3\hat{k}) \times \mathbf{B} \quad (8.30)$$

$$= a_1\hat{i} \times \mathbf{B} + a_2\hat{j} \times \mathbf{B} + a_3\hat{k} \times \mathbf{B} \quad (8.31)$$

$$= a_1\hat{i} \times (b_1\hat{i} + b_2\hat{j} + b_3\hat{k}) + a_2\hat{j} \times (b_1\hat{i} + b_2\hat{j} + b_3\hat{k}) + a_3\hat{k} \times (b_1\hat{i} + b_2\hat{j} + b_3\hat{k}) \quad (8.32)$$

$$= a_1b_1\hat{i} \times \hat{i} + a_1b_2\hat{i} \times \hat{j} + a_1b_3\hat{i} \times \hat{k} + a_2b_1\hat{j} \times \hat{i} + a_2b_2\hat{j} \times \hat{j} + a_2b_3\hat{j} \times \hat{k} + a_3b_1\hat{k} \times \hat{i} + a_3b_2\hat{k} \times \hat{j} + a_3b_3\hat{k} \times \hat{k} \quad (8.33)$$

With the following corollaries,

$$\hat{i} \times \hat{i} = 0 \quad (8.34)$$

$$\hat{j} \times \hat{j} = 0 \quad (8.35)$$

$$\hat{k} \times \hat{k} = 0 \quad (8.36)$$

$$\hat{i} \times \hat{j} = -\hat{j} \times \hat{i} = \hat{k} \quad (8.37)$$

$$\hat{j} \times \hat{k} = -\hat{k} \times \hat{j} = \hat{i} \quad (8.38)$$

$$\hat{k} \times \hat{i} = -\hat{i} \times \hat{k} = \hat{j} \quad (8.39)$$

$$(8.40)$$

Eq. 8.33 simplifies down to

$$\mathbf{A} \times \mathbf{B} = a_1b_2\hat{k} - a_1b_3\hat{j} - a_2b_1\hat{k} + a_2b_3\hat{i} + a_3b_1\hat{j} - a_3b_2\hat{i} \quad (8.41)$$

$$= (a_2b_3 - a_3b_2)\hat{i} - (a_1b_3 - a_3b_1)\hat{j} + (a_1b_2 - a_2b_1)\hat{k} \quad (8.42)$$

We recall that a determinant of order 2 is defined by

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1 \quad (8.43)$$

A determinant of order 3 can be defined in terms of determinants of order 2 as

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \quad (8.44)$$

Eq. 8.42 is equivalent to

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \hat{\mathbf{i}} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \hat{\mathbf{j}} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \hat{\mathbf{k}} \quad (8.45)$$

$$= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \quad (8.46)$$

It should be noted that formula 8.46 is by no means a definition of cross product, because obtaining it must assume the [distributivity law](#)^a. We should view 8.46 simply as a convenient tool for making calculations

In addition, Definition like 28 that avoid dependence on explicit representations of vectors in terms of any particular coordinate system are called *invariant* or *coordinate-free*. 8.46 doesn't preserve such invariant because it assumes a Cartesian space^b

^aThe determinant pre-assumes the distributivity of cross product

^bCross product, Wikiversity

Curl

⁸ From Eq. 8.46 we construct the curl:

$$\nabla \times \mathbf{v} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_x & v_y & v_z \end{vmatrix} \quad (8.47)$$

$$= \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \hat{\mathbf{z}} \quad (8.48)$$

The Geometrical Interpretation of the Divergence

The curl is a measure of how much the vector $\mathbf{v}$ "curls around" the point in question

⁸Introduction to Electrodynamics by Griffiths, 3rd, p. 19

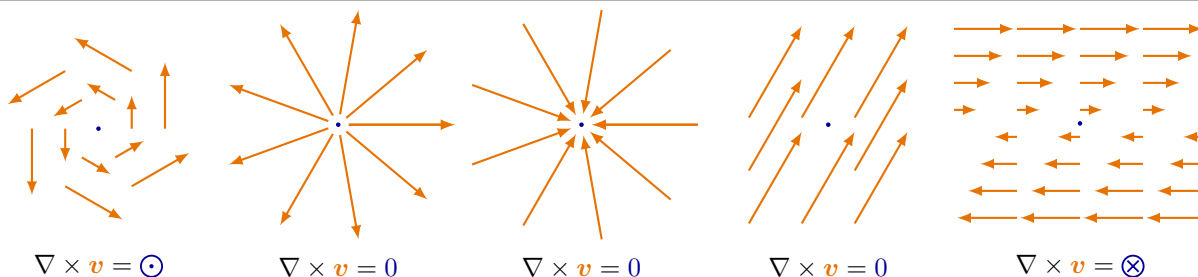


Figure 8.4: Curl of vector fields

8.1.4 Line Integrals

⁹ Throughout our discussion we assume that the functions under discussion have all the continuity and differentiability properties that are needed in any given situation

If a charge is moving in an electromagnetic field with a constant force $\mathbf{F}$ (constant in both direction and magnitude), then we know that the work done by this force is the product of the component of $\mathbf{F}$ in the direction of motion and the distance the charged particle moves, i.e.

$$W = \mathbf{F} \cdot \Delta \mathbf{R} \quad (8.49)$$

where $\mathbf{R}$ is the vector from the initial position of the particle to its final position. Now suppose that $\mathbf{F}$ is not constant, but instead is a vector function that varies from point to point throughout a certain region of the plane, say

$$\mathbf{F} = \mathbf{F}(x, y) = M(x, y)\hat{\mathbf{i}} + N(x, y)\hat{\mathbf{j}} \quad (8.50)$$

The vector-value function 8.50 is usually called *force field*. More generally, a *vector field* in the plane is any vector-valued function that associates a vector with each point (x, y) in a certain plane region R . In this context a function whose values are numbers (scalars) is called a *scalar field*. Every scalar field $f(x, y)$ gives rise to a corresponding vector field

$$\nabla f(x, y) = \frac{\partial f}{\partial x}\hat{\mathbf{i}} + \frac{\partial f}{\partial y}\hat{\mathbf{j}} \quad (8.51)$$

This is called the *gradient field* of f . Some vector fields are gradient fields, but most are not. Those gradient fields, however, are of special importance

Suppose also that this variable force pushes the charged particle along a smooth curve C with a parametric equations

$$x = x(t) \quad \text{and} \quad y = y(t), \quad t_1 \leq t \leq t_2 \quad (8.52)$$

The work done by this force is denoted by

$$\int_C \mathbf{F} \cdot d\mathbf{R} \quad \text{or} \quad \int_C M(x, y)dx + N(x, y)dy \quad (8.53)$$

⁹Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 751



This is called the **line integral**

We approximate the curve by a polygonal path as shown in Fig. 8.5. That is, choose points $P_0 = A, P_1, P_2, \dots, P_{n-1}, P_n = B$ along C in this order; let $\mathbf{R}_k$ be the position vector of P_k and define the n incremental vectors by $\Delta \mathbf{R}_k = \mathbf{R}_{k+1} - \mathbf{R}_k$, where $k = 0, 1, \dots, n-1$

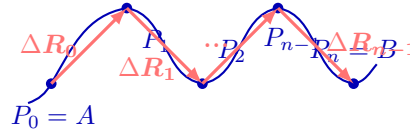


Figure 8.5: Approximating line integral

If we denote by $\mathbf{F}_k$ the value of the vector function $\mathbf{F}$ at P_k and form the sum

$$\sum_{k=0}^{n-1} \mathbf{F}_k \cdot \Delta \mathbf{R}_k \quad (8.54)$$

then the *line integral of $\mathbf{F}$ along C* is defined to be the limit of sums

$$\int_C \mathbf{F} \cdot d\mathbf{R} = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \mathbf{F}_k \cdot \Delta \mathbf{R}_k \quad (8.55)$$

^a It will often be necessary to consider situations in which the path of integration C is a *closed* curve. In this case a line integral is usually written with a small circle on the integral sign, as in

$$\oint_C \mathbf{F} \cdot d\mathbf{R}$$

^aCalculus with Analytic Geometry by George F. Simmons, 2nd, p. 757

The Fundamental Theorem of Calculus

¹⁰We intuitively know that the **definite integral** of a continuous function is the limit of approximating sums, i.e.

$$\int_a^b f(x) dx = \lim_{\max \Delta x_k \rightarrow 0} \sum_{k=1}^n f(x_k^*) \Delta x_k \quad (8.56)$$

The definite integral which is defined here is often called the **Riemann integral**, in honor of the 19th-century German mathematician who was the first to give a careful discussion of integrals of discontinuous functions¹¹

Eq. 8.56 immediately proves the following properties of definite integral¹²:

¹⁰Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 206

¹¹Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 202

¹²Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 214

$$\int_a^b f(x)dx = - \int_b^a f(x)dx \quad (8.57)$$

$$\int_a^b c f(x)dx = c \int_a^b f(x)dx \quad (8.58)$$

$$\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx \quad (8.59)$$

$$\text{if } f(x) \leq g(x) \text{ on } [a, b], \quad \text{then} \quad \int_a^b f(x)dx \leq \int_a^b g(x)dx \quad (8.60)$$

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \quad (8.61)$$

This definition works for integrals of rather *simple* functions such as

$$\int_0^b x dx = \frac{b^2}{2}$$

but not for such complicate integrals as

$$\int_0^1 \frac{x^4}{(7+x^5)^{\frac{1}{3}}} dx$$

The Fundamental Theorem of Calculus links the concept of differentiating a function with the concept of integrating a function, showing that these two operations are essentially inverse of one another. Before the discover of this theorem, however, it was not recognized that these two operations were related. Ancient Greek mathematicians knew how to compute area via infinitesimals, an operation that we would now call integration. The origins of differentiation likewise predate the fundamental theorem of calculus by hundreds of years. **The historical relevance of the fundamental theorem of calculus is not the ability to calculate these operations, but the realization that the two seemly distinct operations are actually closely related**

Theorem 5 (The First Fundamental Theorem of Calculus)

Let f be a continuous real-valued function defined on a closed interval $[a, b]$. Let F be the function defined, for all x in $[a, b]$, by

$$F(x) = \int_a^x f(t)dt \quad (8.62)$$

Then F is uniformly continuous on $[a, b]$ and differentiable on the open interval (a, b) and

$$F'(x) = f(x) \quad (8.63)$$

for all x in (a, b) so F is an antiderivative of f

Proof of Theorem 5 ¹³ For a given function f , define the function $F(x)$ as

$$F(x) = \int_a^x f(t)dt$$

For any two number x_1 and $x_1 + \Delta x$ in $[a, b]$, we have

$$F(x_1 + \Delta x) - F(x_1) = \int_a^{x_1 + \Delta x} f(t)dt - \int_a^{x_1} f(t)dt \quad (8.64)$$

$$= \int_a^{x_1 + \Delta x} f(t)dt + \int_{x_1}^a f(t)dt \text{ (by Eq. 8.57)} \quad (8.65)$$

$$= \int_{x_1}^{x_1 + \Delta x} f(t)dt \text{ (by Eq. 8.61)} \quad (8.66)$$

To be able to go any further, we shall introduce the *Mean value theorem for definite integrals*

Theorem 6 (Mean Value Theorem for Definite Integrals)

If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and g is an integrable function that does not change sign on $[a, b]$, then there exists c in (a, b) such that

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx \quad (8.67)$$

Proof suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and g is a *non-negative* integrable function on $[a, b]$. By the [extreme value theorem](#), there exists m and M such that for each x in $[a, b]$, $m \leq f(x) \leq M$ and $f[a, b] = [m, M]$. Since g is *non-negative*,

$$m \int_a^b g(x)dx \leq \int_a^b f(x)g(x)dx \leq M \int_a^b g(x)dx \quad (8.68)$$

If $g(x) = 0$,

$$0 \leq \int_a^b f(x)g(x)dx \leq 0 \quad (8.69)$$

$$\int_a^b f(x)g(x)dx = 0 \quad (8.70)$$

so for any $c \in [a, b]$

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx = 0 \quad (8.71)$$

If $g(x) \neq 0$,

$$m \leq \frac{1}{\int_a^b g(x)dx} \int_a^b f(x)g(x)dx \leq M \quad (8.72)$$

By the [intermediate value theorem](#), f attains every value of the interval $[m, M]$ so for some c in $[a, b]$:

¹³Proof of the first part, Fundamental theorem of calculus, Wikipedia

$$f(c) = \frac{1}{\int_a^b g(x)dx} \int_a^b f(x)g(x)dx \quad (8.73)$$

that is,

$$f(c) \int_a^b g(x)dx = \int_a^b f(x)g(x)dx \quad (8.74)$$

Finally, if g is negative on $[a, b]$, we can still get the same result.



By having $g(x) = 1$ in [mean value theorem for integration](#), we have

$$\int_a^b f(x)dx = f(c)(b - a) \quad (8.75)$$

We can now move on with Eq. 8.66. There exists a real number $c \in [x_1, x_1 + \Delta x]$ such that

$$F(x_1 + \Delta x) - F(x_1) = \int_{x_1}^{x_1 + \Delta x} f(t)dt = f(c)\Delta x \quad (8.76)$$

so

$$\frac{F(x_1 + \Delta x) - F(x_1)}{\Delta x} = f(c) \quad (8.77)$$

$$\lim_{\Delta x \rightarrow 0} \frac{F(x_1 + \Delta x) - F(x_1)}{\Delta x} = \lim_{\Delta x \rightarrow 0} f(c) \quad (8.78)$$

that is,

$$F'(x_1) = f(x_1) \quad (8.79)$$

Corollary 1: Corollary of Theorem 5

If f is a real-valued continuous function on $[a, b]$ and F an antiderivative of f in $[a, b]$, then

$$\int_a^b f(t)dt = F(b) - F(a) \quad (8.80)$$

Note that the corollary assumes continuity on the whole interval. This result is strengthened slightly in Theorem 7

Proof of Corollary 1 ¹⁴

¹⁵ Suppose F is an antiderivative of f , which is continuous on $[a, b]$. Let $G(x)$ also be an antiderivative of f :

¹⁴

¹⁵[Proof of the corollary, Fundamental theorem of calculus](#), Wikipedia

$$G(x) = \int_a^x f(t)dt \quad (8.81)$$

By Theorem 6, we have

$$F'(x) = f(x) \quad (8.82)$$

$$G'(x) = f(x) \quad (8.83)$$

It is easy to see then

$$F - G = c \quad (8.84)$$

where c is a constant. That is, there is a number c such that

$$G(x) = F(x) + c \quad (8.85)$$

for all $x \in [a, b]$

Let $x = a$, we have

$$F(a) + c = G(a) = \int_a^a f(t)dt = 0 \quad (8.86)$$

which means

$$c = -F(a) \quad (8.87)$$

or

$$G(x) = F(x) - F(a) \quad (8.88)$$

Therefore

$$\int_a^b f(t)dt = G(b) = F(b) - F(a) \quad (8.89)$$



Theorem 7 (The Second Fundamental Theorem of Calculus: Newton-Leibniz Theorem)

Let f be a real-valued function on a closed interval $[a, b]$ and F a continuous function on $[a, b]$ which is an antiderivative of f in (a, b) :

$$F'(x) = f(x) \quad (8.90)$$

If f is **Riemann integrable** on $[a, b]$, then

$$\int_a^b f(x)dx = F(b) - F(a) \quad (8.91)$$

Here we **do not assume** f is continuous

Proof of Theorem 7 ¹⁶ Let f be **Riemann integrable** on $[a, b]$ and let f admit an antiderivative F on (a, b) such that F is continuous on $[a, b]$.

Begin with the quantity $F(b) - F(a)$. Let there be numbers $x_0, \dots, x_n$ such that

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b \quad (8.92)$$

It follows that

$$F(b) - F(a) = F(x_n) - F(x_0) \quad (8.93)$$

Now, we add each $F(x_i)$ along with its additive inverse, so that the resulting quantity is equal:

$$F(b) - F(a) \quad (8.94)$$

$$= F(x_n) + [-F(x_{n-1}) + F(x_{n-1})] + \dots + [-F(x_1) + F(x_1)] - F(x_0) \quad (8.95)$$

$$= [F(x_n) - F(x_{n-1})] + [F(x_{n-1}) - F(x_{n-2})] + \dots + [F(x_2) - F(x_1)] + [F(x_1) - F(x_0)] \quad (8.96)$$

$$= \sum_{i=1}^n [F(x_i) - F(x_{i-1})] \quad (8.97)$$

Since F is differentiable on interval (a, b) and continuous on $[a, b]$, it is also differentiable on each interval (x_{i-1}, x_i) and continuous on each interval $[x_{i-1}, x_i]$. According to the **Mean Value Theorem**, for each i there exists a c_i in (x_{i-1}, x_i) such that

$$F(x_i) - F(x_{i-1}) = F'(c_i)(x_i - x_{i-1}) \quad (8.98)$$

Plugging this equation into Eq. 8.97, we get

$$F(b) - F(a) = \sum_{i=1}^n [F'(c_i)(x_i - x_{i-1})] = \sum_{i=1}^n [f(c_i)\Delta x_i] \quad (8.99)$$

We are describing the area of a rectangle, with the width times the height, and we are adding the areas together. Each rectangle, by virtue of the Mean Value Theorem, describes an approximation of the curve section it is drawn over. Also Δx_i need not be the same for all values of i . What we will do is to approximate the curve with n rectangles. As the widths of the partitions get smaller and n increases, we get closer to the actual areas

Since f is Riemann integrable:

¹⁶Proof of the second part, Fundamental theorem of calculus, Wikipedia

$$\lim_{\|\Delta x_i\| \rightarrow 0} F(b) - F(a) = \lim_{\|\Delta x_i\| \rightarrow 0} \sum_{i=1}^n [f(c_i) \Delta x_i] \quad (8.100)$$

$$F(b) - F(a) = \lim_{\|\Delta x_i\| \rightarrow 0} \sum_{i=1}^n [f(c_i) \Delta x_i] \quad (8.101)$$

$$F(b) - F(a) = \int_a^b f(x) dx \quad (8.102)$$



Independence of Path & Conservative Fields

¹⁷ We have already known that the vector field is the gradient of a scalar field. The Fundamental Theorem of Calculus of single-variable in the case of 2 variables can be derived as follows:

$$\int_C \mathbf{F} \cdot d\mathbf{R} = \int_C \nabla f \cdot d\mathbf{R} \quad (8.103)$$

$$= \int_a^b \left(\nabla f \cdot \frac{d\mathbf{R}}{dt} \right) dt \quad (8.104)$$

$$= \int_a^b \left[\left(\frac{\partial f}{\partial x} \hat{x} + \frac{\partial f}{\partial y} \hat{y} \right) \cdot \frac{d(x\hat{x} + y\hat{y})}{dt} \right] dt \quad (8.105)$$

$$= \int_a^b \left(\frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} \right) dt \quad (8.106)$$

$$= \int_a^b \frac{d}{dt} f(x, y) dt \quad (8.107)$$

$$= f(b) - f(a) \quad (8.108)$$

Theorem 8 (Fundamental Theorem of Calculus for Line Integrals)

If a vector field $\mathbf{F}$ is the gradient of some scalar field f in a region R , so that $\mathbf{F} = \nabla f$ in R , and if C is any piecewise smooth curve in R with initial and final points A and B , then

$$\int_C \mathbf{F} \cdot d\mathbf{R} = f(B) - f(A) \quad (8.109)$$

The right side of the Eq. 8.109 depends only on the points A and B and not at all on the path C that joins them. The line integral on the left side of Eq. 8.109 therefore has the same value for all paths C from A to B . This can be expressed by saying that the **line integral of a gradient field is independent of the path**



. Next, it is clear from Eq. 8.109 that **if C is a closed path, then**



$$\oint_C \mathbf{F} \cdot d\mathbf{R} = 0 \quad (8.110)$$

¹⁷ Calculus with Analytic Geometry by George F. Simmons, 2nd, p. 758

These arguments show that

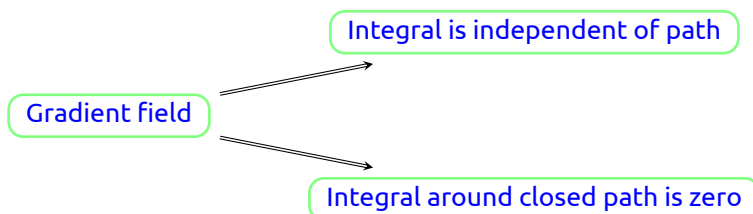
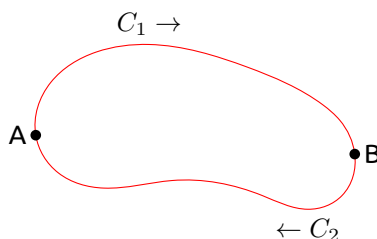


Figure 8.6: The symbol $\Rightarrow$ means “implies”

We are going to show that these 3 properties are equivalent, in the sense that each implies the other two.

Suppose that the line integral of the vector field $\mathbf{F}$ is independent of the path. We shall prove that the line integral of the vector field $\mathbf{F}$ around a closed path is zero.

We examine the figure below, in which two points A and B are chosen on the closed path C . These points divide C into paths C_1 , from A to B , and C_2 , from B to A .



Since both C_1 and $-C_2$ are paths from A to B , the assumption of independence of path implies that

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{R} = \underbrace{\int_{-C_2} \mathbf{F} \cdot d\mathbf{R}}_{\text{Eq. 8.57}} = - \int_{C_2} \mathbf{F} \cdot d\mathbf{R} \quad (8.111)$$

It then follows that

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{R} + \int_{C_2} \mathbf{F} \cdot d\mathbf{R} = \oint_C \mathbf{F} \cdot d\mathbf{R} = 0 \quad (8.112)$$



Then we can easily reverse this argument to show that the integral from A to B is independent of the path

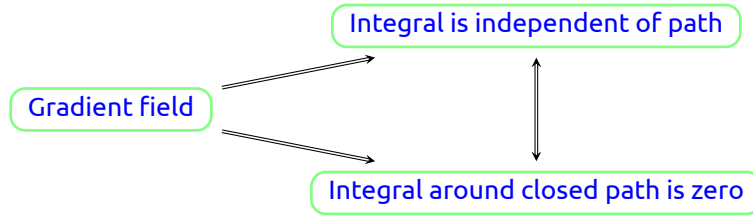
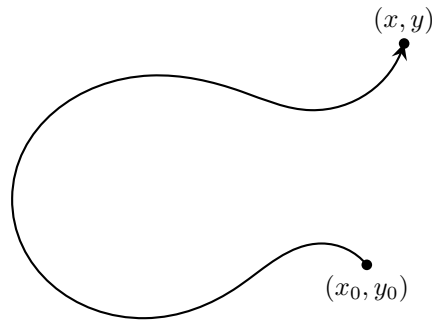


Figure 8.7: The updated Fig. 8.6

To complete the proof of the equivalence of the 3 properties, it suffices to show that if F is a vector field whose line integral is independent of path, then $F = \nabla f$ for some scalar field f . To do this, we choose a fixed point (x_0, y_0) and an arbitrary point (x, y) . Given any path from C from (x_0, y_0) to (x, y) , because we assume the hypothesis of independence of path, we can unambiguously define the function $f(x, y)$ by

$$f(x, y) = \int_C F \cdot dR = \int_{(x_0, y_0)}^{(x, y)} F \cdot dR \quad (8.113)$$



F has the usual form of $F = M(x, y)\mathbf{i} + N(x, y)\mathbf{j}$ so that

$$f(x, y) = \int_{(x_0, y_0)}^{(x, y)} M(x, y)dx + N(x, y)dy \quad (8.114)$$

Let's first hold y fixed and move along the x -direction by Δx . By [The Fundamental Theorem of Calculus](#), [Eq 8.114](#) implies

$$f(x + \Delta x, y) - f(x, y) = \int_{(x, y)}^{(x + \Delta x, y)} M dx \quad (8.115)$$

so

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} \int_{(x, y)}^{(x + \Delta x, y)} M dx = M \quad (8.116)$$

Similarly

$$\frac{\partial f}{\partial y} = N \quad (8.117)$$

so

$$\frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} = \left(\frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} \right) f = \nabla f = M \mathbf{i} + N \mathbf{j} = \mathbf{F} \quad (8.118)$$

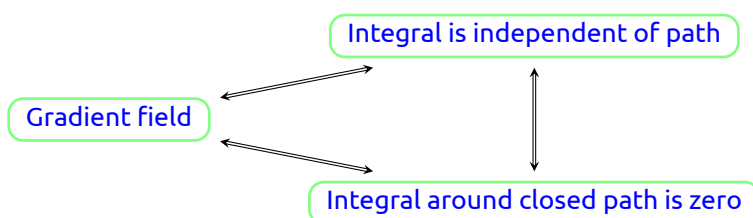


Figure 8.8: The updated Fig. 8.7

A force that has one of 3 properties above is called *conservative*.

8.1.5 Green's Theorem

Proof of Green's Theorem

Green did not actually derive the form of "Green's theorem" which appears in the [form we see today](#); rather, he derived a form of the "divergence theorem", which appears on [pages 10 - 12](#) of his *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*. A proof of the theorem was finally provided in 1851 by [Bernhard Riemann](#) in his inaugural dissertation: Bernhard Riemann (1851) *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse* (Basis for a general theory of functions of a variable complex quantity), (Göttingen, (Germany): Adalbert Rente, 1867); see pages 8 - 9.^a

Having formulating the Riemannian Geometry, Riemann laid the foundations of the mathematics of **General Relativity**^b

A [separate part](#) has been dedicated to Riemannian Geometry which shall include the rigorous proof of Green's theorem

Given that, the purpose of this section is solely for revealing the nature of the link between line integrals and double integrals^c

^aGreen's theorem, Wikipedia

^bBernhard Riemann, Wikipedia

^cCalculus with Analytic Geometry by George F. Simmons, 2nd, p. 764

Let's look at a rectangular path like the one shown below



Draw rectangular box

$$\oint_C Mdx + Ndy = \iint_R \left[\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y} \right] dA \quad (8.119)$$

Chapter 9

Maxwell's Equations

Does the Electromagnetic Field *physically* exist?

"There exists a model of the universe which includes a field known as the Electromagnetic Field. This model does a remarkably good job of predicting the observations we make in the world. It does so good at making such predictions that it is often phrased as 'existing in the world'"^a

^a<https://philosophy.stackexchange.com/a/28010>

9.1 Gauss's Law for Electric Fields

There are two kinds of electric field:

1. the *electrostatic* field produced by electric charge
2. the *induced* electric field produced by a changing magnetic field

Gauss's law for electric fields deals with the electrostatic field. It relates the spatial behavior of the electrostatic field to the charge distribution that produces it. The integral form is generally written like this:

9.1.1 The Electric Field

To understand Gauss's law, we first have to understand the concept of the electric field. In some physics and engineering books, no direct definition of the electric field is given; instead we see a statement that an electric field is "said to exist" in any region in which electrical forces act. But what exactly is an electric field? This question has deep philosophical significance and it is not easy to answer¹. It traces all the way back to a person named Michael Faraday who is believed to discover the concept of an electric field.

While Faraday did not develop a complete mathematical description of the electric field, his concept laid the groundwork for later scientists to quantify the electric field using mathematical equations. The book by Michael Faraday that introduced the concept of the electric field is called *Experimental Researches in Electricity*, particularly in its **Eleventh Series**, a chapter of the book.

Faraday conducted an experiment using two long coils. Let's call them coils *A* and *B*. Coil *A* was connected to a battery source while coil *B* was connected to a galvanometer, which measures the current in *B*. He discovered that, when coils were long enough and battery source was strong enough, the galvanometer

¹Fleisch, *A Student's Guide to Maxwell's Equations*, p. 3

signaled a slight current passing through coil B at the moment of connecting coil A the battery. In addition, when coil A and the battery was disconnected, the same amount of current in coil B was detected again, but this time the current was in the opposite direction.²

² *Experimental Researches In Electricity, Vol. 1, Faraday, Michael, 6 - 11*

Chapter 10

Euclid's Elements

1. *Ἡιτήσθω ἀπὸ παντὸς σημείου ἐπὶ πᾶν σημεῖον εὐθείαν γραμμὴν ἀγαγεῖν.*
2. *Καὶ πεπερασμένην εὐθεῖαν κατὰ τὸ συνεχὲς ἐπ' εὐθείας ἐκβαλεῖν.*
3. *Καὶ παντὶ κέντρῳ καὶ διαστήματι κύκλον γράφεσθαι.*
4. *Καὶ πάσας τὰς ὀρθὰς γωνίας ἴσας ἀλλήλαις εἶναι.*
5. *Καὶ ἐὰν εἰς δύο εὐθείας εὐθεῖα ἐμπίπτουσα τὰς ἐντὸς καὶ ἐπὶ τὰ αὐτὰ μέρη γωνίας δύο ὀρθῶν ἐλάσσονας ποιῇ, ἐκβαλλομένης τὰς δύο εὐθείας ἐπ' ἄπειρον συμπίπτειν, ἐφ' ἃ μέρη εἰσὶν αἱ τῶν δύο ὀρθῶν ἐλάσσονες.*

Part V

Quantum Mechanics

Chapter 11

Introduction

11.1 The Schrödinger Equation

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + V\Psi \quad (11.1)$$

where i is the square root of -1 and $\hbar$ is the Planck's constant:

$$\hbar = \frac{h}{2\pi} = 1.054573 \times 10^{-34} Js \quad (11.2)$$

11.1.1 Statistical Interpretation

Born's **statistical interpretation** says that $|\Psi(x, t)|^2$ gives the *probability* of finding the particle at point x at time t :

$$\int_a^b |\Psi(x, t)|^2 dx = \left\{ \begin{array}{l} \text{probability of finding the particle} \\ \text{between } a \text{ and } b, \text{ at time } t. \end{array} \right\} \quad (11.3)$$

The statistical interpretation introduces a kind of **indeterminacy** into quantum mechanics, i.e. we cannot predict with certainty the outcome of a simple experiment to measure its position - all quantum mechanics has to offer is *statistical* information about the *possible* results. A particle simply does not have a precise position prior to measurement

11.2 Probability

Because of the statistical interpretation, probability plays a central role in quantum mechanics, in which the *average* is usually the quantity of interests. In this context, the average has come to be called the **expectation value**.

Expectation value is a misleading term, since it suggests that this is the outcome we would be most likely to get if we made a single measurement (*what would be the most probable value, not the average value*)

For discrete variables, the average value of some *function* of j is given by

$$\langle f(j) \rangle = \sum_{j=0}^{\infty} f(j) P(j) \quad (11.4)$$

The amount of “spread” in a distribution, with respect to the average is described by a quantity called **variance** of the distribution:

$$\sigma^2 \equiv \langle (\Delta j)^2 \rangle \quad (11.5)$$

where σ itself is called the **standard deviation**.

A much faster way to compute σ than by direct application of Equation 11.5 is using a useful theorem on variances:

$$\begin{aligned} \sigma^2 &= \langle (\Delta j)^2 \rangle = \sum (\Delta j)^2 P(j) = \sum (j - \langle j \rangle)^2 P(j) \\ &= \sum (j^2 - 2j\langle j \rangle + \langle j \rangle^2) P(j) \\ &= \sum j^2 P(j) - 2\langle j \rangle \sum j P(j) + \langle j \rangle^2 \sum P(j) \\ &= \langle j^2 \rangle - 2\langle j \rangle \langle j \rangle + \langle j \rangle^2 = \langle j^2 \rangle - \langle j \rangle^2. \end{aligned}$$

Taking the square root, the standard deviation itself can be written as

$$\sigma = \sqrt{\langle j^2 \rangle - \langle j \rangle^2} \quad (11.6)$$

For continuous variables, if The probability of x lying between a and b (a *finite* interval) is given by the integral of a **probability density function**, $\rho(x)$:

$$P_{ab} = \int_a^b \rho(x) dx \quad (11.7)$$

and the rules we deduced for discrete distribution translate in the obvious way:

$$\int_{-\infty}^{+\infty} \rho(x) dx = 1, \quad (11.8)$$

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \rho(x) dx, \quad (11.9)$$

$$\langle f(x) \rangle = \int_{-\infty}^{+\infty} f(x) \rho(x) dx, \quad (11.10)$$

$$\sigma^2 \equiv \langle (\Delta x)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2. \quad (11.11)$$

11.2.1 Normalization

We return now to the statistical interpretation of the wave function (Equation 11.3), which says that $|\Psi(x, t)|^2$ is the probability density for finding the particle at point x at time t . It follows from Equation 11.8 that

$$\int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = 1 \quad (11.12)$$

How do we make sure a wave function satisfies both equations 11.1 and 11.12?

Suppose $\Psi(x, t)$ is a solution for equation 11.1 and we have

$$\int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = C \quad (11.13)$$

where C is some finite constant.

We shall see in [in a bit](#), solutions in the form of equation 11.13 are guaranteed to exist in Hilbert Space, where wave functions reside in quantum mechanics.

Then we can make the right side of the equation 1 with

$$\int_{-\infty}^{+\infty} \left| \frac{1}{C} \Psi(x, t) \right|^2 dx = 1 \quad (11.14)$$

A glance at equation 11.1 reveals that if $\Psi(x, t)$ is a solution, so too is $\frac{1}{C} \Psi(x, t)$. What we must do, then, is to pick this undetermined multiplicative factor so as to ensure that equation 11.12 is satisfied. This process is called **normalizing** the wave function.

The fact that equation 11.12 is integrating over x *only* implies normalization of the wave function is **preserved** and we can prove this. To begin with,

$$\frac{d}{dt} \int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = \int_{-\infty}^{+\infty} \frac{\partial}{\partial t} |\Psi(x, t)|^2 dx \quad (11.15)$$

By the product rule,

$$\frac{\partial}{\partial t} |\Psi|^2 = \frac{\partial}{\partial t} (\Psi^* \Psi) = \Psi^* \frac{\partial \Psi}{\partial t} + \frac{\partial \Psi^*}{\partial t} \Psi \quad (11.16)$$

Now the Schrödinger equation says that

$$\frac{\partial \Psi}{\partial t} = \frac{i\hbar}{2m} \frac{\partial^2 \Psi}{\partial x^2} - \frac{i}{\hbar} V \Psi \quad (11.17)$$

and hence also

$$\frac{\partial \Psi^*}{\partial t} = -\frac{i\hbar}{2m} \frac{\partial^2 \Psi^*}{\partial x^2} + \frac{i}{\hbar} V \Psi^* \quad (11.18)$$

so

$$\frac{\partial}{\partial t} |\Psi|^2 = \frac{i\hbar}{2m} \left(\Psi^* \frac{\partial^2 \Psi}{\partial x^2} - \frac{\partial^2 \Psi^*}{\partial x^2} \Psi \right) = \frac{\partial}{\partial x} \left[\frac{i\hbar}{2m} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \frac{\partial \Psi^*}{\partial x} \Psi \right) \right] \quad (11.19)$$

The integral in Equation 11.15 can now be evaluated explicitly:

$$\frac{d}{dt} \int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = \frac{i\hbar}{2m} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \frac{\partial \Psi^*}{\partial x} \Psi \right) \Big|_{-\infty}^{+\infty} \quad (11.20)$$

But $\Psi(x, t)$ must go to zero as x goes to $(\pm)$ infinity—otherwise the wave function would not be normalizable.¹⁵ It follows that

$$\frac{d}{dt} \int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = 0 \quad (11.21)$$

and hence that the integral is *constant* (independent of time); if Ψ is normalized at $t = 0$, it textitstays normalized for all future time.



11.2.2 Momentum

Let's look at how fast a particle moves by taking the derivative of particle's position with respect to t . According to the [statistical interpretation](#) and equation 11.9, for a particle in state Ψ the expectation value of x is

$$\langle x \rangle = \int_{-\infty}^{+\infty} x |\Psi(x, t)|^2 dx \quad (11.22)$$

What exactly does equation 11.22 mean?

It emphatically does *not* mean that if we measure the position of one particle over and over again, $\int x |\Psi|^2 dx$ is the average of the results we will get.

On the contrary: The first measurement (whose outcome is indeterminate) will collapse the wave function to a spike at the value actually obtained, and the subsequent measurements (if they are performed quickly) will simply repeat that result.

Rather, $\langle x \rangle$ is the average of measurements performed on particles *all in the state* Ψ , which means that either we must

1. find some way of returning the particle to its original state after each measurement, or
2. prepare a whole **ensemble** of particles, each in the same state Ψ

and measuring the positions of all of them would give us the average of *these* results.

In short, *the expectation value is the average of measurements on an ensemble of identically-prepared systems*, not the average of repeated measurements on one and the same systems.

11.3 Formalism

The collection of *all* functions of x constitutes a vector space, but for our purpose it is much too large. To represent a possible physical state, the wave function Ψ must be [normalized](#). The set of all **square-integrable** functions, on a specified interval,

$$f(x) \quad \text{such that} \quad \int_a^b |f(x)|^2 dx < \infty, \quad (11.23)$$

constitutes a (much smaller) vector space.

Bibliography

Fleisch, Daniel. *A Student's Guide to Maxwell's Equations*. Cambridge University Press, 2008. isbn: 978-0-521-87761-9 (cit. on p. [73](#)).