

Originally prepared for METU - *Phys209 & Phys210*

Mathematical Methods in Physics

Soner Albayrak

MAY 4, 2024

Summary

These are the notes of the lectures prepared for the “PHYS209 Mathematical Methods in Physics I” and “PHYS210 Mathematical Methods in Physics II” courses at *Middle East Technical University*, 2023-2024 Fall & Spring Semesters —see <https://soneralbayrak.com/teaching/>. These notes are mostly based on other sources and I provided the sources in relevant places. Whenever I get the chance, I will keep updating the notes to keep it up-to-date and self-contained; there are also reminders [in blue](#) for me to add further discussion/comments.

For contact: contact@soneralbayrak.com

Preface

Remarks to the reader

I would like to make a few points crystal clear:

1. I do not update these notes on a regular basis.
2. The contents are correct to the best of my knowledge, but I have not put the extra effort to make sure that everything is book-level correct; nevertheless, I try to put as many references as possible when relevant, so please make the most of it!
3. I believe that the level of this book is appropriate for an average sophomore, but I would dare say that it should be quite useful even for graduate students of theoretical physics.
4. Lastly, I provide several links as references here and there: I agree that this is a bad practice in academia and one should instead convert them to proper references in the bibliography. Nevertheless, it is faster for me to write this way and faster for the reader to just click on them where they appear, so I'll keep this practice. My upfront apologies if the links get broken in time: hopefully a snapshot will have been available here <https://web.archive.org/> (it would be rather amusing if this site itself becomes unavailable).

^aFor instance, I provide all links in their explicit form (not like [google](#) but as <https://www.google.com/>) so that the book on an ereader without a browser or its printed version is as useful as its digital version.

About the format of the book

This book is publicly available online, and anyone can simply download and read it on their computer. Nevertheless, it is a lot easier on eyes to read a book on paper (or epaper), so I suspect many readers will simply print this book. To make this book convenient for all types of readers, I need to make a few design choices^a and the most important of it is the format of this book.

The most convenient paper that most students have easy access to is A4 (or letterpaper which is very similar in size): so if they choose to print this book, their easiest and cheapest option would be to use A4 paper. However that paper has been historically designed for typewriters which use large monospaced fonts, hence is not really appropriate for a digitally prepared book. Indeed, if you check your favorite-to-read book, you will most likely see that it uses a smaller paper size, with proportionally spaced small fonts.

What is wrong with using a large paper? It is empirically known that a document is properly legible if there are around 60-75 characters on a line: if there are more characters, it becomes harder to read and the reader may end up re-reading same line over and over again (doubling). When used with typewriter, A4 paper indeed has appropriate number of characters one a line, but as stated previously, this is not the optimal setup with the digital fonts, hence making A4 paper *too large for digital books*.

How to solve the problem that A4 is too large for a digital book? Obviously, we can use a typewriter font for which A4 is historically intended in the first place! However, such monospaced fonts (Courier being another example) are not aesthetically pleasing and do not belong to modern texts!

Of course, we can go with a modern proportionally spaced fonts but make the font size large enough such that a line has few enough characters for it to be easily legible. Although a better one than using monospaced fonts, this is still a suboptimal solution to the problem at hand...

Another solution which is somehow popular around the institutions is to use double-spacing among the lines. Indeed, regulations for master and doctorate thesis of various universities include compulsory large spacing among the lines, such as one and a half spacing or double spacing: you can also see this in my master and doctorate thesis: <https://arxiv.org/abs/1602.07676>, and <https://arxiv.org/abs/2107.13601>. Although this method can indeed prevent doubling to a degree, it is neither an aesthetic nor an efficient solution.

A somehow better solution than those listed above is to use multiple columns in the document. Indeed, this is the traditional approach in magazines and newspapers, and is immediately applicable in academic papers and manuscripts as well. However, A4 paper is not really big enough to have two columns of text with around 60-75 characters (let alone three or more columns), and although one can go with smaller font sizes to make it more legible digitally, the printed version would still be hard to read either way (proper number of small font characters, or few number of normal font characters).

Common text editors such as Microsoft Word either go with one of the solutions above or do not solve the problem at all. On the contrary, \LaTeX templates default to choose another approach: they stick to a modern font with an appropriate spacing and a single column, but they also increase the margins such that a line has proper number of characters. This is an ideal solution if the resultant text will be read digitally, however it leads to a waste of paper when printed.

In this book, we will not follow any of these design choices. Instead, we will go with the rather unorthodox *Tufte style*,^b an asymmetric allocation of the text in the paper. Indeed, the main text will be in the left of the paper, whereas we have another block of text on the right dedicated to the *sidenotes*,^c margin figures, and margin tables. We choose a rather narrow font family (*libertine*) and arrange the margins such that the main text is of 26 pica width and side text is of 14 pica width: for the 11pt and 9pt font sizes, this corresponds to roughly 66 and 44 characters of the libertine font for main and side text blocks respectively.^e Thus we have an ideally-sized main text block and acceptably-sized side text block for a modern proportionally spaced

^bSee <https://www.ctan.org/tex-archive/macros/latex/contrib/tufte-latex/>

^cIn the traditional layout, one usually uses endnotes, margin notes, or footnotes; in this paper, most “notes” will be sidenotes with occasional footnotes.^d

^dSuch as this one.

font in an A4 paper, and we do this without unnecessarily wasting the paper.^f

About the courses Phys209 & Phys210

As stated in the front page, this “book” is collection of notes prepared for the courses Phys209 & Phys210, with the syllabus provided here: <https://soneralbayrak.com/teaching>. Although the title is somewhat generic, the contents of this book is severely restricted and arranged so as to be an appropriate two one-semester-long courses for *an average sophomore at the Physics program of Middle East Technical University*. I would also like to acknowledge that the level and approach of this book is based solely on my expectations and projections for such a student, and thus it may not be really appropriate in real life; nevertheless, it is what it is.

I left several sources in the syllabus, including the textbooks of the courses on which these notes are somewhat based on. I’ll also make use of other sources; yet, any error or incorrect information on these notes are entirely my fault and not of any of these sources.

^eFor a nice discussion of these points along with the tools to compute approximate expected number of characters per line, see <https://ftp.cc.uoc.gr/mirrors/CTAN/macros/latex/contrib/memoir/memman.pdf>

^fI would like to acknowledge the following nice discussion with which I started to learn more about these *typographical* issues: <https://tex.stackexchange.com/questions/71172/why-are-default-latex-margins-so-big>.

Contents

Preface	i
Remarks to the reader	i
About the format of the book	i
About the courses Phys209 & Phys210	iii
Contents	iv
List of Tables	vi
List of Figures	vii
 I Differential Equations in Physics	 1
1 Definition and Classification of Differential Equations	2
1.1 Preliminaries: some basic terminology	2
1.1.1 Type notation and functions	2
1.1.2 Higher order functions and the derivative	3
1.1.3 Functionals and the integral	4
1.2 Differential equations	5
1.2.1 Basics	5
1.2.2 Classification	7
2 Linear equations with constant coefficients	10
2.1 Linear mappings and kernels	10
2.2 Homogeneous solutions	13
2.2.1 Basics	13
2.2.2 Repeated roots	14
2.2.3 Examples	16
2.3 Laplace transform	16
3 Linear equations with functional coefficients	19
3.1 Homogeneous solution	19
3.2 Particular solution	28
4 Systems of first order linear differential equations	31
5 Eigensystems and Sturm-Liouville theory	36
6 Beyond linear ordinary differential equations	39
6.1 Non-linear ordinary differential equations	39
6.2 Partial differential equations	39
 II Vector Calculus for Physicists	 42
7 Definition and fundamentals of Vector Spaces	43
7.1 Preliminaries: some basic terminology	43

7.1.1	Primer: groups, fields, and algebras	43
7.1.2	Properties of Linear spaces	48
7.2	Inner product spaces	49
7.3	Dual vector spaces	50
7.4	Algebras over fields	51
7.4.1	Tensor algebras	51
7.4.2	Exterior algebras	53
7.4.3	Division algebras	55
8	Differentiation in Vector Calculus	56
8.1	Vector fields and forms on Euclidean Spaces	56
8.2	Grad, div, curl, and Laplacian	58
8.2.1	Exterior derivative and hodge star operator	58
8.2.2	Invariant formulation of differentiation in vector calculus	59
8.3	Helmholtz decomposition of vector fields	62
9	Integration in Vector Calculus	63
9.1	Higher Dimensional Integration	63
9.2	Curves and Line Integrals	64
9.3	Integral theorems	64
9.3.1	Gradient, Divergence, Curl, and Greens theorems	64
9.3.2	Generalized Stokes theorem: one equation to derive them all!	64
10	Curvilinear Coordinate Transformations	66
III	Complex Analysis for Physicists	69
11	Basics of complex calculus	70
11.1	Structures of complex numbers	70
11.1.1	Complex numbers as a Euclidean vector space	70
11.1.2	Further properties of complex numbers	70
11.2	Complex differentiation	70
12	Riemann Surfaces	71
12.1	Riemann sphere	71
12.2	Analysis of elementary functions via Riemann surfaces	71
13	Analytic Structure of Multi-valued “functions”	72
13.1	Branch points, branch cuts, and singularities	72
13.2	Series expansion of complex functions	72
13.3	Analytic continuation and monodromy theorem	72
A	About <i>Wolfram Mathematica</i>	73
	Bibliography	75

List of Tables

1.1 Illustration of various differential equations 9

List of Figures

Part I

Differential Equations in Physics

Definition and Classification of Differential Equations

1.1 Preliminaries: some basic terminology

1.1.1 Type notation and functions

Consider the number 2. It is an integer, which mathematicians denote as $2 \in \mathbb{Z}$, where \mathbb{Z} denotes the set of integers.^g Computer scientists on the other hand would denote this as

$$2 :: \text{Integer} \quad (1.1)$$

where $a :: b$ reads as “ a is of type b ”. Further examples would be

$$3/2 :: \text{Rational} \quad (1.2a)$$

$$1.56 :: \text{Real} \quad (1.2b)$$

$$1 + i :: \text{Complex} \quad (1.2c)$$

Note that an object may be of multiple types. Mathematically, $3 \in \mathbb{Z}$ and $3 \in \mathbb{R}$, meaning

$$3 :: \text{Integer} \quad (1.3a)$$

$$3 :: \text{Real} \quad (1.3b)$$

A somewhat hybrid notation between mathematicians and computer scientists would be

$$3 :: \mathbb{Z} \quad (1.4a)$$

$$3 :: \mathbb{R} \quad (1.4b)$$

We shall use this notation in the rest of the book.^h

Just as we do with the explicit numbers above, we can *define* variables with explicit types; for instance,

$$x :: \mathbb{Z} \quad (1.5)$$

It is up to us to choose what we want for the type, we can even left the type unknown; for instance,

$$y :: A \quad (1.6)$$

means that y is a variable of the type A ,ⁱ where A can be anything.^j

Unlike the numbers or the variables above, the functions have an input and an output, hence their type actually reads differently.^k For instance,

$$f :: \mathbb{Z} \rightarrow \mathbb{Z} \quad (1.7)$$

denotes “a function that acts on integers and produces another integer”. An example would be

$$f = \lambda \rightarrow \lambda^2 \quad (1.8)$$

^g \mathbb{Z} is actually an integral domain.

^hI personally find this notation clearer when we use it with higher order functions such as derivatives.

ⁱThis is called a *type variable*.

^jIt could be a simple field such as \mathbb{Z} or \mathbb{N} , or it could be a more complex object such as $\mathfrak{M}_{2 \times 2}(\mathbb{C})$ which denotes two by two matrices with complex entries.

^kPhysicists tend to refer to multi-valued relations as functions as well: this is a justifiable habit as such relations can always be treated as genuine functions by appropriately restricting their domains.^lWe will stick to this convention in the rest of the book and refer all multi-valued relations (such as an arctan) as functions.

^lMathematically, a function yields a unique output for a given input, therefore so-called multi-valued “functions” are not really functions in their full analytic domain. For instance, the relation $\text{sqrt} = \lambda \rightarrow \sqrt{\lambda}$ is not a function in the complete complex plane, as $\text{sqrt}(4) = \pm 2$. One solution is to choose a *restricted domain* so that the relation actually yields a unique solution for a given input from the domain, hence making the relation a genuine function, e.g. choosing the domain \mathbb{R}^+ for sqrt . In principle, we do not need to make an arbitrary restriction: the strategy would be to analyze the *Riemann surface* of the relation, and then determine the codomain in which the relation yields a unique result; in the case of sqrt , we can state $\text{sqrt} :: \mathbb{C} \rightarrow A$ where $x \in A$ if and only if $0 \leq \arg(x) < \pi$. This means that $\text{sqrt}(re^{i\theta}) = \sqrt{r}e^{i\theta/2}$ for θ chosen in the range $0 \leq \theta < 2\pi$ with $r > 0$; hence, $\text{sqrt}(4) = 2$.

The regions of the codomain in which the multi-valued relation becomes a genuine function are called *sheets*; in the example above, we choose a principle sheet (or a first sheet) for the relation sqrt : we can move on to the *other sheets* by removing the restriction on θ . Indeed, we have on the second sheet $\text{sqrt}(4) = \text{sqrt}(4e^{i2\pi}) = \sqrt{4}e^{i\pi} = -2$, the other solution! One could go on to higher sheets to find even more solutions; in the case of sqrt , n -th sheet is actually identified with $(n-2)$ -th sheet, hence we have only two so-

which gives the integer $f(x) = x^2$ when acted on the integer x .^m Another example would be

$$g(y) = y/3 \quad (1.9)$$

for which we can write downⁿ

$$g :: \mathbb{Z} \rightarrow \mathbb{Q} \quad (1.10a)$$

$$y :: \mathbb{Z} \quad (1.10b)$$

$$g(y) :: \mathbb{Q} \quad (1.10c)$$

Of course, we can also extend our interested regime for the input y and simply state

$$g :: \mathbb{C} \rightarrow \mathbb{C} \quad (1.11a)$$

$$y :: \mathbb{C} \quad (1.11b)$$

$$g(y) :: \mathbb{C} \quad (1.11c)$$

which is still true for $g(y) = y/3$. In fact, we can even write $g :: A \rightarrow B$ if we do not care for the explicit types of input and output.^o

1.1.2 Higher order functions and the derivative

Consider the operation T of “doubling the output of a function”. If we apply this operation T to a function f , then it yields a function g such that $g(x) = 2f(x)$. For instance,

$$f = \lambda \rightarrow \lambda + 5 \quad (1.12a)$$

$$g = \lambda \rightarrow 2\lambda + 10 \quad (1.12b)$$

The question now is this: what is the type of the operation T ?

Clearly, $T = f \rightarrow g$ as it takes the function f as an input and produces the function g as the output. Thus, we can write it as

$$T :: (A \rightarrow B) \rightarrow (C \rightarrow D) \quad (1.13)$$

which means if

$$f :: A \rightarrow B \quad (1.14)$$

then

$$(g = T \cdot f) :: C \rightarrow D \quad (1.15)$$

T is called a *higher order function*: it acts on a function and produces another function.

The derivative operator is a higher order function, i.e.

$$\frac{d}{dx} :: (A \rightarrow B) \rightarrow (A \rightarrow C) \quad (1.16)$$

which means^p

$$x :: A \quad (1.17a)$$

$$f :: A \rightarrow B \quad (1.17b)$$

$$f(x) :: B \quad (1.17c)$$

$$f' :: A \rightarrow C \quad (1.17d)$$

$$f'(x) :: C \quad (1.17e)$$

lutions (as expected from a square root operation).

Somewhat more traditional approach to the Riemann surfaces is the *analysis of branch cuts*. We (1) take one of the solutions of the relation as the output (called *principal value*), (2) determine some lines on the complex plane (branch cuts), (3) impose discontinuity on the cuts such that the relation is a true function in the rest of the complex plane! With the insight from Riemann surfaces, we know that moving across such lines actually takes us from one sheet to another —previous (next) sheet if we pass the branch cut (counter)clockwise. For `sqrt`, the conventionally chosen principle value is $\sqrt{r^2} = r$ for $r \in \mathbb{R}^+$, and branch cut is the line $(-\infty, 0)$: `sqrt`(z) for any other $z \in \mathbb{C}$ can then be uniquely determined to be consistent with these; for instance `sqrt`($-1 \pm i10^{-100}$) $\sim 6 \times 10^{-17} \pm i$ —note the jump!

^mI’d like to note that there is a common misconception (especially in the physics community. $f(x)$ is *not* the function, the function is f . f acts on the input x , and produces the output $f(x)$.

ⁿIf you couldn’t remember, \mathbb{Q} denotes the set of rational numbers.

^oWe use different letters for the type variables (A and B) so that the input and output are not necessarily of the same type. On the contrary, the function $h :: A \rightarrow A$ can only produce integers when acted on integers, reals when acted on reals, and so on.

^pWe are using the common convention $f' := \frac{df}{dx}$ and $f'(x) := \frac{df}{dx}(x)$ for brevity.

For example,

$$f : \mathbb{R} \rightarrow \mathbb{C} \quad (1.18a)$$

$$f = \lambda \rightarrow \lambda^2 + 2i \quad (1.18b)$$

leads to

$$f' : \mathbb{R} \rightarrow \mathbb{R} \quad (1.19a)$$

$$f' = \lambda \rightarrow 2\lambda \quad (1.19b)$$

where the type variables in eqn. (1.16) are $A = C = \mathbb{R}$ and $B = \mathbb{C}$.

The derivatives can shrink the codomain of a function;^q in the above example, the original codomain (that of f) was \mathbb{C} whereas the new codomain (that of f') is \mathbb{R} . Nevertheless, we can always *embed* the smaller codomain into a larger one (e.g. all real numbers can be considered as complex numbers as well), hence we can always take $\frac{d}{dx} : (A \rightarrow B) \rightarrow (A \rightarrow B)$. This shows that the derivative is a higher order function that can be *repeatedly applied*; thus, we say^r

$$\frac{d^n}{dx^n} : (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (1.20a)$$

$$f : A \rightarrow B \quad (1.20b)$$

$$f^{(n)} : A \rightarrow B \quad (1.20c)$$

1.1.3 Functionals and the integral

In the previous section, we have seen that the derivative is a higher-order function, i.e. it takes a function to another function. Naturally, its inverse is also a higher-order function:^s

$$\frac{d^{-1}}{dx^{-1}} : (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (1.21a)$$

$$g : A \rightarrow B \quad (1.21b)$$

$$g^{(-1)} : A \rightarrow B \quad (1.21c)$$

where $g^{(-n)} = \left(\frac{d^{-1}}{dx^{-1}}\right) \cdot g^{(1-n)}$ with $g^{(0)} = g$, in line with the notation for derivatives. Fundamental theorem of calculus then tells us that the output $g^{(-1)}(x) : B$ can be written as

$$g^{(-1)}(x) = \int_0^x g(t)dt \quad (1.22)$$

which is compatible with $\frac{d}{dx} g^{(-1)}(x) = g(x)$.

We have shown above that the *indefinite integral* is a higher order function, but how about a definite integral? How do we determine its type?

We can start by writing down a generic definite integral:

$$\int_0^{\pi/2} \cos(x)dx = 1 \quad (1.23)$$

^qReminder: if $f = A \rightarrow B$, we call A (B) the (co)domain of f .

^rWe will use the notation such that $f^{(n)}$ is the n -th derivative of the function f .

^sIn principle, the anti-derivative can *extend* the codomain of a function, just as derivative shrinks it. We can see this via *integration constant*, which can be anything as long as it is x -independent. We put this subtlety aside as we can always extend the original codomain such that it matches the new one, hence (1.21a).

Clearly, we take a function (cos) and a range over which we do the integration (between 0 and $\pi/2$). We can always specify the integration range via the domain of the function,^t thus

$$\int :: (A \rightarrow B) \rightarrow C \quad (1.24)$$

as the integration turns the function cosine into a number 1.

Operations that turn functions into numbers are called *functionals*, and definite integration is a functional. For instance, the operation to compute the area under a curve is a functional: if we call that operation T , we then have

$$T :: (\mathbb{R} \rightarrow \mathbb{R}) \rightarrow \mathbb{R} \quad (1.25a)$$

$$f :: \mathbb{R} \rightarrow \mathbb{R} \quad (1.25b)$$

$$\left(T \cdot f = \int_{-\infty}^{\infty} f(x) dx \right) :: \mathbb{R} \quad (1.25c)$$

Note that the parentheses in the type definition is important; for instance,

$$T :: (\mathbb{R} \rightarrow \mathbb{R}) \rightarrow \mathbb{R} \quad (1.26a)$$

$$F :: \mathbb{R} \rightarrow (\mathbb{R} \rightarrow \mathbb{R}) \quad (1.26b)$$

denote different objects: T is a functional, which produces a number if given a function as input. F on the other hand produces a function when fed a number, i.e. $F(x) = f$ is a function, whose output can be written as $F(x)(y) = f(y)$. This show that we can actually interpret F as a function of two variables!^u

1.2 Differential equations

1.2.1 Basics

Very broadly, we could define any relation that contains the derivative higher order function $\frac{d}{dx}$ and an unknown function f as a differential equation. For instance,

$$\cos \left(\exp \left(\frac{d}{dx} \right) f(x) + \frac{1}{f(x)} \right) = 0 \quad (1.29)$$

is a differential equation: but it is neither real-world motivated nor easy-to-solve, so let's skip it and focus on more relevant and simpler cases.^v

The simplest differential equation is

$$x :: \mathbb{R} \quad (1.35a)$$

$$\left(\frac{d}{dx} \cdot f \right) :: A \rightarrow B \quad (1.35b)$$

$$\frac{d}{dx} \cdot f = 0 \quad (1.35c)$$

^t For instance, if we would like to do the integration from 0 to 1, we can restrict the function $f(x) :: A \rightarrow B$ to $f(x) :: \text{UnitReal} \rightarrow B$ where $x :: \text{UnitReal}$ means $x \in [0, 1]$.

^u This property can be generalized. A higher order function

$$f :: A_1 \rightarrow (A_2 \rightarrow (\dots \rightarrow (A_{n-1} \rightarrow (A_n \rightarrow B)))) \quad (1.27)$$

produces a function of $n-1$ variable once given a variable as input. That function then produces another function of $n-2$ variable once given a variable as input, and so on. Indeed, it means

$$x_i :: A_i \quad (1.28a)$$

$$f(x_1)(x_2) \dots (x_n) :: B \quad (1.28b)$$

which can easily be re-interpreted as $f(x_1, \dots, x_n) :: B$.

This concept of turning higher-order functions into functions of multiple variables (and vice versa) is called *currying*, see bla bla bla. [Put some sources here.](#)

^v You may be surprised with the expression $\exp(\frac{d}{dx})$. To understand it, let's first view the taking-the- n^{th} -power operation as a higher order function:

$$P_n :: (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (1.30)$$

and

$$(P_0 \cdot f = f) :: A \rightarrow B \quad (1.31a)$$

$$\left(P_n \cdot f = f \cdot (P_{n-1} \cdot f) \right) :: A \rightarrow B \quad (1.31b)$$

meaning

$$x :: A \quad (1.32a)$$

$$(P_n \cdot f)(x) = f(f(\dots f(x))) :: B \quad (1.32b)$$

For instance, $P_2 \cdot \cos = \lambda \rightarrow \cos(\cos(\lambda))$.

We can now define *exponentiation* as a higher order operation:

$$\exp :: (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (1.33a)$$

$$\exp = \sum_{n=0}^{\infty} \frac{1}{n!} P_n \quad (1.33b)$$

One can then immediately compute, say,

$$\exp \left(\frac{d}{dx} \right) x^3 = x^3 + 3x^2 + 3x + 1 \quad (1.34)$$

$$\exp \left(\frac{d}{dx} \right) e^{kx} = e^k e^{kx}$$

and so on.

which states that *there is an unknown function f such that “the derivative higher order function acting on it” leads to the zero function.*^w You may hope to formally solve this equation by applying $\frac{d^{-1}}{dx^{-1}}$ to the both sides and use $\frac{d^{-1}}{dx^{-1}} \cdot \frac{d}{dx} \cdot f = f$, but this actually leads to a circular argument.^x Instead, let us proceed to apply this function to a real variable and write

$$\left(\frac{d}{dx} \cdot f \right) (x) \equiv \frac{df}{dx}(x) \equiv f'(x) = 0 \quad (1.36)$$

for which one usually writes down the result as

$$f(x) = \text{constant} \quad (1.37)$$

immediately. This makes sense, as the derivative of a constant is always zero.

The next simplest example would be the following differential equation

$$\frac{d}{dx} \cdot f = f \quad (1.38)$$

for the unknown function f . Solving this equation is equivalent to answering this question: *what function is equal to its derivative?*

Even though what we know and what we try to solve for are all *functions*, the traditional way of writing down such equations is in terms of *the values of functions*; in other words, we say

$$f'(x) = f(x) \quad (1.39)$$

is the differential equation, and we are trying to find the output $f(x)$ that satisfies this. Indeed, in the rest of the notes, we will mostly stick to this more traditional form.

Let us ask the question again: what is the function that is equal to its derivative? We will provide three equivalent answer.

1. We *define* a function as solution of this equation. Indeed, most of the famous mathematical functions (Hypergeometric, Bessel, Hankel, Gegenbauer, etc.) are *defined* as solutions to various differential equations. Analogously, we define

$$\exp : \mathbb{C} \rightarrow \mathbb{C} \quad (1.40a)$$

$$\exp = x \rightarrow \exp(x) \text{ such that } \frac{d \exp(x)}{dx} = \exp(x) \quad (1.40b)$$

We call this function *exponential* and usually denote it as $\exp(x) = e^x$.^y

2. We first assume that $f(x) \neq 0$, with which we can rewrite eqn. (1.39) as

$$\frac{1}{f'(x)} = \frac{1}{f(x)} \quad (1.41)$$

By chain rule, we have

$$\frac{df(x)}{dx} \frac{dx}{df(x)} = 1 \quad (1.42)$$

^w We use the convention such that 0 can be of any type that yields the ordinary number zero ($0 :: \mathbb{C}$) as the output. In eqn. (1.35c), 0 has the type $A \rightarrow (0 :: \mathbb{C})$, which we call *the zero function*.

^x Naively applying $\frac{d^{-1}}{dx^{-1}}$ would lead to the equation $f = \frac{d^{-1}}{dx^{-1}} \cdot 0$ but this equation is not necessarily equivalent to the original one. Indeed, both $f = \frac{d^{-1}}{dx^{-1}} \cdot 0$ and $f = g + \frac{d^{-1}}{dx^{-1}} \cdot 0$ would lead to the original equation if $\frac{d}{dx} \cdot g = 0$ as well. [burada kernel kavramindan, homojen ve hetetojen denklemlerden bahset.](#)

^yBy using various numerical methods, we can compute the value of this function for arbitrary complex numbers, e.g. $e^0 = 1$, $e^1 \sim 2.72$, $e^{1+i} \sim 1.5 + 2.3i$, and so on.

hence the above equation becomes

$$\frac{dx}{df(x)} = \frac{1}{f(x)} \quad (1.43)$$

If we now replace $x = f^{-1}(y)$ where f^{-1} is the inverse of the function f ,^z we get

$$\frac{df^{-1}(y)}{dy} = \frac{1}{y} \quad (1.45)$$

By integrating this function, we get

$$f^{-1}(y) = \int \frac{dy}{y} \quad (1.46)$$

If we now *define* the function *logarithm* as the right hand side, we arrive at the solution that *the function whose derivative is equal to itself is the inverse of the logarithm function*, which we call the exponential function.^{aa}

Summary In the first approach, we *defined* the exponential function as the solution of the differential equation $f'(x) = f(x)$. We can then *derive* that its inverse (logarithmic function) can be given as the integral of $1/x$.^{ab} In the second approach, we *defined* the logarithmic function as the integral of $1/x$, and then *derived* that its inverse (exponential function) solves the differential equation. Which one we choose is purely conventional.

What did we learn? In math, we *define* many objects as our initial data, and then *derive* other quantities based on those. What we *choose to define* is purely conventional; however, we cannot afford to define too many things and still remain consistent. For instance, in the example above, we actually show that if we give two of the following three statements, the third one is already fixed by the other two: (1) *exponential and logarithm functions are inverse of each other*, (2) *exponential function is the solution of the differential equation $f'(x) = f(x)$* , and (3) *logarithm function is the integration of $1/x$* .

1.2.2 Classification

In the beginning of the section above, we defined differential equations as any relation that contains the derivative operator $\frac{d}{dx}$ and an unknown function $f(x)$.^{ac} There is nothing that stops us from generalizing this to multiple variables;^{ad} indeed, an expression that contains the partial derivatives $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ (along with an unknown function $f(x, y)$) is *also* a differential equation. We then divide all differential equations into two categories:

$$\begin{aligned} \text{A differential equation is called } & \left\{ \begin{array}{c} \text{ordinary} \\ \text{partial} \end{array} \right\} \text{ if there are} \\ & \left\{ \begin{array}{c} \text{one} \\ \text{more than one} \end{array} \right\} \text{ derivatives with respect to } \text{variables.} \end{aligned} \quad (1.48)$$

^zThis means

$$f : \mathbb{C} \rightarrow \mathbb{C} \quad (1.44a)$$

$$f^{-1} : \mathbb{C} \rightarrow \mathbb{C} \quad (1.44b)$$

$$f = x \rightarrow f(x) \quad (1.44c)$$

$$f^{-1} = f(x) \rightarrow x \quad (1.44d)$$

^{aa}We can now check that our very initial assumption $f(x) \neq 0$ is indeed satisfied.

^{ab}We can show this by using the fundamental theorem of calculus.

^{ac}As stated earlier, $f(x)$ is actually *not* the function but the *output* of the function f . Nevertheless, I'll abuse terminology here and there to remain more familiar to physicists.

^{ad}Alternatively, we can generalize to multiple *functions*; for instance,

$$\frac{df(x)}{dx} = g(x), \quad \frac{dg(x)}{dx} = -f(x). \quad (1.47)$$

Such relations are called *systems of differential equations*. We will see more about such systems in § 3.2.

For instance,

$$\frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} + f(x, y) = 0 \quad (1.49)$$

is a partial differential equation. Until the last chapter, we will only focus on *ordinary* differential equations!

We also define the *order* of a differential equation to be the highest number of derivatives in it; for instance,

$$\frac{d^3}{dx^3} f(x) = 0 \quad (1.50)$$

is a third order differential equation,^{ae} whereas

$$\frac{d^3}{dx^3} f(x) + f(x) \frac{d^4}{dx^4} f(x) + x \frac{d}{dx} f(x) = 0 \quad (1.52)$$

is a fourth order one. Note that not all differential equations have to have a finite order.^{af}

The differential equations are also grouped according to the *linearity* of the unknown function f . For instance, the differential equation

$$\frac{d^2}{dx^2} f(x) + f(x) = 1 \quad (1.55)$$

is called a *linear differential equation*, whereas

$$f(x) \frac{d}{dx} f(x) = x^3 \quad (1.56)$$

is a *nonlinear* differential equation.^{ag} An easy way to check if a differential equation is linear or nonlinear is to apply the transformation $f(x) \rightarrow \lambda f(x)$ for the constant λ : if the differential equation is linear in λ (i.e. it can be written as $\lambda(\dots) + (\dots) = 0$), then the differential equation is a linear differential equation; otherwise, it is a nonlinear differential equation.

Nonlinear equations are way harder to solve than the linear equations; in fact, we actually do not know how to solve most of the nonlinear equations! In practice, one usually handles them numerically, which is beyond of the scope of this course. If you are only interested in a particular regime, you can also *linearize* a nonlinear equation around that regime, which is what most physicists do in practice. For instance, consider the nonlinear differential equation

$$\frac{d}{dx} f(x) + \sin(f(x)) = 0 \quad (1.57)$$

If we say that we are only interested in the results $f(x) \ll 1$, then we can linearize this equation as

$$\frac{d}{dx} f(x) + f(x) = 0 \quad (1.58)$$

which has the solution

$$f(x) = ce^{-x} \quad (1.59)$$

^{ae} See if you can convince yourself that

$$f(x) = c_0 + c_1 x + c_2 x^2 \quad (1.51)$$

for the coefficients c_i is the solution to this equation.

^{af} It is perfectly possible to define the differential equation

$$\exp\left(\frac{d}{dx}\right) f(x) = f(x) + 3x^2 + 3x + 1 \quad (1.53)$$

for which

$$f(x) = x^3 \quad (1.54)$$

is a solution (see the footnote v). However, clearly, this differential equation has arbitrarily high numbers of derivatives, hence it is of infinite order.

^{ag} One important feature of linear differential equations is that their solutions obey *the principle of supersposition*; that is, if $f(x)$ and $g(x)$ are two solutions to the linear differential equation, then $c_1 f(x) + c_2 g(x)$ is also a solution for arbitrary constants $c_{1,2}$.

Table 1.1: Illustration of various differential equations

Example differential equation	ordinary?	linear?	homogeneous?
$\frac{d^2 f(x)}{dx^2} + f(x) = 0$	✓	✓	✓
$\frac{d^2 f(x)}{dx^2} + f(x) = x^2$	✓	✓	✗
$f(x) \frac{d^3 f(x)}{dx^3} + \left(\frac{df(x)}{dx} \right)^2 = 0$	✓	✗	✓
$\frac{d^2 f(x)}{dx^2} + \sin(f(x)) = 0$	✓	✗	✗
$\frac{\partial^2 f(x, y)}{\partial x \partial y} + f(x, y) = 0$	✗	✓	✓
$\frac{\partial^2 f(x, y)}{\partial x^2} + f(x, y) = x^2$	✗	✓	✗
$f(x) \frac{\partial^3 f(x, y)}{\partial x^3} + \left(\frac{\partial f(x, y)}{\partial y} \right)^2 = 0$	✗	✗	✓
$\frac{\partial^2 f(x, y)}{\partial x \partial y} + \sin(f(x, y)) = 0$	✗	✗	✗

^{ah} We can actually solve the full nonlinear differential equation eqn. (1.57); the result is

$$f(x) = 2 \operatorname{arccot} \left(\frac{2e^x}{c} \right) \quad (1.60)$$

which matches the linearized result in the regime it is valid, i.e.

$$\lim_{x \rightarrow \infty} 2 \operatorname{arccot} \left(\frac{2e^x}{c} \right) = \lim_{x \rightarrow \infty} ce^{-x} \quad (1.61)$$

which satisfies our necessary condition for $x \gg 1$.^{ah}

A last classification we can do with our differential equations is their *homogeneity*: a differential equation is said to be *homogeneous* if it is invariant under the scaling of the unknown function. This is just a fancy way of saying that the differential equation does not change even if we replace $f(x)$ with $\lambda f(x)$ for an unknown constant λ .

We can summarize the classification of all differential equations with examples as given in Table 1.1

2

Linear equations with constant coefficients

2.1 Linear mappings and kernels

Formally, we could write down the most generic linear ordinary differential equation for the unknown function f as

$$g\left(x, \frac{d}{dx}\right) f(x) = h(x) \quad (2.1)$$

for arbitrary known functions g and h . Indeed, this is a linear equation in the function f , and it has only one kind of derivative, $\frac{d}{dx}$, hence it is an ordinary differential equation.

Let's assume that we are given such an equation for known g and h , and we are trying to solve for f . A naive attempt would be to write down

$$f(x) = \frac{1}{g\left(x, \frac{d}{dx}\right)} h(x) \quad (2.2)$$

which looks like a total nonsense! Nevertheless, we cannot help but realize that it does somewhat work in some cases; for instance, for

$$\frac{d}{dx} f(x) = x^2 \quad (2.3)$$

we can write down

$$f(x) = \left(\frac{d}{dx}\right)^{-1} x^2 \quad (2.4)$$

which we can rewrite as

$$f(x) = \int dx x^2 = \frac{x^3}{3} + \text{constant} \quad (2.5)$$

by observing that integral is *the inverse of derivative*.^{ai}

We need to be careful with such manipulations, but physicists *tend to define things formally*, which allows such expressions. For instance, we could say that *the formal solution* to the differential equation

$$\left(\frac{d^2}{dx^2} + c^2\right) f(x) = 0 \quad (2.6)$$

is

$$f(x) = \left(\frac{d^2}{dx^2} + c^2\right)^{-1} 0 \quad (2.7)$$

For a physicist, there is nothing wrong with writing things like the equation above *as long as we are careful with what we mean!* To spell out what we really mean with such an equation, we need to set up some terminology.

^{ai}Rigorously speaking, we are referring to indefinite integrals (also known as antiderivatives or Newton integrals).

Remember how we defined the derivative higher order function (or its integer powers) in eqn. (1.20a):

$$\frac{d^n}{dx^n} :: (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (2.8a)$$

$$f :: A \rightarrow B \quad (2.8b)$$

$$f^{(n)} :: A \rightarrow B \quad (2.8c)$$

The operation of taking derivatives is a *map of functions to functions*; in fact, it is a *linear map*!^{aj} Linear maps are really useful when we work with vectors, but we will see below that an important notion called *kernel* can be extended from vector spaces to the functions as well.^{ak}

In vector spaces linear transformations are implemented by matrices; for instance, the transformation “clockwise rotation by $\pi/4$ ” on $2d$ vectors can be implemented by the matrix

$$R(-\pi/4) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (2.12)$$

which indeed rotates any vector $\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}$ to its rotated version $R(-\pi/4) \cdot \vec{v}$; for instance, the unit vector pointing to NorthEast direction on a map —i.e. $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ — gets rotated to the vector pointing to the East by this 45 degrees of clockwise rotation:

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (2.13)$$

In fact a *general counterclockwise rotation by an angle θ* can be implemented by the matrix

$$R(-\pi/4) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \quad (2.14)$$

The *kernel of a map* (or equivalently the kernel of the matrix that implements that map) is the set of vectors that are mapped to *zero vector*; for instance, we can show that the only such vector for the rotation matrix is the zero vector itself; in other words,

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.15)$$

is true only if $a = b = 0$; thus, we write

$$\ker [R(\theta)] = \{\vec{0}\} \quad (2.16)$$

which means *the only vector that can be rotated to the zero vector is the zero vector itself*. When said this way, it clearly makes sense!

Let’s look at another example: we define the matrix S as

$$S = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \quad (2.17)$$

^{aj}We can easily see the linearity by noting the relation

$$\frac{d^n}{dx^n} (c_1 f(x) + c_2 g(x)) = c_1 \frac{d^n}{dx^n} f(x) + c_2 \frac{d^n}{dx^n} g(x) \quad (2.9)$$

for arbitrary coefficients c_1 and c_2 .

^{ak} The analogy is as follows: functions are like vectors, and linear transformations due to derivatives are like matrix multiplications. Indeed, a matrix M (say $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$) acting on a vector v (say $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$) is a linear mapping, just as the derivative $\frac{d}{dx}$ turning the function x^2 into $2x$.

The analogy extends to the equations. We could solve $M \cdot w = v$ for the unknown vector w , similar to how we solve $\frac{d}{dx} f(x) = x^2$ for the function f . In fact such analogies can be made more precise if we realize that a function f is in some sense an infinite dimensional vector. Indeed, in a neighborhood containing the point c in which the function f is analytic, we can just do a Taylor expansion and rewrite $f(x)$ as

$$f(x) = \sum_{n=0}^{\infty} f_n x^n \quad (2.10)$$

where f_n can be viewed as an infinite-dimensional vector $f_n = (f_0, f_1, \dots)$.^{af}

^{af} If we take a step back, we can actually realize that the converse is also true (in fact, it is *generically* true): *any vector v is simply a function from integers to the domain of the components of the vector*.

What do we mean by that? Consider the vector $\vec{v} = \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$. This vector is equivalent to the set of relations $v_1 = 1$, $v_2 = 0$, and $v_3 = -3$. But that is simply a function

$$v :: \mathbb{N} \rightarrow \mathbb{R} \quad (2.11a)$$

$$v = n \rightarrow \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n = 2 \\ -3 & \text{if } n = 3 \\ \text{undefined} & \text{otherwise} \end{cases} \quad (2.11b)$$

This process can be generalized to any finite or infinite dimensional vector.

If we now look at the *kernel of this linear transformation*, we find a non-trivial result; in fact, we can immediately write down

$$\ker[S] = \left\{ \vec{0}, \begin{pmatrix} 2a \\ -a \end{pmatrix} \right\} \quad (2.18)$$

which means not only the zero vector gets mapped to zero vector, but also any vector of the form $\begin{pmatrix} 2a \\ -a \end{pmatrix}$ becomes zero under the action of this matrix. Indeed, we see that

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 2a \\ -a \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.19)$$

What does this mean? And what is the action of this linear transformation? Just like the rotation matrix rotates any input vector, this S matrix also transforms the input vectors, but it actually *squeezes* them. Indeed, we see that for any vector pointing in any direction, the action of this transformation squeezes them into the $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ direction. We can see this explicitly:

$$S \cdot \vec{v}_{\text{input}} = \vec{v}_{\text{output}} \quad (2.20a)$$

$$\vec{v}_{\text{input}} = \begin{pmatrix} a \\ b \end{pmatrix} \quad (2.20b)$$

$$\vec{v}_{\text{output}} = (a + 2b) \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (2.20c)$$

Summary: We have seen with examples that some linear transformations (such as rotation) has a *trivial kernel*,^{am} whereas other transformations (such as squeezing) may have a nontrivial kernel.

Quick check in vector spaces: Whether a linear transformation has a trivial kernel or not can immediately be checked in the case of vector spaces by computing the *determinant* of the matrix that implements that transformation. If the determinant is zero (e.g. $\det S = 0$), then the kernel is nontrivial; otherwise ($\det R = 1$) the kernel is trivial.

The importance of nontrivial kernel: If the kernel is nontrivial, then the transformation is not uniquely invertible. For instance, if we have

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (2.21)$$

Then $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is a solution, but so is $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$ or $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. In fact, the full family of solutions is given as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 2a \\ -a \end{pmatrix} \quad (2.22)$$

^{am}We say that a kernel is trivial if it only includes the identity element ($\vec{0}$ vector in the case of vector spaces).

On the other hand, the rotation having a trivial kernel makes sure that we have a unique answer; for instance,

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (2.23)$$

has the unique answer

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 3 \end{pmatrix} \quad (2.24)$$

Back to the differential equations: The story with the matrices immediately carries over to the differential equations: the differential operators have nontrivial kernels, and these results are called *homogeneous solutions*. For the general differential equation

$$g\left(x, \frac{d}{dx}\right) f(x) = h(x) \quad (2.25)$$

the solution then becomes

$$f(x) = p(x) + \ker \left[g\left(x, \frac{d}{dx}\right) \right] \quad (2.26)$$

where $p(x)$ is called the *particular solution*, and the elements of the kernel are the homogeneous solutions.

2.2 Homogeneous solutions

2.2.1 Basics

We have discussed in § 1.2.1 that the solution to the differential equation $f'(x) = f(x)$ is given as^{an}

$$f(x) = e^x \quad (2.27)$$

We can actually generalize this to^{ao}

$$\left(\frac{d}{dy} - \lambda \right) e^{\lambda y} = 0 \quad (2.29)$$

which says the *linear ordinary differential equation with constant coefficient*

$$\left(\frac{d}{dx} - \lambda \right) f(x) = 0 \quad (2.30)$$

has the *solution*

$$f(x) = e^{\lambda x} \quad (2.31)$$

In the fancy language, we can now write this result as

$$\ker \left[\left(\frac{d}{dx} - \lambda \right) \right] = \{0, e^{\lambda x}\} \quad (2.32)$$

^{an} As we discussed in that section, this result is either a definition or a derived result depending on our conventions.

^{ao} One way to show this is the judicious use of the chain rule as follows:

$$\begin{aligned} \frac{d}{dx} e^x &= e^x \xrightarrow{\text{define } x=\lambda y} \frac{d}{dx} e^{\lambda y} = e^{\lambda y} \\ &\xrightarrow{\text{use chain rule}} \frac{dy}{dx} \frac{d}{dy} e^{\lambda y} = e^{\lambda y} \\ &\xrightarrow{\text{use } y=x/\lambda} \frac{1}{\lambda} \frac{d}{dy} e^{\lambda y} = e^{\lambda y} \\ &\xrightarrow{\text{rewrite}} \left(\frac{d}{dy} - \lambda \right) e^{\lambda y} = 0 \end{aligned} \quad (2.28)$$

which means that

$$\left(\frac{d}{dx} - \lambda\right) f(x) = h(x) \Rightarrow f(x) = p(x) + ce^{\lambda x} \quad (2.33)$$

for the arbitrary variable c , where we will discuss the computation of particular solution $p(x)$ later.

One immediate observation we can make is that $e^{\lambda x}$ would still be a solution if there were more terms to the left of the equation; in other words,

$$g\left(x, \frac{d}{dx}\right) \left(\frac{d}{dx} - \lambda\right) f(x) = 0 \quad (2.34)$$

is still satisfied for $f(x) = e^{\lambda x}$. This becomes particularly interesting if $g\left(x, \frac{d}{dx}\right)$ is a product of $\left(\frac{d}{dx} - a\right)$, i.e.

$$\left(\frac{d}{dx} - r_1\right) \left(\frac{d}{dx} - r_2\right) \cdots \left(\frac{d}{dx} - r_n\right) f(x) = 0 \quad (2.35)$$

Clearly $e^{r_n x}$ is a solution, but as these terms commute with each other, we can immediately write down the full solution as^{ap}

$$f : \mathbb{C} \rightarrow \mathbb{C} \quad (2.36a)$$

$$f = x \rightarrow \sum_{i=1}^n c_i e^{r_i x} \quad (2.36b)$$

for arbitrary constants c_i .

Differential equations are usually given in the form

$$\left(a_0 + a_1 \frac{d}{dx} + a_2 \frac{d^2}{dx^2} + \cdots + a_n \frac{d^n}{dx^n}\right) f(x) = 0 \quad (2.37)$$

which can be brought to the form eqn. (2.35) by simply finding the roots of the equation^{aq}

$$a_0 + a_1 r + a_2 r^2 + \cdots + a_n r^n = 0 \quad (2.38)$$

If the coefficients a_i are simply complex numbers (or real numbers as a special case of complex numbers), we can always find n complex roots $r_i!$ ^{ar}

2.2.2 Repeated roots

Consider the differential equation

$$\left(\frac{d}{dx} - r_1\right) \left(\frac{d}{dx} - r_2\right) f(x) = 0 \quad (2.39)$$

which has the solution $f(x) = c_1 e^{r_1 x} + c_2 e^{r_2 x}$. If we now do a change of parameters as

$$r_1 = r, \quad r_2 = r + \delta, \quad c_1 = a - \frac{b}{\delta}, \quad c_2 = \frac{b}{\delta} \quad (2.40)$$

^{ap}This follows from the principle of superposition, see footnote ag.

^{aq}This equation is called *characteristic equation* of the given system.

^{ar} The field of complex numbers is algebraically closed, hence such polynomials *always* have solutions. In contrast, the field of real numbers is *not* algebraically closed; for instance, $x^2 + 1 = 0$ has no real root. For more information on these, see *fundamental theorem of algebra*.

our statement becomes

$$\left(\frac{d}{dx} - r\right) \left(\frac{d}{dx} - (r + \delta)\right) f(x) = 0 \quad f(x) = ae^{rx} + b \frac{e^{(r+\delta)x} - e^{rx}}{\delta} \quad (2.41)$$

If we now take the limit $\delta \rightarrow 0$ and recognize the definition of derivative, we arrive at

$$\left(\frac{d}{dx} - r\right)^2 f(x) = 0 \quad \rightarrow \quad f(x) = ae^{rx} + bxe^{rx} \quad (2.42)$$

The way we arrived at this curious result is not satisfactory: we did a particular transformation in eqn. (2.40) and we do not have a strong reason to choose that transformation. For instance, if we instead choose

$$r_1 = r, \quad r_2 = r + \delta, \quad c_1 = a - b, \quad c_2 = b \quad (2.43)$$

and then take the limit $\delta \rightarrow 0$, we end up with

$$\left(\frac{d}{dx} - r\right)^2 f(x) = 0 \quad \xrightarrow{???} \quad f(x) = ae^{rx} \quad (2.44)$$

We missed the second piece of $f(x)$ in eqn. (2.42).

What is the resolution of this discrepancy? We have two potential scenarios: **(a)** xe^{ax} is a *spurious solution*,^{as} or **(b)** eqn. (2.44) misses one of the solutions.

We can check it straightforwardly that the option **(b)** is the correct case,^{at} indicating that our choice of reparametrization of the variables in terms of infinitesimal variable δ affects which solutions we obtain. This then begs the question: *can we potentially have more solutions?*

We have mathematical arguments why a second order differential equation should have two solutions,^{au} so we can already infer that eqn. (2.42) is the full solution; however, let's see another method to derive why this is the case.

Define a new function g such that $f(x) = g(x)e^{rx}$.^{av} If we insert this into the original differential equation, we immediately see that

$$\left(\frac{d}{dx} - r\right)^2 f(x) = 0 \quad \xrightarrow{f(x)=g(x)e^{rx}} \quad \frac{d^2}{dx^2} g(x) = 0 \quad (2.46)$$

which tells us that *the most general result* is $f(x) = (ax + b)e^{rx}$. In fact, this derivation generalizes, i.e.

$$\left(\frac{d}{dx} - r\right)^n f(x) = 0 \quad \xrightarrow{f(x)=g(x)e^{rx}} \quad \frac{d^n}{dx^n} g(x) = 0 \quad (2.47)$$

yielding $f(x) = (a_1 + a_2x + \dots a_nx^{n-1})e^{rx}$.

With the discussion above, we can now write down the most general homogeneous solution to a linear ordinary differential equation with constant coefficients:

$$\left(\frac{d}{dx} - r_1\right)^{m_1+1} \left(\frac{d}{dx} - r_2\right)^{m_2+1} \dots \left(\frac{d}{dx} - r_n\right)^{m_n+1} f(x) = 0 \quad (2.48a)$$

$$\Rightarrow \quad f(x) = \sum_{i=1}^n \left[\left(\sum_{k=0}^{m_i} c_{ik} x^k \right) e^{r_i x} \right] \quad (2.48b)$$

^{as}Spurious solutions are fake results that emerge as solutions even though they actually do not solve the problem.

^{at} We only need to check

$$\left(\frac{d}{dx} - r\right)^2 (xe^{rx}) = 0 \quad (2.45)$$

^{au}Maybe expand on this more.

^{av}Note that we can do this *without a loss of generality*!

for arbitrary coefficients c_{ij} .

2.2.3 Examples

RLC circuits, damper-spring systems, traffic models, etc.

2.3 Laplace transform

Consider the following higher order function:^{aw}

$$\mathcal{I}\mathcal{T} :: (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \quad (2.50a)$$

$$\mathcal{I}\mathcal{T} = (x \rightarrow f(x)) \rightarrow \left(s \rightarrow \int_{\alpha}^{\beta} K(x, s) f(x) dx \right) \quad (2.50b)$$

where $\mathcal{I}\mathcal{T}$ is an *integral transform*, i.e. it maps a function to another one by using the integration operation. The function K above is called *the kernel of the transformation*: different kernels (along with different integration ranges) lead to different integral transforms.

The Laplace transform is a special kind of an integral transformation:

$$\mathcal{L} :: (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \quad (2.51a)$$

$$\mathcal{L} = (x \rightarrow f(x)) \rightarrow \left(s \rightarrow \int_0^{\infty} e^{-xs} f(x) dx \right) \quad (2.51b)$$

which plays a immense role in the analysis of linear ordinary differential equations with constant coefficients because such equations become algebraic under this transformation. To see this, consider how the laplace transform interacts with the derivative operation: replace f with g' above, and integrate by parts

$$\mathcal{L} = (x \rightarrow g'(x)) \rightarrow \left(s \rightarrow \left[s \int_0^{\infty} e^{-xs} g(x) dx - g(0) + \lim_{x \rightarrow \infty} e^{-xs} g(x) \right] \right) \quad (2.52)$$

We will assume that the last piece is zero, which is a necessary condition for the Laplace transform to be well-defined in the first place.^{ax} Thus

$$\mathcal{L} \cdot g' = s \rightarrow (s(\mathcal{L} \cdot g)(s) - g(0)) \quad (2.53)$$

or in a more conventional notation, we state

$$\frac{dg(x)}{dx} \xrightarrow{\text{Laplace transform}} sG(s) - g(0) \quad (2.54)$$

where $G(s)$ is the laplace transform of $g(x)$.

One can repeat this process iteratively for higher numbers of derivative; in fact, we can immediately write down the Laplace transform of n -the derivative of a function:

$$(\mathcal{L} \cdot g^{(n)})(s) = s^n (\mathcal{L} \cdot g)(s) - \sum_{i=0}^{n-1} s^{n-i-1} g^{(i)}(0) \quad (2.55)$$

^{aw} Note that the letters on the left hand side of an arrow are *placeholders*, i.e. they do not inherently carry information. Such parameters are called dummy variables in math (or scooping variables in computer science) and they are ubiquitous in math and physics; for instance, the integrals $\int dx f(x)$ and $\int dy f(y)$ are the same expression as x and y are dummy variables. Similarly, the expressions $x \rightarrow f(x)$ and $y \rightarrow f(y)$ are equivalent.

It gets complicated with the higher order functions as they include multiple arrows; in such cases, the left hand side of *each* arrow contains only placeholders for *the right hand side of that particular arrow*. For example, let us rewrite eqn. (2.50) in a colorful way:

$$\mathcal{I}\mathcal{T} :: (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \quad (2.49a)$$

$$\mathcal{I}\mathcal{T} = (x \rightarrow f(x)) \rightarrow \left(s \rightarrow \int_{\alpha}^{\beta} K(x, s) f(x) dx \right) \quad (2.49b)$$

Variables of the same color can be changed as they are dummy variables for the same color arrow (in the case of the color magenta, the variables are dummy variables of the integration operation). For instance, following expressions are all equivalent:

$$\mathcal{I}\mathcal{T} = (x \rightarrow f(x)) \rightarrow \left(s \rightarrow \int_{\alpha}^{\beta} K(x, s) f(x) dx \right)$$

$$\mathcal{I}\mathcal{T} = (y \rightarrow f(y)) \rightarrow \left(z \rightarrow \int_{\alpha}^{\beta} K(x, z) f(x) dx \right)$$

$$\mathcal{I}\mathcal{T} = (y \rightarrow g(y)) \rightarrow \left(z \rightarrow \int_{\alpha}^{\beta} K(s, z) g(s) ds \right)$$

Note that the letters K , α , β are not dummy variables as they are externally fixed. Nevertheless, we *can* turn them into dummy variables of the equal sign = by defining them in the left hand side of =; e.g.

$$\mathcal{I}\mathcal{T}_{K, \alpha, \beta} = (y \rightarrow g(y)) \rightarrow \left(z \rightarrow \int_{\alpha}^{\beta} K(s, z) g(s) ds \right)$$

$$\mathcal{I}\mathcal{T}_{T, \gamma, \lambda} = (y \rightarrow g(y)) \rightarrow \left(z \rightarrow \int_{\gamma}^{\lambda} T(s, z) g(s) ds \right)$$

are equivalent expressions —just like $f(x) = x^2$ and $f(y) = y^2$ being equivalent expressions.

^{ax}Otherwise, the integral in the definition does not converge.

We can now justify our previous statement of **Laplace transform converts linear ordinary differential equations with constant coefficients into algebraic ones!** Start with the most generic such differential equation:

$$\sum_{i=0}^n a_i f^{(i)}(x) = g(x) \quad (2.56)$$

which is *homogeneous* if $g(x) = 0$ and nonhomogeneous otherwise. If we take the Laplace transform of this equation, we end up with

$$\sum_{i=0}^n a_i \left[s^i F(s) - \sum_{k=0}^{i-1} s^{i-k-1} f^{(k)}(0) \right] = G(s) \quad (2.57)$$

where $F(s) := (\mathcal{L} \cdot f)(s)$ and $G(s) := (\mathcal{L} \cdot g)(s)$ are defined for brevity. By using algebra, we can rewrite this equation in the form

$$F(s) = \frac{\sum_{i=0}^{n-1} f^{(i)}(0) \left[\sum_{k=1+i}^n a_k s^{k-i-1} \right]}{\sum_{i=0}^n a_i s^i} + \frac{G(s)}{\sum_{i=0}^n a_i s^i} \quad (2.58)$$

Let us comment on this result a little bit. **Firstly**, we can immediately state that the solution $f(x)$ to the differential equation in eqn. (2.56) is simply the *inverse Laplace transform* of $F(s)$. Even though this is a well-defined transformation that we can introduce, we actually do not need it: we will discuss other methods to obtain $f(x)$ from $F(s)$. **Secondly**, we can actually see that the first piece is the homogeneous solution to the differential equation, and the second piece is the particular solution: Laplace transform allowed us to solve both of them at once!

Consider the simple case of $n = 2$:

$$F(s) = \frac{f(0)(a_1 + a_2 s) + f^{(1)}(0)a_2}{a_0 + a_1 s + a_2 s^2} + \frac{G(s)}{a_0 + a_1 s + a_2 s^2} \quad (2.59)$$

If r_1 and r_2 are two distinct roots of $a_0 + a_1 s + a_2 s^2 = 0$, we can simply write down this expression as *bla bla bla blato be written later, probably next year*.

We have covered several topics in class but I will not be able to type them in time. So I'm postponing that to next year; after all, all of those topics are already in the textbook — chapter 6 of *Elementary Differential Equations and Boundary Value Problems* by Boyce and DiPrima (10th edition). The summary is as follows:

1. Derive laplace transforms of common functions
2. Solving homogeneous differential equations using Laplace transformation, and "proof" of why characteristic equation method works

3. Discussion of particular solutions in Laplace domain; example: RLC circuit with an AC input
4. Discussion of how it is tiresome to repeat the computations for each nonhomogeneous piece and why we need a universal solution true for any non-homogeneous piece. For this, we need to find an operation that maps to multiplication in Laplace domain
5. Derive $H(s) = F(s)G(s)$ means $h(x)$ =convolution of $f(s)$ and $g(x)$, i.e. derive convolution
6. Introduce impulse response: it is the solution to the same differential equation with nonhomogeneous part $H(s) = 1$ in laplace domain (with $i(0) = i'(0) = \dots = 0$).
7. Introduce Dirac-delta distribution, discuss why it is a generalized function but not a well defined function. Show that it maps to 1 as needed.
8. Write down the most generic solution to a linear ordinary differential equation with constant coefficients. Show examples.

Linear equations with functional coefficients

3.1 Homogeneous solution

Summary of what we have discussed in class (to be typed later, maybe next year):

1. Differential equations with functional coefficients *do not necessarily have generic analytic solutions*: we do not know how to solve them except the isolated cases!
2. Explicit solutions of $(x \frac{d}{dx} + a) f(x) = 0$ and $(x^k \frac{d}{dx} + a) f(x) = 0$ for $k \neq 1$.
3. How these differential operators can be chained, and how it leads to the following form:

$$\left(x^k \frac{d}{dx} + a_1\right) \cdots \left(x^k \frac{d}{dx} + a_n\right) f(x) = 0 \quad (3.1)$$

for both $k = 1$ & $k \neq 1$.

4. How these equations are secretly related to the differential equations with constants coefficients through a reparametrization, i.e.

$$\begin{aligned} \frac{d}{dy} &= \frac{1}{1-n} x^n \frac{d}{dx} \text{ for } y = x^{-n+1} \\ \frac{d}{dy} &= x \frac{d}{dx} \text{ for } y = \log(x) \end{aligned} \quad (3.2)$$

5. Lesson: The differential equations with constant coefficients are nice guys with straightforward general solutions: do your best to check if a given differential equation can be brought to that form! Usually, the physics of the problem gives us insight as to whether that would be possible.
6. Discuss why $k = 1$ case above is treated differently, show a little bit about scale invariance, and informally mention the Mellin transform: we do not need to know this for this course (but it is super relevant in modern physics)!
7. Introduce Euler equations: show how this is solvable simple because of the scale invariance and how it is actually a subset of similar higher order equations.
8. Discuss change of variables to make equations constant coefficients:

$$\begin{aligned} f''(x) + p(x)f'(x) + q(x)f(x) &= 0 \\ \Rightarrow \\ (u')^2 f''(u) + (u''(x) + u'(x)p(x))y'(u) + q(x)y(u) &= 0 \end{aligned} \quad (3.3)$$

for a change of parameter $x \rightarrow u(x)$. For this to be constant coefficient, we need $u(x) = \int \sqrt{q(x)} dx$ and $\frac{u''(x)+u'(x)p(x)}{(u')^2} = \text{constant}$.

9. An example equation where change of variables would work:
 $ty'' + (t^2 - 1)y' + t^3y = 0$.
10. Reparametrization is harder to do generically for higher orders! For those equations (also for second order equations without reparametrization), check if the function $f(x)$ is missing. If that is missing, we can lower the order of differential equation by writing it in terms of a new function $g(x) = f'(x)$. If both $f(x)$ and $f'(x)$ are missing, use $g(x) = f''(x)$, and so on!
11. Examples (page 135 of textbook): $xf''(x) + f'(x) = 0$, $x^2f''(x) + 2xf'(x) = 2$. Solve these!
12. Another thing to check is if the equation is *exact*, i.e. if it can be rewritten as a total derivative:

$$\left[p_n(x) \frac{d^n}{dx^n} + \dots + p_1(x) \frac{d}{dx} + p_0(x) \right] f(x) = 0$$

$$\xrightarrow{???}$$

$$\frac{d}{dx} \left(\left[q_{n-1}(x) \frac{d^{n-1}}{dx^{n-1}} + \dots + q_1(x) \frac{d}{dx} + q_0(x) \right] f(x) \right) = 0$$

If that is the case, then we can turn an order- n (non)homogeneous differential equation into an order- $(n-1)$ nonhomogeneous one.

13. For a second order differential equation $p(x)f''(x) + q(x)f'(x) + r(x)f(x) = 0$, the condition $p''(x) - q'(x) + r(x) = 0$ is sufficient for it to be exact.
14. Examples (page 157 of textbook): $f''(x) + xf'(x) + f(x) = 0$, $x^2f''(x) + xf'(x) - f(x) = 0$.
15. **Reduction of order:** if we already know k solutions of an order n differential equation, we can use that information to transform the system into an order $n - k$ differential equation with no known solutions. This is rather useful as lower differential equations are easier to solve, and it becomes extremely useful if we know one solution of a second order differential equation as first order differential equations are always formally solvable (we will discuss this in more detail later).
16. Examples (page 174 of textbook): $xf''(x) - f'(x) + 4x^3f(x) = 0$ with $f_1(x) = \sin(x^2)$, $(x-1)f''(x) - xf'(x) + f(x) = 0$ with $f_1(x) = e^x$.

17. We introduced the Levi-Civita symbol $\epsilon : \{\mathbb{Z}^+, \dots, \mathbb{Z}^+\} \rightarrow \mathbb{Z}$ and discussed its properties.

$$\epsilon : \{\mathbb{Z}^+, \dots, \mathbb{Z}^+\} \rightarrow \mathbb{Z} \quad (3.4a)$$

$$\epsilon = \{a_1, \dots, a_n\} \rightarrow \begin{cases} 1 & \text{if } (a_1 a_2 \dots a_n) \text{ is an even permutation of } (12 \dots n) \\ -1 & \text{if } (a_1 a_2 \dots a_n) \text{ is an odd permutation of } (12 \dots n) \\ 0 & \text{otherwise} \end{cases} \quad (3.4b)$$

Example: $(132) \rightarrow (123)$: we need 1 permutation for (132) : $\epsilon_{132} = -1$. $(2314) \rightarrow (2134) \rightarrow (1234)$: we need 2 permutations for 2314: $\epsilon_{2314} = 1$.

Properties: $\epsilon_{\dots a \dots a \dots} = 0$, $\epsilon_{\dots a \dots b \dots} = -\epsilon_{\dots b \dots a \dots}$

18. We introduced the function $\det : \mathfrak{M}_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$ in terms of Levi-Civita symbol ϵ and discussed its properties.

$$\det : \mathfrak{M}_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C} \quad (3.5a)$$

$$\det = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} a_{1i_1} \dots a_{ni_n} \quad (3.5b)$$

19. We discussed linear independence of solutions and introduced the Wronskian determinant to check if given set of solutions span the solution space.

The summary of the discussion is as follows. Assume that we are given an order- n linear ordinary differential equation $g\left(x, \frac{d}{dx}\right) f(x) = h(x)$, and assume that we have found n -solutions $f_i(x)$. If these solutions are linearly independent, they span the solution space and can be used to match any initial condition uniquely, i.e.

$$\sum_{i=1}^n c_i f_i(x_0) = f(x_0) \quad (3.6a)$$

$$\sum_{i=1}^n c_i f'_i(x_0) = f'(x_0) \quad (3.6b)$$

$$\dots \quad (3.6c)$$

$$\sum_{i=1}^n c_i f_i^{(n-1)}(x_0) = f^{(n-1)}(x_0) \quad (3.6d)$$

for the unique set of numbers c_i . As a matrix equation, this means

$$\begin{pmatrix} f_1(x_0) & f_2(x_0) & \dots & f_n(x_0) \\ f'_1(x_0) & f'_2(x_0) & \dots & f'_n(x_0) \\ \dots & \dots & \dots & \dots \\ f_1^{(n-1)}(x_0) & f_2^{(n-1)}(x_0) & \dots & f_n^{(n-1)}(x_0) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix} = \begin{pmatrix} f(x_0) \\ f'(x_0) \\ \dots \\ f^{(n-1)}(x_0) \end{pmatrix} \quad (3.7)$$

We can find out the unique c_i only if we can invert the matrix, which is only possible if it is full rank, which requires its determinant to be nonzero. That determinant is called Wronskian determinant and its value tells us if the given set of solutions span the solution space or not.

20. Started talking about Taylor series, how it can be interpreted as an expansion over an infinite dimensional vector space, and how series expansion can turn a differential equation into infinitely many algebraic equations. This is similar to turning vector equations into multiple scalar equations by expanding vectors on a basis and working with the components instead.
21. Solved explicitly the differential equation $f''(x) - xf(x) = 0$. Note that

- Not with constant coefficients
- not in $D_1 \cdot D_2 \cdot f = 0$ form
- cannot do reparametrization as $\frac{q'+2pq}{2q^{3/2}}$ is not constant for $p = 0$ and $q = x$
- $f(x)$ is not missing and we do not know one of the solutions (cannot reduce order)
- equation is not exact, hence cannot be rewritten as a non-homogeneous lower-order equation

So we have to use series expansion!

22. Discussed expansions around different points, i.e. $f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$
23. Introduced the classification of expansion points: given the differential equation

$$\left[P_n(x) \frac{d^n}{dx^n} + P_{n-1}(x) \frac{d^{n-1}}{dx^{n-1}} + \cdots + P_0(x) \right] f(x) = 0 \quad (3.8)$$

for the analytic functions $P_i(x)$, a point x_0 is called “ordinary point” if $\frac{P_i(x)}{P_n(x)}$ is analytic for all i , and is called “singular point” otherwise. For instance $x = 0$ is an ordinary point of $\left(x \frac{d}{dx} + \sin(x)\right) f(x) = 0$.

24. Around ordinary points, Taylor series expansion works and one can get all solutions correctly.
25. Rewrite the differential equation above as

$$\left[\frac{d^n}{dx^n} + Q_{n-1}(x) \frac{d^{n-1}}{dx^{n-1}} + \cdots + Q_0(x) \right] f(x) = 0 \quad (3.9)$$

If $Q_{n-i}(x)$ has a pole of order at most i at x_0 for all i , then the singular point is called “regular singular point”. Otherwise, it is called “essential singular point” (or non-regular singular point).

26. Around regular singular points, we can use Taylor series expansion with an unknown monomial as an overall factor; i.e. we can take $f(x) = (x - x_0)^r \sum_{n=0}^{\infty} a_n (x - x_0)^n$ for the unknowns a_n and r . This is called Frobenius method.
27. We do not use series expansions around essential singular points; even if there is a way to do that, I do not know!
28. In class, we solved explicitly the diff eqn.

$$x^2 f''(x) - x f'(x) + (1 + x)f(x) = 0 \quad (3.10)$$

around $x = 0$.

29. Discussion of how series solutions are *local solutions* in the complex plane: they have (usually) finite radius of convergence, and one needs to construct different series solutions to access values of the function in different locations of the complex plane. In some cases, we might get lucky as we can recognize the series series as a particular representation of a more general function such as Bessel function; in such cases, with one series solutions, we can discover more general properties of the solution for the given differential equation. However, if we are not lucky, we need to construct other series solutions for different points, and we cannot really see the whole picture.
30. Another problem with the series solutions is that they do not make use of the global properties of the unknown function such as its symmetries. For instance, if I'm trying to solve a differential equation and I know that the solution function should have a symmetry (such as $f(x + a) = f(x)$ for some a), this information should in principle help me constraint the solution further. But as series expansions focus on local properties (such as analyticity in and around expansion point), they do not make use of such global information.
31. If we know some global properties of a function (such as it being spherically symmetric), we may be better off with a different kind of expansion. In fact, there are infinitely many different expansions (in group theoretical language, this is because we can use unitary irreducible representation of any group as a basis). To understand that, we need to discuss the concept of unitarity.

32. We started a new discussion: concept of unitarity. To understand that, we introduced the following definitions:

$$* :: \mathbb{C} \rightarrow \mathbb{C} \quad (3.11a)$$

$$* = z \rightarrow z^* = \operatorname{Re} z - i \operatorname{Im} z \quad (3.11b)$$

$$T :: \mathfrak{M}_{n \times n}(\mathbb{C}) \rightarrow \mathfrak{M}_{n \times n}(\mathbb{C}) \quad (3.11c)$$

$$T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & & & \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} \quad (3.11d)$$

$$\dagger :: \mathfrak{M}_{n \times n}(\mathbb{C}) \rightarrow \mathfrak{M}_{n \times n}(\mathbb{C}) \quad (3.11e)$$

$$\dagger = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11}^* & a_{21}^* & \dots & a_{n1}^* \\ a_{12}^* & a_{22}^* & \dots & a_{n2}^* \\ \dots & & & \\ a_{1n}^* & a_{2n}^* & \dots & a_{nn}^* \end{pmatrix} \quad (3.11f)$$

where $*$, T , \dagger are called to output *the complex conjugate*, *the transpose*, and *the hermitian conjugate* of the input respectively; for illustration, A^\dagger is called the hermitian conjugate of the matrix A .

33. A *unitary ordinary number* is a complex number with unit length, i.e. z with $|z| = zz^* = 1$. To matrices, this can be generalized with the hermitian conjugation: *a unitary matrix A is a matrix such that $A \cdot A^\dagger = \mathbb{I}$ for the unit matrix \mathbb{I} .*
34. Unitarity can be generalized beyond those inputs: an invertible object U is called *unitary* if it satisfies the condition $U^\dagger = U^{-1}$. U can be wilder objects in principle, such as an infinite dimensional matrix or a general operator; for instance, $U = \exp(i \frac{d}{dx})$ is a unitary operator for $x \in \mathbb{R}$.
35. Concept of unitarity is important, because there exists unitary matrices with functional entries which can be used as a basis to expand any given function. This is similar to Taylor series expansion: there, we used some sort of orthogonality of x^m and x^n for $m \neq n$ and used $\{x^i\}$ as a basis over which we expand $f(x)$. We state that similar basis (in fact, infinitely many of them) exist and we can expand any given function in terms of such basis consisting of an infinite set of particular unitary matrices of functional entries.
36. The modern way to understand such expansions is through *group theory*! Since we will not learn about these details, we only present a very important and general theorem: we will not dwell on its details and we will not try to do actual computations with that. We only present this to emphasize that Fourier transform (or spherical harmonics expansion, or Mellin transform, or many more) are special examples of a very general and fundamental branch of mathematics called *harmonic analysis*!

The main result of harmonic analysis is as follows:

$$f(g) = \int_{\hat{G}} d\pi \operatorname{tr} \left(\hat{\pi}(g)^{-1} \hat{f}(\pi) \right) \quad (3.12a)$$

$$\hat{f}(\pi) = \int_G dh \hat{\pi}(g) f(h) \quad (3.12b)$$

37. We will *not* discuss the details of above formula, and no one needs to know the following for this course! What we need to know is that there exist such a general result, and most of the stuff we see around (such as Fourier transform) are special cases of this general result! Nevertheless, for completeness, I list the ingredients as follows:

- G : space of a group, with dh being the measure in this space invariant under the action of the group (Haar measure)
- \hat{G} : space of the unitary irreducible representations of G , with $d\pi$ being the Plancheral measure
- $\hat{\pi} : g \rightarrow \operatorname{End}(V_\pi)$ is a map from the group element g to the space of endomorphisms on the vector space of the representations of the group. Such endomorphisms can be implemented as a matrix acting on this vector space, therefore $\hat{\pi}$ is a simple matrix on the representation space (which is why we take a trace in the first integral)
- $f(g)$ is a function on the group space
- $\hat{f}(\pi)$ is generalization of *Fourier coefficients*: this is a matrix in the representation space

38. We can make the following analogy:

Object	Basis	Decomposition	Components	Extracting components
\vec{v}	$\{\hat{i}, \hat{j}, \hat{k}\}$	$\vec{v} = v_x \hat{i} + v_y \hat{j} + v_z \hat{k}$	$\{v_x, v_y, v_z\}$ (a finite set)	$v_x = \vec{v} \cdot \hat{i}$
$f(x)$	$\{x^i\}$	$f(x) = \sum_{n=0}^{\infty} a_n x^n$	$\{a_n\}$ (countable infinite set)	Cauchy's integral formula, see 2.10
$f(g)$	$\{\hat{\pi}(g)\}$	$f(g) = \int_{\hat{G}} d\pi \operatorname{tr} \left(\hat{\pi}(g)^{-1} \hat{f}(\pi) \right)$	$\{\hat{f}(\pi)\}$ (uncountable infinite set)	$\hat{f}(\pi) = \int_G dh \hat{\pi}(g) f(h)$

39. Simplest example of Harmonic analysis is the Fourier transform.

In this case, $\hat{\pi}$ is a one-dimensional unitary function $\hat{\pi} = e^{-ikx}$, π are spanned by one continuous real parameter k (meaning $\hat{G} = \mathbb{R}$), and the original function space is one dimensional real line as well (hence $G = \mathbb{R}$ with $g = x$). One can derive that the Plancheral measure is $\frac{dk}{2\pi}$, hence we have

$$f(x) = \int_{\mathbb{R}} \frac{dk}{2\pi} (e^{-ikx})^{-1} \hat{f}(k) \quad (3.13a)$$

$$\hat{f}(k) = \int_{\mathbb{R}} dx e^{-ikx} f(x) \quad (3.13b)$$

hence we can specialize the table above for Fourier transform as

Object	Basis	Decomposition	Components
\vec{v}	$\{\hat{i}, \hat{j}, \hat{k}\}$	$\vec{v} = v_x \hat{i} + v_y \hat{j} + v_z \hat{k}$	$\{v_x, v_y, v_z\}$ (a finite set)
$f(x)$	$\{x^i\}$	$f(x) = \sum_{n=0}^{\infty} a_n x^n$	$\{a_n\}$ (countable infinite set)
$f(x)$	$\{e^{ikx}\}$	$f(x) = \int_{-\infty}^{\infty} dk e^{ikx} \hat{f}(k)$	$\left\{ \hat{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x) dx \right\}$ (uncountable infinite set)

40. We can now use periodicity information if we know that a function f satisfies $f(x) = f(x + a)$: the identification $x \sim x + a$ means some of our basis components should be absent as we dictate $e^{ikx} \sim e^{ik(x+a)}$. We see that this is possible only if k is actually discrete, i.e.

$$k = \frac{2\pi}{a}n \text{ for } n \in \mathbb{Z} \quad (3.14)$$

This leads to discrete time fourier series:

$$f(x) = \frac{1}{a} \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi n}{a}x} \hat{f}(n) \quad (3.15a)$$

$$\hat{f}(n) = \int_0^a dx e^{-i\frac{2\pi n}{a}x} f(x) \quad (3.15b)$$

41. In class, we introduced four different Fourier analysis:

- Fourier transform:

$$f : \mathbb{C} \rightarrow \mathbb{C} \quad (3.16a)$$

$$\hat{f} : \mathbb{C} \rightarrow \mathbb{C} \quad (3.16b)$$

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \hat{f}(k) \quad (3.16c)$$

$$\hat{f}(k) = \int_{-\infty}^{\infty} dx e^{-ikx} f(x) \quad (3.16d)$$

- Fourier series:

$$f : [a, a + T] \rightarrow \mathbb{C} \quad (3.17a)$$

$$\hat{f} : \mathbb{Z} \rightarrow \mathbb{C} \quad (3.17b)$$

$$f(x) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi n}{T}x} \hat{f}(n) \quad (3.17c)$$

$$\hat{f}(n) = \int_b^{b+T} dx e^{-i\frac{2\pi n}{T}x} f(x) \quad (3.17d)$$

for arbitrary $a, b \in \mathbb{R}$. Note that this is also applicable for periodic functions: for any periodic function with a period T , we can use Fourier series as instructed above!

- Discrete-time Fourier transform:

$$f : \mathbb{Z} \rightarrow \mathbb{C} \quad (3.18a)$$

$$\hat{f} : [a, a + T] \rightarrow \mathbb{C} \quad (3.18b)$$

$$f(n) = \frac{1}{T} \int_b^{b+T} dx e^{i \frac{2\pi n}{T} x} \hat{f}(x) \quad (3.18c)$$

$$\hat{f}(k) = \sum_{n=-\infty}^{\infty} e^{-i \frac{2\pi n}{T} k} f(n) \quad (3.18d)$$

$$(3.18e)$$

where \mathbb{Z}_N denotes the set $\{0, 1, \dots, N-1\}$.

- Discrete Fourier series:

$$f : \mathbb{Z}_N \rightarrow \mathbb{Z}_N \quad (3.19a)$$

$$\hat{f} : \mathbb{Z}_N \rightarrow \mathbb{Z}_N \quad (3.19b)$$

$$f(n) = \frac{1}{N} \sum_{m=0}^{N-1} e^{i \frac{2\pi nm}{N}} \hat{f}(m) \quad (3.19c)$$

$$\hat{f}(m) = \sum_{n=0}^{N-1} e^{-i \frac{2\pi nm}{N}} f(n) \quad (3.19d)$$

$$(3.19e)$$

for arbitrary $a, b \in \mathbb{R}$.

42. Summary: finite range/periodic in one domain \leftrightarrow discrete in the other domain
43. Drew plots of these transformations for visualization in class.
44. Introduced two higher order functions E and O :

$$E : (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \quad (3.20a)$$

$$E = (x \rightarrow f(x)) \rightarrow \left(x \rightarrow f_E(x) = \frac{f(x) + f(-x)}{2} \right) \quad (3.20b)$$

$$O : (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \quad (3.20c)$$

$$O = (x \rightarrow f(x)) \rightarrow \left(x \rightarrow f_O(x) = \frac{f(x) - f(-x)}{2} \right) \quad (3.20d)$$

with which any single-argument function satisfies $f = E \cdot f + O \cdot f$, or with a more common notation, $f(x) = f_E(x) + f_O(x)$. As any function can also be decomposed into its real and imaginary part, we arrive at

$$f(x) = f_{RE}(x) + f_{RO}(x) + i f_{IE}(x) + i f_{IO}(x) \quad (3.21)$$

45. Since Fourier transform is linear, we can reconstruct $(\text{F.T.} \cdot f)(k)$ from $(\text{F.T.} \cdot f_{RE})(k)$ and so on.

46. $f_{RE}(k)$ etc. and their Fourier transforms satisfy nice properties. We derived a few of them in the class, for instance taking complex conjugation and changing the dummy variable $x \rightarrow -x$ leads to

$$[(F.T. \cdot f_{RE})(k)]^* = (F.T. \cdot f_{RE})(k) \quad (3.22)$$

Likewise, taking $k \rightarrow -k$ and $x \rightarrow -x$ leads to

$$(F.T. \cdot f_{RE})(-k) = (F.T. \cdot f_{RE})(k) \quad (3.23)$$

hence we conclude Fourier transform of f_{RE} is itself even and real. On the contrary, Fourier transform of f_{RO} is purely-imaginary and odd.

47. For real functions, $(F.T. \cdot f_R)(-k) = [(F.T. \cdot f_R)(k)]^*$, hence we only need to know positive frequencies, consistent with our everyday experience.
48. Consider the Fourier transform of the function

$$f : \left[-\frac{T}{2}, \frac{T}{2}\right] \rightarrow \mathbb{C} \quad (3.24a)$$

$$f = x \rightarrow 1 \quad (3.24b)$$

We computed Fourier series expansion of this in class, found $\hat{f}(n) = T\delta_{n0}$ and checked consistency. We then considered the limit $T \rightarrow \infty$ and argued that Kronecker-delta function turns into Dirac-delta distribution.

49. We also considered the function

$$f : \mathbb{R} \rightarrow \mathbb{C} \quad (3.25a)$$

$$f = x \rightarrow \begin{cases} 1 & |x| \leq \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad (3.25b)$$

We computed the Fourier transformation of this, introduced the sinc function, discussed its properties, talked about its usage in single-slit experiment, optics, and signal process. We then also discussed how $T \rightarrow \infty$ takes sinc function to the Dirac-delta distribution.

3.2 Particular solution

Summary of what we have discussed in class (to be typed later, maybe next year):

1. The particular solution to a differential equation is unique: we haven't proven this but it is actually most easily seen if the unknown function is expanded in a basis where the action of the differential operator becomes *diagonal*; i.e., the differential equation then becomes algebraic and there is a unique solution for an algebraic equation $ax = b$ for x .

2. As particular solutions are unique, we can get away with guessing it if we can: the simplest way to find the particular solution is to guess it and then check that it satisfies the differential equation.
3. The next simplest thing we can try is to guess the *functional form* of the particular solution with some arbitrary coefficients and then fix them imposing the differential equation. This approach is called *method of undetermined coefficients*. As an example, for

$$xf'(x) - 2f(x) = 6x^4 \quad (3.26)$$

we can guess $f_p(x) = ax^b$: inserting it into the differential equation, we find that $f_p(x) = 2x^4$.

4. For differential equations with constant coefficients, we can find the particular solution more systematically as we have reviewed in the beginning of the semester: convolution of nonhomogeneous piece with the impulse response.
5. A similar systematic approach exists for more general differential equations: it is called *method of variation of parameters*. Let us consider a general linear ordinary differential equation

$$\left(\frac{d^n}{dx^n} + a_{n-1}(x) \frac{d^{n-1}}{dx^{n-1}} + a_0(x) \right) f(x) = h(x) \quad (3.27)$$

If we find the homogeneous solutions $f_1(x), \dots, f_n(x)$, then we know that the most general solution can be written as

$$f(x) = \sum_{i=1}^n c_i f_i(x) + f_p(x) \quad (3.28)$$

for arbitrary coefficients c_i . Here, we have one unknown function $f_p(x)$ and we trade it for n unknown functions by rewriting this equation as

$$f(x) = \sum_{i=1}^n c_i(x) f_i(x) \quad (3.29)$$

for undetermined functions $c_i(x)$, i.e. we *varied the parameters*. We can clearly do this, as we had 1 unknown functional degree of freedom and now we have n unknown functional degrees of freedom. In fact, we can impose $n - 1$ constraints, which we choose as

$$\sum_{i=1}^n c'_i(x) f_i^{(k-1)}(x) = 0 \quad \text{for } k = 1, 2, \dots, n-1 \quad (3.30)$$

If we now insert (3.29) into (3.27) and use these constraints and the fact that $f_i(x)$ are homogeneous solutions, we end up with

$$\sum_{i=1}^n c'_i(x) f_i^{(n-1)}(x) = g(x) \quad (3.31)$$

The equations (3.30) and (3.31) can be combined to solve for $c'_i(x)$ as

$$\begin{pmatrix} c'_1(x) \\ c'_2(x) \\ \dots \\ c'_n(x) \end{pmatrix} = \begin{pmatrix} f_1(x) & f_2(x) & \dots & f_n(x) \\ f'_1(x) & f'_2(x) & \dots & f'_n(x) \\ \dots & \dots & \dots & \dots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \dots & f_n^{(n-1)}(x) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ \dots \\ g(x) \end{pmatrix} \quad (3.32)$$

6. In summary, inverse of the Wronskian matrix and the nonhomogeneous piece is sufficient to find $c'_i(x)$. By integrating these, we get both the homogeneous solution (through the integration constants) and particular solution.
7. In practice, computer programs (such as Mathematica) are the best way to compute matrix inverses (avoid pen and paper if you can). However, it is better if we learn the math behind such implementations. To understand the computation of a matrix inverse, we need to define a new operation:

$$\text{adj} : \mathcal{M}_{n \times n}(\mathbb{C}) \rightarrow \mathcal{M}_{n \times n}(\mathbb{C}) \quad (3.33a)$$

$$\text{adj} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} \quad \text{where} \quad (3.33b)$$

$$b_{i_n k_n} = \frac{1}{(n-1)!} \epsilon_{i_1 \dots i_n} \epsilon_{k_1 \dots k_n} a_{i_1 k_1} \dots a_{i_{n-1} k_{n-1}} \quad (3.33c)$$

where adj yields the *adjugate* of the given matrix. We can now give the inverse of a matrix as

$$A^{-1} = \frac{\text{adj}(A)}{\det A} \quad (3.34)$$

where we have defined and discussed determinant before:

$$\det : \mathcal{M}_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C} \quad (3.35a)$$

$$\det = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} a_{1i_1} \dots a_{ni_n} \quad (3.35b)$$

8. Computed adjugate and inverse of a 2×2 matrix in the class.

4

Systems of first order linear differential equations

Summary of what we have discussed in class (to be typed later, maybe next year):

1. We reviewed in class that we have learned various methods to solve differential equations. In this chapter, we will see one last method: *conversion of generic linear ordinary differential equations to first order system of differential equations*.
2. Learning how to solve first order differential equations of matrices is important for multiple reasons:
 - a) Any linear ordinary differential equation can be rewritten as a first order differential equation of a column matrix.
 - b) There exists coupled systems which can only be solved through differential equations of matrices
 - c) There is a conceptual reason to consider such differential equations of matrices: phase space, using momenta, etc.
3. We introduced systems of linear ordinary differential equations: multiple unknown functions, one independent variable.
4. As an example, we considered a one-dimensional system of two masses attached to a wall via two springs (something like $| \sim \square \sim \square$), and showed that this system is described by a system of two differential equations. The unknown functions are two positions of the masses, the independent variable is the time, and the key point is that this is a coupled system: the differential equations cannot be solved independently! Indeed, the system is best described as

$$\frac{d^2}{dt^2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (4.1)$$

5. One can solve this differential equation by various ways. One common method is to find the normal modes of this system by diagonalizing the square matrix: this will tell us the combinations of x_1 and x_2 which oscillate independently: for those variables, we have two independent differential equations that can be solved separately. We will not discuss this further in this class.
6. Another approach to solve such a differential equation is to bring it to a first order form. We motivate this by using our physical intuition: momentum (in addition to position) also describes a

current state of the system, so we should work with four variables instead of two! Indeed, we can rewrite the above differential equation as

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ p_1 \\ x_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{m_1} & 0 & 0 \\ -(k_1 + k_2) & 0 & k_2 & 0 \\ 0 & 0 & 0 & \frac{1}{m_2} \\ k_2 & 0 & -k_2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ p_1 \\ x_2 \\ p_2 \end{pmatrix} \quad (4.2)$$

7. This approach is generalizable to a set of differential equations for m unknown functions where the highest order derivative for each unknown function is n_1, n_2, \dots, n_m . This system is equivalent to a first-order differential equation for a vector of $\sum_{i=1}^m n_i$ components!

8. As example, we solved $f''(x) + 3f'(x) + 2f(x) = 0$ by converting it into the matrix equation

$$\frac{d}{dx} V(x) + A \cdot V(x) = 0 \quad (4.3)$$

for

$$A = \begin{pmatrix} 0 & -1 \\ 2 & 3 \end{pmatrix} \quad (4.4)$$

and

$$V(x) = \begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} \quad (4.5)$$

We immediately say that the solution is

$$V(x) = (\exp\{-Ax\}) \cdot C \quad (4.6)$$

for the arbitrary column matrix C .

9. We computed $\exp(Ax)$ explicitly in class using its definition:

$$\exp(Ax) = \mathbb{I} + xA + \frac{x^2}{2!} A \cdot A + \frac{x^3}{3!} A \cdot A \cdot A + \dots \quad (4.7)$$

10. In class, we discussed that if A is t -independent, then we can use the generalization

$$[f'(t) = af(t) \rightarrow f(t) = \exp(at)c] \Rightarrow [V'(t) = A \cdot V(t) \rightarrow V(t) = \exp(At) \cdot C] \quad (4.8)$$

which means for any first-order differential equation with constant A is immediately solvable this way. A similar generalization exists even if A is t -dependent, but it is nontrivial:

$$f'(t) = a(t)f(t) \rightarrow f(t) = \exp\left(\int a(t)dt\right) c \quad (4.9)$$

however

$$V'(t) = A(t) \cdot V(t) \rightarrow V(t) \neq \exp\left(\int A(t)dt\right) \cdot C \quad (4.10)$$

The correct version is

$$V'(t) = A(t) \cdot V(t) \rightarrow V(t) = \mathcal{T} \left\{ \exp \left(\int A(t) dt \right) \right\} \cdot C \quad (4.11)$$

where \mathcal{T} is called *time-ordering operator*! There is a purely mathematical derivation of this via Volterra integral equation (we will see this), but the physical implications are rather important in its usage in quantum mechanics: the measurements at different times (by $A(t)$) should be time-ordered, i.e. causality has to be preserved! You will learn more about this when you solve Schrödinger's equation (a first order diff. equation of operators, which can be represented with matrices)!

11. We solved in class $f''(x) + 3f'(x) + 2f(x) = 0$ as an example. We already know that the characteristic equation is $(r+2)(r+1) = 0$ hence the answer is $f(x) = c_1 e^{-x} + c_2 e^{-2x}$. Nevertheless, let's see how we can get this answer through matrix computation.

We realize that this equation can be rewritten as

$$\frac{d}{dx} \begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} \quad (4.12)$$

hence we can immediately write down the answer as

$$\begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} = e^{\begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} x} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad (4.13)$$

In class, we calculated the exponentiation of this matrix by computing first few terms in the Taylor expansion, finding the pattern, and then resumming all terms.

12. We emphasize that this approach does not work if the matrix is not constant: as we mentioned above, we instead need the *time-ordering*. In class, we schematically derived this as follows.

Start with

$$\frac{d}{dt} V(t) = A(t) \cdot V(t) \quad (4.14)$$

Any non-homogeneous piece can be ignored as the particular solutions can always be found via variation of parameters method *after* homogeneous solutions are computed. By integrating this equation, we get *Volterra integral equation*

$$V(t) = V(0) + \int_0^t A(t') \cdot V(t') dt' \quad (4.15)$$

Now consider a modified version of this as

$$V(\epsilon, t) = V(0) + \epsilon \int_0^t A(t') \cdot V(\epsilon, t') dt' \quad (4.16)$$

where $V(t) = V(1, t)$. Let us expand $V(\epsilon, t)$ around $\epsilon = 0$ as follows:

$$V(\epsilon, t) = V_{(0)}(t) + \epsilon V_{(1)}(t) + \epsilon^2 V_{(2)}(t) + \dots \quad (4.17)$$

Note that $V(0, t) = V_{(0)}(t) = V(0)$ so this series is definitely well defined around $\epsilon \sim 0$, but we do not yet know its radius of convergence so we might in principle not be able to take $\epsilon \rightarrow 1$ to obtain $V(t)$ at the end. Nevertheless, we will see that the radius of convergence is actually infinite as the expansion will turn out to be that of the exponential function. For now, let us proceed by inserting this into the modified Volterra equation:

$$(V(0) + \epsilon V_{(1)}(t) + \epsilon^2 V_{(2)}(t) + \dots) = V(0) + \epsilon \int_0^t A(t') \cdot (V(0) + \epsilon V_{(1)}(t') + \epsilon^2 V_{(2)}(t') + \dots) dt' \quad (4.18)$$

which leads to

$$V_{(n+1)}(t) = \int_0^t A(t') \cdot V_{(n)}(t') dt' \quad (4.19)$$

if we match different orders of ϵ . But inserting this back, we obtain

$$V(\epsilon, t) = V(0) + \left[\epsilon \int_0^t A(t') dt' \right] \cdot V(0) + \left[\epsilon^2 \int_0^t dt' \int_0^{t'} dt'' A(t') \cdot A(t'') \right] \cdot V(0) + \dots \quad (4.20)$$

which can be rewritten as

$$V(\epsilon, t) = \left(\mathbb{I} + \left[\epsilon \int_0^t A(t') dt' \right] + \left[\epsilon^2 \int_0^t dt' \int_0^{t'} dt'' A(t') \cdot A(t'') \right] + \dots \right) \cdot V(0) \quad (4.21)$$

One can show that the integration ranges $\int_0^t dt' \int_0^{t'} dt''$ and $\int_0^t dt'' \int_0^{t''} dt'$ actually cover two triangles that add upto a square in the $t' - t''$ plane, given by the integration range $\int_0^t dt' \int_0^t dt''$. Therefore, we see that

$$\int_0^t dt' \int_0^t dt'' f(t', t'') = \int_0^t dt' \int_0^{t'} dt'' f(t', t'') + \int_0^t dt'' \int_0^{t''} dt' f(t', t'') \quad (4.22)$$

which can be rewritten by changing the dummy variable as

$$\int_0^t dt' \int_0^t dt'' f(t', t'') = \int_0^t dt' \int_0^{t'} dt'' (f(t', t'') + f(t'', t')) \quad (4.23)$$

Observe that if we define

$$f(t, t') = \begin{cases} g(t, t') & \text{if } t > t' \\ g(t', t) & \text{if } t' > t \end{cases} \quad (4.24)$$

then we get

$$\frac{1}{2} \int_0^t dt' \int_0^{t'} dt'' f(t', t'') = \int_0^t dt' \int_0^{t'} dt'' g(t', t'') \quad (4.25)$$

Note that f is just the function g in a way that its arguments are rearranged in the decreasing order. If we then define *time-ordering* \mathcal{T} as

$$\mathcal{T} : (\mathbb{R}^n \rightarrow \mathbb{C}) \rightarrow (\mathbb{R}^n \rightarrow \mathbb{C})$$

$$\mathcal{T} = \left((t_1, \dots, t_n) \rightarrow f(t_1, \dots, t_n) \right) \rightarrow \left((t_1, \dots, t_n) \rightarrow f \cdot \mathcal{O}_> \cdot (t_1, \dots, t_n) \right) \quad (4.26)$$

where $\mathcal{O}_>$ is the ordering function acting on *list of real numbers*

$$\begin{aligned} \mathcal{O}_> &: \mathbb{R}^n \rightarrow \mathbb{R}^n \\ \mathcal{O}_> &= (t_1, \dots, t_n) \rightarrow (t_{i_1}, \dots, t_{i_n}) \text{ such that } t_{i_a} \leq t_{i_b} \text{ if } a < b \end{aligned} \quad (4.27)$$

we then see that

$$\int_0^t dt' \int_0^{t'} dt'' g(t', t'') = \frac{1}{2} \int_0^t dt' \int_0^{t'} dt'' \mathcal{T} \cdot g(t', t'') \quad (4.28)$$

One can actually observe the same pattern in higher order versions, hence

$$\int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_n g(t_1, \dots, t_n) = \frac{1}{n!} \int_0^t dt_1 \dots \int_0^{t_1} dt_n \mathcal{T} \cdot g(t_1, \dots, t_n) \quad (4.29)$$

One can use this result in (4.21) and after some manipulation, we obtain

$$V(\epsilon, t) = \mathcal{T} \left(e^{\epsilon \int_0^t A(t') dt'} \right) \cdot V(0) \quad (4.30)$$

where \mathcal{T} is understood to apply products of A when exponential is expanded. Since the Taylor expansion of exponential function has infinite radius of convergence, we can take $\epsilon \rightarrow 1$ and obtain the final result:

$$V(t) = \mathcal{T} \left(e^{\int_0^t A(t') dt'} \right) \cdot V(0) \quad (4.31)$$

In Physics, this is mostly known as Dyson series.

13. In class, I emphasized that this is a hard topic and the students are not expected to understand it fully. No question was asked about this in any homework or examination.

5

Eigensystems and Sturm-Liouville theory

1. Getting back to the simple case of constant $A(t)$, we solved some explicit examples in class. We considered an RLC circuit of resistor, capacitor, and inductor in parallel (and nothing else). We have shown in class that this system is coupled, hence one needs to work with matrices to find the, say, current over inductor as a function of time if there was initial energy in the system. Indeed, the system is described by

$$\frac{d}{dt} \begin{pmatrix} I_L(t) \\ V(t) \end{pmatrix} = \begin{pmatrix} 0 & 1/L \\ -1/C & -1/(RC) \end{pmatrix} \begin{pmatrix} I_L(t) \\ V(t) \end{pmatrix} \quad (5.1)$$

We solved in class for $C = 1mF$, $L = 200H$, and $R = 1/6kOhm$ via exponentiation of the matrix. As a teaser, I also showed that the differential equations for the variables $S_1 = I_L + 10^{-3}V$ and $S_2 = I_L + 510^{-3}V$ are actually decoupled: these are called normal modes, and we will learn about how to get these (finding the eigenvalues and eigenfunctions of a system) below!

2. How to compute exponential of a matrix? As an example, consider

$$\begin{aligned} e^{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \frac{1}{2} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \frac{1}{3!} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \dots \\ &= \begin{pmatrix} f_{11}(a, b, c, d) & f_{12}(a, b, c, d) \\ f_{21}(a, b, c, d) & f_{22}(a, b, c, d) \end{pmatrix} \end{aligned} \quad (5.2)$$

This is tedious!

Observe that

$$A = U \cdot D \cdot U^{-1} \quad e^A = U \cdot e^D \cdot U^{-1} \quad (5.3)$$

which one can check by expanding matrices. If D is diagonal, then

$$e^A = U \cdot \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & e^{\lambda_n} \end{pmatrix} \cdot U^{-1} \quad (5.4)$$

where λ_i are diagonal entries of the matrix D .

3. We now try to solve the question: how do we find the matrix U for a given A ? If we write U in terms of its column vectors as $U = (\vec{u}_1 \quad \vec{u}_2 \quad \dots \quad \vec{u}_n)$, then $A \cdot U = U \cdot D$ implies

$$A \cdot (\vec{u}_1 \quad \vec{u}_2 \quad \dots \quad \vec{u}_n) = (\lambda_1 \vec{u}_1 \quad \lambda_2 \vec{u}_2 \quad \dots \quad \lambda_n \vec{u}_n) \quad (5.5)$$

which implies

$$A \cdot \vec{u}_i = \lambda_i \vec{u}_i \quad \forall i \quad (5.6)$$

u_i (we'll drop the vector sign for simplicity) are called *right eigenvectors* of A , and λ_i are called *eigenvalues* of A . We can find an analogous equation in terms of *left eigenvectors* of A ($\vec{v}_i \cdot A = \vec{v}_i \lambda_i$), where \vec{v} are actually row vectors of U (unlike \vec{u} , which are column vectors of U). We'll only work with right eigenvectors so we'll drop the adjective right from now on.

4. Observe that finding the matrix U , which is also called *modal matrix*, is equivalent to solving the equation (5.6), which can be rewritten as

$$(A - \lambda_i \mathbb{I}) \cdot u_i = 0 \quad (5.7)$$

If $(A - \lambda_i \mathbb{I})$ is invertible, then we contract the equation above with $(A - \lambda_i \mathbb{I})^{-1}$ from the left, which gives $u_i = (A - \lambda_i \mathbb{I})^{-1} \cdot 0$ hence the only solution is the trivial solution $u_i = 0$. Therefore, for nontrivial u_i , we need $(A - \lambda_i \mathbb{I})$ to be non-invertible, which means its determinant is zero:

$$\det(A - \lambda_i \mathbb{I}) = 0 \quad (5.8)$$

This gives us a polynomial in λ_i , which is called *characteristic polynomial* of A . For the matrix A that describes a single linear ordinary differential equation with constant coefficients, the characteristic polynomial reduces to the characteristic equation that we have learned earlier.

5. In class, we have solved two examples: the RLC circuit earlier, and $f''(x) + 3f'(x) + 2f(x) = 0$. We found their eigensystem (and hence normal modes).
6. Discussed physical importance of this *similarity transformation*, i.e.

$$A = U \cdot D \cdot U^{-1} \quad (5.9)$$

where U is called a similarity transformation and the diagonal matrix D is called the spectrum. This mathematical terminology is in line with our physical intuitions and usage of the term spectrum. Indeed, in optics, electromagnetic theory, atomic & nuclear physics, and chemical physics, we use the term spectrum to refer to the different frequencies of light in the emission (or absorption) spectrum of some system/material. These different frequencies are actually the eigenvalues of the system as monochromatic light with different frequencies are *eigenfunctions* that form a basis on which any electromagnetic disturbance can be decomposed. This is most easily seen in a prism where light gets decomposed into monochromatic light (remember Pink Floyd :D): this is simply expansion of white light in the basis of eigenfunctions of the light (e^{ikx}). In fact, the Harmonic analysis that we have learned before (Fourier transform being

one example) are generally expansions over eigenfunctions of a differential operator! (that differential operator being Casimir operator and eigenfunctions being unitary irreducible representations of the relevant group, but this is a story for another time!)

7. Extending this whole discussion of spectrum and matrix transformation to infinite dimensions is equivalent to extending our eigenvalue equation to functions from finite dimensional column vectors: we now have

$$D \cdot f(x) = \lambda f(x) \quad (5.10)$$

for a differential operator D , where $f(x)$ is its eigenfunction and λ is its eigenvalue.

8. The analysis of eigensystem of operators is immensely important, but to make further progress, we need to define *inner products*!
9. Remind dot products over real vector spaces $(\cdot) : (\mathbb{R}^n, \mathbb{R}^n) \rightarrow \mathbb{R}$, extend them to functions $\mathbb{R} \rightarrow \mathbb{R}$ by defining the inner product $\langle \cdot, \cdot \rangle : (\mathbb{R} \rightarrow \mathbb{R}, \mathbb{R} \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ as $\langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x)dx$, and then generalize this further to complex functions as

$$\langle \cdot, \cdot \rangle_{\omega} : (A \rightarrow \mathbb{C}, A \rightarrow \mathbb{C}) \rightarrow \mathbb{C} \quad (5.11a)$$

$$\langle f, g \rangle_{\omega} = \int_A \left(f(x) \right)^* g(x) \omega(x) dx \quad (5.11b)$$

for $A \subseteq \mathbb{R}$.

10. In class, we defined adjoint of an operator wrt an inner product, discussed its properties, showed that self-adjoint operators have a real spectrum and orthogonal eigenfunctions. Add further discussion regarding symmetric vs self-adjoint operators.
11. Sturm-Liouville problem is basically analysis of the eigensystem of *second order self adjoint differential operators*.

Beyond linear ordinary differential equations

1. Talked about abstract concept of a system S , with input I , and with output O , where their types are \mathcal{S} , \mathcal{I} , and \mathcal{O} . Talked about how systems can be represented in different frameworks, as differential equations, as algebraic equations, or etc. Discussed how certain transformations actually act like a map between different representations of a system, e.g. Fourier transform can take a differential representation in one domain to an algebraic representation in another domain.
2. Discussed how systems can have properties (such as bounded input bounded output, casual, linear, etc.), and how we have been focused on linear systems with single input (hence can be represented by an linear ordinary differential equations), and how natural generalizations to nonlinear systems or multiple inputs lead to non-linear and/or partial differential equations.

6.1 Non-linear ordinary differential equations

1. Discussed exact equations, separable equations, Bernoulli equations, and solved examples.

6.2 Partial differential equations

1. Introduced *tuples*, which are ordered lists:

$$(x_1, x_2, \dots, x_n) :: T_1 \times T_2 \times \dots \times T_n \quad (6.1)$$

where \times denotes a product, and $T_1 \times T_2 \times \dots \times T_n$ is called a *product type*. \times creates a new product type from individual types; its inverse is *projection* denoted by π_i which works such that

$$\pi_i :: (T_1 \times T_2 \times \dots \times T_n) \rightarrow T_i \quad (6.2a)$$

$$\pi_i = (x_1, x_2, \dots, x_n) \rightarrow x_i \quad (6.2b)$$

What does it mean to take products of types? What do we really mean by, say, $\mathcal{M}_{n \times n}(\mathbb{C}) \times \mathbb{R}$?

- In computer science, vast applications via so-called algebraic data types
- In logic, product of types are actually related to \wedge (and) operation via Curry-Howard correspondence
- In abstract math, product of any family of objects is defined as “*the most general thing*” from which individual objects can be extracted (category theory).

We are really used to taking products of types in Physics; for instance, Cartesian plane is $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is simply product of two real types. However, there is actually subtlety with such products in more general cases; for instance, the mathematical framework for quantum information theory is the so-called **Hilb** category which does not admit a Cartesian product (the only available products are non-cartesian), hence we do not actually have projection operations in this category. The well known physical result of this mathematical fact is that you cannot copy information in a quantum computer!

2. If the elements of a tuple are of the same type, we use the easier notation $(x_1, \dots, x_n) :: T^n$.
3. Multi-variable functions are functions from tuples:

$$f :: \mathbb{R}^3 \rightarrow \mathbb{R} \quad (6.3a)$$

$$f = (x_1, x_2, x_3) \rightarrow f(x_1, x_2, x_3) \quad (6.3b)$$

4. Introduced D-notation:

$$D :: \mathbb{Z}_+ \rightarrow (\mathbb{R}^n \rightarrow \mathbb{C}) \rightarrow (\mathbb{R}^n \rightarrow \mathbb{C}) \quad (6.4a)$$

$$D = i \rightarrow ((x_1, \dots, x_n) \rightarrow f(x_1, \dots, x_n)) \rightarrow \left((x_1, \dots, x_n) \rightarrow \frac{\partial f(x_1, \dots, x_n)}{\partial x_i} \right) \quad (6.4b)$$

$$\text{e.g. } D_2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial y^2}$$

5. Chain rule:

$$\frac{d}{dx} f(g_1(x), \dots, g_k(x)) = \sum_{i=1}^k \left(\frac{d}{dx} g_i(x) \right) D_i f(g_1(x), \dots, g_k(x)) \quad (6.5)$$

6. Consider two functions f and g :

$$\begin{aligned} f &:: (X \times Y) \rightarrow Z \\ g &:: X \rightarrow (Y \rightarrow Z) \end{aligned} \quad (6.6)$$

We can *choose* g such that $g(x)(y) = f(x, y)$. This is called *currying*, i.e. we curried f into this new form as g . The reverse (rewriting a function like g as a function like f) is called *uncurrying*.

7. Discussed several important examples of partial differential equations:

$$D_1 f(t, x) - \alpha^2 D_2^2 f(t, x) = 0 \quad (\text{heat conductance equation}) \quad (6.7a)$$

$$D_1^2 f(t, x) - c^2 D_2^2 f(t, x) = 0 \quad (\text{wave equation}) \quad (6.7b)$$

$$D_1^2 f(t, x) + D_2^2 f(t, x) = 0 \quad (\text{Laplace equation}) \quad (6.7c)$$

8. We solved these equations both via method of separation of variables, and also proposing solutions of the form $f(x, t) = g(x+at)$.
9. We discussed how partial differential equations can have undetermined functions in the answer just like how ordinary differential equations have undetermined coefficients in the answer!

Part II

**Vector Calculus for
Physicists**

7

Definition and fundamentals of Vector Spaces

7.1 Preliminaries: some basic terminology

7.1.1 Primer: groups, fields, and algebras

The vector spaces are some structures build on fields, which are themselves built on groups. The groups are built on sets. So we start with sets.

Our practical reason to start with sets is that we need to properly understand them to understand vectors, but one can even say that set theory is the basis of all mathematics! Indeed, historically, 20th century Mathematicians did try to build the foundations of math on sets, but Russell's paradox squeezed in and ruined the whole program, at least to the degree it was envisioned back then. Today, even more fundamental approaches (such as category theory or type theory) try to explain whole of math, but this is beyond our scope: sets are good enough to be the foundation of everything we will care in this book, so we will stick with them to prepare the background for vector spaces.

Sets: A set is a collection of objects, and its elements determine what the set is. Importantly, we know all of the elements of a set!^{ay}

Sets can be denoted as explicit list of elements within curly brackets;^{az} for instance

$$A :: \text{Set} \quad (7.4a)$$

$$A = \{1, 2, 3\} \quad (7.4b)$$

Equation (7.4b) shows that the set A consists of the elements 1, 2, and 3. Equation (7.4a) on the other hand tells us that A is of type *Set*, i.e. A is a set: for more details about this notation, please check out Section 1.1.1.

Sets can have infinitely many elements and we can use dots to infer the rest from the given; for instance, the set $\{\dots, -1, 1, 3, 5, \dots\}$ can denote the set of all odd integers. Although rather convenient, this method is inherently ambiguous; a more proper way to denote sets of infinite elements is through *set comprehensions*.

A set comprehension is a way to build a set, and it is usually used with *predicates*.^{bb} There are two conventional ways to write down a set A this way:

- $A = \{x \in \text{Domain} \mid \text{conditions}\}$
- $A = \{f(x) \mid \text{conditions}\}$ which is just an abbreviated version of the method above in the sense that this form actually means

$$A = \{y \in \text{Domain} \mid \exists x \text{ s.t. } y = f(x) \ \& \ \text{conditions}\} \quad (7.6)$$

^{ay}This is actually a rather important point for deep mathematical reasons that we will not visit, so what follows in this footnote is beyond the scope of this book (rather disappointing for me).

The fact that we cannot know everything in a quantum system suggests that we should probably not use sets for quantum information, and this is actually true! The right mathematical object to use in quantum system does not have the type *Set* but has the type *Hilb*, or more explicitly, quantum systems belong to a different *category* than the category of sets! An explicit discussion of these would require us to cover category theory, but an interested reader can consult here [add sources](#).

^{az}Note that sets are orderless^{ba} and that duplication of elements do not change a set, e.g.

$$A, B, C :: \text{Set} \quad (7.1a)$$

$$A = \{1, 2, 3, 4\} \quad (7.1b)$$

$$B = \{1, 4, 3, 2\} \quad (7.1c)$$

$$C = \{1, 2, 3, 4, 3, 2, 1\} \quad (7.1d)$$

implies $A = B = C$.

^{ba} Ordered sets are denoted with *round* brackets, e.g.

$$A :: \text{Ordered Set} \quad (7.2a)$$

$$A = (1, 2, 3) \quad (7.2b)$$

where $(1, 2, 3) \neq (1, 3, 2)$. Ordered sets can be constructed via usual sets, e.g.

$$\begin{aligned} (1, 2, 3) &\leftrightarrow \{\{1\}, \{1, 2\}, \{1, 2, 3\}\} \\ (1, 3, 2) &\leftrightarrow \{\{1\}, \{1, 3\}, \{1, 2, 3\}\} \end{aligned} \quad (7.3)$$

^{bb}Predicates are functions with the codomain *Boolean*.^{bc}In other words, a function is a predicate if it only yields two outputs: *True*, or *False*. For instance, the function “isOddInteger” defined as

$$\text{isOddInteger} :: \text{Boolean} \quad (7.5a)$$

$$\text{isOddInteger} = x \rightarrow \left(\frac{x+1}{2} \in \mathbb{Z} \right) \quad (7.5b)$$

is a *predicate*, e.g. $\text{isOddInteger}(1) = \text{True}$, $\text{isOddInteger}(2) = \text{False}$, $\text{isOddInteger}(5/2) = \text{False}$, etc.

For instance, the rigorous way to denote the odd integers, compared to a convenient casual notation $\{\dots, -3, -1, 1, 3, \dots\}$, is $\{a \in \mathbb{Z} \mid (\exists n \in \mathbb{Z}) a = 2n + 1\}$ by the first approach.^{bd bf} In comparison, the second approach makes it simpler: $\{2n + 1 \mid n \in \mathbb{Z}\}$.

I will assume that the readers are familiar with the comparisons of sets^{bh} and miscellaneous details of sets.^{bi} I will also skip discussing the details of operations between sets.^{bj} Lastly, I would like to mention the relevant concepts such as *relations*, *partitions*, and *quotients*: one should know about them for a good general mathematical background, but they will unfortunately not be covered in this book.

Groups: Consider a set S and impose the existence of two functions o and i such that

$$S :: \text{Set} \quad (7.11a)$$

$$o :: (S, S) \rightarrow S \quad (7.11b)$$

$$i :: S \rightarrow S \quad (7.11c)$$

where S is of type set, and the function o has the domain “tuple of S ” and the codomain S —check out section 6.2 for more details about tuples. In simpler terms, the function o takes two inputs (both being elements of the set S) and yields an element of the set S ; in comparison, the function i takes an element of S to another (not necessarily different) element of S .

The set S with these two functions o and i form a group if the following statements are true:

1. $\boxed{(\exists e \in S)(\forall s \in S) o(e, s) = o(s, e) = s}$. In words, this says that “There exists an element e of the set S and $o(e, s) = o(s, e) = e$ for any element s of this set”. The element e is called the *identity element* with respect to the group operation o .^{bk}
2. $\boxed{(\forall s \in S) o(s, i(s)) = o(i(s), s) = e}$. In words, this says that “The group operation o acting on the tuple $(s, i(s))$ or $(i(s), s)$ yields the identity element e , for any element s of the set S ”. In practice, this condition specifies the action of i on any element s , and $i(s)$ is called the *inverse* of the element s .
3. $\boxed{(\forall a, b, c \in S) o(a, o(b, c)) = o(o(a, b), c)}$. This is called *associative law*, and it is best visualized when we use *infix notation*: if we denote $o(a, b)$ as $a * b$, then this condition simply reads as $a * (b * c) = (a * b) * c$. Therefore, the parentheses are not important, and we can even loosely denote without parentheses: $a * (b * c) = (a * b) * c = a * b * c$.

We will *define* a group to be the pair (S, o) , i.e.

$$\text{Group} = \left(\text{Set}, (\text{Set}, \text{Set}) \rightarrow \text{Set} \right) \quad (7.12)$$

^{bc}This is the definition due to *Gottlob Frege*, the father of axiomatic first-order logic. You may find other definitions in different branches of math: an interested reader can check Wikipedia for more.

^{bd}This would read as “the set of integer a ’s for which there exists an integer n such that $a = 2n + 1$ is true”.^{be}

^{be}We can understand this more deeply as follows. Consider the *predicate* P defined as

$$\begin{aligned} P &:: \mathbb{Z} \rightarrow \text{Boolean} \\ P &= a \rightarrow (\exists n \in \mathbb{Z}) a = 2n + 1 \end{aligned} \quad (7.7)$$

The set of odd integers are then constructed by using this predicate with the list comprehension, i.e.

$$\text{odd integers} = \{a \in \mathbb{Z} \mid P(a)\} \quad (7.8)$$

This also shows how the list comprehension works: $S = \{T \mid C\}$ means that the set S consists of all elements of T for which the condition C is true.

^{bf}We remind the reader that \exists reads as “there exists...”, and it is one of the *quantifiers* in logic. The other two important ones that we will make use of are \forall and $\exists!$, which denote “for all...” and “there exists a unique...” respectively. For instance, $(\exists! n \in \mathbb{Z}) n^2 = n$ is the statement “there exists a unique integer n such that $n^2 = n$ ”, and it is False as both 0 and 1 satisfy $n^2 = n$: $\boxed{((\exists! n \in \mathbb{Z}) n^2 = n) = \text{False}}$ ^{bg}

It is no coincidence that I put the dots to the right of the expressions for the quantifiers; one should read them as in a sentence, hence their order matters. For instance

$$\left((\forall a \in \mathbb{Z})(\exists b \in \mathbb{Z}) b = 2a \right) = \text{True} \quad (7.9a)$$

$$\left((\exists b \in \mathbb{Z})(\forall a \in \mathbb{Z}) b = 2a \right) = \text{False} \quad (7.9b)$$

where the first expression is “For all integers, there exists an integer which is twice the former” and the second one is “There exists an integer which is twice all integers”.

^{bg}It may seem weird at first that there is a variable in the left hand side of the equality $((\exists! n \in \mathbb{Z}) n^2 = n) = \text{False}$, which does not appear on the right hand side. The variable n here is a *dummy/scooping/bound* variable hence the right hand side should be independent of it, just like the function $g(y)$ in $\int f(x, y)dx = g(y)$ being independent of x .

such that the above statements are true (satisfaction of those rules will in turn determine the inverse function $i : S \rightarrow S$ as well). For instance, the pair $(\mathbb{Z}, +)$ is a group, where $+$ denotes the arithmetic addition, and the inverse function turns out to be multiplication by -1 : we say that *integers form a group under arithmetic addition*. We can indeed check that the above statements are true:

1. $(\forall a \in \mathbb{Z}) a + 0 = 0 + a = a$ (addition by 0 does not change any integer)
2. $(\forall a \in \mathbb{Z}) a + (-a) = (-a) + a = 0$ (addition by the inverse element takes an integer to zero)
3. $(\forall a, b, c \in \mathbb{Z}) a + (b + c) = (a + b) + c$ (the order of addition does not change the result)

As another example for groups, consider the set $\text{classrooms} = \{P1, P2, P3\}$ and the group operation denoted as $\langle \rangle$ in infix notation. If we are given the information

$$\begin{aligned}
 P1 \langle \rangle P1 &= P3 \\
 P2 \langle \rangle P2 &= P2 \\
 P3 \langle \rangle P3 &= P1 \\
 P1 \langle \rangle P2 &= P2 \langle \rangle P1 = P1 \\
 P1 \langle \rangle P3 &= P3 \langle \rangle P1 = P2 \\
 P2 \langle \rangle P3 &= P3 \langle \rangle P2 = P3
 \end{aligned} \tag{7.13}$$

we can immediately deduce that $P2$ is the identity element, and $P1^{-1} = P3$, $P2^{-1} = P2$, & $P3^{-1} = P1$ (inverse operations). One can check that the necessary conditions are satisfied, hence we state

$$(\text{classrooms}, \langle \rangle) : \text{Group} \tag{7.14}$$

We will leave further details of groups to the interested reader to research! One important exception will be the concept of *commutativity*: if the group operation is commutative, i.e. $a * b = b * a$, then that group is called a *commutative group*. In commutative groups, one conventionally use the infix operator $+$ instead of $*$ for the group operation, $-s$ instead of s^{-1} to denote inverse operation, and 0 instead of 1 to denote identity element. For instance, we can rewrite the example in eqn. (7.13) as $\text{classrooms} = \{P1, P2, P3\} = \{P1, 0, -P1\}$. Note that the minus ($-$) sign and the symbol (0) has nothing to do with the arithmetic negation and the number 0 in general: this is simply a conventional notation.

Rings: Consider a set S and two functions $+$ and \cdot such that

$$S : \text{Set} \tag{7.15a}$$

$$+ : (S, S) \rightarrow S \tag{7.15b}$$

$$\cdot : (S, S) \rightarrow S \tag{7.15c}$$

For more details about this, see the comprehensive footnote [aw](#).

^{bh} $A \subseteq B$ (or $B \supseteq A$) means that all elements of A are also elements of B (A is called a *subset* of B), and $A \subset B$ (or $B \supset A$) means the same thing with the additional condition that $A \neq B$ (A is called a *proper subset* of B).

^{bi}For instance, $|S|$ denotes the number of elements in the set S , e.g. $|\{1, 2, 4\}| = 3$. If the set S is infinite, we write $|S| = \infty$. As expected, $S \subseteq T$ implies $|S| \leq |T|$. I would also like to introduce *power sets*: the power set of a set S is denoted by $\mathcal{P}(S)$ and it is the set whose elements are the subsets of the set S . For example,

$$S, \mathcal{P}(S) : \text{Set} \tag{7.10a}$$

$$S = \{a, b, 1\} \tag{7.10b}$$

$$\mathcal{P}(S) = \{\{\}, \{a\}, \{b\}, \{1\}, \{a, b\}\} \tag{7.10c}$$

$$\{a, 1\}, \{b, 1\}, \{a, b, 1\}\} \tag{7.10d}$$

We note that if the set S is finite, we have $|\mathcal{P}(S)| = 2^{|S|}$, which is indeed the case in the above example ($8 = 2^3$). This relation is why the power set of S is also loosely denoted as 2^S .

^{bj}Six important operations to know are *union*, *intersection*, *difference* (also called *complement*), *Cartesian product*, *disjoint union*, and *quotient by an equivalence relation*. I expect the reader to know the first four operations: the last two will not be relevant for this book, but I advice the interested reader to learn more about them.

^{bk}Suppose for a second that there are two identity elements: e_1, e_2 . Since e_1 is identity element, we need $o(e_1, e_2) = e_1$; but the same argument for e_2 dictates $o(e_1, e_2) = e_2$: we need $e_1 = e_2$. In conclusion, if identity element exists, it is *unique*!

where we will denote the functions in the infix form, e.g. $a + b$ instead of $+(a, b)$. We say that the triplet $(S, +, \cdot)$ is a ring, i.e.

$$(S, +, \cdot) :: \text{Ring} \quad (7.16)$$

if the following statements are true:

1. $(S, +)$ is a commutative group
2. $(\forall a, b, c \in S) a \cdot (b + c) = a \cdot b + a \cdot c$
3. $(\forall a, b, c \in S) (b + c) \cdot a = b \cdot a + c \cdot a$

In other words, rings are simply commutative groups with an additional operation (call multiplication) that distributes over group addition.

The most familiar rings are those where the ring addition and multiplication operations are literally arithmetic addition and multiplication; for instance, $(\mathbb{Z}, +, \cdot)$ forms a ring and so does $(\mathbb{R}, +, \cdot)$. Likewise, rectangular matrices with complex entries form a ring under matrix addition and matrix multiplication, i.e. $(\mathcal{M}_{n \times n}(\mathbb{C}), +, \cdot) :: \text{Ring}$.

Skew field: We say that a triplet $(S, +, \cdot)$ is a skew field, i.e.

$$(S, +, \cdot) :: \text{Skew Field} \quad (7.17)$$

if the following statements are true:

1. $(S, +, \cdot)$ is a ring
2. $(S \setminus \{0\}, \cdot)$ is a group

In other words, skew fields are rings where multiplication operation also forms a group if we remove the identity element of the addition operation. The prototypical example of the skew fields is the *quaternions*.^{bt}

Fields: We say that a triplet $(S, +, \cdot)$ is a field, i.e.

$$(S, +, \cdot) :: \text{Field} \quad (7.18)$$

if the following statements are true:

1. $(S, +, \cdot)$ is a ring
2. $(S \setminus \{0\}, \cdot)$ is a commutative group

Fields are ubiquitous and super important for physicists as we use them everywhere! The rational, real, and complex numbers all form fields under ordinary arithmetic addition and multiplication.^{bm}

$$(\mathbb{Q}, +, \cdot) :: \text{Field}, \quad (\mathbb{R}, +, \cdot) :: \text{Field}, \quad (\mathbb{C}, +, \cdot) :: \text{Field} \quad (7.26)$$

^{bt}Put some discussion of quaternions.

^{bm} For a more exotic example, consider the following set:

$$A = \left\{ (x \rightarrow x + a) \mid a \in \mathbb{R} \right\} \quad (7.19)$$

for the ordinary arithmetic addition. This is a set of *functions*, i.e. elements of the set are functions; for instance, one of the elements of this set is $x \rightarrow x + 1$: this is a function that adds one to its input.

Let us rewrite this set in a more familiar way. For this, we define the *higher order function* “plus”:

$$\begin{aligned} \text{plus} &:: \mathbb{R} \rightarrow (\mathbb{R} \rightarrow \mathbb{R}) \\ \text{plus} &= a \rightarrow (x \rightarrow (x + a)) \end{aligned} \quad (7.20)$$

This higher order function becomes an ordinary function if given an input, i.e. $\text{plus}(1) :: \mathbb{R} \rightarrow \mathbb{R}$, which can then be given one more input that converts it into a real number: $\text{plus}(1)(5) = 6 :: \mathbb{R}$. Indeed, this higher order function simply adds its inputs, i.e. $\text{plus}(a)(b) = a + b$. In terms of this higher order function, we can write down the set A above as

$$A = \left\{ \text{plus}(a) \mid a \in \mathbb{R} \right\} \quad (7.21)$$

Let us now introduce two higher order functions:

$$\oplus, \otimes :: ((\mathbb{R} \rightarrow \mathbb{R}), (\mathbb{R} \rightarrow \mathbb{R})) \rightarrow (\mathbb{R} \rightarrow \mathbb{R}) \quad (7.22)$$

These functions take a pair of functions, and then give another function! Let us use these guys in infix form, e.g. $\text{plus}(a) \oplus \text{plus}(b)$.

If we are given the information that

$$\text{plus}(a) \oplus \text{plus}(b) = \text{plus}(a + b) \quad (7.23)$$

we can immediately say (A, \oplus) forms a commutative group, with the identity element $\text{plus}(0)$. If we are given the further information

$$\begin{aligned} (\text{plus}(a) \oplus \text{plus}(b)) \otimes \text{plus}(c) &= \\ (\text{plus}(a) \otimes \text{plus}(c)) \oplus (\text{plus}(b) \otimes \text{plus}(c)) & \end{aligned} \quad (7.24a)$$

$$\begin{aligned} \text{plus}(c) \otimes (\text{plus}(a) \oplus \text{plus}(b)) &= \\ (\text{plus}(c) \otimes \text{plus}(a)) \oplus (\text{plus}(c) \otimes \text{plus}(b)) & \end{aligned} \quad (7.24b)$$

we can then say that (A, \oplus, \otimes) forms a ring! Finally, if we are also given

$$\text{plus}(a) \otimes \text{plus}(b) = \text{plus}(a \cdot b) \quad (7.25)$$

then we can prove that (A, \oplus, \otimes) is actually a field!

Vector space (Linear space): We define “a vector space over a field F ” as the triple (V, \oplus, \odot) , i.e.

$$V :: \text{Set} \quad (7.27a)$$

$$(F = (S, +, \cdot)) :: \text{Field} \quad (7.27b)$$

$$\oplus :: (V, V) \rightarrow V \quad (7.27c)$$

$$\odot :: (S, V) \rightarrow V \quad (7.27d)$$

$$(V, \oplus, \odot) :: \text{Vector Space} \quad (7.27e)$$

if the following statements are true

1. $(V, \oplus) :: \text{Commutative Group}$
2. $(\forall v \in V) 1 \odot v = v$ where 1 is the identity element of \cdot
3. $(\forall v \in V)(\forall s \in S) s \odot v \in V$
4. $(\forall v \in V)(\forall a, b \in S) (a \cdot b) \odot v = a \odot (b \odot v)$
5. $(\forall v \in V)(\forall a, b \in S) (a + b) \odot v = (a \odot v) \oplus (b \odot v)$
6. $(\forall v, w \in V)(\forall s \in S) s \odot (v \oplus w) = (s \odot v) \oplus (s \odot w)$

Here, the elements of the underlying field F are called *scalars* ($s \in S$) and the elements of the linear spaces are called *vectors* ($v \in V$). indeed, V above is the set of these vectors, \oplus is the *vector addition* operation, and \odot is the *scalar multiplication* operation. We will see more properties of vectors (such as basis, linear independence, etc.) in the next section.

The standard example of a linear space is this:

$$((\mathcal{M}_{n \times m}(S), \oplus), (S, +, \cdot)) :: \text{Linear Space} \quad (7.28)$$

In words, we say that all $m \times n$ matrices from a field $\kappa = (S, +, \cdot)$ form a linear space over κ . In this example, matrices *are* vectors of this linear space and the vector addition \oplus is simply defined as the matrix addition. A more familiar example can be given as follows:

$$\begin{aligned} V &= \{a\mathbf{x} + b\mathbf{y} + c\mathbf{z} \mid a, b, c \in \mathbb{R}\} \\ (a_1\mathbf{x} + b_1\mathbf{y} + c_1\mathbf{z}) \oplus (a_2\mathbf{x} + b_2\mathbf{y} + c_2\mathbf{z}) &= (a_1 + a_2)\mathbf{x} + (b_1 + b_2)\mathbf{y} + (c_1 + c_2)\mathbf{z} \\ s \odot (a\mathbf{x} + b\mathbf{y} + c\mathbf{z}) &= (s \cdot a)\mathbf{x} + (s \cdot b)\mathbf{y} + (s \cdot c)\mathbf{z} \end{aligned} \quad (7.29)$$

where we defined the set V and the explicit action of the functions \oplus and \odot .^{bn}

1. (Linear algebra=linear space with vector multiplication) is introduced and discussed.
2. (Lie algebras=linear algebra with multiplication being antisymmetric) is introduced and discussed.

^{bn} Let us consider a more nontrivial example. Consider the field (A, \oplus, \otimes) discussed in footnote [bm](#). We will define a vector space (W, \square, \odot) over this field. Note that, in this example, \square will denote the vector addition whereas \oplus is reserved for the scalar addition in the field A .

We define the set W as

$$W = \{(a, b) \mid a, b \in \mathbb{R}\} \quad (7.30)$$

whereas we choose the vector addition and scalar multiplications as

$$(a, b) \square (c, d) = (a + c, b + d) \quad (7.31)$$

and

$$\text{plus}(a) \odot (b, c) = (a \cdot b, a \cdot c) \quad (7.32)$$

With these definitions, we can verify that all conditions for (W, \square, \odot) to be a vector space over the field A are satisfied.

3. Made reference to the commutators and discussed how a linear algebra can be turned into a lie algebra via commutators.
4. Cross products in three dimensional vector space is discussed from this perspective: $3d$ vectors form a lie algebra thanks to the cross product.
5. Emphasized how vector spaces can be built on an arbitrary set V as long as the necessary conditions are satisfied. For instance, we can choose

$$V = \left\{ a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} + c \frac{\partial}{\partial z} \mid a, b, c \in \mathbb{R} \right\} \quad (7.33)$$

With arithmetic addition and multiplication, these indeed form a vector space. In fact, this way of definition is far more useful than $ai + bj + ck$ as we will see later.

7.1.2 Properties of Linear spaces

1. Basis for vector spaces is introduced: let $(B \supset V) :: \text{Set}$. B is called a *basis* of V if following statements are true:
 - $(\forall k \in \{1, 2, \dim B\})(\forall e_1, \dots, e_k \in B)(\forall c_1, \dots, c_k \in S)[c_1 = \dots = c_k = 0] \vee [c_1 e_1 + \dots + c_k e_k \neq 0]$
 - $(\forall v \in V)(\exists! a_1, \dots, a_{\dim B} \in S)v = a_1 e_1 + \dots a_{\dim B} e_{\dim B}$

The first statement is the *linear independence* of the basis vectors, whereas second statement is the requirement that the basis *spans* the vector space, i.e. any vector can be decomposed in terms of the basis elements. The coefficients a_i are called the *components* of the vector in this basis.

2. If $\dim B < \infty$, it is called the dimension of the vector space!
3. If $\dim B = \infty$, the vector space is called *infinite dimensional*!
4. Using a basis, a vector can be related to a list/array/tuple of scalars:

$$\left. \begin{array}{l} v :: \text{Vector} \\ a_1, \dots, a_n :: \text{Scalar} \end{array} \right\} \Rightarrow v = (a_1 \quad \dots \quad a_n) \begin{pmatrix} e_1 \\ \dots \\ e_n \end{pmatrix} \quad (7.34)$$

For instance, $(3i+4k) :: \text{Vector}$ whereas $(3, 0, 4) :: (\text{Scalar}, \text{Scalar}, \text{Scalar})$; nevertheless, they represent the same information once the proper

basis is given: $3i + 4k = \begin{pmatrix} 3 & 0 & 4 \end{pmatrix} \begin{pmatrix} i \\ j \\ bk \end{pmatrix}$.

5. If multiple basis exist for a vector space, all have the same dimension and all can be converted to one another (proof omitted).

6. Index notation and Einstein's summation convention is discussed, i.e. $v = v^i e_i$. Note that index placement is important, i.e. $e^i \neq e_i$.
7. One can choose a_i to represent a row or column matrix (and a^i the other one). Note that M_i^j or M_j^i represent a square matrix (as they convert a row/column matrix to another row/column matrix), but M_{ij} or M^{ij} are not square matrices.

7.2 Inner product spaces

1. Normed vector spaces: in general, we do not need to have functions that takes vectors as inputs and spits out scalars as outputs, i.e. no $(\cdot) : (V, V) \rightarrow S$ or $\|\cdot\| : V \rightarrow S$ is necessary for the definition of the vector space.

If we impose that a function *norm* such that

$$\begin{aligned} \text{norm} : V &\rightarrow \mathbb{R} \\ \text{norm} = x &\rightarrow \|x\| \end{aligned} \quad (7.35)$$

exists where

- $(\forall v \in V)[\|v\| \neq 0] \vee [v = 0]$
- $(\forall v \in V)(\forall s \in S)s \odot v = |s| \cdot \|v\|$
- $(\forall v, w \in V)\|v \oplus w\| \leq \|v\| + \|w\|$

and where V is a vector space over the field \mathbb{R} or \mathbb{C} , then we call $(V, \|\cdot\|)$ a *normed vector space*.

2. With norm, we can define distance between two vectors: $d : (V, V) \rightarrow \mathbb{R}$ and $d = (x, y) \rightarrow \|x - y\|$.
3. With norm, we can define angle between two vectors: $\theta : (V, V) \rightarrow \mathbb{R}$ and $\theta = (x, y) \rightarrow \arccos\left(\frac{\|x\|^2 + \|y\|^2 - \|x - y\|^2}{2\|x\|\|y\|}\right)$.
4. Inner product spaces: If we have a vector space over the field \mathbb{R} or \mathbb{C} , and if we impose the existence of the function $\langle \cdot, \cdot \rangle$ for

$$\begin{aligned} F &\in \{\mathbb{R}, \mathbb{C}\} \\ V &:: \text{Vector Space over } F \\ \langle \cdot, \cdot \rangle &:: (V, V) \rightarrow F \end{aligned} \quad (7.36)$$

where the following are true:

- $(\forall v, w \in V) \langle v, w \rangle = \langle w, v \rangle^*$ (conjugate symmetry)
- $(\forall u, v, w \in V)(\forall a, b \in F) \langle au + bv, w \rangle = a \langle u, w \rangle + b \langle v, w \rangle$
- $(\forall v \in V \setminus \{0\}) \langle v, v \rangle > 0$
- $\langle 0, 0 \rangle = 0$

5. Inner product vector spaces can immediately be turned to a normed vector space as well as we can define $\|v\| = \sqrt{\langle v, v \rangle}$ which is consistent with all norm requirements. In this definition, the angle between two vectors become $\cos \theta = \frac{\text{Re}(\langle x, y \rangle)}{\|x\|\|y\|}$.

7.3 Dual vector spaces

1. Dual vector spaces: Let V be a vector space over a field F . If we consider linear functions that take a vector to a scalar, those functions themselves form a vector space: that is called the *dual vector space* V^* . So we have

$$(F = (S, +, \cdot)) :: \text{Field} \quad (7.37a)$$

$$V :: \text{Set} \quad (7.37b)$$

$$\oplus :: (V, V) \rightarrow V \quad (7.37c)$$

$$\odot :: (S, V) \rightarrow V \quad (7.37d)$$

$$(V^* = \{V \rightarrow S\}) :: \text{Set} \quad (7.37e)$$

$$\oplus^* :: (V^*, V^*) \rightarrow V^* \quad (7.37f)$$

$$\odot^* :: (S^*, V^*) \rightarrow V^* \quad (7.37g)$$

$$(7.37h)$$

where

$$(V, \oplus, \odot) = \text{vector space over } F \quad (7.38a)$$

$$(V^*, \oplus^*, \odot^*) = \text{vector space over } F \quad (7.38b)$$

Gave several examples in class, for instance $V = \{a\mathbf{i} + b\mathbf{j} \mid a, b \in \mathbb{R}\}$ and $V^* = \{(x\mathbf{i} + y\mathbf{j}) \rightarrow (ax + by) \mid a, b \in \mathbb{R}\}$.

2. Conventionally, elements of V are called *vectors* and elements of V^* are called *covectors*.
3. For the vector space given in the notation

$$V = \left\{ a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} + c \frac{\partial}{\partial z} \mid a, b, c, \in \mathbb{R} \right\} \quad (7.39)$$

the dual vector space is denoted as

$$V^* = \{adx + bdy + cdz \mid a, b, c \in \mathbb{R}\} \quad (7.40)$$

where $dx \left(\frac{\partial}{\partial x} \right) = 1$, $dx \left(\frac{\partial}{\partial y} \right) = 0$, and similarly.

Note that this convention is more than just a notation: the vector spaces indeed transform functions to functions, and integrations are indeed summations of covectors in a loose sense. We will make these much more concrete later.

4. Discussed in detail and derived that if a basis of the vector space e_i and the basis of the dual vector space e^i transform as

$$\begin{aligned} e_i &\rightarrow M_i^k e_k \\ e^i &\rightarrow e^k N_k^i \end{aligned} \quad (7.41)$$

^{bo}Note that $v^i (w_i)$ is actually only the components of the (co)vector, hence it is only an array of the underlying field, i.e. $w_i :: (S, S, \dots, S)$ and $v^i :: (S, S, \dots, S)$. In contrast, the combination $(w_i e^i) :: V^*$ is a covector and the combination $(v^i e_i) :: V$ is a vector.

then the components of the basis vector v^i and the components of the covector w_i transform as

$$\begin{aligned} v^i &\rightarrow v^k (M^{-1})_k^i \\ w_i &\rightarrow (N^{-1})_i^k w_k \end{aligned} \quad (7.42)$$

If we want the action of the covector basis on vector basis to remain unchanged, i.e.

$$e^i(e_k) \rightarrow e^i(e_k) \quad (7.43)$$

then we need $N = M^{-1}$. With this, we see that

$$\begin{aligned} e_i &\rightarrow M_i^k e_k \\ w_i &\rightarrow M_i^k w_k \\ e^i &\rightarrow e^k (M^{-1})_k^i \\ v^i &\rightarrow v^k (M^{-1})_k^i \end{aligned} \quad (7.44)$$

We see that the array w_i transforms in the same way as the basis vectors of V , i.e. it transforms *covariantly*. We then loosely state that w_i is a *covariant vector*!^{bo} In contrast, v^i transforms in an inverse way compared to e_i , and hence it is loosely called a *contravariant vector*!

We did not discuss topological vector spaces and Hilbert spaces this year: we may discuss it next year!

7.4 Algebras over fields

7.4.1 Tensor algebras

1. Consider a linear space over a field F . We call “*type (r, s) tensor on a vector space V* ” an element of

$$\underbrace{V \otimes V \otimes \dots \otimes V}_r \otimes \underbrace{V^* \otimes V^* \otimes \dots \otimes V^*}_s \quad (7.45)$$

where \otimes is an associative bilinear map, i.e.

$$\begin{aligned} A &= \{V, V^*\} \\ \otimes &:: (A, A) \rightarrow A \otimes A \\ (\forall a, b, c \in A) \quad (a \otimes b) \otimes c &= a \otimes (b \otimes c) \end{aligned} \quad (7.46)$$

2. We conventionally denote

$$T^k V = \underbrace{V \otimes \cdots \otimes V}_k \quad (7.47)$$

Since the bilinear map \otimes takes the product of $v : : T^k V$ and $\omega : : T^\ell V$ to $(v \otimes \omega) : : T^{k+\ell} V$, we need to consider the space of all possible $T^k V$ to have an algebra: this is denoted as $T(V)$ (tensor algebra over V) and is defined as

$$T(V) = \bigoplus_{k=0}^{\infty} T^k V = F \oplus V \oplus (V \otimes V) \oplus \cdots \quad (7.48)$$

A similar definition can be made for $T(V^*)$; the product $T(V) \otimes T(V^*)$ is the algebra over vector field that includes all type of tensors (i.e. $(\forall r, s \in \mathbb{N})$ (r, s) -tensor).

3. In Physics, we usually work with components. Indeed, consider the following example:

$$\begin{aligned} a &: : V \otimes V \otimes V^* \\ a &= a^{ij}_k e_i \otimes e_j \otimes e^k \end{aligned} \quad (7.49)$$

We usually call a^{ij}_k a tensor: in reality, it is only a collection of scalars (underlying field elements of the vector space); nevertheless, once a basis is specified, there is indeed a unique map between a and a^{ij}_k hence this abuse of notation is justified.

4. \otimes provides an *outer* product, i.e. $a \otimes b$ is not part of the types of a or b . For instance, if both a and b are of the type $(1, 0)$, $a \otimes b$ is of the type $(2, 0)$. In general the value $r + s$ for a type (r, s) tensor is called *the rank of the tensor*, and the rank of the tensor $a \otimes b$ is the sum of the ranks of the tensors a and b . For instance:

$$\begin{aligned} a &: : V \otimes V \otimes V^* \\ b &: : V^* \otimes V^* \\ (c = a \otimes b) &: : V \otimes V \otimes V^* \otimes V^* \otimes V^* \\ a &= a^{ij}_k e_i \otimes e_j \otimes e^k \\ b &= b_{\ell m} e^\ell \otimes e^m \\ c &= c^{ij}_{k\ell m} e_i \otimes e_j \otimes e^k \otimes e^\ell \otimes e^m \\ c^{ij}_{k\ell m} &= a^{ij}_k b_{\ell m} \end{aligned} \quad (7.50)$$

5. As stated earlier, we usually work with components in practice (at least in the Physics), so the product of two tensors $a^{i_1 \dots i_r}_{k_1 \dots k_s}$ and $b^{\ell_1 \dots \ell_t}_{m_1 \dots m_u}$ just becomes $c^{i_1 \dots i_r \ell_1 \dots \ell_t}_{k_1 \dots k_s m_1 \dots m_u} = a^{i_1 \dots i_r}_{k_1 \dots k_s} b^{\ell_1 \dots \ell_t}_{m_1 \dots m_u}$.
6. There is a way to change the type and rank of a tensor by an operation called *contraction*. What it does is basically act the basis covector on the basis vector; to see this, consider the tensor

c in the equation 7.50. We can *choose* to contract some of its basis elements to reduce its rank, for instance if we choose to act e^k on e_i , we get $e^k(e_i) = \delta_i^k$ for the kronecker-delta δ . But since $c_{ktm}^{ij} \delta_i^k = c_{itm}^{ij}$, we arrive at

$$\begin{aligned} c' &: V \otimes V^* \otimes V^* \\ c' &= c_{itm}^{ij} e_j \otimes e^\ell \otimes e^m \end{aligned} \quad (7.51)$$

If we instead acted e^k on e_j , we would get a different tensor c'' :

$$\begin{aligned} c'' &: V \otimes V^* \otimes V^* \\ c'' &= c_{jtm}^{ij} e_i \otimes e^\ell \otimes e^m \end{aligned} \quad (7.52)$$

Or we could act e^k on e_i and e^ℓ on e_j : this would give us an entirely different tensor c''' :

$$\begin{aligned} c''' &: V^* \\ c''' &= c_{ijm}^{ij} e^m \end{aligned} \quad (7.53)$$

7. Consider the rotational version of Newton's second law, i.e. "*torque is equal to moment of inertial times angular acceleration*". We conventionally write it as a matrix equation, i.e. $\tau^i = I_j^i \alpha^j$. We see that we could write them as follows:

$$\begin{aligned} \tau, \alpha &: V \\ I &: V \otimes V^* \\ (I \otimes \alpha) &= I_j^i \alpha^k e_i \otimes e_k \otimes e^j \\ \tau &= I_j^i \alpha^j e_i \end{aligned} \quad (7.54)$$

Clearly, we got τ by first taking the outer product of I and α , and then doing a contraction between two of the indices (equivalently, applying the covector e^j on the vector e_k).^{bp}

7.4.2 Exterior algebras

1. Tensor product \otimes is only defined as an associative bilinear mapping. If we define a similar product with the additional property that it is antisymmetric, we get *exterior algebra*.
2. The relevant antisymmetric operator is called *Wedge product* and is denoted as \wedge :

- $(\forall u, v \in V) \quad u \wedge v = -v \wedge u$
- $(\forall u, v, w \in V) \quad u \wedge (v \wedge w) = (u \wedge v) \wedge w$

3. As the product is fully antisymmetric, any wedge product of k -many vectors is zero if $k > \dim(V)$, thus:

$$\Lambda(V) = \bigoplus_{k=0}^{\dim(V)} \Lambda^k V = F \oplus V \oplus (V \wedge V) \oplus \dots \quad (7.55)$$

^{bp} This line of reasoning is valid to a degree for $3d$ -vectors, but in higher dimensions, it is not true. Firstly, τ and α are not vectors but *bivectors* in general dimension (we'll cover bivectors below), hence I is actually not a matrix (a rank-2 tensor) but a linear map of bivectors to bivectors (hence a rank-4 tensor): $I :: V \otimes V \otimes V^* \otimes V^*$.

for

$$\Lambda^k V = \underbrace{V \wedge \cdots \wedge V}_k \quad (7.56)$$

forms the exterior algebra.

4. Elements of $\Lambda^k V$ are called k -vectors (or multivectors in general).
5. In $3d$, the exterior product of two vectors would almost like the cross product of the vectors:

$$\left. \begin{aligned} u &= u_1 e_1 + u_2 e_2 + u_3 e_3 \\ v &= v_1 e_1 + v_2 e_2 + v_3 e_3 \end{aligned} \right\} \rightarrow u \wedge v = (u_1 v_2 - v_2 u_1)(e_1 \wedge e_2) + \dots \quad (7.57)$$

Similarly, $u \wedge v \wedge w$ will be almost same as $\vec{u} \cdot (\vec{v} \times \vec{w})$. These are no coincidences; in fact, we will see that many vector operations can be understood in terms of multivectors.

6. We have seen several area and volume integrations in first year courses: we know that the integrations (such as use of Gauss law in electrodynamics) require *oriented areas*; in practice, we write integrals like $\int f(x, y) dx dy$; however, obviously, the infinitesimal area element $dx dy$ is *not* oriented when written like that. The proper way to write it is actually as $dx \wedge dy$: this is simply a 2 -covector; likewise, the oriented volume element is a 3 -covector $dx \wedge dy \wedge dz$.
7. Hodge duality: Any two vector space with same dimension are isomorphic to each other, i.e. they can be converted to one another. Since $\Lambda^p V$ is a vector space with dimensionality $\frac{d!}{p!(d-p)!}$,^{bq} there is an isomorphism between p -vectors and $(d-p)$ -vectors: a natural way to achieve this is the *Hodge duality*. If v is a p -vector, then its Hodge dual (denoted as $\star v$) is a $(d-p)$ -vector!
8. We physicists mostly work only with components (components of the vector, multivector, tensor, etc.) instead of the actual object, so for us it is sufficient to know how to relate the components of a p -vector and its Hodge dual.^{br} In the convention

$$\begin{aligned} \alpha &= \alpha_{i_1 \dots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k} \\ \star \alpha &= (\star \alpha)_{i_{k+1} \dots i_d} e^{i_{k+1}} \wedge \cdots \wedge e^{i_d} \end{aligned} \quad (7.61a)$$

the relation reads as

$$(\star \alpha)_{i_{k+1} \dots i_d} = \frac{1}{(d-k)!} \alpha_{i_1 \dots i_k} \epsilon^{i_1 \dots i_k i_{k+1} \dots i_d} \delta_{i_{k+1} i_{k+1}} \cdots \delta_{i_d i_d} \quad (7.61b)$$

if we are working in Euclidean spaces, which will be the focus of this course! Here δ is the Kronecker delta symbol.

9. We discussed examples of Hodge dualities in $2d$ and $3d$, e.g.

$$\star 1 = dx \wedge dy, \quad \star dx = dy, \quad \star dy = -dx \text{ in } 2d \quad (7.62)$$

10. We discussed how Hodge duality helps us define the cross product in $3d$: $\star(u \wedge v) = u \times v$.

^{bq}This is the number to choose p elements out of d , as the basis vectors of $\Lambda^p(V)$ are ordered combinations (e.g. $e_1 \wedge e_2$ and $e_2 \wedge e_1$ are not linearly independent).

^{br} More abstractly, we can understand Hodge duality as follows.

Let V be a d -dimensional Inner Product Vector Space. $\langle \cdot, \cdot \rangle : (V, V) \rightarrow \mathbb{C}$. This inner product *induces* an inner product in the vector space $(\star \alpha)_{i_{k+1} \dots i_d}$ as follows:

$$\langle \cdot, \cdot \rangle_k : (\Lambda^k V, \Lambda^k V) \rightarrow \mathbb{C} \quad (7.58)$$

for

$$\langle \alpha_1 \wedge \cdots \wedge \alpha_k, \beta_1 \wedge \cdots \wedge \beta_k \rangle = \det \begin{pmatrix} \langle \alpha_1, \beta_1 \rangle & \cdots & \langle \alpha_1, \beta_k \rangle \\ \vdots & & \vdots \\ \langle \alpha_k, \beta_1 \rangle & \cdots & \langle \alpha_k, \beta_k \rangle \end{pmatrix} \quad (7.59)$$

The Hodge-dual of a k -vector β is then defined in terms of this induced inner product as

$$(\forall \alpha \in \Lambda^k(V)) \quad \alpha \wedge (\star \beta) = \langle \alpha, \beta \rangle e_1 \wedge \cdots \wedge e_n \quad (7.60)$$

where e_i form the orthonormal basis of V .

7.4.3 Division algebras

Following is not covered in class or in homeworks; may be covered next years:

Normed division algebras and Hurwitz's theorem; finite-dimensional associative division algebras over the real numbers and Frobenius theorem; more about quaternions, their usage in spatial rotation (hence applications in engineering), and Feza Gursey's work with quaternions

Differentiation in Vector Calculus

8.1 Vector fields and forms on Euclidean Spaces

1. In class, we discussed 2d and 3d pendulums: at different positions, the velocity at that position belongs to the vector space *at that position*. Even though these vector spaces are isomorphic to each other,^{bs} i.e. they can be transformed to each other, they are actually different vector spaces! This is not only true for pendulums, but for any motion!
2. Let us assume that we are going to describe the kinematics of a particle. We need to specify the position and the velocity of the particle at that position: this is the complete information to understand the trajectory of the particle. In other words, if we know all possible $(r, \vec{v}(r))$ pairs where r parametrizes the particle's position in the system and $\vec{v}(r)$ is the velocity vector in that position, then we can describe the motion completely!

Consider a two dimensional pendulum. We can parametrize the object's position by an angle θ as the trajectory forms a circle. At any position, its velocity lies in the *tangent line* at the pendulum's position to this circle: note that these lines are *different* at different values of θ , so we indeed have infinitely many vector spaces, one for each value of θ . We could choose $\frac{\partial}{\partial \theta}$ as the basis vector; then the relevant pairs would look like $(\frac{\pi}{6}, 2\frac{\partial}{\partial \theta})$ or $(-\frac{\pi}{3}, -\frac{\partial}{\partial \theta})$. The first one says that at $\theta = \pi/6$, the velocity is $2\frac{\partial}{\partial \theta}$ at the tangent space to the circle at $\theta = \pi/6$. We will see later that this basically means the particle is moving such that at $\theta = \pi/6$, it will move to increase its θ . On the contrary, at $\theta = -\pi/3$, it will move to decrease its θ (as the velocity at that point is $-\frac{\partial}{\partial \theta}$). If we are to combine all such $(\theta, v(\theta)\frac{\partial}{\partial \theta})$ pairs, we have the complete behavior of the pendulum!

To understand these better, we need to introduce some terminology.

3. We call a space *Manifold* if it resembles the Euclidean space near each point.^{bt} For instance, \mathbb{R}^k is trivially a Manifold, it is Euclidean space itself. A more nontrivial example is a circle (denoted as S^1) or a sphere (denoted as S^2): near any point on a circle or a sphere, the space *looks like* a Euclidean space!^{bu}
4. At any point on a Manifold M , we have the *tangent space* to the manifold denoted as $T_x M$. For instance, $T_x S$ denotes the tangent line to the circle S at the position x ; likewise, $T_x S^2$ denotes the tangent plane to the sphere S^2 at the position x . Tangent spaces

^{bs}Any two finite dimensional vector spaces with the same dimension are actually isomorphic to each other.

^{bt}We actually need to know *topological spaces* to define manifolds more appropriately; nevertheless, this definition is good enough for this course.

^{bu}This is why flat Earth believers are so confused: the surface of our planet is a Manifold, and at around any point, we see it approximately as a flat plane \mathbb{R}^2 .

are necessarily flat, i.e. they are simply \mathbb{R}^d where d is the dimension of the Manifold.

5. Clearly, $T_x M$ and M are different spaces: this is most easy when the Manifold is not \mathbb{R}^d . For instance, if we consider a pendulum, the position of the object is an element of the Manifold S , and its velocity at any point x is an element of the tangent space $T_x S$. Clearly, the tangent space (simply the tangent line \mathbb{R} in this case) is a linear space, hence its elements are vectors (called tangent vector), which makes sense as we expect the velocity to be a vector!

In other examples, it might look tricky. For instance, for a car moving on a plane, the manifold is \mathbb{R}^2 , and the tangent space at any point is also \mathbb{R}^2 : even though these two spaces are *isomorphic*, they are actually different: position is an element of the manifold \mathbb{R}^2 , whereas the velocity at any position is an element of the tangent space \mathbb{R}^2 at that position.

6. Consider a manifold M and an element $x \in M$. Clearly we have infinitely many tangent spaces $T_x M$ for all $x \in M$. We can then define so-called *tangent bundle* of this manifold as the collection of all such vector spaces! Tangent bundle of the manifold M is denoted as TM and we will define it^{bv} as

$$TM = \{(x, y) \mid x \in M, y \in T_x M\} \quad (8.1)$$

7. At any point on the Manifold, we can also define the dual vector space $T_x^* M$, which is the vector space whose elements are linear functions from $T_x M$ to the underlying field. $T_x^* M$ is called *cotangent space*, and by combining all of them for a given Manifold M , we obtain the *cotangent bundle* denoted as $T^* M$.
8. In physics, we tentatively call a function *field* if its domain is the Manifold.^{bw} For instance, the function $f : M \rightarrow \mathbb{R}$ is a real scalar field on the manifold M ; for instance, *the temperature field* is a function which takes an element of the manifold (the position in the space) and spits out a real number, i.e. the temperature at that position. Likewise the *electric field* is a function which takes an element of the manifold as input and spits out a vector, i.e. the electric field vector at that position.
9. In the partial derivative notation of vectors (see (7.39) for an example), *the vector fields* become a simple generalization of that; for instance,

$$V : \mathbb{R}^3 \rightarrow T\mathbb{R}^3$$

$$V = (x, y, z) \rightarrow \left(x^2 \frac{\partial}{\partial x} + xz \frac{\partial}{\partial y} + \frac{1}{x+y} \frac{\partial}{\partial z} \right) \quad (8.2)$$

is a $3d$ vector field.^{bx}

^{bv} A more appropriate definition of tangent bundle would involve *disjoint union* of individual tangent spaces. As there is no unique way to define disjoint unions (they are only defined up to isomorphisms), we can in fact define TM in different ways as well. We will not be using the explicit definition of TM in this course, so our definition is correct and good enough!

^{bw} Note that this has nothing to do with the mathematical term *field*, which refers to a set with two operations satisfying certain conditions as detailed in § ??.

^{bx} We are being very schematic with our definitions and types of objects when it comes to bundles. Indeed, the way we defined in (8.1), V can have the codomain $T\mathbb{R}^3$ only if its output is actually $\left((x, y, z), \left(x^2 \frac{\partial}{\partial x} + xz \frac{\partial}{\partial y} + \frac{1}{x+y} \frac{\partial}{\partial z} \right) \right)$. In fact, even this is not true: (x, y, z) does not really represent the element of M but rather an element of a *chart* on M , which we do not cover in these notes. Furthermore, as we stated earlier in footnote (bv), the tangent bundle is defined only upto isomorphisms, hence even with charts introduced, we need to be extra careful with our definitions if we want rigor. Instead, we will only use types as *schematic reminders* in this chapter and do not pursue the topic any further.

10. Multivector fields on the cotangent space, or equivalently covector fields on the tangent space, are called *differential forms* (forms for short). For instance,

$$\begin{aligned} \omega &: \mathbb{R}^3 \rightarrow \Lambda^2(T_x\mathbb{R}^3) \\ \omega = (x, y, z) &\rightarrow \left(x^2 dx \wedge dy + xz dx \wedge dz + \frac{1}{x+y} dy \wedge dz \right) \end{aligned} \quad (8.3)$$

is a 2-form on the manifold \mathbb{R}^3 .

11. Tensor fields are also straightforward generalizations of tensors to mappings from manifolds; for example,

$$\begin{aligned} T &: S^2 \rightarrow (T_x\mathbb{R}^2) \otimes (T_x\mathbb{R}^2) \otimes (T_x^*\mathbb{R}^2) \\ T = (\phi, \psi) &\rightarrow \left(\sin(\psi) \frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial \phi} \otimes d\phi + \cos(\psi + \phi) \frac{\partial}{\partial \psi} \otimes \frac{\partial}{\partial \phi} \otimes d\psi \right) \end{aligned} \quad (8.4)$$

is a type (2,1) tensor field on the manifold S^2 .

12. Musical isomorphism: We define two operations that convert back and forth between tangent and cotangent bundles:

$$\flat :: TM \rightarrow T^*M \quad (8.5a)$$

$$\sharp :: T^*M \rightarrow TM \quad (8.5b)$$

which in practice amounts to raising and lowering the indices on the components of the tensors, i.e.

$$(x^i e_i)^\flat = (x_i e^i), \quad (x_i e^i)^\sharp = (x^i e_i) \quad (8.6)$$

In practice, we physicists do not use musical isomorphism terminology: the manifolds we work with are usually something called a *metric space*, meaning that they are endowed with a *metric* which is type (0, 2) tensor field (and the inverse metric which is type (2, 0) tensor field). The metric is then used to go back and forth between the dual spaces. In this course, we stick to Euclidean spaces, hence we do not need to introduce the metric tensor and its properties, and hence we won't: the discussion of musical isomorphism is simpler and sufficient enough!^{by}

13. The application of musical isomorphism can be extended to k -vectors and forms as well, e.g. $(x_{ij} e^i \wedge e^j)^\sharp = (x^{ij} e_i \wedge e_j)$

^{by}The musical isomorphism is actually broader in the sense that it represent the abstract mapping between the bundles, which may not be actuated with a metric. For instance, *fiber derivative* also induces a map from tangent to cotangent bundle, which is rather relevant in the study of *Legendre transformations*.

8.2 Grad, div, curl, and Laplacian

8.2.1 Exterior derivative and hodge star operator

1. We define *exterior derivative* as a map from p -form to a $(p + 1)$ -form, and denote it as $d\omega$ where ω is the p -form. With the

differentials as the basis vectors of the cotangent bundle, it explicitly reads as

$$\omega = \omega_{i_1 \dots i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p} \quad (8.7a)$$

$$d\omega = \frac{\partial \omega_{i_1 \dots i_p}}{\partial x^k} dx^k \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p} \quad (8.7b)$$

^{bz}Note to myself: we discussed these points in detail in class, so I better include a nice discussion here as well, along with some external references.

2. We have discussed several examples of exterior derivative in class. For instance, for the electric potential $V(x, y) = x^2 + y^2$ given in a $2d$ plane, the exterior derivative of it reads as $dV = 2xdx + 2ydy$: if we were to plot them, they are actually vectors radially outward, precisely orthogonal to the equipotential lines (i.e. $V(x, y) = \text{constant}$); in fact, the electric field is actually proportional to this exterior derivative, i.e. $E \sim dV$.
3. The importance of the exterior derivative is that it is *basis-independent*. Physically, this means that exterior derivative of an observer-covariant quantity is itself observer-covariant, which is precisely what we need! In contrast, a vector field defined as $v = \frac{\partial f}{\partial x} dx - \frac{\partial f}{\partial y} dy$ is *not* observer-covariant even if the function f is so. In other words, it cannot represent something physical!^{bz}
4. The Hodge duality introduced for multivectors naturally generalizes to the forms; in fact, we can immediately use the same formula given in eqn. (7.61) for forms as follows:

$$\begin{aligned} \alpha(x_1, \dots, x_d) &= \alpha_{i_1 \dots i_k}(x_1, \dots, x_d) e^{i_1} \wedge \dots \wedge e^{i_k} \\ (\star \alpha)(x_1, \dots, x_d) &= (\star \alpha)_{i_{k+1} \dots i_d}(x_1, \dots, x_d) e^{i_{k+1}} \wedge \dots \wedge e^{i_d} \end{aligned} \quad (8.8a)$$

and

$$(\star \alpha)_{i_{k+1} \dots i_d}(x_1, \dots, x_d) = \frac{1}{(d-k)!} \alpha_{i_1 \dots i_k}(x_1, \dots, x_d) \epsilon^{i_1 \dots i_k \ell_{k+1} \dots \ell_d} \delta_{i_{k+1} \ell_{k+1}} \dots \delta_{i_d \ell_d} \quad (8.8b)$$

5. Schematically speaking, *musical isomorphism*, *Hodge-duality*, and *exterior derivative* are our necessary and sufficient tools to construct *differential maps* from scalar & vector fields to scalar & vector fields. In the next section, we will then use these to define various vector differentiation operations.

8.2.2 Invariant formulation of differentiation in vector calculus

1. We define the gradient (denoted as grad) as follows:

$\text{grad} :: \text{Scalar Field} \rightarrow \text{Vector Field}$

$$\text{grad} = f \rightarrow (df)^\sharp$$

$$\text{grad} = ((x_1, \dots, x_d) \rightarrow f(x_1, \dots, x_d))$$

$$\rightarrow \left((x_1, \dots, x_d) \rightarrow \frac{\partial f(x_1, \dots, x_d)}{\partial x_i} \hat{x}_i \right) \quad (\text{in Cartesian coordinates}) \quad (8.9)$$

2. It is also commonly denoted as ∇ ; for instance, for the scalar field $f(x, y) = x^2 + y$, we write the vector field as $\nabla f(x, y) = 2x \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$. In the freshman notation, this simply is $\nabla f(x, y) = 2x\hat{i} + \hat{j}$.
3. It is easiest to understand the gradient visually: For any scalar field f whose value at any point gives the scalar S (temperature, electric potential, pressure, etc.), draw the equi- S lines (equitemperature, equipotential, etc.) which correspond to $S = \text{constant}$ lines. The vector field ∇f then is the collection of vectors that are *orthogonal* to these equi- S lines. In the case of electric potential scalar field f , ∇f is proportional to the electric field; likewise, in the case of temperature field f , ∇f is related to the heat flow vector field.

4. We define the divergence (denoted as div) as follows:

$\text{div} :: \text{Vector Field} \rightarrow \text{Scalar Field}$

$$\text{div} = v \rightarrow (\star d \star v^b)$$

$$\text{div} = ((x_1, \dots, x_d) \rightarrow v^i(x_1, \dots, x_d) \hat{x}_i)$$

$$\rightarrow \left((x_1, \dots, x_d) \rightarrow \frac{\partial v^i(x_1, \dots, x_d)}{\partial x^i} \right) \quad (\text{in Cartesian coordinates}^{ca})$$

(8.10)

5. It is also commonly denoted as $\nabla \cdot$; for instance, for the vector field $v(x, y) = xy \frac{\partial}{\partial x} - y^2 \frac{\partial}{\partial y}$, we write the scalar field as $\nabla \cdot v(x, y) = -y$.^{cb}
6. Divergence measures *the scalar source* of the vector field: we will learn more about this when we cover the Helmholtz decomposition.
7. We define the curl (denoted as curl) as follows:

$\text{curl} :: \text{Vector Field} \rightarrow \text{Scalar Field} \quad (d = 2)$

$\text{curl} :: \text{Vector Field} \rightarrow \text{Vector Field} \quad (d = 3)$

$\text{curl} :: \text{Vector Field} \rightarrow (d - 2) - \text{Vector Field} \quad (d > 3)$

$$\text{curl} = v \rightarrow (\star dv^b)^\#$$

(8.12)

8. The curl operator reduces to the cross product with the gradient operation in $3d$, i.e. $\text{curl } v = \nabla \times v$; in Cartesian coordinates, it then yields

$$\nabla \times v = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) \hat{x} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) \hat{y} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \hat{z} \quad (8.13)$$

9. Curl measures *the vector source* of the given vector field: we will understand this better with the Helmholtz decomposition.

^{ca} It may look counter-intuitive to see that such a simple definition in Cartesian derivative involve so many operations in its abstract definition. Let us do the explicit derivation to see that it makes sense:

$$v = v^i(x_1, \dots, x_d) \frac{\partial}{\partial x^i}$$

$$v^b = v_i(x_1, \dots, x_d) dx^i$$

$$\star v^b = \frac{1}{(d-1)!} v_i(x_1, \dots, x_d) \epsilon^{ik_2 \dots k_d} \delta_{k_2 \ell_2} \dots \delta_{k_d \ell_d} \times dx^{\ell_2} \wedge \dots \wedge dx^{\ell_d}$$

$$d \star v^b = \frac{1}{(d-1)!} \frac{\partial v_i(x_1, \dots, x_d)}{\partial x^m} \epsilon^{ik_2 \dots k_d} \delta_{k_2 \ell_2} \dots \delta_{k_d \ell_d} \times dx^m \wedge dx^{\ell_2} \wedge \dots \wedge dx^{\ell_d}$$

$$\star d \star v^b = \frac{1}{(d-1)!} \frac{\partial v_i(x_1, \dots, x_d)}{\partial x^m} \epsilon^{ik_2 \dots k_d} \delta_{k_2 \ell_2} \dots \delta_{k_d \ell_d} \epsilon^{m \ell_2 \dots \ell_d}$$

$$\star d \star v^b = \frac{\partial v_i(x_1, \dots, x_d)}{\partial x^m} \delta^{im}$$

$$\star d \star v^b = \frac{\partial v_i(x_1, \dots, x_d)}{\partial x^i}$$

(8.11)

where we used $\epsilon^{ia_1 \dots a_n} \epsilon_{ka_1 \dots a_n} = n! \delta_k^i$.

^{cb} This notation is no coincidence: the divergence can be seen as the dot product between the vector field and the gradient operator ∇ .

10. Discussed the pseudovector nature of curl and did a few examples in class.
11. We now turn to our last operator: the Laplacian (denoted as Δ). In principle, it can take any scalar or multivector field to itself; however, we will only define it for scalar and vector fields:^{cc}

^{cc}Note to myself: maybe discuss Laplace beltrami operator, or at least a few citations?

^{cd}Note to myself: Again, I discussed this whole observer-independence in class, but these notes lack a proper discussion.

$$\begin{aligned}
 \Delta &:: \text{Scalar Field} \rightarrow \text{Scalar Field} \\
 \Delta &= f \rightarrow (\star d \star df) \\
 \Delta &:: \text{Vector Field} \rightarrow \text{Vector Field} \\
 \Delta &= v \rightarrow (d \star d \star v^b - \star d \star dv^b)^{\#}
 \end{aligned} \tag{8.14}$$

We only need this abstract definition to ensure that this operator is observer-covariant, i.e. we can do science with it!^{cd} In practice, we can also define this same operator as simply *the gradient of divergence* and denote it as $\Delta = \nabla \cdot \nabla = \nabla^2$, which can take any tensor field to another tensor field: in Cartesian coordinates, it simply reads as follows:

$$\begin{aligned}
 R &:: TM \otimes \dots \otimes TM \otimes T^*M \otimes \dots \otimes T^*M \\
 R &= R^{i_1 \dots i_r}_{k_1 \dots k_s} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{k_1} \otimes \dots \otimes dx^{k_s} \\
 (\Delta R) &:: TM \otimes \dots \otimes TM \otimes T^*M \otimes \dots \otimes T^*M \\
 \Delta R &= \frac{\partial^2 R^{i_1 \dots i_r}_{k_1 \dots k_s}}{\partial x^m \partial x_m} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{k_1} \otimes \dots \otimes dx^{k_s}
 \end{aligned} \tag{8.15}$$

12. In class, derived several identities such as $\nabla \cdot \nabla \times v = 0$ using their abstract definitions and the identity $d^2 = 0$.
13. In class, we discussed why we use these generalized definitions even though Cartesian component based ones are actually far more simple! The key point is that this way the result is guaranteed to be observer covariant/invariant! To emphasize, we defined a new operator in class (call it \mathcal{D}) which acts on $2d$ -vectors as

$$\mathcal{D} \cdot A = \frac{\partial A_x}{\partial y} + \frac{\partial A_y}{\partial x} \tag{8.16}$$

Even though this operator does generate a scalar field from a vector field, and even though there is nothing wrong with it mathematically, it is actually observer-dependent. In other words, the scalar that it generates changes when we change the coordinate system. Physically, we do not want scalars to be observer dependent (the temperature at a point should not change when I rotate clockwise 90 degrees). In this example, it is easy to see this as we can straightforwardly check it; but in general, it can become quite tedious for longer operators; and it does not instruct us how we should rather construct our operator. When we work with forms, we are guaranteed that any operator out

of exterior derivative, hodge duality, and musical isomorphism is observer independent as it does not refer to a particular basis but is instead defined abstractly.

8.3 Helmholtz decomposition of vector fields

1. Any 3d vector field can be written as

$$E = \text{constant} - \nabla\Phi + \nabla \times V \quad (8.17)$$

In most situations, the constant piece is simply zero (so that the behavior of the vector field around infinity is zero) for which we can write

$$\Phi(r) = \frac{1}{4\pi} \int_{\text{manifold}} \frac{\nabla' \cdot E(r')}{|r - r'|} dV' - \frac{1}{4\pi} \oint_{\text{boundary}} \frac{\hat{n}' \cdot E(r')}{|r - r'|} dS' \quad (8.18)$$

and

$$V(r) = \frac{1}{4\pi} \int_{\text{manifold}} \frac{\nabla' \times E(r')}{|r - r'|} dV' - \frac{1}{4\pi} \oint_{\text{boundary}} \frac{\hat{n}' \times E(r')}{|r - r'|} dS' \quad (8.19)$$

Here, Φ is called the *scalar potential* of E whereas V is called the *vector potential* of E .

2. Even if the vector field E is fixed, the vector potentials are not: the vector field is invariant under the change

$$\Phi \rightarrow \Phi + c, \quad V \rightarrow V + \nabla a \quad (8.20)$$

for a constant c and scalar field a . This transformation is called *gauge transformation*, and invariance of E under this transformation is called *gauge invariance*.

3. In class, we discussed the conservative forces (for which vector potential is zero), importance of these *sources* in understanding electromagnetism, and how gauge transformation is actually rather important in quantum mechanics.
4. In class, we analyzed the vector fields through their flow diagrams: I draw a diagram and we have seen around which parts we have divergence contributions (basically Gauss law), and which parts have rotational contributions (e.g. Ampere's law).
5. I did not discuss or even mention the *Hodge decomposition*, the generalization of this. Note to myself: decide whether to include a brief mention of this or not next year!

Integration in Vector Calculus

9.1 Higher Dimensional Integration

1. Discussed in class that the students have actually learned three kinds of integration so far: indefinite, signed definite, and unsigned definite. The fundamental theorem of calculus relates these three to each other:

$$\int f(x)dx = F(x) \quad (\text{indefinite integral}) \quad (9.1a)$$

$$\int_a^b f(x) = F(b) - F(a) \quad (\text{signed definite integral}) \quad (9.1b)$$

$$\int_{[a,b]} f(x)dx = \begin{cases} \int_a^b f(x)dx & b > a \\ \int_b^a f(x)dx & a > b \end{cases} \quad (\text{unsigned definite integral}) \quad (9.1c)$$

2. When we generalize to higher dimensional integrals, these three actually diverge to different kinds of integrations:^{ce}

$$\text{indefinite} \rightarrow \text{anti-derivative} \quad (9.2a)$$

$$\text{signed definite} \rightarrow \text{integration of forms} \quad (9.2b)$$

$$\text{unsigned indefinite} \rightarrow \text{Lebesgue integration} \quad (9.2c)$$

Anti-derivation is useful for solving the partial differential equations, and Lebesgue integrations are used for integrations on measures. The integration of forms on the other hand are the most common types in the elementary Physics, so we will focus on that in these notes.^{cf}

3. In class, we discussed how one-dimensional ordinary integration can be seen as summation of covectors. I'll write these in more detail with figures in future (hopefully next year). Interested reader can consult to Terence Tao's "Differential Forms and Integration" notes in *The Princeton Companion to Mathematics*.
4. In class, we discussed how ambient space and integration surfaces can be of different dimensions, and how signed definite integration (which generalizes to integration of forms) is actually pairing of a p -form with a p -dimensional integration surface to create a scalar. For instance, in Gauss' law, we integrate a

^{ce}In other words, fundamental theorem of calculus fails for higher dimensional integrations. The branch of math that analyzes how much it fails is called *de Rham cohomology*.

^{cf}Note to myself: add examples for all.

2-form (electric flux) against a 2-dimensional integration surface (a closed Gaussian surface) which yields a scalar (total enclosed charge). Likewise, the infinitesimal work done can be seen as a 1-form and by integrating it against a path (1-dimensional surface), we get a real number (the work done).

5. The simplest generalization of one-dimensional signed definite integral is the line integral in higher dimensions (for which integration surface is a one-dimensional line in a higher dimensional ambient space), which we will focus next. But for these integrals, we need to understand how we can mathematically describe paths, which we turn to next.

9.2 Curves and Line Integrals

1. Concepts: Parametrization, arc length, concept of curvature and torsion, and Frenet–Serret formulas
2. In class, we discussed the above concepts and schematically derived the Frenet–Serret formulas. For more on these topics, see David Tong’s lecture notes on vector calculus: <http://www.damtp.cam.ac.uk/user/tong/vc.html> (see chapter 1).

9.3 Integral theorems

9.3.1 Gradient, Divergence, Curl, and Greens theorems

1. In class, we discussed these theorems and their applications in Physics, see <http://www.damtp.cam.ac.uk/user/tong/vc.html> (chapter 4) or chapter 6.9–6.11 of the textbook for details.

9.3.2 Generalized Stokes theorem: one equation to derive them all!

1. Generalized Stoke’s theorem reads as

$$\int_{\partial\Omega} \omega = \int_{\Omega} d\omega \quad (9.3)$$

which is an extremely simple yet immensely powerful relation: all integral theorems can be derived using this relation.

2. In the above equation, Ω denotes any $(p + 1)$ -dimensional integration surface, $\partial\Omega$ its p -dimensional boundary, ω a p -dimensional differential form, and d the exterior derivative.

3. The antisymmetric nature of forms ensures that $d^2 = 0$; this then leads to

$$\int_{\partial^2 \Omega} \omega = \int_{\partial \Omega} d\omega = \int_{\Omega} d^2 \omega = 0 \quad (9.4)$$

for any form ω of region Ω . This is then only possible if $\partial^2 \Omega = \{\}$, which basically means *the boundary of a region does not have a boundary itself*. We illustrated this in class as it also intuitively makes sense.

4. We derived a few of integral theorems from the Generalized Stokes theorem. For instance, consider the integral of the exterior derivative of a 0 – *form* over a path γ . By generalized Stoke's theorem, we have

$$\int_{\gamma} df(x_1, \dots, x_n) = \int_{\partial \gamma} f(x_1, \dots, x_n) = f(x_1, \dots, x_n) \Big|_{\gamma_i}^{\gamma_f} \quad (9.5)$$

where the boundary of γ is just its end points, hence the integration just becomes the difference of the function evaluated at the final γ_f and the initial γ_i points. As we also have

$$df(x_1, \dots, x_n) = \frac{\partial f}{\partial x_i} dx_i = \nabla f \cdot dx \quad (9.6)$$

we get the Gradient theorem. Similarly, the exterior derivative of a 1 – *form*

$$E(x_1, \dots, x_n) = E_1(x_1, \dots, x_n)dx_1 + \dots E_n(x_1, \dots, x_n)dx_n = \vec{E} \cdot \vec{dx} \quad (9.7)$$

is

$$dE(x_1, \dots, x_n) = \left(\frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2} \right) dx^1 \wedge dx^2 + \dots = (\nabla \times E) \cdot \hat{n} dA \quad (9.8)$$

leading to the curl theorem

Curvilinear Coordinate Transformations

Consider the vector field

$$M(x, y) = x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \quad (10.1)$$

which is a map from the manifold to the tangent bundle, i.e. its output is a vector at the tangent space to the manifold at the input position. We can draw this vector field at the Cartesian plane; for instance, at $(x, y) = (1, 0)$, the vector is $M(1, 0) = \frac{\partial}{\partial y}$, which is a unit vector pointing “north”. Likewise, at $(x, y) = (0, 2)$, $M(0, 2) = -2 \frac{\partial}{\partial x}$ which is a vector of length 2 pointing “west”. If we draw all such vectors in the Cartesian plane, we get the picture that it describes a *counterclockwise rotation* around the origin.

Intuitively, rotation on a plane should be describable with a single parameter, the rotation angle θ . Indeed, if we were to start at any point in the coordinate plane and follow the vector field alongside the vectors, we will observe a simple rotation, describable by the rotation angle. It thus makes sense to switch to a new coordinate system where the vector field was manifestly one-dimensional, i.e. something like $\alpha \frac{\partial}{\partial \beta}$.

Consider two new parameters r and θ as functions of the old Cartesian coordinates:

$$\begin{aligned} r &= f(x, y) & x &= h(r, \theta) \\ \theta &= g(x, y) & y &= j(r, \theta) \end{aligned} \quad (10.2)$$

The vector field then becomes via the chain rule

$$M(x, y) = x \left(\frac{\partial f(x, y)}{\partial y} \frac{\partial}{\partial r} + \frac{\partial g(x, y)}{\partial y} \frac{\partial}{\partial \theta} \right) - y \left(\frac{\partial f(x, y)}{\partial x} \frac{\partial}{\partial r} + \frac{\partial g(x, y)}{\partial x} \frac{\partial}{\partial \theta} \right) \quad (10.3)$$

hence

$$M(t, \theta) = \alpha(r, \theta) \frac{\partial}{\partial r} + \beta(r, \theta) \frac{\partial}{\partial \theta} \quad (10.4)$$

for

$$\alpha(r, \theta) = \left[x \frac{\partial f(x, y)}{\partial y} - y \frac{\partial f(x, y)}{\partial x} \right]_{\substack{x=h(r, \theta) \\ y=j(r, \theta)}} \quad (10.5a)$$

$$\beta(r, \theta) = \left[x \frac{\partial g(x, y)}{\partial y} - y \frac{\partial g(x, y)}{\partial x} \right]_{\substack{x=h(r, \theta) \\ y=j(r, \theta)}} \quad (10.5b)$$

Let us now choose f, g such that $\alpha(r, \theta) = 0$, i.e.

$$\frac{1}{y} \frac{\partial f(x, y)}{\partial y} = \frac{1}{x} \frac{\partial f(x, y)}{\partial x} \quad (10.6)$$

Note that $f(x, y)$ is function with two variables (2 degrees of freedom), and above equation puts 1 constraint: we expect the most general solution to be a *function* of 1 variable. In fact, we can easily check (for

instance using Mathematica, or just by inspection) that the most general solution to this differential equation is

$$f(x, y) = c(x^2 + y^2) \quad (10.7)$$

for the arbitrary function c . This then leads to the condition

$$r = c(h(r, \theta)^2 + j(r, \theta)^2) \quad (10.8)$$

which can be satisfied if we choose

$$\begin{aligned} h(r, \theta) &= \sqrt{c^{-1}(r)} \cos(\theta) \\ j(r, \theta) &= \sqrt{c^{-1}(r)} \sin(\theta) \\ c^{-1}(r) &> 0 \end{aligned} \quad (10.9)$$

Thus, we arrive at the unique parametrization

$$\begin{aligned} r &= c(x^2 + y^2) & x &= \sqrt{c^{-1}(r)} \cos(\theta) \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right) & y &= \sqrt{c^{-1}(r)} \sin(\theta) \end{aligned} \quad (10.10)$$

for an arbitrary function $c^{-1}(r) > 0$, through which

$$M(x, y) = x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \quad (10.11)$$

becomes

$$M(r, \theta) = \frac{\partial}{\partial \theta} \quad (10.12)$$

We usually choose $c^{-1}(r) = r^2$ (hence $c(r) = \sqrt{r}$), which then leads to the standard polar coordinates.

Let us instead consider another vector field, i.e.

$$N(x, y) = x \frac{\partial}{\partial y} + y \frac{\partial}{\partial x} \quad (10.13)$$

We still want to find a parametrization that simplifies this, and all of the arguments above apply all the way upto the differential equation (10.6) which now becomes

$$\frac{1}{y} \frac{\partial f(x, y)}{\partial y} = -\frac{1}{x} \frac{\partial f(x, y)}{\partial x} \quad (10.14)$$

leading to

$$f(x, y) = c(x^2 - y^2) \quad (10.15)$$

for which the parametrization turns into

$$\begin{aligned} r &= c(x^2 - y^2) & x &= \sqrt{c^{-1}(r)} \cosh(\theta) \\ \theta &= \tanh^{-1}\left(\frac{y}{x}\right) & y &= \sqrt{c^{-1}(r)} \sinh(\theta) \end{aligned} \quad (10.16)$$

for an arbitrary function $c^{-1}(r) > 0$, through which

$$N(x, y) = x \frac{\partial}{\partial y} + y \frac{\partial}{\partial x} \quad (10.17)$$

becomes

$$N(r, \theta) = \frac{\partial}{\partial \theta} \quad (10.18)$$

Note that we could have found the correct parametrization for $N(x, y)$ by realizing that $N(x, y) = iM(x, -iy)$ hence we should go to polar coordinates in the $(x, -iy)$ cartesian plane, *not* in (x, y) plane. Nevertheless, we may not always be able to relate one vector field to a known one, hence our generic algebraic way of derivation is the most applicable procedure in general settings.

1. We discussed the above points in class.
2. We briefly discussed the spherical, cylindrical, and polar coordinates. We derived the area and volume elements in these coordinates.

Part III

Complex Analysis for Physicists

Basics of complex calculus

11.1 Structures of complex numbers

11.1.1 Complex numbers as a Euclidean vector space

1. We reviewed the importance of complex numbers in Physics by re-discussing the historical appearance of complex numbers in Physics. Indeed, even though mathematicians had been using complex numbers as tools for centuries, physicists only began using it after the Schrödinger equation:

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = H(t) |\Psi(t)\rangle \quad (11.1)$$

We do not need to understand the details of this equation: this is part of the quantum mechanics course. However, we recognize that it is a first order linear ordinary differential equation for the variable $|\Psi(t)\rangle$, which is actually a vector in an infinite dimensional space.^{cg} As we have discussed in the first part of the semester, we can expand a vector in a basis; in the coordinate basis, this becomes the more familiar version:

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}, t) = H(\vec{x}, t) \Psi(\vec{x}, t) \quad (11.2)$$

In these equations, $H(t)$ is a time-dependent function (called *Hamilton*) which describes the characteristics of the system of which Ψ is called the wavefunction. This much of information is more than enough for our purposes.

The appearance of i in a physical equation was quite shocking to the physics community;^{ch} however, we can actually rewrite it in a way free of i : this equivalent formulation is called *Madelung equations*. This is possible because a complex number is actually equivalent to a 2d vector space over real numbers.

2. We discussed the uniqueness of complex numbers satisfying following conditions:
 - a divisible algebra over real numbers
 - closed under algebra (n^{th} order polynomial has n roots)
 - commutative
3. In class, we basically follow chapter 10 of Hildebrand.

With $\{1, i\}$ as the standard basis, complex numbers are simply a real vector space of dimension two: complex plane as cartesian plane, geometric interpretation of complex numbers, complex absolute value as Euclidean norm, complex conjugation, polar form

^{cg}Note to self: maybe a few words about Hilbert space?

^{ch}Note to self: put some references of Dyson, Dirac, etc. regarding the complex nature of Schrödinger equation.

11.1.2 Further properties of complex numbers

Complex numbers as an algebraically closed field, complex numbers as a commutative algebra over the reals, complex numbers (endowed with the metric $d(z_1, z_2) = |z_1 - z_2|$) as a complete metric space

11.2 Complex differentiation

Cauchy–Riemann equations, holomorphicity, Cauchy’s integral formula

Riemann Surfaces

12.1 Riemann sphere

Extended complex numbers, projective maps

12.2 Analysis of elementary functions via Riemann surfaces

The qualitative analysis of the analytic structure of elementary functions via their Riemann surfaces (square-root, log, etc.),

13

Analytic Structure of Multi-valued “functions”

13.1 Branch points, branch cuts, and singularities

Branch analysis, different sheets, essential singularities, etc.

13.2 Series expansion of complex functions

Taylor and Laurent series, relations to Cauchy’s integral formula, poles and residues, residue theorem

13.3 Analytic continuation and monodromy theorem

About Wolfram Mathematica

A

Bibliography