Originally prepared for *METU - Phys209*

Differential Equations in Physics

Soner Albayrak

February 1, 2024

Summary

These are the notes of the lectures prepared for the "PHYS209 MATHEMATICAL METHODS IN PHYSICS I" course at *Middle East Technical University*, 2023-2024 Fall Semester —see https://soneralbayrak.com/teaching/Phys209. 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).

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.

[a] For 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.

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

[b] See https://www.ctan.org/tex-archive/macros/latex/contrib/tufte-latex/

[c] In the traditional layout, one usually uses endnotes, margin notes, or footnotes; in this paper, most "notes" will be sidenotes with occasional footnotes.[d]

[d] Such as this one.

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

About the course Phys209

As stated in the front page, this "book" is collection of notes prepared for the course Phys209, with the syllabus provided here: https://soneralbayrak.com/teaching/Phys209. Although the title is somewhat generic, the contents of this book is severely restricted and arranged so as to be an appropriate one-semester-long course 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 textbook of the course 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.

[e] For 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

[f] I 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

List of Tables

List of Figures

1

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

$$2 :: \texttt{Integer} \tag{1.1}$$

where $a :: b$ reads as *"a is of type b"*. Further examples would be

$$3/2 :: \texttt{Rational} \tag{1.2a}$$
$$1.56 :: \texttt{Real} \tag{1.2b}$$
$$1 + i :: \texttt{Complex} \tag{1.2c}$$

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

$$3 :: \texttt{Integer} \tag{1.3a}$$
$$3 :: \texttt{Real} \tag{1.3b}$$

A somewhat hybrid notation between mathematicians and computer scientists would be

$$3 :: \mathbb{Z} \tag{1.4a}$$
$$3 :: \mathbb{R} \tag{1.4b}$$

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,

$$x :: \mathbb{Z} \tag{1.5}$$

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

$$y :: \mathrm{A} \tag{1.6}$$

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,

$$f :: \mathbb{Z} \to \mathbb{Z} \tag{1.7}$$

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

$$f = \lambda \to \lambda^2 \tag{1.8}$$

[g] $\mathbb{Z}$ is actually an integral domain.

[h] I personally find this notation clearer when we use it with higher order functions such as derivatives.

[i] This is called a *type variable.*

[j] It 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.

[k] Physicists 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.[ℓ] We will stick to this convention in the rest of the book and refer all multi-valued relations (such as an arctan) as functions.

[ℓ] Mathematically, 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 $\texttt{sqrt} = \lambda \to \sqrt{\lambda}$ is not a function in the complete complex plane, as $\texttt{sqrt}(4) = \pm 2$. 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 $\texttt{sqrt}$. 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 $\texttt{sqrt}$, we can state $\texttt{sqrt} :: \mathbb{C} \to \mathrm{A}$ where $x \in \mathrm{A}$ if and only if $0 \le \arg(x) < \pi$. This means that $\texttt{sqrt}\left(re^{i\theta}\right) = \sqrt{r}e^{i\theta/2}$ for θ chosen in the range $0 \le \theta < 2\pi$ with $r > 0$; hence, $\texttt{sqrt}(4) = 2$.

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 $\texttt{sqrt}$: we can move on to the *other sheets* by removing the restriction on θ. Indeed, we have on the second sheet $\texttt{sqrt}(4) = \texttt{sqrt}\left(4e^{i2\pi}\right) = \sqrt{4}e^{i\pi} = -2$, the other solution! One could go on to higher sheets to find even more solutions; in the case of $\texttt{sqrt}$, 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

$$g(y) = y/3 \tag{1.9}$$

for which we can write down[n]

$$g :: \mathbb{Z} \to \mathbb{Q} \tag{1.10a}$$
$$y :: \mathbb{Z} \tag{1.10b}$$
$$g(y) :: \mathbb{Q} \tag{1.10c}$$

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

$$g :: \mathbb{C} \to \mathbb{C} \tag{1.11a}$$
$$y :: \mathbb{C} \tag{1.11b}$$
$$g(y) :: \mathbb{C} \tag{1.11c}$$

which is still true for $g(y) = y/3$. In fact, we can even write $g :: \mathrm{A} \to \mathrm{B}$ 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,

$$f = \lambda \to \lambda + 5 \tag{1.12a}$$
$$g = \lambda \to 2\lambda + 10 \tag{1.12b}$$

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

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

$$T :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{C} \to \mathrm{D}) \tag{1.13}$$

which means if

$$f :: \mathrm{A} \to \mathrm{B} \tag{1.14}$$

then

$$(g = T \cdot f) :: \mathrm{C} \to \mathrm{D} \tag{1.15}$$

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.

$$\frac{\mathrm{d}}{\mathrm{d}x} :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{C}) \tag{1.16}$$

which means[p]

$$x :: \mathrm{A} \tag{1.17a}$$
$$f :: \mathrm{A} \to \mathrm{B} \tag{1.17b}$$
$$f(x) :: \mathrm{B} \tag{1.17c}$$
$$f' :: \mathrm{A} \to \mathrm{C} \tag{1.17d}$$
$$f'(x) :: \mathrm{C} \tag{1.17e}$$

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 $(-\infty, 0)$: $\mathtt{sqrt}(z)$ for any other $z \in \mathbb{C}$ can then be uniquely determined to be consistent with these; for instance $\mathtt{sqrt}\left(-1 \pm i10^{-100}\right) \sim 6 \times 10^{-17} \pm i$ —note the jump!

[m] I'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 $f(x)$.

[n] If you couldn't remember, $\mathbb{Q}$ denotes the set of rational numbers.

[o] We 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 $h :: \mathrm{A} \to \mathrm{A}$ can only produce integers when acted on integers, reals when acted on reals, and so on.

[p] We are using the common convention $f' := \frac{\mathrm{d}f}{\mathrm{d}x}$ and $f'(x) := \frac{\mathrm{d}f}{\mathrm{d}x}(x)$ for brevity.

For example,

$$f :: \mathbb{R} \to \mathbb{C} \tag{1.18a}$$
$$f = \lambda \to \lambda^2 + 2i \tag{1.18b}$$

leads to

$$f' :: \mathbb{R} \to \mathbb{R} \tag{1.19a}$$
$$f' = \lambda \to 2\lambda \tag{1.19b}$$

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 $\frac{\mathrm{d}}{\mathrm{d}x} :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B})$. This shows that the derivative is a higher order function that can be *repeatedly applied*; thus, we say[r]

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B}) \tag{1.20a}$$
$$f :: \mathrm{A} \to \mathrm{B} \tag{1.20b}$$
$$f^{(n)} :: \mathrm{A} \to \mathrm{B} \tag{1.20c}$$

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]

$$\frac{\mathrm{d}^{-1}}{\mathrm{d}x^{-1}} :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B}) \tag{1.21a}$$
$$g :: \mathrm{A} \to \mathrm{B} \tag{1.21b}$$
$$g^{(-1)} :: \mathrm{A} \to \mathrm{B} \tag{1.21c}$$

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

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

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

We have shown above that the *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:

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

[q] Reminder: if $f = \mathrm{A} \to \mathrm{B}$, we call A (B) the (co)domain of f.

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

[s] In 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

$$\int :: (\mathrm{A} \to \mathrm{B}) \to \mathrm{C} \tag{1.24}$$

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

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

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

$$T :: (\mathbb{R} \to \mathbb{R}) \to \mathbb{R} \tag{1.26a}$$
$$F :: \mathbb{R} \to (\mathbb{R} \to \mathbb{R}) \tag{1.26b}$$

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 $\frac{\mathrm{d}}{\mathrm{d}x}$ and an unknown function f as a differential equation. For instance,

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

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

$$x :: \mathbb{R} \tag{1.35a}$$
$$\left(\frac{\mathrm{d}}{\mathrm{d}x} \cdot f\right) :: \mathrm{A} \to \mathrm{B} \tag{1.35b}$$
$$\frac{\mathrm{d}}{\mathrm{d}x} \cdot f = 0 \tag{1.35c}$$

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

[u] This property can be generalized. A higher order function

$$f :: \mathrm{A}_1 \to (\mathrm{A}_2 \to (\cdots \to (\mathrm{A}_{n-1} \to (\mathrm{A}_n \to \mathrm{B})))) \tag{1.27}$$

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

$$x_i :: \mathrm{A}_i \tag{1.28a}$$
$$f(x_1)(x_2)\dots(x_n) :: \mathrm{B} \tag{1.28b}$$

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

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

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

$$P_n :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B}) \tag{1.30}$$

and

$$(P_0 \cdot f = f) :: \mathrm{A} \to \mathrm{B} \tag{1.31a}$$
$$\left(P_n \cdot f = f \cdot (P_{n-1} \cdot f)\right) :: \mathrm{A} \to \mathrm{B} \tag{1.31b}$$

meaning

$$x :: \mathrm{A} \tag{1.32a}$$
$$(P_n \cdot f)(x) = f(f(\dots f(x))) :: \mathrm{B} \tag{1.32b}$$

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

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

$$\exp :: (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B}) \tag{1.33a}$$
$$\exp = \sum_{n=0}^{\infty} \frac{1}{n!} P_n \tag{1.33b}$$

One can then immediately compute, say,

$$\begin{aligned} \exp\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)x^3 &= x^3 + 3x^2 + 3x + 1 \\ \exp\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)e^{kx} &= e^k e^{kx} \end{aligned} \tag{1.34}$$

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

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

for which one usually writes down the result as

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

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

The next simplest example would be the following differential equation

$$\frac{d}{dx} \cdot f = f \tag{1.38}$$

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

$$f'(x) = f(x) \tag{1.39}$$

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

$$\exp :: \mathbb{C} \to \mathbb{C} \tag{1.40a}$$

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

   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

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

   By chain rule, we have

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

[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 $A \to (0 :: \mathbb{C})$, 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.

[y] By 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

$$\frac{\mathrm{d}x}{\mathrm{d}f(x)} = \frac{1}{f(x)} \tag{1.43}$$

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

$$\frac{\mathrm{d}f^{-1}(y)}{\mathrm{d}y} = \frac{1}{y} \tag{1.45}$$

By integrating this function, we get

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

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]

[z]This means

$$\begin{aligned} f &:: \mathbb{C} \to \mathbb{C} && (1.44a)\\ f^{-1} &:: \mathbb{C} \to \mathbb{C} && (1.44b)\\ f &= x \to f(x) && (1.44c)\\ f^{-1} &= f(x) \to x && (1.44d) \end{aligned}$$

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

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.

[ab]We can show this by using the fundamental theorem of calculus.

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 $\frac{\mathrm{d}}{\mathrm{d}x}$ 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:

$$\begin{gathered}\text{A differential equation is called} \begin{Bmatrix} \text{ordinary} \\ \text{partial} \end{Bmatrix} \text{if there are} \\ \text{derivatives with respect to} \begin{Bmatrix} \text{one} \\ \text{more than one} \end{Bmatrix} \text{variables.}\end{gathered} \tag{1.48}$$

[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,

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = g(x)\,, \quad \frac{\mathrm{d}g(x)}{\mathrm{d}x} = -f(x)\,. \tag{1.47}$$

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

For instance,

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

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,

$$\frac{\mathrm{d}^3}{\mathrm{d}x^3} f(x) = 0 \tag{1.50}$$

is a third order differential equation,[ae] whereas

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

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

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

is called a *linear differential equation*, whereas

$$f(x)\frac{\mathrm{d}}{\mathrm{d}x} f(x) = x^3 \tag{1.56}$$

is a *nonlinear* differential equation.[ag] An easy way to check if a differential equation is linear or nonlinear is to apply the transformation $f(x) \to \lambda f(x)$ 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

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) + \sin(f(x)) = 0 \tag{1.57}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) + f(x) = 0 \tag{1.58}$$

which has the solution

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

[ae] See if you can convince yourself that

$$f(x) = c_0 + c_1 x + c_2 x^2 \tag{1.51}$$

for the coefficients c_i is the solution to this equation.

[af] It is perfectly possible to define the differential equation

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

for which

$$f(x) = x^3 \tag{1.54}$$

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?
$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + f(x) = 0$	✓	✓	✓
$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + f(x) = x^2$	✓	✓	✗
$f(x)\frac{\mathrm{d}^3 f(x)}{\mathrm{d}x^3} + \left(\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right)^2 = 0$	✓	✗	✓
$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + \sin(f(x)) = 0$	✓	✗	✗
$\frac{\partial^2 f(x,y)}{\partial x \partial y} + f(x,y) = 0$	✗	✓	✓
$\frac{\partial^2 f(x,y)}{\partial x^2} + f(x,y) = x^2$	✗	✓	✗
$f(x)\frac{\partial^3 f(x,y)}{\partial x^3} + \left(\frac{\partial f(x,y)}{\partial y}\right)^2 = 0$	✗	✗	✓
$\frac{\partial^2 f(x,y)}{\partial x \partial y} + \sin(f(x,y)) = 0$	✗	✗	✗

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

[ah] We can actually solve the full nonlinear differential equation eqn. (1.57); the result is

$$f(x) = 2\,\mathrm{arccot}\left(\frac{2e^x}{c}\right) \tag{1.60}$$

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

$$\lim_{x\to\infty} 2\,\mathrm{arccot}\left(\frac{2e^x}{c}\right) = \lim_{x\to\infty} ce^{-x} \tag{1.61}$$

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

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

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, $\frac{\mathrm{d}}{\mathrm{d}x}$, 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

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

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

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = x^2 \tag{2.3}$$

we can write down

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

which we can rewrite as

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

by observing that integral is *the inverse of derivative.*[ai]

[ai] Rigorously speaking, we are referring to indefinite integrals (also known as antiderivatives or Newton integrals).

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

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

is

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

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.

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

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} \; : : \; (\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{B}) \tag{2.8a}$$

$$f \; : : \; \mathrm{A} \to \mathrm{B} \tag{2.8b}$$

$$f^{(n)} \; : : \; \mathrm{A} \to \mathrm{B} \tag{2.8c}$$

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

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

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:

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

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

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

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,

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

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

$$\ker\left[R(\theta)\right] = \{\vec{0}\} \tag{2.16}$$

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

$$S = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \tag{2.17}$$

[aj] We can easily see the linearity by noting the relation

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(c_1 f(x) + c_2 g(x)\right) = c_1 \frac{\mathrm{d}^n}{\mathrm{d}x^n} f(x) + c_2 \frac{\mathrm{d}^n}{\mathrm{d}x^n} g(x) \tag{2.9}$$

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 $\frac{\mathrm{d}}{\mathrm{d}x}$ 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 $\frac{\mathrm{d}}{\mathrm{d}x} f(x) = x^2$ 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

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

where f_n can be viewed as an infinite-dimensional vector $f_n = (f_0, f_1, \dots)$.[aℓ]

[aℓ] 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

$$v \; : : \; \mathbb{N} \to \mathbb{R} \tag{2.11a}$$

$$v = n \to \begin{cases} 1 \text{ if } n = 1 \\ 0 \text{ if } n = 2 \\ -3 \text{ if } n = 3 \\ \text{undefined otherwise} \end{cases} \tag{2.11b}$$

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

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

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

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

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:

$$S \cdot \vec{v}_{\text{input}} = \vec{v}_{\text{output}} \tag{2.20a}$$

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

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

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.

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

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

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{2.21}$$

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

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

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

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

has the unique answer

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

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

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

the solution then becomes

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

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]

$$f(x)=e^x \tag{2.27}$$

We can actually generalize this to[ao]

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

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

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

has the *solution*

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

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

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

[an] As we discussed in that section, this result is either a definition or a derived result depending our conventions.

[ao] One way to show this is the judicious use of the chain rule as follows:

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

which means that

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

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,

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

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

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

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]

$$f :: \mathbb{C} \to \mathbb{C} \tag{2.36a}$$

$$f = x \to \sum_{i=1}^{n} c_i e^{r_i x} \tag{2.36b}$$

for arbitrary constants c_i.

Differential equations are usually given in the form

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

which can be brought to the form eqn. (2.35) by simply finding the roots of the equation[aq]

$$a_0 + a_1 r + a_2 r^2 + \cdots + a_n r^n = 0 \tag{2.38}$$

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

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

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

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

[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

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

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

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

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

$$r_1 = r\,, \quad r_2 = r+\delta\,, \quad c_1 = a - b\,, \quad c_2 = b \tag{2.43}$$

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

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

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 out 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

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

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

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

yielding $f(x) = (a_1 + a_2 x + \ldots a_n x^{n-1})e^{rx}$.

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

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

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

[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

$$\left(\frac{\mathrm{d}}{\mathrm{d}x} - r\right)^2 (xe^{rx}) = 0 \tag{2.45}$$

[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]

$$\mathfrak{IT} :: (\mathbb{C} \to \mathbb{C}) \to (\mathbb{C} \to \mathbb{C}) \tag{2.50a}$$

$$\mathfrak{IT} = (x \to f(x)) \to \left| s \to \int_\alpha^\beta K(x,s) f(x) dx \right| \tag{2.50b}$$

where $\mathfrak{IT}$ 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 tranformation*: different kernels (along with different integration ranges) lead to different integral transforms.

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

$$\mathcal{L} :: (\mathbb{C} \to \mathbb{C}) \to (\mathbb{C} \to \mathbb{C}) \tag{2.51a}$$

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

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

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

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

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

or in a more conventional notation, we state

$$\frac{\mathrm{d}g(x)}{\mathrm{d}x} \xrightarrow{\text{Laplace transform}} sG(s) - g(0) \tag{2.54}$$

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:

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

[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 $x \to f(x)$ and $y \to f(y)$ 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:

$$\mathfrak{IT} :: (\mathbb{C} \to \mathbb{C}) \to (\mathbb{C} \to \mathbb{C}) \tag{2.49a}$$

$$\mathfrak{IT} = (x \to f(x)) \to \left(s \to \int_\alpha^\beta K(x,s) f(x) dx \right) \tag{2.49b}$$

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:

$$\mathfrak{IT} = (x \to f(x)) \to \left(s \to \int_\alpha^\beta K(x,s) f(x) dx \right)$$

$$\mathfrak{IT} = (y \to f(y)) \to \left(z \to \int_\alpha^\beta K(x,z) f(x) dx \right)$$

$$\mathfrak{IT} = (y \to g(y)) \to \left(z \to \int_\alpha^\beta K(s,z) g(s) ds \right)$$

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.

$$\mathfrak{IT}_{K,\alpha,\beta} = (y \to g(y)) \to \left(z \to \int_\alpha^\beta K(s,z) g(s) ds \right)$$

$$\mathfrak{IT}_{T,\gamma,\lambda} = (y \to g(y)) \to \left(z \to \int_\gamma^\lambda T(s,z) g(s) ds \right)$$

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:

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

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

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

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

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

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 $n = 2$:

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

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 bla*to 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 $i(0) = i'(0) = \cdots = 0$).

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.

3

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 $\left(x\frac{\mathrm{d}}{\mathrm{d}x}+a\right)f(x)=0$ and $\left(x^k\frac{\mathrm{d}}{\mathrm{d}x}+a\right)f(x)=0$ for $k\neq 1$.

3. How these differential operators can be chained, and how it leads to the following form:
$$\left(x^k\frac{\mathrm{d}}{\mathrm{d}x}+a_1\right)\cdots\left(x^k\frac{\mathrm{d}}{\mathrm{d}x}+a_n\right)f(x)=0 \tag{3.1}$$
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.
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y}&=\frac{1}{1-n}x^n\frac{\mathrm{d}}{\mathrm{d}x}\text{ for }y=x^{-n+1}\\ \frac{\mathrm{d}}{\mathrm{d}y}&=x\frac{\mathrm{d}}{\mathrm{d}x}\text{ for }y=\log(x)\end{aligned} \tag{3.2}$$

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:
$$\begin{gathered}f''(x)+p(x)f'(x)+q(x)f(x)=0\\ \Rightarrow\\ (u')^2f''(u)+(u''(x)+u'(x)p(x))y'(u)+q(x)y(u)=0\end{gathered} \tag{3.3}$$

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

9. An example equation where change of variables would work: $ty'' + (t^2 - 1)y' + t^3 y = 0$.

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): $xf''(x) + f'(x) = 0$, $x^2 f''(x) + 2xf'(x) = 2$. Solve these!

12. Another thing to check is if the equation is *exact*, i.e. if it can be rewritten as a total derivative:

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

$$\xRightarrow{???}$$

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

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): $f''(x) + xf'(x) + f(x) = 0$, $x^2 f''(x) + xf'(x) - f(x) = 0$.

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): $xf''(x) - f'(x) + 4x^3 f(x) = 0$ 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 $\epsilon :: \{\mathbb{Z}^+, \dots, \mathbb{Z}^+\} \to \mathbb{Z}$ and discussed its properties.

$$\epsilon :: \{\mathbb{Z}^+, \dots, \mathbb{Z}^+\} \to \mathbb{Z} \tag{3.4a}$$

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

Example: $(132) \to (123)$: we need 1 permutation for (132): $\epsilon_{132} = -1$. $(2314) \to (2134) \to (1234)$: 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 $\det :: \mathfrak{M}_{n\times n}(\mathbb{C}) \to \mathbb{C}$ in terms of Levi-Civita symbol ϵ and discussed its properties.

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

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

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.

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

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

$$\dots \tag{3.6c}$$

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

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

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

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 $\mathcal{D}_1 \cdot \mathcal{D}_2 \cdot f = 0$ 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. $f(x) = \sum\limits_{n=0}^{\infty} a_n(x-x_0)^n$

23. Introduced the classification of expansion points: given the differential equation

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

    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 points of $\left(x\frac{\mathrm{d}}{\mathrm{d}x} + \sin(x)\right) f(x) = 0$.

24. Around ordinary points, Taylor series expansion works and one can get all solutions correctly.

25. Rewrite the differential equation above as

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

    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.

$$x^2 f''(x) - x f'(x) + (1 + x) f(x) = 0 \tag{3.10}$$

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:

$$
\begin{aligned}
* &:: \mathbb{C} \to \mathbb{C} && (3.11a)\\
* &= z \to z^* = \operatorname{Re} z - i \operatorname{Im} z && (3.11b)\\
T &:: \mathfrak{M}_{n\times n}(\mathbb{C}) \to \mathfrak{M}_{n\times n}(\mathbb{C}) && (3.11c)\\
T &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \to \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} && (3.11d)\\
\dagger &:: \mathfrak{M}_{n\times n}(\mathbb{C}) \to \mathfrak{M}_{n\times n}(\mathbb{C}) && (3.11e)\\
\dagger &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \to \begin{pmatrix} a^*_{11} & a^*_{21} & \dots & a^*_{n1} \\ a^*_{12} & a^*_{22} & \dots & a^*_{n2} \\ \dots \\ a^*_{1n} & a^*_{2n} & \dots & a^*_{nn} \end{pmatrix} && (3.11f)
\end{aligned}
$$

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, $U = \exp\left(i\frac{\mathrm{d}}{\mathrm{d}x}\right)$ 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:

$$f(g) = \int_{\hat{G}} d\pi \, \mathrm{tr}\Big(\hat{\pi}(g)^{-1}\hat{f}(\pi)\Big) \tag{3.12a}$$

$$\hat{f}(\pi) = \int_{G} dh\hat{\pi}(g)f(h) \tag{3.12b}$$

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
    - $\hat{\pi} \; :: \; g \to \mathrm{End}(V_\pi)$ 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 210
$f(g)$	$\{\hat{\pi}(g)\}$	$f(g) = \int_{\hat{G}} d\pi \, \mathrm{tr}\Big(\hat{\pi}(g)^{-1}\hat{f}(\pi)\Big)$	$\{\hat{f}(\pi)\}$ (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

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

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

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)
$f(x)$	$\{x^i\}$	$f(x)=\sum_{n=0}^{\infty}a_nx^n$	$\{a_n\}$ (countable infinite set)
$f(x)$	$\{e^{ikx}\}$	$f(x)=\int_{-\infty}^{\infty}dke^{ikx}\hat{f}(k)$	$\left\{\hat{f}(k)=\int_{-\infty}^{\infty}e^{-ikx}f(x)\right\}$ (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.

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

This leads to discrete time fourier series:

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

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

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

- Fourier transform:

$$f::\mathbb{C}\to\mathbb{C} \tag{3.16a}$$

$$\hat{f}::\mathbb{C}\to\mathbb{C} \tag{3.16b}$$

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

$$\hat{f}(k)=\int_{-\infty}^{\infty}dxe^{-ikx}f(x) \tag{3.16d}$$

- Fourier series:

$$f::[a,a+T]\to\mathbb{C} \tag{3.17a}$$

$$\hat{f}::\mathbb{Z}\to\mathbb{C} \tag{3.17b}$$

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

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

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:

$$f \; :: \mathbb{Z} \to \mathbb{C} \tag{3.18a}$$
$$\hat{f} \; :: \; [a, a+T] \to \mathbb{C} \tag{3.18b}$$
$$f(n) = \frac{1}{T} \int_{b}^{b+T} dx e^{i\frac{2\pi n}{T}k} \hat{f}(k) \tag{3.18c}$$
$$\hat{f}(k) = \sum_{n=-\infty}^{\infty} e^{-i\frac{2\pi n}{T}k} f(n) \tag{3.18d}$$
$$\tag{3.18e}$$

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

- Discrete Fourier series:

$$f \; :: \mathbb{Z}_N \to \mathbb{Z}_N \tag{3.19a}$$
$$\hat{f} \; :: \; \mathbb{Z}_N \to \mathbb{Z}_N \tag{3.19b}$$
$$f(n) = \frac{1}{N} \sum_{m=0}^{N-1} e^{i\frac{2\pi nm}{T}} \hat{f}(m) \tag{3.19c}$$
$$\hat{f}(m) = \sum_{n=0}^{N-1} e^{-i\frac{2\pi nm}{T}} f(n) \tag{3.19d}$$
$$\tag{3.19e}$$

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

42. Summary: finite range/periodic in one domain ↔ discrete in the other domain

43. Drew plots of these transformations for visualization in class.

44. Introduced two higher order functions E and O:

$$E \; :: (\mathbb{C} \to \mathbb{C}) \to (\mathbb{C} \to \mathbb{C}) \tag{3.20a}$$
$$E = (x \to f(x)) \to \left(x \to f_E(x) = \frac{f(x) + f(-x)}{2} \right) \tag{3.20b}$$
$$O \; :: (\mathbb{C} \to \mathbb{C}) \to (\mathbb{C} \to \mathbb{C}) \tag{3.20c}$$
$$O = (x \to f(x)) \to \left(x \to f_E(x) = \frac{f(x) - f(-x)}{2} \right) \tag{3.20d}$$

    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

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

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 $x \to -x$ leads to
$$[(\text{F.T.} \cdot f_{RE})(k)]^* = (\text{F.T.} \cdot f_{RE})(k) \tag{3.22}$$
Likewise, taking $k \to -k$ and $x \to -x$ leads to
$$(\text{F.T.} \cdot f_{RE})(-k) = (\text{F.T.} \cdot f_{RE})(k) \tag{3.23}$$
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, $(\text{F.T.} \cdot f_R)(-k) = [(\text{F.T.} \cdot f_R)(k)]^*$, hence we only need to know positive frequencies, consistent with our everyday experience.

48. Consider the Fourier transform of the function
$$f :: \left[-\frac{T}{2}, \frac{T}{2}\right] \to \mathbb{C} \tag{3.24a}$$
$$f = x \to 1 \tag{3.24b}$$
We computed Fourier series expansion of this in class, found $\hat{f}(n) = T\delta_{n0}$ and checked consistency. We then considered the limit $T \to \infty$ and argued that Kronecker-delta function turns into Dirac-delta distribution.

49. We also considered the function
$$f :: \mathbb{R} \to \mathbb{C} \tag{3.25a}$$
$$f = x \to \begin{cases} 1 & |x| \leq \dfrac{T}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.25b}$$
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 $T \to \infty$ takes sinc functio 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

$$x f'(x) - 2 f(x) = 6x^4 \tag{3.26}$$

   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

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

   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

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

   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

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

   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 know we have n unknown functional degrees of freedom. In fact, we can impose $n-1$ constraints, which we choose as

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

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

$$\sum_{i=1}^{n} c_i'(x) f_i^{(n-1)}(x) = g(x) \tag{3.31}$$

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

$$\begin{pmatrix} c_1'(x) \\ c_2'(x) \\ \dots \\ c_n'(x) \end{pmatrix} = \begin{pmatrix} f_1(x) & f_2(x) & \dots & f_n(x) \\ f_1'(x) & f_2'(x) & \dots & f_n'(x) \\ \dots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \dots & f_n^{(n-1)}(x) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ \dots \\ g(x) \end{pmatrix} \tag{3.32}$$

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:

$$\mathrm{adj} :: \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathcal{M}_{n\times n}(\mathbb{C}) \tag{3.33a}$$

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

$$b_{i_n k_n} = \frac{1}{(n-1)!}\epsilon_{i_1\dots i_n}\epsilon_{k_1\dots k_n} a_{i_1k_1}\dots a_{i_{n-1}k_{n-1}} \tag{3.33c}$$

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

$$A^{-1} = \frac{\mathrm{adj}(A)}{\det A} \tag{3.34}$$

where we have defined and discussed determinant before:

$$\det :: \mathfrak{M}_{n\times n}(\mathbb{C}) \to \mathbb{C} \tag{3.35a}$$

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

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 diffential 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

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{4.1}$$

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

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

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,..., 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

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

for

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

and

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

We immediately say that the solution is

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

for the arbitrary column matrix C.

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

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

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

$$\left[f'(t) = af(t) \rightarrow f(t) = \exp(at)c\right] \Rightarrow \left[V'(t) = A \cdot V(t) \rightarrow V(t) = \exp(At) \cdot C\right] \tag{4.8}$$

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:

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

however

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

The correct version is

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

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 $f(x) = c_1e^{-x}+c_2e^{-2x}$. Nevertheless, let's see how we can get this answer through matrix computation.

   We realize that this equation can be rewritten as

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

   hence we can immediately write down the answer as

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

   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 wotk 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

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t) = A(t) \cdot V(t) \tag{4.14}$$

   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*

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

   Now consider a modified version of this as

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

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

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

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 $\epsilon \to 1$ 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:

$$\left(V(0) + \epsilon V_{(1)}(t) + \epsilon^2 V_{(2)}(t) + \ldots\right) = V(0) + \epsilon \int_0^t A(t') \cdot \left(V(0) + \epsilon V_{(1)}(t') + \epsilon^2 V_{(2)}(t') + \ldots\right) dt' \tag{4.18}$$

which leads to

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

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

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

which can be rewritten as

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

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

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

which can be rewritten by changing the dummy variable as

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

Observe that if we define

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

then we get

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

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

$$\begin{aligned} &\mathcal{T} :: (\mathbb{R}^n \to \mathbb{C}) \to (\mathbb{R}^n \to \mathbb{C}) \\ &\mathcal{T} = \Big((t_1,\dots,t_n) \to f(t_1,\dots,t_n)\Big) \to \Big((t_1,\dots,t_n) \to f \cdot \mathcal{O}_> \cdot (t_1,\dots,t_n)\Big) \end{aligned} \tag{4.26}$$

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

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

we then see that

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

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

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

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

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

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 $\epsilon \to 1$ and obtain the final result:

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

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-Lioville 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

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

We solved in class for $C = 1mF$, $L = 200H$, and $R = 1/6kOhm$ via exponentiation of the matrix. As teasor, 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 mods, 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

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

This is tedious!

Observe that

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

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

$$e^A = U \cdot \begin{vmatrix} e^{\lambda_1} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2} & \ldots & 0 \\ \ldots & & & \\ 0 & \ldots & 0 & e^{\lambda_n} \end{vmatrix} \cdot U^{-1} \tag{5.4}$$

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 $U = \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{pmatrix}$, then $A \cdot U = U \cdot D$ implies

$$A \cdot \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\vec{u}_1 & \lambda_2\vec{u}_2 & \ldots & \lambda_n\vec{u}_n \end{pmatrix} \tag{5.5}$$

which implies

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

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

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

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:

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

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.

$$A = U \cdot D \cdot U^{-1} \tag{5.9}$$

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 right gets decomposed into monochromotic 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

$$\mathcal{D} \cdot f(x) = \lambda f(x) \tag{5.10}$$

for a differential operator $\mathcal{D}$, 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 $(.) :: (\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}$, extend them to functions $\mathbb{R} \to \mathbb{R}$ by defining the inner product $\langle , \rangle :: (\mathbb{R} \to \mathbb{R}, \mathbb{R} \to \mathbb{R}) \to \mathbb{R}$ as $\langle f, g \rangle = \int\limits_{\mathbb{R}} f(x) g(x) dx$, and then generalize this further to complex functions as

$$\langle , \rangle_\omega :: (A \to \mathbb{C}, A \to \mathbb{C}) \to \mathbb{C} \tag{5.11a}$$

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

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

6

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:

$$(x_1, x_2, \dots, x_n) \ :: \ T_1 \times T_2 \times \cdots \times T_n \tag{6.1}$$

   where $\times$ denotes a product, and $T_1 \times T_2 \times \cdots \times T_n$ 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

$$\pi_i \ :: \ (T_1 \times T_2 \times \cdots \times T_n) \to T_i \tag{6.2a}$$

$$\pi_i = (x_1, x_2, \dots, x_n) \to x_i \tag{6.2b}$$

   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:

$$f :: \mathbb{R}^3 \to \mathbb{R} \tag{6.3a}$$
$$f = (x_1, x_2, x_3) \to f(x_1, x_2, x_3) \tag{6.3b}$$

4. Introduced D-notation:

$$D :: \mathbb{Z}_+ \to (\mathbb{R}^n \to \mathbb{C}) \to (\mathbb{R}^n \to \mathbb{C}) \tag{6.4a}$$
$$D = i \to ((x_1, \dots, x_n) \to f(x_1, \dots, x_n)) \to \left((x_1, \dots, x_n) \to \frac{\partial f(x_1, \dots, x_n)}{\partial x_i}\right) \tag{6.4b}$$

e.g. $D_2 f(x, y) = \dfrac{\partial f(x, y)}{\partial y}$

5. Chain rule:

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

6. Consider two functions f and g:

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

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:

$$D_1 f(t, x) - \alpha^2 D_2^2 f(t, x) = 0 \qquad \text{(heat conductance equation)} \tag{6.7a}$$
$$D_1^2 f(t, x) - c^2 D_2^2 f(t, x) = 0 \qquad \text{(wave equation)} \tag{6.7b}$$
$$D_1^2 f(t, x) + D_2^2 f(t, x) = 0 \qquad \text{(Laplace equation)} \tag{6.7c}$$

8. We solved these equations both via method of seperation 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!

A

About Wolfram Mathematica

Bibliography