

DO NOT WORRY ABOUT YOUR DIFFICULTIES IN MATHEMATICS. I CAN ASSURE YOU MINE ARE STILL GREATER.

ALBERT EINSTEIN

I LOVE ONLY NATURE, AND I HATE MATHEMATICIANS.

RICHARD FEYNMAN

THE ONLY WAY TO LEARN MATHEMATICS IS TO DO MATHEMATICS

PAUL HALMOS



FABRICE P. LAUSSY

WOLVERHAMPTON  
LECTURES OF PHYSICS  
III  
MATHEMATICS

UNIVERSITY OF WOLVERHAMPTON

Copyright © 2020 Fabrice P. Laussy

<http://laussy.org>

# Contents

<i>Lecture 1: The Platonic and Physical universes</i>	9
<i>Lecture 2: Arithmetics and Algebra</i>	19
<i>Lecture 3: Logic, proofs and notations</i>	33
<i>Lecture 4: Complex numbers</i>	41
<i>Lecture 5: Vectors</i>	51
<i>Lecture 6: Functions</i>	61
<i>Lecture 7: Infinites</i>	75
<i>Lecture 8: Infinitesimals</i>	85
<i>Lecture 9: Inverses and compositions of functions.</i>	93
<i>Lecture 10: Taylor Polynomials</i>	101
<i>Lecture 11: Kets and Bra.</i>	115

<i>Lecture 12: Areas.</i>	127
<i>Lecture 13: Basis.</i>	141
<i>Lecture 14: Linear functions.</i>	155
<i>Lecture 15: Inverses &amp; Determinants.</i>	171
<i>Lecture 16: Eigensystems.</i>	183
<i>Lecture 17: Fourier Series.</i>	193
<i>Lecture 18: Differential equations.</i>	207
<i>Lecture 19: Wronskians</i>	219
<i>Lecture 20: Jacobians.</i>	233
<i>Lecture 21: The equations of Physics.</i>	249

*Dedicated to François Dumas,  
who was teaching me some of this  
when I was the student.*





## Lecture 1: The Platonic and Physical universes

Exact Science starts with counting. Today, our Science is exact to the point that one can measure with an accuracy of 0.0000000000000001, but before we get to such highly advanced concepts, let us start at the beginning, where counting means quantifying how many objects we have in a collection:



Here we have one die and four pawns. On the faces of the die, we can count four, five and six dots. There are five objects in total. The important thing here, is that one is “counting” different things. This is a highly advanced concept that takes much time for children to capture (although some birds also master it): although we are dealing with different objects, there arises something common or unique about them when it comes to “how many” there are. The “what” recedes and the “how many” proceeds, taking the shape of so-called *numbers*:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots \quad (1)$$

These objects exist independently of any physical reality, that is, of any object to map to. If you think of 3 in your mind, maybe you picture three little blurry light balls, or depending on the context that brings you to think of that, of three seats, three friends, three hours, three opportunities, etc. Most likely you picture the shape 3 with some associated feeling that it’s more than 2 and less than 4.

There are two immediate things we can do with numbers, adding them, which corresponds to grouping the objects they map with. This is very easy and about the first operation a child learns to master.

Note that one can keep adding, for instance,  $1 + 1 = 2$ , adding 1 to that brings us to  $2 + 1 = 3$  or  $1 + 1 + 1 = 3$ . And here comes the first form of the other immediate thing: *multiplication*, as an *repeated addition*:

$$\underbrace{a + a + \cdots + a}_{n \text{ times}} = n \times a \quad (2)$$

Maybe it is easier seen in this form:

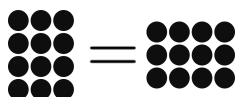
$$a \times n = \left. \begin{array}{cccc} 1 & + & 1 & \cdots & + & 1 \\ + & 1 & + & \vdots & \ddots & + & 1 \\ + & 1 & + & 1 & \cdots & + & 1 \end{array} \right\} n \text{ times}$$

$\underbrace{\hspace{10em}}_{a \text{ times}}$

in which case it is clear that

$$a \times n = n \times a. \quad (3)$$

If it's not clear, think of a particular example:



This is  $3 \times 4 = 4 \times 3$ , and satisfy yourself that this holds for any numbers  $a$  and  $n$ .

This (Eq. (3)) is called *commutativity* of the multiplication (or product). Commutation of operations is important, and may look trivial but we will see repeatedly how it relates to deep features related to ordering things.<sup>1</sup>

Multiplication is so common that we typically get rid of the sign  $\times$  when this does not lead to confusion (of course we don't write  $32 = 23 = 6$ ). So we write:

$$an = na. \quad (4)$$

We should avoid notations like  $3 \cdot 2 = 6$  or  $3 * 2$  or even, if you're attentive,  $3 \times 2$ , as these notations are used for other meanings and are also not as pretty.<sup>2</sup>

Repeated addition is one way to think about multiplication, but not the only one. Another is to see it as a rescaling, as we will see in the next Lecture. This is also particularly useful for Physicists as it introduces notions of *dimensionality*.

We now illustrate further how numbers are, beyond a mere notation or a bookkeeping device, abstract objects that truly exist by themselves, in the sense that they have properties of their own, obey some rules, follows some logic and even some patterns, many that we still do not understand. Such facts about abstract numbers can be explored and discovered. This is certainly not something that we

<sup>1</sup> Is addition commutative? What is the geometric interpretation of this fact? Are every operations commutatives?

<sup>2</sup> What is wrong with  $3 \times 2$ ?

make up. This is something that exist independently from us, in a universe which, although we cannot locate it in space, certainly exist. We call it the *Platonic universe*. The discipline that deals with this place is Mathematics.

Not everything is known about the properties of these objects. For instance, there is a link between the number of distinct entries formed by the sums and the products of any list of numbers. This is known as the Erdős–Szemerédi theorem, which states that there exist positive constants  $c$  and  $\varepsilon$  such that

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{1+\varepsilon} \quad (5)$$

where  $|A|$  is the *cardinality* (that is, the number of elements) of the collection of objects gathered in a so-called “set”  $A$ . We have also introduced the notations  $A + A = \{a + b : a, b \in A\}$  (so-called “sumset”) and  $A \cdot A = \{ab : a, b \in A\}$ . It was conjectured by Erdős and Szemerédi that  $\varepsilon$  is very close to 1. So far, this has not been proved, but we know that  $\varepsilon$  is very close to  $1/3 + 5/5277$ . So much remains to understand (and discovered) from the most basic properties of numbers!

Even more significantly, a very important properties of integer, is that some of them are products of others. For instance:

$$15 = 3 \times 5.$$

But not all of them. The numbers 3 and 5, for instance, are not products of other numbers. Such numbers, which are not products of other integers, are called *prime numbers*. They are very important. And also much remains to be discovered and understood about them. For instance, it is conjectured (but still unknown) that every even integer greater than 2 can be expressed as the sum of two primes (this is known as Goldbach’s conjecture). Another example relates to those primes that just skip the even number that separates them (all primes except 2 are odd)<sup>3</sup>. For instance 5 and 7 are twin primes. There is a conjecture that the number of such twin primes is infinite. All this is not known, although strongly suspected, especially since computers can check numerically for a lot of numbers, but since they cannot look at *all* the numbers (since there is an infinite number of them), they cannot prove it. They could, however, *disprove* it, by finding a counter-example.

<sup>3</sup> Prove this.

We have emphasized things we do not know yet, to show the limitation of our knowledge. But a lot is known as well. For instance, it is known (since Euclid), that there is an infinite number of primes.

Actually, looking at it in this way, this looks pretty much like real objects (like apples or dices), which also follow rules (of gravity, how

they collide, how they respond to pressure, etc.) The discipline that deals with “real objects” is Physics.

The difference being the two disciplines is that some objects exist here in front of us while the others belong to some ethereal, dreamy-universe. The latter have a purity, an exactness, an absolute and incorruptible which is out of reach of the rough, approximate, relative and corruptible physical objects.

It has been a great deal to find how to properly define numbers in Mathematics and today’s most popular acception is in terms of “set theory”, the theory of the sets previously introduced, which is a system able to provide a foundation of all of Mathematics in terms of axioms that derive from the notion of collections of objects (sets). Incredibly, we can start everything with only the empty set (notation  $\emptyset$ ), which is a collection of nothing, so we could say that this is

$$\emptyset = \{ \} \quad (6)$$

where we put between bracket what the set is collecting (here, well, nothing). With this we can define the set whose content is the empty set, and this we call 1:

$$1 = \{ \emptyset \} . \quad (7)$$

Then we can have the set which content is 1 along with the empty set, and this we call 2:

$$2 = \{ 0, 1 \} . \quad (8)$$

Don’t laugh, we managed to get our first numbers out of only the empty set:  $2 = \{ \emptyset, \{ \emptyset \} \}$ . This process can be carried over with the concept of number materialising as the cardinality (number of elements) of the set<sup>4</sup> This is called the Zermelo-Fraenkel theory and if you wonder why that is at all necessary, you should get interested into famous problems such as the axiom of choice and paradoxes like Russel’s paradox, that bring Mathematicians to worry about such concepts tackled in such a bizarre way.

<sup>4</sup> Define 3 and 4 and give their expressions in terms of the empty set only.

As Physicists we are not immediately concerned with that. So in the interest of time, we will assume a working concept and personal intuition of natural numbers, which is something that you have since you are 5 or so. This collection, or set, of numbers, we call it  $\mathbb{N}$ , the set of *natural numbers*:  $\mathbb{N} = \{ \emptyset, \{ \emptyset \}, \{ \emptyset, \{ \emptyset \} \}, \dots \}$ . Incidentally,  $\emptyset$  in  $\mathbb{N}$  we shall write 0 and call it “zero”. So for us:

$$\mathbb{N} = \{ 0, 1, 2, \dots \} . \quad (9)$$

From these objects, we can generalize to other types of numbers, starting with negative numbers. In abstract Mathematics, they arise from the notion of “inverse”, which is confusing because this term we shall use for something else and refer to the “opposite” instead. The

term is because it applies to all abstract objects (not only numbers) and requires that there is an abstract operation (not only addition) that such that two elements from this structure (which is called a “group”) can be combined to yield a particular element called the identity. That sounds like multiplication but group features applies equally to addition, in which case the inverse is the opposite and the identity is zero. Groups are important in Physics but we will really need them much later on. So we’ll come back to our working notions and intuitive understanding of this, which is that of “*negative numbers*”.

Indeed, These are easy to understand, in our times of bank accounts, that easily go below zero. This set of numbers—the natural numbers augmented with their opposite—we call  $\mathbb{Z}$  (“Zahlen”, form the German for numbers). These are called the *integer numbers*.<sup>5</sup>

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\} \quad (10)$$

This seems to introduce another operation: subtraction. But this is actually merely summing with a negative number, i.e.,

$$a - b \equiv a + (-b). \quad (11)$$

Here we introduced a convenient notation,  $\equiv$  instead of  $=$ , to mean “is defined by” rather than “is equal to”. It is also “equal to”, by definition!

The same previous ideas hold for multiplication. Each natural number  $q$ , *except zero*, has a so-called “inverse” (this time in the common sense of the term) which is the quantity that we need to sum to itself  $q$  times to get 1:

$$\underbrace{\frac{1}{q} + \frac{1}{q} + \dots + \frac{1}{q}}_{q \text{ times}} = 1. \quad (12)$$

For instance, we need to add 0.5 two times to get 1, and we need to add 0.125 eight times to get 1. If instead to sum to 1 we want to sum to  $p$ , we then form the number

$$\frac{p}{q} \quad (13)$$

which is, again, the quantity that we need to add  $q$  times to itself to get  $p$ . These numbers are useful, because they can get very close to any quantity we wish to approximate. For instance, say that we are looking for the number whose square—or product by itself—is the number 2 (this is known as the square root of 2, and is written  $\sqrt{2}$ ). Then one can check that

$$\frac{141421}{100000} \quad (14)$$

<sup>5</sup> Is the order in which “objects” appear in a set important? No. Find a way to write  $\mathbb{Z}$  that does not feature this annoying “double-infinite” of Eq. 10.

(which by the way, we would write as 1.41421) is very close, because

$$\left(\frac{141421}{100000}\right)^2 = \frac{19999899241}{10000000000} \approx 1.99999 \dots \quad (15)$$

This is not exact, indeed we have approximated the 5th number after the decimal as a 9 where it really is a 8, but itself followed by a nine so closer to .00009 than to .00008. It is not exact but we can get closer, if we want, for instance:

$$\left(\frac{14142135623}{10000000000}\right)^2 = \frac{19999999979325598129}{100000000000000000000} \approx 1.999999999 \dots \quad (16)$$

which is this time all nine to the ninth decimal (then, this time, it's an 8 that follows).<sup>6</sup> As such, these numbers interpolate between the integers.  $\sqrt{2}$  is a number between 1 and 2 which is slightly less than halfway them. The decimal expansion is a nice representation, but sometimes unpractical. For instance,  $1/3 = 0.333333 \dots$  never ends. It does, however, repeat. All  $p/q$  numbers eventually repeat, however long is the pattern:

$$1/7 = 0.142857142857142857142857142857142857 \dots$$

We call these numbers  $p/q$  which are ratios of integers, the *rational*s. The set of all rationals is called  $\mathbb{Q}$ . The same rational number can have various representation, for instance,  $6/20$  and  $15/50$  are both the same number, namely, 0.3.

Here, it is handy to use prime numbers to have a unique way to write what is eventually a unique number. The unique way to represent this is to write the *numerator* and *denominator* as products of primes:

$$0.3 = \frac{2 \times 3}{2 \times 2 \times 5} = \frac{3}{10}. \quad (17)$$

The numbers that are common to both the numerator and denominator make up the so-called *greatest common divisor* or *gcd*:

$$\text{gcd}(6, 20) = 2. \quad (18)$$

After simplification by the gcd, we find the irreducible expression<sup>7</sup> for this fraction  $6/20 = 3/10$ .

A first great commotion in Science was the discovery that not all numbers are rationals. This was made (one story says) by a member of the Pythagorean school (a so-called Hippasus), who made the discovery on a boat. This was so shocking that he was sentenced to death on the spot, and tossed to the sea. The simplest proof goes as follows:

Consider

$$\frac{a}{b} = \sqrt{2} \quad (19)$$

<sup>6</sup> Do you agree that the number squared in Eq. 16 is closer to  $\sqrt{2}$  than the number in Eq. (14)? Can you provide a still better approximation than Eq. (16)?

<sup>7</sup> Give the irreducible form of the approximation for  $\sqrt{2}$  given by Eq. (14).

with  $\gcd(a, b) = 1$ , i.e., we simplified the fraction already. Squaring both sides (or by definition of the square root), this means:

$$\frac{a^2}{b^2} = 2 \quad (20)$$

and, multiplying both sides by  $b^2$

$$a^2 = 2b^2. \quad (21)$$

We know already that  $a^2$  is even: by definition, it is a multiple of 2. But now, it is easy to prove that if a number, say,  $c$ , is odd, then  $c^2$  is odd as well. Indeed  $c = 2k + 1$ , by definition, and  $c^2 = 4k^2 + 4k + 1$  is also odd (since it is of the type  $2\kappa + 1$ , namely, with  $\kappa = 2k^2 + 2k$ ). Therefore, since  $a^2$  is even, also  $a$  is even (otherwise, if it'd be odd,  $a^2$  would be odd, as we have just shown). If  $a$  is even, then there exist  $d$  such that  $a = 2d$  and, of course,  $a^2 = 4d^2$ . Inserting this back into Eq. (21), we find  $4d^2 = 2b^2$  or, simplifying,

$$b^2 = 2d^2. \quad (22)$$

But this should look familiar... cf. Eq. (21) again. Going back through the same logic, this tells us that  $b$  is even. But how can  $a$  and  $b$  be both even, since  $\gcd(a, b) = 1$ ? This is a contradiction, meaning that the hypothesis of Eq. (19) is not possible. There does not exist such numbers.  $\sqrt{2}$  is not the ratio of two integers. But this number has to exist. From Pythagoras' theorem, this is the length of the diagonal of a square whose side is unity. How could such a distance not exist?

We have to introduce another set of such numbers that cannot be written as  $p/q$ . We call it the set of *irrational* numbers, since they arise precisely as a failure of existing in this form. The set of rational and irrational numbers, we call the set of *real numbers*, and write it  $\mathbb{R}$ . This is our collection of numbers so far:

$$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R} \quad (23)$$

Note that each set (collection) is included in the next one. We write  $\mathbb{N} \subset \mathbb{Z}$  to say that all positive integers are in the set of integers. We use instead the symbol  $\in$  to refer to a particular member of the set, e.g.,  $n \in \mathbb{N}$  means  $n$  is a natural number (positive). Literally, it states that  $n$  is in  $\mathbb{N}$ .

We will proceed to discover amazing and incredible properties about these collections of objects. We shall soon see, for instance, that there are much more irrational than rational numbers.

## Exercises

### Sexy primes

Twin primes are of the type  $(p, p + 2)$ . The twin-primes conjecture also extends to “cousin primes” (of the type  $(p, p + 4)$ ) as well as “sexy primes” (of the type  $(p, p + 6)$ ). They are called “sexy” because “sex” in latin means six. Find all the sexy primes below 30 (there should be 5 couples; you’ll find several triplets and even one quadruplet and one quintuplet).

### Goldbach’s conjecture

Show that the Goldbach’s conjecture does not hold for all integers (it is stated only for even integers). For even integers, check it for as high as you can go (the record is  $4 \times 10^{18}$ ). There is a weak Goldbach’s conjecture that postulates odd numbers (larger than 5) to exist as the sum of three primes. Show that if Goldbach’s conjecture (for even numbers) holds, the weak version (for odd ones) follows.

### Associativity and commutativity

We have “proven” that multiplication is commutative, cf. Eq. (3). Can you prove that it is *associative*, i.e.,  $a(bc) = (ab)c$ ? Since this is the case, we write, with no risk of ambiguity, simply  $abc$ . While you’re at it, prove associativity and commutativity of addition too. Can you prove the distributivity of multiplication over addition?

$$a(b + c) = ab + ac. \quad (24)$$

How about subtraction and division?

### Additive combinatorics

Consider the sets  $A = \{1, 2, 3, 4, 5\}$ . Build the sumset  $A + A$  and the product set  $A \cdot B$ . What does the Erdős–Szemerédi theorem tells you in this case? What do we get if we define  $nA$  as  $\underbrace{A + \cdots + A}_{n \text{ times}}$  but with  $A$  now a set?



## Problems

### Hexadecimal basis

Our basis is 10 because we have ten fingers. We do not always use the basis 10. For instance, for time keeping, we use the basis of Amerindians, which is base 60 (there are 60 minutes in one hour). Do you see how to count in, say, basis 16? (this is called **hexadecimal**, and is useful in computer science). Since we lack numbers after 10, we use letters, so the first hexadecimal numbers are (16 of them):

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F$$

So the next number, back to 10, is 16 in decimals. Then 11, 12, ..., 19 (by this time we should be at... 25).

- What is the next number after 20? (equivalently, how much is the decimal 26 in hexadecimal?)
- How much is  $AF$  (hexadecimal) in decimals?
- Can you see the rule to convert numbers between two bases?

### Group theory

Look up in some book or online for the definition of a group, which will come in terms of “axioms” (or rules) that regulate how elements of the group do behave. If the number of elements is finite, we speak of a finite group. Show that  $\mathbb{Z}$  with the addition as its operation is a group. Can you think of a finite group? Can you think of one which does not consist of numbers?



## Lecture 2: Arithmetics and Algebra

We have seen in the previous Lecture how Mathematics is concerned about relationships between Platonic objects. So far we have dealt with the simplest and most common objects: numbers. The science of adding and multiplying such objects is called *arithmetic*. This started with the Platonic school who obtained important results such as the fundamental theorem of Arithmetics, which states that every integer can be written in one and only way as a product of primes. This is a result we still use on a daily basis (to simplify fractions to their irreducible form). The art of arithmetic was then developed by Indian scholars, in ca. the 5th century. The climax of their technique is probably the decimal notation, whereby the same numbers

$$0, 1, 2, 3, 4, 5, 7, 8 \text{ and } 9 \quad (25)$$

are used to represent all possible quantities. The numbers themselves, incidentally, are from the Arabs and this is muslim scholars who brought the decimal system to the occident in ca. the 14th century. This choice of ten is due to the fact that we have 10 fingers and started counting by associating our fingers on the one hand (no pun intended) with the counted objects on the other hand. Instead, romans (for instance) would write I, II, III, IIII, etc., at which point the introduction of new letters is necessary (even Phoenicians who stucked to that principle up to 9 introduced new symbols for larger numbers), with V for five, X for ten, L for fifty, C for a hundred, D for five-hundred and M for a thousand, and a handy convention of subtracting a smaller number when on the left of a larger one, so that IV is used insted of IIII. Roman numbers are still used nowadays, for instance clocks often read:

$$I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII \quad (26)$$

But this is clearly unpractical for large numbers.<sup>8</sup> It shows, incidentally, the sort of magnitudes which Romans, that formed a mighty empire, had to deal with. In their system, the largest number that can be written is:

$$MMMCMXCIX \quad (27)$$

<sup>8</sup> The book held by the statue of liberty reads IV July MDCCLXXVI. Which date is that? (optionally, what date is that?)

that is, 3 999 (if sticking to rules that make each number unique, e.g., we would proscribe IMMMM for using four letters to produce a number)

Indians, on the other hand, would simply shift the numbers with zeros to keep track of so-called *orders of magnitudes*, which can then be described exhaustively with the same numbers (25). This is known as a “positional system”. Namely, after 9, we use two numbers to get to the same number, so starting from the first numbers 0 and 1, we make 10. Up to 99, we still need only two numbers only, then we need three, starting from 100, up to the next (3rd) order of magnitude 1000. The 5th order of magnitude gathers all numbers between 10000 (smallest number with 5 digits) and 99999 (biggest number with 5 digits).

We can also go in the other direction, i.e., scaling down rather than up, that brings us from 1 down to 0.1, 0.01, 0.001, etc. Here it is handy to use the power notation to count the numbers of zeros if there are so many than counting them brings us back to the Roman/Phoenician ordeal. We write

$$10^n = 1 \underbrace{0 \cdots 0}_{n \text{ zeros}} . \quad (28)$$

To go in the other direction, we use the – sign:

$$10^{-n} = 0. \underbrace{0 \cdots 0}_{n \text{ zeros}} 1 . \quad (29)$$

This is known as the “scientific notation”. That is a tidy way to keep track of the orders of magnitude: it’s the number  $n$  itself! How many orders of magnitudes are there? In the platonic universe, there is no limit. In the physical universe, if we speak in terms of lengths, for instance, today’s knowledge is that the largest distance is the size of the visible universe, and that is estimated at about  $10^{27}$  meters, while the smallest distance is Planck’s length  $\ell_P$  and is about  $10^{-35}$  meters, so that makes 63 (including  $10^0$ ) orders of magnitudes!

A book by Kees Boeke, “Cosmic View”, explores through 40 changes of orders of magnitude, the infinitely small and infinitely large on both sides of a girl sitting on a chair with a cat. Several movies have since been released to animate this idea in a smooth progression, starting with the Eames couple of architect <https://www.youtube.com/watch?v=0fKBhvDjuy0>, who in turn inspired several others (such as “Cosmic Voyage” in 1996 <https://www.youtube.com/watch?v=44cv416bKP4> or “Cosmic Eye” <https://www.youtube.com/watch?v=8Are9dDbw24> in 2012)<sup>9</sup>. Asimov also wrote a book describing orders of magnitudes for other quantities besides distance (“The Measure of the Universe”).

<sup>9</sup> How many orders of magnitudes have been added in the successive remakes of the 1977 movie?

Like Romans, this notation is quite revealing of the sorts of magnitudes we need to work with. We are commonly dealing with numbers such as  $10^{10}$  nowadays (from a Feynman's lecture):

There are  $10^{11}$  stars in the galaxy. That used to be a huge number. But it's only a hundred billion. It's less than the national deficit! We used to call them astronomical numbers. Now we should call them economical numbers.

And we are indeed able to account for a large numbers of objects. For instance it is estimated that there are  $\approx 3 \times 10^{12}$  (three trillion) trees on Earth, with more trees in Russia alone than stars in the galaxy. Avogadro's number  $\approx 6 \times 10^{23}$  defines the unit for the "amount of substance" by counting how many atoms are there in 12 g of carbon. Archimedes estimated that it would require  $10^{63}$  grains of sand to fill-up the universe. A modern estimate is of about  $10^{80}$  atoms in the universe<sup>10</sup>

The number  $10^{100}$ , known as a "Googol", is sometimes used as a representative of a large number, to which one can compare other gigantic quantities, that appear to be much smaller<sup>11</sup> The Platonic Universe, more than the Physical Universe, can readily give rise to really large numbers. For instance it is estimated that there are  $10^{120}$  possible games of chess (Shannon number, but only  $10^{50}$  positions).

The scientific notation thus allows us to represent fairly easily all possible numbers, and with arbitrary precision ( $\pi$  for instance is known up to 31 415 926 535 897 digits, or 31 trillions).

While base 10 is the most common, other bases can be used, such as the sexagesimal (base 60) for time counting again (one hour equal sixty minutes). In fact, other bases can be more practical than base 10, which has been almost universally by all cultures and civilisation. The duodecimal system, or base 12, for instance, gives as "successors" of nine (for 9) two numbers still in the same order of magnitudes, A and B, and then 10, which is *not* "ten" but "one-zero" as the successor of B, i.e.,

$$B_{(12)} = 11_{(10)} \quad (30)$$

$$10_{(12)} = 12_{(10)} \quad (31)$$

since  $10_{(12)} = 1 \times 12^1 + 0 \times 12^0$ . In this way we can easily convert from duodecimal to decimal, e.g.,

$$123_{(12)} = 1 \times 12^2 + 2 \times 12^1 + 3 \times 12^0 = 171, \quad (32)$$

or, using the "new numbers":

$$AB2_{(12)} = 10 \times 12^2 + 11 \times 12^1 + 2 \times 12^0 = 1574. \quad (33)$$

<sup>10</sup> Eddington believed there are 15 747 724 136 275 002 577 605 653 961 181 555 468 044 717 914 527 116 709 366 231 425 076 185 631 031 296 protons in the universe and the same number of electrons. This became known as the Eddington number. Give a power of ten approximation of this number.

<sup>11</sup> What percentage of a Googol is  $10^{90}$ ?

The other way around requires Euclidean division, which retains the remainder, since we need to find how the initial quantity fits in the various orders and in which quantities:

$$a = bq + r \quad \text{with } 0 \leq r < b. \quad (34)$$

Indeed, since

$$12^3 = 1728 \quad (35a)$$

$$12^2 = 144 \quad (35b)$$

$$12^1 = 12 \quad (35c)$$

$$12^0 = 1 \quad (35d)$$

we find, e.g.,  $2020 = 1 \times 1728 + 292 = 1 \times 12^3 + 2 \times 144 + 4 = 1 \times 12^3 + 2 \times 12^2 + 0 \times 12 + 4 \times 12^0$  so that

$$2020_{(10)} = 1204_{(12)}. \quad (36)$$

Divisions work similarly with negative powers, e.g.,

$$\left(\frac{1}{2}\right)_{(12)} = 6 \times 12^{-1} = 0.6_{(12)} \quad (37)$$

$$\left(\frac{1}{3}\right)_{(12)} = 4 \times 12^{-1} = 0.4_{(12)} \quad (38)$$

$$\left(\frac{1}{4}\right)_{(12)} = 3 \times 12^{-1} = 0.3_{(12)} \quad (39)$$

which shows that, thanks to 12 having so many prime divisor, fractions in the duodecimal systems are actually simpler (compare to  $1/3 = 0.33333 \dots$  in the decimal system). Now of course not all fractions are that simple. Instead of  $1/5 = 0.2$  we find for  $\left(\frac{1}{5}\right)_{(12)}$ , since

$$\frac{12}{5} = 2.4 \implies \frac{1}{5} = 2 \times 12^{-1} + 0.4 \times 12^{-1} \quad (40)$$

for the next power we carry on:

$$\frac{1}{5} = 2 \times 12^{-1} + 0.4 \times 12 \times 12^{-2} \quad (41a)$$

$$= 2 \times 12^{-1} + 4 \times 12^{-2} + 0.8 \times 12^{-2} \quad (41b)$$

$$= 2 \times 12^{-1} + 4 \times 12^{-2} + 0.8 \times 12 \times 12^{-3} \quad (41c)$$

$$= 2 \times 12^{-1} + 4 \times 12^{-2} + 9 \times 12^{-3} + 0.6 \times 12^{-3} \quad (41d)$$

so the procedure is simple: we multiply the remainder by 12, keep the whole part that makes one digit more, and reiterate. In this way, we find:

$$\left(\frac{1}{5}\right)_{(12)} = 0.24972497249 \dots_{(12)} \quad (42)$$

repeating. Similarly

$$\left(\frac{1}{6}\right)_{(12)} = 0.2_{(12)} \quad (43)$$

and we also give the next one as it provides a nondecimal digit

$$\left(\frac{1}{7}\right)_{(12)} = (0.186A35186A3\dots)_{(12)} \quad (44)$$

Note that  $(1/A)_{(12)}$  is a tenth.  $(1/10)_{(12)}$  is a twelfth.<sup>12</sup> Rationals in one basis remain rationals in any other basis, meaning that they have ultimately repeating patterns. Irrationals also remain irrationals.<sup>13</sup> The principles apply equally from any basis to the other.<sup>14</sup>

Muslim scholars eventually noted that the rules that apply to numbers also have an internal consistency. So they started to abstract the numbers away. Namely, they replaced them by letters. Pythagoras' theorem, for instance, possibly the most famous identity in Mathematics, states that if a right-angle triangle has lengths  $a$ ,  $b$  and (hypotenuse)  $c$ , then:

$$a^2 + b^2 = c^2. \quad (45)$$

You can check for instance that a triangle with sides 3 and 4 has hypotenuse 5 (it's not always as round as that).<sup>15</sup>

The point of Eq. (45) written in this form is that it allows us to think in a more generic way than if we pin the quantities to specific values. Note that we still use numbers in this equation. One could generalize this even more, for instance, consider the equality:

$$a^n + b^n = c^n \quad (46)$$

for some integer  $n$ . This simple looking formula (which reduces to Pythagoras' theorem for  $n = 2$  and relates the lengths of right-angle triangles) happens to be one of the most celebrated brain-twister of Mathematics, the so-called Fermat's last theorem, since Fermat wrote in margin of his copy of *Arithmetica* that he had an elegant proof that no integers exist that satisfy this equation for integers  $n > 2$ , but, so Fermat wrote, the proof is too long to write in this margin. It is only in 1995 that a correct proof was established (by Andrew Wiles) after generations of people maddeningly trying to reproduce Fermat's insight (it is believed he did not have such an elegant proof and was himself mistaken).

The power of algebra is that it allows to think in terms of the general concepts rather than particular cases. For instance, if we ask the question, what is the length of a right-angle triangle with hypotenuse  $c$  and with the other side of length  $a$ , from Eq. (45), we find:

$$b = \sqrt{c^2 - a^2} \quad (47)$$

<sup>12</sup> Give the duodecimal expansion of  $1/A$ ,  $1/B$  and  $1/11$ .

<sup>13</sup> What are the duodecimal expansion of  $\pi$ ? Is  $1.4B_{(12)}$  a better two-digits approximation of  $\sqrt{2}$  than its decimal version?

<sup>14</sup> What is the base-13 ABC number in undecimal (base-11)?

<sup>15</sup> Provide other examples of Pythagoras' theorem which work with only integers (these are known as "Pythagorean triples" and are the main object of study of a Babylonian tablet from c. 1800BC, known as the Plimpton 322).

which we obtained by adding  $-a^2$  on both sides of the equation, then taking the square root. Such manipulations, of balancing terms on both sides of the equation, led to the denomination of “algebra”, or “al-jabr”, meaning, the “reunion of broken parts”.

Even when we were doing arithmetic, previously, we already borrowed concepts of algebra. For instance, rationals have been introduced as numbers of the form

$$\frac{p}{q} \quad (48)$$

with  $p, q \in \mathbb{Z}$  ( $q \neq 0$ ). Let us do some arithmetic of letters (that is, some algebra).

We have already convinced ourselves (think about it and do it if it's not yet the case) about the following properties of addition

- Associativity:  $(a + b) + c = a + (b + c)$  (so we can write  $a + b + c$ ).
- Commutativity:  $a + b = b + a$ .

and multiplication:

- Associativity:  $(ab)c = a(bc)$  (so we can write  $abc$ ).
- Commutativity:  $ab = ba$ .

The two together work according to the rule of distributivity:

$$a(b + c) = ab + ac. \quad (49)$$

From this, we can work out the rest. For instance:

$$\begin{aligned} (a + b)^2 &= (a + b)(a + b) = a(a + b) + b(a + b) \\ &= a^2 + ab + ba + b^2 = a^2 + 2ab + b^2 \end{aligned} \quad (50)$$

You must know this one on the tip of your hands... but until you do, work it out. You'll learn it before you get tired of deriving it.

Next is the cubic expansion, it works the same:

$$(a + b)^3 = (a + b)(a + b)(a + b) \quad (51)$$

we could work it out like this or see that this is  $(a + b)^2(a + b)$  (order doesn't matter, associativity). And we know the first one, so that's  $(a^2 + 2ab + b^2)(a + b) = a^3 + 2a^2b + ab^2 + a^2b + 2ab^2 + b^3$  which we can simplify further by grouping terms:

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3. \quad (52)$$

So on and so forth. Actually at this stage (you can practice with a few more examples)<sup>16</sup> we would want the general case rather than

<sup>16</sup> Expand by brute-force calculation  $(a + b)^4$ , either as  $(a + b)(a + b)(a + b)(a + b)$  or  $(a + b)^3(a + b)$  or  $(a + b)^2(a + b)^2$ ; which sounds the easiest? You can try  $(a + b)^5$  too.



particular ones, and go full-algebraic:

$$(a+b)^n = \underbrace{(a+b) \cdots (a+b)}_{n \text{ times}}. \quad (53)$$

Let's think about this expression. What terms can we get? We have to pick up one from each parenthesis, and we make the product. So clearly the generic term is of the form:

$$a^k b^{n-k} \quad (54)$$

since  $k + n - k = n$  the number of parenthesis that make up the products. By commutativity, some of these terms will appear several times, e.g.

$$aaabba \quad (55)$$

in the expansion of  $(a+b)^6$  is the same as:

$$baabaa \quad (56)$$

so there'll be at least two  $a^4 b^2$ . How many are there finally? This is a problem in combinatorics, which is a topic for another lecture. For now we can simply introduce as a notational device, the number of ways that one can choose  $k$  objects out of  $n$ , as a so-called *binomial coefficient* which we write as  $\binom{n}{k}$  and that be computed according to some formula, which we will provide later.<sup>17</sup> With this notation, we are able to get the complete expression:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \cdots \\ \cdots + \binom{n}{n-2} a^2 b^{n-2} + \binom{n}{n-1} a b^{n-1} + \binom{n}{0} b^n. \quad (57)$$

It's good but a bit bulky. There is however a beautiful symmetry and repeating pattern, so we can simplify considerably the expression, by using the  $\Sigma$  notation for summation:

$$(a+b)^n = \sum_{k=0}^n a^k b^{n-k}. \quad (58)$$

Do you see what we did here? We used  $k$  as a so-called *dummy index*. It's not something that actually shows up in the final, expanded result, it's just an intermediate, handy variable, to capture and synthesize the generic pattern. The range of values this takes is given by the boundaries below and above  $\Sigma$ . The rules are intuitive.<sup>18</sup>

At this stage it should be straightforward to see that, for  $n, m \in \mathbb{N}$ , we have:<sup>19</sup>

$$a^n a^m = a^{n+m}. \quad (60)$$

<sup>17</sup> For now you can obtain them by explicit calculation, e.g., from question 16 above, what are  $\binom{4}{3}$  and  $\binom{5}{2}$ ?

<sup>18</sup> Compute the following sums:

$$\sum_{i=-5}^5 i^2, \quad \sum_{i=-5}^5 2^i, \quad \sum_{j=0}^3 \sum_{0 < i < j} ij.$$

Show that (assume finite sums only):

$$\sum_i \sum_j a_i b_j = \sum_i a_i \sum_j b_j. \quad (59)$$

Which algebraic phenomenon is going on there?

<sup>19</sup> Prove it.

It's also easy to check that:<sup>20</sup>

$$(a^n)^m = a^{nm}. \quad (61)$$

Now let's carry on with a big idea of Mathematics (in general) and Algebra (in particular), which is to extend concepts beyond their realm of application. Namely, what could be (if it makes sense at all) the number:

$$a^n, n \in \mathbb{Z} \quad (62)$$

or to pinpoint where the difficulty lies, let's ask, what is  $a^{-n}$  with  $n \in \mathbb{N}$  (negative power!) Going down the route of multiplying  $a$  by itself minus  $n$  times doesn't seem to bring us anywhere. But let's assume that Eq. (60) holds... then we have:

$$a^{-n} a^n = a^{-n+n} = a^0 = 1 \quad (63)$$

so that "balancing the equation" by bringing  $a^n$  on the other side through division, we get:

$$a^{-n} = \frac{1}{a^n} \quad (64)$$

So negative powers are powers of the inverse! This proves, or justify, Eq. (29) that was so far really another conventional notation.

Let's carry on and ask, what is  $a^n$  for  $n \in \mathbb{Q}$ , that is, really, what is

$$a^{\frac{p}{q}}, \quad p, q \in \mathbb{N} \quad (65)$$

with  $q \neq 0$ . Let's assume  $p = 1$  for now, as it's then just to deal with  $a^{\frac{1}{q}}$  instead. Still using Eq. (60), we find that:

$$\underbrace{a^{\frac{1}{q}} a^{\frac{1}{q}} \cdots a^{\frac{1}{q}}}_{q \text{ times}} = a^{\underbrace{\frac{1}{q} + \cdots + \frac{1}{q}}_{q \text{ times}}} = a \quad (66)$$

So  $a^{\frac{1}{q}}$  is the number we need to multiply to itself  $q$  times to get  $a$ . This is known as the  $q$ -th root of  $a$ , and we write it as:

$$\sqrt[q]{a} \equiv a^{\frac{1}{q}}. \quad (67)$$

Note that the lhs is a notation. We can carry on and on and on, for instance, what is

$$a^r, \quad r \in \mathbb{R} - \mathbb{Q} \quad (68)$$

that is, an irrational power? One can always get arbitrarily close to an irrational with a sequence of ever more accurate rationals. For

<sup>20</sup> Prove it.

instance

$$\begin{aligned}
 \pi &\approx 3.14 \\
 &\approx 3.141 \\
 &\approx 3.1415 \\
 &\approx 3.14159 \\
 &\dots \\
 &\approx 3.141592653589793238462643
 \end{aligned} \tag{69}$$

The last one is  $3141592653589793238462643/10^{24}$  (which is also the irreducible form as the gcd of numerator and denominator is 1). So, if you wonder what

$$2^\pi \tag{70}$$

is, then in good approximation you can say that it's close to:

$$\frac{3141592653589793238462643}{2 \cdot 10^{24}} \approx \frac{10^{24}}{\sqrt{2^{3141592653589793238462643}}}. \tag{71}$$

Not something very nice to do but this sort of stuff we can pass over to a computer and it's clear to us what this is. If we need more accuracy, we can chuck in more digits of  $\pi$ .

So far we did all this with integer numbers. Let us explore how this extends to rationals. Let us start with the simplest operation, to multiply fractions:

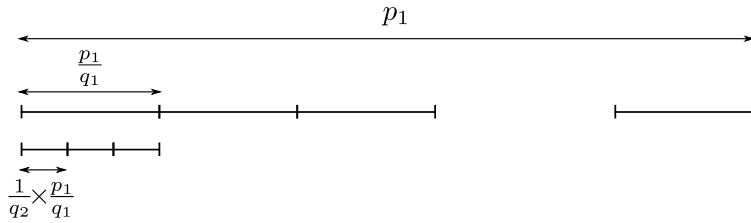
$$\frac{p_1}{q_1} \times \frac{p_2}{q_2}. \tag{72}$$

Note that we used subscripts rather than different letters, because like this we can emphasize the  $p$ s as the numerators and the  $q$ s as the denominators. It'll be useful if we need to turn to  $\Sigma$  notation as these subscripts can become dummy indices.

Remember that  $p_1/q_1$  is the number that needs to be added to itself  $q_1$  times to make  $p_1$ . So  $p_2 \times \frac{p_1}{q_2}$  is scaling up this quantity, and adding  $q_2$  times these new bits of  $p_2 \times \frac{p_1}{q_2}$ , we will then get  $p_1 p_2$  (the rescaled final total). So we have shown that:

$$p_2 \times \frac{p_1}{q_1} = \frac{p_1 p_2}{q_2}. \tag{73}$$

Now let us consider  $\frac{1}{q_2} \times \frac{p_1}{q_1}$ . This is the quantity that we have to add  $q_2$  times to itself so that it yields  $\frac{p_1}{q_1}$ . This little quantity rescaled in this way, if we add it to itself  $q_1 q_2$  times, will thus give  $p_1$  as sketched below:



this means, by definition of the fraction, that

$$\frac{1}{q_2} \times \frac{p_1}{q_1} = \frac{p_1}{q_1 q_2}. \quad (74)$$

If we bring Eqs. (73) and (74) together, also using associativity and commutativity, we have just demonstrated then that:

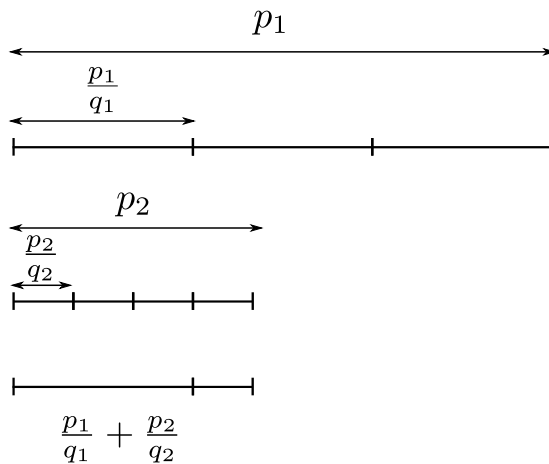
$$\frac{p_1}{q_1} \times \frac{p_2}{q_2} = \frac{p_1 p_2}{q_1 q_2}. \quad (75)$$

adding to itself  $p_2$  times the number which needs to be added to itself  $q_2$  times to make  $p_1$ . Clearly, this number is the one that, added to itself  $q_2$  times, make  $p_1 p_2$ , that is:

Adding two fractions gives a more complicated process for an equally simple (or not?) reasoning:

$$\frac{p_1}{q_1} + \frac{p_2}{q_2}. \quad (76)$$

We know that if we add  $\frac{p_1}{q_1}$  to itself  $q_1$  times we'll get  $p_1$  and that if we add  $\frac{p_2}{q_2}$  to itself  $q_2$  times we'll get  $p_2$ . Adding these two numbers together a given number of time gives us what, exactly?



If we add it to itself  $q_1$  times we know what to do with the  $p_1/q_1$  bit but not with the other one, and vice-versa. So let's be clever and generous, and add  $p_1/q_1$  to itself  $q_1 q_2$  times, then we will get  $p_1$  (that's the  $q_1$  times adding-up) times  $q_2$  (because we do this  $q_2$  times).

The same with  $p_2/q_2$ , we also add it to itself  $q_1q_2$  times, and we get  $p_2 \times q_1$ . Together, they sum to  $p_1q_2 + p_2q_1$  and we got this by adding the same quantity  $q_1q_2$  times in both cases. Therefore, we have just demonstrated that:

$$\frac{p_1}{q_1} + \frac{p_2}{q_2} = \frac{p_1q_2 + p_2q_1}{q_1q_2}. \quad (77)$$

This is how fractions add up together. We have explained why, in a way that a Physicist would understand it. You can probably find other ways to get this result, that is just one of them. Most people would not even care, and would take Eqs. (75) and (77) as rules, or axioms, to be obeyed blindly. That's fine too, although it helps us to understand where things are coming from.

We conclude with a very important result of Algebra. A product is equal to zero *if and only if* at least one of the terms is equal to zero. Can you convince yourself that this is the case? Once you agree to that, it then becomes easy to solve the following, so-called *quadratic equation*:

$$ax^2 + bx + c = 0 \quad (78)$$

where we use, following Descartes, beginning-of-alphabet letters  $a, b, c$  for what is supposed to be a constant or any type of known variable, and end-of-alphabet letters  $x, y, z$  for unknown variables (here there is only one unknown). The trick is to transform Eq. (78) into a product through the so-called *completion of the square*. Namely, we make  $x$  disappear by making it a byproduct of the squaring of  $x$  (namely, the double product). First, we can simplify  $a$  in the leading exponent (since  $a \neq 0$ ; if it is, the solution is trivial,  $x = -c/b$ ):

$$ax^2 + bx + c = x^2 + \frac{b}{a}x + \frac{c}{a} = 0 \quad (79)$$

Then we look how to make appear the  $x$  terms:

$$\left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} - \frac{b^2}{4a^2} = 0 \quad (80)$$

Simplifying the expression for the new terms (those not linked to  $x$ ):

$$\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2} = 0 \quad (81)$$

This starts to look like a difference of two squares, which is easy to bring into a product form, since  $p^2 - q^2 = (p - q)(p + q)$ . If only  $b^2 - 4ac$  was a square... but  $a, b$  and  $c$  are merely numbers, and so is  $b^2 - 4ac$ . Now, any positive number is a square (the square of its square root). So if we introduce

$$\Delta \equiv b^2 - 4ac \quad (82)$$

(this is called the “*discriminant*”), and if this  $\Delta \geq 0$ , then we definitely have a difference of two squares:

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{\Delta}}{2a}\right)^2 = 0 \quad (83)$$

$$\left(x + \frac{b}{2a} - \frac{\sqrt{\Delta}}{2a}\right) \left(x + \frac{b}{2a} + \frac{\sqrt{\Delta}}{2a}\right) = 0. \quad (84)$$

Fortunately, they have the same denominator, so from the rule above (a product is equal to zero if one of its member is equal to zero), we have that the quadratic equation has two possibilities to hold, i.e., it has two solutions:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (85)$$

This is a very important result, which one needs to know, even without coming to university. That is how important it is. Note that to make it work, we had to involve a difference of two squares in Eq. (81), since we don’t know (at least, not yet) what to do with a sum of two squares  $p^2 + q^2$ . If we would represent the equation (78) in a graph, with  $x$  on the horizontal axis and  $ax^2 + bx + c$  on the vertical axis, we would find the solutions to  $ax^2 + bx + c = 0$  as the intersection of the curve (it’s a parabola) with the horizontal axis. When such a curve does not intersect the horizontal axis, because its lowest point is above 0 or on the opposite its highest point is below 0, then there is no solution. In this case, the discriminant is negative. We will see in the next Lecture the so-called *fundamental theorem of algebra* that shows that this is not, in fact, the case.

## Problems

### Sigma notation

Let us assume a real number  $r \in \mathbb{R}$  with numeral decomposition

$$\cdots d_3 d_2 d_1 d_0 . d_{-1} d_{-2} d_{-3} \cdots \quad (86)$$

for instance, if  $r = \pi$ , we have:

$$d_2 = 0 \quad (87a)$$

$$d_1 = 0 \quad (87b)$$

$$d_0 = 3 \quad (87c)$$

$$d_{-1} = 1 \quad (87d)$$

$$d_{-2} = 4 \quad (87e)$$

Write expression (86) with the  $\sum$  notation. Provide the general expression for any base (e.g., duodecimal).

### Combinatorics

Show that there are  $n!$  to arrange  $n$  objects; this is called a *permutation* and  $n! \equiv 1 \times 2 \times 3 \times \cdots \times n$  is called a *factorial*. For instance, there are 6 ways to permute  $\{a, b, c\}$ , namely:

$$(a, b, c) \quad (a, c, b) \quad (b, a, c) \quad (b, c, a) \quad (c, a, b) \quad (c, b, a). \quad (88)$$

Note that we use curly brackets for sets, where the order of the elements doesn't matter, and parenthesis for "lists" (or "vectors" as we'd call that in Physics), where the order matters.

Check that the number of choosing  $k$  objects in a collection of  $n$  is:

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!} \quad (89)$$

(the left-hand side is a notation, the right-hand side is a result.)

Compute the following table (we take that  $0! = 1$ ):

$$\begin{array}{cccc} \binom{0}{0} & & & \\ \binom{1}{0} & \binom{1}{1} & & \\ \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & \\ \binom{3}{0} & \binom{3}{1} & \binom{3}{2} & \binom{3}{3} \\ \binom{4}{0} & \binom{4}{1} & \binom{4}{2} & \binom{4}{3} \end{array} \quad (90)$$

Do you see the relationship between these numbers? There is a way to get the numbers from any row from the row above it, which one? Write it as a formula. This is called *Pascal's triangle*, and is very useful to compute binomial expansions. Add a few more lines to this triangle using the easy (row-above) rule to compute:

$$(a + b)^7. \quad (91)$$

### Irrational powers with your bare hands

Powers of 2 are easy, and they are very useful (thus familiar) in computer science. Compute  $2^n$  for  $0 \leq n \leq 22$  (this can be done even without a computer).

Check that  $\pi \approx 22/7$ , so we can estimate  $2^\pi$  as which number gives  $2^{22}$  when multiplied to itself 7 times. Observe that 2 multiplied to itself 21 times is close (by a factor 2, which, in Physics, is usually okay) to  $2^{22}$ , and  $21 = 3 \times 7$ . Use these facts to give a rough estimate of  $2^\pi$  without using a calculator. The simplest result is approximating  $\pi$  to 3, so try to give a better estimate (for instance a lower and an upper bound) Compare your best guess with the exact result (obtained with a calculator)

*Graphical and algebraic method to solve the quadratic equation*

Solve the equation

$$2x^2 - 7x + 3 = 0 \quad (92)$$

first graphically (plotting the curve corresponding to this equation and finding the intersect with the  $x$  axis), then algebraically (applying the formula (85)).

*Hexadecimal basis*

Our basis is 10 because we have ten fingers. We do not always use the basis 10, although it seems to. For instance, for time keeping, we use the basis of Amerindians, which is base 60 (there are 60 minutes in one hour). We still use the Arab numerals (25) though instead of adding other symbols (which would be inconvenient as we would need fifty more). There are still traces of the underlying basis, for instance the fact that we wouldn't use the time 16:61 (that's 17:01).

There are two bases which are commonly used in computer science, and since we use computers a lot in Physics, these are basis that we may encounter time to time.

One is the basis 16? (this is called **hexadecimal**, the other is even more important, the basis 2 (**binary**)).

For the hexadecimal, we use letters for the quantities after 10, so the first hexadecimal numbers are (16 of them):

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F \quad (93)$$

So like in the duodecimal case, the next number after exhausting the basis, is 10, which is 16 in base 10. Then 11, 12, ..., 19 (by this time we should be at... 25 in base 10).

Here is how part of the current text is encoded by the computer:

```
00007880: 7220 696e 7374 616e 6365 2074 6865 2066  r instance the f
00007890: 6163 7420 7468 6174 2077 6520 776f 756c  act that we woul
000078a0: 646e 2774 2075 7365 2074 6865 2074 696d  dn't use the tim
000078b0: 6520 3136 3a36 3120 2874 6861 7427 7320  e 16:61 (that's
000078c0: 3137 3a30 3129 2e0a 0a54 6865 7265 2061  17:01)...There
```

Explain this structure.



## Lecture 3: Logic, proofs and notations

The previous lecture introduced the notion of a Platonic universe populated with abstract objects, and its own rules. In Physics, the rules that dictate the behaviour of the Universe are known as “Physical laws”, and are, for instance, the conservation of energy, special relativity, quantum mechanics, etc. Some are known to be approximations, like Newton’s equations. Others are believed to be exact, like conservation rules. It is not yet known, however, what set of fundamental laws can be used to derived everything else. It is not even known if this is possible to have one theory for everything, as two main branches of modern physics, relativity and quantum mechanics, appear to be incompatible. In the Platonic universe, on the other hand, one starts from such a small set of fundamental laws, which are called “*axioms*”, and from which one derives everything else. The rules in the Platonic universe often consist in stating whether a property or statement regarding its objects, is true, or false. Such a statement is called a “*theorem*”.

Mathematicians introduce special notations to describe both the relations between mathematical objects and the nature of the statements, which we will call, by the way, “relations” (or “propositions”, “statements”, etc.) We are going to introduce some of these symbols now, starting with the negation:

$$\neg \tag{94}$$

If we have a relation  $R$ , then  $\neg R$  is the negation of this relation. It means that if  $R$  is true then  $\neg R$  is false, and vice-versa.

The main two ways to bring relations together is to look at the logical operations

*and*    *and*    *or*

which we will write as

$$\wedge \quad \text{and} \quad \vee . \tag{95}$$

We also call them “conjunction” and “disjunction”, for their connection to the intersection  $\cap$  and union  $\cup$  in set theory, to which we will come back later.

In this way we can start to make mathematical statements. For instance:

$$R \vee (\neg R) \quad (96)$$

is always true, since whether  $R$  is true or not, either it or its negation is true. A very important logical construct is so useful that we introduce a new notation for it. It reads:

$$S \vee (\neg R) \quad (97)$$

which we will abbreviate as:

$$R \implies S \quad (98)$$

and read it as “ $R$  implies  $S$ ” or “if  $R$  then  $S$ ” (implying “if  $R$  is true then  $S$  is true”). From the point of view of  $R$ , this states that  $R$  is a “*sufficient*” condition for  $S$  to be true. It means that *if*  $R$  is true, then  $S$  is also true, but if  $R$  is not true,  $S$  can or can not be true while  $R \implies S$  is still true. From the point of view of  $S$ , this states that  $S$  is a “*necessary*” condition for  $R$  to be true, meaning that  $R$  can be true *only if*  $S$  is true. For instance:

$$(p^2 \text{ is even}) \implies (p \text{ is even}) \quad (99)$$

is true. We demonstrated this in the last lecture. Note that the other way around (the so-called “converse”) in this particular case is also true, namely:

$$(p \text{ is even}) \implies (p^2 \text{ is even}).$$

This we did not need last time so we did not make any statement about it, but it is trivial to prove: if  $p$  is even then  $p = 2k$  for some  $k$ , and therefore  $p^2 = 4k^2 = 2(2k^2)$ , i.e., a multiple of two of some number ( $2k^2$ ). When an implication and its converse are both true, i.e., when

$$(R \implies S) \wedge (S \implies R) \quad (100)$$

is true, then we write

$$R \iff S \quad (101)$$

and call this an “*equivalence*”, and read it as “*if and only if*” (abbreviated as “*iff*”).

From the definition of Eq. (98), note that the implication itself is true if  $R$  is false, e.g.,

$$(4 \text{ is prime}) \implies 2 + 2 = 5 \quad (102)$$

is true. It is true because the premise “(4 is prime)” is false, so that regardless of whether  $2 + 2 = 5$  is true or not (it is not true in Mathematics, but this has been discussed a lot in other contexts, see Orwell’s 1984), then the implication itself is true.

The following distributivity between these logical operators hold:

$$(P \wedge (Q \vee R)) \Leftrightarrow ((P \wedge Q) \vee (P \wedge R)) \quad (103a)$$

$$(P \vee (Q \wedge R)) \Leftrightarrow ((P \vee Q) \wedge (P \vee R)) \quad (103b)$$

This can be demonstrated by going through a comprehensive listing of all possibilities, invoking the concept of a “truth table”, where each statement can take the value T (true) or F (false). In Boolean logic, one use the values 1 and 0 instead. The negation swaps the “state”

$P$	$\neg P$
T	F
F	T

The conjunction (AND  $\wedge$ ) and disjunction (OR  $\vee$ ) yield two very famous “logic tables”:<sup>21</sup>

$P$	$Q$	$P \wedge Q$	$P \vee Q$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F

<sup>21</sup> Write the others famous ones, namely NAND ( $\neg(P \wedge Q)$ ), NOR ( $\neg(P \vee Q)$ ), XOR or “exclusive-OR” ( $(P \vee Q) \wedge \neg(P \wedge Q)$ ) and XNOR ( $(P \wedge Q) \vee (\neg P \wedge \neg Q)$ ).

With three propositions as in Eq. (103), we have  $2^3 = 8$  possible cases, and we can work out the various terms of the equations in turn:

$P$	$Q$	$R$	$Q \vee R$	$P \wedge (Q \vee R)$	$P \wedge Q$	$P \wedge R$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	T
T	F	T	T	T	F	T	T
T	F	F	F	F	F	F	F
F	T	T	T	F	F	F	F
F	T	F	T	F	F	F	F
F	F	T	T	F	F	F	F
F	F	F	F	F	F	F	F

The fifth column is the lhs (left-hand-side) of Eq. (103) and the last (eighth) columns its rhs. Since they are everywhere equal, the equality is demonstrated.<sup>22</sup>

The simple relation (98) is one of the key mechanisms to prove, or demonstrate, new relations, or results, from known ones. From its definition (97) we can derive an important tool, namely, let us write  $S \vee (\neg R)$  as  $(\neg R) \vee S$  and  $(\neg R) \vee \neg(\neg S)$ , which is, with the arrow notation introduced:

$$\neg R \implies \neg S. \quad (104)$$

<sup>22</sup> Prove similarly Eq. (103b).

Since we have merely rewritten the initial statement, this is clearly an equivalence. The result is so important that we write it again in full:

$$(R \implies S) \iff (\neg R \implies \neg S). \quad (105)$$

This is actually how we demonstrated Eq. (99), by proving the (equivalent) statement:

$$(p \text{ is odd}) \implies (p^2 \text{ is odd}). \quad (106)$$

If it ever happens that  $R$  and  $\neg R$  are both true, then that is a “contradiction”, which would invalidate all of Mathematics. We hope there are no contradictions in Mathematics. We can also mention the “tautology”  $R \implies R$ .<sup>23</sup>

<sup>23</sup> Show that  $R \implies \neg R$  is true.

Other important logical concepts are quantifiers. We will use two of them, the Existential  $\exists$  and universal (for All)  $\forall$ , which condition relations or propositions on variables. Here are examples:

$$(\exists x \in \mathbb{R} - \mathbb{Q})(x^2 = 2) \quad (107)$$

which at a higher, abstract level, says that  $\sqrt{2}$  is irrational. A more literate reading is that there exists a number in the set of irrationals whose square is two. The statement  $(\exists x \in \mathbb{Q})(x^2 = 2)$  is false.

Now for a universal statement:

$$(\forall x \in \mathbb{R})(x^2 \geq 0) \quad (108)$$

which says that squares of real numbers are positive, or, with a more literal reading, for every real number, the square of this number is positive. We will see how challenging the rightmost proposition  $(x^2 \geq 0)$  led to a major discovery in calculus.

When there are several quantifiers, their order is usually important. For instance, the concept of “asymptote” is that of a function  $f$  that gets arbitrarily close to a value  $a$  without ever having taking this value. This is formulated as follows:<sup>24</sup>

$$(\forall \epsilon > 0)(\exists x \in \mathbb{R})(|f(x) - a| < \epsilon). \quad (109)$$

<sup>24</sup> Apply to the particular case  $f(x) = 1/x$ . What is the asymptote in this case?

The other ordering of quantifiers

$$(\exists x \in \mathbb{R})(\forall \epsilon > 0)(|f(x) - a| < \epsilon) \quad (110)$$

means the very different statement that  $f(x)$  actually takes the value  $a$ , since there exists a value for which the function is closer to  $a$  than any possible positive number, and this implies that  $f(x) = a$  for this  $x$ . Proof: if we would have  $f(x) = b$  with  $b \neq a$ , choosing  $\epsilon = |b - a|/2$  would bring us to  $1 < 1/2$  which is false, while the assumption is that Eq. (110) is true.<sup>25</sup>

<sup>25</sup> Still with  $f(x) = 1/x$ , find the values of  $a$  which make Eq. (110) true. Compare with Eq. (109).

In other cases, different orders provide different, and equally important, definitions, as is the case for continuity for instance. The following re-orderings, however, are allowed:

$$(\forall x)(\forall y)R \iff (\forall y)(\forall x)R \quad (111a)$$

$$(\exists x)(\exists y)R \iff (\exists y)(\exists x)R \quad (111b)$$

$$(\exists x)(\forall y)R \implies (\forall y)(\exists x)R \quad (111c)$$

Logical constructs can be negated using De Morgan's law, which distribute negation over disjunction and conjunction:

$$\neg(R \vee S) \iff (\neg R \wedge \neg S) \quad (112)$$

$$\neg(R \wedge S) \iff (\neg R \vee \neg S) \quad (113)$$

Let us use these results to demonstrate interesting logical negations:<sup>26</sup>

$$\neg(R \implies S) \iff \neg(S \vee \neg R) \iff \neg S \wedge R \quad (114)$$

The negation of quantifiers are  $\neg\forall \iff \exists$  and  $\neg\exists \iff \forall$ . In this way, the universal quantifier can actually be derived from the existential one, and vice-versa:<sup>27</sup>

$$(\forall x)R \iff \neg[(\exists x)(\neg R)], \quad (115)$$

$$(\exists x)R \iff \neg[(\forall x)(\neg R)]. \quad (116)$$

Therefore, if  $(\forall x)R$  is false, then  $(\exists x)(\neg R)$  is true. The negation of "all men are mortal" is thus "there exists immortal people". This is to be differentiated from "no men are mortal" or "all men are immortal".

In general, to negate a logical statement, we negate all quantifiers and the statements, keeping their order. For instance<sup>28</sup>

$$\neg[(\forall x \in A)(\exists y \in B)P(x, y)] \iff (\exists x \in A)(\forall y \in B)(\neg P(x, y)). \quad (117)$$

Similarly:<sup>29</sup>

$$\neg[(\exists x \in A)(\forall y \in B)P(x, y)] \iff (\forall x \in A)(\exists y \in B)(\neg P(x, y)) \quad (118)$$

There are close relationships between sets and logical arguments. Here too, new symbols appear profusely. For instance, instead of  $\neg(a \in A)$  we write  $a \notin A$ , e.g.,  $\sqrt{2} \notin \mathbb{Q}$ .

A first important relation between sets include the "subset":

$$(A \subset B) \iff (a \in A) \implies (a \in B). \quad (119)$$

We have seen for instance how  $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ .

From this, one can deduce the following:<sup>30</sup>

<sup>26</sup> Apply to true and false implications, such as (99) and (102)

<sup>27</sup> Convince yourself that this is the case.

<sup>28</sup> Negate the statement (109) and discuss its meaning as negating asymptotic behaviour.

<sup>29</sup> Negate (110) and discuss.

<sup>30</sup> While they can appear intuitive, they can of course be proved rigorously, as you should check for yourself.

$$(A \subset B) \wedge (B \subset C) \implies A \subset C, \quad (120)$$

$$A = B \iff (A \subset B) \wedge (B \subset A). \quad (121)$$

We define  $A - B$  the set of elements  $a \in A$  such that  $a \notin B$ , i.e.,

$$(a \in A - B) \iff (a \in A) \wedge (a \notin B), \quad (122)$$

and call this the complementary of  $B$  in  $A$ . We can prove that, for  $A$  and  $B$  two subsets of  $C$ :<sup>31</sup>

$$A \subset B \iff C - B \subset C - A \quad (123)$$

Let us look at a trivial looking relation, if  $B \subset A$  then

$$A - (A - B) = B, \quad (124)$$

but it should not be understood as a trivial algebraic simplification since there is no subtraction implied, or in fact even defined, here. There is only a handy notation used for what has otherwise a precise meaning: set-complementarity. Let us prove it. We shall use the definition  $A = B$  for set equalities to mean  $a \in A \iff a \in B$ . Let us then assume  $a \in A - (A - B)$ , which means, by Eq. (122),

$$(a \in A) \wedge \neg(a \in A - B) \quad (125)$$

or, taking the negation of Eq. (122)

$$(a \in A) \wedge [(a \notin A) \vee (a \in B)] \quad (126)$$

which by distributivity (103) yields

$$[(a \in A) \wedge (a \notin A)] \vee [(a \in A) \wedge (a \in B)] \quad (127)$$

the left-hand side of which is a contradiction, so is false, but as it enters in a disjunction, it can simply be dropped (if it would enter in a conjunction it would make the whole statement wrong). So we are left with  $(a \in A) \wedge (a \in B)$  which, since  $B \subset A$ , reduces to  $a \in B$  by Eq. (119), which proves  $a \in A - (A - B) \implies a \in B$  but since all the steps of this demonstration are equivalence, this also proves the converse, thus proving the equivalence and by Eq. (123) proves Eq. (124).<sup>32</sup>

As previously noted, the concept of logical “and” is associated to intersection, and that of “or” to union:

$$x \in A \cap B \iff (x \in A) \wedge (x \in B), \quad (128a)$$

$$x \in A \cup B \iff (x \in A) \vee (x \in B) \quad (128b)$$

One can easily prove commutativity and associativity of set unions and intersections<sup>33</sup> as well as the following distributive relations:<sup>34</sup>

<sup>31</sup> Prove it and study what this says for the sets  $\mathbb{N}, \mathbb{Z}$  and  $\mathbb{Q}$ .

<sup>32</sup> What happens to Eq. (124) if  $B$  is not a subset of  $A$ ?

<sup>33</sup> Prove it.

<sup>34</sup> Prove them.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad (129)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \quad (130)$$

De Morgan's relationship read:<sup>35</sup>

<sup>35</sup> Prove them.

$$(X - A) \cap (X - B) = X - (A \cup B), \quad (131)$$

$$(X - A) \cup (X - B) = X - (A \cap B). \quad (132)$$

A final but very subtle point regarding logic and the material of this Lecture, which we need to state but that we cannot hope to fully clarify today, is that a relation can "not be true", without its negation being true. If the relation is false, then its negation is true, but there is a difference in logic between "not being true" and "being not true". Note the inadequacy of language to logical meaning, here we have the example where "can not  $\neq$  cannot" and there are plentiful examples, such as the choice "cheese *or* dessert" at a restaurant that has the meaning of a logical xor.

Such relations which are neither true nor false are called undecidable, they cannot be demonstrated from the axioms.  $R \implies \neg R$  remains true even if  $R$  and  $\neg R$  are not known to be true (undecidable). From the axioms,  $R \vee S$  being true does not imply that one proposition,  $R$  or  $S$ , is true. One can of course, in such a case, complete the set of axioms by deciding whether it is true or not true, but there will always be, in axiomatic systems of enough complexity, such as those required to describe arithmetics, some incompleteness, some statements which proof does not follow from the set of axioms. This is the famous Gödel's incompleteness theorem, which is really deep and, interestingly, bring us back to our earlier description of a starting point in the Platonic Universe to derive in it everything from there, as compared to the physical universe which may have no ultimate consistency or unified theory. Actually, the Platonic universe seems to suffer even more from such incompleteness, as in its case, it has even been demonstrated rigorously.





## Lecture 4: Complex numbers

The algebraic problem that involves the square of a quantity posed itself a long time ago and the historical record tells us that it was first investigated at depth by the Babylonians. In such remote times, before Al Khwarizmi's formalisation of algebra, the method was largely geometrical. In fact, Al Khwarizmi himself lacked important concepts such as negative numbers, and apparently also the zero, so his method still relied chiefly on geometrical arguments (namely completion of the square with actual figures), but using algebraic concepts of transforming an equation. A Spanish scholar, Abraham bar Hiyya, brought this Arabic knowledge into the occident and one of his works (*Liber Embadorum*) keeps the first known general solution to the quadratic equation.

The next big problem in algebra was the cubic equation, where the cube of the variable enters:

$$x^3 + bx^2 + cx + d = 0 \quad (133)$$

The main progress there came from Italy, in the late 15th and 16th century. The mathematician Scipione del Ferro had found a solution for the cubic equation<sup>36</sup>

$$x^3 + mx = n \quad (134)$$

that he, however, kept secret, until his death where the technique passed to his student (Antonio Fior). When it circulated that someone knew how to solve the cubic equation of the type  $x^3 + mx = n$  but kept the technique secret, Tartaglia released publicly his own method to solve a particular case  $x^3 + mx^2 = n$  and was subsequently challenged by Fior on problems of the  $x^3 + mx = n$  type. Tartaglia obtained the full solution that he also kept secret! He was lured by Cardano to share the discovery, under the promise not to disclose it until Tartaglia would publish it himself. Cardano who later realised del Ferro knew first about cubic solutions did not keep his promise and in 1545, in his *Ars Magna*, the first Latin treatise on Algebra, he published the solution, duly crediting del Ferro and Tartaglia for their respective inputs, as well as his own student Ferrari who, in the

<sup>36</sup> Show that knowing how to solve Eq. (134) is enough to solve the general case (133) by using  $y = x + b/3$ .

mean time, also solved the *quartic* equation, with a fourth power of the variable. Let us look, with modern notations, at the idea to solve the cubic. From the cubic expansion

$$(a - b)^3 = a^3 - 3a^2b + 3ab^2 - b^3 \quad (135)$$

one can write, factoring  $ab$  in the two middle terms of the rhs:

$$(a - b)^3 + 3ab(a - b) = a^3 - b^3 \quad (136)$$

which shows that  $a - b$  is solution of Eq. (134)

$$x^3 + 3abx = a^3 - b^3 \quad (137)$$

provided that

$$3ab = m, \quad (138a)$$

$$a^3 - b^3 = n. \quad (138b)$$

So these Renaissance Italian people managed to rewrite the original problem into a new one. The secret is that the new one can be solved, substituting Eq. (138a) into (138b) we find:

$$a^3 - \frac{m^3}{27a^3} = n \quad (139)$$

or, multiplying by  $a^3$  and rearranging:

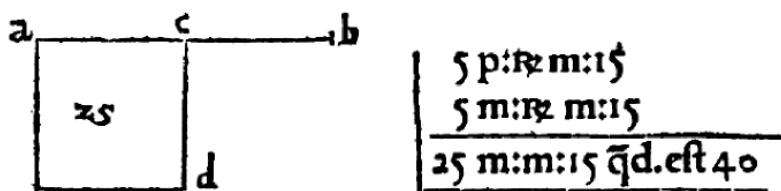
$$a^6 - na^3 - \frac{m^3}{27} = 0 \quad (140)$$

which is quadratic in  $a^3$ , hence solvable already since several centuries. This complete the solution.<sup>37,38</sup>

Interestingly, still at this time, negative numbers were not widely used. Consequently, cardano was providing different techniques to solve  $x^3 + ax = b$  and  $x^3 = ax + b$  (with  $a, b > 0$ ) and, in all, thirteen different types of cubic equations, each in a chapter of *Ars Magna*. Repeatedly, he observed however the occurrences of negative numbers, in a way as to make sense. His most important, ground-breaking observation lies in Chapter XXXVII, where he reduces his struggles with negative numbers to the algebraic procedures that provide answers which, although nonsensical for their purpose, are mathematically consistent. The problem he considers is the typical one that Babylonians were dealing with on their clay tablets and involves the quadratic equation only: divide a line of length 10 in two parts so that the area of the rectangle formed by them is 40:

<sup>37</sup> Provide closed-form solutions.

<sup>38</sup> Find the solutions of  $x^3 - 2x = 1$ .



This is impossible (the area is at most 25, cf. Problems), but a blind application of the formula that we have seen previously for the corresponding equation

$$x(10 - x) = 40 \quad (141)$$

gives solutions:

$$x = 5 \pm \sqrt{-15}, \quad (142)$$

where he introduced  $\sqrt{-15}$  the number which multiplied by itself, gives the negative number  $-15$ . This seems to be useless, because the numbers are not distances that we can locate on a ruler, but, surprisingly, that still works, since

$$(5 + \sqrt{-15})(5 - \sqrt{-15}) = 25 - (\sqrt{-15})^2 = 25 + 15 = 40 \quad (143)$$

Although it caused much controversies (and vexed feelings), the usefulness of such a trick in an age where Mathematics was more about exploration than about contemporary rigor, eventually led to their acceptance. Euler introduced  $i$  as a notation for the number which squares to  $-1$  (the imaginary unit) and made the topic flourish, culminating with the so-called Euler formula to which we shall return in a coming lecture. We first have to build ourselves “complex calculus” by mere “blind application” of the rules of algebra with the added rule:

$$i^2 = -1. \quad (144)$$

Interestingly, amazingly even, this is the only new rule—new axiom if you want Mathematical parlance—that we need to build a new branch of Mathematics, without destroying everything by resulting in contradictions or absurdities. Instead, we discover a whole new series of powerful results, that sometimes connect to others, well known, with remarkable elegance and depth. Really, Cardano just had opened a new door in the Platonic universe.

Let us explore indeed algebra (addition and multiplication) of numbers with the added component  $i$  (or, if using particular cases, arithmetic). This introduces a new—possibly the most important—set of numbers, namely, the set of *complex numbers*, with notation  $\mathbb{C}$ . These are numbers of the type

$$z = x + iy, \quad x, y \in \mathbb{R}. \quad (145)$$

with  $x$  and  $y$  are two “normal” (real) numbers, that we call the *real* ( $x$ ) and *imaginary* ( $y$ ) parts of  $z$ . The terminology is funny, and is rooted into history. There is little actually “complex” about these numbers, in fact, as we will come to appreciate, they are on the opposite much simpler than the “real” numbers. Let us check that everything is consistent and self-contained, i.e., that we do not create new types of numbers by manipulating those in  $\mathbb{C}$  with the known rules of algebra, in which case the definition (145) would be incomplete. In particular,  $z_1$  and  $z_2$  being two such numbers, we have:

$$z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2), \quad (146a)$$

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1). \quad (146b)$$

This follows from assuming that associativity and commutativity of real numbers also applies to complex numbers, as well as the distribution of multiplication over addition and, generally, all the rules that we need. Let us demonstrate Eq. (146b) slowly:

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) \quad (147a)$$

$$= x_1 x_2 + x_1(iy_2) + (iy_1)x_2 + (iy_1)(iy_2) \quad (147b)$$

$$= x_1 x_2 + i(x_1 y_2 + y_1 x_2) + i^2 y_1 y_2 \quad (147c)$$

where Eq. (147a) is a mere substitution, Eq. (147b) is the distribution  $(a + b)(c + d) = ac + ad + bc + bd$ , Eq. (147c) invokes associativity and commutativity, e.g.,  $(iy_1)(iy_2) = iy_1 iy_2 = ii y_1 y_2 = i^2 y_1 y_2$ , and factorizes the common  $i$ . Then substituting  $i^2 = -1$  this gives Eq. (146b).<sup>39</sup> So the addition and multiplication of two complex numbers is also a complex number, in the sense that it can be written in the form  $a + ib$  with  $a, b \in \mathbb{R}$ . How about the fraction of two complex numbers, though? This is maybe not immediately clear that we can write

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2}, \quad (148)$$

also in the form  $a + ib$ ! To show that this is indeed possible, we introduce an important concept, *complex-conjugation*, denoted by aposing a star  $*$  to a complex number, and that consists in changing the sign of the term that comes with  $i$ , i.e.,

$$z^* = (x + iy)^* = x - iy. \quad (149)$$

This seems a silly thing to do but this is actually a very important operation because it gives us access to the constituents ( $x$  and  $y$ ) of  $z$  from  $z$  itself, namely:

$$\Re(z) \equiv x = (z + z^*)/2, \quad (150a)$$

$$\Im(z) \equiv y = (z - z^*)/(2i), \quad (150b)$$

<sup>39</sup> Show similarly Eq. (146a) (maybe you can do this in your head?)

where we introduced the so-called “real-part” and “imaginary-part” of  $z$ . For instance one can easily see when a complex number, is not:

$$z \in \mathbb{R} \iff z = z^*. \quad (151)$$

More importantly, one can also force the imaginary part out of the number and reduce it to a real number by taking the product with its complex conjugate:<sup>40</sup>

$$zz^* = x^2 + y^2 \quad (152)$$

which is an important combination for which we introduce a new notation

$$|z|^2 \equiv zz^* \quad (153)$$

and name (the “modulus square”), as the modulus  $|z| \equiv \sqrt{zz^*}$  has the properties reminiscent of a “distance” or “magnitude” for the complex number  $z$ , namely <sup>41,42,43</sup>

$$|z| \in \mathbb{R}^+, \quad (154a)$$

$$|z| = 0 \iff z = 0, \quad (154b)$$

$$|z_1 + z_2| \leq |z_1| + |z_2|. \quad (154c)$$

These are indeed the properties of “distances”, that transform numbers into a length of some sort. For the numbers we have met so far, the number itself serves as its own distance, unless it is negative, in which case, we use the “absolute value” to simply chops the sign off, e.g., the distance of  $-3$  is  $|-3| = 3$ , so that  $|x|$  is the distance of  $x$  for  $x \in \mathbb{N}, \mathbb{Z}, \mathbb{Q}$  and  $\mathbb{R}$ . We wouldn’t even need it for  $\mathbb{N}$  but that helps us remember we speak of a distance. Consequently, we keep the same notation for  $\mathbb{C}$  but the definition is different, as this time we have the more delicate task to chomp the  $i$  out, so we use Eq. (153).

With this important concept, one can now check, still using old-fashioned algebra, that<sup>44,45</sup>

$$\frac{z_1}{z_2} = \frac{z_1 z_2^*}{|z_2|^2}, \quad (155)$$

and therefore that, indeed,  $z_1, z_2 \in \mathbb{C} \implies z_1/z_2 \in \mathbb{C}$ . Namely, back to Eq. (148), using Eq. (155), we show that  $z_1/z_2$  is complex, i.e., is of the type  $\alpha + i\beta$  for  $\alpha, \beta \in \mathbb{R}$ , simply by computing explicitly

$$\alpha = \frac{x_1 x_2 - y_1 y_2}{x_1^2 + x_2^2}, \quad (156)$$

$$\beta = \frac{-x_1 y_2 + x_2 y_1}{x_1^2 + x_2^2}, \quad (157)$$

and since the denominator is nonzero from Eq. (154b) (assuming  $z_2 \neq 0$ , one cannot divide by zero in  $\mathbb{C}$  either), this completes our proof that  $z_1/z_2 \in \mathbb{C}$ .

<sup>40</sup> Prove this.

<sup>41</sup> Prove these, including Eq. (154c) which is the so-called triangle inequality and that we will prove later for real numbers, but assuming it is true for them, then you can show it is also true for complex numbers!

<sup>42</sup> Compare the square of a sum of complex numbers  $(z_1 + z_2)^2$  to the modulus square of a sum  $|z_1 + z_2|^2$ .

<sup>43</sup> Show that  $|z| = |z^*|$ ,  $|z_1 z_2| = |z_1| \times |z_2|$  and  $|z_1/z_2| = |z_1|/|z_2|$ . Show, however, that  $|z_1 + z_2| \neq |z_1| + |z_2|$ .

<sup>44</sup> Prove this.

<sup>45</sup> Compute  $1/(1-i)^k$  for  $k = 1, 2, 3$ .

By recurrences, we see that any sums, products and their inverses remain complex numbers, that is, of the type  $x + iy$  (for instance,  $i^3 = -i$  and also  $1/i = -i$ ).<sup>46</sup> They form a closed set of numbers under these operations, for which we can also always find an inverse (except zero that has no inverse for the product operation). Mathematicians say that “the set  $\mathbb{C}$  of Complex number is a field”. For us, it is indeed a consistent set of numbers under addition and multiplication.

But how about more complicated operations, like powers and roots? What about, for instance  $\sqrt{i}$ ? By definition, this is the number whose product with itself gives  $i$ . Which number is that? By asking this question, we ask ourselves, what are  $x, y \in \mathbb{R}$  such that

$$(x + iy)^2 = i. \quad (158)$$

By the way, two complex numbers are equal iff their real and imaginary parts are equal.<sup>47</sup> Spelling this out, defining  $z_k = x_k + iy_k$  for  $k = 1, 2$ , it means:

- If  $x_1 = x_2$  and  $y_1 = y_2$ , then  $z_1 = z_2$ ,
- If  $z_1 = z_2$  then  $x_1 = x_2$  and  $y_1 = y_2$ .

Back to our square root problem. From Eq. (158), we will want to use that  $z_1 = z_2 \Rightarrow x_1 = x_2$  and  $y_1 = y_2$ . So let us compute:

$$\begin{cases} x^2 - y^2 = 0, \\ 2xy = 1, \end{cases} \quad (159)$$

with solutions  $\pm(1 + i)/\sqrt{2}$ .<sup>48</sup> Here we have two solutions. This was also the case for the real case:

$$(\pm\sqrt{2})^2 = 2 \quad (160)$$

but it was convenient for convention to chose the positive one as the unique solution. In the complex case, this is not so straightforward, and we shall see in the future that it turns out to be an important feature. We will leave it at that for now but it will be a recurrent theme that functions of complex numbers return complex numbers. We have no need for new quantities to make the whole edifice stand up, only  $i$ .

What goods are complex numbers? In physics, as we will see, they are indispensable. The quantum world, for instance, is written in complex numbers. The reason is due to the fact that, beyond a length, or magnitude,  $|z|$ , they also carry a notion of “phase”, or angle. One insightful way to represent, or plot, complex numbers, is the so-called *Cartesian form*, which means that we represent a number in the form

<sup>46</sup> Compute  $(1 + i)^5$  using the binomial formula and  $i^2 = -1$ .

<sup>47</sup> Prove this.

<sup>48</sup> Prove this; remember that the difference of squares is also a product of the sum and difference and a product equals zero iff one or the other terms equals zero.

of Eq. (145) as a point in the so-called *complex plane*. It's a plane because complex numbers are two-dimensional: one axis is real, the other imaginary (such mystifying worst for otherwise straightforward concepts). Every point of the Cartesian plane can also be given a so-called *polar representation*, that is, instead of a horizontal and vertical distance, one give a total distance and an angle:

$$x + iy = r(\cos \theta + i \sin \theta), \quad (161)$$

where  $r = |x + iy|$  and  $\theta$  such that  $\tan \theta = y/x$  provided  $x \neq 0$  in which case  $\theta = \pm\pi/2$  (depending on the sign of  $y$ ).<sup>49</sup> This shows incidentally that  $zz^*$  really is the natural notion of "distance" or magnitude for a complex number since it corresponds to Pythagoras theorem. Later on, we will see that such polar decomposition allows us to perform extremely advanced calculations with angles in a very easy and straightforward way.<sup>50</sup> As such, complex number will be very powerful to address problems of trigonometry, whenever an angle is involved. Indeed, rather than "complex" they might have been christened "trigonometric numbers", "geometrical numbers" or "phase numbers".

We conclude today's lecture with a fairly recent (beginning of the 20th century) discovery related to complex numbers, which should make the Platonic universe more palpable and substantial. We only added one tiny rule to our algebra, namely, Eq. (144). Still, the consequences of this innocent move bring us to unravel unexpected features of the Platonic universe, whose complexity (this time in the literal sense of the term) and richness defies the imagination. Consider this simple *iterative* map to construct a so-called *sequence* of complex numbers, starting from  $z_0 = c$ , we define  $z_1 = z_0^2 + c$  (that is,  $z_1 = c^2 + c$ ),  $z_2 = z_1^2 + c$  (that is  $z_2 = (c^2 + c)^2 + c$ ) and so-on and so-forth, with general rule:

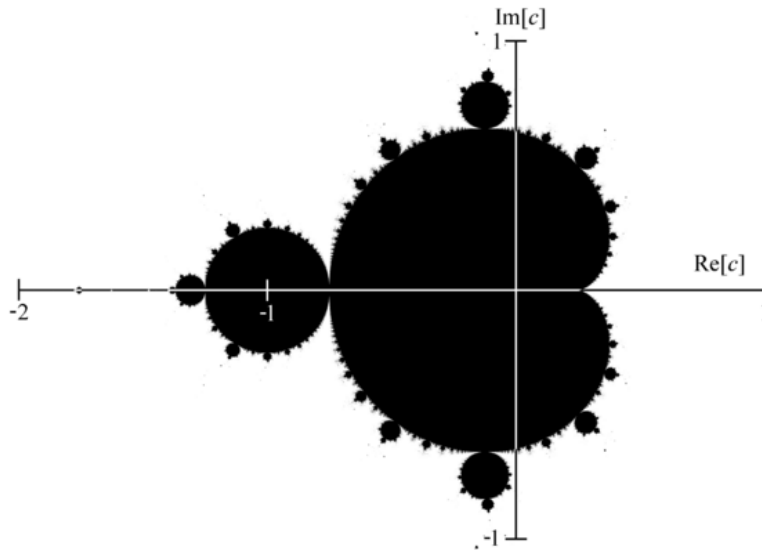
$$z_{n+1} = z_n^2 + c. \quad (162)$$

Now two things can happen with these numbers. Either they remain in a bound region of space, either they drift to infinity. That is, either one can define a circle which will contain all  $z_n$ , or no such circle exist. If we color each point  $c = x + iy$  in black if they are bound, and in white if they are not, we find the following structure:<sup>51</sup>

<sup>49</sup> Provide the couples  $(r, \theta)$  for the numbers  $i^k$  with  $k = 0, 1, 2, 3, 4, \sqrt{i}$  and  $(1 + 2i)$ .

<sup>50</sup> As an exploratory exercise, take a few complex numbers  $z$  of your choice and compute  $z^2$ . What happens to their respective magnitude lengths and angles after squaring? See Problems for more advanced explorations.

<sup>51</sup> Check whether the points  $i^k$  for  $k = 0, 1, 2, 3$  belong to the Mandelbrot set. Do your results seem to agree with the figure?



It is of surprisingly complexity (and beauty). There are a lot of little structures that pop out of the general shape. The most amazing is that if we “zoom”, by focusing on a little area nearby, say, one of the little dendrites, we find that new structures appear, and that, however deep we delve into this, more and more landscapes appear. This is the basis for *fractal* geometries, discovered in the last century. The above figure is known as the *Mandelbrot set*, after Benoit Mandelbrot. It is surprising that simple algebra (iteration of squaring and adding a constant) on simple numbers (merely extending them with a new object  $i$  such that  $i^2 = -1$ ) reveals such an amazing universe. Its existence by itself, independent from our own physical world and/or imagination, is compelling. This is the Platonic universe in its most obvious, and possibly spectacular, manifestation.

### *Biographical notes*

**Gerolamo Cardano**, 1501—1576, Italian polymath, one of the most influential mathematicians of the Renaissance, with major inputs in probability theory, in particular for his introduction of the binomial coefficients and theorem.

**Benoit Mandelbrot**, 1924—2010 was a sort of contemporary Cardano. Also a polymath, he discovered fractal structures in pretty much the same way that Cardano discovered complex numbers.



## Problems

### Cardano area

Cardano's problem takes the algebraic form  $(10 - x)x = 40$  or, written as a canonical quadratic equation:

$$x^2 - 10x + 40 = 0. \quad (163)$$

Consider the length is not 10 but a variable parameter  $L$ . Write the discriminant  $\Delta$  for Eq. 163 and find for which values of  $L$  the problem admits positive solutions. What can be said about the two solutions? What happens when  $\Delta = 0$ ? For each cases,  $\Delta < 0$  (say with  $L = 10$ ),  $\Delta = 0$  and  $\Delta > 0$ , plot Eq. (163) as a function of  $x$  and locate the algebraic solutions on the  $x$  axis.

### Cardano cruise

Check the answers to Cardano's problem (Eq. (142)). He was not the first to be involved with the quadratic equation, this goes back as far as the Egyptians, Chinese and Babylonians, who met quadratic equations in problems related to areas (*Quadratic* equation means there is a square (*quad*) involved). We will repeatedly meet it in Physics problems, and are very happy when this is the case, because it means we can easily solve the problem). Often, we don't know beforehand. Here is a typical example:

A 3 hour river cruise goes 15 km upstream and then back again. The river has a current of 2 km an hour. What is the boat's speed and how long was the upstream journey?

To interest Cardano, one would like to bring this problem into the realm of impossibilities. Is there a minimum time for which the boat cannot complete the journey? What does your intuition tell you? How about the algebra? What is breaking up first in this case? Real analysis or classical mechanics?

### Complex circuit

Consider two resistors in parallel, with resistance  $R_1 - R_0$  and  $R_1 + R_0$ . For a given  $R_1$ , how would you chose  $R_0$  to have the total resistance of the circuit be  $R_1/4$ ,  $R_1/2$ ,  $R_1$ ?

### Complex geometry

We will see that complex numbers are super-powerful when it comes to geometry. In fact, they are basically geometric numbers. To give you a hint of that, consider a simple shape in the complex plane, for

instance, the triangle with edges:

$$\mathcal{T} = \{0, i, 1 + i\} \quad (164)$$

which we represent as a set of three points, that you can plot in the complex plane (and connect with lines to guide they eye). Then construct the following shapes:

- The conjugate of  $\mathcal{T}$ , i.e.,  $\mathcal{T}^* \equiv \{z^*, z \in \mathcal{T}\}$ .
- The opposite of  $\mathcal{T}$ , i.e.,  $-\mathcal{T} \equiv \{-z, z \in \mathcal{T}\}$ .
- The displaced  $\mathcal{T}$ , e.g.,  $\mathcal{T} + (2 + i) \equiv \{z + (2 + i), z \in \mathcal{T}\}$ .
- $i\mathcal{T}$ , i.e.,  $\{iz, z \in \mathcal{T}\}$ .
- $\frac{1+i}{\sqrt{2}}\mathcal{T}$ , i.e.,  $\{\frac{1+i}{\sqrt{2}}z, z \in \mathcal{T}\}$ .
- The imaginary part of  $\mathcal{T}$ , i.e.,  $\Im(\mathcal{T}) \equiv \{\Im(z), z \in \mathcal{T}\}$ .
- The square of  $\mathcal{T}$ , i.e.,  $\mathcal{T}^2 \equiv \{z^2, z \in \mathcal{T}\}$ .
- The cube of  $\mathcal{T}$ , i.e.,  $\mathcal{T}^3 \equiv \{z^3, z \in \mathcal{T}\}$ .
- The inverse of  $\mathcal{T}$ , i.e.,  $1/\mathcal{T} \equiv \{1/z, z \in \mathcal{T}\}$ .

In each case, how would you describe the corresponding transformation? Since  $z = 0$  is in  $\mathcal{T}$ , and one cannot divide by zero, you'll get into a problem with the last set. How would you go around it?

### *Roots of $i$*

We have seen that the solution of Eq. (159) is

$$\pm(1 + i)/\sqrt{2}, \quad (165)$$

which you can check explicitly. Can you compute (or in any other way find)  $\sqrt[3]{i}$ ? (of course checking your answer).

## Lecture 5: Vectors

We have seen a lot of Mathematical objects already, and all fell in the category of “numbers”. Namely, we have discussed  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$  and  $\mathbb{C}$ . The last ones were a particular type of numbers, in the sense that they brought something more to the notion of “length” or “size” of the numbers, which is clear from all the other types and is captured by the notion of *ordering*, that is, the ability we have to say that one number is smaller than an other, e.g.,

$$\pi < 5, \quad (166)$$

or

$$-7 < -2. \quad (167)$$

The latter may require more thought. If you think in terms of a quantity which we easily associate to possibly negative values, such as the amount of money one has on their bank account, then Eq. (167) refers to the fact that someone with a credit of  $-\pounds 2$  is richer (or less poor) than someone with a debt of  $-\pounds 7$ . Interestingly, this is how negative numbers came to be conceptualised in Mathematics, through debts.

So far so good. But how do we order complex numbers? What would you say is “larger” or “bigger”,  $1$  or  $i$ ?

$$i < 1 \quad \text{or} \quad 1 < i? \quad (168)$$

(to be sure, Eq. (168) is meaningless). The thing is that complex number bring another dimension to their structure, namely, a *phase*. Their length is captured by the modulus square, which we have defined last lecture as, for any complex number  $z = a + ib$ :

$$|z|^2 = z^*z = a^2 + b^2. \quad (169)$$

So in this sense, both  $1$  and  $i$  have the same length. As do many other numbers, namely, all numbers on the unit circle. What distinguish them then is, as we said, the angle on this circle, which, in Physics, we call a “phase”. So complex number, in some sense, can be seen as numbers with a magnitude (length) and a direction (phase). They are more than that, but they are also simply that. As such, they are the

simplest example of objects with such a structure, which are called, *vectors*. Because we need two numbers ( $a$  and  $b$  in the Cartesian form, or, equivalently,  $R$  and  $\theta$  in the polar form), they are two-dimensional (2D) vectors.

Here we will leave complex numbers for a while, and turn instead to the general case of an  $n$ -dimensional vector. This means an object for which we need  $n$  numbers (we'll start with real vectors, that means,  $n$  real numbers to define it). It is handy to list these numbers in a column. If we note

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}} \quad (170)$$

the set of all  $n$ -dimensional real vectors, then we can write  $\mathbf{v} \in \mathbb{R}^n$  as

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \quad (171)$$

The numbers  $v_i$ , for  $1 \leq i \leq n$ , are called the *coordinates* of the vector. Now we can do some vector algebra. We can easily add vectors:

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}. \quad (172)$$

If we add a vector  $k$  times to itself, for  $k \in \mathbb{N}$ , we find:

$$k\mathbf{v} = \begin{pmatrix} kv_1 \\ \vdots \\ kv_n \end{pmatrix}, \quad (173)$$

and using the same reasoning as for the algebra of rational and real numbers, you should be able to convince yourself that Eq. (173) holds for  $k \in \mathbb{R}$  too.

From the results above, we can do something quite neat, namely, decompose Eq. (171) as what we will later call a *linear superposition* of more fundamental vectors, as follows:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + v_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (174)$$

The vectors with a 1 at the  $k$ th entry are called the *canonical basis* vectors:

$$\mathbf{e}_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}. \quad (175)$$

They are a basis because every vector can be written in terms of them (weighted by the coordinates  $v_k$ ) and they are canonical because they are not the only choice one can make to make such a decomposition, but they are by far the simplest one. The number of such vectors we need is given by the dimension of the space.

Let us carry on with algebra of vectors. Here will surely come a great surprise and/or disappointment. The products of vectors we would like to define, likewise, as a product of corresponding elements. This operation exists, it is called a *hadamard product*, and comes with a strange symbol for the product:

$$\mathbf{u} \odot \mathbf{v} = \begin{pmatrix} u_1 v_1 \\ u_2 v_2 \\ \vdots \\ u_n v_n \end{pmatrix} \quad (176)$$

but it turns out *not* to be a very useful operation with vectors. There is a good reason why this product is not useful and other types of products are used, instead. This has to do with the structure of a so-called *vector space*, which we will study systematically when we embark onto Linear Algebra, which is the algebra of “vector-like” objects.

Instead of Hadamard product, one use a much more important quantity, known as a *scalar product*, for which we use as a notation  $\cdot$  and that produces a *scalar* (whence the name) as follows:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i. \quad (177)$$

The scalar product with itself, namely

$$\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^n v_i^2 \quad (178)$$

serves to define the so-called *norm* of a vector, which is its length (like the modulus for complex numbers):

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}. \quad (179)$$

You'll find it very difficult to resist the temptation to write

$$\mathbf{v}^2 \equiv \mathbf{v} \cdot \mathbf{v} \quad (180)$$

and indeed we use this a lot, although it's important to understand it's a notation (or convention) and that  $\mathbf{v}^2$  is *not*  $\mathbf{v}\mathbf{v}$  (without the  $\cdot$ ) which is not defined! Look how Eq. (179)  $\|\mathbf{v}\| = \sqrt{\mathbf{v}^2}$  looks silly when we start to abuse notations!<sup>52</sup>

We can check however that the algebra with the scalar product works "as expected", namely:<sup>53</sup>

- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ .
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$ .

These are easy to prove applying the very definition (177)<sup>54</sup> <sup>55</sup> For instance, we prove the last identity in  $\mathbb{R}^n$ . The vector  $\mathbf{u} + \mathbf{v}$  has  $k$ th coordinate  $u_k + v_k$  by definition, so its scalar product with  $\mathbf{w}$  is

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \sum_{k=1}^n (u_k + v_k)w_k \quad (181)$$

but each element  $(u_k + v_k)w_k$  of the sum is  $u_k w_k + v_k w_k$  by algebra of real numbers (distributivity) and the sum is commutative so

$$\sum_{k=1}^n (u_k w_k + v_k w_k) = \sum_{k=1}^n u_k w_k + \sum_{k=1}^n v_k w_k \quad (182)$$

but by definition, the rhs is  $\mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$ , QED. ("QED" stands for "Quod Erat Demonstrandum" meaning, "which was to be proved" and implies that our job is done there). These canonical basis vectors also allow us to pick-out the coordinates of the vectors, indeed:

$$v_k = \mathbf{v} \cdot \mathbf{e}_k \quad (183)$$

so that one can write, a bit redundantly but this is useful in many occasions:

$$\mathbf{v} = \sum_{k=1}^n (\mathbf{v} \cdot \mathbf{e}_k) \mathbf{e}_k \quad (184)$$

(don't let yourself be intimidated by the notation; this is merely Eq. (174) where we have substituted Eq. (183)).

A vector which has length one is said to be "normalized". If a vector has not unit length, we can still "normalize it", which is something we'll be doing a lot, especially in quantum mechanics:

$$\text{Normalized } \mathbf{v} \text{ is } \frac{\mathbf{v}}{\|\mathbf{v}\|}. \quad (185)$$

<sup>52</sup> Why silly?

<sup>53</sup> The first line says that the scalar product is commutative. What could we say about its associativity?

<sup>54</sup> Do it.

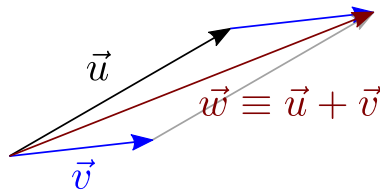
<sup>55</sup> Compute, for the canonical basis  $\hat{i}, \hat{j}$  and  $\hat{k}$  the scalar products  $\hat{i} \cdot \hat{j}$ ,  $\hat{j} \cdot \hat{k}$  and  $\hat{k} \cdot \hat{i}$  (the other possibilities, like  $\hat{i} \cdot \hat{i}$  or  $\hat{j} \cdot \hat{j}$  should follow straightforwardly).

While the length of a vector is easily defined, the “direction” is a different matter altogether, especially as it’s a direction in a  $n$ -dimensional space, which is quite tricky to visualize.

There are particular cases of  $n$ -dimensional vectors which are very important because of their familiarity and relevance to our physical space. They are the so-called *geometrical vectors*. They are basically vectors of dimension 2 and/or 3. For them we have a special notation:

$$\vec{u}, \vec{v} \quad (186)$$

and since this typically refer to space, instead of 1, 2, 3, we label their coordinates as  $x, y, z$ . These vectors, unlike the  $n$ -dimensional general case, are easily represented in the space. The basic rule is that only their length and direction matter, so their origin, in particular, is irrelevant. This allows us to translate vectors in space. This makes it easy, for instance, to add them, following the “parallelogram law”:<sup>56</sup>



Note the geometrical interpretation of  $\vec{w} - \vec{u}$ : it is the vector  $\vec{w}$  (in particular, that has the same head as this vector) of which we removed  $\vec{u}$  (that is, its tail starts at the head of  $\vec{u}$ ).

A common convention for the canonical basis of geometrical vectors, is to call them:

$$\hat{i}, \hat{j}, \hat{k} \quad (187)$$

where we use a hat  $\hat{\phantom{x}}$  rather than an arrow  $\vec{\phantom{x}}$  to mean that the vector is normalized. Then every geometrical vector reads:

$$\vec{v} = v_x \hat{i} + v_y \hat{j} + v_z \hat{k}. \quad (188)$$

We can of course, like for all vectors, compute their scalar product, in which case:<sup>57</sup>

$$\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z \quad (190)$$

(in 3D, in 2D we drop the  $z$  component).

There is an important equivalent way to look at scalar products for geometric vectors, namely:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \theta \quad (191)$$

where  $\theta$  is the angle between  $\vec{u}$  and  $\vec{v}$ . Note that even in 3D, there is only one angle as two vectors can always be fit in the same plane.

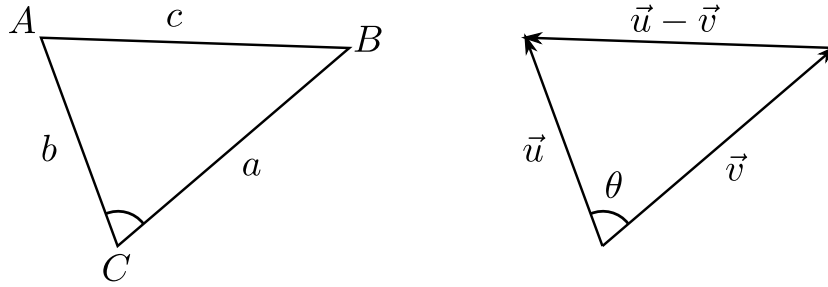
<sup>56</sup> Convince yourself that this follows from Eq. (172) applied to geometrical vectors.

<sup>57</sup> Use the results of footnote 55 to obtain Eq. (190) from the direct computation of

$$(u_x \hat{i} + u_y \hat{j} + u_z \hat{k}) \cdot (v_x \hat{i} + v_y \hat{j} + v_z \hat{k}). \quad (189)$$

Eq. (191) is proven from the law of cosines (which requires a proper and self-contained definition of trigonometric functions, in particular the cosine, so we postpone the proof itself for a future lecture and take it as true for now). Namely, in a triangle ABC, with the letter  $l$  opposite the vertex  $L$  for  $l \in \{a, b, c\}$ , we have:

$$a^2 + b^2 - 2ab \cos C = c^2 \quad (192)$$



in which case, from the figure above, we have:

$$\|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos\theta = \|\vec{u} - \vec{v}\|^2 \quad (193)$$

but expanding the right-hand side, we find, by definition:

$$\|\vec{u} - \vec{v}\|^2 = (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) \quad (194)$$

and expanding this following the algebra of vectors:

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\vec{u} \cdot \vec{v} \quad (195)$$

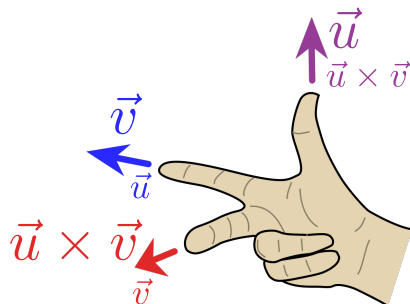
which, inserted back into Eq. (193) and after simplification, yields Eq. (191). This is an important way to understand scalar product, this is something that involves an *overlap*. When two quantities are not seeing each other, we say that they are “orthogonal”. Their scalar product is then zero.

Finally, we turn to a last important operation with geometrical vectors. It is much less important than the scalar product, which applies to all types of vectors, but it is fundamental for physics because many things (such as the position, or velocity or other dynamical variables for a mechanical object, or the electromagnetic field) are described by 3D vectors. This is the so-called *vector product* (it is also called the “cross product”), which indeed turns two vectors into another vector. It has a nice “opposite” relationship to the scalar product. First, it is defined as:

$$\vec{u} \times \vec{v} = \|\vec{u}\| \|\vec{v}\| \sin(\theta) \hat{n} \quad (196)$$

where  $\hat{n}$  is the (normalized) vector perpendicular to the plane that contains  $\vec{u}$  and  $\vec{v}$  in the direction given by the right-hand rule:





It does not matter which fingers you chose (big letters starting on the thumb or small letters starting on the index) as long as they are in the good cyclic order.

Its computation in terms of coordinates requires the notion of a *determinant*, which is an important quantity related to objects we have not yet seen, called matrices, that are tables of numbers rather than just columns (like vectors). This makes the product of elements but never taking two elements on the same column or the same row, and we put a minus sign in front if we don't make the selection directly rightward and downward an even number of times (assuming cyclic boundaries):<sup>58 59</sup>

$$\vec{u} \times \vec{v} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix} = \hat{i}(u_y v_z - v_y u_z) + \hat{j}(u_z v_x - u_x v_z) + \hat{k}(u_x v_y - v_x u_y). \quad (199)$$

Why is all this important? Why do we need such complicated mathematical objects beyond mere numbers? Well, because the universe can be described by such objects. For instance, light is made of two vectors, an electric vector  $\vec{E}$  and a magnetic vector  $\vec{B}$ . The flow of energy and momentum of light is described by a so-called “Poynting vector” that is a cross product of electricity and magnetism:

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}. \quad (200)$$

The intensity of light will be measured through the scalar product  $|\vec{E}|^2$ . Fine, we need 3D vectors, fair enough, but why do we need to bother about  $n$ -dimensional ones, and about complex numbers? As we will see next Semester in quantum mechanics, something as simple and elementary as an electron, for instance, is described by such a  $n$ -dimensional vector... what's more, with complex values:

$$\psi = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \quad (201)$$

<sup>58</sup> Check the distributivity of the vector product over vector addition, i.e., show that

$$\vec{a} \times (\vec{b} + \vec{c}) = \vec{a} \times \vec{b} + \vec{a} \times \vec{c}. \quad (197)$$

This is a nice case where the algebra is not commutative. Namely, compare  $\vec{a} \times \vec{b}$  and  $\vec{b} \times \vec{a}$ . How are they related? What name would you give to this?

Compute, for the canonical basis  $\hat{i}$ ,  $\hat{j}$  and  $\hat{k}$ , the vector products  $\hat{i} \times \hat{j}$ ,  $\hat{j} \times \hat{k}$  and  $\hat{k} \times \hat{i}$  (the other possibilities, like  $\hat{i} \times \hat{i}$  or  $\hat{j} \times \hat{i}$  should follow straightforwardly).

<sup>59</sup> From vector-product algebra, prove Eq. (199), i.e., compute directly:

$$(u_x \hat{i} + u_y \hat{j} + u_z \hat{k}) \times (v_x \hat{i} + v_y \hat{j} + v_z \hat{k}). \quad (198)$$

with  $\alpha_i \in \mathbb{C}$  and  $n \in \mathbb{N}$  depending on the surroundings of our electrons. Actually, in some cases,  $n$  can even be unbounded, or infinite, and we have an infinite-size vector! What's more, we'll soon see that such vectors can also be "continuous". At any rates, the universe is thus made that it is described by very abstract objects (such as "complex-valued infinite-size vectors"). But that's how it is. And to understand its meaning, structure and beauty, we have to be comfortable with algebra of these objects. Which is actually the easy part.

## Problems

### Triangle inequalities

Inspiring yourself from the law of cosines (also known as Al Kashi's theorem), prove the triangle equalities  $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$  and  $|\|\vec{u}\| - \|\vec{v}\|| \leq \|\vec{u} - \vec{v}\|$ .

### Jacobi identity

The vector product is not associative (show it). Instead, it satisfies the so-called *Jacobi identity*, which is:

$$\vec{a} \times (\vec{b} \times \vec{c}) + \vec{b} \times (\vec{c} \times \vec{a}) + \vec{c} \times (\vec{a} \times \vec{b}) = \vec{0}. \quad (202)$$

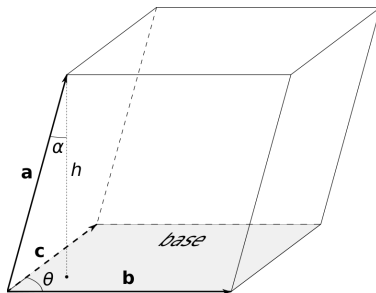
Prove this.

### Bringing scalar and vector products together

Show that the following expression

$$\vec{a} \cdot (\vec{b} \times \vec{c}) \quad (203)$$

gives the volume of a parallelepiped, whose sides are defined by the vectors  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$ . Actually, this can be negative, so this refers to the "signed" volume which can be regarded as a convention to chose the order of the labeling of the parallelepiped.



Show that the following properties are true, both from the point of view of vector algebra, and from interpreting them in geometrical terms (volume of a parallelepiped).

- $\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b}),$
- $\vec{a} \cdot (\vec{b} \times \vec{c}) = (\vec{a} \times \vec{b}) \cdot \vec{c},$

The following has no interpretation in terms of a volume, as far as I know, so you have to rely on algebra only:

$$a \cdot (\vec{b} \times \vec{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \quad (204)$$

*Remarkable identity*

An important relation (which will be used in electromagnetism) is:

$$\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{u} \cdot \vec{v})\vec{w}. \quad (205)$$

Prove it.



## Lecture 6: Functions

Now that we have a fairly large collection of mathematical objects ( $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{R}^n$ ), it is time to do things with them, beyond playing with their mere algebra; that is, it is time to “act” on them. This brings us to the notion of a *function*:

$$\begin{aligned} f : \mathbb{A} &\longrightarrow \mathbb{B}, \\ x \in \mathbb{A} &\rightarrow f(x) \in \mathbb{B}. \end{aligned} \tag{206}$$

This is a new mathematical object  $f$ , this time, definitely not a number, but still, with similar underlying principles, starting with the fact that one can consider their algebra: how do functions add and multiply together?

Let us first delve deeper into what a function is and does. A function could be any mapping between  $\mathbb{A}$  (called the “*domain*” of the function) and  $\mathbb{B}$  (the “*codomain*”) and these sets themselves could be anything. For instance, the “age” function could be defined on the set

$$\{ \text{Nuha, Kelly, Luke, Oliver, Sebastian, Thomas, Humza, Edward, Keyshawn, Grzegorz, Gary, Manjinder, Katarzyna, Ebun} \}$$

which is a function from  $\mathbb{A}$  = set of students in the class to  $\mathbb{B}$  =  $\mathbb{R}^+$ . So, for instance,  $\text{age}(\text{Humza})=24$  (this is a fictional example). A function in Mathematics is always single-valued, which means that every value of the domain takes only one value in the codomain (we will see concepts of “multi-valued functions” but they are beyond the strict mathematical definition of a function). If, furthermore, every value from  $A$  takes a different value in  $B$ , we say that the function is *injective*. This is the formal definition:

$$(x \neq y) \Rightarrow (f(x) \neq f(y)) \tag{207}$$

or, equivalently, by negation of the implication

$$(f(x) = f(y)) \Rightarrow (x = y). \tag{208}$$

If every value in  $B$  is taken by at least one value in  $A$ , we say that the function is *surjective*. This is the formal definition:

$$(\forall y \in \mathbb{B})(\exists x \in \mathbb{A})(y = f(x)). \tag{209}$$

A function that is both injective and surjective is called *bijjective*.<sup>60</sup>

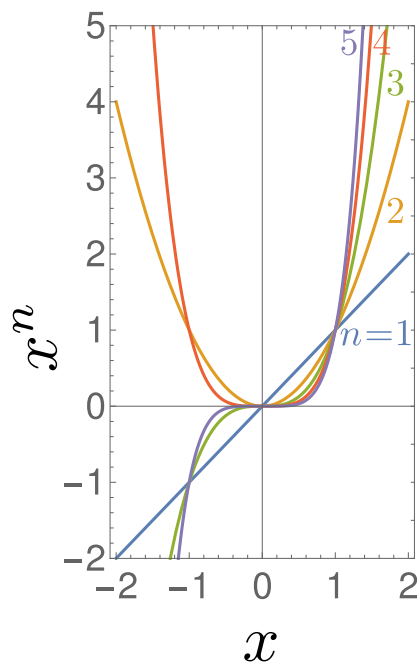
Whatever the sets, there is a function that is always defined, the *constant function*:

$$(\forall x \in \mathbb{A})(\exists y \in \mathbb{B})(f(x) = y). \quad (210)$$

Other famous types of functions that also exist in many cases, are particular cases, for instance:

- The zero function:  $(\forall x \in \mathbb{A})(f(x) = 0)$  (this is a particular case of the previous case when  $0 \in \mathbb{B}$ ).
- The identity function:  $(\forall x \in \mathbb{A})(f(x) = x)$  (this is possible only when  $\mathbb{B} \subset \mathbb{A}$ ).

The set of all pairs  $(x, f(x))$  is called the *graph* of the function. When  $\mathbb{A}$  and  $\mathbb{B}$  are sets of (non-complex) numbers, it makes it easy to visualize the function, by mean of a plot. Here, for instance, are the first *power functions*  $x \rightarrow x^n$  (for  $1 \leq n \leq 5$ ):



All functions intersect at  $x = 0$  and  $x = 1$ . This plots make it clear, for instance, that:

- If  $x < 1$ ,  $n < m \Rightarrow x^n > x^m$ ,
- If  $x > 1$ ,  $n < m \Rightarrow x^n < x^m$ .

Functions that are algebraic combinations of a variable  $x$  (mixing addition and multiplications) constitute a big share of what we understand, or deal with, as functions. When this involves only

<sup>60</sup> Give examples of functions that are injective (but not surjective), surjective (but not injective) and bijective on the set of students.

positive-integer-power functions, we speak of *polynomials* (the power functions above are then particular cases known as *monomials*). These are functions of the form:

$$P(x) = \sum_{k=0}^n \alpha_k x^k, \quad (211)$$

e.g., these are examples of polynomials, of increasing “order” or “degree” (the highest power):<sup>61</sup>

$$H_0(x) = 1, \quad (212a)$$

$$H_1(x) = 2x, \quad (212b)$$

$$H_2(x) = 4x^2 - 2, \quad (212c)$$

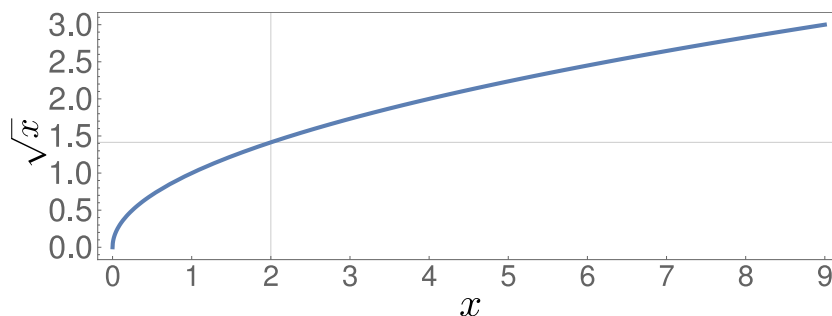
$$H_3(x) = 8x^3 - 12x, \quad (212d)$$

$$H_4(x) = 16x^4 - 48x^2 + 12, \quad (212e)$$

$$H_5(x) = 32x^5 - 160x^3 + 120x. \quad (212f)$$

This particular sequence is one of the many families of polynomials, and one that plays a particular role in physics (they are known as Hermite polynomials<sup>62</sup> and appear in the description of the Hydrogen atom).

And here is the square root function  $\sqrt{x}$ :



We have highlighted the value  $\sqrt{2} \approx 1.41$ , as can be confirmed by eye inspection.<sup>63</sup> Note that in this case, the domain is restricted to  $\mathbb{R}^+$ .

We could extend it to the complex plane, but first we’ll have to find a good way to represent complex numbers, which is tricky, as they are two-dimensional, so we leave this for later (actually for next year!)

Another fundamental algebraic operation which we have defined (ultimately) on real numbers is the power of a number,  $a^x$  is  $a$  multiplied to itself  $x$  times when  $x \in \mathbb{N}$  and for  $x = p/q$  when in  $\mathbb{Q}$ , this is  $\sqrt[q]{a}$  multiplied itself  $p$  times. We can *interpolate* (which means, so-far still in an intuitive way, “extends” or “stitch smoothly”) for real values and now look at this operation as a function in itself, the so-called “*exponential*” functions, here shown for  $2^x$  (red) till  $3^x$  (blue) with intermediate case  $(2 + k/10)^x$  for integer  $1 \leq k \leq 9$  in thin, also-interpolating, colors:

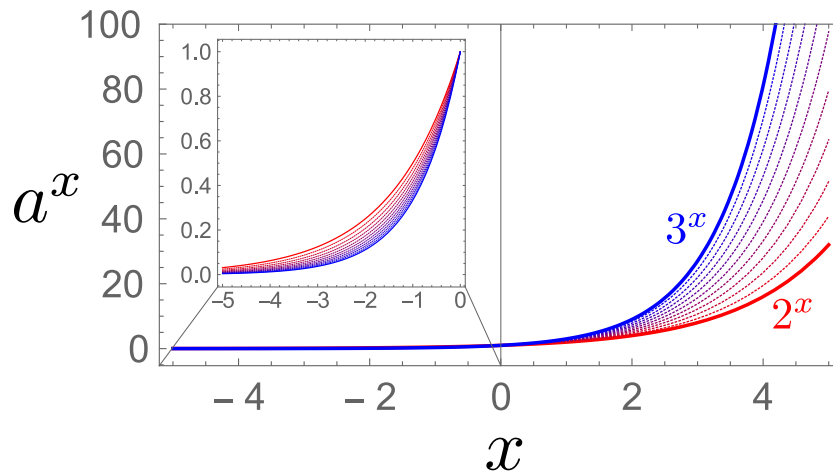
<sup>61</sup> Polynomial equations of order 2 yield the quadratic equation. Those of order 3 the cubic equation (etc.) Using graphic methods, explain how a quadratic can have two, one or no real solutions. A cubic equation always have at least one real root. “Prove” this using a graphical method.

<sup>62</sup> We will see in quantum mechanics how Hermite polynomials satisfy several properties which make them relevant in theoretical Physics. With the tools we have available so far, we will limit to state the so-called “recurrence” relation they satisfy, meaning that one order  $n$  depends on smaller orders, in this case, two orders below. Namely, check that

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x) \quad (213)$$

and then “predict” the form of  $H_6(x)$ .

<sup>63</sup> From this graph, give an estimate of  $\sqrt{7}$ . Compare with a calculator.



We have zoomed in the caption the negative part, where we see that, this time,<sup>64</sup> for  $a > b$ :

- If  $x > 0$ ,  $a^x > b^x$ ,
- If  $x < 0$ ,  $a^x < b^x$ .

These functions are very important and we will come back to them not only in this course but in all other courses of Physics. At this stage there is little more to say than to observe how they grow very rapidly.<sup>65</sup>

As part of our Mathematical baggage, we will have to collect a lot of functions, be familiar with them, know their properties, etc. Where do these functions come from? One possible way is to blindly provide their graph (i.e., for all  $x$ , give  $f(x)$ ).<sup>66</sup> More insightful is to associate to them a mechanism to construct or compute the value of the function to each possible value it can take. This is particularly important to the Physicist who is the person who brings “meaning” through “interpretation” so we are often interested in describing the inner mechanism of a function, rather than take it at face value that this is really a binding mechanism (between a variable and its value through a function).

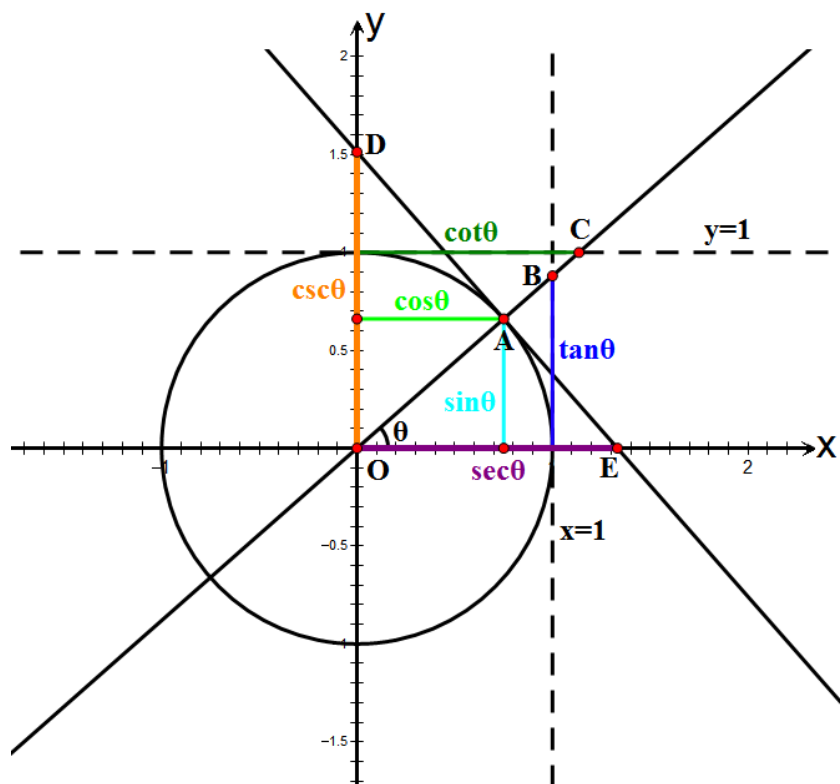
For instance, the trigonometric functions, are defined as distances projected on the axes of the trigonometric circle, as defined below:

<sup>64</sup> For a given  $a \in \mathbb{R}^+$ , how do  $a^{x_1}$  and  $a^{x_2}$  compare for  $x_i \in \mathbb{R}$ ,  $i = 1, 2$ ?

<sup>65</sup> From the graph, whose domain is  $[-5, 5]$  and codomain  $[0, 100]$ , give an estimate for  $y^5 = 100$ . Note that, by definition, this is also  $\sqrt[5]{100}$  so you can compare with a calculator. It is difficult to see that from the graph which  $y$  axis is bounded but from the graph (not the calculator) how much would you say is  $3^3$ ? (and then you can compute)

<sup>66</sup> This is already very useful. Keep investigating: from the graph, locate  $2^3$  and  $3^2$  and provide estimates for  $2.7^{\pi}$  and  $-2.5^{2.5}$ . Compare with the exact values.





From geometry, one can already derive interesting (and important) relationships between these functions, for instance:

- Pythagoras:  $\sin^2 \theta + \cos^2 \theta = 1$ .
- Thales:  $\sin \theta / \cos \theta = \tan \theta$ .

and more to be reviewed when we turn to trigonometry. You are invited to take the time it takes to work out the problem below, to understand the difference between a definition and a property:

#### TRIGONOMETRIC FUNCTIONS

From the definition of  $\sin$ ,  $\cos$ ,  $\tan$ ,  $\sec$ ,  $\cot$  and  $\csc$  in the trigonometric circles given above (as projections of various distances on various axes), plot “experimentally” (that is, measuring directly with a ruler on the plot, all these functions — *no calculator or computer allowed, only pen, paper, compass and ruler*). The more points you take, the better should be the resolution. As this is an experiment, you can actually place error bars. Would you know how to?

Now, directly from the graphs themselves, compute  $\sin^2 \theta$ ,  $\cos^2 \theta$ ,  $\sin^2 \theta + \cos^2 \theta$  and  $\sin \theta / \cos \theta$ .

We can also provide an explicit formula to define a function. Here is an example which is important in Physics, the so-called Lorentzian:

$$l(x) = \frac{1}{1+x^2}, \quad (214)$$

You can check that  $l(0) = 1, l(1) = 1/2, l(2) = 1/5$ , etc., out of which you can easily produce the function's graph.<sup>67</sup> This is a so-called "closed-form" expression, because the result is given directly by a formula, no need to make fancy constructions or sophisticated computations. So far we have little tools to build new functions but we will acquire them.

<sup>67</sup> What does its graph look like?

Here are more involved examples based only on things we have already seen:

$$f_n(x) = \left( \sum_{k=0}^n \frac{1}{k!} \right)^x, \quad g_n(x) = \sum_{k=0}^n \frac{x^k}{k!}, \quad h_n(x) = \left( 1 + \frac{x}{n} \right)^n. \quad (215)$$

These functions involve a parameter  $n \in \mathbb{N}$ , or we could say that have introduced a family or a series of functions,  $f_1, f_2$ , etc. These are still, technically, closed-form, because although the formula are more complicated in their full-fledged form, they are still basic expressions, e.g.,<sup>68</sup>

$$f_3(x) = \left( 1 + 1 + \frac{1}{2} + \frac{1}{6} \right)^x, \quad (216a)$$

$$g_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}, \quad (216b)$$

$$h_3(x) = \left( 1 + \frac{x}{3} \right)^3. \quad (216c)$$

<sup>68</sup> Compute these for  $x \in \{0, \pm\frac{1}{2}, \pm 1\}$  and plot them.

It starts to become a different concept when we ask what happens to these functions as we consider higher and higher values of  $n$ . It is particularly clear for  $f_n$  and  $g_n$  that as  $n$  increases, the new terms bring in smaller and smaller corrections. For instance, the definition of  $f_n$  is that this is the exponential of a constant

$$e_n \equiv \sum_{k=0}^N \frac{1^k}{k!} \quad (217)$$

which we can easily, although tediously (not with a computer), com-

pute:

$$e_0 = 1, \quad \text{Remember that } 0! = 1 \text{ and } 1^0 = 1 \quad (218a)$$

$$e_1 = 1 + 1 = 2, \quad (218b)$$

$$e_2 = 1 + 1 + \frac{1}{2} = 2.5 \quad (218c)$$

$$e_3 = 1 + 1 + \frac{1}{2} + \frac{1}{6} = 2.666\dots \quad (218d)$$

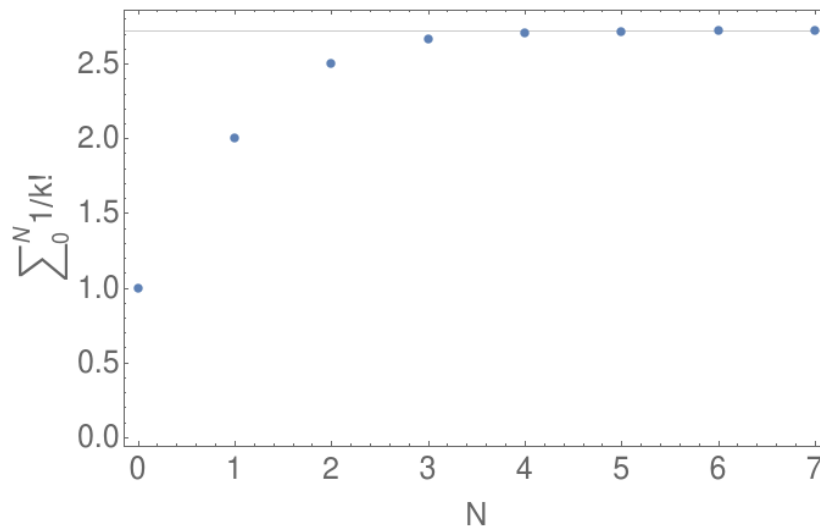
$$e_4 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} = 2.70833\dots \quad (218e)$$

$$e_5 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = 2.71667\dots \quad (218f)$$

$$e_6 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} = 2.71806\dots \quad (218g)$$

$$e_7 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \frac{1}{5040} = 2.71825\dots \quad (218h)$$

we will stop here, because even though the number still change as we add more terms to the sum, this is for decimals so far away that this is not visible on a plot of the function anymore:



This phenomenon is called *convergence*. We will study it in detail in a coming lecture. For now, we take it as a practical way to construct new objects. We will call *limit* the process of dealing with a version of  $e_n$  with  $n$  large enough so that we don't have any effect from any particular choice of  $n$ . In practical terms, any operation we undertake with  $e_n$ , if the result change when we take a larger value for  $n$ , then we take this new value, and check again with a larger  $n$ , until the value does not change anymore. In which case we say that the result has converged. We can then get rid of  $n$  since, by definition, it does not play a role, and simply write  $e$ . We could also

write  $e_\infty$  but that's one symbol more and one which we did not introduce yet. This stands for "infinity" and means "as large as needed".

We can now look at  $f, g, h$  defined in this way:

$$f(x) = \left( \sum_{k=0}^{\infty} \frac{1}{k!} \right)^x, \quad g(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad h(x) = \lim_{n \rightarrow \infty} \left( 1 + \frac{x}{n} \right)^n, \quad (219)$$

where in the latter case, for  $h$ , we introduce a special notation to draw our attention on the "limiting" process that achieves  $h$  by a limitless increase of  $n$  (we cannot simply substitute  $\infty$  in the formula as it appears in several places and then the order in which it is to be considered becomes ambiguous; on the other hand, it doesn't hurt to simplify

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \equiv \sum_{k=0}^{\infty}. \quad (220)$$

Now, in Eqs. (219), we have non-closed form for three functions.

That requires computation, evaluation, even maybe the complicated recipes involved are not possible to get through or, in the worst cases, not even well-defined (not converging). Here, all quantities converge, in particular, the quantity

$$e \equiv \sum_{k=0}^{\infty} \frac{1}{k!} \quad (221a)$$

$$\approx 2.7182818284590452353602874713526624977572470936999 \dots$$

$$(221b)$$

is well defined and is an important quantity, which is known as the "Euler number" although we simply call it "e". It is less famous, but equally important than  $\pi$  for reasons that will become apparent later. Incidentally,  $e \in \mathbb{R} - \mathbb{Q}$ .

For now, since we focus on functions rather than on this number, we will look at the beautiful, important and surprising result:

$$f = g = h. \quad (222)$$

Incidentally, two functions  $f$  and  $g$  are equal iff for all  $x$ ,  $f(x) = g(x)$ . In which case we do not need to name the variable, and can equate directly the functions, as shown above. We can then also start to contemplate a "space of function", say  $\mathbb{F}$  (this notation is not standard), on which we can define an algebra of function, e.g., for  $f, g \in \mathbb{F}$ ,  $f + g$  and  $fg \in \mathbb{F}$ . We will come back to this important observation later. For now, it is enough to focus on this remarkable fact that we have three—clearly very different—mechanisms to provide values out of the variable  $x$ , that yield one and the same function. This function is, from the first ( $f$ ) definition, an exponential, actually some function very close to the  $2.7^x$  graph above. The official name is the *natural*

exponential but we simply call it the “exponential”  $\exp(x)$ . Therefore, we have:

$$\exp(x) = e^x, \quad (223)$$

and

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad (224)$$

as well as<sup>69</sup>

$$\exp(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n. \quad (225)$$

Note that in the three above equalities,  $\exp$  on the lhs is just a name or label for this (important) function, while the expressions on the right are formulas, mechanisms to define how the function attach a given value to any input value of its variable.

We will demonstrate Eq. (222) first by showing that  $h = g$  and then that  $h = f$ . This will thus establish that  $f = g$ . In both cases we take  $h$  as our starting point as the Binomial expansion is a good starting point to reach our targets. Indeed, in the first case, we write:

$$\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \frac{x^k}{n^k} \quad (226a)$$

$$= 1 + n \frac{x}{n} + \frac{n(n-1)}{2} \frac{x}{n} \frac{x}{n} + \frac{n(n-1)(n-2)}{3!} \frac{x}{n} \frac{x}{n} \frac{x}{n} + \dots \quad (226b)$$

$$= 1 + x + \frac{x}{2} \frac{n-1}{n} + \frac{x^3}{3!} \frac{n-1}{n} \frac{n-2}{n} + \dots \\ \dots + \frac{x^k}{k!} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} + \dots \quad (226c)$$

and we do not have to worry at writing the full sum as  $n$  gets arbitrarily large so we can focus on the first elements, which will always read as such regardless of our choice of  $n$ . There is a part which is  $n$ -independent, of the type  $k^k/k!$ , and then there is a product of terms  $\frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n}$  which can be made as close to 1 as we want, so in a limiting process, we are left with

$$\left(1 + \frac{x}{n}\right)^n \rightarrow \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (227)$$

where the arrow is a shortcut for  $\lim$ , and with, by the way, for the particular case  $x = 1$

$$\left(1 + \frac{1}{n}\right)^n \rightarrow e \quad (228)$$

by definition (cf. Eq. 217). We will use this in our second proof, for which we first rewrite  $h_n$  as

$$\left(1 + \frac{x}{n}\right)^n = \left(1 + \frac{1}{\frac{n}{x}}\right)^{\frac{n}{x}x} = \left[\left(1 + \frac{1}{v}\right)^v\right]^x \quad (229)$$

<sup>69</sup> Before we demonstrate Eqs. (223–225), it may be useful or cautious to convince ourselves that this looks correct numerically. Have a look.

where, since  $x$  is fixed in principle,  $v \equiv n/x$  can be taken also as essentially a large integer. For instance, assume that  $x$  is rational and thus of the type  $x = p/q$  with  $p, q \in \mathbb{N}$ , then  $v = nq/p$  and we can take as our series of  $n$  multiples of  $p$  so that  $v$  will now be exactly a large integer. For irrational  $x$  we would have to take another limit, something which we do not want to enter into right now so we assume it will be enough to get as close as we want to any real value. For indeed, also using the fact that the power of a limit is the limit of the power, since  $(1 + 1/v)^v \rightarrow e$  (cf. Eq. 228), then  $[(1 + 1/v)^v]^x \rightarrow e^x$ , QED.

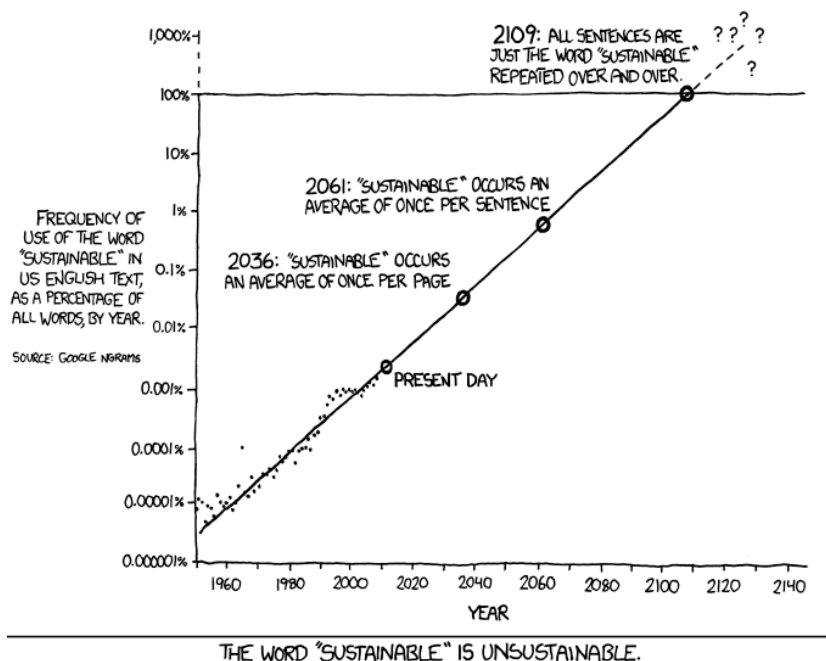
We have, to be sure, used slightly less rigorous Mathematical standards than are possible, but in fact our demonstration is in line with the types of arguments used by Euler himself. What matters most to us is a good grasp of the exponential function and how it arises in different forms, and how those are ultimately connected. The exponential is one of the most important function of Physics. Start studying it. We'll meet it all the time. And with it, we are back to the Platonic universe. Here we have (at least) three different ways to define some object which appears to exist regardless of which rule we settle on to compute it. This is a recurrent theme.

### *Further problems*

### *Meaning of functions*

Explain this joke<sup>70</sup> (explaining a joke usually spoils it to those who listen, but it may bring additional enjoyment to those who speak).

<sup>70</sup> <https://xkcd.com/1007>



### Domains of functions

When you are given a function, often, you are not given its domain and codomain, and it is then for you to find out.  $\sqrt{x}$  for instance is not defined over  $\mathbb{R}$ , unless we want to go into the complex plane, but let us assume it is a real-valued function of a real variable. You then have to find  $\mathbb{A}$  and  $\mathbb{B}$ . Do that for this and the other functions defined in the text, and check by plotting them.

### The Natural Exponential

Check numerically that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad (230)$$

not only for  $x = 1$  but for all  $x$ . Plot your approximations.

### Funny functions

Here are some definitions for functions we shall occasionally use, which we leave for you to play with: The ceil  $\lceil x \rceil$  and floor  $\lfloor x \rfloor$  functions are defined from  $\mathbb{R} \rightarrow \mathbb{Z}$  and give the least/greatest integer less than or equal to  $x$ , respectively, e.g.,  $\lceil \pi \rceil = 4$  and  $\lfloor \pi \rfloor = 3$ . Plot them. Plot  $\lfloor x^2 \rfloor$ .

The Heavyside function  $\Theta$  from  $\mathbb{R}$  to  $\{0, 1\}$  is defined as:

$$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases} \quad (231)$$

Plot it. Plot  $\Theta(\sin(x))$ .

The sign (sgn) function is defined from  $\mathbb{R}$  to  $\{-1, 0, 1\}$  as

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \quad (232)$$

Plot it and check that  $\text{sgn}(x) = x/|x|$  (except for  $x = 0$ ). Show that  $x = \text{sgn}(x) \cdot |x|$ ,  $\text{sgn}(x^n) = \text{sgn}(x)^n$ ,  $\text{sgn}(x) = 2\Theta(x) - 1$ . Show that:

$$\text{sgn}(x) = \left\lfloor \frac{x}{|x| + 1} \right\rfloor - \left\lfloor \frac{-x}{|-x| + 1} \right\rfloor. \quad (233)$$

### More polynomials

While all polynomials are of the form (211) and that encompasses all the possible cases, particular types of polynomials, such as the Hermite ones above, sharing a particular structure or satisfying given properties, form families of particular interest and importance. We will meet several such families (Legendre polynomials, Bernstein polynomials, etc.) Here is another friendly family, the Fibonacci polynomials, defined recursively as:

$$F_0(x) = 0, \quad F_1(x) = 1, \quad (234)$$

and then for higher  $n \in \mathbb{N}$ :

$$F_n(x) = xF_{n-1}(x) + F_{n-2}(x). \quad (235)$$

Write down the first few Fibonacci polynomials. You might know about the Fibonacci numbers, that are related to a problem of rabbits proliferation. These are obtained as  $F_n(1)$ . Write them down and see if you spot their pattern (if you don't know it yet). Another, less known, series is the so-called Pell numbers, related to the closest rational approximations of  $\sqrt{2}$ . These are obtained as  $F_n(2)$ . Write them down. As a final exploration of these polynomials which, unlike other families, we will not regularly meet again, consider the coefficients  $\alpha_k^{(n)}$  of the  $k$ th power in the  $n$ th order Fibonacci polynomial, i.e., let us define

$$F_n(x) = \sum_{k=0}^n \alpha_k^{(n)} x^k. \quad (236)$$



Then it can be shown that  $a_k^{(n)}$  is the number of ways of writing  $n - 1$  as an ordered sum (meaning that  $1 + 2$  is counted separately from the otherwise identical  $2 + 1$ ), involving only 1 and 2, so that 1 is used exactly  $k$  times. For example  $a_3^{(6)} = 4$ , so that 5 can be written in four ordered ways with only 1 and 2, where 1 appears three times, namely:

$$1 + 1 + 1 + 2, \quad 1 + 1 + 2 + 1, \quad 1 + 2 + 1 + 1, \quad 2 + 1 + 1 + 1. \quad (237)$$

Using this “rule”, find how many ordered sums can be written for the number 7 that involve exactly two 2 with the rest consisting of 1. In how many ways can 7 be written with any number of 1 and 2? Check by independently listing all the possibilities.

### *Bell-shaped functions*

We will meet repeatedly in Physics two types of “bell-shaped” functions, defined (and named) as follows:

- The Gaussian function:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right). \quad (238)$$

- The Lorentzian function:

$$h(x) = \frac{1}{\pi\gamma} \left[ \frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right]. \quad (239)$$

What is the meaning of  $x_0$  and  $\sigma$  (Gaussian),  $\gamma$  (Lorentzian)? Compare these two functions by plotting them. Do you agree that they look qualitatively the same? Yet, we shall see that they behave in incredibly different ways!

### *A function in Optics*

In Optics, we have seen one *law of Physics*, the law of refraction, that relates the angle  $\theta_i$  impinging at the interface of two transparent media with indices  $n_i$ ,  $1 \leq i \leq 2$ :

$$n_1 \sin \theta_1 = n_2 \sin \theta_2. \quad (240)$$

This law comes here in the form of an *equation*. It can turn into a function as we extract information out of it. Namely, let us ask the most natural question that can be formulated from this Law: how is light refracted? This is better answered by plotting the refracted angle  $\theta_2$  as a function of the incident angle  $\theta_1$ . What are the expected (possible) domains and codomains in this case? Plot both cases  $n_1 > n_2$  and  $n_2 > n_1$  (the case  $n_1 = n_2$  is trivial, but address it too). You will see that either the actual domain or the actual codomain change. The restricted codomain results in an important physical phenomenon. Which one?

*A function in Mechanics*

In Mechanics, we find the trajectory  $(x(t), y(t))$  of a projective with initial velocity  $\vec{v}_0 = v_0 \cos(\theta)\hat{i} + v_0 \sin(\theta)\hat{j}$  as:

$$x(t) = v_0 t \cos(\theta), \quad (241a)$$

$$y(t) = v_0 t \sin(\theta) - \frac{1}{2}gt^2, \quad (241b)$$

with  $g$  the acceleration of gravity (and assuming the floor to be at  $y = 0$ ). From these equations, plot the trajectories of an object thrown at various angles (sideway, upward, etc.) Does this match with your physical intuition of what should happen?

Check from your plots that the maximum height of the projectile is given by  $h = v_0^2 \sin^2(\theta)/(2g)$ . Find, also from the plots, the horizontal distance traveled by the projectile when it hits the floor. Check that it is given by this formula:

$$d = \frac{v \cos \theta}{g} \left( v \sin \theta + \sqrt{(v \sin \theta)^2 + 2gy_0} \right). \quad (242)$$

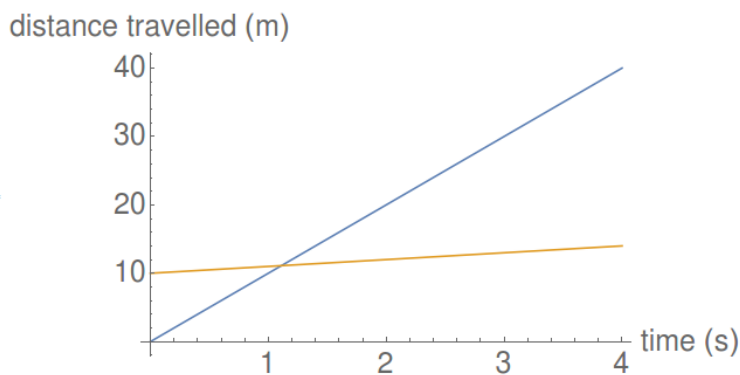
Now seeing  $d$  as a function of  $\theta$ , at which angle should you send a projectile so as to maximize its landing distance? Note that this depends on  $y$ . In which way, and why?

## Lecture 7: Infinites

Zeno was a pre-Socratic philosopher known for a series of deep and subtle paradoxes, of which Achilles chasing the turtle is the most famous. It is meant to show that speed and movement are illusory, in the form of a fast runner (Achilles) who can never catch up with a much slower runner (the turtle), if the latter starts with a head up. The paradox reads as follows. Imagine that Achilles runs  $10 \text{ m s}^{-1}$  while the turtle runs only  $1 \text{ m s}^{-1}$ , but starts with a 10 m head up.

Achilles then runs until it arrives to the point where the turtle was initially. This he achieves in 1 s (the time it takes to cover 10 m if you go at a speed of  $10 \text{ m s}^{-1}$ ). The point is, during that interval of time, the turtle has also moved forward, namely, it has progressed by 1 m, so still with a lead on Achilles, who then has to run to this new distance, which he achieves in 0.1 s, but, here again, the turtle will have progressed in this time, namely, of 10 cm, which Achilles will have to cover (in 0.01 s), and so on: in the interval of time it takes to Achilles to reach the situation previously held by the turtle, the latter has moved further on, thus perpetually escaping Achilles.

It is clear that Achilles will catch and also overtake the turtle, for instance from a plot of the functions of their distance at any given time, with Achilles in blue and the turtle in orange.



The point where the two lines intersect is where/when Achilles and the turtle meet (same point in space at the same time). When the

blue line is above the orange line, Achilles has overtaken the turtle. As one can see, their relative distance keep increasing from this point onward.

This is also obvious from the equations that describe this problem:

$$\begin{cases} d_1 = 10t, \\ d_2 = t + 10. \end{cases} \quad (243)$$

We can find the time when they meet as the one that satisfies  $d_1 = d_2$ , which we can easily solve as

$$t = \frac{10}{9}. \quad (244)$$

So how does this fit with Zeno's paradox? From the fact that an infinite sequence of time windows can add up to a finite time. Indeed, the time needed to reach the common point is

$$t_{\text{meet}} = \sum_{k=0}^{\infty} \left(\frac{1}{10}\right)^k. \quad (245)$$

The sum in Eq. (245) is called a *geometric series*. We will compute it by taking the limit of the partial sum

$$S_N \equiv \sum_{k=0}^N p^k = 1 + p + p^2 + p^3 + \dots + p^N. \quad (246)$$

Now there is a trick here:

$$\begin{aligned} & (1 + p + p^2 + p^3 + \dots + p^N)(1 - p) \\ &= \left(1 + p + p^2 + p^3 + \dots + p^N\right) - \left(p + p^2 + p^3 + \dots + p^{N+1}\right) \end{aligned} \quad (247)$$

by distribution. If we simplify terms that appear twice, we find that

$$1 - p^{N+1} \quad (248)$$

are the only one left (leftmost and rightmost from the respective parentheses). As a conclusion, bringing Eqs. (247) and (248) together, we find:

$$S_N = \frac{1 - p^{N+1}}{1 - p}. \quad (249)$$

Let us now take the limit, that is, let us see if  $S_N$  converges ("stabilizes") as we take  $N$  larger and larger. If  $p > 1$ , it doesn't, as this is clearly something that becomes bigger and bigger, so this explodes. If  $p = 1$ , we have the problem of dividing by zero, but also multiplying by zero, and the quantity is not well defined (if we come back to what it means, we are trying to compute  $1 + 1 + 1 + 1 + 1 + \dots$ , which is probably infinite). However, if  $p < 1$ ,  $p^{N+1}$  becomes always

smaller and smaller, until its effect on 1 becomes negligible, so we can just drop it. Therefore:

$$\lim_{N \rightarrow \infty} S_N = \frac{1}{1-p}. \quad (250)$$

This is a very important case of a series, that you will need to remember:<sup>71</sup>

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}. \quad (252)$$

Back to Achilles, now we can compute Eq. (245) and find  $t_{\text{meet}} = \frac{1}{1-(1/10)} = 10/9$ , as found previously. This is the resolution of the paradox: an infinite amount of contributions can pile up to a finite amount. This is also the limit of Greek's power of penetration of Mathematical concepts: handling the infinity. A proper command of it started with Newton who developed calculus.

For Achille's problem (at which time or at which point does he catch the turtle), this is an overkill to think of the problem in this way. But in many problems of Physics, such infinite sums are often invoked, say that we compute a field, that affects the charges that bath in the field, which in turn change the field, which in turn change the charges, etc., this can be computed exactly by summing over an infinite number of terms of actions and back-actions, in a way that would not be so straightforward otherwise. In quantum field theory, we also describe events as a series of Feynman diagrams that each correspond to a particular interaction and that add up to the exact physical process. So infinite sums will be part of our toolkit and we must know about them.

Before we start manipulating infinite sums or any type of infinity, we must understand better the nature of infinity. The sum in Eq. (252) is over integers of the function  $f(k) = p^k$ . Could we sum this function not only for the values it takes on  $\mathbb{N}$ , but also on  $\mathbb{Q}$  or  $\mathbb{R}$ . How would we go about this? You might have seen previously that summing over "continuous lines" rather than over "discrete numbers" bring us to replace the  $\sum$  by a  $\int$ . Here we have to ask ourselves up to which point, if at all, infinity becomes "too much". How big is really  $\mathbb{N}$  or  $\mathbb{Q}$ , etc.? The good way to think about it, is a fairly recent discovery (19th century) and comes from Cantor, who found the beginning of the thread, and by pulling it, came to the conclusion that one can count infinities, and that there are an infinite number of different types of infinities. "I see it but I don't believe it", was his first reaction.

Here are some of the statements from Cantor, which we first give to ponder about, and then we prove them (once we reveal "the good way to think about it"):

<sup>71</sup> As well as their most immediate variants. Compute the following series:

$$S_1 = \sum_{k=0}^{\infty} (p^n)^k, \quad \text{for } n \in \mathbb{N}, \quad (251a)$$

$$S_2 = \sum_{k=0}^{\infty} (-1)^k p^k = 1 - p + p^2 - p^3 + \dots \quad (251b)$$

- There are as many natural numbers (integers) than odd numbers.
- There are as many integers than primes.
- There are as many integers than rational numbers.
- There are more real numbers than rational numbers.

How much of this do you buy? How can it be that  $1, 2, 3, 4, 5, \dots$  up to infinity has the same “size” than the same list of which we drop every over element,  $2, 4, 6, \dots$ , and that both have the same “size” than the list of primes, “ $2, 3, 5, 7, 11, 13, 17, \dots$ ” (with only 7 numbers at this stage of contig when the “full list of integers” has 17). One should not stop, though, and consider the full set.

Let’s now see the trick. How to count things? The big insight of Cantor is that a good definition when dealing with infinities is for two sets to have the same numbers of elements if there exists a bijection between them.

This is clear (and in fact trivial) for sets with finite numbers of elements. There are the same number of boys and girls in the dance hall if they can all be paired together in a way that nobody is left out and no couple has more than two people. This is also how children count things, by bringing them in a bijection with their fingers. But this nicely extends to infinite sets. Two (possibly infinite) sets have the same number of elements if one can find a one-to-one mapping (bijection) between them. For instance, that proves the first point above, because  $n \rightarrow 2n$  is bijective.

1) The “multiply by 2” function  $m_2$  is injective. We know from last lecture that  $m_2$  is injective iff  $n \neq m \implies m_2(n) \neq m_2(m)$ . But if  $n \neq m$ , then clearly  $2n \neq 2m$  so  $m_2(n) \neq m_2(m)$ , QED. Or the other way around, if  $2n = 2m$ , then, simplifying by 2, we get  $n = m$ , which is QED again as we started from  $m_2(n) = m_2(m)$ .

2) The “multiply by 2” function  $m_2$  is surjective. In our case, that means that any number  $2n$  comes from some  $m_2(k)$  for an integer  $k$ , but  $m_2(k) = 2k$ , and clearly there is a  $k$  such that  $2k = 2n$ , namely,  $k = n$ . QED.

As you see, the arguments are trivial (in this case).<sup>72</sup> In any way, that shows that  $\mathbb{N}$  and  $2\mathbb{N}$  (or, clearly, any  $k\mathbb{N}$  for any  $k \in \mathbb{N}$ ) have the same “size”, or, as we say in set theory, “cardinality”. It is a bit counter-intuitive because it means, for instance, that one can remove elements from an infinite set, and it is still of the same size as before. If you think about it, this is reasonable. You can remove a handcup of water from the sea, you did not really change the size or the amount of the sea. And the sea is not infinite, it is still huge. The larger is the sea, the more insignificant is your withdrawal. If it is infinite, its size remains exactly the same. A bit more surprising is that you can

<sup>72</sup> Prove that  $\#\mathbb{N} = \#\mathbb{Z}$ .

remove an infinite number of elements from an infinite set, and it can still remain infinite. But here again, intuition suggests that, not how much you remove, but how much you leave, matters.

This cardinality for the smallest type of infinite has the name “Aleph zero”, after the first letter of the Hebrew alphabet:

$$\boxed{\#\mathbb{N} = \aleph_0}. \quad (253)$$

A first genuine surprise, possibly, is that there is no smaller infinite set than the integers (or any other set that can be mapped to it bijectively). Indeed, for any infinite set  $\mathbb{A}$  of cardinality at most  $\aleph_0$ , we can define a mapping  $f$  by picking up one element  $a_1 \in \mathbb{A}$  and define  $f(1) = a_1$ , then pick up another element  $a_2 \in \mathbb{A}$  and define  $f(2) = a_2$ , etc., this process never ending since  $\mathbb{A}$  is infinite. That defines an injection from  $\mathbb{N}$  to  $\mathbb{A}$  and as a result  $\#\mathbb{N} \leq \#\mathbb{A}$  for all  $\mathbb{A}$  which proves that  $\#\mathbb{N}$  is the smallest infinite. In particular, all subsets of  $\mathbb{N}$  are finite or countable. Indeed, any such set  $\mathbb{A}$  can be ordered  $\mathbb{A} = \{a_1, a_2, \dots\}$  with  $i < j \implies a_i < a_j$ . Clearly, unless the set is finite, the function  $f(a_i) = i$  is a bijection. Since there is an infinite number of primes, as a corollary, the cardinality of primes is  $\aleph_0$ . In essence, any set that can be counted by ordering its element in the order of first, second, etc., is countable. Indeed, if there is a bijection, then one can “count” the set with fingers (just an infinite number of them). For example, the sieve of Eratosthenes is a mechanism to build an explicit bijection between primes and integers.

This also allows us to prove that the cardinality of  $\mathbb{Q}$  is  $\aleph_0$ . Let us consider positive rationals. We can make up unambiguously this table:

$$\begin{array}{cccccccc}
 & 1 & 2 & 3 & 4 & \dots & q & \\
 1 & (1,1) & (1,2) & (1,3) & (1,4) & \dots & (1,q) & \dots \\
 2 & (2,1) & (2,2) & (2,3) & (2,4) & \dots & (2,q) & \dots \\
 3 & (3,1) & (3,2) & (3,3) & (3,4) & \dots & (3,q) & \dots \\
 & \vdots & \dots & \dots & \ddots & \dots & \vdots & \vdots \\
 p & (p,1) & (p,2) & (p,3) & (p,4) & \dots & (p,q) & \vdots \\
 & \vdots & \dots & \dots & \dots & \ddots & \vdots & \vdots
 \end{array} \quad (254)$$

which contains all possible  $p/q$  combinations. From this we can reconstruct a one-to-one mapping with all rationals, by transforming  $(p, q)$  into  $p/q$  provided that  $\gcd(p, q) = 1$ . If it is not, this means that we do not have the irreducible (unique) representation of the rational, so we just skip this entry. Then we walk through the table of existing entries (those that have not been skipped as reducible forms) by following successive diagonals  $p + q = n$  for  $n \in \mathbb{N}$ . That is, we

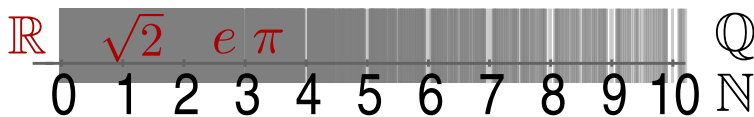
have this mapping:

$$\begin{array}{cccccc}
 1 \rightarrow 1 & 3 \rightarrow 1/2 & 5 \rightarrow 1/3 & 9 \rightarrow 1/4 & 11 \rightarrow 1/5 & \\
 2 \rightarrow 2 & (\text{skipped}) & 8 \rightarrow 2/3 & (\text{skipped}) & 16 \rightarrow 2/5 & \\
 4 \rightarrow 3 & 7 \rightarrow 3/2 & (\text{skipped}) & 15 \rightarrow 3/4 & 20 \rightarrow 3/5 & \\
 6 \rightarrow 5 & (\text{skipped}) & 14 \rightarrow 4/3 & (\text{skipped}) & 25 \rightarrow 4/5 & \\
 10 \rightarrow 5 & 13 \rightarrow 5/2 & 19 \rightarrow 5/3 & 23 \rightarrow 5/4 & (\text{skipped}) & 
 \end{array} \quad (255)$$

Etc., the bijection (let us call it  $\mathcal{C}$ , from  $\mathbb{N}$  to  $\mathbb{Q}^+$ ) is now well defined, e.g.,  $\mathcal{C}(25) = 4/5$  (from the table above). It is clearly injective (two different integers correspond to two distinct elements in the table) and surjective (each ratio  $p/q$  is sure to be found somewhere). Note that we took only positive rationals but it is straightforward to extend the argument to all rationals (including negatives). For instance one can map one-to-one  $\mathbb{Z}$  to  $\mathbb{Q}$  and since  $\mathbb{N}$  is trivially mapped one-to-one to  $\mathbb{Z}$ , we can write down the big result:<sup>73,74</sup>

$$\boxed{\#\mathbb{N} = \#\mathbb{Q}} \quad (256)$$

That's surprising, certainly, but a great result. There are no more integers than rationals. This is surprising because integers are neatly spaced, so one can imagine counting them with one's hand (provided we have enough fingers, but we can borrow some, or start counting with our feet), while rationals are all over the place. This is the set of all rationals obtained by the process above from a table with  $p$  and  $q$  each going up to 100. Of the  $100^2$  possible numbers, only 6087 are reducible (the computer tells us that), and that's how many lines you get below. There are still visible gaps around the largest integers, but these would fill up too as we increase  $p$  and  $q$ , and we would get an appearance of continuity or smoothness. But this is an illusion,  $\mathbb{Q}$  is actually full of gaps, or holes, it is like a very fine mesh but, incredibly enough, of "countable" numbers, meaning one can count them in some order! (the 1st, the 2nd, etc.)



We have also put on this axis some of the numbers we know are not of the form  $p/q$ , i.e., that are not in the table, however big it is (or, for that matter, infinite). Even more surprising, the set of real numbers,  $\mathbb{R}$ , is *not countable* (or "uncountable"), meaning, there is no bijection between it and  $\mathbb{N}$ . This is surprising because this tells us that  $\mathbb{N}$  and  $\mathbb{Q}$  are much more alike than  $\mathbb{Q}$  and  $\mathbb{R}$ , while our intuition would seem to tell us the opposite!

<sup>73</sup> Write down the complete and detailed proof (with signs, etc.)

<sup>74</sup> Show that irrationals  $\mathbb{R} - \mathbb{Q}$  are not countable. Give an example of a countable set of irrationals.



The proof comes from what is known as Cantor's diagonal argument. It is as beautiful, elegant and simple than it is deep and important. Consider the set  $\mathcal{S}$  of infinite sequences of binary numbers, e.g.,

$$s_1 = (0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, \dots) \in \mathcal{S} \quad (257)$$

is such a sequence and we can label it as  $s_{11} = 0, s_{12} = 0, s_{13} = 1$ , etc. What Cantor shows with his diagonal argument, is that, regardless of the ordering one makes, there is always at least one sequence  $s_0$  which is not in any countable set of  $\mathcal{S}$ . The proof is by construction of such a sequence, namely, let us define:

$$s_0 = (\bar{s}_{11}, \bar{s}_{22}, \bar{s}_{33}, \dots) \quad (258)$$

where  $\bar{0} = 1$  and  $\bar{1} = 0$ . Clearly  $s_0 \in \mathcal{S}$ . Yet it is nowhere to be found in an ordered list because it is different from all possible members of the set of  $s_k$ . This is easy to show. Two elements of  $\mathcal{S}$  are identical if all their digits are the same. Consider then that  $s_0$  is the  $N$ th element (it has to be somewhere) of an ordered list. This clearly cannot be, since we defined  $s_{0N} = \bar{s}_{NN}$  and thus  $s_0 \neq s_N$  since their digits do differ in at least one place. This shows that it does not matter how we order the elements of  $\mathcal{S}$ , we cannot do it in a countable way. There are more elements that can be counted (by the integers). The set  $\mathcal{S}$  can be mapped one-to-one with  $[0, 1]$  (binary expansion), so this basically proves that  $[0, 1]$  and thus  $\mathbb{R}$  is not countable since one can easily find bijection between these two sets.<sup>75</sup>

<sup>75</sup> Provide one.

It can be shown, on the other hand, that there are bijections between  $\mathcal{S}$ , or  $\mathbb{R}$ , with the "power set"  $\mathcal{P}(\mathbb{N})$ . The *power set* of a set  $S$  is the set of all its subsets (including the empty set  $\emptyset \equiv \{\}$  and  $S$  itself. For instance:

$$\mathcal{P}(\{x, y, z\}) = \{\emptyset, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\} \quad (259)$$

One can show<sup>76</sup> that if  $\#S = N$ , then  $\#\mathcal{P}(S) = 2^N$ .

<sup>76</sup> Show that  $\#\mathcal{P}(A) = 2^{\#A}$  for any finite  $A$ .

*Cantor's theorem:* For any set  $S$ , the cardinality of  $S$  is strictly less than the cardinality of  $\mathcal{P}(S)$ . This holds for infinite sets too. This shows that there are an infinite numbers of infinities:

$$\aleph_0, 2_0^{\aleph_0}, 2^{2_0^{\aleph_0}}, \dots \quad (260)$$

We will not prove this as we do not need such a general statement. We will content with this particular case:

*The powerset of integers  $\mathcal{P}(\mathbb{N})$  is not countable.* The proof is also based on a diagonal argument. To each subset  $A \in \mathcal{P}(\mathbb{N})$ , we associate  $g(A) \in [0, 1]$  constructed in base 2 in such a way that the  $n$ th digit of  $g(A)$  is 1 if  $n \in A$  and is zero otherwise. For example  $g(\{1, 3, 4, 5, 7\}) = 0.1011101$ .<sup>77</sup> We further write  $x_{[i]}$  the  $i$ th

<sup>77</sup> Which number is that in base 10 decimals?

digit of  $x$ , so  $g(\{1, 3, 4, 5, 7\})_{[3]} = 1$ . We now assume that  $\mathcal{P}(\mathbb{N})$  is countable, which means, by definition, that there exists a bijection  $f : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N})$  such that  $f(n) = A_n$  and we thus construct the number

$$x = 0.\overline{g(A_1)_{[1]}}\overline{g(A_2)_{[2]}}\cdots\overline{g(A_n)_{[n]}}\cdots \quad (261)$$

that has for first digit the complementary of the first digit of  $g(A_1)$ , for second digit the complementary of the second digit of  $g(A_2)$  and more generally,  $x$  has for  $k$ th digit the complementary of the  $k$ th digit of  $g(A_k)$ . Diagonal argument. Now there obviously exists  $X \in \mathcal{P}(\mathbb{N})$  such that  $g(X) = x$ . But, equally clearly, we cannot find  $n \in \mathbb{N}$  such that  $f(n) = X$  since by definition the  $n$ th digit of  $x$  will be the complementary of  $g(X)_{[n]}$ . This proves that there is no room in  $\mathbb{N}$  to make a surjection into  $\mathcal{P}(\mathbb{N})$ . Its cardinality is larger.

It can be proved (we don't here) that there is a bijection between  $\mathcal{P}(\mathbb{N})$  and  $\mathbb{R}$ . Actually, it has long been unknown whether there are smaller infinities than  $\#\mathbb{R}$  but larger than  $\aleph_0$ . The negative is known as the “continuum hypothesis” (i.e., there is no larger set than  $\mathbb{N}$  but smaller than  $\mathbb{R}$ ). It was later proven that this is undecidable, i.e., cannot be proved by axiomatic methods but needs to be formulated as a possibility, or not, to yield different version of the Platonic universe. This is beyond our interest. Our interest—and what we have proved—is really that:

$$\aleph_0 \neq 2_0^{\aleph_0} \quad (262)$$

i.e., the integers and rationals are infinite in a different way than real numbers are. Namely, this brings for us two important concepts:

- Discrete quantities (finite or infinite).
- Continuum.

The real numbers are like a continuous fluid, they have no gaps, they are smooth. This will require quite a different treatment for them.

We will repeatedly come back to deal with infinities. In Physics, they play a considerable role, in the form of *singularities*, starting with the big bang. We will also see how to classify infinities thanks to complex numbers, and also why certain quantities diverge with no apparent reasons. For instance, the infinities in the series (252) is quite clear. You can also check the following:<sup>78</sup>

$$\frac{1}{1-x^2} = 1 + x^2 + x^4 + x^6 + x^8 + \cdots, \quad (263a)$$

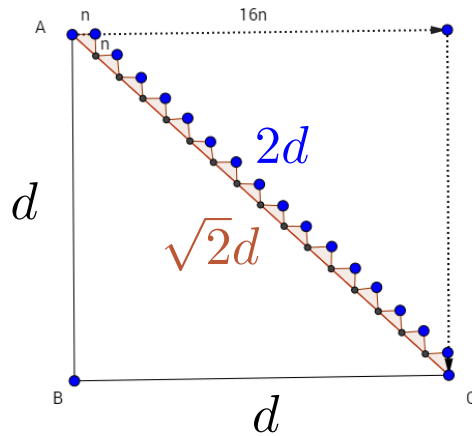
$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 + \cdots. \quad (263b)$$

They both feature a divergence at  $x = \pm 1$ . This is clear in the case of Eq. (263a) since the function itself (on the lhs) diverges. But since

<sup>78</sup> Do it.

$1/(1+x^2)$  is everywhere well defined, why would its right-hand side also diverge there? (some functions like exponentials can be decomposed as polynomials over their entire domain of definition). This will be revealed when we take a deeper look at this function in the complex plane, then the singularity will become naked.

We conclude this Lecture with a modern observation (by Hermann Weyl) why some quantities, like space or time, should be treated as continuous, rather than discrete. This is known as the tile argument: imagine computing the distance of the diagonal of a square (hypotenuse of a right-angle triangle) by little steps on a staircase, as shown below:



Here we took 16 stairs, each of length  $n$ . By construction, the sides of the triangle is  $d = 16n$ , and we know from Pythagoras that the diagonal has length  $\sqrt{2}d$ . Now, measuring not the diagonal, for which we might not have a ruler, but the distance on the staircase, we find the diagonal is really  $2d$  (we go through the same horizontal distance and the same vertical distance). Of course it's longer on the staircase than taking the straight line. The idea then is to take smaller stairs so that the error made is smaller, indeed there is a point where the difference will not be visible to the eye. But  $n$  does not appear in  $2d$ , this distance is the same regardless of the size of the steps! Weyl used that as an argument to show that space is continuous, i.e., a different type of "infinite" than discrete little tiles (that are countables); it is, instead, smooth, and we cannot break it in little chunks. We will repeatedly meet with these two types of infinities: discrete, and continuous, and the mathematical tools to deal with them are typically different.

### *Biographical notes*

Herman Weil: 1885–1955, Mathematician, theoretical physicist and philosopher; in Physics, he put forward the concept of Weyl fermions and wormholes, among others.

### *Further Problems*

#### *Achilles' lamp*

If Achilles' problem is too simple for you, consider this variation:

Consider a lamp with a toggle switch. Flicking the switch once turns the lamp on. Another flick will turn the lamp off. Now suppose someone turns the lamp on and starts a timer. At the end of one minute, he turns it off. At the end of another half minute, he turns it on again. At the end of another quarter of a minute, he turns it off. At the next eighth of a minute, he turns it on again, and he continues thus, flicking the switch each time after waiting exactly one-half the time he waited before flicking it previously.

What is the sum of the infinite series of the time intervals? (this is the point at which Achilles catches the turtle, and is an easy question). Now comes the not-so-easy-question: when this time is reached, is the lamp on or off?

(it is not called Achilles' lamp but *Thomson's lamp*, by the way, after a 20th century philosopher who pondered on this question in 1954.)

#### *Cantor's theorem*

Cantor's theorem states that the cardinality of the power set of any set is strictly larger than the cardinality of the set. This is trivial for finite sets. Actually, the proof is fairly simple also for infinite sets (once you get used to reasoning with sets). It consists in showing that any function  $f$  from  $A$  to  $\mathcal{P}(A)$  is not surjective. Consider  $B = \{x \in A \mid x \notin f(x)\}$  (remember that  $f$  is a function from elements of  $A$  to subsets of  $A$ , so  $f(x)$  is indeed a subset of  $A$  and one can ask if  $x$  itself belongs there). Show that if  $f$  is surjective (remember the definition of that) one reaches a contradiction by finding an element  $y \in A$  satisfying both  $y \in B$  and  $y \notin B$ , which is impossible, proving that  $f$  cannot be surjective. Show as well that there exist injective functions from  $A$  to  $\mathcal{P}(A)$ . Therefore, conclude that  $\#A < \#\mathcal{P}(A)$ . Note that the notation  $\aleph_1$  is sometimes used for  $2^{\aleph_0}$  and  $\aleph_N$  for the cardinality of  $\mathcal{P}(S)$  when  $S$  has cardinality  $\aleph_{N-1}$ . Also,  $\omega$  is used for  $\aleph_1 = 2^{\aleph_0}$ . But all this is notations, what matters are the concepts.

## Lecture 8: Infinitesimals

Let us come back to the function

$$f(x) = \frac{1}{1+x}. \quad (264)$$

This is really the inverse function shifted to  $-1$ . Certainly it has been shifted for good reasons, namely, what matters really is the behaviour of this function around  $x = 0$  since just like world maps place Europe at the center due to a Eurocentric bias, one typically uses a choice of axis that puts the important part of the function at the center. A point of interest is typically centered at  $x = 0$ , as we are usually interested in Physics with what happens at the origin. This shift of the inverse function certainly occurred because we want to focus on this region in the first place. The other important or striking features are now far-away: there is a “singularity”, i.e., a point where the function becomes infinite, at  $x = -1$ , since in this case we divide by zero. For values of  $x$  above  $-1$ , the function is positive, for values below, it is negative (from the sign of the denominator). Also, for large values of  $|x|$ , the functions becomes very small (tends to zero).

But let us come back at the center. How to describe what happens locally, in a so-called neighborhood of  $x = 0$ ? (that is, in the region of points very close to zero). We have seen last Lecture that this expression is actually:

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-x)^k \quad (265)$$

so as long as  $|x| < 1$ , we can expand it as:

$$\frac{1}{1+x} \approx 1 - x + x^2 - x^3 + \dots \quad (266)$$

and since we are interested by the function for small values of  $x$ , given that  $x^n$  is much smaller than  $x^m$  if  $n > m$ , then it is clear that:

- The function is centered at  $\mathbf{1}$  (that's  $f(0) = 1/(1-0)$  so this one was easy anyway).
- It departs from  $\mathbf{1}$  linearly, like a line, since it is then  $1 - x$ .

- As  $x$  gets larger, it bends up,  $1 - x + x^2$ .

etc. To make it clear that we are dealing with small values of  $x$ , we replace this variable by another one, epsilon  $\epsilon$ , and we give it a new name, namely, it is an *infinitesimal*, and we write

$$\frac{1}{1 + \epsilon} \approx 1 - \epsilon \quad (267)$$

You can check:

$$\frac{1}{0.9} \approx 1.1 \quad (268)$$

since  $0.9 = 1 + (-0.1) = 1 - (-0.1)$ . The next order correction is  $1 - \epsilon + \epsilon^2$  is  $1 - (-0.1) + (-0.1)^2 = 1.11$  and this case we can even go to all orders:

$$\frac{1}{0.9} = 1.111111\bar{1} \quad (269)$$

where the bar means this repeats for ever.<sup>79</sup> Another less trivial case:

$$\frac{1}{1.15} \approx 0.85(+)$$

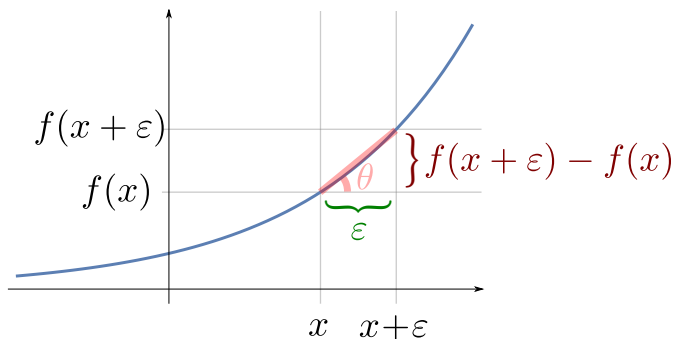
we put a (+) to mean the next correction will increase this value, so it's a lower bound (the real value is 0.869565)

$$\frac{1}{0.99983} = \frac{1}{1 - 0.00017} \approx 1.00017(+)$$

The real value is 1.000170028, etc. This starts to look pretty exact. It is *really exact* when  $\epsilon$  is exactly zero, of course, in which case we have a trivial statement  $1/(1 + 0) = 1 + 0$ . But we will see that infinitesimals can cancel in a way that is not trivial. This is the basis for "calculus" and "analysis".

The most important use of infinitesimals is in their linear approximation of any function  $f$ :

$$f(x + \epsilon) \approx f(x) + \epsilon \tan \theta \quad (272)$$



or, equivalently, by taking the ratio of opposite side with the adjacent side (red over green in the graph above), which is the tangent of the

<sup>79</sup> Use this to show that  $99/100 = 1.010101\bar{01}$  (tips: write  $99/100$  as  $1/(1 - 1/100)$ ). How about  $10^k/(10^k - 1)$  for various  $k$ ?

angle. For the tangent calculated as this ratio to be exactly the same as the tangent to the curve at this point, that is, the straight line that touches the curve at only one point while still going in the same direction, one needs  $\varepsilon$  to be arbitrarily close to zero, so that the angle be exactly that at the one point considered (for instance in the graph above, you can see that  $\varepsilon$  not being an infinitesimal, the red straight line overestimate the tangent to the blue curve, the actual tangent is slightly less steep). So the formula reads:<sup>80</sup>

$$f'(x) \equiv \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon}, \quad (274)$$

where we take the limit  $\lim_{\varepsilon \rightarrow 0}$  to get rid of the infinitesimal in the final result. We can thus rewrite Eq. (272) as:

$$f(x + \varepsilon) \approx f(x) + \varepsilon f'(x). \quad (275)$$

Note that  $f'$  now itself becomes a function of  $x$ , since we can ask the question of its linear behaviour around any value  $x$  and the answer we get in the form of its tangent, or slope, there, clearly is a function of  $x$ .

Let us practice, with  $f(x) = x^2$ . By definition, Eq. (274) yields:

$$(x^2)' = \lim_{\varepsilon \rightarrow 0} \frac{(x + \varepsilon)^2 - (x)^2}{\varepsilon}. \quad (276)$$

Expanding and simplifying the numerator, we find

$$(x^2)' = \lim_{\varepsilon \rightarrow 0} \frac{2\varepsilon x + \varepsilon^2}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} (2x + \varepsilon) \quad (277)$$

and clearly, since  $\varepsilon$  can be made as small as wanted, we then have the result:

$$(x^2)' = 2x. \quad (278)$$

Let us now turn to the general case  $f(x) = x^n$  for  $n \in \mathbb{N}$ . Following the same procedure:

$$(x^n)' = \lim_{\varepsilon \rightarrow 0} \frac{(x + \varepsilon)^n - x^n}{\varepsilon} \quad (279a)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{x^n + nx^{n-1}\varepsilon + \binom{n}{2}x^{n-2}\varepsilon^2 + \dots + \binom{n}{n-1}x\varepsilon^{n-1} + \varepsilon^n - x^n}{\varepsilon} \quad (279b)$$

$$= \lim_{\varepsilon \rightarrow 0} \left( nx^{n-1} + \binom{n}{2}x^{n-2}\varepsilon + \dots + \binom{n}{n-1}x\varepsilon^{n-2} + \varepsilon^{n-1} \right) \quad (279c)$$

$$= nx^{n-1} \quad (279d)$$

since all the terms except the first one can be made arbitrarily small.

As a result:

$$\boxed{(x^n)' = nx^{n-1}} \quad (280)$$

<sup>80</sup> If  $\varepsilon \rightarrow 0$  then also  $-\varepsilon \rightarrow 0$ . Use this to provide the equivalent definition of the derivative:

$$f'(x) \equiv \lim_{\varepsilon \rightarrow 0} \frac{f(x) - f(x - \varepsilon)}{\varepsilon}. \quad (273)$$

We can compute many derivatives like this. Let us take the inverse:  $f(x) = 1/x$ .

$$f'(x) = \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon} \quad (281a)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left( \frac{1}{x + \varepsilon} - \frac{1}{x} \right), \quad (281b)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left( \frac{x}{(x + \varepsilon)x} - \frac{(x + \varepsilon)}{(x + \varepsilon)x} \right), \quad (281c)$$

$$= \lim_{\varepsilon \rightarrow 0} \left( \frac{-1}{(x + \varepsilon)x} \right) \quad (281d)$$

$$= -\frac{1}{x^2} \quad (281e)$$

so that

$$\boxed{\left(\frac{1}{x}\right)' = -\frac{1}{x^2}} \quad (282)$$

Now with the derivative of  $f(x) = \sqrt{x}$ . This can be computed by using  $(a + b)(a - b) = a^2 - b^2$ , which removes the square roots:

$$f'(x) = \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{x + \varepsilon} - \sqrt{x}}{\varepsilon}, \quad (283a)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{x + \varepsilon} - \sqrt{x}}{\varepsilon} \frac{\sqrt{x + \varepsilon} + \sqrt{x}}{\sqrt{x + \varepsilon} + \sqrt{x}}, \quad (283b)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{x + \varepsilon - x}{\sqrt{x + \varepsilon} + \sqrt{x}}, \quad (283c)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{1}{\sqrt{x + \varepsilon} + \sqrt{x}}, \quad (283d)$$

$$= \frac{1}{2\sqrt{x}} \quad (283e)$$

So that, finally:

$$\boxed{(\sqrt{x})' = \frac{1}{2\sqrt{x}}} \quad (284)$$

We have also seen the exponentials as part of our family of functions. The simplest is to use the natural exponential, algebra of powers that turn powers of sum into products of powers, i.e.,  $e^{x+\varepsilon} = e^x e^\varepsilon$  and the definition  $e^\varepsilon = \sum_{k=0}^{\infty} \varepsilon^k / k! = 1 + \varepsilon + \varepsilon^2/2 + \dots$  from which:

$$(e^x)' = \lim_{\varepsilon \rightarrow 0} \frac{e^{x+\varepsilon} - e^x}{\varepsilon}, \quad (285a)$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{e^x (e^\varepsilon - 1)}{\varepsilon}, \quad (285b)$$

$$= e^x \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon + (\varepsilon^2/2) + \dots}{\varepsilon}, \quad (285c)$$

$$= e^x \quad (285d)$$



that is, the derivative of the exponential is itself. That is an important property that makes this important function even more important:

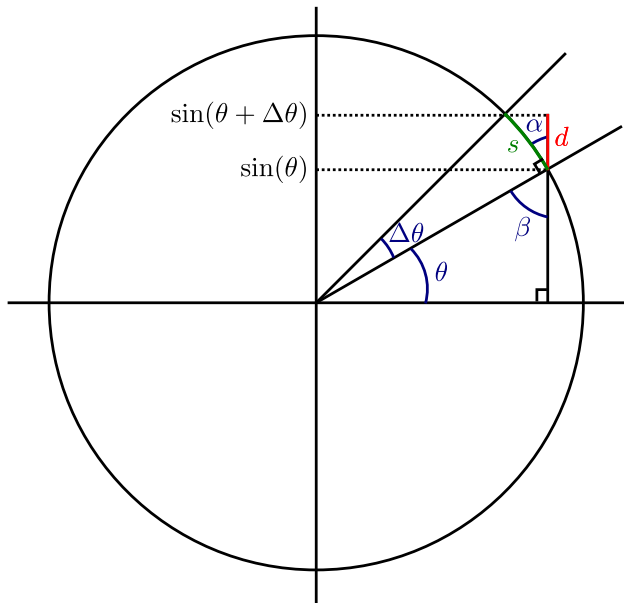
$$\boxed{(\exp)' = \exp .} \quad (286)$$

A last derivation, which is interesting because it relies even more on geometric arguments, involves trigonometric functions. What is the derivative of, say, the sine function?

Let's come back to the original definition Eq. (274), which we will write in terms of angle to get a better grasp of the quantities which vary, namely, angles:

$$f'(\theta) = \lim_{\Delta\theta \rightarrow 0} \frac{\sin(\theta + \Delta\theta) - \sin\theta}{\Delta\theta} \quad (287)$$

as well as the definition of the sine, namely, the vertical projection of an angle in the trigonometric circle (we have two angles in Eq. (287),  $\theta$  and  $\theta + \Delta\theta$ , so we project two sines)



Let us now read the various quantities on this figure. The numerator of Eq. (287) is the distance  $d$  (shown in red) and  $s$  is  $\Delta\theta \times R$  where  $R = 1$  is the radius of the trigonometric circle, meaning that this angle, or the denominator of Eq. (287), is also the arc-distance  $s$  (in green). So the ratio is  $\cos(\alpha)$  in the right-angle triangle in the approximation where  $s$  is not an arc but a straight line, which will become exact when  $\Delta\theta \rightarrow 0$  (and is a good approximation otherwise, as seen on the figure). Now, it remains to find which angle is  $\alpha$ . Since the radius of a circle makes a right-angle ( $\pi/2$ ) with the tangent (or the green line  $s$ ), we have  $\alpha + \frac{\pi}{2} + \beta = \pi$ . From the sum of the angles in a triangle, we also have that  $\theta + \frac{\pi}{2} + \beta = \pi$ . These two together

show that  $\alpha = \theta$ . Therefore,  $d/s = \cos \theta$  (in the limit when  $s \rightarrow 0$ , which is the case when  $\Delta\theta \rightarrow 0$ . Therefore:

$$\boxed{(\sin \theta)' = \cos \theta} \quad (288)$$

An important value of the derivative, which is often why we compute them in the first place, is when it cancels. That is, we are looking for the point  $x_0$  such that  $f'(x_0) = 0$ . In this case, infinitesimal variations around  $x_0$  does not affect the function itself, since, according to Eq. (275), we have  $f(x_0 + \epsilon) \approx f(x_0)$  or, said otherwise (now in terms of Eq. (272)), we have  $\theta = 0$  for the slope of the tangent. Namely, the curve is locally horizontal there. Such a point is called a *stationary point*. The function doesn't change there, or, to be more accurate, it changes to second-order (like the square of  $\epsilon$ , which is much smaller than  $\epsilon$ ).

For the function itself at this point  $x_0$ , it can be either a local minimum, a local maximum or a so-called saddle point. We say "local" because we are sampling in the neighbourhood of the point. The function can still be smaller or larger somewhere else. And it is minimum, maximum or a saddle-point depending on whether it changes signs from negative to positive, in which case we have a minimum (the function first decreases, gets stationary and then increases) or the other way around.<sup>81</sup> We call such a point an "extremum" if we do not know whether it is a minimum or maximum.<sup>82</sup> In the case where the derivative has the same sign on both sides, the point is still stationary but not an extremum anymore, and that's then a saddle point. This is the case for instance of  $x_0 = 0$  for the cube function<sup>83</sup> Some people call such a point a "minimax" point instead; the term saddle will make sense when we turn to the case of multi-variables functions.

We can also assess the nature of the zero of a derivative by iterating the previous ideas to the derivative itself. Since the derivative of a function is itself a function, it follows automatically that one can compute the derivative of the derivative. We call this the "second-order" derivative  $(f')'$  or simply  $f''$ . It is, by definition:

$$f''(x) = \lim_{\epsilon \rightarrow 0} \frac{f'(x + \epsilon) - f'(x)}{\epsilon} \quad (289a)$$

$$= \lim_{\epsilon \rightarrow 0} \frac{\lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon) - f(x + \epsilon - \epsilon)}{\epsilon} - \lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon) - f(x)}{\epsilon}}{\epsilon} \quad (289b)$$

where we have used both Eqs. (273) and (274).<sup>84</sup> Here, we could take various limits for  $\epsilon$  and  $\epsilon$  (and indeed various are taken, for instance when looking at finite-difference approximations for computer methods, which is something we will cover next semester). The simplest is  $\epsilon = \epsilon$ , in which case Eq. (289a) simplifies to:<sup>85</sup>

<sup>81</sup> Work it out

<sup>82</sup> Check that the zeros of the sine correspond to the point where the cosine is extremum. How does this relate to the present discussion?

<sup>83</sup> Check it.

<sup>84</sup> By using the "conventional" definition for the derivative for both terms, show that another definition reads:

$$f''(x) = \lim_{\epsilon \rightarrow 0} \frac{f(x + 2\epsilon) - 2f(x + \epsilon) + f(x)}{\epsilon}$$

This is the so-called "backward" definition while Eq. (290) is called "central". Provide the "forward" definition of the second-order derivative.

<sup>85</sup> Use Eq. (273) as the starting point instead of Eq. (289a), to arrive directly to Eq. (290).

$$f''(x) = \lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon) - 2f(x) + f(x - \epsilon)}{\epsilon^2} \quad (290)$$

where we have used the fact that two limits, if they exist, can be added and the limit of the sum is the sum of the limits. The numerator of Eq. (290), written as  $[f(x + \epsilon) - f(x)] - [f(x) - f(x - \epsilon)]$  shows that the sign of the second-order derivative tells us whether the function bends upward (is concave) or downward (is convex). That is a criterion that is often used to decide of the concavity or curvature. Points where  $f''(x) = 0$  are called “inflection points” (provided  $f''$  also changes signs, in which case it changes curvature at this point). Therefore, the sign of  $f''$  can be used to decide whether  $x_0$  such that  $f'(x_0)$  is a local minimum or maximum.

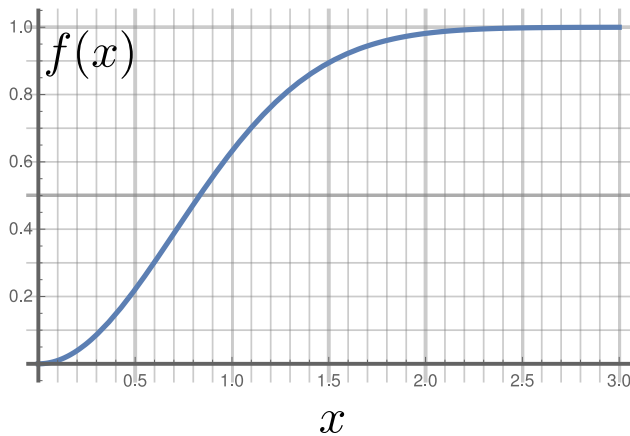
The process can be continued for ever (assuming the function remains differentiable, which will be the case for most functions that we deal with in Physics), which leads to the third  $f'''$ , fourth  $f''''$ , etc., derivatives. The notation  $f^{(k)}$  for the  $k$ -th order derivative is common for  $k \geq 2$ .

## Exercises

### Empirical derivatives

Plot the sine function and construct by hand its derivative by computing the ratio  $(f(\theta + \Delta\theta) - f(\theta)) / \Delta\theta$  for various  $\theta$  (the more points, the better the resolution). Plot the function  $f'(\theta)$  found in this way, and compare with Eq. (288).

Then consider this function, which underlying form or expression is not a priori known. Compute its derivative from the graph itself:



### Derivatives of boredom

Show that the derivative of a constant is zero. Show that  $(\alpha x)' = \alpha$ .

### Derivatives of sums and products

For two functions  $f$  and  $g$ , show that  $(f + g)' = f' + g'$  and that  $(\alpha f)' = \alpha f'$ .

The derivative of  $fg$  for a product of functions requires a bit of trickery. Starting from  $f(x + \epsilon)g(x + \epsilon) - f(x)g(x)$ , try to add (and remove, to keep the final result the same) terms that allow you to let appear quantities related to  $f'$  and  $g'$  and establish their relationship with  $(fg)'$ .

Check the formula you obtained by computing  $(x^n)'$  from Eq. (280) and  $(x^k x^l)'$  with  $k + l = n$  from the rule you have just derived.

### Derivatives of rational powers

Show that  $(x^{\frac{p}{q}})' = \frac{p}{q} x^{\frac{p}{q} - 1}$  for any  $p, q \in \mathbb{N}$ . What do you suggest for  $(x^r)'$  with  $r \in \mathbb{R}$ ? Note that this formula is extremely powerful as it allows you to compute derivative of roots and/or inverse powers at no extra costs. Namely, compute the derivatives of  $1/x$  (as  $x^{-1}$ ),  $\sqrt{x}$  (as  $x^{1/2}$ ) and  $1/\sqrt{x^3}$  (work out the power yourself).

$(\cos \theta)'$

Work out the derivative of the cosine based on its geometrical definition. Can you work out in this way derivatives of other trigonometric functions?

### Derivatives of derivatives

Note that the derivative  $f'$  of a function  $f$  is itself a function of the same variable (since this gives the slope at all points where the function is defined). Therefore, one can iterate, and compute the derivative of a derivative,  $(f')'$ , which we call a *second-order* derivative  $f''$ . And this can be iterated for ever,  $(f'')' = f'''$ , etc., although we tend to write  $f^{(n)}$  (with a parenthesis not to confuse with the power) the  $n$ th derivative. Compute the  $n$ th derivative of the power function  $x^r$  (for any real  $r$ ).

We will soon see about derivatives of composed functions. But we can already compute repeated derivatives of simple compositions, e.g., it is trivial that  $(f + g)^{(n)} = f^{(n)} + g^{(n)}$ . Can you work out the product? It is given by Leibniz' rule:

$$(fg)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(n-k)} g^{(k)}. \quad (291)$$

## Lecture 9: Inverses and compositions of functions.

To increase our collections of function, we will now consider two simple ways to generate new functions out of well known ones.

The first way is so natural that we have already invoked it without mentioning it. This is the *composition of functions*, namely, how to combine two functions  $f$  and  $g$  into a new one  $h(x) = f(g(x))$ . That is, one function becomes the variable for the other one. The notation for this is  $\circ$ , and we write

$$h = f \circ g, \quad (292)$$

so that

$$\boxed{(f \circ g)(x) \equiv f(g(x))}. \quad (293)$$

For instance, if  $f(x) = \sin x$  and  $g(x) = x^2$ , we find  $(f \circ g)(x) = \sin(x^2)$ . From this example, it is clear that, in general:

$$\boxed{(f \circ g) \neq (g \circ f)} \quad (294)$$

since in the particular case at hand,  $(g \circ f)(x) = \sin^2 x$  (make sure that you see the difference).<sup>86</sup> The commutation relation Eq. (294) can be possible, for instance  $f_n(x) = x^n$  for  $n \in \mathbb{N}$  is such that  $f_n \circ f_m = f_m \circ f_n$  since  $(x^n)^m = (x^m)^n = x^{n+m}$ .<sup>87</sup>

<sup>86</sup> Prove it with other counter-examples.

<sup>87</sup> Provide still other cases where  $f \circ g = g \circ f$ .

It will be useful to know how to differentiate compositions of functions. In the following proof, we assume that  $f$  and  $g$  are differentiable (that is, their derivatives exist) and that  $g'(x) \neq 0$ . By definition:

$$(f \circ g)'(x) = \lim_{\varepsilon \rightarrow 0} \frac{(f \circ g)(x + \varepsilon) - (f \circ g)(x)}{\varepsilon} \quad (295)$$

which we can rewrite to let appear the underlying functions directly, namely, multiplying and dividing by  $g(x + \varepsilon) - g(x)$ :

$$(f \circ g)'(x) = \lim_{\varepsilon \rightarrow 0} \frac{f(g(x + \varepsilon)) - f(g(x))}{g(x + \varepsilon) - g(x)} \frac{g(x + \varepsilon) - g(x)}{\varepsilon} \quad (296)$$

where we also used the definition (296). Now, since we assume that  $g$  has a derivative at  $x$ , we know that

$$g(x + \varepsilon) = g(x) + \varepsilon g'(x) \quad (297)$$

for  $\varepsilon$  small enough. Therefore, we can rewrite Eq. (296) as:

$$(f \circ g)'(x) = \lim_{\varepsilon \rightarrow 0} \frac{f(y+v) - f(y)}{v} \frac{g(x+\varepsilon) - g(x)}{\varepsilon} \quad (298)$$

where we defined

$$y \equiv g(x) \quad \text{and} \quad v \equiv \varepsilon g'(x). \quad (299)$$

Note that  $v$  (upsilon) tends to zero if  $\varepsilon$  tends to zero since  $g'(x)$  is just a (nonzero) constant. Using the limit of a product being the product of the limit, we now have:

$$(f \circ g)'(x) = \lim_{v \rightarrow 0} \frac{f(y+v) - f(y)}{v} \lim_{\varepsilon \rightarrow 0} \frac{g(x+\varepsilon) - g(x)}{\varepsilon} \quad (300)$$

since both limits of the product exist, as was initially assumed. The first term of the product is  $f'(y) = f'(g(x))$  and the second is  $g'(x)$ .

Therefore:

$$\boxed{(f \circ g)' = (f' \circ g)g'}. \quad (301)$$

This is an important formula, that allows us to compute the derivative of intricate composite functions. For instance,  $\exp(-x^2) = (f \circ g)(x)$  with  $f = \exp$  and  $g$  the square function, so that  $f'(x) = \exp(x)$  and  $g'(x) = 2x$  and the application of Eq. (301) yields  $(\exp(-x^2))' = -2x \exp(-x^2)$ . You can practice in the Exercises.<sup>88</sup>

By using this rule, one can complete the important list of rules of differentiations for functions (where  $\alpha$  is a constant and  $n \in \mathbf{Z}$ ):<sup>89</sup>

$$(\alpha f)' = \alpha f' \quad (302a)$$

$$(f + g)' = f' + g' \quad (302b)$$

$$(fg)' = f'g + fg' \quad (302c)$$

$$(f^n)' = n f' f^{n-1} \quad (302d)$$

$$(1/f)' = -f'/f^2 \quad (302e)$$

$$(f/g)' = (f'g - fg')/g^2 \quad (302f)$$

$$(\sqrt{f})' = f'/(2\sqrt{f}) \quad (302g)$$

For instance, Eq. (302d) is obtained by using  $h(x) = x^n$ , in which case  $(h \circ f)' = (f^n)' = (h' \circ f)f' = n f^{n-1} f'$  as shown (commuting  $f'$  and  $f^{n-1}$ ).<sup>90,91</sup>

In the particular case where one function undoes the other, we have an important particular case: the *inverse function*. Given a function  $f$ , we call  $f^{-1}$  (but will note this  $g$  for now) its inverse iff:

$$f \circ g = g \circ f = \text{Id}, \quad (303)$$

where  $\text{Id}$  is the identity,  $\text{Id}(x) = x$ . Note that in this sense  $f^{-1} \neq 1/f$  although the notation is sometimes used to mean the latter too,

<sup>88</sup> Differentiate  $1/(1+x)$ ,  $\sqrt{1+x}$  and  $1/\sqrt{x}$ . More cases are given in the exercises.

<sup>89</sup> The first cases have been addressed in the problems of the last lecture using the definition of the derivative. You can extend the procedure to other cases of the list in this way now.

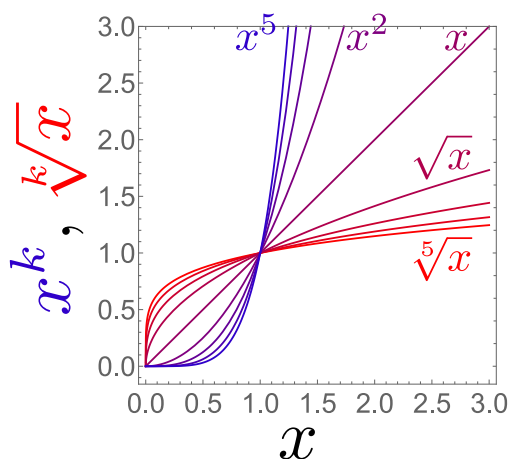
<sup>90</sup> Use the composition rule to demonstrate Eqs. (302) for the other cases.

<sup>91</sup> Using Eq. (302f), show that  $\tan' = 1/\cos^2$ .

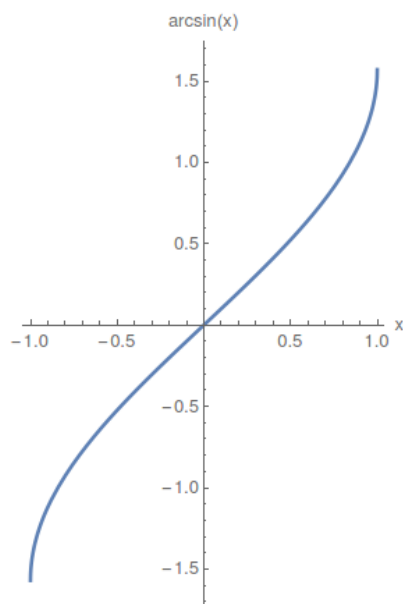
so be careful conventions!<sup>92</sup> For instance, the inverse function of  $f(x) = x^2$  is  $h(x) = \sqrt{x}$ . Note that the domain of definition can be different, since we must ensure that the function remains single-valued. So while  $f : \mathbb{R} \rightarrow \mathbb{R}$ , we only have  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  in the square/square root case. Here the algebraic notation  $x^{1/p}$  for the  $p$ -th root of  $x$  makes sense, since:

$$\sqrt[p]{x^p} = \sqrt[p]{x^p} = (x^{1/p})^p = (x^p)^{1/p} = x^{p/p} = x. \quad (304)$$

These functions, plotted as described in Lecture 6, complete our family of functions.<sup>93</sup>



This is, as another example, the inverse of the sine, i.e., the arcsin function:



which is defined from  $[-1, 1]$  (since this is the codomain of sin) and

<sup>92</sup> What is the “inverse” of the identity or  $f(x) = x$  and of the (real) inverse  $f(x) = 1/x$ ?

<sup>93</sup> From this graph, how much do you estimate is  $\sqrt[3]{e}$ ?

itself takes the values between  $[-\pi/2, \pi/2]$ , which is the largest interval where the sine is injective.

If one has to compute with an inverse function, often, this will involve the properties of the function itself. This is particularly true for derivatives. So we now turn to the practical question of computing the derivative of the inverse  $h$  of a function  $f$ . So let us assume:

$$y = f(x), \quad (305a)$$

$$x = h(y). \quad (305b)$$

We compute  $h'(y)$  as a straightforward application of the derivative of composite functions Eq. (301) applied to Eq. (303):

$$(h \circ f)'(x) = (h(f(x)))' = (h' \circ f)(x)f'(x) \quad (306a)$$

$$= h'(f(x))f'(x) = h'(y)f'(x) = 1 \quad (306b)$$

where the rhs of Eq. (306b) is because by definition  $h(f(x)) = x$  and  $x' = 1$ . Therefore:

$$\boxed{h'(y) = \frac{1}{f'(h(y))}}. \quad (307)$$

Let us compute, for instance, the derivative of arcsin:

$$y = f(x) = \sin(x), \quad (308a)$$

$$x = h(y) = \arcsin(y) \quad (308b)$$

From Eq. (307), it follows that:

$$(\arcsin y)' = \frac{1}{\cos(\arcsin y)} \quad (309)$$

We can go farther by introducing Pythagoras,  $\cos^2 + \sin^2 = 1$  to replace the cos by a sin, which will simplify the arcsin through Eq. (303):

$$\cos(\arcsin y) = \sqrt{1 - \sin^2 \arcsin y} = \sqrt{1 - (\sin(\arcsin y))^2} = \sqrt{1 - y^2}, \quad (310)$$

so that, finally:

$$\boxed{(\arcsin y)' = \frac{1}{\sqrt{1 - y^2}}}. \quad (311)$$

Of all the important functions, we are missing one that we said might be the most important for Physics, the exponential. Its inverse, the so-called *logarithm*, is indeed the queen of the inverse functions, for its magical properties in transforming not-so-easy products (e.g.,  $8 \times 64$ ) into simpler additions (namely,  $3 + 6$ ). The latter is 9, whose inverse operation—back to product—partner is 1024 which, you can check, is  $8 \times 64$ . This was obtained from base 2 logarithms:



1	2	3	4	5	6	7	8	9	10
2	4	8	16	32	64	128	256	1024	2048

We go from the upper row to the lower one by exponentiation of 2, i.e.,  $2^k$ . The reverse operation is the said logarithm, here in base 2:

$$k = \log_2 2^k. \quad (312)$$

Unfortunately, the elegant and powerful notation one way  $a^x$  is ugly and clumsy the other way:

$$x = \log_a a^x. \quad (313)$$

The Euler-number base logarithm we write simply  $\ln$ :

$$\ln \equiv \log_e. \quad (314)$$

(here more details on definition of logarithms  $l(xy) = l(x) + l(y)$  [p29 analysis history]) By definition:

$$\ln \circ \exp = \exp \circ \ln = \text{Id}. \quad (315)$$

A nice surprise is that the derivative of the log is itself a very important function and maybe quite an unexpected one! Defining

$$y = f(x) = \exp(x), \quad (316a)$$

$$x = h(y) = \ln(y) \quad (316b)$$

we find, by application of (301),

$$h'(y) = \frac{1}{\exp'(\ln(y))} = \frac{1}{y} \quad (317)$$

since the exponential is its own derivative and  $\exp \circ \ln = \text{Id}$ . Thus:

$$\boxed{(\ln x)' = \frac{1}{x}}. \quad (318)$$

## Exercises

### *An important reciprocal function*

Study the reciprocal of the tangent (it is called the arctan). That is, find its domain, codomain and its derivatives, in particular its slope at the origin (you will need to adapt the proof yielding Eq. (311) to do so). The shape is characteristic of so-called "sigmoid" functions.

### *A stupid way to differentiate*

Compute  $(x^6)'$  as the derivative of  $f \circ g$  and as the derivative of  $g \circ f$  with  $f(x) = x^3$  and  $g(x) = x^2$ . Is everything according to your expectation?

*Derivative of the ln*

Use the formula for the derivative of the inverse function to prove Eq. (318). Use this knowledge to provide the Taylor expansion of the logarithm.

*Derivatives of intricate composite functions*

Give the formula for  $(f \circ g \circ h)'$ . Use it (and possibly even higher iterates) to compute the derivatives of the following intricate composite functions:

$$f_a(x) = \sqrt{1 - \sqrt{1 + x^2}}, \quad (319a)$$

$$f_b(x) = \frac{1}{\sqrt{1 + \exp(-x^2)}}, \quad (319b)$$

$$f_c(x) = e^{\sin(x^2)}, \quad (319c)$$

$$f_d(x) = \ln(\sqrt{1 + x^2}), \quad (319d)$$

$$f_e(x) = \ln\left(\sqrt{\frac{1+x}{1-x}}\right). \quad (319e)$$

*Something to remember*

Sometimes derivative of complicated compositions of functions result in remarkably simple results. This is something to keep in mind would we need the “reverse” operation (called “integration”) of these simple functions. For instance, compute  $(\ln \circ \cos)'$  and keep this in mind for your collection of integrals of trigonometric functions.

*Good to remember*

We have seen previously how to compute the derivative of a sum and a product of two functions  $u$  and  $v$ , namely,  $(u + v)' = u' + v'$  and  $(uv)' = u'v + uv'$  (this was demonstrated with the definition of the derivative). Now we can add the important following formulas to compute quickly derivative of functions that the most popular compositions:

1.  $(1/u)' = -u'/u^2$ .
2.  $(u/v)' = (u'v - uv')/u^2$ .
3.  $u^n = nu'u^{n-1}$ .
4.  $\sqrt{u} = u'/(2\sqrt{u})$ .

Prove the above results (and learn them) using Eq. (301). Now it's easy to find that, e.g.,  $(\cos^3 x)' = -3 \sin x \cos^2 x$ . If you forget a particular formula, go back to Eq. (301) instead. Note that 1. is a

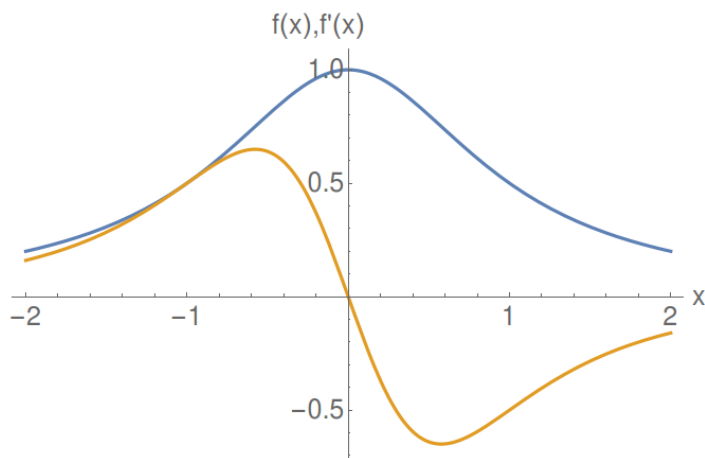
particular case of 2. and 4. is a particular case of 3., so there is really little to learn.

*Larger than one's derivative*

The function  $f(x) = 1/(1+x^2)$  has an interesting feature:

$$f'(x) \leq f(x). \quad (320)$$

Prove it. Can the equality be satisfied, if so, for which values of  $x$ ?





## Lecture 10: Taylor Polynomials

We have seen how  $1/(1+x) = 1 - x + x^2 - x^3 + \dots$  and how this provides useful approximations to the function in terms of a polynomial, or, more accurately, a *power series* (power refers to the  $x^k$  term and series refers to an infinite sum).

Here there is a powerful idea, rewriting something complicated (at any rate, general) in term of a set of generic objects, which are the “monomials”  $x^k, k \in \mathbb{N}$ . We are slowly going towards a major idea of modern Mathematics, which is to encapsulate the complicated things into separate objects (here the monomials) and deal with the simpler ones, which are merely numbers that “weight” these objects (here the coefficients 1 for  $x^0$ ,  $-1$  for  $x^1$ , 1 for  $x^2$ , etc. So the function  $x \rightarrow 1/(1+x)$  can be seen as the vector (remember, a vector is really a column of numbers):

$$x \rightarrow \frac{1}{1+x} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ \vdots \end{pmatrix}. \quad (321)$$

This is an infinite vector. But so what? It is countably infinite, so nothing we cannot deal with with our fingers only. And even if we could let it to the power of abstraction of our brain to imagine what is going on in the “ $\vdots$ ”, in most cases one can even provide the generic term of the  $k$ -th entry, in this case,  $(-1)^k$ , so there is nothing ill- or un-defined about this. Note that by definition, all polynomials are finite vectors. Since there is no maximum order in the set of all polynomials, we still infinite vectors though. Isn't an infinite actually simpler in this case? Should we, for whatever reason, require a maximum size for our vectors, we could always “truncate” the space, and ignore everything that goes over the limit. But again, is that really a simplification? In computer versions of some problems, it often is. In the Platonic universe, it seldom is.

Of course such a concept of replacing functions by vectors will be useful if we can generalize it to all functions, and not only this one (or its immediate variants), which we have derived from a trick of distributivity of the product over the addition:  $(1+x)(1-x+x^2-x^3+\dots+x^n) = 1+x^{n+1}$  (and taking the limit  $n \rightarrow \infty$ ). We can find expressions for other functions using similar tricks (we already used  $1/(1-x) = 1+x+x^2+x^3+\dots$ ). For instance:

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \dots \quad (322)$$

But how to generalize it to all possible functions?

Here we will see a very powerful formula that achieves such a polynomial expansion for all functions (not literally *all functions* that exist, but basically all the useful functions for Physics, which we shall call “analytic”; they will be studied in more details next year). This is known as Taylor’s formula and reads as follows:

$$f(x) = f(a) + (x-a)f'(a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots + \frac{f^{(k)}(a)}{k!}(x-a)^k + h_k(x)(x-a)^k, \quad (323)$$

or directly as expression of the general term, since  $f^{(0)} = f$  and  $f^{(1)} = f'$ :

$$f(x) = \sum_{l=0}^k \frac{f^{(l)}(a)}{l!} (x-a)^l + h_k(x)(x-a)^k. \quad (324)$$

Here it is written for the general case of the expansion around  $a$ . When we take (as is often the case) the particular case  $a = 0$ , this some people prefer to call a “Maclaurin” series, especially if we carry this on up to infinity:

$$f(x) = \sum_k \frac{f^{(k)}(0)}{k!} x^k. \quad (325)$$

You will use that profusely in the Exercises

#### IMPORTANT CASES

Taylor series are very important for approximations. You need to know them off the top of your head. It will come through practice. Here's practice time. Using the formula for the derivative of composite functions, compute the Taylor series for the following cases:

- |                   |                     |
|-------------------|---------------------|
| 1. $\sin(x)$      | 7. $1/(1-x)$        |
| 2. $\cos(x)$      | 8. $1/\sqrt{1+x}$   |
| 3. $\tan(x)$      | 9. $1/\sqrt{1+x^2}$ |
| 4. $\sqrt{1+x}$   | 10. $1/(1+x)^{3/2}$ |
| 5. $\sqrt{1+x^2}$ | 11. $\ln(1-x)$      |
| 6. $1/(1+x)$      | 12. $\ln(1+x)$      |

Do this to practice your derivating skills. Note that some functions, like numbers 6 and 7 above, we have already dealt with through other ingenious ways, but it's good practice to check it also works the general way.

#### GENERALIZED BINOMIAL COEFFICIENTS

Show that the Taylor expansion of  $(1+x)^\alpha$  gives rise to the so-called *generalized Binomial coefficients*

$$\binom{\alpha}{n} \equiv \prod_{k=1}^n \frac{\alpha - k + 1}{k} \quad (326)$$

in the power series

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k. \quad (327)$$

Show that the binomial formula where  $\alpha = n \in \mathbb{N}$  is a particular case of Eq. (327). Check the cases  $\alpha = \pm 1/2$  (cf. previous exercise).

A direct application of the formula (325) is not always the simplest way to compute a Taylor (or Maclaurin) expansion, although it is the most straightforward as it is automatic. In addition to the tricks already seen, we can add for instance the following tricks (which get justified in a lecture on power series, next year): power series can be

integrated and differentiated term by term, that is, if

$$f(x) \equiv \sum_{k=0}^{\infty} \alpha_k x^k \quad (328)$$

then

$$f(x)' = \sum_{k=1}^{\infty} k \alpha_k x^{k-1}, \quad (329a)$$

$$\int f(x) dx = \sum_{k=0}^{\infty} \frac{\alpha_k}{k+1} x^{k+1}. \quad (329b)$$

For example, since  $1/(1+x)' = -1/(1+x)^2$  then we have:

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + 5x^4 + \dots \quad (330)$$

a Similar results can be obtained easily with the integration part.<sup>94</sup> We will meet several other examples.

But let us come back to Taylor's formula, because there is much we left to discuss. Note that there is term  $h_k(x)$ , which, unless we characterize it, makes the full expression worthless (as this could be anything). The important property of this term is that it is very small in the neighborhood of  $a$ :

$$\lim_{x \rightarrow a} h_k(x) = 0 \quad (332)$$

which means that the so-called *remainder*  $h_k(x)(x-a)^k$  goes to zero as  $x$  goes to  $a$  faster than all the previous other terms. It means, for us Physicists, that this is a part which we can ignore (or approximate away). The part without the remainder is the *Taylor polynomial* that provides the promised transformation of a general function  $f$  into a sum of monomials (here, around  $a$ ):

$$f(x) = P_k(x) + h_k(x)(x-a)^k \quad (333)$$

with

$$P_k(x) = \sum_{l=0}^k \frac{f^{(l)}(a)}{l!} (x-a)^l. \quad (334)$$

Various  $k$  yield various (increasing) orders of approximations. Did you recognize the first-order approximation?

$$f(x) = f(a) + (x-a)f'(a). \quad (335)$$

This is the linear  $(x-a)$  approximation to  $f$ , through the derivative! In this respect, the second-order derivative can be seen as the best quadratic  $((x-a)^2)$  approximation to  $f$ . Write  $x-a = \pm\epsilon$ , so that, to

<sup>94</sup> Use Eq. (329b) to compute the Taylor expansion of the arctan, remembering from the previous lecture that  $(\arctan)' = 1/(1+x^2)$ . Since  $\arctan(1) = \pi/4$  (check it from the definition of this function), use the previous result to derive this famous (so-called Leibniz') result:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots + \frac{(-1)^n}{2n+1}. \quad (331)$$



second-order

$$f(a + \varepsilon) = f(a) + \varepsilon f'(a) + \frac{\varepsilon^2}{2} f''(a) \quad (336a)$$

$$f(a - \varepsilon) = f(a) - \varepsilon f'(a) + \frac{\varepsilon^2}{2} f''(a) \quad (336b)$$

and thus, summing both terms, we arrive to, to second-order:

$$f''(a) = \frac{f(a + \varepsilon) - 2f(a) + f(a - \varepsilon)}{\varepsilon^2} \quad (337)$$

which is the expression we derived last lecture.

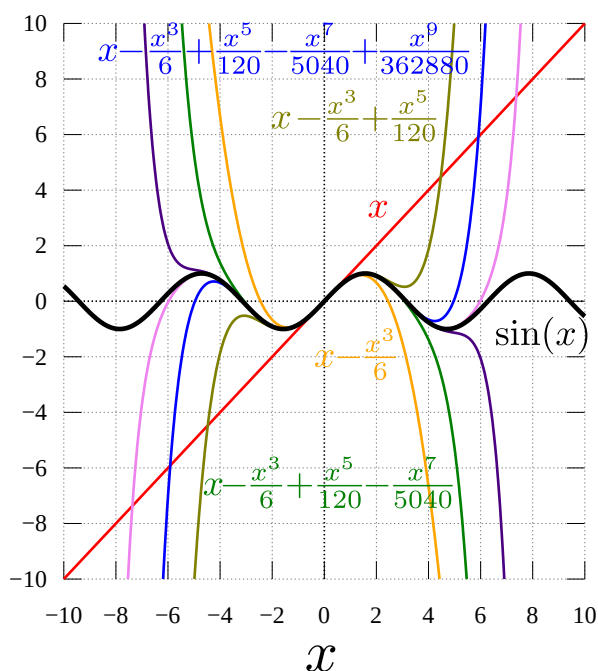
Let us compute one important case for practice. We have computed last time (the cos left for you as an exercise)

$$\sin(x)' = \cos(x) \quad \text{and} \quad \cos(x)' = -\sin(x) \quad (338)$$

so repeated applications of Eq. (338) inserted into Eq. (325) yields:

$$\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \frac{x^9}{362880} + \dots + (-1)^k \frac{x^{2k+1}}{(2k+1)!} + \dots \quad (339)$$

Isn't it marvelous? The complicated, wiggly sine function, can be decomposed as increasing powers of its variable!



Here you can see how increasing powers provide better and better approximations, both for small  $x$  but also in reaching increasing spans of overlap with the sine function itself.

If we think back in terms of vectors, cf. Eq. (321), we then have:

$$\sin(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -\frac{1}{6} \\ 0 \\ \frac{1}{24} \\ \vdots \\ 0 \\ \frac{(-1)^k}{(2k+1)!} \\ \vdots \end{pmatrix}. \quad (340)$$

Note that to first order, the sine function is a line. That is the (so-called *paraxial*) approximation you are familiar with from Optics:

$$\sin(\theta) \approx \theta. \quad (341)$$

Who would not admit this is a considerable simplification? That is Physics in plain view here. We could also carry on and define new functions. For instance, we will often use in Physics the so-called cardinal sine, defined as  $\text{sinc}(x) \equiv \sin x/x$ . There seems to be a 0/0 indeterminacy, but clearly not if we look at its Taylor expansion:

$$\frac{\sin x}{x} = 1 - \frac{x^2}{6} + \frac{x^4}{120} + \cdots + (-1)^k \frac{x^{2k+1}}{(2k)!} + \cdots \quad (342)$$

which is well defined with  $\text{sinc}(0) = 1$ . Taylor expansions are of great uses for quick tricks, efficient calculations, powerful results. We will meet it often in all branches of Physics.

Anyway, we are in a Math course, so let us prove what we assert. There is (should be) no magic.

*Proof of Taylor's formula:* What do we have to prove? You might feel that Mathematicians are powerful and that the result should be derived from scratch. Let us be more practical and start from the result, and confirm it instead as a mean of proof; how this result has been found in the first place is interesting but another story: luck, trial and error, brilliant insight, message from God (Cantor was thinking that!), etc.

So, we have to prove that

$$\lim_{x \rightarrow a} h_k(x) = 0 \quad (343)$$

where, from rearranging Eq. (334)

$$h_k(x) = \begin{cases} \frac{f(x) - P(x)}{(x-a)^k} & \text{if } x \neq a \\ 0 & \text{if } x = a \end{cases} \quad (344)$$

where, it might be useful to recall

$$P(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k \quad (345)$$

(we don't write  $P_k(x)$  to simplify the notation but you know what we mean). The limit in Eq. (343) is of the type<sup>95</sup>  $0/0$  and one needs to find a simplification of some sort. We can use the so-called *L'Hôpital's rule*, that says that if:

$$\lim_{x \rightarrow x_0} f(x) = 0 \quad (346a)$$

$$\lim_{x \rightarrow x_0} g(x) = 0 \quad (346b)$$

then

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}. \quad (347)$$

This is actually easy to prove as well (and is a nice trick to remember, maybe the funny name will help you):

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{g(x) - g(x_0)} \quad (348)$$

because by definition of the limit,  $f(x_0) = g(x_0) = 0$

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{\frac{f(x) - f(x_0)}{x - x_0}}{\frac{g(x) - g(x_0)}{x - x_0}} \quad (349)$$

here we use the fact that if  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$  and  $\lim_{x \rightarrow x_0} g(x) = g(x_0)$ , then  $\lim_{x \rightarrow x_0} f(x)g(x) = f(x_0)g(x_0)$  (the limit of a product is the product of the limits, *if* the limits exist!). So, carrying on:

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}}{\lim_{x \rightarrow x_0} \frac{g(x) - g(x_0)}{x - x_0}} = \frac{f'(x_0)}{g'(x_0)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}. \quad (350)$$

Back to our remainder:

$$\lim_{x \rightarrow a} h_k(x) = \lim_{x \rightarrow a} \frac{f'(x) - P'(x)}{[(x-a)^k]'} \quad (351)$$

which is still  $0/0$  indeterminacy! Indeed,  $P^{(j)}(a) = f^{(j)}(a)$  for  $0 \leq j \leq k-1$ . Same problem, same solution, we iterate L'Hôpital's rule, until we get to the case that is not  $0/0$ , namely:

$$\lim_{x \rightarrow a} h_k(x) = \lim_{x \rightarrow a} \frac{[f(x) - P(x)]^{(k-1)}}{[(x-a)^k]^{(k-1)}} \quad (352)$$

<sup>95</sup> Check it: The denominator  $(x-a)^k$  is clearly zero when  $x \rightarrow a$ . What is  $f(a)$  and  $P(a)$  in the numerator and thus what is  $f(a) - P(a) = \lim_{x \rightarrow a} f(x) - P(x)$ ?

The numerator can be evaluated as follows:

$$P(x) = \sum_{l=0}^k \frac{f^{(l)}(a)}{l!} (x-a)^l \quad (353a)$$

$$P(x)^{(1)} = \sum_{l=1}^k \frac{f^{(l)}(a)}{l!} l(x-a)^{l-1} \quad (353b)$$

$$P(x)^{(2)} = \sum_{l=2}^k \frac{f^{(l)}(a)}{l!} l(l-1)(x-a)^{l-2} \quad (353c)$$

$$\dots \quad (353d)$$

$$P(x)^{(k-1)} = \sum_{l=k-1}^k \frac{f^{(l)}(a)}{l!} l(l-1)\dots(l-k+2)(x-a)^{l-k+1} \quad (353e)$$

where in Eq. (353e) the last term is  $l(l-1)(l-2)\dots(l-l_0+1)$  with  $l_0$  the lower bound of the sum, so after  $k-1$  derivatives, we are left with  $l(l-1)(l-2)\dots(l-k+2)$  when  $l_0 = k-1$ , as written.

Therefore, spelling out the sum of Eq. (353e), this reduces to two terms only:

$$P(x)^{(k-1)} = f^{(k-1)}(a) + f^{(k)}(a)(x-a). \quad (354)$$

The denominator we can compute similarly easily:<sup>96</sup>

<sup>96</sup> Prove this (by induction).

$$[(x-a)^k]^{(k-1)} = k!(x-a) \quad (355)$$

Using Eqs. (354) and (355) in our  $(k-1)$ -times iterated L'Hôpital's rule, we can reduce

$$\lim_{x \rightarrow a} h_k(x) = \lim_{x \rightarrow a} \frac{f^{(k-1)}(x) - P^{(k-1)}(x)}{k!(x-a)} \quad (356)$$

to

$$\lim_{x \rightarrow a} h_k(x) = \frac{1}{k!} \lim_{x \rightarrow a} \left( \frac{f^{(k-1)}(x) - f^{(k-1)}(a)}{x-a} - f^{(k)}(a) \frac{x-a}{x-a} \right) \quad (357a)$$

$$= \frac{1}{k!} (f^{(k)}(a) - f^{(k)}(a)) = 0 \quad (357b)$$

QED.

If we call  $R_k(x) \equiv f(x) - P_k(x)$  the remainder, then one can show (we won't) that there exists  $\zeta$  in  $[x, a]$  such that:

$$R_k(x) = \frac{f^{k+1}(\zeta)}{(k+1)!} (x-a)^{k+1}. \quad (358)$$

(there are other forms, this one is Lagrange's, there is also a Cauchy's form; you can look it up in the literature).

Taylor's theorem applies to so-called analytical functions, which are defined precisely as functions whose Taylor expansion converge

towards the function itself. The main reason why a function is not analytical is because it cannot be differentiated (e.g., it is not smooth, is discontinuous or has a kink). But some non-analytical functions are more subtle, for instance:

$$f(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \quad (359)$$

is a strange beast with many astounding properties. One is that its Taylor expansion brings you nowhere: it is everywhere zero (said otherwise, the Taylor remainder is the full function itself). We can check it explicitly for the first derivative

$$f'(x) = \frac{2e^{-1/x^2}}{x^3} \quad (360)$$

which is however not defined at  $x = 0$ , so we need to compute it using the basic definition:

$$f'(0) = \lim_{\varepsilon \rightarrow 0} \frac{\exp(-1/\varepsilon^2)}{\varepsilon} \quad (361)$$

and change variable so that  $\omega = 1/\varepsilon$ , which gives:

$$f'(0) = \lim_{\omega \rightarrow \infty} \frac{\omega}{\exp(\omega^2)} \quad (362)$$

or, using L'Hôpital rule:

$$f'(0) = \lim_{\omega \rightarrow \infty} \frac{1}{2\omega \exp(\omega^2)} = 0, \quad (363)$$

which gives us  $f'$ 's definition as a counterpart of Eq. (359):

$$f'(x) = \begin{cases} \frac{2 \exp(-1/x^2)}{x^3} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \quad (364)$$

and we compute  $f''$  similarly:

$$f''(x) = \frac{4e^{-\frac{1}{x^2}}}{x^6} - \frac{6e^{-\frac{1}{x^2}}}{x^4} \quad (365)$$

where we are faced with the same problem of  $f''(0)$  not being defined by this expression, forcing us to compute it explicitly. In general, we find that  $f^{(n)}(x) = P_n(1/x) \exp(-1/x^2)$  where  $P_n$  is a polynomial of order  $n$ . The derivative at 0 is then found as:

$$f^{(n)}(0) = \lim_{\varepsilon \rightarrow 0} \frac{P_n(1/\varepsilon) \exp(-1/\varepsilon^2)}{\varepsilon} \quad (366a)$$

$$= \lim_{\omega \rightarrow \infty} \frac{P_{n+1}(\omega)}{\exp(\omega^2)}. \quad (366b)$$

where we defined  $P_{n+1}(\omega) \equiv \omega P_n(\omega)$  the  $(n + 1)$ th polynomial. We could use L'Hôpital's rule again, or see that the exponential grows faster than any polynomial<sup>97</sup> therefore the limit is 0, showing that to all orders, the Taylor expansion of  $\exp(-1/x^2)$  is zero. Thus the series only converges at  $x = 0$ , which is however trivial. It is easy to show that for any  $x \neq 0$ , the function is nonzero<sup>98</sup> and thus the Taylor theorem breaks to all orders. Such a function, by definition, is *not* analytic! Its analyticity breaks at a single point.<sup>99</sup> Some cases can be given (although we won't here) of functions that are continuous yet *nowhere* analytic.

We conclude with another peek at something which we already discussed in Lecture 7 on infinities, and which you will study properly next year, but that is so revealing of the structure of things that it is worth telling you about it now. Does this formula work forever? We've seen that the remainder is small in the neighborhood of  $a$ , but we've seen that bringing more and more terms, we are able to stray farther and farther from  $a$ , as was the case for the sine function, for instance. Is it always the case? No, for example, Eq. (322), with which we started, breaks down dramatically at  $x = \pm 1$ . Why is this? It is because this idea of Taylor expansion does not work only with real numbers, but also (and in fact, even better) with complex numbers. And while Eq. (322) presents no problem for the real variable, do you see what could go wrong in  $\mathbb{C}$ ? It can diverge, there is a "pole" at  $z = i$  for

$$f(z) = \frac{1}{1+z^2}, \quad (367)$$

and, as you will see, if it breaks at one point of the complex plane, it has to break on all points that are a same distance away.

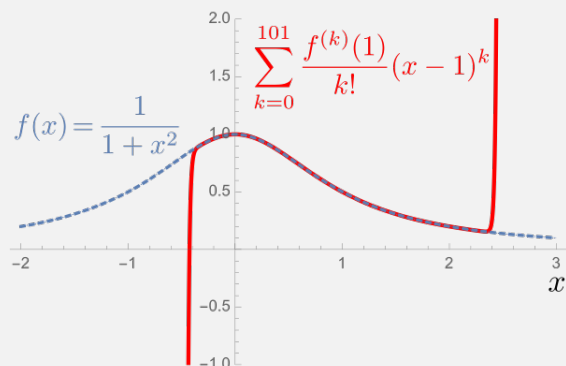
<sup>97</sup> Prove this (for instance using L'Hôpital, but that is not the only way).

<sup>98</sup> Prove this.

<sup>99</sup> Give the Taylor expansion of  $\exp(-1/x^2)$  around  $x_0 = 1/4$ .

### THE INVISIBLE SINGULARITY

Since the Maclaurin expansion for  $1/(1+x^2)$  diverges at  $x = \pm 1$  although the function itself is smooth there, let us obtain its Taylor expansion around  $a = 1$  using Eq. (323) and see if something dramatic happens there. We let you work out the first orders and provide the graphed result for the 101-th order. Where do the divergences appear to be? What is special about them? Can you provide the exact expression based on the preliminary discussion we made in the text?



### Exercises

#### Future friends

These are very important Taylor series which we have not look at in details yet, but on which we shall spend considerable amounts of time. This is your chance to make their acquaintance prior to any bias from knowing what they are:

$$f_1(x) = \sum_{k=0}^{\infty} \frac{(-1)^{k+1} x^k}{k}, \quad (368a)$$

$$f_2(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \quad (368b)$$

Plot these two functions,  $f_1$  on  $[-1, 1]$  and  $f_2$  on  $[-1, 2]$  (for  $f_2$  you can actually take any interval you want), to various orders, e.g., using a computer, or if you are not familiar with computers yet, taking a grid of numbers, e.g.,  $\mathcal{G} = \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$  and computing up to order 5. This will require a bit of organization but with method and order, you should get through it in reasonable time. Once you have these functions, do you see a way to relate them? (believe it or not, they are essentially the same shape).

Now assume the following rule of differentiation:

$$\left(\sum_{k=0}^{\infty} \alpha_k x^k\right)' = \sum_{k=1}^{\infty} k \alpha_k x^{k-1} \quad (369)$$

i.e., that the normal rule of differentiation work even though we have an infinite sum (you will see next year when this is allowed; it is in this case). Compute  $f_1$  and  $f_2$  and from your now ample knowledge of Taylor series, identify which functions are the derivatives. Does it help you identifying  $f_1$  and  $f_2$  themselves?

### *More of the same*

The functions worked out in the text itself are so important, you really need to do them, like, even several times. The following are also important but somehow are more intended for the “connoisseur” and “refined theorist”. You may do them last if you wish.

1.  $\arcsin(x)$
2.  $\arccos(x)$
3.  $\arctan(x)$

To get you started, remember from last lecture that  $(\arcsin(x))' = 1/\sqrt{1-x^2}$ ,  $(\arccos(x))' = -1/\sqrt{1-x^2}$  and  $(\arctan(x))' = 1/(1+x^2)$ .

### *What's wrong with me?*

How about functions like  $\sqrt{x}$  (expanding around  $x = 0$ ). Why do you think we don't put it in the list? It should be clear from the formula. Plot it and see if you can find what is wrong with it geometrically. It is even more clear for functions like  $1/x$  (still at  $x = 0$ ). Why isn't it in the list? It looks more important than  $1/(1+x)$  (we will see how to expand such functions usefully next years, through Laurent series; we will do this when we return to infinities, this time in the complex plane).

### *Not Maclaurin*

We have been dealing with the particular case  $a = 0$  above, as it's actually the most common one. That is to say, we have been using Maclaurin series. Let us practice the general case.

We cannot compute a Taylor series for  $1/x$  around  $x = 0$ , but we can around, e.g.,  $x = 1$ . What is the series then?

Similarly, compute the Taylor series for  $\sin(x)$  around  $a = \pi/4$ .



Note that we did not expand polynomials as they are, basically, already in the Taylor form (which is turning functions into polynomials). But it's not entirely trivial if we shift the origin. Give the Taylor expansion of  $x^3$  around  $a = 1$ . Note that the expansion is finite in this case. Can you see why it has to be the case?

We have shown that  $\exp(-1/x^2)$  is not analytical at 0, meaning that its Maclaurin expansion does not converge to the function itself there. It does everywhere else. Provide for instance the Maclaurin expansion around  $a = 1/4$  for this function (and try to check it converges locally to the weird shape of this strange function: you will see how it cancels the low-order polynomials and magnifies the high-order ones to reproduce this uncanny flatness).

### *L'Hôpital*

Use L'Hôpital's rule to compute:

$$\lim_{x \rightarrow 0} \frac{2 \sin(x) - \sin(2x)}{x - \sin(x)}. \quad (370)$$

Like for the proof of Taylor's theorem, a single application is not enough to lift the o/o indeterminacy. In this case, the rule needs to be applied three times! Your answer should be between between 1 and 10.



## Lecture 11: Kets and Bra.

In this lecture, we start to explore one of the most powerful ideas and certainly the most important topic of University-level Mathematics: the concept of an abstract *vector space*. We have already met geometrical vectors, many times in our previous education and in this course as well (Lecture 5). All together they form a set (the set of geometrical vectors) with properties, such as distances, etc., that make this set behave like a “space” of its own. A vector space has well defined properties, which we will cover with all the details of cold Mathematical rigor later on. For now, as a mean of introduction to these ideas, we will approach the problem through a highly Physical-perspective, in particular with notations and terminology introduced by a Physicist (not a Mathematician), Paul Adrian Dirac, who brought together special relativity (Einstein theory) and quantum mechanics (Schrödinger, Heisenberg and others theory). This is an important notation which we will use profusely next semester onward and that you will find everywhere in the literature. But we are of course concerned with something more important than the notation itself, namely, to the concept, which boils down to the fact that a vector space is a general structure that applies to a whole set of different types of Mathematical objects, not only geometrical arrows, or their immediate extension of  $n$ -column vectors. That will be the first benefit of using Dirac notation, namely, to make clear that a vector space does not have to be the one of little geometric arrows but applies to literally all sorts of objects, as long as they obey the rules of the structure. Thus, instead of

$$\vec{u} \tag{371}$$

or  $\mathbf{u}$  for  $n$ -dimensional vectors, we will now write

$$|u\rangle \tag{372}$$

and that’s the first step we make towards abstraction. We will see the advantages and power of this elegant notation later on. Dirac called that a “ket” (we’ll also see why later on).

Maybe the most vivid illustration of the abstractness brought by this notation is that the following expression is a legitimate equation of physics:

$$\left| \begin{array}{c} \text{cat} \\ \text{state} \end{array} \right\rangle = \frac{1}{\sqrt{2}} \left( \left| \begin{array}{c} \text{cat} \\ \text{state} \end{array} \right\rangle + \left| \begin{array}{c} \text{cat} \\ \text{state} \end{array} \right\rangle \right) \quad (373)$$

and one of the most important one, namely, the principle of superposition in quantum mechanics, here illustrated with so-called “cat states”. It will be for a quantum mechanics course to explain their physical meaning. Here we will content to stress that one can make calculations with such an expression, and derive results (cf. Exercises).

In this first Lecture exploring these notions, we will focus on an important aspect of a vector space, related to the norm, or size, of its vectors. For geometrical arrows, following Pythagoras theorem, we have seen that it is given by the square root of the scalar product with itself:

$$\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} \quad (374)$$

since indeed  $\vec{u} \cdot \vec{u} = u_x^2 + u_y^2 + u_z^2$  for  $\vec{u} = (u_x, u_y, u_z)^T$ , we have the expected distance in geometrical space. Note by the way that we used here the “transpose”  $^T$  which consist in laying down the column vector into a row vector. Dirac’s notation already justifies itself as the transpose of a ket becomes a “bra” and is written  $\langle u|$ , which is precisely what the transpose seems to be doing, but now it does it to our abstract vector:

$$\langle u| \equiv |u\rangle^T \quad (375)$$

In addition to the norm of a single vector, we have also seen how the scalar product gives a useful measure of the overlap between two vectors, through their scalar product:

$$\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z \quad (376)$$

so the norm is really obtained by looking at the overlap of a vector with itself. Now let’s turn to a function space  $\mathcal{F}$ , that is, vectors are now real functions of the real variable  $f(x)$  (we’ll deal with complex cases later on). To make the transition smooth, let’s approach the problem through *discretization*, which means we will not consider  $x \in \mathbb{R}$  just right now but instead consider only a finite mesh of points, and work with  $n$ -dimensional vectors  $\mathbb{R}^n$  (meaning columns with  $n$  numbers). For instance, the sine function sampled over, say,  $[-24, 24]$ , and for the ease of computation, taking small multiples of  $\pi$ , namely, working with:

$$f(x) = \sin\left(\frac{\pi}{6}x\right) \quad (377)$$

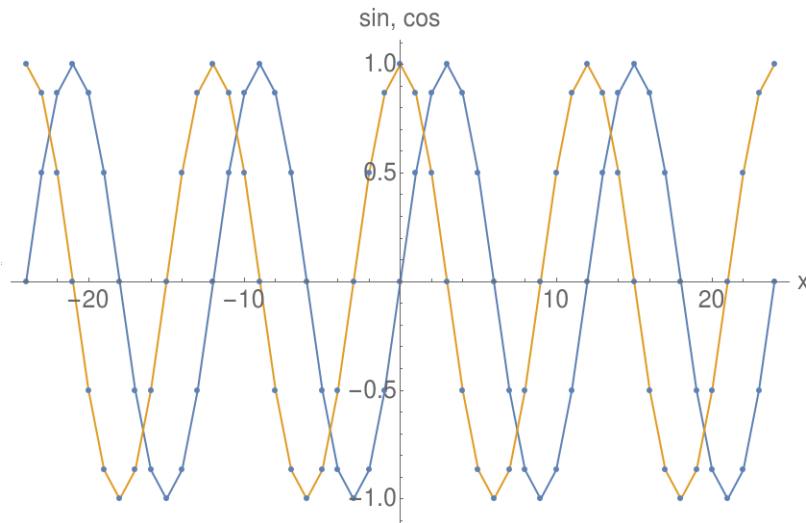
(you can of course consider other cases, we took 24 to go round the circle several times, namely, four, in integer steps). Since  $\sin(\pi/6) = 1/2$  and  $\sin(\pi/3) = \sqrt{3}/2$ , as is easily worked out (cf. exercises), we can write this as a vector:

$$|\sin_{\pi/6}\rangle = \begin{pmatrix} 0 \\ 1/2 \\ \sqrt{3}/2 \\ 1 \\ \sqrt{3}/2 \\ 1/2 \\ 0 \\ -1/2 \\ -\sqrt{3}/2 \\ -1 \\ -\sqrt{3}/2 \\ \vdots \\ -1 \\ -\sqrt{3}/2 \\ -1/2 \\ 0 \end{pmatrix} \quad (378)$$

Similarly, we can construct:

$$|\cos_{\pi/6}\rangle = \begin{pmatrix} 1 \\ \sqrt{3}/2 \\ 1/2 \\ 0 \\ -1/2 \\ -\sqrt{3}/2 \\ -1 \\ -\sqrt{3}/2 \\ \vdots \\ -1 \\ -\sqrt{3}/2 \\ -1/2 \\ 0 \\ 1/2 \\ \sqrt{3}/2 \\ 1 \end{pmatrix} \quad (379)$$

This is how they look:



These are tall vectors (they are 49 components long!) The bra form is more suited to the way we write text in occidental culture:

$$\langle \sin_{\frac{\pi}{6}} | = \left( 0 \quad \frac{1}{2} \quad \frac{\sqrt{3}}{2} \quad 1 \quad \frac{\sqrt{3}}{2} \quad \frac{1}{2} \quad \dots \quad -1 - \frac{\sqrt{3}}{2} \quad -\frac{1}{2} \quad 0 \right), \quad (380)$$

$$\langle \cos_{\frac{\pi}{6}} | = \left( 1 \quad \frac{\sqrt{3}}{2} \quad \frac{1}{2} \quad 0 \quad -\frac{1}{2} \quad -\frac{\sqrt{3}}{2} \quad \dots \quad \frac{1}{2} \quad \frac{\sqrt{3}}{2} \quad 1 \right). \quad (381)$$

We have seen the scalar products of two vectors  $\mathbf{a}$  and  $\mathbf{b}$  in  $\mathbb{R}^n$ , it is not a big deal:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i. \quad (382)$$

We can actually see it, not as a “dot product” of vectors but as a “table product”  $(\mathbf{a}^T) \times \mathbf{b}$  where the left table is read horizontally and the right table is read vertically. The way a row multiplies a column is with this “reading” order, i.e., according to Eq. (382):

$$\mathbf{a}^T \mathbf{b} = \mathbf{a} \cdot \mathbf{b} \quad (383)$$

which means, we go through the left-hand row components in turn and multiply them with the corresponding right-hand column components, and we make the total sum. We will later see that there is a special name for these tables, namely, we will call them *matrix*, so that a vector is a particular case of  $n \times 1$  matrices and a transposed vector is a particular case of  $1 \times n$  matrices. We will also see in a future lecture what  $\mathbf{a} \mathbf{b}^T$  corresponds to. This is a scalar product of vectors (turning two vectors into a “scalar”, i.e., a number). Mathematicians like to call that an *inner product*, and they write it as:

$$\langle \mathbf{a}, \mathbf{b} \rangle \quad \text{or} \quad (\mathbf{a}, \mathbf{b}). \quad (384)$$

This is called a “bracket” product. In Physics, following Dirac, we write it as the product of the bra with the ket:

$$\langle \mathbf{a} | \mathbf{b} \rangle \tag{385}$$

so that, in particular:

$$\begin{aligned} \langle \cos_{\pi/6} | \sin_{\pi/6} \rangle &= \\ 0 + \frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4} + 0 - \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4} + 0 + \dots + 0 - \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4} + 0 &= 0. \end{aligned} \tag{386}$$

you can check there is the same number of  $\frac{\sqrt{3}}{4}$  than  $-\frac{\sqrt{3}}{4}$  and several (seventeen) zeros, all adding up to zero.

When the scalar product (or inner product) of two vectors is zero, we say that they are *perpendicular*, meaning, as for the case of geometrical arrows, that they don’t see each others. So the sine and cosine are perpendicular, in pretty much the same way that  $\hat{i} = 1, 0$  and  $\hat{j} = 0, 1$  are perpendicular. They point at different directions. This is a deep concept. It will take us time to appreciate it and through examples and applications, we will slowly start to understand why this is indeed the case. So far, what we have seen is that the “overlap” between our sine and cosine zero:

$$\langle \cos_{\pi/6} | \sin_{\pi/6} \rangle = 0 \tag{387}$$

and while we have not proved it yet, but you can check on other cases,<sup>100</sup> we have in fact, quite generally:

$$\langle \cos | \sin \rangle = 0. \tag{388}$$

which is interesting and maybe intuitively appealing. On the trigonometric circle itself, the sine and cosine are, after all, literally orthogonal to each others! We will come back to a more general and precise meaning of this fact.

Of course we can also compute the norm of the functions, by taking their overlap with themselves:

$$\|f\| \equiv \sqrt{\langle f | f \rangle} \tag{389}$$

from which we find that:<sup>101</sup>

$$\| \sin(\frac{\pi}{6}x) \| = \sqrt{24}, \tag{390a}$$

$$\| \cos(\frac{\pi}{6}x) \| = \sqrt{25}. \tag{390b}$$

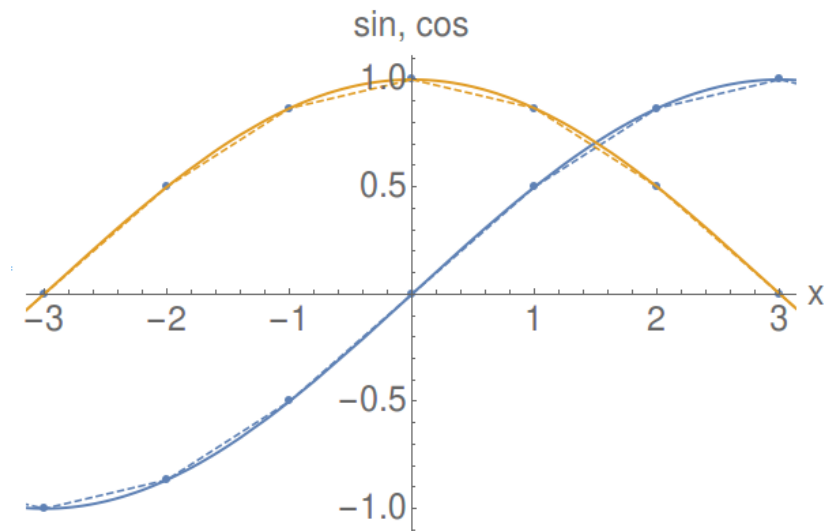
Interesting. The sine and cosine have slightly different norms... Does this sound correct? Why such an asymmetry between two functions

<sup>100</sup> Knowing that  $\cos(\pi/4) = \sin(\pi/4) = \sqrt{2}/2$ , check for instance that  $\langle \cos_{\pi/4} | \sin_{\pi/4} \rangle = 0$ . The cases  $|\sin_0\rangle$  and  $|\sin_{\pi/2}\rangle$  are even more straightforward. Why? To carry this on, you’ll need to turn to the problems.

<sup>101</sup> What do we find for the other modulations  $(0, \pi/4, 1)$ ?

that, although perpendicular, look otherwise the same: they are oscillating (just in different “dimensions”)? The number is also not anecdotal. It is clearly linked to the number of points we took (so it seems the cosine is seeing “one more point”!)

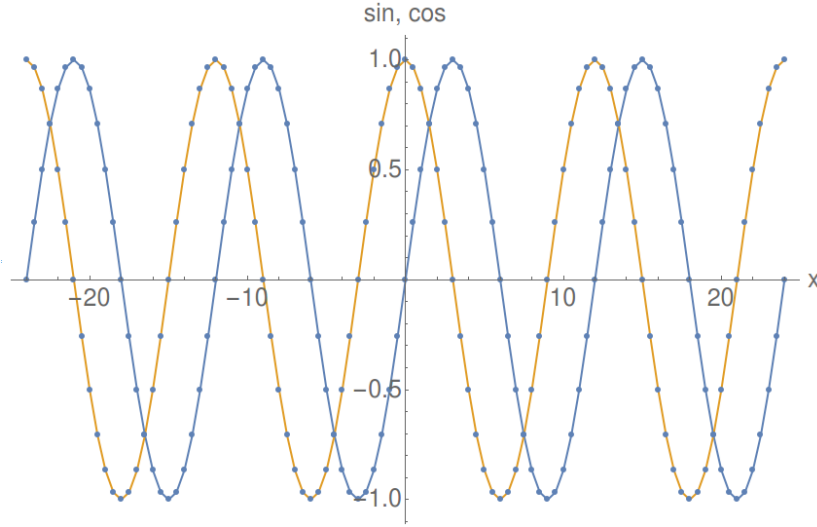
Indeed, remember what we are doing: we are in a discretized space. We only took 48 points to sample a function which is defined over  $2^{N_0}$  points! Not surprisingly, it cannot be the full story. We are making a mistake or at least an approximation somewhere. Here you see it in full view on the zoomed portion between  $-\pi$  and  $\pi$ :



It might not seem a big deal but this little difference between the discretized vector (between the points) and the full, exact function defined everywhere (solid line) is what account for this mismatch by one full unit in Eqs. (390). What is the solution then?

An obvious remedy is to take more sample points. We only took integers. Let us take the half integers as well:





Now it looks really good. However, we find, for their norm (with  $4 \times 24 + 1 = 97$  sampling points):

with 97 points :

$$\left\| \sin\left(\frac{\pi}{6}x\right) \right\| = \sqrt{48}, \tag{391a}$$

$$\left\| \cos\left(\frac{\pi}{6}x\right) \right\| = \sqrt{49}. \tag{391b}$$

To find that, we have to use the sines and cosines of multiples of  $\pi/12$ , which, we will let you check in Exercises, are given by:

$$\left\{ 0, \frac{\sqrt{3}-1}{2\sqrt{2}}, \frac{1}{2}, \frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{2}, \frac{1+\sqrt{3}}{2\sqrt{2}}, 1, \frac{1+\sqrt{3}}{2\sqrt{2}}, \frac{\sqrt{3}}{2}, \frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{\sqrt{3}-1}{2\sqrt{2}}, 0, \dots \right\} \tag{392}$$

(this is for  $\sin(k\pi/12)$  for  $k \in \mathbb{N}$ ). We should not be too surprised this increase happens, because we are doubling the number of points in the sum, adding more positive numbers (squares of a real variable) so of course the total can only increase, and in proportion to the number of points. So since we doubled the number of points, it's only fair to halve the total sum, in which case, since we take the square root:

with 97 points *normalized* :

$$\left\| \sin\left(\frac{\pi}{6}x\right) \right\| = \sqrt{48/2} \approx 4.89898, \tag{393a}$$

$$\left\| \cos\left(\frac{\pi}{6}x\right) \right\| = \sqrt{49/2} \approx 4.94975. \tag{393b}$$

Of course the same happens (and the same re normalization by the extra number of points should be made) as we discretize more and

more. This is by taking steps of  $\Delta x = 1/10$  rather than  $1/2$ :

with 481 points :

$$\| \sin(\frac{\pi}{6}x) \| = \sqrt{10 \times 24}, \quad (394a)$$

$$\| \cos(\frac{\pi}{6}x) \| = \sqrt{10 \times 24 + 1}, \quad (394b)$$

where we now used:

$$\sin(\frac{\pi}{10}) = \frac{\sqrt{5} - 1}{4}, \quad (395)$$

and by normalization:

with 481 points *normalized* :

$$\| \sin(\frac{\pi}{6}x) \| = \sqrt{24}, \quad (396a)$$

$$\| \cos(\frac{\pi}{6}x) \| = \sqrt{241/10} \approx 4.90918, \quad (396b)$$

So they seem to converge both to

$$\| \sin(\frac{\pi}{6}x) \| = \| \cos(\frac{\pi}{6}x) \| = \sqrt{24} \quad (397)$$

with the interesting feature that the sine is independent of the discretization, so it seems to be the exact result. What is 24? It is boundary  $\pm 24$  over which we defined the function. Does it seem okay that the norm of the function is proportional to the length of its domain? Clearly, yes.

We have been lucky with the sine but it would be nice to check the correct limit in general. Calling  $a \equiv -24$  and  $b \equiv +24$  the boundaries of our function space, we have computed for various discretizations  $n = 48$  (integer steps between  $a$  and  $b$ ),  $n = 2 \times 48$  and then  $n = 10 \times 48$  the following norm (squared):

$$\| \sin(\frac{\pi}{6}x) \|^2 = \sum_{k=0}^n \sin^2(\frac{\pi}{6}[a + k\frac{b-a}{n}]) \frac{b-a}{n} \quad (398)$$

where the “normalising” factor  $\frac{b-a}{n}$  is indeed 1, 1/2 and 1/10 as used above. As we take more steps, we refine our measurements and meet the now-familiar concepts of a limit again. This has been applied to a product of functions (which is by itself also a function), so if we look at the more basic case of a single function, such a sum of a sampled function weighted (averaged) by the length of the step is a particular case of a famous operation in Mathematics, known as the *Riemann sum*. Its limit when the step goes to zero is called a *Riemann integral* or, because we’ll mainly focus on this type of integrals, simply, an *integral*:

$$\boxed{\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{k=0}^n f\left(a + k\frac{b-a}{n}\right) \frac{b-a}{n}}. \quad (399)$$

Maybe this is more clear rewriting explicitly in terms of  $\Delta x$ :

$$\int_a^b f(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{k=0}^n f(a + k\Delta x) \Delta x \tag{400}$$

as it makes a clearer link with the  $\int$  notation. We will see next lecture how, in geometrical terms, this corresponds to the area below the curve of  $f$ . For now, coming back to a Riemann sum applied to inner products, we have for fine-enough discretization of our function space:

$$\langle f | g \rangle = \int f(x)g(x) dx \tag{401}$$

If you know or remember enough of calculus, we let you check that indeed,  $\langle \sin | \sin \rangle = \langle \cos | \cos \rangle = \sqrt{24}$  exactly as well as Eq. (388), now computed directly with Eq. (401). Otherwise you will have to wait for our own lectures covering these questions. We don't usually work on the interval  $[-24, 24]$ . The more common version of the sine and cosine normalisation on a  $2\pi$  interval (i.e., between  $a$  and  $b$ ) is<sup>102</sup>

$$\langle \sin_\pi | \sin_\pi \rangle = \langle \cos_\pi | \cos_\pi \rangle = \frac{b-a}{2}. \tag{402}$$

<sup>102</sup> Why is this also true for  $\langle \sin_\theta | \sin_\theta \rangle$  with  $\theta = \pi/6, \pi/4, \pi/3$ , etc?

By the way, we said that Mathematicians write this as  $\langle f, g \rangle$  and call it a bracket product. We Physicists write it as a product of a "bra" and a "ket" and also call it a *bracket*. Just, it's done in a more playful and funny way.

### Problems

#### Important trigonometric steps

You must know these important trigonometric values:  $\sin(\pi/6) = 1/2$ ,  $\sin(\pi/4) = \sqrt{2}/2$  and  $\sin(\pi/3) = \sqrt{3}/2$  (and of course  $\sin(0)$  and  $\sin(\pi/2)$  as well, out of which you can reconstruct all the "important" angles, e.g.,  $\sin(4\pi/3)$ ,  $\cos(5\pi/6)$ , etc.) This is found from an equilateral triangle, which, by definition (all angles are the same) has angle  $60^\circ$  (since  $180/3 = 60$ ).

#### Small trigonometric steps

To increase our sampling of the sine function, we have used

$$\sin\left(\frac{\pi}{12}\right) = \frac{\sqrt{3}-1}{2\sqrt{2}}. \tag{403}$$

While you must know  $\sin(\pi/6) = 1/2$ ,  $\sin(\pi/4) = \sqrt{2}/2$  and  $\sin(\pi/3) = \sqrt{3}/2$ , etc., you don't have to know Eq. (403) by heart, but you must know how to find it, if required.

Using the trigonometric identity<sup>103</sup>  $\sin(2x) = 2 \sin(x) \cos(x)$ , as well as Pythagoras, prove Eq. (403).

From Eq. (403), compute the vectors whose  $k^{\text{th}}$  component is  $\sin(k\pi/12)$  (call that  $|\sin\rangle$ ) and  $\cos(k\pi/12)$  (call that  $|\cos\rangle$ ) respectively, for  $0 \leq k \leq 12$ .

Then, compute  $\langle \sin | \cos \rangle$ ,  $\langle \sin | \sin \rangle$ ,  $\langle \cos | \cos \rangle$ . Do it exactly (the result should be exact fractions. If you're lucky—and don't fear the number 13—you should find integer solutions).

Compare with  $\int_0^{12} \sin(x)^2 dx$ ,  $\int_0^{12} \cos(x)^2 dx$  and  $\int_0^{12} \sin(x) \cos(x) dx$ .

If you are interested to know how to compute Eq. (395), you must either use ingenious (but complicated) geometric tricks, or wait that we cover trigonometric functions to see how to find this easily.

<sup>103</sup> Which you must know as well; we will see when we come back to trigonometric functions where this comes from.

### Algebra with cats

If  $\langle \uparrow | \uparrow \rangle = \langle \downarrow | \downarrow \rangle = 1$  and  $\langle \uparrow | \downarrow \rangle = 0$ , when what are  $\langle \uparrow \downarrow | \uparrow \downarrow \rangle$ ,  $\langle \uparrow \downarrow | \uparrow \rangle$  and  $\langle \uparrow \downarrow | \downarrow \rangle$ ? Do you understand now why there is a square root of 2 in Eq. (373)?

### Inner products of monomials

Assume the set of polynomials on  $[-1, 1]$ . Find the inner (braket) product between any two monomials  $x^n$  and  $x^m$  for  $n, m \in \mathbb{N}$ . When are two monomials orthogonal?

### Inner products of polynomials

Compute the inner products between two polynomials:

$$\left\langle \sum_{k=0}^n \alpha_k x^k \left| \sum_{l=0}^m \beta_l x^l \right. \right\rangle. \quad (404)$$

### Properties of the inner product

Vectors are defined together with scalars. Whenever you have vectors, there are scalars not too far. These can be rationals, reals, complex. You need them so that a scalar product (changing vectors into scalars), can be defined. For now we will work with real scalars.

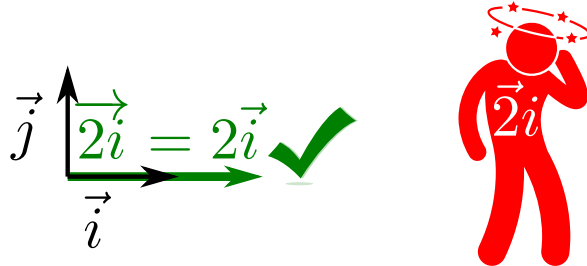
We can bring the scalar outside the vector. For geometrical vectors:

$$\vec{\alpha u} = \alpha \vec{u}, \quad (405)$$

this is called *rescaling*. With the ket notation:

$$|\alpha u\rangle = \alpha |u\rangle. \quad (406)$$

This later thing tells you that if you have a function  $u(x)$ , then  $\alpha u(x)$  is also a function (left-hand side) which is  $\alpha$  times the function  $u(x)$ . Don't bring the vector outside! That is, don't write  $|\alpha u\rangle = u|\alpha\rangle$ . While  $|\alpha\rangle$  means something (that is the constant function  $\alpha$ ), the other way around is like writing  $\vec{2i} = \vec{2}i$  which is meaningless!



Remember that a vector is normalized when its norm is 1. Normalize the monomials from the previous problem.

Prove the following important properties of the scalar product (we will write them using Dirac's notation and assuming real functions of the real variables,  $s$  but you are encouraged to also consider them in the particular cases of geometric vectors  $\vec{u}, \vec{v}$ ):

1. Symmetry:  $\langle f|g\rangle = \langle g|f\rangle$ .
2. Linearity of scaling:  $\langle \alpha f|g\rangle = \alpha \langle f|g\rangle$ .
3. Linearity of addition:  $\langle f+h|g\rangle = \langle f|g\rangle + \langle h|g\rangle$ .
4. (and same for the 2nd argument for the last two linearity conditions).
5. Positive-definiteness: If  $f \neq 0$ ,  $\langle f|f\rangle > 0$  and  $\langle f|f\rangle = 0 \Rightarrow f = 0$ .

Show that, for instance:

$$\langle f+g|f+g\rangle = \langle f|f\rangle + 2\langle f|g\rangle + \langle g|g\rangle. \quad (407)$$

And consider Eq. (404) in the light of these results.

We will see more properties of the inner products and how they generalize when complex scalars are involved. For instance, this nice property,  $\langle \alpha f|g\rangle = \alpha \langle f|g\rangle$  for  $\alpha \in \mathbb{R}$ , that you can easily prove, will be generalized later on, so we don't insist too much on it for now.



## Lecture 12: Areas.

Let us consider basic measures of basic shapes, namely perimeters and area of circles and triangles. The perimeter (or length of the contour) is easy for a triangle with sides  $a$ ,  $b$  and  $c$ , since they are straight lines:  $a + b + c$ .

The perimeter of a circle is even simpler since it is basically a definition: we call, say,  $\tau$  the circumference of a circle of unit radius (the trigonometric circle). Now if you imagine stretching linearly, that is, in the same proportion in both  $x$  and  $y$  (or any other two perpendicular directions), the medium on which these objects are drawn, you will observe that they look exactly the same, just, they are zoomed (in or out, depending on whether you stretch or compress). Such identical geometrical shapes that just differ in their scaling are called “similar”. For the circle, the perimeter thus scales linearly with the radius, by the very linear nature of the transformation that links two similar circles. Therefore, the perimeter of a circle of radius  $R$  is “simply”  $\tau R$  (simply because that is, thus, essentially a definition). We don’t use  $\tau$  but  $\pi \equiv \tau/2$ , so that a circle of radius  $R$  has perimeter:

$$2\pi R. \quad (408)$$

Now how about areas? That’s easy for triangles since we can turn them into parallelograms and then turn parallelograms into square. Therefore, the area of a triangle is:

$$\mathcal{A}_\Delta = \frac{1}{2}ah_a \quad (409)$$

where  $a$  is the base and  $h_a$  the height of a triangle relative to this base.<sup>104</sup> This is trivially determined in terms of one auxiliary quantity not readily accessible from the triangle, for reasons that should now be clear. There is a formula (and many equivalents) that provides the area directly in terms of the lengths, which is known as Heron’s formula and reads:

$$\mathcal{A}_\Delta = \sqrt{s(s-a)(s-b)(s-c)} \quad \text{where} \quad s \equiv \frac{a+b+c}{2}. \quad (410)$$

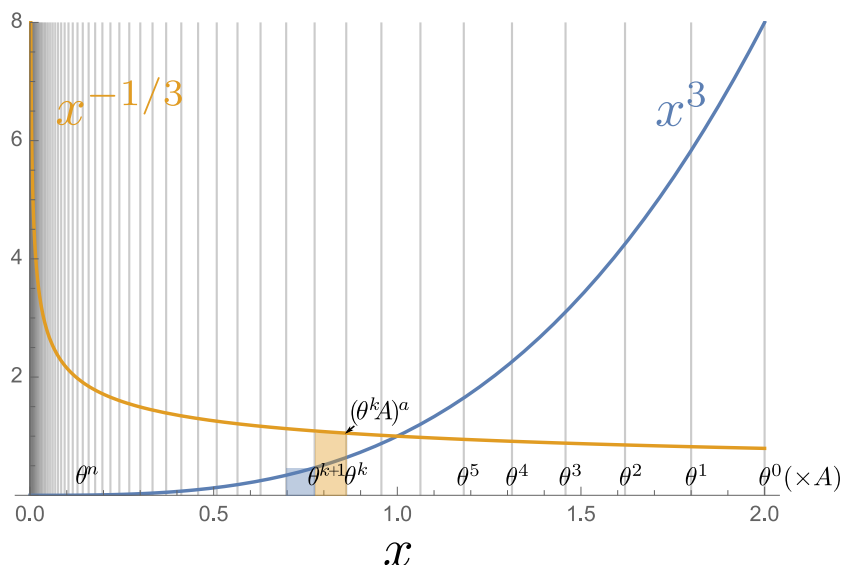
The circle is not so straightforward. Archimedes was the first to compute the area of a circle by a so-called “method of exhaustion”,

<sup>104</sup> Show that, alternatively,  $\mathcal{A}_\Delta = bh_b/2 = ch_c/2$ .

which consisted in approximating the circle by polygons, whose area is easy, as this can be expressed in terms of triangles.

How about other shapes? Archimedes, using basically the same technique (with triangles) could measure the area below a parabola, but could not extend it to other curves. Fermat found the following clever way to do it for the areas below the curve  $x^a$  between 0 and  $A$ . He considers a scaling factor  $\theta \approx 1$  but smaller than 1, ultimately to get as close to 1 as desired in a limiting process, so that he can partition the axis into the following set of intervals, here<sup>105</sup> with  $A = 2$ :

<sup>105</sup> Which value of  $\theta$  has been taken for this particular case?



that is also shown intercepting two power functions,  $x^3$  (blue) and  $x^{-1/3}$  (orange). He then computes the integral by summing the areas of the triangles formed by the gridlines and the value of the function as intercepted on the right part (two examples are shown, for  $\theta = 8$  and 9). The area of such a rectangle is, in all cases:

$$(\theta^k A - \theta^{k+1} A) \times (\theta^k A)^a \quad (411)$$

so that the total area is, in good approximation (the better the approximation the closer is  $\theta$  to 1) the sum over these rectangles, or, after simplification and bringing out of the sum terms not depending on  $k$ :

$$\mathcal{I} \approx A^{a+1} (1 - \theta) \sum_{k=0}^{\infty} \theta^{(1+a)k}. \quad (412)$$

Now, since  $\theta^{(1+a)k} = (\theta^{1+a})^k$ , we have a geometric series, which we know how to sum:

$$\mathcal{I} \approx A^{a+1} \frac{1 - \theta}{1 - \theta^{1+a}}. \quad (413)$$



we will consider the range of validity later, let us first compute the *exact* area by turning to the limit  $\theta \rightarrow 1$ :

$$\mathcal{I} = \lim_{\theta \rightarrow 1} A^{a+1} \frac{1 - \theta}{1 - \theta^{a+1}} = A^{a+1} \lim_{\theta \rightarrow 1} \frac{-1}{-(a+1)\theta^a} = \frac{A^{a+1}}{a+1}, \quad (414)$$

where we have applied l'Hopital's rule to obtain the second limit. A beautiful, general result. Back to its range of validity: the series  $\sum_k (\theta^{a+1})^k$  converges provided  $|\theta^{a+1}| < 1$ . Since  $\theta > 0$ , we need only worry for the quantity not to overcome 1. Taking the logarithm of both sides, we must have  $\ln(\theta^{a+1}) < 0$  which corresponds to  $(a+1)\ln(\theta) < 0$  but since  $\theta < 1$  then  $\ln(\theta) < 0$  therefore we need  $a+1 > 0$  or

$$a > -1. \quad (415)$$

That's a strict equality. So we can "integrate"  $1/\sqrt[3]{x}$  in this way (the curve shown in the figure) but not  $1/x$ . Why? What is so special about it? Symmetry. Consider first the following problem as a warmup.

#### AREA BELOW THE SQUARE ROOT

Knowing the area below the parabola (from Archimedes or from Fermat, cf. Eq. (414), give by geometric arguments the area below the square root. Check with Fermat's formula.

The area of  $x^a$  changes shapes. For  $a > 0$ , it partitions the rectangle of width  $A$  and height  $A^a$  into two areas, one below and the other above the curve, that are each computed by inverse functions, as in the above problem. If  $-1 < a < 0$ , the curve has no maximum (or it is  $\infty$ ) but the area below the curve remains finite, because it increases less fast than it goes towards 0 so rectangles remain of small areas that can be summed all together in an infinite series. If  $a < -1$ , the rectangles increase faster than they decrease widths so their combined areas diverge in the final sum. Alternatively, we could put a cutoff in  $x$  but take as large values on the  $x$  axis as we wish. The case  $a = -1$  is that where the balance is exact: rectangle neither decrease in size nor increase, but remain of constant area, which still fails to produce a convergent series (this is an infinite of the type  $1 + 1 + 1 + 1 + \dots$ ). This appears on both the  $x$  and  $y$  axes. We shall come back to this feature of the inverse later on.

This trick by Fermat works great for power functions because of the geometric series, but a more systematic way to carry out such measurements of area for arbitrary curves is possible. One was designed by Cauchy and formalised by Riemann, following the method we introduced in the previous Lecture when measuring the length of functions in a vector space through generalisation of the scalar

product, where we introduced the Riemann sum and took the limit:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} S_n \quad (416)$$

where we introduced

$$S_n \equiv \sum_{k=0}^n f(a + k\Delta x) \Delta x \quad (417a)$$

$$\Delta x \equiv \frac{b-a}{n}. \quad (417b)$$

Note that the notation invites to define

$$a + k_x \Delta x \rightarrow x \quad \text{and} \quad \Delta x \rightarrow dx, \quad (418)$$

where we now write  $k_x$  (could be also  $k(x)$ ) since  $k$  is the discrete variable associated to the continuous  $x$  (in the limit  $n \rightarrow \infty$ ). We can similarly define the so-called “*integral*” of  $f$  as the continuous function:

$$F_a(y) \equiv \lim_{\Delta x \rightarrow 0} \sum_{k_x=0}^{k_y} f(a + k_x \Delta x) \Delta x, \quad (419)$$

where, of course,  $y \equiv \lim_{\Delta x \rightarrow 0} (a + k_y \Delta x)$ , in which case:

$$F_a(y) \equiv \int_a^y f(x) dx. \quad (420)$$

We call  $F_a$  a cumulative function, as it indeed cumulates values of  $f$ . Now, taking the difference of two successive terms of the Riemann summation, we find

$$S_k - S_{k-1} = f(a + k\Delta x) \Delta x \quad (421)$$

or, rearranging and taking the limit

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{S_k - S_{k-1}}{\Delta x} \quad (422a)$$

$$= \lim_{\Delta x \rightarrow 0} \frac{F_a(x) - F_a(x - \Delta x)}{\Delta x} \quad (422b)$$

where we have taken a loose definition of  $F$  in Eq. (422b) since, strictly speaking, we did not take the limit yet.<sup>106</sup> As a conclusion:

$$f(x) = F'_a(x). \quad (423)$$

Note that the choice of  $a$  vanishes when taking the difference since the integral has the property

$$F_a(x) = F_b(x) + C_{ab} \quad (424)$$

where  $C_{ab}$  is a constant.<sup>107</sup> When taking the difference in Eq. (422a),

<sup>106</sup> Work out a more accurate proof.

<sup>107</sup> Prove it by specifying  $C_{ab}$ .

the same result is thus obtained regardless of the choice of the starting point ( $a$ ,  $b$  or still something else). Therefore, we do not even need to mention it and we can simply write  $F$ , the so-called “primitive” of  $f$ :

$$\boxed{F(x) = \int f(x) dx} \quad (425)$$

which is now in a so-called “indefinite integral” form, as boundaries are not specified (we also give the same variable to both sides, which is an abuse of notations since  $x$  is a proper-variable on the left and a dummy one on the right).<sup>108</sup> Then Eq. (423) takes the “more-to-the-point” form:

$$\boxed{f(x) = F'(x)}. \quad (426)$$

This result is so important, we call it the *fundamental theorem of calculus*. It links derivatives to integrals. This is actually the main way to compute integral, by realizing that primitives are anti-derivatives. Indeed, since  $(x^{n+1})' = (n+1)x^n$ , it is easy to see that<sup>109</sup>

$$\int x^n dx = \frac{1}{n+1} x^{n+1}. \quad (427)$$

Truely, we should write:

$$\int x^n dx = \frac{1}{n+1} x^{n+1} + \text{Cste}. \quad (428)$$

where Cste is the constant linked to our choice of lower-boundary for the integral. Since the derivative of a constant is zero, this disappears in the differentiation, but in full generality, it should also reappear in the integration. When integrating equations, we shall find this term is often important. If we specify the boundaries, then we have the “definite integral” which is not a function but a number, namely:

$$\boxed{\int_a^b f(x) dx = F(b) - F(a)} \quad (429)$$

Indeed, from Eq. (420),  $\int_a^b f(x) dx = F_a(b) - F_a(a)$  since  $F_a(a) = \lim_{\Delta x \rightarrow 0} f(a)\Delta x = 0$ , but then, by Eq. (424), for any choice of  $c$  we have  $F_a(b) - F_a(a) = F_c(b) + C_{ac} - [F_c(a) + C_{ac}] = F_c(b) - F_c(a)$  so that we can get rid of the origin altogether and simply write Eq. (429).<sup>110</sup> A convenient notation when dealing with definite integrals is to write:

$$\int_a^b f(x) dx = F(x) \Big|_a^b = F(b) - F(a). \quad (431)$$

An indefinite integral can be seen itself as a function (of functions), since you feed it a function and it returns another function. We state this fact because as a function, integration is a linear one:

$$\int (\alpha u + \beta v) dx = \alpha \int u dx + \beta \int v dx. \quad (432)$$

<sup>108</sup> Explain what does this mean.

<sup>109</sup> Connect this to the Fermat results above.

<sup>110</sup> What if  $b > a$  in Eq. (429). Show in particular that

$$\int_a^b f(x) dx = - \int_b^a f(x) dx \quad (430a)$$

$$\int_a^c f(x) dx + \int_c^b f(x) dx = \int_a^b f(x) dx \quad (430b)$$

for any  $c$ .

This means for instance that all polynomials are trivial:

$$\int \sum_k \alpha_k x^k dx = \sum_k \frac{\alpha_k}{k+1} x^{k+1}. \quad (433)$$

Not quite so trivial but at least do-able in general are the so-called *rational functions*, which are those that are a fraction whose numerator and denominator are a polynomial:

$$\frac{P(x)}{Q(x)}. \quad (434)$$

We assume that the degree (highest power) of  $P$  is smaller than that of  $Q$  (it can always be brought in this form through polynomial Euclidean division). The technique is to use partial fraction decomposition, that is, to rewrite, e.g.,

$$\frac{P(x)}{(a_1x + b_1)(a_2x + b_2) \cdots (a_kx + b_k)} \quad (435)$$

(with the order of  $P$  less than  $k$ ) as

$$\frac{A_1}{a_1x + b_1} + \frac{A_2}{a_2x + b_2} + \cdots + \frac{A_k}{a_kx + b_k} \quad (436)$$

and, for degenerate roots

$$\frac{P(x)}{(ax + b)^k} \quad (437)$$

(with the order of  $P$  less than  $k$ ) as

$$\frac{A_1}{ax + b} + \frac{A_2}{(ax + b)^2} + \cdots + \frac{A_k}{(ax + b)^k}. \quad (438)$$

There are as many terms as needed in the expansion to cover all the cases, so if in Eq. (435), one term is  $(a_lx + b_l)^m$  then we add all the sequences of terms in the form of Eq. (438). Once we have the coefficients  $A_k$ , then the integration is trivial since

$$\int \frac{A_k}{a_kx + b_k} dx = \frac{A_k}{a_k} \int u^{-1} du = \frac{A_k}{a_k} \ln(a_kx + b_k) \quad (439a)$$

$$\int \frac{A_k}{(a_kx + b_k)^l} dx = \frac{A_k}{a_k} \int u^{-l} du = \frac{A_k}{a_k(1-l)} (a_kx + b_k)^{1-l} \quad \text{if } l \neq 1. \quad (439b)$$

The coefficients can be computed by solving linear equations in the most general case, or, if there is no degeneracy, multiplying both sides by  $(a_lx + b_l)$  and setting  $x = -b_l/a_l$ . For example, let us compute:

$$\int \frac{x^2}{x^2 - 1} dx = \int \frac{x^2 - 1}{x^2 - 1} dx + \int \frac{1}{x^2 - 1} dx \quad (440a)$$

$$= x + \int \frac{1}{(x-1)(x+1)} dx \quad (440b)$$

where the first step was to lower the order of the numerator. We now have to make the partial fraction expansion of

$$\frac{1}{(x-1)(x+1)} = \frac{A}{x-1} + \frac{B}{x+1} \quad (441)$$

Multiplying both sides by  $x-1$  and setting  $x=1$ , we find:

$$A = \frac{1}{x+1} = \frac{1}{2} \quad (442)$$

and multiplying by  $x+1$  and setting  $x=-1$  we also find

$$B = \frac{-1}{2} \quad (443)$$

so that, finally:

$$\int \frac{x^2}{x^2-1} dx = x + \frac{1}{2}(\ln(x-1) - \ln(x+1)) + c = x + \frac{1}{2} \ln \frac{x-1}{x+1} + c. \quad (444)$$

The main strategy from there onward is your knowledge of functions. For instance, once you came to realize that

$$(\ln(\sin(x)))' = \frac{\cos(x)}{\sin(x)} \quad (445)$$

because you stumbled upon the derivative of this composite function, then you know that:

$$\int \cot(x) dx = \ln(\sin(x)) + c \quad (446)$$

which you can store in your wider baggage of general knowledge.

Which baggage is that? Let us survey:

$$\int e^x dx = e^x \quad (447a)$$

$$\int \cos x dx = \sin x \quad (447b)$$

$$\int \sin x dx = -\cos x \quad (447c)$$

$$\int \frac{dx}{1+x^2} = \arctan x \quad (447d)$$

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x \quad (447e)$$

**POWERFUL TOOLS DON'T ALWAYS GIVE SIMPLER RESULTS**

Compute the derivative of  $\frac{x}{2}\sqrt{1-x^2} + \frac{1}{2}\arcsin x$ . You now know the primitive of a very particular shape. Which one? (hint: think of the geometrical figure you obtain by appending the mirror images, for both the  $x$  and  $y$  axes.) Use the fundamental theorem of calculus and your understanding of inverse functions to recover the exact expression for one of the most important areas ever computed (originally not in this way, though).

Unfortunately, there are much less tricks for integration than for derivation, the latter being fairly automatic while integration requires more inspiration and resources. The composition rule is one of the few rare general methods of integration, and follows from the fundamental theorem: still assuming  $F' = f$ , since  $(F \circ g)' = (f \circ g)g'$  then:

$$\int (f \circ g)g' = F \circ g \quad (448)$$

which we can rewrite, using definite integrals, as

$$\int_a^b (f(g(x)))g'(x) dx = F \circ g \Big|_a^b = F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(y) dy \quad (449)$$

or, since  $\frac{dg(x)}{dx} = g'(x)$

$$\int_a^b (f(g(x)))g'(x) dx = \int_a^b f(g(x))dg(x) \quad (450)$$

so that equating Eqs. (449) and (450), we find:

$$\int_a^b (f(g(x)))g'(x) dx = \int_{g(a)}^{g(b)} f(y) dy \quad (451)$$

which is best remembered as integration by “change of variables”, namely  $y = g(x)$ ; but do not forget to change the boundaries! For example:

$$\mathcal{I} = \int_0^2 \frac{x}{1+x^2} dx \quad (452)$$

can be found by introducing  $y = x^2$  so that  $dy = 2xdx$  and

$$\mathcal{I} = \frac{1}{2} \int_0^4 \frac{1}{1+y} dy = \frac{1}{2} \ln(1+y) \Big|_0^4 = \frac{1}{2} \ln 5. \quad (453)$$

Note that, if not clear as is, we could also have made another change of variable  $z = 1 + y$  and pass by the further intermediate step  $\frac{1}{2} \int_1^5 \frac{dz}{z}$ . Note that it would have been easy to solve Eq. (446) by the substitution technique, namely, we can introduce  $u = \sin x$  and since

$du/dx = u'(x) = \cos x$ :

$$\int \frac{\cos x}{\sin x} dx = \int \frac{d \sin x}{\sin x} \quad (454a)$$

$$= \int \frac{du}{u} = \ln u = \ln \sin(x). \quad (454b)$$

There are often more than one way to do it! (when there is one). This substitution technique can also be useful for cases that you'd give no further thoughts, such as:

$$\int_a^b 3x^2(x^3 + 1)^5 dx \quad (455)$$

where we put boundaries to maintain your attention to the fact that when we change the variable, we must correspondingly also change the boundaries, to correspond to the new variable. Anyway, the point is that Eq. (455) is a polynomial, so you'd naturally solve it using Eq. (433) (do it!) However that involves expanding the polynomial itself, which involves some hefty algebra. It is faster to notice that, with  $u = x^3 + 1$ , this becomes:

$$\int_a^b 3x^2(x^3 + 1)^5 dx = \int_{a^3+1}^{b^3+1} u^5 du \quad (456a)$$

$$= \frac{1}{6} u^6 \Big|_{a^3+1}^{b^3+1} \quad (456b)$$

$$= \frac{1}{6} \left\{ (b^3 + 1)^6 - (a^3 + 1)^6 \right\} \quad (456c)$$

Here is a case where substitution seems more necessary than useful:

$$\int x^4 \sin(x^5) dx. \quad (457)$$

We put  $u = x^5$  so that  $du = 5x^4 dx$  and

$$\int x^4 \sin(x^5) dx = \frac{1}{5} \int \sin u du = -\frac{\cos u}{5} = -\frac{\cos x^5}{5}. \quad (458)$$

Or consider:

$$\int x \sqrt{2x^2 + 1} dx \quad (459)$$

with  $u = 2x^2 + 1$  we have  $du = 4x dx$  so

$$\int x \sqrt{2x^2 + 1} dx = \frac{1}{4} \int \sqrt{u} du = \frac{1}{6} u^{3/2} = \frac{1}{6} \sqrt{(2x^2 + 1)^3}. \quad (460)$$

The next trick is maybe the most important of all, and relies on the derivative of a product (and the fundamental theorem):

$$(fg)' = f'g + fg' \quad (461)$$

so that  $f'g = (fg)' - fg'$  and then, in turn:

$$\int f'g = fg - \int fg'. \quad (462)$$

This may not seem such a big gain but by so-transferring the integral from one function ( $f$ ) to the other ( $g$ ) we may actually considerably simplify the integration of a product, in particular when that involves a polynomial, since we can then lower the degree until, by repeated applications, it disappears completely, leaving us with the integration of  $f$  alone. Consider for instance,  $\int xe^x dx$  we write

$$f = x \quad f' = 1 \quad (463a)$$

$$g = e^x \quad g' = e^x \quad (463b)$$

so that:

$$\int xe^x dx = xe^x - \int e^x = (x-1)e^x \quad (464)$$

plus a constant. Remember that integration requires imagination and inspiration. For instance, let us compute the primitive of the logarithm:

$$\int \ln(x) dx. \quad (465)$$

This can be done integrating by part, which is funny because there is only one function! We write:

$$f = \ln x, \quad f' = \frac{1}{x}, \quad (466a)$$

$$g = x, \quad g' = 1, \quad (466b)$$

so that:

$$\int \ln x dx = x \ln x - \int dx = x(\ln x - 1) \quad (467)$$

We often have to iterate the integration by parts rule, in which case there's a neat bookkeeping device known as "tic-tac-toe"<sup>111</sup> that consists in storing the successive derivatives and antiderivatives in columns (1st column has the derivatives, 2nd has the primitives) and take sum of diagonal products, alternating the sign (that we can store in a 3r column). This is easier seen than explained. Say you want to compute:

$$x^4 \sin x \quad (468)$$

(which we will need to compute the variance of the position of a quantum particle in a box), then you write derivatives on the left,

<sup>111</sup> From some movie, "Stand and deliver".



primitive on the right

$$x^4 \qquad \sin x \qquad + \qquad (469a)$$

$$4x^3 \qquad -\cos x \qquad - \qquad (469b)$$

$$12x^2 \qquad -\sin x \qquad + \qquad (469c)$$

$$24x \qquad \cos x \qquad - \qquad (469d)$$

$$24 \qquad \sin x \qquad + \qquad (469e)$$

$$0 \qquad -\cos x \qquad - \qquad (469f)$$

$$\qquad \qquad \qquad + \qquad (469g)$$

and the primitive is found by making diagonal products as follows:

$$\int x^4 \sin x \, dx = -x^4 \cos x + 4x^3 \sin x + 12x^2 \cos x - 24x \sin x - 24 \cos x + c. \quad (470)$$

Sometimes iterating the integration by parts brings us back to where we started, which allows us to find the solution. An important example is:

$$\int \sin^2 x \, dx \quad (471)$$

which we can solve by taking:

$$f = \sin x, \qquad f' = \cos x, \qquad (472a)$$

$$g = -\cos x, \qquad g' = \sin x, \qquad (472b)$$

so that

$$\int \sin^2 x \, dx = -\sin x \cos x + \int \cos^2 x \, dx = -\sin x \cos x + \int (1 - \sin^2 x) \, dx \quad (473)$$

where we have used  $\sin^2 x + \cos^2 x = 1$ , so that, bringing all  $\sin^2$  terms on the same side:

$$\int \sin^2 x \, dx = \frac{1}{2}(x - \sin x \cos x) = \frac{x}{2} - \frac{\sin(2x)}{4}, \quad (474)$$

if we prefer to get rid of the product (using  $\sin 2x = 2 \sin x \cos x$ ).

Note that integrals of functions of the type:

$$\int \sin^n x \cos^m x \, dx \quad (475)$$

with either  $n$  or  $m$  (or both) odd, can be easily integrated by substitution, by extracting one sine or cosine to integrate the rest, getting rid of the "unwanted" trigonometric functions through Pythagoras. For example:

$$\begin{aligned} \int \sin^3 x \, dx &= \int \sin^2 x \sin x \, dx = -\int (1 - \cos^2 x) d \cos x = \\ &= -\int (1 - u^2) du = -\cos x + \frac{1}{3} \cos^3 x. \end{aligned} \quad (476)$$

As you see, lots of creativity in a simple game of integration.

We conclude with that shape that gave Fermat so many problems. We have seen that  $(\ln x)' = 1/x$ , therefore

$$\int \frac{dx}{x} = \ln x \quad (477)$$

(read +Cste), which is why we could not integrate it before since:

$$\int_0^y \frac{dx}{x} = \ln x \Big|_0^y \quad (478)$$

but the logarithm diverges at 0, and we speak of a logarithmic divergence in a case like (478) where we try to integrate an inverse! We can however (and obviously) integrate over intervals not enclosing 0, in which case:

$$\int_a^b \frac{dx}{x} = \ln b - \ln a = \ln \frac{b}{a} = \ln \frac{\alpha b}{\alpha a} = \ln(\alpha b) - \ln(\alpha a) = \int_{\alpha a}^{\alpha b} \frac{dx}{x} \quad (479)$$

the two extremes of which say that the area enclosed by two intervals, one of which has been stretched, are the same, which is not surprising because this  $x$  stretching is compensated by an  $y$  compression, so the area is conserved (you can see this better back to Riemann sums if needed where the compensation is easy to see). This shows how the inverse function exhibits the sort of “similarity” we started the lecture with but of a more sophisticated type.

## Problems

### Heron's formulae

There are many variations to Heron's formula, Eq. (410). Show that it can also be given as:

$$\mathcal{A}_\Delta = \frac{1}{4} \sqrt{(a+b+c)(-a+b+c)(a-b+c)(a+b-c)} \quad (480a)$$

$$= \frac{1}{4} \sqrt{(a^2+b^2+c^2)^2 - 2(a^4+b^4+c^4)} \quad (480b)$$

$$= \frac{1}{4} \sqrt{4a^2b^2 - (a^2+b^2-c^2)^2} \quad (480c)$$

### Familiar Primitives

Can you go down the list and provide all primitives? (check by derivating your proposals, assume  $\alpha \in \mathbb{R}$ .)

1.  $x^n$  for  $n \in \mathbb{N}$ .
2.  $x^n$  for  $n \in -\mathbb{N}^*$ .
3.  $\sqrt{x}$ .
4.  $\cos(\alpha x)$ .

5.  $\sin(ax)$ .

6.  $e^{ax}$ .

*New Primitives*

Differentiate the following functions and propose formulas for new primitives (we remind that  $(\arcsin x)' = 1/\sqrt{1-x^2}$ ):

1.  $\tan(x)$ .

3.  $-\ln(\cos(x))$ .

2.  $\sqrt{1-x^2} + x \arcsin(x)$ .

4.  $-x + x \ln x$ .

*Definite integrals*

Compute the areas between  $y = 0$  and the following curves in the range  $a = 0$  and  $b = 1$ , counting as negative those areas that fall below the  $y = 0$  line:

1. 1

5.  $\sin(\pi x)$

9.  $\sqrt{x}$

2.  $x$

6.  $\cos(\pi x)$

10.  $\tan(x)$

3.  $x^2$

7.  $\sinh(x)$

11.  $\ln(x)$

4.  $x^n$  for  $n \in \mathbb{N}$ .

8.  $\cosh(x)$

12.  $\arcsin(x)$

*Back to where we started*

Use integration by part to compute

$$\int e^\theta \cos(\theta) d\theta \quad \text{and} \quad \int e^\theta \sin(\theta) d\theta. \quad (481)$$

*768 in the middle*

Also by part, compute:

$$\int x^4 e^{x/2} dx. \quad (482)$$

This one is lengthy but automatic. Check your result. If you like catchy titles, you can also do that by evaluating the primitive at the origin.

*Beauty and the Beast*

This is a polynomial, so it's in principle straightforward, but the expansion is tedious. Do it the hard way to practice, and compare with an integration by part to see how much faster this is:

$$\int (-2x + 3)(x + 7)^5 dx. \quad (483)$$

For the following one, compare brute-force and change of variable:

$$\int 9x^2(3x^3 + 5)^3 dx. \quad (484)$$

As you can see, it pays to be clever.

*By parts*

Use integration by parts to compute:

$$\int x^k \sin(x) dx \quad (485)$$

for  $0 \leq k \leq 4$  (we will need them in quantum mechanics; so if you keep your result somewhere, this is saving you from solving it again later). How about:

$$\int x^k e^x dx, \quad (486)$$

for any  $k \in \mathbb{N}$ .

*Tic-Tac-Trick*

Using the Tic-Tac-Toe mnemonic, and the following table of derivatives and “antiderivatives”

$$x^2 \quad \frac{x}{\sqrt{(1-x^2)^3}} \quad + \quad (487a)$$

$$2x \quad \frac{1}{\sqrt{1-x^2}} \quad - \quad (487b)$$

$$2 \quad \arcsin(x) \quad + \quad (487c)$$

$$0 \quad \sqrt{1-x^2} + x \arcsin(x) \quad - \quad (487d)$$

$$+ \quad (487e)$$

compute

$$\int \frac{x^3}{\sqrt{(1-x^2)^3}} dx. \quad (488)$$

No arcsin allowed in your final result! Do you see any other way to compute this integral? (with integrals of this sort, the substitution  $x \rightarrow \sin \theta$  can give good results. In this case, it would bring you to the solution, passing by  $\int \frac{\sin^3 \theta}{\cos^2 \theta} d\theta$  which in itself also requires some work), so this knowledge of the primitives seems the fastest way forward).

## Lecture 13: Basis.

We now delve deeper into our vector spaces. Exploring the function space further (we'll be in  $[-1,1]$  this time), we have seen that the sine and cosine are orthogonal,  $\langle \sin | \cos \rangle = 0$  (taking, e.g.,  $\sin(\pi x)$ ), which is interesting and maybe intuitively appealing. We still need to get to a better understanding of what that means. If you look at the monomials, they can also be orthogonal:

$$\langle x^n | x^m \rangle = \int_{-1}^1 x^{n+m} dx = \frac{1}{n+m+1} x^{n+m+1} \Big|_{-1}^1 = \frac{1 - (-1)^{n+m+1}}{n+m+1} \quad (489)$$

and this is either  $2/(n+m+1)$  if  $n+m+1$  is odd and 0 if  $n+m+1$  is even. In the latter case, that is, if  $n+m$  is odd, the scalar product cancels and we have orthogonal functions. This happens iff monomials have different parities (Exercise). So this "independence" seems to be linked to different parities. This was the case of sine and cosine, by the way. So we have one strong aspect of orthogonality here: the *parity*, i.e., being odd or even. You could of course complain that most functions are neither so this seems a bit specific. But actually, most functions are in fact a bit of both, since any function whatsoever can be decomposed as<sup>112</sup>

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}, \quad (490)$$

where  $f$  is here written as an even plus odd function. Some functions only happen to be fully in one direction (odd) or the other (even). Most functions are indeed looking at both, otherwise, orthogonal aspects!<sup>113</sup> The symmetric and antisymmetric part of the exponential are important functions, known as hyperbolic sine  $\sinh$  and hyperbolic cosine  $\cosh$ :

$$\exp(x) = \cosh(x) + \sinh(x), \quad (491)$$

with, from Eq. (490)

$$\cosh(x) \equiv \frac{e^x + e^{-x}}{2} \quad \text{and} \quad \sinh(x) \equiv \frac{e^x - e^{-x}}{2}. \quad (492)$$

<sup>112</sup> Show this.

<sup>113</sup> Parity is thus an important, fundamental aspect of functions. Show the following: the sum of two odd functions is odd, sum of two even is even, product of two even or two odd is even, sum of even and odd is neither, product of even and odd is odd, composition of two even functions is even, of two odd is odd, of even and odd is odd, composition of any function with an even one is even (but not vice-versa).

But that is only scratching the surface, as parity is a binary thin (odd, even), while we will see there is an infinite (for function spaces) number of orthogonal functions to any monomial, or to any function in general. What are these? The function space is a big place, we cannot hope to understand it by looking at all its constituents. Instead, like explorer who made first contact with some specimens or individuals from the area, we ask: “bring us to your leaders”, that is, to the important functions that represent all the others. We shall stick to analytical functions in the following, to keep the discussion simple and civilized (avoiding infinities, discontinuities, etc.)

The important vectors in a vector space are called the *basis vectors*. They are the set of vectors from which one can reconstruct all the other vectors of the space. The number of these vectors gives the *dimensionality* of the vector space. If there is a finite number of these, like in the case of the geometrical vectors, we are therefore in a *finite-dimensional vector space*:

$$\{\hat{i}, \hat{j}\}, \quad \text{2D space} \quad (493a)$$

$$\{\hat{i}, \hat{j}, \hat{k}\}, \quad \text{3D space} \quad (493b)$$

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5\}, \quad \text{5D space} \quad (493c)$$

etc. The set of basis vector can be infinite, such as is the case for a function space, for which a basis is, e.g.,

$$\mathcal{B} = \{1, x, x^2, \dots, x^k, \dots\} \quad (494)$$

the monomials. The basis here is countably infinite. It is not a basis for *all* the functions, because  $1/x$  for instance cannot be written in terms of monomials, but if we restrict to analytical functions, i.e., those that can be written as a Taylor series, then clearly (494) is a basis, since for any such function  $f$ , we can write it equivalently as a vector  $(\alpha_1, \alpha_2, \dots)^T$ , which collects the coefficients of  $f$  in its Taylor expansion, or, more aptly the *coordinates* of  $f$  in its basis of monomials. In Mathematical terms:

$$(\forall f \in \mathcal{B})(\exists (\alpha_k)_k \in \mathbb{R}^\infty)(f = (\alpha_1, \alpha_2, \dots)^T) \quad (495)$$

This is true from Taylor’s formula:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k. \quad (496)$$

So monomials form a basis. Note, *a basis*, not “the basis”, as you can chose other representatives. Let us see how this works out in the familiar geometrical 2D space. A good basis is Eq. (493a). It is so good that we call it the “canonical basis”, in the sense that it is the natural, compelling, or obvious one to chose.

Any vector can be written as a *linear combination* of vectors from the basis. For instance, the vector  $\vec{s} = 3\hat{i} - 2\hat{j}$  is 3 times  $\hat{i}$  plus (combined to) 2 times  $\hat{j}$ . So we can write:

$$\vec{s} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}. \quad (497)$$

We need two numbers to define the vector, so it lives in a two-dimensional space.

But we could choose another basis, for instance, if we choose the “ugly basis”

$$\mathcal{U} = \{\vec{u} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \vec{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\} \quad (498)$$

then every vector can be written as a linear superposition of these too. How? A general theme of linear algebra is that if we solve the problem for the basis vectors, we solved it for all vectors. So let us find how  $\hat{i}$  and  $\hat{j}$  themselves satisfy this, i.e., is it true they can be written as linear combinations of  $\vec{u}$  and  $\vec{v}$ ? We know that  $\vec{u} = \hat{i} - \hat{j}$  and that  $\vec{v} = \hat{i} + 2\hat{j}$  so it is easy to invert this, which gives:

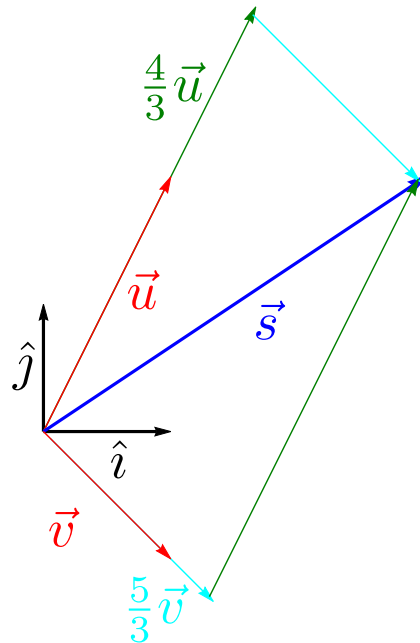
$$\hat{i} = \frac{2}{3}\vec{u} + \frac{1}{3}\vec{v} \quad (499a)$$

$$\hat{j} = -\frac{1}{3}\vec{u} + \frac{1}{3}\vec{v} \quad (499b)$$

From which we find:

$$\vec{s} = \frac{4}{3}\vec{u} + \frac{5}{3}\vec{v}. \quad (500)$$

This is how the various vectors involved combine each others:



You see how  $\vec{s}$  is obtained both as  $(4/3)\vec{u} + (5/3)\vec{v}$  and as  $(5/3)\vec{v} + (4/3)\vec{u}$ , which are the same thanks from commutativity of addition, but leads to two different (parallelogram) constructions in space. You should make sure you are happy and comfortable with these statements;

What makes the canonical basis “nice” is that its vectors are *orthonormal*, that is, they are:

1. Orthogonals to each other.
2. Normed (length 1).

That makes it easy to pick-up a coefficient, say, if you want the  $\hat{i}$  coefficient of  $\vec{s}$  then you simply compute

$$\hat{i} \cdot \vec{s} = \hat{i} \cdot (3\hat{i} + 2\hat{j}) = 3\hat{i} \cdot \hat{i} + 2\hat{i} \cdot \hat{j} = 3 \quad (501)$$

since  $\hat{i} \cdot \hat{i} = 1$  (normalization) and  $\hat{i} \cdot \hat{j} = 0$  (orthogonality). In an orthonormal basis, any vector is then easily expressed in its “column form”:

$$\vec{s} = (\hat{i} \cdot \vec{s})\hat{i} + (\hat{j} \cdot \vec{s})\hat{j} \quad (502)$$

or, equivalently:

$$\vec{s} = \begin{pmatrix} \hat{i} \cdot \vec{s} \\ \hat{j} \cdot \vec{s} \end{pmatrix}, \quad (503)$$

or, equivalently in  $\mathbb{R}^n$ :

$$\mathbf{s} = \begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{s} \\ \mathbf{e}_2 \cdot \mathbf{s} \\ \vdots \\ \mathbf{e}_n \cdot \mathbf{s} \end{pmatrix}. \quad (504)$$

From that we can easily see that, if we have an orthonormal basis of functions  $b_i$ , then:

$$|f\rangle = \begin{pmatrix} \langle b_0|f\rangle \\ \langle b_1|f\rangle \\ \vdots \\ \langle b_k|f\rangle \\ \vdots \end{pmatrix}. \quad (505)$$

The monomial basis is an “ugly” basis for functions. Monomials are neither normed nor all orthogonal to each others. Our program for the rest of the lecture is then to find such an orthonormal basis of functions. And then we can write:

$$f(x) = \langle b_0|f\rangle b_0(x) + \langle b_1|f\rangle b_1(x) + \dots \quad (506)$$

or, if we want to stay fully within:

$$|f\rangle = \langle b_0|f\rangle |b_0\rangle + \langle b_1|f\rangle |b_1\rangle + \dots \quad (507)$$



Of course, this we would write with the  $\Sigma$  notation:

$$\boxed{|f\rangle = \sum_{k=0}^{\infty} \langle b_k|f\rangle |b_k\rangle .} \quad (508)$$

Make sure you understand this esoteric notation  $\langle b_k|f\rangle |b_k\rangle$ . We have two objects here:  $\langle b_k|f\rangle$  which is a scalar, this is the coefficient of the vector, and  $|b_k\rangle$ , the said vector. Since we can move the scalar around (remember,  $2\vec{i} = \vec{i}2$ ), Eq. (508) can be written as

$$|f\rangle = \sum_{k=0}^{\infty} |b_k\rangle \langle b_k|f\rangle \quad (509)$$

and given that  $|f\rangle$  does not depend of the sum's dummy index  $k$ , it can be factored out from it, namely:

$$|f\rangle = \left( \sum_{k=0}^{\infty} |b_k\rangle \langle b_k| \right) |f\rangle . \quad (510)$$

Using algebra and this powerful Dirac notation, we have naturally introduced a new object:

$$|b_k\rangle \langle b_k| \quad (511)$$

This is known as a “*projector*”, i.e., it takes the vector  $|f\rangle$  and turns it into a vector  $|b_k\rangle$  rescaled by how much  $|f\rangle$  was overlapping with  $|b_k\rangle$ , which is, as we already know,  $\langle b_k|f\rangle$ . Note that, from Eq. (510) (which is true for all  $|f\rangle$ ) we can see that the sum of all projectors is unity:

$$\sum_{k=0}^{\infty} |b_k\rangle \langle b_k| = \mathbf{1} . \quad (512)$$

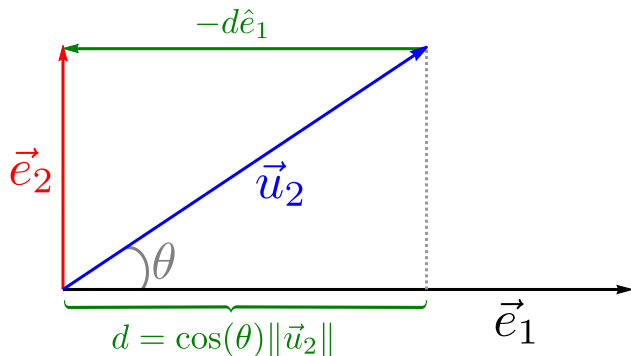
To find an orthonormal basis, we will first find an orthogonal one (normalization is easy, just divide by the square root of the norm) and this we do by removing projected parts of one vectors onto the others. We will illustrate the process with familiar vectors (geometrical ones) at which point it will be easy to generalize it all vectors (abstract ones). We start from what we have:

$$\vec{e}_1 \equiv \vec{u}_1 . \quad (513)$$

For this vector, which is the only one we have so far, there is little else to do than to normalize it, which is easy (divide by the norm):

$$\hat{e}_1 \equiv \frac{\vec{e}_1}{\|\vec{e}_1\|} . \quad (514)$$

Then the second vector  $u_2$  is not linearly dependent on  $\vec{e}_1$ , since it is part of a basis, but it can have some component in the direction of  $\vec{e}_1$ , which is the problem, as we want it to be completely independent. Therefore, we need to remove this component, which is the quantity  $d$  in the diagram below:



And from elementary relations in a right-angle triangle, we find:

$$d = \cos(\theta) \|\vec{u}_2\| \quad (515)$$

so we need to remove this vector,  $-d\hat{e}_1$  from  $\vec{u}_2$ . Note that must not forget to use the normalized  $\vec{e}_1$  (that's what the hat means):

$$\vec{e}_2 \equiv \vec{u}_2 - \cos(\theta) \|\vec{u}_2\| \hat{e}_1 \quad (516)$$

so we only need get rid of the angle  $\theta$  to express everything in terms of vectors, which is easy to do because  $\cos(\theta)$  appears explicitly in the scalar product:

$$\vec{u}_2 \cdot \vec{e}_1 = \cos(\theta) \|\vec{u}_2\| \|\vec{e}_1\| \quad (517)$$

so finally, inserting Eq. (517) into Eq. (516) and using Eq. (514), we find:

$$\boxed{\vec{e}_2 \equiv \vec{u}_2 - (\hat{e}_1 \cdot \vec{u}_2) \hat{e}_1} \quad (518)$$

which is clear in itself:  $\vec{e}_2$  is  $\vec{u}_2$  minus the projection of  $\vec{u}_2$  on  $\vec{e}_1$ .

That's elementary vector algebra but, as it typical of linear algebra, can involve a lot of symbols. So one needs much practice to get used to it.

The same happens for the 3rd vector, we remove successively all its components along the previous vectors in the list, using the same procedure, so that, this yields:

$$\vec{e}_1 \equiv \vec{u}_1 \quad (519a)$$

$$\vec{e}_2 \equiv \vec{u}_2 - (\hat{e}_1 \cdot \vec{u}_2) \hat{e}_1 \quad (519b)$$

$$\vec{e}_3 \equiv \vec{u}_3 - (\hat{e}_2 \cdot \vec{u}_3) \hat{e}_2 - (\hat{e}_1 \cdot \vec{u}_3) \hat{e}_1 \quad (519c)$$

$$\vdots \quad (519d)$$

$$\vec{e}_k \equiv \vec{u}_k - \sum_{i=1}^{k-1} (\hat{e}_i \cdot \vec{u}_k) \hat{e}_i. \quad (519e)$$

This is called the *Gram-Schmidt* orthogonalization procedure. You'll probably find it enlightening if you practice (Exercises). The

same happens for functions. We let you orthogonalize (and normalize, thus, orthonormalise) geometric vectors in exercises, but we do orthogonalize monomials together. The procedure is exactly the same, but notations are different, so let us first rewrite Eqs. (519) with Dirac's notations. The main difference is that while we can write  $\hat{u}$  the normalized  $\vec{u}$  vector, there is no such handy notation for Dirac vectors (where hats are kept for something else, namely, operators). Instead, we write  $|b_0\rangle / \sqrt{\langle b_0|b_0\rangle}$  the normalized vector

$$|b_0\rangle \equiv |1\rangle, \quad (520a)$$

$$|b_1\rangle \equiv |x\rangle - \frac{\langle b_0|x\rangle}{\langle b_0|b_0\rangle} |b_0\rangle \quad (520b)$$

$$|b_2\rangle \equiv |x^2\rangle - \frac{\langle b_0|x^2\rangle}{\langle b_0|b_0\rangle} |b_0\rangle - \frac{\langle b_1|x^2\rangle}{\langle b_1|b_1\rangle} |b_1\rangle, \quad (520c)$$

$$|b_3\rangle \equiv |x^3\rangle - \frac{\langle b_0|x^3\rangle}{\langle b_0|b_0\rangle} |b_0\rangle - \frac{\langle b_1|x^3\rangle}{\langle b_1|b_1\rangle} |b_1\rangle - \frac{\langle b_2|x^3\rangle}{\langle b_2|b_2\rangle} |b_2\rangle, \quad (520d)$$

$$\vdots \quad (520e)$$

$$|b_k\rangle \equiv |x^k\rangle - \sum_{i=0}^{k-1} \frac{\langle b_i|x^k\rangle}{\langle b_i|b_i\rangle} |b_i\rangle \quad (520f)$$

Actually with Dirac's notation it is even more enlightening to rewrite Eqs. (520) as:

$$|b_0\rangle \equiv |1\rangle, \quad (521a)$$

$$|b_1\rangle \equiv |x\rangle - \frac{|b_0\rangle\langle b_0|}{\langle b_0|b_0\rangle} |x\rangle \quad (521b)$$

$$|b_2\rangle \equiv |x^2\rangle - \frac{|b_0\rangle\langle b_0|}{\langle b_0|b_0\rangle} |x^2\rangle - \frac{|b_1\rangle\langle b_1|}{\langle b_1|b_1\rangle} |x^2\rangle, \quad (521c)$$

$$\vdots \quad (521d)$$

$$|b_k\rangle \equiv |x^k\rangle - \sum_{i=0}^{k-1} \frac{|b_i\rangle\langle b_i|}{\langle b_i|b_i\rangle} |x^k\rangle \quad (521e)$$

and if we now factorize  $|x^k\rangle$  in Eq. (521e), we have:

$$\boxed{|b_k\rangle \equiv \left( \mathbf{1} - \sum_{i=0}^{k-1} \frac{|b_i\rangle\langle b_i|}{\langle b_i|b_i\rangle} \right) |x^k\rangle.} \quad (522)$$

You can probably even understand this expression: our Gram-Schmitted  $k$ -th vector is obtained from the  $k$ -th vector of the initial basis by stripping it down from all its projections from the previous orthonormalized vectors.

Here, maybe for the first time, we have an example where Dirac's notation makes things simpler with abstract vectors than the usual notation with familiar and simpler vectors!

Anyway, enough of too-abstract, general considerations. Let us compute! We can write directly the functions and keep Dirac's notation for the overlaps, that we'll soon enough transform into integrals (Eqs. 523 are to be read simultaneously with Eqs. 524):

$$b_0(x) \equiv 1, \quad (523a)$$

$$b_1(x) \equiv x - \frac{\langle 1|x \rangle}{\langle 1|1 \rangle} = x, \quad (523b)$$

$$b_2(x) \equiv x^2 - \frac{\langle 1|x^2 \rangle}{\langle 1|1 \rangle} - \frac{\langle x|x^2 \rangle}{\langle x|x \rangle} x = x^2 - \frac{1}{3}, \quad (523c)$$

$$b_3(x) \equiv x^3 - \frac{\langle 1|x^3 \rangle}{\langle 1|1 \rangle} - \frac{\langle x|x^3 \rangle}{\langle x|x \rangle} x - \frac{\langle x^2 - \frac{1}{3}|x^3 \rangle}{\langle x^2 - \frac{1}{3}|x^2 - \frac{1}{3} \rangle} (x^2 - \frac{1}{3}) = x^3 - \frac{3}{5}x, \quad (523d)$$

$$b_4(x) \equiv x^4 - \frac{\langle 1|x^4 \rangle}{\langle 1|1 \rangle} 1 - \frac{\langle x|x^4 \rangle}{\langle x|x \rangle} x - \frac{\langle x^2 - \frac{1}{3}|x^4 \rangle}{\langle x^2 - \frac{1}{3}|x^2 - \frac{1}{3} \rangle} (x^2 - \frac{1}{3}) - \frac{\langle x^3 - \frac{3}{5}x|x^4 \rangle}{\langle x^3 - \frac{3}{5}x|x^3 - \frac{3}{5}x \rangle} (x^3 - \frac{3}{5}x) = x^4 - \frac{6}{7}x^2 + \frac{3}{35}. \quad (523e)$$

since

$$\langle 1|x \rangle = \int_{-1}^1 x dx = \frac{1}{2}x^2 \Big|_{-1}^1 = 0 \quad (524a)$$

$$\langle 1|1 \rangle = \int_{-1}^1 dx = 1 \Big|_{-1}^1 = 2 \quad (524b)$$

$$\langle 1|x^2 \rangle = \langle x|x \rangle = \int_{-1}^1 x^2 dx = \frac{1}{3}x^3 \Big|_{-1}^1 = \frac{2}{3} \quad (524c)$$

$$\langle 1|x^3 \rangle = \langle x|x^2 \rangle = \int_{-1}^1 x^3 dx = \frac{1}{4}x^4 \Big|_{-1}^1 = 0 \quad (524d)$$

$$\langle x|x^3 \rangle = \int_{-1}^1 x^4 dx = \frac{1}{5}x^5 \Big|_{-1}^1 = \frac{2}{5} \quad (524e)$$

$$\langle x^2 - \frac{1}{3}|x^3 \rangle = \text{integral of odd function} = 0 \quad (524f)$$

$$\langle x^2 - \frac{1}{3}|x^2 - \frac{1}{3} \rangle = \int_{-1}^1 (x^2 - \frac{1}{3})^2 dx = \frac{8}{45} \quad (524g)$$

$$\langle x^2 - \frac{1}{3}|x^4 \rangle = \int_{-1}^1 (x^2 - \frac{1}{3})x^4 dx = \frac{16}{105} \quad (524h)$$

$$\langle x^3 - \frac{3}{5}x|x^4 \rangle = \text{integral of odd function} = 0 \quad (524i)$$

$$\langle x^3 - \frac{3}{5}x|x^3 - \frac{3}{5}x \rangle = \int_{-1}^1 (x^3 - \frac{3}{5}x)^2 dx = \frac{8}{175} \quad (524j)$$

where integrals in Eqs. (524g), (524h) and (524j) are, at least in principle, straightforward, since they are mainly primitives of polynomials.<sup>114</sup> Note that we didn't need compute Eq. (524j) for our current results but we would need it for extending the procedure to higher orders.<sup>115</sup> So lots of integrals there, but easy (technical) stuff. Note

<sup>114</sup> Check the result.

<sup>115</sup> Find  $b_5(x)$ .

that they are not normalized. You can do that (Exercises), but actually a different convention is used, namely, such polynomials are so-called “standardized”, meaning that they are rescaled so that  $b_k(1) = 1$  for all  $k$ . In this form, they are known as *Legendre polynomials* and are important in Physics, for instance to describe multipole expansion, such as the electric field of a point charge, or for their role in solving the Schrödinger equation of Hydrogen.

Now it is easy to decompose any function in terms of polynomials, using the Legendre basis and Equation (508), e.g., for  $\sin(\pi x)$  to order 3:

$$|\sin\rangle = \frac{\langle 1|\sin\rangle}{\langle 1|1\rangle} + \frac{\langle x|\sin\rangle}{\langle x|x\rangle}x + \frac{\langle x^2 - (1/3)|\sin\rangle}{\langle x^2 - (1/3)|x^2 - (1/3)\rangle}(x^2 - (1/3)) \quad (525)$$

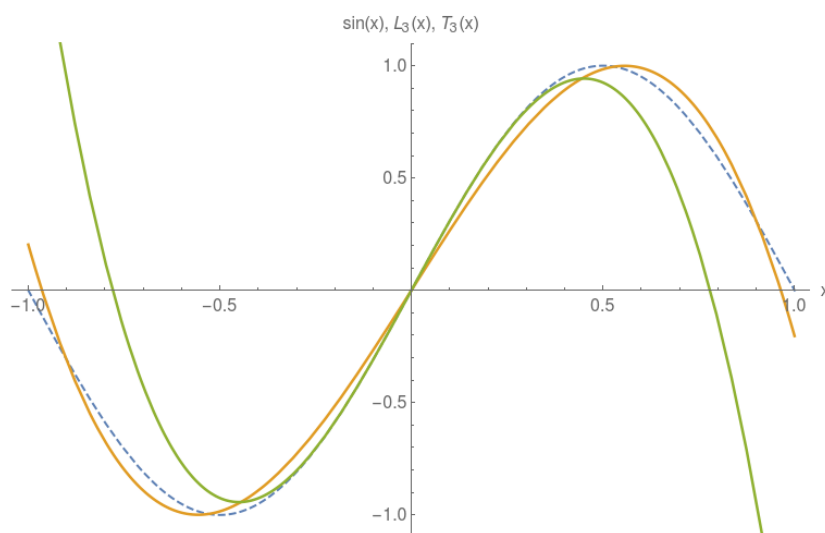
so go ahead, compute the integrals (Exercises), you should find:

$$L_3(x) = \left( \frac{3}{\pi} - \frac{21(\pi^2 - 15)}{2\pi^3} \right)x + \frac{35(\pi^2 - 15)}{2\pi^3}x^3 \quad (526)$$

which is  $2.69229x - 2.8956x^3$ . Compare this to Taylor’s polynomial, also to order 3:

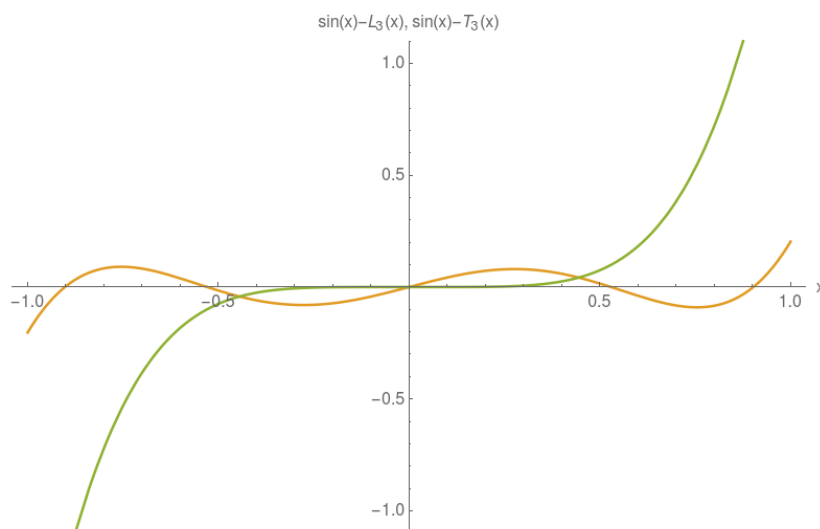
$$T_3(x) = \pi x - \frac{\pi^3}{6}x^3 \quad (527)$$

which is much simpler to compute. It is  $3.14159x - 5.16771x^3$  so quite different weights for the basis vectors! This is how they compare:



The exact sine is the dashed blue, the Taylor approximation Eq. (527) is the green and the Legendre approximation Eq. (526) is the orange. Clearly the Legendre approximation gives a much better *overall* agreement. The Taylor expansion is still typically more useful

to a Physicist because it gives a much better *local* agreement (towards zero). This you can see in the error each function makes:



And in Physics, it is typically better to be very good for a reduced span than fairly okay for a broader range.

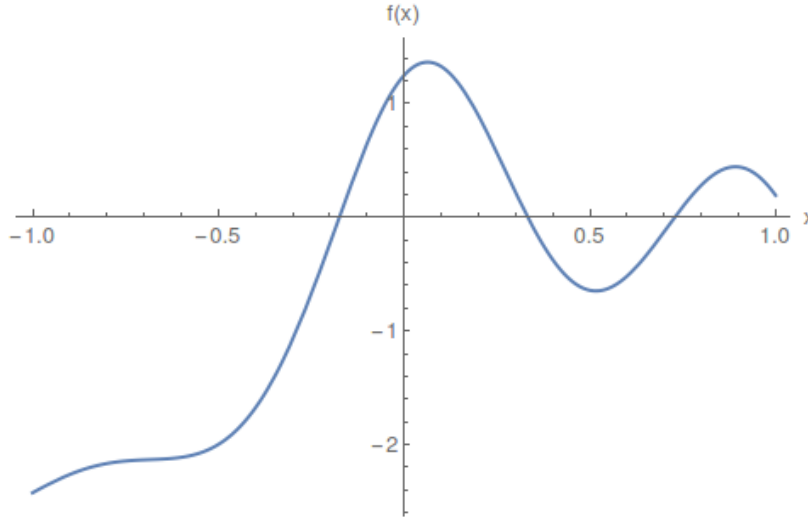
## Problems

### Odd properties

Show that  $n + m$  is odd iff  $n$  and  $m$  have different parity. Since an integer which is not odd is even, this means that  $n + m$  is even iff  $n$  and  $m$  have the same parity. What about the product? (when is  $nm$  odd/even?) Summarize these results as logic table (the product, for instance, is an OR table).

We define a function as odd if  $f(-x) = -f(x)$  and even if  $f(x) = f(-x)$ . Unlike integers, if a function is not odd, it is not compulsorily even, and vice-versa. Are there functions that can be odd and even simultaneously? (answer is yes)

Decompose this function into its odd and even components: (you have to work with what you get, which is here just the graph of it)



### Maclaurin parity

Show that odd functions only feature odd powers in their Maclaurin expansion and similarly that even functions only feature even powers.

### Training basis

It's good exercise to try your hands at manipulating various basis. On top of the "ugly basis"  $\mathcal{U}$  from the text, let us use the training basis  $\mathcal{T}$ :

$$\mathcal{T} = \{\vec{t}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \vec{t}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}\} \quad (528)$$

Normalize the vectors. Express  $\vec{s} = 3\hat{i} - 2\hat{j}$  in the training basis. Express the training basis vectors in the ugly basis, and vice-versa. Explain why

$$\{\vec{u}, \vec{t}_1\} \quad (529)$$

is *not* a basis, while, e.g.,  $\{\vec{v}, \vec{t}_2\}$  is.

When we have various basis, if we want to express vectors as column of numbers, we need to keep track of which basis they are expressed in. This can be done by keeping the basis as a subscript, e.g.,

$$\hat{i} + \hat{j} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{u} + \vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{U}}, \quad \vec{t}_1 + \vec{t}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{T}}. \quad (530)$$

(we do not put a subscript on the canonical basis, because it is *the* special one). Of course:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{T}} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{U}}. \quad (531)$$

In fact, express  $(1, 1)_U^T$  as a column vector in the canonical and training basis, then  $(1, 1)_T^T$  as a column vector in the canonical and training basis and finally  $(1, 1)^T$  as a column vector in the ugly and training basis. It is then easy to show that the difference of the two sides of Eq. (531) is not zero.

It's probably worth to check your geometric skills as well. Plot  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}_U$  on the graph (for instance directly on your Lectures note; you will have to add the training basis vectors) and then  $\begin{pmatrix} 3 \\ 4 \end{pmatrix}_T$  and then construct the sum of these two vectors geometrically. Then with a ruler estimate its length as well as its coordinates in the canonical basis. Compare to the exact result obtained from algebra:

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}_U + \begin{pmatrix} 3 \\ 4 \end{pmatrix}_T. \quad (532)$$

### Trigonometric basis

Consider the following basis in  $\mathbb{R}^2$ :

$$\mathcal{R}_\theta = \{ \vec{e}_1 = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \}. \quad (533)$$

Is it an orthonormal basis?

### Gram-Schmidt in 2D

Orthonormalize this basis:

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\} \quad (534)$$

### Gram-Schmidt in 3D

Orthonormalize this basis:

$$\mathcal{B}_3 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\} \quad (535)$$

### Gram-Schmidt in 4D

Orthonormalize this basis:

$$\mathcal{B}_4 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right\} \quad (536)$$



### Legendre Polynomials

Compute the next-order Legendre Polynomial, i.g., Eq. (523f)  $b_5(x) = x^5 - \dots$ . Standardize ( $b_k(1) = 1$ ) and Normalize ( $\langle b_k | b_k \rangle = 1$ ) all of them. Check that the “normalization” condition for the standardized functions are

$$\langle b_k | b_l \rangle = \frac{2}{2k+1} \delta_{kl}. \quad (537)$$

### Legendre Exponential

Find the best Legendre approximation to the exponential function.

### Out of powers

All the polynomials of orders less than  $n$  can be exactly decomposed in term of Legendre polynomials of orders up to  $n$ . For instance, express  $x^3$  in terms of Legendre polynomials, i.e., find  $\alpha_k$  such that

$$x^3 = \sum_{k=0}^4 \alpha_k b_k(x). \quad (538)$$

If one considers a polynomial of higher order, then it becomes approximates again. For instance, express  $x^5$  in terms of Legendre polynomials (i.e., replace  $x^3$  by  $x^5$  in Eq. (538)) and check that in the canonical basis, this yields:

$$x^5 \approx -\frac{5}{21}x + \frac{10}{9}x^3. \quad (539)$$

If you worked out  $b_5$ , you can consider  $x^6$  or  $x^7$  instead.

What happens in the case of a Taylor expansion?



## Lecture 14: Linear functions.

We have seen functions and we have seen vectors, as well as how vectors can be decomposed in terms of other “representative” vectors (the basis vectors). Now we bring all these notions together and show the considerable simplifications that arise for a large class of functions, namely, *linear functions*.

A function  $f$  (of vectors) is linear iff:

1. it is “homogeneous”, meaning that the scaling of a vector can be taken out of the function:

$$f(\alpha \mathbf{v}) = \alpha f(\mathbf{v}). \quad (540)$$

2. it is “additive”, meaning that the addition of vectors can be taken out of the function:

$$f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}). \quad (541)$$

These two conditions together can be brought together in a single condition:

$$\boxed{f(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v})}. \quad (542)$$

It is easy to prove (do it) that Eq. (542) implies Eqs. (540) and (541), and vice-versa.<sup>116</sup>

<sup>116</sup> We asked and ask again: do it!

Linear functions look boring, because the particular case people have in mind before they deal with abstract vector space is from  $\mathbf{R}$  to  $\mathbf{R}$  (numbers can be treated as 1D vectors) and then it reduces to:

$$f(x) = ax, \quad (543)$$

where  $a$  is a constant. Note that not even  $f(x) = ax + b$  is linear (if  $b \neq 0$ )! It’s a line, sure, but it satisfies neither Eq. (540) nor (541). Indeed,  $f(\alpha x) = a(\alpha x) + b = \alpha(ax + b) - \alpha b + b = \alpha f(x) + b(1 - \alpha)$  which is true only in the trivial cases  $0 = 0$  ( $\alpha = 0$ ) and  $f(x) = f(x)$  ( $b = 1$ ) but is false in general. Also,  $f(x + y) = a(x + y) + b = ax + b + ay + b - b = f(x) + f(y) - b$  which is also false unless  $b = 0$ .

So which need do we have to devote a full lecture to such silly functions? We will devote much more than one lecture. They are very

important. And in general abstract spaces, they are not so simple as Eq. (543). Geometrically, they can be seen as stretching and rotating the Euclidean space (Exercises), possibly reducing its dimensionality. In general, they correspond to actions or operations which are not dependent on the order with which you perform the task, or on the previous history, or on any absolute reference point. Many things, especially fundamental ones, turn out to be linear.

Here is a concrete example within a full Mathematical context, the special case in the arrow-vector space (space of geometrical arrows) of our projection operator  $|a\rangle\langle a|$  on the (abstract) vector  $|a\rangle$ :<sup>117,118</sup>

$$\Pi_{\vec{a}}(\vec{v}) \equiv (\hat{a} \cdot \vec{v})\hat{a}. \quad (544)$$

Here we are projecting any vector given to the function on the vector  $\vec{a}$ .  $\Pi_{\vec{a}}$  is linear. Proof: We can either do it with Eqs. (540) and (541) in two steps or with Eq. (542) in one step, depending on how complicated this is:

$$\Pi_{\vec{a}}(\alpha\vec{u} + \beta\vec{v}) = [\hat{a} \cdot (\alpha\vec{u} + \beta\vec{v})]\hat{a}, \quad (545a)$$

$$= [\alpha\hat{a} \cdot \vec{u} + \beta\hat{a} \cdot \vec{v}]\hat{a}, \quad (545b)$$

$$= \alpha\Pi_{\vec{a}}(\vec{u}) + \beta\Pi_{\vec{a}}(\vec{v}). \quad (545c)$$

QED. A linear function has no other constraints than Eq. (542), in particular, its domain and codomain don't have to be the same. For instance, the "pick-up coefficient"  $P_k(\vec{v}) \equiv \vec{v} \cdot \hat{e}_k$  of the vector  $\vec{v}$  expressed in the basis  $\mathcal{A} = \{\hat{e}_1, \hat{e}_2, \dots\}$  takes a vector and returns a scalar.<sup>119</sup>

Another basic way of constructing linear functions is to "compose" them. This is both easy and important so we leave this one to you.

#### COMPOSITION OF LINEAR IS LINEAR

Show that if  $f : A \rightarrow B$  and  $g : C \rightarrow A$  are both linear functions, then  $(f \circ g) : C \rightarrow B$  is also linear.

Nevertheless, many functions are naturally nonlinear. Here is a function of a vector which is *not* linear: the norm! Proof:  $\|\alpha\vec{u}\| = \sqrt{\alpha\vec{u} \cdot \alpha\vec{u}} = |\alpha|\sqrt{\vec{u} \cdot \vec{u}} = \alpha\|\vec{u}\|$ . So this almost works. However, the presence of the absolute value breaks the sought property, so this is not linear. It fails in an even more dramatic way on the additivity property:

$$\|\vec{u} + \vec{v}\| = \sqrt{(\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v})} \quad (546a)$$

$$= \sqrt{\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} + 2\vec{u} \cdot \vec{v}} \quad (546b)$$

which is clearly not the same as  $\|\vec{u}\| + \|\vec{v}\| = \sqrt{\vec{u} \cdot \vec{u}} + \sqrt{\vec{v} \cdot \vec{v}}$  if  $\vec{u} \cdot \vec{v} \neq uv$ . If it is, then the vectors are aligned (we would say "colinear"). In general, this does not hold, as you can show that by taking

<sup>117</sup> Play with, e.g.,  $\Pi_{2\hat{i}+\hat{j}}$ .

<sup>118</sup> Note  $\hat{a}$ . Give the exact "abstract" counterpart of Eq. (544) if  $|a\rangle$  is not normalized.

<sup>119</sup> Play with  $P_k$  and show that it is linear. Provide its abstract vector counterpart.

a counter-example, for instance when  $\vec{u}$  and  $\vec{v}$  are orthogonal, since then this gives  $\sqrt{a+b} = \sqrt{a} + \sqrt{b}$  and clearly  $\sqrt{\cdot}$  is not linear as a function  $\mathbb{R} \rightarrow \mathbb{R}$ ). Actually, in general, one has:

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|. \quad (547)$$

This is known as the *triangle inequality*.

Why are linear functions so important? Because they make a great team together with basis vectors. Namely, if you know what a function does to all the basis vectors, you know what it does to the entire space.

Proof: It is mainly to show that Eq. (542) extends to arbitrary linear combinations (rescaled sums):

$$f\left(\sum_{k=1}^n \alpha_k \vec{v}_k\right) = \sum_{k=1}^n \alpha_k f(\vec{v}_k). \quad (548)$$

We leave the proof of this to you (Exercise). This is true for any vectors, so in particular for the basis

$$\mathcal{A} = \{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\} \quad (549)$$

if we know all

$$\vec{f}_k \equiv f(\vec{a}_k), \quad (550)$$

for  $1 \leq k \leq n$ , then clearly we know what is  $f(\vec{u})$  for any  $\vec{u}$ . We just decompose  $\vec{u}$  as

$$\vec{u} = \sum_{k=1}^n u_k \vec{a}_k \quad (551)$$

from which we find, from Eq. (548):

$$f(\vec{u}) = \sum_{k=1}^n u_k \vec{f}_k. \quad (552)$$

This means the function  $f$  in a  $n$ -dimensional vector space essentially reduces to  $n$  vectors (the  $\vec{f}_k$ ). Note again that the vectors  $\vec{f}_k$  are not necessarily of the same type as the  $\vec{a}_k$ . They can live in different spaces, that is, the domain and codomain of the function do not have to be the same or have the same dimension:

$$f : A \rightarrow B, \quad (553)$$

of codomain the space  $B$  with basis

$$\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m\}. \quad (554)$$

For instance,  $P_{\hat{e}_k}$  was precisely such a function from  $\mathbb{R}^2$  or  $\mathbb{R}^3$  to  $\mathbb{R}$  ("transforming" a vector into a number). When working in the same

space (or within “subspaces”) they generalize the idea of “stretching and rotations” of the Euclidean space.

Just like there is a nice and natural way to write a vector like Eq. (551) as a column of its coefficients (for a given basis):

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad (555)$$

there is also a nice and natural way to “write” a linear function, i.e., storing all the vectors  $\vec{f}_k \equiv f(\vec{a}_k)$ . These are themselves columns of numbers when written in their basis:

$$\vec{f}_k \equiv f(\vec{a}_k) = \sum_{i=1}^m f_{ik} \vec{b}_i = \begin{pmatrix} f_{1k} \\ f_{2k} \\ \vdots \\ f_{mk} \end{pmatrix}, \quad (556)$$

so we keep the full and complete information about the linear application by writing

$$f \equiv \left( \left( \begin{pmatrix} f(\vec{a}_1) \\ \vdots \\ f(\vec{a}_1) \end{pmatrix} \right) \left( \begin{pmatrix} f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \end{pmatrix} \right) \cdots \left( \begin{pmatrix} f(\vec{a}_n) \\ \vdots \\ f(\vec{a}_n) \end{pmatrix} \right) \right). \quad (557)$$

Removing the spurious parenthesis, this gives us a table of numbers:

$$f \equiv \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix}. \quad (558)$$

The table (558) we call a “matrix”. Note that, by construction, it has  $m$  rows (given by  $\dim(\mathcal{B})$ ) and  $n$  columns (given by  $\dim(\mathcal{A})$ ). We call it a  $m \times n$  matrix. This is what a matrix is, by the way, it’s not merely a table of numbers, it is, really, a linear function. We also say “linear application”, because we like to see  $f(\vec{u})$  as “applying”  $f$  to the vector. In this case, we call  $f$  an “operator” (that’s another name for a function), and to stress the point, to attract attention on the operator character of this object, sometimes we put a caret on top. This is useful when we deal with abstract vectors, because we want to distinguish operators from scalars (constants). Then we would write something like:

$$\hat{f}|u\rangle = f(\vec{u}). \quad (559)$$

The left-hand side of Eq. (559) is literally

$$\hat{f}|u\rangle = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}. \quad (560)$$

A matrix is also, beyond merely a table of numbers, an “operator”. From Eqs. (552) and (556), we:

$$f(\vec{u}) = \sum_{k=1}^n u_k \sum_{i=1}^m f_{ik} \vec{b}_i = \sum_{i=1}^m \sum_{k=1}^n f_{ik} u_k \vec{b}_i = \begin{pmatrix} \sum_k^n f_{1k} u_k \\ \sum_k^n f_{2k} u_k \\ \vdots \\ \sum_k^n f_{mk} u_k \end{pmatrix} \quad (561)$$

we see that the product of a matrix and a vector is obtained by scanning horizontally the matrix and making the product with the corresponding vertical vector component. Alternatively, one can see this is also weighting each column from the left-hand table by the corresponding coefficient (line-counting) from the right-hand column.

**That is how we multiply a matrix with a vector.**

We have seen above that the composition of linear functions is itself linear. Let us assume the following configuration:

$$A \xrightarrow{f} B \xrightarrow{g} C \quad (562)$$

meaning that given two matrices for  $\hat{f}$  and  $\hat{g}$  there exists another matrix  $\hat{h}$  such that  $h = g \circ f$ . We now find this matrix. We now have three vector spaces  $A$ ,  $B$  and  $C$  with respective bases:

$$A = \{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}, \quad (563a)$$

$$B = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m\}, \quad (563b)$$

$$C = \{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_l\}, \quad (563c)$$

of respective dimensions  $n$ ,  $m$  and  $l$  (possibly identical, possibly these are even the same spaces).

In these spaces, the linear function  $f$  is the matrix given by Eqs. (557–558) and  $g$  is also a matrix, given by

$$\hat{g} \equiv \left( \left( \begin{pmatrix} g(\vec{b}_1) \\ \vdots \\ g(\vec{b}_m) \end{pmatrix} \right) \left( \begin{pmatrix} g(\vec{b}_2) \\ \vdots \\ g(\vec{b}_m) \end{pmatrix} \right) \cdots \left( \begin{pmatrix} g(\vec{b}_m) \\ \vdots \\ g(\vec{b}_m) \end{pmatrix} \right) \right) \quad (564)$$

that is, a matrix of dimensions  $l \times m$ :

$$\hat{g} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \cdots & \ddots & \vdots \\ g_{l1} & g_{l2} & \cdots & g_{lm} \end{pmatrix}, \quad (565)$$

where, for all  $1 \leq k \leq m$ :

$$\vec{g}_k \equiv g(\vec{b}_k) = \sum_{j=1}^l g_{jk} \vec{c}_j = \begin{pmatrix} g_{1k} \\ g_{2k} \\ \vdots \\ g_{lk} \end{pmatrix}. \quad (566)$$

This is the definition of  $g$ : its columns are the vectors of  $g$  applied to the basis vectors.

We want to compute  $g \circ f$ , so it's enough, since it is linear, to know what it does to the basis vectors of its domain (i.e., the domain of  $f$ ), that is, we need to compute, for all  $1 \leq k \leq n$ :

$$(g \circ f)(\vec{a}_k) \quad (567)$$

but this is "simply"

$$g(f(\vec{a}_k)) = g(\vec{f}_k) \quad (568a)$$

$$= g\left(\sum_{i=1}^m f_{ik} \vec{b}_i\right) \quad (568b)$$

$$= \sum_{i=1}^m f_{ik} g(\vec{b}_i) \quad (568c)$$

$$= \sum_{i=1}^m f_{ik} \sum_{j=1}^l g_{ji} \vec{c}_j \quad (568d)$$

$$= \sum_{j=1}^l \sum_{i=1}^m g_{ji} f_{ik} \vec{c}_j \quad (568e)$$

$$= \sum_{j=1}^l h_{jk} \vec{c}_j. \quad (568f)$$

Eq. (568a) is from Eq. (550).

Eq. (568b) is from Eq. (556).

Eq. (568c) is because  $g$  is linear.

Eq. (568d) is Eq. (566) (changing  $k$  for  $i$  as indices get shuffled around)

Eq. (568e) is changing the order of summation and commuting the coefficients.



Eq. (568f) is introducing a new object:

$$h_{jk} \equiv \sum_{i=1}^m g_{ji} f_{ik} \quad (569)$$

What Eq. (568f) tells us is that  $g \circ f$  is a matrix whose column-vectors for the  $k$ th basis vector of its domain  $((g \circ f)(\vec{a}_k))$  (lhs of Eq. (568a)) in the basis of its codomain  $(\vec{c}_j)$  are given by  $h_{jk}$ . By definition, this means  $h_{jk}$  is the  $j$ th row,  $k$ th column of the linear application  $g \circ f$ , so that its matrix is found to be  $\hat{h}$  with components:

$$\boxed{\hat{h}_{jk} = \sum_{i=1}^m \hat{g}_{ji} \hat{f}_{ik}} \quad (570)$$

meaning that it is simply

$$\boxed{\hat{h} = \hat{g} \hat{f}} \quad (571)$$

where the product of matrices is defined as given by Eq. (570). This is called “concatenation”. In this way, you see that product of matrices  $(l \times m) \times (m \times n)$  gives a  $l \times n$  matrix: the middle index got “concatenated”. Which is what should happen if we just remove the middle space! **That is how we multiply matrices together.** Now we go directly from  $\mathcal{A}$  to  $\mathcal{C}$ . More importantly, we have now proved that *compositions of linear functions are given by the product of their matrices!* Extremely powerful, because matrix product are easy to compute.

We now look at one things that linear applications are good at doing, and which is very important in physics: the change of basis (in Physics this becomes a “change of the reference frame”). For instance, for problems which have polar symmetry, it is better to work in a rotated basis.

Everything that we did was in some choice of a basis for the respective spaces, namely, Eqs. (563). That is what gave us in turn column-form expressions such as (555) or (556). Of course this is central to the specific expression that the linear applications take in the form of matrices, e.g., Eq. (558). In itself, changing basis does not change a vector or an application, although it changes its component when laying them down in a column or a table.

Let us consider one vector  $\vec{u}$  written in two bases, where it is expressed as a unique linear combination:

$$\vec{u} = \sum_{i=1}^n \alpha_i \vec{a}_i, \quad (572a)$$

$$\vec{u} = \sum_{j=1}^n \epsilon_j \vec{e}_j. \quad (572b)$$

We can write these vectors as columns of their coordinates, but since we have several basis, now we must be careful to remember which

basis we are working with. We can keep track by writing it down as a subscript, i.e.:

$$\vec{u} \rightarrow \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{\mathcal{A}} \quad \text{and} \quad \vec{u} \rightarrow \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{\mathcal{E}}. \quad (573)$$

But here, *careful*, the same vector ( $\vec{u}$ ) expressed with different coordinates, shouldn't lead to equate these two columns:

$$\text{Wrong!} \quad \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{\mathcal{A}} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{\mathcal{E}}, \quad (574)$$

so we cannot put an equal sign in Eq. (573), since two vectors are equal when their components are equal, i.e., we would have  $\alpha_i = \epsilon_i$  for all  $i$ . In fact the relationship between the  $\alpha_i, \epsilon_j$  coefficient is an important one which we are interested in. The *change of basis matrix* provides this useful service of telling you how to pass from one, to the other. We write it as  $C_{\mathcal{A} \rightarrow \mathcal{E}}$ , and this is thus defined as

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{\mathcal{A}} = C_{\mathcal{E} \rightarrow \mathcal{A}} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{\mathcal{E}} \quad (575)$$

Note that we can also make the change in the other direction:

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{\mathcal{E}} = C_{\mathcal{A} \rightarrow \mathcal{E}} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{\mathcal{A}}. \quad (576)$$

which, if inserted in Eq. (576) (or the other way around) leads to:

$$C_{\mathcal{A} \rightarrow \mathcal{E}} C_{\mathcal{E} \rightarrow \mathcal{A}} = C_{\mathcal{E} \rightarrow \mathcal{A}} C_{\mathcal{A} \rightarrow \mathcal{E}} = \mathbb{1} \quad (577)$$

with  $\mathbb{1}$  the identity. Since this holds for all vectors  $\vec{u}$ , by definition, we then have

$$C_{\mathcal{E} \rightarrow \mathcal{A}} = C_{\mathcal{A} \rightarrow \mathcal{E}}^{-1} \quad (578)$$

and vice-versa:

$$C_{\mathcal{A} \rightarrow \mathcal{E}} = C_{\mathcal{E} \rightarrow \mathcal{A}}^{-1}, \quad (579)$$

where the inverse is, so far, in the sense of composite of functions (not yet in the algebraic sense, although we shall see next Lecture that for linear functions these two notions coincide!)

So it remains to compute this change-of-basis matrix. Let us go from Eq. (572a) to (572b) by substituting the  $\vec{e}_j$ . These, as all vectors, can be written in terms of the  $\vec{a}_i$  vectors since they are a basis, so that, for all  $1 \leq j \leq n$ , there exist  $c_{ij}$  such that:

$$\vec{e}_j = \sum_{i=1}^n c_{ij} \vec{a}_i. \quad (580)$$

Substituting Eq. (580) into Eq. (572b), we find:

$$\vec{u} = \sum_{j=1}^n \epsilon_j \sum_{i=1}^n c_{ij} \vec{a}_i, \quad (581a)$$

$$= \sum_{i=1}^n \sum_{j=1}^n c_{ij} \epsilon_j \vec{a}_i, \quad (581b)$$

which, by comparison (and identification) with Eq. (572a), yields:

$$\boxed{\alpha_i = \sum_{j=1}^n c_{ij} \epsilon_j} \quad (582)$$

This is the  $i$ th row equation of Eq. (575), so we have now characterized the matrix change-of-basis-from- $\mathcal{E}$ -to- $\mathcal{A}$ . It is, according to Eq. (582), composed of the components of the  $\mathcal{E}$  vectors expressed in the  $\mathcal{A}$  basis, cf. Eq. (580), that is to say:

$${}_{\mathcal{E} \rightarrow \mathcal{A}} C \equiv \left( \begin{array}{c} \left( \begin{array}{c} \vec{e}_1 \\ \end{array} \right)_{\mathcal{A}} \\ \left( \begin{array}{c} \vec{e}_2 \\ \end{array} \right)_{\mathcal{A}} \\ \cdots \\ \left( \begin{array}{c} \vec{e}_n \\ \end{array} \right)_{\mathcal{A}} \end{array} \right), \quad (583a)$$

$$(583b)$$

$${}_{\mathcal{A} \rightarrow \mathcal{E}} C \equiv \left( \begin{array}{c} \left( \begin{array}{c} \vec{a}_1 \\ \end{array} \right)_{\mathcal{E}} \\ \left( \begin{array}{c} \vec{a}_2 \\ \end{array} \right)_{\mathcal{E}} \\ \cdots \\ \left( \begin{array}{c} \vec{a}_n \\ \end{array} \right)_{\mathcal{E}} \end{array} \right). \quad (583c)$$

This times there are as many columns than rows, because we are operating in the same vector space (changing basis there). This is a very important type of matrices, called, the *square matrices*.

We will work out examples of how vectors transform when changing bases in the tutorials. For now, we carry on to consider the case of what happens to a linear function, that is defined from one space

(say of dimension  $n$ ) into another (of dimension  $m$ ), when we change basis in the respective spaces!

This is the generic scheme:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \mathcal{A}, \mathcal{E} & & \mathcal{B}, \mathcal{D} \end{array} \quad (584)$$

Here too, because we work with different basis, we have to be mindful which vectors the function is working with when considering its components. While we wrote simply  $\hat{f}$  for the matrix form of  $f$ , when changing bases, we need be mindful of this and will thus write

$$\hat{f}_{\mathcal{A} \rightarrow \mathcal{B}} \quad (585)$$

the matrix that takes vectors expressed in basis  $\mathcal{A}$  and brings them to the space of vectors expressed in basis  $\mathcal{B}$ .

By definition of a linear application, it is fully defined when we know its action on all the basis vectors. This is what we have already seen, but now we keep track of which basis we are working with:

$$\hat{f}_{\mathcal{A} \rightarrow \mathcal{B}} = \left( \left( \begin{array}{c} f(\vec{a}_1) \\ \mathcal{B} \end{array} \right) \left( \begin{array}{c} f(\vec{a}_2) \\ \mathcal{B} \end{array} \right) \cdots \left( \begin{array}{c} f(\vec{a}_n) \\ \mathcal{B} \end{array} \right) \right). \quad (586)$$

If we want to express the codomain vectors in the  $\mathcal{D}$  basis, instead, we have:

$$\hat{f}_{\mathcal{A} \rightarrow \mathcal{D}} = \left( \left( \begin{array}{c} f(\vec{a}_1) \\ \mathcal{D} \end{array} \right) \left( \begin{array}{c} f(\vec{a}_2) \\ \mathcal{D} \end{array} \right) \cdots \left( \begin{array}{c} f(\vec{a}_n) \\ \mathcal{D} \end{array} \right) \right). \quad (587)$$

but since for all  $1 \leq i \leq n$

$$\left( \begin{array}{c} f(\vec{a}_i) \\ \mathcal{D} \end{array} \right) = \begin{array}{c} \mathcal{C} \\ \mathcal{B} \rightarrow \mathcal{D} \end{array} \left( \begin{array}{c} f(\vec{a}_i) \\ \mathcal{B} \end{array} \right) \quad (588)$$

Now, one can see that:

$$\begin{aligned} \left( \begin{pmatrix} f(\vec{a}_1) \\ \vdots \\ f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_n) \end{pmatrix} \right)_{\mathcal{D}} &= \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} + \cdots + \begin{pmatrix} 00 & 00 \\ 00 & 00 \\ \vdots & \vdots \\ 00 & 00 \end{pmatrix} \right\} \quad (589) \end{aligned}$$

where we decompose the right-hand-side matrix into  $n$  matrices with only one nonzero column (all the rest of the matrix being zero). For each matrix in isolation, it is easy to see that, e.g., for the 2nd column:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_2) \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (590)$$

where the last equality follows from the rules of matrix algebra.<sup>120</sup> <sup>120</sup> Show it.

Then, for all vectors put back together to reassemble  $\hat{f}_{A \rightarrow C}$ , we have the matrix expression

$$\boxed{\hat{f}_{A \rightarrow D} = \begin{pmatrix} C \\ B \rightarrow D \end{pmatrix} \hat{f}_{A \rightarrow B}} \quad (591)$$

This is the relationship to change the codomain basis.

In contrast, if we work with vectors expressed in the  $\mathcal{E}$  basis of the domain space, then, by the same definition of a linear function, we have:

$$\hat{f}_{\mathcal{E} \rightarrow \mathcal{B}} = \left( \begin{pmatrix} f(\vec{e}_1) \\ \vdots \\ f(\vec{e}_2) \\ \vdots \\ f(\vec{e}_n) \end{pmatrix} \right)_{\mathcal{B}} \quad (592a)$$

$$\hat{f}_{A \rightarrow \mathcal{B}} = \left( \begin{pmatrix} f(\vec{a}_1) \\ \vdots \\ f(\vec{a}_2) \\ \vdots \\ f(\vec{a}_n) \end{pmatrix} \right)_{\mathcal{B}} \quad (592b)$$

While two column vectors expressed in different basis are not equal, cf. Eq. (574), since they refer to the same vector, the linear function must return the same value (in some basis, say  $\mathcal{B}$ ) for them:

$$f\left(\begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix}\right) = f\left(\begin{pmatrix} \vec{a}_i \\ \mathcal{E} \end{pmatrix}\right) \quad (593)$$

where we write  $\begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix}$  and  $\begin{pmatrix} \vec{a}_i \\ \mathcal{E} \end{pmatrix}$  the column-vectors for  $\vec{a}_i$  in these two bases. Of course the first one is full of zeros except at the  $i$ th row where it is 1. Equation (593) is true when providing the correct-basis matrix, i.e.,

$$\hat{f}_{\mathcal{A} \rightarrow \mathcal{B}}\left(\begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix}\right) = \hat{f}_{\mathcal{E} \rightarrow \mathcal{B}}\left(\begin{pmatrix} \vec{a}_i \\ \mathcal{E} \end{pmatrix}\right) \quad (594)$$

the result being a column vector in  $\mathcal{B}$ , namely, those in Eq. (592b), as is trivial for the lhs of Eq. (594). For the rhs, we note that the  $\mathcal{A}$  basis vectors, like all other vectors, are expressed in the  $\mathcal{E}$  basis by applying the change-of-basis matrix:

$$\begin{pmatrix} \vec{a}_i \\ \mathcal{E} \end{pmatrix} = \overset{\mathcal{C}}{\mathcal{A} \rightarrow \mathcal{E}} \begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix} \quad (595)$$

which, substituted in Eq. (594), yields

$$\hat{f}_{\mathcal{A} \rightarrow \mathcal{B}}\left(\begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix}\right) = \hat{f}_{\mathcal{E} \rightarrow \mathcal{B}} \overset{\mathcal{C}}{\mathcal{A} \rightarrow \mathcal{E}} \begin{pmatrix} \vec{a}_i \\ \mathcal{A} \end{pmatrix} \quad (596)$$

and since this is true for all basis vectors, it is, by linear superposition, true for all vectors, so that:

$$\boxed{\hat{f}_{\mathcal{A} \rightarrow \mathcal{B}} = \hat{f}_{\mathcal{E} \rightarrow \mathcal{B}} \overset{\mathcal{C}}{\mathcal{A} \rightarrow \mathcal{E}}.} \quad (597)$$

This is the change in the domain space. Note the order of matrix multiplication, this time, this is right-most, as expected, because we transform the vectors before they reach the application, why if we change their domain space, we transform them after, so the multiplication is leftmost. Bringing Eq. (591) and Eq. (597) together, we have, finally:

$$\boxed{\hat{f}_{\mathcal{E} \rightarrow \mathcal{D}} = \overset{\mathcal{C}}{\mathcal{B} \rightarrow \mathcal{D}} \hat{f}_{\mathcal{A} \rightarrow \mathcal{B}} \overset{\mathcal{C}}{\mathcal{A} \rightarrow \mathcal{E}}} \quad (598)$$

which should be easy to remember. If we have a linear transform from  $\mathcal{A}$  to  $\mathcal{B}$ , and we want to express it from  $\mathcal{E}$  to  $\mathcal{D}$ , first we change from  $\mathcal{A}$  to  $\mathcal{E}$  (domain space), apply the transformation, and then change from  $\mathcal{B}$  to  $\mathcal{D}$  (codomain space). Easy.

## Exercises

### Linearity of linear things

Prove Eq. (548) (you can do this by induction). Prove that the sum of two linear functions is linear. Prove that a linear combination of linear functions is linear.

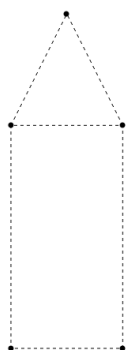
What do you think of the following statement: "The inverse of a linear function is linear?" (this we will question next lecture).

### Our house

Consider the following set of points:

$$\text{House} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 3 \end{pmatrix} \right\} \quad (599)$$

which in the 2D plane makes this pattern:



(dashed lines are to guide the eye). Add more points to break the symmetry and make the pattern even more recognizable (e.g., add a chimney or a window). Then consider the result of the following linear transforms:

- $M_\alpha \equiv \alpha \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  for  $\alpha \neq 0$ .
- $T_s \equiv \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix}$  for  $0 < s < \infty$ .
- $M_1 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ .
- $S_k \equiv \begin{pmatrix} k & 0 \\ 0 & 1/k \end{pmatrix}$ .
- $M_2 \equiv \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ .
- $L_k \equiv \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$  for  $k \neq 0$ .
- $M_3 \equiv \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ .
- $R_\theta \equiv \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ .

Compute  $O \times \text{House}$  where  $O$  is any of the above transformation applied to all points of the House set. Describe what the transformation achieves.

Compute  $OP$  for any two  $O, P$  linear transformations from the set above and consider their effect on House. Does the order matter, i.e., is  $OP$  the same as  $PO$ ?

### Householder

The “Householder transformation” describes a reflection about a line passing through the origin along the vector  $\vec{l} = l_x\hat{i} + l_y\hat{j}$ . Its matrix reads:

$$H \equiv \frac{1}{\|\vec{l}\|^2} \begin{pmatrix} l_x^2 - l_y^2 & 2l_xl_y \\ 2l_xl_y & l_y^2 - l_x^2 \end{pmatrix}. \quad (600)$$

Check that the transformation achieves what it promises (for instance as a continuation of the previous problem).

### Do it if you can

Consider the three linear applications:

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 3 & 2 & -2 \\ 4 & 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & -1 & 0 & 2 \\ 2 & 1 & 0 & 7 \\ 3 & -2 & -1 & -1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ 1 & 5 & -1 \end{pmatrix}. \quad (601)$$

How many products of two matrices can be made from these three matrices (e.g.,  $AA, AB, AC, \dots, CC$ )? Discard products that cannot be computed. Compute those that can. Be careful to check both, e.g.,  $AB$  and  $BA$ .

### A very abstract, general (and useless) case

We will change basis. Make sure to draw your results whenever possible and track what is going on with the various vectors. In each case we assume the canonical basis:

Consider the three basis of  $\mathbb{R}^2$ :

$$\mathcal{A} = \left\{ \vec{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \vec{a}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \quad (602a)$$

$$\mathcal{B} = \left\{ \vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\} \quad (602b)$$

$$\mathcal{U} = \left\{ \vec{u}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \vec{u}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right\}. \quad (602c)$$



We will also later work with the two basis of  $\mathbb{R}^3$ :

$$\mathcal{C} = \left\{ \vec{c}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{c}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{c}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \quad (603a)$$

$$\mathcal{D} = \left\{ \vec{d}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \vec{d}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \vec{d}_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\}. \quad (603b)$$

- Which of the basis are orthonormal?
- Check that the change of basis matrix from  $\mathcal{B}$  to  $\mathcal{A}$  is:

$${}_{\mathcal{B} \rightarrow \mathcal{A}} \mathbf{C} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \quad (603c)$$

- Find  ${}_{\mathcal{C} \rightarrow \mathcal{A}} \mathbf{C}$ .
- By computing the inverses, provide:

$${}_{\mathcal{A} \rightarrow \mathcal{B}} \mathbf{C} \quad \text{and} \quad {}_{\mathcal{A} \rightarrow \mathcal{C}} \mathbf{C}. \quad (603d)$$

It is easy to write the change of matrix that involve the canonical basis, because all vectors are easy to write in such a basis (that's why it's canonical). But what if we want the change-of-basis matrix between  $\mathcal{B}$  and  $\mathcal{C}$ ? Well, we can still pass by  $\mathcal{A}$ ! Say to go from  $\mathcal{B}$  to  $\mathcal{C}$ , first compute  $\vec{b}_1$  in  $\mathcal{A}$ , which is  ${}_{\mathcal{B} \rightarrow \mathcal{A}} \mathbf{C} \vec{b}_1$ , and then compute this vector in  $\mathcal{C}$ , which is  ${}_{\mathcal{A} \rightarrow \mathcal{C}} \mathbf{C} ({}_{\mathcal{B} \rightarrow \mathcal{A}} \mathbf{C} \vec{b}_1) = ({}_{\mathcal{A} \rightarrow \mathcal{C}} \mathbf{C} {}_{\mathcal{A} \rightarrow \mathcal{C}} \mathbf{C}) \vec{b}_1$  so that, finally:

$$\boxed{{}_{\mathcal{B} \rightarrow \mathcal{C}} \mathbf{C} = {}_{\mathcal{A} \rightarrow \mathcal{C}} \mathbf{C} {}_{\mathcal{A} \rightarrow \mathcal{B}} \mathbf{C}.} \quad (603e)$$

- Find:

$${}_{\mathcal{B} \rightarrow \mathcal{C}} \mathbf{C} \quad \text{and} \quad {}_{\mathcal{C} \rightarrow \mathcal{B}} \mathbf{C}. \quad (603f)$$

- Write the vector  $\vec{w} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$  in all three basis, e.g.,  $\vec{w} = \vec{a}_1 + 2\vec{a}_2$ .

Check graphically.

We now define a linear function from  $\mathbb{R}^2$  to  $\mathbb{R}^3$ . We define it as follows:

$$f(\vec{a}_1) = 3\vec{c}_1 + 2\vec{c}_2 \quad (603ga)$$

$$f(\vec{a}_2) = -\vec{c}_2 + \vec{c}_3 \quad (603gb)$$

$$f(\vec{a}_3) = \vec{c}_1 + \vec{c}_2 + \vec{c}_3 \quad (603gc)$$

- Write down the matrices for:

- |                                |                                |
|--------------------------------|--------------------------------|
| 1. $\hat{f} : A \rightarrow C$ | 4. $\hat{f} : B \rightarrow D$ |
| 2. $\hat{f} : A \rightarrow D$ | 5. $\hat{f} : U \rightarrow C$ |
| 3. $\hat{f} : B \rightarrow C$ | 6. $\hat{f} : U \rightarrow D$ |

- Compute  $f(2\vec{a}_1 + 3\vec{a}_3)$  in the  $\mathcal{C}$  and  $\mathcal{D}$  basis; do this in various ways (using  $\hat{f} : A \rightarrow C$  and  $\hat{f} : A \rightarrow D$  or changing the codomain basis for  $\hat{f} : A \rightarrow C$ ,  $\hat{f} : A \rightarrow C$ ).
- Compute  $f(2\vec{u}_1 + 3\vec{u}_3)$  in the  $\mathcal{C}$  and  $\mathcal{D}$  basis; do this in various ways.
- Compute  $2\vec{a}_1 + 3\vec{a}_3$  in the  $\mathcal{B}$  basis and compute the application of  $f$  on this vector in the  $\mathcal{C}$  and  $\mathcal{D}$  basis. Check that you get the same result as obtained previously.

*An even more abstract, but useful, case*

Consider the three basis in the space of polynomial functions of order up to 4:

$$\mathcal{C} = \{1, x, x^2, x^3, x^4\} \quad (603ha)$$

$$\mathcal{L} = \left\{1, x, \frac{1}{2}(3x^2 - 1), \frac{1}{2}(5x^3 - 3x), \frac{1}{8}(35x^4 - 30x^2 + 3)\right\} \quad (603hb)$$

$$\mathcal{B} = \left\{\frac{1}{16}(y-1)^4, \frac{1}{4}(1-y)^3(1+y), \frac{3}{8}(1-y^2)^2, \frac{1}{4}(1-y)(1+y)^3, \frac{1}{16}(1+y)^4\right\} \quad (603hc)$$

- The  $\mathcal{B}$  is called the Bernstein basis. What are the names of the  $\mathcal{C}$  and  $\mathcal{L}$  basis? (the notation gives a clue).
- Which of these basis is orthogonal? Normalized?
- Find the  $\begin{matrix} \mathcal{C} \\ \mathcal{C} \rightarrow \mathcal{L}' \end{matrix}$ ,  $\begin{matrix} \mathcal{C} \\ \mathcal{C} \rightarrow \mathcal{B}' \end{matrix}$ ,  $\begin{matrix} \mathcal{C} \\ \mathcal{L} \rightarrow \mathcal{B} \end{matrix}$  basis and vice-versa.
- Express  $x$  and  $x^2$  in terms of  $\mathcal{L}$  vectors and Bernstein polynomials.
- Decompose the sine, cosine,  $\sqrt{1+x}$  and  $e^x$  functions in the  $\mathcal{C}$ ,  $\mathcal{L}$  and  $\mathcal{B}$  basis.

## Lecture 15: Inverses & Determinants.

We have seen in the previous lecture that linear applications are conveniently represented as tables of numbers, whose columns are the result of applying the function to the basis vectors of the space, which is all we need to know to characterize the function completely. We have seen that compositions of linear functions are themselves linear and their corresponding matrix is the product of matrices of the composing functions. We now focus on one particular composition of functions, that of undoing, namely, the inverse (cf. Lecture 9). We assume  $f$  to be linear and to have an inverse,<sup>121</sup> i.e., there exists  $f^{-1}$  such that

$$f^{-1} \circ f = f \circ f^{-1} = \mathbb{1}. \quad (9)$$

We show that  $f^{-1}$  is then also linear. Indeed, for any  $X, Y$  values over which it can be defined, there exists  $x$  and  $y$  such that

$$X = f(x) \quad \text{and} \quad Y = f(y) \quad (10)$$

so that  $f^{-1}(\alpha X + \beta Y) = f^{-1}(\alpha f(x) + \beta f(y)) = f^{-1}(f(\alpha x + \beta y)) = \alpha x + \beta y = \alpha f^{-1}(X) + \beta f^{-1}(Y)$  which is the definition of linearity, QED. We have commented that it was unfortunate that the notation  $f^{-1}$  be used for the “composite-inverse”, since it is also used for  $1/f$  which, in general, is not the same thing (e.g.,  $\exp^{-1} = \ln$  which is different from  $1/\exp$ ). It does not even work<sup>122</sup> for  $f(x) = x$ . We now show in which sense it can be understood to work for linear functions<sup>123</sup> and we start with a venerable problem, solving systems of linear equations (a particular case where there are as many equations as unknown; other cases are both interesting and important but we shall overlook them for now). Such a linear system of equations with as many variables than unknown can be written as a Matrix equation:

$$Ax = y \quad (11)$$

where  $A$  is a matrix and  $\mathbf{x}, \mathbf{y}$  are vectors. If we know  $\mathbf{y}$ , what is  $\mathbf{x}$  that satisfies Eq. (11)? This is the (composite)-inverse counterpart of  $f(x) = y \implies y = f^{-1}(x)$ . In 1D space, where  $A = (a)$  itself reduces to a number, such an equation is easily solved, i.e.,  $ax = y$  yields

<sup>121</sup> Not all linear functions do have an inverse: give counter-examples.

<sup>122</sup> Why not?

<sup>123</sup>  $f(x) = x$  is linear, so why does it not work there?

$x = y/a$  (provided that  $a$  is not zero). Similarly, with matrices, one can formally solve Eq. (11) by introducing the *inverse matrix*, which exists in the sense of inverse function, as demonstrated above:

$$\mathbf{x} = A^{-1}\mathbf{y}. \quad (12)$$

Here it is tempting to think of  $A^{-1}$  (inverse function) as  $1/A$ . It is basically what is achieved, and the link is made between the two concepts, but we shall still refrain from using the later notation, as it could be confusing (since, for sure, at some point someone would write  $\frac{y}{A}$  and what meaning is there in  $y\frac{1}{A}$ ?

Importantly, while  $a$  can be inverted for all nonzero numbers, not all matrices can be inverted. First its domain and codomain must have the same dimension, so the matrix must be square. Even if it is square, a matrix can be nonzero and not have an inverse (the zero matrix is one with only zeros inside). We call such a matrix that has no inverse *singular*. The definition of the inverse  $B$  of a matrix  $A$  is one such that:

$$AB = BA = \mathbb{1}, \quad (13)$$

where  $\mathbb{1}$  is the identity matrix defined as:

$$(\mathbb{1})_{ij} = \delta_{ij} \quad (14)$$

for all  $1 \leq i, j \leq n$ , with  $\delta_{ij}$  the Kronecker symbol that is 1 if  $i = j$  and is zero otherwise. Often we will write  $\mathbb{1}$  simply as 1.

We will now study properties of matrices which have an inverse:

*Theorem:* The inverse of a matrix, if it exists, is unique.

*Proof:* Assume  $A$  and  $C$  are two inverses of  $A$ . From Eq. (13) it follows that:

$$BA = \mathbb{1}, \quad (15a)$$

$$AC = \mathbb{1}, \quad (15b)$$

then multiplying Eq. (15b) by  $C$  from the right and Eq. (15a) by  $B$  from the left, we find  $BAC = C = B$  so that  $B = C$ , meaning that the two matrices are the same (so there is only one).

Before we turn to the general case, it will be informative to turn to a particular (and easy, and important) case, that of  $2 \times 2$  matrices. We are given  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  and we want to find  $\begin{pmatrix} x & y \\ z & w \end{pmatrix}$  such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (16)$$

Using the most left-hand side product, we find:

$$ax + bz = 1 \quad (17a)$$

$$ay + bw = 0 \quad (17b)$$

$$cx + dz = 0 \quad (17c)$$

$$cy + dw = 1 \quad (17d)$$

Multiply Eq. (17b) by  $-c$  and Eq. (17d) by  $a$  and add them, to get:

$$(-bc + ad)w = a \implies w = \frac{1}{ad - bc}a \quad (18)$$

which, if

$$ad - bc \neq 0 \quad (19)$$

we can substitute back into Eqs. (17) to find:

$$\boxed{\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.} \quad (20)$$

You must remember this result. We change the sign of the anti-diagonal elements, permute the diagonal elements and divide by the difference of the diagonal products  $ad - bc$ . You can remember which diagonal is permuted and which is taken the opposite of, by remembering that  $\mathbb{1}^{-1} = \mathbb{1}$ . Equation (19) is the condition for a  $2 \times 2$  matrix to be invertible, i.e., to be “non-singular”.

The quantity  $ad - bc$  is a particular combination, which we call the “determinant”. It is an important and recurrent concept which we now study in details. This is linked to *permutations*, that are simply bijections from  $\{1, 2, \dots, n\}$  to itself. Since the set is finite, a permutation is simply a re-arrangement of the numbers. We call  $\mathfrak{S}_n$  the set of all permutations of  $n$  numbers. There are  $n!$  possible permutations of  $\llbracket 1, n \rrbracket$ , with  $\pi_1$  the identity (no actual permutation).

For instance, with 2 numbers, we have  $\pi_1(1) = 1$  and  $\pi_1(2) = 2$  for the first permutation (identity), and  $\pi_2(1) = 2$  and  $\pi_2(2) = 1$ . We can write the permutations:

$$\pi_1 : 1, 2, 3 \quad (21a)$$

$$\pi_2 : 1, 3, 2 \quad (21b)$$

$$\pi_3 : 2, 1, 3 \quad (21c)$$

$$\pi_4 : 2, 3, 1 \quad (21d)$$

$$\pi_5 : 3, 1, 2 \quad (21e)$$

$$\pi_6 : 3, 2, 1 \quad (21f)$$

To any permutation, we associate a so-called *signature*, which is  $(-1)^k$  with  $k$  the number of times one needs to permute two consecutive number to reach the final permutation from the original

ordering. For instance, one reaches  $\pi_2$  by permuting 2 and 3, so

$$\text{sign}(\pi_2) = -1. \quad (22)$$

It is also the case of  $\pi_3$  since this comes from permuting 1 and 2, while  $\pi_4$  is obtained by permuting 1 and 3 from  $\pi_3$ , and it has permutation  $(-1)^2 = 1$ . The order you decide to do the permutations does not matter: the signature of a permutation is unique (Exercises).

Now we can define the determinant. Consider a  $n \times n$  matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad (23)$$

then we defined the *determinant* of  $A$  (and write either  $\det A$  or  $|A|$ ) the following number:

$$\det A = \sum_{\pi \in \mathfrak{S}_n} \text{sign}(\pi) \prod_{i=1}^n a_{i\pi(i)}. \quad (24)$$

For instance, for  $n = 2$ :

$$\det A = a_{11}a_{22} - a_{12}a_{21}, \quad (25)$$

which is the combination we find in the denominator of Eq. (20).

For a  $3 \times 3$  matrix, from Eq. (21) and the signatures we find in the Exercises, we can compute:

$$\begin{aligned} \det A = & a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + \\ & + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned} \quad (26)$$

There is a powerful recursive way to compute determinants, which involve the *minors* and *cofactors* of the matrix. The cofactor  $C$  of a matrix  $A$  is a smaller matrix obtained by computing the determinants of the submatrices  $D$  obtained from  $A$  by deleting one of its row and one of its column, keeping track of the parity of  $i$  and  $j$ . Namely, the  $i$ th row and the  $j$ th column of  $C$  is obtained as

$$C_{ij} \equiv (-1)^{i+j} \det D_{\ast ij \ast} \quad (27)$$

where we defined  $D_{\ast ij \ast}$  as the matrix whose  $i$ th row and  $j$ th column has been removed. As this might sound a bit obscure, we show that

more “graphically”. If  $A$  is the matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(j-1)} & a_{1j} & a_{1(j+1)} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2(j-1)} & a_{2j} & a_{2(j+1)} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)j} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ a_{i1} & a_{i2} & \cdots & a_{i(j-1)} & a_{ij} & a_{i(j+1)} & \cdots & a_{in} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)j} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(j-1)} & a_{nj} & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix} \quad (28)$$

then  $D$  expanded along the  $i$ th row,  $j$ th column becomes:

$$D_{\times ij \times} \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2(j-1)} & a_{2(j+1)} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(j-1)} & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix} \cdot \quad (29)$$

The determinant of this matrix, multiplied by the sign found in this corresponding matrix

$$\begin{pmatrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (30)$$

gives the  $(i, j)$  cofactor  $C_{ij}$ . Namely:<sup>124</sup>

<sup>124</sup> Check this and work out which of the  $\pm$  applies.

$$C_{ij} \equiv \begin{vmatrix} a_{11} & -a_{12} & \cdots & (-1)^j a_{1(j-1)} & (-1)^j a_{1(j+1)} & \cdots & (-1)^{n+1} a_{1n} \\ -a_{21} & +a_{22} & \cdots & (-1)^{j+1} a_{2(j-1)} & (-1)^{j+1} a_{2(j+1)} & \cdots & (-1)^n a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ (-1)^{i-1} a_{(i-1)1} & (-1)^i a_{(i-1)2} & \cdots & \pm a_{(i-1)(j-1)} & \pm a_{(i-1)(j+1)} & \cdots & \mp a_{(i-1)n} \\ (-1)^{i+1} a_{(i+1)1} & (-1)^i a_{(i+1)2} & \cdots & \mp a_{(i+1)(j-1)} & \mp a_{(i+1)(j+1)} & \cdots & \pm a_{(i+1)n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ (-1)^n a_{n1} & (-1)^{n+1} a_{n2} & \cdots & \pm a_{n(j-1)} & \pm a_{n(j+1)} & \cdots & \mp a_{nn} \end{vmatrix}, \quad (31)$$

where we introduced the notation  $|\cdots|$  for the determinant replacing parentheses  $(\cdots)$  for the matrix. Note that the top-left corner always

reads  $\begin{pmatrix} + & - \\ - & + \end{pmatrix}$  regardless of the size of the matrix. From these, you can then obtain the determinant through its expansion theorem, which states that (Exercises):

$$\det A = \sum_{i=1}^n a_{ij}C_{ij} \text{ (for all } j) = \sum_{j=1}^n a_{ij}C_{ij} \text{ (for all } i). \quad (32)$$

The free choice of  $i$  and  $j$  means that one can “expand” the determinant along any row or column, which, in practice, will be chosen to be one with many zeros  $a_{ij}$  as then there is no need to compute the cofactor.

For example, if we want to compute:

$$|A| = \begin{vmatrix} 4 & -1 & 2 & 1 \\ 3 & 0 & 1 & -2 \\ 2 & 1 & 5 & 1 \\ -2 & 1 & 3 & -1 \end{vmatrix} \quad (33)$$

we can expand along the 2nd row (or 2nd column), because there is a zero there which will simplify the calculation (one time). This gives:

$$|A| = -3 \begin{vmatrix} -1 & 2 & 1 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} - \begin{vmatrix} 4 & -1 & 1 \\ 2 & 1 & 1 \\ -2 & 1 & -1 \end{vmatrix} + (-2) \begin{vmatrix} 4 & -1 & 2 \\ 2 & 1 & 5 \\ -2 & 1 & 3 \end{vmatrix}. \quad (34)$$

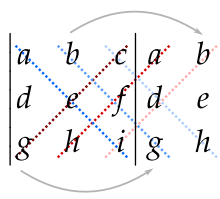
The  $3 \times 3$  determinant we encounter often in Physics (we live in 3D space and we often have to take vector products of 3D vectors). This can be expanded the same way, e.g.,<sup>125</sup>

<sup>125</sup> Check this agrees with Eq. (26).

$$\begin{vmatrix} -1 & 2 & 1 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} = - \begin{vmatrix} 5 & 1 \\ 3 & -1 \end{vmatrix} - 2 \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} + \begin{vmatrix} 1 & 5 \\ 1 & 3 \end{vmatrix} \quad (35a)$$

$$= -(-5 - 3) - 2(-1 - 1) + (3 - 5) = 8 + 4 - 2 = 10. \quad (35b)$$

Or you can use the following pattern: we copy the first two columns and obtain all the possible products without repeating elements on the same column and/or the same row, which is easily done by following diagonals and antidiagonals. We then *add* when we go along the antidiagonals, and *subtract* otherwise:



$$\begin{vmatrix} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{vmatrix} = aei + bfg + cdh - gec - hfa - idb \quad (36)$$



which, for Eq. (35a), also yields:

$$\begin{vmatrix} -1 & 2 & 1 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} = (-1) \times 5 \times (-1) + 2 \times 1 \times 1 + 1 \times 1 \times 3 \\ - 1 \times 5 \times 1 - 3 \times 1 \times (-1) - (-1) \times 1 \times 2 = 10. \quad (37)$$

In Physics,  $3 \times 3$  determinants are useful to compute the vector product through the following trick:

$$\vec{A} \times \vec{B} \equiv \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}, \quad (38)$$

which is a property you will encounter a lot (this Semester in Mechanics, next Semester in Electromagnetism, etc.)

There is also a geometrical interpretation to determinants. In the Euclidean space  $\mathbb{R}^2$ , the determinant gives the stretching factor of the unit square with sides  $\hat{i}$  and  $\hat{j}$  as transformed by a linear application whose matrix has the given determinant.<sup>126</sup> When the determinant is zero, this area vanishes as the result of the transformation collapsing the space into one of lower dimension (a line or a point). This happens when the vectors through the applications do not form a basis for the space, that is, they become linearly dependent. Vectors that are equal are the most linearly dependent type you can get! The same happens in 3D where the determinant gives the volume<sup>127</sup> and also in higher dimensions but then we need to turn to hypervolumes. If you worry that the determinant can be negative, this is interpreted as “inversion” of the area (volume/hypervolume), meaning that you are looking at it from the other side.

<sup>126</sup> Show this

<sup>127</sup> Show this

Determinants have a lot of interesting properties. We let you demonstrate the most important ones by yourself in the Exercises. There is one which we will need to work with matrix inversion so we demonstrate it now:

*Theorem:* A matrix which has two identical rows (or columns) has determinant zero.

*Proof:* We show it by induction, that is, we show that if it is true for a matrix of size  $n$ , then it is also true for a matrix of size  $n + 1$ . Imagine thus a  $n + 1$  matrix  $A$  with two identical columns, say,  $r$  and  $k$ . Now let us compute the determinant according to Eq. (32) keeping both the identical columns:

$$|A| = \sum_i a_{il} C_{il} \quad (39)$$

where  $l \neq r, k$ . Now  $C_{il}$  for all  $i, l$ , is the determinant of a matrix (of size  $n \times n$ ) that still contains two identical columns (truncated for

each  $i$  by one of their elements). By assumption, this means that  $C_{ii}$  is zero. Therefore also  $|A|$ , at the order  $n + 1$ , is zero. We complete the proof by showing that the assumption is indeed correct for  $n = 2$ , which is clear since  $\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0$ . Hence, since it is true for  $n = 2$ , we have shown it becomes true by induction at  $n = 3$ , and iterating, for all  $n$ . QED. Note that this can also be shown from the property demonstrated in the Exercises that permuting two columns changes the sign of the determinant. If two columns are identical, the permutation will result in no change in the matrix but a change in the sign of its determinant, so that  $\det A = -\det A \implies \det A = 0$ .

We now come back to our initial problem of computing the inverse of a  $n \times n$  matrix.

Actually, we have done most of the work now. We first find the matrix that gives the good diagonal elements (all ones). From Eq. (32),  $\det A = \sum_{k=1}^n a_{ik}C_{ik} = \sum_{k=1}^n a_{ik}\tilde{A}_{ki}$  where we have introduced

$$\tilde{A} \equiv C^T \quad (40)$$

with  $C^T$  the *transpose* (i.e., permutating rows and columns, that is, formally,  $(C^T)_{ij} \equiv C_{ji}$ ) of the cofactor matrix  $C$ . We call the matrix  $\tilde{A}$  the “*adjunct*” matrix of  $A$  (transpose of cofactors). Therefore, we have:

$$\frac{(A\tilde{A})_{ii}}{\det A} = \sum_{k=1}^n A_{ik} \frac{\tilde{A}_{ki}}{\det A} = 1. \quad (41)$$

The diagonal elements of  $A\tilde{A}/|A|$  are one.

For the off-diagonal element  $i, j$ , with  $i \neq j$  (off-diagonal), we introduce a new matrix  $B$  which is the same as  $A$  except that on its  $j$ th row we copy the  $i$ th row of  $A$  (note: copy, not permute). Importantly, the cofactors for  $A$  and  $B$  *not* on the  $j$ th row, are the same, since they are obtained precisely by deleting the  $j$ th row (which is where the two differ). So if apply Eq. (32) to  $B$  on its  $j$ th row, we find:

$$\det B = \sum_{k=1}^n b_{jk}C_{jk} \quad (42a)$$

$$= \sum_{k=1}^n a_{ik}C_{jk} \quad (42b)$$

$$= \sum_{k=1}^n a_{ik}\tilde{A}_{kj} \quad (42c)$$

$$= (A\tilde{A})_{ij} \quad (42d)$$

$$= 0 \quad (42e)$$

where Eq. (42a) is Eq. (32) applied to  $B$  with the knowledge that the cofactors of  $B$  are those of  $A$  on the  $j$ th row, Eq. (42b) is by definition of  $B$  ( $j$ th row is  $A$ 's  $i$ th column), Eq. (42c) is the definition of the

adjunct (transpose of cofactors), Eq. (42d) is the definition of a matrix product and Eq. (42e) is because  $B$  has two identical rows (and we are computing its determinant, left-hand side). Equations (42d–42e) together state that the off-diagonal elements of  $A\tilde{A}/\det A$  are zero (we can divide by the nonzero determinant). So, finally: we have the general formula for the inverse of a matrix:

$$\boxed{A^{-1} = \frac{\tilde{A}}{\det A}} \quad (43)$$

where  $\tilde{A}$  is the “adjunct” (or “adjugate”) of the matrix, i.e., the transpose of its cofactors.

This gives an explicit way to find the matrix. This is not, however, the most efficient from an algorithmic point of view. Methods to inverse higher  $n \times n$  matrices are Gauss-Jordan elimination, Gaussian elimination, or LU decomposition, which we will not cover here but that you can read about in the literature. They are used in computer science to make such calculations efficient. For us, and for now, it will be enough to know Eq. (43), of which the important case (20) is a particular case.

### Exercises

#### Transpositions

Show that:

$$(A + B)^T = A^T + B^T. \quad (44)$$

Similarly, it is easy to see that  $(\alpha A)^T = \alpha A^T$ . It is more interesting to prove the (possibly less intuitive, but important) transpose of a product:

$$(AB)^T = B^T A^T. \quad (45)$$

#### Inverses

Prove the important way to compute the inverse of a matrix product:

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}}. \quad (46)$$

This is most-simply proved by showing that the proposed solution actually works!

#### Signatures

Show that the permutations of Eqs. (21) are, respectively, 1,  $-1$ ,  $-1$ , 1, 1,  $-1$ . Find all the permutations of  $\{1, 2, 3, 4\}$  and give their signa-

tures. What are the signatures of:

$$\{7, 2, 5, 1, 3, 4, 6, 9, 8\}, \quad (47a)$$

$$\{5, 1, 6, 2, 3, 8, 9, 4, 7\}. \quad (47b)$$

(clue to check your answers: they are the same).

### *The determinant of a $4 \times 4$ matrix*

Using the results from the previous exercise, write down the generic expression for the determinant of a  $4 \times 4$  matrix (the first term, from  $\pi_1 \in \mathfrak{S}_4$ , is  $a_{11}a_{22}a_{33}a_{44}$ ).

### *Practicing with $2 \times 2$ matrices*

Show that Eq. (20) is also found from solving (we used the other order in the Lecture):

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (48)$$

Compute the following inverses:

$$M_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & 2 \\ -3 & 5 \end{pmatrix}. \quad (49)$$

### *Practicing with $3 \times 3$ matrices*

Compute the following inverses:

$$M_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 3 & -2 & 1 \\ 0 & 1 & 7 \\ 5 & 4 & -6 \end{pmatrix}. \quad (50)$$

### *Solving a matrix equation*

Find  $x$ ,  $y$ ,  $z$  and  $w$  that satisfy this equation:

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & 1 & 0 \\ -\frac{1}{2} & \frac{3}{2} & -\frac{5}{2} & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}. \quad (51)$$

### Properties of determinants

Prove the following statements:

- A matrix and its transpose have the same determinant:

$$\det A = \det A^T. \quad (52)$$

- If  $B$  is obtained from  $A$  by interchanging two different columns, then:

$$\det A = -\det B. \quad (53)$$

- If  $B$  is obtained from  $A$  by multiplying one of the columns of  $A$  by a non-zero constant  $k$ , then

$$\det B = k \det A \quad (54)$$

The determinant is *not linear*, but it has special properties of its own. Prove this interesting property under expansion of addition:

$$\begin{vmatrix} a_{11} & \cdots & a_{1s} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{r1} + a'_{r1} & \cdots & a_{rs} + a'_{rs} & \cdots & a_{rn} + a'_{rn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1s} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rs} & \cdots & a_{rn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{1s} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a'_{r1} & \cdots & a'_{rs} & \cdots & a'_{rn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} & \cdots & a_{nn} \end{vmatrix} \quad (55)$$

Also prove the following

$$\begin{vmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{vmatrix} \quad (56)$$

and

$$\begin{vmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rk} & \cdots & a_{rn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} & \cdots & a_{nn} \end{vmatrix} = \sum_{k=1}^n a_{rk} \begin{vmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} & \cdots & a_{nn} \end{vmatrix}. \quad (57)$$

Use the previous properties to prove the determinant-expansion theorem (also known after Laplace).



## Lecture 16: Eigensystems.

We now turn to the capital feature of linear applications, namely, their intrinsic linear character and fundamental one-dimensional nature. They are already pretty simple, from their property of  $f(\alpha\vec{u} + \beta\vec{v}) = \alpha f(\vec{u}) + \beta f(\vec{v})$ , but they still involve a lot of shuffling around, multiplying matrices means a lot of spanning of horizontal and verticals, computation of an inverse matrix through its adjunct or of a determinant are a nightmare of simplicity due to the huge amount of operations this involves. Actually, computations of *permanents*, which are related to determinants except that they do not include the signature of the permutation, i.e.,  $\text{per } A = \sum_{\pi \in \mathfrak{S}_n} \prod_{i=1}^n a_{i\pi(i)}$  are regarded as among the most difficult objects to compute as the size  $n$  of the matrix  $A$  increases and there is a currently a race to build quantum machines to sample from such permanents to prove “quantum supremacy” (boson sampling problem). Determinants, on the other hand, are actually no more complicated than quantities like the “*trace*”, which is the sum of the diagonal elements, because they are preserved through changes of basis and we can find a basis where the linear application is very simple, and by very simple we mean it has a lot of zeros (which will cancel most of the terms of the tedious operations). The permanent on the other hand, is not conserved through a change of basis, so we need to compute all the  $n!$  terms due to all the possible permutations. And it becomes quickly impossible.

The simplest thing that a linear application can do is rescale a vector, without changing its direction:

$$A\vec{e}_i = \lambda_i\vec{e}_i \quad (58)$$

that is, it doesn’t meddle with other vectors of the space. If you place yourself along that particular direction, the application is one-dimensional to you, and what it does is simply  $f(x) = \lambda x$ .

When Eq. (58) is satisfied, we call  $\vec{e}_i$  an “*eigenvector*” and  $\lambda_i$  the corresponding “*eigenvalue*”. This comes from the German for “proper” (There is a Nobel prize (in Chemistry) called Manfred Eigen; this is not named after him!). So a “proper vector” is the one that remains

the same after passing through the linear application, just picking up its “proper value” in the process. Note that there is a vector that always achieve that, the so-called “*trivial solution*”, because it works all the time but not achieving much. Namely,  $\vec{e}_i = \vec{0}$ . So we demand that an eigenvector never be zero. We will see, however, that the eigenvalue itself can be zero, but never the eigenvector, which can still be such that  $A\vec{e}_i = \vec{0}$  without itself being zero.

We can try to find the eigenvectors and eigenvalues of a linear application  $A$  by solving Eq. (58), which we rewrite as:

$$(A - \lambda_i \mathbf{1}_n) \vec{e}_i = \vec{0}. \quad (59)$$

This is a matrix equation, i.e., a “system of linear equations”:

$$M\vec{x} = \vec{y}. \quad (60)$$

If  $\vec{y} = \vec{0}$ , then we say that the solution is *homogeneous*, which is the case of Eq. (59). If the determinant of  $M$  is *nonzero*, then there is a unique solution, namely,  $M^{-1}\vec{y}$  in general and  $M^{-1}\vec{0} = \vec{0}$  for the homogeneous case. If the determinant is *zero*, then we have an infinite number of solutions, both for the homogeneous and non-homogeneous case. The non-homogeneous equations are actually obtained by finding one specific solution  $\vec{x}_y$  for the non-homogeneous system. Then by linearity, all other solutions are found by adding to this any solution  $\vec{x}_0$  of the homogenous system.

Coming back to Eq. (59), if the matrix  $(A - \lambda_i \mathbf{1}_n)$  is not singular, that is, it can be inverted, then we hit the trivial solution  $\vec{e}_i = (A - \lambda_i \mathbf{1}_n)^{-1} \vec{0} = \vec{0}$  (which we don’t want). Therefore, the matrix must be singular, and since it must be so for all eigenvalues  $\lambda$ , we can remove the index and demand

$$\det(A - \lambda \mathbf{1}_n) = 0. \quad (61)$$

This equation is called the “characteristic polynomial”, as it is a polynomial equation in  $\lambda$  of order  $n$ , since it will involve a  $n$ th power (the diagonal term of the determinant).

Let us find eigenvalues of a  $2 \times 2$  matrix:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0 \quad (62a)$$

$$(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0 \quad (62b)$$

$$\lambda^2 - \lambda(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}a_{21} = 0 \quad (62c)$$

so indeed for a  $2 \times 2$  matrix we have a quadratic equation, which we know how to solve. Note that it turns out to be

$$\lambda^2 - \lambda \text{Tr}A + \det A = 0. \quad (63)$$



The solutions to the quadratic equation gives us the two eigenvalues. The discriminant is:

$$\Delta = (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21}) \quad (64)$$

and therefore:

$$\lambda_{1,2} = \frac{a_{11} + a_{22}}{2} \pm \sqrt{\left(\frac{a_{11} - a_{22}}{2}\right)^2 + a_{12}a_{21}}. \quad (65)$$

Keep in mind this result, it will come back often in Physics. The eigenvalues are the average of the diagonal elements plus and minus a term which involves the average difference and the so-called “coupling terms”  $a_{12}$  and  $a_{21}$  that couple together the  $x$  and  $y$  components of the 2D vector. We let you compute as an exercise the corresponding Eigenvectors.

The same applies to higher-dimension matrices. For a  $n \times n$  matrices, we have  $n$  eigenvalues (possibly “degenerate”, i.e., identical, also possibly zero). Note that these eigenvalues can be complex, for instance in 2D if the discriminant is negative. There might not be as many eigenvectors as eigenvalues, since, remember, we do not accept  $\vec{0}$  as an eigenvector. Here you have a matrix:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (66)$$

which has two eigenvalues, both degenerate and zero,  $\lambda_1 = \lambda_2 = 0$ , but only one eigenvector:  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ . An important case is when there are  $n$  eigenvectors which are linearly independent, that is, none of them can be obtained as a linear superposition of the others. It can be proven (Exercises) that this is the case when the eigenvalues are different. But be careful that, on the other hand, some eigenvalues could be degenerate and their eigenvectors still be linearly independent. It means nondegenerate  $\lambda$  is a sufficient but not necessary condition. Now,  $n$  linearly independent vectors form a basis of the vector space. And we arrive at this important fact: **A matrix is diagonal in its basis of eigenvectors.**

Consider indeed a linear application  $f$  with associated matrix  $\hat{f}_{\mathcal{B} \rightarrow \mathcal{B}}$  in the basis  $\mathcal{B}$ , and let us change to the basis of eigenvectors of  $A$ , which we call  $\mathcal{E}$ :

$$\hat{f}_{\mathcal{E} \rightarrow \mathcal{E}} = \underset{\mathcal{B} \rightarrow \mathcal{E}}{C} \hat{f}_{\mathcal{B} \rightarrow \mathcal{B}} \underset{\mathcal{E} \rightarrow \mathcal{B}}{C} \quad (67)$$

Now by definition of the matrix associated to a linear application:

$$\hat{f}_{\mathcal{E} \rightarrow \mathcal{E}} = \left( \begin{pmatrix} f(\vec{e}_1) \\ \vdots \\ f(\vec{e}_n) \end{pmatrix}_{\mathcal{E}} \right), \quad (68)$$

and since  $f(\vec{e}_i) = \lambda_i \vec{e}_i$ , also by definition of the eigenvectors and eigenvalues, then we have:

$$\hat{f}_{\mathcal{E} \rightarrow \mathcal{E}} = \left( \begin{pmatrix} \lambda_1 \vec{e}_1 \\ \vdots \\ \lambda_n \vec{e}_n \end{pmatrix}_{\mathcal{E}} \right). \quad (69)$$

But, clearly,  $\left( \vec{e}_i \right)_{\mathcal{E}}$ , is a complicated notation for something very simple, this is the column (component) representation of the  $i$ th basis vector, expressed in its own basis, and since, clearly:

$$\vec{e}_i = 0\vec{e}_1 + 0\vec{e}_2 + \cdots + 1\vec{e}_i + \cdots + 0\vec{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad (70)$$

where 1 is at the  $i$ th row, i.e., we have:

$$\hat{f}_{\mathcal{E} \rightarrow \mathcal{E}} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \quad (71)$$

and zero everywhere else. This is a so-called “*diagonal matrix*”. Diagonalization is usually obtained in this way:

$$\boxed{\hat{f}_{\mathcal{E} \rightarrow \mathcal{E}} = C^{-1} \hat{f}_{\mathcal{E} \rightarrow \mathcal{B}} C_{\mathcal{B} \rightarrow \mathcal{E}}}, \quad (72)$$

i.e., if we have a matrix  $A$ , we find its eigenvectors, make the matrix  $B = C_{\mathcal{E} \rightarrow \mathcal{B}}$  whose columns are the eigenvectors (that’s the change-of-basis matrix), then  $B^{-1}AB = D$  with  $D$  diagonal.

From known properties of some operations, like the cyclicity of the trace  $\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB)$  (proven in Exercises), or

the product of determinants  $\det AB = \det A \det B$ , we can now easily prove the properties we trumpeted before:

$$\text{Tr}(A) = \sum_i a_{ii} = \text{Tr}(ABB^{-1}) = \text{Tr}(B^{-1}AB) = \text{Tr}(D) = \sum_i \lambda_i. \quad (73)$$

The determinant is related:

$$\begin{aligned} \det A &= \det ABB^{-1} = \det AB \det B^{-1} = \det B^{-1} \det AB \\ &= \det B^{-1}AB = \det D = \prod_i \lambda_i. \end{aligned} \quad (74)$$

We have proven it for diagonalizable matrices, though. It happens to be true for *all* matrices, based on related ideas (of triangular rather than diagonal matrices), which is also left as further material for you to explore (Exercises), as this is not so crucial for our Physicists needs.

As another example, consider the matrix equation:

$$A\vec{x} = \vec{y} \quad (75)$$

since  $B$  is invertible, you can go through the following steps:

$$ABB^{-1}\vec{x} = \vec{y} \quad (76a)$$

$$B^{-1}ABB^{-1}\vec{x} = B^{-1}\vec{y} \quad (76b)$$

$$D\vec{v} = \vec{w} \quad (76c)$$

where

$$\vec{v} \equiv B^{-1}\vec{x} = \underset{B \rightarrow \mathcal{E}}{C} \vec{x} \quad (77a)$$

$$\vec{w} \equiv B^{-1}\vec{y} = \underset{B \rightarrow \mathcal{E}}{C} \vec{y} \quad (77b)$$

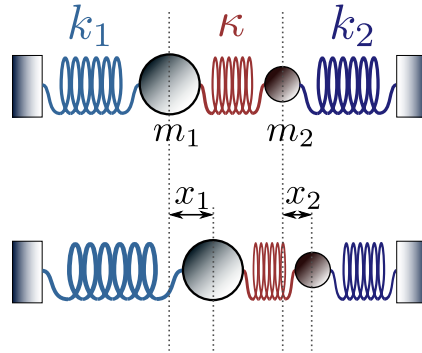
are the vectors expressed in the eigenvectors basis. So our initial matrix equation (75) in its basis becomes a diagonal equation in the eigenvector basis:

$$D\vec{v} = \vec{w}. \quad (78)$$

This is easily solved. We have effectively *decoupled* the various components of the vectors. We can easily manipulate this, the inverse of  $D$ , in particular,  $D^{-1}$  is also diagonal with diagonal elements the inverse eigenvalues  $\lambda_i^{-1}$  (so yes, a matrix is nonsingular iff all its eigenvalues are nonzero). Once we are happy with the solutions, if for some reason we want to come back to our initial basis (typically the canonical basis, which we like for other reasons), then we just apply the inverse change-of-basis matrix.

We will conclude by working out one of the fundamental problems in Physics: two coupled oscillators. This ultimately subtends

sound propagation, electric circuits, light-matter interactions and a variety of interesting problems, including this mechanical system, of two masses with respective couplings to the wall  $k_1$  and  $k_2$ , and coupled to each other with a coupling strength  $k_{12}$ .



Without this coupling term, we have two independent oscillators, which we know how to solve from Mechanics. Calling  $x_i$  the displacement of the  $i$ th oscillator ( $i = 1, 2$ ) as compared to its equilibrium or rest position, we have (Newton's equations  $F = ma$  with  $F = -kx$  for a spring and  $a = \ddot{x}$  the second-derivative of position, giving the acceleration):

$$m_1 \ddot{x}_1(t) = -k_1 x_1(t), \quad (79)$$

$$m_2 \ddot{x}_2(t) = -k_2 x_2(t), \quad (80)$$

which is of the type:

$$\ddot{x}_i = -(k_i/m_i)x_i \quad (81)$$

with solutions (check it):

$$x_i(t) = A_i \cos(\sqrt{k_i/m_i}t) + B_i \sin(\sqrt{k_i/m_i}t) \quad (82)$$

with  $A_i, B_i$  given by the initial conditions. Harmonic motion. Simple.

Bringing in the coupling looks a very complicated business. Now the equations are coupled and the force exerted by one oscillator on the other depends on their relative positions:

$$m_1 \ddot{x}_1(t) = -k_1 x_1(t) + k_{12}(x_2 - x_1), \quad (83)$$

$$m_2 \ddot{x}_2(t) = -k_2 x_2(t) - k_{12}(x_2 - x_1), \quad (84)$$

The two oscillators have different frequencies, one will push on the other as its trying to recede away and if they go in the same direction their action will tend to cancel, so the actual dynamics looks very complicated! And it is to some extent, but because we are not looking at the good objects, namely, the coupled oscillators are mixing behaviours of two fundamentally simpler "objects". These can be seen

by turning the problem into a matrix equation

$$\begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = - \begin{pmatrix} k_1 + k_{12} & -k_{12} \\ -k_{12} & k_2 + k_{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (85)$$

or, even simpler<sup>128</sup>

$$\ddot{\vec{x}} = -M^{-1}K\vec{x}, \quad (86)$$

with  $M$  and  $K$  the mass and coupling matrices.<sup>129</sup> Diagonalization of this equation for<sup>130</sup>

$$A = M^{-1}K \quad (87)$$

as we did in Eqs. (76–77) leads us to:

$$\ddot{\vec{w}} = -D\vec{w} \quad (88)$$

where

$$D \equiv B^{-1}AB \quad (89)$$

is diagonalized by the matrix  $D$  of eigenvectors of  $A$ . Equation (88) is now a differential equation of the type (81) meaning that it has the same solutions as Eq. (82): harmonic oscillations! Just with different frequencies: the eigenfrequencies. In the case where both free oscillators have the same mass,  $m_1 = m_2$  as well as the same frequency,  $k_1 = k_2 \equiv k$ , then the change-of-basis matrix, made up of the eigenvectors of  $A$ , is very simple, and remarkably, does not depend on the physical constants of the problem:<sup>131</sup>

$$B = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (90)$$

with eigenvalues, which we would call, *eigenfrequencies*:<sup>132</sup>

$$\lambda_1 = \frac{k + 2k_{12}}{m} \quad (91)$$

$$\lambda_2 = \frac{k}{m} \quad (92)$$

and thus

$$w_i(t) = A_i \cos(\sqrt{\lambda_i}t) + B_i \sin(\sqrt{\lambda_i}t) \quad (93)$$

which we can write as

$$\vec{w} = (A_1 \cos(\sqrt{\lambda_1}t) + B_1 \sin(\sqrt{\lambda_1}t)) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \quad (94a)$$

$$+ (A_2 \cos(\sqrt{\lambda_2}t) + B_2 \sin(\sqrt{\lambda_2}t)) \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (94b)$$

The two solutions are the real things. We can choose one or the other by setting  $A_i = B_i = 0$  for  $i = 1, 2$ .

<sup>128</sup> Are you surprised or bothered by taking the derivative of a vector? Show that according to the definitions we have given to all these objects, this is clear and works as expected.

<sup>129</sup> Write them down, including  $M^{-1}$ .

<sup>130</sup> Compute  $A$ .

<sup>131</sup> Check it.

<sup>132</sup> Check it.

Back to the “oscillators basis”:

$$\vec{x} = B\vec{w} \quad (95)$$

so that<sup>133</sup>

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (A_1 \cos(\sqrt{\lambda_1}t) + B_1 \sin(\sqrt{\lambda_1}t)) \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \quad (96a)$$

$$+ (A_2 \cos(\sqrt{\lambda_2}t) + B_2 \sin(\sqrt{\lambda_2}t)) \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (96b)$$

This is, actually, not that much more complicated than an isolated oscillator. Here we find the phenomenon of “beating”, familiar in optics.

### Exercises

#### All the eigenvectors

Prove that if  $\vec{e}$  is an eigenvector of  $A$ , then so are all parallel vectors (aligned or anti-aligned)  $\alpha\vec{e}$  for  $\alpha \neq 0$ , with, naturally, the same eigenvalue. Due to this degree of freedom in choosing one, often we will take the normalized eigenvectors.

#### Eigenvectors of a general $2 \times 2$ matrix

We computed the eigenvalues in the Lecture. Now we compute the eigenvectors. We use the same equation (59), but now we know  $\lambda_i$  and solve for  $\vec{e}_i$ . Since the matrix  $A - \lambda_i \mathbf{1}_n$  is singular by definition (or construction), we have an infinite number of solutions (this is consistent with the result from the previous exercise). Show that

$$\alpha x + \beta y = 0 \quad (97)$$

can be solved by choosing  $x = -\beta$  and  $y = \alpha$ . Other choices are possible. As a conclusion, show that

$$\begin{pmatrix} -a_{12} \\ a_{11} - \lambda_i \end{pmatrix} \quad (98)$$

are eigenvectors of  $A$ . Substitute Eqs. (65) and obtain the result only in terms of the coefficients of  $A$ .

#### Eigenvectors of a particular $2 \times 2$ matrix

Find the eigenvalues and eigenvectors of

$$M = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}. \quad (99)$$

<sup>133</sup> Find the solutions for the following initial conditions:

1.  $x_1(0) = 1$  and  $x_2(0) = 0$ .
2.  $x_1(0) = 1$  and  $x_2(0) = -1$ .
3.  $x_1(0) = x_2(0) = 1$ .
4.  $x_1(0) = x_2(0) = 0$ .

Comment in each case.

### Eigenvectors of a very particular $2 \times 2$ matrix

Consider the matrix of rotation:

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (100)$$

Check (again, if you forgot) that  $R_\theta$  applied to a  $(x, y)^T$  vector produces a rotated vector. Now, we said that eigenvectors are those that remain in the same direction as the original vector. How is this possible, to remain aligned, and to have been rotated at the same time? If you can't break this puzzle by thinking, just compute and see what happens, namely, find the eigenvalues (first) and the eigenvectors. Check that they obey the promised property of remaining aligned. Are you happy you resolved the apparent paradox?

### Harmless proofs

We show you one, and let you explore/prove (and check) the other.

*Theorem:* If  $A$  has eigenvalues  $\lambda_i$ , then  $A^2$  has eigenvalues  $\lambda_i^2$  with the same associated eigenvectors.

*Proof:* By definition,  $A\vec{e}_i = \lambda_i\vec{e}_i$ . Applying  $A$  to both sides, we get  $A(A\vec{e}_i) = A\lambda_i\vec{e}_i$ , i.e.,  $A^2\vec{e}_i = \lambda_i A\vec{e}_i = \lambda_i^2\vec{e}_i$ , but this is precisely saying that  $\vec{e}_i$  is an eigenvector of  $A^2$  with eigenvalue  $\lambda_i^2$ . Check that this is the case through actual computations. Extend your reasoning to other powers and inverse matrices.

### The case of the transpose

Show on an example that  $M$  and  $M^T$  have the same eigenvalues but may have different eigenvectors.

### Cyclicality of the trace

Show that

$$\text{Tr}(ABC) = \sum_i \sum_j \sum_k A_{ij} B_{jk} C_{ki}. \quad (101)$$

and use this to show the cyclic properties of the trace, i.e.,  $\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB)$ . Show as a result that the trace, like the determinant, is invariant under a change of basis. Show that this is not the case for instance for the sum of the antidiagonal terms or the sum of a row or column.

*Diagonalization*

Find the eigenvalues and eigenvectors of:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}. \quad (102)$$

Compute the trace and determinant of  $A$ . Is it singular?

Write  $C$  the change-of-basis matrix from the eigenvectors to the canonical basis and compute its inverse  $C^{-1}$  (this matrix is invertible). Compute  $C^{-1}AC$  and check that you find the eigenvalues on the diagonal. Check the trace and determinant of this diagonal matrix. Compute also  $CAC^{-1}$  to make sure the order is important.

*Basis of eigenvectors*

Prove that if all eigenvalues are different, then the corresponding eigenvectors are linearly independent.

 *$N$ -coupled oscillators*

Write the differential equations for  $N$  coupled oscillators, after the model  $N = 2$  in the text. Write the corresponding matrix. Solve first the case  $N = 3$ . Can you work out the formal solutions for the general case? This is the foundation for the theory of solids, among other things, and in particular sound propagation in them.



## Lecture 17: Fourier Series.

We now look in more details at a particular basis of functions, and look at the decomposition of arbitrary functions in this basis. As such, this is no different than what we did with Legendre polynomials. But while Legendre polynomials are a seldom-used basis, the one we will now introduce is extremely important and widespread and comes with its own terminology, of Fourier analysis. The basis functions are, beside, very familiar functions, namely, the sine and cosine. But that is two functions only while we need a countably infinite number! What differs is not the shape but the frequency of these functions. Namely, we will take, on the space  $[-\pi, \pi]$ , the functions

$$\langle x|s_n\rangle = \frac{1}{\sqrt{\pi}} \sin nx \quad \text{and} \quad \langle x|c_n\rangle = \frac{1}{\sqrt{\pi}} \cos nx \quad (103)$$

which we have normalized,<sup>134</sup> so that the functions are orthonormal, i.e.,

$$\langle s_n|s_m\rangle = \delta_{nm}, \quad \langle c_n|c_m\rangle = \delta_{nm} \quad \text{and} \quad \langle c_n|s_m\rangle = 0 \quad (104)$$

for all  $n, m \in \mathbb{N}$  not both zero; for  $n = m = 0$ ,

$$|s_0\rangle = 0 \quad \text{and} \quad |c_0\rangle = 1/\sqrt{2\pi} \quad (105)$$

so that  $|s_0\rangle$  is not part of the basis (since a basis vector cannot be zero) while  $|c_0\rangle$  is but with a different normalizing factor.

While we will not prove it here, it can be shown that the Fourier basis is complete, meaning that any function of the space can be written as a linear superposition of its basis vectors. The closure relation reads:

$$\mathbb{1} = \sum_{n=0}^{\infty} (|s_n\rangle \langle s_n| + |c_n\rangle \langle c_n|) \quad (106)$$

so that we can expand any function  $|f\rangle$  as:

$$|f\rangle = \sum_{n=0}^{\infty} (|s_n\rangle \langle s_n|f\rangle + |c_n\rangle \langle c_n|f\rangle) \quad (107)$$

<sup>134</sup> Check the normalization is correct.

with, by definition:

$$\langle s_n | f \rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \sin(nx) dx \quad (108a)$$

$$\langle c_n | f \rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \cos(nx) dx \quad (108b)$$

for  $n \neq 0$ , with also:

$$\langle c_0 | f \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) dx. \quad (109)$$

Let us look for instance how the diagonal vector—let us write it  $|/\rangle$ —such that  $\langle x | / \rangle = x$  looks like in terms of sines and cosines. Note that  $|/\rangle \neq |x\rangle$ , the latter being the function which is defined only at  $x$  (a sort of continuum version of the Kronecker delta, which is known as the Dirac delta function). According to Eqs. (108), we have to compute:<sup>135</sup>

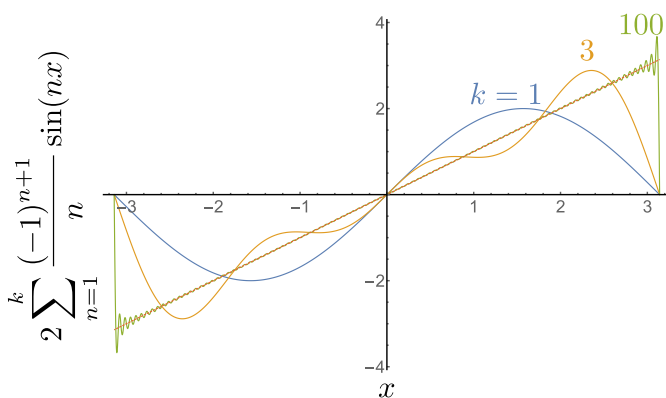
$$\langle s_n | / \rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} x \sin(nx) dx = 2\sqrt{\pi} \frac{(-1)^{n+1}}{n} \quad (110a)$$

$$\langle c_n | / \rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} x \cos(nx) dx = 0 \quad (110b)$$

for  $n \neq 0$ , with also  $c_0 = 0$ . We do not need to compute cos integrals since the function is odd (cos is even but  $x$  is odd so their product is odd) and thus with cancelling integral on a symmetrical interval. Therefore, our statement is:

$$x = 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(nx). \quad (111)$$

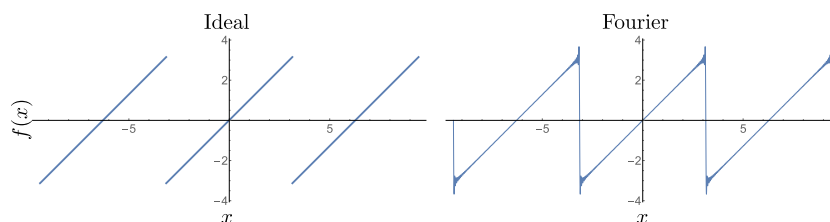
We can check it with the computer:



Notice what happens at the border. This is known as Runge's phenomenon. The point here is that while we work on the space of functions defined on  $[-\pi, \pi]$ , in fact, because all functions from the

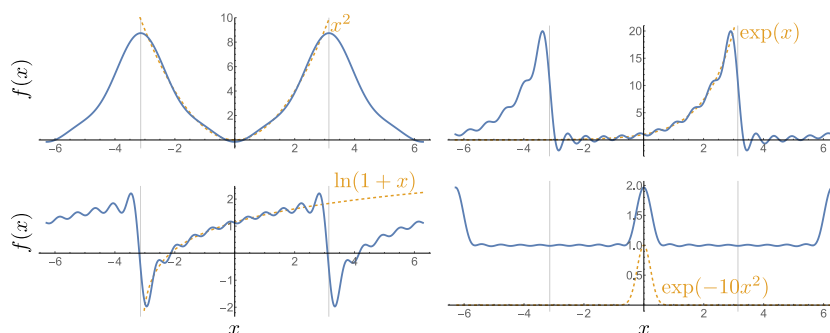
<sup>135</sup> Compute this (you can do it by parts).

basis simply repeat from there on to the rest of the  $x$  axis, what we really do is the decomposition of a periodic function of period  $2\pi$ . The continuation of  $x$  to the rest of the  $x$  axis with this constrain is not  $f(x) = x$  but the see-saw function as follows:



so the procedure is actually struggling with the breaking point. It oscillates there because we are taking finite series (with the computer) but in the limit of infinite sums, the match is perfect. This is known as Gibbs' phenomenon. You might wonder what happens precisely at the discontinuity then? Although not obvious mathematically, the result is the average of both sides (which has a nice symmetric look).

Of course this works for all functions. Here are some plotted results which we leave to you to obtain the coefficients (there are 10 terms in all cases except for  $x^2$  where there are only 3; the Gaussian case is displaced for better visualisation):<sup>136</sup>



<sup>136</sup> That's do-able for the exp and  $x^2$ ; play with the log and Gaussian if you want but a computer would help there.

This is neat, but there is an even more powerful and more succinct way to put it, which involves complex numbers, even though we are dealing with real-valued functions!

This is based on the observation that the Taylor series of the sine and cosine

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \quad (112a)$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \quad (112b)$$

have a peculiar connection to that of the exponential:

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} \dots \quad (113)$$

It is very close, but not quite, the sign is alternating, in fact, oscillating, in the sine and cosine, in a way which does not happen in the exponential. However, we know just the type of object which oscillates with successive powers:

$$i, \quad i^2 = -1, \quad i^3 = -i, \quad i^4 = 1, \quad i^5 = i, \quad \text{etc.} \quad (114)$$

so that if we look at the Taylor expansion of  $\exp(ix)$ , we find, by substituting in Eq. (113):

$$\exp(ix) = 1 + ix - \frac{x^2}{2} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - \frac{x^7}{7!} \cdots \quad (115)$$

so that, collecting the terms with an  $i$  left and those which cancel it altogether:

$$\exp(ix) = \left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - i\frac{x^7}{7!} + \cdots\right) \quad (116)$$

i.e., comparing with Eq. (112)

$$\boxed{e^{ix} = \cos x + i \sin x}. \quad (117)$$

This is known as Euler's formula, and has been described by Feynman as the most important formula of Mathematics. It links through the complex variable seemingly unrelated but extremely fundamental mathematical functions:

$$\cos(x) = \operatorname{Re}(e^{ix}) = \frac{e^{ix} + e^{-ix}}{2}, \quad (118a)$$

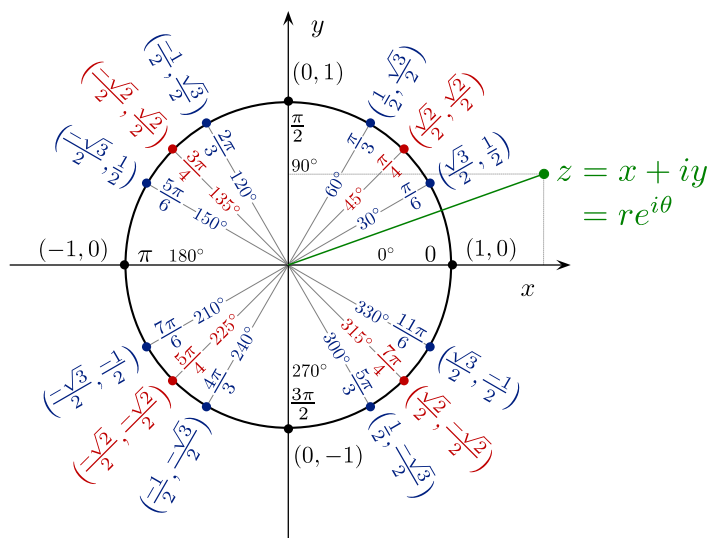
$$\sin(x) = \operatorname{Im}(e^{ix}) = \frac{e^{ix} - e^{-ix}}{2i}. \quad (118b)$$

The important relations (118) can also be obtained from:<sup>137</sup>

$$e^{-ix} = \cos x - i \sin x. \quad (119)$$

<sup>137</sup> Show that this follows straightforwardly from Eq. (117)

This shows that complex numbers really are "trigonometric" numbers! In particular, we can now write any complex number either in its Cartesian representation  $z = x + iy$  or its polar representation  $z = re^{i\theta}$ :



with the following relations:

$$x = r \cos \theta \quad (120a)$$

$$y = r \sin \theta \quad (120b)$$

(projecting on a circle), and the other way around:<sup>138</sup>

$$r = \sqrt{x^2 + y^2} \quad (121a)$$

$$\theta = \arctan(y/x) \quad (121b)$$

<sup>138</sup> What happens to  $\theta$  when  $x = 0$ ?  
Complete or precise Eqs. 121.

The geometric consequences are actually even simpler in polar form.

Consider

$$|z|, \quad z^*, \quad z^2, \quad z^n, \quad \frac{1}{z}, \quad \text{etc.} \quad (122)$$

Complex exponentials are also easier to work with than trigonometric functions thanks to the power algebra  $a^b a^c = a^{b+c}$  and  $(a^b)^c = a^{bc}$ , so remarkable identities which might be very difficult to derive without complex numbers are straightforward with Eq. (117), for instance:

$$e^{2ix} = (e^{ix})^2 = (\cos x + i \sin x)^2 = (\cos^2 x - \sin^2 x) + 2i \cos x \sin x \quad (123)$$

so that, equating real and imaginary part, we find:<sup>139</sup>

$$\cos 2x = \cos^2 x - \sin^2 x \quad (124a)$$

$$\sin 2x = 2 \cos x \sin x \quad (124b)$$

You might know Eq. (124a) from its “all-cos” version obtained by substituting  $\cos^2 + \sin^2 = 1$ :

$$\cos 2x = 2 \cos^2 x - 1. \quad (125)$$

<sup>139</sup> Do the same with the cube and jump directly to the  $n$ th term and the binomial theorem.

These are particular cases of the most general result, that is equally easily obtained:

$$e^{i(a+b)} = e^{ia} e^{ib} = (\cos a + i \sin a)(\cos b + i \sin b) \quad (126a)$$

$$= (\cos a \cos b - \sin a \sin b) + i(\cos a \sin b + \sin a \cos b) \quad (126b)$$

or, equating real and imaginary parts:

$$\cos(a+b) = \cos a \cos b - \sin a \sin b \quad (127a)$$

$$\sin(a+b) = \cos a \sin b + \sin a \cos b \quad (127b)$$

also with a mnemonic way to remember which goes with which sign (cos are real parts so get the  $i^2 = -1$  sign of like terms, two cosines and two sines).

This has merits in countless areas of Mathematics. Consider for instance integration and the extension of our algebra to complex numbers, i.e.,

$$\int e^{i\alpha x} dx = -\frac{i}{\alpha} e^{i\alpha x} \quad (128)$$

(remember,  $1/i = -i$ ) which allows to easily solve a basic question like, what is the primitive of  $\cos^2$ ? We do not have to even think, we compute:

$$\int \cos^2 x dx = \int \left( \frac{e^{ix} + e^{-ix}}{2} \right)^2 dx \quad (129a)$$

$$= \int \frac{e^{2ix} + e^{-2ix} + 2}{4} dx \quad (129b)$$

$$= \frac{1}{2} \left[ \int \cos 2x dx + \int dx \right] \quad (129c)$$

$$= \frac{1}{4} \sin 2x + \frac{x}{2} \quad (129d)$$

plus constant.<sup>140</sup> Here is an even more compelling example that would be tedious with real-valued trigonometric only, but that is fairly straightforward with complex exponentials:

$$\int \sin^2 x \cos 4x dx = \int \left( \frac{e^{ix} - e^{-ix}}{2i} \right)^2 \left( \frac{e^{4ix} + e^{-4ix}}{2} \right) dx \quad (130a)$$

$$= -\frac{1}{8} \int (e^{2ix} - 2 + e^{-2ix}) (e^{4ix} + e^{-4ix}) dx \quad (130b)$$

$$= -\frac{1}{8} \int (e^{6ix} - 2e^{4ix} + e^{2ix} + e^{-2ix} - 2e^{-4ix} + e^{-6ix}) dx. \quad (130c)$$

$$= -\frac{1}{24} \sin 6x + \frac{1}{8} \sin 4x - \frac{1}{8} \sin 2x. \quad (130d)$$

There are many techniques of this type, we illustrate a last one that should give you a fair overview of the might of complex exponentials. We compute the primitive of  $e^x \cos x$  as  $\operatorname{Re} \int e^x e^{ix} dx$  which

<sup>140</sup> Check the result with the fundamental theorem of calculus.

complex integration is trivial:

$$\int e^x e^{ix} dx = \int e^{(1+i)x} dx = \frac{e^{(1+i)x}}{1+i} \quad (131)$$

so that<sup>141</sup>

$$\int e^x \cos x dx = \operatorname{Re} \left( \frac{e^{(1+i)x}}{1+i} \right) \quad (132a)$$

$$= e^x \operatorname{Re} \left( \frac{e^{ix}}{1+i} \right) \quad (132b)$$

$$= e^x \operatorname{Re} \left( \frac{e^{ix}(1-i)}{2} \right) \quad (132c)$$

$$= e^x \frac{\cos x + \sin x}{2}. \quad (132d)$$

<sup>141</sup> Check with the fundamental theorem of calculus.

There are much more benefits, but let us now come back on how it affects the decomposition we started this lecture with. Using complex exponentials as basis functions make the whole procedure simpler and tidier. We now use functions:

$$|n\rangle \equiv \frac{1}{\sqrt{2}}(|c_n\rangle + i|s_n\rangle). \quad (133)$$

This  $1/\sqrt{2}$  is typical of such “superpositions”, it is there to maintain the normalization of a vector which arises from the sum of two vectors.<sup>142</sup> When dealing with sines and cosines independently, not only did we have to keep them apart, but we also had to single out the case  $n = 0$ , cf. Eqs. (104) and (105). Now you see where the  $\sqrt{2}$  was coming from: the complex case! (which is therefore more general and fundamental).

<sup>142</sup> Prove it.

When working with complex-valued functions, there is something important to keep in mind when going to the dual space, i.e., when going from kets to bra. Namely, we have not only to transpose operators (matrices) but also conjugate scalars, so that Eq. (133) becomes:

$$\langle n| = \frac{1}{\sqrt{2}}(\langle c_n| - i\langle s_n|). \quad (134)$$

From Euler’s formula, we get the functional form of  $|n\rangle$  in the  $x$  space:

$$\langle x|n\rangle = \frac{1}{\sqrt{2n}} e^{inx}. \quad (135)$$

For the closure relation, we could try  $\sum_{n=0}^{\infty} |n\rangle \langle n|$  but this gives us

$$\sum_{n=0}^{\infty} |n\rangle \langle n| = \frac{1}{2}(|c_n\rangle + i|s_n\rangle)(\langle c_n| - i\langle s_n|) \quad (136a)$$

$$= \frac{1}{2} [ |c_n\rangle \langle c_n| + |s_n\rangle \langle s_n| + i|s_n\rangle \langle c_n| - i|c_n\rangle \langle s_n| ] \quad (136b)$$

where we see the benefit of the complex-conjugate, it brings us to a form close to Eq. (106), except for the cross-terms, the complex valued terms which mix sines and cosines, and this annoying factor 1/2. We get rid of them all thanks to the trick:

$$|c_{-n}\rangle = |c_n\rangle \quad \text{and} \quad |s_{-n}\rangle = -|s_n\rangle \quad \text{and} \quad (137)$$

so that, in a mixed product, the sign survives. This allows us to establish, from Eq. (106):

$$\sum_{n=-\infty}^{\infty} |n\rangle\langle n| = \mathbb{1}. \quad (138)$$

That's another closure relation we see, this time with complex exponentials, and we start to see the trend: we add all the projectors of the space. Here is another useful closure relation, with  $|x\rangle$ , which however requires an integral rather than a sum, due to  $x$  being a continuous variable, not a discrete one!

$$\int |x\rangle\langle x| dx = \mathbb{1}. \quad (139)$$

Let us see how  $|/\rangle$  transforms in terms of complex exponentials:

$$|/\rangle = \mathbb{1} |/\rangle \quad (140a)$$

$$= \sum_{n=-\infty}^{\infty} |n\rangle\langle n| |/\rangle \quad (140b)$$

$$= \sum_{n=-\infty}^{\infty} |n\rangle\langle n| \mathbb{1} |/\rangle \quad (140c)$$

$$= \sum_{n=-\infty}^{\infty} |n\rangle\langle n| \left( \int |x\rangle\langle x| dx \right) |/\rangle \quad (140d)$$

$$= \sum_{n=-\infty}^{\infty} |n\rangle \left( \int \langle n|x\rangle\langle x|/\rangle dx \right) \quad (140e)$$

$$= \sum_{n=-\infty}^{\infty} |n\rangle \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{-inx} x dx \quad (140f)$$

where we pick up a sign in the complex exponential because it comes as a bra and is, precisely, complex, or, said otherwise,  $\langle n|x\rangle = (\langle x|n\rangle)^* = (e^{inx})^* = e^{-inx}$ . The integral is easy to do:<sup>143</sup>

<sup>143</sup> Do it.

$$\int_{-\pi}^{\pi} e^{-inx} x dx = \frac{2i\pi(-1)^n}{n} \quad \text{if } n \neq 0 \quad (141)$$

and 0 if  $n = 0$ , so that

$$|/\rangle = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{2i\pi(-1)^n}{n\sqrt{2\pi}} |n\rangle \quad (142)$$



or, if you prefer the full expression:

$$x = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{i(-1)^n}{n} e^{inx} \quad (143)$$

where you see how the Dirac notation allows us to keep track of all these pesky constants, when we need to turn to a concrete or given space.

All functions (that satisfy Dirichlet's theorem) can be thus decomposed, for any variable. For instance, in the time domain, we can write a time-varying function  $F(t)$  as a function of complex time-exponentials as  $F(t) = \sum_{n=0}^{\infty} F_n e^{int}$ . As we start to migrate from Maths to Physics, and since we now speak about time, it would be wise to ensure that  $n$  is not merely an integer but a physical quantity with a dimension. Since this is part of an exponential, i.e., a nonlinear function, this needs to be dimensionless.<sup>144</sup> So we would instead write:

<sup>144</sup> Do you see why?

$$F(t) = \sum_{n=0}^{\infty} F_n e^{i\omega_n t} \quad (144)$$

where  $\omega_n$  are frequencies (i.e., something with a dimension of inverse time) but a countable set of frequencies, that is, which can be indexed by a discrete sum. Such an expression is called a discrete Fourier transform.

Let us illustrate all the above on another variation on the oscillator dynamics, this time, instead of coupling two harmonic oscillators, we will come back to one only but add damping and driving. We will not lose time explaining how this very basic and fundamental problem turns out to be extremely important in all areas of physics, where things usually come with both frictions, losses and dissipation of all sorts, as well as external stimuli, pumping and excitations of all sorts. Considering their combined effect on the dynamics of the harmonic oscillator is clearly very important. Newton's equation of motion for the oscillator of mass  $m$  and coupling strength  $k$  now becomes:

$$m\ddot{x} + \gamma\dot{x} + kx = F(t) \quad (145)$$

where  $\gamma$  is the dissipative term (a common type, proportional to the speed of the oscillator) and  $F(t)$  is the time-dependent driving, which is of arbitrary type. Now lo and behold how quickly and efficiently we can get the exact and complete solution to this general problem, to which we will return in next lectures in a more systematic way. There we will see that the solution consists of a transient term, which is the part that describes the time it needs for the oscillator to adapt to the driving, and the stationary solution, when the oscillator's response has locked with the driving. We will focus on the later type.

The point is that while  $F(t)$  is general, it is easy with complex exponentials to solve Eq. (145) when  $F(t)$  is itself a complex exponential:

$$m\ddot{x} + \gamma\dot{x} + kx = F_n e^{i\omega_n t} \quad (146)$$

where  $F_n$  is the magnitude of the “complex” driving and  $\omega_n t$  the phase, at time  $t$ , which goes uniformly around the trigonometric circle with the frequency  $\omega_n$ . What does it mean for a driving to be complex? Basically that we allow it to have a phase. We can also understand the imaginary part of our final solution,  $x$ , to describe a phase. But we could also, if we are conservative, just take the real part or imaginary part of both sides (which we can do by linearity) and then everything is real.

More interesting is the fact that one can guess a solution to be of the type:

$$X(t) = X_n e^{i\omega_n t} \quad (147)$$

where  $X_n$  can be complex, even if we take  $F_n$  real. This shows that the oscillator indeed dephases from the driving, but otherwise has the same frequency. This can be motivated physically. Instead, we just go to check. Substituting Eq. (147) in 146, we find:

$$-mX_n\omega_n^2 e^{i\omega_n t} + i\gamma\omega_n X_n e^{i\omega_n t} + kX_n e^{i\omega_n t} = F_n e^{i\omega_n t} \quad (148)$$

or, after simplifying by  $e^{i\omega_n t}$  and re-arranging:

$$X_n = \frac{F_n}{-m\omega_n^2 + i\gamma\omega_n + k}. \quad (149)$$

This is the solution! Easy. We would typically introduce the new variables  $f_n \equiv F_n/m$  and  $\beta \equiv \gamma/2$  to tidy-up the equation that reads:<sup>145</sup>

$$X_n = \frac{f_n}{\omega_0^2 - \omega_n^2 + 2i\beta\omega_n}. \quad (150)$$

So, for the real-valued-solutions inclined, if we want the real-valued solution corresponding to driving with a cosine, we take the real part of Eq. (150), if we want the solution corresponding to a driving sine, we take the imaginary part, out of which we could, through trigonometric identities, get any-real valued solution for any dephasing. But it is simpler, really, to accept complex exponentials.

Now we can easily obtain the full solution for arbitrary  $F(t)$  thanks to linearity by turning back to Eq. (144). If  $X_1 e^{i\omega_1 t}$  and  $X_2 e^{i\omega_2 t}$  are two solutions of Eq. (146), then  $X_1 e^{i\omega_1 t} + X_2 e^{i\omega_2 t}$  is a solution to the driven-damped oscillator with driving  $F_1 e^{i\omega_1 t} + F_2 e^{i\omega_2 t}$ . By induction (or iteration), it is then clear that

$$X(t) = \sum_{n=-\infty}^{\infty} X_n e^{i\omega_n t} \quad (151)$$

<sup>145</sup> Check. Remember that  $\omega_0 = \sqrt{k/m}$  is the resonant frequency of the harmonic oscillator.

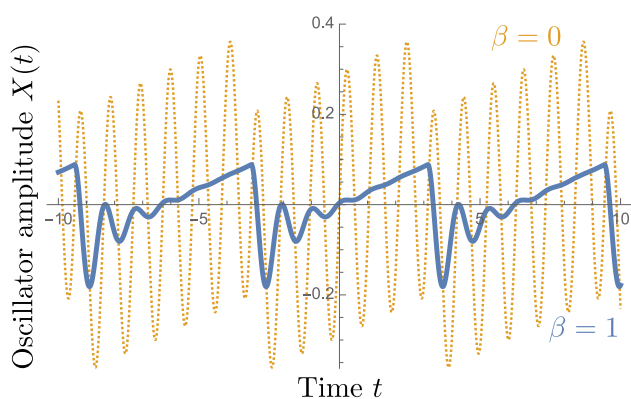
is the general solution, or, explicitly, inserting coefficients from Eq. (150):

$$X(t) = \sum_{n=-\infty}^{\infty} \frac{f_n}{\omega_0^2 - \omega_n^2 + 2i\beta\omega_n} e^{i\omega_n t}. \quad (152)$$

Now that the formalism is complete, that our technique is laid down, we could do some Physics. For instance, here is the actual displacement from a damped oscillator driven with a seesaw force with time period  $2\pi$ :

$$X(t) = \sum_{n=-\infty}^{\infty} \frac{i(-1)^n}{n} \frac{e^{int}}{\omega_0^2 - n^2 + 2i\beta n}, \quad (153)$$

and here are two numerical solutions for an oscillator with resonant frequency  $\omega_0 = 5.9$ , one solution is without damping (orange), the other with damping  $\beta = 1$ . You see that we did not need worry too much about our passage to the complex variable! The solution Eq. (153) is completely real, we don't need take its real or imaginary part, the latter cancels!



And this shows us indeed how the driving is forcing the oscillator to drift in amplitude under the action of its continuous push-to-pull<sup>146</sup> with a change of behaviour as the nature of the force changes, and how this gets damped in presence of dissipation, with oscillations restored as the nature of the force is changed, so that the oscillator has a small time-window to oscillate again. There is a lot of physics to look at there, including the notion of resonance and quality factor, but we leave this to relevant physics lectures (it could hit you from anywhere).

We will conclude with one observation on this powerful technique. Notice how we have been confined to periodic functions, so our solution is not completely general indeed. What would happen if the driving term wouldn't be periodic? For instance, what would happen if we send a Gaussian kick to our oscillator? What we could do is consider a large time window around our pulse so that in this time

<sup>146</sup> Check that the frequency observed on the numerical solution is that of the free oscillator.

window the Fourier series indeed cancel out. That would work, for a while. Then the second pulse would come. So we could increase the time window. Then the frequencies would get closer together, and the sum would differ by always smaller quantities. At this point, we could take the limit and turn our series into an integral. Then we could deal with non-periodic function in a more fundamental way. This is called Fourier-analysis. This is even more powerful than the Fourier series that we have seen today. We will leave it as a more advanced material, for next year, although it is only marginally more technical. You now have all the basic ideas underlying this considerable part of advanced mathematics!

### *Problems*

#### *Remarkable Trigonometric identities*

Show that:

$$\cos 3x = \cos^3 x - 3 \cos x \quad (154a)$$

$$\sin 3x = 3 \sin x - 4 \sin^3 x. \quad (154b)$$

Can you obtain Eqs. (124–125) and this's counterpart for  $n = 4$ ?

#### *de Moivre's formula*

De Moivre's formula is a famous (and spectacular) trigonometric identity:

$$(\cos x + i \sin x)^n = \cos nx + i \sin nx. \quad (155)$$

Prove it, in *two different ways* (it can also be proved easily by induction).

#### *hi hi ( $i^i$ )*

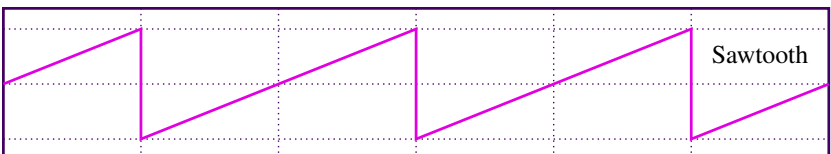
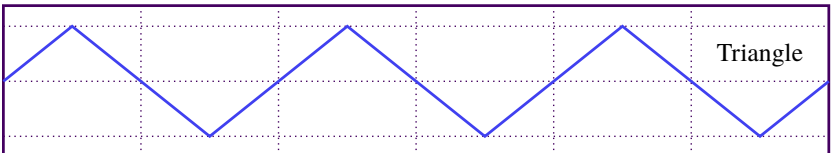
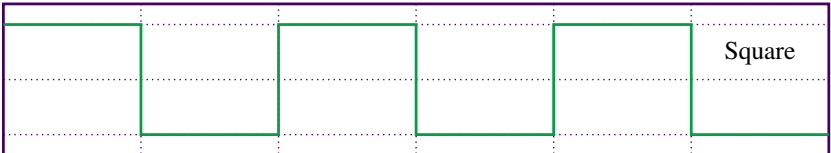
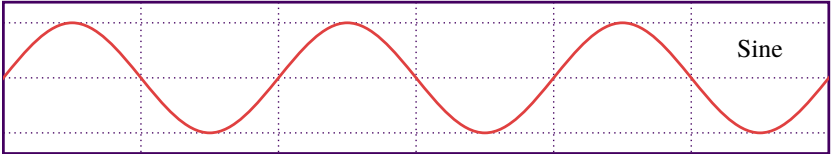
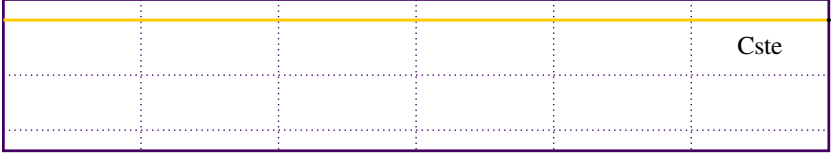
With complex exponentials, one can set to rest questions like, what is the  $i$ th power of a number, and in particular, what is  $i^i$ ? Writing  $i$  as  $\exp(i\pi/2)$ , show that  $i^i = \exp(-\pi/2)$  (and is therefore, surprisingly, real!)

#### $\pi/12$

Check that  $\pi/12 = (\pi/3) - (\pi/4)$  and use this to give an exact value for  $\sin(\pi/12)$  and  $\cos(\pi/12)$ . What other remarkable angles are thus now defined on the trigonometric circle? Could you extend this procedure to still other values?

*Driven oscillator*

Provide the solutions of the driven damped harmonic oscillator in the case where the driving is constant, a square wave and and triangle wave (we have covered the sine and sawtooth waves):





## Lecture 18: Differential equations.

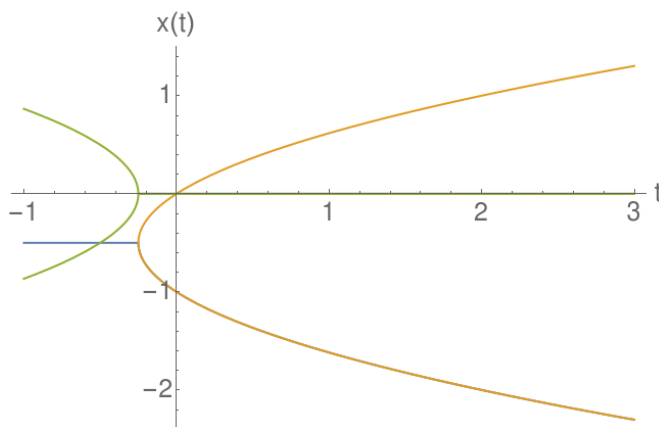
In Physics we need to solve a lot of equations. Now we know the most important things there is to know about linear equations. The world is also nonlinear (including at the fundamental level, for instance general relativity is nonlinear). The most important nonlinear equation is the quadratic equation, i.e., what values of  $x$  satisfy  $ax^2 + bx + c = 0$  and we found that such equations are “easy” in the sense that, with some effort of memory and practice, we can always give all the solutions. This equation is for a single variable so we expect numbers (possibly complex) as an answer. We will find that equations are often not only for variables, but directly for the functions themselves, e.g., which is the function  $x(t)$  such that

$$x(t)^2 + x(t) = t \quad (156)$$

This is still quadratic, and for any given time the formula we know for numbers applies, therefore we can actually also solve that at all times and thus solve it for functions. We find:

$$x(t) = \frac{1}{2}(-1 \pm \sqrt{1 + 4t}). \quad (157)$$

For  $t < -\frac{1}{4}$ , solutions are complex. So we can represent the solution as follows:



where the orange part is the real solutions, the green is the real part of the complex solutions and the the blue is the imaginary part of the

complex solutions. If the problem would only admit real solutions, then only the orange branches should be retained. Note that while Eq. (157) yields two solutions, the graph solution is a seamless curve. This is an important functional-equation, that is related to coupled *dissipative* oscillators so you'll definitely encounter it again.

So we took a big conceptual step but a fairly small technical one. A parameter becomes a variable... no big deal (technically, conceptually this is momentous). There is an even bigger conceptual step made when also derivatives of the function enter the equation, in which case, it becomes a so-called *differential equation* (in the grand scheme of things, this belongs to so-called "Ordinary Differential Equations", or ODE, as compared to PDE that involve partial derivatives, that we will include in the coming lectures). Most equations of Physics are such equations, so it is important to understand and know them. They are also very varied in character, and most of them cannot be solved exactly, so it will take us much time to get used to them. We will focus first on those that can be solved.

The derivative being the rate of change, it is clear that such equations will be commonplace, as they translate (as an equation) a statement regarding a given phenomenon. Consider for instance the snowball equation, which says how a snowball grows, namely, proportionally to its size  $S(t)$  which, of course, changes with time as the snowball rolls and gathers snow proportionally to its area. So the equation reads  $S' \propto S$ . To solve it, we'll specify the constant  $A$  and for clarity we will also write explicitly the variable  $t$ :

$$S'(t) = AS(t). \quad (158)$$

Solving a differential equation is an art, that include a mix of knowledge, technique, luck and inspiration. In this case, we recognize the main behaviour of a particular function, the exponential, so we can guess the solution to be:

$$S(t) = S(0) \exp(At) \quad (159)$$

where  $S(0)$  is a the size of the ball at  $t = 0$ . This is confirmed by direct substitution. This rescales its exponential growth, at the rate  $A$ . The exponential also play a central role in differential equation, as we shall see below.

One of the most important equation of Physics is Newton's  $F = ma$ . For the most celebrated case where  $F$  is gravity, i.e.,  $F = mg$ , since the acceleration is the rate of change  $\frac{d}{dt}$  of the velocity, being itself the rate of change  $\frac{d}{dt}$  of position, we have  $a = \frac{d}{dt} \frac{d}{dt} y = \frac{d^2 y}{dt^2}$  and Newton's equation reads:

$$y''(t) = g. \quad (160)$$



We want a function  $y$  which, when derived twice, is constant. Clearly, this is a quadratic function, to which we can hook linear and constant functions since these will be “killed” by the derivative and thus also satisfy Eq. (160). Therefore, we try:

$$y(t) = \alpha t^2 + \beta t + \gamma \quad (161)$$

and find:

$$y' = 2\alpha t + \beta, \quad (162a)$$

$$y'' = 2\alpha, \quad (162b)$$

which shows that  $\alpha = g/2$ , and  $\beta, \gamma$  can be anything. So the solutions are:

$$y(t) = \frac{g}{2}t^2 + \beta t + \gamma, \quad \beta, \gamma \in \mathbb{R}. \quad (163)$$

In Physics, quantities have a dimension, which is useful to track that everything is done correctly. Here we see that  $\beta t$  being a distance (like  $y$ , on the lhs of the equation), it means that  $\beta$  is a speed (something which multiplied by time gives a distance) and  $\gamma$  is, directly, a distance. We don't get to choose  $g$ , which needs to be an acceleration, and indeed it is the acceleration of (Earth's) gravity,  $g \approx -9.81 \text{ m s}^{-2}$ . So the solution is the trajectory for Newton's apple, with the speed and eight as important parameters for the initial condition:  $\gamma$  sets the initial altitude, and  $\beta$  the initial speed. If we take  $y = 0$  as the floor, then we better start at a positive altitude, or we could throw the apple in the air ( $v_0 > 0$ ) to get it accelerate, reach a maximum, and drop down again.

Those are interesting particular cases. Now let us delve into the more general theory. We will use various notations depending on the context but for abstract purposes, maybe the most meaningful notation is  $y(x)$  ( $y$  is the function,  $x$  the variable). The most general type of so-called differential equation for such functions (of one variable) reads:

$$\boxed{F(x, y, y', y'', \dots, y^{(n)}) = 0}, \quad (164)$$

with  $n \in \mathbb{N}$  is the *order* of the equation, and  $F$  is another function that relates this various quantities together. Eq. (158) is 1st order with  $F(a, b, c) = c - Ab$  while Eq. (160) is 2nd order with  $F(a, b, c, d) = d - g$ . The simplest differential equation is  $F(a, 0, c) = c - f(a)$  or:

$$y'(x) = f(x) \quad (165)$$

which we can formally solve as:

$$y(x) = \int f(x) dx + C. \quad (166)$$

Indeed, we refer to the solving of Eq. (165) as “integrating” the equation. We recognize Eqs. (165) and (166) as the “*fundamental theorem of calculus*” which state that differentiation and integration are the inverse of each other. This is so simple that people would not even call such equations “differential” equations. However, other differential equations of the  $F(a, b, c)$  type, or

$$y' = f(x, y) \quad (167)$$

where the function appears as well as the derivative, can sometimes be brought to this type. This is the case of “*separable equations*” that can be brought in the form:

$$y'(x) = f(x)g(y) \quad (168)$$

as compared to a non-separable equation:

$$y'(x) = h(x, y) \quad (169)$$

with  $h(x, y) \neq f(x)g(y)$ .

Equation (167) is called a “first-order differential equation”. This cannot be solved in the most general case, even when we can separate variables because of the difficulty of finding primitives. But in the separable case, at least we can break the problem in two:

$$\frac{dy}{dx} = f(x)g(y) \quad (170)$$

with all same-type of variables on each side of the equation:

$$\frac{dy}{g(y)} = f(x)dx \quad (171)$$

which can now be integrated independently (as one-variable type of problems):

$$\int \frac{dy}{g(y)} = \int f(x)dx + C. \quad (172)$$

Often we will not be able to solve differential equations exactly, and even when we can, sometimes we can only get a solution not in closed-form, or not explicit, in the sense that the function is nicely on one side of the equation. This

$$xy = \ln y + C \quad (173)$$

for instance, is an example of an *implicit* solution to the differential equation:

$$y' = \frac{y^2}{1 - xy}. \quad (174)$$

You can check that differentiating both sides and re-arranging. You can also explore this solution as an Exercise. Solvable cases are few

and make for nice (but somehow artificial) examples. Here is an example where primitives are simple enough to bring us to the final solution:

$$y' = (xy)^2. \quad (175)$$

We can rewrite it as:

$$\frac{dy}{dx} = x^2 y^2 \quad (176a)$$

$$\int \frac{dy}{y^2} = \int x^2 dx \quad (176b)$$

$$\frac{-1}{y} = \frac{1}{3}x^3 + c \quad (176c)$$

$$y = \frac{-3}{x^3 + c}. \quad (176d)$$

Often we want the solution that satisfies a particular initial condition, e.g., such that  $y(0) = 1$ , so that  $y(0) = -3/c = 1 \implies c = -3$ .

There are tricks to work out some particular cases or reduce their complexity to bring them to cases which can be tackled. We will give one such trick as an illustration, the so-called Bernoulli's equation:

$$y'(x) + P(x)y = Q(x)y^n. \quad (177)$$

Such equations arise for instance in some models of damping:<sup>147</sup>

$$\dot{v} = -\gamma v + \nu v^3. \quad (178)$$

Here the trick is to make the change of variable  $z \rightarrow y^{1-n}$ , because  $z' = (1-n)y'y^{-n}$  and thus, multiplying both sides of Eq. (177) by  $y^{-n}$ , we find:

$$z' + (1-n)P(x)z = (1-n)Q(x) \quad (179)$$

and here we made a considerable simplification, because if we take  $Q = 0$ , then this equation is of a very particular character, one which we have encountered already, namely, it is *linear* (in  $y$ , note that  $P$  can still be highly nonlinear). We will thus now focus on such equations:

$$\boxed{y'(x) + P(x)y(x) = Q(x)}. \quad (180)$$

Indeed, if  $y_1$  and  $y_2$  are two solutions of Eq. (180) with  $Q = 0$ , then  $\alpha y_1 + \beta y_2$  is also a solution of  $y' + Py = 0$ . The fact that  $Q$  is nonzero breaks the linearity, but not in a serious way. We call this a *non-homogeneous* linear equation, and we will see methods to tackle this complication in the general case. The point is that solutions to the homogeneous case will form the basis for solutions of the most general case, including the non-homogeneous one.

<sup>147</sup> Here  $v$ , or speed, is the function of  $t$ , time. Solve this equation once you've learned the trick and compare to the case where  $\nu = 0$ . Why is the cubic term desirable on physical grounds?

The trick to solve a first-order linear equation is to take advantage of the special structure of the exponential under derivation. If we use this particular combination:

$$\frac{d}{dx} \left( e^{\int_{\alpha}^x P(z) dz} y(x) \right) = \left( \frac{d}{dx} e^{\int_{\alpha}^x P(z) dz} \right) y(x) + e^{\int_{\alpha}^x P(z) dz} \frac{d}{dx} y(x) \quad (181a)$$

$$= e^{\int_{\alpha}^x P(z) dz} P(x) y(x) + e^{\int_{\alpha}^x P(z) dz} y'(x) \quad (181b)$$

$$= e^{\int_{\alpha}^x P(z) dz} (y' + Py) \quad (181c)$$

so, multiplying Eq. (180) by  $e^{\int_{\alpha}^x P(z) dz}$  yields:

$$\frac{d}{dx} \left( e^{\int_{\alpha}^x P(z) dz} y \right) = Q(x) e^{\int_{\alpha}^x P(z) dz} \quad (182)$$

which, by integration of both sides, brings us to:

$$e^{\int_{\alpha}^x P(z) dz} y + c = \int Q(x) e^{\int_{\alpha}^x P(z) dz} dx \quad (183)$$

with  $c$  a constant (don't forget it!); note that its actual value is unknown, so we could have written  $-c$  here and pass it on the other side:

$$y = e^{-\int_{\alpha}^x P(z) dz} \left( \int Q(x) e^{\int_{\alpha}^x P(z) dz} dx + c \right) \quad (184)$$

where you see that  $c$  is important because it gets multiplied by the exponential.

As an example, let us solve:

$$y' + \frac{y}{x} = 3x \quad (185)$$

Direct application of Eq. (183) leads us to:

$$y = e^{-\ln(x) + \ln(\alpha)} \left( \int 3x e^{\ln(x) - \ln(\alpha)} dx + c \right) \quad (186a)$$

$$= \frac{\alpha}{x} \left( \int 3x \frac{x}{\alpha} dx + c \right) \quad (186b)$$

$$= \frac{\alpha}{x} \left( \frac{x^3}{\alpha} + c \right) \quad (186c)$$

$$= x^2 + \frac{c\alpha}{x}, \quad (186d)$$

which is a bit cumbersome as we have to carry over the unimportant constant  $\alpha$  (which cancels or gets absorbed in  $c$  eventually). Actually, it is more convenient to first compute

$$\boxed{e^{\int P(x) dx}} \quad (187)$$

as a primitive (not caring about the constant), which is called, by the way, the *integrating factor* and which is, in this case:

$$\int P(x) dx = \int \frac{dx}{x} = \ln(x) \quad \text{and} \quad e^{\int P(x) dx} = x \quad (188)$$

with the effect that the derivative of the integrating factor times  $y$ , that is,  $(ye^{\int P(x) dx})'$ , gives the left-hand side of the original equation,  $y' + Py$ , times the integrating factor (one should check that to make sure there has been no mistake), so one can equate this derivative to the right-hand side times the integrating factor, i.e., to  $Qe^{\int P}$ ; in the example at hand:

$$(xy)' = 3x^2 \quad (189)$$

which is trivially integrated

$$xy = x^3 + c \quad (190)$$

so that

$$y = x^2 + c/x, \quad (191)$$

which we should check satisfies the original equation:

$$y = x^2 + cx^{-1} \quad (192a)$$

$$y' = 2x - cx^{-2} \quad (192b)$$

$$xy' = 2x^2 - cx^{-1} \quad (192c)$$

$$xy' + y = 3x^2 \quad (192d)$$

which is Eq. (185).

Here is another, more sophisticated example:

$$\cos(x)y' + \sin(x)y = 2\cos^3(x)\sin(x) - 1 \quad (193)$$

We first rewrite it in the general form of Eq. (180), which is:

$$y' + \tan(x)y = 2\cos^2(x)\sin(x) - \sec(x) \quad (194)$$

where we used the notation  $\sec x \equiv 1/\cos x$ . So  $P(x) = \tan x$  and the integrating factor (187) is:

$$e^{\int \tan(x) dx} = e^{-\ln \cos(x)} = \sec(x) \quad (195)$$

(we need to remember the primitive of the tangent; if we don't, we can use the substitution method since  $\tan(x)dx = -d \cos(x)/\cos(x)$ ). Multiplying Eq. (194) by the integrating factor, we find:

$$\frac{y'}{\cos(x)} + \frac{\sin(x)}{\cos^2(x)}y = 2\cos(x)\sin(x) - \sec^2(x) \quad (196)$$

and the left-hand-side of Eq. (196) is the derivative of the integrating factor times the function itself, i.e.,

$$(y \sec x)' \quad (197)$$

and integrating both sides of Eq. (196) with this new left-hand-side, we thus gets:

$$y \sec x = \int \sin(2x) dx - \int \sec^2(x) dx \quad (198)$$

which can be integrated exactly (note that we used  $\sin(2x) = 2 \sin x \cos x$ ), since  $\sec^2(x) = 1/\cos^2(x)$  is the derivative of  $\tan(x)$ . Namely:

$$y \sec x = -\frac{1}{2} \cos(2x) - \tan(x) + c \quad (199)$$

so that, finally:

$$y = -\frac{1}{2} \cos(2x) \cos(x) - \sin(x) + c \cos(x). \quad (200)$$

We let you check that this is the solution of Eq. (193).<sup>148</sup>

<sup>148</sup> Check it. Find the solution of Eq. such that  $y(0) = 0$ .

We conclude with a first glimpse into higher-order cases with instances that, like the Bernoulli equations, can be brought back to the first-order case. Consider

$$f(x, y, y', y'') = 0. \quad (201)$$

In the case where either  $x$  or  $y$  is missing, then we can lower the order.

The case of  $y'$  missing is trivial. In this case, we really have

$$f(x, y', y'') = 0 \quad (202)$$

and if we introduce  $z = y'$ , then  $z' = y''$  and we have:

$$f(x, z, z') = 0 \quad (203)$$

which is first-order. The case where  $x$  is missing is more subtle but follows the same principle. We still use  $y' = z$  but now we strive to make  $y$  the independent variable, so we compute:

$$y'' = \frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = z \frac{dz}{dy} \quad (204)$$

with  $z = y' = \frac{dy}{dx}$  so  $f(y, y', y'') = 0$  becomes  $f(y, z, z \frac{dz}{dy}) = 0$  which is also 1st order.

## Exercises

### Just checking

Check that the following functions (explicit or implicit) are solutions of the corresponding differential equation:

$$y = \alpha \sin(2x) + \beta \cos(2x) \quad y'' + 4y = 0 \quad (205a)$$

$$y = \alpha \sinh(2x) + \beta \cosh(2x) \quad y'' - 4y = 0 \quad (205b)$$

$$x^2 = 2y^2 \ln y \quad y' = \frac{xy}{x^2 + y^2} \quad (205c)$$

$$y = \arcsin(xy) \quad xy' + y = y' \sqrt{1 - x^2 y^2} \quad (205d)$$

For Eq. (205b), we used the so-called hyperbolic trigonometric functions

$$\cosh(x) = \frac{e^x + e^{-x}}{2}, \quad (206)$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2}. \quad (207)$$

You can check that  $e^x$  and  $e^{-x}$  are also, in fact, solutions.

### Separation of variable

Solve the equations

$$y' = \frac{\cos x}{y}, \quad (208a)$$

$$y' = \exp(x + y), \quad (208b)$$

$$y' = (x \cos y)^2. \quad (208c)$$

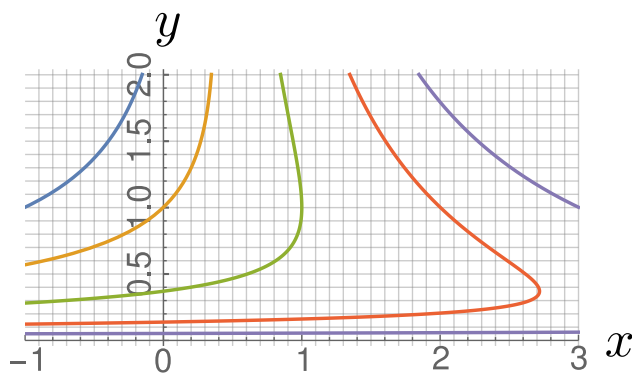
and check your results.

### Graphical solutions

These are some solutions  $xy = \ln y + C$  for  $C \in \{-1, 0, 1, 2, 3\}$  to the equation

$$y' = \frac{y^2}{1 - xy} \quad (209)$$

mentioned in the text.



Check that the various curves do indeed satisfy Eq. (209) from direct measurements on the graph.

### Hermite polynomials

Show that the following differential equation

$$y'' - 2xy' + 2my = 0 \quad (210)$$

has the following (so-called Hermite polynomial) solutions:

$$m = 0, \quad 1 \quad (211a)$$

$$1, \quad 2x \quad (211b)$$

$$2, \quad 4x^2 - 2 \quad (211c)$$

$$3, \quad 8x^3 - 12x \quad (211d)$$

$$4, \quad 16x^4 - 48x^2 + 12 \quad (211e)$$

$$5, \quad 32x^5 - 160x^3 + 120x \quad (211f)$$

Can you guess the higher  $m$  solutions?

### Legendre polynomials

We have introduced Legendre polynomials as an orthogonal basis of polynomials, with  $P_0 = 1$ ,  $P_1 = x$ ,  $P_2 = \frac{1}{2}(3x^2 - 1)$ , etc. They are also obtained as solutions of the following differential equation

$$\frac{d}{dx} \left[ (1-x^2) \frac{dP_n(x)}{dx} \right] + n(n+1)P_n(x) = 0 \quad (212)$$

for integer  $n$ . Check it.

### No clue given

Compute the following integrals (we don't give you a clue which method to use):

- |   |   |
|---|---|
| 1. $\int \frac{(1-x)^2}{1-x^2} dx$          | 6. $\int \sin^3 x \cos^3 x dx$              |
| 2. $\int \frac{x}{\sqrt{25x^2-4}} dx$       | 7. $\int 2xe^{x^2+1} dx$                    |
| 3. $\int \frac{\sin \sqrt{x}}{\sqrt{x}} dx$ | 8. $\int e^{\cos x} \sin x dx$              |
| 4. $\int \frac{\cos x}{1+\sin x} dx$        | 9. $\int \frac{dx}{\sqrt{x}(1+\sqrt{x})^2}$ |
| 5. $\int \sin^2 x \cos^3 x dx$              | 10. $\int \frac{2-x^2}{(1-x^2)(1-4x^2)} dx$ |

### Trick and treat

This is a nice illustration of how clever tricks can be used to solve an integral. We know that a odd function over a symmetric integral is zero. Consider:

$$\int_{-1}^1 \frac{\cos x}{\exp(1/x)+1} dx. \quad (213)$$



This function is neither odd nor even, but it can be decomposed as a sum of an odd and an even function, as we have seen previously (Lecture 11). The integral of the odd part will vanish since the domain is symmetric, so we are left with the even part, that turns out to be a very simple function which is straightforward to integrate exactly. Find the exact value for the definite integral (213) (it is  $\approx 0.84$ ).

### *Free fall*

We have seen already the free-fall equation  $y'' = g$ . If we include air resistance, then a term proportional to the velocity appears in the equation, which becomes:

$$y'' = g - cy'. \quad (214)$$

What are the units of  $c$ ? How does the velocity vary as a function of time? (The quantity  $g/c$  is called the “terminal velocity”, explain why).

### *Linear equations*

Solve the following linear differential equations:

1.  $xy' - 3y = x^4$ .
2.  $xy' + 2y = x^2 - x + 1$ .
3.  $xy' - 2y = x^5 \sin(2x) - x^3 + 4x^4$ .
4.  $y' + y = 1/(1 + e^{2x})$ .
5.  $(x \ln x)y' + y = 3x^3$ .

### *Long-term behaviour*

Solve the following differential equation for  $y(t)$  with  $y(0) \equiv y_0$ :

$$2y' - y = 4 \sin(3t) \quad (215)$$

and study the long-term behaviour. Show that for  $y_0 = -24/37$ , the range of values taken by  $y$  is finite and is unbounded otherwise.

### *Bernoulli equations*

Solve the following differential equations:

1.  $xy' + y = x^4y^3$ .
2.  $xy' + y = xy^2$ .
3.  $xy^2y' + y^3 = x \cos x$ .

*Bernoulli spirit*

Show that

$$y' + Py = Qy \ln y \quad (216)$$

can be solved by the change of variable  $z = \ln y$ . Solve:

$$xy' = 2x^2y + y \ln y. \quad (217)$$

*Lowering the order*

Solve the following equation by lowering the order to 1st-order:

$$xy'' - y' = 3x^2. \quad (218)$$

We now address this even more important case, to which we shall return in next lecture, but that can be solved with techniques introduced in this one:

$$y'' + k^2y = 0. \quad (219)$$

*Falling*

In the free-fall problem above, what happens if air resistance brings a force proportional to the square of the velocity rather than to the velocity? (use separation of variable and the fact that  $(\tanh^{-1}(x))' = 1/(1-x^2)$ ).

*Escaping*

We remind Newton's force of gravity between two massive bodies of mass  $m_1$  and  $m_2$ :

$$F = \frac{m_1 m_2 G}{r^2}. \quad (220)$$

Compute the escape velocity for an object sent from Earth (we remind that  $g = GM_{\oplus}/r_{\oplus}^2 \approx 9.81 \text{ m s}^{-2}$  with  $M_{\oplus}$  the mass of Earth and  $r_{\oplus} \approx 6400 \text{ km}$  the radius of Earth). Clue: You can find it by writing an equation for  $v(r)$  such that  $v(\infty) = 0$  and use separation of variables to integrate the equation).

Show that the maximum height reached by the object if sent at a velocity  $v_0$  that is smaller than the escape velocity  $v_e$  is:

$$h_{\max} = \frac{(v_0/v_e)^2}{1 - (v_0/v_e)^2} R. \quad (221)$$

## Lecture 19: Wronskians

We will devote a lecture to the case of linear second-order differential equations (this could have been the title of this Lecture), that is, of the type:

$$y''(x) + P(x)y'(x) + Q(x)y(x) = R(x) \quad (222)$$

where  $P$  and  $Q$  are, in principle, arbitrary. We do so for two reasons: one is that such equations are important in Physics and if there is one type of ODE that one needs to be familiar with, this is this one. The second is that they offer a beautiful illustration of the might of linear algebra, of its basic concepts such as abstract vector spaces and technical tools such as matrices and determinants. We have already discussed these at length and promised to repeatedly come back to them in our formulation of mathematical solutions to many physical models. Well, this is already the case.

When  $R = 0$  in Eq. (222), the equation becomes homogeneous and we start with this case (for a good reason):

$$y''(x) + P(x)y'(x) + Q(x)y(x) = 0. \quad (223)$$

The differential equation is linear in the sense that if  $y_1$  and  $y_2$  are linearly-independent solutions, then any linear superposition  $\alpha y_1 + \beta y_2$  is also a solution and, besides, includes all solutions. Said otherwise, solutions are vectors in a vector space and  $\{y_1, y_2\}$  are a basis of solutions.<sup>149</sup> To behave as such, it will be important to assure ourselves of the linear independence. For the case of order 2, where we need two independent solutions, it is not a big deal, we merely have to check one is not a multiple of the other. In the general case, however, this is not so easily accomplished, and a particular quantity, the *Wronskian*, is introduced for this purpose. We introduce it now to familiarise ourselves with this object. The Wronskian is the determinant of the matrix formed by successive derivatives of the differential equation in the rows and the various solutions in the columns:

$$W_{y_1, y_2}(x) \equiv \begin{vmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{vmatrix}. \quad (224)$$

<sup>149</sup> What is the basis of this space, which vectors are continuous functions of the continuous variable  $x$ ?

It has the nice property that it is either never zero, or is on the opposite identically zero (that is  $W(x) = 0$  for all  $x$ ). This is easy to prove, by computing:<sup>150</sup>

<sup>150</sup> Do it.

$$W' = y_1 y_2'' - y_2 y_1'' \quad (225)$$

and multiplying Eq. (222) for  $y_1$  by  $y_2$  and, vice-versa, multiplying Eq. (222) for  $y_2$  by  $y_1$  and subtracting, we find:

$$(y_1 y_2'' - y_2 y_1'') + P(y_1 y_2' - y_2 y_1') = 0 \quad (226)$$

which is

$$W' + PW = 0 \quad (227)$$

with solution

$$W = ce^{-\int P(x) dx} . \quad (228)$$

Since an exponential is never zero,  $W$  is either never zero or, if  $c = 0$ , always zero. We can now prove the main result, that comes as an implication:

If two differentiable functions are linearly dependent, then their Wronskian is zero.

Proof: by assumption,  $y_2 = cy_1$  (linear dependence), therefore,

$$W = \begin{vmatrix} y_1 y_2 \\ y_1' y_2' \end{vmatrix} = y_1 (cy_1)' - y_2' (cy_1) = 0 \text{ for all } x.$$

The other way around requires that  $y_1$  and  $y_2$  be solutions of Eq. (223) to hold (see Exercises):

If  $y_1$  and  $y_2$  are nontrivial solutions of Eq. (223) and their Wronskian is zero for some value of  $x_0$ , then they are linearly independent.

Note that when we say “for some value of  $x_0$ ” above, this is for convenience if we can find one such value, but if the Wronskian is zero for one  $x_0$ , it is actually identically zero on the whole domain of  $y_1$  and  $y_2$ , as previously demonstrated. Proof: assuming  $y_1$  is not identically zero itself (otherwise the solution is trivial, that’s what we mean by “trivial”), then there exist an interval where it is nowhere zero. This follows from continuity. On this interval, the quantity

$$\frac{y_1 y_2' - y_1' y_2}{y_1^2} \quad (229)$$

is identically zero, since its numerator is  $W_{y_1, y_2}$ . But the term (229) is  $(y_2/y_1)'$  and being zero everywhere, means  $y_2/y_1 = c$  for  $c$  a constant, i.e.,  $y_2 = cy_1$ . This holds on a subinterval only, not on the full interval, but on this interval, such an equality implies that also the derivative will be equal. Now from the unicity of solutions, there is only one solution to Eq. (223) with a given  $y(x_0)$  and  $y'(x_0)$ , so  $y_2 = cy_1$  on the full domain of  $y_1$  and  $y_2$ . QED.

Now let us proceed and try to solve Eq. (223), meaning, let us try to find the counterpart for this second-order case of the integrating factor trick for the first-order one. Unfortunately, the most general case cannot be tackled in a systematic fashion. However there is one strong result that almost brings us here. Namely, if we know *one* of the two solutions required to build all the solutions, we can find the second linearly independent solution. The trick is even more admirable than for the first-order case. We assume that the solution is of the type:

$$y_2(x) = v(x)y_1(x) \quad (230)$$

where we assume basically nothing since  $v$  is completely unknown, but we now see that we can calculate it by putting Eq. (230) into Eq. (223) and compute:

$$y_2' = vy_1' + v'y_1 \quad (231a)$$

$$y_2'' = vy_1'' + 2v'y_1' + v''y_1 \quad (231b)$$

which yields Eq. (223), after collecting factors of the derivatives of  $v$ , in the form:

$$v(y_1'' + Py_1' + Qy_1) + v''y_1 + v'(2y_1' + Py_1) = 0. \quad (232)$$

Note that the coefficient of  $v$  is Eq. (223) for  $y_1$ , but  $y_1$  is a solution, therefore, we really have  $v''y_1 + v'(2y_1' + Py_1) = 0$  or, bringing all  $v$  terms on one side:

$$\frac{v''}{v'} = -2\frac{y_1'}{y_1} - P \quad (233)$$

which we can integrate on both sides:

$$\ln(v') = -2\ln(y_1) - \int P(x) dx \quad (234)$$

or, by taking the exponential of both sides and integrating again:

$$v' = \frac{1}{y_1^2} e^{-\int P(x) dx} \quad (235a)$$

$$v = \int \frac{1}{y_1^2} e^{-\int P(x) dx} d\chi. \quad (235b)$$

We had to introduce a new dummy variable as we are now integrating all over the place. A more careful and accurate way to write it by introducing dummy variables  $\chi$  and  $v$  (chi and upsilon) to keep  $x$  for the solution, reads:

$$v(x) = \int_{\alpha}^x \frac{1}{y_1^2(\chi)} e^{-\int_{\beta}^{\chi} P(v) dv} d\chi. \quad (236)$$

The lower bounds  $\alpha$  and  $\beta$  play no role. This is not quite as systematic as the first-order's integrating factor since we need to know  $y_1$ .

Here is an example of this method:

$$x^2 y'' + xy' - y = 0. \quad (237)$$

We can find, or guess, or be given one solution, namely,<sup>151</sup>  $y_1 = x$ .

<sup>151</sup> Check that it is a solution.

Therefore, the second solution is found from Eq. (236) as  $y_2 = v y_1$  with:

$$v = \int \frac{1}{x^2} e^{-\int dx/x} dx = \int \frac{1}{x^2} e^{-\ln x} dx = \int x^{-3} dx = \frac{x^{-2}}{-2} \quad (238)$$

so  $y_2 = -1/(2x)$  and the general solution is found as:

$$y = c_1 x + c_2/x \quad (239)$$

(the constant  $c_2$  absorbs the  $-1/2$  factor), as you can check.<sup>152</sup>

<sup>152</sup> Do it.

We now turn to a very important case, which can always be solved exactly, and that we will encounter over and over again (and which we have, in fact, already encountered several times), namely, the case of constant coefficients, that is, when  $P(x) = p$  and  $Q(x) = q$  are constants:

$$y'' + py' + qy = 0. \quad (240)$$

We need two linearly-independent solutions and these follow from  $y = e^{mx}$  since exponentials remain but pick-up a coefficient upon derivation, namely, substituting  $e^{mx}$  in Eq. (240), we find:

$$m^2 e^{mx} + p m e^{mx} + q e^{mx} = 0 \quad (241)$$

or, by dividing both sides by  $e^{mx}$  (an exponential is never zero):

$$m^2 + pm + q = 0 \quad (242)$$

which is a quadratic equation, with  $m$  the variable and  $p, q$  the coefficients; namely, we have transformed the original differential equation (240) [for functions] into a quadratic equation (242) [for scalars]. Eq. (242) is called the "characteristic polynomial". Solutions of Eq. (242) are:

$$m_{1,2} = \frac{-p \pm \sqrt{p^2 - 4q}}{2}. \quad (243)$$

From these solutions, we have three possible scenarios:

1. Discriminant positive  $p^2 - 4q > 0$ . In this case, we have two real-valued and non-degenerate roots, so the linearly independent solutions  $e^{m_{1,2}x}$  yield the general solution

$$y(x) = e^{-px/2} \left( \alpha e^{\sqrt{p^2 - 4q}x/2} + \beta e^{-\sqrt{p^2 - 4q}x/2} \right). \quad (244)$$

2. Discriminant negative  $p^2 - 4q < 0$ . In this case, we have two complex-valued and non-degenerate roots, so the linearly independent solutions  $e^{m_{1,2}x}$  yield the general solution

$$y(x) = e^{-px/2} \left( \alpha e^{i\sqrt{p^2-4q}x/2} + \beta e^{-i\sqrt{p^2-4q}x/2} \right). \quad (245)$$

In this form, the solutions are explicitly complex. If we want to keep their real-valued character, we can use the linear superposition that gives  $e^{i\theta} + e^{-i\theta} = 2 \cos(\theta)$  and  $e^{i\theta} - e^{-i\theta} = 2i \sin(\theta)$ , so that the solutions read:

$$\boxed{y(x) = e^{-px/2} \left( A \cos \left( \sqrt{p^2 - 4q}x/2 \right) + B \sin \left( \sqrt{p^2 - 4q}x/2 \right) \right)} \quad (246)$$

with  $A = (\alpha + \beta)/2$  and  $B = i(\alpha - \beta)/2$ .

3. Discriminant zero  $p^2 - 4q = 0$ . In this case, there is only one solution  $-p/2$ , so one solution  $y_1 = e^{-px/2}$  only is available. The second-one, however, is straightforwardly obtained from Eq. (236), namely:  $y_2 = e^{-px/2} \int [e^{-\int p dx} / (e^{-px/2})^2] dx = xe^{-px/2}$  so the general solution reads:

$$\boxed{y(x) = e^{-px/2}(\alpha + \beta x)}. \quad (247)$$

We can now turn to the *inhomogeneous* case, that features a constant term:

$$y'' + P(x)y' + Q(x)y = R(x). \quad (248)$$

Note that the constant is as far as the “unknown variable”  $y(x)$  is concerned.  $R(x)$  is, in general, a function (of the function-variable  $x$ ). We will see, however, that the theory of the homogeneous case provides the basis for solving Eq. (248). In fact, let us start with a simple but far-reaching observation, if we have two solutions  $Y_1$  and  $Y_2$  of Eq. (248), then  $Y_1 - Y_2$  is a solution of the homogenous version (223). Proof: We compute the lhs of the differential equation for  $Y_1 - Y_2$ :

$$\begin{aligned} (Y_1 - Y_2)'' + P(x)(Y_1 - Y_2)' + Q(x)(Y_1 - Y_2) = \\ (Y_1'' + P(x)Y_1' + Q(x)Y_1) - \{Y_2'' + P(x)Y_2' + Q(x)Y_2\} \end{aligned} \quad (249)$$

by linearity, but, since  $Y_1$  and  $Y_2$  are both solutions of Eq. (248), this is  $R(x) - R(x) = 0$ , therefore  $Y_1 - Y_2$  is a solution of Eq. (223). Since the general solution of the homogeneous equation is  $\alpha y_1 + \beta y_2$  for  $\alpha, \beta$  two constants (and  $y_1, y_2$  linearly independent), the most general solution of the nonhomogeneous equation reads:

$$\boxed{Y = Y_R + \alpha y_1 + \beta y_2} \quad (250)$$

where  $Y_R$  stands for “a particular solution of the non-homogeneous equation”. We prefer  $R$  as a subscript to identify this solution, since  $R(x)$  is precisely the non-homogeneous term, to  $P$  or  $p$  which are the typical names we give the to 1st-derivative coefficients. Therefore, the program for solving Eq. (248) is simple:

- Solve the homogeneous equation (find  $y_1$  and  $y_2$ ).
- Find one particular solution of the non-homogenous equation.

The importance of the homogeneous case will rise from the fact that a particular solution can be constructed from the homogeneous solutions!

Sometimes, however, we can simply guess by tinkering, particularly when coefficients are constant and when the constant term is a simple function (as is often the case in Physics, where this term comes from a driving). For instance, the equation

$$y'' + py' + qy = e^{ax} \quad (251)$$

for which we already know the homogeneous solutions are exponential functions, it makes sense to guess that a particular solution is of the type

$$Ae^{ax}, \quad (252)$$

which we can check by substituting into Eq. (251), to find  $Aa^2 + pAa + qA = 1$ , therefore

$$y_R = \frac{e^{ax}}{a^2 + pa + q} \quad (253)$$

is the sought particular solution, provided that  $a^2 + pa + q \neq 0$ .

If this is zero, then  $a$  is a solution of the characteristic polynomial, meaning that  $e^{ax}$  is actually one of the homogeneous  $y_1, y_2$  solutions of Eq. (223), so is not working as a particular  $Y_R$  solution of Eq. (248). Taking a clue from the previous degenerate case (zero discriminant) where the linearly independent solution was  $xe^{px/2}$ , we then try in this case:

$$y = Axe^{ax} \quad (254a)$$

$$y' = Ae^{ax}(1 + ax) \quad (254b)$$

$$y'' = Aae^{ax}(2 + ax) \quad (254c)$$

which we now substitute into Eq. (251) to find:

$$e^{ax}\{A[(a^2 + pa + q)x + 2a + p]\} = e^{ax} \quad (255)$$

and since  $a^2 + pa + q = 0$  (this was the problem initially, now it's a resource) we find that  $A = 1/(2a + p)$  gives the solution, which reads



$$y_R = \frac{xe^{ax}}{2a + p}, \quad (256)$$

provided, of course, that  $a \neq -p/2$ . If this is the case, the inhomogeneous happens to be the degenerate solution so we hit both  $y_1$  and  $y_2$ , thus we try instead:

$$y = Ax^2e^{ax} \quad (257a)$$

$$y' = Axe^{ax}(2 + ax) \quad (257b)$$

$$y'' = Ae^{ax}(2 + 4ax + a^2x^2) \quad (257c)$$

which we now substitute into Eq. (251) to find:

$$e^{ax}\{A[x^2(a^2 + ap + q) + 2x(2a + p) + 2]\} = e^{ax} \quad (258)$$

which, from the repeated cancellations observed previously, reduces to  $e^{ax}A2$  so that  $A = 1/2$  makes

$$y_R = \frac{e^{ax}}{2} \quad (259)$$

the particular solution with, this time, no risk of division by zero! (there is only two homogeneous solutions we could have met).

This method of guesswork goes by the name of “method of undetermined coefficients”. It is useful for particular types of constant terms, which happen to be those of most common encounters, so worth learning about. For a trigonometric function,  $R(x) = \sin(ax)$  or  $\cos(ax)$ , the trial function becomes:

$$y_R = A \sin(ax) + B \cos(ax) \quad (260)$$

multiplied by  $x$  if hitting a homogeneous solution. It is also useful if  $R(x) = a_0 + a_1x + \dots + a_nx^n$  is a polynomial, in which case

$$y_R = A_0 + A_1x + \dots + A_nx^n \quad (261)$$

is the trial wavefunction with undetermined coefficients  $A_i$ .

This is all clever and good, but a bit tied to the particular cases that work for  $R(x)$ , let alone that practical cases assume constant coefficients. In any case, if this looks simple enough, one can aim for this approach: intelligent guesswork.

We now turn to a more systematic way to construct the particular solution of the inhomogeneous equation from the homogeneous solutions. The method goes by the inspired name of “variation of the constants”.

This method assumes the general solution

$$\alpha y_1(x) + \beta y_2(x) \quad (262)$$

is known (which is always the case for constant coefficients). We now write the  $x$  dependence of  $y$  explicitly, as  $\alpha$  and  $\beta$  are constant, or, more precisely, were constants, since we now upgrade them to functions!

$$\boxed{y_R = \alpha(x)y_1(x) + \beta(x)y_2(x)} \quad (263)$$

(in essence, “varying the constants”). If Eq. (263) is to be a solution of Eq. (248), it must fit when substituted there, which we now check (or demand). We compute:

$$y'_R = \alpha'y_1 + \alpha y'_1 + \beta'y_2 + \beta y'_2. \quad (264)$$

Admittedly this could be confusing to someone entering the class in mid-lecture. We would now proceed to compute  $y''$  but it so happens that we can bring our seemingly heretic promotion of constants to variables to provide useful results if we further impose that

$$\alpha'y_1 + \beta'y_2 = 0 \quad (265)$$

which we will later come back to in order to solve for  $\alpha$  and  $\beta$ . For now, that implies that we have:

$$y_R = \alpha y_1 + \beta y_2 \quad (266a)$$

$$y'_R = \alpha y'_1 + \beta y'_2 \quad (266b)$$

$$y''_R = \alpha' y'_1 + \beta' y'_2 + \alpha y''_1 + \beta y''_2 \quad (266c)$$

which, once substituted in Eq. (248), yields, by collecting terms of  $\alpha$ ,  $\beta$  and their derivatives:

$$\alpha(y''_1 + P y'_1 + Q y_1) + \alpha' y'_1 + \beta(y''_2 + P y'_2 + Q y_2) + \beta' y'_2 = R(x). \quad (267)$$

Now, the  $\alpha$  and  $\beta$  terms cancel from the fact that  $y_1$  and  $y_2$  are solutions to the homogeneous equations, and we are left with:

$$\alpha' y'_1 + \beta' y'_2 = R(x). \quad (268)$$

Together with Eq. (265), this yields the equation for  $\alpha'$  and  $\beta'$ :

$$\begin{pmatrix} y_1 & y_2 \\ y'_1 & y'_2 \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \begin{pmatrix} 0 \\ R(x) \end{pmatrix} \quad (269)$$

The matrix in Eq. (269) is invertible (and we can find a unique solution) if its determinant is nonzero. But this determinant is by now a familiar object, it is the Wronskian, which we already know is nowhere zero since  $y_1$  and  $y_2$  are linearly independent solutions. Therefore, the derivatives are found as

$$\alpha'(x) = \frac{-y_2 R(x)}{W(y_1, y_2)} \quad \text{and} \quad \beta'(x) = \frac{y_1 R(x)}{W(y_1, y_2)}. \quad (270)$$

and from their integration, we arrive to the main result of the variation of the constants:

$$\boxed{\alpha(x) = \int \frac{-y_2(x)R(x)}{W(y_1, y_2)} dx \quad \text{and} \quad \beta(x) = \int \frac{y_1(x)R(x)}{W(y_1, y_2)} dx.} \quad (271)$$

The particular solution is then given by Eq. (263). This does not guarantee the solution since, as we know, primitives are not easy to find in the general case. We got it at least formally, and to this we then only need adding a linear combination of the homogeneous solutions. Here is an example: let us solve

$$y'' - 2y' + y = \frac{e^x}{x^2 + 1}. \quad (272)$$

The general solution to the homogeneous equation is found as:

$$(\alpha + \beta x)e^x \quad (273)$$

so we have  $y_1 = e^x$  and  $y_2 = xe^x$  with Wronskian

$$W_{y_1, y_2} = \begin{vmatrix} e^x & xe^x \\ e^x & e^x(1+x) \end{vmatrix} = e^{2x}. \quad (274)$$

Equations (263) and (271) then give:

$$y_R(x) = e^x \int \frac{-xe^x}{e^{2x}} \frac{e^x}{x^2 + 1} dx + xe^x \int \frac{e^x}{e^{2x}} \frac{e^x}{x^2 + 1} dx \quad (275a)$$

$$= -e^x \int \frac{x}{x^2 + 1} dx + xe^x \int \frac{1}{x^2 + 1} dx \quad (275b)$$

$$= -\frac{e^x}{2} \ln(1 + x^2) + xe^x \arctan(x) \quad (275c)$$

so the general solution is:

$$y(x) = e^x \left[ \left( \alpha - \frac{\ln(1 + x^2)}{2} \right) + x(\beta + \arctan(x)) \right] \quad (276)$$

for  $\alpha, \beta \in \mathbb{R}$ .

We have tackled second-order fairly extensively. The most general type of linear differential equation is:

$$a_n(x)y^{(n)} + \cdots + a_2(x)y'' + a_1(x)y' + a_0(x)y = b(x) \quad (277)$$

where  $a_i(x)$  are differentiable functions. The case of general  $a_k(x)$  is, of course, complicated, but the case of constant coefficients follows similar results as the second-order case considered here.

## Exercises

*Practice, practice, practice*

1.  $y'' + y' - 6y = 0.$
2.  $y'' + 2y' + y = 0.$
3.  $y'' + 8y = 0.$
4.  $2y'' - 4y' + 8y = 0.$
5.  $y'' - 4y' + 4y = 0.$
6.  $y'' - 9y' + 20y = 0.$
7.  $y'' + y' = 0.$
8.  $y'' + 2y' + 3y = 0.$
9.  $y'' = 4y.$
10.  $4y'' - 8y' + 7y = 0.$
11.  $16y'' - 8y' + y = 0.$
12.  $y'' + 4y' + 5y = 0.$
13.  $y'' + 4y' - 5y = 0.$
14.  $y'' + y' + y = 0.$

### Initial value problems

Solve the following:

1.  $y'' - 5y' + 6y = 0$  with  $y(1) = e^2$  and  $y'(1) = 3e^2.$
2.  $y'' - 6y' + 5y = 0$  with  $y(0) = 3$  and  $y'(0) = 11.$
3.  $y'' - 6y' + 9y = 0$  with  $y(0) = 0$  and  $y'(0) = 5.$

### Big Wronskian

Solve the following fourth-order equation:

$$y^{(4)} - 5y^{(2)} + 4y = 0 \quad (278)$$

and show that the solutions you find from the characteristic polynomial with trial solutions  $e^{px}$ , are linearly independent by computing the Wronskian:

$$W = \begin{vmatrix} y_1 & y_2 & y_3 & y_4 \\ y_1' & y_2' & y_3' & y_4' \\ y_1'' & y_2'' & y_3'' & y_4'' \\ y_1''' & y_2''' & y_3''' & y_4''' \end{vmatrix}. \quad (279)$$

### Cauchy–Euler equation

The equation

$$x^2y'' + pxy' + qy = 0 \quad (280)$$

with  $p, q$  constants is known as the Cauchy-Euler equation. Use the change-of-variable  $x = e^z$  to transform this into a tractable equation and solve the following cases:

1.  $x^2y'' + 3xy' + 10y = 0.$
2.  $2x^2y'' + 10xy' + 8y = 0.$
3.  $4x^2y'' = 3y.$

*A general case*

Find the general solution of:

$$y'' - xf(x)y' + f(x)y = 0. \quad (281)$$

Does this help you in solving particular cases?



*A particular case (Legendre of the 2nd kind)*

Legendre's differential equation reads:

$$\frac{d}{dx} \left( (1 - x^2) \frac{dy}{dx} \right) + l(l + 1)y = 0 \quad (282)$$

for some constant  $l$ . When  $l \in \mathbb{N}$ , check that the Legendre polynomials  $P_l$  are solutions. They are called "solutions of the first kind". The other solution needed to form the basis is called "of the second kind". Find it for  $l = 1$  and write the most general solution for this case. How would you go about to solve other  $l \in \mathbb{N}$  cases?



*A particular case (Bessel of the 2nd kind)*

Bessel functions are important functions in Physics that are solutions of the differential equation:

$$x^2 y'' + xy' + (x^2 - \alpha^2)y = 0. \quad (283)$$

When  $\alpha$  is an integer or half-integer, the functions are known as "cylinder functions" or "cylindrical harmonics" and play a major role for oscillating systems which have the polar symmetry. Consider the case  $\alpha = 1/2$ . Verify that  $y_1 = \sin(x)/\sqrt{x}$  is a solution on  $\mathbb{R}^+$ . This is the first-kind. Find the second-kind solution and, therefore, the general solution (complex superpositions  $y_1 \pm iy_2$  are known as "Hankel functions").



*Beware of the Wronskian*

The Wronskian can be identically zero and yet the functions not be linearly independent. Check this with the example of  $y_1 = x^3$  and  $y_2 = x^2|x|$ . Namely, show that:

1. The Wronskian is identically zero.
2. The functions are linearly independent.

Show that they are not both solutions of  $y'' + P(x)y + Q(x) = 0$ , so their harm for the validity of the Wronskian to differential equations is limited. However it illustrates clearly the importance of distinguishing between if and iff conditions. Namely  $f, g$  linearly dependent  $\implies W_{f,g} = 0$  but the other way around is not true (unless if they are solutions of an ODE).

### *Undertermined coefficients*

Solve the following:

1.  $y'' + 3y' - 10y = 6e^{4x}$ .
2.  $y'' + 10y' + 25y = 14e^{-5x}$ .
3.  $y'' + 4y = 3 \sin x$ .
4.  $y'' + y = 2 \cos x$ .
5.  $y'' - 3y' + 2y = 14 \sin(2x) - 18 \cos(2x)$ .
6.  $y'' + y' = 10x^4 + 2$ .

In each case, check your solution!

### *Superposition of solutions*

Linearity allows for some powerful results. Show that if  $y_1$  is a solution of

$$y'' + P(x)y' + Q(x)y = R_1(x) \quad (284)$$

and  $y_2$  is a solution of

$$y'' + P(x)y' + Q(x)y = R_2(x) \quad (285)$$

then  $y_1 + y_2$  is a solution of

$$y'' + P(x)y' + Q(x)y = R_1(x) + R_2(x). \quad (286)$$

Use this superposition principle to solve:

$$y'' + 4y = 4 \cos(2x) + 6 \cos x + 8x^2 - 4x. \quad (287)$$

### *Variation of constants*

Solve the following:

1.  $y'' + 4y' + 4y = \cosh(x)$  (solution is in Wikipedia's "variation of parameters").
2.  $2y'' + 18y = 6 \tan(3x)$ .
3.  $y'' + 2y' + y = \ln x / e^x$ .

### *No great tricks*

Solve the following. They are not exactly in the canonical form but you can surely find your way around easily. Note for instance how the first one is not an obvious match for undetermined coefficients because coefficients are *not* constant.

1.  $xy'' - (1+x)y' + y = x^2$ .

2.  $\sin(x)(y'' + y) = 1$ .

3.  $\tan(2x)(y'' + y) = 1$ .

4.  $\tan^2(x)(y'' + y) = 1$ .

### *No need of constants*

Show that in Eq. (271), one does not need to retain the constants of integrations to form the general solution of the inhomogeneous differential equation.



### *Varying a lot of constants*

Extend the variation of the constants technique that we used for the second-order linear inhomogeneous differential equation to the case of  $n$ th order (linear inhomogeneous differential) equations.



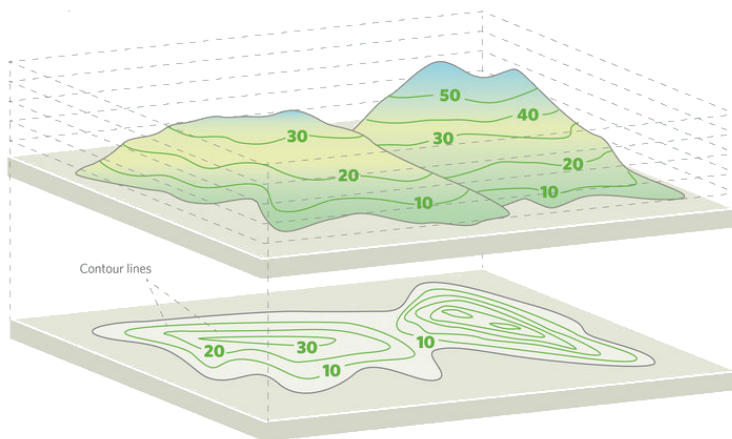


## Lecture 20: Jacobians.

We have already seen how the concept of a function is a general mapping from one set  $\mathbb{A}$  to another  $\mathbb{B}$  and although we have lately focused on functions  $\mathbb{R} \rightarrow \mathbb{R}$  for nonlinear mapping, we have already covered in great details linear mappings  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ . We now study nonlinear functions  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  and we will be interested in how they vary locally, how to approximate them to linear functions, which is by now a familiar problem and thus, generally, how to extend to them the concept of derivatives. Such nonlinear multi-variable functions are ubiquitous in Physics to describe a variety of more or less intuitive concepts. Let us start with a fairly intuitive one: consider the altitude  $h$  of a 2D landscape. It is a function:

$$h : \mathbb{R}^2 \rightarrow \mathbb{R} \quad (288)$$

which associate to each position on a 2D map the corresponding height, as experienced for instance by a hiker who walked there. Here is an example of a possible representation (sketch):



Formally, such a function can usefully be seen as a function of a vector  $\vec{r} = x\hat{i} + y\hat{j}$ , but often we will work directly with the components of the vector, and thus deal with  $h(x, y)$  instead.

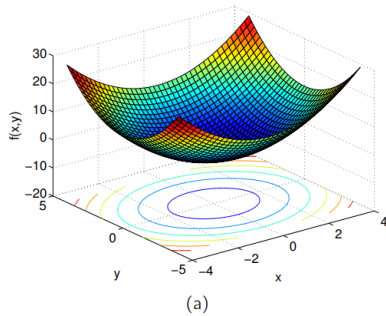
The 3D plot shown above looks familiar enough, it is roughly what we would see with the naked eye by contemplating the scenery. Note

that a  $\mathbb{R}^2 \rightarrow \mathbb{R}$  function becomes a 3D plot, which it is still possible to represent, but we can naturally expect complications for the more general case. Even the 3D plot is fairly clumsy, and partial (we have to choose an angle of visualisation). Various projections are useful, in particular, *isolines*, which are solutions in 2D of the equation:

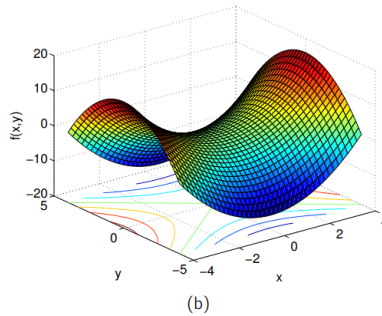
$$h(x, y) = c \quad (289)$$

for various constant values of  $c$  (namely,  $10i$  for  $1 \leq i \leq 5$  for the example above). The previous example was assumed to be taken from real (measured) data. Mathematically, we will mainly work with functions defined from elementary functions. Here are some examples:

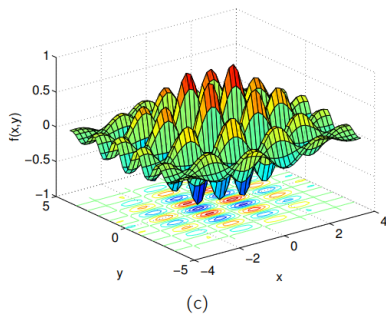
- (a)  $f(x, y) = x^2 + y^3 - 3$ .
- (b)  $(x + y)(x - y)$ .
- (c)  $\exp(-(x^2 + y^2)/10) \sin(2x) \cos(4y)$ .
- (d)  $\sqrt{|x - y|}$ .



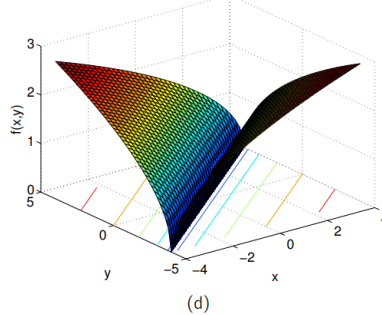
(a)



(b)



(c)



(d)

One can get familiar with the complexity of such objects by reducing them to familiar ones, namely, functions of a single variable. This is achieved by looking at 2D cuts through the 3D structure, which allow to regard them temporarily through the well-known eye of single-variable functions. The rate of change in these cuts are simply (and exactly) those already known, but since this takes place

in a bigger setting, we need more powerful (and general) notations to keep track, as well as a dedicated vocabulary. We speak of *partial derivatives* when we consider the (usual) derivative for one variable:

$$\frac{\partial f}{\partial x}(x, y) \equiv \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon, y) - f(x, y)}{\varepsilon}, \quad (290a)$$

$$\frac{\partial f}{\partial y}(x, y) \equiv \lim_{\varepsilon \rightarrow 0} \frac{f(x, y + \varepsilon) - f(x, y)}{\varepsilon}. \quad (290b)$$

We will often use the notation:

$$\partial_x f \equiv \frac{\partial f}{\partial x} \quad (291)$$

and refer to  $\partial_x$  as an “operator”, in the sense of some object which needs to be applied to a function  $f$  to produce another valid mathematical object (another function, namely, the partial derivative).

Would we work in our discretized space of functions, these operators become, literally, matrices.

Importantly, note that  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$  are themselves also functions of  $x$  and  $y$ , so we can write  $h_x(x, y) \equiv \frac{\partial f}{\partial x}(x, y)$  and  $h_y(x, y) \equiv \frac{\partial f}{\partial y}(x, y)$ , and ask what are the partial derivatives of the newly defined 2D-function  $h_x$  and  $h_y$ . This brings us to the concept of *higher-order partial derivatives*, e.g., for  $h_x$ :

$$\frac{\partial}{\partial x} \frac{\partial f}{\partial x}, \quad (292a)$$

$$\frac{\partial}{\partial y} \frac{\partial f}{\partial x}. \quad (292b)$$

These first one (we also introduce its shorthand notation on the left):

$$\partial_x^2 f \equiv \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \frac{\partial f}{\partial x} \quad (293a)$$

$$\partial_{yx}^2 f \equiv \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \frac{\partial f}{\partial x} \quad (293b)$$

A natural question arises whether  $\partial_{yx}^2 = \partial_{xy}^2$ . Is it the same? This problem is known as that of the “symmetry of the second derivatives”. We let you check in Exercises that this is almost always true. It can be proven in analysis (Schwarz’s theorem, aka Clairaut’s theorem) that if the partial derivatives are defined and continuous, then they are symmetric, i.e., the order does not matter.

Let us discuss the natural question of a possible “total” derivative, not “partial”, i.e., how much does the function  $f(\vec{r})$  changes when we take an epsilon step  $\vec{\varepsilon}$  away? Of course we cannot generalize brutally to  $f'(\vec{r}) \equiv \lim_{\vec{\varepsilon} \rightarrow \vec{0}} [(f(\vec{r} + \vec{\varepsilon}) - f(\vec{r})) / \vec{\varepsilon}]$  because there is no meaning to vector multiplication, and so even less so for division (remember that we only discussed scalar products and vector products and neither

is related to the type of products we have chucked-in here). Let us advance more cautiously and try to see the “total” change in  $f$  as a result of changing a bit of both the variables simultaneously:

$$\delta f = f(x + \delta x, y + \delta y) - f(x, y) \quad (294)$$

This we can rewrite as:

$$\delta f = f(x + \delta x, y + \delta y) - f(x, y + \delta y) + f(x, y + \delta y) - f(x, y) \quad (295a)$$

$$= \frac{f(x + \delta x, y + \delta y) - f(x, y + \delta y)}{\delta x} \delta x + \frac{f(x, y + \delta y) - f(x, y)}{\delta y} \delta y \quad (295b)$$

$$= \frac{\partial f}{\partial x}(x, y + \delta y) \delta x + \frac{\partial f}{\partial y}(x, y) \delta y \quad (295c)$$

The first term can be further processed into  $\frac{\partial f}{\partial x}(x, y + \delta y) = \frac{\partial f}{\partial x}(x, y) + \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(x, y) \delta y$ . Assuming  $\partial_{xy}^2 f$  is finite, by taking the limit  $\delta x \rightarrow dx$  and  $\delta y \rightarrow dy$  (infinitesimals), we arrive to:

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy. \quad (296)$$

Here we can introduce a useful (and important) notation: *the gradient*  $\nabla$  (upside down triangle or Delta  $\Delta$ ) defined as the *vector of partial derivatives*:

$$\nabla f \equiv \begin{pmatrix} \partial_x f \\ \partial_y f \end{pmatrix}. \quad (297)$$

Note that the gradient transforms a scalar field  $f(x, y)$  into a vector field  $(\nabla f)(x, y)$ . Sometimes, a vector is even put on  $\nabla$  to emphasize its vectorial character:  $\vec{\nabla}$ . We deal a lot with this object in Physics. For now, it is enough to observe that by blind obedience to the laws of algebra, remembering that  $\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y$  (in 2D), then Eq. (296) can be rewritten in the simple form:

$$df = \vec{\nabla} f \cdot d\vec{r}. \quad (298)$$

Banach once said that “A mathematician is a person who can find analogies between theorems; a better mathematician is one who can see analogies between proofs and the best mathematician can notice analogies between theories. One can imagine that the ultimate mathematician is one who can see analogies between analogies.” At this stage we have a acquired material to start trying to perceive analogies between different objects, for instance, that operators, like gradients, can themselves be seen as vectors, in an abstract vector space of operators. Let us see as an illustration of such generalization, although we will not comment it much now, let alone prove it, the Taylor expansion for

multi-variable functions by a small  $n$ D displacement  $\vec{\epsilon}$  around the point  $\vec{r}$ :

$$f(\vec{r} + \vec{\epsilon}) = f(\vec{r}) + (\vec{\epsilon} \cdot \nabla)f(\vec{r}) + \frac{(\vec{\epsilon} \cdot \nabla)^2}{2}f(\vec{r}) + \dots + \frac{(\vec{\epsilon} \cdot \nabla)^n}{n!}f(\vec{r}) + \dots \tag{299}$$

where the power of operators is to be understood as repeated applications, i.e.,  $(\vec{\epsilon} \cdot \nabla)^n = (\vec{\epsilon} \cdot \nabla) \dots (\vec{\epsilon} \cdot \nabla)$ , so that, e.g., in 2D and for  $n = 2$ :<sup>153</sup>

$$(\vec{\epsilon} \cdot \nabla)^2 = (\epsilon_x \partial_x + \epsilon_y \partial_y)(\epsilon_x \partial_x + \epsilon_y \partial_y) = \epsilon_x^2 \partial_x^2 + 2\epsilon_x \epsilon_y \partial_x \partial_y + \epsilon_y^2 \partial_y^2. \tag{300}$$

Equation (299) should look both natural, elegant and, of course, understandable.

The first-order Taylor expansion in the case of a single-variable function was the derivative. In this case, the link is not so straightforward and this is because, in fact, that is not the most direct connection, which is brought by another object known as the “Jacobian”, which we obtain by rewriting Eq. (295) as:<sup>154</sup>

$$f(\vec{r} + \vec{\epsilon}) \approx f(\vec{r}) + \mathbf{J}_f(\vec{r}) \cdot \vec{\epsilon} \tag{301}$$

which becomes exact in the limit  $\vec{\epsilon} \rightarrow \vec{0}$ . It generalizes the derivative of a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , which was itself a  $\mathbb{R} \rightarrow \mathbb{R}$  function, to a multivariate function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , nonlinear in principle, into a  $n \times m$  matrix, so a linear application in  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ , with components:<sup>155</sup>

$$\mathbf{J}_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \dots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{302}$$

Because the Jacobian is linear, the derivative is, after all, a different object than the function it describes! Just we don’t get to see this difference with one variable. The Jacobian (being the derivative) is of course a central object in multivariable calculus. Note also that for functions of a vector but that are scalars themselves, i.e.,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , the transpose of the Jacobian is the “gradient”. It shows, by the way, one should not call  $f'(x)$  a gradient... but the “slope” of the  $\mathbb{R} \rightarrow \mathbb{R}$  function, since at such a level of pedantry, otherwise, one should call  $f'(x)$  the Jacobian!

In particular, when  $n = m$ , the determinant of the Jacobian (which is also called, confusingly, the Jacobian) at a point  $\mathbf{r}$  is the factor by which  $f$  locally dilates hypervolumes around  $\mathbf{r}$  the point (since  $f$  acts locally like a linear function, the Jacobian matrix!). For this reason, the Jacobian also appears in the change of variables formula for multivariate integrals (which we will not cover).

<sup>153</sup> Check it.

<sup>154</sup> This can also be found by integrating Eq. (296) between  $\vec{r}$  and  $\text{vecr} + \vec{\epsilon}$  with  $|\vec{\epsilon}|$  small enough so that the partial derivatives can be assumed constant. Show that.

<sup>155</sup> We have built from the  $\mathbb{R}^2 \rightarrow \mathbb{R}$  case; show that the general case given here follows straightforwardly.

We now turn to the problem of composition of functions of several variables, and how derivatives work out in this way. This is known as the “chain rule” (in fact some people also call  $(f \circ g)' = (f' \circ g)g'$  a chain rule, as this is a particular case of the general one which we see now). It is a central concern of Physics where we parameterise things. Let us come back to fix ideas to the hiking problem  $h(x, y)$ , meaning that at point  $(x, y)$ , the elevation is  $h$ . Now assume our little bug going through this plane following the trajectory  $\mathcal{T}(t)$  as a function of time  $t$ , i.e.,  $\mathcal{T}(t) = (x(t), y(t))$  tells us where it is at this time. Then  $h \circ \mathcal{T}$  is a single-valued function of time that tells us the altitude of the bug at any given time. How does this function varies? This depends on the trajectory itself, how fast the bug is moving, but also of the landscape of where it's moving: if it goes fast in slowly-changing areas or, on the opposite, slow in the fast-changing ones, it will in both cases experience a steep change of altitude. Note that we would like to write  $h(t)$  for such a quantity, and we will often rely on such sloppy, but intuitively clear (and desirable) notation.

We have already looked in details at composition of *linear* functions. We now do the same but with arbitrary functions (not compulsorily linear) and will assume the case of  $\mathbb{R}^k$  spaces to fix ideas:

$$\mathbb{R}^n \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^l \quad (303)$$

The general relationship is given by the chain rule, which states that the Jacobian of a composite function is given by the product of the Jacobians of the functions:

$$\mathbf{J}_{f \circ g} = \mathbf{J}_f \mathbf{J}_g \quad (304)$$

where, by definitions:

$$g(\mathbf{x} + \boldsymbol{\epsilon}) = g(\mathbf{x}) + \mathbf{J}_g(\mathbf{x}) \cdot \boldsymbol{\epsilon} \quad (305a)$$

$$f(\mathbf{y} + \boldsymbol{\epsilon}) = f(\mathbf{y}) + \mathbf{J}_f(\mathbf{y}) \cdot \boldsymbol{\epsilon} \quad (305b)$$

$$(f \circ g)(\mathbf{x} + \boldsymbol{\epsilon}) = (f \circ g)(\mathbf{x}) + \mathbf{J}_{f \circ g}(\mathbf{x}) \cdot \boldsymbol{\epsilon} \quad (305c)$$

We sketch the proof, which is similar to the derivative for composition of functions. On the one hand:

$$(f \circ g)(\mathbf{x} + \boldsymbol{\epsilon}) = (f \circ g)(\mathbf{x}) + \mathbf{J}_{f \circ g}(\mathbf{x}) \cdot \boldsymbol{\epsilon} \quad (306)$$

while on the other hand:

$$(f \circ g)(\mathbf{x} + \boldsymbol{\epsilon}) = f(g(\mathbf{x} + \boldsymbol{\epsilon})) = f(g(\mathbf{x}) + \mathbf{J}_g(\mathbf{x}) \cdot \boldsymbol{\epsilon}) \quad (307a)$$

$$= f(g(\mathbf{x})) + \mathbf{J}_f(\mathbf{J}_g(\mathbf{x})) \cdot \boldsymbol{\epsilon} \quad (307b)$$

but since  $\mathbf{J}_f$  and  $\mathbf{J}_g$  are linear functions, then  $\mathbf{J}_f(\mathbf{J}_g(\mathbf{x})) = (\mathbf{J}_f \mathbf{J}_g)(\mathbf{x})$ , which, by comparison to the previous expression (306), leads us to the final result (304)

Let us look at the dimensions:

$$\mathbb{R}^n \xrightarrow[\substack{\mathbf{J}_g \\ (m \times n)}]{g} \mathbb{R}^m \xrightarrow[\substack{\mathbf{J}_f \\ (l \times m)}]{f} \mathbb{R}^l \quad (308a)$$

$$\mathbb{R}^n \xrightarrow[\substack{\mathbf{J}_{f \circ g} = \mathbf{J}_f \mathbf{J}_g \\ (l \times n)}]{f \circ g} \mathbb{R}^l \quad (308b)$$

so Eq. (304) works well since concatenation of the matrix product gives us  $(l \times m) \times (m \times n) = (l \times n)$  as it should.

The full-version of Eq. (304) is:

$$\mathbf{J}_{f \circ g}(\mathbf{x}) = \mathbf{J}_f(g(\mathbf{x}))\mathbf{J}_g(\mathbf{x}) \quad (309)$$

but we tend to “simplify” away or drop the intermediate variables.

The simpler is to illustrate with cases of interest, starting with the one that opened our discussion, formulated with the bug in the hiking landscape:

$$\mathbb{R} \xrightarrow[\substack{\mathbf{J}_T \\ (2 \times 1)}]{T} \mathbb{R}^2 \xrightarrow[\substack{h \\ (1 \times 2)}]{} \mathbb{R} \quad (310a)$$

$$t \mapsto \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \mapsto h(x, y) \quad (310b)$$

$$\mathbb{R} \xrightarrow[\substack{\mathbf{J}_{h \circ T} \\ (1 \times 1)}]{h \circ T} \mathbb{R} \quad (310c)$$

$$t \mapsto h(t) \quad (310d)$$

with

$$\mathbf{J}_T = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} \quad (311a)$$

$$\mathbf{J}_h = \begin{pmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{pmatrix} \quad (311b)$$

$$\mathbf{J}_{h \circ T} = \begin{pmatrix} \frac{dh}{dt} \end{pmatrix} \quad (311c)$$

so that, by application of Eq. (304):

$$\frac{dh}{dt} = \frac{\partial h}{\partial x} \frac{dx}{dt} + \frac{\partial h}{\partial y} \frac{dy}{dt}. \quad (312)$$

Note that we could cancel  $dt$  since this is a full differential and use:

$$dh = \frac{\partial h}{\partial x} dx + \frac{\partial h}{\partial y} dy. \quad (313)$$

There are many possible configurations, and before we turn to the general one, let us consider a small variation of the one we just

addressed:

$$\mathbb{R}^2 \xrightarrow[\substack{\mathbf{J}_g \\ (1 \times 2)}]{g} \mathbb{R} \xrightarrow[\substack{\mathbf{J}_f \\ (1 \times 1)}]{f} \mathbb{R} \quad (314a)$$

$$(x, y) \mapsto z \mapsto f(z) \quad (314b)$$

$$\mathbb{R} \xrightarrow[\substack{\mathbf{J}_{f \circ g} \\ (1 \times 2)}]{f \circ g} \mathbb{R} \quad (314c)$$

$$(x, y) \mapsto f(x, y) \quad (314d)$$

The chain rule (in matrix form) reads:

$$\mathbf{J}_{f \circ g} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \mathbf{J}_g \mathbf{J}_f = \begin{pmatrix} \frac{df}{dz} \end{pmatrix} \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \quad (315)$$

or, equivalently:<sup>156</sup>

$$\frac{\partial f}{\partial x} = \frac{df}{dz} \frac{\partial z}{\partial x}, \quad (318a)$$

$$\frac{\partial f}{\partial y} = \frac{df}{dz} \frac{\partial z}{\partial y}. \quad (318b)$$

This is easily generalized to an arbitrary number of variables: for a function  $f$  of  $n$  variable  $x_i(t)$  each a function of a parameter  $t$ , we find:<sup>157</sup>

$$\frac{df}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt}. \quad (319)$$

Particular cases of this are also easily obtained, e.g., if we have a function  $f(x, y(x))$ , then:

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx}. \quad (320)$$

We are now more than ready for the truly general case (308):

$$\mathbb{R}^n \xrightarrow[\substack{\mathbf{J}_g \\ (m \times n)}]{g} \mathbb{R}^m \xrightarrow[\substack{\mathbf{J}_f \\ (l \times m)}]{f} \mathbb{R}^l \quad (321a)$$

$$(t_1, \dots, t_n) \mapsto (x_1(t_1, \dots, t_n), \dots, x_m(t_1, \dots, t_n)) \mapsto (f_1(x_1, \dots, x_m), \dots, f_l(x_1, \dots, x_m)) \quad (321b)$$

$$\mathbb{R}^n \xrightarrow[\substack{\mathbf{J}_{f \circ g} \\ (l \times n)}]{f \circ g} \mathbb{R}^l \quad (321c)$$

$$(t_1, \dots, t_n) \mapsto (f_1(t_1, \dots, t_n), \dots, f_l(t_1, \dots, t_n)) \quad (321d)$$

Do you imagine dealing with such a beast without matrices? We have

<sup>156</sup> Work out the case

$$f(u) = \sin(u) \quad (316a)$$

$$u(x, y) = xy \quad (316b)$$

In this case, the function  $f(u)$  can be seen as a function of  $x, y$ :

$$f(u(x, y)) = \sin(xy). \quad (317)$$

Since  $f$  is single-variable, we can compute  $f' = df/du$ . What is the role between of the partial derivatives involved undercover?

<sup>157</sup> Prove it using Jacobians.



(chain rule):

$$\mathbf{J}_{f \circ g} = \begin{pmatrix} \frac{\partial f_1}{\partial t_1} & \frac{\partial f_2}{\partial t_1} & \cdots & \frac{\partial f_l}{\partial t_1} \\ \frac{\partial f_1}{\partial t_2} & \frac{\partial f_2}{\partial t_2} & \cdots & \frac{\partial f_l}{\partial t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial t_n} & \frac{\partial f_2}{\partial t_n} & \cdots & \frac{\partial f_l}{\partial t_n} \end{pmatrix} = \mathbf{J}_f \mathbf{J}_g = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_l}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_l}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_l}{\partial x_n} \end{pmatrix} \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_2}{\partial t_1} & \cdots & \frac{\partial x_m}{\partial t_1} \\ \frac{\partial x_1}{\partial t_2} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_m}{\partial t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial t_n} & \frac{\partial x_2}{\partial t_n} & \cdots & \frac{\partial x_m}{\partial t_n} \end{pmatrix} \quad (322)$$

which gives, for all  $1 \leq j \leq l$  and  $1 \leq k \leq n$ :

$$\frac{\partial f_j}{\partial t_k} = \sum_{i=1}^n \frac{\partial f_j}{\partial x_i} \frac{\partial x_i}{\partial t_k}. \quad (323)$$

For  $n = m = l = 2$ , this gives:

$$\frac{\partial f_1}{\partial t_1} = \frac{\partial f_1}{\partial x} \frac{\partial x}{\partial t_1} + \frac{\partial f_1}{\partial y} \frac{\partial y}{\partial t_1} \quad (324a)$$

$$\frac{\partial f_1}{\partial t_2} = \frac{\partial f_1}{\partial x} \frac{\partial x}{\partial t_2} + \frac{\partial f_1}{\partial y} \frac{\partial y}{\partial t_2} \quad (324b)$$

$$\frac{\partial f_2}{\partial t_1} = \frac{\partial f_2}{\partial x} \frac{\partial x}{\partial t_1} + \frac{\partial f_2}{\partial y} \frac{\partial y}{\partial t_1} \quad (324c)$$

$$\frac{\partial f_2}{\partial t_2} = \frac{\partial f_2}{\partial x} \frac{\partial x}{\partial t_2} + \frac{\partial f_2}{\partial y} \frac{\partial y}{\partial t_2} \quad (324d)$$

This is similar to Eq. (319) except that in these cases, one expresses a partial derivative instead of a total derivative.

Now you see how the rule for the derivative of a composite function  $(f \circ g)' = (f' \circ g)g'$  is actually a particular case of the chain rule:

$$\frac{df}{dx} = \frac{df}{dg} \frac{dg}{dx} \quad (325)$$

is another way to write:

$$f'(x) = f'(g(x))g'(x). \quad (326)$$

The chain rule is useful to change variables and compute derivatives in the new variables, which is a typical problem of physics. For the linear functions, this change-of-variable business was taking the form of a change of basis. We are still looking after the same idea, but this time not only for linear transformation. A typical change of variable is:

$$x = r \cos \theta \quad (327a)$$

$$y = r \sin \theta \quad (327b)$$

for which we will compute various differentials, but for now we still keep it general:

$$x = x(u, v) \quad (328a)$$

$$y = y(u, v) \quad (328b)$$

(this would be  $u = r$  and  $v = \theta$  for the polar transformation). If we have a relationship between the old  $(x, y)$  and the new  $(u, v)$  variables, such as Eqs. (328), this is all very good, we look at the following particular case (the function  $f$  could take values in  $\mathbb{R}^l$  without loss of generality)

$$\mathbb{R}^2 \xrightarrow[\substack{J_g \\ (2 \times 2)}]{g} \mathbb{R}^2 \xrightarrow[\substack{J_f \\ (1 \times 2)}]{f} \mathbb{R} \quad (329a)$$

$$(u, v) \mapsto (x, y) \mapsto f(x, y) \quad (329b)$$

$$\mathbb{R}^2 \xrightarrow[\substack{J_{f \circ g} \\ (1 \times 2)}]{F \equiv f \circ g} \mathbb{R} \quad (329c)$$

$$(u, v) \mapsto F(u, v) \quad (329d)$$

i.e., we rewrite the function  $f(x, y) = f(x(u, v), y(u, v)) = F(u, v)$ . Note that we use a new (but hopefully related) letter for the functions in the new variables since it has a completely different form in general. To keep the notion that it's still the same function, we can join them through a "go-between" function  $z$ , which is such that  $z(x, y) = f(x, y)$  and  $z(u, v) = F(u, v)$  and allows the nice trick that  $z = f(x, y) = F(u, v)$  without having  $f = F$  which is not true. The original problem of finding the derivatives

$$\frac{\partial F}{\partial u} \quad \text{and} \quad \frac{\partial F}{\partial v} \quad (330)$$

is immediate by using the chain rule directly.<sup>158</sup>

<sup>158</sup> Do it.

If, on the other hand, what we have is  $u(x, y)$  and  $v(x, y)$  and need to compute Eq. (330) from that, then we need either to solve  $x = x(u, v)$  and  $y = y(u, v)$  and proceed as before, which might be difficult (if at all possible). Or, we can use the fact that  $f = F \circ g^{-1}$  to turn to this configuration, where  $g^{-1}$  is the inverse in the sense of function composition:

$$\mathbb{R}^2 \xrightarrow[\substack{J_{g^{-1}} \\ (2 \times 2)}]{g^{-1}} \mathbb{R}^2 \xrightarrow[\substack{J_F \\ (1 \times 2)}]{F} \mathbb{R} \quad (331a)$$

$$(x, y) \mapsto (u, v) \mapsto F(u, v) \quad (331b)$$

$$\mathbb{R}^2 \xrightarrow[\substack{J_f \\ (1 \times 2)}]{f = F \circ g^{-1}} \mathbb{R} \quad (331c)$$

$$(x, y) \mapsto f(x, y) \quad (331d)$$

so that the chain rule Eq. (304) yields:

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \quad (332)$$

which is correct, but remember that we want Eqs. (330). This is, however, easily obtained from the linear, or matrix, equation Eq. (332) by taking the (matrix) inverse of the Jacobian:

$$\begin{pmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}^{-1} \quad (333)$$

and while the method is general, in the 2D case which we have chosen, the computation of the inverse is immediate. In all cases, it relies on the determinant of the Jacobian (remember, also called the Jacobian) and the adjunct of the Jacobian:

$$\begin{pmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \frac{\begin{pmatrix} \frac{\partial v}{\partial y} & -\frac{\partial u}{\partial y} \\ -\frac{\partial v}{\partial x} & \frac{\partial u}{\partial x} \end{pmatrix}}{\frac{\partial(u,v)}{\partial(x,y)}}, \quad (334)$$

where we have introduced the funny notation for Jacobian determinant, which in the  $2 \times 2$  case (you can generalize easily), reads

$$\frac{\partial(u,v)}{\partial(x,y)} \equiv \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}. \quad (335)$$

Anyway, Eq. (334) gives us the sought derivatives:

$$\frac{\partial F}{\partial u} = \frac{\frac{\partial v}{\partial y} \frac{\partial f}{\partial x} - \frac{\partial v}{\partial x} \frac{\partial f}{\partial y}}{\frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial v}{\partial x} \frac{\partial u}{\partial y}}, \quad (336a)$$

$$\frac{\partial F}{\partial v} = \frac{-\frac{\partial u}{\partial y} \frac{\partial f}{\partial x} + \frac{\partial u}{\partial x} \frac{\partial f}{\partial y}}{\frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial v}{\partial x} \frac{\partial u}{\partial y}}. \quad (336b)$$

Let us give as an example the computation of derivatives in polar coordinates, cf. Eqs. (327), (from which one can compute the gradient and other operators, as shown in Exercises). We are now in the situation

$$\mathbb{R}^2 \xrightarrow[\substack{J_g \\ (2 \times 2)}]{g} \mathbb{R}^2 \xrightarrow[\substack{J_f \\ (1 \times 2)}]{f} \mathbb{R} \quad (337a)$$

$$(r, \theta) \mapsto (x, y) \mapsto f(x, y) \quad (337b)$$

$$\mathbb{R}^2 \xrightarrow[\substack{J_F \\ (2 \times 2)}]{F \equiv f \circ g} \mathbb{R} \quad (337c)$$

$$(r, \theta) \mapsto F(r, \theta) \quad (337d)$$

with

$$x = r \cos \theta, \quad y = r \sin \theta \quad (338)$$

so that

$$\mathbf{J}_f = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \quad (339a)$$

$$\mathbf{J}_g = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \quad (339b)$$

$$\mathbf{J}_{f \circ g} = \begin{pmatrix} \frac{\partial F}{\partial r} & \frac{\partial F}{\partial \theta} \end{pmatrix} \quad (339c)$$

which gives, according to the chain rule  $\mathbf{J}_{f \circ g} = \mathbf{J}_f \mathbf{J}_g$ :

$$\frac{\partial F}{\partial r} = \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} \quad (340a)$$

$$\frac{\partial F}{\partial \theta} = -r \sin \theta \frac{\partial f}{\partial x} + r \cos \theta \frac{\partial f}{\partial y}, \quad (340b)$$

or getting rid of the functions to retain the expression in operator form (you might find this way of writing often, or using the same function there for “convenience”):

$$\frac{\partial}{\partial r} = \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} \quad (341a)$$

$$\frac{\partial}{\partial \theta} = -r \sin \theta \frac{\partial}{\partial x} + r \cos \theta \frac{\partial}{\partial y}. \quad (341b)$$

If we want higher-order derivatives, we simply iterate:

$$\frac{\partial^2 F}{\partial r^2} \equiv \frac{\partial}{\partial r} \frac{\partial F}{\partial r} \quad (342a)$$

$$= \frac{\partial}{\partial r} \left( \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} \right) \quad (342b)$$

$$= \cos \theta \frac{\partial}{\partial r} \frac{\partial f}{\partial x} + \sin \theta \frac{\partial}{\partial r} \frac{\partial f}{\partial y} \quad (342c)$$

and since  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$  are themselves functions of  $x, y$ , we can apply the chain rule to them as well, to compute:

$$\mathbf{J}_{(\partial_{x,y}f) \circ g} = \mathbf{J}_{\partial_{x,y}f} \mathbf{J}_g \quad (343)$$

where (grouping the cases  $x$  and  $y$  together when possible):

$$\mathbf{J}_{(\partial_{x,y}f) \circ g} = \begin{pmatrix} \frac{\partial^2 f}{\partial r \partial \{x,y\}} & \frac{\partial^2 f}{\partial \theta \partial \{x,y\}} \end{pmatrix} \quad (344a)$$

$$\mathbf{J}_{\partial_x f} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \end{pmatrix} \quad \text{and} \quad \mathbf{J}_{\partial_y f} = \begin{pmatrix} \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \quad (344b)$$

so that

$$\frac{\partial}{\partial r} \frac{\partial f}{\partial x} = \cos \theta \frac{\partial^2 f}{\partial x^2} + \sin \theta \frac{\partial^2 f}{\partial y \partial x} \quad (345a)$$

$$\frac{\partial}{\partial r} \frac{\partial f}{\partial y} = \cos \theta \frac{\partial^2 f}{\partial x \partial y} + \sin \theta \frac{\partial^2 f}{\partial y^2} \quad (345b)$$

$$\frac{\partial}{\partial \theta} \frac{\partial f}{\partial x} = -r \sin(\theta) \frac{\partial^2 f}{\partial x^2} + r \cos(\theta) \frac{\partial^2 f}{\partial y \partial x} \quad (345c)$$

$$\frac{\partial}{\partial \theta} \frac{\partial f}{\partial y} = -r \sin(\theta) \frac{\partial^2 f}{\partial x \partial y} + r \cos(\theta) \frac{\partial^2 f}{\partial y^2} \quad (345d)$$

which, inserted back into Eqs. (342) yields:

$$\frac{\partial^2 F}{\partial r^2} = \cos^2 \theta \frac{\partial^2 f}{\partial x^2} + 2 \cos \theta \sin \theta \frac{\partial^2 f}{\partial x \partial y} + \sin^2 \theta \frac{\partial^2 f}{\partial y^2}. \quad (346)$$

The second-order  $\theta$  derivative is obtained similarly

$$\frac{\partial^2 F}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left( -r \sin \theta \frac{\partial f}{\partial x} + r \cos \theta \frac{\partial f}{\partial y} \right) \quad (347)$$

but since one has to differentiate the sine and cosine, the algebra is a bit more lengthy:

$$\frac{\partial^2 F}{\partial \theta^2} = -r \cos(\theta) \frac{\partial f}{\partial x} - r \sin(\theta) \frac{\partial}{\partial \theta} \left( \frac{\partial f}{\partial x} \right) - r \sin(\theta) \frac{\partial f}{\partial y} + r \cos(\theta) \frac{\partial}{\partial \theta} \left( \frac{\partial f}{\partial y} \right) \quad (348a)$$

$$\begin{aligned} &= -r \cos(\theta) \frac{\partial f}{\partial x} - r \sin(\theta) \frac{\partial f}{\partial y} + \\ &\quad + r^2 \sin^2(\theta) \frac{\partial^2 f}{\partial x^2} - 2r^2 \sin(\theta) \cos(\theta) \frac{\partial^2 f}{\partial y \partial x} + r^2 \cos^2(\theta) \frac{\partial^2 f}{\partial y^2} \end{aligned} \quad (348b)$$

You could also do it in the other direction, to express  $\frac{\partial}{\partial \{x,y\}}$  in terms of  $\frac{\partial}{\partial \{r,\theta\}}$  (Problems) using one method or the other, or even do that in 3D. These are things you shall need to do at some point to solve problems with polar or cylindrical symmetries. It's a lot of algebra, but now you have the tools to do it.

## Exercises

### Isolines

Sketch some isolines for the following functions:

1.  $f(x, y) = x + y$ .
2.  $f(x, y) = xy$ .
3.  $f(x, y) = (x + y)^2$ .
4.  $f(x, y) = \sqrt{x^2 + y^2}$ .
5.  $f(x, y) = \exp(-x^2 - y^2)$ .
6.  $f(x, y) = \sin(\pi x) \sin(\pi y)$ .

Can you visualize them in 3D? (maybe even sketching them).

### *Symmetry of the second derivatives*

For every function from the previous questions, compute  $\partial_x f$ ,  $\partial_y f$ ,  $\partial_x^2 f$ ,  $\partial_y^2 f$  and the cross-derivatives  $\partial_{xy}^2 f$ ,  $\partial_{yx}^2 f$ , checking whether the latter are equal for all  $(x, y)$ . Also do so for the following functions:

1.  $f(x, y) = 2x^3y^4$ .
2.  $f(x, y) = x^p y^q$  for  $(p, q) \in \mathbb{N}^2$ .
3.  $f(x, y) = (x + y)^3$ .
4.  $f(x, y) = x \cos(\pi y) - y \sin(\pi x)$ .
5.  $f(x, y) = \tan(xy)$ .
6.  $f(x, y) = 1/\sqrt{x^2 + y^2}$ .

### *Gradients*

Compute the gradients of the functions introduced above.

### *Partial differential equation*

Consider  $f(x, y) = \ln(x) + \ln(y)$ . Derive an expression for  $(\partial_x f)(\partial_y f)$  in terms of  $f$ . Can you as a result propose a nonlinear partial differential equation of which  $f$  is a solution?

### *Perfect gas*

We will see in Thermodynamics that perfect gases obey the equation:

$$pV = RT \quad (349)$$

with  $R$  a constant and  $p$  (pressure),  $V$  (volume) and  $T$  (temperature) variables of the gas. Compute:

$$\frac{\partial p}{\partial V} \frac{\partial V}{\partial T} \frac{\partial T}{\partial p} \quad (350)$$

(it's not 1 as you would naively assume by simplifying the  $\partial$  terms, which you can never do, *quite on the opposite*).



## Quadratic surfaces

Quadratic surfaces are the solutions of the following equation in  $\mathbb{R}^3$  up to second-order (whence the name) in the variables:

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0 \quad (351)$$

for  $A, \dots, J$  some constants. Show that

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (352)$$

is the equation for a cylinder. How about?

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad (353)$$

In particular, what happens when  $a = b = c$ . The following is the equation for a cone (it really looks like a sandclock):

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}. \quad (354)$$

How about the following:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{z^2}{c^2}, \quad (355a)$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1 + \frac{z^2}{c^2}. \quad (355b)$$

There are a lot of quadratic surface and even listing them all would be a voluminous work. Let us conclude with this interesting case (a “saddle”), which we will encounter in stability problems:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}. \quad (356)$$

A particular family of quadratic surface, known as the “conics”, is usefully studied through the “matrix representation of conic sections” (which we will not cover in this course but you are welcome to study them).

### Gradient in polar coordinates

The gradient in Cartesian coordinates being defined as

$$\nabla = \hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y}, \quad (357)$$

and since  $r = \sqrt{x^2 + y^2}$  and  $\theta = \arctan(y/x)$ , show that:

$$\frac{\partial r}{\partial x} = \cos \theta \quad \frac{\partial r}{\partial y} = \sin \theta \quad (358a)$$

$$\frac{\partial \theta}{\partial x} = -\frac{\sin \theta}{r} \quad \frac{\partial \theta}{\partial y} = \frac{\cos \theta}{r} \quad (358b)$$

By using the chain rule, show that, as a consequence, in polar (2D) coordinates:

$$\nabla = \hat{e}_r \frac{\partial}{\partial r} + \hat{e}_\theta \frac{1}{r} \frac{\partial}{\partial \theta}. \quad (359)$$

### Laplacian in polar coordinates

The so-called Laplacian  $\nabla^2$  is defined in 2D-cartesian coordinates as  $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ . Can you justify the notation? Check that its 2D-polar coordinates expression is:

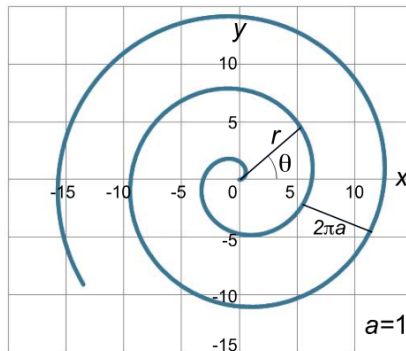
$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}. \quad (360)$$

### Archimedean spiral

Consider this function:

$$r = a\theta \quad (361)$$

for some constant  $a > 0$ . Satisfy yourself that the graph of this function is as follows:



Using the chain rule, compute  $y'(x)$  for this so-called Archimedean spiral (note that as a function  $y$  of  $x$ , it is not single-valued, so the derivative is defined locally only). Show that it is independent of  $r$ .



### Fermat's spiral

Fermat's spiral is defined as

$$r = \sqrt{\theta}. \quad (362)$$

Plot it and compare with Archimede's spiral. Show that, in this case

$$\frac{dy}{dx} = \tan(\theta + \arctan(2\theta)). \quad (363)$$



## *Lecture 21: The equations of Physics.*

We have now sufficient mathematical baggage to contemplate some of the main equations of Physics. We have seen differential equations of one variable and functions of several variables. Of course, some equations involve both differentials and several variables. We call such equations “partial differential equations” (PDE; we call the single-variable differential equations ODE where “O” stands for “ordinary”). Most equations of physics are partial differential equations. We will see two of them in some amounts of details, namely, the wave equation and the heat equation, and overview other foundational equations which support entire fields of physics, that you will study in details in dedicated courses. This will allow us to illustrate how the tools we have developed interconnect and relate to each others and form, literally, the foundations of Physics.

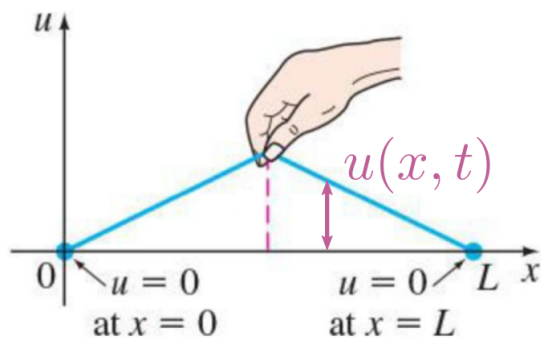
The wave equation is one of the most important equations of Physics. It is a fundamental and that describes various types of waves in various media. It can be derived from various particular cases. The more familiar form for us is maybe how it arises in mechanical systems, by applying Newton’s equation

$$\vec{F} = m\vec{a} \tag{364}$$

to an extended system, namely, a string, plucked at its two extremities. How do we model such a system? For a single point particle, we would keep track of its position at each moment of time. What for a string? Similarly, we can keep track of the position of each element of the string at all times. The simplest and most useful way to do so is to model the string through its vertical position at every point  $x$ , and also, of course, at all times  $t$ . Therefore, we need to deal with a quantity

$$u(x, t). \tag{365}$$

This is typically called a “field”, here a 1D field because we keep track of only one space dimension, namely, the elevation of the string, but we wouldn’t use this vocabulary in this case, although that’s what the structure of the problem is.



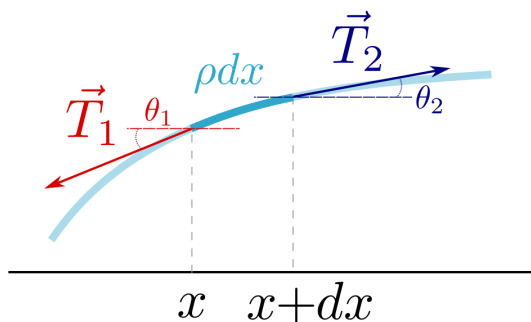
Let us compute the force first, in this case, considering only the tension in the string (neglecting gravity, etc.) A tension is a force (with units of N) transported or exerted along the length of a medium, especially a force carried by a flexible medium, such as a string, a rope or a cable. The word “tension” comes from the Latin word for “to stretch” (also naming the flexible cords that carry muscle forces to other parts of the body, the tendons). The tension being parallel to the string, considering a little element of string, we have for the net force:

$$\vec{F} = \vec{T}_1 - \vec{T}_2 \quad (366)$$

which, projected on the  $x, y$  axis, yields:

$$F_x = T_1 \cos \theta_1 - T_2 \cos \theta_2 \quad (367a)$$

$$F_y = T_1 \sin \theta_1 - T_2 \sin \theta_2 \quad (367b)$$



But since we assume that angles are vanishing, to first-order in the Taylor expansion, with  $\cos \theta \approx 1$  and  $\sin \theta \approx \theta$ , this simplifies to:

$$F_x = T_1 - T_2 \quad (368a)$$

$$F_y = T_1 \theta_1 - T_2 \theta_2 \quad (368b)$$

The first equation, Eq. (368a) is simple enough to solve, since the absence of longitudinal motion implies  $F_x = 0$ , i.e.,  $T_1 = T_2$  which

we will call simply  $T$ . The second involves  $\theta$ , which we'd better replace in terms of other more central parameters of the problems, namely,  $u$ . Clearly, since this is the slope of the string, we have  $\theta_1 = u'(x)$  and  $\theta_2 = u'(x + dx)$ . We can compute  $u'(x) = \frac{du}{dx}$  from  $du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial t}dt$ . Since  $x$  and  $t$  are independent variables,  $dt/dx = 0$ , and Eq. (368b) becomes:

$$F_y = T \left( \frac{\partial u}{\partial x}(x, t) - \frac{\partial u}{\partial x}(x + dx, t) \right) \quad (369)$$

that we rewrite as

$$F_y = Tdx \frac{\frac{\partial u}{\partial x}(x, t) - \frac{\partial u}{\partial x}(x + dx, t)}{dx} \quad (370)$$

which is the rate of change of  $\partial u/\partial x$  in the  $x$  variable for a fixed (held constant)  $t$ , i.e., the partial derivative of  $\partial u/\partial x$ , i.e., the 2nd-order partial derivative in  $x$ :

$$F_y = aTdx \frac{\partial^2 u}{\partial x^2}(x, t). \quad (371)$$

It would be – from the usual definition of the partial derivative but this depends how we define the positive  $y$  axis, so it can be absorbed there, as long as the calculation of the acceleration use the same convention, taken such as the object accelerates along the applied force. Now for the rhs of Eq. (364). The mass  $m$  is the mass of the string in the little (differential) element of string of length  $dx$ . Surely the mass is vanishing there:

$$m = \rho dx \quad (372)$$

The acceleration is, since the string oscillates only in the  $y$  direction, the time-derivative of the speed, which is itself

$$v_y = \frac{\partial u(x, t)}{\partial t} \quad (373)$$

i.e.,

$$a_y = \frac{\partial v_y}{\partial t} = \frac{\partial^2 u}{\partial t^2}(x, t) \quad (374)$$

so all together,  $ma_y = \rho dx \frac{\partial^2 u}{\partial t^2}$  which yields:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\rho}{T} \frac{\partial^2 u}{\partial t^2} \quad (375)$$

(this is  $F_y = ma_y$  after dividing by  $dx$  and bringing all constants on one side). The constant  $\rho/T$  has units of  $(\text{kg}/\text{m})/\text{N}=(\text{s}/\text{m})^2$  that is, inverse speed square  $1/v^2$ , so we write:

$$\boxed{\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2}} \quad (376)$$

which is dimensionally correct since the lhs has no units ( $\text{m}^2/\text{m}^2$ ) as is the rhs since  $\partial^2 u/\partial t^2$  has units of ( $\text{m}^2/\text{s}^2$ ) which cancel the dimensions of  $v^{-2}$ . This gives us a microscopic expression for the speed of waves in a string:  $v = \sqrt{T/\rho}$ .

We introduce some notations, starting with the d'Alembertian:

$$\square \equiv \frac{1}{v^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \quad (377)$$

so that in the space of solutions, the wave equation takes the funny form:

$$\square = 0. \quad (378)$$

In 3D, the d'Alembertian reads:

$$\square \equiv \frac{1}{v^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \quad (379)$$

where  $\nabla^2 = \nabla \cdot \nabla$  is an even more important operator, known as the *Laplacian*, which we can easily compute from the dot product:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (380)$$

Actually, the wave equation is easy to solve! To start with, it is a linear equation, in the sense that if  $u_1$  and  $u_2$  are solutions, then any linear superposition  $\alpha u_1(x, t) + \beta u_2(x, t)$ , for any  $\alpha, \beta \in \mathbb{R}$ , is also a solution. This is proved simply by computing:

$$\frac{\partial^2}{\partial x^2} [\alpha u_1 + \beta u_2] = \alpha \frac{\partial^2}{\partial x^2} u_1 + \beta \frac{\partial^2}{\partial x^2} u_2 \quad (381a)$$

$$= \alpha \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} u_1 + \beta \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} u_2 \quad (381b)$$

$$= \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} [\alpha u_1 + \beta u_2] \quad (381c)$$

We introduce the variables:

$$\zeta \equiv x + vt \quad (382a)$$

$$\eta \equiv x - vt \quad (382b)$$

so that, in particular:

$$\frac{\partial \zeta}{\partial x} = 1, \quad \frac{\partial \eta}{\partial x} = 1, \quad (383a)$$

$$\frac{\partial \zeta}{\partial t} = v, \quad \frac{\partial \eta}{\partial t} = -v, \quad (383b)$$

and apply the chain rule to  $z = U(\zeta, \eta) = u(x, t)$ :

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial \zeta} \frac{\partial \zeta}{\partial x} + \frac{\partial z}{\partial \eta} \frac{\partial \eta}{\partial x} \quad (384a)$$

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial \zeta} \frac{\partial \zeta}{\partial t} + \frac{\partial z}{\partial \eta} \frac{\partial \eta}{\partial t} \quad (384b)$$

which, from Eqs. (383), gives, after substitution  $u$  and  $U$  back for the go-in-betweeners  $z$ :

$$\frac{\partial u}{\partial x} = \frac{\partial U}{\partial \xi} + \frac{\partial U}{\partial \eta}, \quad (385a)$$

$$\frac{\partial u}{\partial t} = v \left( \frac{\partial U}{\partial \xi} - \frac{\partial U}{\partial \eta} \right). \quad (385b)$$

We iterate this to get the 2nd-order derivatives (remember that  $\partial z/\partial \xi$  and  $\partial z/\partial \eta$  are also functions of  $\xi, \eta$  so the same procedure applies. For the space 2nd-order derivative:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x} \frac{\partial u}{\partial x} \quad (386a)$$

$$= \frac{\partial}{\partial x} \frac{\partial U}{\partial \xi} + \frac{\partial}{\partial x} \frac{\partial U}{\partial \eta} \quad (386b)$$

$$= \frac{\partial^2 U}{\partial \xi^2} + \frac{\partial^2 U}{\partial \eta \partial \xi} + \frac{\partial^2 U}{\partial \xi \partial \eta} + \frac{\partial^2 U}{\partial \eta^2} \quad (386c)$$

$$= \frac{\partial^2 U}{\partial \xi^2} + 2 \frac{\partial^2 U}{\partial \xi \partial \eta} + \frac{\partial^2 U}{\partial \eta^2}. \quad (386d)$$

Eq. (386a) is the definition of the 2nd-order partial derivative, Eq. (386b) is substituting Eq. (385a), Eq. (386c) is the Chain-rule and Eq. (386d) is assuming commutativity of the partial derivatives and after collecting likewise terms.

Similarly, for the time 2nd-order derivative:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial t} \frac{\partial u}{\partial t} \quad (387a)$$

$$= v \left( \frac{\partial}{\partial t} \frac{\partial U}{\partial \xi} - \frac{\partial}{\partial t} \frac{\partial U}{\partial \eta} \right) \quad (387b)$$

$$= v \left( v \frac{\partial^2 U}{\partial \xi^2} - v \frac{\partial^2 U}{\partial \eta \partial \xi} - \left[ v \frac{\partial^2 U}{\partial \xi \partial \eta} - v \frac{\partial^2 U}{\partial \eta^2} \right] \right) \quad (387c)$$

$$= v^2 \left( \frac{\partial^2 U}{\partial \xi^2} - 2 \frac{\partial^2 U}{\partial \xi \partial \eta} + \frac{\partial^2 U}{\partial \eta^2} \right) \quad (387d)$$

with the same process except that this time we involve Eq. (385b) which introduces  $\pm v$ . With the new variables, Eq. (376) becomes:

$$\frac{\partial^2 U}{\partial \xi \partial \eta} = 0 \quad (388)$$

which is easy to solve. Indeed, Eq. (388) says that  $\partial U/\partial \eta$  is a constant of  $\xi$ , i.e.,  $\partial U/\partial \eta = C$  but this constant can be a function of  $\eta$  itself, since  $\xi$  and  $\eta$  are unrelated variables, therefore:

$$\frac{\partial U}{\partial \eta} = C(\eta). \quad (389)$$

We integrate Eq. (389), to find:

$$U = \int C(\eta) + B \quad (390)$$

where  $B$  is a constant of  $\eta$ , but this constant can in fact be a function of  $\xi$ , for the same reason as previously. Also, calling  $A$  the primitive of  $C$ , we have:

$$U(\xi, \eta) = A(\eta) + B(\xi) \quad (391)$$

or, in the original variables (using different letters for the functions expressed in other variables):

$$\boxed{u(x, t) = f(x + vt) + g(x - vt)}. \quad (392)$$

This is the Mathematical solution. From this point onward, we would go on and look at the Physics of the solution, how phenomena emerge, for instance, the notions of group velocity and phase velocity, wave packets, etc. We leave this to your Physics lectures (in particular, in this case, optics).

The wave equation is second-order in time. The same equation but with a first-order in time dependence is also an important one, it is known as the *Heat equation*. Its physical meaning is that of temperature flowing from hot to cold places, the rate of change being given by the average temperature difference. This is the equation for which Fourier introduced his technique of function decomposition, so we already have the tool to solve it, including the separation of variables, which applies also (and particularly effectively) to PDE, turning them into ODE. Namely, the heat equation reads:

$$\boxed{\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}}, \quad (393)$$

where  $T(x, t)$  is the temperature (SI units of Kelvin) at time  $t$  (in seconds) and position  $x$  (in meters) of a 1D object (a rod) which we take such that  $0 \leq x \leq 1$ . The constant  $\alpha > 0$  is called the thermal diffusivity and quantifies how well heats moves in a material. Temperature measures the amount of “heat” (thermal energy). In the following, we shall assume we are in a system of units such that  $\alpha = 1$ . In thermodynamics, we will see that heat is transferred from regions of higher temperature to regions of lower temperature. We assume that the temperature profile of the rod is given at  $t = 0$  and is fixed at all times at the extremities (in contact with a reservoir), i.e., we assume that

$$T(x, 0) = T_0(x), \quad (394a)$$

$$T(0, t) = T(1, t) = 0, \quad (394b)$$

for some function  $T_0$ . The temperature is given in  $^{\circ}\text{C}$  and can be negative. Solving the heat equation means providing the temperature of all points  $x$  of the rod at any time  $t$ , i.e., we want the full evolution of  $T(x, t)$ . To do so, we first assume that  $T(x, t)$  can be written in the form of a product of two functions of a single variable each (rather than one function of two variables). We call these functions  $T$  and  $\Theta$ , so that:

$$T(x, t) = X(x)\Theta(t). \quad (395)$$

We compute, first the lhs of Eq. (393):

$$\partial_t T = \partial_t(X\Theta) = (\partial_t X)\Theta + X\partial_t\Theta = X\partial_t\Theta = X\Theta' \quad (396)$$

since  $X$  is time independent and since  $\partial_t\Theta = \Theta'$ . Then, for the rhs:

$$\alpha\partial_x^2 T = \alpha\partial_x^2(X\Theta) = \alpha\partial_x\partial_x(X\Theta) = \alpha\partial_x[(\partial_x X)\Theta] = \alpha[\partial_x^2 X]\Theta = X''\Theta \quad (397)$$

since  $\Theta$  is  $x$  independent (so we put  $\partial_x\Theta = 0$  directly whenever it appeared) and  $\partial_x^2 X = X''$ . Equating the two sides:

$$X\Theta' = X''\Theta \quad (398)$$

Dividing both sides by  $X$  and  $\Theta$ :

$$\boxed{\frac{\Theta'}{\Theta} = \frac{X''}{X}}. \quad (399)$$

The boundary condition Eq. (394b) implies that  $X$  must satisfy:<sup>159</sup>

$$X(0) = X(1) = 0. \quad (400)$$

<sup>159</sup> Show it.

In Eq. (399), the left-hand side only depends on time  $t$  while the right-hand side only depends on position  $x$ . As changing time is independent from changing position, and vice-versa, it means that both combinations  $\Theta'/\Theta$  and  $X''/X$  are time and position independent, i.e., they are constant. We call  $-\lambda$  this common constant (both terms being equal; the sign is for convenience). This thus gives two ODE:

$$X''(x) = -\lambda X(x), \quad (401a)$$

$$\Theta'(t) = -\lambda\Theta(t), \quad (401b)$$

both being linear and homogeneous, so of the easiest kinds that we have covered earlier in the course!

If we assume that  $\lambda < 0$ , it can then be written as  $\lambda \equiv -k^2$  (since  $k \in \mathbb{R}$  is positive), and  $X(x) = Ae^{kx} + Be^{-kx}$ , with  $A, B$  two constants that, to satisfy the boundary conditions Eq. (400), must be such that  $A = B = 0$ , i.e., such solutions can be discarded (as trivial). Therefore, assuming instead that  $\lambda > 0$ , we find that

$$X(x) = A \cos(\sqrt{\lambda}x) + B \sin(\sqrt{\lambda}x) \quad (402)$$

is a solution of Eq. (401a), with boundary conditions such that

$$B \sin(\sqrt{\lambda}x) = 0. \quad (403)$$

Clearly, Eq. (403) cannot be solved by finding  $B$ . Instead, it must be solved by choosing the right  $\lambda$ , namely:

$$\sqrt{\lambda} = n\pi, \quad (404)$$

for  $n \in \mathbb{N}^*$ . Therefore, we have a full family of different solutions, that form a basis:

$$S_n(x) \equiv \mathcal{N}_n \sin(n\pi x) \quad (405)$$

for  $n \in \mathbb{N}^*$  ( $n = 0$  is the trivial solution of a constant equal to zero, and so we discard it). We have introduced a normalization constant  $\mathcal{N}_n$  to normalize them, which we find to be:<sup>160</sup>

$$\mathcal{N}_n = \sqrt{2} \quad \text{for all } n \geq 1. \quad (406)$$

From our Fourier lecture, we know that any function  $f$  can be written as a Fourier series:

$$f(x) = \sum_{k=0}^{\infty} c_k \sin(n\pi k), \quad (407)$$

for some coefficients  $c_k$ , which depend on the function.

Now to solve Eq. (401b), and keeping in mind that we now know  $\lambda$  from the  $x$ -part of the equation (namely, Eq. (404)), we have:

$$\Theta_n(t) = c_n e^{-n^2\pi^2 t} \quad (408)$$

where  $c_n$  are constants.

Since the Heat equation is linear, meaning that if  $T_1(x, t)$  and  $T_2(x, t)$  are solutions, then so is  $\alpha T_1 + \beta T_2$  for any two scalars  $\alpha, \beta \in \mathbb{R}$ , and since  $S_n(x)\Theta_n(t)$  is a solution of Eq. (393), then an arbitrary linear superposition of solutions in the form  $S_n\Theta_n$  is also a solution, i.e.:

$$T(x, t) = \sum_{n=1}^{\infty} \alpha_n \mathcal{N}_n \sin(n\pi x) e^{-n^2\pi^2 t}. \quad (409)$$

To complete the solution, we need to provide the coefficients  $\alpha_n$ .

These are fixed by the initial condition, i.e., from Eq. (409) at  $t = 0$ , for which we take the inner product of both sides with  $S_k(x) = \mathcal{N}_k \sin(k\pi x)$ . Using the orthonormality of the basis  $S_n$ , the coefficient  $\alpha_k$  are found according to the Fourier procedure, as:

$$\alpha_k = \mathcal{N}_k \int_0^1 T(x, 0) \sin(n\pi x) dx. \quad (410)$$

For instance, if the temperature profile is initially  $T(x, 0) = x(1 - x)$ , then, using integration by parts, we find that

$$\alpha_k^{(1)} = \int_0^1 4x(1 - x) \sqrt{2} \sin(k\pi x) dx = \frac{8\sqrt{2}(1 - (-1)^k)}{k^3\pi^3}, \quad (411)$$

<sup>160</sup> Show it



for  $k \in \mathbb{N}^*$ . This gives us the semi-analytical solution for this particular initial condition:

$$T^{(1)}(x, t) = \sum_{k \text{ odd}} \frac{16}{k^3 \pi^3} \sin(k\pi x) e^{-k^2 \pi^2 t}. \quad (412)$$

That is an infinite series. At this point we would use a computer to compute the sum and plot it. That is when we are lucky (or have worked enough) to have a solution. In most cases, we would actually integrate numerically the equation, which we could do even if we have the exact, or closed-form solution, to check. This brings us to another topic, scientific computing, or how to defer our questions to brute numerical computation. Here's an example of how the integration of Eq. (401b) could be achieved with a computer. We return to the now familiar approximation of the derivative:

$$\Theta(t + \Delta t) = \Theta(t) - \lambda \Theta(t) \quad (413)$$

from which one can see that knowing  $\Theta$  at time  $t$  allows us to know it at time  $t + \Delta t$ , and iterating, we could thus reconstruct the value for all  $t$ . This is how we could do it in the Julia programming language, using not  $\Theta$  but  $y[i]$  for the value of the function at the  $i$ -th point of the grid we use to discretize the problem (to avoid non-ASCII characters, although Julia supports them), with a timestep  $\Delta t$  that we encode in  $h$ :

$$y[i+1] = y[i] + h * f((i-1)*h, y[i])$$

where  $f$  is the function defined to encode the ODE  $y' = f(t, y)$ , exactly as we have done it with the mathematical theory. A full working code would read:

---

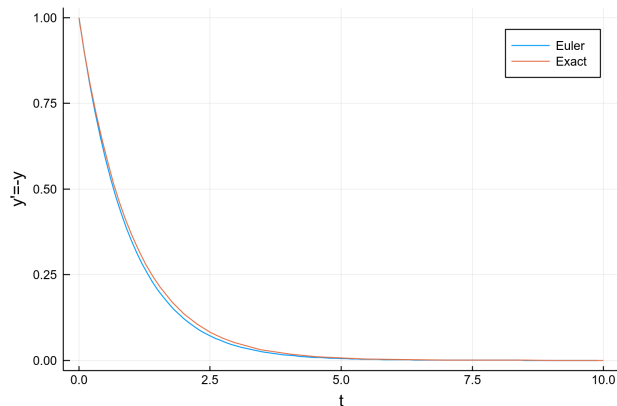
```

1 npts = 100; # number of points
2 y=0.0*collect(1:npts); # predefine the array
3 h=.1; # timesteps, so total time = npts*h=10
4 # The differential equation to solve
5 function f(t,y)
6     -y
7 end
8 y[1]=1; # initial condition
9 # The numerical method (Euler's)
10 for i=1:npts-1
11     y[i+1]=y[i]+h*f((i-1)*h,y[i])
12 end
13
14 plot(((1:npts).-1)*h, [y[i] for i=1:npts], label="Euler", xlabel="t", ylabel="y'=-y")
15 plot!(x->exp(-x),0, 10, label="Exact")

```

---

and returns



Here there are various things that we will need to look at into details: mainly, the syntax of the language (e.g., what is `[y[i] for i=1:npts]` doing for instance?) and the method itself, which is particularly naive and not efficient: you can see how, although qualitatively correct, it differs from the exact solution. It is still famous enough to have its own name: it is called the explicit Euler method. We shall learn why it is not exact and how to implement better methods, quantifying as well their performance.

The wave and heat equations are real-valued equations. Even if we use complex numbers, say, for convenience, the quantities they describe are real numbers. Schrödinger's equation is technically a heat equation for an intrinsically complex-valued field:

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\nabla^2}{2m} + V\right)\psi \quad (414)$$

where  $\psi(\mathbf{r}, t)$  is the so-called *wavefunction* whose modulus square gives a probability density. Because it has the same structure as the Heat equation, it is solved in basically the same-way, first by separation of variables and then by methods of linear algebra and typically Fourier analysis. The physics comes from the potential in question,  $V(\mathbf{r})$  which can be, for instance, a square potential (zero in some interval and infinite at the boundaries), or quadratic (harmonic oscillator), a defect (Dirac  $\delta$  function) or even no potential at all (free particle). All these are first studied in 1D and this is what we do in first-year quantum mechanics. An important 3D potential is Coulomb's potential  $V(\mathbf{r}) = e/(4\pi\epsilon_0|\mathbf{r}|)$  of a point-charge  $e$ , which describes the Hydrogen atom. The difficulty there will not be so much in the technical mathematical tools—we have basically seen them all already—but in their interpretation, in what they have to say about the world, with strange concepts such as quantum superpositions (which are linear superpositions of vectors), entanglement

and nonlocality. In fact, the wavefunction  $\psi$  will be seen to be better described as an abstract vector  $|\psi\rangle$  and physical quantities as operators, e.g., the Hamiltonian  $H = -\hbar^2\nabla^2/(2m) + V$  (the energy) or the momentum  $p = -i\hbar\nabla$ , are operators that apply on the wavevectors, or ket.

Maxwell's equations are equations for the electric  $\mathbf{E}(\mathbf{r}, t)$  and magnetic  $\mathbf{B}(\mathbf{r}, t)$  fields. Here not only the variables but also the functions themselves are vectors. So we deal with vector fields, and this introduces new concepts to characterize their geometry, such as the divergence, or a measure of "flow" of the field, which is the trace of the Jacobian, or sum of partial derivatives, and formally obtained by taking the scalar product with the nabla operator  $\nabla$ :

$$\nabla \cdot F = \frac{\partial F_x}{x} + \frac{\partial F_y}{y} + \frac{\partial F_z}{z} \quad (415)$$

or the curl, or measure of the "rotation" of the field, which is obtained

$$\nabla \times F = \left( \frac{\partial F_z}{y} - \frac{\partial F_y}{z} \right) \hat{i} + \frac{\partial F_z}{z} \quad (416)$$

An important result, Helmholtz's theorem but also known as the fundamental theorem of vector calculus, states that any field (with suitable, physical-looking, properties) can be decomposed as the sum of a gradient and a curl:

$$\vec{F} = -\nabla\phi + \nabla \times \vec{A}. \quad (417)$$

which means, for reasons that are clarified when studying the geometrical meaning of these operators, that one can describe completely a field by specifying the dynamics of its divergence and curl. This is what Maxwell equations do for the electric and magnetic fields, in the form of first-order linear PDE:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (418a)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (418b)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (418c)$$

$$\nabla \times \mathbf{B} = -\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J} \quad (418d)$$

as a function of the electric density  $\rho$  and current densities  $\mathbf{J}$ . We first study static fields, i.e.,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (419a)$$

$$\nabla \times \mathbf{E} = 0 \quad (419b)$$

which is known as “electrostatics”, and

$$\nabla \cdot \mathbf{B} = 0 \quad (420a)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (420b)$$

which is known as “magnetostatics”. These are beautiful illustrations of how vector fields can behave: in one case, in the presence of a charge, and the fluids flow, in the other case when a current circulates, and so does the field.

But the prettiest part is yet to come. The term  $\frac{1}{\epsilon_0 \mu_0} \frac{\partial \mathbf{E}}{\partial t}$  was added by Maxwell for self-consistency of the separate terms. If you further combine these equations between each others, as Maxwell did, you, you arrive to the following second-order PDE:

$$\left( \epsilon_0 \mu_0 \nabla^2 - \frac{\partial^2}{\partial t^2} \right) \mathbf{E} = \mathbf{0}, \quad (421a)$$

$$\left( \epsilon_0 \mu_0 \nabla^2 - \frac{\partial^2}{\partial t^2} \right) \mathbf{B} = \mathbf{0}. \quad (421b)$$

These should actually be familiar by know. They are two wave equations. The electric and magnetic fields can actually propagate, at the speed, the wave equation tells us,  $\sqrt{\epsilon_0 \mu_0}$  which, numerically, was about the known speed of light. Therefore, so Maxwell deduced, light is an electromagnetic field! One of the most enlightning discoveries of the human mind!

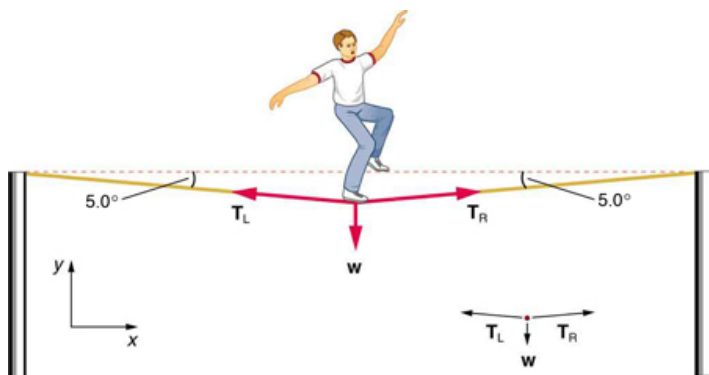
## Exercises

### Speed

If a steel wire 2m in length weighs 0.9N and is stretched by a tensile force of 300N, what is the corresponding speed of transverse waves?

### Tension

Calculate the tension in a wire that supports a 70.0kg tightrope walker who cause the wire to sag by  $5^\circ$ .



### Solutions or not

Assuming  $\omega/k = v$ , which of the following functions are solutions of the wave equation? (Justify your answers).

$$u_a(x, t) = \cos(kx - \omega t), \quad (422a)$$

$$u_b(x, t) = \cos(kx) \cos(\omega t), \quad (422b)$$

$$u_c(x, t) = \cos(kx) \exp(\omega t), \quad (422c)$$

$$u_d(x, t) = \exp(kx - \omega t), \quad (422d)$$

$$u_e(x, t) = e^{i(kx - \omega t)}, \quad (422e)$$

$$u_f(x, t) = e^{i(kx - i\omega t)}, \quad (422f)$$

$$u_g(x, t) = \exp(-(kx - \omega t)^2 / \sigma_x^2), \quad (422g)$$

$$u_h(x, t) = f(kx - \omega t), \quad (422h)$$

where  $f$  in Eq. (422h) is a differentiable function.

### Two-ways transport

Show that the following:

$$\left[ \frac{\partial}{\partial t} - v \frac{\partial}{\partial x} \right] \left[ \frac{\partial}{\partial t} + v \frac{\partial}{\partial x} \right] u = 0 \quad (423)$$

is another way to write the wave equation, i.e., expand the product (keeping in mind these are operators, they do not commute in general but here we assume they do because of commutativity of partial derivatives), and recover the familiar wave equation. Show that  $f(x \pm vt)$  (for arbitrary well-behaved function  $f$ ) is a solution of the so-called “transport” equation:

$$\frac{\partial}{\partial t} \pm v \frac{\partial}{\partial x} = 0. \quad (424)$$

Note that left-propagating solutions are not solutions of the right-propagating equation, and vice-versa. For this you need the wave equation.

### 2D Wave equation

The 2D wave equation for the displacement  $u(x, y)$  reads

$$\frac{\partial^2 u}{\partial t^2} = v^2 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \quad (425)$$

Show that its expression in polar coordinates is:

$$\frac{\partial^2 u}{\partial t^2} = v^2 \left( \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} \right) \quad (426)$$

Assuming polar symmetry, i.e., that the wave does not depend on  $\theta$ , show that the wave equation simplifies to:

$$\frac{\partial^2 u}{\partial t^2} = v^2 \left( \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} \right). \quad (427)$$

By making the change of variables  $u(r, t) = R(r)T(t)$ , show that the equation can be written in the form

$$\frac{T''(t)}{v^2 T(t)} = \frac{1}{R(r)} \left( R''(r) + \frac{1}{r} R'(r) \right) \quad (428)$$

which gives rise to two ordinary (single-variable) differential equations:

$$T''(t) = Kv^2 T(t), \quad (429a)$$

$$rR''(r) + R'(r) - KrR(r) = 0. \quad (429b)$$

The  $T$  part is trivial to solve.



The  $R$  part admits so-called Bessel equations as a solution. You can pursue this problem further and solve the full 2D polar-symmetric wave equation.