Mathematical Methods for Physicists

Henrik Jeldtoft Jensen
Department of Mathematics
Imperial College London

(Dated: June 28, 2010)

Contents

I. INTRODUCTION

The following is a set of notes to be read in conjunction with the lectures delivered to the 2nd, 3rd and 4th year Students of Physics at Imperial College London.

Let us make it clear from the beginning; mathematics is the language needed to be able to formulate the laws of Nature and it is the language needed in order to be able to think about the subtleties of Nature. As with any other language it is difficult to separate the content of a message from the notation and the syntax. Try to formulate the content of Newton's 2nd law *without* using any mathematical notation.

Physicists need mathematics in order to be able to talk and to think. The basic mathematics covered in this course is not only for the theoretical physicists. It is a prerequisite for anyone who wants to be able to read about physics in books or scientific papers or to be able to follow a physics seminar. I have met students who thought they were better physicists because they didn't waist their time or energy on mathematics. This is a misunderstanding. Physics, as any other reasonably developed science, can only be appreciated by the use of mathematics. Not only is mathematics needed to be able to carry out simple analytic analysis of problems. More importantly, one is only able to understand the concepts of physics, if one possesses some degree of mathematical flexibility.

It is as impossible to learn mathematics solely by watching others perform, as it is to learn to play a musical instrument exclusively by listening to other people play. This is not a problem. The mind expanding experience of digesting mathematical concepts through the contemplation of the lecture notes and the exercises will easily capture the open minded person and lead to a sound engagement with one of humankind's oldest and most profound activities. The musing of mathematical ideas and constructions can be done anywhere. All what is needed is to activate this most marvellous of all instruments: our brain. And sometimes pencil and paper comes in handy too. In fact it is hardly possible to understand a mathematical text by *reading* it. It is much better to *copy* every single detail and to think carefully about what it all means while coping.

The Cambridge book Riley, Hobson and Bence: *Mathematical Methods for Physics and Engineering*, together with the *Students Solution Manual for Mathematical Methods for Physics and Engineering* by Riley and Hobson is magnificent. If one gets hold of this pair of books and read and work with the problems, one will become greatly accomplished in the math used by physics.

But there exist many excellent text books on mathematics written for physicists and other scientist. I list a few which I know and have found readable.

Introductory level:

I1 G. Arfken, "Mathematical Methods for Physicists", Academic Press.

I2 H. Jeffreys and B.S. Jeffreys, Methods of Mathematical Physics, Cambridge University Press.

I3 D. A. McQuarrie, Mathematical Methods for Scientists and Engineers", University Science Books.

I4 J. Mathews and R.L Walker, "Mathematical Methods of Physics", Benjamin/Cummings Publishing Company.

Comprehensive more advanced level:

A1 Courant and Hilbert, "Methods of Mathematical Physics".

A2 Morse and Feshbach, " Methods of Theoretical Physics".

Here follows a few books that might help one to appreciate the nature of mathematics.

E1 J. Dieudonné, "Mathematics - The music of reason", Springer.

E2 G. M. Phillips, "Two Millennia of Mathematics", Springer

E3 M. Aigner and G. M. Ziegler, "Proofs from THE BOOK", Springer

E4 J. R. Brown, "Philosophy of Mathematics", Routledge.

II. COMPLEX FUNCTIONS

Complex functions are not any more complicated than real functions as soon as one has got used to them. That complex numbers are called complex is a matter of history, not an indication that they are more remote and less relevant or less real than real numbers are.

That complex numbers are intricately connected to real and natural numbers should be immediately clear after a few moments of consideration of the famous and spectacular formula.

$$e^{\pi i} = -1. \tag{1}$$

The irrational number π denotes the ratio between the circumference and the diameter of a circle. The oldest record of this number is from 1650 BC by an Ahmes, Egyptian scibe. The irrational number e was first mentioned in writing in 1618 by Napier in his work on logarithms. The natural way to think about e is to think of a quantity that increases in time with a rate proportional to itself. I.e. a quantity $N(t)$ (think of the number fragments produced by nuclear fission) that evolves according to the equation

$$\frac{dN(t)}{dt} = N(t). \tag{2}$$

It is easy to see that this equation implies an exponential growth of $N(t)$. Replacing the right hand side by differences

$$\frac{N(t+\delta) - N(t)}{\delta} = N(t) \Rightarrow \tag{3}$$

$$N(t+\delta) = [1+\delta]N(t) \tag{4}$$

and repeating this relation we get

$$N(q\delta) = [1+\delta]^q N(0) \tag{5}$$

for any natural number q. The exact solution to Eq. (2) subject to the initial condition $N(0) = 1$ is denote by $N(t) = e^t$, and hence the number e emerges from the study of *rate equations*. An alternative natural way of introducing e is through the discussion of logarithms.

The imaginary unit was introduced in order to be able to give meaning to roots in equations of the form $x^2 = -1$. Imaginary numbers have been refered to ever since Heron of Alexandria (1st century AD) but became only common place after 1500. Today we know that complex numbers are not just a question of finding roots in mathematical equations they are essential tools for the understanding of the most profound aspects of reality as captured by Quantum Mechanics.

The point we want to make is that Eq. (1) relates three numbers from three different ages of human history in a way that is far from obvious, seen in relation to how these numbers were discovered. Who would have expected that the square root of -1 and the ratio between the circumference and the diameter of a circle are so intimately related? Complex numbers and complex functions are as real and as necessary as the natural numbers.

Complex numbers and complex functions are generalisations of their real counter parts. They grow out of the real numbers and functions through the simplest possible generalisations. The best way to approach complex numbers and functions is first to think carefully about definitions and concepts already familiar from real analysis and then follow the consequences as these definitions are carried over into the complex regime. Differentiation is an important and transparent example of this procedure.

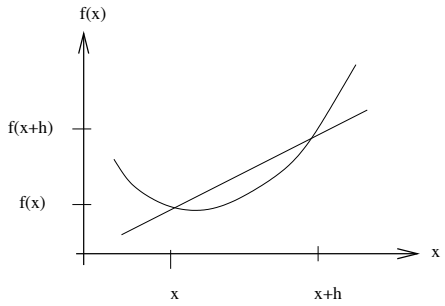

FIG. 1 The line approaches the tangent as $h \to 0$. The slope of the tangent is defined as the derivative.

A. Differentiation of Complex Functions

Recall how one differentiates a real function. The derivative of the function $f(x)$ at the point x_0 can be interpreted as the slope of the tangent at this point. The reason for this is clear. Consider the slope of the line from the point $(x, f(x))$ to the point $(x+h, f(x+h))$. See Fig. 1. The derivative is obtained as the value of the slope of this line in the limit $h \to 0$. Now note the following important point. We only consider $f(x)$ to be differentiable if the same value for the slope is obtained irrespectively of *how* the limit $h \to 0$ is assumed. We can take the limit while demanding $h > 0$, or we can take the limit while insisting that h remains non-positive. The function is only differentiable at x_0 if this makes no difference. [1]

The essential point is that the limit

$$\lim_{\delta x \to 0} \frac{f(x_0 + \delta x) - f(x)}{\delta x} \equiv \frac{df}{dx}|_{x_0} \tag{6}$$

is *independent* of how the limit is taken. The derivative of a complex function $f(z)$ at the point z_0 is defined in exactly the same way

$$\frac{df}{dz}|_{z_0} = \lim_{\delta z \to 0} \frac{f(z_0 + \delta z) - f(z)}{\delta z}, \tag{7}$$

with the same insistence that the limit must not depend on the manner in which the limit is taken, This has remarkably restrictive consequences in the case of complex functions. From the requirement that the same value is obtained for df/dz when we let δz vary parallel to the x-axis as when we let δz vary parallel to the y-axis it follows, (as will be discussed in the lectures) that the real part and the imaginary part of $f(z)$ are linked together in the following intriguing way.

$$\begin{aligned} \frac{\partial U}{\partial x} &= \frac{\partial V}{\partial y} \\ \frac{\partial V}{\partial x} &= -\frac{\partial U}{\partial y}. \end{aligned} \tag{8}$$

We used the following notation $z = x + iy$ together with

$$\begin{aligned} f(z) &= \mathrm{Re} f(z) + i \mathrm{Im} f(z) \\ &= U(x,y) + iV(x,y). \end{aligned} \tag{9}$$

The relations in Eq. (8) are called the Cauchy-Riemann conditions. Not only does a differentiable function satisfy these conditions, we will find that if a complex function satisfies Eq. (8) then the limit in Eq. (7) will be independent of how δz is taken to zero, and, hence, a function that satisfies Eq. (8) will also be differentiable.

Thus, to check if a complex function is differentiable one simply check if its real and imaginary parts fulfil the Cauchy-Riemann condition.

---

[1] One can define the derivative from the left or from the right and they might be different if the function has a cusp at x_0, but in that case the function is not differentiable, only differentiable from the left or the right respectively.

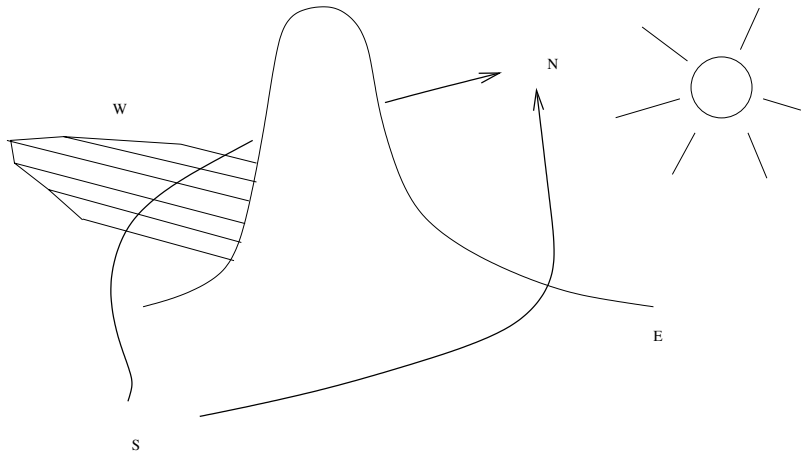

FIG. 2 Many ways around the mountain, but only two different classes

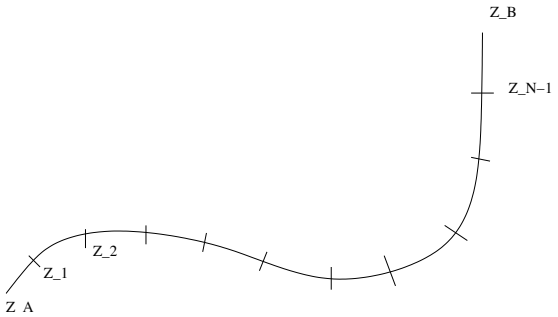

FIG. 3 Integration along a path in the complex plane.

B. Residue theory

Complex differentiable functions posses a spectacular sensitivity to theirs global environment. Think of it this way. You are going to cross a big plain from South to North, see Fig. (2). In the centre there is a high mountain. It is early morning and the sun is in the East. You can either pass to the East of the mountain or to the West. Whether you pass two kilometres or three kilometres to the East of the mountain won't make much difference to you. But there is a big difference between passing the mountain to East or to the West. If you decide on the Western route you'll find yourself in the shadow of the mountain for a while, whereas if you stay on the Eastern route you'll be warmed by the sun during the entire trip.

It is the same with integrals in the complex plane of complex functions. The integral is define much in the same way as the Riemann sum used to define the integral of a real function. Assume that the path P leads from point z_A to point z_B in the complex plane. Chop the path up in $N-1$ segments from z_A to z_1, and from z_1 to z_2 etc. see Fig. 3. We introduce $dz_1 = z_1 - z_a$, $dz_2 = z_2 - z_1$,..,$dz_N = z_B - z_{N-1}$. The following sum

$$I_N = \sum_{i=1}^{N} f(z_i) dz_i, \tag{10}$$

is used to define the integral from z_A to z_B along the path P by assuming the limit $N \to \infty$, i.e. we define

$$\int_P f(z)dz \equiv \lim_{N\to\infty} I_N. \tag{11}$$

Consider now the integral along different paths leading from point z_A to point z_B in the complex plane. If the complex function $f(z)$ is differentiable everywhere in the plane then, as we will see in the lectures,

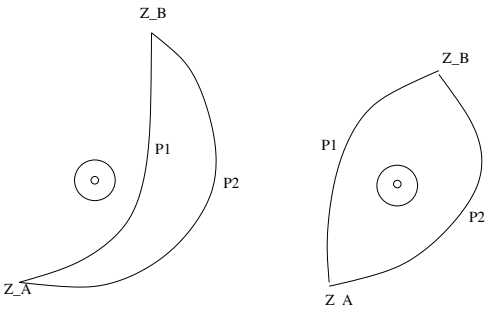

FIG. 4 Path not including or including a sigularity.

the integral of $f(z)$ from z_A to z_B will not depend on which path one follows. The situation is, however, different if $f(z)$ has a singularity of the form called a pole. If two paths P_1 and P_2 pass the same way around this pole then the integrals along the two paths are identical. The two integrals are on the other hand different if the two paths enclose the pole. See Fig. 4.

III. FOURIER TRANSFORMATION

We are used to decomposing the 3 dimensional position of a particle $\mathbf{r}$ as the sum

$$\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3 \tag{12}$$

along three independent direction vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. Now think of a function from the real numbers to, say, the complex numbers[2] $f : \mathbb{R} \to \mathbb{C}$. Joseph Fourier (1768-1830) discovered that a function $f(x)$ can be represented as a sum much in the same way as in Eq. (12). The difference is that the sum will typically include infinitely many terms and that instead of direction vectors we need to use functions. It is convenient to use exponential functions $x \mapsto e^{ikx}$ of different wave vectors k. In a schematic fashion we can write

$$f(x) = \sum_k \hat{f}(k) e^{ikx}. \tag{13}$$

The important point to pay attention to is that the function $f(x)$ according to (13) is specified by a set of "coordinates" $\hat{f}(k)$, one for each value of k, and a given set of *reference* functions.

The details are as follows. The sum in Eq. (13) is, for functions defined on the entire real axis, replaced by an integral, but that is simply because we may need to include functions e^{ikx} for all possible $k \in \mathbb{R}$ values - an integral and a sum are in our context essentially the same thing. The following two equations express $f(x)$ as a superposition of exponentials and give a recipe for how to determine the expansion coefficients $\hat{f}(k)$.

Forward Fourier transformation:

$$\hat{f}(k) = \int_{-\infty}^{\infty} dx f(x) e^{-ikx}. \tag{14}$$

Inverse Fourier transformation:

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{f}(k) e^{ikx}. \tag{15}$$

---

[2] One might also think of functions from $\mathbb{R}$ to $\mathbb{R}$, that won't make much difference.

The inverse transform can be considered equivalent to the determination of the coefficients in Eq. (12). In the case of vectors we determine the coordinates, say along the 2nd basis direction $\mathbf{e}_2$ from the scalar product. If $\mathbf{e}_2$ has length 1 and $\mathbf{e}_i$ are mutually orthogonal, we have $y = \mathbf{r} \cdot \mathbf{e}_2$. The integral between $f(x)$ and e^{-ikx} in Eq. (15) is like a scalar product between $f(x)$ and e^{-ikx}.

A. The power of the Fourier transform

Fourier transformation is a very powerful tool when analysing physical problems. We often have to solve both ordinary as well as partial differential equations. In condensed matter physics and statistical mechanics theoreticians as well as experimentalists use response functions and correlation functions to characterise the behaviour. The mathematical analysis relies on some rather impressive properties of the Fourier transform. E.g. that differentiation of a function corresponds to a multiplication of its Fourier transform. That certain types of integrals of functions correspond to multiplication of Fourier transforms. We list a few of the details below.

1. Differential equations

Consider the ordinary differential equation

$$\frac{df^2}{dx^2} + f = g(x). \tag{16}$$

We know that the solution is obtained from the homogeneous Eq.

$$\frac{d^2 f}{dx^2} + f = 0$$

by adding a particular solution to the inhomogeneous Eq. (16). We also know how to find the solution to the homogeneous Eq. by solving the auxiliary eq. $\lambda^2 + 1 = 0$. But how do we obtain the solution to the inhomogeneous Eq. (16 directly. In introductory courses all what is normally said is that one tries to guess. Instead one may use Eq. (15). First we notice that

$$\begin{aligned} \frac{df}{dx} &= \frac{d}{dx}\int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{f}(k) e^{ikx} \\ &= \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{f}(k) \frac{d}{dx} e^{ikx} \\ . &= \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{f}(k)[ik] e^{ikx}. \end{aligned} \tag{17}$$

From this we conclude that the Fourier transform of the derivative of a function is equal to ik times the Fourier transform of the function:

$$\frac{\hat{df}}{dx}(k) = ik\hat{f}(k). \tag{18}$$

Thus, if we substitute the left hand side of Eq. (15) for $f(x)$ and the equivalent for $g(x)$ we obtain from Eq. (16)

$$(ik)^2 \hat{f}(k) + \hat{f}(k) = \hat{g}(k). \tag{19}$$

Or equivalently

$$\hat{f}(k) = \frac{\hat{g}(k)}{1 - k^2}. \tag{20}$$

It is now just a matter of performing the inverse Fourier transformation to obtain the sought of function

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{\hat{g}(k)}{1 - k^2} e^{ikx}. \tag{21}$$

2. Response functions

Let us consider the gravitational potential $\phi(\mathbf{r})$ at position $\mathbf{r}$ produced by a set of N masses $m(\mathbf{r}_i)$ located at the positions $\mathbf{r}_i$ with $i = 1, 2, ..., N$,

$$\phi(\mathbf{r}) = \sum_{i=1}^{N} \frac{-G_0}{|\mathbf{r} - \mathbf{r}_i|} m(\mathbf{r}_i). \tag{22}$$

Here G_0 is the gravitational constant. If we instead consider a continuous mass distribution the equation will become an integral equation:

$$\phi(\mathbf{r}) = \int_V \frac{-G_0}{|\mathbf{r} - \mathbf{r\prime}|} \rho(\mathbf{r\prime}) d\mathbf{r\prime}. \tag{23}$$

Where $\rho(\mathbf{r\prime})$ is the mass density at position $\mathbf{r\prime}$. The structure of Eq. (23) is frequently encountered in physics and is of the form

$$\phi(\mathbf{r}) = \int_V G(\mathbf{r} - \mathbf{r\prime}) q(\mathbf{r\prime}) d\mathbf{r\prime}. \tag{24}$$

This type of Eq. describes how the effect of a "charge" at position $\mathbf{r\prime}$ is "transmitted" through space by the *response* function $G(\mathbf{r} - \mathbf{r\prime})$ to the position $\mathbf{R}$ where we "monitor" the response $\phi(\mathbf{r})$, accumulated from all the charges present. Note that the response function depends only on the relative position vector $\mathbf{R} - \mathbf{r\prime}$. In Eq. (23) the gravitational response function is obviously given by $G(\mathbf{r} - \mathbf{r\prime}) = -G_0/|\mathbf{r} - \mathbf{r\prime}|$ and the gravitational charge is simply the mass density $q(\mathbf{r\prime}) = \rho(\mathbf{r\prime})$. The response Eq. (24) is particularly simple when Fourier transformed. Assume one dimension for simplicity. Substitute the expression in Eq. (15) into Eq. (24)

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{\phi}(k) e^{ikr} &= \int_V \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{G}(k) e^{ik(r-r\prime)} \int_{-\infty}^{\infty} \frac{dk\prime}{2\pi} \hat{q}(k\prime) e^{ik\prime r\prime} dr\prime \\ &= \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{G}(k) e^{ikr} \int_{-\infty}^{\infty} \frac{dk\prime}{2\pi} \hat{q}(k\prime) \int_V e^{i(k\prime - k)r\prime} dr\prime. \end{aligned} \tag{25}$$

In the lectures we introduce Dirac's delta function and discuss that

$$\int dr\prime e^{i(k\prime - k)r\prime} = 2\pi\delta(k\prime - k). \tag{26}$$

We shall also see that the Dirac delta function when integrated simply zeros its argument:

$$\int_{-\infty}^{\infty} dx f(x) \delta(x - x_0) = f(x_0). \tag{27}$$

We can make use of this property to finish the calculation in Eq. (25)

$$\int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{\phi}(k) e^{ikr} = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{G}(k) \hat{q}(k) e^{ikr} \tag{28}$$

from which we conclude that

$$\hat{\phi}(k) = \hat{G}(k) \hat{q}(k). \tag{29}$$

The conclusion is that if we know the Fourier transform of the response function and of the charges, all we need to do is to multiply these together in order to obtain (the Fourier transform) of the resulting physical effect.

3. Correlation functions

In the study of natural phenomena we are often concerned with how much one part of a system influences another. Or we want to know how much of what happens at a specific moment in time, is influenced by what

has happened previously and how much the present events will influence the future. Correlation functions are used to quantify this sort of interdependence.

Let us for concreteness focus on temporal correlations. Assume $f(t)$ describes the time dependence of some observable, say the total power output from the sun. We all know that the activity of the sun fluctuates and we can ask how, typically, the energy production at time t is related to the energy production at some other time $t+T$. We are not so much interested in the specific moments at time t and $t+T$; our interest is more in how the energy production at two time instances separated by a time T is in general related as T varies. We will consider the signal $A = f(t)$, at time t, and the signal $B = f(t+T)$, at the later (or earlier) time $t+T$, as two stochastic events A and B. We then construct a measure that will be equal to zero if the stochastic process A is independent of the stochastic process B. The correlation coefficient between A and B is one such measure.

From probability theory we know that if we have two stochastic processes A and B which can assume values a and b[3], the probability that the variable A assumes the value a is denoted by $P_A(a)$ and similar $P_B(b)$ denotes the probability that the variable B assumes the value b.

The average value of A is denoted by $\langle A\rangle$ and given by

$$\langle A\rangle = \sum_a aP_A(a). \tag{30}$$

and similar for the average $\langle B\rangle$. The simultaneous event $A = a$ together with $B = b$ occurs with probability $P_{A,B}(a,b)$. The relationship between the individual probabilities $P_A(a)$ and $P_B(b)$ and the joint, or simultaneous, probability $P_{A,B}(a,b)$ can in general be anything, but if the events A and B are unrelated, i.e. independent as in the example in the footnote, we have that

$$P_{A,B}(a,b) = P_A(a)P_B(b). \tag{31}$$

Consider now the correlation coefficient defined by

$$C_{A,B} = \langle AB\rangle - \langle A\rangle\langle B\rangle. \tag{32}$$

If A and B are independent, i.e. fulfil Eq. (31) the correlation coefficient will be zero. This can be shown by use of Eq. (30) and Eq. (31):

$$\begin{aligned} C_{AB} &= \sum_a\sum_b abP_{A,}(a,b) - [\sum_a aP_A(a)][\sum_b bP_B(b)] \\ &= \sum_a\sum_b abP_A(a)P_B(b) - [\sum_a aP_A(a)][\sum_b bP_B(b)] \\ &= [\sum_a aP_A(a)][\sum_b bP_B(b)] - [\sum_a aP_A(a)][\sum_b bP_B(b)] \\ &= 0. \end{aligned} \tag{33}$$

We used the assumption concerning independence, i.e. Eq. (31) in the second step.

In general we don't kow the probabilities $P_A(a)$ or $P_B(b)$ but will have to extract them from measurements. This could, e.g., be done by repeating the measurements of the observables. Imagine that, say, A is measured N times with the outcomes a_i for $i = 1, 2, ..., N$. We will then estimate the average value of A by

$$\langle A\rangle = \frac{1}{N}\sum_{i=1}^{N} a_i. \tag{34}$$

---

[3] Perhaps it is helpful to think of the following simple example. A might be the number of eyes on a dice, i.e. $a = 1, 2, 3, 4, 5, 6$ and B could relate to throwing a coin with a -1 painted on one face and $+1$ painted on the other face, i.e. $b = -1, +1$. The probabilities are $P_A(a) = 1/6$ for all values of a and $P_B(b) = 1/2$ for all values of b. And since the two events clearly are independent the simultaneous probability distribution is $P_{A,B}(a,b) = P_A(a)P_B(b) = (1/6)(1/2)$ for all values of a and b.

To obtain the correlation coefficient between A and B empirically we would have to perform N measurements of A and N measurements of B and finally N measurements of A and B simultaneously. As an example think of trying to settle whether a person's height and distance between the ears are related to each other. We would then first determine the average height, then the average distance between ears and finally the average of the product of height times distance between ears.

The empirical estimate of the correlation coefficient obtained by this procedure is given by

$$\begin{aligned} C_{AB} &= \frac{1}{N}\sum_{i=1}^{N} a_i b_i - [\frac{1}{N}\sum_{i=1}^{N} a_i][\frac{1}{N}\sum_{i=1}^{N} b_i] \\ &= \frac{1}{N}\sum_{i=1}^{N}[a_i - \langle A\rangle][b_i - \langle B\rangle]. \end{aligned} \tag{35}$$

In the second equality we made use of Eq. (34) for both A and B.

We now return to the temporal signal $f(t)$. We measure the signal again and again at times separated by T time units. We think of the measurements of f at times $t = 1, 2, 3, ...$ as a_i and the measurements of F at the time $t + T$ as b_i. We will now use Eq. (35) to obtain an estimate of the correlation between the two measurements for the signal separated by T time units. To make life simple we will *neglect* the $1/N$ factors in Eq. (35). A justification for this is that we are interested in the functional dependence of the correlations on the time interval T and not so much interested in the actual specific value of the correlation coefficient. Since the correlation coefficient will depend on T we talk about the correlation function. Moreover, since we are correlating the signal with itself we talk about the autocorrelation function given by:

$$\begin{aligned} C(T) &= \sum_t [f(t) - \langle f(t)\rangle][f(t+T) - \langle f(t+T)\rangle] \\ &= \int dt [f(t) - \langle f(t)\rangle][f(t+T) - \langle f(t+T)\rangle] \\ &= \int dt [f(t) - \langle f(t)\rangle][f(t+T) - \langle f(t)\rangle]. \end{aligned} \tag{36}$$

In the last equality we made use of the fact that the average of value $f(t)$ and $f(t + T)$ are identical.

It is hopefully by now clear why the autocorrelation function is an important object for the study of memory effects or causality effects in a signal. Much time is spend experimentally and theoretically on the determination of correlation functions or the equivalent *power spectrum* whichever is most easy to determine. The power spectrum of the signal $f(t)$ is defined as

$$S_f(\omega) = |\hat{f}(\omega)|^2. \tag{37}$$

That is the square of absolute value of the Fourier transform of the signal. The reason it doesn't matter which of the two $C(T)$ of $S_f(\omega)$ one gets hold of is that, as we show in the lectures, they are related in the following simple way

$$S_f(\omega) = \hat{C}(\omega). \tag{38}$$

In words: the power spectrum is equal to the Fourier transform of the correlation function.

IV. LINEAR VECTOR SPACES

The prototype space is $\mathbb{R}^3$, that is

$$\mathbb{R}^3 = \{(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) | \mathrm{x_i} \in \mathbb{R}, \mathrm{i} = 1, 2, 3\}. \tag{39}$$

A basis in $\mathbb{R}^3$ is

$$\mathbf{e}_1 = (1,0,0),\ \mathbf{e}_2 = (0,1,0),\ \mathbf{e}_3 = (0,0,1). \tag{40}$$

since an arbitrary vector $\mathbf{v} = (x_1, x_2, x_3)$ in $\mathbb{R}^3$ can be written as a linear combination

$$\mathbf{v} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3. \tag{41}$$

The two vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ *span* a subspace of $\mathbb{R}^3$:

$$\text{span}(\mathbf{e}_1, \mathbf{e}_2) = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}. \tag{42}$$

This subspace is equivalent to the two dimensional plane.

From the theory of Fourier series we know that the set of functions

$$\mathbf{e}_0(x) = 1,\ \mathbf{e}_1(x) = \cos x,\ \mathbf{e}_2(x) = \cos 2x, ... \tag{43}$$

together with the functions

$$\tilde{\mathbf{e}}_1(x) = \sin x,\ \tilde{\mathbf{e}}_2(x) = \sin 2x,\ \tilde{\mathbf{e}}_3(x) = \sin 3x, ... \tag{44}$$

constitute a basis in $C^0([-\pi,\pi],\mathbb{R})$ since any continuous function $f : [-\pi,\pi] \to \mathbb{R}$ can be written as

$$f(x) = a_0\mathbf{e}_0(x) + \sum_{n=0}^{\infty}[a_n\mathbf{e}_n(x) + b_n\mathbf{e}_n(x)]. \tag{45}$$

Consider the scalar product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)dx. \tag{46}$$

By use of this scalar product one can show that the basis

$$\mathbf{e}_0(x), \mathbf{e}_1(x), ..., \tilde{\mathbf{e}}_1(x), \tilde{\mathbf{e}}_2(x), ... \tag{47}$$

consists of orthogonal (but not normalised) vectors. For example

$$\langle \mathbf{e}_n, \mathbf{e}_m\rangle = \int_{-\pi}^{\pi} \cos(nx)\cos(mx)dx \propto \delta_{n,m}. \tag{48}$$

Normed vector spaces
We are used to being able to measure the distance between two points $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$ in $\mathbb{R}^3$ by use of Pythagoras' theorem, namely as $[(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2)]^{1/2}$. This is of course the same as the length of the vector $\vec{XY} = (y_1 - x_1, y_2 - x_2, y_3 - x_3)$ or in other words the distance between point X and point Y is equal to the norm of the vector $\vec{XY}$:

$$\|\vec{XY}\| = [(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2)]^{1/2}. \tag{49}$$

We introduce generalised norms. Consider the following example. The subspace C^0_{odd} of $C^0([-\pi,\pi],\mathbb{R})$ consisting of odd functions. We know from the theory of Fourier series that these functions can be written as

$$f(x) = \sum_{n=1}^{\infty} b_n \sin(nx), \tag{50}$$

and we can use

$$\|f\| = \sum_{n=1}^{\infty} |b_n|^2 \tag{51}$$

as a norm. I.e., we can consider the normed vector space

$$C^0_{odd} = \text{span}\{\sin x, \sin 2x, \sin 3x, ...\}. \tag{52}$$

In analogy to *open* and *closed* intervals, as for instance $]0,1[$ compared to $[0,1]$, we can ask if everything, we can converge towards from within C^0_{odd}, is also included in C^0_{odd}.

We formulate this in terms of whether a Cauchy sequence (a sequence for which the tail is confined within a region of ever decreasing size) within the considered space is convergent to a limit which is also an element of the same space. If all Cauchy sequences in a space are convergent to a limit within that space then we say that the space is *complete* (or we call it a Banach space, same thing). I.e. a Banach (or complete) space corresponds to the closed interval $[0,1]$ whereas an incomplete space corresponds to the open interval $]0,1[$.

Now back to the space C^0_{odd}. We consider the sequence

$$g_n(x) = \sum_{m=1}^{n} b_n \sin(mx) \tag{53}$$

We want to construct $g_n(x) \in C^0_{odd}$ such that they converge to a discontinuous function, i.e. a function outside C^0_{odd}. Choose

$$g(x) = \begin{cases} 1 & \text{if } 0 < x < \pi \\ 0 & \text{if } x = 0 \\ -1 & \text{if } -\pi < x < 0 \end{cases} \tag{54}$$

Using the theory of Fourier series we can check that the sequence of functions $g_n(x)$ will converge towards $g(x)$ if we choose

$$b_m = \frac{1}{m}(1 - (-1)^m). \tag{55}$$

The sequence g_n is a Cauchy sequence within C^0_{odd} using the norm in Eq. (51). This we see from the fact that the norm

$$\|g_{n_1} - g_{n_2}\| = \sum_{m=n_1}^{n_2} |\frac{1}{m}(1 - (-1)^m)|^2 \tag{56}$$

(assume $n_2 > n_1$) can be made arbitrarily small by choosing n_1 sufficiently large. Hence, we have that g_n is a Cauchy sequence within C^0_{odd} but the sequence does not converge to a limit within C^0_{odd}. The reason for this is that the function towards which g_n converges is $g(x)$ in Eq. (54 is discontinuous and therefore not a member of C^0_{ood}. Conclusion: C^0_{odd} is *not* a complete space.

Example
Show that any Cauchy sequence $a_n \in [0,1]$ is convergent within $[0,1]$.
Proof:
Since a_n is a Cauchy sequence, we have (see Sec. IX for notation)

$$\forall \epsilon \; \exists n_0 \mid n, m > n_0 \Rightarrow |a_n - a_m| < \epsilon. \tag{57}$$

(we use $\|x\| = |x|$, the absolute value) and we want to show that an $a \in [0,1]$ exists such that

$$\forall \epsilon > 0 \; \exists n_0 \mid n > n_0 \Rightarrow |a_n - a| < \epsilon. \tag{58}$$

To show that Eq. (57) implies Eq. (58) we make use of the properties of an interval of real numbers. Namely the fact that the interval $[0,1]$ contains *all* numbers between 0 and 1 *including* 0 and 1. From Eq. (57) we have that the tail of the sequence a_n is limited to a smaller ad smaller subinterval of $[0,1]$ the larger n becomes: According to Eq. (57) we can choose $\epsilon_1 > 0$ and a corresponding n_1 such that $|a_n - a_m| < \epsilon_1$ when $n, m > n_1$. Now choose $p_1 = n_1 + 1$. Define the interval $I_1 = [a_{p_1} - \epsilon_1, a_{p_1} + \epsilon_1]$, we have $a_n \in I_1$ for all $n > p_1$. Then choose $\epsilon_2 = \epsilon_1/2$ and a corresponding n_2 such that $|a_n - a_m| < \epsilon_2$ when $n, m > n_2$. Let $p_2 = n_2 + 1$ and define the interval $I_2 = [a_{p_2} - \epsilon_2, a_{p_2} + \epsilon_2]$. Again $a_n \in I_2$ for all $n > p_2$. We continue this

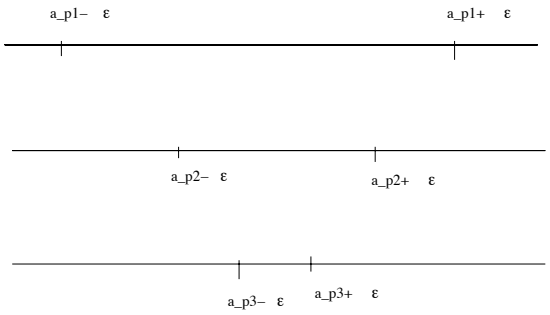

FIG. 5 The limit is confined to a smaller and smaller interval

procedure, see Fig. 5 forever and construct a sequence of intervals within intervals of shorter and shorter length:

$$[0,1] \supset I_1 \supset I_2 \supset I_3 \supset \cdots. \tag{59}$$

Since the length of the interval I_k decreases to zero and $a_n \in I_k$ for all $n > p_k$, the sequence a_n *must* converge to a number a larger than 0 and smaller than 1, since $[0,1]$ is an interval we know that $a \in [0,1]$ and hence that the sequence is strongly convergent to a limit within the set $[0,1]$. This completes the proof.

Scalar products and projections
For ordinary vectors in $\mathbb{R}^2$ or in $\mathbb{R}^3$ we use the scalar product to find the projection $\mathbf{b}_{\|}$ of a vector $\mathbf{b}$ along a vector $\mathbf{a}$. We have

$$\mathbf{b}_{\|} = (\mathbf{b}\cdot\mathbf{a})\frac{\mathbf{a}}{|\mathbf{a}|}. \tag{60}$$

Consider quantum mechanics. Assume a particle is in a quantum state described by the normalised wave function ψ. In order to calculate the probability that a measurement of the energy will yield the value $E = E_n$, for some particular energy level E_n, we need to know what the component of ψ is along the normalised eigenstate ϕ_n corresponding to the eigenvalue E_n. I.e., we must calculate the scalar product $\langle\psi,\phi_n\rangle$ to obtain the probability $|\langle\psi,\phi_n\rangle|^2$ for measuring E_n.

ϵ and δ expressions
As a way to get used to the $\epsilon-\delta$ formalism we list here a few examples:

Continuity
A function $f(x)$ is said to be continuous at a point x_0 if and only if

$$\forall\epsilon > 0\ \exists\delta > 0 \mid |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon. \tag{61}$$

Convergence
A sequence of numbers x_n is said to converge to the point x if and only if

$$\forall\epsilon > 0\ \exists n_0 \mid n > n_0 \Rightarrow |x - x_n| < \epsilon. \tag{62}$$

Example:
The sequence $a_n = 1 + 1/n$ converges to 1.
Proof: given $\epsilon > 0$ we can choose n_0 as the smallest integer larger than $1/\epsilon$ to obtain that

$$n > n_0 \Rightarrow |1 - x_n| < \epsilon. \tag{63}$$

V. CALCULUS OF VARIATION

How do we find the most efficient way of doing something. We first need to decide exactly what type of efficiency we are interested in. We might want to calculate the fastest way of getting from one point to

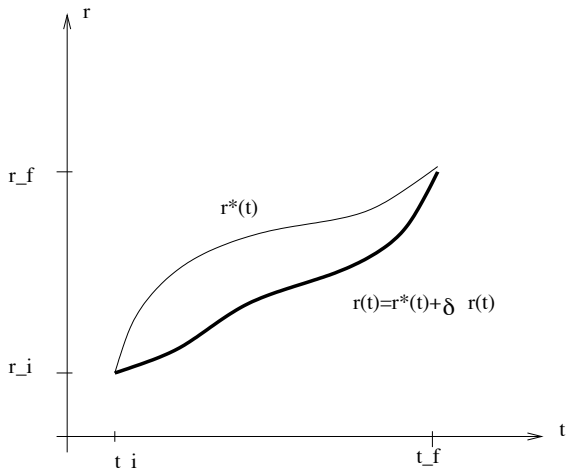

FIG. 6 Variations about a trajectory

another. Or we might be interested in which shape of a string suspended at its endpoints minimises its total potential energy.

A *very* important example is to find the physical path followed by a particle as it travels from an initial position $\mathbf{r}_i$ to a final position $\mathbf{r}_f$ through a potential field $U(\mathbf{r})$. Sir William Rowan Hamilton (1805-1865) realised that the particle travels in such a way as to minimise the difference between the kinetic energy and the potential energy, i.e. a path that minimises the time integral of the Lagrangian. At every time instance you monitor your kinetic energy, $T[\dot{\mathbf{r}}(t)]$, as well as your potential energy, $U[\mathbf{r}(t)]$). You calculate the difference $L(t) = T(t) - U(t)$ and add it all up to obtain what is called the action

$$S = \int_{t_i}^{t_f} L[\mathbf{r}, \dot{\mathbf{r}}]dt. \tag{64}$$

In order to minimise S one should spend long time in regions where the potential is large and short times where the potential assumes small values. Hence a small action is obtained if we can design our trajectory $\mathbf{r}(t)$ and velocity $\mathbf{v}(t) = d\mathbf{r}/dt$ such that we move with low velocity through regions of large potential energy. The question is how do we determine $\mathbf{r}(t)$ from $U(\mathbf{r})$ in order to obtain the smallest possible value of S?

We need to establish an equation for $\mathbf{r}(t)$ from the requirement that S is minimal for that particular function $\mathbf{r}(t)$. For ease of notation we confine ourselves to one dimension[4], and specify the position by the scalar $r(t)$. Imagine that we somehow had found the trajectory which minimises $S[r(t)]$ and denote this specific functional relation between position and time by $r^*(t)$. We will then have that for all other choices $r(t)$ of ways to get from r_i to r_f the action of $r^*(t)$ is smaller than the action of $r(t)$, or

$$S[r^*(t)] < S[r(t)]. \tag{65}$$

Imagine we choose another trajectory $r(t)$ which only slightly differs from $r^*(t)$. This is easiest illustrated by plotting the position as function of time as in Fig. 6. The difference between $r^*(t)$ and $r(t)$ we denote by $\delta r(t)$, i.e,

$$r(t) = r^*(t) + \delta r(t). \tag{66}$$

The difference between $S[r(t)]$ and $S[r^*(t)]$ will be small if $\delta r(t)$ is small. In fact we can imagine to write the difference between $S[r(t)]$ and $S = [r^*(t)]$ as an expansion in deviation $\delta r(t)$ between the two paths. We will the have something that schematically would look like

$$\begin{aligned} \delta S &= S[r(t)] - S[r^*(t)] \\ &= a_1 \delta r(t) + a_2 [\delta r(t)]^2 + \cdots . \end{aligned} \tag{67}$$

---

[4] This is really not an essential restriction, but we avoid to have to think about the three dimensional vector $\mathbf{r}$

The expansion in Eq. (67) is to be thought as being analogous to the Taylor expansion of a function $f(x)$ about a specific point x^*:

$$f(x) - f(x^*) = a_1 x + a_2 x^2 + \cdots. \tag{68}$$

We know how to calculate the coefficients a_i in Eq. (68) from the standard Taylor expansion about x^*. We also know that *if* the point x^* is a minimum for $f(x)$ then the linear term in Eq. (68) will be absent, i.e. $a_1 = 0$. Put in the usual way, we determine the minima [5] x^* of a function $f(x)$ as the solution to the equation

$$\frac{df}{dx} = 0. \tag{69}$$

We are going to determine the trajectory of least action $r^*(t)$ in a similar way. We need to find a way to express the coefficient a_1 in Eq. (67) in terms of the trajectory, $a_1 = a_1[r(t)]$ and then determine $r^*(t)$ from the equation

$$a_1[r(t)] = 0. \tag{70}$$

This can be done by use of the formula for Taylor expansion of a real function of more than a single variable. We have in general for a function $f(x, y)$ of two variables

$$f(x,y) = f(x^*, y^*) + \frac{\partial f}{\partial x}_{|(x^*,y^*)}[x - x^*] + \frac{\partial f}{\partial y}_{|(x^*,y^*)}[y - y^*] + \cdots \tag{71}$$

We now apply this Taylor expansion to the integrand $L[r, \dot{r}]$ in Eq. (64) with $r(t)$ in the form given in Eq. (66). We expand $L[r(t), \dot{r}(t)]$ for each t. Imagine to keep t fixed while we perform the expansion.

$$\begin{aligned} L[r(t), \dot{r}(t)] &= L[r^*(t) + \delta r(t), \dot{r}^*(t) + \delta r(t)] \\ &= L[r^*(t), \dot{r}^*(t)] + \left(\frac{\partial L}{\partial r}\right)_* \delta r(t) + \left(\frac{\partial L}{\partial \dot{r}}\right)_* \delta \dot{r}(t). \end{aligned} \tag{72}$$

The notation $()_*$ indicates that the derivatives of the Lagrangian is to be calculated at the unperturbed position and velocity $r^*(t)$ and $\dot{r}^*(t)$ similar to the way things are done in Eq. (71). Next we notice that

$$\begin{aligned} \delta\dot{r}(t) &= \delta\frac{dr}{dt}(t) \\ &= \frac{dr^*}{dt}(t) - \frac{dr}{dt}(t) \\ &= \frac{d}{dt}[r^*(t) - r(t)] \\ &= \frac{d}{dt}\delta r(t). \end{aligned} \tag{73}$$

We can now use Eq. (72) to obtain

$$\begin{aligned} \delta S &= \int_{t_i}^{t_f} \{L[r^*(t) + \delta r(t), \dot{r}^*(t) + \delta r(t)] - L[r*(t), \dot{r}^*(t)]\} \\ &= \int_{t_i}^{t_f} \left\{ \left(\frac{\partial L}{\partial r}\right)_* \delta r(t) + \left(\frac{\partial L}{\partial \dot{r}}\right)_* \delta \dot{r}(t) \right\} dt. \end{aligned} \tag{74}$$

The relation in Eq. (73) allows us to perform integration by parts on the last term in Eq. (74). Namely,

$$\int_{t_i}^{t_f} \left\{ \left(\frac{\partial L}{\partial \dot{r}}\right)_* \delta \dot{r}(t) \right\} dt. = \left[\left(\frac{\partial L}{\partial \dot{r}}\right)_* \delta r(t) \right]_{t_i}^{t_f} - \int_{t_i}^{t_f} \left\{ \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}}\right)_* \right\} \delta r(t) dt. \tag{75}$$

---

[5] Extremal points in general.

We assume, as in the Figure, that $\delta r(t_i) = \delta r(t_f) = 0$ and arrive at the following expression

$$\delta S = \int_{t_i}^{t_f} \left\{ \left(\frac{\partial L}{\partial r} \right)_* - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right)_* \right\} \delta r(t) dt. \tag{76}$$

Since δS must be equal to zero for all perturbations $\delta r(t)$ if $r^*(t)$ is an extremal path we conclude that the integrand in Eq. (76) must vanish:

$$\left(\frac{\partial L}{\partial r} \right)_* - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right)_* = 0. \tag{77}$$

This equation is called the Euler-Lagrange equation for the action in Eq. (64). In general whenever one wants to minimise or maximise a certain cost function, the starting point is to determine the cost function and then derive the Euler-Lagrange equation. All the fundamental equations of physics can be obtained from the action, i.e. the time integral of the Lagrangian. The Maxwell equations for electrodynamics can be derived in this way, the Einstein equations for gravitation can be obtained this way and so can the equations for QCD and for the electro-weak forces. The Euler-Lagrange Eq. (77) is in fact equivalent to Newton's 2nd law when the Lagrangian in Eq. (64) is the usual mechanical Lagrangian:

$$L[r(t), \dot{r}(t)] = \frac{1}{2} m \dot{r}^2 - U(r), \tag{78}$$

where m denotes the mass of the particle and $U(r)$ the potential energy. We immediately obtain Newton's law from Eq. (77) by noticing that for $L[r, \dot{r}]$ of the form in Eq. (78) we have

$$\begin{aligned} \frac{\partial L}{\partial r} &= -\frac{\partial U(r)}{\partial r} \\ \frac{\partial L}{\partial \dot{r}} &= m\dot{r}, \end{aligned} \tag{79}$$

which upon substitution into Eq. (77) leads to

$$-\frac{\partial U(r)}{\partial r} - \frac{d}{dt} m\dot{r} = 0. \tag{80}$$

VI. SUFFIX NOTATION

When dealing with vectors and matrices and tensors, calculations easily become very cumbersome if one trys to bulldoze one's way through in coordinate notation without an efficient notation. Suffix notation is meant to supply this efficient and economical notation. After a bit of practise the machinery will hopefully become a cherished device - especially if the study of electromagnetism or other field theories are undertaken.

As often in physics we start out with a contribution by Albert Einstein, namely the so called repeated index convention. As an example consider the familiar scalar product between the two vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, both of three dimensions:

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 = \sum_{i=1}^{3} a_i b_i \equiv a_i b_i. \tag{81}$$

Einstein's contribution consists in ignoring the summation sign and simply as of convention assume summation over any index that is repeated. Einstein may have made more important contributions to our understanding of the univers, but hardly any more straight forward.

There are two other bits of notation that turn out to be very handy when analysing vector identities. The first is the Kronecker delta (after the German mathematician Leopold Kronecker 1823-1891):

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{82}$$

One notice that $\delta_{i,j}$ can be thought of as a discrete version of the Dirac delta function $\delta(x-x_0)$ in the sense that also $\delta(x-x_0)=0$ for all values of x and x_0 except when $x=x_0$. Another way of think about δ_{ij} is as the elements of the unit matrix

$$\mathbf{I}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\{\delta_{ij}\}. \tag{83}$$

The permutation symbol is slightly more involded to define

$$\epsilon_{ijk}=\begin{cases}0 & \text{if any two } i,j,k \text{ are the same}\\ 1 & \text{if } i,j,k \text{ is a cyclic permutation of 1,2,3 (an even permutation)}\\ -1 & \text{if } i,j,k \text{ is an anti-cyclic permutation (an odd permutation)}\end{cases} \tag{84}$$

One often finds great use of ϵ_{ijk} in relation to vector cross products and determinants of matrices.

To get some feeling of how these symbols are manipulated we will present the proof of the following identity. The proof should definitely be read with a pencil in hand. The best way of absorbing the idea is to simply copy the proof bit by bit.

$$\epsilon_{ijk}\epsilon_{pqk}=\delta_{ip}\delta_{jq}-\delta_{iq}\delta_{jp}. \tag{85}$$

Proof:
STEP 1.
By a bit of perseverance one can show that for any matrix $\mathbf{A}=\{a_{ij}\}$

$$\det\mathbf{A}=\epsilon_{ijk}a_{1i}a_{2j}a_{3k}. \tag{86}$$

This is done in the following way.
Proof Step 1:
We note that in the sum $\epsilon_{ijk}a_{1i}a_{2j}a_{3k}$ there are $3\times3\times3=27$ terms. However only 6 are non-zero, namely:

$$\begin{aligned}\epsilon_{ijk}a_{1i}a_{2j}a_{3k} &= a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}\\ &\quad -a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}\\ &= a_{11}\begin{vmatrix}a_{22}&a_{23}\\a_{32}&a_{33}\end{vmatrix}-a_{12}\begin{vmatrix}a_{21}&a_{23}\\a_{31}&a_{33}\end{vmatrix}+a_{13}\begin{vmatrix}a_{21}&a_{22}\\a_{31}&a_{32}\end{vmatrix} \qquad (87)\\ &= \begin{vmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{vmatrix}\\ &= \det\mathbf{A} \qquad (88)\end{aligned}$$

STEP 2.
It is not very difficult to generalise this result to the more general

$$\epsilon_{ijk}a_{pi}a_{qj}a_{rk}=\det\mathbf{A}\epsilon_{pqr}. \tag{89}$$

Proof Step 2:
One simply need to distinguish between the three different cases:

i) If pqr is a cyclic permutation of 123, say $pqr=231$, then

$$\begin{aligned}\epsilon_{ijk}a_{2i}a_{3j}a_{1k} &= \epsilon_{ijk}a_{1k}a_{2i}a_{3j}\\ &= \epsilon_{kij}a_{1k}a_{2i}a_{3j}\\ &= \det\mathbf{A}. \qquad (90)\end{aligned}$$

where the last equality follows from Step 1 above.

ii) If pqr is anti-cyclic, say $pqr = 132$, then

$$\begin{aligned} \epsilon_{ijk}a_{1i}a_{3j}a_{2k} &= \epsilon_{ijk}a_{1i}a_{2k}a_{3j} \\ &= -\epsilon_{ikj}a_{1i}a_{2k}a_{3j} \\ &= -\mathrm{det}\mathbf{A}. \end{aligned} \tag{91}$$

iii) If pqr are not all different, say $pqr = 112$,

$$\begin{aligned} \epsilon_{ijk}a_{1i}a_{1j}a_{2k} &= -\epsilon_{jik}a_{1j}a_{1i}a_{2k} \\ &= -\epsilon_{ijk}a_{1i}a_{1j}a_{2k}, \end{aligned} \tag{92}$$

where the last equality follows from the renaming $i \mapsto j$ and $j \mapsto i$. Since from $x = -x$ follows that $x = 0$ we conclude that $\epsilon_{ijk}a_{1i}a_{1j}a_{2k} = 0$.

STEP 3.
We are now ready for the final piece of the proof Eq. (85). A pencil is needed to digest the following. According to Eq. (89) we have

$$\begin{aligned} \epsilon_{ijk}\delta_{pi}\delta_{qj}\delta_{rk} &= \epsilon_{pqr}\mathrm{det}\delta \\ &= \epsilon_{pqr}\mathrm{det}\mathbf{I} \\ &= \epsilon_{pqr}. \end{aligned} \tag{93}$$

I.e. we have $\epsilon_{pqr} = \epsilon_{ijk}\delta_{pi}\delta_{qj}\delta_{rk}$. Moreover, from Eq. (86) we have second equality below, namely

$$\begin{aligned} \epsilon_{pqr} &= \epsilon_{ijk}\delta_{pi}\delta_{qj}\delta_{rk} \\ &= \begin{vmatrix} \delta_{p1} & \delta_{p2} & \delta_{p3} \\ \delta_{q1} & \delta_{q2} & \delta_{q3} \\ \delta_{r1} & \delta_{r2} & \delta_{r3}. \end{vmatrix} \end{aligned} \tag{94}$$

Now we make use of Eq. (94) twice to get

$$\begin{aligned} \epsilon_{ijk}\epsilon_{pqr} &= \begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} \begin{vmatrix} \delta_{p1} & \delta_{p2} & \delta_{p3} \\ \delta_{q1} & \delta_{q2} & \delta_{q3} \\ \delta_{r1} & \delta_{r2} & \delta_{r3} \end{vmatrix} \\ &= \begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} \begin{vmatrix} \delta_{1p} & \delta_{1q} & \delta_{1r} \\ \delta_{2p} & \delta_{2q} & \delta_{2r} \\ \delta_{3p} & \delta_{3q} & \delta_{3r} \end{vmatrix} \\ &= \begin{vmatrix} \delta_{il}\delta_{lp} & \delta_{il}\delta_{lq} & \delta_{il}\delta_{lr} \\ \delta_{jl}\delta_{lp} & \delta_{jl}\delta_{lq} & \delta_{jl}\delta_{lr} \\ \delta_{kl}\delta_{lp} & \delta_{kl}\delta_{lq} & \delta_{kl}\delta_{lr}. \end{vmatrix} \end{aligned} \tag{95}$$

The last equality follows from $\mathrm{det}\mathbf{C} = \mathrm{det}\mathbf{A}\mathrm{det}\mathbf{B}$ when $\mathbf{C} = \mathbf{AB}$. We can perform the sums over the repeated index l and obtain

$$\epsilon_{ijk}\epsilon_{pqr} = \begin{vmatrix} \delta_{ip} & \delta_{iq} & \delta_{ir} \\ \delta_{jp} & \delta_{jq} & \delta_{jr} \\ \delta_{kp} & \delta_{kq} & \delta_{kr} \end{vmatrix}. \tag{96}$$

Finaly put $k = r$ and as usual assume summation over the repeated index k:

$$\epsilon_{ijk}\epsilon_{pqr} = \begin{vmatrix} \delta_{ip} & \delta_{iq} & \delta_{ik} \\ \delta_{jp} & \delta_{jq} & \delta_{jk} \\ \delta_{kp} & \delta_{kq} & \delta_{kk} \end{vmatrix}$$

$$
\begin{aligned}
&= \begin{vmatrix} \delta_{ip} & \delta_{iq} & \delta_{ik} \\ \delta_{jp} & \delta_{jq} & \delta_{jk} \\ \delta_{kp} & \delta_{kq} & 3 \end{vmatrix} \\
&= \delta_{ip}\delta_{jq}3 + \delta_{iq}\delta_{jk}\delta_{kp} + \delta_{jp}\delta_{kq}\delta_{ik} \\
&\quad -\delta_{ik}\delta_{jq}\delta_{kp} - \delta_{iq}\delta_{jp}3 - \delta_{jk}\delta_{kq}\delta_{ip} \\
&= 3(\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}) + \delta_{iq}\delta_{jp} + \\
&\quad \delta_{jp}\delta_{iq} - \delta_{ip}\delta_{jq} - \delta_{ip}\delta_{jq} \\
&= \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}. \qquad (97)
\end{aligned}
$$

Which completes the proof of Eq. (85).

To become better acquainted with this sections vector notation it may be useful to look through Jefferyes and Jeffreys book Chap. 2 (See p. 2 for reference).

VII. TRANSFORMATION UNDER ROTATION

The very important point to keep in mind is that a vector is a **geometrical** object: a direction in space. From this simple fact follow all the transformation properties discussed in the lectures. Most of which are likely to be known from A-levels.

A look at Chap. 3 in Jeffreys and Jeffreys book (See p. 2 for reference) would probably help one to be come acquainted with tensors.

VIII. NUMERICAL ANALYSIS

See separate set of notes.

IX. SOME MATHEMATICAL NOTATION

Example: Continuity of a fuction:

$$\forall \epsilon > 0 \exists \delta > 0 \mid |x_0 - x| < \delta \Rightarrow |f(x_0) - f(x)| < \epsilon$$

Translations: $\forall$ means "for all"
$\exists$ means "there exists"
$\mid$ means "such that"

Example: Set theory:
If $S_1 = \{1, 2, 3, 4\}$ and $S_2 = \{2, 3\}$ then $S_2 \subset S_1$ since $2 \in S_2$ and $3 \in S_2$ are also in S_1.

Translations: $\{a, b\}$ denotes the set consisting of the elements a and b.
$a \in S$ means that a is an element in the set S.

$S_2 \subset S_1$ means that the elements of the set S_2 are all members in S_1
and that S_1 contains elements that are not found in S_2.
If we write $S_2 \subseteq S_2$ the two sets might be identical.

More set theory:

$[0, 1]$ denotes the **closed** interval $0 \leq x \leq 1$.
$]0, 1[$ denotes the **open** interval $0 < x < 1$.
$[0, 1[$ denotes the half open/closed interval $0 \leq x < 1$.