

MATHEMATICS INTRO FOR PHYSICS

YORK UNIVERSITY

1. SCALARS AND VECTORS

1.1. **Preliminaries.** Many physics variables use the real-number system to provide a quantitative description. Temperature is one such example: on the Celsius scale we encounter positive and negative real numbers, which allow ordering: we can always compare two temperature readings and state whether one is higher than the other (or lower, or equal). A complete description requires specification of a unit: this distinguishes physics from mathematics. When we adopt the international unit system SI (meters, kilograms, seconds, Ampere) we can avoid writing down the unit. However, when doing so we should make clear that we are using SI to avoid confusion.

In a physics measurement one is limited by the accuracy of equipment. Thus, an answer like $x = 4$ m has to be treated carefully. In mathematics the implication would be that x equals to the integer number 4 (the integers are embedded in the real numbers). No physics measurement can be as precise as that. Therefore, we quote results to a certain number of significant digits. A meaningful physics statement can be: $x = 4.00$ m - which would imply that we are certain about the result being above $x = 3.995$ m and below $x = 4.005$ m. When calculating answers to physics problems in this course pay attention to how many significant digits are provided by the input, use one more digit for intermediate calculations, and quote your final answer to no more than this one additional digit. Ideally, you will round the result to the correct number of significant digits.

Some physics variables require two, three, (or perhaps more) real numbers to specify them. From navigation or geography (GPS) you know that specifying your location on the surface of the earth requires a longitude and a latitude. In three-space we locate objects by specifying their (x, y, z) coordinates. Such objects are called vectors. In this course you will make heavy use of vectors in two dimensions, and also to a lesser extent in three. In upper-year courses you will encounter higher-dimensional vector spaces. Some physics variables are represented by matrices, i.e., n -by- n arrays of real numbers - again, this you will encounter in second year and up.

The next section will explain how to deal with objects represented by vectors. Later an extension of calculus ideas will be provided to show how they apply to vectorial quantities. This will be needed from the start, since position, velocity, and acceleration are three-vectors, and they are related by calculus operations.

The final section will introduce integration as or area-under-a-curve measurement. This we need, since many physics laws are based on the area under a curve representing one of the physics variables. The power of physics is to understand these relationships for relatively simple systems. Modern approaches to many disciplines (from biology, kinesiology to economics) are driven by understanding laws that are built on these principles.

1.2. Unit vectors. Vectors can be specified by two different methods. One, more general, method is to specify them by their length, and by their orientation in space. When one adopts this method it is clear that they can be moved around in space and remain the same. The second method depends on a choice of coordinate system. If we defines orientations that specify what we mean by horizontal, vertical, and depth displacements (such as the edges of a rectangular room we find ourselves in), then we can represent vectors by a triple of numbers:

$$\vec{r} = (x, y, z) . \quad (1.1)$$

While this representation depends on a choice of what the (x, y, z) axes are (and therefore appears to be less general), we nevertheless use it, since it is easy to work with. In physics one prefers the following notation: one defines the axes of the coordinate system by unit-length vectors $\hat{i}, \hat{j}, \hat{k}$ and uses vector addition to assemble the position vector according to:

$$\vec{r} = x\hat{i} + y\hat{j} + z\hat{k} . \quad (1.2)$$

We can work in the plane by assuming $z \equiv 0$, and illustrate in Fig (1) what we mean by the three objects in this case, namely the relevant unit vectors \hat{i}, \hat{j} , and the components of the position vector \vec{r} along the axes (x along \hat{i} , and y along \hat{j}).

1.3. Vector addition. The representation of the two-dimensional (2d) position vector as an addition of contributions along \hat{i} and \hat{j} defined a mathematical operation that has both an algebraic definition and a geometric interpretation. The algebraic definition of vector addition simply states that we add the x - and y -components separately for any two vectors $\vec{r}_1 = x_1\hat{i} + y_1\hat{j}$, $\vec{r}_2 = x_2\hat{i} + y_2\hat{j}$:

$$\vec{r}_1 + \vec{r}_2 \equiv (x_1 + x_2)\hat{i} + (y_1 + y_2)\hat{j} . \quad (1.3)$$

The geometric realization of vector addition consists of forming a parallelogram by moving the arrow representing the second vector such that its tail coincides with the tip of the first vector arrow. The resulting vector is given by the arrow connecting the tail of the first with the arrow of the second vector. The order of which vector to start with doesn't matter, reversing the order of the geometric construction completes the parallelogram. Note that this geometric interpretation consolidates the notion that vectors can be moved in space: they really are defined by their length (magnitude) and direction.

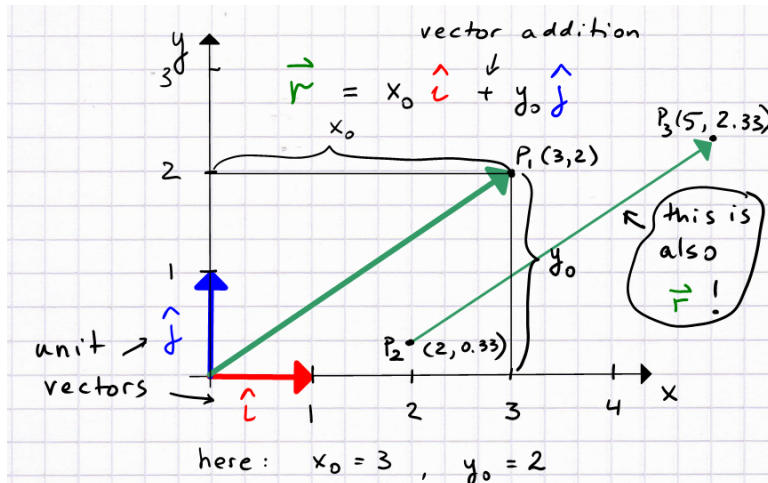


FIGURE 1. A vector in the plane is resolved into x - and y -components using the \hat{i} and \hat{j} unit vectors. The tip of the vector is at $P_1(3,2)$ when the tail is put at the origin. If the tail was put at $P_2(2,0.33)$, then the *same* vector \vec{r} would be represented by the shifted arrow ending at $P_3(5,2.33)$.

1.4. **Vector magnitude.** The Cartesian coordinate system defined by $\hat{i}, \hat{j}, \hat{k}$ has perpendicular axes. The Pythagorean theorem allows us to calculate the length of a vector given by its components such as in Eq. (1.2). In the plane this is very obvious from the triangle:

$$r = \sqrt{x^2 + y^2} . \tag{1.4}$$

Note that r is a (positive) scalar. We need to be careful in our notation to distinguish between the vector and its magnitude. A physics example: a car is moving with a speed of 96 km/h. This statement does not say which direction the car moves in! The statement: a car moves with velocity

$$\vec{v} = 3.0\hat{i} + 2.0\hat{j} \text{ m/s} \tag{1.5}$$

states specifically the direction of motion. This is a vectorial statement, and provides a complete understanding of the motion. If this velocity vector does not change with time, we can understand the motion of the car by tracking the change in the position and applying this velocity vector for each instant in time.

The speed of the car which moves with the velocity (vector) given by Eq. (1.5) is obtained by applying the magnitude calculation. We denote the components of the velocity vector as v_x and v_y , i.e., $\vec{v} = v_x\hat{i} + v_y\hat{j}$, and compute in analogy with Eq. (1.4):

$$v = \sqrt{v_x^2 + v_y^2} . \tag{1.6}$$

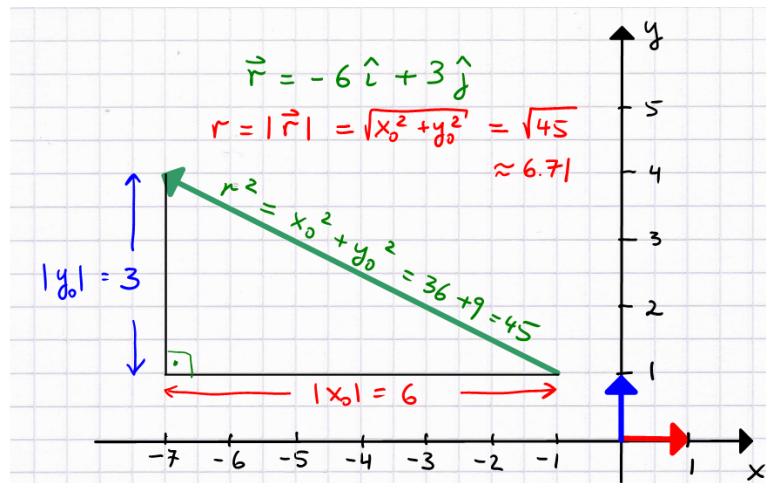


FIGURE 2. The magnitude (or length) of a vector is calculated from its Cartesian components using the Pythagorean theorem.

For our example (1.5) the speed is $v = 3.6$ m/s. For intermediate calculations you should keep an additional digit, i.e., it would be prudent to record $v = 3.61$ m/s. In fact, the on-line problem-solving environment used in our courses (WebAssign) assumes that final results are answered with one extra digit (it is not coded properly for significant digits, and works with a small-percentage window around the ‘exact’ answer to mark your result - sometimes the correctly rounded answer will fall outside this acceptance window).

1.5. Dot product. Note that you can compare two vectors by their magnitudes. This is possible, since the magnitudes are real (non-negative) numbers, and these numbers are ordered. The vectors themselves cannot be ordered. You can have for two vectors \vec{r}_1 and \vec{r}_2 the relation $x_1 > x_2$ and $y_1 < y_2$, which means that comparing the two vectors directly is meaningless. You can compare their lengths, however. A number of physics laws involving vectorial quantities require us to form products of vectors. The first product, called the scalar product, or dot product is available in many vector spaces, e.g., in the plane. It associates a scalar number with any pair of vectors. There are two ways to define it, an algebraic one (in the Cartesian, or component representation), and a geometric one. Let us start with the algebraic definition:

$$\vec{a} \cdot \vec{b} \equiv a_x b_x + a_y b_y . \quad (1.7)$$

We will avoid dots to denote multiplication of scalars to emphasize that the dot product is formed from two vectors. The right-hand side of (1.7) is a scalar, which is a real number without any restrictions (it can be negative). The definition of a vector magnitude (length) can be traced back to the dot product of a vector with itself (in which case one obtains the square of the vector length, which is positive). The dot product (1.7) is easily generalized to more than two dimensions. In this

course we will use at most $N = 3$ dimensions, but in special relativity one needs $N = 4$, and later (in the context of function spaces, Fourier series, etc.) physicists use higher N , even $N \rightarrow \infty$:

$$\vec{a} \cdot \vec{b} \equiv \sum_{i=1}^N a_i b_i . \tag{1.8}$$

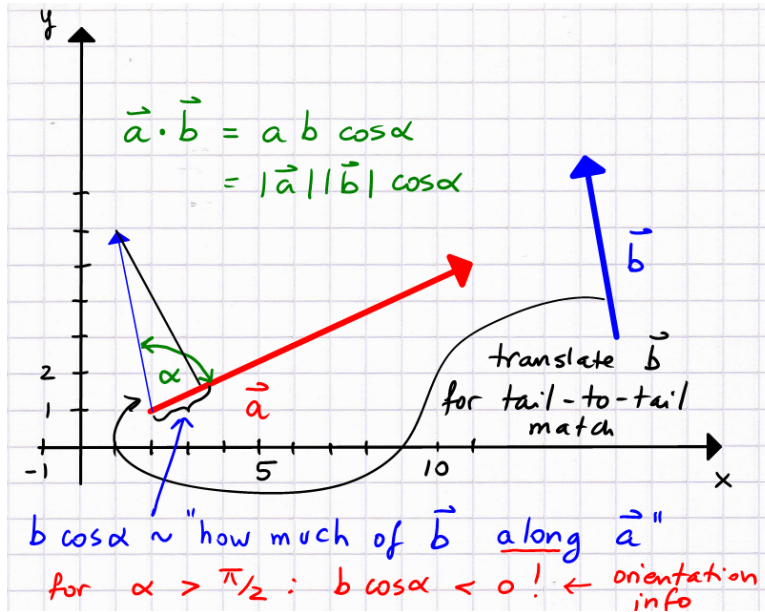


FIGURE 3. The dot product or scalar product of two vectors is a scalar number (with sign) which measures how much the second vector points along the first vector, and then scales this number by the length of the first vector. Since the dot product doesn't care about the order of the factors, the interpretation also works the other way around.

The geometric version of the dot product is quite intuitive:

$$\vec{a} \cdot \vec{b} \equiv ab \cos \alpha , \tag{1.9}$$

where α is the angle between vectors \vec{a} and \vec{b} , and note that on the right hand side of (1.9) the magnitudes of these vectors appear. In the real plane it is straightforward to show that (1.8) and (1.9) are equivalent. In vector spaces with $N > 3$ one can use the combination of these equations to define the angle between two vectors.

The significance of the dot product in physics, especially apparent in the form of (1.9) is that it allows to formulate laws between vectorial quantities in which not only the magnitude of the factors plays a role, but the relative orientation of the two vectors entering the product. In particular, it is obvious from the relation $\cos \frac{\pi}{2} = 0$ that the dot product of two perpendicular vectors vanishes.

From the geometric version (1.9) we see that the dot product of two vectors becomes negative when the angle between them is obtuse (larger than 90 degrees). To find the angle α we place the vectors tail to tail. A graph of the cosine function shows that for $\frac{\pi}{2} < \alpha < \pi$ the result is negative. A useful geometrical interpretation of the dot product is that it provides the product of the component of \vec{b} along \vec{a} , which is $b\cos(\alpha)$, times the magnitude a . The construction works also the other way around, since the dot product is symmetric in its two factors. It does not matter whether the angle α is measured from \vec{a} towards \vec{b} , or the other way around. This works out, since the cosine function has the symmetry property $\cos\alpha = \cos(-\alpha)$. In many other instances we will be careful when measuring angles, by agreeing that they are positive when increasing in the counter-clockwise (CCW) direction.

An important physics application of the scalar product can be found when calculating the work ΔW done by a force \vec{F} along some displacement $\Delta\vec{r}$. The work is a scalar quantity and takes into account only that component of the force that acts along the displacement.

1.6. Vector components and component vectors. We now return to the notion of assembling a vector from its components along the axes to prevent some confusion. The motion on an inclined plane shown here is a typical problem where one can simplify matters by a clever choice of coordinate system. In the standard set-up where \hat{i} is chosen along the horizontal, and \hat{j} along the vertical the motion is two-dimensional, i.e., would require functions $x(t)$ and $y(t)$ for a mathematical description.

Yet, the motion is one-dimensional along the surface, since the normal force from the surface cancels the gravitational force component pressing the object into the surface. Thus, it is clever to choose the direction of motion as one basis vector, and the perpendicular direction as the other. In order to take advantage of this trick we need to know how to resolve vectors into components along any possible direction.

In terms of given \hat{i} and \hat{j} any vector \vec{a} can be written as the sum of component vectors:

$$\vec{a} = \vec{a}_x + \vec{a}_y = a_x\hat{i} + a_y\hat{j} . \quad (1.10)$$

Here \vec{a}_x and \vec{a}_y are component vectors that make up \vec{a} by vector addition, while a_x and a_y are scalars, which are called the x - and y -components of \vec{a} . A rather confusing point is that we might think about a_x and a_y as being the magnitudes of \vec{a}_x and \vec{a}_y . However, this is only true if vector \vec{a} points in the first quadrant when its tail is at the origin! In general, the components a_x and a_y can be (and will be) negative. This leads sometimes to notational confusion: often, when using trigonometry or geometry to find out relations, we need the magnitudes, rather than the components, in which case we should, e.g., specify $|a_x|$, or at least provide a sketch of the situation to make things clear.

The vector components can be found by using the representation (1.10) and dotting into the equation (on both sides) with \hat{i} or \hat{j} (a trick used often in physics, so get used to it). This is called taking the projection of the equation along \hat{i} or \hat{j} . For example:

$$\hat{i} \cdot \vec{a} = \hat{i} \cdot (a_x \hat{i}) + \hat{i} \cdot (a_y \hat{j}) = a_x . \tag{1.11}$$

Here we used the properties that $\hat{i} \cdot \hat{i} = 1$, and $\hat{i} \cdot \hat{j} = 0$. Figure out these properties using both the algebraic and the geometric versions of the dot product!

1.7. Magnitude and direction from the components. The representation of vectors in terms of components is easy to work with, but has the drawback that we need to make a choice of axes (basis vectors $\hat{i}, \hat{j}, \hat{k}$), and specify what they are when describing \vec{a} in terms of (a_x, a_y, a_z) . We stated already in Eq. (1.6) that the Pythagorean in 2d (and the dot product in 3d) let us calculate the magnitude $a = |\vec{a}|$. How do we get the direction of the vector?

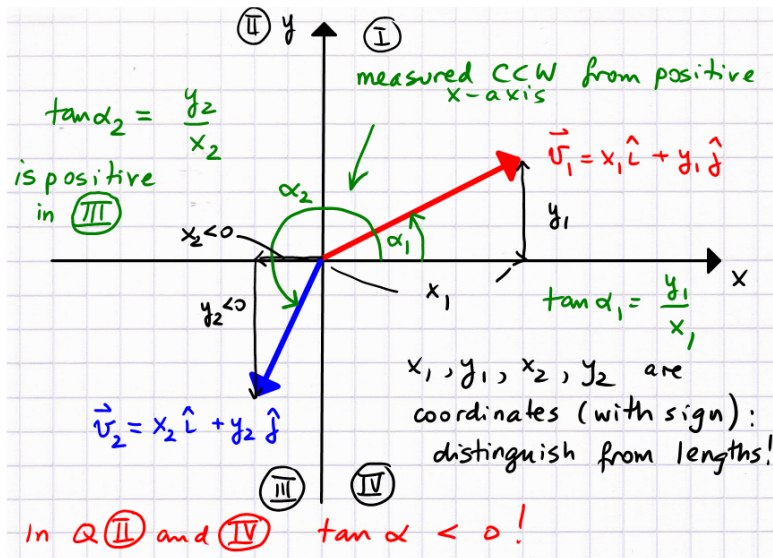


FIGURE 4. The direction of a vector is obtained from its components. Care has to be taken: the arctan = \tan^{-1} function as implemented on calculators returns answers in the $-\pi/2 < \alpha < \pi/2$ range. Inspection of the signs of the components x and y tells which of the four quadrants the vector points into. An addition of π or 2π corrects the answer. The arctan function with a single argument cannot distinguish between quadrants I and III.

In the plane, and if the vector points into the first quadrant, things are easy. From geometry/trigonometry we know that

$$\tan \theta = \frac{a_y}{a_x} . \tag{1.12}$$

Applying the idea of inverse functions, we solve (1.12) for θ . The arctan, or inverse tan function is implemented on calculators, and formally we write:

$$\theta = \arctan \frac{a_y}{a_x} = \tan^{-1} \frac{a_y}{a_x} . \quad (1.13)$$

You should explore this inverse function on your calculator to make sure to understand that it has nothing to do with $1/\tan x$. An inverse function undoes the effect of a given function; for another important pair of such functions you can study e^x and $\ln x$.

Things, are, however, a little trickier than just finding the key for arctan on your calculator. Depending on whether the argument is positive or negative, the calculator arctan function can only answer in the range $0 \dots \frac{\pi}{2}$, or $-\frac{\pi}{2} \dots 0$, i.e., the angle will always be in the first or fourth quadrant. Using negative angles to denote locations in the fourth quadrant is not such a bad thing, in principle (for significant digits it is much more reasonable to write -0.1° , than 359.9°). However, sometimes vectors do point into the second or third quadrants, and the calculator is misleading us! Many programming languages provide an arctan function with two independent arguments to solve this problem. However, in this course you'll have to do it yourselves. You have to look at the sign of a_x and a_y to determine the quadrant. A negative argument into the arctan function can happen when EITHER a_x OR a_y is negative. This happens in quadrants II and IV. The calculator always goes for quadrant IV, as stated above. You need to add π (or 180°) if $a_x < 0$ and $a_y > 0$. Likewise, the case when BOTH a_x AND a_y are negative, your calculator can't recognize this, since the signs cancel in the division before the arctan is evaluated. The angle given by the calculator will be in the first quadrant, and again, you'll need to correct manually by adding π (or 180°).

It is conventional to measure angles from the positive x -axis as positive in the CCW sense. Draw a sketch to state your case clearly.

1.8. Vector product or cross product. Midway through the course you will find this mathematical tool to help you understand three-dimensional space, and you will learn how to think properly about angular momentum and torque. Later in magnetism the tool will become indispensable. There is a product of two vectors which span (define) a plane, that results in a vector, which, however, points in a perpendicular direction to this plane. The two vectors \vec{a} and \vec{b} will be assumed to not point in the same (or exactly opposite) directions - otherwise they would not define a plane. To define this vector (or cross) product $\vec{c} \equiv \vec{a} \times \vec{b}$ we need to specify magnitude and direction. The magnitude is given by

$$c = ab \sin \alpha , \quad (1.14)$$

where α is chosen such that it falls into the range $0 < \alpha < 180^\circ$. The latter guarantees that the above is positive, i.e., describes a magnitude. Concerning the direction, we have two choices left: perpendicular to the plane defined by \vec{a} and \vec{b}

could be pointing up or down. To determine it uniquely, we define a right-hand-rule (left-handers, beware!): align the thumb with vector \vec{a} , the index finger with \vec{b} , and curl the middle finger naturally inside to be at a right angle to thumb and index finger; the middle finger will give you the orientation of the resulting vector \vec{c} . Doing this trick with the unit vectors \hat{i} , and \hat{j} (thumb and index) defines the z -axis for a right-handed coordinate system, i.e., it implements

$$\hat{k} = \hat{i} \times \hat{j} \tag{1.15}$$

We never use left-handed coordinate systems in this course: therefore, learn how to get it right. Convince yourself that using your left hand will result in \hat{k} would point in the opposite direction.

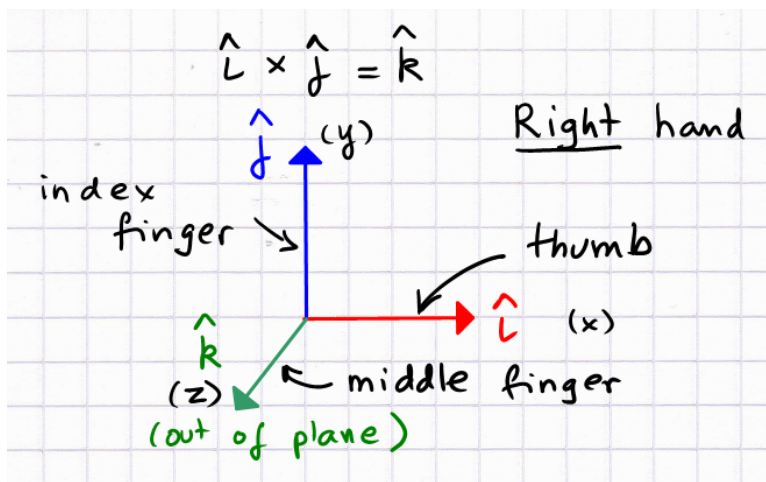


FIGURE 5. The right-hand rule uses the first three fingers spread out, with the middle finger pointing perpendicularly to the plane spanned by thumb and index finger to define a right-handed three-dimensional system of unit vectors. Use of the left hand would make the z -axis point into the screen - avoid this! The middle finger is curled naturally inward. This right-hand rule is used to find the direction for the cross product result $\vec{a} \times \vec{b}$, where \vec{a} and \vec{b} are represented by thumb and index finger, respectively. Order matters here: $\vec{b} \times \vec{a} = -\vec{a} \times \vec{b}$!!!

There is a formal cross-product rule for evaluation using determinants. It is used typically from second-year on.

A geometric interpretation of the magnitude of the cross product is that given vectors \vec{a} and \vec{b} , and the angle α as measured from \vec{a} towards \vec{b} , then $b \cos \alpha$ is the projection of vector \vec{b} onto the perpendicular direction to \vec{a} (on the same side as \vec{b}). The vector cross product is maximal in magnitude when \vec{b} is perpendicular to \vec{a} .

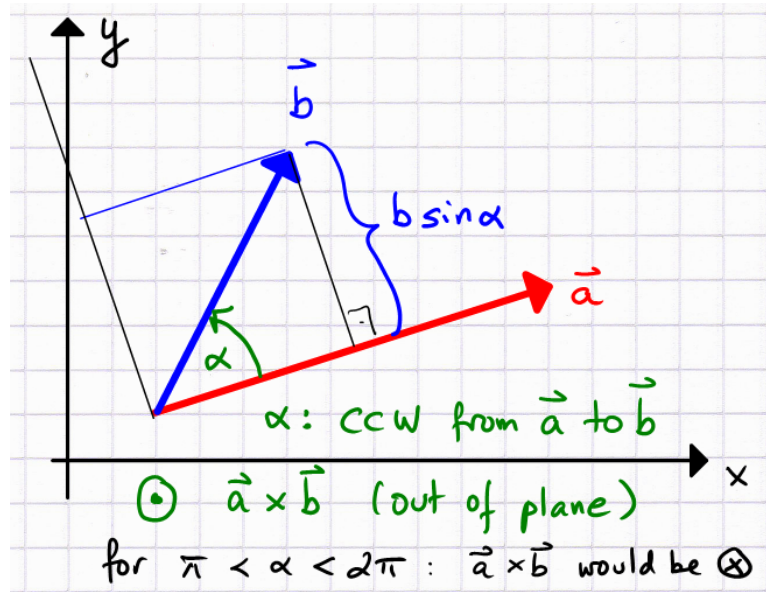


FIGURE 6. The vector cross product provides a measure of how much vector \vec{b} contributes in a direction perpendicular to that of \vec{a} ; this contribution, $b \cos \alpha$ is then multiplied by the magnitude $a = |\vec{a}|$. Reversing \vec{a} and \vec{b} results in the same magnitude $ab|\sin \alpha|$, but the resultant vector points in the opposite direction. For angles $\pi < \alpha < 2\pi$ the right-hand rule is awkward to use, but try it out for α close to 2π to learn how it works! In this case \vec{b} would be below \vec{a} in the figure.

2. CALCULUS

2.1. Preliminaries. Our textbook works at the pre-calculus level. Yet, the course is supposed to be calculus-based. Why do we do this, and how can we put the calculus in? There are multiple reasons for our choice. The main reason is that the physics laws are valid and can be formulated without taking limits. The formalism of calculus was developed by Newton, Leibnitz and others to derive nice symbolic results in mechanics. The physics principles, such as Newton's laws, the gravitational law, or the laws of electricity and magnetism by themselves have to be valid when using finite intervals. Finite-interval arithmetic is used commonly in computer programs to solve problems for which a symbolic answer cannot be found using Calculus tools. Thus, the pre-calculus version of physics has acquired some modern validation.

The main problem with so-called calculus based texts is that very quickly they move from simple ideas such as 'the slope of a curve is given by the first derivative' to 'calculate the area under the curve'. This happens very early in classical mechanics, yet in mathematics it is covered later under integral calculus (and not

covered in your typical high-school calculus course). To make things worse, one needs quickly a new type of integral (line or curve integral to calculate the work done by a force acting on a moving object in two or three dimensions), then the moment of inertia (a volume integral with potentially complicated boundaries). In electricity and magnetism we need another type of integral to measure how a force field permeates some surface. It doesn't stop there: we also want to solve differential equations, and we even write down some differential equations with partial derivatives (the wave equation) before the course is over. Past experiences with a deeply conceptual book that is based on calculus (by Randy Knight) have generated student complaints 'why are we doing new mathematics all the time?'.

Modern books addressing College Physics (college in the USA is not a community college, but really another name for university, usually without a graduate program) have found a way to emphasize the important laws without over-burdening the student with advanced mathematics (which is really covered only in the second year of university mathematics). The book by N. Giordano matches Knight's text in emphasis on the concepts. Very elegantly, it deals with topics such as Gauss' law in electricity without the mathematical burden. The present summary of calculus ideas is provided to put the calculus into the course where it is most appropriate and straightforward, namely the mechanics part. The first-year student in Ontario (and most likely elsewhere in Canada) is ready for this, since it really relies only on high-school calculus as a prerequisite. In a very straightforward way does this chapter introduce the idea of how area is measured under a curve. It also reviews some of the trigonometric and the exponential function to highlight their derivative properties - something that has been eliminated from the high-school curriculum when grade 13 (or OAC) was finally eliminated.

Let us make just one important comment here: the reason why physicists work liberally with objects such as Δt , and even treat the symbol dt as if it were a finite Δt , is that physics measurements (or physical reality) only exists for finite time intervals $\Delta t = t_2 - t_1$. Below a certain scale, the notion of classical mechanics begins to fail, the boundaries of a material object become fuzzy, since the laws of quantum physics take over. It is a beautiful result that in reality one can reduce Δt by orders of magnitude when defining the instantaneous velocity of an object while finding that $v(\frac{t_1+t_2}{2}) = \frac{x_2-x_1}{t_2-t_1}$ becomes independent of $t_2 - t_1$, but a measurement cannot take the limit $\Delta t \rightarrow 0$. Thus, physical reality cannot depend on the question whether a limit exists or not. The mathematical convenience of expressing results, however, does depend very much on this ability.

2.2. Derivative. In one-dimension the position of a particle is completely described by the function $x(t)$. In kinematics we study how this function is related to the variables that control its behaviour, namely the velocity $v(t)$, and the acceleration $a(t)$. In dynamics we are trying to answer the question how the net force acting on the particle at any location x governs the behaviour of $x(t)$ based

upon two initial conditions that make the motion unique. For a given time in-

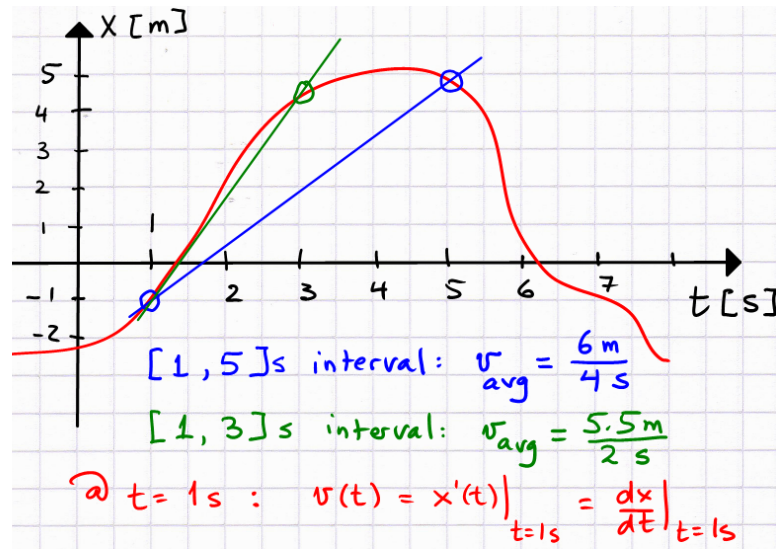


FIGURE 7. The trace of position versus time can be used to compute the average velocity for a given time interval. In the limit of vanishingly small time interval, $\Delta t \rightarrow 0$, this leads to the instantaneous velocity defined as the first derivative of $x(t)$ which is represented by the slope of the graph at time t . In physics measurements Δt remains always finite, but for sufficiently small Δt the velocity doesn't change anymore within the significant number of digits.

terval $\Delta t = t_2 - t_1$, which we control by using, e.g., a stopwatch we measure the displacement Δx , which is given by $\Delta x = x(t_2) - x(t_1) \equiv x_2 - x_1$. The average velocity for this time interval then is defined as

$$v_{\text{avg}} \equiv \frac{\Delta x}{\Delta t}. \quad (2.1)$$

Consider now the example of a car starting from a stop with the throttle fully depressed and moving along a perfectly straight road. We expect the speed of the car to increase rapidly at first, and to level off eventually, as the maximum speed is approached. Since the motion is in one dimension, and never changes direction speed and velocity will be the same (we choose the direction of motion as the positive x -axis). The velocity itself is a function of time. The graph displaying $v(t)$ becomes boring after t_0 when the top speed is reached. The time t_0 will be on the order of a minute. If we look at the part of the curve where $v(t)$ changes with time we expect that the average velocity will be different if we choose $\Delta t = 2$ s vs $\Delta t = 10^{-2}$ s = 10 ms. Why? Our intuition tells us that on the 2 s time scale the car is accelerating, and enough time has passed to observe a measurable change in velocity before and after this time interval. When we reduce the time interval

significantly below 1 s, however, the change in velocity Δv will become rather small, and we are measuring the instantaneous velocity at time t :

$$v(t) \equiv \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} \equiv \frac{dx}{dt} . \quad (2.2)$$

Geometrically, we observe that the velocity can be interpreted as the rate of change of position with time, or as the slope in the $x(t)$ graph shown below. The slope in the $v(t)$ graph given above represents the instantaneous acceleration $a(t)$. Initially, it is approximately constant: the engine delivers a maximum forward thrust (actually a backward thrust by the drive wheels against the ground, and a reaction force from the ground propels the car forward). This thrust is opposed by small frictional forces in the bearings. Eventually, air resistance builds, since the car picks up speed. When the velocity reaches its maximum value all of the forward thrust is compensated by air drag (and the small frictional force from the bearings).

Differential calculus is sufficient to understand how to obtain velocity from position, and acceleration from velocity. Our ultimate goal, however, is to do the reverse: to start with acceleration at any time t , and move towards determining the velocity change Δv over some time interval Δt , and from the determined $v(t)$ then reconstruct the displacement Δx for some Δt . Why is this our goal? Newton's second law relates the net force acting on a particle of mass m to the acceleration. Thus, if we understand which forces act on a particle at any position x , then we know the acceleration, and from it we can then solve for the motion. This is the ultimate goal of dynamics, and completes the understanding of how motion of a particle comes about. Before we move on to explain the opposite operation to differentiation, we make some remarks about how the differential calculus ideas are generalized in the case of multi-dimensional motion.

2.3. Derivatives of vectors. Given a vector that depends on a parameter (time in our example), $\vec{r}(t)$, we can generalize Eq. (2.2)

$$\vec{v}(t) \equiv \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{r}}{\Delta t} , \quad (2.3)$$

using the displacement vector $\Delta \vec{r}$. We can reduce these expressions to known quantities by using the Cartesian representation:

$$\vec{v}(t) = \frac{dx}{dt} \hat{i} + \frac{dy}{dt} \hat{j} + \frac{dz}{dt} \hat{k} , \quad (2.4)$$

We have the equivalent of the one-dimensional mathematics in all three components, since

$$\vec{v}(t) = v_x \hat{i} + v_y \hat{j} + v_z \hat{k} . \quad (2.5)$$

Therefore, we need to apply our high-school calculus ideas to the one-dimensional motions $x(t)$, $y(t)$, (and possibly $z(t)$), and then combine it to get the full picture.

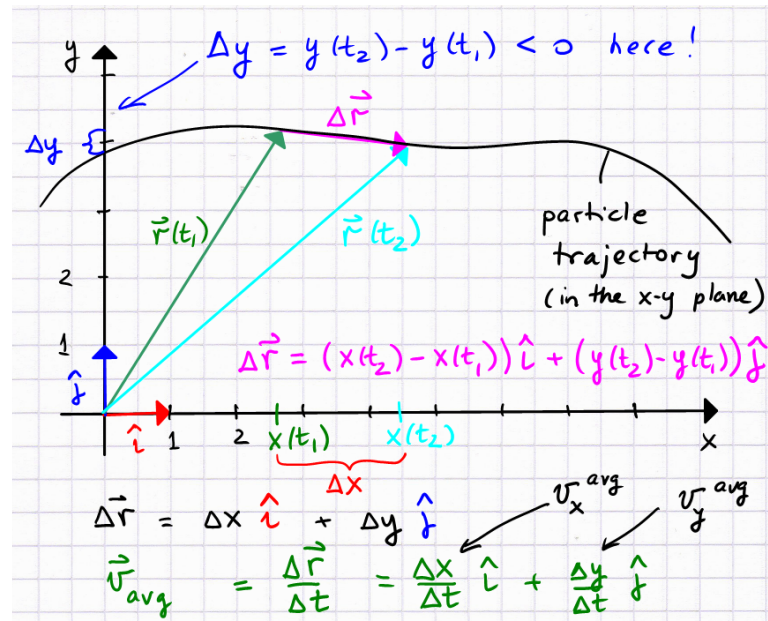


FIGURE 8. The derivative of a vector with respect to a parameter is defined in terms of the derivatives of the component functions $x(t)$ and $y(t)$. The graph illustrates this for finite time interval Δt , before the limit is taken. The unit vectors \hat{i} and \hat{j} are fixed in time for a non-moving coordinate system.

Whether it will be easy to carry out the program or not will depend on the forces acting on the particle, since they may depend on all three coordinates x, y, z in some non-trivial way.

2.4. Area under a curve. We showed that the relationship between velocity and position follows from the definition of velocity; when looking at the velocity graph vs time, it should tell us what the slope is at any time in the corresponding position vs time graph. How can we figure out an $x(t)$ graph from a given $v(t)$ graph?

Let us begin from Eq. (2.1) and solve it for the displacement (while dropping the avg label by assuming that Δt is sufficiently small):

$$\Delta x = v(t)\Delta t. \quad (2.6)$$

Let us assume that the particle is located at time t at position $x(t)$, at which it has velocity $v(t)$ which will not change appreciably over the short interval Δt . At the end of this interval the particle is located at $x(t + \Delta t) = x(t) + \Delta x$. From the graph shown for a linearly increasing $v(t)$ (which corresponds to constant-acceleration motion for some larger time interval) we can infer that the statement

$$x(t + \Delta t) = x(t) + v(t)\Delta t \quad (2.7)$$

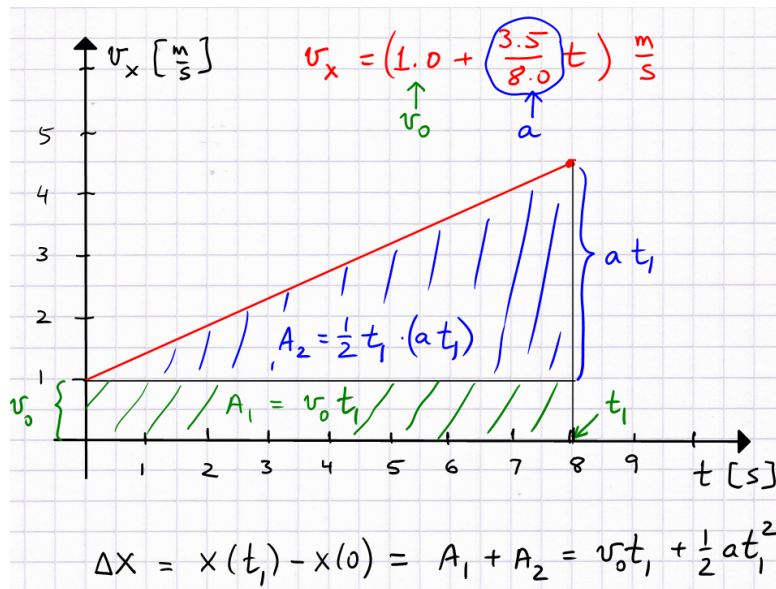


FIGURE 9. The velocity as a function of time for constant-acceleration motion. The particle starts at $t = 0$ with velocity v_0 . The displacement of the particle is observed to be the area between the trace of $v(t)$, the t -axis and the vertical lines $t = t_i = 0$, and $t = t_f$. The area is made up of two parts: a rectangle corresponding to the contribution from the initial velocity, and a triangle which is due to the constant acceleration.

establishes that the displacement is given by the area under the curve $v(t)$ for the segment $t \rightarrow t + \Delta t$. More precisely, the area bounded by the vertical lines $t, t + \Delta t$, by the t -axis and by the trace of $v(t)$ yields the displacement. Eq. (2.7) makes it clear that we start from a given $x(t)$, and the area law tells us to which position we advance after the time Δt . Wait, you might think, there is an inaccuracy here: the area under the curve is a rectangle plus a triangle, whereas you only accounted for the rectangle part! The small rectangle is of higher order, it is proportional to $\Delta t \Delta v$, and will drop out in a limiting process.

The same area-under-a-curve trick will work to get the velocity function $v(t)$ from a known $a(t)$. Solving for the motion from a given acceleration function $a(t)$ involves the repeated application of ‘integration’. Let us illustrate this by the constant-acceleration example known as a ‘physics law’ from high school, which merely becomes an almost trivial example in university:

We start with known acceleration $a = \text{const}$. The area under the horizontal line from fixed initial time t_0 to arbitrary time t is found to be

$$v(t) = v(t_0) + a(t - t_0) . \tag{2.8}$$

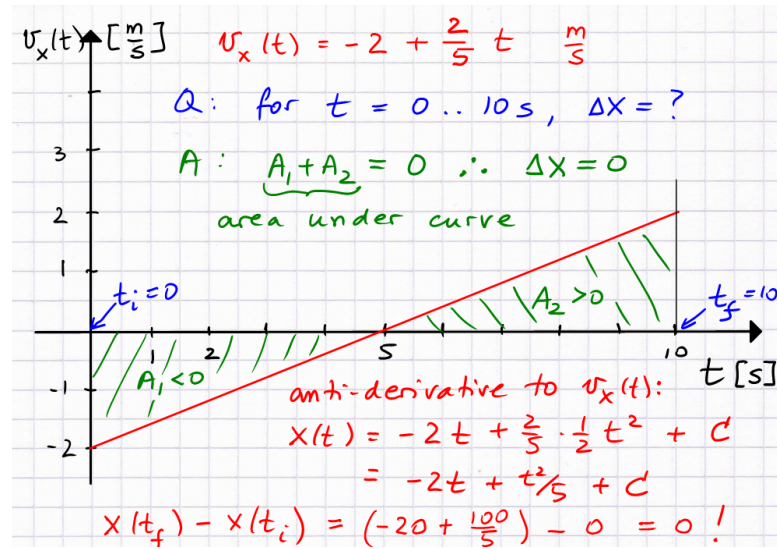


FIGURE 10. A similar example to the previous figure, except that the initial velocity is negative. For the choice of $v(t)$ and time interval $t_f - t_i$ the contributions from the initial velocity and from the acceleration just cancel. The area law requires to measure the area with sign which is exactly what definite integration does. The particle was moving to the left at $t = 0$, comes to a stop at $t = 5$ s due to the positive acceleration, and returns to the initial position at $t = 10$ s.

Note that the area-law does not tell us what $v(t_0)$ is! It only tells us how the velocity changes from the value at time t_0 . Since we never use absolute time, we could set our stopwatch to $t_0 = 0$. Then Eq. (2.8) for a given a value would represent a family of curves that start at $t = 0$ at any height $v(0) = v_0$, and are straight lines with slope a . We would state the answer as $v(t) = v_0 + at$. The meaning of v_0 is that of an initial condition. We can start the free-fall motion of a rock by releasing it from our hand ($v_0 = 0$), or by tossing it (up or down), therefore giving it some initial velocity.

Let us pick one such choice in the graph (with $a > 0$ and $v_0 > 0$). The area under this curve should give us the displacement of the particle. The area between $t_0 = 0$ and arbitrary time t , the t -axis and the graph consists of two parts: a rectangular piece of area $v_0 t$, and a triangular piece given by ‘one half times base times height’, i.e., $\frac{1}{2}t(at)$. We arrive at

$$\Delta x = v_0 t + \frac{1}{2}at^2. \quad (2.9)$$

Again, the area law has nothing to say about an initial condition, namely the initial position $x_0 = x(t_0) = x(0)$, and the full answer becomes

$$x(t) = x_0 + v_0 t + \frac{1}{2} a t^2 . \quad (2.10)$$

You can verify that differentiation of (2.10) results in the velocity (2.8), and further differentiation in the statement that the acceleration is a constant function of time $a(t) = a$.

Area-under-a-curve laws appear in other branches of physics in various contexts. They also appear in mathematical descriptions of biology, economics, engineering, or other fields where physicists have pioneered a deeper understanding of conceptual relationships. The calculus challenge of re-tracing the steps that lead to $x(t)$ when we have some function $a(t)$ requires the notion of integrals or anti-derivatives. It is needed when areas can no longer be obtained from simple geometric shapes. Note that the area formula for a circle is derived in calculus. In high school you were simply using a given formula $A = \pi R^2$ (a rabbit out of a hat?).

2.5. Anti-derivative. Given our simple example, where the position was given by a general quadratic function (2.10), and the velocity by a linear function (2.8) where we noted that the velocity follows by differentiation of the position function, we should be able to reverse the process. Thus, in general, when we try to find a position function that corresponds to a given velocity function, we are looking for the function whose derivative is $v(t)$. In integral calculus you will find that this is a hard process, and that simple functions can have complicated (or unknown) functions as their anti-derivatives. Sometimes they are sufficiently important that we call them ‘special’, study their properties and equip calculators and programming languages to provide easy access to them.

In integral calculus you prove the important theorem that if $f(t)$ and $F(t)$ satisfy $f(t) = \frac{dF}{dt}$, then the area under the curve $f(t)$ from t_1 to t_2 (between the vertical lines) is given by $F(t_2) - F(t_1)$. Note that area is computed with sign here, positive and negative areas (by being above and below the t -axis) do cancel. In this course you may have to apply this rule occasionally. In upper-year courses you will use it on many occasions.

In mathematics you prove what the derivative functions are for some elementary functions. We need these properties in first-year physics for sine, cosine, the exponential function (apart from the simple power law). While you should commit some of this to memory, you may be given a table. $\sin' = \cos$, $\cos' = -\sin$, $\exp' = \exp$, where the prime denotes differentiation with respect to the argument. If the velocity is given by the sine function, then the position will look like an inverted sine function (plus a constant).

2.6. Chain rule. The differentiation rule for composite functions is of utmost importance in physics, yet it was perhaps practiced too little in your mathematics education. A composite function can be denoted as $f(g(t))$, i.e., your ‘outermost’

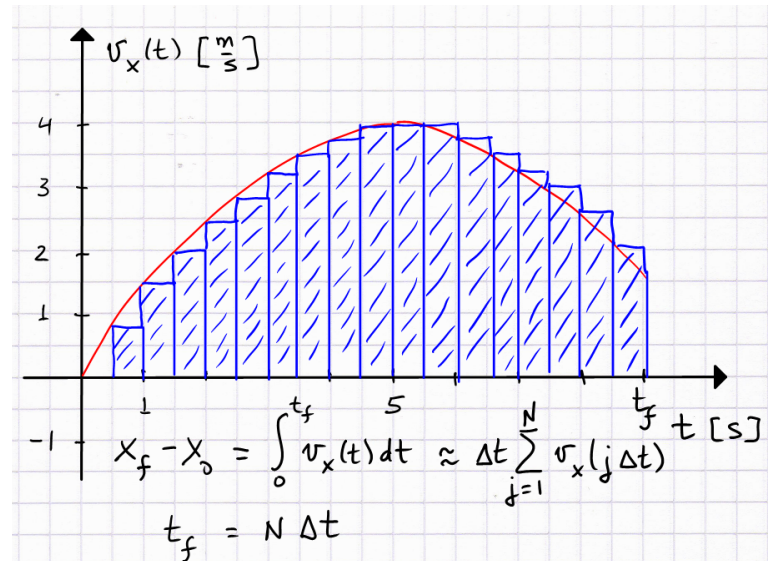


FIGURE 11. The trace of the velocity over time for a particle released by a compressed spring: the particle starts from rest, increases its velocity, reaches a maximum velocity when the spring is in its natural state (equilibrium), then it starts expanding the spring and slows down. The area under the curve which yields the displacement is approximated by a sum obtained from sampling the particle velocity at $\Delta t = 0.5$ s intervals. The sum of the rectangles will approach the area under the curve when the sampling rate is increased, i.e., the interval Δt will be decreased towards zero. In integral calculus this area is computed from the anti-derivative.

function has as its argument itself a function of the independent variable. An example could be $f(t) = \cos(\sqrt{t^2 + 1})$. How do we find the derivative with respect to t ? The rule is

$$\frac{df}{dt} = \frac{df}{dg} \frac{dg}{dt} . \quad (2.11)$$

The proof is in the calculus course. We just need to know how to use it. For our example it means: use the cosine derivative pretending the argument was the independent variable. Then multiply by the derivative of the inner function with respect to t .

This rule will be needed even in simple examples in physics for the following reason: one cannot calculate functions of dimensionful variables, i.e., $\sin t$, or $\exp t$ are OK in mathematics, but not in physics. Simple power-law functions are an exception to this rule. The simplest combination of two powers, such as $t^0 + t^1 = 1 + t$ shows that in physics this is a forbidden operation, unless one makes the convention that t is a number in some unit system. In general, the statement

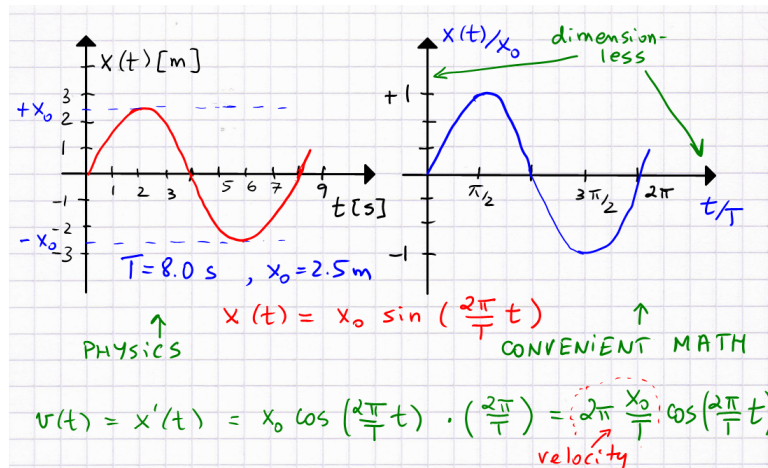


FIGURE 12. The trace of an oscillator solution (a mass attached to a spring is swinging to the right and left without damping) in a physics representation in SI units, as well as in dimensionless units. The length is scaled out using the amplitude x_0 , while time is measured in units of the oscillation period T . The argument of the sine function has to be dimensionless (an angle measured in radians). The velocity calculation requires the chain rule, and it is this factor that ensures that the derivative has dimensions of length divided by time.

$f(t) = 1 + t$ would mean that we are adding a dimensionless number (here 1) to the variable t , i.e., time which is dimensionful. take $f(t) = t + t^2$: what would be the meaning of adding seconds to seconds squared???

For this reason functions such as sine, cosine, exp must have arguments in physics that are made dimensionless. We may have some time scale t_0 , and a meaningful function $\exp(-t/t_0)$. In circular and oscillatory motion we will use frequency ν or circular frequency $\omega = 2\pi\nu$, or the period of the motion $T = 1/\nu = 2\pi/\omega$ to apply meaningfully:

$$\cos \omega t = \cos (2\pi \nu t) = \cos (2\pi t/T) . \tag{2.12}$$

While differentiating such expressions (or finding anti-derivatives) the chain rule will bring out dimensionful factors, which are needed, in fact, to turn a function describing a position into a velocity, etc.

2.7. Functions defined by their derivative properties. Laws in physics (and in other disciplines) are often relating variable functions to their derivatives (rate of change), or anti-derivatives. Answers to such laws (solutions) are then functions of the independent variable. We will look at three functions that come up in first-year physics (and also in chemistry and even biology). The class of exponential

functions is defined by the relationship

$$\frac{df}{dt} = f(t) . \quad (2.13)$$

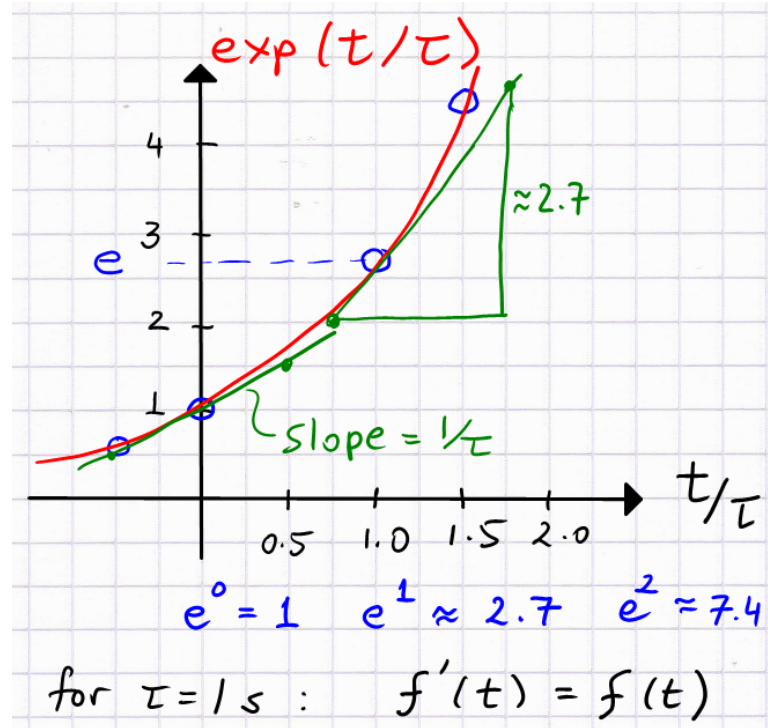


FIGURE 13. The exponential function is used in science to model growth that depends on the population size. The rate of change (slope) for any value of the argument is equal to the function value. Instead of using Euler's number $e \approx 2.7$ as base one could define the function $a^{t/\tau}$, e.g., with $a = 2$. For this function shifting the argument by one unit to the right would correspond to doubling the population instead of multiplying by $e \approx 2.7$.

Here t is a mathematical parameter (it could be time in dimensionless units), and a simpler way to write the equation is $f' = f$. Why are we interested in a function with such a property? The equation states that for any value of t the rate of change of the function (the slope of the graph) equals the function value. One of the most fundamental relations in population growth is the statement that the replication rate is proportional to the population. This assumes that there is space and food for unlimited growth. A bigger population results in more offspring. The population is measured by the number $N(t)$, and we state that how this number

changes with time depends is given by

$$\frac{dN}{dt} = aN(t) . \tag{2.14}$$

If t was dimensionful, we might argue that the dimensions of the proportionality constant a should be t^{-1} to make things right. Let us, however, write the equation in a physically meaningful way: we introduce a time constant τ , characteristic of the process we are investigating. The law is then written as

$$\frac{df}{dt} = af\left(\frac{t}{\tau}\right) . \tag{2.15}$$

If we write this in terms of dimensionless time $\xi = t/\tau$, i.e., time is measured as a multiple of the time constant τ , we see that the right hand side of (2.15) is $af(\xi)$, while the left-hand side by the chain rule becomes $\frac{df}{d\xi} \frac{d\xi}{dt}$. The second factor is simply $\frac{d\xi}{dt} = \frac{1}{\tau}$. We thus arrive at

$$\frac{df}{d\xi} = \tau af(\xi) . \tag{2.16}$$

Therefore, the proportionality constant in the population growth model (2.14) $a = 1/\tau$ is expressible as the inverse of a time constant. The mathematical statement (2.13) is for the special case when $\tau = 1$ unit. What is the meaning of this time constant τ ?

The mathematical statement (2.13) does not determine a function uniquely, it defines the ‘class’ of exponential functions. To select one of these functions from the class one has to supplement (2.13) with a statement of what the value of f is at some argument. All exponential functions pass through the point $f(0) = 1$. We need an additional selector. In physics (and most other science applications) we use a strange number to select our preferred f , namely Euler’s number $e \approx 2.7183$, which is an irrational number defined by a limit (a transcendental number such as π , since it does not follow from an algebraic equation). When looking at traces of electric signals on an oscilloscope we will be happy with the approximation $e \approx 3$.

Exponential growth is characterized by the (dimensionless-time) function $exp(t) \equiv e^t$, which means that as we move along the time axis by one unit the population increases by a factor of almost three. This is true whether we go from $t = 0$ to $t = 1$ (unit), from $t = 1$ to $t = 2$, or anywhere on the graph. Thus, the meaning of the time constant $\tau = 1/a$ is clear now: it is the length of time over which the population grows by a factor of 2.718..., or roughly a factor of 3.

We can also discuss radioactive decay: the number of decays observed per unit time is proportional to the number of parent nuclei still present in the source:

$$\frac{dN}{dt} = -aN(t) . \tag{2.17}$$

The negative sign is needed, since the rate of change is negative: after some time τ the number of undecayed nuclei will be less; for a finite time interval Δt this

would be expressed as $\Delta N = N_{\text{fin}} - N_{\text{in}}$, and since $N_{\text{fin}} < N_{\text{in}}$ we see that the derivative on the left-hand side of (2.17) is negative. In dimensionless units the function describing this behaviour is

$$\exp(-t) \equiv e^{-t}, \quad (2.18)$$

and the meaning of the time constant $\tau = 1$ unit in this case is that it takes time τ for the population to drop by almost a factor of three.

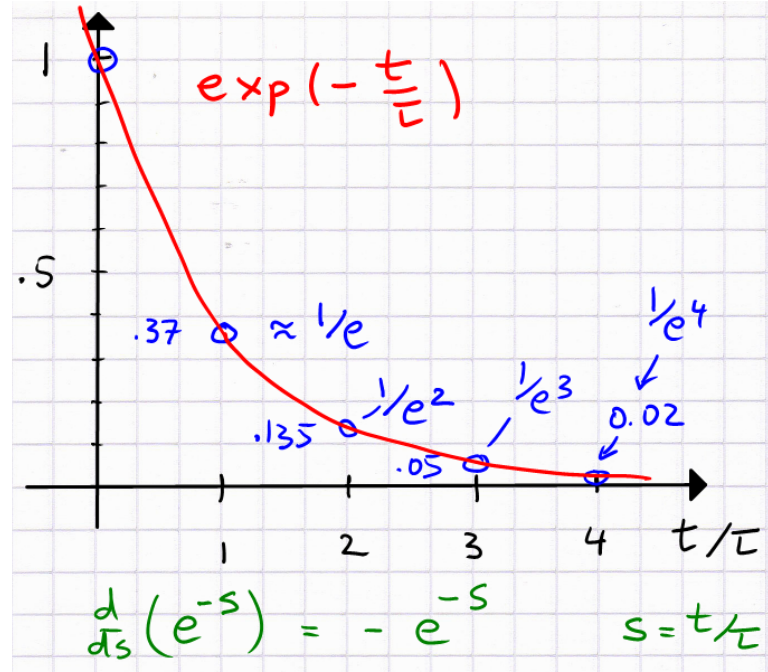


FIGURE 14. Exponential decay is a process where the (negative) rate of change of some quantity is proportional to the amount of quantity left. Within one characteristic lifetime the amount is reduced by a factor of $e^{-1} \approx 0.37$. For each further passage of a lifetime the amount is again reduced by the same factor.

We will use the exponential function to describe how mechanical energy is converted in a real-life oscillator into heat (the energy loss is proportional to the motional energy), or how electrical oscillations die down due to damping (electrical energy is converted into heat in a resistor). The oscillations themselves are described by a class of functions that satisfy the law (again in dimensionless time):

$$\frac{d^2 f}{dt^2} = -f(t). \quad (2.19)$$

Both the sine and cosine functions solve this equation, based upon their relationships under differentiation: $\sin' = \cos$ and $\cos' = -\sin$. The details of this relation

are discussed in the lectures on oscillations, but the basic properties are needed to understand the mathematical description of circular motion.

2.8. **Complex numbers.** Hundreds of years ago mathematicians wondered about the question why some polynomials of order n have n roots in the real numbers, while others didn't. It happens already for the quadratic. The equation

$$z^2 + 1 = 0 \tag{2.20}$$

has no solution for real z . The idea to extend the real-number system to complex numbers by defining the imaginary unit $i^2 = -1$ (leading to two solutions to (2.20), namely $z = \pm i$) leads to a number field that is closely related to the two-dimensional real plane. The complex number

$$z \equiv x + iy \tag{2.20}$$

is defined in terms of independent real x and y . The beauty of this number system is that any n^{th} order polynomial has exactly n roots. While physics variables to be measured have to be representable by real numbers (there is no meter in this world that can show a complex number for an answer!) complex numbers are used in physics. It is a convenience in electricity (electrical engineers really use them!) and a necessity in quantum mechanics - even though the measurable quantities come out as real-valued. We simply end this sub-section by stating a

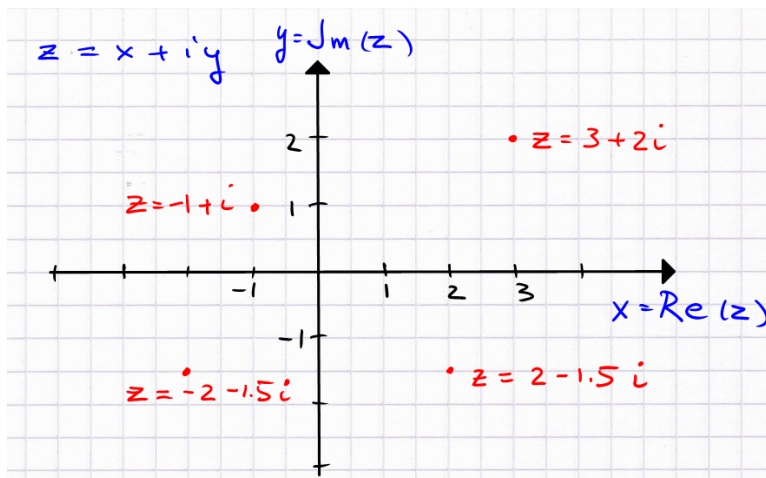


FIGURE 15. The graphical representation of the complex number plane illustrates that the real and imaginary parts of a complex number are independent of each other, very much like the Cartesian x - and y -components of a vector in the plane.

powerful relationship, namely Euler's theorem which shows the relation between

the exponential and basic trigonometric functions (which you might have wondered about in the previous subsection):

$$e^{i\alpha} = \cos \alpha + i \sin \alpha \quad (2.21)$$

2.9. Taylor series. It is possible to verify Euler's theorem (2.21) by observing the Taylor series expansions for the three functions involved:

$$\sin t \equiv t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} \pm \dots, \quad (2.22)$$

where the factorial $n! \equiv 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$, and

$$\cos t \equiv 1 - \frac{t^2}{2!} + \frac{t^4}{4!} \pm \dots, \quad (2.23)$$

while

$$\exp t \equiv 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \dots \quad (2.24)$$

The proof of these expansions comes late in your Calculus education, since it is an application of integral calculus. The meaning, however, is important in physics: a function can be approximated locally by starting with the function value, and having some derivative information for the local point $t = t_0$:

$$f(t_0 + h) = f(t_0) + \frac{f'(t_0)h}{1!} + \frac{f''(t_0)h^2}{2!} + \frac{f'''(t_0)h^3}{3!} + \dots \quad (2.25)$$

The main idea of the Taylor series (2.25) is to replace a complicated function f by a polynomial. Eqs. (2.22-24) state that relation for the three 'complicated' functions used in this course using $t_0 = 0$ as the expansion point. Computer arithmetic only knows elementary operations (add, multiply and their inverses). Calculators and floating-point units in CPUs use expansions based upon (2.25) to compute numbers for a given argument t . When modelling nature we use expansions such as (2.22-25) to simplify laws. An example in this course is when the pendulum equation is reduced to a simple oscillator for small-amplitude oscillations.

This summary of what will be used in the first-year courses should serve as motivation to study mathematics ideas in some depth. It is meant as a shortcut so that one can apply them without being help up by the (sometimes tedious) process of proving them.