Per Amund Amundsen
Department of Mathematics and Natural Science
UNIVERSITY OF STAVANGER

Notes for FYS500 Classical Mechanics 21.08 2017

Additions and comments to *Classical Mechanics* by H. Goldstein & al. (3 ed. 2002).

Introduction

Before we start with the textbook, it may be useful to make a quick review of some basic material from elementary mathematics and mechanics. This will also serve to establish the notation we shall use in the course.

In *Classical Mechanics* we assume, with Isaac Newton, that we live in an 3-dimensional Euclidean space, and that time passes uniformly at the same constant rate everywhere. This is only an approximation to the real world, but it is an excellent one in many cases. We must restrict ourselves to events which take place *locally*, say within a radius of a few billion light-years and well away from black holes, on time scales much shorter than 10 billion years, and to objects moving slowly relatively to the speed of light. The last restriction will be lifted when we come to the *Special Theory of Relativity* in chapter 7.

Another well-known restriction on Classical Mechanics is that it can only be regarded as an approximation to the more fundamental *Quantum Mechanics*, which applies on atomic length scales. However, for distances longer than a few atomic diameters, say 0.2 nanometers, or so, Classical Mechanics become surprisingly accurate for most processes. This is to some extent guaranteed by the equations of motion of non-relativistic Quantum Mechanics, because the *average values* of quantum mechanical dynamical variables satisfy the corresponding classical equations [Ehrenfest's principle].

Mathematical preliminaries

We know that in an Euclidean space we can introduce *orthogonal coordinate systems*, such that the position of any point $\mathcal{P}$ can be uniquely specified by giving three coordinates, which are real numbers, say x, y and z. We have also learned that it is extremely convenient to consider these three numbers together as a *single* entity, a *vector*. In standard linear algebra notation, we can write it as a *column* vector:

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} . \tag{0.1a}$$

In mechanics it is common to write vectors with boldface type. In handwriting we use an arrow above the symbol, $\vec{r}$, or an underline, $\underline{r}$.

We call the individual coordinates the *components* of the vector. It is often useful to consider the *coordinate vector*, $\mathbf{r}$, as the fundamental object, and to have a notation which makes it clear that the coordinates are the components of the same vector. We shall therefore often use naming conventions like $x = r_x = r_1$, $y = r_y = r_2$ and $z = r_z = r_3$ for the components of a vector. Thus:

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} . \tag{0.1b}$$

Two vectors are equal if and only if they have the same components.

Instead of writing vectors as column vectors, we can equally well have written them as *row* vectors. In linear algebra, this is done by introducing the *transpose* of a vector $\mathbf{r}$:

$$\mathbf{r}^{\mathsf{T}} = [x, y, z] = [r_x, r_y, r_z] = [r_1, r_2, r_3] . \tag{0.2}$$

We see that there is a trivial one-to-one correspondence between a vector and its transpose, and most of the time no confusion will arise by not distinguishing between them. Since it is typographically more convenient, we shall use the row vector notation of eq. (0.2) most of the time, dropping the $\intercal$.

Vectors are added and multiplied by numbers by performing the operations for each component separately, so if $\mathbf{r} = [r_1, r_2, r_3]$ and $\mathbf{s} = [s_1, s_2, s_3]$, then:

$$\mathbf{r} + \mathbf{s} = [r_1 + s_1, r_2 + s_2, r_3 + s_3] = \mathbf{s} + \mathbf{r}\,, \tag{0.2a}$$

$$c\mathbf{r} = [cr_1, cr_2, cr_3]\,, \tag{0.2b}$$

where c is a number. We can obviously also subtract vectors and divide by numbers in the same manner. We can exploit this to introduce three **basis vectors**:

$$\mathbf{i} = \mathbf{e}_x = \mathbf{e}_1 = [1,0,0]\,; \qquad \mathbf{j} = \mathbf{e}_y = \mathbf{e}_2 = [0,1,0]\,; \qquad \mathbf{k} = \mathbf{e}_z = \mathbf{e}_3 = [0,0,1]\,. \tag{0.3}$$

By using eqs. (0.2) repeatedly, we see that we can expand any vector $\mathbf{r}$ in terms of these basis vectors as:

$$\begin{aligned}\mathbf{r} = [x,y,z] &= x[1,0,0] + y[0,1,0] + z[0,0,1] = x\,\mathbf{i} + y\,\mathbf{j} + z\,\mathbf{k} \\ &= r_x\mathbf{e}_x + r_y\mathbf{e}_y + r_z\mathbf{e}_z = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3 = \sum_{i=1}^{3} r_i\mathbf{e}_i = \sum_i r_i\mathbf{e}_i \equiv r_i\mathbf{e}_i\,.\end{aligned} \tag{0.4}$$

In the next to the last step, we have introduced the convention that if the number of components of a vector is obvious, we need not bother to record it. In the last step, writing $\mathbf{r} = r_i\mathbf{e}_i$, we have introduced:

> **Einstein's summation convention**: *If a vector index appears exactly twice in a product, a sum over all possible values of that index is implied.*

Goldstein uses this convention from chapter 4 on. All the above ways of writing coordinate vectors are in common usage, and we will switch freely between them, in order to get used to all of them. We shall also adopt the usual physicist's convention that if $\mathbf{a} = a_i\mathbf{e}_i$ is a N-dimensional vector, and we talk about a_i, we mean all the a_i for $i = 1, \ldots N$, unless something else is explicitly stated. The same applies to a basis vector $\mathbf{e}_i$.

In the above discussion, the number $N = 3$ of space dimensions does not play any important mathematical role. If we index the vector components and basis vectors by numbers, we can define vectors in the same manner for any number of components $N = 1, 2, 3, \ldots$ (the case $N = 1$ is, of course, trivial). Any set (collection) of objects for which eqs. (0.2) are well defined and valid, is called a *vector space*. The components, r_i, can be taken to be any kind of numbers we like, integers, rational numbers, reals, complex numbers or something else, as long as addition and multiplication are well defined and *commutative*, *i.e.* $r_ir_j = r_jr_i$. In Classical Mechanics we shall mostly need $N = 3$, the simpler case $N = 2$ and the trivial case $N = 1$, all with real components, but vector spaces of higher dimensions prove very useful in many other contexts. Thus we shall have use for N=4 in special relativity. Indeed, we may even let $N = \infty$, with components that are complex numbers. The resulting vector space is known as the *Hilbert space*, well known from non-relativistic quantum mechanics. In the following, we shall continue to use $N = 3$ for definiteness, but the results will remain valid for any N, unless something else is explicitly stated.

A real Euclidean space has more structure than just being a vector space. As is well known, it is equipped with a (commutative) *scalar product* (dot product), defined as:

$$\mathbf{r}\cdot\mathbf{s} = \mathbf{r}^{\mathsf{T}}\mathbf{s} = [r_1, r_2, r_3]\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = r_1s_1 + r_2s_2 + r_3s_3 = \sum_{i=1}^{3} r_is_i = \sum_i r_is_i = r_is_i = \mathbf{s}\cdot\mathbf{r} = \mathbf{s}^{\mathsf{T}}\mathbf{r}\,, \tag{0.5}$$

where the Einstein convention is used in the third last step. This can be used to express the *length*, r, of a vector $\mathbf{r}$ as:

$$r^2 = \mathbf{r}^2 = \mathbf{r}\cdot\mathbf{r} = \mathbf{r}^{\mathsf{T}}\mathbf{r} = x^2 + y^2 + z^2 = r_x^2 + r_y^2 + r_z^2 = r_1^2 + r_2^2 + r_3^2 = \sum_{i=1}^{3} r_i^2 = \sum_i r_i^2 = r_ir_i\,. \tag{0.6a}$$

$$r = |\mathbf{r}| = \sqrt{\mathbf{r}^2} = \sqrt{x^2 + y^2 + z^2} = \ldots = \sqrt{r_ir_i}\,. \tag{0.6b}$$

Geometrically, this is nothing but the *Pythagorean Theorem.*† We see that any vector $\mathbf{r}$, except the *null vector* $\mathbf{0} = [0, 0, 0]$, have length $r = |\mathbf{r}| > 0$. Note that we have used the same letter written with different typefaces to distinguish between the *vector* $\mathbf{r}$, written in **bold**, and its *length* $r = |\mathbf{r}|$, written in *italic*. In handwriting, we just drop the vector symbol for the length. We shall often employ this convention in these notes. Also note that when using the Einstein convention, we do *not* write $r = \sqrt{r_i^2}$, because in the expression r_i^2 the index, i, is not appearing twice, as it is in $r_i r_i$.

If $\mathbf{r}$ and $\mathbf{s}$ are two vectors of length $r = |\mathbf{r}|$ and $s = |\mathbf{s}|$, the angle between them, θ ($0 \leq \theta \leq \pi$), can be found from the well known formula:

$$\mathbf{r} \cdot \mathbf{s} = rs \cos\theta \,. \tag{0.7}$$

This formula is valid in any number, N, of dimensions. If $\theta = \pi/2$ we have $\mathbf{r} \cdot \mathbf{s} = 0$, and we say that the vectors are *orthogonal*, or *perpendicular*, to each other. We also write this as $\mathbf{r} \perp \mathbf{s}$.

From eqs. (0.4) and (0.5) we then find the length of the sum and difference of two vectors from:

$$|\mathbf{r} \pm \mathbf{s}|^2 = (\mathbf{r} \pm \mathbf{s})^2 = \mathbf{r}^2 + \mathbf{s}^2 \pm 2\mathbf{r} \cdot \mathbf{s} = r^2 + s^2 \pm 2rs \cos\theta \,. \tag{0.8}$$

Geometrically, this is the well known *Extended Pythagorean Theorem.*

Two vectors, $\mathbf{r}$ and $\mathbf{s}$, are *colinear* if $\mathbf{s} = c\,\mathbf{r}$ for some number $c \neq 0$. From eqs. (0.6) we find in this case $s = |\mathbf{s}| = \sqrt{c^2 \mathbf{r}^2} = \sqrt{c^2}|\mathbf{r}| = |c|r$. In the last step we have to be a little careful, and write $= \sqrt{c^2} = |c|$, because of our convention that square roots of positive numbers are taken to be positive, while c might be negative. From eq. (0.7) we then have:

$$\mathbf{r} \cdot \mathbf{s} = c\mathbf{r}^2 = cr^2 = rs \cos\theta = |c| r^2 \cos\theta \,,$$

so $\cos\theta = c/|c| = \pm 1$. If $c > 0$ we have $\cos\theta = 1$ and $\theta = 0$, and the vectors are parallel, pointing in the same direction. If $c < 0$, $\cos\theta = -1$ and $\theta = \pi$. We say that the vectors are *anti-parallel*, pointing in opposite directions. From eq. (0.8) we find that for parallel vectors we have $|\mathbf{r} \pm \mathbf{s}| = |r \pm s| = |1 \pm c| r$, for anti-parallel ones $|\mathbf{r} \pm \mathbf{s}| = |r \mp s| = |1 \mp c| r$. In many situations, we do not distinguish parallel and anti-parallel vectors, and call all colinear vectors for parallel. We write $\mathbf{r} \parallel \mathbf{s}$.

A vector $\mathbf{e}$ of unit length, $|\mathbf{e}| = 1$, is called a *unit vector*. From the definitions (0.3) and (0.5), we see that the basis vectors are all unit vectors, $|\mathbf{i}| = |\mathbf{j}| = |\mathbf{k}| = |\mathbf{e}_i| = 1$ ($i = 1, 2, 3$, or $i = 1 \ldots N$ in the general case). Different basis vectors are found to be *orthogonal* from eq. (0.5): $\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = \mathbf{e}_i \cdot \mathbf{e}_j = 0$ ($i \neq j$). We normally prefer to work with such *orthonormal* coordinate systems, *i.e.* coordinate systems where the basis vectors are orthogonal unit vectors.

For any vector $\mathbf{r}$, except the null vector, $r = |\mathbf{r}| > 0$, and we can write:

$$\mathbf{r} = r\,\widehat{\mathbf{r}} \qquad \Longleftrightarrow \qquad \widehat{\mathbf{r}} = \frac{1}{r}\mathbf{r} \qquad \Longrightarrow \qquad |\widehat{\mathbf{r}}| = \frac{1}{r}|\mathbf{r}| = 1 \,. \tag{0.9}$$

Thus $\widehat{\mathbf{r}}$ is a unit vector, parallel to $\mathbf{r}$ (since $r > 0$), and is said to point in the *direction* of $\mathbf{r}$. We use this notation also for more complicated vector expressions, like:

$$\widehat{\mathbf{r} - \mathbf{s}} = \frac{\mathbf{r} - \mathbf{s}}{|\mathbf{r} - \mathbf{s}|} \,.$$

Using the definition of the scalar product and that basis vectors are orthogonal unit vectors, we can express the components of a vector $\mathbf{r}$ as:

$$x = [x, y, z] \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{r} \cdot \mathbf{i} \,, \qquad y = \mathbf{r} \cdot \mathbf{j} \,, \qquad z = \mathbf{r} \cdot \mathbf{k} \,. \tag{0.10a}$$

or more generally:

$$r_i = \mathbf{r} \cdot \mathbf{e}_i \,, \qquad (i = 1, 2, 3) \,, \tag{0.10b}$$

(or $i = 1 \ldots N$ in the case of a vector in N dimensions).

As they stand, the above results only apply to Euclidean spaces, with a scalar product given by eq. (0.5), or, equivalently, the length, also called the *norm*, given by eqs. (0.6). But it is often useful to introduce other norms, which may not have a natural interpretation as a length, in a vector space. Such spaces are then called *normed spaces*. One often defines the norm to be *positive definite*, *i.e.* $|\mathbf{r}| > 0$ for any vector $\mathbf{r} \neq \mathbf{0}$. But this is not always done. Indeed, we shall se that the space-time of special relativity has the geometry of the *Minkowski space*, with a scalar product that is not positive definite.

† In this course the symbol $\sqrt{x}$ for $x > 0$ stands for the *positive* square root of x

Coordinate systems

We have defined Euclidean spaces, and more generally vector spaces, in terms of vector components. This implies that we have chosen a specific *system of coordinates*. But this is not necessary. To the contrary, an important point about coordinate systems is that they are introduced by *us* to simplify our analysis. We can choose them freely in order to make our analysis as simple as possible. Thus, measurable physical quantities, like the *distance* between two points, or the *angle* between two (straight) lines, do not depend on our choice of coordinate system. Quantities that do not change when we change coordinates are called *scalars*. Thus lengths and angles are such scalars, and from eq. (0.7) it then follows that the *scalar product* is indeed also a scalar. Other scalars which we frequently encounter in physics are *volume, area, mass, density, electric charge, pressure, temperature, ...* . Mathematically, scalars are represented by pure numbers, but as physicists we shall almost always have a measurement unit associated with them.

If we chose to work with a different coordinate system, the coordinates of a given point will change, so that $x, y, z \longrightarrow x', y', z'$ (or $r_i \longrightarrow r_i'$), where x', y' and z' (or r_i') in general are different numbers from x, y and z (or r_i).[†] The precise relation between the primed and the unprimed coordinates of course depend on our choice of coordinate systems.

The simplest case of coordinate transformations is a *translation*, which amounts to a change of *origin* for the coordinates. This is simply done by adding a *constant vector*, $\mathbf{a} = [a_x, a_y, a_z]$ to our coordinate vectors:

$$\mathbf{r}' = [x', y', z'] = \mathbf{r} + \mathbf{a} = [x + a_x, y + a_y, z + a_z]\,, \qquad \Longleftrightarrow \qquad r_i' = r_i + a_i\,. \tag{0.11}$$

Obviously, $\mathbf{r} = \mathbf{r}' - \mathbf{a}$. We se that a translation transforms the original origin, O, with coordinate vector $\mathbf{r}_O = [0, 0, 0]$, to $\mathbf{r}_O' = \mathbf{r}_O + \mathbf{a} = \mathbf{a}$. Thus the vector $\mathbf{a}$ is just the position of the original origin O expressed in the new coordinate system. Similarly, the new origin, O', with new coordinates $\mathbf{r}_{O'}' = [0, 0, 0]$, has coordinates $\mathbf{r}_{O'} = \mathbf{r}_{O'}' - \mathbf{a} = -\mathbf{a}$ in the original coordinate system. Note that when we work with the components of vectors which also have indices themselves, we may have to watch out to keep the notation unambiguous. Thus if we introduce $\mathbf{r}_1 = [x_1, y_1, z_1]$ and want to have a more compact notation, we write $\mathbf{r}_1 = (r_1)_i \mathbf{e}_i = r_{1i}\,\mathbf{e}_i$ in the Einstein notation.

The length of a coordinate vector is obviously changed by a translation. In particular $|\mathbf{r}_O| = 0$ while $|\mathbf{r}_O'| = |\mathbf{a}| > 0$, unless $\mathbf{a}$ is the null vector. But the physical interpretation of the length of a coordinate vector $\mathbf{r} = [x, y, z]$ is that $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$ is the distance of the point from the origin of the coordinates, and that distance will of course change with a change of the origin. If we compute the distance, d, between two points with coordinate vectors $\mathbf{r} = [r_x, r_y, r_z]$ and $\mathbf{s} = [s_x, s_y, s_z]$, so $d = |\mathbf{r} - \mathbf{s}|$, we find in the primed coordinate system:

$$d' = |\mathbf{s}' - \mathbf{r}'| = |\mathbf{s} + \mathbf{a} - \mathbf{r} - \mathbf{a}| = |\mathbf{s} - \mathbf{r}| = d\,. \tag{0.12}$$

Thus we see that distances between points are indeed left unchanged, or *invariant*, by a translation, and so is a scalar.

The other important class of coordinate transformations is the *rotations* about some fixed point, which we can always arrange to be the origin, since we can always perform a translation first, if necessary. Rotations leave lengths and angles invariant, and consequently also scalar products. If we rotate our coordinate system, the basis vectors will also rotate, so $\mathbf{e}_i \longrightarrow \mathbf{e}_i'$. But since angles and lengths are not changed, the rotated basis vectors, $\mathbf{e}_i'$, still represent an orthogonal coordinate system, on par with the original set $\mathbf{e}_i$.

Even without actually performing any rotations, we often exploit the freedom to choose the orientation of the coordinate systems to simplify a problem. This is in particular useful if we manage to reduce the number of coordinates needed to describe a system, like if we can reduce a three-dimensional problem to an effectively two-dimensional one. In such cases we may choose coordinates so that one of them, often z, remains constant, preferably 0, throughout the problem, and so all vectors reduce to two-dimensional ones. As we shall learn, this can always be done for motion in the Earth's gravitational field, and also for the two-body problems both for Newton's law of gravity and for Coulomb's law for electric charged particles.

We shall defer any further discussion of rotational invariance to chapter 4.

[†] Note that in this course, as is usual in theoretical physics, a *prime* on a variable, like in x', does *not* in general signify a derivative, even if x is a function. Hence $x'(t)$ will typically be some function of t, which is distinguished from some other function $x(t)$. Normally we shall use the Leibnitz notation for derivatives, so the derivative of $x(t)$ will be written as $\mathrm{d}x/\mathrm{d}t$, except for time derivatives, where we shall mostly use Newton's dot-notation: $\dot{x} = \mathrm{d}x/\mathrm{d}t$. Only sometimes in mathematical manipulation will it be convenient to let $x'(t)$ represent the derivative of $x(t)$.