

The Seven Wonders of the World

Lecture notes for ING 175

Version 0.1, updated 27 March 2024



Typeset with L^AT_EX

Cover image: the “Pale Blue Dot” image of the Earth taken by Voyager 1,
from right outside Pluto’s orbit

<https://science.nasa.gov/resource/voyager-1s-pale-blue-dot/>

Introduction

The loss implied in such an acquisition can be estimated only by those who have been compelled to unlearn a science that they might at length begin to learn it.

J. C. Maxwell 1878

Until some decades ago, the 18th-century physical notions typically taught in introductory Bachelor physics courses were enough to prepare an engineer for future specializations and jobs. Students who wanted to venture into modern theories, such as Relativity, were required to **re-learn** some of the most important physical notions – *Energy*, mass, time, entropy above all – which in these theories are quite different from the 18th-century ones. But at that time the modern theories still mostly had only theoretical, not practical, importance. So the re-learning efforts of the curious students could perhaps be justified.

That situation has changed today. Modern theories and modern physical notions are an essential part of many everyday technologies, such as nuclear reactors and the [Global Positioning System¹](#) (see the entertaining discussion in Taylor & Wheeler 2000 Project A, and also Petit & Wolf 2005; Fliegel & DiEsposti 1996; Ashby 2002; Müller et al. 2008). Modern theories and modern physical notions are required for the development of new technological possibilities, from [quantum computers²](#) to solar sails, such as NASA's [Advanced Composite Solar Sail System³](#). An engineering student (including communication, computer, and data engineering) may likely end up in a job that requires an understanding of modern physical notions. The diffusion of [large language models⁴](#) will moreover lead to a future demand for engineers who actually **understand** those physical notions, not little monkeys who have been trained to mechanically manipulate

"Numerous relativistic issues and effects play roles in the global positioning system, on which millions of drivers, hikers, sailors, and pilots depend to find out where they are. The GPS system is, in effect, a realization of Einstein's view of space and time. Indeed, the system cannot function properly without taking account of fundamental relativistic principles." Ashby 2002

some equations while throwing some technobabble around. Automated large language models are faster, cheaper, and more precise in doing this kind of monkey activities. So why should one hire a human to do the same?

While moving from the older to the newer notions often requires re-learning efforts and conceptual re-orientations, the move in the opposite direction is less demanding. The modern physical notions are more encompassing than their 18th-century parents. Their understanding leads to an understanding of their older counterparts as approximate and special cases. Students, moreover, have often been hearing quite early about the new notions from mass media; for instance about the equivalence of mass and energy, or about the discrepancy in time reckoning by different observers. Owing to this early exposure, students sometimes ask very intelligent and deep questions – such as “*should the mass of the body be included in its internal energy?*” – when exposed to the old notions.

It is therefore high time that introductory Bachelor physics courses be based on modern physical notions. Students should not be required to waste time and mental effort to learn something that they must unlearn and re-learn, only because of the inertia of academia and teachers.

Some teachers say “it would be too difficult for students to understand modern ideas, because they are too familiar with the old ones. This is why we need to teach the old and slightly incorrect ideas first, and correct them later”. I think that this kind of reasoning is scientifically unacceptable and leads to a vicious circle. Students are unfamiliar with new notions only because they were raised by a generation who was taught old ones. New notions become familiar only after a couple of generations learn them at an early stage. This is obvious if you consider that notions such as “energy”, “electromagnetic field”, “vector” are very familiar today, but were absolutely *unfamiliar* a century or so ago. If we had always taught what’s familiar, then we would still be teaching about the *air, earth, water, fire* elements, and that the Sun revolves around the Earth. Arguments in favour of teaching old notions are for the most part just pretexts for laziness.

The present lecture notes are an experiment and attempt to introduce classical mechanics and thermodynamics from modern physical notions and viewpoints. The core equations remain the same, but the students should have a broader conception of their meaning and of the symbols that appear in them.

URLs for chapter *Introduction*

1. <https://www.gps.gov>
2. <https://www.ibm.com/topics/quantum-computing>
3. <https://www.nasa.gov/mission/acs3/>
4. <https://www.ibm.com/topics/large-language-models>

Chapter 0

Guide to the text

Maths prerequisites

- Working familiarity with algebra, its operations, and their properties, including square roots; knowledge of exponentials and logarithms is useful.
- Working familiarity with solving equations and inequalities, linear and non-linear.
- Working familiarity with the study of functions of one real variable.
- Working familiarity with derivatives.
- Understanding of what an integral is, even if you won't be required to solve integrals.
- Working familiarity with vector calculus.
- Some familiarity with functions of many variables.
- Understanding of what partial derivatives are.

Physics prerequisites

Just some vague reminiscences of secondary/high-school physics should be enough.

It can be beneficial if you are familiar with basic physics notions such as *velocity*, *mass*, *force*, and similar ones; and if you vaguely remember about Newton's laws.

Structure of this text

0.0.1 Graphical devices

The text includes the following graphical devices:

- Important notions and definitions are also given in **boldface**.
- The side margins often report clickable references to [previous topics](#),  § 0.0.1 page 5 emphasized in blue.
- Important-notion boxes:

Some important notion or definition

This is a definition or explanation of Something.

- Warnings and important points that require careful thinking:

Careful!

Something you must be careful about.

- Exercises:

Exercise 0.1

This isn't really an exercise

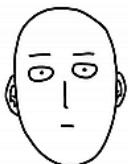
- Discussions and connections with more advanced physics:

How things really are in quantum physics

Just for your curiosity. You don't need to remember any of this.

0.0.2 Side pictures and quotes

Pictures, graphs, or quotes related to the material are displayed on the right.



This is an image of Saitama, which actually has nothing to do with the text on the left.

0.0.3 Hyperlinks and bibliography

Some words are hyperlinks, like this one about [One Punch Man](#)¹; you also recognize them because they have a little footnote number. The links'

URLs are listed at the end of each chapter, just in case you're reading a printed copy and wonder what the link was.

The text gives bibliographic references, like “Einstein 1905a”, to scientific literature. The references are listed in the final Bibliography on page 158.

These references are not part of the course and don't need to be “studied”. They are given for two reasons:

- For your own curiosity.
- To back up what's written in the text. In science you should not believe something just because you've read it somewhere. You should, as much as possible, *go and check for yourself how the logic behind the statement is proved and what is the experimental evidence behind the statement.*

0.0.4 Notation and terminology

Mathematical notation, as well as notation for physical dimensions, strictly follows the standards given by the International System of Units (SI)², listed for example in iso 2009 and iso 2019.

“Believe nothing, O monks, merely because you have been told it, or because it is traditional, or because you yourselves have imagined it. Do not believe what your teacher tells you merely out of respect for the teacher.”

(attributed to Gautama Buddha)

URLs for chapter 0

1. <https://onepunchman.fandom.com>
2. <https://www.nist.gov/pml/special-publication-811>

Chapter 1

Physics?

If you think about it, many things we ordinarily do every day are some sort of magic. Think of how you can instantaneously see and speak with a person living on another continent, in real time, using just a small widget in the palm of your hand. Think of how you can instantaneously see where you are on the Earth, using the same widget. Think of how fast you can go to another country, by flying in a huge metal thing. Think of how you can command and interact with a purely fictitious animated world when you play on your computer. The list can go on forever. Other things are luckily less ordinary, but still inspire a lot of awe: think of the devastating power unleashed by something roughly as small as a tennis ball, in an atomic bomb.

We can do these astonishing things thanks to our understanding of how the world works. That's Physics.

Many things can be said and have been said about science and physics. Rather than repeating what's been already written in many excellent books, I invite you to take a break here and go read their introductions. Choose as you please; compare what they say; don't limit yourself to popular books.

.....

1.1 Several possible formalisms or “languages”

Physics can be expressed and written from wildly different points of view, using wildly different principles. Let's call these “different physics languages”; a more technical name is “physics formalisms”. One may approach a physics phenomenon or problem in terms of *Lagrangeans*, or

Hamiltonians, or *fibre bundles*, or *categories*, or *action principles*, or many other formalisms. These formalisms or languages are not completely separated; we know how to translate among them. In "doing" physics, one may jump among formalisms, because some ideas may be easier to express, or some results easier to find, in one formalism than another. No matter which physics formalism you choose, the results and the concrete applications are still the same. The choice is to a great extent subjective, based on your aesthetic tastes. You see that in "doing" physics you can express your personality and put your own artistic touch; this is why it's such a cool subject (and other scientific subjects are like this too).

In these notes I'm choosing one particular formalism: the one that for me is the most easily *visualizable*; because I believe that visualization can be beneficial in learning new things. Or maybe I'm choosing it just because I like it best. I encourage you to explore how the physics you've learned is expressed in other physics formalisms; maybe you'll like another physics formalism better.

The formalism we'll be using might be called "field theory". Roughly speaking it takes as starting point the ideas of space and time, or better spacetime, in which there are different kinds of "stuff". It expresses the regularity and patterns that we observe in physical phenomena as "budgets" about the different kinds of stuff, and of relations between these kinds. Please don't take the description just given too literally; it's just meant to give you a very vague idea of the field-theoretical viewpoint.

It goes without saying that all these "physics languages" are to a great extent mathematical.

One reason is that numbers allow us to convey information in a concise and precise way. Imagine you have to tell someone, who doesn't know Bergen, where in Bergen you are right now, to within 10 m. You can do that with a description, "... and there's a building called so-and-so which looks like so-and-so...", which would be lengthy and tricky. Or you can just give two numbers: latitude and longitude:

60.369 40, 5.3518 .

And in these two numbers all digits are important; for instance, the latitude is not 60.369 47.

But the most important reason is that mathematics allows us to describe and follow the patterns and variety of physical phenomena in a greatly concise and precise way. And to develop their relationships in a rigorous

$$\delta \int L dt = 0 \quad L = \frac{1}{2} mv^2$$

$$\mathbf{F} = \frac{d}{dt} m \mathbf{v} \quad \mathbf{F} = 0$$

Example of two different formalisms (red, blue) expressing the same physical phenomenon.

"this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics"

Galilei 1623

way. All our present technology would have been impossible to discover, and would be impossible to realize, without the mathematical language of physics.

I invite you again to read what many good texts say about the relationship between physics and mathematics. No point repeating here what is said better elsewhere.

1.2 Quantities: primitive and derived

One topic must be briefly discussed because it's important for understanding the notes that follow. It's the distinction between *primitive* and *derived* quantities.

I shall assume that you already know what a **physical quantity** is. Examples are: position, duration, velocity, pressure, energy, temperature.

A **derived quantity** is one that is defined in terms of other quantities. For example, velocity v (more precisely: average velocity) is defined as the ratio between a distance d (a vector) and a time duration t :

$$v := \frac{d}{t}$$

where the symbol “:=” means “is defined as” or “is defined by”. This means that in principle we could avoid using the word “velocity” and the symbol “ v ” altogether, and instead always speak about distance and duration, using their symbols. It would lead to very long sentences and formulae and would be extremely inconvenient, but it could be done. The definition of a derived quantity often tells us how that quantity can be measured.

A derived quantity is defined in terms of other quantities, and these may in turn be derived quantities, that is, defined in terms of still other quantities, and so on. But at some point this chain of definitions must come to an end, otherwise we would go around in circles.

A **primitive quantity** is one that we do not define in terms of other quantities. Primitive quantities are the building blocks from which we define all others. That they are not defined in terms of others doesn't mean that we cannot try to explain them. But such explanations must be taken as informal and heuristic. Primitive quantities are often explained through metaphors and by appealing to intuition. You must always be wary of such explanations, because they may fail you spectacularly in some situations.

Often we have a choice about which quantities should be primitive and which should be derived. For instance, *energy* can be defined, in a

“There is nothing that can be said by mathematical symbols and relations which cannot also be said by words. The converse, however, is false. Much that can be and is said by words cannot successfully be put into equations, because it is nonsense.”

Truesdell 1966



“you have to be in some framework that you allow something to be true. Otherwise you're perpetually asking ‘why’”. (see [video¹](#)).

somewhat complicated way, in terms of quantities like *work* and *heat*, which would then need to be taken as primitive. Or we can take *energy* as primitive, and define *work* and *heat* in terms of it. This second choice can be more convenient to develop a physical theory. It often happens that a quantity is very convenient for building a theory, if used as primitive; but difficult to understand intuitively. Vice versa, a quantity can be very intuitive but lead to a complicated theory. Among quantities which we'll take as primitive are: *time*, *space* and *length*, *matter*, *energy*, *momentum*, *entropy*, *temperature*, and several others. All will be discussed soon.

1.3 Physical dimensions and units

Measurement is the process by which we determine the value of a physical quantity. Measurements can be extremely complex, and can extremely different even if they are about the same quantity. Consider the ways we can measure the mass of a football, compared to the ways we can measure the mass of the Sun.

To each quantity we associate a **physical dimension**. The term ‘dimension’ here has nothing to do with physical extension, as in “the dimensions of this box”; be careful not to confuse the two. Usually it’s clear which one is meant from the context. Physical dimensions help us avoid making operations that don’t make sense with some quantities. For example, it doesn’t make sense to sum up the volume of a body of water with its temperature; and indeed the volume has dimension *length*³, whereas temperature has dimension *temperature*, and quantities with different dimensions cannot be added up.

With each physical dimension we can associate a measurement **unit**, which expresses a basic standard for comparing the measurement results of similar quantities. For example, we could use the *minute* or the *second* as units to measure *time*.

One can choose a basic set of physical dimensions from which to define all others, and for these a set of standard units. Here we shall follow the [International System of Units \(SI\)](#)² (see also [NIST Special Publication 811](#)³).

The topics of measurement and physical dimensions, which are studied in *metrology* and in *dimensional analysis*, could occupy an entire course by themselves. I shall assume that you already know their basics notions and that you read about the SI.

The measurement of some physical quantities consists in just one number with associated physical dimension; we shall call such quantity a

scalar. The measurement of other physical quantities consists instead in a triplet of numbers with associated physical dimension; we shall call such quantity a **vector**.

! What's scalar or vector depends on the theory

Scalar and *vector* have very specific and slightly different meanings in different theories, so don't take the definitions used here as universal. For example, in these notes and in Newtonian mechanics we call *energy density* a scalar, but in general relativity it cannot be called a scalar.

1.4 Informal tips about units and maths

1.4.1 Importance of units

Units are very important and must always be written for several reasons.

First, a number without units doesn't tell us anything. If I tell you "the place is at a distance 100 from here", you have no idea how far the place is. "100" what? 100 metres? 100 kilometres? These are completely different distances.

Second, units give us useful information about mathematical objects and their physical relationships and measurement. If you see the expression "3 m/s", then there's a strong possibility that that's a velocity. If you see the expression "5 J/m²", then you have a hint that it could be measured by measuring an energy and then dividing by an area.

Third, because of the previous reason, keeping track of units often allows us to quickly catch errors in solving a physical problem.

1.4.2 Variables and units

When a physical quantity is denoted by a symbol or variable, keep in mind that a unit is "contained" in the symbol, so to speak. For example if the variable t denotes a time, then it includes some time unit, say seconds. This becomes apparent when we write the value of the symbol, for instance " $t = 120 \text{ s}$ ". The unit is not predetermined, but it must correspond to the dimension of that quantity. We could for instance write " $t = 2 \text{ min}$ " instead; the two expressions are completely equivalent.

This fact must be kept in mind when combining symbols. For example, if d is a distance and t is a time, then writing $v = d/t$ tells us that v is a

velocity, and it has appropriate units that come from d and t , for instance m/s.

Units otherwise behave just like *literal constants* for all mathematical purposes, just like the letter ‘ a ’ in the expression ‘ $a x$ ’. This is why they can, for instance, be simplified:

$$3 \text{ mol/s} \cdot 5 \text{ s} = 3 \frac{\text{mol}}{\text{s}} \cdot 5 \text{ s} = 15 \text{ mol} .$$

1.4.3 Mathematical functions and units

Particular care must be taken with trigonometric and exponential functions, such as $\sin()$, $\cos()$, $\tan()$, $\exp()$, $\log()$; **these functions only admit a dimensionless argument** (which for the trigonometric ones corresponds to *radians*). So there cannot be units like ‘s’ or ‘m’ within these functions: we must make sure that any units present within cancel out.

This makes sense, because we wouldn’t know how to interpret the argument otherwise. Suppose you read “ $\cos(60 \text{ s})$ ” somewhere: how much is that? If we say “just discard the unit”, we would have

$$\cos(60 \text{ s}) \stackrel{?}{=} \cos(60) \approx -0.95$$

but wait: $60 \text{ s} \equiv 1 \text{ min}$, so we could equivalently write “ $\cos(1 \text{ min})$ ”. Then, according to the hypothetical rule “just discard the unit”, we would have

$$\cos(60 \text{ s}) \equiv \cos(1 \text{ min}) \stackrel{?}{=} \cos(1) \approx +0.54$$

a completely different result!

For this reason an expression like ‘ $\cos(t)$ ’, with t denoting time, doesn’t make sense: there’s a time unit in the argument of $\cos()$. If we want to express an oscillation with time, we must write instead something like

$$\cos\left(\frac{t}{T}\right)$$

where T is the period of the oscillation, a symbol which also includes a time unit, which simplifies with the one in t . If the period of the oscillation is $T = 1 \text{ s}$ then we can also simply write

$$\cos(t/\text{s})$$

This expression is now unambiguous: suppose that $t = 60 \text{ s} \equiv 1 \text{ min}$, then

$$\begin{aligned} \cos(t/\text{s}) &= \cos(60 \text{ s}/\text{s}) = \cos(60) \approx -0.95 \\ &= \cos(1 \text{ min}/\text{s}) = \cos(1 \cdot 60 \text{ s}/\text{s}) = \cos(60) \approx -0.95 \end{aligned}$$

Also remember that **the results of trigonometric and exponential functions are dimensionless numbers**, so an expression like ‘ $3 \cos(\dots)$ ’ denotes a pure number, with no units. If you want to express that the result is a length, the appropriate units must appear. We can for instance write

$$L \cos(\dots)$$

where L is a length, and therefore includes some kind of length unit such as ‘m’. If this length is, say, $L = 2 \text{ m}$ we can also simply write

$$2 \cos(\dots) \text{ m}$$

1.4.4 Units and derivatives

When we follow the rules above, all other mathematical operations automatically take care of everything. The derivative, for instance, is calculated in the usual way, treating any visible units as literal constants. Let’s see a concrete example. This expression

$$x(t) = 2 \cos(t/\text{s}) \text{ m}$$

says that the position of some object oscillates with time, between the values -2 m and $+2 \text{ m}$. When $t = 0 \text{ s}$, the position is $x = +2 \text{ m}$. The position $x = -2 \text{ m}$ is reached when the argument of $\cos()$ is π , that is

$$t/\text{s} = \pi \quad \Rightarrow \quad t \approx 3.14 \text{ s}.$$

The **velocity** of the object is given by the derivative of this expression with respect to t . Let’s calculate it treating all unit symbols as literal constants:

↳ § 2.4 page 26

$$\frac{dx(t)}{dt} = \frac{d}{dt} \left(2 \cos(t/\text{s}) \text{ m} \right) = 2 \underbrace{\left[-\sin(t/\text{s}) \cdot \frac{1}{\text{s}} \right]}_{\text{chain rule}} \text{ m} = -2 \sin(t/\text{s}) \text{ m/s}$$

and you see that the correct units for velocity have automatically appeared.

1.5 What is “fundamental” physics?

But what’s the “ultimate” goal of physics? What’s “fundamental” physics? The answer to this question is again subjective – also in this case physics lets you express your proclivities and personality. In the history of physics one can probably identify two main conceptions of “fundamental” physics.

For some physicists it is about finding the ultimate building blocks, so that one day we can say “... and these are the constituents, and they obey these equations”. The history of physics seems to show that this goal is overturned every few generations. And yet every generation says “Now we almost have the complete picture – it’s right behind the corner. It’s true that previous generations thought they almost had it, and turned out to be wrong. But *this time* is different, this time we have the real deal!”. The theoretical and particle physicist [Geffrey Chew⁴](#) depicted this situation as in fig. 1.1. For this reason some physicists are a little sceptical about this goal; maybe it’s a never-ending structure, with surprises at every deeper look.

So for other physicists fundamental physics is about finding some point of view or mathematical structure that is rich enough to make useful predictions, and yet flexible enough to accommodate any new patterns or objects that we might discover. In a manner of speaking, it is about finding “patterns of patterns” or “laws about physical laws”.

The two conceptions above are not mutually exclusive, and both are always pursued, even if time-changing fashions may emphasize the one or the other.

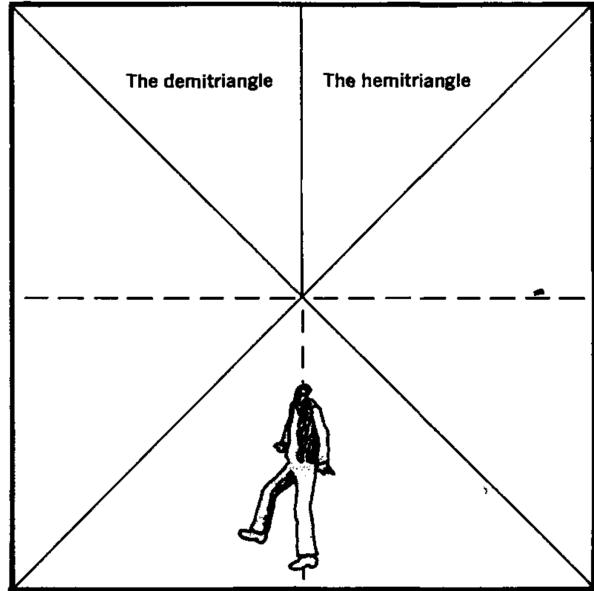
In these notes we take a point of view slightly closer to the second conception. This will also be reflected in the main division between physical laws that we’ll draw in Chapter 5.



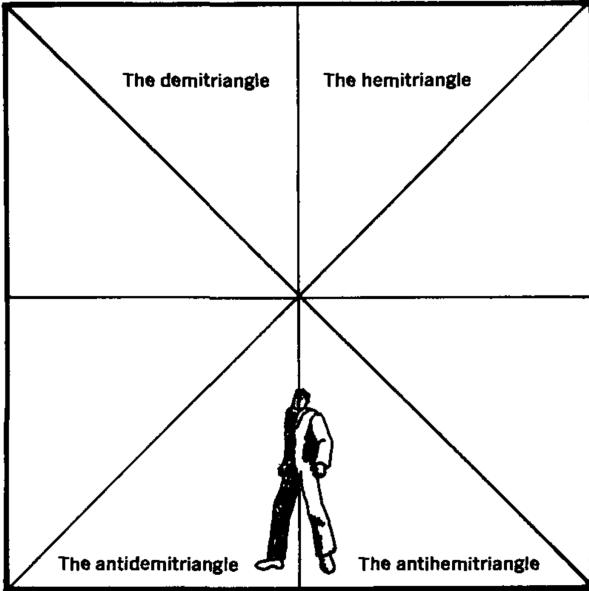
10 000 BC. The inhabitants of the paper square have no conception of the true nature of the universe they inhabit.



1900 AD. Physicists of the square discover a basic subdivision of their universe. They call it the "triangle" and consider it to be the fundamental building block of the universe.



1930 AD. Physicists discover that the triangle can be split. Its parts are termed the "hemitriangle" and the "demitriangle." These are thought to be the fundamental building blocks of the universe.



1950 AD. Mirror images of the hemitriangle and the demitriangle are discovered. These are termed "antihemitriangle" and "antidemitriangle."

Figure 1.1 (Continues on p. 17) *The progress of "fundamental" physics*, from Chew 1970 as reproduced in Truesdell 1987



1960 AD. Physicists' conception of their universe is further clouded by new discoveries: the rhombus, the parallelogram, the antiparallelogram, the nonalateral and many others. It is unclear what these discoveries signify.



1970 AD. A new configuration, the "hemidemisemitriangle," is hypothesized, out of which all known configurations of the universe can be constructed. The hemidemisemitriangle is thought to be the fundamental building block of the universe.



1975 AD. The hemidemisemitriangle is discovered. The following year the hemidemisemitriangle is split.



2000 AD. The inhabitants of this paper square have no conception of the true nature of the universe they inhabit.

URLs for chapter 1

1. <https://www.youtube.com/watch?v=nYg6jzotiAc&t=893s>
2. <https://www.nist.gov/pml/owm/metric-si/si-units>
3. <https://www.nist.gov/pml/special-publication-811>
4. <https://www.physics.lbl.gov/rememberinggeoffreychew/>

Chapter 2

Building blocks: Time and space

If we want to describe the *motion* of a material point, we give the values of its coordinates as a function of time. However, we should keep in mind that for such a mathematical description to have physical meaning, we first have to clarify what is to be understood here by “time”. We have to bear in mind that all our propositions involving time are always propositions about *simultaneous events*. If, for example, I say that “the train arrives here at 7 o’clock”, that means, more or less, “the pointing of the small hand of my clock to 7 and the arrival of the train are simultaneous events”.

A. Einstein 1905a

2.1 Time

Time is a primitive quantity. We understand the notion of time intuitively, even if it’s difficult to explain (that’s why it’s taken as primitive). In 1905, with the theory of relativity, part of our everyday intuition about this notion was seriously shaken. For many years afterwards our old intuition could still be used in practice and in applications. But the new, correct intuition is becoming more and more important in everyday life and technologies. For example, GPS navigation – which we use everyday from leisure activities such as hiking or sightseeing to more critical ones such as aeroplane landing – critically depends on the correct notion and intuition of time.

Let’s see how our traditional intuition goes astray with a concrete experiment. Here’s Alice, Bob, and Charlie. They have extremely precise

clocks built in exactly the same way. They stay very close to one another and synchronize their clocks. Still keeping close, they go around, maybe on an aeroplane or space ship, and all the time they check their clocks. They notice that their clocks stay perfectly synchronized all the time, no matter where they go and what they do.

At some point they separate, each one going around independently. One of them might stay in place, another might take a helicopter, and another might go for a trip on Mars and back.

Alice and Bob at some point meet again, and compare their clocks. They see that their clocks aren't synchronized anymore; the difference could be as small as microseconds, or as large as years. In fact, if this time discrepancy is large, they would notice that they themselves have aged differently; so time discrepancy doesn't affect the clocks only. Let's say for concreteness that Alice's clock is ahead of Bob's, or equivalently that Bob's is behind Alice's. Note the following aspects:

First, neither Alice or Bob can say "my clock was wrong": neither has noticed anything strange about the "passage of time".

Second, they might wonder what's the time on Charlie's clock. But Charlie is at some distance away. They could decide to contact Charlie via radio, say, and ask "what shows your clock right *now*?". But they would notice that there's a delay, even if extremely small, in the radio transmission; so it's unclear to what time would Charlie's answer apply. If we say "let's account for the radio-signal speed", we see that there's a logical problem: speed is distance divided by time, and here we have a problem in exactly determining what's the "correct" time! So we would be reasoning in circles. Besides, even neglecting these difficulties, Charlie's answer could reveal a time that completely different from Alice's and from Bob's – it could be years ahead or behind both of theirs!

Third, if they now stay together, they will see that their clocks remain exactly synchronized, besides the discrepancy they noticed when they met. This discrepancy doesn't increase or decrease. They may even retrace together Alice's and Bob's previous trips; their clocks still remain synchronized.

The experience just described will occur again any time two or more of them meet. There could be a hundred observers like Alice, Bob, Charlie, initially at the same place and synchronized. Whenever two or more of them meet after having been separated, they will notice discrepancies in their clocks. But their clocks will have exactly the same time lapses as long as they stay together. See the illustration in fig. 2.1



Figure 2.1 Illustration of the experiences of Alice (dashed 😎), Bob (solid 😊), Charlie (dot-dashed 😌) with time. The page represents a two-dimensional spacetime, and is followed from bottom to top.

Bottom: Alice, Bob, Charlie stay close and observe their clocks are perfectly synchronized from 12:00 to 12:10, then they separate.

Right: Charlie visits a region near a strong mass-energy source. Upon meeting again with Bob, the two notice their clocks differ: 16:00 for Bob, 12:30 for Charlie. Yet this clock difference stays the same while they travel together for 10 min.

Left: Alice wanders around travelling at high speed with respect to the fixed stars. At some point she wonders what's the time "now" for Bob and Charlie. Obviously this question doesn't make sense, because (1) when Bob and Charlie are together their clocks differ – impossible to say what's "the" time at their position; (2) not clear which instant in Bob & Charlie's trajectory should be considered as "now" (yellow dashed lines).

Top: When all three meet again, their clocks have completely different readings; and they themselves have aged differently. But their clocks run again at the same rate as long as they stay close.

Consider for a moment an imaginary world in which these experiments had given a different kind of result. According to Newtonian mechanics, whenever two or more initially synchronized observers like Alice, Bob, Charlie had met, their clocks would have always shown exactly the same time. If one year, 23 days, 8 hours, 9 minutes, and 3.045 399 283 240 992 663 02 seconds have passed for you since you last met Alice, you'd see that exactly the same amount of time has passed for her when you two meet again. If you think about it, in this case it would have been somewhat natural to think "right now, the clocks of far-away Alice, Bob, Charlie must show the same time as mine" (even though you have no real experimental way of confirming that).

But that's an imaginary world. In our world is the more complicated situation described initially that holds. Only one conclusion can be drawn from these experimental results: **Time is not some sort of universal quantity. It is, so to speak, "local" to a person or clock, or to a group of persons or clocks that stick together.** This also means that *it doesn't make sense to ask questions like "what can be the time for far-away Charlie, right now?"*

The time measured by a specific observer is called the **proper time** of that observer. Luckily we know more about how the proper times of separated observers can differ when they meet again. It turns out – according to our current understanding – that the time differences depend, roughly speaking, on how fast the observers are moving with respect to one another and to matter around the universe, and on how much energy is contained in the regions they travel. The general theory of relativity gives us the equations determining any such proper-time differences.

The situation depicted in the experiments above is real. It can be measured, for example, comparing initially synchronized clocks that have been put in aeroplanes flying in different directions. Most importantly, it affects everyday relevant technologies such as the Global Positioning System. Formulae from general relativity appear in your phone's GPS software; see for instance § 20.3.3.3 of the Interface Control Document IS-GPS-200 at <https://www.gps.gov/technical/icwg/>. It must also be taken into account in the establishment and synchronization of time in our everyday equipments:

International Atomic Time (TAI) is based on more than 250 atomic clocks distributed worldwide that provide its stability, whereas a small number of primary frequency standards provide its accuracy. Universal Coordinated Time, which is the basis of all legal time scales, is derived from TAI. To

"In 1976, the International Astronomical Union introduced relativistic concepts of time and the transformations between various time scales and reference systems. [...] Now [...] it is necessary to base all astrometry, reference systems, ephemerides, and observational reduction procedures on consistent relativistic grounds. This means that relativity must be accepted in its entirety, and that concepts, as well as practical problems, must be approached from a relativistic point of view."

Kovalevsky & Seidelmann 2004

allow the construction of TAI and the general dissemination of time, clocks separated by thousands of kilometres must be compared and synchronized. [...] The achieved performances of atomic clocks and time transfer techniques imply that the definition of time scales and the clock comparison procedures must be considered within the framework of general relativity.

(Petit & Wolf 2005)

In most everyday situations for us, who live on or nearby Earth and move at speeds much smaller than c with respect to one another, the discrepancies between our proper times are so small that cannot be measured with ordinary clocks or with our internal clocks. Consider a person walking 10 m away from you and then immediately walking back to you, at 1 m/s. The time elapsed for you will be 20 s, but for that person will be 19.999 999 999 999 999 889 s, a difference of 10^{-16} s, which is the error of an atomic clock. If human beings still exist in some decades or centuries, with space travel they will probably have to deal more and more with proper-time discrepancies also in everyday life.

For the most part of the rest of these notes, we won't need to deal with differences in proper time. But I recommend that you keep present how time really works, and that these small time discrepancies exist and occur all the time along your *worldline*.

Time has physical dimension of time and we shall for the most part measure it using the unit *second*, symbol 's'.

2.2 Space

Together with the notion of time, also the notion of space loses some of its traditional intuition. Several observers in motion with respect to one another will generally disagree on the dimensions of an approximately rigid object in their vicinity. For objects that are far away from an observer, the very notion of "distance" becomes tricky has different and non-equivalent definitions; one must be very careful on which definition is being used.

We shall not delve further into these peculiarities of time and space. Keep simply in mind that phenomena happen in **spacetime**, and that there's no way to attribute a universal time, nor a universal position in space, to a physical event. There is one absolute: whoever locally measures the speed of light, will find the value

$$c := 299\,792\,458 \text{ m/s} \quad (2.1)$$

"The plot for Cesium [...] characterizes the best orbiting clocks in the GPS system. What this means is that after initializing a Cesium clock, and leaving it alone for a day, it should be correct to within [...] 4 nanoseconds. Relativistic effects are huge compared to this."

Ashby 2003

This value is exact by definition, and serves as a way to define a local notion of space and distance.

Space has physical dimension of length and we shall for the most part measure it using the unit *metre*, symbol ‘m’.

2.3 Coordinate systems and events

It is necessary to have a way for distinguishing physical events and phenomena and locating them in spacetime. This is achieved through a **coordinate system**.

A coordinate system assigns four numerical labels to every point in spacetime. Often these labels have some kind of physical meaning – such as the proper time elapsed for a specific clock, or the distance from some event as measured by a specific observer – but they don’t need to.

A coordinate system also solves the problems coming from proper-time and space discrepancies among different observers. We can assign to every physical event a *coordinate time* and a *coordinate spatial position*, which are the same for all observers, because decided by agreement. Coordinate time doesn’t have a strict physical meaning, and will generally be different from the proper times registered by different observers. It can nevertheless be used for “doing physics”, and it is the time we shall most often use in our equations. A coordinate time commonly used for Earth-physics purposes is [Universal Coordinated Time \(UTC\)](#)¹. The clock on your phone, and on devices that get synchronized via internet, shows UTC, not your proper time. An observer on Earth at 0 m over sea level, and not moving, measures a proper time exactly equal to UTC (besides small variation coming from the movements of Solar System bodies). Observers at other altitudes or moving with respect to Earth’s surface notice that their proper times are slightly different from UTC.

Up to now we have often used the word ‘event’, informally taking its meaning for granted. Let’s be more precise: we call **event** or **spacetime point** an extremely small region of space – a point – that only lasts for a very small lapse of time – an instant. The name ‘event’ is used because typically we approximately identify such a point and instant by means of a physical phenomenon of limited spatial extension and short duration, such as the collision of two subatomic particles. The sudden burst of a very small soap bubble can be considered as an event in some circumstances; but something like “a tennis ball” cannot be considered as an event, mainly because a tennis ball exists for quite a long time, not just for a short instant.



A lattice of clocks and meters, defining a spacetime coordinate system (from Taylor & Wheeler 2000)

From a four-dimensional spacetime point of view, a tennis ball could be characterized as a line: a **worldline**.

We shall often denote the four coordinates of a coordinate system by the letters

$$(t, x, y, z)$$

where t is a coordinate time, usually UTC, and (x, y, z) determine a spatial position. The triplet of spatial coordinates is often denoted by the vector \mathbf{r} :

$$\mathbf{r} := (x, y, z) \quad \text{or} \quad \mathbf{r} := [x, y, z] \quad \text{or} \quad \mathbf{r} := \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

use round brackets ‘()’ or square brackets ‘[]’, and horizontal or vertical notation as you prefer.

It is always important to specify how the coordinate system you’re using is defined. The definition of the spatial coordinates (x, y, z) is typically different from problem to problem. We shall typically use coordinates that form $\frac{\pi}{2}$ rad $\equiv 90^\circ$ angles with one another; but their directions and their origin – that is, where they have value $x = y = z = 0$ m – always depend on the problem, so make sure you always specify them.

Whenever we speak of a “region of space” or of a “surface in space”, we mean a 3D or 2D region at some specific coordinate time t .

Some physical phenomena happen approximately along a line, in one dimension; or on a surface, in two dimensions. In these cases we can omit two or one of the spatial coordinates, assuming they have some constant, unimportant values; and we can simply write, for instance, (t, x) or (t, x, y) as our coordinates.



SPACETIME HEALTH TIP: REMEMBER TO CANCEL OUT YOUR ACCUMULATED TURNS AT THE END OF EACH DAY TO AVOID WORLDLINE TORSION.

<https://xkcd.com/2882>

Exercise 2.1

Consider a clock at rest on the Earth’s surface, at a distance r_e from the Earth’s centre; and a clock on a satellite, for instance a GPS satellite or the [International Space Station](#)², right above the first clock, at a distance r_s from the Earth’s centre. An observer by the clock on Earth measuring a time lapse Δt_e will see the clock on the satellite has having run for a time lapse Δt_s , and vice versa (note that this “vice versa” only holds in this specific situation!). The relation

between two time lapses is approximately given by

$$\frac{\Delta t_s}{\Delta t_e} = \frac{\sqrt{1 - 2\frac{G}{c^2} \frac{M}{r_e}}}{\sqrt{1 - 2\frac{G}{c^2} \frac{M}{r_s}}}$$

where $G \approx 6.7 \times 10^{-11} \text{ m}^3/(\text{kg s}^2)$, $c = 3.0 \times 10^8 \text{ m/s}$, and the Earth's mass $M = 6.0 \times 10^{24} \text{ kg}$.

1. Take the case of a GPS satellite, with $r_e = 6.4 \times 10^6 \text{ m}$ and $r_s = 2.6 \times 10^7 \text{ m}$ ([NASA data³](#)). If you, on the ground, measure a time lapse of $\Delta t_e = 10 \text{ years}$, what's the difference, in seconds, with the time lapse Δt_s you see on the satellite?
2. If the time lapses are large compared with the time needed to go from ground to orbit or vice versa, then $\Delta t_s/\Delta t_e$ is also the ratio between the real *ageing* of a person who's been in orbit and one who's been on the ground, when they meet again.

Now consider the case with a black hole instead of Earth. The formula above still apply approximately.

In the movie [Interstellar⁴](#), two astronauts go on Miller's planet, at a distance r_e from the black hole Gargantua, and stay there for 3 hours, leaving one astronaut in orbit at a distance $r_s \approx \infty$ (the distance is large enough that it can be approximated as infinity). When they meet again, the latter astronaut has aged 23 years.

Given that Gargantua's mass is $M = 2.0 \times 10^{38} \text{ kg}$, calculate the distance r_e of Miller's planet from the black hole.



2.4 Velocity and acceleration

In some situations the spatial coordinates $\mathbf{r} = (x, y, z)$ may turn out to be functions of the time coordinate t ; the typical example is when we describe how the spatial position of a small body changes with coordinate time. We can write this functional dependence in different ways, for instance

$$\mathbf{r}(t) \quad \text{or} \quad [x(t), y(t), z(t)].$$

So \mathbf{r} is a vector function of time, which simply means that we have a collection of three functions of time.

If we take the derivative of each coordinate with respect to the time t , we obtain the **coordinate velocity**

$$\mathbf{v}(t) := \frac{d}{dt} \mathbf{r}(t) = \left[\frac{d}{dt} x(t), \frac{d}{dt} y(t), \frac{d}{dt} z(t) \right]$$

which is also a vector.

Dot-notation for time derivative

The derivative of some quantity with respect to coordinate time is often denoted by a **dot** over the quantity. So we can also write

$$\mathbf{v}(t) = \dot{\mathbf{r}}(t) = [\dot{x}(t), \dot{y}(t), \dot{z}(t)]$$

The coordinate velocity is usually different from the *physical velocity*, which an observer would measure using proper time and space, for instance using bouncing light rays. In many everyday situations the difference between coordinate and physical velocity is so small that it can be neglected. But in situations involving subatomic particles at high speed, for example, one must take into account that the two velocities are different.

Taking the time derivative once more we obtain the coordinate acceleration, also a vector:

$$\mathbf{a}(t) := \frac{d}{dt} \mathbf{v}(t) = \frac{d^2}{dt^2} \mathbf{r}(t) = \left[\frac{d^2}{dt^2} x(t), \frac{d^2}{dt^2} y(t), \frac{d^2}{dt^2} z(t) \right]$$

which is also a vector.

Acceleration in relativity theory

In relativity theory, acceleration acquires a special physical significance, because it includes the effect of gravity, and its calculation does not involve just a time derivative. For instance, let's say that you are standing still on the ground, and let's use a coordinate system where x points in front of view, y to your left, and z points upwards. Then your coordinate velocity is $\mathbf{v} = (0, 0, 0)$ m/s also according to relativity theory. But the spatial part of your acceleration is approximately $(0, 0, 9.8)$ m/s², not zero!

The definitions and values of acceleration according to relativity theory and according to Newtonian mechanics are therefore quite different even in everyday situations. In these notes we'll mean the Newtonian definition of acceleration, unless stated otherwise.

Exercise 2.2

1. Here are the three components of a time-dependent velocity vector; the variable t is the time, and therefore has dimension time. Introduce units 's' and 'm' appropriately in such a way that the expression is dimensionally correct:

$$\mathbf{v}(t) = [4, \cos(3t), -\exp(8/t)]$$

2. The position vector of a satellite is given below. Calculate the satellite's velocity and acceleration vectors:

$$\mathbf{r}(t) = \begin{bmatrix} 2.0 \times 10^7 \cos\left(\frac{t}{13751\text{s}}\right) \\ 2.2 \times 10^7 \sin\left(\frac{t}{13751\text{s}}\right) \\ 0 \end{bmatrix} \text{ m}$$

3. What is the satellite's velocity at $t = 6875$ s? What is the magnitude of the velocity (that is, the speed)?

URLs for chapter 2

1. [https://www.nist.gov/pml/time-and-frequency-division/time-realization/utc
nist-time-scale-0/](https://www.nist.gov/pml/time-and-frequency-division/time-realization/utc-nist-time-scale-0/)
2. <https://www.nasa.gov/international-space-station/>
3. [https://www.nasa.gov/directorates/somd/space-communications-navigation-p
rogram/gps/](https://www.nasa.gov/directorates/somd/space-communications-navigation-program/gps/)
4. <https://www.imdb.com/title/tt0816692/>

Chapter 3

Building blocks: “Stuff”

For Euler, clarity was the hallmark of truth. [...] To him we owe also the brilliant imagination of the internal pressure in generality, the pressure field as equipollent to the action of the fluid outside any imaginary closed diaphragm upon that within. [...] I remark upon it in emphasis of the role of imagination and the importance of quantities which can only be thought of and cannot in themselves be measured.

C. A. Truesdell 1956

3.1 Seven primitive quantities with three basic properties

Besides time and space, our physics formalism includes around seven more quantities that we take as primitive:

Seven primitive quantities

matter
electric charge
magnetic flux
energy
momentum
angular momentum
entropy

Technically they are called *fields*, for reasons we shall shortly see.

Recall that primitive quantities cannot be defined: we can only try to understand them intuitively, for example through their properties. Some

of the seven primitive quantities are easier to grasp intuitively than others. But all seven primitive quantities have three basic properties in common. First of all, of each we can ask, or rather, measure:

The two basic measurements that can be made on the seven quantities

1. How much of this quantity is in a particular three-dimensional region of space at a particular time instant?
2. How much of this quantity flows through a particular two-dimensional surface during a particular time lapse?

We can ask these questions of any region of space and any time lapse, and the surface in the second question can be moving and deforming. The results of the two measurements above are numbers, which in general can be positive or negative, for scalar quantities; or vectors for vector quantities.

Often we consider the second question, about the flow through a surface, in the case of a very short lapse of time, and divide the total flow by that time lapse. So we have an alternative form of the second measurement:

Flux of a substance through a surface

- 2b. How much of this quantity is flowing through a particular two-dimensional surface per unit time, at a particular time instant?

This is called the **flux** of the quantity through that surface.

The third property common to the seven quantities regards the two measurements above.

Extensivity of the seven quantities

If we consider two or more separate volumes, the amount of quantity in the total volume is equal to the sum of the amounts in the separate volumes.

Analogously for the flux through separate surfaces.

We say that each of the seven quantities is an **extensive quantity**

The basic measurements above can't in general be made, and don't even make sense, for some other quantities. For instance, we cannot ask "what's the total amount of *temperature* in this region?", or "how much velocity is flowing through this surface?".

Thanks to the three properties above, each of the seven quantities can be intuitively visualized as some kind of “stuff” that can be present at each spacetime point, fills regions of space, and flows through surfaces. This visualization is useful, but also comes with some warnings which we’ll discuss later.

What’s remarkable about matter, electric charge, magnetic flux, energy, momentum, angular momentum, and entropy, is that *they are common to all our main physical theories*, approximate or not: from Newtonian mechanics to general relativity and quantum theory; from subatomic scales to cosmological scales. The physical meaning and mathematical characterization of these quantities can be slightly different depending on the physical theory and scale. For example, in quantum theory they are mathematically represented by so-called operators rather than functions; and at molecular scales entropy has a meaning connected with probability theory. Yet, these seven quantities are really universal to our present way of doing physics and of describing physical phenomena around and within us.

Let us make a first acquaintance with these seven quantities. The discussion that follows is meant as an introduction; we shall repeat and say more about each quantity in later chapters.

Keep in mind that it is very difficult, if not impossible, to answer questions like “what *really* is...?” or “why *really* does...?”.

3.2 Matter

Matter is a *scalar* quantity, with SI dimension [amount of substance](#)², and measured in units of [moles \(mol\)](#)³. Its flux is therefore measured in units of *moles per second* (mol/s). In statistical mechanics and particle physics, matter is often simply counted and so measured in dimensionless units, rather than moles.

Matter is probably the easiest quantity to grasp intuitively; what we ordinarily call “stuff” is matter. It is usually classified into several kinds; the classification depends on the physical phenomena and physical theory one works with, and is related to whether the kinds of matter can be considered separately conserved, as we’ll discuss in Chapter 7.

In everyday phenomena not involving [radioactivity](#)⁴ or [nuclear energy](#)⁵, the different kinds of matter approximately correspond to the [non-radioactive chemical elements](#)⁶: [hydrogen](#)⁷, [helium](#)⁸, [lithium](#)⁹, and so



Richard Feynman explains the difficulty of “why” questions in a [funny and insightful video](#)¹.

on. Note that some common everyday devices, such as smoke detectors, do involve radioactivity.

In phenomena involving radioactivity or nuclear energy, the different kinds of matter correspond approximately to *baryons*¹⁰, such as protons and neutrons; and *leptons*¹¹, such as electrons. In particle physics, even more subtle classifications of matter are made, into kinds that seem to be conserved, such as electronic-leptons, muonic-leptons, and others. This kind of research is still open, but it seems that the total amount of baryonic and leptonic matter, independently of the kinds into which it can be classified, is always conserved. Note that we're using the term 'matter' in a sense that includes anti-matter, such as positrons.

The total amount of matter in a region *can be negative*. "Negative matter" is what's usually called *anti-matter*¹². Anti-matter appears in small amounts in everyday life, for example in connection with common radioactivity processes. It is also created and used in medicine, in *positron-emission tomography (PET)*¹³ scans.

In these notes we shall usually not consider distinctions between different kinds of matter, making some exceptions for discussions about chemical reactions and nuclear phenomena.

Matter: notation

The amount of matter in a region is usually denoted with N , and flux of matter with J . In chemistry we usually specify what kind of matter we are speaking about, writing for instance $N_{\text{Ca}} = 5.3 \text{ mol}$, to indicate an amount of 5.3 mol of *calcium*¹⁴ atoms.

Matter is different from mass or energy

It is important to clearly distinguish matter from *mass* or *energy*. Mass can be considered a property of matter, but the two are different. In nuclear reactions, for instance, the mass of some amount of matter may change, even if the amount of matter stays the same.

As far as we know, the total amount of energy associated with an amount of matter is always positive, whether the amount of matter is positive or negative (antimatter). This is the reason why antimatter "falls" just like positive matter, a fact that has been experimentally confirmed: see Anderson et al. 2023.

Exercise 3.1

According to statements on symmetrymagazine.org¹⁵ and quantum-diaries.org¹⁶,

The average banana (rich in potassium) produces a positron roughly once every 75 minutes.

Unfortunately the original site where this statement was discussed, and the corresponding calculation made, seems not to exist anymore.

1. Do a little research and find out whether this statement is true.
2. From your research, approximately quantify the flux of positrons around an ordinary banana, expressing it in particles/s.



How many positrons do bananas produce?

3.3 Electric charge and magnetic flux

Electric charge is a *scalar* quantity, with SI dimension [electric charge](#)¹⁷ (equivalent to electric current×time), and measured in units of [coulombs](#) (C)¹⁸. Flux of electric charge is called *electric current*, and measured in units of [amperes](#) ($A = C/s$)¹⁹.

Electric charge is a quantity that is easily grasped in our everyday experience, and doesn't need much comments.

Magnetic flux is a *vector* quantity, with SI dimension [magnetic flux](#)²⁰, and measured in units of [webers](#) (Wb)²¹.

The electromagnetic field is most commonly represented by vectors associated to each point in space. But it can also be interpreted and visualized as a collection of moving, oriented tubes or lines, either closed or extending indefinitely; somewhat analogously to how we visualize matter and charge, as moving blobs or points, but with one more dimension. This interpretation goes back to Faraday (1846), Maxwell (1855), and later Dirac (1955) among others, and today is conveniently used in some fields such as [solar physics](#)²², for example to study [sunspots](#)²³ (see Ryutova 2018). From this point of view, the *voltage* turns out to be the flux of the magnetic flux; it is indeed measured in *volts* (V), equivalent to *webers/second*.

Electromagnetism and this particular visualization of it are very fascinating topics, but we shall not discuss them in these notes.



"sketch of the magnetic lines of force in a magnetic filament extending up through the photosphere." Parker 1974a

3.4 Energy

Energy is a *scalar* quantity, with SI dimension energy, and measured in units of *joules* (J); its flux is measured in units of *watts* (W = J/s).

Equivalently we can speak of mass, with SI dimension mass, and measured in units of *kilograms* (kg)²⁴; its flux is measured in *kilograms per second* (kg/s).

The notion of energy is extremely important today, and central in many world-wide important discussions and worries – think of the “energy crisis”, “renewable energy”, and so on. It is somewhat funny that despite this importance it’s actually difficult to answer ‘what *is* energy, really?’. Often we speak about energy as something that “flows”, is “transported”, “converted”, “stored”, and similar visualizations. This intuition will be enough in these notes. The notion of *mass* is also very intuitive in our everyday life; we associate it with the difficulty in setting objects into motion, or with the weight of objects.



From Relativity Theory and experimentally we know that *energy and mass are the same quantity*, and in these notes we shall emphasize this experimental fact.

3.4.1 Energy and mass are the same

Let’s see some examples of why it is impossible to make a real distinction between energy and mass. The following examples have been simplified in some of their aspects, but their main point is valid.

Heated gas Imagine we have a box with a given amount of gas, say 1 mol of *oxygen*²⁵ molecules. Using an extremely precise weighing scale, we observe that the mass of the gas is, say, exactly

$$0.031\,999\,540\,000\,000\,000 \text{ kg} .$$

Now we heat the gas, providing 60 J, while making sure that not a single molecule of oxygen gets in or out of the box. The temperature of the gas increases by around 3 K. We observe that the weight measured by the scale increases while we heat the gas, reaching the new value

$$0.031\,999\,540\,000\,000\,668 \text{ kg} .$$

Clearly the mass has increased, but no molecules were added! The additional mass is the 60 J that we provided to the gas. Energy has weight, energy is mass.

Stretched or moving rubber band Take a common rubber band, and imagine again that we have an extremely precise weighing scale. The rubber band, unstretched, has a mass of exactly

0.000 500 000 000 000 000 kg .

Now we stretch the band a little. According to Newtonian mechanics it acquires ‘elastic energy’, which increases its initial internal energy by, say, 0.3 J. Now we weigh the rubber band again, while stretched, and observe a mass of approximately

0.000 500 000 000 000 003 338 kg .

The extremely small difference of around 3×10^{-18} kg from the initial mass is exactly the elastic energy that we provided by stretching. Energy has weight; energy is mass.



When we stretch a rubber band, its mass increases slightly – even if the amount of rubber remains exactly the same.

Now set the unstretched band in motion, so that according to Newtonian mechanics it acquires a kinetic energy of, say, 0.2 J. If we could weigh the band while in motion (but without moving the weighing scale), we would measure a mass of approximately

0.000 500 000 000 000 002 225 kg .

The small difference from the initial mass is the new kinetic energy of the band.

This case is actually connected with the example of the gas above. If we observed the gas at a molecular level, we would interpret the energy of 60 J provided to it as additional kinetic energy of its molecules. The increase in weight was exactly this additional kinetic energy.

Fission and atomic bombs The atomic bomb²⁷ is a dark example of the fact that mass is energy. In the case of nuclear fission, if we weigh the amount of nuclear material, say within a box, before and after fission, we observe that its mass has decreased. But we also observe that a great amount of energy has been released out of the box.



Hydrogen Bomb Test, 1954 (National Museum of Nuclear Science & History²⁶)

Electric heater As a final example consider a 1000 W electric heater, which is radiating 1000 J in one second. The heater is also losing around 0.000 000 000 01 kg of mass every second owing to this heat radiation – although it's also acquiring the same amount of mass as electromagnetic energy.

The equivalence between energy and mass is given by the famous formula $E = mc^2$, where c is the speed of light, eq. (2.1). In their respective units this gives

$$1 \text{ kg} = 89\,875\,517\,873\,681\,764 \text{ J} \quad (\text{exactly})$$

$$1 \text{ J} \approx 0.000\,000\,000\,000\,000\,011\,126\,5 \text{ kg}$$

To grasp these numbers, consider that the mass of the rubber band in the example above, 0.5 g, is comparable to the energy released by the [atomic bomb over Hiroshima](#)²⁸.

"we are led to the more general conclusion: The mass of a body is a measure of its energy content; if the energy changes by L , the mass changes in the same sense by $L/9 \cdot 10^{20}$, if the energy is measured in ergs and the mass in grams." Einstein 1905b

3.4.2 The practical use of the words 'mass' and 'energy'

From the examples above it becomes clear that energy and mass are two names for the same thing. But it also becomes clear that in our daily experience we deal with energy-mass in two different ways:

On the one hand, we deal with huge (atom-bomb-like) amounts of energy packed in very small volumes: the huge amount of energy that goes together with objects such as pens, keys, bicycles, cars, houses, and so on. We move, push, pull these huge energy amounts from one place to another, and even put them in our pockets. These energy amounts change a little all the time; see the examples above. But the change is so small as to be often undetectable with ordinary scales, and negligible for practical purposes. We call this energy 'mass' and measure it with a unit – kg – that doesn't lead to ridiculously large numbers. And we also agree to neglect the imprecision and fluctuation in its measurement, say any imprecision under 0.000 001 %.

On the other hand, we also deal with the small energy changes and exchanges in all these objects. These exchanges that are very important for our daily life: they keep us warm, make our cells work, make our laptops work. In dealing with these energy exchanges, we don't care about the huge energy reservoirs they come from. So we agree to measure them with a unit – J – that doesn't lead to ridiculously small numbers. And we also agree not to be precise about the total amount in the reservoir from which these energy bits come from.

As an analogy, consider when we speak about the amount of people in different countries. We can say that in Norway there are 5 millions, and in India 1500 millions, so in India there are 300 times more people. By this we don't mean that in Norway there are *exactly* 5 000 000 people and

that India has *exactly* 300.000 000 times more people. These numbers are changing slightly all the time; but we don't care about differences of 10 or even 10 000 people. At the same time, think of when you have three dear friends or relatives visiting you from abroad: the amount of 3 people is now for you very important; and you don't care about how much this amount is compared to the total population of your country.

The distinction above is of course not clear-cut. In dealing with some physical phenomena, for example with few molecules or in particle physics, the distinction become too blurry and not useful anymore. And indeed in these situations one often uses the terms 'mass' and 'energy' interchangeably, as well as a common unit for both (for instance *electronvolts*²⁹).

In these notes we shall often use the expressions 'energy-mass' and 'mass-energy' to remember that these two words denote the same thing.

3.4.3 Different 'forms' of energy

We often speak of different *forms* of energy-mass. The most important for us will be **internal energy**, **kinetic energy**, **gravitational potential energy**, to be discussed later; another important one is *electromagnetic energy*.

In Chapter 5 we shall see that the differences among these energy forms come from the way they are calculated from other quantities, such as matter or magnetic flux and electric charge. For example, if we know that in a volume there's an amount of a particular kind of matter, then we know that there must also be an amount of energy given by a particular formula. And if that matter is moving, then we have to add to the total an extra amount of energy given by another formula. And if in that volume there's a gravitational field (that is, a particular kind of spacetime curvature), then another extra amount must be added, given by yet another formula. Similarly if we know that an electromagnetic field is in that volume.

We also speak of different forms of flux of energy. The most important for us will be **heat** and **work**, and possibly **convection**. The difference is again in how these fluxes are calculated depending on whether there are also fluxes of matter.

3.4.4 Amounts and forms of energy are coordinate- and observer-dependent

An aspect of energy that must always be kept in mind is that **the amount of energy depends on the coordinate system we're using**. If someone

points at a specific region of space at a particular instant, and asks “how much energy is there?”, we *cannot* give an answer until a coordinate system is specified. Once the coordinate system has been chosen, then a precise and unambiguous answer can be given. The same is true for the flow of energy through a surface. This also means that observers using different coordinates will usually assign different amounts of energy to the same regions of spacetime.

This is an important difference between energy on one side, and matter and electric charge on the other side. *For matter and electric charge, the questions above can be answered unambiguously independently of any coordinate system.* But not so for energy. The reason of this quirky difference is ultimately connected with the fact that time and space are also observer-dependent.

This coordinate-dependence is not a problem: we must always specify our coordinate system anyway, in order to agree on the time and position of physical events. But it can cause problems when we calculate *changes* in energy. If we first calculate or measure some amount of energy in a coordinate system, then we calculate or measure another amount at some later time in a *different* coordinate system, the difference between the two amounts has no meaning whatsoever. In fact it could happen that there's no energy change at all in one coordinate system or in the other: any change we found was just an artefact of mixing up coordinates.



“How much energy is there in this volume at this instant?” This question cannot be answered until we have specified which coordinate system we're using.

! Amounts of energy-mass are coordinate-dependent

Never change coordinate system in the middle of energy calculations!

Also the distinction among different forms of energy is coordinate-dependent. For instance, in one coordinate system we can say that a given volume contains only internal energy; but in another coordinate system that same volume can be said to contain internal and kinetic energy.

The distinction among different forms of energy also depends on the observation scale and on the theory used. For instance, we can observe and model the gas in a container on a scale of metres, seeing it as a uniform flux; in this case we say that there's internal energy in the container. But if we observe and model the gas as a collection of molecules, on a microscopic scale, then we say that there's internal energy (of the single molecules) and kinetic energy in the container (the total energy being the same, unless we have changed coordinate system).



Hydrocarbon fuel particles³⁰. The small blobs have size of around 2×10^{-8} m.

The same is true of energy flux: what we call ‘heat’ on one observation scale is ‘work’ on a finer scale.

Energy-mass: notation

The amount of energy in a region is usually denoted with E , or with m if we describe it as mass. Internal energy is denoted U , kinetic energy E_k , potential energy E_p .

The total flux of energy will be denoted by Φ ; the flux in the form of heat, by Q ; and in the form of work, by W .

Exercise 3.2

In an hour, 14 people exit through a door. Taking the average human weight to be 62 kg (Walpole et al. 2012), what’s the average *energy* flux, in J/s, through that door?

3.5 Momentum

Momentum, also called *linear momentum* or *translational momentum* to distinguish it from angular momentum, is a *vector* quantity. Its SI dimension and units can be written in several equivalent ways; we shall keep in mind especially these three:

$$\begin{array}{ccc} \text{force} \times \text{time} & \equiv & \text{mass} \times \text{length/time} \\ \text{N} \cdot \text{s} & & \text{kg} \cdot \text{m/s} \end{array} \quad \equiv \quad \begin{array}{c} \text{energy} \times \text{time/length} \\ \text{J} \cdot \text{s/m} \end{array}$$

Since it is a vector quantity, it is usually expressed with three numbers, typically the x -, y -, and z -components. In simplified problems where only one or two dimensions are relevant, only the relevant components are reported.

Momentum is a subtle quantity, even subtler than energy. Textbooks that focus on Newtonian mechanics *define* it as the product of the mass and the velocity of a body, usually written “ $\mathbf{p} = m\mathbf{v}$ ”. This relation, however, is only valid in special circumstances, and cannot be used in many everyday technological applications, especially when electromagnetism is involved. And that relation is actually only an approximation even in the circumstances where it’s used.

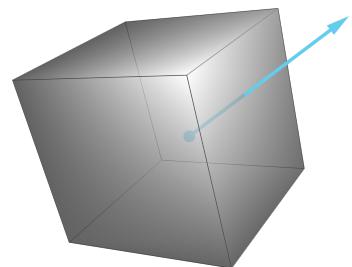
It is therefore convenient to separate our idea of momentum from the idea of “objects” moving, keeping in mind that the latter idea is just a particular case of momentum. Yet, *momentum is indeed associated with translational motion* of matter and of electromagnetic fields. *Translational motion* is the kind of motion that leads to a *new position* in space. For instance, when you walk from one place to a different one, you have performed translational motion (note that translational motion doesn’t need to be in a straight line).

Just as energy is often mentally visualized as a sort of fluid (although this visualization comes with many warnings), also momentum can actually be visualized as a sort fluid; but you must imagine it as a “fluid of vectors”. Given a particular volume at a particular instant in time, and given a coordinate system, we can speak of the total amount of momentum within that volume. This amount is represented by a *vector*. You can imagine a continuous collection of vectors filling the volume, possibly with different directions and small magnitudes; the total momentum is the sum of all these vectors. This visualization obviously comes with many warnings, but it can be very useful if we are careful.

And just as energy, the total amount of momentum is *coordinate-dependent*. So we have the same warning here:

! Amounts of momentum are coordinate-dependent

Don’t change coordinate system in the middle of calculations of momentum!



The amount of momentum within a volume is represented by a (3D) vector

Flux of momentum is what we call **force**. A static force, like the one you exert when you hold a bag, is often mentally visualized as a static vector. In Chapter 4 we shall discuss a different visualization, in which force is represented as a sort of flow of vectors.

! Momentum: notation

The amount of momentum in a region is usually denoted with **P** .
The flux of momentum is also called force and denoted with **F** .

! Momentum and energy flux are the same

According to Relativity Theory, momentum *is* energy flux (such as heat), and energy flux *is* momentum. If we represent momentum with the symbol **P** , and energy flux with the symbol **Q** , the equivalence between

them is given by

$$\mathbf{Q} = \mathbf{P}c^2 \quad (3.1)$$

Compare this formula with $E = mc^2$. From this point of view, you can think of momentum as “energy in motion”. This is consistent with our discussion about mass-energy: since mass is energy, the Newtonian expression “ $m\mathbf{v}$ ” indicates energy in motion, or a flux of energy. On a sunny day, if you close your eyes and feel the Sun’s heat on your face, what you are feeling is actually a flow of momentum. And when you kick a ball, you’re setting a huge bundle of energy in motion, and that’s why the ball has acquired momentum.

3.6 Angular momentum

Angular momentum, also called *moment of momentum* or *rotational momentum*, is a *vector* quantity. Its SI dimension and units can be written in several equivalent ways; we shall keep in mind especially these three:

$$\begin{array}{ccc} \text{force} \times \text{length} \times \text{time} & \equiv & \text{mass} \times \text{length}^2/\text{time} \\ \text{N} \cdot \text{m} \cdot \text{s} & \equiv & \text{kg} \cdot \text{m}^2/\text{s} \end{array} \equiv \text{energy} \times \text{time} \quad \text{J} \cdot \text{s}$$

It is usually expressed with three numbers, typically the x -, y -, and z -components.

Angular momentum is probably an even subtler quantity than momentum. Just as momentum is associated with translational motion, angular momentum is associated with *rotational* motion. Rotational motion is the kind of motion that leads to a *new orientation* in space, rather than to a new position. For instance, if you turn to your left or to your right while standing in place, you have performed rotational motion.

There isn’t a clear-cut distinction between translational and rotational motion: usually they involve each other to some degree. A translational motion can be interpreted as a rotation around a point that is very far away; and a rotation of an extended object can be interpreted as small translational motions of its parts.

This is the reason why in many situations we can calculate angular momentum in terms of momentum. If the momentum in a *small* volume is denoted by the vector $\mathbf{P} = (P_x, P_y, P_z)$, and the position vector by

$\mathbf{r} = (x, y, z)$, then the angular momentum $\mathbf{L} = (L_x, L_y, L_z)$ with respect to the origin of coordinates, in that same volume, is given by the vector product

$$\mathbf{L} = \mathbf{r} \times \mathbf{P}$$

or equivalently
$$\begin{cases} L_x = y P_z - z P_y \\ L_y = z P_x - x P_z \\ L_z = x P_y - y P_x \end{cases} \quad (3.2a)$$

Instead of calling the components " (L_x, L_y, L_z) ", we can also call them " (L_{yz}, L_{zx}, L_{xy}) ", as some books do. The last names make the formulae above easier to remember:

$$\begin{cases} L_{yz} = y P_z - z P_y \\ L_{zx} = z P_x - x P_z \\ L_{xy} = x P_y - y P_x \end{cases} \quad (3.2b)$$

Choose whichever you prefer.

Angular momentum is something that is associated not only with ordinary bodies (matter), but also with electromagnetic fields. Just like momentum, also angular momentum can be visualized as a "fluid of vectors". Given a particular volume at a particular instant in time, and given a coordinate system, we can speak of the total amount of angular momentum within that volume. This amount is represented by a *vector*, and is *coordinate-dependent*:

⚠️ Amounts of angular momentum are coordinate-dependent

Don't change coordinate system in the middle of calculations of angular momentum!



The amount of angular momentum within a volume is represented by a (3D) vector. (Curious about the little circulation symbol around the vector? Then read on.)

The flux of angular momentum is also called the *torque*, and bears a relation to the flux of momentum similar to the formula above. If the force – flux of momentum – is denoted by the vector $\mathbf{F} = (F_x, F_y, F_z)$ and the position vector by $\mathbf{r} = (x, y, z)$, then the flux of angular momentum, or torque, $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)$ with respect to the origin of coordinates is given by

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

or equivalently
$$\begin{cases} \tau_x = y F_z - z F_y \\ \tau_y = z F_x - x F_z \\ \tau_z = x F_y - y F_x \end{cases} \quad \text{or} \quad \begin{cases} \tau_{yz} = y F_z - z F_y \\ \tau_{zx} = z F_x - x F_z \\ \tau_{xy} = x F_y - y F_x \end{cases} \quad (3.3)$$

You may wonder: "Do we really need angular momentum? after all it just looks like something constructed from momentum". The answer is yes, we really need it, for two reasons. First, angular momentum obeys an important universal law which is independent from those obeyed by energy and by momentum (Truesdell 1968a tells some of the story of how this was discovered). Second, for some physical phenomena, for example involving liquid polymers³¹, elementary particles, or electromagnetic radiation, the angular momentum includes an additional part, called *spin* or *intrinsic angular momentum*, that is *not* related to linear momentum. In the present notes we shall not use this more general kind of angular momentum.



Some liquid polymers (**top**: Liquid Diethoxymethane Polysulfide) need to be described with a special kind of angular momentum, owing to their molecular structure (**bottom**).

3.6.1 Angular momentum as a twisted vector

In order to represent angular momentum we can use a kind of vectors different from the arrow-like ones (called *polar* vectors) with which you are probably familiar. They are called **twisted vectors**, or also *pseudo-vectors* or *axial* vectors or *outer-oriented* vectors. Twisted vectors represent rotations, and therefore have an orientation, not along them, but *around* them:

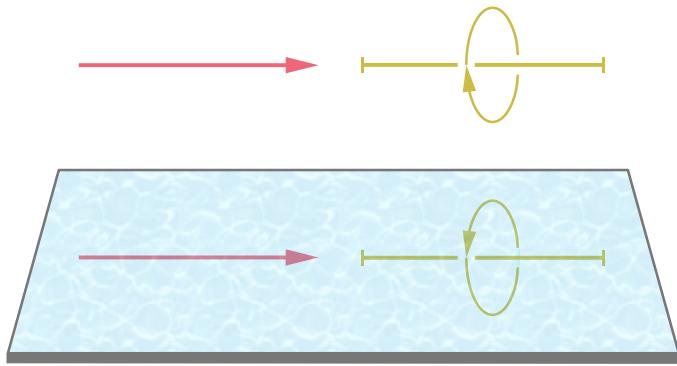


Their length still represents the magnitude of the vector. They make it immediately clear what is the axis of rotation, and what is the sense of rotation.

The sum of twisted vectors is analogous to the sum of ordinary vectors, with the parallelogram rule:



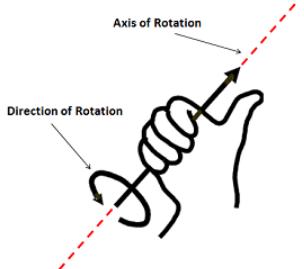
Ordinary vectors and twisted vectors behave very differently if we look at their images through a mirror parallel to their axis: the orientation of ordinary vectors appears unchanged, whereas the orientation of twisted vectors appears *reversed*:



this phenomenon reflects the behaviour of rotations under reflections.

For some mysterious reason many books are afraid of using twisted vectors, and rely on ordinary vectors instead, introducing the “right-hand rule” to determine the sense of rotation from the arrow of the ordinary vector. If you’ve ever asked yourself “why the right hand, and not the left hand?”, the answer is that it’s purely a convention; one could have introduced a left-hand rule instead. Using twisted vectors we don’t need these arbitrary conventions and mnemonics: the sense of rotation is unequivocally indicated by the twisted vector.

Use whichever vector representation you prefer. In these notes we shall use a hybrid notation, as in the side picture of p. 43, so as to make everybody (or no one) happy.



Exercise 3.3

A GPS satellite has, at a given instant, the following position and momentum:

$$\mathbf{r} = [1.4 \times 10^{+7}, 1.6 \times 10^{+7}, 0] \text{ m}$$

$$\mathbf{P} = [-3.1 \times 10^{+6}, 3.4 \times 10^{+6}, 0] \text{ N} \cdot \text{s}$$

Assuming that the satellite's volume can be considered as small for some purpose, calculate the satellite's angular momentum.

What are energy, momentum, angular momentum?

From the discussions and formulae above, it seems that energy-mass, momentum, angular momentum are quite closely related to one another. For all three, the amount in a volume or through a surface is undefined unless we specify a coordinate system. And we shall see later that all three satisfy balance laws but not necessarily conservation laws.

Relativity Theory indeed shows that energy, momentum, angular momentum are different aspects of one single geometric object, called *energy-momentum tensor*. They are like its "shadows", that we can observe by looking at it from different points of view in time and space. This is also why their values get intermixed if we change our system of coordinates.

General Relativity gives a new meaning to these quantities: they are *particular curvatures of spacetime*. They express how spacetime is curved in different directions. So whenever we measure, say, the energy or the momentum of some object or of some electromagnetic radiation, we are actually measuring how much that object or radiation is curving spacetime in a particular way.



Energy, momentum, angular momentum are measures of particular curvatures of spacetime.

3.7 Entropy

Entropy is a *scalar* quantity, with SI dimension energy/temperature, and measured in units of *joules/kelvins (J/K)*³². From this definition it would seem that entropy is derived from temperature. However, although temperature is taken as primitive by the SI, the [definition of temperature](#)³³ actually depends on a fixed value of *Boltzmann's constant*³⁴, which has the dimension of entropy.

Entropy is probably the most difficult quantity to grasp intuitively. Many seemingly intuitive descriptions given in some textbooks are, unfortunately, unhelpful and even misleading. One particularly misleading intuition is that entropy would be a “measure of disorder”. Besides the fact that “disorder” is very vague and subjective, it turns out that some physical phenomena, for example with liquid crystals³⁵, can be considered more “disordered”, and yet have *lower* entropy, than others. See also the example in the side figure. We shall discuss more about such phenomena later on.

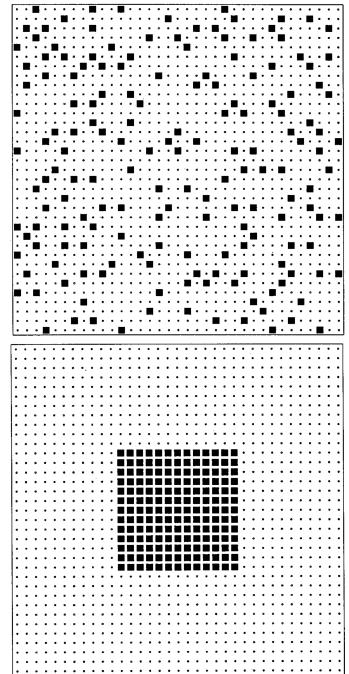
In these notes we shall rely on the idea that *entropy expresses a limit on the flux of energy into matter*. Said in simpler but more imprecise words, entropy is a bound on how fast we can heat up a body. We shall develop this idea further later.

One reason why entropy is difficult to grasp intuitively is that it has very different physical and mathematical aspects depending on the spatial scales and physical theory that we use to describe physical phenomena.

In many “continuum” phenomena, that is, phenomena where the molecular constitution of matter is not visible or not taken into account, entropy is treated as a “stuff-like” quantity similar to energy or electric charge. But there are difficulties also in this case. For some phenomena, for example involving non-elastic materials such as a simple paper clip, it is possible to introduce several entropies having different values – and not just because of a change in measuring scale – all of which can serve their purpose perfectly fine.

In molecular phenomena involving statistical mechanics, on the other hand, entropy is no longer a physical notion, but a *probabilistic* and *statistical* one, related to guesses and inferences that we make about the physical phenomenon. Yet from many points of view it has roles similar to those of the entropy used in continuum phenomena.

We shall see later that the physical laws for entropy have also a different status with respect to the laws for the other six main quantities: they are, so to speak, “laws about laws”.



Microscopic configurations of a lattice gas. Left: a configuration coming from a *low-entropy* state. Right: a configuration coming from a *high-entropy* state (Styer 2000).

Entropy: notation

The amount of entropy in a region is usually denoted with S . We shall see that the flux of entropy is tightly related to heat, and we won't need a special symbol for it.

3.8 Metric

A very important quantity constitutes an eighth fundamental building block of all our physical theories: the **metric**. It is quite different from the seven fundamental quantities, from both a physical and a geometrical point of view. For instance it is *not* an extensive quantity: we can't ask "what's the total amount of metric in this region?" – it would be a meaningless question.

The metric characterizes our measurements of space and time. It's the object that allows us to calculate how much *physical* time has elapsed, or the *physical* distance of an object. In General Relativity it allows us to calculate the curvature of spacetime.

In the Newtonian approximation, that is, for speeds smaller than the speed of light and low energy densities (hence weak gravitational fields and small spacetime curvature), the metric is just a static, uniform background object, and spacetime has no curvature. This is why we can speak of an 'absolute time' and 'absolute distances' in this approximation. In these notes we shall for the most part use this Newtonian approximation.

In General Relativity the metric is a dynamic object instead: it can change in time, and can vary from region to region. These changes are determined by the seven fundamental quantities, and the metric, in turn, determines changes in the seven quantities.

3.9 Auxiliary quantities

Besides the seven principal quantities, other auxiliary quantities appear in some physical theories. Important examples are **temperature** and **strain**. Most auxiliary quantities are not extensive. For instance we cannot ask "what's the total amount of temperature in this region?". We shall later discuss and use some auxiliary quantities, especially temperature.

The dimensions, units, and scalar or vector character of all quantities mentioned so far are summarized in table 3.1.

Quantity	SI Dimension	Unit
Time	time	<i>second</i> s
Length	length	<i>metre</i> m
Temperature	temperature	<ikelvin< i=""> K</ikelvin<>
Matter	amount of substance	<i>mole</i> mol
Electric charge	electric charge	<i>coulomb</i> C
Magnetic flux	magnetic flux	<i>weber</i> Wb
Energy	energy, mass	<i>joule</i> J, <i>kilogram</i> kg
Momentum	force · time, mass · length/time, energy · time/length	N · s, kg · m/s, J · s/m
Angular momentum	force · length · time, mass · length ² /time, energy · time	N · m · s, kg · m ² /s, J · s
Entropy	energy/temperature	J/K

Table 3.1 Dimensions and units of the main physical quantities used in these notes.
 Quantities in **boldface** are vectors, the others are scalars

URLs for chapter 3

1. <https://www.youtube.com/watch?v=nYg6jzotiAc&t=893s>
2. <https://doi.org/10.1351/goldbook.A00297>
3. <https://doi.org/10.1351/goldbook.M03980>
4. <https://www.iaea.org/newscenter/news/what-are-radioactive-sources>
5. <https://www.iaea.org/newscenter/news/what-is-nuclear-energy-the-science-of-nuclear-power>
6. <https://www.ciaaw.org/radioactive-elements.htm>
7. <https://pubchem.ncbi.nlm.nih.gov/element/Hydrogen>
8. <https://pubchem.ncbi.nlm.nih.gov/element/Helium>
9. <https://pubchem.ncbi.nlm.nih.gov/element/Lithium>
10. <http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/hadron.html#c6>
11. <http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/lepton.html#c1>
12. <https://www.britannica.com/science/antimatter>
13. <https://www.britannica.com/topic/positron-emission-tomography>
14. <https://pubchem.ncbi.nlm.nih.gov/element/Calcium>
15. <https://www.symmetrymagazine.org/2009/07/23/antimatter-from-bananas>
16. <https://www.quantumdiaries.org/2009/07/21/positrons-from-bananas/>
17. <https://doi.org/10.1351/goldbook.E01923>
18. <https://doi.org/10.1351/goldbook.C01365>
19. <https://doi.org/10.1351/goldbook.A00300>
20. <https://doi.org/10.1351/goldbook.M03684>
21. <https://doi.org/10.1351/goldbook.W06666>
22. <https://doi.org/10.1093/acrefore/9780190871994.013.21>
23. <https://spaceplace.nasa.gov/solar-activity/>
24. <https://doi.org/10.1351/goldbook.K03391>
25. <https://pubchem.ncbi.nlm.nih.gov/element/Oxygen>
26. <https://nuclearmuseum.pastperfectonline.com/Archive/716477C1-5E7A-485C-8BE1-857919471563>
27. <https://www.britannica.com/science/nuclear-fission>
28. <https://www.britannica.com/story/atomic-bombing-of-hiroshima>
29. <https://home.cern/tags/13-tev>
30. <https://doi.org/10.4209/aaqr.2019.04.0177>
31. <https://www.britannica.com/science/polymer>
32. <https://doi.org/10.1351/goldbook.C01365>
33. <https://doi.org/10.1351/goldbook.K03374>
34. <https://doi.org/10.1351/goldbook.B00695>
35. <https://www.britannica.com/science/liquid-crystal>

Chapter 4

Fluxes and volume contents

For the sake of persons of these different types, scientific truth should be presented in different forms, and should be regarded as equally scientific, whether it appears in the robust form and the vivid colouring of a physical illustration, or in the tenuity and paleness of a symbolical expression.

J. Clerk Maxwell 1870

Recall that the [main seven quantities](#) – matter, electric charge, magnetic flux, energy, momentum, angular momentum, and entropy – have three common properties related to their measurement:

◀ § 3 page 30

- (1) We can measure the amount of quantity within a three-dimensional region, at a specific time instant
- (2a) We can measure the amount of quantity flowing through a two-dimensional surface during a time lapse...
- (2b) ... or alternatively we can measure the amount of quantity flowing through a two-dimensional surface *per unit time*, at a particular time instant
- (3) The amount in a volume consisting of separate volumes is equal to the total of the separate amounts. Similarly for the flux through separate surfaces.

Let's give definite names to the measurements (1) and (2b):

volume content and flux

We can call measurement (1) the **volume content** or **volume integral** of the quantity.

We call measurement (2b) the **flux** of the quantity.

These two notions and measurements are very intuitive; that's also why it's convenient to base our physics upon them. In this chapter we straighten some details about their definition and also about our intuition.

4.1 Control volumes and control surfaces

A first question that comes to mind is: how are the volumes and surfaces chosen?

The choice of volume (and therefore of the surface bounding it) for a volume content, and the choice of surface for a flux are *completely arbitrary*, and they can be *completely imaginary*. Since they are under our control, they are called **control volumes** and **control surfaces**.

For example, consider a classroom and the people in it. In your imagination you can divide the classroom into two halves, say the front and the rear half. You can then ask or measure simply by counting: (1) how many people are, right now, in the rear half; (2) how many people are crossing the imaginary division between the front and rear half during one minute, starting from now.

A control volume and a control surface don't need to be static: they can move and deform.

In the case of a control volume, movement doesn't matter: the volume content in a control volume *does not depend on the instantaneous motion of the volume*. In fact we can even imagine a control volume that exists for just one time instant.

In the case of a control surface the situation is different. The flux through it *depends also on the motion of the surface*. As a trivial example, consider a glass surface, and a person on one side of it, moving with a high velocity directed towards the surface. Will the person crash on the glass? We can't say for sure. The glass surface could be a glass wall in a building, which is not moving; in this case the person will likely crash on it. Or it could be the windscreens of a car, which is moving together with the person, who's the driver; in this case the person won't crash on it.



The golf ball is moving leftwards. Will it hit the metal surface? We don't know unless we know how the surface is moving.

So we can't just imagine a surface that exists for one time instant: we need to imagine it for a very short time lapse, and be able to say how it's moving. If someone asks you what's the flux through a control surface at a given instant, but they don't tell you what's the motion of the surface, then the flux is unknown.

The fact that we can choose control volumes and control surfaces arbitrarily gives us a lot of power in solving physics problems and in making predictions. Typically they are chosen so as to simplify the equations that describe the physical situation, simulate physical phenomena in a more precise way, and focus on details of interest.

When we study some solid object, like a football, a rocket, or a planet, we typically choose a control volume that tightly encloses the object. When we study something flowing or moving, such as a fluid material or an electromagnetic field, we typically divide the space of interest into small control volumes and surfaces, constructing a *mesh*; this mesh can even be refined in regions that are of special interest.



Clever choice of control volume and surfaces allow us to model and predict complex motions of fluids; see also animation¹ (Wojtan et al. 2009)

4.2 Volume contents

4.2.1 Scalar quantities

A volume content (or volume integral) for a scalar quantity, for example energy, can be represented like this:



we have eliminated one spatial dimension for simplicity, considering the analogous two-dimensional idea. The volume is in light grey, delimited by a closed darker grey boundary, and we're indicating that the volume content, that is, the amount of energy within, is **8 J**.

As a visualization device, this representation can be useful. But let's straighten out some of its aspects:

- Recall that this is a snapshot at a given time instant. So there are 8 J of energy in the volume at that instant, but we don't know the situation earlier or later: there could be a different amount of energy, the region

might be at a different position and have a different shape, or it might not even exist.

- Recall that some scalar quantities, such as electric charge and in some situations matter (antimatter), can have negative amounts.
- We must not surmise that the amount of quantity is uniformly distributed within the volume. In fact there could be negative amounts of it in some subvolumes and positive in others. In particular, even if there is a zero amount of quantity in a volume, some subvolumes could have non-zero amounts: some positive and some negative, so that the total is zero.



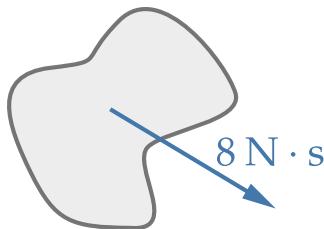
A region with a negative amount of charge

Exercise 4.1

The volume content of matter in a particular volume is equal to 36 mol.
Can we conclude that the volume doesn't contain antimatter?

4.2.2 Vector quantities

A volume content for a vector quantity, for example momentum, can be represented as follows (we still simplify our visualization to two dimensions):



Momentum is a vector quantity, so the total amount in the volume above is a vector. The picture shows the direction and orientation of this vector, and the magnitude of $8 \text{ N} \cdot \text{s}$ is explicitly reported.

Vector magnitudes and opposite vectors

Remember that the *magnitude* of a vector is always positive, and that
 $\overleftarrow{\text{v}} = -1 \cdot \overrightarrow{\text{v}}$

The visual representation above is useful, if we keep in mind remarks analogous to the scalar case:

- This is a time snapshot.
- The application point of the vector representing the volume content is unimportant: for instance, it doesn't need to be placed at the centre of the volume. The vector refers to the volume as a whole, not to some specific point within.
- Different subvolumes could have amounts represented by different vectors; only the total vector is represented above.

This last remark will be especially important when we discuss some physical phenomena involving rotation. As an example, look at the side picture: the volume content for the whole region is *zero*, but its left and right subregions have *non-zero and opposite* volume contents.

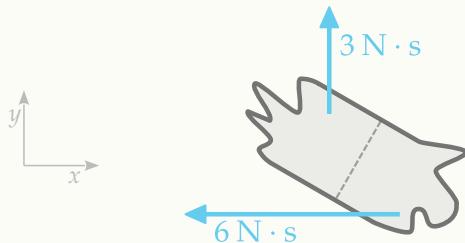


The whole region has zero volume content. The left and right subregions have non-zero and opposite volume contents.

Exercise 4.2

Recall the **third property of our seven main quantities**: The amount in a volume consisting of separate volumes is equal to the total of the separate amounts.

We have a region consisting of two subregions; the amounts of momentum in each subregion are shown below.



1. Write the total momentum in each subregion in component form, (P_x, P_y) , according to the coordinate system shown.
2. Calculate the momentum in the whole region; represent it graphically as vector and write it in component form.

§4 page 51

Adding vectors in General Relativity

We are used to the idea of adding vectors placed at different points in space: we only have to first move each – keeping it parallel to itself – to a common point, and then add them all at that point with the usual rule.

This operation *cannot* be done in General Relativity: the notion of parallelism doesn't apply anymore in a simple way, owing to the curvature

of spacetime. The addition would lead to different results depending on how we transported the vectors. So in General Relativity we can only sum vectors that are placed at the same spacetime point.

How can this operation be possible in Newtonian mechanics and in practical applications, then? After all, General Relativity surely applies here! The answer is that the discrepancies of vector transportation are small enough in the neighbourhood of the Earth, as the curvature of spacetime is very small here.

4.3 Fluxes of scalar quantities

4.3.1 The direction, reckoning, and representation of scalar fluxes

Earlier we took the intuitive example of a flow of people through an open door; we might ask, for instance, how many people crossed the door in a minute. But one more detail about this flow is important: in which direction did the persons cross the door? For example, if the door leads to a classroom, we may need to know whether the people who crossed the door got in or out, so as to know if there are seats left in the classroom.

In order to do this we can: 1. Assign a crossing direction to the door, calling for instance ‘positive’ the direction from outside to inside the classroom. 2. Count as ‘positive’ each person who crosses the door in the positive direction, and count as ‘negative’ each person who crosses the door in the opposite direction. The total tells us the *net* number of people who *entered* the classroom. If the total is positive, then more people got in than out; if the total is negative, then more people got out than in.

One important aspect of this example and terminology is the following symmetry:

- What crossing direction is called ‘positive’ is fully arbitrary, just a matter of agreement.
- If we decide to call ‘positive’ the other crossing direction, then the total will change sign. But the physical situation is of course still the same.

Therefore the sentences “+5 persons entered the room” and “−5 persons exited the room” are saying exactly the same thing. This somewhat trivial fact about fluxes will, later on, turn out to be connected with a famous law. So let’s remember:

 Every flux is equivalent to a flux of opposite amount in the opposite direction

A flux in a particular surface-crossing direction is equivalent to a flux of *opposite sign* in the *opposite* crossing direction.

Another important aspect of the example above is that if we're told “ -5 persons exited the room”, we don't know how exactly this happened: it could be that 5 persons got into the room during that minute; or that 10 persons got in and 5 got out; and so on. This is why we call ‘ -5 ’ the *net* amount.

Finally, consider a similar example but with a quantity that can ordinarily also be negative, such as electric charge. Call ‘positive’ the crossing direction from outside to inside the room. If we're told that a net charge amount of -5 crossed the door in the positive direction in one minute, then this could have happened in several ways:

- a charge of -5 was brought into the room
- a charge of $+5$ was brought out of the room
- a charge of -2 was brought into the room during the first 30 s, and a charge of $+3$ was brought out in the remaining 30 s
- a charge of -2 was brought into the room during the whole minute, and a charge of $+3$ was brought out at the same time
- ... and many other possible combinations.

Indeed, ordinary electricity was thought for some time to be associated with movements of negative *and* positive charges in opposite directions.

The purpose of the examples and scenarios above is just to make you aware of some aspects of what we shall call “flux”, which are trivial but important when considering fluxes of physical quantities.

How can we graphically represent the flux of a quantity, in such a way as to take care of these aspects? Consider these three representations:



The grey straight line represents a surface (simplified to two dimensions) through which we're measuring a flux. In the first and second picture, the wavy arrows represent arbitrary crossing directions called ‘positive’.

“LEX III. Actioni contrariam semper & æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales & in partes contrarias dirigi.”

Newton 1726a

“Fechner [in 1845] supposed every current to consist in a streaming of electric charges, the vitreous charges travelling in one direction, and the resinous charges, equal to them in magnitude and number, travelling in the opposite direction with equal velocity.” Whittaker 1951

The blue arrow represents that we're calling 'positive' the left-to-right direction; the red arrow represents that we're calling 'positive' the right-to-left direction. The signed number represents the net amount according to the positive direction – so the first and second picture represent the same thing. One possible drawback of these two pictures is that they may suggest that a given amount is actually moving from left to right or vice versa; but we have seen that in general we don't know this. The third picture tries to avoid this misleading suggestion by not showing any arrows; it is meant to represent that on the left side the amount of quantity has changed by **-5**, and on the right side by **+5**.

The third representation above has one more advantage. Remember that **a flux through a surface may occur because the surface itself is moving**. The wavy arrows in the first two representations above may misleadingly suggest that some amount of quantity is "moving" in their direction, or that the surface itself is moving in that direction. The third, arrow-less representation is less misleading.

↳ §4.1 page 52

In these notes we shall settle on the third representation above, but feel free to use the one you prefer – as long as you are aware of all the important aspects of a flux.

! What a flux does and doesn't tell

A flux tells us the net change in the amounts of a quantity on the two sides of a surface. These two amounts are equal in magnitude but have opposite signs.

A flux does *not* tell us:

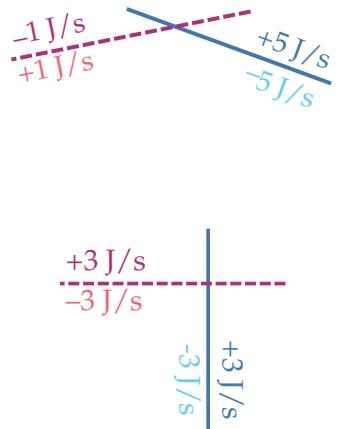
- whether the quantity "was in motion"
- ... and if the quantity was in motion, what was its sign
- whether the surface was in motion

4.3.2 How does a scalar flux change, if we change the surface?

We know that the flux through a surface consisting of two smaller surfaces is the sum of the fluxes through the smaller surfaces. But what if we consider a *different* surface, maybe intersecting the original one? It's important to keep in mind that **a flux refers to a particular surface, and can be very different if we consider a different surface, even if it's close to the original one**.

Consider for instance the picture on the side. We have two intersecting surfaces (as usual simplified by removing one dimension), both almost horizontal. The energy flux through the **solid blue surface**, in a roughly upward direction, is **+5 J/s**. The energy flux through the **dashed red surface**, again in a roughly upward direction, is instead **-1 J/s**.

As another example, the new side picture shows two similar intersecting surfaces, one fully vertical in **solid blue**, and one fully horizontal in **dashed red**. The energy flux through the first in a rightward direction is **+3 J/s**, and so is the energy flux in an upward direction through the second surface.



4.3.3 Flux units: scalar quantities

Remember that the flux of a quantity is defined as an amount of that quantity *per time lapse*. Therefore the physical dimension of the flux is

$$\frac{\text{quantity}}{\text{time}}$$

and its units will be the units of the quantity divided by seconds:

units of fluxes of the four scalar quantities				
quantity	matter	electric charge	energy	entropy
flux units	mol/s	C/s	J/s	J/(K · s)
equivalent units		ampere A	watt W	

After our discussion about the peculiarity of fluxes it's quite easy to work with the fluxes of the four main scalar quantities: matter, electric charge, energy, entropy. Let us add some reminders and remarks about the fluxes of matter and energy.

4.3.4 Matter flux

Remember that *antimatter* “counts as -1 ” for calculating amounts of matter. If 1 mol of positrons (anti-electrons) crosses a surface from left to right in 1 s, the left-to-right flux equals -1 mol/s – note the minus sign. The fact that antimatter is given special names can lead to ambiguities. For instance, if someone asks “what’s the left-to-right flux of *positrons*?”, maybe we should answer “1 mol/s”, since the question concerned specifically positrons. It’s somewhat like asking “what’s the flux of *negative* electric

charge?". In these ambiguous situations is best to add some explanatory words.

4.3.5 Energy flux

We already mentioned that [energy flux can be categorized into different kinds](#), depending on whether there are fluxes of other quantities through the same surface. We study the exact definitions and formulae later on. The total flux is given by the sum of all these kinds. For instance, through a horizontal surface we can have a downward energy flux of 3 J/s as *heat*, and a downward flux of -1 J/s as *work*. The total downward flux is then 2 J/s . The energy flux that you will calculate in the fourth exercise below is called *energy convection*.

◀ § 3.4.3 page 38

Exercise 4.3

For each question, answer in an *unambiguous* way and sketch a picture representing the flux.

- The two sides of a particular surface are called 'up' and 'down'. During 0.2 s , an energy of $+3 \text{ J}$ flows from the up-side to the down-side, and an energy of -4 J flows from the down-side to the up-side. How much is the flux of energy through the surface?
- Through the same surface, at a later time, 2 mol of neutrons flow from the up- to the down- side in 0.01 s , and 2 mol of neutrons flow from the down- to the up-side during the same time. How much is the flux of matter through the surface?
- The two sides of a surface are called 'in' and 'out'. During 0.01 s there is a flow of 1000 electrons from the in-side to the out-side, and also a flow of 1000 *positrons* (anti-electrons) in the same direction. How much is the flux of matter through the surface?
- The side picture shows a [surface](#) moving from left to right at a (constant) velocity of 1 m/s . The space to its right has two static regions with some amount of [energy](#) as shown (there's no energy behind to the left of the surface). How much is the flux of energy through the surface in 1 s ?



4.4 Fluxes of vector quantities

4.4.1 The direction, reckoning, and representation of vector fluxes

The flux of a vector quantity is also a vector, because it is given by an amount of that quantity, which is a vector, divided by time, which is a scalar.

The discussion about the arbitrary choice of a crossing direction and the minus signs that appear when we reverse it also applies, in an analogous way, to vector fluxes. Also the discussion about the minus sign remains the same; we must only remember that a minus sign changes the sense or orientation of a vector:

$$\nwarrow = -1 \cdot \searrow$$

 § 4.3.1 page 56

Take for instance a horizontal surface. Suppose that we call ‘positive’ the *upward* crossing direction, and that with this convection the flux is represented by the vector



Then if we decide to call ‘positive’ the *downward* crossing direction instead, the same flux is represented by the opposite vector



The situation in the case of vector fluxes can require a little more thinking and attention, because it’s easy to get confused between the crossing direction for the surface, and the direction suggested by the vector flux – which are completely separate things. For this reason a good graphical representation of a vector flux is this, analogous to the third representation for scalar fluxes:

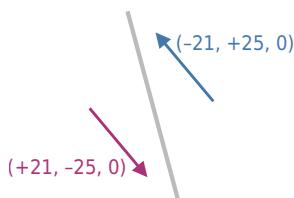
 § 4.3.1 page 57



The grey straight line represents a surface through which we're measuring the flux of a vector quantity; on the left side of the surface, the amount of quantity has changed by  and on the right side by . Obviously these two changes have equal magnitude but are opposite.

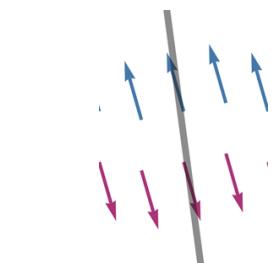
As a mental image of a vector flux, you can imagine a flow of vectors crossing the surface in one direction, together with a flow of opposite arrows crossing the surface in the opposite direction. [See this link²](#) as an example.

Since a vector is represented by three numbers (which can be positive or negative), a vector flux can also be interpreted as the collection of three distinct numerical fluxes:



the picture says that the x -component of the quantity has changed on the right side of the surface by -21 , and on the left side by $+21$; the y -component has changed on the right side by $+25$, and on the left side by -25 ; and the z -component hasn't changed on either side.

Another aspect of vector fluxes that we must try not to get confused about is the application point of the vector representing the flux (the base point of the arrow). This is a time snapshot. Just like in [the case of vector volume contents](#), the application point of the vector representing the flux is unimportant. The vector refers to one side of the surface as a whole.



Imagine the blue upper arrows moving from left to right, and the red lower arrows moving from right to left. [Animated version here³](#).

◀ § 4.2.2 page 55

Exercise 4.4

A horizontal surface is given, and there is a flux of a vector quantity through it (for the moment we neglect units):

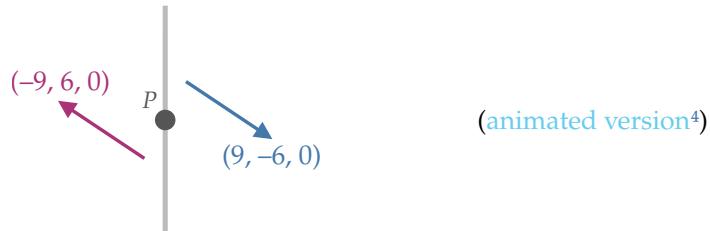
1. If we take the *downward* crossing direction as 'positive', the flux xyz -components are $(5, 5, 0)$. Represent this flux graphically, in the way discussed in the present section. Use the coordinate system  where y points upward.
2. Taking the same crossing direction, represent graphically the flux $(0, -2, 0)$ instead.

3. Taking the same downward crossing direction, we are now told that there is a flux with components $(1, -2, 3)$. What are the components of this flux if we take the *upward* crossing direction as positive?

4.4.2 How does a scalar flux change, if we change the surface?

We saw that the flux of a scalar quantity can be very different if we take a slightly different surface. The same is true of the flux of a vector quantity: in particular, **the vectors representing the fluxes through two slightly different surfaces can point in completely different directions**.

Here is an example. Take a fixed point P . Now take a small vertical surface passing through P . The flux of a vector quantity (momentum for example) through this quantity can be as in this picture:



it has components $(9, -6, 0)$, with magnitude around 10.8, if we take the rightward crossing direction as positive.

Now forget about that surface, and take instead a small horizontal surface passing through the same point P . The flux – of the same quantity – through this surface can be as in this picture:



it has components $(-6, -3, 0)$, with magnitude around 6.7, if we take the upward crossing direction as positive.

Clearly the vectors representing the fluxes through these two surfaces are different: they point in different directions, and have even different magnitudes.

If we have a vector flux through a particular surface, and we're asked about the flux through a different surface, we might be tempted to "move"

through the new surface the vectors representing the flux through the old one. *We must fully resist this temptation.* It only leads to mistakes.

! Each surface has a unique flux

A flux, scalar or vector, through a particular surface, in general doesn't tell you anything about the flux though another surface, even if the other surface is only slightly different from the first one.

4.4.3 Flux units: vector quantities

Also for vector quantities the physical dimension of the flux is

$$\frac{\text{quantity}}{\text{time}}$$

and its units will be the units of the quantity divided by seconds:

units of fluxes of the three vector quantities

quantity	momentum	ang. momentum	magnetic field
flux units	N	N m	Wb/s
equivalent units			volt V

4.5 Flux of momentum: force

We already mentioned that **flux of momentum is what we call 'force'**. Owing to the importance of the notion of *force* in the many branches of physics which rely on Newtonian mechanics, we must discuss this connection in depth. This connection, as well as the connection to Newton's laws, will become even clearer when we discuss the balance of momentum in Chapter 8.

↳ §3.5 page 40

4.5.1 Units

Recall that **momentum can be measured in newton-seconds, 'N · s'**. Its flux, being a momentum per unit time, is therefore measured in newtons, 'N'.

↳ §3.5 page 40

4.5.2 Representing and visualizing force as flux of momentum

The notion of force is very intuitive. We associate it to the sensations that we feel in our skin, flesh, and even bones when, for instance, we push against a wall, twist a door knob, push backwards on the ground with our feet to run, or other similar actions. This force is typically represented by a vector, having the direction and orientation of the “push” or “pull”, and magnitude expressing its intensity. *Such a force vector is exactly the vector expressing the flux of momentum.* The two are the same. A force can therefore also be visualized as a flow of momentum. This mental representation can be illuminating in some physical problems.

As a concrete example, imagine a person pushing against a wall. In terms of force, we say that *the person is exerting a force on the wall, and the force vector on the wall has a person→wall orientation.* This is usually depicted, for example, like this:



In terms of momentum flux, we imagine a surface separating the person and the wall. If we take a person→wall crossing direction, the flux vector also has a person→wall orientation. Because of the symmetry of flux, if we take a wall→person crossing direction instead, then the flux vector also changes direction. This momentum flux can be depicted (zooming in) like this:



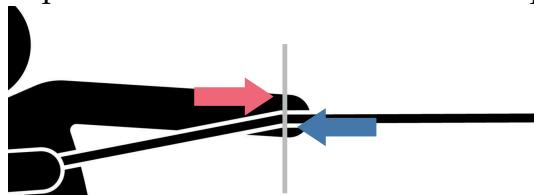
The picture says: on the side of the wall, momentum is changing by an amount having **person→wall orientation**; and, by symmetry of the flux, on the side of the person’s head, momentum is changing by an amount having opposite, **wall→person orientation**.

Let’s take an example involving pulling instead of pushing. Imagine a person pulling a rope fastened somewhere. In terms of force we say that

the person is exerting a force on the rope, and the force vector on the rope has a rope→person orientation. This can be depicted like this:



In terms of momentum flux, we imagine a vertical surface between the person's hand and the rope. If we take a hand→rope crossing direction, the flux vector has a rope→hand orientation – note the difference from the previous example. Because of the symmetry of flux, if we take a rope→hand crossing direction instead, then the flux vector has opposite direction: hand→rope. This momentum flux can be depicted like this:



The picture says: on the side of the rope, momentum is changing by an amount having [rope→person orientation](#); and, by symmetry of the flux, on the side of the hand, momentum is changing by an amount having opposite, [hand→rope orientation](#).

A final example illustrates a situation in between the previous two. Consider the foot of a running person, as it pushes on the ground. In terms of momentum flux, we imagine a horizontal surface between the runner's foot and the ground. If we take a downward crossing direction, the flux vector is oriented towards the back of the foot. Because of the symmetry of flux, if we take an upward crossing direction, then the flux vector has opposite direction, towards the front of the foot, the same as the running direction. This momentum flux can be depicted like this:



The picture says: on the side of the ground, momentum is changing by an amount having [leftward orientation](#); and, by symmetry of the flux, on

the side of foot, momentum is changing by an amount having rightward orientation.

4.5.3 Newton's Third Law!

From the examples above, we see that thinking of force as momentum flux automatically leads to *Newton's third law*: if one side is gaining/losing momentum with a given orientation, by symmetry the other side is gaining/losing momentum with the opposite orientation. So if one side is experiencing a force with a given orientation, the other side must be experiencing a force with the opposite orientation.

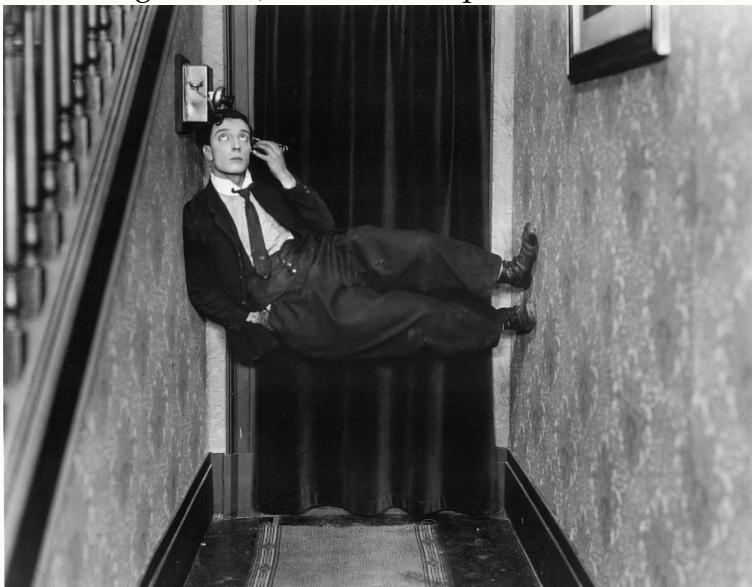
We see that Newton's third law is the expression of the symmetry of fluxes *for the specific case of the flux of momentum (force)*. But we realize that this property is more general: it applies not only to force, but also to the flux of all other quantities, even scalar ones.

"LAW III. To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

(Original Latin on p. 57)
Newton 1726a

Exercise 4.5

Using your intuition, try to guess the various momentum fluxes (except their magnitudes) between this person and the walls:



(Buster Keaton in 'The Electric House'⁶)

4.6 Pressure, tension, shear force

The examples of the previous section demonstrated a variety of possible orientations of the momentum-flux vector with respect to the surface through which it occurs. Obviously all orientations are possible. Special names are given, however, to a momentum flux having three specific orientations: *pressure*, *tension*, and *shear force*.

Take a surface and call its sides *A* and *B*. Now consider a flux of momentum according to the crossing direction $A \rightarrow B$.

4.6.1 Pressure

If the momentum-flux from *A* to *B* is a vector also oriented from *A* to *B*, then we call the momentum flux a **pressure**, or *compressive momentum flux* or *compressive force*. Here are two examples of pressure:



([animated version⁷](#))

Pressure is the kind of momentum flux that we exert when we *push* on an object, and that air exerts on all objects it surrounds.

4.6.2 Tension

If the momentum-flux from *A* to *B* is a vector oriented from *B* to *A*, then we call the momentum flux a **tension**, or *tensile momentum flux* or *tensile force*. Here are two examples of tension:



([animated version⁸](#))

Tension is the kind of momentum flux that we experience in our bones when we *pull* an object, and that occurs in any section of a stretched rubber band.

4.6.3 Shear force

If the momentum-flux from *A* to *B* is a vector oriented transversally, along the surface itself, then we call the momentum flux a **shear force**, or *shearing momentum flux*. Here are two examples of shear force:



(animated version⁹)

Shear force is the kind of momentum flux that we experience under our feet when we walk or run, and that occurs between a car's wheels and the ground.

In general, a momentum flux won't have any of the three special directions above, but rather a combination of them.

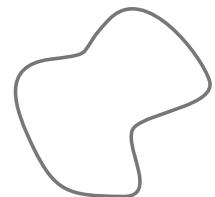
Exercise 4.6

Using your intuition, try to identify the various momentum fluxes that occur in the different parts of a tower crane. Which of the fluxes are (approximately) compressive, tensile, and shearing?



4.7 Closed surfaces, influxes, effluxes

We shall often consider **closed** surfaces, that is, surfaces that don't have a rim or border or holes, like the surface of a sphere or of a cube. A closed surface delimits a specific three-dimensional volume, and we can naturally speak of its **interior** and its **exterior**. An example (simplified by removing one dimension as usual) is the surface in the side picture.



We can give two crossing directions to a closed surface: *inward*, from exterior to interior; or *outward*, from interior to exterior. A flux through the surface is usually called **influx** if we are considering the inward crossing direction, and **efflux** or *outflux* if we are considering the outward crossing direction. Obviously, by the symmetry of fluxes,

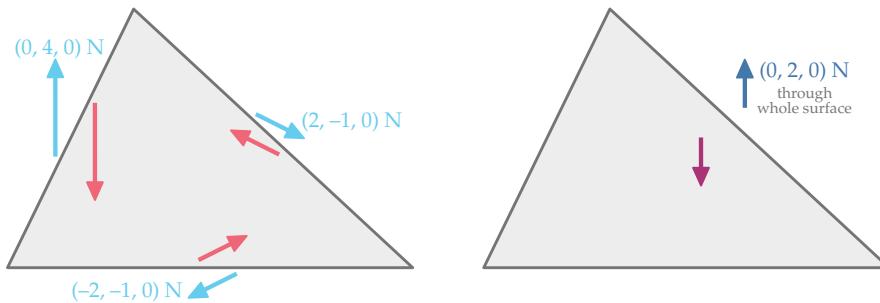
$$\text{influx} \equiv -\text{efflux} \quad \text{efflux} \equiv -\text{influx}$$

The influx and efflux are fluxes *through the whole surface*. Consider for instance these fluxes of energy and of momentum:



On the left we have an efflux of $+5 \text{ J/s}$ on the whole outer side of the surface. We don't know whether these 5 J/s are being evenly distributed all around the surface, or just at particular spots of it. On the right we have an efflux of 8 N , with a vector pointing approximately rightward. Again we don't know what are the flux vectors of individual pieces of the surface: the vector in the picture is just their total.

Let's see another example of this fact. The picture below on the left show the **outward fluxes** through three parts of a closed surface; the picture on the right shows the **total efflux** through the same surface:



The individual fluxes and the total efflux are consistent, since

$$(2, -1, 0) \text{ N} + (0, 4, 0) \text{ N} + (-2, -1, 0) \text{ N} = (0, 2, 0) \text{ N}$$

Exercise 4.7

1. Are the four partial fluxes shown on the left consistent with the total efflux shown on the right? Why?:



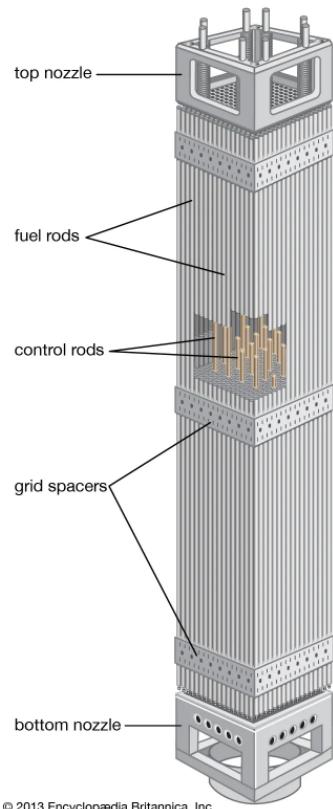
2. Take an imaginary cylindrical surface enclosing one **control rod**¹⁰ in a **nuclear-fission reactor**¹¹ (side picture). Let's say that in a reactor there are 20 such rods. Approximately 5×10^{19} neutrons are liberated in a second in the whole reactor by the fission fuel, but $2/3$ of these are *absorbed* by the control rods.

How much, on average, is the **efflux** of neutrons (matter) through the surface of one control rod?

Express the result first in neutrons/s, and then in mol/s, using the Avogadro constant

$$N_A = 6.022\,140\,76 \times 10^{23} \text{ particles/mol} .$$

Be careful about the signs!



© 2013 Encyclopædia Britannica, Inc.

4.8 Integrated fluxes

A flux is defined as the amount of a quantity crossing a surface in a short time lapse Δt , divided by that time lapse. Denoting the flux by, say, J , this definition also means that the amount of quantity crossing the surface in a short time Δt is equal to $J \Delta t$.

Take now a surface that exists between two time instants t_0 and t_1 ; during this time lapse it could also be moving and changing shape. Choose a crossing direction through the surface. At each intermediate time instant t we can then measure the flux of a quantity crossing the surface in that direction, at that instant; denote it by $J(t)$.

The total amount of quantity that crosses the surface between times t_0 and t_1 can be found by integrating $J(t)$. That is, we divide the time interval into very short time lapses of length Δt ; for each short time lapse we know that the amount that crosses the surface is $J(t) \Delta t$. The total is obtained by adding these small amounts. As Δt is considered shorter and shorter, this sum is by definition an integral:

Integrated flux of a quantity through a surface between two times

The total amount of quantity crossing a surface (relative to a specified crossing direction) between times t_0 and t_1 is given by

$$\int_{t_0}^{t_1} J(t) dt . \quad (4.1)$$

The meaning of the integral above should be clear for any scalar quantity, for which the flux is also a scalar. In the case of a vector quantity, for instance momentum, the flux is also a vector, represented by three components. The integral of a vector is obtained by calculating the integral for each component, obtaining three results, which are the components of a new vector. Geometrically this corresponds to summing a large number of very small vectors.

Take the case of momentum, whose flux (force) we denote $\mathbf{F} = (F_x, F_y, F_z)$. The integral of this flux is then

$$\int_{t_0}^{t_1} \mathbf{F}(t) dt := \left(\int_{t_0}^{t_1} F_x(t) dt, \int_{t_0}^{t_1} F_y(t) dt, \int_{t_0}^{t_1} F_z(t) dt \right). \quad (4.2)$$

The integrated flux can be zero even with non-zero flux

The result of the integral defining the integrated flux can be zero. This means that no *net* amount of quantity crossed the surface between t_0 and t_1 . Yet the flux $J(t)$ can be non-zero, even at all times; of course it needs to be positive at some times, and negative at others.

As a simple example, consider a room's door. During one minute, three people enter through the door; during the next minute, three people (not necessarily the same) exit through the door. The total flow of people is zero, but the flux was non-zero during the first minute, and non-zero during the second minute.

Exercise 4.8

1. What is the physical dimension of the integrated flux of a quantity?
2. Suppose that we calculate the integral above for a particular surface in the case of matter, finding a total of $\int_{t_0}^{t_1} J(t) dt = 7 \text{ mol}$. Now we change our mind and choose the opposite crossing direction for that surface. How does the result above change?

4.9 Fluxes and velocities

The idea of flux naturally evokes the idea of movement, and therefore of *velocity*. Is there a relationship between flux and velocity?

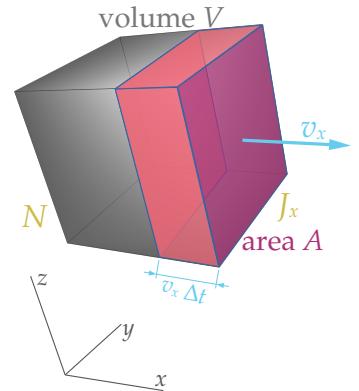
The answer is yes: velocity is in fact essentially defined from a flux and a volume content. Consider for example how we measure the velocity of an object: we are actually keeping track of a flow of matter – the matter that makes up the object – from a region of space to another.

The rigorous definition of velocity from flux is involved, so here we'll just see a simplified and approximate example of how this definition works. In the following chapters you will not need to remember any particular mathematical definition; just keep in mind that there's a tight connection between velocity, flux, and volume content.

Take an extensive scalar quantity like matter, charge, energy, or entropy. For concreteness let's take matter. Choose a coordinate system (t, x, y, z) and consider a point in space at a specific time. Around this point, choose a very small cuboid region, delimited by six small rectangular surfaces: two parallel to the yz -coordinate plane, two to the zx one, and two to the xy one (see side picture). The cuboid has volume V , and the amount of matter (volume content or integral) in it is N . The two yz surfaces (one of them is in **dark red** in the picture) have area A , approximately the same for both; and there is a flux of matter J_x crossing either of these two surfaces in the positive- x direction. These two parallel surfaces have approximately the same area and the same flux because the cuboid region is very small.

The x -component of the coordinate velocity of matter in this region is then *defined* as

$$v_x := \frac{J_x/A}{N/V} \quad (4.3)$$



with analogous definitions for the y - and z -components.

The velocity $\mathbf{v} = (v_x, v_y, v_z)$ so defined has the following intuitive property. If you choose *any* very small surface centred at this point, and move it with this velocity (and in the direction specified by the velocity), then the matter flux through it is zero. This reflects the intuitive understanding that if a surface is moving together with the matter, at the same speed, then we shouldn't observe any flux through it.

Exercise 4.9

Try to prove the formula (4.3) relating flux and velocity in an intuitive way, referring to the picture above.

As a starting point, consider this question: if the amount of matter N in the volume V is moving with velocity v_x in the positive- x direction, how much of it will cross the area A during time $Δt$?

Velocities of objects in general relativity

One consequence of the relationship between velocities and fluxes is that we can define such a velocity for any extensive quantity. So we have a velocity matter from the flux of matter, but also a “velocity of energy” from the flux of energy.

In Newtonian approximation these two velocities are equal, so we do not need to distinguish them. In situations where the Newtonian approximation is not valid, on the other hand, we have to take into account the “velocity of matter” and the “velocity of energy” separately. This difference is important for instance in the study of plasma in stars and in numerical general relativity. There is an ongoing discussion as to which of the two velocities is more convenient to use; see for instance Kandus & Tsagas 2008, especially the section *Eckart frame versus Landau frame*, which refers to the choice between these two velocities.

4.10 Symbols for volume contents and fluxes

Let us summarize the symbols for volume contents and fluxes of the seven quantities, together with their units, that are used in these notes:

Symbols and units of volume contents and fluxes

quantity	vol. integral [unit]	flux [unit]
matter	N [mol]	J [mol/s]
energy	E [J]	Φ [J/s or W]
momentum	\mathbf{P} [N s]	\mathbf{F} [N]
angular momentum	\mathbf{L} [N m s]	$\boldsymbol{\tau}$ [N m]
entropy	S [J/K]	H [J/(K s)]
electric charge	(not used) [C]	(not used) [C/s or A]
magnetic field	(not used) [Wb]	(not used) [Wb/s or V]

URLs for chapter 4

1. <http://www.youtube.com/watch?v=M5xnAdVPbgQ>
2. <https://pglpm.github.io/7wonders/media/vectorfluxanimsin.webp>
3. <https://pglpm.github.io/7wonders/media/vectorfluxanimsin.webp>
4. <https://pglpm.github.io/7wonders/media/skewflux1.webp>
5. <https://pglpm.github.io/7wonders/media/skewflux2.webp>
6. <https://www.imdb.com/title/tt0013099/>
7. <https://pglpm.github.io/7wonders/media/pressure.webp>
8. <https://pglpm.github.io/7wonders/media/tension.webp>
9. <https://pglpm.github.io/7wonders/media/shearforce.webp>
10. https://energyeducation.ca/encyclopedia/Control_rod
11. <https://www.britannica.com/technology/nuclear-reactor>

Chapter 5

Physical laws

Every branch of physical science is based on two sets of fundamental equations. The first set is that of basic laws of physics, which are postulated to hold valid for all bodies under all conceivable circumstances [. .]. The second set of fundamental equations are the constitutive equations: these are relationships which are not supposed to hold for all bodies, but only to describe the behavior of some restricted class of bodies, or possibly of a larger class of bodies for a more restricted class of phenomena.

G. Astarita 1990

5.1 Two kinds of laws

In Chapter 3 we introduced seven main quantities in terms of which we describe physical phenomena. How do we describe the patterns, regularities, and diversities that we observe in these phenomena? This is the role of *physical laws*. Different physics formalisms express physical laws in different ways, even if the actual physical phenomenon is exactly the same. In the formalism that we are using we distinguish between two kinds of physical laws:

Two kinds of physical laws

- **Fundamental laws** represent universal patterns that we observe everywhere, and that are taken as the basic mathematical equations in a physical theory.

- **constitutive relations or properties** only apply to particular phenomena, only on particular time & space scales, only in particular theories. They express the diversity that we observe around us, and different kinds of approximations that we use in describing physical phenomena.

This distinction it's important. The fundamental laws can be applied with some confidence to every new phenomenon we observe, and often allow us to make predictions of at least a qualitative character. As far as we know there is a very small number of fundamental laws. We only modify them as a last resort (and so far this has rarely or never been necessary). We have a lot of freedom in modifying constitutive relations instead, and in proposing new ones to account for newly observed physical phenomena.

5.1.1 Balance laws

There is a small set of laws that either are fundamental laws or can be derived from fundamental laws, and that have a special mathematical form and physical meaning. They are called *balance laws*:

Balance and conservation laws

Balance laws, or simply *balances*, have this name because they express a sort of trade-off or “budget” about physical quantities.

Conservation laws are a special and important kind of balance laws.



<https://xkcd.com/2904>

‘Conservation’ vs ‘balance’ in other texts

Be aware that many texts do *not* distinguish between ‘conservation law’ and ‘balance law’, and often use ‘conservation’ to mean both; which is unfortunate, since balance is the most general of the two. Always try to infer from the context how these two terms are used.

In this chapter we shall see how to mathematically express balance and conservation laws, as well as some kinds of constitutive relations.

5.2 Conservation laws

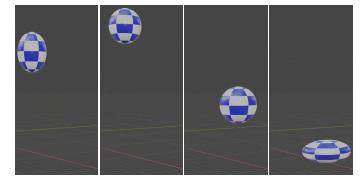
Conservation laws tightly connect *volume contents* and *fluxes*. They are intuitively and visually easy to grasp. The archetype is as follows:

↳ § 4 page 51

Consider a *closed control surface* at a (coordinate) time t_0 , which delimits a control volume. A spherical surface, delimiting a ball-shaped region, is an example. Actually the closed control surface doesn't need to be connected: it could consist of several separate closed control surfaces, each delimiting a different volume.

↳ § 4.7 page 69

Starting from time t_0 , imagine the closed control surface moving and deforming through time, changing shape and size but remaining closed – no holes or cuts – until a later time t_1 . At this final time we have a final closed control surface that can be different from the initial one, and at a different position. If we wanted, we could start with a spherical surface delimiting a region the size of a tennis ball, and end with a cubical surface delimiting a region the size of the Sun. The temporal sequence of closed control surfaces delimits a sequence of control volumes; see the sequence on the side picture below as an example.



Four snapshots of a moving and deforming closed control surface ([click here for animated version¹](#))

Now consider any one of the *seven primitive quantities*, except magnetic flux. Recall that these quantities are *extensive*: we can measure the amount of this quantity in any control volume, and its flux across any control surface. For concreteness let's take *matter*, for which we denote volume contents by N and flux by J . We make the following measurements:

↳ § 3.1 page 30

- $N(t_0)$: total amount of matter contained within the control volume at time t_0 .
- $J(t)$: *influx* (*flux crossing inwards*) of matter (and antimatter!) through the closed surface at any particular time t between t_0 and t_1 .
- $N(t_1)$: total amount of matter contained within the control volume at time t_1 .

↳ § 4.7 page 69

Each of the three amounts above is, in this example, a scalar that can be positive, negative, or zero. For instance we could have:

$$t_0 = 00:00:00 \quad t_1 = 00:00:18$$

$$N(t_0) = 3.50 \text{ mol} \quad J(t) = -0.06 t \text{ mol/s}^2 \quad N(t_1) = -6.22 \text{ mol}$$

note that $J(t)$ is given as a function of the time t .

With this setup we can introduce two kinds of balance laws.

Conservation law in integral form

An extensive quantity with volume content $N(t)$ and influx $J(t)$ is said to satisfy a **conservation law**, or to be **conserved**, if the following equality holds *always*:

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t) dt \quad (5.1)$$

The meaning of the equation above is intuitive: the amount of quantity in the final control volume, $N(t_1)$, must be equal to the amount in the initial control volume, $N(t_0)$, plus the total amount (which could be negative) that flowed in through the surface between those times, that is, the integrated flux $\int_{t_0}^{t_1} J(t) dt$. Said otherwise, any amount of quantity that appears in (or disappears from) the control volume, must come in (or go out) through the surface; it can't appear or disappear out of nowhere.

A conservation law allows us to make several kinds of predictions or deductions. For example:

- If we know the amount of quantity within the control surface at t_0 and the net amount that crossed the surface between t_0 and t_1 , then we can predict the amount within the control surface at t_1 : this is given by the equation above.
- If we know the amount of quantity within the control surface at t_0 and the one at t_1 , then we can deduce the net amount that crossed the surface between the times t_0 and t_1 :

$$\int_{t_0}^{t_1} J(t) dt = N(t_1) - N(t_0) .$$

- If we know the amount of quantity within the surface at t_1 and the net amount that crossed the surface between t_0 and t_1 , we can deduce the amount initially within the surface, at t_0 :

$$N(t_0) = N(t_1) - \int_{t_0}^{t_1} J(t) dt .$$

These kinds of predictions and deductions are very powerful because we are fully free to decide the shape and motion of the control surface and volume, as well as the times t_0 and t_1 .

We unconsciously use several conservation laws all the time in our everyday life. If a bike tyre is suddenly deflated, we conclude that that air

must have got out of it through its surface, which must therefore have a hole or a defective valve; we don't conclude that air "just disappeared".

Note that the discussion so far is equally valid for a scalar or a vector quantity (we shall consider vector quantities more in detail in a moment).

5.2.1 Differential form of conservation laws

The conservation law

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t) dt$$

is said to be written in *integral form*, because it contains the integrated flux, necessary to calculate the net amount of quantity that crosses the control surface.

Change the symbol for time t_0 simply into t , and take t_1 to come after a *very short* lapse of time Δt after t :

$$t_0 \text{ becomes } t, \quad t_1 \text{ becomes } t + \Delta t$$

Then the integrated flux can be approximated by $\int_t^{t+\Delta t} J(t) dt \approx J(t) \Delta t$ (because the flux $J(t)$ is approximately constant in the short time lapse Δt).

The conservation law can then be approximated as

$$N(t + \Delta t) \approx N(t) + J(t) \Delta t$$

Now bring the term $N(t)$ to the left side, and divide the whole expression by Δt :

$$\frac{N(t + \Delta t) - N(t)}{\Delta t} \approx J(t) \Delta t$$

Consider smaller and smaller Δt : then the fraction $\frac{N(t + \Delta t) - N(t)}{\Delta t}$ becomes, by definition, a derivative, and we obtain a new form for the conservation law:



Some mathematical conditions must be satisfied to make these steps; otherwise the derivative below must be carefully defined in a generalized sense.

Conservation law in differential form

A quantity is said to satisfy a **conservation law** if the following equality holds always:

$$\frac{dN(t)}{dt} = J(t) \quad (5.2)$$

This form says that the rate of change of the volume content of a quantity

must equal *influx* of that quantity, that is, the rate at which the quantity is coming in through the control surface.

Both forms, integral and differential, of a conservation law are useful, in different situations. The integral form is obviously useful if we know the total flow during a time lapse but we don't know the flux at each instant. The differential form is often useful for simulating physical processes and make predictions.

5.2.2 Example with a moving control surface

The tube within the tyre of a bicycle has, at a given time, 12.6 mol of air. The bicycle and its tyres moves around, and thirty minutes later the tyre is flat, the inner tube having only 0.5 mol of air.

We consider an imaginary, closed control surface wrapping the tube. The control surface deforms just like the tube deforms. The initial control surface has therefore a tubular shape, but the final control surface doesn't.

We assume that air satisfies a conservation law. From the description above we have

$$t_0 = 0 \text{ s}, \quad t_1 = 1800 \text{ s}, \quad N(t_0) = 12.6 \text{ mol}, \quad N(t_1) = 0.5 \text{ mol}.$$

The conservation law (5.1) allows us to calculate the integrated *influx*:

$$\begin{aligned} \int_{t_0}^{t_1} J(t) dt &= N(t_1) - N(t_0) \\ &= 0.5 \text{ mol} - 12.6 \text{ mol} \\ &= -12.1 \text{ mol} \end{aligned}$$



The negative-signed result says that 12.1 mol of air have crossed the moving control surface, in an outward direction, during thirty minutes. Obviously this happened because the tyre tube must have a hole. Note that this is a physical hole in the physical tube; our control surface is imaginary and doesn't have any holes. It's just a sort of "border checkpoint" where anything can in principle move through; we only keep count of what's crossing, how much, and how fast.

We see that the fact that air satisfies a conservation law allows us to determine the integrated flux through the control surface. But it doesn't allow us to determine the flux $J(t)$ at every individual time between t_0 and t_1 . For example, we don't know if there was a larger flux at the beginning than at the end; the tyre may have gone completely flat just after 60 s; in that case the flux $J(t)$ was zero for $t > 60$ s.



There's a physical hole in the physical tube, but no holes in the imaginary control surface that wraps the tube

5.2.3 Example with a static control surface

In the previous example we chose a closed control surface that moved and deformed with some object of interest (the tyre tube). But we can also choose a different control surface, for instance a static one.

Consider a block of 53.4 mol of ice having cylindrical shape with 1 m diameter and 1.226 m height. It is falling downward owing to gravity (and in a vacuum). At a particular time it occupies a particular position; after 0.5 s it is situated, for an instant, immediately underneath the initial position, as illustrated in the side picture.

Suppose we want to know the net quantity of ice that passed, in a downward direction, across the circular surface separating the initial and final positions. We can calculate this quantity by assuming that ice (as matter) satisfies a conservation law.

We choose a *static* closed control surface of cylindrical shape. This control surface is located in such a way that it wraps the block of ice at the final time; at the initial time, therefore, the control surface is empty. See the side picture, where the control surface is represented by the **dashed yellow line**. This static control surface consists of three parts: a circular surface at the bottom, a side surface, and a circular surface at the top – which is the one that interests us.

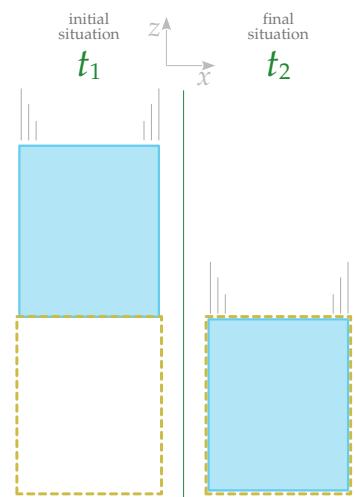
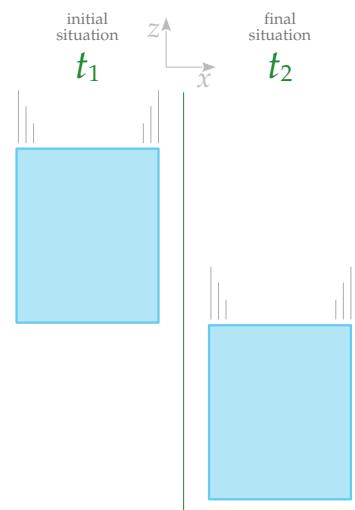
According to the description above we have

$$t_0 = 0 \text{ s}, \quad t_1 = 0.5 \text{ s}, \quad N(t_0) = 0 \text{ mol}, \quad N(t_1) = 53.4 \text{ mol}.$$

From the conservation law (5.1) we calculate the integrated **influx**:

$$\begin{aligned} \int_{t_0}^{t_1} J(t) dt &= N(t_1) - N(t_0) \\ &= 53.4 \text{ mol} - 12.6 \text{ mol} \\ &= 53.4 \text{ mol} \end{aligned}$$

Note that this is the integrated influx through the *whole* cylindrical surface. The conservation law doesn't tell us anything about the integrated influx through *parts* of the surface. In the present case we provide the extra knowledge that the fluxes through the side and bottom surfaces are zero, between t_0 and t_1 . Then, by the **extensivity property**,



↳ § 3.1 page 31

$$\begin{aligned} \int_{t_0}^{t_1} J(t) dt &= \int_{t_0}^{t_1} [J_{\text{top}}(t) + J_{\text{side}}(t) + J_{\text{bot}}(t)] dt \\ 53.4 \text{ mol} &= \int_{t_0}^{t_1} [J_{\text{top}}(t) + 0 \text{ mol/s} + 0 \text{ mol/s}] dt \end{aligned}$$

from which we find, as was intuitively clear, that the net amount of ice that crossed the top surface in a downward direction is 53.4 mol.

Exercise 5.1

Solve the following exercises not just by using intuition, but by explaining step-by-step how you use a conservation law, in integral or differential form, to obtain the result:

- what is the relevant time interval (if using the integral form)
- how you define the closed control surfaces and their movement, as well as any subdivisions of the surfaces
- what are the volume content and integrated-flux values you know, and which you want to find

1. For the present exercise we assume that energy satisfies a conservation law.

An apartment's room has two identical electric heaters along a wall. An electric heater can be considered as a piece of surface across which energy flows into the room: the energy is entering (through the electric cable) in the form of electromagnetic energy, and is converted into internal energy (mainly of the room's air) by means of the heater. Suppose that each heater corresponds to an influx of 200 J/s.

The room also has a window, which is the only other part of the room's boundary where energy can flow in or out.

In one hour we measure that the total amount of energy in the room has not changed. How much is the integrated energy *influx* through the window during that time?

2. For the present exercise we assume that argon² atoms satisfy a conservation law.

A [distillation column³](#) in a chemical plant has a constant influx of 2 mol/s of argon atoms at one inlet, and a constant outflux of argon atoms, possibly at a different rate, at an outlet. An amount of 500 mol of argon is measured in the chamber at a given time, and one minute later the amount is measured to be 620 mol. How much is the argon *outflux* (*not* the integrated outflux) at the outlet?



5.3 Balance laws

Our intuitive understanding of a conserved quantity is that it cannot be “created” or “destroyed”. This leads to the definition of a more general law where this condition is not necessarily satisfied.

The setup is just like before: consider an initial and final closed control surfaces and volumes at t_0 , t_1 , and a sequence of closed control surfaces in between. This time let’s take *energy-mass* as a concrete example, with initial volume content $E(t_0)$, final volume content $E(t_1)$, and influx $\Phi(t)$.

Balance law

A quantity with volume content $E(t)$ and influx $\Phi(t)$ is said to satisfy a **balance law**, or to be **balanced**, if the following equality holds *always*:

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t) dt + \int_{t_0}^{t_1} R(t) dt \quad (5.3)$$

where $R(t)$, called the **supply** or **source**, is the amount of quantity created per unit time within the closed control surface at time t .

If the quantity is a scalar, then $R(t)$ can also be negative, in which case we may call it a *sink* and we say that some of the quantity is being destroyed at time t .

The meaning of the equation above is intuitive: the amount of quantity in the final control volume, $E(t_1)$, must be equal to the amount in the initial control volume, $E(t_0)$, plus the total amount that flowed in through the surface between those times, $\int_{t_0}^{t_1} \Phi(t) dt$, plus the total amount that was created in the volume, $\int_{t_0}^{t_1} R(t) dt$.

According to the definition above, a *conservation law* is a *special, powerful case* of a balance law: one for which $R(t) = 0$ *always and everywhere*. That is, there never are sources or sinks of the quantity: the quantity cannot be *created or destroyed*: it can only “move around”.

A conservation law is powerful because the supply $R(t)$ is already known and has an extremely simple value (zero). It is also powerful because it allows us to predict the amount of quantity in the final volume by making measurements *only on the surface* during the time lapse. A general balance law, instead, requires us to also make measurements at every point *within the volume* during the time lapse: we must do this to check whether some quantity was created or destroyed within the volume.



A conservation law only involves what happens within the initial and final control volume (represented by the red hatched disks), and on the sequence of control surfaces in between (represented by the curved red contours), but not what happens within the control volumes in between. A balance law instead also involves what happens within the control volumes in between.

For these reasons a balance law is in some respects more trivial than a conservation law. If we measure that $E(t_0) - E(t_1) - \int_{t_0}^{t_1} \Phi(t) dt$ is not zero (so there's no conservation law), we can always say that some amount of quantity must have been created or destroyed within the control volume between t_0 and t_1 . An extensive quantity can therefore always be said to satisfy a balance. A balance is not trivial, however, if we have some specific physical law that tells us in advance how the amount created or destroyed at each instant, $R(t)$, can be calculated.

A conservation law is therefore powerful also because it allows us to predict the amount of quantity in the final volume by making measurements *only on the surface* during the time lapse. A general balance law, instead, requires us to also make measurements at every point *within the volume* during the time lapse.

• Supplies are very different from fluxes

One could object: "why don't you just put the flux $\Phi(t)$ and the supply $R(t)$ together, adding the two integrals in equation (5.3)? Wouldn't you get

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} [\Phi(t) + R(t)] dt = 0$$

which looks like a conservation law?"

Unfortunately the mathematical expression above wouldn't be a conservation law, despite its appearance. The point is this: the flux Φ involves only what's happening *on the control surface*; the supply R , on the other hand, involves what's happening *within the control volumes, at each time*. A conservation law instead does *not* require us to know what's happening within the control volumes, except at the initial and final time.

5.3.1 Balance law for vector quantities

Let us now consider an extensive vector quantity like *momentum*. Denote its initial and final volume contents by $\mathbf{P}(t_0)$ and $\mathbf{P}(t_1)$, and its influx (the force) by $\mathbf{F}(t)$. These are all time-dependent vectors:

$$\mathbf{P}(t) = \begin{bmatrix} P_x(t) \\ P_y(t) \\ P_z(t) \end{bmatrix} \quad \mathbf{F}(t) = \begin{bmatrix} F_x(t) \\ F_y(t) \\ F_z(t) \end{bmatrix} .$$

The balance law has an analogous form:

$$\mathbf{P}(t_1) = \mathbf{P}(t_0) + \int_{t_0}^{t_1} \mathbf{F}(t) dt + \int_{t_0}^{t_1} \mathbf{G}(t) dt \quad (5.4)$$

where the supply

$$\mathbf{G}(t) = \begin{bmatrix} G_x(t) \\ G_y(t) \\ G_z(t) \end{bmatrix}$$

is the amount of momentum created per unit time within the control volume at time t . This balance corresponds to three equations, one per component:

$$\begin{cases} P_x(t_1) = P_x(t_0) + \int_{t_0}^{t_1} F_x(t) dt + \int_{t_0}^{t_1} G_x(t) dt \\ P_y(t_1) = P_y(t_0) + \int_{t_0}^{t_1} F_y(t) dt + \int_{t_0}^{t_1} G_y(t) dt \\ P_z(t_1) = P_z(t_0) + \int_{t_0}^{t_1} F_z(t) dt + \int_{t_0}^{t_1} G_z(t) dt \end{cases} \quad (5.5)$$

5.3.2 Differential form of balance laws

By considering a very short time lapse Δt , with reasoning analogous to the case of a conservation law (5.2), we can write a balance law in a different form:

Balance law in differential form

A quantity is said to satisfy a **balance law** if the following equality holds *always*:

$$\frac{dE(t)}{dt} = \Phi(t) + R(t) \quad (5.6)$$

For a vector quantity such as momentum, the differential form again corresponds to three equations, one per component:

$$\begin{cases} \frac{dP_x(t)}{dt} = F_x(t) + G_x(t) \\ \frac{dP_y(t)}{dt} = F_y(t) + G_y(t) \\ \frac{dP_z(t)}{dt} = F_z(t) + G_z(t) \end{cases} \quad (5.7)$$

5.3.3 Example with a moving control surface

Take a coordinate system (x, y, z) where y has a horizontal direction and z an upward direction. At a particular time instant a tennis ball has momentum $[0, 1.70, 0.98]$ N s. It flies for two seconds. During this time there's a continuous and constant supply of momentum within the ball, equal to $[0, 0, -0.579]$ N. No net momentum is leaving or entering the ball in any other way (there is a flux of momentum corresponding to air pressure, but it is such that its total over the whole surface is approximately zero).



We assume that momentum satisfies a balance law. Choose an imaginary, closed control surface that tightly wraps the tennis ball and moves with it. From the description above we have

$$t_0 = 0 \text{ s}, \quad t_1 = 2 \text{ s},$$

$$\mathbf{P}(t_0) = \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s}, \quad \mathbf{F}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N (const.)}, \quad \mathbf{G}(t) = \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N (const.)}.$$

The balance law (5.4) allows us to find the amount of momentum in the tennis ball at time $t_1 = 2$ s:

$$\begin{aligned} \mathbf{P}(t_1) &= \mathbf{P}(t_0) + \int_{t_0}^{t_1} \mathbf{F}(t) dt + \int_{t_0}^{t_1} \mathbf{G}(t) dt \\ &= \mathbf{P}(t_0) + \mathbf{F} \cdot (t_1 - t_0) + \mathbf{G} \cdot (t_1 - t_0) \quad (\text{because } \mathbf{F}, \mathbf{G} \text{ are constant}) \\ &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N} \cdot 2 \text{ s} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \cdot 2 \text{ s} \\ &= \begin{bmatrix} 0 \\ 1.70 \\ -0.18 \end{bmatrix} \text{ N s} \end{aligned}$$

We see that two seconds later the ball's momentum has the same x - and y -component, but its z -component points downward – which means that the ball is moving downward.

The description above actually has enough information for applying the balance law in differential form, formula (5.7), because the value of the influx \mathbf{F} and supply \mathbf{G} are given at every time t (they are constant). We

find that, at any time t between t_0 and t_1 , the rate of change of momentum is

$$\begin{aligned}\frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}(t) + \mathbf{G}(t) \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \\ &= \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N}\end{aligned}$$

So we could find the momentum at time t_1 , or at any other time between t_0 and t_1 , by an easy integration:

$$\begin{aligned}\mathbf{P}(t_1) &= \mathbf{P}(t_0) + \int_{t_0}^{t_1} \frac{d\mathbf{P}(t)}{dt} dt \\ &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s} + \int_{t_0}^{t_1} \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} dt \\ &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \cdot (t_1 - t_0) \quad (\text{because the integrand is constant}) \\ &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \text{ N s} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N} \cdot 2 \text{ s} \\ &= \begin{bmatrix} 0 \\ 1.70 \\ -0.18 \end{bmatrix} \text{ N s}\end{aligned}$$

5.3.4 Example with a static control surface

Consider again the [example of a cylindrical block of ice](#) moving downward, in a vacuum, during a lapse of time of 0.5 s, depicted in the side picture below. The block of ice has a fully downward momentum of magnitude 4.72 N s at the final time.

↳ § 5.2.3 page 82

We are now interested in the net amount of *momentum* that crosses, in a downward direction, the circular surface separating the initial and final location.

We could use the same strategy adopted in that example: considering a static control surface at the final position of the block of ice. From the description, our data are

$$t_0 = 0 \text{ s}, \quad t_1 = 0.5 \text{ s}, \quad \mathbf{P}(t_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ Ns}, \quad \mathbf{P}(t_1) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns},$$

But we don't have enough data to apply the balance law (5.4) for finding the integrated influx. It is also necessary to know the supply of momentum $\mathbf{G}(t)$, or at least its integral. This is an example of what we said previously: **balance laws in general require more information** than conservation laws.

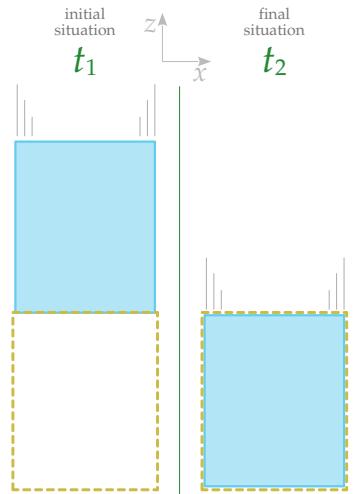
Suppose we are told that the net amount of momentum generated *within the control volume*, during the 0.5 s, has magnitude 1.57 N s and fully downward direction. That is,

$$\int_{t_0}^{t_1} \mathbf{G}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{ Ns}$$

Now we can use the balance law (5.4) to find the integrated influx:

$$\begin{aligned} \int_{t_0}^{t_1} \mathbf{F}(t) dt &= \mathbf{P}(t_1) - \mathbf{P}(t_0) - \int_{t_0}^{t_1} \mathbf{G}(t) dt \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ Ns} - \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{ Ns} \\ &= \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \text{ Ns} \end{aligned}$$

↳ § 5.3 page 84



Let us finish this example with a quick analysis using a *moving control surface*, wrapping the block of ice. We are told that the block initially

has zero momentum. Moreover, there is no flux of momentum through a surface that moves along with the block. Our data are therefore

$$t_0 = 0 \text{ s}, \quad t_1 = 0.5 \text{ s},$$

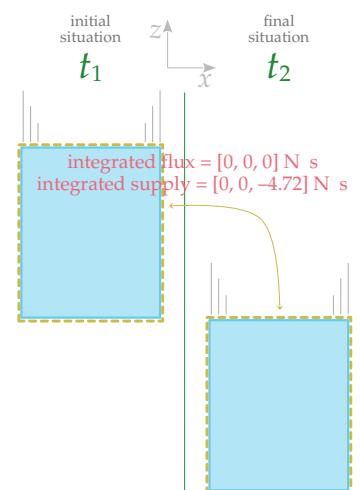
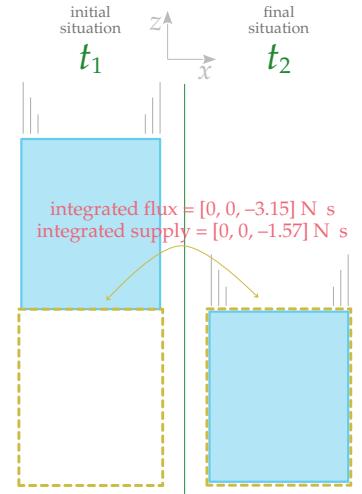
$$\underline{\mathbf{P}}(t_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ Ns}, \quad \underline{\mathbf{P}}(t_1) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns}, \quad \underline{\mathbf{F}}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N (const.)}.$$

Note that the volume contents and fluxes are now represented by different symbols (we use an underline, but we could use any other graphical symbol, or simply change the letters themselves). This is because **they refer to different control surfaces and volumes**. We expect their values to be different from the previous ones, which referred to static surfaces. Indeed we already notice that

$$\int_{t_0}^{t_1} \underline{\mathbf{F}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \text{ Ns} \neq \int_{t_0}^{t_1} \underline{\mathbf{F}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ Ns}.$$

We have enough data to find the total amount of momentum generated within the moving control volume:

$$\begin{aligned} \int_{t_0}^{t_1} \underline{\mathbf{G}}(t) dt &= \underline{\mathbf{P}}(t_1) - \underline{\mathbf{P}}(t_0) - \int_{t_0}^{t_1} \underline{\mathbf{F}}(t) dt \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ Ns} - \int_{t_0}^{t_1} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ N dt} \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns} \end{aligned}$$



We notice again the difference:

$$\int_{t_0}^{t_1} \underline{\mathbf{G}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \text{ Ns} \neq \int_{t_0}^{t_1} \underline{\mathbf{G}}(t) dt = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \text{ Ns}$$

! Fluxes and supplies depend on the particular control surface and its motion

The integrated flux $\int_{t_0}^{t_1} \mathbf{F}(t) dt$ and integrated supply $\int_{t_0}^{t_1} \mathbf{G}(t) dt$ that appear in conservation and balance laws are strictly dependent on the sequence of closed control surfaces and volumes that we choose.

Therefore, if we analyse the same physical phenomenon with a new set of control surfaces, in general we cannot expect the results of calculations with the old set to be valid for the new one.

Exercise 5.2

Consider once more the block of ice analysed in the previous example.

This time choose a *static* closed control surface that coincides with the *initial* position of the block. This surface therefore includes all ice at time t_0 , but is empty – and has zero momentum – at time t_1 .

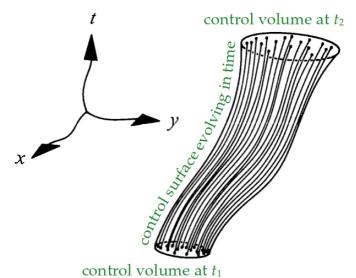
Calculate the integrated supply for this new control surface between t_0 and t_1 .

(Hint: consider the result from the exercise and the **symmetry of fluxes** for the flux through the bottom part of the surface, as well as the fact that the fluxes through the side and top surfaces are zero.)

2 Balance laws in General Relativity

Conservation and balance laws appear even simpler from the point of view of Relativity Theory. From a four-dimensional spacetime perspective, a 3D volume is a region, called *hypersurface*, having one less dimension than spacetime. But a sequence of 2D surfaces through time is also just a region having one less dimension than spacetime – two spatial dimensions and one temporal one. Thus the distinction between the 3D region at t_0 , the sequence of 2D surfaces between t_0 and t_1 , and the 3D region at t_1 disappear: they are seen to be just different parts of the same three-dimensional hypersurface. We perceive some parts of this hypersurface as belonging “to the same time”, showing their three dimensions all at once; and other parts as extending through time, showing only two dimensions at any time. In fact, different observers make this division in different ways.

And from a spacetime perspective, the amount of a quantity $E(t_0)$ or $E(t_1)$ within a 3D region is seen as a flux through time; so its apparent difference from the flux Φ also disappear.



Spacetime representation of the evolution of a closed control surface, containing some pointlike objects (adapted from Misner et al. 2017)

What about the magnetic flux?

In the case of magnetic flux, the idea of a conservation law is analogous, but is formulated with one less spatial dimension: we consider a closed 1D curve at t_0 , one at t_1 , and a sequence in between these times. The magnetic flux turns out to be a quantity for which it's possible to ask how much of it is "linked" to a closed curve, and how much of it is crossing a closed curve. One way to understand this is to imagine magnetic flux as a bundle of tubes or lines that are either closed or extend to infinity. It is a very fascinating quantity, and one may wonder if other quantities exist which satisfy similar balances even with one less dimension. But we shall not pursue magnetic flux and related topics in these notes.

5.4 Constitutive relations

In the previous sections we have seen several examples of use of conservation and balance laws. They allowed us to determine or predict the amount of quantities in given spatial regions or crossing through surfaces.

The examples also showed that extra information is always needed in order to use the balance laws. For instance, in the [example with the tennis ball](#) we said that there was no flux of momentum through the moving control surface (the flying ball). In the [example with the block of ice](#) we said "where there's no matter, there's no momentum"; and we also said that the flux of momentum was zero on some surfaces, but non-zero on others. *Well, how did we know?*

This kind of information and physical properties were *not* something we could deduce from the conservation or balance laws.

Indeed those were examples of physical properties that apply, not universally, but only to some specific phenomena, and only to some theory-dependent degree of approximation. They were examples of *constitutive relations*:

 [§ 5.3.3 page 87](#)

 [§ 5.3.4 page 88](#)

Constitutive relation

A **constitutive relation**, also called *constitutive equation*, or *closure equation*, or *constitutive property*, is a physical relationship or property that is only true for specific physical phenomena.

A constitutive relation often has different expressions, of varying levels of approximation, in different theories.

Constitutive relations express the diversity that we observe around us, for example the fact that a body of water can easily change shape upon movement, as opposed to a block of concrete, which we rely on for its rigidity. The differences between states of matter – solid, liquid, gas, plasma, and there are others – arise from different constitutive relations. Constitutive relations also mark the difference between specialized or approximate physical theories, for example between Newtonian mechanics, which applies only for low speeds and low energy concentrations (hence weak gravitational fields and small spacetime curvature), and general relativity, which applies on all scales, including cosmological scales.

When we read that a new physical phenomenon has been discovered, usually that means that a new *constitutive relation* has been discovered. Depending on the specific scientific field you'll work in, you'll learn some constitutive relations in more detail than others.

Constitutive relations come in a great variety of mathematical forms. Some of them are simple algebraic relations between the volume content of one quantity and the flux of another. Others involve spatial or time derivatives. Other still involve integrals in space or in time.

5.4.1 Examples

Let us briefly discuss some of the constitutive equations that have silently been used in the examples of the previous sections. All the constitutive relations below are valid only in “Newtonian approximation”, that is, when speeds are much lower than the speed of light and gravitation is weak.

Constitutive relation for mass-energy and matter. If a small control volume contains an amount of matter N , then it also contains an amount of mass-energy

$$m = \rho N$$

where ρ , called *molar mass*, is (approximately) a constant that depends on the kind of matter. This constitutive relation is the reason why an amount of matter is often quantified in terms of mass.

Constitutive relation for momentum and matter. Take a small control volume containing an amount of matter N moving with velocity \mathbf{v} (and recall that [there's a connection between velocity and flux](#)



Four states of matter (image by Spirit469⁴)

of matter). Then this control volume also contains an amount of momentum

$$\mathbf{P} = m\mathbf{v} \equiv \rho N \mathbf{v}$$

This is the famous relation that old textbooks take as the *definition* of momentum.

Constitutive relation for supply of momentum. If a small control volume contains an amount of mass-energy m , then it also has a supply of momentum

$$\mathbf{G} = mg = \rho N \mathbf{g}$$

where \mathbf{g} is a vector that depends on the gravitational field (it expresses spacetime curvature). This constitutive relation expresses the *gravitational force*.

For physical phenomena on the Earth's surface, the vector \mathbf{g} is approximately constant: it points towards the Earth's centre and has magnitude $g := |\mathbf{g}| \approx 9.8 \text{ N/kg} \equiv 9.8 \text{ m/s}^2$.

For physical phenomena further away from Earth, for instance when dealing with satellite motion, the vector \mathbf{g} points towards the Earth's centre and has magnitude

$$g = \frac{GM_E}{r^2}$$

where $G = 6.67 \text{ N m}^2/\text{kg}^2$ is the gravitational constant, $M_E = 5.97 \times 10^{24} \text{ kg}$ is Earth's mass⁶, and r is the distance of the control volume from the Earth's centre.



(image from NASA⁵)

Constitutive relation for energy-mass of matter. Take a small control volume on Earth's surface containing an amount of matter N , and such that there is no flux of matter across its surface. Then this control volume also contains an amount of energy-mass

$$E = E_0 + U + \frac{1}{2}mv^2 + mgh$$

where the first term is a constant zero level of energy, the second term U is called *internal energy*, the third is called *kinetic energy*, and the fourth *gravitational potential energy*. The internal energy U is typically a function of the amount of matter N .

Matter enters this formula also through the constitutive relation $m = \rho N$. The zero value E_0 is usually unimportant because we mostly work with energy changes. This value is huge in ordinary situations.

↳ § 3.4 page 35

5.5 Numerical simulations

Recall again the integral and differential forms of a balance law (we use the symbols for energy here, but the reasoning is valid for any other quantity):

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t) dt + \int_{t_0}^{t_1} R(t) dt \quad \frac{dE(t)}{dt} = \Phi(t) + R(t)$$

integral form differential form

If $R = 0$, then this is a conservation law.

From either of these forms we can find an approximate mathematical expression to make predictions for a short instant of time. Let's see this from the integral form first.

Suppose that the lapse of time between t_0 and t_1 :

$$\Delta t = t_1 - t_0$$

is extremely short. So short that the flux $\Phi(t)$ and the supply $R(t)$ can't change appreciably during this time lapse, and we can take them as constant:

$$\Phi(t) \approx \Phi(t_0) \quad R(t) \approx R(t_0)$$

(if they do change appreciably, then we consider an even shorter Δt). If they are practically constant, then their integrals become

$$\begin{aligned} \int_{t_0}^{t_1} \Phi(t) dt &\approx \int_{t_0}^{t_1} \Phi(t_0) dt = \Phi(t_0)(t_1 - t_0) = \Phi(t_0)\Delta t \\ \int_{t_0}^{t_1} R(t) dt &\approx \int_{t_0}^{t_1} R(t_0) dt = R(t_0)(t_1 - t_0) = R(t_0)\Delta t \end{aligned}$$

Recall that

$$\int_a^b \text{const} dx = \text{const} \cdot (b - a)$$

And finally, considering that $t_1 = t_0 + \Delta t$, the balance law in integral form can be written approximately as

$$E(t_0 + \Delta t) \approx E(t_0) + \Phi(t_0)\Delta t + R(t_0)\Delta t$$

$$= E(t_1) \quad \approx \int_{t_0}^{t_1} \Phi(t) dt \quad \approx \int_{t_0}^{t_1} R(t) dt$$

Factoring Δt we obtain:

Finite-difference approximation

$$E(t_0 + \Delta t) \approx E(t_0) + [\Phi(t_0) + R(t_0)]\Delta t \quad (5.8)$$

called a **finite-difference approximation**.

This equation says that if we know the value of the volume content E , the influx Φ , and the supply R (which is zero for a conservation law) at time t_0 , then we can predict the value of the volume content a short time later, $E(t_0 + \Delta t)$.

For example, suppose that the total amount of energy in a given control volume, at time $t_0 = 20$ s is $E(t_0) = 500$ J. At that same time there's a net influx of $\Phi(t_0) = -30$ J/s into the control volume. And let's suppose that the supply is zero, $R(t_0) = 0$ J/s. The amount of energy in the control volume 0.1 s later, that is, at time $t_0 + \Delta t = 20$ s + 0.1 s = 20.1 s, is then

$$\begin{aligned} E(20.1 \text{ s}) &\approx E(t_0) + [\Phi(t_0) + R(t_0)] \Delta t \\ &\approx E(20 \text{ s}) + [\Phi(20 \text{ s}) + R(20 \text{ s})] \cdot 0.1 \text{ s} \\ &\approx 500 \text{ J} + [-30 \text{ J/s} + 0 \text{ J/s}] \cdot 0.1 \text{ s} \\ &\approx 497 \text{ J}. \end{aligned}$$

This basic idea is the same as the one behind [Euler's method](#)⁷.

! A control-surface sequence is understood

As you recall from the [definition of conservation and balance law](#), we must have chosen a *sequence of closed control surfaces*. How this sequence is chosen must be either clearly stated or understood from the context. If this sequence is not specified, the volume contents and fluxes above don't have any clear meaning.

↳ § 5.1 page 79

Exercise 5.3

- At time $t = 0$ s the amount of oxygen in a control volume is 0 mol, and at that instant there is an influx of 8 mol/s. Assume that oxygen satisfies a conservation law, and calculate its amount in the control volume at time $t' = 0.01$ s.
- Try to obtain the finite-difference approximation starting from the differential form of the balance law instead:

$$\frac{dE(t)}{dt} = \Phi(t) + R(t)$$

Use the fact that the derivative at a given time t_0 can be approximately calculated as

$$\frac{dE(t_0)}{dt} \approx \frac{E(t_0 + \Delta t) - E(t_0)}{\Delta t}.$$

5.5.1 Vector quantities

The finite-difference approximation is also valid for the balance law of a vector quantity like momentum. We only have to remember that we have three equations – one per vector component – instead of one:

$$\mathbf{P}(t_0 + \Delta t) \approx \mathbf{P}(t_0) + [\mathbf{F}(t_0) + \mathbf{G}(t_0)] \Delta t$$

or
$$\begin{cases} P_x(t_0 + \Delta t) \approx P_x(t_0) + [F_x(t_0) + G_x(t_0)] \Delta t \\ P_y(t_0 + \Delta t) \approx P_y(t_0) + [F_y(t_0) + G_y(t_0)] \Delta t \\ P_z(t_0 + \Delta t) \approx P_z(t_0) + [F_z(t_0) + G_z(t_0)] \Delta t \end{cases}$$

(5.9)

Exercise 5.4

- Let's follow the flight of a tennis ball by choosing a sequence of closed control surfaces, with corresponding control volumes, that tightly wrap it.

At time $t_0 = 0$ s the amount of momentum and the supply of momentum in the tennis ball are

$$\mathbf{P}(t_0) = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \text{ Ns} \quad \mathbf{G}(t_0) = \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \text{ N}$$

and the influx of momentum is zero. Calculate the momentum within the control volume 0.01 s later.

- The tennis ball has a mass-energy $m = 0.059$ kg. Assume the [constitutive relation for momentum](#):

$$\mathbf{P} = m \mathbf{v}$$

What was the velocity of the tennis ball at time t_0 ? How much is it 0.01 s later?



◀ §5.4.1 page 93

5.5.2 Iterating: numerical time integration and boundary conditions

The evolution equation (5.11) can obviously be used iteratively: once we have the volume content $E(t + \Delta t)$ at time $t + \Delta t$, we can use it to find the

value at a slightly later time $t + \Delta t + \Delta t$, and so on:

$$\begin{aligned}
 E(t + \Delta t) &\approx E(t) + [\Phi(t) + R(t)] \Delta t \\
 E(t + 2\Delta t) &\approx E(t + \Delta t) + [\Phi(t + \Delta t) + R(t + \Delta t)] \Delta t \\
 E(t + 3\Delta t) &\approx E(t + 2\Delta t) + [\Phi(t + 2\Delta t) + R(t + 2\Delta t)] \Delta t \\
 &\dots
 \end{aligned} \tag{5.10}$$

```

p.vx = p.vx + dt * p.fx / this.mass;
p.vy = p.vy + dt * (p.fy / this.mass + t

// integrate position
p.x = p.x + dt * p.vx;
p.y = p.y + dt * p.vy;

```

Code snippet⁸ from the [Particle System webpage⁹](#) iterating an evolution equation for momentum

Numerical time integration

The numerical calculation of a quantity at successive time steps, with algorithms similar to the one illustrated above, is called **numerical time integration**, often simply shortened to *integration*.

5.5.3 Boundary conditions and constitutive equations

In order to do the numerical time integration above, you notice that we need to know the new values of influx Φ and supply R at all subsequent time steps. These values are *not* given by the evolution equation (5.11).

In fact, the evolution equation from a balance or conservation law always gives us the *volume content* at a new time. So it would seem that we have no ways to predict a *flux* or a *supply*. Where do we get these from? There are two possibilities:

- They are simply assigned at every time step, completely or in part. This may be possible because they are known, measured, or controlled.

Boundary conditions

The quantities that we need to specify *at each time* for the prediction or simulation of a physical phenomenon are called **boundary conditions**, and their values *boundary values*.

For example, for isolated moving objects such as the tennis ball in the last exercise, it is known that the influx of momentum is zero if the object is moving in vacuum, or sometimes negligible when moving through air; and the constant value of the momentum supply (the gravitational force) is also known. They are both given as boundary conditions for that specific problems.

- They can be calculated, thanks to *constitutive relations*, from the values of volume contents of other quantities – which are in turn predicted with an evolution equation.

An example is the [Newtonian constitutive relation for momentum and matter \$\mathbf{P} = m\mathbf{v}\$](#) , briefly discussed previously. It actually relates the *flux* of matter ([which determines the velocity \$\mathbf{v}\$](#)) to the *volume content* of momentum. We shall see this relationship more clearly later in §[7.2](#), and explore and use more examples of constitutive relations for the purpose of simulating physical systems.

↳ §[5.4.1](#) page 93

↳ §[4.9](#) page 72

Exercise 5.5

In Exercise [5.4](#) you calculated the evolution of the tennis ball's momentum – all its three components – for one timestep of 0.01 s.

1. Write a script, in your preferred programming language, that implements the time-stepped evolution algorithm [\(5.10\)](#) and evolves all three components of momentum. Assume:

- total duration of simulation is 2 s
- timestep is $\Delta t = 0.01\text{s}$
- momentum at initial time $t_0 = 0\text{ s}$ is $\mathbf{P}(t_0) = [3, 0, 3]\text{ Ns}$
- momentum influx is zero at all times
- momentum supply is $\mathbf{G} = [0, 0, -0.579]\text{ N}$ at all times

The script should output the values of the three momentum components of the tennis ball at all times; plot each of them against time.

2. Recall the constitutive relation $\mathbf{P} = m\mathbf{v}$ for the tennis ball, with $m = 0.059\text{ kg}$. Use the results of your script to plot the three components of the velocity \mathbf{v} against time.
3. Make your script more general: it should initially ask for (or require as arguments, if implemented as a function):
 - the desired total simulation time
 - the timestep Δt to be used in the algorithm
 - the initial momentum $\mathbf{P}(t_0)$
 - the mass m of the tennis ball (or other flying object)

and then use the constant momentum supply

$$\mathbf{G} = m \cdot [0, 0, -9.81]\text{ N/kg}$$

5.5.4 Numerical time integration of position

In many applications we are interested in how the position \mathbf{r} of a small object, or of a small portion of an object, changes in time. The numerical time integration that we have discussed for the volume contents of physical quantities can be applied in an analogous way to the position vector \mathbf{r} and its velocity.

The velocity \mathbf{v} of a point is the time derivative of its position: $\mathbf{v}(t) = \frac{d\mathbf{r}(t)}{dt}$. For a small time step Δt we can approximately write

$$\mathbf{v}(t) \approx \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}$$

Multiply by Δt and bring to the right side all terms that refer to time t . We obtain

Time-stepped evolution equation for position

$$\mathbf{r}(t_0 + \Delta t) \approx \mathbf{r}(t_0) + \mathbf{v}(t_0) \Delta t \quad \text{or} \quad \begin{cases} x(t_0 + \Delta t) \approx x(t_0) + v_x(t_0) \Delta t \\ y(t_0 + \Delta t) \approx y(t_0) + v_y(t_0) \Delta t \\ z(t_0 + \Delta t) \approx z(t_0) + v_z(t_0) \Delta t \end{cases} \quad (5.11)$$

This formula says that if we know the position and the velocity at some time, we can approximately predict the position a short time later. This formula can also be iterated [as we did with a general balance law](#). This way we numerically time-integrate the position $\mathbf{r}(t)$ and can therefore keep track and predict the motion of an object.

↳ § 5.5.2 page 97

The numerical time integration of a position vector $\mathbf{r}(t)$ is often done together with that of momentum $\mathbf{P}(t)$. Thanks to the constitutive relation $\mathbf{P} = m\mathbf{v}$, knowledge of the momentum at a later timestep allows us to also know the velocity at that timestep. Here is how such an iteration might look like. Suppose that the initial $\mathbf{r}(t)$ and $\mathbf{P}(t)$ are given, and also the

influx \mathbf{F} and supply \mathbf{G} are known at all times:

$$\begin{aligned}
 \mathbf{v}(t_0) &= \mathbf{P}(t_0)/m \\
 \mathbf{r}(t_0 + \Delta t) &\approx \mathbf{r}(t_0) + \mathbf{v}(t_0) \Delta t \\
 \mathbf{P}(t_0 + \Delta t) &\approx \mathbf{P}(t_0) + [\mathbf{F}(t_0) + \mathbf{G}(t_0)] \Delta t \\
 \mathbf{v}(t_0 + \Delta t) &= \mathbf{P}(t_0 + \Delta t)/m \\
 \mathbf{r}(t_0 + 2\Delta t) &\approx \mathbf{r}(t_0 + \Delta t) + \mathbf{v}(t_0 + \Delta t) \Delta t \\
 \mathbf{P}(t_0 + 2\Delta t) &\approx \mathbf{P}(t_0 + \Delta t) + [\mathbf{F}(t_0 + \Delta t) + \mathbf{G}(t_0 + \Delta t)] \Delta t \\
 \mathbf{v}(t_0 + 2\Delta t) &= \mathbf{P}(t_0 + 2\Delta t)/m \\
 \mathbf{r}(t_0 + 3\Delta t) &\approx \mathbf{r}(t_0 + 2\Delta t) + \mathbf{v}(t_0 + 2\Delta t) \Delta t \\
 \mathbf{P}(t_0 + 3\Delta t) &\approx \mathbf{P}(t_0 + 2\Delta t) + [\mathbf{F}(t_0 + 2\Delta t) + \mathbf{G}(t_0 + 2\Delta t)] \Delta t \\
 \mathbf{v}(t_0 + 3\Delta t) &= \mathbf{P}(t_0 + 3\Delta t)/m \\
 &\dots
 \end{aligned} \tag{5.12}$$

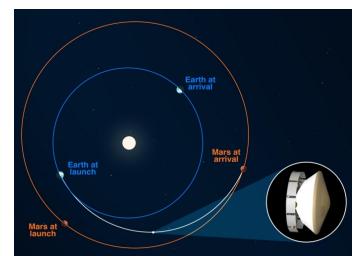
Keep in mind that these are vector equations, so each one corresponds to three components.

For some physical phenomena – such as planets orbiting around a star or satellites around a planet – the momentum influx \mathbf{F} or supply \mathbf{G} may also depend on the position \mathbf{r} . The scheme above works also in these cases.

5.5.5 Numerical time integration of position: example

Let's see an example with a falling object. For simplicity we only consider the z-coordinate, pointing upward, the other two coordinates having some constant value. Assume these conditions:

- initial time is $t_0 = 0$ s
- initial position is $z(t_0) = 5$ m
- initial momentum is $P_z(t_0) = 30$ N s
- mass-energy of object is $m = 4$ kg
- momentum supply, gravity, is $G_z = -m \cdot 9.81$ N/kg = -39.23 N
- timestep is $\Delta t = 0.1$ s



Sending probes, such as the [Perseverance rover](#)¹⁰, to other planets requires a careful prediction of their trajectory ([image: NASA](#)¹¹)

Then numerical integration up to $t = 0.2$ s would look as follows:

$$\begin{aligned} v_z(0 \text{ s}) &= P_z(0 \text{ s})/m \\ &= 30 \text{ N s}/4 \text{ kg} = 7.50 \text{ m/s} \end{aligned}$$

$$\begin{aligned} z(0.1 \text{ s}) &\approx z(0 \text{ s}) + v_z(0 \text{ s}) \Delta t \\ &\approx 5 \text{ m} + 7.50 \text{ m/s} \cdot 0.1 \text{ s} = 5.75 \text{ m} \end{aligned}$$

$$\begin{aligned} P_z(0.1 \text{ s}) &\approx P_z(0 \text{ s}) + [F_z(0 \text{ s}) + G_z(0 \text{ s})] \Delta t \\ &\approx 30 \text{ N s} - 39.23 \text{ N} \cdot 0.1 \text{ s} = 26.08 \text{ N s} \end{aligned}$$

$$\begin{aligned} v_z(0.1 \text{ s}) &= P_z(0.1 \text{ s})/m \\ &= 26.08 \text{ N s}/4 \text{ kg} = 6.52 \text{ m/s} \end{aligned}$$

$$\begin{aligned} z(0.2 \text{ s}) &\approx z(0.1 \text{ s}) + v_z(0.1 \text{ s}) \Delta t \\ &\approx 5.75 \text{ m} + 6.52 \text{ m/s} \cdot 0.1 \text{ s} = 6.40 \text{ m} \end{aligned}$$

$$\begin{aligned} P_z(0.2 \text{ s}) &\approx P_z(0.1 \text{ s}) + [F_z(0.1 \text{ s}) + G_z(0.1 \text{ s})] \Delta t \\ &\approx 26.08 \text{ N s} - 39.23 \text{ N} \cdot 0.1 \text{ s} = 22.15 \text{ N s} \end{aligned}$$

$$\begin{aligned} v_z(0.2 \text{ s}) &= P_z(0.2 \text{ s})/m \\ &= 22.15 \text{ N s}/4 \text{ kg} = 5.54 \text{ m/s} \end{aligned}$$

...



Exercise 5.6

Modify the script you made for part 1. of Exercise 5.5, so that it also evolves the position $\mathbf{r}(t)$ of the tennis ball.

Redo the numerical simulation you did for that exercise, with the same numerical parameters, and with an initial position for the tennis ball $\mathbf{r}(t_0) = [0, 0, 2]$ m.

Plot $z(t)$ and $P_z(t)$ against the time t , similarly to the two margin plots above.

Plots of $z(t)$ and $P_z(t)$ vs t obtained by continuing the numerical integration for 2 s

5.5.6 Applicability of numerical time integration

We have illustrated the numerical-evolution procedure, formulae (5.10) and (5.12), focusing on momentum and position. But obviously this procedure can be applied to any quantities that satisfy balance laws.

The time-stepping scheme discussed in the previous sections is very crude and quickly leads to increasing numerical errors; see for instance

the examples on the [Mass-Spring Model webpage¹²](#) of *Physics Simulation in Visual Computing*¹³. More refined and complex schemes are used for concrete applications.

Yet, the simple time-stepping scheme that we have explored remains the core of most numerical-evolution procedures, and allows us to understand how they essentially work.

5.5.7 Example script for numerical time integration

Here is an example script, written in [Octave¹⁴](#) (should also work in MATLAB), that is a solution for Exercise 5.6 p. 102. Blue lines are strictly related to numerical time integration; grey lines take care of saving and plotting data. You can use this script as a starting point for exercises that require scripting in the next chapters.

```

1 %% (base SI units used throughout)
2 %% Numerical simulation of object motion in 2D with gravity
3 %% Coordinates (y,z)
4 %% Parameters and initial values:
5 Py0 = 3; Pz0 = 0.75; % initial momentum
6 y0 = 0; z0 = 2; % initial position
7 Fy = 0; Fz = 0; % momentum influx (constant)
8 Gy = 0; Gz = -0.579; % momentum supply (constant)
9 m = 0.059; % tennis ball's mass
10 t0 = 0; % initial time
11 t1 = t0 + 2; % final time, can also be earlier than initial
12 dt = 0.01; % time step, negative if backward time integration
13 %% adjust final time if not multiple of timestep
14 t1 = t1 + mod(t1-t0,dt);
15 %% check if timestep and time interval are consistent
16 if (t1-t0)*dt <= 0
17     error('time interval inconsistent with timestep');
18 end
19 %% Save values of all quantities at some steps during the simulation,
20 %% for subsequent analysis or plotting
21 %% (saving at all timesteps could be too costly)
22 Nsaves = 200; % number of timepoints to save during the simulation
23 %% Calculate time interval for saving
24 dsave = (t1-t0)/(Nsaves-1);
25 if abs(dsave) < abs(dt)
26     error('time interval between saves is smaller than timestep')
27 end
28 %% Initialize vectors to contain saved values
29 tSave = nan(Nsaves,1);
30 ySave = nan(Nsaves,1); zSave = nan(Nsaves,1);

```

[Download tennisball.m¹⁵](#)

```

31 PySave = nan(Nsaves,1); PzSave = nan(Nsaves,1);
32 vySave = nan(Nsaves,1); vzSave = nan(Nsaves,1);
33 %% Save initial values
34 i = 1; % index that keeps count of savepoints
35 tSave(i) = t0;
36 ySave(i) = y0; zSave(i) = z0;
37 PySave(i) = Py0; PzSave(i) = Pz0;
38 vySave(i) = Py0/m; vzSave(i) = Pz0/m;
39 %% Initialize plot
40 cols = get(0, 'DefaultAxesColorOrder');
41 plot(ySave(1), zSave(1), '.', 'Color', cols(1,:)); axis('tight');
42 xlabel('y/m'); ylabel('z/m'); hold on;
43 %%
44 %% Numerical time integration
45 %% initialize
46 t = t0;
47 y = y0; z = z0;
48 Py = Py0; Pz = Pz0;
49 vy = Py/m; vz = Pz/m;
50 %% loop
51 while sign(dt)*t < sign(dt)*t1 % possible backward time integration
52   %% update time
53   t = t + dt;
54   %% update momentum
55   Py = Py + (Fy + Gy)*dt;
56   Pz = Pz + (Fz + Gz)*dt;
57   %% update velocity (from constitutive relation P=mv)
58   vy = Py/m;
59   vz = Pz/m;
60   %% update position
61   y = y + vy*dt;
62   z = z + vz*dt;
63   %% Check whether to save & plot at this step
64   if min(abs([0 dsave] - mod(t-t0, dsave))) <= abs(dt)/2
65     i = i+1;
66     tSave(i) = t;
67     ySave(i) = y; zSave(i) = z;
68     vySave(i) = vy; vzSave(i) = vz;
69     PySave(i) = Py; PzSave(i) = Pz;
70     plot(y, z, '.', 'Color', cols(1,:));
71     pause(0.001);
72   end
73 end
74 %% Plot trajectory
75 plot(ySave, zSave, 'Color', cols(1,:)); axis('tight');

```

URLs for chapter 5

1. https://pglpm.github.io/7wonders/media/volume_moving.gif
2. <https://pubchem.ncbi.nlm.nih.gov/element/Argon>
3. <https://encyclopedia.che.ingen.umich.edu/distillation-columns/>
4. https://commons.wikimedia.org/wiki/File:Four_Fundamental_States_of_Matter.png
5. https://nssdc.gsfc.nasa.gov/imgcat/html/group_page/ER.html
6. <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
7. https://encyclopediaofmath.org/wiki/Euler_method
8. https://github.com/InteractiveComputerGraphics/physics-simulation/blob/5fbfb1022ac8b055bf52b197da22b0c2e4122d5f8/examples/particle_system.html#L181
9. https://interactivecomputergraphics.github.io/physics-simulation/examples/particle_system.html
10. <https://mars.nasa.gov/mars2020/>
11. <https://www.jpl.nasa.gov/news/nasas-perseverance-rover-is-midway-to-mars>
12. https://interactivecomputergraphics.github.io/physics-simulation/examples/spring_plot.html
13. <https://interactivecomputergraphics.github.io/physics-simulation/>
14. <https://octave.org/>
15. <https://pglpm.github.io/7wonders/code/tennisball.m>

Chapter 6

The Seven Wonders of the world

[Marco Polo:] “... You take delight not in a city’s seven or seventy wonders, but in the answer it gives to a question of yours.”

[Kublai Khan:] “Or the question it asks you, forcing you to answer, like Thebes through the mouth of the Sphinx.”

I. Calvino 1979

6.1 Seven universal laws

The reason for introducing the [seven primitive quantities](#), which are common to all our main physical theories, is that each of them obeys a [balance law](#). More precisely: *matter*, *electric charge*, and *magnetic flux* actually obey conservation laws, like (5.1). Whereas *energy*, *momentum*, *angular momentum*, and *entropy* satisfy balance laws, like (5.3), that may have a supply; only in special circumstances these quantities also obey conservation laws.

↳ § 3 page 30

↳ § 5 page 76

What’s remarkable about these seven balance laws?

- They are known, so far, to be satisfied by *all* physical phenomena, from subatomic scales to cosmological scales. No exceptions are known.
- They are satisfied in *all* our main physical theories, approximate or not: from Newtonian mechanics to special relativity, from general relativity to quantum theory.
- Each one can be expressed by the *same* mathematical equation in all of these theories.

They are therefore, as far as we know, universal:

 **The seven universal balance laws**

- Conservation of matter**
- Conservation of electric charge**
- Conservation of magnetic flux**
- Balance of energy**
- Balance of momentum**
- Balance of angular momentum**
- Balance of entropy**

These balances are truly the Seven Wonders of the World, even more long-lasting than the traditional “seven wonders”¹ or the “new seven wonders”².

These laws appear in all our physical theories and physics formalisms. In some physical theories, such as Newtonian thermo-mechanics, all these balance laws are, or can easily be taken to be, the fundamental laws on which the theory is built. In other theories some of these laws are taken as fundamental, while others are *derived* from more fundamental laws – but nevertheless all these balances are still satisfied. In general relativity, for instance, the conservation of matter, electric charge, magnetic flux, and the balance of entropy are taken as fundamental, but the balances of energy, momentum, angular momentum are a *consequence* of the so-called Einstein equations.

It is therefore extremely useful to learn these universal balances, no matter what kind of specialized theory and physical phenomena you may end up working with in the future. You can apply these balances to any kind of physical phenomenon or problem: construction of bridges, control of chemical reactions, operation of GPS navigation and satellites, monitoring of nuclear power plants, sending robots to Mars, collisions of subatomic particles, cosmology. Every physical phenomenon involves at least one of these seven balances in its physical description.

The Einstein equations, which can be deceptively simply written as

$$\mathbf{G} = \frac{8\pi G}{c^4} \mathbf{T},$$

include the balances of energy, momentum, angular momentum as special consequences.

6.2 General form of the universal laws

It is important to always keep in mind the setup that lies behind each universal balance law.

Recall that a balance or conservation law always involves a choice of coordinates (t, x, y, z) (including a coordinate time) and *sequence of closed control surfaces*: it states that a particular “budget” must hold when we check what’s within these surfaces and what flows through them.

↳ § 5.2 page 78

Also recall that we are free to choose this sequence of closed control surfaces as we please. A balance law is so powerful because it holds no matter what our choice of surfaces is.

The “budget” involves three amounts:

volume content: how much of a quantity is within a control volume;

influx: how much of a quantity is flowing in through a closed control surface per unit time;

supply: how much of a quantity is being created or destroyed within a control volume per unit time; this is zero for a conservation law.

The budget is expressed in the following general forms:

General form of the universal balance laws (again)

$$\text{vol. integral}(t_1) = \text{vol. integral}(t_0) + \int_{t_0}^{t_1} \text{influx}(t) dt + \int_{t_0}^{t_1} \text{supply}(t) dt$$

integral form

$$\frac{d \text{vol. integral}(t)}{dt} = \text{influx}(t) + \text{supply}(t)$$

differential form

If the supply(t) must always be zero, then we have a *conservation law*.

The entropy balance is similar, except that its supply is left unspecified and must always be non-negative.

The expressions above are valid in Newtonian mechanics, general relativity, and even quantum theory if the symbols are interpreted as so-called ‘statistical operators’, which encode probabilistic properties.

The individual balances and symbols for matter, momentum, energy, angular momentum, entropy are also presented in table 6.1 p. 111.

Concise mathematical form of the universal balances

The seven balances can be expressed very concisely if we use the language of *differential forms*³. These are geometric objects that associate a number to any curve, surface, or volume of our choice. In terms of differential forms, even the difference between the ‘integral’ and ‘differential’ forms of the balance laws disappear. The balances for matter, momentum, energy, angular momentum, entropy then assume these very concise expressions:

$$dN = -A \quad dE = R \quad d\mathbf{P} = \mathbf{G} \quad d\mathbf{L} = \mathbf{M} \quad dS \geq 0$$

similarly for the conservation of electric charge and magnetic flux.

If you want to learn more about differential forms take a look at the books by Burke 1987; 1995 and Bossavit 1991.

6.3 Roles of the seven balances in description and prediction

The seven universal balances govern every physical phenomenon. Yet this doesn’t mean that all of them are always used explicitly in the description or prediction of physical phenomena.

- For some physical phenomena, all the seven universal balances enter our calculations.
- For other physical phenomena, some of the seven balances do not appear *explicitly* in our calculations; yet they still enter *implicitly* in the way we choose to set up or describe the phenomenon.

For example, we may choose control volumes or control surfaces in such a way that some conservation laws are automatically satisfied. A typical case is the choice of control surface around a given object (amount of matter), which guarantees that the law of conservation of matter is automatically satisfied. As another example, sometimes we simplify a physical phenomenon to one spatial dimension only; think of when we throw a ball vertically in the air, and only consider its height from the ground. In such a case we can make some predictions using only the balance of energy, apparently avoiding the balance of momentum. But in reality, the fact that the ball can be considered as moving vertically is possible because momentum is balanced in the horizontal directions. The balance of momentum is therefore still necessary for this prediction, but it has been silently taken care of.



The description of how a common lighter works, thanks to piezoelectricity⁴, requires more or less all seven universal laws to be explicitly accounted for.

- For still other physical phenomena, some of the seven balances may not be required because we do not need the kind of physical information they provide. We saw an example of this with the [flat-tire problem](#), where we only used the conservation of matter but we weren't interested in what happened to other quantities such as energy or momentum, or in how the tyre was moving.

↳ § 5.2.2 page 81

Some physical phenomena may be predicted using one particular subset of the seven balances, or alternatively a different subset, as we please. For instance a given problem might be solved using conservation of matter and balance of momentum, or alternatively by using conservation of matter and balance of energy. We shall see examples of all these possibilities in the rest of these notes.

But the fact that the seven universal balances govern every physical phenomenon doesn't mean that they can be used alone, by themselves. In the vast majority of cases they need to be augmented by appropriate constitutive equations. We saw examples of this fact in the chapter about [physical laws](#).

↳ § 5 page 76

In the next chapters we shall discuss the seven universal balance laws in more detail. For almost each of them we shall give its mathematical expression, discuss some constitutive relations that are commonly used with it, and examine some example applications.

Note that some constitutive relations will reappear again and again in different chapters, because they connect different quantities, and so must be discussed in connection of each of the quantities they connect.

	integral form	differential form
matter	$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t) dt - \int_{t_0}^{t_1} A(t) dt$	$\frac{dN(t)}{dt} = J(t) - A(t)$
momentum	$\mathbf{P}(t_1) = \mathbf{P}(t_0) + \int_{t_0}^{t_1} \mathbf{F}(t) dt + \int_{t_0}^{t_1} \mathbf{G}(t) dt$	$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}(t) + \mathbf{G}(t)$
energy	$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t) dt + \int_{t_0}^{t_1} R(t) dt$	$\frac{dE(t)}{dt} = \Phi(t) + R(t)$
angular momentum	$\mathbf{L}(t_1) = \mathbf{L}(t_0) + \int_{t_0}^{t_1} \boldsymbol{\tau}(t) dt + \int_{t_0}^{t_1} \mathbf{M}(t) dt$	$\frac{d\mathbf{L}(t)}{dt} = \boldsymbol{\tau}(t) + \mathbf{M}(t)$
entropy	$S(t_1) \geq S(t_0) + \int_{t_0}^{t_1} H(t) dt$	$\frac{dS(t)}{dt} \geq H(t)$

Table 6.1 Five of the seven universal balance laws. These formulae are valid in Newtonian mechanics, general relativity, and even quantum theory if their symbols are interpreted as ‘statistical operators’.

URLs for chapter 6

1. <https://education.nationalgeographic.org/resource/seven-wonders-ancient-world/>
2. <https://www.britannica.com/list/new-seven-wonders-of-the-world>
3. http://encyclopediaofmath.org/index.php?title=Differential_form
4. <https://www.britannica.com/science/piezoelectricity>

Chapter 7

Conservation & balances of matter

"What you do in this world is a matter of no consequence," returned my companion, bitterly. "The question is, what can you make people believe that you have done? . . ."

Sherlock Holmes (A. C. Doyle) 1887

7.1 Formulation and generalities

 Balance and conservation of matter

Volume content: N Flux: J Supply: $-A$

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t) dt - \int_{t_0}^{t_1} A(t) dt \quad \frac{dN(t)}{dt} = J(t) - A(t)$$

integral form

differential form

(7.1)

The supply A is usually called **activity**, and $A(t) \equiv 0$ in case of conservation.

Conservation of matter is a law that we intuitively take for granted and use continuously in our life. The very notion of ‘object’ – including living objects – is possible thanks to this fundamental regularity of nature: we can speak of objects because they exist for some time and we can follow them as they move in space.

7.1.1 Balance vs conservation of matter

The law of *conservation* of matter holds, as far as we know, for all kinds of matter *when considered together.*

Wilczek & Zee 1979

"All these operators conserve $B - L$ [number of baryons B minus number of leptons L], so in any superunified theory we expect $B - L$ to be conserved... "

More specifically it seems to hold, even in extreme physical conditions, if we count baryonic and *anti*-leptonic matter as 'positive' and leptonic and anti-baryonic matter as 'negative'. As an example, consider a closed surface containing two neutrons (baryon number +1), and some energy, in such a way that the total electric charge, momentum, and angular momentum are zero. Then it's today considered impossible that the two neutrons could just disappear, and two protons (baryon number +1 each) plus two electrons (lepton number +1 each) appear in their stead, even if energy, momentum, and so on were exactly the same. This is because otherwise we would have first $N(t_0) = +2$ and then $N(t_1) = +2 - 2 = 0$, without any flux of matter: $J(t) = 0$; conservation of matter would be broken.

In many common and important physical situations, conservation laws still hold for different kinds of matter *individually*. This is why we can consider the amounts of different chemical elements – hydrogen, helium, and so on – to be individually conserved in chemical reactions happening in a chemical plant.

In other circumstances, however, conservation of these individual kinds of matter does not hold anymore; for example in the previously mentioned **phenomena involving radioactive decay and nuclear energy**. Yet in these circumstances we can still apply a *balance* law, with a non-zero supply or sink.

◀ § 3.2 page 32

7.2 Examples of constitutive relations

7.2.1 Relation between matter and mass-energy

The most common constitutive relation used together with conservation of matter is the one relating an amount of matter N with an amount of mass-energy m .

As far as we know, if a control volume contains an amount of matter N , then it always also contains an amount of mass-energy m (but the opposite is not always true: a control volume can contain mass-energy and no matter). This amount of mass-energy is typically huge when measured in joules. For instance, a volume with $N = 1 \text{ mol}$ of water (roughly $2 \times 10^{-5} \text{ m}^3$ for liquid water) contains approximately $m = 1.62 \times 10^{15} \text{ J}$ of mass-energy.

The amount of mass-energy depends on the kind of matter and on other quantities in that volume, and usually changes with time. These changes, however, are typically **extremely small compared to the total amount** – say much less than 0.000 000 01 % in the example above with water! So, as a very good approximation, we can consider this amount as practically constant. The amount of mass-energy m is therefore proportional to the amount of matter N , and we write

↳ § 3.4.2 page 37

Mass-energy of matter

$$m = \rho N$$

where the proportionality factor ρ , called **molar mass**, depends on the kind of matter and can be taken as practically constant in many applications.

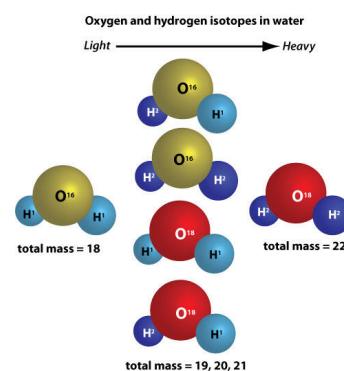
For instance, in many physical phenomena involving water we assume the constitutive relation above, with a **water molar-mass constant¹** of approximately $\rho_{\text{H}_2\text{O}} = 0.0180 \text{ kg/mol}$. If a volume contains $N = 20 \text{ mol}$ of water, we usually attribute to it a mass-energy of

$$\begin{aligned} m &= \rho_{\text{H}_2\text{O}} N \\ &= 0.0180 \text{ kg/mol} \times 20 \text{ mol} \\ &= 0.36 \text{ kg}. \end{aligned}$$

The exact value of the molar-mass constant depends not only on the substance but also on the context and application: the substance may actually consist of a mixture of different kinds of matter. ‘Air’ for example is a mixture of different chemical elements, and their proportion in the mixture may depend on physical conditions such as temperature, and on geographical position. ‘Water’ is a mixture of different **isotopes³** (molecules differing in the number of neutrons), and again the mixture proportions may vary with physical conditions.

The constant relating mass-energy and amount of matter is the same for volume contents and for fluxes. So to a matter flux J we can associate a mass-energy flux ρJ . For instance, if through a surface there’s a flux of $J = 30 \text{ mol/s}$ of water, then we associate to it a mass flux of

$$\rho_{\text{H}_2\text{O}} J = 0.0180 \text{ kg/mol} \times 30 \text{ mol/s} = 0.54 \text{ kg/s}.$$



There are several different isotopes of water, with different masses (image from U.S. Geological Survey²)

7.2.2 Conservation of mass as a proxy for conservation of matter

In applications where we can consider the molar mass as practically constant, we can then use ‘mass’ as a *proxy* for matter, and express the balance of matter (7.1) as “conservation of mass” instead: we only need to multiply volume content N , flux J , and supply A by the molar-mass ρ appropriate to that kind of matter.

! “Conservation of mass” is a proxy for conservation of matter

Keep in mind that many books speak of “conservation of mass”, but they’re using mass as an approximate proxy for matter, in the sense explained above.

Exercise 7.1

At a particular time, a party balloon contains 0.012 kg of Helium. A minute later, the same balloon contains 0.010 kg of helium. Assume conservation of Helium, as well as a [molar-mass for Helium⁴](#) $\rho_{\text{He}} = 4 \times 10^{-3} \text{ kg/mol}$.

1. How much is the *integrated efflux* of Helium, **in moles**, through the balloon’s surface during the one-minute lapse of time?
2. Assume that the efflux of Helium was constant in time during the one-minute time lapse. How much was the efflux of Helium, in moles/second?



7.2.3 Radioactive decay

For radioactive substances, for which conservation does not hold for individual kinds of matter, we still have a balance law where the supply $-A$ has a specific constitutive equation:

■ Matter supply in radioactive decay

$$-A(t) = -\lambda N(t)$$

where λ is positive and called the **decay constant** of that particular substance. The supply with changed sign, A , is usually called **activity⁵**.

If there is no influx of matter, $J(t) = 0$, the balance law for the substance then takes the form

$$\frac{dN(t)}{dt} = -\lambda N(t)$$

called the *law of radioactive decay*⁶.



When we see this symbol we know that there's matter that's only balanced, not conserved – which involves danger

7.3 Examples of applications

7.3.1 Rigid-body and particle mechanics

In many applications, the law of conservation of matter can be used in such a subtle way that we almost don't notice that we're actually using it.

This happens when we choose closed control surfaces through which there is no flux of matter: $J(t) = 0$. For example, in studying solid objects, we choose control surfaces that tightly “wrap” and follow the object; think of what we did in the tennis-ball examples and exercises of the chapter on [physical laws](#). With this choice, the amount of matter N within the control surfaces doesn't change in time, thanks to the law of conservation of matter. So this law is hidden in the fact that we're taking that amount of matter as constant.

We saw an example of this procedure in the [numerical evolution of the motion of a falling object](#). Take again a look at the timesteps in formula (5.12): at each timestep the mass m was assumed to be constant, not changing with time. But as explained above, [this mass is proportional to the amount of matter](#): $m = \rho N$. So this assumption was guaranteed, implicitly, by the law of conservation of matter.

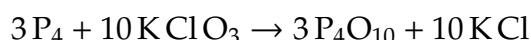
↳ § 5.3.3 page 87

↳ § 5.5.5 page 101

↳ § 7.2.2 page 116

7.3.2 Chemistry

One of the main assumptions in chemistry is the ‘permanence of atoms’. This assumption imposes important restrictions in the [stoichiometry](#)⁷ of chemical reactions, that is, in determining the amounts of products that can appear from given amounts of reactants. For instance, the [match-head reaction](#)⁸



expresses that if 30 moles of oxygen (O) atoms appear among the reactants, they must also appear among the products; same for the 12 moles of phosphorus (P) atoms⁹, the 10 moles of potassium (K) atoms¹⁰, and so on.

This assumption is simply the statement of conservation of matter, *separately* for each chemical element: since the reaction doesn't let any other matter go in or out, the flux J for each chemical element must be zero. Therefore the amount N of each chemical element must be constant: $dN(t)/dt = J = 0$.

The assumption of the permanence of atoms is only approximate and no longer valid in phenomena for which nuclear physics or particle physics become relevant.

7.3.3 Climate

The laws of conservation and balances of matter turn out to be very useful also in problems related to climate.

On the Earth's surface and atmosphere we can assume a law of conservation of matter for each of the [stable isotopes](#)¹¹ of the chemical elements, for instance the [two stable carbon isotopes](#)¹² and the [three stable oxygen isotopes](#)¹³. We can therefore follow these isotopes as they flow between different physical systems, such as the atmosphere, the oceans, and the biosphere, especially plants – in practice we are using huge control volumes.

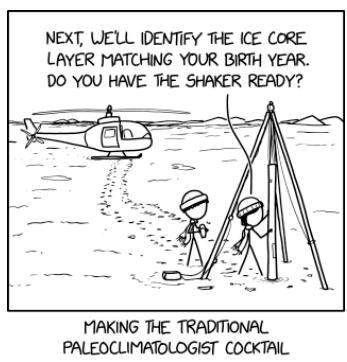
For the radioactive isotopes, the [law of radioactive decay](#) applies, and from it we can deduce the age of different materials and objects such as ice or wood.

This is how we are able to say that human usage of fossil fuel is an important factor in the increase of carbon dioxide (CO_2) in the atmosphere during the past 200 years or so. Take a look at the more detailed explanations given by the [Global Monitoring Laboratory](#)¹⁴.

↳ § 14.2.4 page 172

Exercise 7.2

Measuring the relative amounts of carbon within an air pocket near the surface of a block of ice, you find that one mole of air contains 1×10^{-12} mol of the radioactive isotope ^{14}C . Making an analogous measurement for an air pocket deeper in the ice, you find instead a relative abundance of 2×10^{-13} mol of ^{14}C . How old is the deeper section of ice?



<https://xkcd.com/2902>

Find out the ice's age t_{ice} by numerically evolving the balance equation

$$\frac{dN(t)}{dt} = -\lambda N(t)$$

with $\lambda = 0.000\,122 \text{ yr}^{-1}$, until you reach the amount $N(t_{\text{ice}}) = 2 \times 10^{-13} \text{ mol}$. Use the procedure of eq. (5.11), starting from time $t_0 = 0 \text{ yr}$ and $N(t_0) = 1 \times 10^{-12} \text{ mol}$, and assuming that the flow $J(t)$ of ^{14}C is zero. You can use a timestep $\Delta t = (1/365) \text{ yr}$.

↳ § 5.5 page 95

7.3.4 Nozzle flow

The law of conservation of matter in its explicit form is at the heart of fluid-dynamic problems. Consider the flow of a fluid (liquid or gas) for instance through a pipe or through a jet engine. When we say that a flow is **steady** we mean that the volume contents and the fluxes taken for whatever control volumes and surfaces do not change in time (though they may change in space). This condition can be viewed as a constitutive relation.

Consider a control volume, for instance the one indicated in light blue in the side picture. The amount of fluid N in this volume is constant in time: $\frac{dN(t)}{dt} = 0$. By the law of conservation of matter, the total influx must then be zero:

$$0 = \frac{dN(t)}{dt} = J(t)$$

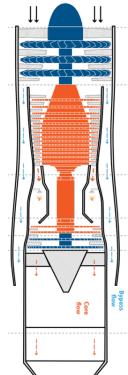
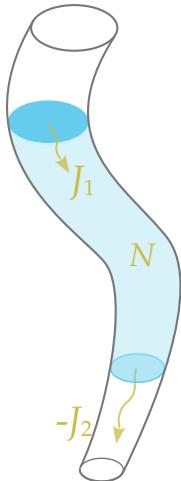
and it is given by the influxes through three surfaces: the side surface, the one at the top, and the one at the bottom. The influx through the side surface is zero. Let us denote the influx through the top surface by J_1 , and the influx through the bottom surface by J_2 . We must therefore have

$$0 = J(t) = J_1(t) + J_2(t) \implies J_1(t) = -J_2(t)$$

that is, the influx through the top surface must equal the efflux through the bottom one. This is a very powerful deduction: consider that it is valid even if the flow of the fluid is turbulent, and it is valid in more general situations such as at different sections of a jet engine.

If the surfaces are small enough, we can also use the connection between flux and velocity:

$$v_1 = \frac{J_1/A_1}{N/V} \quad v_2 = \frac{-J_2/A_2}{N/V}$$



(image from JetX¹⁵)

↳ § 4.9 page 72

where A_1, A_2 are the areas of the top and bottom surfaces, and v_1, v_2 the *downward* velocities through them (hence the minus sign for the bottom surface). From $J_1 = -J_2$ we then find this important relationship:

$$v_1 A_1 = v_2 A_2$$

that is, if the area through which the flux occur decreases, then the velocity of the fluid through it increases, and vice versa. This is what we often observe in water running from our taps. We can feel with our fingers that the water stream is slightly faster at the bottom, right before it hits the basin; and at this point the stream usually also thinner than at the top (the presence of an aerator or of turbulence can mask this effect).



The thickness and velocity variations of tap water are a consequence of the law of conservation of matter

URLs for chapter 7

1. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/H20/h1H2>
2. <https://www.usgs.gov/media/images/water-isotopes-diagram>
3. <https://www.britannica.com/science/isotope>
4. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/He>
5. <https://goldbook.iupac.org/terms/view/A00114>
6. <https://www.britannica.com/science/radioactivity/Rates-of-radioactive-transitions#ref496415>
7. <https://doi.org/10.1351/goldbook.S06026>
8. <https://chem.washington.edu/lecture-demos/match-head-reaction>
9. <https://pubchem.ncbi.nlm.nih.gov/element/Phosphorus>
10. <https://pubchem.ncbi.nlm.nih.gov/element/Potassium>
11. <https://www.britannica.com/science/isotope/The-discovery-of-isotopes#ref496311>
12. <https://education.jlab.org/itselemental/ele006.html>
13. <https://education.jlab.org/itselemental/ele008.html>
14. <https://gml.noaa.gov/outreach/isotopes/>
15. <https://www.jet-x.org/a8.html>

Chapter 8

Balance of momentum

LEX II.

Mutationem motus proportionalem esse vi motrici impressæ, & fieri secundum lineam rectam qua vis illa imprimitur.

I. Newton 1726a

"LAW II. The change of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed."

8.1 Formulation and generalities

Balance of momentum

$$\begin{aligned} \mathbf{P}(t_1) &= \mathbf{P}(t_0) + \int_{t_0}^{t_1} \mathbf{F}(t) dt + \int_{t_0}^{t_1} \mathbf{G}(t) dt & \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}(t) + \mathbf{G}(t) \\ \text{integral form} & & \text{differential form} & \\ & & & (8.1) \end{aligned}$$

Among the seven universal balances, the balance of momentum is probably the one most used in applications where motion or stability are important. Newton's famous "second law" is included in this balance as a special case, if we disregard the specific association that Newton made between momentum and velocity.

Amounts of momentum depend on the coordinate system

Remember that the amount of momentum \mathbf{P} in a volume, the flux \mathbf{F} through a surface, and the supply \mathbf{G} in a volume all depend on the

specific coordinate system (t, x, y, z). If we use a different coordinate system, these amounts will be different *for the same volume and surface*. For example, a tennis ball can contain a huge amount of momentum with respect to one coordinate system, and zero momentum with respect to another!

The importance of the balance of momentum comes to a great extent from two kinds of constitutive relations:

- a small number of constitutive relations that connect the volume content \mathbf{P} of momentum with the motion and flux of matter and of the electromagnetic field;
- an amazingly wide variety of constitutive relations that connect, in the most diverse ways, the flux \mathbf{F} of momentum with many properties of matter, such as its extension, deformation, motion, temperature, and the simultaneous presence of charge and electromagnetic field.

Through this balance we can therefore describe and predict motions. We saw a concrete example of this application in the [numerical time integration of the motion of a falling object](#).

But this balance is also essential in the opposite problem: when we need to study things that *don't and shouldn't move*, like a building or a bridge. In this case the balance is used to study which momentum fluxes \mathbf{F} and supply \mathbf{G} are necessary to ensure that the volume content \mathbf{P} of momentum is constantly zero.



Momentum balance keeps buildings from collapsing
(photo: Bryggen, Bergen, from UNESCO World Heritage Centre¹)

8.2 Examples of constitutive relations

8.2.1 Newtonian momentum: relation between momentum and matter

One of the most important constitutive relations is the one connecting momentum with the flux of matter. We have mentioned it several times in the previous chapters, and used it in [examples of numerical integration](#).

Consider a *small* control volume containing an amount of matter N . From the [constitutive equation between matter and mass-energy](#) we know that this volume also contains an amount of mass-energy $m = \rho N$. The matter in this volume also has a velocity \mathbf{v} (possibly zero), [related to its flux](#). The **Newtonian formula for the momentum of matter** then says that the control volume also contains an amount of momentum \mathbf{P} that is proportional to the mass-energy and the velocity:

↳ § 5.4 page 97

↳ § 7.2.1 page 114

↳ § 4.9 page 72

Newtonian constitutive relation for momentum

If a control volume contains an amount of matter N having mass-energy m and velocity \mathbf{v} , then it also contains an amount of momentum

$$\mathbf{P} = m\mathbf{v} = \rho \frac{V}{A} \begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} \quad (8.2)$$

This relation is only valid for very small areas and volumes (we should take the limit).

You are probably familiar with the first equality above; old textbook present it as the *definition* of momentum. We know today that this equality is only approximate: it's only valid for small speeds compared to the speed of light, in weak gravitational fields, and when changes in energy-mass are small compared to the total.

8.2.2 Hookean spring: relation between momentum flux and distance

We are all familiar with elastic bands, springs, and similar objects having the following approximate properties: (a) one of their dimensions is somehow singled out, maybe because more extended, with respect to the other two; (b) if we try to modify their length along that particular dimension, they exert a tension or, in some cases, a pressure; (c) the tension or compression are stronger, the more the extension in that dimension is modified; (d) these objects return to an initial configuration when we don't try to modify their extension.

We know that [tension and compression are particular kinds of forces, that is, momentum fluxes](#). Objects like elastic bands and springs therefore create particular momentum fluxes that are related to their spatial extension. Their behaviour is therefore described by a *constitutive relation about the flux of momentum \mathbf{F}* .



◀ § 4.6 page 68

We can approximately treat these objects as control volumes for which the following constitutive properties apply:

Hookean spring

A **Hookean spring** is a control volume with the following properties:

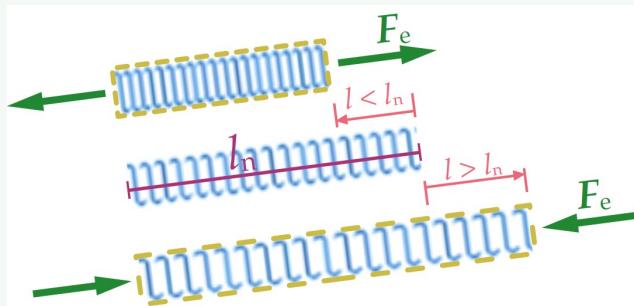
- One dimension has a natural length l_n that is much larger than the other two, which are usually neglected.

- The total amount and supply of momentum \mathbf{P} within the control volume are negligible.
- Across the (small) surfaces orthogonal to main dimension there is an **efflux** of momentum equal to

$$\mathbf{F}_e = -k \Delta \mathbf{l} \quad (8.3)$$

this formula is called **Hooke's law**.

k is a constant, called the **elastic constant** of the spring; $\Delta \mathbf{l}$ is a vector along the main dimension, with magnitude $\Delta l = |l - l_n|$ equal to difference between the present length l and the natural length l_n ; it points *away* from the control volume if $l > l_n$, and *toward* the control volume if $l < l_n$.



Note the minus sign in Hooke's law. It means that if the present length is larger than the natural one, $l > l_n$, then there is an influx of momentum directed towards the spring, that is, the spring exerts a *tensile* force towards itself; and if the present length is smaller than the natural one, $l < l_n$, then there is an influx of momentum directed away from the spring, that is, the spring exerts a *compressive* force away from itself.

In modelling some physical problems we can put $l_n = 0$ m: this represents that the natural length of the spring is very small compared to usual amounts by which the spring is stretched. Note that in this case the spring always exerts a tension, never a pressure.

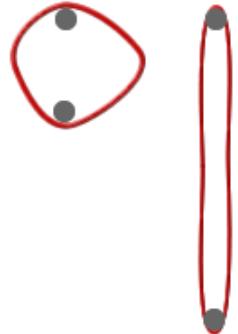
Exercise 8.1

What are the physical dimensions of the elastic constant k ? in which units could we measure it?

8.2.3 Non-hookean springs

Items like rubber bands and strings deviate from Hooke's law in a notable way: they exert a purely tensile force, that is, they effect a tensile flux of momentum, only if they are stretched beyond their natural, relaxed length.

This kind of behaviour can of course be modelled, at least qualitatively, by a constitutive equation such as the following:



Constitutive equation for rubber bands and strings

Suppose that the two extremities of a rubber band have position vectors \mathbf{r}_a and \mathbf{r}_b , and that the natural length of the rubber band is l_n . Then the momentum effluxes at the extremities \mathbf{r}_a and \mathbf{r}_b are approximately given by

$$\mathbf{F}_{ab} = \begin{cases} 0, & \text{if } |\mathbf{r}_a - \mathbf{r}_b| \leq l_n \\ -k (|\mathbf{r}_a - \mathbf{r}_b| - l_n) \frac{\mathbf{r}_a - \mathbf{r}_b}{|\mathbf{r}_a - \mathbf{r}_b|}, & \text{if } |\mathbf{r}_a - \mathbf{r}_b| > l_n \end{cases} \quad (8.4)$$

Note that $\frac{\mathbf{r}_a - \mathbf{r}_b}{|\mathbf{r}_a - \mathbf{r}_b|}$ is just a *unit vector* directed from the position \mathbf{r}_b to the position \mathbf{r}_a .

A string can be approximately modelled by such a constitutive equation with a large constant k .

This constitutive equation is difficult to be used with analytical methods, but can be amenable to numerical time integration

8.2.4 Pairwise forces

A Hookean or non-Hookean “spring” is modelled as having no mass-energy or momentum of its own. It is essentially a one-dimensional device that effects an *instantaneous* flux of momentum between its two extremities.

We can clearly employ this general idea to model other, wildly different, physical phenomena:

Pairwise forces

Consider a phenomenon where these three main conditions are met:

- a transfer of momentum occurs between two *separate* control volumes
- this transfer can approximately be treated as *instantaneous*

- mass-energy and momentum involved in the transfer (such as the mass and momentum of a real spring) can be neglected

Under these conditions we have a constitutive relation of the general form

$$\mathbf{F}_{ab} = f(|\mathbf{r}_a - \mathbf{r}_b|) \frac{\mathbf{r}_a - \mathbf{r}_b}{|\mathbf{r}_a - \mathbf{r}_b|}$$

where $f()$ can in principle be any real function.

We call this a **pairwise long-distance force**.

Among the most important examples of such a constitutive relation is **Newton's law of gravitation**:

$$\mathbf{F}_{ab} = -G \frac{m_a m_b}{|\mathbf{r}_a - \mathbf{r}_b|^2} \frac{\mathbf{r}_a - \mathbf{r}_b}{|\mathbf{r}_a - \mathbf{r}_b|} \quad (8.5)$$

Another example is the force of the so-called *Lennard-Jones potential*:

$$\mathbf{F}_{ab} = \frac{\epsilon}{|\mathbf{r}_a - \mathbf{r}_b|} \left[12 \left(\frac{\sigma}{|\mathbf{r}_a - \mathbf{r}_b|} \right)^{12} - 6 \left(\frac{\sigma}{|\mathbf{r}_a - \mathbf{r}_b|} \right)^6 \right] \frac{\mathbf{r}_a - \mathbf{r}_b}{|\mathbf{r}_a - \mathbf{r}_b|} \quad (8.6)$$

Which is often used to model the flux of momentum among molecules of some fluids and solids.

According to general relativity there cannot be an *instantaneous* transfer of momentum between two spatially separate control volumes. The momentum transfer modelled by pairwise forces actually occurs in a lapse of time, mediated by matter or electromagnetic field present between the two control volumes.

8.2.5 Gravity and momentum supply near a planet's surface

According to general relativity, there can be creation of momentum in a control volume, that is, there can be a momentum supply \mathbf{G} . This supply depends on two aspects: our choice of coordinate system, and the nearby presence of large amounts of energy-mass, momentum, and of their fluxes. General relativity also makes it clear that it doesn't make sense to distinguish between these two aspects in a small control volume: if we measure a supply of momentum \mathbf{G} in a very small region, it could be because of our coordinate system, or because of the nearby presence of energy-mass or momentum. In fact, by changing our coordinate system we can always make the supply to be zero *in a small region* – but in general not everywhere.

Examples of these kinds of supplies are the ‘gravitational force’, the ‘centrifugal force’, and other forces called ‘inertial’. For example, when we’re travelling in a car that speeds up or slows down, we feel a horizontal force pushing us against our seat or pulling us away from it. That force is a supply of momentum. We feel that supply because our “force-receptors”, the [mechanoreceptors](#)² in our skin and bones, *measure momentum flux with respect to a coordinate system that is at rest with respect to our body*, that is, a coordinate system where our body has constant coordinates, independent of time. A person on the street that sees us passing by would say that there’s no momentum supply (besides the vertical one discussed below), simply because that person is using a different coordinate system.

When we consider physical phenomena that happen close to a planet’s surface and involve spatial extensions that are small compared to the planet’s size, and we choose a coordinate system fixed with the planet’s ground, then we have special supply of momentum:



The supply of momentum appearing in a slowed-down car can be deadly real

Gravitational force near a planet’s surface

Take a coordinate system (t, x, y, z) fixed with the ground, with z the vertical coordinate with an upward positive direction. A control volume containing an amount of mass-energy m also has a supply of downward momentum, the **gravitational force**, given by

$$\mathbf{G} = m \mathbf{g}, \quad \text{with} \quad \mathbf{g} = [0 \text{ N/kg}, 0 \text{ N/kg}, -g], \quad (8.7)$$

where g , the *gravitational acceleration*, depends on the planet. On Earth it is approximately

$$g \approx 9.8 \text{ N/kg} \equiv 9.8 \text{ m/s}^2. \quad (8.8)$$

This supply of momentum is approximately constant in time.

8.3 Examples of applications

8.3.1 Statics

As previously mentioned there are situations in which we must study or find which momentum fluxes \mathbf{F} and supply \mathbf{G} can make the amount of momentum in a control volume to be zero at all time. This is the domain of the discipline of [statics](#)³. Let’s see a couple of concrete examples.

An object, such as a book, is resting on a table. Which momentum fluxes occur in such a situation?

Let's choose a coordinate system (see side picture) and a static control surface that wraps the object. The total amount of momentum in this control volume is zero and constant:

$$\mathbf{P}(t) = [0, 0, 0] \text{ N s} \text{ (constant).}$$

We also know that any control volume close to Earth's surface has a constant *supply* of momentum, proportional to the mass-energy it contains. Let's say that the supply in this case is

$$\mathbf{G}(t) = [0, 0, -2] \text{ N} \text{ (constant).}$$

Then *what can the momentum fluxes across different parts of the control surface be?*

A straightforward application of momentum balance in differential form tells us that the **net influx** of momentum must be

$$\begin{aligned} \mathbf{F}(t) &= \frac{d\mathbf{P}(t)}{dt} - \mathbf{G}(t) \\ &\quad [0,0,0] \text{ because constant} \quad [0,0,-2] \text{ N} \\ &= [0, 0, 2] \text{ N} \text{ (constant).} \end{aligned}$$

Note that this is the only piece of information that the balance of momentum, applied to the chosen closed control surface, can give us. The **influxes through different parts of the control surface could be very different** from this value, which is only their total; but the balance of momentum by itself cannot give us these partial fluxes. We need additional information, which can only come from constitutive relations.

↳ § 4.7 page 69

In the present case there's a constitutive relation about the momentum flux between a substance like air and the book, which says that such flux is approximately zero:

$$\mathbf{F}_{\text{air-book}}(t) \approx [0, 0, 0] \text{ N}.$$

The only remaining part of the control surface is the one between the book and the table. So the total influx $\mathbf{F}(t)$ that appears in the momentum balance must be the sum of these two:

$$\mathbf{F}(t) = \mathbf{F}_{\text{air-book}}(t) + \mathbf{F}_{\text{table-book}}(t).$$



From the last three equations we finally find

$$\mathbf{F}_{\text{table-book}}(t) \approx [0, 0, 2] \text{ N} \text{ (constant).}$$

In other words, there's a flux of *upward* momentum from the table to the book. This is the flux that compensates the *downward* momentum supply in the book, keeping the book's total momentum to zero. This kind of flux is called **normal force**, and we'll discuss it again later on.

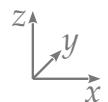
Exercise 8.2

Consider two identical books, on top of each other, resting on a table. Choose two static closed control surfaces of cuboid shape, each wrapping one book, and having one side in common (where the two books touch). Assume that the momentum supply in each control volume is $[0, 0, -2]$ N, constant in time, and make the same assumptions as before regarding the momentum flux across the parts of the surfaces in contact with air.

Use the balance of momentum with the two control surfaces to find:

1. the flux of momentum between the two books
2. the flux of momentum between the bottom book and the table

answer the questions above not just by using intuition, but by explaining step-by-step how you use the balance law and the given assumptions, in a mathematical deduction.



8.3.2 Atmospheric pressure

There is a continuous flux of momentum across any surface that has air on at least one of its sides. Not only any surface where air is in contact with objects, but also any imaginary surface, say in the sky over an open field, with air on both sides. This flux is called *atmospheric pressure*⁴ and has some important constitutive properties:

- It is always *compressive*, that is, its direction is always orthogonal to the surface, and its orientation is the same as the crossing orientation.
- Its magnitude, for a surface of 1 m^2 at sea level, is approximately equal to $1 \times 10^5 \text{ N}$ (corresponding to a weight of about 10 000 kg!).

 § 4.6.1 page 68

- Its magnitude slowly decreases with altitude; that is, a 1 m^2 surface 1 km above the ground will have an atmospheric pressure (momentum flux) lower than a 1 m^2 surface close to the ground.

In examining momentum fluxes for objects on Earth we often neglect the presence of air and atmospheric pressure. We did so in the previous example about a book on a table. The reason is that the *total* momentum flux from atmospheric pressure around an object is usually approximately zero. This happens on account of three factors: (a) there's always at least a thin layer of air around an object; even an object lying on a table or on the ground has a thin layer of air underneath; (b) the atmospheric-pressure momentum flux is always orthogonal to every small piece of surface; (c) its magnitude is approximately the same at the same altitude. Together, these factors lead to a zero total when we sum up all the atmospheric-pressure fluxes across the different parts of a surface that encloses an object.

Indeed we suddenly feel the heavy presence of atmospheric pressure when the first or the last of those factors doesn't hold anymore.

For example you may sometimes have made the mistake of laying a wide slab of glass, say from a window, upon another, only to find out that you couldn't easily lift it up anymore: the two slabs were sort of "stuck" together. The reason is that the glass surface is so smooth that effectively there's no space for a thin layer of air between the two glass slabs. So the total momentum flux from air to the upper slab does no longer sum up to zero: there's only a compressive flux of downward momentum from air to the slab across its upper surface. As mentioned, for a 1 m^2 slab this flux is equivalent to a weight of 10 000 kg! You can't lift the upper slab, not because it's "stuck" to the lower one, but because it suddenly has a net force of 10 tonnes on top (which presses the two glass slabs together even more, and therefore drives out more air between them). The only thing you can do is to try to let air again between the two slabs, for instance moving them to a vertical position, so that there will again be atmospheric pressure on both sides of each slab.

The same phenomenon occurs, in a more useful way, with a suction cup: the atmospheric pressure on its two sides is very different, and therefore it receives a net influx of momentum pointing towards the wall. This, in turn, leads to an influx of upward momentum owing to friction (which we'll discuss later), which compensates for the momentum supply from gravity. Thus the cup can stay in place without falling and even support some object.



In some cases the magnitude of the atmospheric pressure is different because of altitude, and this can be exploited for flying, as we discuss below. But in order to do that, try to solve this exercise first:

Exercise 8.3

In this exercise we apply a reasoning very similar to that of Exercise 8.2.

Consider a body of air at rest. Imagine a closed control surface which encloses some of this air. This is analogous to the [example with the book on the table](#); now instead of a book we simply have a volume of air, and instead of a table we have air also underneath the volume. For definiteness let's say that the volume is a cuboid with small height h and horizontal base area A , so the volume is $h A$

Air is matter, and therefore this control volume has a constant supply of downward momentum owing to gravity. On Earth's surface we can say (this is a constitutive relation) that a volume $h A$ of air has a momentum supply with z -component approximately equal to

$$G_z = -h \cdot A \cdot 11.8 \text{ N/m}^3 .$$

Assume that the momentum flux on the lateral surface of the volume is zero.

1. Apply the momentum balance, and extensivity of momentum, to find the **difference in magnitude** between the momentum influx through the bottom of the cuboid and the momentum influx through the top of the cuboid.
2. Pressure is defined as momentum flux *per unit area*. How much is the pressure difference between the bottom and top of the cuboid of air?
3. Try to generalize and write an approximate formula of how pressure changes with altitude.

! The results obtained in this exercise are approximate, because the supply of momentum \mathbf{G} actually changes with altitude as well, and depends on other quantities such as temperature. But they do have the correct order of magnitude.

 § 8.3.1 page 128

8.3.3 Airborne flight

Airborne flight is a conceptually extremely simple application of momentum balance. In terms of momentum, the objective of flight is to compensate or over-compensate the constant downward-momentum supply \mathbf{G} that an object, not touching any other solid object, receives because of gravity. One way to compensate this supply is by creating a *net* influx of upward-momentum, or equivalently a *net* efflux of downward-momentum, between the object and the air or atmosphere that surrounds it. This is airborne flight.

The ways in which such a momentum flux between object and air is realized can be very different:

Buoyancy In the preceding section we saw that the magnitude of the compressive momentum flux – atmospheric pressure – between air and any object *decreases* with height. So if the object is large enough compared to its weight, it automatically receives a net influx of upward momentum that can balance its weight. A calculation shows that this net upward force is proportional to the volume of the object, the mass per unit volume of the surrounding air (or atmosphere or other fluid), and the gravitational acceleration; this formula is called '[Archimedes's principle](#)'⁵. Flight by buoyancy can therefore be achieved by making the object large, light, or both. This is the principle upon which helium-filled party balloons, hot-air balloons, airships, and also submarines, are based.



Airship *Airlander* from [Hybrid Air Vehicles](#)⁶

Soaring Air is not always at rest (with respect to the Earth's surface). Owing to energy flow in the atmosphere and Earth's rotation, bodies of air – air currents – can be moving in different directions, including upward. An upward-moving body of air contains a net upward momentum. If an object stops the vertical motion of this air, for example deflecting it horizontally, then by the balance of momentum there must be a flux of upward momentum from the body of air to the object. This upward momentum can compensate the gravity supply of downward momentum, and therefore the object can float or rise. This mechanism is called [soaring or gliding](#)⁷ and is used by birds and gliders; it's also the mechanism that lifts a light object such as a piece of paper or a feather.



A hang-glider (image from [Jæren Luftsportsklubb](#)⁸)

Propelled flight We can try to create a flow of downward momentum from an object to the surrounding air, even if air is initially at rest. As a

result, a body of air surrounding the object will acquire a net, partially downward movement. By the [symmetry of fluxes](#) this also means that momentum with an upward component flows from the air to the object, and this component can compensate the object's weight. This is achieved by birds by flapping their wings, and by aeroplanes and helicopters through horizontally moving wings or blades.

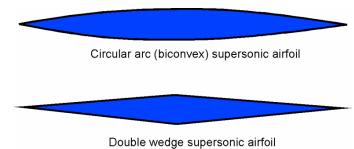
↳ § 4.3.1 page 57

It's not the difference in air pressure

Some texts say that wings can sustain an aeroplane because their cross-section has an asymmetric shape, slightly more bulged upward than downward, leading to a difference in the air pressure between the top and bottom of the wing. This is not true. In fact, aeroplanes with symmetric wings fly as well.

the popular theory of lift generation found in many textbooks is completely wrong! The upper surface doesn't have to be longer than the lower surface to generate lift. The lift occurs because the airfoil turns the flow of air and both the lower and upper surface contribute to the turning.

Wing geometry, NASA Glenn Research Center⁹



Sections of two kinds of wings for supersonic flight
(Leishman 2024 § 26)

8.3.4 Rockets

The flight mechanisms discussed in the previous section rely on the presence of air or some kind of atmosphere around the object, with which a momentum flux can occur. This is not possible in outer space, and such flux may not be enough to lift an object even if air is surrounding it.

Another way of producing an influx of upward momentum is to release matter having downward momentum. As we saw in an [example with a falling block of ice](#), if through a surface there's a flux of matter, then there's also a flux of momentum. So if an object manages to realize a strong enough flux of matter through a surface at its bottom, the object itself will experience an influx of upward momentum, which can overcome the object's weight. This is the momentum-flux mechanism employed by rockets.

↳ § 5.3.4 page 88

Obviously the amount of matter in the control volume defining the object will be decreasing, since there's also a negative influx of matter, and the law of conservation of matter must hold as well.

8.3.5 Statics again: cable cars

Consider a cable car suspended on a horizontal cable, and not moving. Which momentum fluxes occur in this situation?

This problem has some similarities with the previous one about the book on a table. Choose a control volume containing the cable car.

- The total momentum in this volume is constantly zero, because the cable car isn't moving
- There's a constant momentum supply because of gravity; let's say that in this case it's

$$\mathbf{G}(t) = [0, 0, -8 \times 10^4] \text{ N} \text{ (constant)}$$



*Scenic Skyway*¹⁰ cable car, Australia

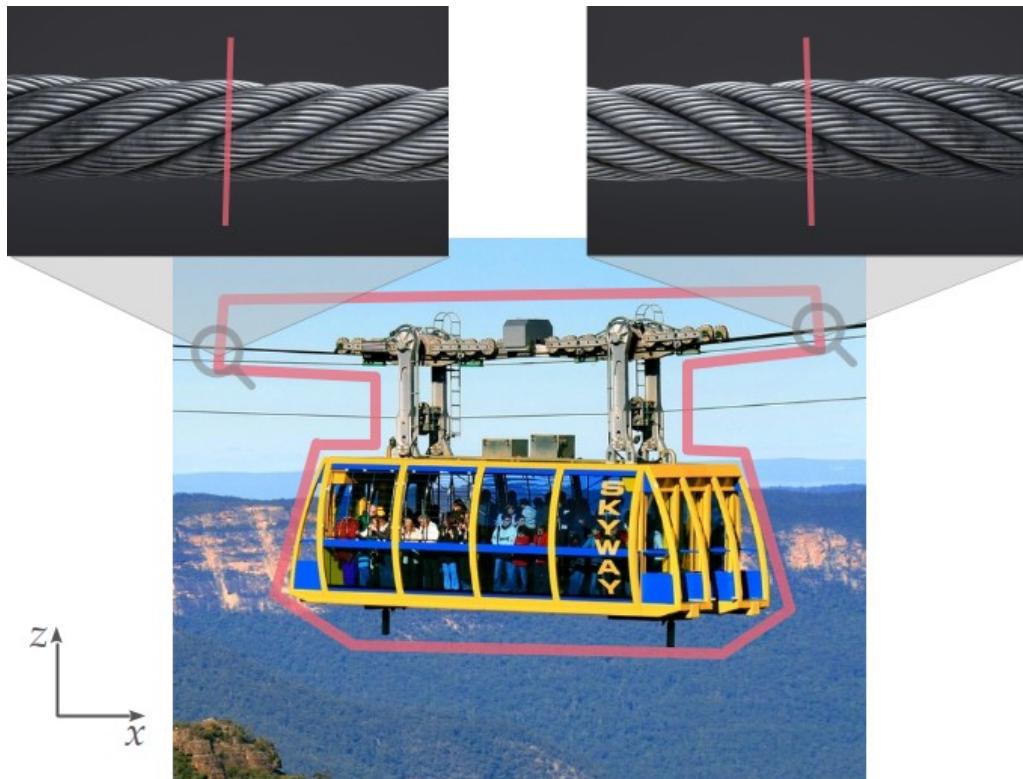
- By the balance of momentum, there must then be an influx of upward-momentum of the same magnitude as the supply:

$$\begin{aligned} \mathbf{F}(t) &= \frac{d\mathbf{P}(t)}{dt} - \mathbf{G}(t) \\ &\quad [0,0,0] \text{ because constant} \quad [0,0,-8 \times 10^4] \text{ N} \\ &= [0,0,8 \times 10^4] \text{ N} \text{ (constant)}. \end{aligned}$$

- Also in this case the influx through the parts of the control surface in contact with air is practically zero:

$$\mathbf{F}_{\text{air-car}}(t) \approx [0,0,0] \text{ N}.$$

All the influx of momentum must therefore come through the cable. But there are some interesting aspects on how this happens. In this case it's interesting to choose a static closed control surface that wraps the car and part of the cables, "cutting" the cables in imagination, as illustrated by the red polygon in the picture below:



This picture also depicts, as zoom-ins, the two parts of the imaginary control surface that “cut” the cable (for simplicity imagine that there’s only one cable). The total influx \mathbf{F} must happen through these two small regions of the surface, in the cable. An interesting question is: *How much is the momentum influx through each one?*. Call these two momentum influxes \mathbf{F}_{left} through surface cutting the cable on the left; and $\mathbf{F}_{\text{right}}$ through surface cutting the cable on the right. Since all these fluxes are constant in time, let’s drop the argument ‘(t)’; and let’s drop the y -components, which are all zero.

We know that

$$\mathbf{F}_{\text{left}} + \mathbf{F}_{\text{right}} = \mathbf{F} = \begin{bmatrix} 0 \\ 8 \times 10^4 \end{bmatrix} \text{ N}$$

from this equation we find

$$(\mathbf{F}_{\text{left}})_x = -(\mathbf{F}_{\text{right}})_x \quad (\mathbf{F}_{\text{left}})_z = 8 \times 10^4 \text{ N} - (\mathbf{F}_{\text{right}})_z$$

where ‘(\dots) $_x$ ’ denotes the x -component, and similarly for the z -component. But this doesn’t tell us how much \mathbf{F}_{left} and $\mathbf{F}_{\text{right}}$ are individually. For example we could have

$$\mathbf{F}_{\text{left}} \stackrel{?}{=} \begin{bmatrix} 5 \\ -2 \times 10^4 \end{bmatrix} \text{ N} \quad \mathbf{F}_{\text{right}} \stackrel{?}{=} \begin{bmatrix} -5 \\ 20 \times 10^4 \end{bmatrix} \text{ N}$$

and the total influx would be the correct one.

We shall see later that the *balance of angular momentum* applied to this problem leads to a further constraint: $(\mathbf{F}_{\text{left}})_z = (\mathbf{F}_{\text{right}})_z$. So we can conclude that

$$(\mathbf{F}_{\text{left}})_z = (\mathbf{F}_{\text{right}})_z = 4 \times 10^4 \text{ N}$$

that is, each half of the cable is “taking half of the weight”, as intuitively expected.

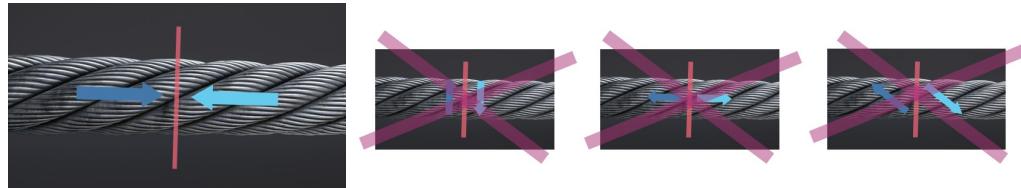
The last part of our mystery, about the x -components, can be solved thanks to an additional constitutive property:

Momentum flux allowed in cables and ropes

A cable, rope, or similar object can (approximately) only transmit **tensile momentum flux**, aligned along its axis.

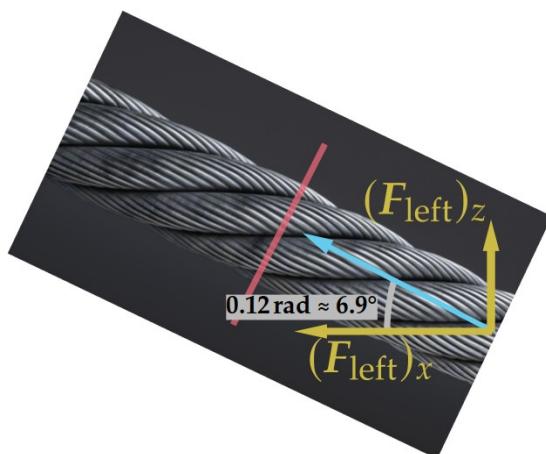
 § 4.6.2 page 68

Pictorially this means that only the momentum flux represented in the left picture below (tension) is physically possible:



the other three (from left to right: shear, pressure, mixed) are not.

We therefore need to know the directions of the axes of the two parts of the cable. Let's suppose that each part has an inclination of $0.12 \text{ rad} \approx 6.9^\circ$, but in opposite directions. This must then also be the angle between the influx \mathbf{F}_{left} and its horizontal component $(\mathbf{F}_{\text{left}})_x$, as in this picture (the angle has been exaggerated for clarity):



From trigonometry we find

$$(\mathbf{F}_{\text{left}})_x = -(\mathbf{F}_{\text{left}})_z / \tan(0.12 \text{ rad}) = -4 \times 10^4 \text{ N} / 0.12 \approx -3.3 \times 10^5 \text{ N}$$

and therefore $(\mathbf{F}_{\text{right}})_x \approx 3.3 \times 10^5 \text{ N}$:

$$\mathbf{F}_{\text{left}} \approx \begin{bmatrix} -3.3 \times 10^5 \\ 4 \times 10^4 \end{bmatrix} \text{ N} \quad \mathbf{F}_{\text{right}} \approx \begin{bmatrix} 3.3 \times 10^5 \\ 4 \times 10^4 \end{bmatrix} \text{ N}$$

The magnitude of these tensions is also approximately $|\mathbf{F}_{\text{left}}| = |\mathbf{F}_{\text{right}}| \approx 3 \times 10^5 \text{ N}$ or an equivalent weight of 35 tonnes (this result seems in the correct order of magnitude, comparing with the data in Brownjohn 1998).

In solving the cable-car problem above we used the balance of momentum, but note and keep in mind how that balance alone wasn't enough. We also had to use:

- the balance of angular momentum
- constitutive relations regarding:
 - the momentum supply
 - the amount of momentum flux between air and the cable car
 - the kind of momentum flux allowed in a cable

8.3.6 Hookean spring and harmonic oscillator

The physical system composed of two bodies of matter connected by a spring, with the spring modelled by the [Hookean constitutive properties](#), is one of particular importance in physics, for several reasons. Its equations can be solved analytically and describe a [harmonic oscillator](#)¹¹, with a behaviour that can be predicted without numerical time-integration methods. The *mathematical form* of its equations can also be applied to many other physical phenomena involving quantities other than momentum. Hooke's law can sometimes be used as a linear approximation in more complicated systems. And if we consider particular additional forces applied to one or both of the two bodies, we can study simple examples of the important phenomena of [resonance](#)¹² and [damping](#)¹³.

↳ § 8.2.2 page 124

The Hookean spring and the harmonic oscillator are therefore discussed in detail in essentially all physics textbooks and in specialized treatises. For this reason I won't discuss them here. I recommend that you read the [chapter on the harmonic oscillator](#)¹⁴ and [that on resonance](#)¹⁵ in Feynman's Lectures, for example.

In these notes, instead, I'd like to discuss more in detail the role of the balance of momentum and of constitutive relations in this system, as these aspects are often quickly glossed over in other texts.

Exercise 8.4

Before continuing, try to on your own the analysis of two small bodies connected by a Hookean spring, as previously defined. Consider in particular:

- How would you set up a coordinate system and closed control surfaces to describe this physical phenomenon? how many control volumes?
- How many applications of the balance of momentum would you need to make?
- What would be the volume contents, fluxes, and supplies in these balances?
- What would be the relevant constitutive relations?
- What would be the initial data and the [boundary conditions](#)?

↳ § 5.5.3 page 98

We consider two bodies of matter, call them a and b , of small extension compared to the distance between them. We attach them at the two ends of a Hookean spring, which for simplicity here we take as having a negligible natural length; that is, $l_n \approx 0$ m. We choose a coordinate system at rest with the Earth's surface, so that the only momentum supply, if considered, is given by the constitutive equation for gravitational force.

In terms of control surfaces, we have three of them: two tightly enveloping one body each, and a third enveloping the spring. Let us analyse each one in turn, writing down its relevant balances. Note that all quantities below, except the masses, depend on the time t .

Body a : The two control volume for body a is small and can be characterized simply by its position vector \mathbf{r}_a . The mass-energy in this control volume is m_a . [By construction, conservation of matter automatically holds for this control volume.](#)

↳ § 7.3.1 page 117

This control volume contains an amount of momentum $\mathbf{P}_a = m_a \mathbf{v}_a$, according to the [Newtonian constitutive equation](#).

↳ § 8.2.1 page 123

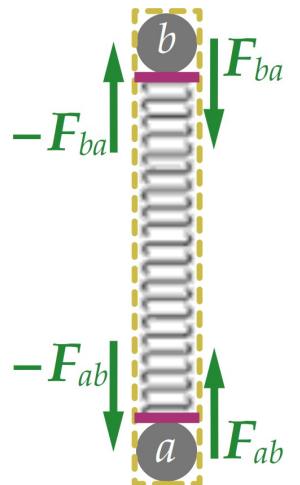
We assume that the only momentum flux across the closed control surface occurs on the small portion of surface in common with the control surface for the spring (see [thick red lines](#) in the side picture below). Across that small surface there is an **influx** \mathbf{F}_{ab} . We are using this notation for the fluxes: ' \mathbf{F}_{ab} ' means that it's the influx into body a , coming from the spring that connects it to body b .

If we considered gravity effects, this control volume would also have a constant momentum supply $\mathbf{G}_a = m_a \mathbf{g}$, according to the [constitutive equation for gravitational force](#). Let's omit this supply for simplicity.

↳ § 8.2.5 page 128

For this control volume we have therefore the following momentum balance and constitutive relations:

$$\begin{array}{ll} \text{momentum balance} & \frac{d\mathbf{P}_a(t)}{dt} = \mathbf{F}_{ab}(t) + \mathbf{G}_a(t) \\ \text{const. relation} & \mathbf{P}_a(t) = m_a \mathbf{v}_a(t) \\ \text{const. relation} & \mathbf{G}_a(t) = 0 \end{array}$$



Schematics of the analysis. The control volumes are indicated by dashed yellow lines. The surfaces in common between spring and the two bodies are depicted by two thick red lines. The fluxes $\mathbf{F}_{ab}, \mathbf{F}_{ba}$ happen across these surfaces, but their vectors are placed on the side in the illustration.

Body b: An analogous analysis can be made for the control volume of body b . It has position \mathbf{r}_b , total mass-energy m_b , total momentum $\mathbf{P}_b = m_b \mathbf{v}_b$.

The only momentum flux occurs across the portion of control surface in common with that of the spring; there we have an **influx** \mathbf{F}_{ba} . For simplicity we omit the momentum supply $\mathbf{G}_b = m_b \mathbf{g}$ from gravity.

For this control volume we have the following momentum balance and constitutive relations:

$$\begin{array}{ll} \text{momentum balance} & \frac{d\mathbf{P}_b(t)}{dt} = \mathbf{F}_{ba}(t) + \mathbf{G}_b(t) \\ \text{const. relation} & \mathbf{P}_b(t) = m_b \mathbf{v}_b(t) \\ \text{const. relation} & \mathbf{G}_b(t) = 0 \end{array}$$

Spring: The control volume for the spring, according to the simplifications typical of Hookean springs, has two negligible dimensions and can be represented by a long, narrow tube extending between the extremities at \mathbf{r}_a and \mathbf{r}_b . The two surfaces at these extremities are in common with two surfaces of the control volumes for the bodies a and b , one surface each. The amount of matter and matter flux in the control volume for the spring are considered to be zero, again according to the simplifications for Hookean springs. The same is true for the total momentum \mathbf{P} and supply \mathbf{G} , also zero.

At the two narrow surfaces in contact with bodies a and b (see side picture), the control surface has momentum **influxes** $-\mathbf{F}_{ab}$ and $-\mathbf{F}_{ba}$: the minus sign comes from the [symmetry of fluxes](#). The total momentum influx is therefore $\mathbf{F} = -\mathbf{F}_{ab} - \mathbf{F}_{ba}$ by extensivity.

↳ § 4.3.1 page 57

The Hookean constitutive relation (8.3) says that the momentum efflux across one surface, say the one in contact with body a , must be

$$\mathbf{F}_{ab} = -k(\mathbf{r}_a - \mathbf{r}_b)$$

because $\mathbf{r}_a - \mathbf{r}_b$ is the main length of the control volume for the spring.

The momentum balance and constitutive relations for the control volume of the spring are therefore:

momentum balance	$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}(t) + \mathbf{G}(t)$
total influx by extensivity	$\mathbf{F}(t) = -\mathbf{F}_{ab}(t) - \mathbf{F}_{ba}(t)$
Hooke const. relation	$\mathbf{P}(t) = 0$
Hooke const. relation	$\mathbf{G}(t) = 0$
Hooke const. relation	$\mathbf{F}_{ab}(t) = -k[\mathbf{r}_a(t) - \mathbf{r}_b(t)]$

The momentum balance for the spring reduces to a very simple equation: $0 = \mathbf{F} = -\mathbf{F}_{ab} - \mathbf{F}_{ba}$. We thus find that the effluxes at the extremities of the spring must be opposite:

$$\mathbf{F}_{ab} = -\mathbf{F}_{ba}$$

this result is therefore a consequence of momentum balance, *not* of the ‘principle of action and reaction’.

The setup for the description of this physical system is thus complete. This setup was probably intuitively clear, but I’d like you to stop for a moment and note all the subtle steps and details that it involves. We often *cannot* reason intuitively in the analysis of more complex physical systems, and need to spell out their description step by step, to avoid neglecting important details. It is therefore a good exercise to do this kind of analysis for a simpler system, as we just did.

In this analysis, note how some control volumes have parts of their control surfaces in common. These common surfaces are important because they connect the fluxes of momentum and other quantities between the various control volumes. In modelling extended solids and fluids we often use a *grid* of control volumes, so that the surface of one control volume has regions in common with the surrounding control volumes.

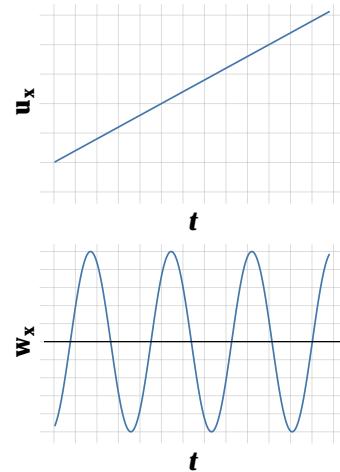
The set of equations above is usually solved analytically by introducing two alternative vector variables:

$$\mathbf{u} := \frac{m_a \mathbf{r}_a + m_b \mathbf{r}_b}{m_a + m_b} \quad \mathbf{w} := \mathbf{r}_a - \mathbf{r}_b$$

the first, \mathbf{u} , is called the *centre of mass* of this physical system. In terms of these variables the whole set of equations above reduces to these two linear differential equations with constant coefficients (each equation is a set of three, one for each component):

$$\frac{d^2\mathbf{u}(t)}{dt^2} = 0 \quad \frac{d^2\mathbf{w}(t)}{dt^2} + k \frac{m_a + m_b}{m_a m_b} \mathbf{w}(t) = 0$$

that can be solved analytically.



Every component of $\mathbf{u}(t)$ is a linear function of time, and every component of $\mathbf{w}(t)$ has a harmonic dependence on time with period $\frac{1}{2\pi} \sqrt{\frac{m_a m_b}{k(m_a+m_b)}}$. The slopes and intercepts of $\mathbf{u}(t)$ and the amplitudes and phases of $\mathbf{w}(t)$ depend on the initial conditions.

Exercise 8.5

Try to refresh your knowledge of linear differential equations with constant coefficients, and find the general solution for the equations above.

Then express this solution in terms of the positions \mathbf{r}_a and \mathbf{r}_b .

8.3.7 Hookean spring: numerical time integration

Even though the physical system of two masses connected by a Hookean spring can be solved analytically, it is interesting to examine what its numerical time-integration scheme is; also because such a scheme can be readily generalized to constitutive relations for [non-Hookean springs](#), for which analytical solutions may not exist.

For such a scheme we need to [step forward in time the momentum](#) of both masses using their momentum balances. The momentum balance of the spring is reduced to the equation $\mathbf{F}_{ba} = -\mathbf{F}_{ab}$, as discussed in the previous section.

[§ 8.2.3 page 126](#)

[§ 5.5.1 page 97](#)

The basic time-stepping scheme looks as follows:

$$\begin{aligned}
 \mathbf{F}_{ab}(t) &= -k [\mathbf{r}_a(t) - \mathbf{r}_b(t)] \\
 \mathbf{F}_{ba}(t) &= -\mathbf{F}_{ab}(t) \\
 \mathbf{P}_a(t + \Delta t) &\approx \mathbf{P}_a(t) + [\mathbf{F}_{ab}(t) + \mathbf{G}_a(t)] \Delta t \\
 \mathbf{P}_b(t + \Delta t) &\approx \mathbf{P}_b(t) + [\mathbf{F}_{ba}(t) + \mathbf{G}_b(t)] \Delta t \\
 \mathbf{r}_a(t + \Delta t) &\approx \mathbf{r}_a(t) + \mathbf{v}_a(t) \Delta t \\
 \mathbf{r}_b(t + \Delta t) &\approx \mathbf{r}_b(t) + \mathbf{v}_b(t) \Delta t \\
 \mathbf{v}_a(t + \Delta t) &= \mathbf{P}_a(t + \Delta t) / m_a \\
 \mathbf{v}_b(t + \Delta t) &= \mathbf{P}_b(t + \Delta t) / m_b
 \end{aligned} \tag{8.9}$$

Exercise 8.6

We consider a system like the one discussed in the previous two sections: two small objects attached to a Hookean spring, in the simple case of **two dimensions**, so that all vectors have two components only, say the y and z -components. We also assume that the natural length l_n of the spring is zero.

1. Write a script, in your preferred programming language, that implements a numerical time integration scheme according to the time-stepping scheme (8.9) above, from an initial time t_0 to a final time t_1 with timestep Δt .

As a starting point you can use the [script tennisball.m](#).

↳ §5.5.7 page 103

Your script should save some or all of the values of $t, \mathbf{r}_a, \mathbf{r}_b, \mathbf{P}_a, \mathbf{P}_b$ for subsequent analysis. It should also let you specify the quantities m_a, m_b, k , as well as the initial values of $\mathbf{r}_a, \mathbf{r}_b, \mathbf{v}_a, \mathbf{v}_b$, as parameters (that is, no specific numerical values for them are hard-coded in the script).

2. Run the script with the following values:

$$\begin{aligned}
 t_0 &= 0 \text{ s} & t_1 &= 10 \text{ s} & \Delta t &= 0.001 \text{ s} \\
 m_a &= m_b = 2 \text{ kg} & k &= 5 \text{ N/m} \\
 \mathbf{r}_a(t_0) &= [-3, 0] \text{ m} & \mathbf{r}_b(t_0) &= [3, 0] \text{ m} \\
 \mathbf{v}_a(t_0) &= [0, 0] \text{ m/s} & \mathbf{v}_b(t_0) &= [0, 0] \text{ m/s}
 \end{aligned}$$

Plot the y -coordinate of body a against time t . What kind of time dependence do you observe? can you explain it intuitively?

Now plot the trajectories of the two bodies, that is, z_a against y_a , and z_b against y_b . What do you observe? can you explain it intuitively?

Plot, against time t , the y - and z -components of the *total* momentum $\mathbf{P}_a + \mathbf{P}_b$ for the system composed by the two bodies and the spring. How do these component change? Why?

3. Run the script with the same parameter as before but the following initial velocity values:

$$\mathbf{v}_a(t_0) = [1, 1] \text{ m/s} \quad \mathbf{v}_b(t_0) = [-1, -1] \text{ m/s}$$

Plot again the y -coordinate of body a against time t . Is the time dependence different from the previous simulation? How do the trajectories of the two bodies look like this time?

Plot again the components of the total momentum against time. How do they change? Why have they the same time dependence as before?

4. Run the script with the following initial velocity values:

$$\mathbf{v}_a(t_0) = [2, 2] \text{ m/s} \quad \mathbf{v}_b(t_0) = [1, 1] \text{ m/s}$$

Plot the trajectories of the two bodies. What do you observe this time?

How do the components of the total momentum differ from the previous simulation? Try to explain why.

5. Play with all the parameter values and initial conditions, and see what happen. Before simulating, try to intuitively predict what the behaviour of the system will be.

8.3.8 Non-Hookean spring: numerical time integration

The system of two masses connected by a *non-Hookean spring* with constitutive relation (8.4) is analytically challenging. Its numerical solution, however, only involves a change of a couple of lines to the code that you wrote for Exercise 8.6. A small mathematical change in a constitutive

◀ § 8.2.3 page 126

relation can lead to a much richer set of behaviours, as the examples in the side plots below show.

Exercise 8.7

Modify the script you wrote for Exercise 8.6 so that the spring is modelled by the non-Hookean constitutive relation (8.4) (which can be implemented with an `if`-statement), instead of the Hookean one.

Then numerically time-integrate the system with the following four different sets of parameters and initial values. Find out which of these correspond to the trajectories shown in the side plots:

Set 1:

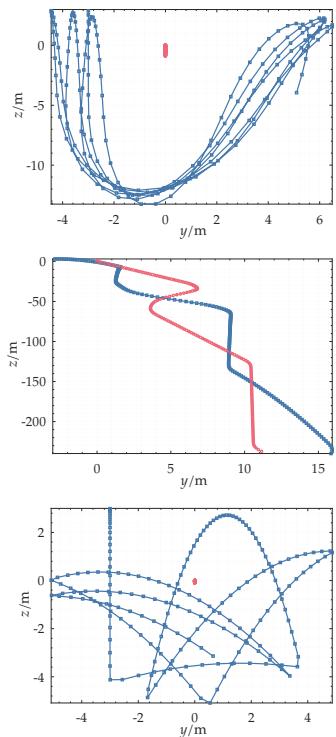
$$\begin{aligned} t_0 &= 0 \text{ s} & t_1 &= 10 \text{ s} & \Delta t &= 0.001 \text{ s} \\ m_a &= 1 \text{ kg} & m_b &= 1 \text{ kg} & k &= 5 \text{ N/m} & l_n &= 5 \text{ m} \\ \mathbf{r}_a(t_0) &= [-3, 3] \text{ m} & \mathbf{r}_b(t_0) &= [0, 0] \text{ m} \\ \mathbf{v}_a(t_0) &= [3, 1] \text{ m/s} & \mathbf{v}_b(t_0) &= [0, 0] \text{ m/s} \\ \mathbf{G}_a(t_0) &= m_a \cdot [0, -9.8] \text{ N/kg} & \mathbf{G}_b(t_0) &= [0, 0] \text{ N} \end{aligned}$$

Set 2:

$$\begin{aligned} t_0 &= 0 \text{ s} & t_1 &= 10 \text{ s} & \Delta t &= 0.001 \text{ s} \\ m_a &= 2 \text{ kg} & m_b &= 2 \text{ kg} & k &= 5 \text{ N/m} & l_n &= 5 \text{ m} \\ \mathbf{r}_a(t_0) &= [-3, 0] \text{ m} & \mathbf{r}_b(t_0) &= [3, 0] \text{ m} \\ \mathbf{v}_a(t_0) &= [0, 0] \text{ m/s} & \mathbf{v}_b(t_0) &= [0, 0] \text{ m/s} \\ \mathbf{G}_a(t_0) &= [0, 0] \text{ N} & \mathbf{G}_b(t_0) &= [0, 0] \text{ N} \end{aligned}$$

Set 3:

$$\begin{aligned} t_0 &= 0 \text{ s} & t_1 &= 10 \text{ s} & \Delta t &= 0.001 \text{ s} \\ m_a &= 1 \text{ kg} & m_b &= 5000 \text{ kg} & k &= 5000 \text{ N/m} & l_n &= 5 \text{ m} \\ \mathbf{r}_a(t_0) &= [-3, 3] \text{ m} & \mathbf{r}_b(t_0) &= [0, 0] \text{ m} \\ \mathbf{v}_a(t_0) &= [0, 0] \text{ m/s} & \mathbf{v}_b(t_0) &= [0, 0] \text{ m/s} \\ \mathbf{G}_a(t_0) &= m_a \cdot [0, -9.8] \text{ N/kg} & \mathbf{G}_b(t_0) &= [0, 0] \text{ N} \end{aligned}$$



Examples of trajectories of two masses connected by a non-Hookean spring (), for different values of parameters and initial conditions.

Set 4:

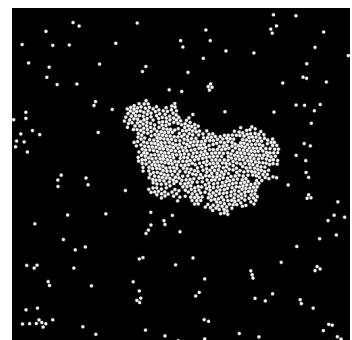
$$\begin{aligned}
 t_0 &= 0 \text{ s} & t_1 &= 30 \text{ s} & \Delta t &= 0.001 \text{ s} \\
 m_a &= 1 \text{ kg} & m_b &= 5000 \text{ kg} & k &= 5 \text{ N/m} & l_n &= 5 \text{ m} \\
 \mathbf{r}_a(t_0) &= [-3, 3] \text{ m} & \mathbf{r}_b(t_0) &= [0, 0] \text{ m} \\
 \mathbf{v}_a(t_0) &= [0, 0] \text{ m/s} & \mathbf{v}_b(t_0) &= [0, 0] \text{ m/s} \\
 \mathbf{G}_a(t_0) &= m_a \cdot [0, -9.8] \text{ N/kg} & \mathbf{G}_b(t_0) &= [0, 0] \text{ N}
 \end{aligned}$$

Play with other parameters and initial values!

8.3.9 Many-body systems

We can obviously consider more than two objects connected by a number of springs or other pairwise forces. The analytical description of such a system becomes quickly intractable. For numerical time integration, however, we only need to add additional lines of code to compute the additional momentum fluxes and timestep the momentum for the individual bodies; this can be done by appropriate for-loops for the pairwise forces and for the momenta.

↳ § 8.2.4 page 126



Snapshot of a molecular-dynamics simulation from [Interactive Molecular Dynamics](#)¹⁶

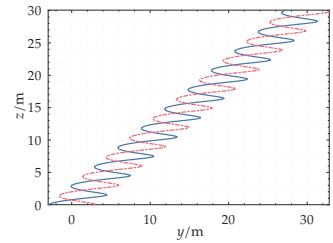
8.4 Choice of control surfaces and volumes

In the preceding examples with the two bodies & spring, we chose to describe the system by means of three closed control surfaces. We did so because we were interested in the detailed motion of the two bodies with respect to each other, and we were *not* interested in what was happening *within* each body.

In a real physical realization, for instance with two tennis balls connected by a spring, each tennis ball would slightly deform upon receiving momentum flux from the spring. If we were interested in such deformations, we would need to describe each tennis ball by a set of control surfaces and volumes, small enough so as to keep a detailed track of the different momentum contents and momentum fluxes within the ball. This detailed

description would also require knowledge of constitutive relations for the ball's material. Analogously for the spring, if we wished to describe it as a real spring made of different, deforming parts.

Yet in other situations we may instead not even be interested in the relative motion of the tennis balls. For instance, in Exercise 8.6 p. 143, with the Hookean spring, one set of initial values leads to the trajectories shown in the plot on the side. The distance between the two bodies cyclically changes between zero and a maximum value. If we imagine to zoom out from this plot, the positions of the two masses become almost indistinguishable, that is, $\mathbf{r}_a \approx \mathbf{r}_b$; and we see that the system as a whole is essentially moving on a straight line. A zoomed-out plot of $\mathbf{r}_a(t)$ or $\mathbf{r}_b(t)$ against time t would also show that its velocity is essentially constant.



In such a situation we would have chosen just *one* closed control surface containing the two bodies and the spring. By the [extensivity property](#), the total mass-energy m_{tot} , momentum content \mathbf{P}_{tot} , momentum flux \mathbf{F}_{tot} , and momentum supply \mathbf{G}_{tot} for this closed control surface can be obtained by adding up those of the three original control surfaces, spring+ $a+b$:

$$\begin{aligned} m_{\text{tot}} &= 0 + m_a + m_b & \mathbf{P}_{\text{tot}} &= 0 + \mathbf{P}_a + \mathbf{P}_b \\ \mathbf{F}_{\text{tot}} &= (-\mathbf{F}_{ab} - \mathbf{F}_{ba}) + \mathbf{F}_{ab} + \mathbf{F}_{ba} & \mathbf{G}_{\text{tot}} &= 0 + \mathbf{G}_a + \mathbf{G}_b \end{aligned}$$

↳ §3.1 page 31



Note that the total momentum influx is zero: $\mathbf{F}_{\text{tot}} = 0$. Indeed the only non-zero momentum fluxes occur at the surfaces between each body and the spring – but in the present description these two surfaces are not considered (just like the momentum fluxes that in reality occur *within* each body were not considered when we chose three control volumes).

Momentum balance also applies to the total control volume:

$$\mathbf{P}_{\text{tot}}(t_1) = \mathbf{P}_{\text{tot}}(t_0) + \int_{t_0}^{t_1} \mathbf{F}_{\text{tot}}(t) dt + \int_{t_0}^{t_1} \mathbf{G}_{\text{tot}}(t) dt$$

If the total momentum influx and supplies are zero, then we find $\mathbf{P}_{\text{tot}}(t_1) = \mathbf{P}_{\text{tot}}(t_0)$, which explains why we found that the velocity for the whole system could be considered as constant.

↳ §5.5.7 page 103

8.5 Example script for non-Hookean spring

Here is an example script that is a solution for Exercise 8.7 p. 145. It is a generalization of the [tennisball.m](#) script. Blue lines are strictly related

to numerical time integration; grey lines take care of saving and plotting data.

```

1 %% SI units used throughout
2 %% Simulation of two bodies connected by non-Hookean spring in 2D
3 %% Coordinates (y,z)
4 ma = 1; % mass of object a
5 mb = 1; % mass of object b
6 ya0 = -1; za0 = 1; % initial position of object a
7 yb0 = 0; zb0 = 0; % initial position of object a
8 vya0 = 0; vza0 = 0; % initial velocity of object a
9 vyb0 = 0; vzb0 = 0; % initial velocity of object b
10 Gya = 0; Gza = -9.8*ma; % gravity supply on object a
11 Gyb = 0; Gzb = 0; % gravity supply on object b
12 k = 4; % spring constant
13 ln = 5; % natural length of rubber band
14 t0 = 0; % initial time
15 t1 = 10; % final time, can also be earlier than initial
16 dt = 0.001; % time step, negative if backward time integration
17 %% adjust final time if not multiple of timestep
18 t1 = t1 + mod(t1-t0,dt);
19 %% check if timestep and time interval are consistent
20 if (t1-t0)*dt <= 0
    error('time interval inconsistent with timestep');
21 end
22 %% Save values of all quantities at some steps during the simulation,
23 %% for subsequent analysis or plotting
24 %% (saving at all timesteps could be too costly)
25 Nsaves = 200; % number of timepoints to save during the simulation
26 %% Calculate time interval for saving
27 dsave = (t1-t0)/(Nsaves-1);
28 if abs(dsave) < abs(dt)
29     error('time interval between saves is smaller than timestep')
30 end
31 %% Initialize vectors to contain saved values
32 tSave = nan(Nsaves,1); % time
33 yaSave = nan(Nsaves,1); zaSave = nan(Nsaves,1);
34 ybSave = nan(Nsaves,1); zbSave = nan(Nsaves,1);
35 PyaSave = nan(Nsaves,1); PzaSave = nan(Nsaves,1);
36 PybSave = nan(Nsaves,1); PzbSave = nan(Nsaves,1);
37 %% Save initial values
38 i = 1; % index that keeps count of savepoints
39 tSave(1) = t0;
40 yaSave(1) = ya0; zaSave(1) = za0;
41 ybSave(1) = yb0; zbSave(1) = zb0;
42 PyaSave(1) = ma*vya0; PzaSave(1) = ma*vza0;
43 PybSave(1) = mb*vyb0; PzbSave(1) = mb*vzb0;
```

[Download rubberband2D.m¹⁹](#)

```

45 %% Initialize plot
46 cols = get(0, 'DefaultAxesColorOrder');
47 plot(yaSave(1), zaSave(1), 's','Color',cols(1,:)); axis('tight');
48 plot(ybSave(1), zbSave(1), 'o','Color',cols(1,:)); axis('tight');
49 xlabel('{\it y}/m'); ylabel('{\it z}/m'); hold on;
50 %%
51 %% Numerical time integration
52 %% Initialize
53 t = t0;
54 ya = ya0; za = za0;
55 yb = yb0; zb = zb0;
56 vya = vya0; vza = vza0;
57 vyb = vyb0; vzb = vzb0;
58 Pya = ma*vya0; Pza = ma*vza0;
59 Pyb = mb*vyb0; Pzb = mb*vzb0;
60 %% loop
61 while sign(dt)*t < sign(dt)*t1 % possible backward time integration
62   %% update time
63   t = t + dt;
64   %% non-Hookean constitutive relation
65   l = norm([ya-yb, za-zb]); % present length
66   if l < ln
67     Fyab = 0;
68     Fzab = 0;
69   else
70     Fyab = -k*(ya-yb)*(l-ln)/l;
71     Fzab = -k*(za-zb)*(l-ln)/l;
72   end
73   %% update momentum
74   Fyba = -Fyab;
75   Fzba = -Fzab;
76   Pya = Pya + (Fyab + Gya)*dt;
77   Pza = Pza + (Fzab + Gza)*dt;
78   Pyb = Pyb + (Fyba + Gyb)*dt;
79   Pzb = Pzb + (Fzba + Gzb)*dt;
80   %% update velocity
81   vya = Pya/ma;
82   vza = Pza/ma;
83   vyb = Pyb(mb;
84   vzb = Pzb(mb;
85   %% update position
86   ya = ya + vya*dt;
87   za = za + vza*dt;
88   yb = yb + vyb*dt;
89   zb = zb + vzb*dt;
90   %% Check whether to save & plot at this step
91   if min(abs([0 dsave] - mod(t-t0, dsave))) <= abs(dt)/2

```

```
92 i = i+1;
93 tSave(i) = t;
94 yaSave(i) = ya; zaSave(i) = za;
95 ybSave(i) = yb; zbSave(i) = zb;
96 PyaSave(i) = Pya; PzaSave(i) = Pza;
97 PybSave(i) = Pyb; PzbSave(i) = Pzb;
98 plot(ya, za, 's','Color',cols(1,:));
99 plot(yb, zb, 'o','Color',cols(2,:));
100 pause(0.001);
101 end
102 end
103 %% Plot trajectory
104 plot(yaSave,zaSave,'-','Color',cols(1,:));
105 plot(ybSave,zbSave,'-','Color',cols(2,:));
```

URLs for chapter 8

1. <https://whc.unesco.org/en/list/59>
2. <https://openbooks.lib.msu.edu/neuroscience/chapter/touch-the-skin/>
3. <https://www.britannica.com/science/statics>
4. <https://www.britannica.com/science/atmospheric-pressure>
5. <https://www.britannica.com/science/Archimedes-principle>
6. <https://www.hybridairvehicles.com/>
7. <https://www.britannica.com/topic/locomotion/Gravitational-gliding#ref497010>
8. <https://www.jaerenluftsport.no>
9. <https://www1.grc.nasa.gov/beginners-guide-to-aeronautics/wing-geometry/>
10. <https://www.scenicworld.com.au/experience/scenic-skyway>
11. <https://www.britannica.com/science/mechanics/Simple-harmonic-oscillation#ref612067>
12. <https://www.britannica.com/science/resonance-vibration>
13. <https://www.britannica.com/science/damping>
14. https://www.feynmanlectures.caltech.edu/I_21.html
15. https://www.feynmanlectures.caltech.edu/I_23.html
16. <https://physics.weber.edu/schroeder/md/>
17. <https://doi.org/10.1351/goldbook.MT06969>
18. <https://physics.weber.edu/schroeder/md/>
19. <https://pglpm.github.io/7wonders/code/rubberband2D.m>

Chapter 9

Balance of energy

I turned the page. The answer was, for the wind-up toy, "Energy makes it go." And for the boy on the bicycle, "Energy makes it go." For everything, "Energy makes it go."

Now that doesn't *mean* anything. Suppose it's "Wakalixes." That's the general principle: "Wakalixes makes it go." There's no knowledge coming in. The child doesn't learn anything; it's just a *word!* [. . .]

It's also not even true that "energy makes it go," because if it stops, you could say, "energy makes it stop" just as well.

R. P. Feynman 1989

9.1 Formulation and generalities

Balance of energy

Volume content: E Flux: Φ Supply: R

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t) dt + \int_{t_0}^{t_1} R(t) dt$$

integral form

The balance of energy is extremely important in physical phenomena that underlie many modern (post-industrial revolution¹) technologies. It often must be explicitly accounted for together with the balance of momentum, and often it's the main governing balance when the balance

of momentum can be neglected. It's the balance that we use when we put thick clothes on in order to keep warm, or when we watch a video or do computations on a laptop.

In discussing the [balance of momentum](#) we saw that there were just a few constitutive relations for its volume content, and a plethora of constitutive relations for its flux. In the case of energy there is a great variety of constitutive relations that connect both its volume content E and its flux Φ to many other quantities: matter, momentum, angular momentum, and [auxiliary quantities](#) such as metric and temperature.

↳ § 8.1 page 122

The balance of energy is therefore very important for numerical time integration. Recall that balance laws allow us to predict *volume contents* at a later time, but not fluxes or supplies. Fluxes and supplies must [either be given as boundary conditions, or predicted by constitutive relations](#) that connect them to the volume contents of other quantities. The balance and constitutive equations for energy often have this specific "gluing" role.

↳ § 3.9 page 48

↳ § 5.5.3 page 98

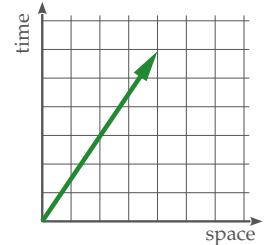
9.1.1 Forms of energy

In physics textbooks and literature we find many different energies and energy fluxes and many different energy balances. These texts often discuss sequences of mathematical steps that show how we can get from one form of energy balance to another.

Relativity theory gives a simple explanation of such a variety: *the definition of energy depends on the choice of a reference velocity and a reference clock*. This choice of velocity & clock can moreover be different at each spacetime point. It is this variety of choices that leads to the variety of energies.

Keep in mind that the choice of velocity & clock for the definition of energy is *independent from the choice of a coordinate system*. That is, once we have chosen coordinates (t, x, y, z) , we can still choose arbitrary reference velocities and reference times.

Given a coordinate system (t, x, y, z) , three possible choices are usually made when electromagnetic fields are negligible:



A reference 'velocity & clock' is simply a vector in four-dimensional spacetime

! Some texts say that different definitions of energy correspond to different choices of coordinate systems, but that is not correct.

- At each point in space and time we can choose a zero coordinate velocity – that is, we don't move with respect to the coordinates (x, y, z) – and we can measure time according to coordinate time t . We call **total energy** or **coordinate energy** the energy defined with respect to this velocity & clock, and denote its volume content by E .

☰ If matter is present, we can choose the matter's velocity, and measure time according to the proper time of a clock moving together with that matter. We call the energy so defined the **internal energy** of matter, and denote its volume content by U . Note that this energy cannot be defined if matter is not present. This definition of energy is often adopted in the description of fluids.

⌚🌐 We can choose a zero coordinate velocity, as in the first case, but measure time according to *proper time* rather than coordinate time t . Energy defined with respect to this reference velocity & clock is usually different from the previous two definitions. The difference between this energy and internal energy is called *kinetic energy*; we denote its volume content by E_k . The difference between total energy and this energy is called *potential gravitational energy*; we denote its volume content by E_p .

↳ § 2.1 page 19

We thus have the relation

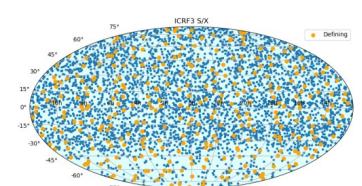
$$E = U + E_k + E_p \quad (9.2)$$

Each of these differently defined energies satisfies a balance law with different fluxes and supplies. An interesting property of *internal energy* is that the gravitational field is completely absent in its volume content, flux, and supply. Its balance is usually called the **first law of thermodynamics**. Otherwise, what appears as volume content in one definition also typically appears, slightly disguised, as flux or supply in another definition.

"a distinction is made between the principle of conservation of energy and the first law of thermodynamics, the two being identical only when matter is at rest" Eckart 1940

Note that we do not need to use all of these differently defined energies; one is enough. One definition can be mathematically more convenient to use than another, depending on the application or problem. In these notes we shall simply use the *total energy*, because of the following interesting or convenient features:

- It is defined anywhere, even where no matter is present
- Its volume content E and flux Φ typically contain terms that refer to the motion of matter and on the gravitational field
- Its supply is practically zero with coordinate systems and coordinate time typically used for physical phenomena on Earth ([International Terrestrial Reference System, ITRS³](#)) or in the solar system ([International Celestial Reference System, ICRS⁴](#)). In other words, it practically satisfies a *conservation law* in these common coordinate systems



Map of some distant astronomical objects used to define the International Celestial Reference Frame (from [The ICRF²](#))

9.1.2 Is energy conserved?

No, according to our present knowledge. First of all, recall that many texts use the term ‘conservation’ in the sense of ‘balance’ as defined in the present notes. So when they speak of “energy conservation” they don’t exclude the presence of an energy supply, the creation of energy within a volume.

↳ § 5.1.1 page 77

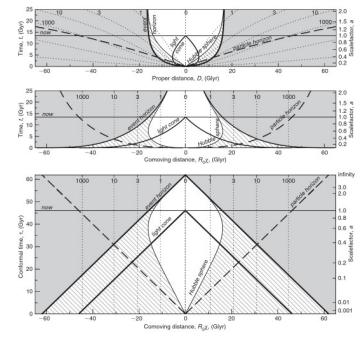
In fluid mechanics, for instance, we usually describe physical phenomena using the internal-energy definition, and as mentioned above this energy definition satisfies a balance law, not a conservation law.

But doesn’t the energy that appears as “supply” come from somewhere else, so it also disappears from somewhere else and its total is the same? Also in this case the answer is no. As far as we know, there is constant creation of energy on cosmological scales, no matter how we choose the reference velocity & clock for its definition. The total energy of the universe (provided such a notion can be mathematically defined) is therefore not constant. The reason is that the metric of the universe is not constant along the time direction – the so-called [expansion of the universe⁵](#) – and an energy that is strictly conserved can only be defined if the metric is constant.

Something analogous is true for momentum and angular momentum: they are strictly conserved, that is, their supplies are zero, only if the spacetime metric remains the same in different spatial directions and under rotations. We have therefore this interesting difference:

- On planetary scales the spacetime metric is not quite constant in space but is almost constant in time; so momentum and angular momentum are not conserved, but energy approximately is.
- On cosmological scales the spacetime metric is approximately constant in space but not in time; so momentum and angular momentum are approximately conserved, but energy is not.

The present discussion about energy conservation, however, is not meant to suggest that we could find ways for our cars or laptop to magically operate by themselves. First, the amount of energy created in our solar system and even in our galaxy is negligible. For devices operating in the solar system, any increase in energy content ultimately comes as an energy influx from the Sun. Second, any energy supply is determined by particular constitutive relations. These relations involve influxes of matter and of momentum, so they don’t allow for the creation of energy in an



Depiction of change of spacetime metric on cosmological scales, in three different coordinate systems (from Davis & Lineweaver 2004)

isolated system. This is also the reason why we can't have any "magical" production of internal energy, which as explained above *does* have a supply, ordinarily taken care of in fluid mechanics.

URLs for chapter 9

1. <https://www.britannica.com/event/Industrial-Revolution>
2. <https://hpiers.obspm.fr/icrs-pc/newww/icrf/index.php>
3. <https://itrf.ign.fr/en/background>
4. https://aa.usno.navy.mil/faq/ICRS_doc
5. <https://www.jpl.nasa.gov/edu/news/2023/7/24/exploring-the-mystery-of-our-expanding-universe/>

Bibliography

(“de X” is listed under D, “van X” under V, and so on, regardless of national conventions.)

- Anderson, E. K., Baker, C. J., Bertsche, W., Bhatt, N. M., Bonomi, G., Capra, A., Carli, I., Cesar, C. L., et al. (2023): *Observation of the effect of gravity on the motion of antimatter*. Nature **621**⁷⁹⁸⁰, 716–722. doi:[10.1038/s41586-023-06527-1](https://doi.org/10.1038/s41586-023-06527-1).
- Ashby, N. (2002): *Relativity and the global positioning system*. Phys. Today **55**⁵, 41–47. doi:[10.1063/1.1485583](https://doi.org/10.1063/1.1485583).
- (2003): *Relativity in the global positioning system*. Living Rev. Relativity **6**, 1–42. doi:[10.12942/lrr-2003-1](https://doi.org/10.12942/lrr-2003-1).
- Astarita, G. (1990): *Thermodynamics: An Advanced Textbook for Chemical Engineers*, 2nd pr. (Springer, New York). doi:[10.1007/978-1-4899-0771-4](https://doi.org/10.1007/978-1-4899-0771-4). First publ. 1989.
- Bossavit, A. (1991): *Differential Geometry: for the student of numerical methods in electromagnetism*. https://www.researchgate.net/publication/200018385_Differential_Geometry_for_the_student_of_numerical_methods_in_Electromagnetism.
- Brownjohn, J. M. W. (1998): *Dynamics of an aerial cableway system*. Eng. Struct. **20**⁹, 826–836. doi:[10.1016/S0141-0296\(97\)00113-2](https://doi.org/10.1016/S0141-0296(97)00113-2).
- Burke, W. L. (1987): *Applied Differential Geometry*, repr. (Cambridge University Press, Cambridge). doi:[10.1017/CBO9781139171786](https://doi.org/10.1017/CBO9781139171786). First publ. 1985.
- (1995): *Div, Grad, Curl Are Dead*. http://people.ucsc.edu/~rmont/papers/Burke_DivGradCurl.pdf. See also <http://www.ucolick.org/~burke>.
- Calvino, I. (1979): *Invisible Cities*. (Picador, London). Transl. by William Weaver. First publ. in Italian 1972.
- Chew, G. F. (1970): *Hadron bootstrap: triumph or frustration?* Phys. Today **23**¹⁰, 23–28. doi:[10.1063/1.3021778](https://doi.org/10.1063/1.3021778).
- Clifford, W. K. (1876): *On the space-theory of matter*. Proc. Camb. Philos. Soc. **2**, 157–158. <https://archive.org/details/proceedingscamb06socigoog>.
- Davis, T. M., Lineweaver, C. H. (2004): *Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe*. Publ. Astron. Soc. Aust. **21**¹, 97–109. doi:[10.1071/AS03040](https://doi.org/10.1071/AS03040).
- Dirac, P. A. M. (1955): *Gauge-invariant formulation of quantum electrodynamics*. Can. J. Phys. **33**¹¹, 650–660. doi:[10.1139/p55-081](https://doi.org/10.1139/p55-081).
- Doyle, A. C. (1887): *A Study in Scarlet*. Repr. in Doyle (1998; 2014).

Bibliography

- Doyle, A. C. (1998): *Camden House: The Complete Sherlock Holmes*. <http://ignisart.com/camdenhouse/canon>. First publ. 1887–1927.
- (2014): *The Complete Sherlock Holmes*. <https://sherlock-holm.es>. Version 3.1. First publ. 1887–1927.
- Dyson, F. J. (2007): *Why is Maxwell's theory so hard to understand?* In: *Antennas and propagation, 2007. eucap 2007*, ed. by The Institution of Engineering and Technology (IEEE, Edinburgh): 1–6. <http://www.damtp.cam.ac.uk/user/tong/em/dyson.pdf>.
- Eckart, C. (1940): *The thermodynamics of irreversible processes. III. Relativistic theory of the simple fluid*. Phys. Rev. **58**¹⁰, 919–924. doi:[10.1103/PhysRev.58.919](https://doi.org/10.1103/PhysRev.58.919).
- Einstein, A. (1905a): *Zur Elektrodynamik bewegter Körper*. Ann. Phys. (Berl.) **17**, 891–921. Transl. in Einstein (1989a) Doc. 23 pp. 140–171.
- (1905b): *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* Ann. Phys. (Berl.) **18**, 639–641. Transl. in Einstein (1989a) Doc. 24 pp. 172–174.
- (1989a): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909. (English translation)*. (Princeton University Press, Princeton). Transl. of Einstein (1989b) by Anna Beck and Peter Havas.
- (1989b): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909*. (Princeton University Press, Princeton). Ed. by John Stachel. Transl. in Einstein (1989a).
- Faraday, M. (1846): *Thoughts on ray-vibrations*. Philos. Mag. **28**¹⁸⁸, 345–350. doi:[10.1080/1478644608645431](https://doi.org/10.1080/1478644608645431).
- Feynman, R. P. (1989): “*Surely You’re Joking, Mr. Feynman!*”: *Adventures of a Curious Character*, repr. (Bantam, New York). “As told to Ralph Leighton”, ed. by Edward Hutchings. First publ. 1985.
- Fliegel, H. F., DiEsposti, R. S. (1996): *GPS and relativity: an engineering overview*. PTTI Proc. **28**, 189–200. https://ui.adsabs.harvard.edu/link_gateway/1997ptti.conf..189F/ADS_PDF, <https://apps.dtic.mil/sti/citations/ADA516975>.
- Galilei, G. (1623): *Il Saggiatore*. http://www.liberliber.it/biblioteca/g/galilei/il_saggiatore/html/index.htm; parts transl. in Seeger (1966).
- iso (2009): *ISO 80000-1:2009: Quantities and units 1: General*. International Organization for Standardization.
- (2019): *ISO 80000-2:2019: Quantities and units 2: Mathematics*. International Organization for Standardization.
- Kandus, A., Tsagas, C. G. (2008): *Generalized Ohm’s law for relativistic plasmas*. Mon. Not. R. Astron. Soc. **385**², 883–892. doi:[10.1111/j.1365-2966.2008.12862.x](https://doi.org/10.1111/j.1365-2966.2008.12862.x).
- Kovalevsky, J., Seidelmann, P. K. (2004): *Fundamentals of Astrometry*. (Cambridge University Press, Cambridge). doi:[10.1017/CBO9781139106832](https://doi.org/10.1017/CBO9781139106832).
- Leishman, J. G. (2024): *Introduction to Aerospace Flight Vehicles*. (Embry-Riddle Aeronautical University, Daytona Beach, USA). doi:[10.15394/eaglepub.2022.1066](https://doi.org/10.15394/eaglepub.2022.1066). First publ. 2022.
- Maxwell (Clerk Maxwell), J. (1855): *On Faraday’s lines of force*. Trans. Cambridge Philos. Soc. **10**, 27–83. First read 1855–1856; repr. in Maxwell (2010a) doc. VIII pp. 155–229.
- (1870): *Mathematics and physics*. Rep. Br. Assoc. Adv. Sci. **40**, 1–9. <https://archive.org/details/reportofbritishab71brit>. Repr. in Maxwell (2010b) doc. XLI pp. 215–229. See also Dyson (2007).
- (1878): *Tait’s “Thermodynamics”*. Nature **17**^{431, 432}, 257–259, 278–280. doi:[10.1038/017257a0](https://doi.org/10.1038/017257a0). Repr. in Maxwell (2010b) doc. XCI pp. 660–671.

Bibliography

- Maxwell (Clerk Maxwell), J. (2010a): *The Scientific Papers of James Clerk Maxwell*. Vol. 1, repr. (Cambridge University Press, Cambridge). Ed. by W. D. Niven. doi:10.1017/CBO9780511698095, <https://archive.org/details/scientificpapers01maxwuoft>. First publ. 1890.
- (2010b): *The Scientific Papers of James Clerk Maxwell*. Vol. 2, repr. (Cambridge University Press, New York). Ed. by W. D. Niven. doi:10.1017/CBO9780511710377, <https://archive.org/details/scientificpapers02maxwuoft>. First publ. 1890.
- Misner, C. W., Thorne, K. S., Wheeler, J. A. (2017): *Gravitation*, repr. (Princeton University Press, Princeton and Oxford). With a new foreword by David I. Kaiser and a new preface by Charles W. Misner and Kip S. Thorne. First publ. 1970. <https://archive.org/details/GravitationMisnerThorneWheeler>.
- Müller, J., Soffel, M., Klioner, S. A. (2008): *Geodesy and relativity*. J. Geod. **82**³, 133–145. doi:10.1007/s00190-007-0168-7.
- Newton, I. (1726a): *Philosophiae naturalis principia mathematica*, 1st ed. (Societatis Regiae, London). https://archive.org/details/philiaenatu00newt_0, <https://archive.org/details/philiaenat00newt>. Transl. in Newton (1846; 1974a,b).
- (1726b): *Philosophiae naturalis principia mathematica*, “Tertia aucta & emendata” ed. (Guil. & Joh. Innys, London). <http://archive.org/details/principiareprint00newtuoft>, <http://archive.org/details/principia00newtuoft>. First publ. 1687; transl. in Newton (1846; 1974a,b).
- (1846): *Newton's Principia: The Mathematical Principles of Natural Philosophy*, first American, “carefully rev. and corr.” ed. (Daniel Adee, New York). Transl. of Newton (1726b) by Andrew Motte. With Newton's system of the world and a life of the author by N. W. Chittenden.
- (1974a): *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his system of the world. Vol. One: The Motion of Bodies*. (University of California Press, Berkeley). Transl. of Newton (1726b) by Andrew Motte, rev. and supplied with an historical and explanatory appendix by Florian Cajori.
- (1974b): *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his system of the world. Vol. Two: The System of the World*. (University of California Press, Berkeley). Transl. of Newton (1726b) by Andrew Motte, rev. and supplied with an historical and explanatory appendix by Florian Cajori.
- Parker, E. N. (1974a): *Hydraulic concentration of magnetic fields in the solar photosphere. II. Bernoulli effect*. Astrophys. J. **190**, 429–436. doi:10.1086/152894. See also Parker (1974b).
- (1974b): *Hydraulic concentration of magnetic fields in the solar photosphere. I. Turbulent pumping*. Astrophys. J. **189**, 563–568. doi:10.1086/152835. See also Parker (1974a).
- Petit, G., Wolf, P. (2005): *Relativistic theory for time comparisons: a review*. Metrologia **42**³, S138–S144. doi:10.1088/0026-1394/42/3/S14, http://geodesy.unr.edu/hanspeterplag/library/geodesy/time/met5_3_S14.pdf.
- Ryutova, M. (2018): *Physics of Magnetic Flux Tubes*, 2nd ed. (Springer, Heidelberg). doi:10.1007/978-3-319-96361-7. First publ. 2015.
- Seeger, R. J., ed. (1966): *Galileo Galilei, His life and His Works*. (Pergamon, Oxford).
- Styer, D. F. (2000): *Insight into entropy*. Am. J. Phys. **68**¹², 1090–1096. doi:10.1119/1.1287353, <http://www.nd.edu/~powers/ame.20231/styer2000.pdf>.
- Taylor, E. F., Wheeler, J. A. (2000): *Exploring Black Holes: Introduction to General Relativity*. (Addison Wesley Longman, San Francisco).

Bibliography

- Truesdell III, C. A. (1956): *Experience, theory, and experiment*. In: *Proceedings of the sixth hydraulics conference*, ed. by L. Landweber, P. G. Hubbard (University of Iowa, Iowa City, USA): 3–18. doi:10.17077/006159. Repr. with comments in Truesdell (1987) Chapter. 1 pp. 3–20.
- (1964): *Die Entwicklung des Drallsatzes*. Z. Angew. Math. Mech. **44**^{4/5}, 149–158. First delivered as a lecture 1963. See also Truesdell (1968a).
- (1966): *Six Lectures on Modern Natural Philosophy*. (Springer, Berlin). doi:10.1007/978-3-662-29756-8.
- (1968a): *Whence the law of moment of momentum?* In: Truesdell (1968b): ch. V:pp. 239–271. doi:10.1007/978-3-642-86647-0_5. First delivered as a lecture 1963. See also Truesdell (1964).
- (1968b): *Essays in the History of Mechanics*. (Springer, Berlin). doi:10.1007/978-3-642-86647-0.
- (1987): *An Idiot's Fugitive Essays on Science: Methods, Criticism, Training, Circumstances*, 2nd pr., rev. and augmented. (Springer, New York). First publ. 1984.
- Walpole, S. C., Prieto-Merino, D., Edwards, P., Cleland, J., Stevens, G., Roberts, I. (2012): *The weight of nations: an estimation of adult human biomass*. BMC Public Health **12**, 439. doi:10.1186/1471-2458-12-439.
- Whittaker, E. T. (1951): *A History of the Theories of Aether and Electricity. Vol. 1: The Classical Theories*, rev. and enl. ed. (Thomas Nelson and Sons, London). <https://archive.org/details/e-t-whittaker-a-history-of-the-theories-of-aether-and-electricity-vol-1-london-n>. First publ. 1910.
- Wilczek, F., Zee, A. (1979): *Operator analysis of nucleon decay*. Phys. Rev. Lett. **43**²¹, 1571–1573. doi:10.1103/PhysRevLett.43.1571.
- Wojtan, C., Thürey, N., Gross, M., Turk, G. (2009): *Deforming meshes that split and merge*. ACM Trans. Graph. **28**³, 1–10. doi:10.1145/1531326.1531382.

Chapter 10

Balance of angular momentum

Balance of angular momentum

$$\begin{aligned} \mathbf{L}(t_1) &= \mathbf{L}(t_0) + \int_{t_0}^{t_1} \boldsymbol{\tau}(t) dt + \int_{t_0}^{t_1} \mathbf{M}(t) dt & \frac{d\mathbf{L}(t)}{dt} &= \boldsymbol{\tau}(t) + \mathbf{M}(t) \\ &\text{integral form} & &\text{differential form} \\ &&&(10.1) \end{aligned}$$

Chapter 11

Balance of entropy

Balance of entropy

$$S(t_1) \geq S(t_0) + \int_{t_0}^{t_1} H(t) \, dt \quad \begin{matrix} \text{integral form} \\ \text{differential form} \end{matrix} \quad \frac{dS(t)}{dt} \geq H(t)0 \quad (11.1)$$

Chapter 12

Conservation of electric charge and magnetic flux

Chapter 13

Constitutive relations

Chapter 14

Conservation of matter

"What you do in this world is a matter of no consequence," returned my companion, bitterly. *"The question is, what can you make people believe that you have done? . . ."*

Sherlock Holmes (A. C. Doyle) 1887

14.1 Formulation and generalities

Choose an arbitrary sequence of closed control surfaces between coordinate times t_0 and t_1 , enclosing a sequence of control volumes. Denote by $N(t)$ and $J(t)$ the amount of matter in the control volume and the influx through the control surface at time t . Then the law of **conservation of matter** is expressed by either of the equations

Conservation of matter

Conservation of matter is a law that we intuitively take for granted and use continuously in our life. The very notion of ‘object’ – including living objects – is possible thanks to this fundamental regularity of nature: we can speak of objects because they exist for some time and we can follow them as they move in space.

14.1.1 Balances of matter

The law of *conservation* of matter holds, as far as we know, for all kinds of matter *considered together*. Wilczek, Zee 1979

More specifically, it seems to hold in extreme physical conditions if we count baryonic and anti-leptonic matter as ‘positive’ and leptonic and anti-baryonic matter as negative.

When we consider different kinds of matter, such as different chemical elements, individual conservation laws still hold and can be applied in many common and important physical situations.

For some kinds of matter, conservation does not hold, however; for example in the previously mentioned phenomena involving radioactive decay and nuclear energy. For these kinds of matter we can still apply a *balance* law, with a non-zero supply or sink.

“All these operators conserve $B - L$, so in any superunified theory we expect $B - L$ to be conserved...”

14.2 Examples of constitutive relations

14.2.1 Relation between matter and mass-energy

The most common constitutive relation used together with conservation of matter is the one relating a constant amount of mass-energy m to a particular amount of matter N :

Molar mass

$$m = \rho N \quad (14.2)$$

where the proportionality constant $\rho = m/N$ is called **molar-mass constant**, and depends on the kind of matter considered.

For instance, in many physical phenomena involving water we assume the constitutive relation above, with a of [water molar-mass constant¹](#) of approximately $\rho_{\text{H}_2\text{O}} = 0.0180 \text{ kg/mol}$. If a volume contains $N = 20 \text{ mol}$ of water, we usually attribute to it a mass-energy of

$$\begin{aligned} m &= \rho_{\text{H}_2\text{O}} N \\ &= 0.0180 \text{ kg/mol} \times 20 \text{ mol} \\ &= 0.36 \text{ kg}. \end{aligned}$$

The exact value of the molar-mass constant depends not only on the substance but also on the context and application: the substance may

actually consist of a mixture of different kinds of matter. ‘Air’ for example is a mixture of different chemical elements, and their proportion in the mixture may depend on physical conditions such as temperature, and on geographical position. ‘Water’ is a mixture of different isotopes³ (molecules differing in the number of neutrons), and again the mixture proportions may vary with physical conditions.

The constant relating mass-energy and amount of matter is the same for volume contents and for fluxes. So to a matter flux J we can associate a mass-energy flux ρJ . For instance, if through a surface there’s a flux of $J = 30 \text{ mol/s}$ of water, then we associate to it a mass flux of

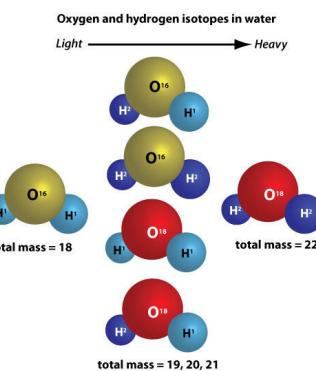
$$\rho_{\text{H}_2\text{O}} J = 0.0180 \text{ kg/mol} \times 30 \text{ mol/s} = 0.54 \text{ kg/s.}$$

Using this constitutive relation we can re-express the conservation of matter in terms of mass. Suppose that for a given kind of matter the proportionality constant is ρ ; then conservation of matter implies

$$\rho N(t_1) - \rho N(t_0) - \int_{t_0}^{t_1} \rho J(t) dt = 0 \quad \frac{d\rho N(t)}{dt} - \rho J(t) = 0$$

This is the ‘law of conservation of mass’. Keep in mind that this is not a universal law: it’s only an approximate law, only valid as long as the constitutive relation between matter and mass-energy is acceptable.

Phenomena for which this constitutive equation is no longer valid appear for example in particle physics, nuclear physics, and astrophysics.



There are several different isotopes of water, with different masses (image from U.S. Geological Survey²)

Exercise 14.1

At a particular time, a party balloon contains 0.012 kg of Helium. A minute later, the same balloon contains 0.010 kg of helium. Assume conservation of Helium, as well as a constitutive relation between amount and mass with [Helium molar-mass constant⁴](#) $\rho_{\text{He}} = 4 \times 10^{-3} \text{ kg/mol}$.

1. How much is the *integrated efflux* of Helium, **in moles**, through the balloon’s surface during the one-minute lapse of time?
2. Assume that the efflux of Helium was constant in time during the one-minute time lapse. How much was the efflux of Helium, in moles/second?



14.2.2 Relation between matter and momentum

Closely related to the previous constitutive relation is the one relating an amount of matter and an amount of momentum, briefly discussed before. Let us examine this constitutive relation more thoroughly.

Fix a coordinate system (t, x, y, z) . At a given point in space, take a small rectangular control surface of area A , with sides parallel to the coordinate axes y and z . See side picture below. The surface must *not* be moving in that coordinate system. Through this surface, in the positive- x direction, there's a flux of matter that we denote J_x .

Construct two more surfaces of area A through the same point in a analogous way: one parallel to the zx axes and one to the xy axes. The matter fluxes through them are J_y and J_z .

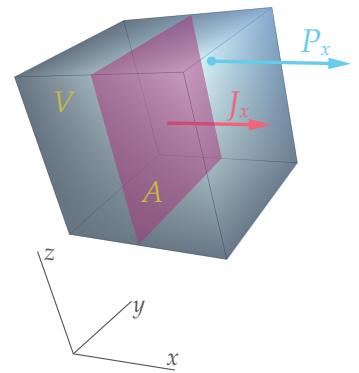
Finally take a control volume of cuboid shape that fits the three surfaces. This cuboid has volume V , and contains an amount of matter N and an amount of momentum \mathbf{P} . Recall the association between the matter flux through the first surface and the velocity component $v_x \equiv \frac{J_x/A}{N/V}$, and similarly for the other two surfaces.

Assume that the constitutive relation between matter and mass, previously discussed, is valid, and associates a mass-energy $m = \rho N$ to the amount of matter N .

Newtonian constitutive relation for momentum

The **Newtonian relation for the momentum of matter** is a constitutive relation saying that the flux and velocity of matter and the volume content of momentum are proportional:

$$\mathbf{P} = m\mathbf{v} = \rho \frac{V}{A} \begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} \quad (14.3)$$



This relation is only valid for very small areas and volumes (we should really take a limit), so small that the flux of matter or the volume content of momentum wouldn't change if we choose a surface and a volume in slightly different positions.

An analogous constitutive relation also exists between *flux* of momentum and flux and volume content of matter. Its most general formulation involves many mathematical terms, so let's consider a simplified situation instead.

Convective flux of momentum (special case)

Through the yz -surface above there's a flux of momentum

$$\mathbf{F}_{\text{conv}} = \rho J_x \mathbf{v} \equiv \rho \frac{V}{A} \frac{J_x}{N} \begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} \quad (14.4)$$

called a **convective flux**. Analogous relations hold for the surfaces parallel to the zx and xy coordinates.



Usually there are additional, non-convective flows of momentum (forces) associated with matter, as we shall see later.

These constitutive relations are only approximate and become incorrect when the speeds involved are close to the speed of light or the energy concentrations are so high to create strong gravitational fields (high spacetime curvature), or when some kind of electromagnetic phenomena are also involved.

In particular, according to the Newtonian constitutive relation, *momentum and velocity of matter are always parallel*, because they are two vectors that differ by a multiplicative scalar. In general this is not exactly true: momentum can have a different direction from velocity. In general relativity one must often take this fact into account.

More symmetric relation between mass & momentum and matter

When we work with *densities*, which express the amounts of a quantity per unit volume or per unit surface, the relation between mass-energy & momentum on one side, and matter & its flux on the other, becomes more explicit and symmetric. We have

$$\mathbf{M} = \rho n \quad \mathbf{p} = \rho \mathbf{j}$$

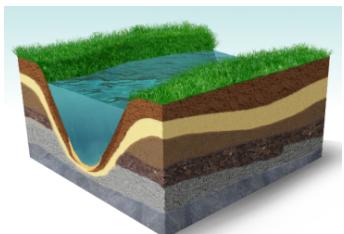
where \mathbf{M} and \mathbf{p} are the mass and momentum densities, and n and \mathbf{j} are the matter and matter-flux densities.

Exercise 14.2

The side picture illustrates the section of a river. The river has an approximately *steady flow* of water; this means that if we choose a static control volume anywhere in the river, such as the one in the depicted section, the amount of water N therein is *constant in time*.

Assume that (a) water satisfies conservation of matter; (b) the amount of water in the illustrated volume is $N = 4 \times 10^5 \text{ mol}$; (c) water only flows across the imaginary vertical surfaces that delimit the river section on the front and back of the picture, cutting across the river.

1. What can you say about the two fluxes across the two imaginary vertical surfaces, each considered as crossed from right to left in the picture? Can you find their numerical values?
2. You are told that the illustrated volume of water has total momentum $P = 21\,600 \text{ N s}$, directed from left to right perpendicularly to the imaginary vertical surfaces. Assuming the Newtonian constitutive relation for momentum, and the molar-mass constant $\rho_{\text{H}_2\text{O}} = 0.0180 \text{ kg/mol}$, can you find the values of the fluxes from the previous question?



(Illustration by Cathryn Briggs⁵)

14.2.3 Gas pressure

In the previous section we saw a constitutive relation between matter and *flux of momentum*, which we called ‘convective flux’. It is a flux that’s always present whenever there’s matter in motion.

In some circumstances, matter gives rise to a flux of momentum even when it is not in (visible) motion. In such cases it is said to be in a *gaseous state*⁶. The most cogent example is the air around us, which has a continuous flux of momentum, of approximately 10^5 N on every square-metre of surface, called *atmospheric pressure*⁷. This flux is always *compressive* and occurs through any surface: any imaginary surface within the body of air itself, and any real surface between air and any other object.

For some kinds of matter in a gaseous state, the magnitude of their compressive flux of momentum $\mathbf{F}_{\text{press}}$ can be approximated by the formula:

Ideal-gas law

$$\frac{F_{\text{press}}}{A} = R T \frac{N}{V}$$

called the **ideal-gas law**, where A is the area through which the momentum flux occurs, V is the volume of a small region adjacent to the area, N is the amount of matter in that volume, T is its temperature, and

$$R = 8.314\,462\,618\,153\,24 \text{ N m/(K mol)} \quad (\text{exactly})$$

is the **molar gas constant**⁸. The ratio F_{press}/A of force and area is called *pressure*.



A barometer for measuring atmospheric pressure (image from [National Geographic](#)⁹)

This law is very important in thermodynamics and chemistry, even if it's only an approximation.

14.2.4 Radioactive decay

For radioactive substances, for which conservation of mass does not hold, we still have a balance law where the supply, which we can denote with the symbol $-A$, has a specific constitutive equation:

Matter supply in radioactive decay

$$-A(t) = -\lambda N(t)$$

where λ is positive and called the **decay constant** of that particular substance. The supply with changed sign, A , is usually called **activity**¹⁰.

If there is no influx of matter, $J(t) = 0$, the balance law for the substance then takes the form

$$\frac{dN(t)}{dt} = -\lambda N(t)$$

called the **law of radioactive decay**¹¹.



When we see this symbol we know that there's matter that's only balanced, not conserved – which involves danger

14.3 Examples of applications

14.3.1 Rigid-body and particle mechanics

In many applications we set up our description of the physical phenomenon of interest in such a way that the law of conservation of matter is automatically satisfied. This is achieved by choosing a sequence of closed control surfaces for which the flux of matter is zero, and therefore the amount of matter in the control volumes is constant. This procedure is often carried without many comments – to the point of being almost forgotten at times.

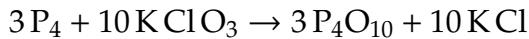
We saw an example of this procedure in the numerical evolution of the motion of a falling object. Take again a look at the timesteps in formula (5.12): you see that at each timestep we calculate the velocity \mathbf{v} of matter from the momentum \mathbf{P} , simply dividing by the mass m . *This mass is assumed to be constant, not changing with time*. But this mass is proportional to the amount of matter: $m = \rho N$. We are therefore assuming that the amount of matter N doesn't change with time: $dN(t)/dt = 0$. This assumption is indeed guaranteed by the balance of matter (14.1), by choosing a sequence of control surfaces that have no net flux of matter through them: $J(t) = 0$; in other words, they tightly wrap and follow the moving object.

$$\begin{aligned} \mathbf{v}(t) &= \mathbf{P}(t)/M \\ \mathbf{r}(t + \Delta t) &\approx \mathbf{r}(t) + \mathbf{v}(t) \Delta t \\ \mathbf{P}(t + \Delta t) &\approx \mathbf{P}(t) + [\mathbf{F}(t) + \mathbf{G}(t)] \Delta t \\ \mathbf{v}(t + \Delta t) &= \mathbf{P}(t + \Delta t)/M \end{aligned}$$

Snippet of formula (5.12)

14.3.2 Chemistry

One of the main assumptions in chemistry is the ‘permanence of atoms’. This assumption imposes important restrictions in the [stoichiometry](#)¹² of chemical reactions, that is, in determining the amounts of products that can appear from given amounts of reactants. For instance, the [match-head reaction](#)¹³



expresses that if 30 moles of oxygen (O) atoms appear among the reactants, they must also appear among the products; same for the 12 moles of [phosphorus \(P\) atoms](#)¹⁴, the 10 moles of [potassium \(K\) atoms](#)¹⁵, and so on.



This assumption is simply the statement of conservation of matter, *separately* for each chemical element: since the reaction doesn't let any other matter go in or out, the flux J for each chemical element must be zero. Therefore the amount N of each chemical element must be constant: $dN(t)/dt = J = 0$.

This assumption is often called ‘conservation of mass’, but, as previously explained, “mass” is in this case only a proxy for matter. The assumption

↳ § 14.1 page 166

of the permanence of atoms is only approximate and no longer valid in phenomena for which nuclear physics or particle physics become relevant.

14.3.3 Climate

The laws of conservation and balances of matter turn out to be very useful also in problems related to climate.

On the Earth's surface and atmosphere we can assume a law of conservation of matter for each of the [stable isotopes](#)¹⁶ of the chemical elements, for instance the [two stable carbon isotopes](#)¹⁷ and the [three stable oxygen isotopes](#)¹⁸. We can therefore follow these isotopes as they flow between different physical systems, such as the atmosphere, the oceans, and the biosphere, especially plants – in practice we are using huge control volumes.

For the radioactive isotopes, the [law of radioactive decay](#) applies, and from it we can deduce the age of different materials and objects such as ice or wood.

This is how we are able to say that human usage of fossil fuel is an important factor in the increase of carbon dioxide (CO_2) in the atmosphere during the past 200 years or so. Take a look at the more detailed explanations given by the [Global Monitoring Laboratory](#)¹⁹.

↳ § 7.2.3 page 116

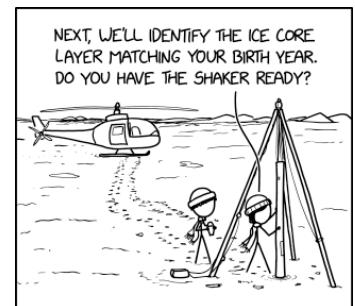
Exercise 14.3

Measuring the relative amounts of carbon within an air pocket near the surface of a block of ice, you find that one mole of air contains 1×10^{-12} mol of the radioactive isotope ^{14}C . Making an analogous measurement for an air pocket deeper in the ice, you find instead a relative abundance of 2×10^{-13} mol of ^{14}C . How old is the deeper section of ice?

Find out the ice's age t_{ice} by numerically evolving the balance equation

$$\frac{dN(t)}{dt} = -\lambda N(t)$$

with $\lambda = 0.000\,122 \text{ yr}^{-1}$, until you reach the amount $N(t_{\text{ice}}) = 2 \times 10^{-13}$ mol. Use the procedure of [eq. \(5.11\)](#), starting from time $t_0 = 0 \text{ yr}$ and $N(t_0) = 1 \times 10^{-12}$ mol, and assuming that the flow $J(t)$ of ^{14}C is zero. You can use a timestep $\Delta t = (1/365) \text{ yr}$.



<https://xkcd.com/2902>

↳ § 5.5 page 95

14.3.4 Fluid dynamics

The law of conservation of matter in its explicit form is at the heart of fluid-dynamic problems. Consider the flow of a fluid (liquid or gas) for instance through a pipe or through a jet engine. When we say that a flow is **steady** we mean that the volume contents and the fluxes taken for whatever control volumes and surfaces do not change in time (though they may change in space). This condition can be viewed as a constitutive relation.

Consider a control volume, for instance the one indicated in light blue in the side picture. The amount of fluid N in this volume is constant in time: $\frac{dN(t)}{dt} = 0$. By the law of conservation of matter, the total influx must then be zero:

$$\underbrace{\frac{dN(t)}{dt}}_{=0} - J(t) = 0 \quad \Rightarrow \quad J(t) = 0$$

and it is given by the influxes through three surfaces: the side surface, the one at the top, and the one at the bottom. The influx through the side surface is zero. Let us denote the influx through the top surface by J_1 , and the influx through the bottom surface by J_2 . We must therefore have

$$0 = J(t) = J_1(t) + J_2(t) \quad \Rightarrow \quad J_1(t) = -J_2(t)$$

that is, the influx through the top surface must equal the efflux through the bottom one. This is a very powerful deduction: consider that it is valid even if the flow of the fluid is turbulent, and it is valid in more general situations such as at different sections of a jet engine.

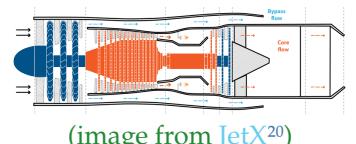
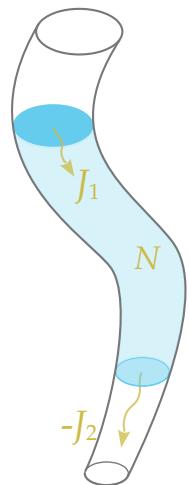
If the surfaces are small enough, we can also use the connection between flux and velocity:

$$v_1 = \frac{J_1/A_1}{N/V} \quad v_2 = \frac{-J_2/A_2}{N/V}$$

where A_1, A_2 are the areas of the top and bottom surfaces, and v_1, v_2 the *downward* velocities through them (hence the minus sign for the bottom surface). From $J_1 = -J_2$ we then find this important relationship:

$$v_1 A_1 = v_2 A_2$$

that is, if the area through which the flux occur decreases, then the velocity of the fluid through it increases, and vice versa. This is what we often observe in running water from our taps.



(image from JetX²⁰)



The thickness and velocity variations of tap water are a consequence of the law of conservation of matter

URLs for chapter 14

1. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/H2O/h1H2>
2. <https://www.usgs.gov/media/images/water-isotopes-diagram>
3. <https://www.britannica.com/science/isotope>
4. <https://webbook.nist.gov/cgi/inchi/InChI%3D1S/He>
5. <https://dribbble.com/squarefountain>
6. <https://www.britannica.com/science/gas-state-of-matter>
7. <https://www.britannica.com/science/atmospheric-pressure>
8. <https://www.britannica.com/science/universal-gas-constant>
9. <https://education.nationalgeographic.org/resource/atmospheric-pressure/>
10. <https://goldbook.iupac.org/terms/view/A00114>
11. <https://www.britannica.com/science/radioactivity/Rates-of-radioactive-transitions#ref496415>
12. <https://doi.org/10.1351/goldbook.S06026>
13. <https://chem.washington.edu/lecture-demos/match-head-reaction>
14. <https://pubchem.ncbi.nlm.nih.gov/element/Phosphorus>
15. <https://pubchem.ncbi.nlm.nih.gov/element/Potassium>
16. <https://www.britannica.com/science/isotope/The-discovery-of-isotopes#ref496311>
17. <https://education.jlab.org/itselemental/ele006.html>
18. <https://education.jlab.org/itselemental/ele008.html>
19. <https://gml.noaa.gov/outreach/isotopes/>
20. <https://www.jet-x.org/a8.html>

Chapter 15

Balance of energy

Chapter 16

Balance of momentum

I hold in fact

- (1) *That small portions of space are in fact of a nature analogous to little hills on a surface which is on the average flat; namely, that the ordinary laws of geometry are not valid in them.*
- (2) *That this property of being curved or distorted is continually being passed on from one portion of space to another after the manner of a wave.*
- (3) *That this variation of the curvature of space is what really happens in that phenomenon which we call the motion of matter, whether ponderable or etherial.*
- (4) *That in the physical world nothing else takes place but this variation, subject (possibly) to the law of continuity.*

W. K. Clifford 1876

Chapter 17

Balance of angular momentum

Chapter 18

Densities and flux-densities

How do we mathematically represent the seven main quantities? We said that, for each, we can ask *how much in this region?* and *how much through this surface during this time?* We must therefore use mathematical objects that allow us to answer these two questions. These objects are the *density* and the *flux vector*. The idea behind them is quite simple.

Density

Take a point in spacetime with coordinates (t_0, x_0, y_0, z_0) . At the instant t_0 , imagine a very small cuboid centred at (x_0, y_0, z_0) , with sides aligned with the coordinate axes x, y, z and of lengths $\Delta x, \Delta y, \Delta z$. Then we have

$$\text{amount of quantity in cuboid} = n(t_0, x_0, y_0, z_0) \Delta x \Delta y \Delta z . \quad (18.1)$$

where $n(t_0, x_0, y_0, z_0)$ is the **density** at the spacetime point (t_0, x_0, y_0, z_0) .

The density tells us how much quantity there's in a unit of volume. As the notation suggests, it is a function of the coordinates, that is, of time and spatial position. This functional dependence reflects the fact that we can have a larger amount of a quantity concentrated in some regions at some times, than in other regions or at other times.

In order to calculate the total amount of the quantity in an arbitrary 3D region, we simply divide it into very small cuboids similar to the one above. For each cuboid we calculate the respective amount; this amount will generally be different from cuboid to cuboid, because the value of the

density $n(t, x, y, z)$ will be different at each cuboid's centre. Then we sum up all these amounts.

Isn't this an integral?

You probably recognize this procedure as the description of integration. Indeed we can write:

$$\text{total amount of quantity in 3D region} = \iiint_{\text{3D region}} n(t, x, y, z) dx dy dz . \quad (18.2)$$

Clearly we must know the value of the density n at all points within the region, in order to calculate the total amount. Note that if the quantity is absent in some subregion, then $n(\dots) = 0$ there.

Do not confuse flux with movement

Flux

For the flux of the quantity through a surface we consider three different cases:

x-Flux

Take again a point in spacetime with coordinates (t_0, x_0, y_0, z_0) . Keeping x_0 fixed, imagine a very small rectangular surface centred at (x_0, y_0, z_0) , with sides aligned with the coordinate axes y, z and of lengths $\Delta y, \Delta z$. Imagine that this surface exists for a lapse of time Δt around the time t_0 . Then we have

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta y \Delta z \\ \text{towards positive } x \\ \text{during time lapse} \end{array} \right\} = j_x(t_0, x_0, y_0, z_0) \Delta t \Delta y \Delta z . \quad (18.3)$$

where $j_x(t_0, x_0, y_0, z_0)$ is the **x-flux** at the spacetime point (t_0, x_0, y_0, z_0) .

The **x-flux** tells us how much quantity is flowing in a unit of time through a unit of surface parallel to y, z . Also the **x-flux** is a function of the coordinates, because the flux could be larger through some surfaces at some times, than through other surfaces at other times. If the quantity is

flowing in the negative x direction, then j_x will be negative; and clearly if no quantity is flowing through the small rectangle, then $j_x = 0$.

In an analogous way we define a y -flux and a z -flux:

y -Flux

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta z \Delta y \\ \text{towards positive } y \\ \text{during time lapse} \end{array} \right\} = j_y(t_0, x_0, y_0, z_0) \Delta t \Delta z \Delta y . \quad (18.4)$$

z -Flux

$$\left. \begin{array}{l} \text{amount of quantity} \\ \text{through rectangle } \Delta x \Delta y \\ \text{towards positive } z \\ \text{during time lapse} \end{array} \right\} = j_z(t_0, x_0, y_0, z_0) \Delta t \Delta x \Delta y . \quad (18.5)$$

The three fluxes can be grouped into the **flux vector**

$$\mathbf{j}(t, x, y, z) := (j_x(t, x, y, z), j_y(t, x, y, z), j_z(t, x, y, z)) \quad (18.6)$$

In order to calculate the total amount of the quantity through an arbitrary sequence of 2D surfaces during a time lapse, we simply divide the total time into very brief time lapses, and for each of these we approximate the surface with very small rectangles aligned along the three axes. For each such lapse and rectangle we calculate the respective flux amount, using j_x , j_y , or j_z depending on the orientation of the rectangle. Then we sum up all these amounts.

T

This is also the description of an integration, which can be written

total amount of quantity through sequence of surfaces =

$$\begin{aligned} & \iiint j_x(t, x, y, z) dt dy dz + \\ & \iiint j_y(t, x, y, z) dt dz dy + \\ & \iiint j_z(t, x, y, z) dt dx dy \end{aligned} \quad (18.7)$$

We shall not have to calculate density integrals (18.2) and flux integrals (18.7) in these notes. In concrete physics problems these integrals are often difficult to calculate, and require advanced mathematical techniques and specialized software. But if you'll ever end up working in fields such as computational fluid dynamics, atmospheric or ocean modelling, or numerical relativity, then you shall probably encounter them again.

The definitions above of density and flux are appropriate if the quantity in question is a scalar. For vector quantities such as momentum and angular momentum, the density is a vector, and each of the three fluxes is a vector as well.

18.1 Balance laws expressed with derivatives

In Chapter 18 we saw that each of the main seven quantities is represented by a density and by a flux vector, which allow us to calculate the amount of quantity within a 3D region at a particular time, and flowing through a 2D surface during a time interval. These amounts are exactly what $N(t_0)$, $N(t_1)$, J express. Therefore, a conservation or balance law must also give some relationship between density and flux vector.

The way to find such relationship is intuitive and mathematically simple, if only a bit lengthy. The idea is to formulate the balance law for a very simple case, the same simple set-up that we used in Chapter 18 to introduce density and flux.

Take a point in spacetime with coordinates (t, x, y, z) . We apply the conservation law (5.1) to the following situation:

$N(t_0)$: Choose an initial time $t_0 = t - \frac{\Delta t}{2}$, slightly before t , and a cuboid region centred at (x, y, z) with sides $\Delta x, \Delta y, \Delta z$. The total amount of the quantity within this region is given, adapting formula (18.1), by

$$N(t_0) = n\left(t - \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z. \quad (18.8)$$

$N(t_1)$: Choose a final time $t_1 = t + \frac{\Delta t}{2}$, slightly after t , and a cuboid region centred at (x, y, z) with sides $\Delta x, \Delta y, \Delta z$. The total amount of the quantity within this region is given, adapting formula (18.1), by

$$N(t_1) = n\left(t + \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z. \quad (18.9)$$

J : We calculate the flux separately through the six rectangular surfaces bounding the small cuboid region:

J_x : First choose a rectangular surface centred slightly to one side of the point (x, y, z) : at $(x - \frac{\Delta x}{2}, y, z)$, with sides parallel to y, z and lengths $\Delta y, \Delta z$. We keep this surface constant for the small duration Δt . The x -flux through this rectangle, towards positive x , during Δt , is according to formula (18.3)

$$J_x^{\text{in}} = j_x(t, x - \frac{\Delta x}{2}, y, z) \Delta t \Delta y \Delta z .$$

Now choose the rectangular surface parallel to the previous one, but on the other side of the point (x, y, z) , centred at $(x + \frac{\Delta x}{2}, y, z)$. The x -flux through this rectangle, towards positive x , during Δt , is

$$J_x^{\text{out}} = j_x(t, x + \frac{\Delta x}{2}, y, z) \Delta t \Delta y \Delta z .$$

Note that the flux J_x^{in} points to the interior of the 3D region, towards its centre (x, y, z) ; whereas J_x^{out} points to the exterior, away from the centre. The total flux into the region is therefore given by their difference:

$$\begin{aligned} J_x &= J_x^{\text{in}} - J_x^{\text{out}} \\ &= [j_x(t, x - \frac{\Delta x}{2}, y, z) - j_x(t, x + \frac{\Delta x}{2}, y, z)] \Delta t \Delta y \Delta z \end{aligned} \quad (18.10)$$

J_y, J_z : An analogous reasoning can be made to find the total flux through the other four surfaces, two parallel to z, y and two to x, y . We find

$$\begin{aligned} J_y &= J_y^{\text{in}} - J_y^{\text{out}} \\ &= [j_y(t, x, y - \frac{\Delta y}{2}, z) - j_y(t, x, y + \frac{\Delta y}{2}, z)] \Delta t \Delta z \Delta x \end{aligned} \quad (18.11)$$

$$\begin{aligned} J_z &= J_z^{\text{in}} - J_z^{\text{out}} \\ &= [j_z(t, x, y, z - \frac{\Delta z}{2}) - j_z(t, x, y, z + \frac{\Delta z}{2})] \Delta t \Delta x \Delta y \end{aligned} \quad (18.12)$$

The total flux into the region during the time interval Δt is finally the sum of the three above:

$$J = J_x + J_y + J_z \quad (18.13)$$

Up to now we have only chosen arbitrary surfaces, regions, and a time interval. Now let's assume that the corresponding amounts and fluxes

satisfy a balance law like (5.1). Putting together the puzzle pieces we find

$$N(t_1) - N(t_0) - J = 0 \quad \implies$$

$$\left. \begin{aligned} & n\left(t + \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z - n\left(t - \frac{\Delta t}{2}, x, y, z\right) \Delta x \Delta y \Delta z \\ & - [j_x(t, x - \frac{\Delta x}{2}, y, z) - j_x(t, x + \frac{\Delta x}{2}, y, z)] \Delta t \Delta y \Delta z \\ & - [j_y(t, x, y - \frac{\Delta y}{2}, z) - j_y(t, x, y + \frac{\Delta y}{2}, z)] \Delta t \Delta z \Delta x \\ & - [j_z(t, x, y, z - \frac{\Delta z}{2}) - j_z(t, x, y, z + \frac{\Delta z}{2})] \Delta t \Delta x \Delta y \end{aligned} \right\} = 0$$

This expression is quite long, but it should be intuitively understandable if you try to identify the individual summed terms.

Now we take the last, long expression – which is an equality – and divide its left and right sides by $\Delta t \Delta x \Delta y \Delta z$. Note that many of the “ Δ ” factors simplify out. We arrive at this:

$$\left. \begin{aligned} & \frac{n\left(t + \frac{\Delta t}{2}, x, y, z\right) - n\left(t - \frac{\Delta t}{2}, x, y, z\right)}{\Delta t} \\ & + \frac{j_x\left(t, x + \frac{\Delta x}{2}, y, z\right) - j_x\left(t, x - \frac{\Delta x}{2}, y, z\right)}{\Delta x} \\ & + \frac{j_y\left(t, x, y + \frac{\Delta y}{2}, z\right) - j_y\left(t, x, y - \frac{\Delta y}{2}, z\right)}{\Delta y} \\ & + \frac{j_z\left(t, x, y, z + \frac{\Delta z}{2}\right) - j_z\left(t, x, y, z - \frac{\Delta z}{2}\right)}{\Delta z} \end{aligned} \right\} = 0 .$$

Still a long expression; but examine each fraction: we have a difference between a term calculated at some point, and one calculated at some point plus an increment “ Δ ”; this difference is then divided by that increment. This is the definition of *partial derivative*, that is, the derivative with respect to one variable, while keeping the other variables fixed. The short symbol for this is “ $\frac{\partial \dots}{\partial \dots}$ ”, so we can simply write

Conservation law in derivative form

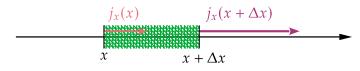
$$\frac{\partial n}{\partial t} + \frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z} = 0 \quad (18.14)$$

Let's make a simple check to see if this formula makes sense, in a simple

one-dimensional case where we disregard the y, z coordinates (let's say their fluxes are zero), so that the balance law becomes

$$\frac{\partial n}{\partial t} + \frac{\partial j_x}{\partial x} = 0.$$

Suppose that the derivative of the x -flux is positive: $\frac{\partial j_x}{\partial x} > 0$. This means that the x -flux increases as we move from x to $x + \Delta x$, as shown in the side figure. The x -flux at x is bringing some amount of quantity into the region of interest, the x -flux at $x + \Delta x$ is taking a larger amount out of that region. Therefore, during the short time Δt in which this flux takes place, the total amount of quantity within the region will decrease. But the total amount is given by $n(t)$, which is therefore decreasing with time. This means that its time derivative is negative: $\frac{\partial n}{\partial t} < 0$. This agrees with balance law above: the sum of the two derivatives must be zero, so if one is positive, the other must be negative.



Balance laws like (18.14) are the basis of many important computational and simulation methods in a huge variety of physical applications: simulation of the ocean around an offshore oil platform, of wind in a wind farm, of air around an aeroplane's wing, of earthquakes, of electromagnetic-wave propagation, of oscillations in a suspension bridge, of percolation of fluids through terrain, of chemical reactions, of energy transfer... The list could go on for pages!

The basic idea for using a balance law for a simulation is easy to understand. Take again the simpler one-dimensional formula without y, z coordinates, and rewrite it remembering the basic meaning of the derivative:

$$\frac{n(t + \Delta t, x) - n(t, x)}{\Delta t} + \frac{j_x(t, x + \frac{\Delta x}{2}) - j_x(t, x - \frac{\Delta x}{2})}{\Delta x} = 0$$

Suppose we know all quantities at time t ; we can then find the density n at time $t + \Delta t$. Leave it on the left side and bring all other terms on the right side, multiplying them by Δt :

$$n(t + \Delta t, x) = \Delta t \left[n(t, x) - \frac{j_x(t, x + \frac{\Delta x}{2}) - j_x(t, x - \frac{\Delta x}{2})}{\Delta x} \right]$$

which is essentially [Euler's method](#)¹.

This formula can then be used iteratively to find the density n at later times. As a simple simulation set-up, we consider a region of space of interest and divide it into cells of width Δx . We also treat time in steps of Δt . Then the iterative procedure goes as follows:

1. *Initial values:* Assign the value of n at each x -cell at the initial time.
2. *For-loop in x :* for each x -cell, calculate the value of n at $t + \Delta t$, using the formula above.
3. Increase time by Δt .
4. Go to 2.

For example, suppose we have two x -cells of width $\Delta x = 1$: one centred at x_1 and the other at $x_2 = x_1 + 1$. There are three boundaries: the leftmost at $x_1 - \frac{1}{2}$, one between the cells at $x_1 + \frac{1}{2} \equiv x_2 - \frac{1}{2}$, and the rightmost at $x_2 + \frac{1}{2}$. Then we have the following steps:

1. Assign the values $n(t_0, x_1), n(t_0, x_2)$ of n at x_1 and x_2 at time t_0 .
2. Assign the values

$$j_x(t_0, x_1 - \frac{1}{2}), j_x(t_0, x_1 + \frac{1}{2}), j_x(t_0, x_2 + \frac{1}{2})$$

of the fluxes at the three boundaries at time t_0 .

3. Calculate the values of n at x_1 and x_2 at time $t_0 = t_0 + \Delta t$, for instance

$$n(t_0, x_1) = \Delta t \left[n(t_0, x_1) - \frac{j_x(t_0, x_1 + \frac{1}{2}) - j_x(t_0, x_1 - \frac{1}{2})}{\Delta x} \right].$$

Note that we have all quantities required on the right side.

4. Assign the values

$$j_x(t_0, x_1 - \frac{1}{2}), j_x(t_0, x_1 + \frac{1}{2}), j_x(t_0, x_2 + \frac{1}{2})$$

of the fluxes at the three boundaries at time t_0 .

5. Calculate the values of n at x_1 and x_2 at time $t_1 = t_0 + \Delta t$, for instance

$$n(t_1, x_1) = \Delta t \left[n(t_0, x_1) - \frac{j_x(t_0, x_1 + \frac{1}{2}) - j_x(t_0, x_1 - \frac{1}{2})}{\Delta x} \right].$$

Note that we have again all quantities required on the right side.

6. And so on.

This routine requires us to know the flux at each x -cell and at each time. But we shall soon see that this requirement can be bypassed.

The rudimentary routine above is the basis from which more precise and complex simulation routines are developed.

Exercise 18.1

Write a script, in your preferred programming language, that implements the simulation routine above in an x -grid with four cells $x = 1, \dots, x = 4$. Take a grid size $\Delta x = 1$ and time step $\Delta t = 1$. We can imagine that the quantity in question is electric charge. Use the following initial values for $n(t, x)$ at $t = 0$:

$$n(t=0, x=1) = 7, \quad n(t=0, x=2) = 0, \quad n(t=0, x=3) = 0, \quad n(t=0, x=4) = 7$$

and assume the flux is, at each time t :

$$\begin{aligned} j_x(t, x=0.5) &= 0, \\ j_x(t, x=1.5) &= +2, \quad j_x(t, x=2.5) = 0, \quad j_x(t, x=3.5) = -2, \\ j_x(t, x=4.5) &= 0. \end{aligned}$$

Questions and tasks:

1. Which time evolution for n do you observe? Does it make sense, given the fluxes above?
2. Feel free to try out or generalize your script with:
 - different initial values
 - different fluxes
 - an explicit formula for the flux, for instance

$$j_x(t, x) = 8x^3 - 60x^2 + 118x - 45$$

- time-dependent fluxes
- more cells
- two or three dimensions, including y - or z -fluxes

URLs for chapter 18

1. <https://mathworld.wolfram.com/EulerForwardMethod.html>

Chapter 19

Matter

Matter is probably the easiest quantity to grasp intuitively; it's what we can really call "stuff". It is important to clearly distinguish matter from *mass*: mass can be considered a property of matter, but the two are different: for example, the mass of some amount of matter may change, even if the amount of matter stays the same.

There are different kinds of matter, each of which satisfies a balance relation. The distinction into different kinds depends on the physical theory. In most everyday situations, the distinction corresponds to different [chemical elements¹](#), and each satisfies its own balance. These balances are the basis of [stoichiometry²](#). If we observe phenomena such as nuclear fission or fusion, however, we notice that the balances of chemical elements are not really satisfied. With such phenomena we make a different distinction of types of matter, for instance *baryons* and *leptons*, and each satisfies again its own balance. It is unclear whether these balances might be broken in other physical phenomena at smaller scales. In these notes we shall usually consider chemical elements as the different kinds of matter, making some exceptions in discussion of nuclear phenomena.

19.1 Electric charge and magnetic flux**19.2 Magnetic flux****19.3 Energy****19.4 Momentum****19.5 Angular momentum****19.6 Entropy**

URLs for chapter 19

1. <https://doi.org/10.1351/goldbook.C01022>
2. <https://doi.org/10.1351/goldbook.S06026>

Chapter 20

Thermodynamics

20.1 Notes on “quasi-static” processes

In general, none of the statements "reversible \Rightarrow quasi-static" or "quasi-static \Rightarrow reversible" is true.

A counterexample to the second implication are systems with internal state variables, which cannot be made non-dissipative, no matter how slowed-down they are. See the discussion and mathematical analysis in Astarita § 2.5.

A counterexample to the first implication is a system of spins in a crystal lattice. It is possible to *reversibly* bring the system from an equilibrium state to another with opposite temperature by reversing the external magnetic field *as fast as possible* – and therefore *not* through a quasi-static process. In fact it is key here that the process be *not-quasi-static*, but as fast as possible, because a slow change of the external magnetic field would lead to an irreversible process with dissipation. For more details see the discussion in Buchdahl, Lecture 20.

The point is that for some systems a *fast* change can actually prevent the onset of dissipative phenomena, and so the process needs to be fast if we want it to be reversible. Adiabatic processes often also need to be fast (as a curious historical fact, Truesdell & Bharatha, Preface p. xii, remark that “In introducing what we today call an ‘adiabatic process’, Laplace called it ‘a sudden compression’, in which he was followed by Carnot”).

In fact, clearly non-quasi-static phenomena such as *explosions* can in some circumstances be described by *reversible* processes! This is possible if the explosion involves many shock waves, as explained by Oppenheim, chap. 1 p. 63:

If there is more than one shock, the losses in available energy are diminished, so that in the limit, with an infinite number of shocks, they become negligible, and the process acquires the character of a thermodynamically optimal, i.e., reversible, change of state. The study of explosion processes reveals that, indeed, they are associated not with one but with a multitude of shocks.

For explosions see also the mathematical analysis by Dunwoody: *Explosion and implosion in a mixture of chemically reacting ideal gases*, where again reversible-process equations are used.

A caveat about reversible and quasi-static associations is given by Erickson (§ 1.2):

Some tend to associate nearly reversible processes with those taking place very slowly – the "quasi-static" processes. This probably stems, at least in part, from experience with classical theories of heat conduction, viscosity, and so on. However, a ball made of silly putty behaves almost reversibly when bounced rapidly and various other high polymers have similar predilections. So, it seems prudent to be open-minded in considering what may be reversible processes for particular systems.

He later discusses (§ 3.1) the case of bars subjected to dead loads, for which we can have reversible processes under sudden jumps in elongation. He concludes (p. 46) that "the sudden jump provides an example of a process that is reversible but not reasonably considered to be quasi-static".

But there's an important question that underlies our discussion: what do we actually mean by "quasi-static"? We need to specify a time scale, otherwise the term is undefined. For example, a geological process (say, tectonic motion) can be considered as quasi-static – or even completely static – on time scales of minutes or days; but it is not quasi-static on time scales of millions of years.

Whether a process is reversible or not, within any tolerance needed, is an experimental question. We can measure any relevant quantities, say pressure p and exchanged heat q , under the process, and compare them with those, p^* and q^* , determined by the equations for a reversible process. We may find for example that at all times

$$\left| \frac{p - p^*}{p^*} \right| < 0.001 , \quad \left| \frac{q - q^*}{q^*} \right| < 0.001$$

and conclude that the process is reversible, if relative discrepancies of 0.1% or less are negligible in our concrete application.

But suppose that someone tells us "if you want the process to be reversible, you must make sure that it is quasi-static". Alright, but how much is "quasi-static"? Is it OK if the piston moves with a speed of 1 cm/s? or is that too much? How about 1 mm/s? – In fact we may find that for some kind of fluid 1 cm/s is absolutely acceptable for the process to be reversible, whereas for another kind of fluid that speed would lead (at the same temperature) to an irreversible process.

You see how this imprecise situation can lead to circular definitions: "if the process is irreversible, then it means it isn't quasi-static" – but then we are actually *defining* "quasi-static" in terms of "reversible"! Any statement of the kind "reversible \Rightarrow quasi-static" or "quasi-static \Rightarrow reversible" then becomes not a matter of experimental verification, but of pure *semantics*. At this point we can simply get rid of "quasi-static" terminology since it doesn't bring any new physics to the table. This circularity is admitted for example by Callen in discussing irreversible gas expansion (Problem 4.2-3 p. 99):

The fact that $dS > 0$ whereas $dQ = 0$ is inconsistent with the presumptive applicability of the relation $dQ = T dS$ to all quasi-static processes. We define (by somewhat circular logic!) the continuous free expansion process as being "essentially irreversible" and *non-quasi-static*.

A similar criticism can be read in Astarita, § 2.9, p. 62, where he also provides a mathematical quantification of quasi-static, similar to the one given above for reversibility:

Often this point is circumvented by bringing in another difficult concept, that of a quasi-static transformation, which proceeds "through a sequence of equilibrium states." Quasi-static is an impressive word, but the only meaning which can be attached to it is the less impressive word "slow" – and how can one speak of slowness without implying the concept of time? How slow is slow enough? If one chooses to develop a thermodynamic theory (rather than a thermostatic one), the answer is easy. For instance, in the case of a system where the state is V, T, \dot{V} [the last is the rate of change of V], one needs to assume that

[the non-equilibrium pressure] $p(V, T, \dot{V})$ is a Taylor-series expandable at $\dot{V} = 0$ to obtain [that

$$p = p^* + \frac{\partial p}{\partial \dot{V}} \Big|_{\dot{V}=0} \dot{V} + O(\dot{V}^2),$$

where $p^* = p(V, T, 0)$ is the pressure at equilibrium]. One then reaches the conclusion that if the condition

$$\dot{V} \ll \frac{p^*}{\partial p / \partial \dot{V} \Big|_{\dot{V}=0}}$$

is satisfied, then indeed the difference between p and p^* is negligibly small as compared to p^* , and thus the process can be regarded as a quasi-static one.

Criticisms against the fuzzy notion of "quasi-static" have appeared in many other works. Truesdell & Bharatha (Preface p. xii), make the historical remark that "the 'quasi-static process' was barely mentioned for the first time in 1853 and was altogether foreign to the early work [in thermodynamics]". See also the mathematical analysis by Serrin: *On the elementary thermodynamics of quasi-static systems and other remarks*.

I also want to point out that "quasi-static" in some works has specific meanings somewhat unrelated to the discussion above. For example that the rate of increase of the total kinetic energy K of the system is negligible, so that the law of energy balance, which in its full generality is

$$\frac{d(U + K)}{dt} = Q + W$$

(that is, the rate of increase of internal energy U and kinetic energy is equal to the heat rate Q and work rate W provided to the system) can be approximated by

$$\frac{dU}{dt} = Q + W.$$

Or that similar inertial terms in the motion of the system are negligible. See for example the book by Day, chap. 2.

But note that such definitions of "quasi-static" have, again, *no* a-priori relation with reversibility.

Finally, the equation $dS = Q/T$ is only valid for a process that is:

- reversible (by definition),

- closed (no exchange of mass),
- with a homogeneous surface temperature,
- without bulk heating (such as instead happens in a microwave oven).

Under the last three conditions we have in general that $dS \geq Q/T$; when the equality sign is satisfied, then the process is *defined* as reversible. See Astarita, § 1.5, or Müller & Müller, for the different forms of the second law under different circumstances. This equation may be valid in quasi-static and non-quasi-static processes, as explained above.

$$\frac{dN(t)}{dt} - J(t) = 0$$

$$\frac{dN(t)}{dt} - J(t) = -\mathcal{A}(t)$$

$$\frac{dE(t)}{dt} - \Phi(t) = R(t)$$

$$\frac{d\mathbf{P}(t)}{dt} - \mathbf{F}(t) = \mathbf{G}(t)$$

$$\frac{d\mathbf{L}(t)}{dt} - \boldsymbol{\tau}(t) = \mathbf{M}(t)$$

$$\frac{dS(t)}{dt} - \frac{\Phi(t)}{T(t)} \geq 0$$