The Seven Wonders of the World

Notes on 21st-century physics

P.G.L. Porta Mana

Working draft version 0.2, updated 14 February 2025

pglpm.github.io/7wonders

Typeset with LaTeX

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/

Contents

Preface

> I don't know what's the matter with people: they don't learn by understanding; they learn by some other way – by rote, or something. Their knowledge is so fragile!
>
> Finally, I said that I couldn't see how anyone could be educated by this self-propagating system in which people pass exams, and teach others to pass exams, but nobody knows anything.
>
> R. P. Feynman 1989

These notes are aimed at undergraduate students in Physics and in Engineering programmes, but graduate students may also find them useful.

You are probably aware of the variety of physics branches, such as mechanics, thermodynamics, chemistry, electromagnetics, fluid mechanics, statics, nuclear physics, and many others. Maybe you are acquainted with some of them. Probably you are also aware of the existence of different physical theories, like Newtonian Mechanics, General Relativity, Quantum Theory, which give different explanations and formulae for the same physical phenomena. Some physical theories are said to be more exact or more approximate than others; Newtonian mechanics, for instance, is an approximation of General Relativity. Does this wild variety of branches and theories also mean a wild variety of principles, methods, mathematical formulae – which you'll have to learn if you want or need to study any particular branch or theory?

Yes, it does.

But there is also a core of very few principles that apply *universally* to every branch and every theory known today, be it mechanics or thermodynamics, Newtonian Mechanics or General Relativity. If you learn these principles, you'll be able to *immediately* work with, and understand,

at least the general features of *every* new physical discipline, phenomenon, or technology that you might meet.

The main goal of these notes is to make you acquainted with this core of few physical principles.

How few are these principles? Around *seven* (the exact number depends on how we arbitrarily group or separate them). This is the reason for the title of these notes. These seven principles are quite amazing for various reasons:

- They apply to *every* physical phenomenon, as already mentioned.
- Their meaning is very intuitive: each of them expresses a sort of budget.
- They are responsible for, so to speak, "driving the universe forward in time". More precisely, they are the basic principles that allow us to simulate and make predictions about physical phenomena.
- Their mathematical formulation is *exactly the same* in all our main approximate and exact theories, like Newtonian mechanics and General Relativity. This means that if you learn how to apply them to a tennis ball, then you are also able to apply them to a black hole.

These universal principles are obviously expressed mathematically. Their consequences and applications can be studied, to some degree, by using analytical methods; that is, by methods involving mathematical operations that we can do by hand. But their most fascinating and practical applications need numerical methods, that is, methods involving some programming and computer simulation.

In my opinion, if you learn how to apply these universal principles in simple simulations, then you also better understand their meaning and the way they work. For this reason these notes mainly take a computational approach. We shall learn how to implement these universal principles in simple computer code, and see the fun variety of physical phenomena they lead to. Programs that accompany these notes can be found at

`https://pglpm.github.io/7wonders/`

Thanks

I am deeply indebted to the students attending the courses ING 175 and ING 174 in 2024 at the Western Norway University of Applied Sciences

(HVL) in Bergen. They allowed me to experiment presenting physics as done in these notes, and gave invaluable feedback about what's difficult and what's easy, what's unclear and what's clear, with many suggestions for improvement, and also moral support with their curiosity and enthusiasm. In particular I'd like to thank, in alphabetic order: Amalie Solberg Magnussen Rege, Bodil Markhus, Carl Jakob Lis Sterner, Dag Åsmund Ørnes, Iver Thoresen Malme, Jonas Pytte, Josefine Björk Jarlesdottir Gjerde, Lars Magnus Birketveit, Leonard Rogardt Heldal, Miriam Osland, Nicolai Lindløkken, Nikolai Ringereide, Oda Skagsoset Kristiansen, Rebecca Sahlem, Rudi Nathaniel Stødle, Severin Johannessen, Thea Tyssøy Jensen, Tim August Birkeland Christiansen, Vegard Aa Albretsen, William Sæther. Mats Øinas deserves a special thank for his thorough feedback, questions, suggestions, and encouragement.

A heartfelt thank goes to Yu-Chung Wang for his assistance in teaching and his feedback and continuous moral support.

0 Overview

"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

The plan of these notes

After some very general remarks about physics, we recall the notions of **physical quantity**, unit, and physical dimension. This is really just a reminder; it's assumed that you are already familiar with these general notions.

We then survey two cardinal physical notions: **time** and **space**. We briefly examine their fascinating nature as we today understand it. The notion of **coordinate system** is also introduced.

Thereafter we take an overview of the main seven physical quantities we shall work with: **matter**, **energy-mass**, **momentum**, **angular momentum**, **entropy**, **electric charge**, **magnetic flux**, discussing some features common to all of them. Other important quantities such as **temperature** are also introduced.

The most important feature of the main seven quantities is that we can measure their amount within any volume, and the amount that passes through any surface. We therefore study the intuitive ideas of **control volume** and **volume content**, **control surface** and **flux**, and **supply**.

At this point the stage is ready for the introduction of seven physical laws which are just expressions of **balances** of volume contents, fluxes, and supplies of the seven quantities introduced earlier. These seven balance laws are universal. They are connected by mathematical expressions called **constitutive relations**, which are not universal but depend on the particular

physical phenomenon, and on the particular physical theory we use to describe it.

The balance laws are the ones that allow us to predict how the values of different quantities change with time. We show this by exploiting them to build simple computer code that can in principle be applied to any physical system. The role that constitutive relations have in connecting the seven balances also becomes quite clear in the code.

Five of the seven universal balances are then examined in turn. For each balance, we study some constitutive relations that are often used together with it, as well as some of its physical applications.

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 like *velocity*, *mass*, *force*, and similar ones.

Maths prerequisites

- Working familiarity with algebra, its operations and their properties.
- 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.

Programming prerequisites

- Knowing what a computer program is.
- Working familiarity with variables in a computer program.
- Working familiarity with `for`- and `while`-loops.

- Working familiarity with outputting and plotting the results of a computer program.

Structure of the text

👉 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, emphasized in blue.

› § 0 page 12

- 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.

👉 Side figures and quotes

Figures, 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.

👉 Cross-references

The text gives cross-references to the main section where a topic was discussed or will be discussed. Cross-references appear as section number and page listed on the margin, like this cross-reference to conservation laws.

› § 5.4 page 114

👉 Hyperlinks and bibliography

Some pieces of text are hyperlinks, like this one about One Punch Man[1]. You recognize them from their different colour and from the little footnote number that follows them. The links' URLs are also listed at the end of each chapter, in case you're reading a printed copy of these notes.

The text gives bibliographic references, like "Einstein 1905a", to scientific literature. The references are listed in the final Bibliography on page 283.

These references 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*.

"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)

👉 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.

URLs for chapter 0

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

1 Physics, quantities, units

> Philosophy is written in 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, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it.
>
> G. Galilei 1623

1.1 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.

1.2 What is "fundamental" physics?

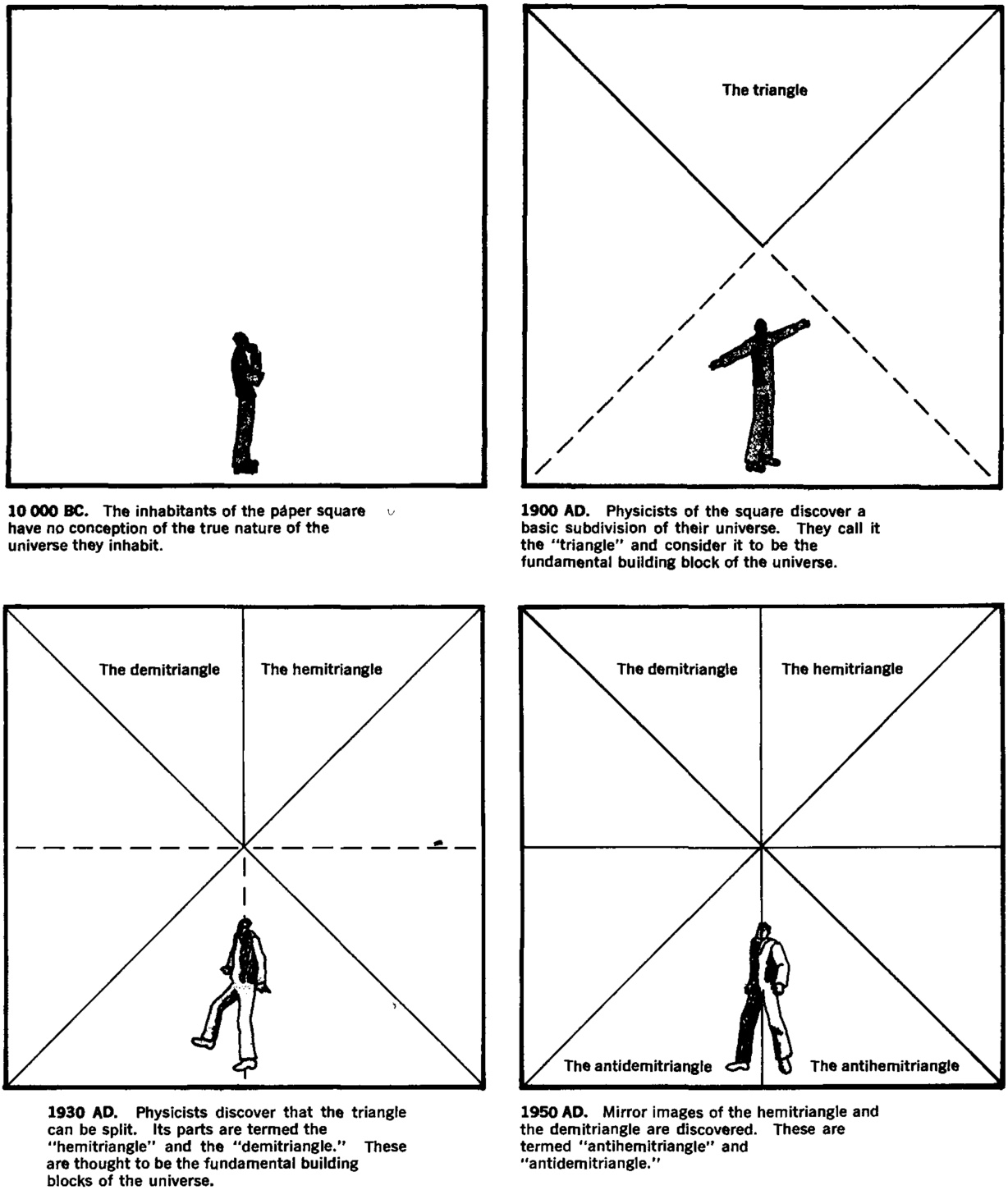
But what's the "ultimate" goal of physics? What's "fundamental" physics? The answer to this question is again subjective – also in this case physics lets you express your proclivities and personality. In the history of physics one can probably identify two main conceptions of "fundamental" physics.

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 Chew[1] 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 two kinds of physical laws that we'll draw in Chapter 5.



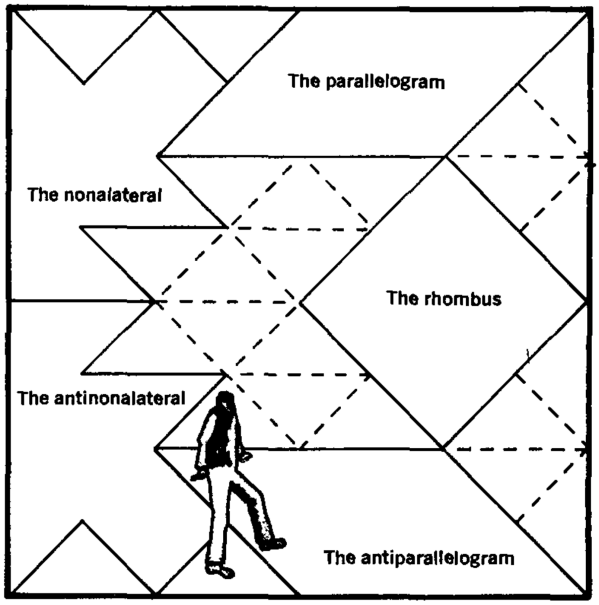
10 000 BC. The inhabitants of the paper square have no conception of the true nature of the universe they inhabit.

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.

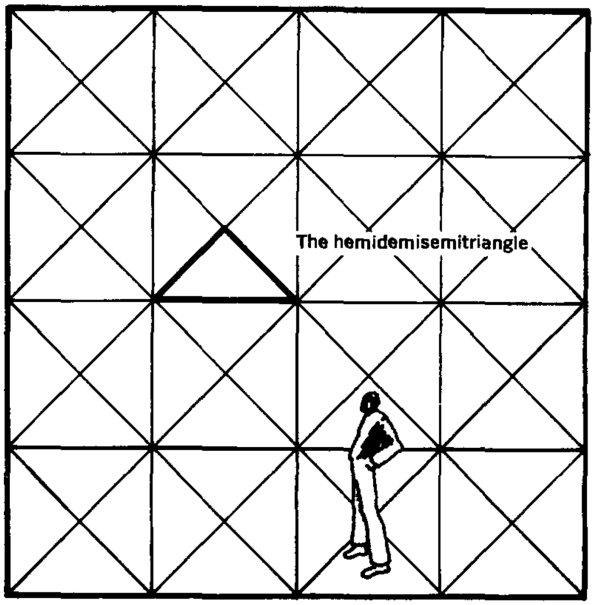
1930 AD. Physicists discover that the triangle can be split. Its parts are termed the "hemitriangle" and the "demitriangle." These are thought to be the fundamental building blocks of the universe.

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

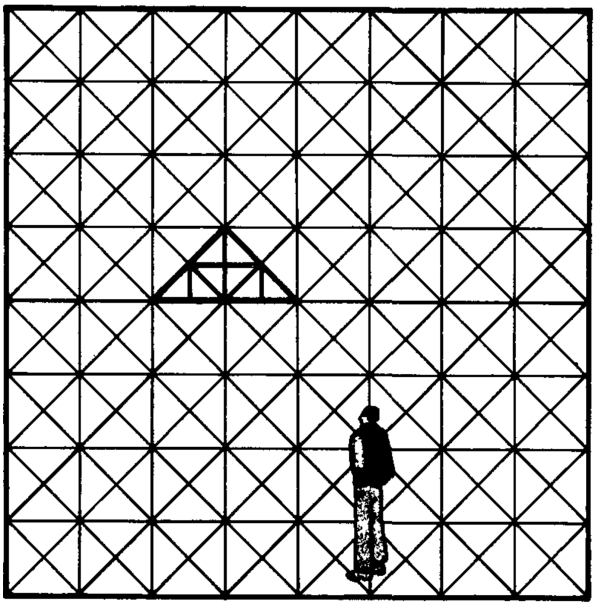
Figure 1.1 (Continues on p. 18) *The progress of "fundamental" physics*, from Chew 1970 as reproduced in Truesdell 1987



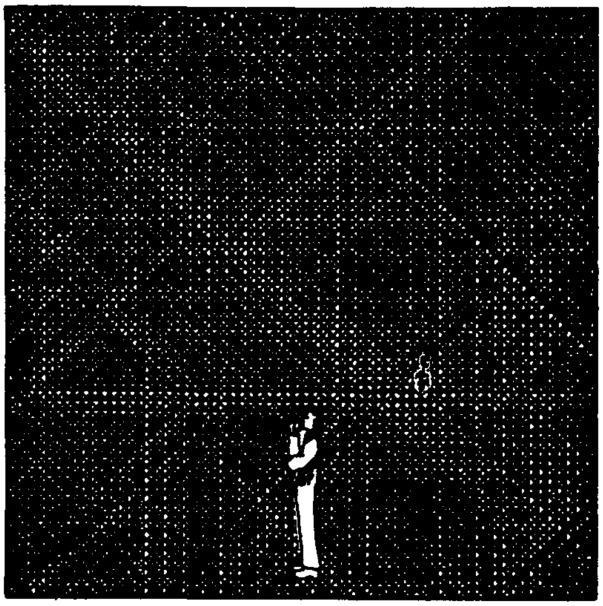
1960 AD. Physicists' conception of their universe is further clouded by new discoveries: the rhombus, the parallelogram, the antiparallelogram, the nonalateral and many others. It is unclear what these discoveries signify.



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.

1.3 Physical laws

What's a physical law?

Our lives rely on many kinds of patterns and regular experiences – so common that we take them for granted and don't even think about them most of the time. If you leave an object in your room in the morning, say a book or a pair of keys, and go out, then when you come back you expect the object to be exactly were you saw it, if you know that nobody entered your room in the meantime. While you go on the street you don't expect to suddenly levitate or flay away. While you write on your notepad, you're not afraid that it might suddenly disappear into thin air. The way we use everyday devices of all kinds relies on predictable behaviours and responses on their part.

Some things, we don't expect to be possible (image: *Flying Lesson* by S. & R. ParkeHarrison[2]).

All these patterns are regular behaviours that we observe and use are the essential core of physical laws.

A **physical law**, in the more technical meaning of this term, goes beyond the simple observation or use of such experiences in at least two ways:

First, a physical law goes from qualitative to **quantitative** observations and expressions. Instead of saying "the book is still on the table, where I left it", it might state "the mass-centre of the book was at position $(x, y, z) = (0.6, 3.1, 1.2)\,\mathrm{m}$ at times $t_0 =$ 07:50:31 and $t_1 =$ 17:14:40". The use of 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

$$60.369\,40,\ 5.3518\ .$$

"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

And in these two numbers all digits are important; for instance, the latitude is not 60.369 4**7**. The use of mathematics leads to amazing precision and predictive powers. It is this precision, which increases every day, that allows for the operation of technologies so important in our everyday lives: from mobile phones to aeroplanes, from laptops to cars.

Second, a physical law goes from very specific experiences to more **general** patterns, which may encompass seemingly very diverse phenomena. Instead of saying "this book is still here" it might say "this kind of

matter is conserved in any volume and for any amount of time". Instead of seeing the falling of an apple and the reciprocal movement of planets and sun as two different kinds of phenomena, it summarizes both together as examples of only one physical phenomenon. This generalization leads to the recognition and understanding of many new patterns.

In being more quantitative and general, physical laws often become more **abstract** than everyday observations. Objects like a book and a body of water get included into the more abstract notion of 'matter'. The movement of different kinds of objects and the propagation of light get included into the more abstract notion of 'momentum'.

To study and use physical laws we must therefore become familiar with and proficient in seeing and describing physical situations in a more abstract way, and to translate them into mathematical terms. This is what we shall learn in the next Chapters 2, 3, 4, so that in Chapter 5 we shall finally learn and start to use physical laws.

Many different mathematical formalisms

The mathematics used to express physical laws can come in wildly diverse forms, using wildly diverse principles. This leads to what we may call "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, we may jump among formalisms, because some physical situations 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. In "doing" physics you can actually 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.

$$\delta \int L \, \mathrm{d}t = 0 \qquad L = \frac{1}{2} m v^2$$

$$\boldsymbol{F} = \frac{\mathrm{d}}{\mathrm{d}t} m \boldsymbol{v} \qquad \boldsymbol{F} = 0$$

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

In these notes I choose 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 choose 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 use 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 as relations between them. 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.

1.4 Physical quantities

Generalities; scalar and vector quantities

As discussed in the previous section, in order to formulate and use physical laws we need to view all physical phenomena in and around us from a more abstract point of view, which more easily lends itself to quantification. This is done by using the concept of **physical quantity**.

The Joint Committee for Guides in Metrology (JCGM)[3] defines 'quantity' as follows[4]:

> property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed by means of a number and a reference.

As you see it's a somewhat vague definition, but it clearly refers to the possibility of quantification. Some physical quantities are connected with familiar concepts that we use every day, like time, position, distance, velocity, temperature, energy-mass – but we must be careful because their use in physics is more rigorous and often has technical features that the everyday familiar use has not. Other physical quantities, like momentum, magnetic flux, entropy, are more abstract and disconnected with everyday concepts. Their intuition and use therefore require practice and carefulness.

As stated in the definition above, all physical quantities are expressed as a number together with a *reference*, also called a **unit of measurement** or simply *unit*. The unit is a basic standard for comparing the measurement of a quantity. For example, when we say that a table has a *length* of two metres, written '2 m', we mean that is it as long as two of those standard lengths that we call 'metre'. We shall further discuss units of measurement in § 1.5. The number and unit together are handled by usual mathematical rules; so we can for example add or multiply physical quantities. Not all mathematical operations are allowed on all quantities, though.

Some physical quantities can be specified by giving only one number together with a unit. Such a quantity is called a **scalar**. Other physical quantities instead require the specification of several numbers, usually three, with associated units. Such a quantity is called a **vector**, and can be graphically represented by a vector. There are also quantities that must be specified by several numbers and units, collected together into a sort of matrix. Such a quantity is called a **tensor**.

> **What's scalar or vector or tensor depends on the theory**
>
> *Scalar*, *vector*, *tensor* have specific and slightly different meanings in different theories. It is therefore important that you don't take the classification used in these notes as universal.
>
> For example, in these notes and in Newtonian mechanics we call *energy density* a scalar, but in General Relativity it isn't a scalar: it is one component of a vector, or even of a tensor.

Derived and primitive quantities

We must briefly discuss a way of categorizing physical quantities which is important for understanding our approach to study physical laws. It's the distinction between *derived* and *primitive* quantities. This is an arbitrary distinction; that is, we can decide, within some bounds, which quantities we consider as 'derived' and which as 'primitive'.

A **derived quantity** is one that we decide to define in terms of other quantities. For example, we can define *speed* v (more precisely: average speed) as the ratio between a *distance* d and a *time duration* t:

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

where the symbol ":=" means "is defined as" or "is defined by". Note how we are already treating and representing 'speed', 'distance', 'time duration' as mathematical objects, and doing mathematical operations on them.

Treating *speed* as a derived quantity means that we could in principle avoid using the word 'speed' and the symbol 'v' altogether, and instead always speak about 'distance' and 'duration', always using the symbols 'd' and 't' in the combination 'd/v', instead of using v. Doing so would probably lead to very long sentences and formulae, and therefore be

extremely inconvenient; but in principle 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, 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 decide *not* to define in terms of other quantities. That it is not defined in terms of other quantities doesn't mean that we cannot try to explain it. But such explanation must be taken as informal and heuristic. Primitive quantities are often explained through metaphors or by appealing to intuition; but you must always be a little wary of such explanations, because they may fail you spectacularly in some situations. Primitive quantities are the building blocks from which we define all other quantities.

"you have to be in some framework that you allow something to be true. Otherwise you're perpetually asking 'why'". (see video[5]).

We have therefore a choice about which quantities to take as primitive and which to take as derived. For instance we can define *energy* in terms of quantities like *work* and *heat*, which in turn need to be defined in terms of others, or to be taken as primitive. Or we can take *energy* as primitive, and define *work* and *heat* in terms of it and of other quantities. The latter choice can be more convenient for stating a physical law and for developing a physical theory. It often happens that a quantity, if taken as primitive, is very convenient for building a theory, but difficult to understand intuitively. Vice versa, a quantity can be very intuitive, and therefore convenient as a primitive; but it leads to a complicated theory. We also have some choice in deciding *how many* quantities to take as primitive. Taking as few as possible primitive quantities can actually make the development of a physical theory more difficult, requiring very complicated definitions of the derived quantities.

In our study of physical laws we shall take these quantities as primitive:

time
space

matter
electric charge
magnetic flux
energy-mass
momentum
angular momentum
entropy

Time and *space* are taken as primitive quantities – for obvious reasons – in almost all current physical theories. We shall discuss them in Chapter 2. The other seven physical quantities, which we shall discuss in Chapter 3, are chosen because of several advantages:

- ✔ They can be understood physically, and treated mathematically, in a similar way. Therefore as we get familiar in thinking about and handling any one of them, we automatically get also familiar with all the others.
- ✔ The way in which they can be understood and handled is intuitive and lends itself to mental visualization.
- ✔ They lead to physical laws that have an almost identical expression. And, as mentioned before, this expression represent a sort of "budget" and is therefore intuitive.
- ✔ These laws are common to all our physical theories, and they can be expressed in a mathematical form that's the same in any theory.

The price to pay for the advantages above is that some of these quantities may be less familiar than others; but the advantages seem to outweigh this disadvantage.

We shall also use other quantities, some of which are very familiar, like *temperature* or *pressure*. But the quantities listed above are our fundamental building blocks.

1.5 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-energy of a football, compared to the ways we can measure the mass-energy 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 reflect the mathematical relationships that exist between physical quantities. For instance, we take *distance* to have dimension length, and *time interval* to have dimension *time*; if we now define *speed* as *distance* divided by a *time interval*, then the physical dimension of *speed*

is length/time. Physical dimensions help us to avoid doing mathematical operations that don't make sense with some quantities. For example, it doesn't make sense to sum up the volume of a glass of water with its temperature.

With each physical dimension we can associate a unit of measurement, which as mentioned in § 1.4 expresses a basic standard for comparing the measurement results for quantities having that physical dimension. 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 physical quantities and their 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/m^2", then you have a hint that it could be measured by measuring an energy-mass and an area, and then dividing them. Third, keeping track of units often allows us to quickly catch errors in solving a physical problem.

If a physical quantity is defined in terms of other quantities, then its unit is usually given in terms of the defining quantities. For example, if we define *speed* as *length* divided by *time*, then its unit is 'metres per second', 'm/s'. Some combinations of units receive special unit names. For instance, *power* is defined as energy-mass divided by time; its unit is therefore 'J/s' ('joules per second'). But this compound unit is usually called 'watt', symbol 'W'. In other words, 'one watt' and 'one joule per second' are the same:

$$1\ \mathrm{W} \equiv 1\ \mathrm{J/s}\ .$$

The topics of measurement, physical dimensions, and units, which are studied in *metrology* and in *dimensional analysis*, could occupy an entire course by themselves! I assume that you will look in the documents by the SI for more complicated details. We shall speak further about units in the next chapters. The main quantities and units used in the present notes are summarized in table 3.1 on page 73 and table 4.1 on page 77.

How to pronounce the quotient and the product of units

According to the rules of the SI, the quotient of two units is pronounced 'per' in English; for instance 'm/s' is pronounced 'metres per second'. The product of two units is simply pronounced by concatenating the

names of the units; for instance 'N · s' is pronounced 'newton seconds'. For other rules about printing and reporting units take a look at the NIST Check List[6].

1.6 Mathematics with quantities and units

Variables and functions

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 "$t = 120$ s". The unit is not predetermined, but it must correspond to the dimension of that quantity. We could for instance write "$t = 2$ 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 be simplified; for instance:

$$3\,\mathrm{mol/s} \cdot 5\,\mathrm{s} = 3\,\frac{\mathrm{mol}}{\cancel{\mathrm{s}}} \cdot 5\ \cancel{\mathrm{s}} = 15\,\mathrm{mol}\,.$$

Particular care must be taken with trigonometric and exponential functions, like 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 s)" somewhere: how much is that? If we say "just discard the unit", we would have

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

but wait: 60 s ≡ 1 min, so we could equivalently write "cos(1 min)". Then, according to the hypothetical rule "just discard the unit", we would have

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

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 $T = 1$ s then we can also simply write

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

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

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

Also remember that **the results of trigonometric and exponential functions are dimensionless numbers** as well, so an expression like '3 cos(...)' 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, $L = 2$ m we can also simply write

$$2\,\cos(\dots)\,\mathrm{m}$$

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

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

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

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

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:

› § 2.8 page 47

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big(2\ \cos(t/\mathrm{s})\,\mathrm{m}\Big) = 2\underbrace{\Big[-\sin(t/\mathrm{s})\cdot\frac{1}{\mathrm{s}}\Big]}_{\text{chain rule}}\,\mathrm{m} = -2\ \sin(t/\mathrm{s})\,\mathrm{m/s}$$

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

URLs for chapter 1

1. https://www.physics.lbl.gov/rememberinggeoffreychew/
2. https://www.parkeharrison.com/bodies-of-work/architect-s-brother/earth-elegies
3. Joint Committee for Guides in Metrology (JCGM)
4. https://jcgm.bipm.org/vim/en/1.1.html
5. https://www.youtube.com/watch?v=nYg6jzotiAc&t=893s
6. https://www.nist.gov/pml/special-publication-811/nist-guide-si-check-list-reviewing-manuscripts

2 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 and proper time

Time is a primitive quantity. We understand the notion of time intuitively, even if it is difficult to explain – that's why it is taken as primitive. In 1905, with the Theory of Relativity, part of our everyday intuition about the notion of time 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. GPS navigation, for example – which we use every day in leisure activities like hiking or sightseeing, as well as in more critical ones like aeroplane landing – crucially depends on the correct notion, intuition, and measurement of time. Luckily the new intuition of time is also becoming more and more widespread thanks to films and mass-media; think of movies like *Interstellar*[1].

Let's see, by means of a thought experiment, how our traditional intuition about time goes astray. Here's Alice, Bob, and Charlie. They have extremely precise clocks, built in exactly the same way. They synchronize

their clocks and stay very close to one another. Keeping close, they go around, maybe on an aeroplane or on a space ship, and they constantly compare their three 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, and each one goes around independently. One of them might stay in place, another might take a helicopter, and another might go for a trip to Mars and back.

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

First, both Alice and Bob can say "my clock has been working fire, so it should be correct": neither has noticed anything strange about the passage of time.

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

Third, they might wonder what the time is on Charlie's clock. But Charlie is at some distance away. They could decide to contact Charlie, say via radio, and ask "what does your clock show, 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 Charlie's answer would 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. But even neglecting these difficulties, Charlie's answer could reveal a time that is completely different from Alice's and from Bob's – it could be years ahead or behind both of theirs!

The experience just described will occur again any time Alice, Bob, Charlie, or any two of them, meet. There could be a hundred observers like Alice, Bob, Charlie, initially at the same place and with synchronized clocks. Whenever two or more of them meet after having been separated, they will notice discrepancies in their clocks (and in the ageing of their

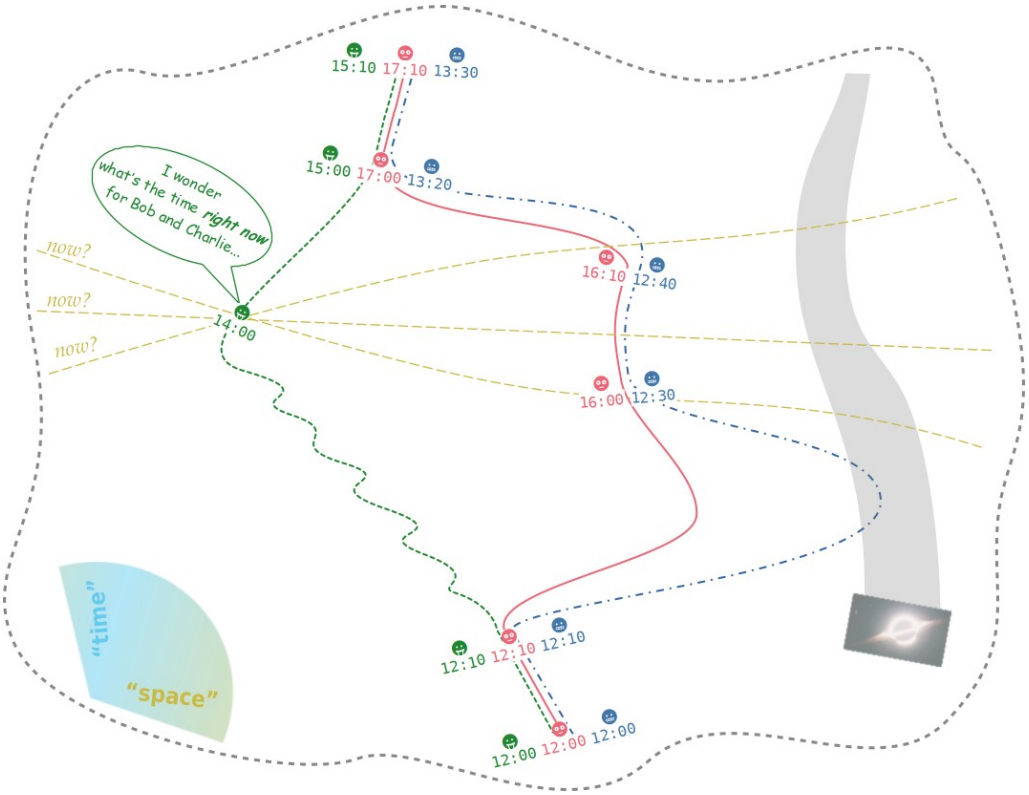


Figure 2.1 A *spacetime diagram* that illustrates of the experiences of Alice (dashed), Bob (solid), Charlie (dot-dashed) with time. The area within the grey dotted line represents a two-dimensional spacetime. Space is more horizontal than vertical; time is more vertical than horizontal, and flows upward. Note that we can't precisely say, for instance, "the vertical direction is time", because our problem is to see whether we can really separate time and space at all.

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

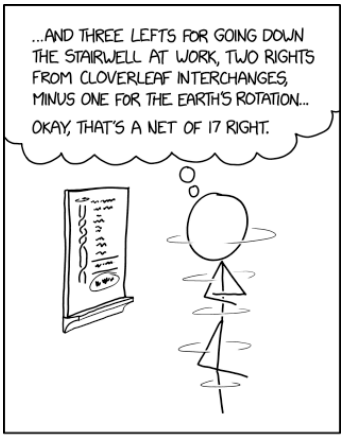
Right: Charlie visits a region near a strong mass-energy source. Upon meeting again with Bob, the two notice their clocks differ: 16:00 for Bob, 12:30 for Charlie. Yet this clock difference stays the same while they travel together for 10 min.

Left: Alice wanders around, travelling at high speed with respect to the fixed stars. When her clock displays 16:00, she wonders what's the time "right now" for Bob and Charlie. But 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) it is not clear which instant in Bob & Charlie's trajectory should be considered as "now" for Alice (yellow dashed lines).

Upper part: When Alice, Bob, Charlie meet again, their clocks have completely different readings; and they have aged differently. But their clocks run again at the same rate as long as they stay close again.

bodies). But their clocks will have exactly the same time lapses as long as they stay together.

This situation is illustrated in fig. 2.1, which is an example of **spacetime diagram**. The figure represents the temporal dimension and one spatial dimension merged together in a two-dimensional image. The motions of Alice, Bob, Charlie 'in space' are therefore represented as lines in this two-dimensional spacetime. They are called *worldlines*. Each point in a worldline has an associated time, the time measured by the person or object following that worldline.



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

https://xkcd.com/2882

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 were to meet again, their clocks would always show identical times. If one year, 23 days, 8 hours, 9 minutes, and 3.045 399 283 240 992 663 02 seconds had passed for Bob since he last met Alice, he would see that exactly the same amount of time had passed for Alice since their previous meeting. If you think about it, in this case it would have beeen somewhat natural for them to think "right now, the clock of far-away Charlie must show the same time as ours" (even though they have no real experimental way of confirming that).

But that's an imaginary world. In our world what occurs is the more complicated situation with time discrepancies described initially. 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 does not make sense to ask questions like "what is the time for far-away Charlie, *right now*?", or "what is happening at some other place *right now*?". There is no universal 'now'; the notion of *now* is local.

The time measured by a specific observer is called the **proper time** of that observer. Luckily we know how to calculate how much the proper times of separated observers will differ when the observers meet again. According to our current understanding, it turns out that the time differences depend, roughly speaking, on how fast the observers are moving with respect to one another and with respect to the distribution of energy-mass in the universe, and on how much energy-mass is contained in the regions they travel across. The general theory of relativity gives us the equations that determine the proper-time differences.

Time discrepancies can be measured, for example, by comparing

initially synchronized clocks that have been put in aeroplanes flying in different directions. The first measurement of this kind was made by Hafele and Keating in 1971. They synchronized four caesium atomic clocks with a reference clock, and then flew the four atomic clocks around the world on commercial jet flights, first eastward, then westward. At the end of the eastward trip, the clocks showed a time *behind* the reference one by around 6×10^{-8} s. At the end of the westward trip, their time was *ahead* the reference one by around 3×10^{-7} s. These measurements were in agreement with General Relativity's prediction, within experimental error.

Hafele, Keating, and their clocks aboard aeroplane (from *Time*, October 1971[2])

Most importantly, these time discrepancies affect everyday technologies such as the Global Positioning System. Formulae from General Relativity appear in your phone's GPS software; see for instance § 20.3.3.3.3 of the Interface Control Document `IS-GPS-200` at `https://www.gps.gov/technical/icwg/`. Time discrepancies must also be taken into account in the establishment and synchronization of time in our everyday equipment:

> 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)

Note therefore the following curious situation. In setting up a time to meet a friend, you don't need to worry about the discrepancies between your and your friend's proper times: if your friend walks 1000 m away from you and then immediately back to you, at 1 m/s, then the time elapsed for you will be 1000 s (around 17 min), but the time elapsed for your friend will be 999.999 999 999 999 994 437 s. That's a difference of less than 10^{-14} s, clearly negligible for the two of you, so you don't need to worry with General Relativity formulae in setting up your meeting time. Yet, if you set up a meeting place via GPS, then the true nature of time and General Relativity formulae become important: if they were not accounted for, you and your friend might end up off your meeting place by 100 m.

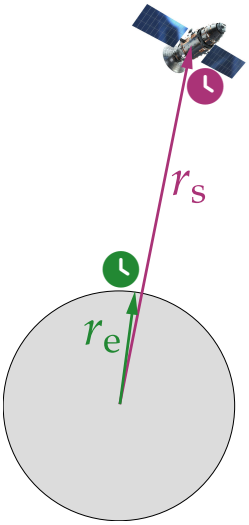
Time lapse or duration has SI dimension 'time', and we shall usually measure it using the unit *second*, symbol 's'.

Exercise 2.1

Consider two clocks: one at rest on the Earth's surface, at a distance r_e from the Earth's centre; the other on a GPS satellite or, say, on the International Space Station[3], 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 that the clock on the satellite has run for a time lapse Δt_s. The relation between two time lapses is approximately given by

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

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



1. Take the case of a GPS satellite, with $r_e = 6.4 \times 10^6$ m and $r_s = 2.6 \times 10^7$ m (NASA data[4]). If you, on the ground, measure a time lapse of $\Delta t_e = 10$ 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 can still be applied as an approximation.

   In the film *Interstellar*[5], two astronauts travel to Miller's planet, at a distance r_e from the black hole Gargantua, while leaving a third astronaut in orbit at a distance $r_s \approx \infty$ (the distance is large enough that it can be approximated as infinity). The two astronauts stay on Miller's planet for 3 hours. When they meet the latter astronaut in orbit again, the latter has aged *23 years*.

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

2.2 Coordinate time

The fact that the time elapsed for you can be different from that elapsed for a satellite can therefore make it difficult to coordinate activities and to operate some technologies.

But luckily there's a way to bypass proper-time discrepancies. Instead of referring to my proper time or to your proper time, we can agree on assigning a somewhat arbitrary numerical time label to every event: this is called a **coordinate time**.

Coordinate time is generally different from the proper times measured by different observers. It can nevertheless be used for "doing physics", and it is the time we shall most often use in our equations. The coordinate time commonly used for Earth-physics purposes is Universal Coordinated Time UTC[6], or the International Atomic Time TAI[7] for astronomy purposes:

> 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 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)

UTC and TAI have the same time lapse, but UTC differ by irregular readjustments of its "zero", the leap seconds[8].

The clock on your phone, and on devices 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 lapse equal to UTC or TAI (besides small variations coming from the irregularity and internal motions of the Earth). But observers at other altitudes and observers moving with respect to Earth's surface can measure that their proper-times lapses are slightly different from UTC and TAI.

In applications related to interplanetary spacecraft navigation and cosmological observations, the Barycentric Coordinate Time TCB is used. Its time lapse is slightly faster than the one of UTC: for each second that passes in UTC, 1.000 000 014 8 s pass in TCB on average. Two coordinate times to be used around an on the Moon, Lunar Coordinate Time TCL[9] and Lunar Time LT, are on the making in the year 2025. Lunar Time will lapse

faster by 5.6×10^{-5} s per day compared to our UTC. The establishment of all these kinds of coordinate times shows how important the difference between coordinate and proper time is for our current technology.

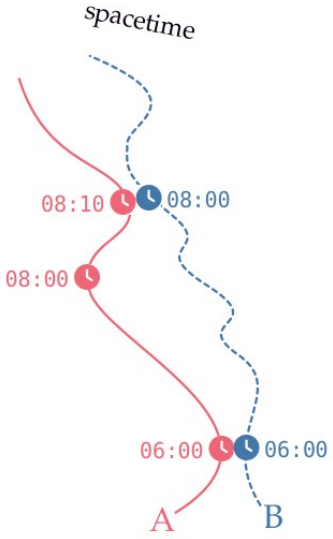
When we use coordinate time, some important physics formulae turn out to be the same no matter whether we use General Relativity or an approximate theory such as Newtonian Mechanics. Thanks to this fact, for the most part of these notes we will not need to deal with proper-time details. But it is important for you to keep in mind how time really works, and the small time discrepancies that exist and occur all the time along your *worldline*.

2.3 Space, length, distance

Together with the notion of time, also the notions of space, length, distance lose some of their traditional intuition. Traditionally when we speak of the distance or the length of a moving object at a given time, we mean the distance 'at the same instant of time'. But we have seen that it does not make sense to ask "what is the time for the object, right now?". And when we speak of lengths or distances, we mean measurements 'on a straight line'. But spacetime is curved.

To see the complications in defining 'distance' and 'length', imagine two objects or people that are moving with respect to each other, getting closer and farther away. Call them A and B; let's say that A represents you. You have a clock, and B is equipped with another clock identical to yours. These clocks were synchronized (to time 06:00) in the past when you and B were in contact with each another.

We can represent the motions of A and B as *worldlines* in spacetime, as in the side figure (similarly to what we did in fig. 2.1). The figure shows the points in spacetime where your clock displays the times 06:00, 08:00, and 08:10, and where B's clock displays 06:00 and 08:00. Note in particular that you and B touch each other and are synchronized at 06:00; then you touch each other again later, when your clock displays 08:10 and B's clock displays 08:00.



Suppose that when your clock displays time 08:00 you ask "how distant is B from me *right now*?". How can we measure or define such distance?

One possibility is to say that we measure the length of a straight line joining you at *your* time 08:00, with B at *its* time 08:00. Could we define this as the distance between you and B at 08:00? Well, there's something

bizarre: when B's clock shows 08:00, B is in contact with you! So we would essentially be measuring the distance between you and yourself at two different times of your clock. This doesn't seem to be very sensible. Also, should this distance be zero, then? But when your clock displayed 08:00, B was *not* in contact with you, so it wouldn't make sense to say that its distance from you was zero.

So let's discard the measurement procedure above, which seems to inconsistent results. Another possibility is to measure the length of a straight line joining you at your time 08:00, with B at some other time on its clock. But which time should we choose? Any choice seems arbitrary.

And there's one more problem. Let's say that we decide to measure the distance between you when your clock shows 08:00, and B when its clock shows 07:50 or some other arbitrary time when it is not in contact with you. We should measure this distance on a straight line. But what's a 'straight line'? Spacetime is curved. A straight line in the figure above is not necessarily a straight line in spacetime. The notion closest to a 'straight line' in a curved space is that of a *geodesic*[10]. It turns out that there may be several geodesics connecting you (at 08:00) and B (at 07:50).

Similar problems appear if we ask questions about *lengths*. Suppose there is a rubber band stretched between you (remember you're A) and B. When your clock shows 08:00 you ask "how long, *right now*, is the rubber band joining me and B?" We encounter the same difficulties as above in trying to answer this question.

A spectacular case of the effect of spacetime curvature is the phenomenon of gravitational lensing[13]. Owing to the curvature of spacetime, light and other electromagnetic radiation emanating from the object reach us along different paths in spacetime. All these paths are "straight lines" (geodesics). Coming from different directions, to us these look like distorted, duplicated images of the object, as schematized in the top side figure. The curvature is generated by some large distribution of energy-mass, such as a galaxy, between us and the object. A beautiful example of this phenomenon is given by the quasar RX J1131-1231[14], bottom side image. Spacetime curvature separates the light arriving to us from this quasar into four yellow & red spots, three at the top and one at the bottom in the image. The curvature is generated by a galaxy visible as the blue spot at the centre.

From *The ABC's of Distances*[11].

From ESA/Webb[12].

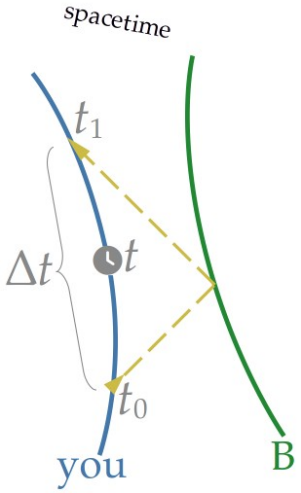
A consequence of these peculiar situations and of the curvature of spacetime is that many notions of distances and length can be defined and

measured, which are generally *not* equivalent to one another. Cosmology, for example, uses a plethora of different distances[15]. One must therefore be careful about which definition of distance is being used. In the next sections we shall focus on two of them: *radar distance* and *coordinate distance*.

2.4 Radar distance

The definition of distance that's regarded as the most "physical" is *radar distance*, defined as follows.

Consider again a situation in which an object B is moving with respect to you. Yours and B's motions in spacetime, around the time when your clock displays time t, are illustrated in the next side figure. Your *worldline* in spacetime is the blue one on the left; the worldline of object B is the green one on the right. At your proper time t_0 you send a light pulse towards object B. The pulse travels in empty space, and upon hitting object B it immediately bounces back to you. It reaches you at your proper time t_1. The worldline of the light pulse is shown in dashed yellow in the figure.



A proper time $\Delta t = t_1 - t_0$ has elapsed for you between emission and reception of the light pulse, and the time exactly in between emission and reception is $t = (t_1 + t_0)/2$. The **radar distance** d of object B from you at time t is defined as

$$d := \tfrac{1}{2} c\, \Delta t \tag{2.1}$$

where c is the speed of light in vacuum, a universal physical constant:

$$c = 299\,792\,458 \text{ m/s} \quad \text{(exactly)}. \tag{2.2}$$

The SI unit of length, the metre[16], is based on the measuring procedure above. Common laser distance meters also work by the same procedure, and therefore yield radar distance. As the name indicates, this is also the distance measured by radars.

A laser distance meter (the light beam is not visible in reality).

In using radar distance, however, we must be wary of the following peculiarities:

- Radar distance makes sense only if the time lapse Δt is small enough compared to variations in the relative motion of the observer and the object, so that this motion is approximately uniform. For this reason this distance cannot be used if the object is too far away: the farther away it is, the longer it takes for a light beam to travel to and fro. Radar distance can be used between the Earth and other Solar System planets; but it cannot be used for galaxies or other distant cosmological objects.

- **Radar distance is not symmetric**: the radar distance of B from A at A's time t is generally different from the radar distance of A from B at B's time t.

- **The value of radar distance depends on the relative motion** between A and B. Imagine that a friend of yours is located very close to you at time t, but is moving with respect to you. Upon measuring object B's radar distance, your friend will generally find a value different from yours. The discrepancy between you and your friend's measured values will be the larger, the higher is the relative velocity between you two. Several observers in motion with respect to one another will generally disagree on the dimensions of an approximately rigid objects in their vicinity.

The dependence on relative motion also affects, at high speeds, how we *see* objects, which appears more and more deformed. You can find beautiful visualizations, both static and animated, at *Relativity visualized*[18].

How a street in Tübingen would look like (except for colour and some other features) if we travelled through it at around 240 000 000 m/s (from *Relativity visualized*[17])

Luckily, for relative speeds that are not too high compared to the speed of light, the radar distances measured by different observers differ by amounts that are negligible in everyday circumstances. As an example, consider a car moving on a road at 100 km/h, that is around 28 m/s. The car's driver measures the length of the car to be 4 m by radar distance. A pedestrian that sees the car passing by instead measures its length to be 3.999 999 999 999 98 m by radar distance. This is obviously a negligible difference.

Length and distance have SI dimension 'length', and we shall usually measure them using the unit *metre*, symbol 'm'.

Exercise 2.2

Imagine that you and a friend of yours are measuring your distance from a wall, using a laser distance meter each. You and the wall are static with respect to Earth's surface. Your friend is moving with a speed v towards the wall, and is right beside you at the exact moment of the measurement.

In this specific situation, if d_{you} is the distance measured by you, and d_{friend} the distance measured by your friend, the two are related by the formula

$$d_{\text{friend}} = d_{\text{you}} \cdot \sqrt{1 - v^2/c^2} \,,$$

where c is the speed of light, given in eq. (2.2). Note that this formula

is also valid if your friend is moving away from the wall, rather than towards it.

1. Suppose you find that the distance of the wall from you is 200 m. Your friend's speed is 300 m/s. How much is the distance from the wall to your friend, who's right beside you, as measured by your friend? (You'll need a high-precision calculator and 18 significant digits.)
2. Now instead suppose you find that the distance of the wall from you is 500 m. Your friend measured (when right beside you) a distance of 499 m. How fast was your friend moving?

2.5 Coordinate systems

"Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Minkowski 1908

From our discussion about time and space we conclude that physical events happen in spacetime, and there is no unique way to attribute a universal time or a universal position in space to a physical event.

In the previous sections we used the word 'event', informally taking its meaning for granted. Let's be more precise now. We call **event** or **spacetime point** a very small region of space that lasts for a very short lapse of time, so that it can be considered as a point in a four-dimensional space. When we say 'small region' or 'short time lapse', it doesn't matter which definition of distance or time lapse we're using.

The word 'event' is used because typically we identify such a spacetime point by means of a physical phenomenon of limited spatial extension and duration. How "limited" should these extension and duration be? It depends on the kind of physical phenomenon we're interested in. The sudden burst of a soap bubble can be considered as an event in comparison to geological distances and times; but it cannot be considered as an event if we're studying subatomic particles.

The peculiarities of spacetime can make it difficult to communicate the positions of objects and events by relying on distances. In giving indications about the location of a shop we can say "it's 200 metres down the road" without ambiguity. But in situations where much higher precision is needed and extreme motions or gravitational fields are involved, we would need to know the velocity of the person we're talking to, because the distances measured by us and by that person could be very different.

In the case of time, we bypassed the problem that time lapse depends on the observer's motion by introducing a coordinate time. This way each event gets an arbitrary but agreed-upon time label. We can bypass the analogous problem that length and distance depend on the observer's motion, by introducing a set of arbitrary *spatial coordinates* for each coordinate time.

› § 2.2 page 36

All these coordinates together form a *coordinate system*, also called *reference system*:

Coordinate system or reference system

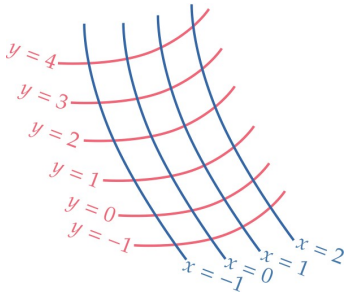
A **coordinate system**, also called **reference system**, is the assignment, by agreement, of four numerical labels to every point in spacetime: one *coordinate time* and three *spatial coordinates*. We use symbols such as (t, x, y, z), or (t, r, θ, ϕ), or others, for these coordinates. These four coordinates are obviously the same for all observers, because they are decided by agreement.

A coordinate system is usually defined on a limited region of spacetime.

The location where all spatial coordinates have value zero is called the **origin** of the spatial coordinates.

Often the coordinates have physical meaning – like 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. Typically we use three spatial coordinates. In special situations, such as locating points on the Earth's surface neglecting altitude, only two or even one spatial coordinate can be enough.

A coordinate system can be visualized as a grid made by a set of lines or planes, one set for each coordinate, which allow us to read the coordinates of any point. The side figure shows an example with spatial coordinates (x, y) in two dimensions, for a specific coordinate time. It is of course assumed that the grid can be refined as much as needed. We are all familiar with the coordinate system (λ, ϕ) of *latitude* and *longitude* to identify locations on Earth's surface, and used internally by the location systems of mobile phones.



Since a coordinate system is arbitrary, we often choose one adapted to the physical phenomenon under study. It's very common to choose a coordinate system (t, x, y, z) that has time coordinate $t = 0$ s at the beginning of our observation of the phenomenon, and spatial coordinates $(x, y, z) = (0, 0, 0)$ m at a location close to where the phenomenon happens.

The spatial coordinates may be chosen so as to have particular physical properties, which in turn may lead to simpler expressions for some physical laws as we shall see later. For instance, the coordinate lines of might be straight lines (geodesics), in which case we speak of *rectilinear coordinates*. If they are not, then we call them *curvilinear coordinates*; the coordinates in the previous side figure are curvilinear.

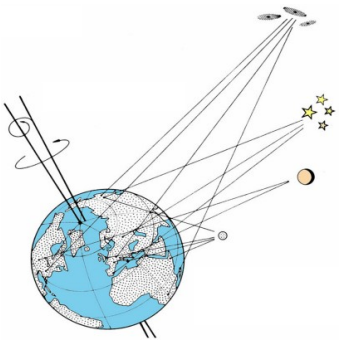
› § 5.9 page 135

If the spatial coordinate lines are *orthogonal* to one another at every point, that is, they intersect at $\pi/2\,\mathrm{rad}$, then we call them **orthogonal coordinates**. This is a very useful property, which simplifies many physics formulae. In these notes we shall always use orthogonal coordinates. Note that a coordinate system can be curvilinear *and* orthogonal.

In astronomy, in space and satellite communication, and in geodesy[19] (the science of accurately measuring and understanding the Earth's geometric shape, orientation in space, and gravity field), several important coordinate systems[20] are used. For example:

- The *International Celestial Reference System* (ICRS) is a coordinate system with orthogonal spatial coordinates, which are almost rectilinear. Their origin is close to the centre of the Sun. Several distant cosmological objects, like quasars[21], have fixed spatial coordinates in this reference system. Its coordinate time is called Barycentric Coordinate Time (TCB).
- The *Geocentric Celestial Reference System* (CGRS) is a coordinate system with orthogonal spatial coordinates, which are almost rectilinear. Their origin is close to the centre of the Earth. Distant cosmological objects have almost constant spatial coordinates in this reference system; this means that the Earth rotates with respect to it. Its coordinate time is called Geocentric Coordinate Time (TGB), which is very similar to the TAI.
- The *International Terrestrial Reference System* (ITRS) is a coordinate system similar to the CGRS, but with the important difference that the Earth is approximately static in this reference system; this means that distant cosmological objects rotate with respect to it. It has the same coordinate time as the ICRS.
- The *World Geodetic System 1984* (WGS 84)[22] is very similar to the ITRS, besides a discrepancy of some centimetres. It is used by the system of GPS satellites.

But how do we determine the position and time of an event in these coordinate systems? The procedure can be extremely complicated in fact. The starting point is the assignment of some predefined coordinates to objects that seem to have fixed spatial coordinates in the coordinate system; for instance distant cosmological objects like quasars[23] in the case of the ICRS and CGRS, or particular reference stations[24] on Earth's surface in the case of the ITRS and WGS 84. The coordinate of other events are then calculated from measurements of proper times, radar distances, and angles from the reference objects, using formulae from general relativity. As we mentioned in § 2.1, even your mobile phone participates in these complex calculations.



The assignment of coordinates on and around Earth depends on nominal values assigned to distant astronomical objects (from Capitaine 2010)

Exercise 2.3

On Earth's surface we often use the system of two coordinates called *latitude* λ and *longitude* ϕ, measured in degrees. Coordinate lines of constant latitude are called *parallels*; those of constant longitude are called *meridians*.

1. Find what's at these three coordinate pairs:

$$(\lambda, \phi) = (60.369\,002^\circ, 5.350\,336^\circ)$$
$$(\lambda, \phi) = (35.658\,587^\circ, 139.745\,424^\circ)$$
$$(\lambda, \phi) = (-13.163\,069^\circ, -72.545\,265^\circ)$$

2. Are latitude and longitude *orthogonal* coordinates? Explain why or why not.

2.6 Spatial coordinate distance and length

If we have chosen a coordinate system, we can define a notion of distance called *coordinate distance* between two points at any coordinate time t. The idea is simple: we measure the length of the shortest path in spacetime joining the two points, and all intermediate points on the path must have the same coordinate time t. Of course it can happen that there is more than one path having shortest length.

Coordinate distance is different from radar distance. It has two advantages: it doesn't depend on relative motion, and can be defined also between objects that are very far apart. In regions of spacetime with

low curvature and slow relative motions, however, radar distance and coordinate distance based on standard coordinate times are approximately the same, and approximately independent of any motions.

Note also that the speed of light defined with respect to coordinate distance need not have the value c, or even be constant! You might have heard or read about distant cosmological objects, like quasars, that are said to be receding from us at speeds faster than light's. How is that possible? The reason is that the 'speeds' they're talking about are defined with respect to coordinate distance, not radar distance.

Cartesian coordinates

We call **Cartesian coordinates** a set of spatial coordinates (x, y, z) with a very special property: the coordinate distance between two locations A and B having spatial coordinates (x_A, y_A, z_A) and (x_B, y_B, z_B) is simply given by

$$d_{AB} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}\,. \tag{2.3}$$

Perfect Cartesian coordinates do not exist, because spacetime is curved. But it is possible to choose coordinates that are approximately Cartesian in limited regions of spacetime. The coordinate systems ICRS, GCRS, ITRS discussed in § 2.5 are *not* Cartesian: they include the effects of curvature generated by the Earth and other bodies in the Solar System.

In most of these notes we will not have to worry about the discrepancies between radar distance and coordinate distance, and about the motion of the observer or instrument that is measuring a distance. So we shall use the term 'distance' without specifications. And we shall often use Cartesian coordinates.

Exercise 2.4

In the Geocentric Celestial Reference System, defined by the International Astronomical Union, the distance between two very close locations A and B having spatial coordinates (x_A, y_A, z_A) and (x_B, y_B, z_B) close to Earth's surface is approximately given by

$$d_{GCRS,AB} = \left(1 + gR/c^2\right)\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2} \tag{2.4}$$

where $g \approx 9.8\,\mathrm{m/s^2}$ is the gravitational accelaration, $R \approx 6371 \times 10^3\,\mathrm{m}$ is

Earth's radius, and c is the speed of light. This is also the radar distance of B from A, if A is not moving with respect to the coordinate system.

1. Verify that the formula above is dimensionally correct: both left and right side should have dimension length.
2. Assume that your coordinates are (x_A, y_A, z_A), and consider an object B at a x-coordinate difference of 100 m from you, that is, with coordinates

$$x_B = x_A + 100\,\text{m}\,, \qquad y_B = y_A\,, \qquad z_B = z_A\,.$$

How much is the difference between the distances d_{AB} and $d_{GCRS,AB}$, calculated between you and the object?

(The inaccuracy in the specification of g and R is actually much larger than the difference you just found.)

2.7 Coordinate position

Let us agree on some notation and terminology that will be used in these notes.

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

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

Unless stated otherwise, the coordinate time t will be taken to be UTC, and the spatial coordinates (x, y, r) will be taken to be Cartesian. As mentioned in the previous section, the definition of the spatial coordinates is usually different from problem to problem; so it is always important to specify how the coordinate system you're using is defined.

› § 2.2 page 36
› § 2.6 page 45

Position or location vector

The triplet of spatial coordinates is called the **position** or **location** vector and is often denoted by $\boldsymbol{r}$:

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

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

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; think for instance of a small falling object. Other phenomena happen approximately on a surface, in two dimensions; think for instance of a swinging pendulum. In these cases we can omit two or one of the spatial coordinates, and simply assume that the omitted ones have some constant, unimportant values. In such cases we can simply write, for instance, (t, x) or (t, x, y) as our coordinates.

2.8 Coordinate velocity and acceleration

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

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

So $\boldsymbol{r}$ 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*:

Coordinate velocity

The **coordinate velocity** is a vector defined as the derivative of the position vector $\boldsymbol{r}(t)$ with respect to coordinate time t:

$$\boldsymbol{v}(t) \coloneqq \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{r}(t) \quad \text{or} \quad \begin{bmatrix} v_x(t) \\ v_y(t) \\ v_z(t) \end{bmatrix} \coloneqq \begin{bmatrix} \dfrac{\mathrm{d}}{\mathrm{d}t}x(t) \\ \dfrac{\mathrm{d}}{\mathrm{d}t}y(t) \\ \dfrac{\mathrm{d}}{\mathrm{d}t}z(t) \end{bmatrix}\,.$$

The word **speed** means the *magnitude* of the velocity:

$$|\boldsymbol{v}| \equiv \sqrt{{v_x}^2 + {v_y}^2 + {v_z}^2}\,.$$

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

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

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, so we shall simply use the word 'velocity'. But in situations involving subatomic particles at high speed, for example, one must take into account that the two velocities are different.

Taking the derivative of the velocity with respect to coordinate time, that is, the second derivative of position, we obtain the *coordinate acceleration*:

Coordinate acceleration

The **coordinate acceleration** is a vector defined as the derivative of the velocity vector $\boldsymbol{v}(t)$ with respect to coordinate time t:

$$\boldsymbol{a}(t) \coloneqq \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{r}(t) = \left[\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t)\,,\ \frac{\mathrm{d}^2}{\mathrm{d}t^2}y(t)\,,\ \frac{\mathrm{d}^2}{\mathrm{d}t^2}z(t)\right].$$

Acceleration in relativity theory

In relativity theory, acceleration has a special physical significance because it includes the effect of gravity. We distinguish between *physical acceleration* and *coordinate acceleration*. The calculation of physical acceleration requires more than 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 you, y to your left, and z upward. Then your coordinate velocity is $\boldsymbol{v} = (0, 0, 0)\,\mathrm{m/s}$, also according to relativity theory. But your spatial *physical* acceleration is approximately $(0, 0, 9.8)\,\mathrm{m/s^2}$, not zero!

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

Integration of velocity and acceleration

Since velocity is the time derivative of position, we can find the position from the velocity by the inverse operation of integration. So we have

$$\boldsymbol{r}(t) = \boldsymbol{r}(t_0) + \int_{t_0}^{t} \boldsymbol{v}(t)\,\mathrm{d}t \qquad \text{or} \qquad \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} x(t_0) \\ y(t_0) \\ z(t_0) \end{bmatrix} + \begin{bmatrix} \int_{t_0}^{t} v_x(t)\,\mathrm{d}t \\ \int_{t_0}^{t} v_y(t)\,\mathrm{d}t \\ \int_{t_0}^{t} v_z(t)\,\mathrm{d}t \end{bmatrix}.$$

Note from the formulae above that if we know the expression of the velocity $\boldsymbol{v}(t)$, we still cannot know the position $\boldsymbol{r}(t)$ unless we *also* know the value $\boldsymbol{r}(t_0)$ of the position at some particular time t_0.

In a analogous way we can find the velocity $\boldsymbol{v}(t)$ from the acceleration $\boldsymbol{a}(t)$ by integration, provided we also know the value $\boldsymbol{v}(t_0)$ of the velocity at some particular time t_0:

$$\boldsymbol{v}(t) = \boldsymbol{v}(t_0) + \int_{t_0}^{t} \boldsymbol{a}(t)\,\mathrm{d}t \qquad \text{or} \qquad \begin{bmatrix} v_x(t) \\ v_y(t) \\ v_z(t) \end{bmatrix} = \begin{bmatrix} v_x(t_0) \\ v_y(t_0) \\ v_z(t_0) \end{bmatrix} + \begin{bmatrix} \int_{t_0}^{t} a_x(t)\,\mathrm{d}t \\ \int_{t_0}^{t} a_y(t)\,\mathrm{d}t \\ \int_{t_0}^{t} a_z(t)\,\mathrm{d}t \end{bmatrix}.$$

Exercise 2.5

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' in such a way that the expression is dimensionally correct:

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

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

$$\boldsymbol{r}(t) = \begin{bmatrix} 2.0 \times 10^7 \cos\left(\frac{t}{13\,751\,\mathrm{s}}\right) \\ 2.2 \times 10^7 \sin\left(\frac{t}{13\,751\,\mathrm{s}}\right) \\ 0 \end{bmatrix} \mathrm{m}$$

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

4. Suppose you know the expression of the *speed* $v(t) \equiv |\boldsymbol{v}(t)|$ of an object, and also the object's position $\boldsymbol{r}(t_0)$ at time t_0. From this information, can you find the position $\boldsymbol{r}(t)$ at any other time t?

2.9 Angles

To be written.

URLs for chapter 2

1. https://www.imdb.com/title/tt0816692/
2. https://time.com/vault/issue/1971-10-18/page/93
3. https://www.nasa.gov/international-space-station/
4. https://www.nasa.gov/directorates/somd/space-communications-navigation-program/gps/
5. https://www.imdb.com/title/tt0816692/
6. https://www.nist.gov/pml/time-and-frequency-division/time-realization/utc nist-time-scale-0/
7. https://gssc.esa.int/navipedia/index.php?title=Atomic_Time
8. https://webtai.bipm.org/ftp/pub/tai/Circular-T/cirthtm/cirt.442.html
9. https://doi.org/10.1103/Physics.17.140
10. https://mathworld.wolfram.com/Geodesic.html
11. https://www.astro.ucla.edu/~wright/distance.htm
12. https://esawebb.org/images/potm2406a
13. https://science.nasa.gov/mission/hubble/science/science-behind-the-discoveries/hubble-gravitational-lenses
14. https://www.nasa.gov/image-article/distant-quasar-rx-j1131
15. https://doi.org/10.48550/arXiv.astro-ph/9905116
16. https://www.nist.gov/si-redefinition/meter
17. https://www.spacetimetravel.org/
18. https://www.spacetimetravel.org/
19. https://oceanservice.noaa.gov/geodesy
20. https://gssc.esa.int/navipedia/index.php?title=Reference_Systems_and_Frames
21. https://esahubble.org/wordbank/quasar
22. https://earth-info.nga.mil/?dir=wgs84&action=wgs84
23. https://esahubble.org/wordbank/quasar
24. https://itrf.ign.fr/en/solutions/ITRF2020-u2023#frame-definition

3 Main physical quantities

> For Euler, clarity was the hallmark of truth. [. . .] To him we owe also the brilliant imagination of the internal pressure in generality [. . .] 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

The discovery, formulation, and use of physical laws requires us to look at the world from a more abstract and quantifiable point of view. As previously discussed, we shall achieve this by interpreting all physical phenomena around and within us in terms of time & space, which we have studied in Chapter 2, and of seven physical quantities:

› § 1.3 page 19

› § 1.4 page 21

Seven primitive quantities

matter	**energy-mass**
electric charge	**momentum**
magnetic flux	**angular momentum**
	entropy

We take these seven quantities as primitive, and shall build our physical laws upon them. Each of these seven quantities satisfies a universal physical law; other physical laws express relationships among these quantities.

Recall that primitive quantities cannot be defined: we can only try to understand them intuitively. This is the goal of present chapter: to make a brief acquaintance with the seven primitive quantities and with some

› § 1.4 page 21

3.2 Two basic properties

Our seven primitive quantities have two basic properties in common:

First property: three measurements

For each quantity – with a slight change for the magnetic flux – we can ask about or measure three kinds of amount:

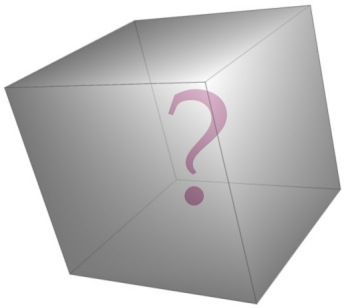
Three basic measurements of the seven quantities

M1. How much of this quantity is **contained** in a particular three-dimensional region of space at a particular time instant?

M2. How much of this quantity **flows** through a particular two-dimensional surface, in a given direction, during a particular time lapse?

M3. How much of this quantity is **produced** in a particular three-dimensional region of space during a particular time lapse?

t = 19:32:15



For six main quantities we can ask: how much of it is in a given volume, at a given time?

We can ask these questions for any region of space, any time instant, any time lapse. The regions of space can be moving and deforming. The results of the three measurements above are scalars for scalar quantities, and vectors for vector quantities.

Questions M2. and M3. can also be asked in a different way. Consider a very short lapse of time, and divide the net flow and the net production by that lapse of time. This way we have an alternative form of the second and third measurements, as flow or production divided by time:

Second and third measurement: new version

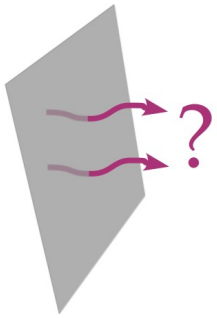
M2b. At which rate is this quantity flowing through a particular two-dimensional surface in a given direction, at a particular time instant?

We call this the **flux** of the quantity through that surface.

M3b. At which rate is this quantity being produced in a particular three-dimensional region of space, at a particular time instant?

We call this the **supply** or **source** of the quantity in that region.

between t = 19:32:15
and t = 19:32:45



For six main quantities we can ask: how much of it is flowing through a given surface in a given direction, during a given time lapse?

In the case of magnetic flux we can ask the three questions above in 2 and 1 dimensions, rather than in 3 and 2 dimensions, as we shall see in Chapter 9.

Second property: extensivity

The second property common to all seven quantities tells how the measurements above combine for several regions of space and several surfaces:

Extensivity or additivity

- If we consider two or more non-overlapping volumes, the amount of quantity contained in the total volume is equal to the sum of the amounts contained in the individual volumes.
- If we consider two or more non-overlapping surfaces, the amount of quantity flowing through the total surface is equal to the sum of the amounts flowing through the individual surfaces.
- If we consider two or more non-overlapping volumes, the amount of quantity produced in the total volume is equal to the sum of the amounts produced in the individual volumes.

We say that each of the seven quantities is **extensive** or **additive**.

The basic measurements above cannot in general be made, and do not 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 two properties above, each of the seven quantities can be intuitively visualized as some kind of "stuff" that fills regions of space or flows through surfaces. This visualization is useful, but also comes with some warnings which we shall discuss later.

What's remarkable about matter, electric charge, magnetic flux, energy-mass, 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. And in all these theories they possess the two basic properties discussed above. The mathematical characterization of these quantities can be slightly different depending on the physical theory and spatial or temporal resolution. For example, in quantum theory

a quantity is mathematically represented by a so-called 'operator'. And on molecular scales, entropy has a meaning connected with probability theory. Yet, these seven quantities are universal in our present way of doing physics and of describing and understanding physical phenomena all 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. But remember that it is very difficult, if not impossible, to answer questions like "what is *really* the quantity. . . ?".

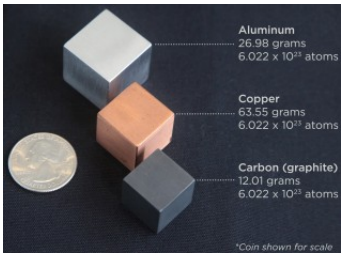
› § 1.4 page 21

3.3 Matter

Matter: units and notation

Matter, which includes *amount of substance*[2] in chemistry, is a *scalar* quantity. The unit for the amount of matter is the *mole*[3] mol; the unit for matter flux and supply is *mole per second* mol/s. In statistical mechanics and particle physics, matter is often simply counted and thus measured in dimensionless units, rather than in moles.

The amount of matter in a volume is usually denoted N. The flux of matter through a surface is denoted J; and supply of matter in a volume, A. In chemistry we usually specify what kind of matter we are speaking about, writing for instance $N_{\mathrm{Ca}} = 5.3$ mol, to indicate an amount of 5.3 mol of calcium[4] atoms.



One mole of different substances (image: NIST[1]).

Matter is probably the easiest quantity to grasp intuitively: it is what we ordinarily call "stuff". It is usually classified into several kinds. The classification depends on the physical phenomena and theory one works with. A building engineer, for instance, could classify "matter" into different kinds of *materials* – such as wood, concrete, steel, sand, plastic, and so on – keeping track of the amount of each material in different regions of space, its movement, its rate of production and transformation. Each material has different physical properties.

A chemist could classify matter into different *substances* – such as water, hydrogen, oxygen, carbon dioxide, and so on – again keeping track of their amounts, movements, production. According to this classification,

the "materials" of the building engineer would be mixtures of the different substances. But note that there is no clear boundary between one classification and the other.

A chemist could also classify matter into different kinds of *atoms* – such as hydrogen[5], helium[6], lithium[7], and the other kinds that appear in the periodic table[8] – and seeing substances and materials as combinations of these different atomic kinds of matter. This classification is special because these different kinds have, at least approximately, the property of being conserved: their amounts in a container or in a region of space can only change if these kinds are entering through an opening in the container or through the boundary of the region of space. In other words, they cannot be created or destroyed. This conservation property is only approximate, however. Radioactive atoms[9] can transmute from one kind to another. This possibility is crucial and must be taken into account in phenomena involving radioactivity[10] and nuclear energy-mass[11].

› § 5.3 page 111

A chemist or a particle physicist may classify matter into fewer different kinds: protons, neutrons, electrons, anti-protons, anti-neutrons, anti-electrons (also called *positrons*), seeing the different atomic kinds as being made of these six basic ones. These kinds may be conserved even when kinds of atoms are not.

But a nuclear or particle physicist knows that the conservation properties of the six kinds above is also only approximate, and there are also other kinds, produced only in special circumstances.

We therefore go down into more and more subtle classifications. This kind of research is still open, but it seems that the total amount of *baryonic*[12], (including protons and neutrons) and *leptonic*[13] (including electrons) matter is always conserved.

According to the definition of matter that we're adopting, the total amount of some kind of matter in a region can in principle be *negative*. A negative amount simply denotes the presence of anti-matter[14]. 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)[15] scans. In ordinary chemical applications, however, all amounts of matter within a region are usually positive or zero.

In positron-emission tomography there is creation of amounts of matter (leptonic matter) that can be considered negative: their *lepton number* is negative (image: Helse Bergen[16]).

Why do we need to worry about how matter gets classified depending on the application? Because for describing a physical system and predicting its behaviour we usually have to use at least one physical law *for each kind*

of matter. So the more kinds of matter we have to keep track of, the more equations we will have.

Ambiguity of the term 'matter'

In these notes we use the term 'matter' in the generic sense discussed above. But be aware that in some disciplines this term may have a much more specialized and slightly different meaning. It may *not* even be used at all. In chemical applications, for instance, one typically speaks specifically of 'compounds', 'mixtures', 'substances', 'elements', rather than 'matter'. A particle physicists speaks of *matter* and *anti-matter*, but in the present notes the term 'matter' refers to both.

Matter is different from mass-energy

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

As far as we know, the total amount of mass-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 downward" just like positive matter, a fact that has been experimentally confirmed: see Anderson et al. 2023.

Exercise 3.1

According to statements on symmetrymagazine.org[17] and quantumdiaries.org[18],

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

Unfortunately the original site where the 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?

3.4 Electric charge

Electric charge: units and notation

Electric charge is a *scalar* quantity. The unit for the amount of electric-charge is the *coulomb*[19] C. The flux of electric charge is called *electric current*, its unit is the *ampere*[20] A = C/s.

To be completed in a later version

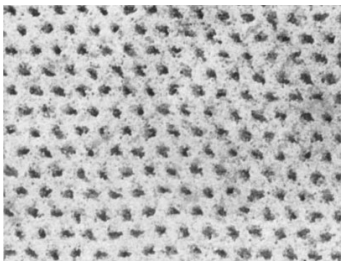
3.5 Magnetic flux

Magnetic flux: units and notation

Magnetic flux is a *scalar* quantity. The unit for magnetic flux is the *weber*[21] Wb. The "flux" of magnetic flux is called *electric potential difference* or *electric tension*; its unit is the *volt*[22] V = Wb/s.

The magnetic flux is usually calculated by means of a *vector* quantity called *magnetic flux density*; its unit is the *tesla*[23] T = Wb/m^2.

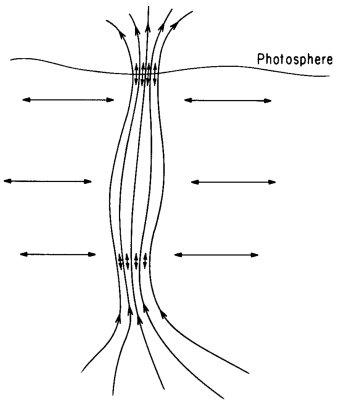
The electric potential difference is usually calculated by means of a *vector* quantity called *electric field strength*; its unit is the *volt per metre* (V/m).



"magnetic-flux lines emerging from the surface of a Type II superconductor" Essmann & Träuble 1971

As we shall see in more detail in Chapter 9, magnetic flux differs from the other six main quantities in that it answers the two "how much?" questions *in one lower dimension*: "How much magnetic flux is in this surface?" and "How much magnetic flux crosses this line in the unit of time?". It also requires a slightly different notion of orientation of a surface. The "flux of magnetic flux", or *electric potential difference*, is therefore a flow connected to a line.

The magnetic flux density and the electric field strength together are usually called "electromagnetic field", which is therefore commonly represented by two vectors associated to each point in space. But it can also be interpreted and visualized as a collection of spinning magnetic tubes or lines, either closed or extending indefinitely, which move around. This visualization is somewhat analogous 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



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

as solar physics[24], for example to study sunspots[25] (see Ryutova 2018). In particular situations, for example in some superconductors subjected to external magnetic fields, the magnetic-flux lines can literally be seen and even tracked as they move around (see previous side figure).

To be completed in a later version

3.6 Energy-mass

Energy-mass: units and notation

Energy is a *scalar* quantity. The unit for the amount of energy is the *joule* J; the unit for energy flux and supply is *joule per second* J/s, also called *watt* W = J/s.

Equivalently we can speak of mass. The unit for the amount of mass is the *kilogram*[26] kg; the unit for mass flux and supply is *kilogram per second* kg/s.

The amount of energy in a volume is usually denoted E, or m if we describe it as mass. The flux of total energy through a surface is denoted Φ; and supply in a volume, R.

The *chemical* energy-mass content in an ordinary AA battery is around 10 000 J. The *total* energy-mass content, including rest energy-mass, is around 10^{15} J.

The notion of energy is extremely important today, and central in many world-wide discussions and worries – think of today's "energy crisis", the need for "renewable energy", and so on. It is somewhat funny that despite its 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 "resistance" we feel when 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.

Energy is mass, mass is energy

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

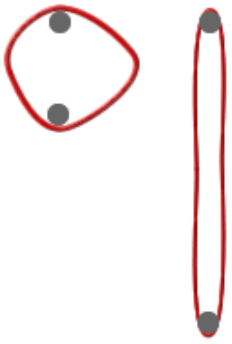
Heated gas. Imagine we have a box with a given amount of gas, say 1 mol of oxygen molecules. Using an extremely precise weighing scale, we observe that the mass of the gas is, say, exactly

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

Now we heat the gas, providing 60 J of energy, 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 actually 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 of energy that we provided to the gas by heating. 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.

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

$$0.000\,500\,000\,000\,000\,000\,000 \text{ kg} .$$

Now we stretch the band a little. By doing so we give energy to the band, which is said to acquire 'elastic energy'. Let's say we have given 0.3 J to the band in this way. Now we weigh the rubber band again, while stretched. We observe a mass of approximately

$$0.000\,500\,000\,000\,000\,003\,338 \text{ kg} .$$

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.

Now set the unstretched band in motion. Owing to the motion, the band is said to have acquired 'kinetic energy'; let's say an amount 0.3 J. If we could weigh the band while in motion (but without moving the weighing scale), we would observe again a mass of approximately

$$0.000\,500\,000\,000\,000\,003\,338 \text{ kg} .$$

The small difference from the initial mass is the additional kinetic energy of the band. Energy has weight; energy is mass.

Fission and atomic bombs. The atomic bomb[28] is a dark example of the fact that mass is energy. In phenomena of nuclear fission, we notice a decrease in the weight, measured at rest, of nuclear material before and after the phenomenon of fission. But we also observe that a great amount of (kinetic) energy is released. This amount is exactly equal to the apparently missing weight.

Hydrogen Bomb Test, 1954 (National Museum of Nuclear Science & History[27])

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 0.000 000 000 000 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 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. The equivalence between energy and mass is given by the famous formula $E = mc^2$, where c is the speed of light, eq. (2.2). In their respective units this gives

$$1\ \text{kg} = 89\,875\,517\,873\,681\,764\ \text{J} \quad \text{(exactly)}$$
$$1\ \text{J} \approx 0.000\,000\,000\,000\,000\,011\,126\,5\ \text{kg}$$

"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

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[29].

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-mass packed in very small volumes: the huge amounts of energy-mass that go together with objects and stuff like pens, keys, bicycles, cars, houses, water, and so on. We move, push, pull these huge energy-mass amounts from one place to another, and even put them in our pockets. We ourselves are huge bundles of energy-mass moving around. These amounts of energy-mass change a little, all the time, as in the examples with the rubber band above. But these changes are so small as to be often undetectable with ordinary weight scales, and negligible for practical purposes. We use the word 'mass' for any such huge amount of 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 01 %[30]. So we say "the rubber band has a mass of 0.0005 kg", rather than "the rubber band has an energy-mass of 45 000 000 000 000 J".

On the other hand, we also deal with the small energy changes and exchanges in all these objects. These energy exchanges that are very important for our daily life: they keep us warm, keep our cells active, 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, think of 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 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, if we have three dear friends or relatives visiting us from abroad, then the amount of 3 people is now for us very important – even if it is a very small amount compared to the total population of a country.

The distinction above is of course not clear-cut. In dealing with some physical phenomena, for example with few molecules or with subatomic particles, the fictitious but pragmatic distinction between mass and energy becomes too blurry and not useful anymore. In discussing these phenomena, indeed, one often uses the terms 'mass' and 'energy' interchangeably, as well as a common unit for both, such as the *electronvolt*[31].

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

Different 'forms' of energy-mass

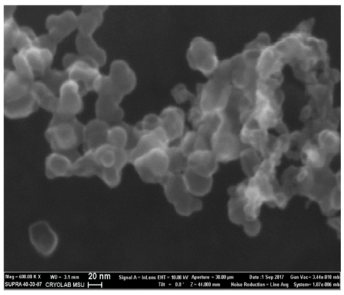
We often speak of different *forms* of energy-mass. The most important forms for us will be **internal energy-mass**, **kinetic energy-mass**, **gravitational potential energy-mass**, **electromagnetic energy-mass** to be discussed later.

The differences among these forms of energy-mass arise from the way they are calculated from other quantities, as we shall see later. For example, if in a volume there's an amount of a particular kind of matter, then in that volume there must also be an amount of energy-mass, given by a formula that involves the amount of matter. And if that matter is moving, then we have to add an extra amount of energy-mass given by another formula which involves the velocity. And if in that volume there's a gravitational field (that is, a particular kind of spacetime curvature), then another extra amount of energy-mass must be added, given by yet another formula involving the gravitational field. Similarly if we know that an electromagnetic field is in that volume.

› § 11.2 page 217

We also speak of different forms of flux of energy-mass. The most important for us will be **heat** and **mechanical power**. The difference is again in how these fluxes are calculated depending on whether there are also fluxes of matter and of other quantities.

The distinctions between different forms of energy-mass also depend on the observation scale and the theory used. Take, for instance, the water in a glass resting on table. We can observe and describe that water on a macroscopic scale of centimetres, seeing it as a still, uniform fluid. On this macroscopic scale we say that the water has internal energy-mass, or that there is internal energy-mass in the glass. But we can also observe and model that same water as a collection of molecules, on a microscopic scale of nanometres (10^{-9} m). On this microscopic scale, we speak of internal energy-mass *and* kinetic energy-mass, because the molecules are in constant motion. The *total* amount of energy-mass is the same on the centimetre-scale and on the microscopic scale, but its partition into different "forms" depends on the scale: only internal on the macroscopic scale, and internal plus kinetic on the microscopic scale.



Hydrocarbon fuel particles[32]. The small blobs have size of around 2×10^{-8} m.

The same is true of flux of energy-mass. What we call 'heat' on one observation scale appears as a flux of energy-mass not associated with the motion of matter. But on a finer scale it is instead called 'work', and it appears as an energy-mass flux associated with the microscopic motion of matter.

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 flux of *energy-mass* t, in J/s, through that door?

3.7 Momentum

Momentum: units and notation

Momentum, also called *linear momentum* or *translational momentum* to distinguish it from angular momentum, is a *vector* quantity. The amount of momentum can be expressed in several equivalent units; we shall keep in mind especially these three:

$$\underset{\mathrm{N\cdot s}}{\textit{newton second}} \equiv \underset{\mathrm{kg\cdot m/s}}{\textit{kilogram metre per second}} \equiv \underset{\mathrm{J\cdot s/m}}{\textit{joule second per metre}}$$

Flux of momentum is also called *contact force* or *surface force*. Supply of momentum is also called *body force* or *volume force*. They can be expressed in several equivalent units:

$$\underset{\mathrm{N}}{\textit{newton}} \equiv \underset{\mathrm{kg\cdot m/s^2}}{\textit{kilogram metre per squared second}} \equiv \underset{\mathrm{J/m}}{\textit{joule per metre}}$$

Since momentum and momentum flux are vector quantities, they are usually expressed with three numbers, typically their x-, y-, and z-components.

The amount of momentum in a volume is usually denoted $\boldsymbol{P}$. The flux of momentum (surface force) is denoted $\boldsymbol{F}$; and supply of momentum (volume force), $\boldsymbol{G}$.

A walking person has, with respect to the ground, an amount of horizontal momentum of around 70 N s (image: Antonio Romei[33]).

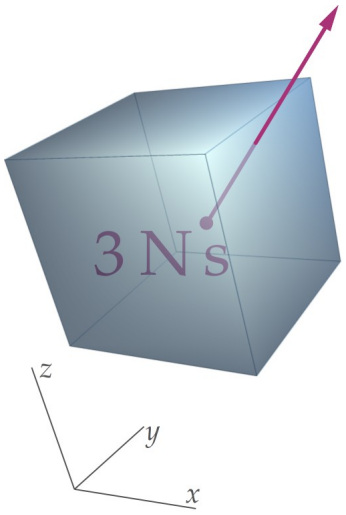
A 10 cm beam of light from a 60 W torch has an amount of momentum around 10^{-16} N s.

Momentum is a subtle quantity, even subtler than energy-mass. Textbooks that focus on Newtonian mechanics *define* it as the product of the mass and the velocity of a body, usually written "$\boldsymbol{p} = m\boldsymbol{v}$". This relation, however, is only valid in special circumstances, and cannot be used in many everyday technological applications, especially when electromagnetism or high speeds are involved. And that relation is actually only an approximation even in the circumstances where it's used. For example, in a small region where there are matter and electromagnetic fields, momentum is approximately given by $\boldsymbol{P} = m\boldsymbol{v} + \boldsymbol{E}\times\boldsymbol{H}/c^2$, where the quantities $\boldsymbol{E}$ (electric field strength) and $\boldsymbol{H}$ (magnetic field strength or free-current potential) are related to the electromagnetic properties of matter. And at high speeds, the momentum of matter is better approximated by $\boldsymbol{P} = (m + E/c^2)\boldsymbol{v}$, where m is the rest-energy-mass and E is the remaining energy-mass of matter.

It is therefore beneficial to separate our idea of momentum from the

"mass-times-velocity" formula, keeping in mind that the latter is just a particular case of momentum. Instead, think of momentum as *something 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). So if something – be it matter or a magnetic-flux line – is changing its position in space, then that something has momentum. Indeed in many languages – for example Chinese, French, Italian, Japanese, Norwegian, Spanish, Swedish – the term for momentum is literally 'quantity of motion'.

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.



The amount of momentum within a volume at a given instant is represented by a vector

Production of momentum, and flux of momentum in particular, is what we call **force**. Whenever we exert a force on an object, for instance when we pull a door or simply hold a bag, we are transferring momentum between us and that object. The force of gravity, or weight, that we feel every day in our bodies, is a continuous production of momentum that happens because of the huge mass-energy of the Earth. The force that presses you against your seat when you sit in an accelerating car or in an aeroplane taking off, is also a production of momentum. The term 'force' is therefore synonymous with 'flux of momentum' or 'production of momentum'.

Momentum and energy-mass flux are proportional

According to Relativity Theory, momentum is always proportional to energy-mass flux, and energy-mass flux is always proportional to momentum. If we represent the magnitude of momentum in a volume V with P, and the flux of total energy-mass through an area A with Φ, their proportionality can be approximately expressed as

$$\frac{\Phi}{A} \approx \frac{P}{V} c^2 \,. \tag{3.1}$$

Compare this formula with $E = m c^2$. From this point of view, you can think of momentum as "energy-mass in motion".

Exercise 3.3

Assume the relation $P = mv$ between the magnitude of momentum P, mass-energy m, and its speed v. When you're walking with a speed of 1 m/s, how much momentum does your body contain?

3.8 Angular momentum

Angular momentum: units and notation

Angular momentum, also called *moment of momentum* or *rotational momentum*, is a *vector* quantity. The amount of angular momentum can be expressed in several equivalent units; we shall keep in mind especially these three:

$$\underset{\mathrm{N\cdot m\cdot s}}{\textit{newton metre second}} \equiv \underset{\mathrm{kg\cdot m^2/s}}{\textit{kilogram squared metre per second}} \equiv \underset{\mathrm{J\cdot s}}{\textit{joule second}}$$

Flux of angular momentum is also called *contact torque*. Supply of momentum is also called *body torque*. They can be expressed in several equivalent units:

$$\underset{\mathrm{N\cdot m}}{\textit{newton metre}} \equiv \underset{\mathrm{kg\cdot m^2/s^2}}{\textit{kilogram squared metre per squared second}} \equiv \underset{\mathrm{J}}{\textit{joule}}$$

Since angular momentum and its flux are vector quantities, they are usually expressed with three numbers, typically their x-, y-, and z-components.

The amount of angular momentum in a volume is usually denoted $\boldsymbol{L}$. The flux of angular momentum (surface torque) is denoted $\boldsymbol{\tau}$; and supply of angular momentum (volume torque), $\boldsymbol{M}$.

A spinning dancer has, with respect to the ground, an amount of angular momentum of around 10 N m s (image: RG-Dance[34]).

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.

Just as momentum is associated with translational motion, angular momentum is *something associated with rotational motion* of matter and of electromagnetic fields. Rotational motion is the kind of motion that leads

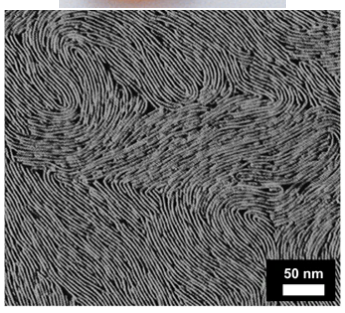
to a *new orientation* in space, rather than to a new position. For instance, if you turn to your left or to your right while standing in place, you have performed a rotational motion.

You can get a feeling of angular momentum and torque by playing with one of those hand-held gyroscopes that can be used for wrist exercise (side figure). But the physical role of angular momentum is actually in front of us every day: its balance is the chief reason of such important phenomena as the alternation of day and night and the alternation of the seasons.

Yet, angular momentum is seen as an even subtler quantity than momentum, and usually is less in the spotlight than momentum. This happens for several reasons.

One reason is that the distinction between momentum and angular momentum is not clear-cut, and depends on our choice of coordinates. For instance, what can be considered as a component of momentum in one coordinate system, can also be considered as a component of angular momentum in another coordinate system. This reflects the fact that 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 a collection of small translational motions of its parts.

Another, very important reason is that the physical laws that involve angular momentum can be expressed in a different way, as sorts of symmetries between momentum fluxes and between energy-mass flux and momentum. With this alternative approach we don't even need to introduce the quantity 'angular momentum' at all! This approach is indeed followed for many physical phenomena and applications, for example those that involve fluid motion. But for other phenomena it's much more convenient to speak of angular momentum and to use its balance law. In phenomena involving liquid polymers[35], elementary particles, or electromagnetic radiation, for example, it's convenient to use angular momentum and to let it include an additional part, called *spin* or *intrinsic angular momentum*, which is associated with rotational motion that is invisible to the naked eye or the particular instruments used to measure motion.



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

We shall further discuss these topics in Chapter 12. In any case keep in mind that *there is* a universal law that is *independent* from the one obeyed

by momentum. Even if we can in principle omit the notion of angular momentum, we *cannot* avoid the use of this additional universal law.

Exercise 3.4

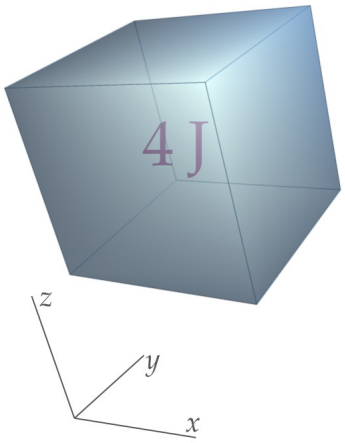
Assume the relation $L = mvR$ between the angular-momentum magnitude L, mass-energy m, speed v, and orbital radius R of a planet. For the Earth[36] we have approximately

$$m = 6.0 \times 10^{24}\ \text{kg} \qquad v = 3.0 \times 10^{4}\ \text{m/s} \qquad R = 1.5 \times 10^{11}\ \text{m}$$

How much is the Earth's angular momentum? (Note that we are neglecting the angular momentum coming from its own rotation.)

3.9 Energy-mass, momentum, angular momentum are coordinate-dependent

An aspect of energy-mass, momentum, angular momentum that must always be kept in mind is that **their amounts 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-mass 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 flux of energy-mass through a surface, and for the amounts, fluxes, supplies of momentum and of angular momentum. This coordinate-dependence also means that observers using different coordinates will usually assign different amounts of energy-mass, momentum, angular momentum to the same regions of spacetime.



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

This is an important difference between energy-mass, momentum, angular momentum on one side, and matter, electric charge, magnetic flux on the other side. *For matter, electric charge, magnetic flux, the questions about their contents and fluxes can be answered unambiguously independently of any spatial coordinate system.*

This coordinate-dependence is not a problem: we must in practice always specify our coordinate system anyway, in order to agree on the time and position of physical events. But it can cause problems if we calculate or measure some amount of energy-mass at a given time and place using a coordinate system, and we calculate or measure some amount

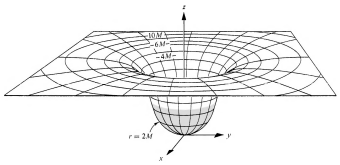
at another time or place using a *different* coordinate system. Combining these two amounts, or using them in the same physical law, has no meaning whatsoever.

Amounts of energy-mass, momentum, angular momentum are coordinate-dependent

Never change coordinate system in the middle of calculations about energy-mass, momentum, or angular momentum!

What are energy-mass, 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-mass, 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. This topic will be discussed in Chapter 13.
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-mass 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-mas, momentum, angular momentum are measures of particular curvatures of spacetime.

3.10 Entropy

Entropy: units and notation

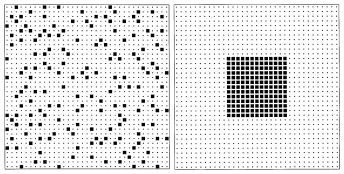
Entropy is a *scalar* quantity. The unit for entropy content is the *joule per kelvin* J/K; the unit for entropy flux is *joule per kelvin per second* J/(K s).

The amount of entropy in a volume is usually denoted S. The flux of entropy through a surface is denoted Π.

From its SI physical dimension and unit, entropy would seem to be derived from temperature. However, although temperature is taken

as primitive quantity by the SI, the definition of temperature[37] actually depends on a fixed value of *Boltzmann's constant*[38], 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* description is to say that entropy is a "measure of disorder". Besides the fact that "disorder" is a very vague and subjective notion, it turns out that some physical phenomena, for example involving liquid crystals[39], 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.



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

In these notes we shall rely on the idea that *entropy expresses a limit on the flux of a particular kind of energy-mass*. Said in simpler but more imprecise words, entropy is a bound on how fast we can heat something up. 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-mass 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 *probabilitistic* 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".

3.11 Auxiliary quantities

Besides the seven principal quantities, other auxiliary quantities appear in some physical theories. Important examples are *temperature*, *metric*, *strain*, *magnetization*. Some auxiliary quantities are not extensive; for instance we cannot ask "what's the total amount of temperature in this region?".

The dimensions, units, and scalar or vector character of all quantities mentioned so far are summarized in table 3.1 on page 73.

Temperature

We have an intuitive understanding of the notion of *hotness* and *coldness*. Temperature quantifies these notions. The physical bases and measurement procedures for this quantification are far from trivial, but we shall take them for granted in the present notes.

For some physical phenomena, especially those involving gases, we know that temperature is related to the invisible motion of microscopic parts of matter, such as molecules. But there are also physical phenomena for which our microscopic understanding of temperature is more complex, and in some cases still unclear.

Temperature is a *scalar* quantity. There are several definitions and scales of measurement for temperature. Of special importance is **thermodynamic temperature**[40], also called *absolute temperature*, which is measured in *kelvins* (K). Thermodynamic temperature has the special property of being always positive in most physical phenomena (there are exceptions, especially in some phenomena where statistical mechanics[41] becomes relevant).

In these notes we shall use thermodynamic temperature, denoting it T. Its relation with Celsius temperature T_C, which is measured in *degrees Celsius* °C, is given by

$$T_C = T - 273.15\,\mathrm{K} \tag{3.2}$$

that is, a redefinition of the "zero" value; for instance $25.00\,^\circ\mathrm{C} = 298.15\,\mathrm{K}$. Note that temperature *differences* are the same for the two temperature scales: $\Delta T_C = \Delta T$, because the constant zero-value cancels out.

Temperature is useful because it enters in many physical laws that involve energy-mass, but it's easier to measure than energy-mass. Temperature generally depends on the time and place, so it can be a function of the coordinates: $T(t, x, y, z)$.

What is temperature?

Inventing Temperature by H. Chang (2004) gives a brilliant account of the history of invention of temperature, as well as an interesting portrait of how scientific concepts are born and develop.

Is There a Temperature? by T. S. Biró (2011) discusses fascinating physical phenomena for which our microscopic understanding of temperature is still incomplete.

3.12 Metric

A very important quantity is a fundamental building block of all our physical theories: the **metric**. It is quite different from the seven fundamental quantities, from a physical and also from a geometrical point of view.

The metric characterizes our measurements of space and time. It's the object that allows us to calculate how much *physical* time has elapsed, the *physical* distance between two objects, the volume (say, in cubic metres) of a three-dimensional region of space, and the area (say, in square metres) of a surface. In General Relativity the metric allows us to calculate the curvature of spacetime.

The metric itself, in the technical sense of the so-called "metric tensor", is *not* an extensive quantity. We can't ask "what's the total amount of metric in this region?"; that's a meaningless question. There are, however, other quantities which can be derived from the metric and which are extensive. Important examples are the so-called "volume density" and "area density", which are the ones that allow us to calculate extended volumes and areas.

In the Newtonian approximation, that is, for speeds smaller than the speed of light and low energy-mass densities (hence weak gravitational fields and small spacetime curvature), the metric is just a static, uniform object, the same everywhere in spacetime; and spacetime is *flat*, that is, it 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 with coordinate time, and can vary from one point in space to another. These changes are determined by the seven main quantities, and the metric, in turn, determines changes in the seven quantities.

Quantity	**SI Dimension**	**Unit**
Time	time	*second* s
Length	length	*metre* m
Matter	amount of substance	*mole* mol
Electric charge	current · time	*coulomb* C
Magnetic flux	mass · length^2/(current · time^2)	*weber* Wb
Energy-mass	mass · length^2/time^2, mass	*joule* J, *kilogram* kg
Momentum	mass · length/time	N · s, kg · m/s, J · s/m
Angular momentum	mass · length^2/time	N · m · s, kg · m^2/s, J · s
Entropy	mass · length^2/(time^2 · temperature)	J/K
Temperature	temperature	*kelvin* K

Table 3.1 SI dimensions and units of the main physical quantities used in these notes. Their fluxes have the dimensions divided by time, and therefore units divided by seconds. Quantities in **boldface** are vectors, the others are scalars.

URLs for chapter 3

1. https://www.nist.gov/image/moleedit2jpg
2. https://doi.org/10.1351/goldbook.A00297
3. https://www.nist.gov/si-redefinition/redefining-mole
4. https://pubchem.ncbi.nlm.nih.gov/element/Calcium
5. https://pubchem.ncbi.nlm.nih.gov/element/Hydrogen
6. https://pubchem.ncbi.nlm.nih.gov/element/Helium
7. https://pubchem.ncbi.nlm.nih.gov/element/Lithium
8. https://iupac.org/what-we-do/periodic-table-of-elements/
9. https://www.ciaaw.org/radioactive-elements.htm
10. https://www.iaea.org/newscenter/news/what-are-radioactive-sources
11. https://www.iaea.org/newscenter/news/what-is-nuclear-energy-the-science-of-nuclear-power
12. http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/hadron.html#c6
13. http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/lepton.html#c1
14. https://www.britannica.com/science/antimatter
15. https://www.britannica.com/topic/positron-emission-tomography
16. https://www.helse-bergen.no/avdelinger/radiologisk-avdeling/senter-for-nukleermedisin-og-pet/tilvising-til-pet
17. https://www.symmetrymagazine.org/2009/07/23/antimatter-from-bananas
18. https://www.quantumdiaries.org/2009/07/21/positrons-from-bananas/
19. https://doi.org/10.1351/goldbook.C01365
20. https://www.nist.gov/si-redefinition/ampere-introduction
21. https://doi.org/10.1351/goldbook.W06666
22. https://doi.org/10.1351/goldbook.V06634
23. https://doi.org/10.1351/goldbook.T06283
24. https://doi.org/10.1093/acrefore/9780190871994.013.21
25. https://spaceplace.nasa.gov/solar-activity/
26. https://doi.org/10.1351/goldbook.K03391
27. https://nuclearmuseum.pastperfectonline.com/Archive/716477C1-5E7A-485C-8BE1-857919471563
28. https://www.britannica.com/science/nuclear-fission
29. https://www.britannica.com/story/atomic-bombing-of-hiroshima
30. https://www.nist.gov/si-redefinition/kilogram-disseminating-new-kilogram
31. https://home.cern/tags/13-tev
32. https://doi.org/10.4209/aaqr.2019.04.0177
33. https://www.flickr.com/photos/33486695@N06/13566555795
34. https://www.rg-dance.com/richardalstondancecompany/
35. https://www.britannica.com/science/polymer
36. https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html
37. https://doi.org/10.1351/goldbook.K03374
38. https://doi.org/10.1351/goldbook.B00695
39. https://www.britannica.com/science/liquid-crystal
40. https://www.iso.org/obp/ui/#iso:std:iso:80000:-5:ed-2:v1:en:tab:1
41. https://www.britannica.com/science/statistical-mechanics

4 Volume contents, fluxes, supplies

> When we regard energy as residing intrinsically in a body, we may measure its intensity by the amount contained in unit of volume. [. . .][T]he only way we have of defining the motion of the fluid is by considering it as a flux [. . .]. This distinction is still more necessary when we come to heat and electricity. The flux of heat or of electricity cannot be even thought of in any way except as the quantity which flows through a given area in a given time.
>
> J. Clerk Maxwell 1869

4.1 Content, flux, supply

The concept of physical quantity allows us to look at the word in ways that can be quantified and expressed with numbers and mathematics. This is how we can formulate physical laws. In the previous chapter we made our acquaintance with seven physical quantities. They were chosen as our building blocks because their quantification is easy to grasp intuitively and to visualize. In the present chapter we study this quantification more rigorously, and start developing the necessary mathematics.

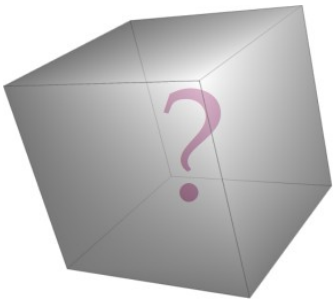
› § 1.3 page 19

For each of the seven primitive quantities, except magnetic flux, we can measure three kinds of amount: **volume content**, **flux**, and **supply**:

Volume content

Volume content or **volume integral** is the amount of quantity contained within a three-dimensional region, at a specific time instant.

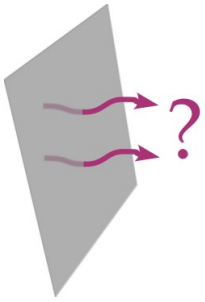
The volume content of a quantity does *not* depend on how the volume is moving. The volume content has the same physical dimension as the quantity.



Flux

Flux is the amount of quantity flowing through a two-dimensional surface in a given direction, *per time*, at a particular time instant.

The flux of a quantity through a surface depends on how that surface is moving and deforming. The flux has the physical dimension of that quantity divided by time.



Supply

Supply or **source** is the amount of quantity being produced or destroyed within a three-dimensional region *per time*, at a particular time instant.

The supply of a quantity in a region depends on how that region is moving and deforming. The supply has the physical dimension of that quantity divided by time.

For the magnetic flux we shall introduce analogous kinds of amount, but in one less dimension.

The volume content, flux, and supply of the four scalar quantities – matter, electric charge, energy-mass, entropy – are also *scalars*; that is, each is expressed by one number and a unit. The volume content, flux, and supply of the two vector quantities – momentum and angular momentum – are also *vectors*; that is, each is usually expressed by three numbers and units.

4.2 Symbols, notation, and extensivity

The three kinds of amount, for each quantity, are denoted by special symbols. Table 4.1 on page 77 summarizes the symbols, units, and scalar or vector character for the volume content, flux, supply of the seven quantities, as used in these notes. Let's see some examples.

Consider a bottle on a table. You measure the amount of water molecules in it at a given time, and find that it's 51.3 mol. We can express this as follows:

$$N_{H_2O} = 51.3 \text{ mol} .$$

Quantity	*Vol. content [unit]*	*Flux [unit]*	*Supply*
matter	N [mol]	J [mol/s]	$\mathcal{A}$
electric charge	$\boldsymbol{Q}$ [C]	$\mathcal{I}$ [C/s or A]	
energy-mass	E [J] m [kg]	Φ [J/s or W] [kg/s]	R
momentum	$\boldsymbol{P}$ [N s]	$\boldsymbol{F}$ [N]	$\boldsymbol{G}$
angular momentum	$\boldsymbol{L}$ [N m s]	$\boldsymbol{\tau}$ [N m]	$\boldsymbol{M}$
entropy	S [J/K]	Π [J/(K s)]	

Table 4.1 Symbols and units for volume content, flux, supply of six main quantities. Vector quantities are in **boldface**. Supplies have the same units as fluxes.

Quantity	*Flux [unit]*	*Circuitation [unit]*	*Flux supply*
magnetic flux	$\mathcal{B}$ [Wb]	$-\mathcal{E}$ [Wb/s or V]	

Table 4.2 Symbols and units for flux and circuitation of magnetic flux.

The symbol 'N' is used for the volume content of *matter*. The kind of matter in this case is water, so we append its chemical formula H_2O as a subscript; but we could also have used the word 'water' or the letter 'w'. We could also append the word 'bottle' to indicate that we're speaking about the volume content of the bottle.

As a second example, consider two ordinary batteries inside some device; one is on the left, the other on the right in the battery compartment. You measure the amount of chemical energy-mass in each battery at a given time, and find 9873 J for the battery on the left, and 4221 J for the one on the right. We can express this as follows:

$$E_{\mathrm{L}} = 9873\,\mathrm{J}\,, \qquad E_{\mathrm{R}} = 4221\,\mathrm{J}\,.$$

The symbol 'E' is used for the volume content of *energy-mass*. In this case we're speaking about chemical energy-mass, and we assume this is clear from the context, so we don't indicate this in the symbols. But we must distinguish the volume contents of the two batteries, so we use the subscripts 'L' and 'R'; but we could have used 'l' and 'r', or 'left' and 'right'.

As a third example, consider a comb that has acquired electricity, say from rubbing with hair. You measure the net volume content of electric charge in the comb when your stopwatch shows 5 s, and find $-0.000\,000\,4$ C (note that the charge is more concentrated at the boundary of the volume). Then you measure it again 10 seconds later and find that there is no net charge. We can express this as follows:

$$\begin{aligned} t_0 &= 5\,\mathrm{s} \qquad & Q(t_0) &= -0.000\,000\,4\,\mathrm{C}\,, \\ t_1 &= 15\,\mathrm{s} \qquad & Q(t_1) &= 0.000\,000\,0\,\mathrm{C}\,. \end{aligned}$$

The symbol 'Q' is used for the volume content of *electric charge*. We assume that it's clear from the context that we're speaking about the charge of the comb, so we don't indicate this in the symbols. The volume content Q of the charge is changing with time, that is, it is a *function* of time. Thus we indicate its value at different times by explicitly writing its argument in round brackets.

The examples above show that it depends on the context what kinds of subscript or additional signs that we use together with the symbols for volume content, flux, supply. We have some freedom in which additional signs to use; what's important is to make the message unambiguous.

Volume content, flux, and supply always refer to some particular time, so strictly speaking they are functions of time and would need a

time argument like '(t)'. But in situations where we are considering one particular time only, or where the quantities are constant in time, we can for brevity omit the time argument.

Extensivity or additivity

The first kind of mathematical operation that we can do with our seven quantities is related to the other important property common to all of them, *extensivity* or *additivity*:

› § 3.2 page 54

- The content in a volume consisting of non-overlapping volumes is equal to the *sum* of the contents in the individual volumes.
- The flux through a surface consisting of non-overlapping surfaces is equal to the *sum* of the fluxes through the individual surfaces.
- The supply in a volume consisting of non-overlapping volumes is equal to the *sum* of the supplies in the individual volumes.

When we speak of sum, we mean an ordinary sum for scalar quantities, and a *vector* sum for vector quantities.

As an example, let's consider again the two batteries. If we take them together, we're considering a new volume, consisting of the volumes of the two batteries. We can denote the content of energy-mass in this volume with E_{tot} to distinguish it from the contents of the left battery E_{L} and of the right battery E_{R}. The content in the combined volume is given by

$$\begin{aligned} E_{\text{tot}} &= E_{\text{L}} + E_{\text{R}} \\ &= 9873\,\text{J} + 4221\,\text{J} = 14\,094\,\text{J} \end{aligned}$$

and this equation expresses the principle of extensivity.

4.3 Control volumes and control surfaces

For each of six main quantities we can therefore say how much of that quantity is in a given volume, how much is flowing through a given surface per time, and how much is being created in a given volume per time. But how should we choose such volumes and surfaces? Are we constrained in their choice? Do they need to have particular shapes or sizes or positions? Do they need to follow the contours of particular physical objects?

These volumes and surfaces will play an important role in formulating our main physical laws; so one might think that they must be chosen in very special ways. But that isn't the case!

What's surprising and extremely useful is that **the choice of volume** for a volume content or supply, **and the choice of surface** for a flux, **are completely arbitrary**. Moreover, **these volumes and surfaces can be completely imaginary**.

Since they are under our control, and since they allow us to keep under control how the amounts of our main quantities change, we call them *control volumes* and *control surfaces*:

Control volume

A **control volume** is an arbitrary three-dimensional region of space. This region can have any position, shape, and size, and these can change smoothly in time.

A control volume can also consist of several disconnected three-dimensional regions, and it can be completely imaginary.

Control surface

A **control surface** is an arbitrary two-dimensional surface. This surface can have any position, shape, and size, and these can change smoothly in time.

A control surface can also consist of several disconnected surfaces, and it can be completely imaginary.

As explicitly stated in the definitions above, control volumes and control surfaces don't need to be static: they can move and deform.

The fact that we can choose control volumes and control surfaces arbitrarily gives us a lot of power in solving physics problems and in simulating and predicting the behaviour of physical phenomena. Control volumes and surfaces are typically chosen so as to simplify the equations that describe the physical situation and to focus on details of interest.

Up to now our discussion has been somewhat abstract. This abstractness reflects the amazing flexibility of the notions of control volumes and control surfaces. We now explore some more concrete examples.

Examples

Let's start with an informal example of control volume and control surface.

Consider a classroom with some people within. In your imagination you can divide the classroom into two halves, say the front half and the rear half. You effect this division by imagining a two-dimensional, immaterial surface, going from the floor to the ceiling, and from one side wall to the opposite one. The exact position and shape of this imaginary surface are up to you, as long as the surface has a clear crossing direction.

You can then simply measure how many people are in the rear half at a particular time instant, say time 14:43:27 of a specific date. This measurement could be done for instance by taking a photograph at exactly that instant, and then counting the people in the rear half on the photograph.

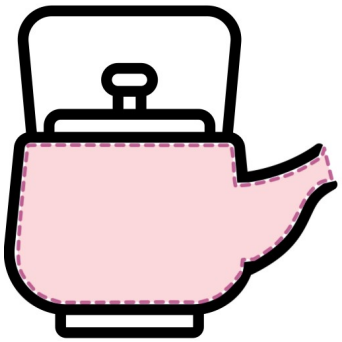
What you just did was to choose a *control volume* at a given time, and to measure its "volume content" of people. Note how the control volume in this case was defined partly by real objects (half of the ceiling and floor, half of the side walls, the rear wall) and partly by an imaginary object (your imagined surface dividing the classroom).

A classroom with an imaginary, slightly skew surface (in red) dividing the front and rear half of the room. (Room design: Raihanali[1].)

You can also focus on your imaginary surface only, rather than the room, and with the help of a video camera you can keep track of how many people cross the imaginary surface between times 14:43:27 and 14:43:30. You could for instance observe that in those three seconds 10 people cross the imaginary surface from the rear half to the front half of the room, and 4 people cross it from the front rear. So a net number of 6 people cross from the rear to the front half in three seconds, or 2 people/second on average.

What you just did was to consider a *control surface*, and to measure the "flux" of people through it. In this case the control surface was defined by an imaginary object (your imagined surface). But you could also consider a different, larger control surface, consisting for instance of the imaginary surface and one side wall. This control surface would be defined partly by a real object (the side wall) and partly by an imaginary object (your imagined surface).

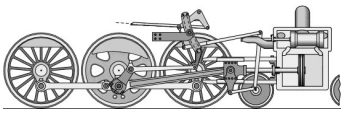
In the previous example the control volume and control surfaces had very simple shapes. We can choose more complex shapes if needed. Take for instance a kettle, and suppose we want to keep track of how much water, steam, air, and maybe energy-mass go into or out of it. The interior of the kettle is our control volume. It is delimited by the interior surface of the metal that makes up the kettle, and also by an imaginary surface at the end of the spout. Note that as we move the kettle around, so do these



Simplified section of a kettle, with a control volume (in red) delimited by a closed control surface (dashed red line).

control volume and control surface. If we make a dent on the kettle with a hammer, then the control volume and control surface will follow that dent.

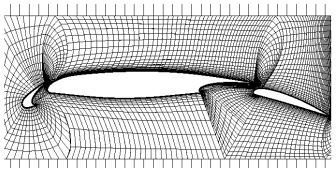
In a similar way we can consider as control volumes more complicated objects and spaces – like a car, a rocket, a bacterium, a planet, the chamber in a piston – with appropriate control surfaces that "wrap" them. Keep in mind that these control volumes and surfaces bend and move as needed. The locomotive system illustrated on the side can be considered as a control volume, with a complex movable control surface that perfectly wraps every element of it.



From 009 Lynton and Barnstaple Railway County Gate[2]

4.4 Choices of control volumes and surfaces

The full freedom we have in choosing control volumes and surfaces is extremely powerful for physics and engineering applications. Let's discuss two particular kinds of choice that are often made in two broad types of physical situations, and that we shall also often use in these notes.



Section view of grid of control volumes and surfaces for studying the flow of air around an aeroplane wing (from Schieffer 2013). The extended white areas are sections of the wing and of its slat and flap[3].

› § 7 page 158

First choice. When we study solid objects, or objects made of solid parts – like a football, a book, a car, an aeroplane, a rocket, and in some situations a human being or a planet – then we use the object itself as a control volume. The control surface tightly "wraps" the object.

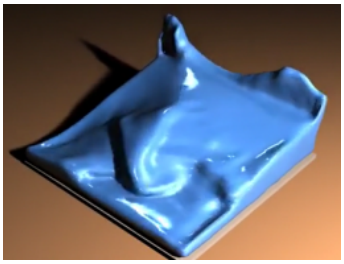
Because of this particular choice, we often say, for instance, 'the momentum of the tennis ball', but what it's meant is 'the momentum in the volume that encloses the tennis ball'.

There are two main reasons for this choice. First, this choice makes one particular physical law, the balance of matter, automatically satisfied. So we don't need to consider it when describing the physical phenomenon of interest. Second, the movement and deformation of control volumes and surfaces so chosen are not too complicated to represent mathematically; because the object is solid, so its kinds of movement are limited.

Second choice. When we study something flowing or moving in a more fluid way – like a body of water, the atmosphere, fuel or gas in a cylinder, plasma, or an electromagnetic field – then we divide the space of interest into small, *static* control volumes and surfaces. We essentially construct a *mesh*.

With this choice we do have to explicitly consider the physical law for the balance of matter, making the equations a little more complicated. But at the same time the mathematical representation of these control volumes and surfaces is extremely simple, because they're static.

There are also hybrid choices between the two special choices above, with some control volumes and surfaces that are static, and others that move and deform in a more fluid way. These hybrid choices are used for instance in simulations of fluids where the behaviour of their surface – waves, detaching drops, and similar – are important.

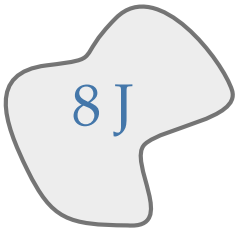


Clever hybrid choices of control volumes and surfaces allow us to simulate complex motions of fluid surfaces. See also animation[4] (Wojtan et al. 2009).

4.5 Volume content

Scalar quantities

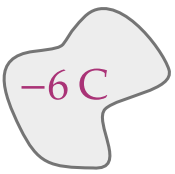
A volume content (or volume integral) for a scalar quantity, for example energy-mass, can be represented like this:



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-mass 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-mass in the volume at that instant, but we don't know the situation earlier or later: there could be a different amount of energy-mass, the region might be at a different position and have a different shape, or it might not even exist.
- Recall that some scalar quantities, like 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.



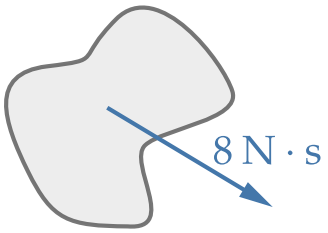
A region with a negative amount of charge

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?

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 figure shows the direction and orientation of this vector, and the magnitude of 8 N s is explicitly reported.

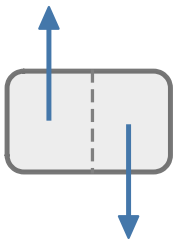
Vector magnitudes and opposite vectors

Remember that the *magnitude* of a vector is always positive, and that ↖ = −1 · ↘

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.

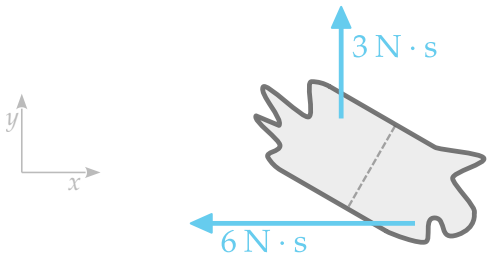
This last remark is especially important when we discuss momentum and angular momentum. As an example, look at the side figure: the volume content for the whole region is *zero*, but its left and right subregions have *non-zero and opposite* volume contents.



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

Exercise 4.2

Recall *extensivity*, the second property of our seven primitive 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.

Adding vectors in General Relativity

We are used to the idea of adding vectors placed at different points in space: we only have to move each vector, 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 so simply 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. But it is still possible to add the momentum of two different spatial regions, simply because *momentum is defined with respect to a coordinate system*. This coordinate system selects, so to speak, a unique way to transport the momentum vectors to a common point. We are again reminded of the fact that momentum is a coordinate-dependent quantity.
In General Relativity momentum is not really a vector, but just a special triplet of quantities.

› § 3.9 page 68

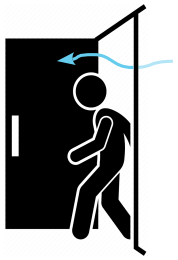
4.6 Flux of scalar quantities

To get an intuitive grip of the notion of flux, consider a flow of people through an open door. The door is our control surface. We can ask how

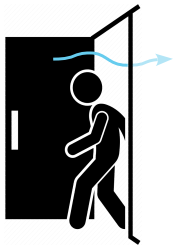
many people crossed the door during a minute. But one more detail about this flow is important: *in which direction did the people cross the door?* This detail is important because, for example, the door leads to a classroom and we need to keep track of how many seats are free. We therefore need to know whether each person who crossed the door was actually *entering* or *leaving* the classroom.

In order to do this we can proceed as follows:

1. Assign a crossing direction to the door, for instance the direction from outside to inside the classroom.
2. Count as 'positive' each person who crosses the door in the chosen crossing direction, and as 'negative' each person who crosses the door in the opposite direction.



According to the crossing direction indicated by the blue wiggly arrow, the person crossing the door counts as '+1'.



According to the crossing direction indicated by the blue wiggly arrow, the person crossing the door counts as '−1'.

The total of this counting tells us the *net* number of people who crossed the door in the chosen direction. If we chose a crossing direction *from outside to inside* the classroom, then this total is the net number of people who *entered* the classroom. If we chose a crossing direction *from inside to outside* the classroom, then this total is the net number of people who *left* the classroom.

Therefore **a flux represents only a *net* amount crossing a control surface**. Note also that this net amount can be *negative*. For instance if we chose a crossing direction from outside to inside the classroom, and the net amount is −3, then it means that *more people got out than in*: 9 persons may have entered the room during that minute, and 12 persons left. Or maybe no person entered the room, and 3 persons left. In either case, the final situation is that those who got out during that minute were three more than those who got in.

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

- It is completely arbitrary which crossing direction we choose.
- If we choose the other crossing direction, then the net amount will switch sign.

The physical situation is of course the same. The sentences

"+5 persons *entered* the room"

and

"−5 persons *left* the room"

are saying exactly the same thing.

Now consider a similar example, but instead of people, think of a quantity that can ordinarily also be *negative*, such as electric charge. Let's choose the door-crossing direction from outside to inside the room. If we're told that a net amount −5 C of charge crossed the door in the chosen direction in one minute, then this could have happened in several ways:

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

So the statement that "the flux of electric charge into the room was −5 C in one minute" does not tell us which of the situations above occurred.

In fact, ordinary electricity in wires was thought for some time to be associated with movements of negative *and* positive charges in opposite directions. Today we know that it consists in the movement of negative charges only.

"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 purpose of the previous examples is to make you aware of some fundamental aspects of what we call "flux". These aspects are trivial but important when considering fluxes of physical quantities:

Fundamental aspects and symmetry of flux

- A flux in a particular surface-crossing direction only tells us the *net* amount of substance that crosses the surface in that direction per time.
- A flux can be *negative*.
- **Symmetry of flux**: A flux in a particular surface-crossing direction is equivalent to a flux of *opposite* sign in the *opposite* crossing direction.

This is called *Cauchy's fundamental lemma* in the technical literature.

What a flux value does not tell

- A flux value does not tell us the amount that crossed during shorter times or through different parts of the surface.

- A flux value does not tell us whether the transfer of the quantity through the surface occurs because the quantity is flowing, or because the surface is moving, or both.

4.7 Representation of scalar fluxes

How can we graphically represent the flux of a quantity, in a way that takes care of all its fundamental aspects?

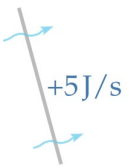
First let's consider a surface through which we're measuring a flux, at a particular instant of time. Here it is represented as line, removing one spatial dimension to simplify the drawing:

Keep in mind that the surface could have any other shape – as long as it can be given a clear crossing direction – and could also be in motion.

Let's indicate a crossing direction through the surface by one or more wiggly arrows:

Keep in mind that these arrows *are not vectors*! They don't have a 'magnitude' or 'components'. They simply indicate a sense in which we imagine the surface to be crossed. We could also have used only one wiggly arrow or three instead of two.

Let's take a scalar quantity such as energy. A flux of energy +5 J/s through the surface, in the first crossing direction, can then be depicted as follows:

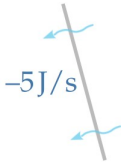


This picture says that a net amount of 5 J is crossing the surface, per second, from the left side to the right side. This also means that per second a net amount of 5J is "disappearing" from the left side of the surface and "appearing" on the right side.

Now consider the opposite crossing direction, depicted like this:

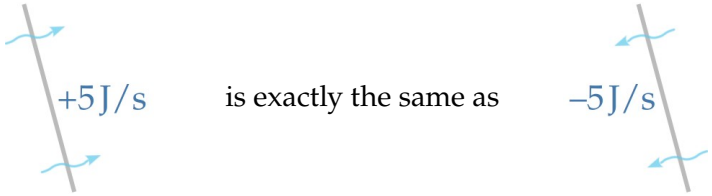
Because of the symmetry of flux, we can say that the flux of energy equals −5 J/s in this opposite direction. We depict this as follows:

› § 4.6 page 87



This picture says that a net amount of −5 has crossed the surface from the right to the left side, in a unit of time. This also means that a net amount of −5 has "disappeared" from the left side of the surface and "appeared" on the right side, in a unit of time.

But this is indeed exactly the same situation as before. *Both pictures therefore represent the same flux*:



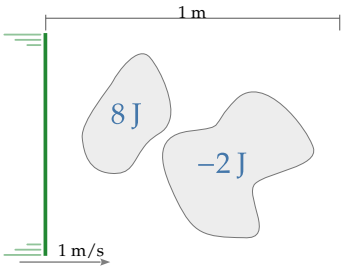
It is extremely important that you remember that the two kinds of picture above are completely equivalent. You can always mentally switch from one to the other. A flux in one crossing direction is exactly the same as a flux with opposite sign in the opposite direction.

The two equivalent pictures do *not* say that a given amount is *only* moving from left to right, or vice versa. We have seen that in general we don't know this. Both pictures say that on the left side the amount of quantity is changing at a rate of −5 J per second, and on the right side by +5 J per second. In these notes we shall usually display only one of these two equivalent representations.

Exercise 4.3

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

1. The two sides of a particular surface are called 'up' and 'down'. During 0.2 s, an energy-mass of +3 J flows from the up-side to the down-side, and an energy-mass of −4 J flows from the down-side to the up-side. How much is the flux of energy-mass through the surface?
2. 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?
3. 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?
4. The side figure shows a control 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-mass as shown (there's no energy-mass behind to the left of the surface). How much is the flux of energy-mass through the surface in 1 s?



4.8 Flux of vector quantities and its representation

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. We can think of it as three fluxes of three scalar quantities.

The intuition and mental representation of the flux of a vector quantity through a control surface may be less straightforward than for a scalar quantity. Think again of the previous examples with people or electric charges crossing a door or control surface. In the case of flux of a vector quantity, we may imagine that what's crossing the control surface are vectors. We are going to discuss some possible graphical representations.

The remarks about the choice of crossing direction and about the symmetry of flux, which we made for scalar quantities, also apply in

analogous ways to the flux of vector quantities. For instance, if the flux of momentum through a surface in a particular crossing direction is

$$[-3, 4, 0]\,\mathrm{N}\,,$$

then if we choose the opposite crossing direction the flux is

$$-[-3, 4, 0]\,\mathrm{N} = [+3, -4, 0]\,\mathrm{N}\,.$$

If we think of vectors as arrows, we must only remember that a minus sign changes their orientation:

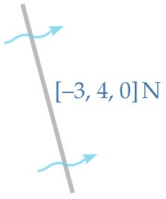
$$\nwarrow \;=\; -1 \cdot \searrow$$

We can devise a graphical representation of the flux of a vector quantity similar to that for the flux of a scalar quantity.

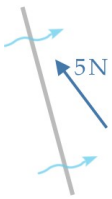
› § 4.7 page 88

First, it's important to indicate the crossing direction, and we can do that again with one or more wiggly arrows; for instance:

Now we have to indicate how much is the flux. In the case of a scalar quantity we simply reported the value, including the unit. For the flux of a vector quantity we have three values, so one possibility is to simply report them. Suppose we are speaking about momentum and the flux in the given crossing direction is $[3, -4, 0]\,\mathrm{N}$; we can then write this explicitly:



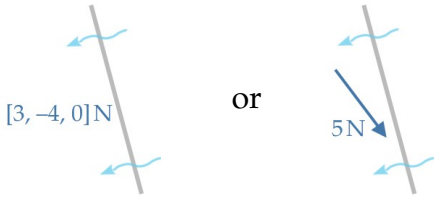
Another alternative is to draw a vector representing these components:



This picture says that a net vector amount of momentum 

is crossing, per second, the surface from the left side to the right side. To help our intuition we can imagine the vector "moving" across the control surface in the direction indicated by the wiggly arrow; an animated representation of this can be found at this link[5].

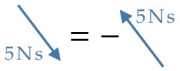
In the opposite crossing direction the flux gets a minus sign, because of the symmetry of fluxes. The corresponding graphical representation is



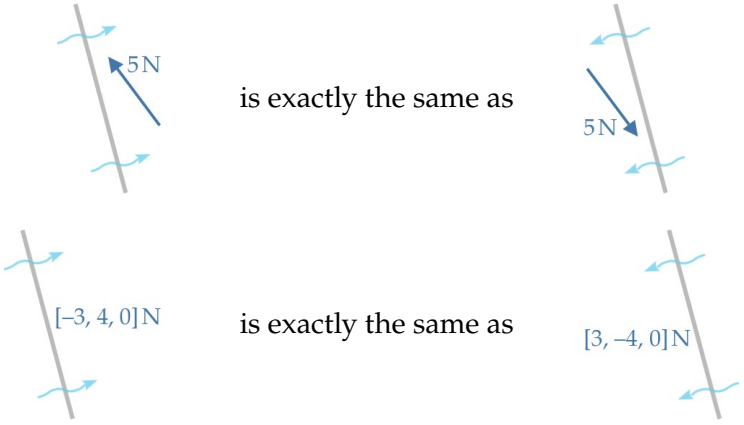
note how the components have flipped sign, and how the vector has flipped direction keeping the same magnitude. The pictures above say that a net vector amount of momentum $[3, -4, 0]\,\mathrm{N\,s}$, or graphically 

, is crossing, per second, the surface from the right side to the left side.

But this is exactly the same flux as before, because

$$[3, -4, 0]\,\mathrm{N\,s} = -[-3, 4, 0]\,\mathrm{N\,s}$$ 

In other words, *the following two pictures represent the same vector flux*:

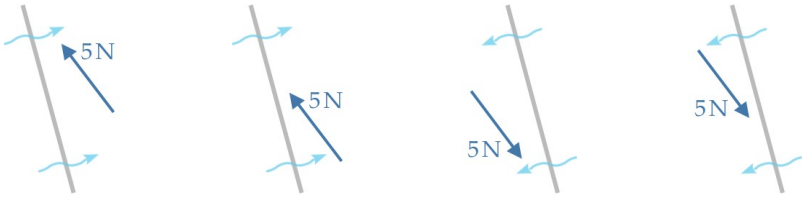


An important aspect of vector fluxes which we must try not to get confused about is the application point of the vector, that is, the base point of the arrow. Just as for vector volume contents, **the application point of**

› § 4.5 page 84

the vector representing the flux is unimportant; the vector refers to one side of the surface as a whole. Graphically:

these four pictures represent exactly the same flux



Exercise 4.4

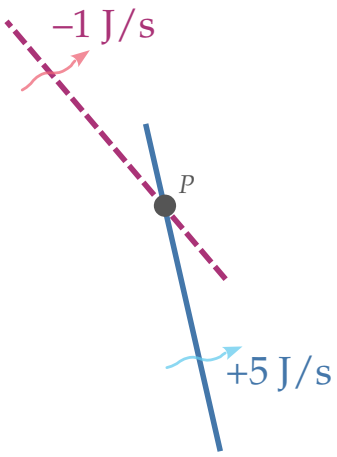
A horizontal surface is given, and there is a flux of a vector quantity through it; for the moment we neglect units:

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  were 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?

4.9 Fluxes through different surfaces

What happens to the value of a flux through a surface, if we consider a *different* surface, maybe intersecting the original one? It's important to keep in mind that

- A flux refers only to a specific surface.
- The fluxes through two distinct surfaces can be very different, even if the two surfaces are quite close.
- The flux depends on the motion of the surface. So if we consider the same surface but with a different instantaneous motion, then the flux may be very different.

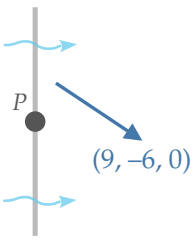


Consider for instance the picture on the side. It depicts two intersecting surfaces (as usual simplified by removing one dimension) and two chosen crossing directions on them. The crossing directions are both roughly rightward. Yet the energy-mass flux through the solid blue surface is $+5\,\mathrm{J/s}$, whereas the energy-mass flux through the dashed red surface is $-1\,\mathrm{J/s}$.

There is, however, a relation between the fluxes through surfaces that share a common point. If we know the flux through *three different, small, static* surfaces having a common point, then we can find the flux through any other *small, static* surface passing through that same point. This possibility leads to the representation of flux through a small surface as a *vector*, called 'flux-density vector'. In these notes we shall not consider flux densities and their vector representation.

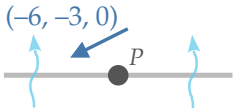
We saw that the flux of a scalar quantity can be very different if we take a slightly different surface, or the same surface with a different motion. The same is true of the flux of a vector quantity: in particular, **the vectors representing the fluxes through two slightly different surfaces can point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

Here is an example. Take a fixed point P. Now take a small vertical surface passing through P, and choose a crossing direction from left to right. The flux of a vector quantity (momentum for example) through this quantity could be as in this picture:



this flux has components $(9, -6, 0)$, with magnitude around 10.8.

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



it has components $(-6, -3, 0)$, with magnitude around 6.7.

Clearly the two vector fluxes through these two surfaces are completely different: they point in different directions and have different magnitudes.

Each surface has a unique flux

A flux, scalar or vector, through a particular surface, in general doesn't tell us anything about the flux though another surface, nor about the flux through the same surface but with a different instantaneous motion.

4.10 Production of momentum: force

We already mentioned that **production 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 10.

› § 3.7 page 64

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

› § 3.7 page 64

Surface forces and volume forces

The notion of force is very intuitive. We associate it to the sensations that we feel in our skin, flesh, bones when, for instance, we push against a wall, twist a door knob, push backwards on the ground with our feet to run or jump. This is what we call a **surface force** or **contact force**.

We also associate force to the sensation we feel in our whole body when we sit in an accelerating car, or go forth and back in a swing, or sit on an aeroplane that's taking off; in these cases the sensation is not limited to a surface. This is what we call a **volume force** or **body force**.

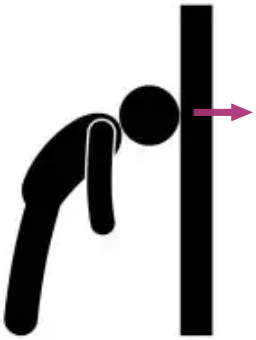
Surface forces and volume forces have some properties in common, because both are production of momentum. Both are represented by a vector having the direction and orientation of the "push" or "pull", and a magnitude expressing its intensity, which is the rate at which momentum is produced.

But they also differ in important respects: surface forces are *flux* of momentum through a surface, whereas volume forces are *supply* of momentum in a volume. Surface forces therefore satisfy the principle of symmetry of flux, whereas volume forces do not always do, or do so only approximately.

› § 4.6 page 87

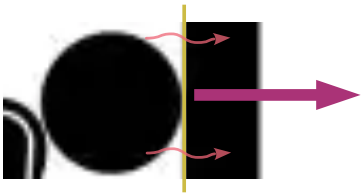
4.11 Flux of momentum is surface force

What we call *surface force* is therefore a *flux of momentum*. This equivalence can be illuminating in some physical problems.



As a concrete example, imagine a person pushing against a wall, as in the side figure. In terms of force, we say that *the person's head is exerting a surface force on the wall*. The vector that represents this pushing force, the purple arrow in the figure, has a person→wall orientation.

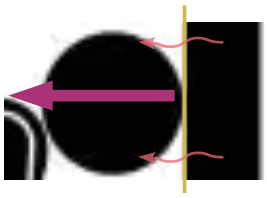
To understand force in terms of momentum flux, let's consider an imaginary control surface separating the person and the wall; the yellow line in the figures below. If we choose a person⇝wall crossing direction, indicated by the wiggly red arrows in the following figure, then the momentum flux indicated by the the purple arrow has a person→wall orientation:



This figure says: on the side of the wall, an amount momentum is being added at a given rate; this momentum has a person→wall orientation.

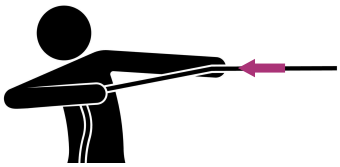
Because of the symmetry of flux, if we instead choose a wall⇝person crossing direction for the control surface, then the momentum flux has opposite orientation:

› § 4.6 page 87



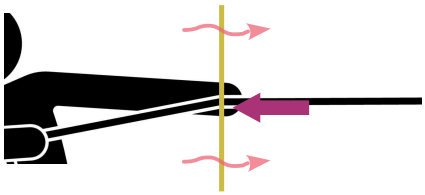
This figure says: on the side of the person's head, momentum is being added at a given rate; this momentum has a wall→person orientation.

Since contact force and momentum flux are the same thing, the last figure tells us that *the wall is exerting a surface force on the person's head*. The vector that represents this opposite pushing force has a wall→person orientation. – But this is *Newton's Third Law*! This example therefore shows that Newton's third law is the expression of the symmetry of flux. We shall discuss this fact again soon.



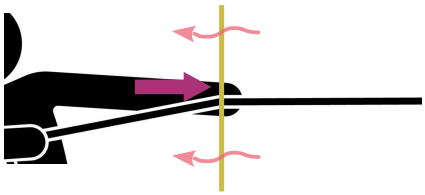
Now let's consider an example involving pulling instead of pushing. Imagine a person pulling a rope fastened somewhere, as in the side figure. In terms of force we say that *the person is exerting a surface force on the rope*; the vector that represents this force has a rope→person orientation.

In terms of momentum flux, let's consider an imaginary vertical control surface between the person's hand and the rope; the yellow line in the figures below. If we choose an arm⇝rope crossing direction, indicated by the wiggly red arrow, then the momentum flux has a rope→arm orientation:



This figure says: on the side of the rope, an amount momentum is being added at a given rate; this momentum has a rope→arm orientation. Note the difference from the previous example involving *pushing*; compare the two momentum fluxes.

But again this momentum flux and the previous figure are completely equivalent to the following one, because of the symmetry of flux:



This figure says: on the side of the arm, an amount momentum is being added at a given rate; this momentum has an arm→rope orientation. Again we find Newton's third law: the rope is exerting a surface force on the person's arm; the vector that represents this opposite pushing force has a rope→arm 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, take an imaginary horizontal control surface between the runner's foot and the ground; the yellow line in the figures below. If we choose a foot→ground crossing direction, the momentum flux is oriented diagonally, downward and backwards, towards the rear of the foot. Because of the symmetry of flux, if we choose a ground→foot crossing direction, then the momentum flux has opposite diagonal direction, upward and forward, towards the front of the foot. This momentum flux can be depicted in these two equivalent ways:

The figures say: on the side of the ground, an amount momentum is being added at a given rate; this momentum has an diagonal, downward and backward orientation. On the side of the foot, an amount momentum is being added at a given rate; this momentum has an diagonal, upward and forward orientation. Both are aspects of the same momentum flux.

Netwon's Third Law!

"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."
Newton 1726a

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

› § 4.6 page 87

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

Exercise 4.5

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

(Buster Keaton in 'The Electric House'[6])

4.12 Pressure, tension, shear force

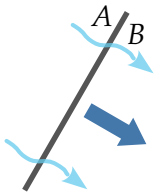
The examples of the previous section demonstrated a variety of possible orientations of the momentum-flux vector with respect to the surface through which it occurs. All orientations of momentum flux are possible. But to momentum flux with three special orientations we give special names: *pressure*, *tension*, and *shear force*.

Take a small surface; it doesn't matter whether it's vertical, horizontal, or whether it has some other inclination. Call the surface's sides A and B. Consider the crossing directions $A \rightsquigarrow B$ and $B \rightsquigarrow A$.

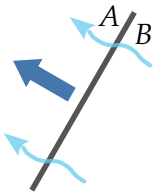
Pressure

Pressure, or *compressive momentum flux* or *compressive force*, is a flux of momentum through a surface, which points *away from* the surface.

So in the crossing direction $A \rightsquigarrow B$, pressure is represented by a vector oriented from A to B. Equivalently, in the crossing direction $B \rightsquigarrow A$ pressure is represented by a vector oriented from B to A:



or equivalently



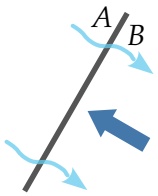
(animated version[7])

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.

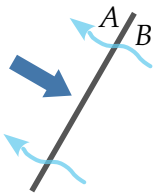
Tension

Tension, or *tensile momentum flux* or *tensile force*, is a flux of momentum through a surface, which points *towards* the surface.

So in the crossing direction $A \rightsquigarrow B$, tension is represented by a vector oriented from B to A. Equivalently, in the crossing direction $B \rightsquigarrow A$ tension is represented by a vector oriented from A to B:



or equivalently



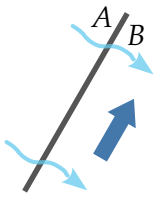
(animated version[8])

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.

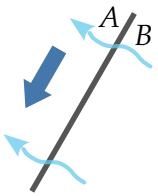
Shear force

Shear force, or *shearing momentum flux*, is a flux of momentum through a surface, *parallel* to the surface.

In the crossing direction $A \rightsquigarrow B$ and in the crossing direction $B \rightsquigarrow A$, shear force is represented by a vector parallel to the surface. Obviously the orientation is opposite on the two sides, owing to the symmetry of flux:



or equivalently



(animated version[9])

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.

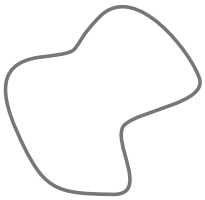
In general, a momentum flux won't have any of the three special directions above, but rather a combination of them.

Exercise 4.6

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

4.13 Closed control surfaces, influxes, effluxes

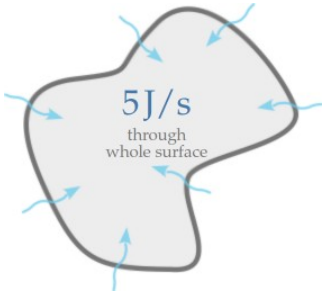
We shall often consider **closed** control surfaces, that is, control surfaces that don't have a rim or border or holes, like the surface of a sphere or of a cube. A closed surface delimits a specific three-dimensional volume, and we can therefore speak of its **interior** and its **exterior**. An example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

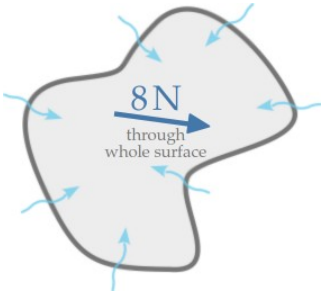


The two crossing directions of a closed surface therefore take on special names: *inward*, from exterior to interior; and *outward*, from interior to exterior. A flux through the surface is usually called **influx** if we are considering the inward crossing direction, and **efflux** or *outflux* if we are considering the outward crossing direction. Obviously, by the symmetry of flux,

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

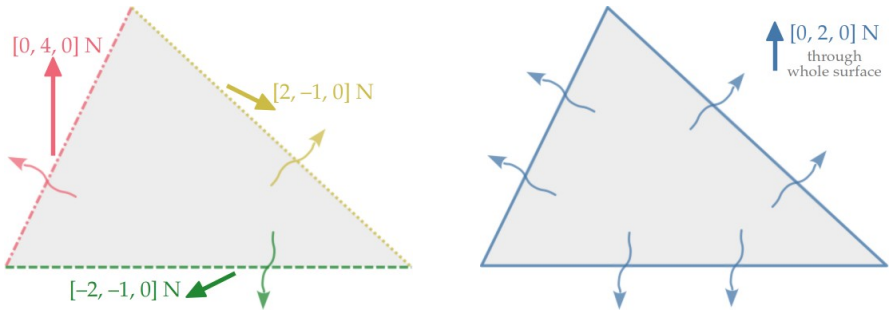
The influx and efflux are fluxes *through the whole surface*. Consider for instance these influxes of energy-mass and of momentum:





In the left picture we have a net influx of 5 J/s on the whole inner side of the control surface. Remember that we don't know whether these 5 J/s are being evenly distributed over the surface, or just at particular spots of it, or whether they are the net result of positive and negative amounts on different parts of the surface. In the right picture we have a vector influx of 8 N, the vector pointing approximately rightward. Again we don't know what are the flux vectors for parts of the surface: the vector in the picture is just the grand total.

Let's see another example of this fact. The picture on the left below shows the outward fluxes through three parts (dotted yellow, dashed green, dot-dashed red) of a closed control surface. The picture on the right shows the total efflux through the same surface:

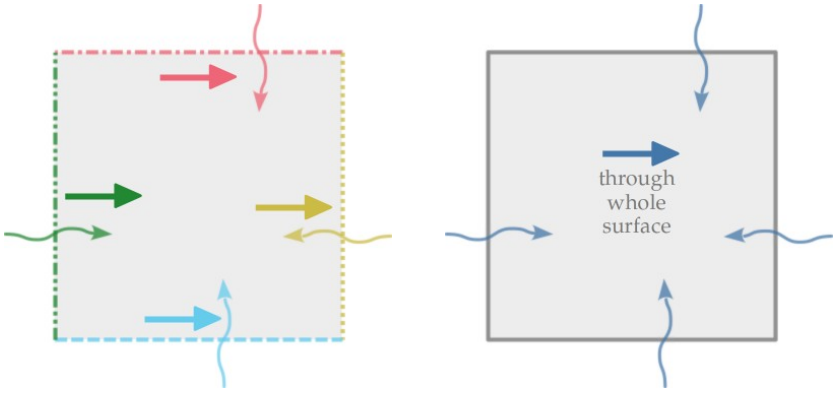


The individual fluxes and the total efflux are consistent because

$$[2, -1, 0]\,\mathrm{N} + [-2, -1, 0]\,\mathrm{N} + [0, 4, 0]\,\mathrm{N} = [0, 2, 0]\,\mathrm{N}$$

Exercise 4.7

1. Are the four partial influxes shown on the left (with different colours and line styles) consistent with the total influx shown on the right? Why or why not?



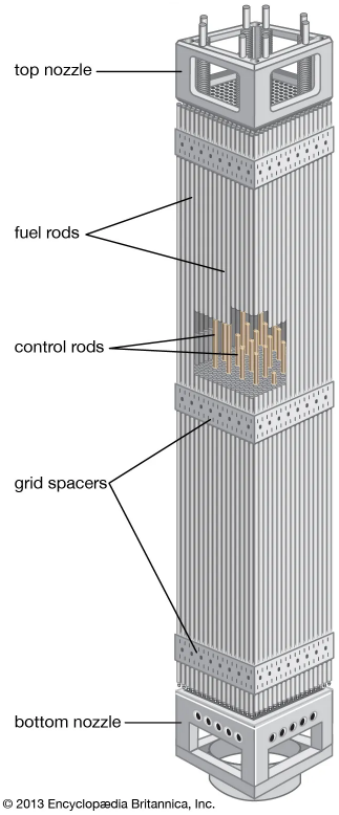
2. Take an imaginary cylindrical surface enclosing one control rod[10] in a nuclear-fission reactor[11] (see side figure). Let's say that in a reactor there are 20 such rods. Approximately 5×10^{19} neutrons are liberated in a second in the whole reactor by the fission fuel, but 2/3 of these are *absorbed*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

How much, on average, is the **ef**flux 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

$$N_{\mathrm{A}} = 6.022\,140\,76\times10^{23}\ \text{particles/mol}\ .$$

Be careful about the signs!



4.14 Time-integrated fluxes

A flux is defined as the amount of a quantity crossing a control 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 $J\,\Delta t$.

Now consider a control surface between two time instants t_0 and t_1; during this time lapse it could be moving and changing shape. Choose a crossing direction through the surface. At each intermediate time instant t we can measure the flux of a quantity crossing the surface in that direction, at that instant; denote this flux 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 then obtained by adding these small amounts. As we consider shorter and shorter Δt, this sum is by definition an integral:

Time-integrated flux

The total amount of quantity flowing through a control surface in a specified crossing direction, between times t_0 and t_1, is called the

time-integrated flux and is given by

$$\int_{t_0}^{t_1} J(t)\,\mathrm{d}t\ , \tag{4.1}$$

where $J(t)$ is the flux of the quantity at time t.

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 short vectors.

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

$$\int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t \coloneqq \left[\int_{t_0}^{t_1} F_x(t)\,\mathrm{d}t\ ,\ \int_{t_0}^{t_1} F_y(t)\,\mathrm{d}t\ ,\ \int_{t_0}^{t_1} F_z(t)\,\mathrm{d}t\right]. \tag{4.2}$$

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 flowed through 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, in order for the integral to be zero.

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 control surface in the case of matter, finding a total of $\int_{t_0}^{t_1} J(t)\,\mathrm{d}t = 7\,\mathrm{mol}$. Now we change our mind and choose the opposite crossing direction for that surface. How does the result above change?

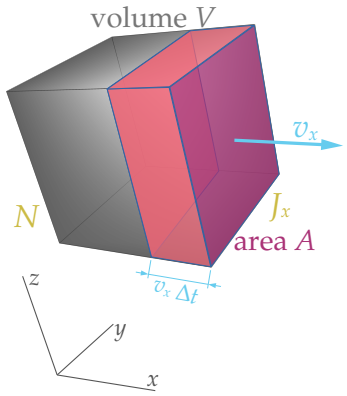
4.15 The relation between fluxes and velocities

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

The answer is yes: the velocity of a quantity is essentially defined its flux and its 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 somewhat involved, so here we'll just see a simplified and approximate example of how this definition works.

Take a scalar extensive quantity like matter, electric charge, energy-mass, 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 coordinate time. Around this point, choose a very small static cuboid region, as in the side figure, delimited by six small rectangular static surfaces: two parallel to the yz-coordinate plane, two to the zx one, and two to the xy one. The cuboid has volume V, and the volume content of matter 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.



This control volume is small, static, and with sides parallel to the coordinate axes.

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

$$v_x := \frac{J_x/A}{N/V} \tag{4.3}$$

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

The velocity $\boldsymbol{v} = [v_x, v_y, v_z]$ so defined has the following intuitive property. If you choose *any* very small surface centred at this point, and move it with velocity $\boldsymbol{v}$, 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

1. 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?

2. A small cuboid region has a volume of 1×10^{-9} m^3, and its sides parallel to the xy axes have each area 1×10^{-3} m^3. Through each of these sides there is a flux of energy-mass $\Phi_z = 3$ J/s in the positive z-direction. The cuboid region contains $E = 0.5$ J of energy-mass. How much is the velocity of energy-mass in the z-direction?

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-mass* from the flux of energy-mass.

In Newtonian approximation these two velocities are approximately equal, so we do not need to distinguish them. But in situations where the Newtonian approximation is not valid, we have to take into account the velocity of matter and the velocity of energy-mass 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.

URLs for chapter 4

1. https://www.turbosquid.com/3d-models/classroom-1726208
2. http://www.009.cd2.com
3. https://www.grc.nasa.gov/WWW/k-12/VirtualAero/BottleRocket/airplane/flap.html
4. http://www.youtube.com/watch?v=M5xnAdVPbgQ
5. https://pglpm.github.io/7wonders/media/vectorfluxanim2.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

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 Some classifications of physical laws

Physical laws, very generally speaking, are mathematical relations between physical quantities. Given information about some quantities, physical laws allow us to deduce information about other quantities, or about the same quantities at other times or spatial regions. As previously discussed, we use many physical laws every day, in a qualitative and approximate way, without even thinking about them.

› § 1.3 page 19

Physical laws can be classified or categorized in many different ways; for instance by the quantities they involve, or by the kind of mathematics they use. So we speak of 'laws of mechanics' and 'electromagnetic laws'; or of 'differential laws' and 'integral laws'; and so on.

One classification distinguishes between *fundamental* vs *derived* physical laws. This distinction is similar to the one between primitive and derived quantities. A physical law is 'fundamental' if it is taken as empirically valid, and as the starting point to make predictions or calculate other kinds of consequences. A physical law is 'derived' if it can be deduced from other fundamental laws: in a manner of speaking, it doesn't really say

› § 1.4 page 21

anything new that wasn't already a consequence of the fundamental laws. But it may still be a very useful shortcut.

Here is an extremely simple imaginary example. Suppose we have three different physical quantities denoted by a, b, c. One fundamental physical law states that a and b are equal; another fundamental law states that b and c are equal too:

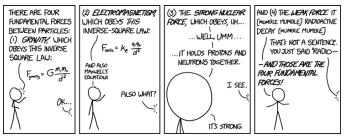
$$a = b\,, \quad b = c \qquad \text{(fundamental laws)}$$

Combining these two fundamental laws we obtain a further law: a is equal to c:

$$a = b\,, \;\; b = c \quad \Longrightarrow \quad a = c \qquad \text{(derived law)}$$

This latter physical law doesn't tell us anything new that wasn't already implicit in the first two together. But in some situations it is useful to simply remember directly that $a = c$.

The distinction between fundamental and derived physical laws is not objective, but mostly a matter of convenience and even of personal taste. We can often promote a derived law to fundamental law, demoting some other fundamental law to a derived-law status. In the example above, we could take '$a = b$' and '$a = c$' as fundamental laws; then '$b = c$' becomes a derived law, because it can be obtained from the other two. It is important to be aware of this flexibility in what's fundamental and what's derived. You'll find physics and engineering texts that present a physical theory as consisting in a particular collection of fundamental laws, and other texts that present the same physical theory as consisting in a collection of slightly different fundamental laws. There is no contradiction there: it means that some laws taken as derived in some texts, but as fundamental in others, and vice versa.



https://xkcd.com/1489

Yet, the choice fundamental laws is not without consequence. A particular choice of fundamental laws, rather than others, may cover more physical situations. It may also suggest new physical ideas and generalizations, leading to the discovery of new physical phenomena.

5.2 Universal laws vs constitutive relations

Another important distinction can be made between **universal** physical laws and **constitutive** physical relations. This distinction is determined not by convenience and personal taste, but by experiment:

Universal laws vs constitutive relations

- **Universal laws** represent universal physical patterns that we observe in all possible physical phenomena we manage to investigate
- **Constitutive relations** represent physical patterns and physical properties that are peculiar to – and therefore are only valid for – specific phenomena, or specific scales of time and space, or specific kinds of control volumes and surfaces, or specific physical theories. Constitutive relations are also called *constitutive equations* or *closure equations*.

One of the meanings of the word *constitutive* is 'that makes a thing what it is' (*Oxford English Dictionary* 2009).

The distinction above is different from the one between 'fundamental' and 'derived'. Let's try to understand this difference by means of our previous simple example.

Suppose we observe that the law '$a = b$' is always valid in all physical phenomena we explore, under all possible extreme conditions, circumstances, regions of space, and time. Also suppose we observe that '$b = c$' is instead only valid in specific physical phenomena and conditions, but not in others. For instance, we may observe that it is only valid when we make experiments with gases and low speeds, but not with solids or high speeds. The conclusion (until we find a disproof) is that the physical law '$a = b$' is *universal*, whereas the physical law '$b = c$' is *constitutive*: constitutive of gases and low speeds. Yet both laws are fundamental, because we cannot deduce '$b = c$' from '$a = b$': the law about b and c says something new, although something that is true only in some circumstances. In concrete applications we may therefore need to use both '$a = b$' and '$b = c$'.

But are there really universal physical laws, which can be applied to *every* physical phenomenon, without exclusions or exceptions?

The answer is *yes*. We shall soon meet them, and we shall see that they are tightly connected with the seven primitive quantities discussed in Chapter 3. Our study of physics will indeed completely hinge on these universal laws. They are applied with extreme confidence to every new phenomenon we observe, and they often allow us to make predictions of at least a qualitative character without the need of constitutive relations. We would modify these universal laws only as a last resort; so far this has rarely or never been necessary. On the other hand, we have a large freedom in modifying constitutive relations, and in proposing new ones to account for newly observed physical phenomena.

› § 5.8 page 131

As real examples, the *balance of momentum*

› § 10.1 page 168

$$\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t} = \boldsymbol{F}(t) + \boldsymbol{G}(t)$$

is a universal physical law: it applies – exactly as written – to any control volume, small as an atom or large as a group of galaxies; to all objects and to electromagnetic fields; to General Relativity and to Newtonian mechanics, and with some reinterpretation of the symbols also to Quantum Theory. Instead the *Newtonian relation between momentum content and matter flux*

› § 10.2 page 169

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

is a constitutive relation: it applies to enough small control volumes, but not to large ones; it applies to matter, but not to electromagnetic fields; it applies at low speeds and weak gravitational fields, but not at high speeds or strong gravitational fields; it is used in Newtonian mechanics, but not in General Relativity.

In §*** we shall discuss further important differences between constitutive relations and a particular kind of universal laws called *balance laws*, which we discuss next.

'Fundamental' vs 'universal' in other texts

Be aware that the terms 'fundamental' and 'universal' are typically not used in a technically precise sense. Some texts may even use 'fundamental' in the sense of 'universal' as done here. What is important for us to remember is that there are physical laws that are used in *all* situations, and physical laws that can be used only in a restricted range of situations.

Exercise 5.1

In our imaginary example, the law '$a = b$' is fundamental and universal, and the law '$b = c$' is fundamental and constitutive. Is the derived law '$a = c$' universal? or is it constitutive?

5.3 Balance and conservation laws

There are physical laws which have a special meaning and mathematical form: they express a sort of trade-off or "budget" among different amounts.

Their basic idea is therefore quite intuitive. They are called **balance laws**, and a special subgroup of them are called **conservation laws**.

Balance laws concern *extensive* quantities, like our seven primitive quantities. Recall that for extensive quantities we may ask: 'how much is there in a particular control volume?', 'how much is flowing through a particular control surface?', 'how much is being produced in a control volume?'

› § 3.1 page 52

Balance and conservation laws

A **balance law**, or simply *balance*, expresses a relation between the volume content, flux, and supply of an extensive physical quantity.

A **conservation law** is a special and important kind of balance law, in which the supply is always zero.

It turns out that *we can formulate almost all known universal laws as balance laws*. For this reason we now study balance laws in detail.

'Conservation' vs 'balance' in other texts

Be aware that some physics and engineering texts do *not* distinguish between 'conservation law' and 'balance law', and often use "conservation" to mean both. This use is unfortunate, because balance is the most general of the two. Always try to infer from the context how these two terms are used.

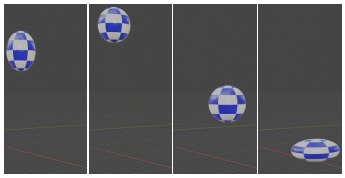
General setup for balance and conservation laws

The general setup to formulate a balance or a conservation law is as follows.

We choose a *closed* control surface that exists between coordinate times t_0 and t_1. Being closed, this control surface delimits a control volume. This surface and volume can move and deform arbitrarily in between the two times; what's important is that the control surface remains closed: no holes or cuts can form. At the final time t_1 the closed control surface and volume can be very different from how they were at the initial time t_0, and even at different positions. We are completely free in our choice of location, size, shape, motion of this control surface and volume.

A very simple example could be a spherical surface delimiting a ball-shaped region. But we could as well consider a surface that initially is spherical and delimits a region the size of a tennis ball, and at the end is

› § 4.13 page 101



t_0 t_1

Four snapshots of a moving and deforming closed control surface (animated version[1])

cubical and delimits a region the size of our galaxy. The closed control surface doesn't even need to be connected: it could consist of several separate closed control surfaces, each delimiting a separate volume.

Now consider an extensive quantity; specifically take any one of our primitive quantities, except magnetic flux. For concreteness let's take some kind of *matter*, but keep in mind that it could be energy-mass, or momentum, or any of the other six. Recall that we denote the volume content of matter by N, the flux by J, the supply by A.

› § 3.1 page 52

Make the following measurements:

- $N(t_0)$: total *volume content* of the quantity in the control volume at time t_0.
- $\int_{t_0}^{t_1} J(t)\,\mathrm{d}t$: total *time-integrated* ***in****flux* of the quantity through the closed control surface between times t_0 and t_1.
- $\int_{t_0}^{t_1} A(t)\,\mathrm{d}t$: total *time-integrated supply* of the quantity through the closed control surface between times t_0 and t_1.
- $N(t_1)$: total *volume content* of this quantity in the control volume at time t_1.

Note that in the second measurement we measure the **in**flux, not the efflux, through the closed control surface. We must pay close attention to the crossing direction and sign of this flux.

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

$$t_0 = 0\,\mathrm{s} \qquad N(t_0) = 10.5\,\mathrm{mol}\,,$$

$$\int_{t_0}^{t_1} J(t)\,\mathrm{d}t = -9.7\,\mathrm{mol}\,,$$

$$\int_{t_0}^{t_1} A(t)\,\mathrm{d}t = 0\,\mathrm{mol}\,,$$

$$t_1 = 18\,\mathrm{s} \qquad N(t_1) = 0.8\,\mathrm{mol}\,.$$

But if the quantity in question were a vector, like momentum, then each of the three amounts above would be a vector, with any combination of positive, negative, or zero components.

With this setup we first study a special kind of balance law.

5.4 Conservation laws

Conservation laws tightly connect *volume contents* and *fluxes*:

› § 4 page 75

Conservation law

A quantity is said to satisfy a **conservation law**, or to be **conserved**, if the following equality between its volume content and time-integrated influx holds for *any closed control surface and volume*, and *any coordinate times* t_0 and t_1:

$$\text{volume content}(t_1) = \text{volume content}(t_0) + \int_{t_0}^{t_1} \text{influx}(t)\,\mathrm{d}t \tag{5.1}$$

For example, in the case of matter with volume content N and influx J, the conservation law would be

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t)\,\mathrm{d}t\ .$$

The meaning of a conservation law is intuitive. Let's take again the example of some kind of matter. 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 that flowed in through the surface between those times, the integrated flux $\int_{t_0}^{t_1} J(t)\,\mathrm{d}t$. 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. Note that these amounts can be positive, negative, or zero.

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

- If we know the amount of quantity in the control volume at t_0, and the net amount that entered through the control surface between t_0 and t_1, then we can predict the amount in the control volume at t_1: it is given explicitly by the equation above.
- If we know the amount of quantity in the control volume at t_0, and the one at t_1, then we can deduce the net amount that entered through the control surface between the times t_0 and t_1:

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

- If we know the amount of quantity in the control volume at t_1, and the net amount that entered through the control surface between t_0 and t_1, we can deduce the initial amount in the control volume at time t_0:

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

These were just examples. Since a conservation law involves different times and a time integral, it also allows us to make deductions about how volume contents or fluxes depend on time, and to predict the time when a volume content or flux has a particular value.

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 later sections.

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

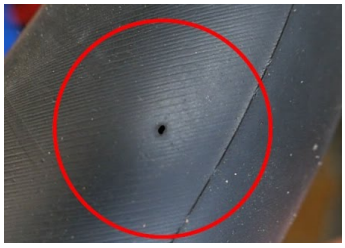
$$t_0 = 0\,\mathrm{s}\ , \quad t_1 = 1800\,\mathrm{s}\ , \quad N(t_0) = 12.6\,\mathrm{mol}\ , \quad N(t_1) = 0.5\,\mathrm{mol}\ .$$

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

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

The result, with a negative sign, says that 12.1 mol of air have crossed the moving control surface in an *outward* direction, during the thirty minutes.

Obviously this happened because the tyre tube has a hole somewhere. 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 track of what's crossing it, how much, and how fast.

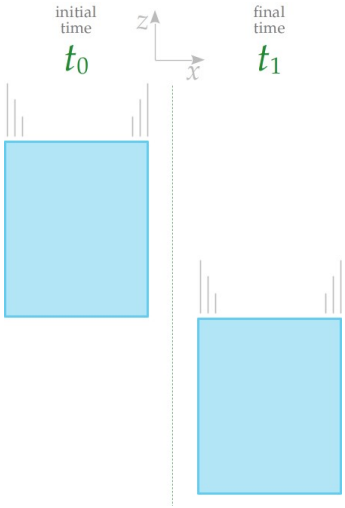


There's a physical hole in the physical tube, but no holes in the imaginary control surface that wraps the tube

The fact that air satisfies, in this example, 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 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.

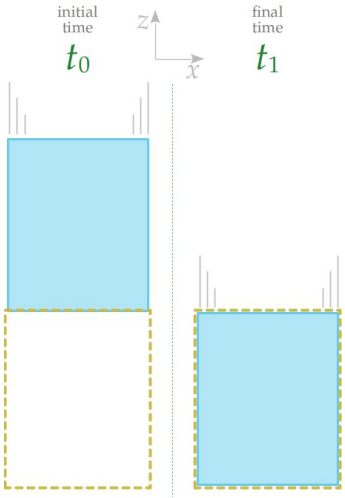
Example with a static control surface

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 closed control surface that is static and doesn't follow or wrap any object.



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, in a vacuum, owing to gravity. At a particular time t_0 it occupies a particular position. At a time t_1, 0.5 s later, it is situated for an instant immediately underneath the initial position, as illustrated in the side figure.

Let's arbitrarily choose a static, closed control surface of cylindrical shape, located in such a way that it wraps the block of ice at the final time. At the initial time this control surface is therefore empty. See the side figure below, where this imaginary control surface is represented by the dashed yellow line.



Suppose we want to know the net amount of ice that entered this control surface between times t_0 and t_1. We can calculate this quantity by assuming that ice – as matter – satisfies a conservation law. 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 time-integrated **in**flux:

$$\begin{aligned}\int_{t_0}^{t_1} J(t)\,\mathrm{d}t &= N(t_1) - N(t_0)\\ &= 53.4\,\text{mol} - 0\,\text{mol}\\ &= 53.4\,\text{mol}\end{aligned}$$

and this is the net amount of ice that entered the control surface during the 0.5 s.

Note that what we found is the time-integrated influx through the *whole* closed control surface. The conservation law doesn't tell us anything about the integrated influx through *parts* of the surface. In the present case we can divide the surface into three parts: a circular surface at the bottom, a side surface, and a circular surface at the top. If we provide the extra knowledge that the fluxes through the side and bottom surfaces are zero, then by the extensivity property we have

› § 3.2 page 54

$$\int_{t_0}^{t_1} J(t)\,\mathrm{d}t = \int_{t_0}^{t_1} \big[J_{\text{top}}(t) + J_{\text{side}}(t) + J_{\text{bot}}(t)\big]\,\mathrm{d}t$$
$$53.4\ \text{mol} = \int_{t_0}^{t_1} \big[J_{\text{top}}(t) + 0\ \text{mol/s} + 0\ \text{mol/s}\big]\,\mathrm{d}t$$

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 5.2

Solve the following exercise not just by using intuition, but ***by explaining step-by-step how you use a conservation law*** *to obtain the result:*

- *What is the relevant time interval?*
- *How do you define the closed control surface and and its movement, as well as any subdivisions of the surface?*
- *What are the values of volume content and of time-integrated flux known to you? Which ones do you want to find?*

For the present exercise we assume that energy-mass 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-mass flows into the room: the energy-mass is entering around the electric wires in the form of electromagnetic energy-mass, and is converted into internal energy-mass (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-mass can flow in or out.

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

5.5 Balance laws

Our intuitive understanding of a conserved quantity is that it cannot be "created" or "destroyed"; it can only "move around". But there are quantities that are not conserved. Such quantities satisfy a *balance* law. They could be produced or disappear in a control volume; recall that we do have a term and measurement for this kind of production or disappearance: the supply.

› § 3.2 page 53
› § 5.3 page 112

To formulate a general balance law we use the same setup as for a conservation law. We take a closed control surface, delimiting a control volume, between coordinate times t_0 and t_1. Having chosen one of the seven quantities except the magnetic flux, we measure its *volume content* at time t_0, its *time-integrated influx* between times t_0 and t_1, its *time-integrated supply* between times t_0 and t_1, and its *volume content* at time t_1.

Balance law

A quantity is said to satisfy a **balance law**, or to be **balanced**, if the following equality holds for *any closed control surface and volume*, and *any coordinate times* t_0 and t_1:

$$\text{volume content}(t_1) = \text{volume content}(t_0) + \int_{t_0}^{t_1} \text{influx}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \text{supply}(t)\,\mathrm{d}t \tag{5.2}$$

For example, in the case of energy-mass with volume content E, influx Φ, supply R, the balance law would be

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t\ .$$

The meaning of a balance law 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: $\int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t$, plus the total amount that was created in the volume: $\int_{t_0}^{t_1} R(t)\,\mathrm{d}t$.

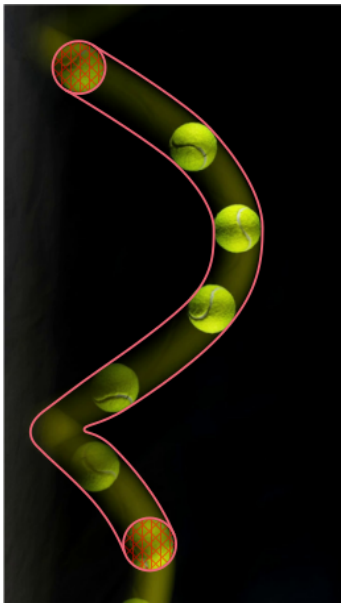
We see that a *conservation* law is a special, powerful case of a balance law. Let's make this connection explicit:

Connection between balance and conservation laws

A quantity is said to satisfy a *conservation law* if it satisfies a *balance law* and **its supply is always zero**, for any closed control surface and volume and any coordinate times t_0 and t_1.

A conservation law is powerful because it means that the supply is always known in advance and has an extremely simple value: zero. The important consequence of this fact is that it allows us to predict the amount of quantity in a final volume by knowing what happens *only on the boundary of the volume* during the time lapse. A general balance law, instead, requires us to know also what happens at every point *within the volume* during the time lapse: we need to know whether some amount of quantity was created or destroyed within the volume.

For these reasons a balance law is in some respects more trivial than a conservation law. If we measure that $E(t_0) - E(t_1) - \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t$ 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 law. A balance is not trivial, however, if we have some other physical law that tells us in advance how the amount created or destroyed at each instant, $R(t)$, can be calculated.



A conservation law involves only knowledge about the initial and final control volumes (represented by the red hatched disks), and about the closed control surface during the time lapse (curved red contours); but not about the control volume during the time lapse. A balance law instead involves also this additional information.

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 (5.2)? Wouldn't you get

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \big[\Phi(t) + R(t)\big]\,\mathrm{d}t = 0$$

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, instead, involves what's happening *within the control volume*. A conservation law does *not* require us to know what's happening within the control volumes, except at the initial and final time.

Remember, moreover, that fluxes always satisfy a symmetry principle, but *supplies in general do not satisfy any analogous symmetry principle*.

› § 4.6 page 87

Balance law for vector quantities

Let us now consider an extensive vector quantity like *momentum*, which will be very important in our future investigations. Its initial volume content is denoted $\boldsymbol{P}$; its influx, also called surface force, $\boldsymbol{F}(t)$; and its supply, also called body force, $\boldsymbol{G}$. They are time-dependent vectors, defined in a coordinate system (t, x, y, z):

$$\boldsymbol{P}(t) = \begin{bmatrix} P_x(t) \\ P_y(t) \\ P_z(t) \end{bmatrix} \qquad \boldsymbol{F}(t) = \begin{bmatrix} F_x(t) \\ F_y(t) \\ F_z(t) \end{bmatrix} \qquad \boldsymbol{G}(t) = \begin{bmatrix} G_x(t) \\ G_y(t) \\ G_z(t) \end{bmatrix} .$$

The balance law for a vector quantity like momentum has exactly the same expression we already know:

$$\boldsymbol{P}(t_1) = \boldsymbol{P}(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t \tag{5.3}$$

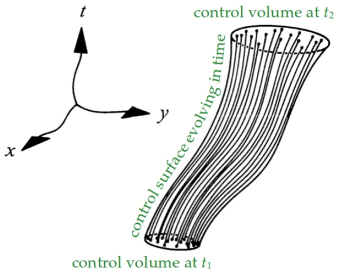
with the only difference that the quantities involved are vectors. So it corresponds to a system of three balance equations, one per component:

$$\begin{cases} P_x(t_1) = P_x(t_0) + \displaystyle\int_{t_0}^{t_1} F_x(t)\,\mathrm{d}t + \int_{t_0}^{t_1} G_x(t)\,\mathrm{d}t \\ P_y(t_1) = P_y(t_0) + \displaystyle\int_{t_0}^{t_1} F_y(t)\,\mathrm{d}t + \int_{t_0}^{t_1} G_y(t)\,\mathrm{d}t \\ P_z(t_1) = P_z(t_0) + \displaystyle\int_{t_0}^{t_1} F_z(t)\,\mathrm{d}t + \int_{t_0}^{t_1} G_z(t)\,\mathrm{d}t \end{cases} \tag{5.4}$$

Balance laws in General Relativity

Conservation and balance laws appear 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 moving 2D surface followed through time is also a spacetime region that has one less dimension than spacetime: two spatial dimensions and one temporal one. Thus the distinctions among the 3D region at t_0, the moving 2D surface 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 of it 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*, and with a different notion of orientation: we consider a control surface that exists between times t_0 and t_1; this surface has a closed control curve as boundary. The magnetic flux turns out to be a quantity for which it's possible to ask how much of it is intersecting a surface, 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 will be discussed in depth in Chapter 9.

5.6 Examples

Let us now study a couple of example applications of balance laws.

In order to apply a balance law we must choose one or more closed control surfaces and corresponding control volumes. Fix a coordinate system (t, x, y, z). We previously discussed that two typical choices of control volumes and surfaces with respect to a coordinate system are: (a) a moving one, usually "wrapping" some solid object; (b) a static one. Let's see how a balance law is applied in these two cases. For the quantity to study we choose momentum, so that we can get immediately get acquainted with handling vector equations.

› § 4.4 page 82

Moving control surface

Let's say that coordinates x, y have horizontal directions, and z an upward direction. Consider a flying tennis ball. Choose an imaginary, closed control surface that perfectly wraps the tennis ball and moves with it.

At a particular time instant t_0 the tennis ball has momentum $[0, 1.70, 0.98]\,\mathrm{N\,s}$; by this we mean that the control volume corresponding to the ball contains that amount of momentum.

While the tennis ball is flying, we assume that the net influx of momentum through the control surface is zero, for instance because the ball is in a vacuum. But there's a supply of momentum within the volume, constant in time, equal to $[0, 0, -0.579]\,\mathrm{N}$. This supply, as we'll see later, exist because the tennis ball is in the gravitational field of the Earth.

The tennis ball flies for two seconds. How much is the momentum of the tennis ball at the end of this time lapse? Let's call this time t_1.

From the description above we have these data:

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

$$\boldsymbol{P}(t_0) = \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \mathrm{N\,s}\,,$$

$$\boldsymbol{F}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\ (const.\ in\ time)}\,, \qquad \boldsymbol{G}(t) = \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \mathrm{N\ (const.\ in\ time)}\,.$$

The balance law (5.3) allows us to find the amount of momentum in the tennis ball at time $t_1 = 2\,\mathrm{s}$:

$$\begin{aligned} \boldsymbol{P}(t_1) &= \boldsymbol{P}(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t \\ &= \boldsymbol{P}(t_0) + \boldsymbol{F}\cdot(t_1 - t_0) + \boldsymbol{G}\cdot(t_1 - t_0) \qquad \text{(because } \boldsymbol{F}, \boldsymbol{G} \text{ are constant in time)} \\ &= \begin{bmatrix} 0 \\ 1.70 \\ 0.98 \end{bmatrix} \mathrm{N\,s} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N}\cdot 2\,\mathrm{s} + \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \mathrm{N}\cdot 2\,\mathrm{s} \\ &= \begin{bmatrix} 0 \\ 1.70 \\ -0.18 \end{bmatrix} \mathrm{N\,s}\,. \end{aligned}$$

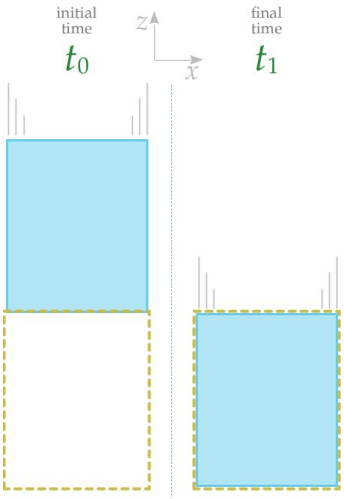
Therefore two seconds later the ball's momentum has the same x- and y-component as it had initially; but now its z-component points

downward – which means that the ball is also moving downward, not only horizontally.

Static control surface

Take again the previous example of a cylindrical block of ice moving downward, in a vacuum, during a lapse of time of 0.5 s, depicted again in the side figure. In the previous example we discussed the net amount of matter that crosses a static, closed control surface at the final location of the ice block. Now we are instead interested in the net amount of *momentum* that crosses the same surface. We shall therefore use the balance of momentum (5.3).

› § 5.4 page 116



Suppose we have this information:

$$t_0 = 0\,\mathrm{s}\,, \quad t_1 = 0.5\,\mathrm{s}\,, \quad \boldsymbol{P}(t_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\,s}\,, \quad \boldsymbol{P}(t_1) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s}\,.$$

The expression for $\boldsymbol{P}(t_0)$ says the initial momentum content in the control volume is zero. This makes sense, because the control volume is initially empty of matter. The expression for $\boldsymbol{P}(t_1)$ says that the final momentum content in the control volume is non-zero: it directed fully downward, with magnitude 4.72 N s. This also makes sense, because at this time in the control volume there's matter moving downward.

Can we find the time-integrated influx or supply of momentum during the time lapse of 0.5 s? Strictly speaking, from the data mathematically expressed above, the answer is no: to find the time-integrated influx we would need the time-integrated supply, and vice versa. This situation illustrates what we said previously: balance laws generally require more information than conservation laws.

Suppose we are told that the time-integrated supply of momentum *within the control volume*, during the time lapse, is directed downward and has magnitude 1.57 N s. That is,

$$\int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \mathrm{N\,s}\,.$$

Pay attention to the fact that this is *not a flux* of momentum: this is not momentum that "enters" from the top of the control surface because the ice block is entering there. This is momentum created (by gravity) in the parts of the volume where there is matter.

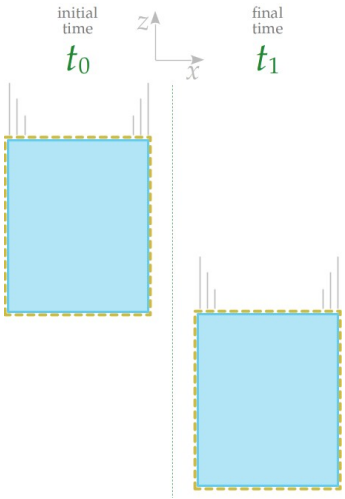
The influx through the top surface can now be found using the balance law (5.3):

$$\begin{aligned}\int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t &= \boldsymbol{P}(t_1) - \boldsymbol{P}(t_0) - \int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\,s} - \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \mathrm{N\,s} \\ &= \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \mathrm{N\,s}\end{aligned}$$

This is the momentum that "enters" through the closed control surface. We can intuitively deduce that it is specifically entering through the top part of the control surface.

Another example with a moving control surface

Let us consider the same physical situation as in the previous example, but now let's choose a *moving* control surface instead, as in the first example with the tennis ball. We choose a closed control surface that tightly wraps the block of ice at all times between t_0 and t_1.



An important remark: since we are now choosing a different control surface and volume from the previous example, the values of momentum contents and fluxes may be different from the previous ones as well – they refer to different regions of space. To avoid getting confused, it's good to denote the amounts in the present example with different symbols. Let's underline them for instance; we could also use some other graphical symbol, or simply change the letters themselves.

Suppose that we want to know how much is the time-integrated supply of momentum in the new, moving control volume, between times t_0 and t_1. In order to find it from the balance of momentum we need to know:

(a) the initial momentum content, (b) the final momentum content, (c) the time-integrated influx.

We are told that the ice block initially has zero momentum (because it's released, and therefore has zero velocity, exactly at that instant), and at the final time it has a downward momentum of magnitude 4.72 N s. We are also told that there is no net flux of momentum, at any time, through the control surface that moves along with the ice block. Our data are therefore

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

$$\underline{\boldsymbol{P}}(t_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\,s}\,, \qquad \underline{\boldsymbol{P}}(t_1) = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s}\,,$$

$$\underline{\boldsymbol{F}}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\ (const.\ in\ time)}\,.$$

We have enough data to find the time-integrated supply of momentum generated within the moving control volume, using the balance of momentum:

$$\begin{aligned} \int_{t_0}^{t_1} \underline{\boldsymbol{G}}(t)\,\mathrm{d}t &= \underline{\boldsymbol{P}}(t_1) - \underline{\boldsymbol{P}}(t_0) - \int_{t_0}^{t_1} \underline{\boldsymbol{F}}(t)\,\mathrm{d}t \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\,s} - \int_{t_0}^{t_1} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N}\,\mathrm{d}t \\ &= \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s} \end{aligned}$$

As we remarked above, even if the physical event is exactly the same, the momentum flux and supply in the present analysis are different from the previous analysis, because *they refer to different imaginary control surfaces and volumes*. Compare the time-integrated influxes of the two analyses:

$$\int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t = \begin{bmatrix} 0 \\ 0 \\ -3.15 \end{bmatrix} \mathrm{N\,s} \quad \neq \quad \int_{t_0}^{t_1} \underline{\boldsymbol{F}}(t)\,\mathrm{d}t = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathrm{N\,s}\,.$$

and the time-integrated supplies:

$$\int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t = \begin{bmatrix} 0 \\ 0 \\ -1.57 \end{bmatrix} \mathrm{N\,s} \quad \neq \quad \int_{t_0}^{t_1} \underline{\boldsymbol{G}}(t)\,\mathrm{d}t = \begin{bmatrix} 0 \\ 0 \\ -4.72 \end{bmatrix} \mathrm{N\,s}$$

Fluxes and supplies depend on the control surface and volume

The time-integrated flux and time-integrated supply that appear in balance laws are strictly dependent on the closed control surface and volume that we choose.

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

Exercise 5.3

Consider once more the block of ice analysed in the previous two examples.
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 at time t_1. The initial and final momentum contents are zero.
Calculate the integrated supply for this new control surface between t_0 and t_1.
*(Hint: consider the results from the example of § 5.6, and use the **symmetry of flux** to find the flux through the bottom part of the control surface of the present exercise. Recall also that the fluxes through the side and top surfaces are zero.)*

5.7 Balance laws: differential expression

A balance law such as

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t$$

or a conservation law such as

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t)\,\mathrm{d}t$$

are said to be written in *integral form* because they contain time-integrated fluxes or supplies. These time integrations are necessary for calculating the net amount of quantity that enters through the control surface or is generated in the volume during the time lapse considered.

We can rewrite these laws in a different mathematical form that doesn't show time integrals, but shows time derivatives.

This rewriting is essentially an expression of the fundamental theorem of calculus[2]. Consider the case of the conservation law for simplicity. If N is a function of t, and we call J its derivative with respect to t, then we can also say that N is a primitive function[3] for J, and the definite integral of J is given by a difference of N at different arguments. In symbols:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = J(t) \qquad \Longleftrightarrow \qquad \int_{t_0}^{t_1} J(t)\,\mathrm{d}t = N(t_1) - N(t_0)\,.$$

But the equation on the right is just a conservation law written in a slightly different way! This means that it can equivalently be expressed by the equation on the left.

Let's interpret this relation from the point of view of the physical measurement of volume content and flux, and for full generality let's consider a balance law.

Change the symbol for the time t_0 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

$$\int_t^{t+\Delta t} \Phi(t)\,\mathrm{d}t \approx \Phi(t)\,\Delta t$$

that is, the flux at time t multiplied by the time lapse. The reason for this approximation is that the flux Φ is approximately constant during the short time lapse Δt, and equal to its initial value at time t. Similarly for the supply:

$$\int_t^{t+\Delta t} R(t)\,\mathrm{d}t \approx R(t)\,\Delta t\,.$$

The balance law above can then be approximated as follows:

$$E(t + \Delta t) \approx E(t) + [\Phi(t) + R(t)]\,\Delta t\,.$$

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

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

$$\frac{E(t+\Delta t) - E(t)}{\Delta t} \approx \Phi(t) + R(t) \,.$$

Finally, consider smaller and smaller Δt: in the limit the ratio $\frac{E(t+\Delta t)-E(t)}{\Delta t}$ becomes, by definition, a derivative; and the approximation of the integrals becomes an exact equality. So we find:

Balance and conservation laws in differential form

A quantity is said to satisfy a **balance law** if the following equality holds for any closed control surface and volume, and any coordinate time t:

$$\frac{\mathrm{d}\,\text{volume content}(t)}{\mathrm{d}t} = \text{influx}(t) + \text{supply}(t) \tag{5.5}$$

Analogously, a quantity is said to satisfy a **conservation law** if the following equality holds for any closed control surface and volume, and any coordinate time t:

$$\frac{\mathrm{d}\,\text{volume content}(t)}{\mathrm{d}t} = \text{influx}(t) \tag{5.6}$$

For instance, the balance and conservation laws written above can be re-expressed as

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t \qquad \text{or} \qquad \frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t)$$

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t)\,\mathrm{d}t \qquad \text{or} \qquad \frac{\mathrm{d}N(t)}{\mathrm{d}t} = J(t) \,.$$

This differential expression says that the *rate of change of the volume content* of a quantity must equal the sum of influx of that quantity through the control surface and the supply of that quantity in the control volume.

When to use the integral or the differential expression?

First of all, let's remark again that the integral expression and the differential expression of balance and conservation laws are mathematically equivalent (under some conditions that can become important in more advanced applications). So in choosing the one or the other we are not choosing between different physical laws.

But using the one or the other expression as our starting point can be more convenient in some situations and less convenient in others. Some obvious guidelines:

- If the data in a problem include time-integrated fluxes or supplies, then the integral expression is probably more convenient, because it directly uses these integrals.
- If a problem asks for some volume contents at a given time, and gives as data the volume contents at some other time, then the integral expression is probably more convenient.
- If a problem involves fluxes and supplies that are zero, then the differential expression may be more practical, because it says that the volume content is then constant in time: its time derivative is zero.
- If a problem asks for the time dependence of some quantity, then the differential expression is probably more convenient, because it involves a generic time instant.

But keep in mind that these are only guidelines. Some problems may be easily solved with either expression; some problems may require both expressions to be used: one expression for one quantity and the other for another quantity.

Example

As an example of use of the differential expression, let's consider again the problem with the tennis ball. The physical situation, choice of control surface and volume, and the known information and question are the same as before.

› § 5.4 page 115

We solved this problem using the integral expression of the balance law for momentum. But it can also be solved using its differential expression:

$$\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t} = \boldsymbol{F}(t) + \boldsymbol{G}(t)$$

or explicitly in components

$$\begin{cases} \dfrac{\mathrm{d}P_x(t)}{\mathrm{d}t} = F_x(t) + G_x(t) \\ \dfrac{\mathrm{d}P_y(t)}{\mathrm{d}t} = F_y(t) + G_y(t) \\ \dfrac{\mathrm{d}P_z(t)}{\mathrm{d}t} = F_z(t) + G_z(t) \end{cases}$$

The differential expression can be convenient because the value of the influx $\boldsymbol{F}$ and supply $\boldsymbol{G}$ are constant, and zero, at every time t. Therefore the rate of change of momentum at any time t between t_0 and t_1 is

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

We can find the momentum at any time t between and including t_0 and t_1 by an easy integration:

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

Exercise 5.4

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

- *What is the relevant time interval, if using the integral expression?*

- *How do you define the closed control surface and and its movement, as well as any subdivisions of the surface?*
- *What are the values of volume content and of time-integrated flux known to you? Which ones do you want to find?*

For the present exercise we assume that argon[4] atoms satisfy a conservation law.
A distillation column[5] in a chemical plant has an **in**flux of 2 mol/s of argon atoms constant in time at one inlet, and an **ef**flux of argon atoms, constant in time but 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 efflux at the outlet, at any time?

5.8 Seven universal balance 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* and *electric charge* obey conservation laws, eq. (5.1); 'magnetic flux' obeys an analogous conservation law but in one less dimension, as we'll discuss later. *Energy*, *momentum*, *angular momentum*, *entropy* instead satisfy balance laws, eq. (5.2); only in special circumstances do these four quantities have zero supply.

› § 3.1 page 52

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 appear implicitly or explicitly in *all* our main physical theories, approximate or not: from Newtonian mechanics to special relativity, from general relativity to quantum theory.
- Each of these balances can be expressed by the *same* mathematical equation in all of these theories.

In other words these seven balances are, as far as we know, *universal*.

› § 5.2 page 109

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 valid in any system of coordinates, for any kind of control volumes and surfaces, at any time. They are valid in Newtonian thermo-mechanics, electromagnetism, General Relativity, and even quantum theory if the symbols are interpreted as so-called statistical operators, which encode probabilistic properties.

The mathematical expressions of these balances are reported in table 5.1 on page 134.

These balances are truly the *Seven Wonders of the World*, even more long-lasting than the traditional "seven wonders"[6] or the "new seven wonders"[7].

In some approximate 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, only some of these laws are taken as fundamental, while the rest are derived from other fundamental laws; but nevertheless all these balances are still universally 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, which are taken as more fundamental.

› § 5.1 page 108

The *Einstein equations*, which can deceivingly simply be written

$$\boldsymbol{G} = \tfrac{8\pi G}{c^4}\boldsymbol{T}$$

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

It is therefore important and 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 will apply these balances to any kind of physical phenomenon or engineering or physics field you'll work with: construction of bridges, control of chemical reactions, operation of GPS navigation and satellites, monitoring of nuclear power plants, sending robots to Mars, design of fuel cells, collisions of subatomic particles,

cosmology, or who knows what else. Every physical phenomenon involves at least one of these seven balances in its physical description.

> **Concise mathematical form of the universal balances**
>
> The seven balances can be expressed very concisely if we use the language of *differential forms*[8]. These are geometric objects that associate a number to any curve, surface, or volume of our choice. The balances for matter, momentum, energy, angular momentum, electric charge, magnetic flux, entropy then take on these very concise expressions:
>
> $$\mathrm{d}N = \mathcal{A} \quad \mathrm{d}\boldsymbol{Q} = \mathcal{I} \quad \mathrm{d}\mathcal{B} = -\mathcal{E} \quad \mathrm{d}E = R \quad \mathrm{d}\boldsymbol{P} = \boldsymbol{G} \quad \mathrm{d}\boldsymbol{L} = \boldsymbol{M} \quad \mathrm{d}S \geq 0$$
>
> where the symbols on the left of the equations are taken in a four-dimensional sense.
>
> If you want to learn more about differential forms, take a look at the books by Burke 1987; 1995 and Bossavit 1991.

The seven balances for 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 some other physical phenomena, some of the seven balances do not appear *explicitly* in our calculations. But 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 (*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 silently been taken care of.

The description of how a common lighter works, thanks to piezoelectricity[9], requires more or less all seven universal laws to be explicitly accounted for.

	integral form	differential form
matter	$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t)\,\mathrm{d}t + \int_{t_0}^{t_1} A(t)\,\mathrm{d}t$	$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = J(t) + A(t)$
electric charge	$Q(t_1) = Q(t_0) + \int_{t_0}^{t_1} \mathcal{I}(t)\,\mathrm{d}t$	$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \mathcal{I}(t)$
magnetic flux	$\mathcal{B}(t_1) = \mathcal{B}(t_0) - \int_{t_0}^{t_1} \mathcal{E}(t)\,\mathrm{d}t$	$\frac{\mathrm{d}\mathcal{B}(t)}{\mathrm{d}t} = -\mathcal{E}(t)$
momentum	$\boldsymbol{P}(t_1) = \boldsymbol{P}(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t$	$\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t} = \boldsymbol{F}(t) + \boldsymbol{G}(t)$
energy	$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t$	$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t)$
angular momentum	$\boldsymbol{L}(t_1) = \boldsymbol{L}(t_0) + \int_{t_0}^{t_1} \boldsymbol{\tau}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{M}(t)\,\mathrm{d}t$	$\frac{\mathrm{d}\boldsymbol{L}(t)}{\mathrm{d}t} = \boldsymbol{\tau}(t) + \boldsymbol{M}(t)$
entropy	$S(t_1) \geq S(t_0) + \int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t$	$\frac{\mathrm{d}S(t)}{\mathrm{d}t} \geq \Pi(t)$

Table 5.1 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'.

- 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-tyre problem, where we only used the conservation of matter but we weren't interested in what happened to other quantities like energy or momentum, or in how the tyre was moving.

› § 5.4 page 115

Some physical phenomena may be equally predicted using either one particular subset of the seven balances, or 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 later applications.

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 relations*, which we discuss in the next section.

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

5.9 Constitutive relations

In the previous sections we studied the mathematical form of balance and conservation laws, and found out that seven balances are of special importance.

From their formulation and from the examples, you noticed that each balance law connects the volume content, flux, supply of one extensive quantity at different times. For instance, the balance law for energy

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t$$

connects the volume contents E *of energy*, the flux Φ *of energy*, the supply R *of energy*. It does not involve, say, the volume content N of matter, or the flux $\boldsymbol{F}$ of momentum.

But we know that there must also exist connections among the amounts of *different* quantities. This fact is implicit in many expressions we use

everyday, like "the pressure of air" (pressure is *momentum* flux, air is *matter*), or "the *energy* of the battery" (the battery is made of *matter*, and involves *electric charge*). In previous sections we often made statements such as "the momentum is zero, because there's no matter".

Physical laws that connect different kinds of quantity are called *constitutive relations*; we briefly discussed them before. As was mentioned in that discussion, the term 'constitutive' actually refers not to the characteristic of connecting different quantities, but to the fact that each of these laws typically applies only to specific situations:

› § 5.2 page 109

Constitutive relation

A **constitutive relation**, also called *constitutive equation*, or *closure equation*, or *constitutive property*, is a physical relationship or that is true only under specific conditions; for example: only for specific physical phenomena, or only on specific scales of space and time, or only for control volumes or surfaces of particular sizes or shapes, or only for specific ranges of measurement precision. Therefore they are often used in specific physical theories.

Constitutive relations express the amazing diversity of physical phenomena that we observe around and within us. For example the fact that a body of water can easily change shape, as opposed to a block of concrete; or that we can store electric energy in a batter but not in a piece of wood. The differences between states of matter[11] – solid, liquid, gas, plasma, and there are others – arise from different constitutive relations.

Four states of matter, arising from different constitutive relations (image: Spirit469[10])

Constitutive relations also express approximations that we use in describing physical phenomena in particular conditions, and therefore mark the difference between specialized or approximate physical theories; for example between Newtonian mechanics, which applies only for low speeds and low energy-mass concentrations (hence weak gravitational fields and small spacetime curvature), and General Relativity, which applies on most if not all known scales, including cosmological ones.

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,

or flux, or supply of one quantity, and the volume content, or flux, or supply of another. Some constitutive relations involve derivatives; some involve integrals.

Many constitutive relations connect volume contents, fluxes, supplies of different primitive quantities, not directly, but through the intermediary of derived or auxiliary quantities such as areas and volumes, temperature, velocity, pressure, polarization, magnetization. We shall see some examples below.

› § 3.11 page 71

A very powerful characteristic of many constitutive relations is that they connect the volume content of a primitive quantity at a given time with the flux or supply of another primitive quantity *at the same time*. For instance, a constitutive law may allow us to find the flux of matter across some control surface at time t: $J(t)$, from the content of momentum in a neighbouring control volume at the same time t: $\boldsymbol{P}(t)$; this example is further discussed below. This characteristic greatly extends the predictive power of balance laws when used together with constitutive relations, as we shall see in Chapter 6.

Examples

Let us briefly reveal some constitutive equations that have tacitly been used in examples of the previous sections. This is only an overview; we shall study these constitutive relations at length in the next chapters. All these constitutive relations are valid only in "Newtonian approximation", that is, when speeds are much lower than the speed of light; gravitation is weak, as it is on Earth; and any present electromagnetic fields are enough weak. Note how all relations are only valid for particular kinds of control volumes or surfaces.

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 *rest 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. One of the most used constitutive relations connects the amount of matter in a small control

volume, or the flux of matter through its surface, with the amount of momentum in the volume. It can be stated in several ways. Here are two alternatives: the first valid for a *moving* control volume, the second for a *static* one.

First formulation. Take a small control volume moving with velocity $\boldsymbol{v}$, with speed much smaller than light's. If this control volume contains an amount of matter N and there's *no flux* of matter through its surface, then it also contains, to a very good approximation, an amount of momentum

$$\boldsymbol{P} = m\boldsymbol{v} \ .$$

Old textbooks take this formula as the *definition* of momentum.

Second formulation. Take a small *static* cuboid control volume of cubical shape and sides parallel to the coordinate axes. The volume has size V and each of its sides area A. If the fluxes of matter through the sides orthogonal to the three coordinates, in a positive direction, are $[J_x, J_y, J_z]$, then the control volume contains momentum

$$\boldsymbol{P} = \rho\,[J_x, J_y, J_z]\,\frac{V}{A}$$

again only if weak electromagnetic fields are present. This relation is connected to the one between velocity and the flux of matter.

› § 4.15 page 105

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

$$\boldsymbol{G} = m\boldsymbol{g}$$

This constitutive relation expresses the *gravitational volume force*. The vector $\boldsymbol{g}$ is called the **acceleration of free fall**. It expresses the gravitational field, that is, spacetime curvature. It generally depends on time, position, and the motion of the matter or electromagnetic field having mass-energy m, and the coordinate system.

For physical phenomena close to Earth's surface, in a coordinate system (t, x, y, z) fixed the ground and with z pointing upwards, the vector $\boldsymbol{g}$ can be taken to be approximately constant; it points towards the ground:

$$\boldsymbol{g} = -g \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \text{with} \quad g \approx 9.8\,\mathrm{N/kg} \equiv 9.8\,\mathrm{m/s^2} \ .$$

In more precise applications in geodesy[12] this vector may depend on latitude, longitude, and even time, owing to Earth's internal motion. Further away from Earth's, the expression of $\boldsymbol{g}$ becomes more complex.

$$\begin{aligned}
&\sum_{j\neq i} \frac{\mu_j(\mathbf{r}_j-\mathbf{r}_i)}{r_{ij}^{\,3}}\left\{1-\frac{2(\beta+\gamma)}{c^2}\sum_{l\neq i}\frac{\mu_l}{r_{il}}-\frac{2\beta-1}{c^2}\sum_{k\neq j}\frac{\mu_k}{r_{jk}}\right.\\
&+\gamma\left(\frac{\dot{s}_i}{c}\right)^2+(1+\gamma)\left(\frac{\dot{s}_j}{c}\right)^2-\frac{2(1+\gamma)}{c^2}\dot{\mathbf{r}}_i\cdot\dot{\mathbf{r}}_j\\
&\left.-\frac{3}{2c^2}\left[\frac{(\mathbf{r}_i-\mathbf{r}_j)\cdot\dot{\mathbf{r}}_j}{r_{ij}}\right]^2+\frac{1}{2c^2}(\mathbf{r}_j-\mathbf{r}_i)\cdot\ddot{\mathbf{r}}_j\right\}\\
&+\frac{1}{c^2}\sum_{j\neq i}\frac{\mu_j}{r_{ij}^{\,3}}\left\{(\mathbf{r}_i-\mathbf{r}_j)\cdot\left[(2+2\gamma)\dot{\mathbf{r}}_i-(1+2\gamma)\dot{\mathbf{r}}_j\right]\right\}(\dot{\mathbf{r}}_i-\dot{\mathbf{r}}_j)\\
&+\frac{3+4\gamma}{2c^2}\sum_{j\neq i}\frac{\mu_j\ddot{\mathbf{r}}_j}{r_{ij}}
\end{aligned}$$

The expression for $\boldsymbol{g}$ used at NASA for satellite and spacecraft motion (Moyer 2000).

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 J across its surface. In a coordinate system (t, x, y, z) where z points upward, this control volume also contains an amount of energy-mass

$$E = mc^2 + U + \frac{1}{2}mv^2 + mgz \ .$$

The first term in this sum is called *rest energy-mass* (a huge amount) of energy in ordinary situations), the second term U is called *internal energy*, the third is called *kinetic energy*, the fourth *gravitational potential energy*. The rest mass-energy m and the internal energy U depend on the amount of matter N.

› § 3.6 page 59

5.10 Summary of differences between the seven balance laws and constitutive relations

Let's summarize the most important typical differences between the seven universal balance laws for our seven primitive quantities on one side, and constitutive relations on the other side:

Universal balance law	Constitutive relation
Seven	A virtually infinite number
Valid for every phenomenon	Valid for a specific phenomenon
Valid for any control volume and surface	Often valid only for control volumes or surfaces of specific size, shape, location
Connects contents, fluxes, supplies of the same quantity	Often connects contents, fluxes, supplies of different quantities
Connects contents, fluxes, supplies at different times	Often connects contents, fluxes, supplies at the same times
Used in all modern theories	Used in specific approximate theories

5.11 Newton's laws

Many books present, mainly out of tradition, a view of physical laws based on Newton's three axioms, or laws of motion, stated in his *Philosophiæ*

naturalis principia mathematica. It is therefore convenient to get acquainted with their statements. In the present section we examine Newton's laws and discuss their limitations as a foundations of today's physical theories.

The statements below are from the translation by Cajori (Newton 1974); the original Latin, from the third edition of Newton's *Principia* (Newton 1726b), is reported on the side. We must keep in mind that these statements are inextricably connected with the rest of definitions and reasoning given in the *Principia*, and with the knowledge of the time in which they were written. This makes it difficult, or maybe even meaningless, to fully interpret them. See for instance the analysis by Smith 2024.

First law

LAW I. Every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed upon it.

LEX I. Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus illud a viribus impressis cogitur statum suum mutare.

This is usually called the 'law of inertia', and it was accepted already sometime before Newton. Today this law is often viewed as a consequence of the second law below, together with Newton's definition of momentum: if there are no forces, then there is no change in 'motion', and if 'motion' is mass times velocity, then the velocity is constant, in magnitude and direction. But there are also differing views, which see it as an independent statement about the properties of space or of absolute motion, or about the differences between "real" and "fictitious" forces.

From a 21st-century point of view, centred on the relativity of motion to coordinate systems and frames, these debates about the first law lose some of their physical importance.

Second law

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.

LEX II. Mutationem motus proportionalem esse vi motrici impressæ, & fieri secundum lineam rectam qua vis illa imprimitur.

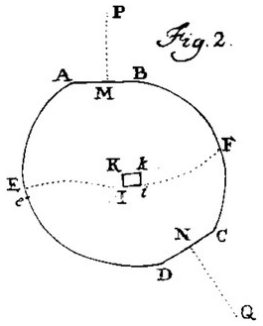
This law will be further discussed in Chapter 10. It is sometimes mathematically reported as '$\boldsymbol{F} = m\boldsymbol{a}$', but Newton never wrote this formula, which moreover may not fully represent the second law.

With some liberty we may interpret this law as the balance of momentum, but with two important warnings.

First, as far as I know, Newton did not make any distinction between surface forces and volume forces. The absence of this important distinction caused some difficulties in applying Newton's laws to bodies capable of

deformation, like fluids. These difficulties were solved when the Bernoullis and especially Euler in the 1700s introduced more clearly the concept of control surface and surface force, eventually developed in full generality by Cauchy in the 1800s. Modern mechanical engineering and fluid mechanics would have been impossible without this conceptual distinction.

Second, in the *Principia* Newton defines 'quantity of motion' as "arising from the velocity and quantity of matter conjointly", that is, as '$m\boldsymbol{v}$', so in this context the second law is specific to matter. But we know today that the exact expression for matter is different, and that electromagnetic fields have momentum as well.



Imaginary surface *EIiF* in a fluid, conceived by Euler to describe pressure. From Euler 1761.

Third law

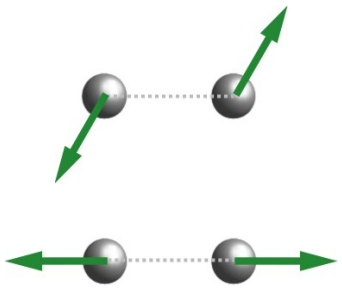
LAW III. To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

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

This is usually called the law of *action and reaction*. This law is only valid for *surface* forces, that is, for momentum flux; but not for *volume* forces, that is, momentum supply. For volume forces this law is generally not true, or only valid in special approximations and coordinate choices. We already mentioned that this law can be viewed as the principle of symmetry of flux for momentum. Therefore an analogous law is true for the fluxes of all other quantities, not only of momentum.

› § 4.11 page 98

In the 1900s literature there was a discussion on whether the third law could allow pairs of *volume* forces on two bodies as in the top side figure: equal and opposite but with a direction that doesn't necessarily align with a line connecting the two bodies. Or whether the third law would only allow for a direction as depicted in the bottom side figure.



Today we know that this discussion is somewhat pointless: as already said, the third law is in generally not true for supplies of momentum. Moreover in a curved spacetime it becomes unclear what a 'line' connecting two bodies is. For *surface* forces – momentum flux – this situation doesn't arise, because the flux concerns the surface as a whole, so there is no "application point".

Further limitations

Besides the shortcomings mentioned in the comments above, there are further limitations that make Newton's laws insufficient today as a general

basis to understand physical phenomena.

One basic limitation is that they do not cover all balances necessary to describe most mechanical phenomena. The balances of angular momentum and of energy are separate laws, *independent* of Newton's laws. The need for a separate balance of angular momentum was definitively recognized again by the Bernoullis and Euler in the 1600–1700s. The recognition of the balance of energy as a separate law led to modern thermomechanics. General relativity later showed that a further set of three balance equations must be taken into account in considering relativistic mechanical phenomena. So Newton's laws account for three out of ten universal equations necessary in mechanics.

The other obvious limitation is that Newton's laws are about 'bodies', that is, *matter*. The idea of an electromagnetic field was unknown to Newton and until the 1800s. The combination of mechanical and electromagnetic phenomena leads to conceptual difficulties with Newton's laws. For instance, if a charged body experiences an electromagnetic force or 'action', then *what* is experiencing the 'reaction'?

Another unfortunate limitation, not of Newton's laws per se but of the way they are often presented, is their mathematical vagueness. You can find Newton's second law variously written as

$$\boldsymbol{F} = m\,\boldsymbol{a}\,, \quad \frac{\mathrm{d}m\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{F}\,, \quad \frac{m\,\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{F}\,, \quad \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = \boldsymbol{F}\,, \quad \frac{\partial \rho \boldsymbol{v}}{\partial t} = -\nabla\cdot\boldsymbol{T} + \rho\boldsymbol{g}\,,$$

and other variations, some of which are more general than others. The reason of this variety is that equations like the ones above come from the combinations of one law – the balance of momentum – with different constitutive relations.

If you simply translate '*force exerted on...*' and similar expressions to '*influx into...*' or '*supply into...*', paying attention on whether they're surface or volume forces, then you'll be able to frame any physical problems in terms of control volumes, control surfaces, and balances.

URLs for chapter 5

1. https://pglpm.github.io/7wonders/media/volume_moving.gif
2. https://encyclopediaofmath.org/wiki/Newton-Leibniz_formula
3. https://encyclopediaofmath.org/wiki/Primitive_function
4. https://pubchem.ncbi.nlm.nih.gov/element/Argon
5. https://encyclopedia.che.engin.umich.edu/distillation-columns/
6. https://education.nationalgeographic.org/resource/seven-wonders-ancient-world/
7. https://www.britannica.com/list/new-seven-wonders-of-the-world
8. http://encyclopediaofmath.org/index.php?title=Differential_form
9. https://www.britannica.com/science/piezoelectricity
10. https://commons.wikimedia.org/wiki/images/File:Four_Fundamental_States_of_Matter.png
11. http://hyperphysics.phy-astr.gsu.edu/hbase/Chemical/chemeas.html#c4
12. https://oceanservice.noaa.gov/geodesy

Inference, prediction, simulation 6

> [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 Numerical time integration and simulations

Prediction and forecast

We have mentioned 'prediction' several times in the present notes: predicting the value of a quantity, predicting a physical behaviour, and so on. What do we mean by 'prediction', more exactly? This word is used in mainly two ways in physics; one more general and the other more specific.

In the general sense, 'predicting' something means managing to find out some piece of information, not by direct observation or measurement, but somehow arriving at it from other information available. For example we can predict that there's a person in a particular room because the light in that room is on, and is never on when it's empty. So we know there's a person there, not because we have taken a look inside and seen the person, but thanks to other information.

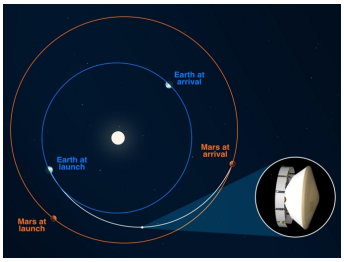
As a more physical example, we may be able to predict that the pressure inside a bike tyre has a particular value, just from knowing the volume of the tyre, the amount of air in it, and the air's temperature – but without directly measuring the pressure with a manometer. Another example is what you did in Exercise 2.1- **1.**: you predicted the time lapse of a GPS satellite's clock, by using the time lapse of your own clock and the laws of General Relativity.

In the two examples above, the predictions are about a piece of information – pressure, time lapse – that occurs at the same time as the information we have. (We are speaking about *coordinate* time; recall that it doesn't make sense to say that two events occur at the same *physical* time, unless they occur at the same place.)

› § 2.1 page 30

In a more specific (and technically more correct) sense, we use 'prediction' to mean that we find out some piece of information occurring at a *later* coordinate time, arriving at it from information occurring at *earlier* coordinate times. Often this means that it would even be impossible to observe that piece of information, because it hasn't happened yet! In this case we can also use the less pretentious term *forecast*.

'Prediction' in this sense, or forecast, is extremely important in our technologies and in human activity in general. We often wish to forecast tomorrow's weather. In designing a bridge, we want to forecast whether it will be able to sustain certain loads. In order to send a probe to Mars, we need to predict where Mars will be, and what trajectory our probe will follow.



Sending probes, such as the Perseverance rover[1], to other planets requires a careful prediction of the planet's and the probe's trajectories (image: NASA[2])

The special role of the universal balance laws, and finite-difference approximations

When we look at physics from the point of view of prediction, we realize the very special position and role of the universal balance laws. They state that a budget holds among several quantities, and these quantities *refer to different times*. With enough information about some quantities at some earlier times, they therefore determine the values of some quantities at later times.

We shall now discuss the basics of how the universal balances can be applied, together with constitutive relations, to make concrete numerical predictions in actual physical situations.

Recall again the integral and differential forms of a balance law. In the following discussion we shall use the symbols for energy, but the reasoning is valid for any other quantity:

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t)$$

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 about a short instant of time ahead. Let's see how this is done from the integral form.

Suppose that the lapse of time Δt 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 $R(t)$ don't change appreciably during this time lapse, and we can take them as approximately constant in time:

$$\Phi(t) \approx \Phi(t_0) \qquad R(t) \approx R(t_0) \qquad \text{for all } t \text{ between } t_0 \text{ and } t_1$$

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

$$\int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t \approx \int_{t_0}^{t_1} \Phi(t_0)\,\mathrm{d}t = \Phi(t_0)\,(t_1 - t_0) = \Phi(t_0)\,\Delta t$$
$$\int_{t_0}^{t_1} R(t)\,\mathrm{d}t \approx \int_{t_0}^{t_1} R(t_0)\,\mathrm{d}t = R(t_0)\,(t_1 - t_0) = R(t_0)\,\Delta t$$

Recall that

$$\int_a^b \text{const}\,\mathrm{d}x = \text{const} \cdot (b - a)$$

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

$$E(t_0 + \Delta t) \approx E(t_0) + \Phi(t_0)\,\Delta t + R(t_0)\,\Delta t$$
$$E(t_0 + \Delta t) = E(t_1) \qquad \Phi(t_0)\,\Delta t \approx \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t \qquad R(t_0)\,\Delta t \approx \int_{t_0}^{t_1} R(t)\,\mathrm{d}t$$

Factoring Δt we obtain:

Finite-difference approximation

$$E(t_0 + \Delta t) \approx E(t_0) + [\Phi(t_0) + R(t_0)]\,\Delta t \tag{6.1}$$

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 $t_0 = 20\,\mathrm{s}$ is $E(t_0) = 500\,\mathrm{J}$. At that same time there's a net influx of $\Phi(t_0) = -30\,\mathrm{J/s}$ into the control volume, and the supply is zero,

$R(t_0) = 0\,\mathrm{J/s}$. The amount of energy in the control volume 0.01 s later, that is, at time $t_0 + \Delta t = 20\,\mathrm{s} + 0.01\,\mathrm{s} = 20.01\,\mathrm{s}$, is then approximately

$$\begin{aligned} E(20.01\,\mathrm{s}) &\approx E(t_0) + [\Phi(t_0) + R(t_0)]\,\Delta t \\ &\approx E(20\,\mathrm{s}) + [\Phi(20\,\mathrm{s}) + R(20\,\mathrm{s})] \cdot 0.01\,\mathrm{s} \\ &\approx 500\,\mathrm{J} + [-30\,\mathrm{J/s} + 0\,\mathrm{J/s}] \cdot 0.01\,\mathrm{s} \\ &\approx 499.7\,\mathrm{J}\,. \end{aligned}$$

This basic idea is the same as the one behind Euler's method[3].

A control-surface sequence is understood

As you recall from the definition of a conservation or balance law, we must have chosen a *sequence of closed control surfaces*. How this sequence is chosen must either be clearly stated, or be 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 114

Keep in mind that the finite-difference approximation (6.1) that we derived is not the only possible one. They key step in our derivation was the approximation of the integrals appearing in the integral form of the balance law. Other approximation approaches are possible, and they lead to finite-difference approximations of slightly different forms. Entire books are devoted to these approximation schemes, and if you'll work in some physics or engineering fields you'll learn several of them in more detail. Some of these alternative forms lead much better approximations and numerical predictions; but they are more complicated and therefore we don't discuss them here.

Exercise 6.1

1. At time $t = 0\,\mathrm{s}$ 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 $t' = 0.01\,\mathrm{s}$.
2. Try to obtain the finite-difference approximation starting from the differential form of the balance law instead:

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t)$$

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

$$\frac{\mathrm{d}E(t_0)}{\mathrm{d}t} \approx \frac{E(t_0 + \Delta t) - E(t_0)}{\Delta t} .$$

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:

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

$$\text{or} \quad \begin{cases} P_x(t_0 + \Delta t) \approx P_x(t_0) + [F_x(t_0) + G_x(t_0)]\,\Delta t \\ P_y(t_0 + \Delta t) \approx P_y(t_0) + [F_y(t_0) + G_y(t_0)]\,\Delta t \\ P_z(t_0 + \Delta t) \approx P_z(t_0) + [F_z(t_0) + G_z(t_0)]\,\Delta t \end{cases} \tag{6.2}$$

Exercise 6.2

1. 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 $t_0 = 0\ \mathrm{s}$ the amount of momentum and the supply of momentum in the tennis ball are

   $$\boldsymbol{P}(t_0) = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \mathrm{N\,s} \qquad \boldsymbol{G}(t_0) = \begin{bmatrix} 0 \\ 0 \\ -0.579 \end{bmatrix} \mathrm{N}$$

   and the influx of momentum is zero. Calculate the momentum within the control volume $0.01\ \mathrm{s}$ later.

2. The tennis ball has a mass-energy $m = 0.059\ \mathrm{kg}$. Assume the constitutive relation for momentum:

   › § 5.9 page 137

   $$\boldsymbol{P} = m\,\boldsymbol{v}$$

   What was the velocity of the tennis ball at time t_0? How much is it $0.01\ \mathrm{s}$ later?

Iterating: numerical time integration and boundary conditions

The evolution equation (6.1) can 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:

$$\begin{aligned} E(t + \Delta t) &\approx E(t) + [\Phi(t) + R(t)]\,\Delta t \\ E(t + 2\Delta t) &\approx E(t + \Delta t) + [\Phi(t + \Delta t) + R(t + \Delta t)]\,\Delta t \\ E(t + 3\Delta t) &\approx E(t + 2\Delta t) + [\Phi(t + 2\Delta t) + R(t + 2\Delta t)]\,\Delta t \\ &\dots \end{aligned} \tag{6.3}$$

```
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 snippet[4] from the `Particle System` webpage[5] 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*.

Boundary conditions and constitutive equations

In order to perform the numerical time integration above, 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 (6.1).

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. 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 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 problem.

- They can be calculated, by means of *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 $\boldsymbol{P} = m\boldsymbol{v}$, briefly discussed previously. It relates the *flux* of matter (which determines the velocity $\boldsymbol{v}$) to the *volume content* of momentum. We shall discuss this relationship more clearly in § 7.2, and explore and use more examples of constitutive relations for the purpose of simulating physical systems.

› § 5.9 page 137

› § 4.15 page 105

Exercise 6.3

In Exercise 6.2 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 (6.3) and evolves all three components of momentum. Assume:
   - total duration of simulation is 2 s
   - timestep is $\Delta t = 0.01s$
   - momentum at initial time $t_0 = 0$ s is $\boldsymbol{P}(t_0) = [3, 0, 3]$ N s
   - momentum influx is zero at all times
   - momentum supply is $\boldsymbol{G} = [0, 0, -0.579]$ 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 $\boldsymbol{P} = m\boldsymbol{v}$ for the tennis ball, with $m = 0.059$ kg. Use the results of your script to plot the three components of the velocity $\boldsymbol{v}$ 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 $\boldsymbol{P}(t_0)$

- the mass m of the tennis ball (or other flying object)

and then use the constant momentum supply

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

Numerical time integration of position

In many applications we are interested in how the position $\boldsymbol{r}$ of a small object, or of a small part of an object, or of a small volume or surface, 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 $\boldsymbol{r}$ and its velocity.

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

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

Multiply by Δt and keep on the right side only the terms that refer to time t. We obtain

Time-stepped evolution equation for position

$$\boldsymbol{r}(t_0 + \Delta t) \approx \boldsymbol{r}(t_0) + \boldsymbol{v}(t_0)\,\Delta t \qquad \text{or} \qquad \begin{cases} x(t_0 + \Delta t) \approx x(t_0) + v_x(t_0)\,\Delta t \\ y(t_0 + \Delta t) \approx y(t_0) + v_y(t_0)\,\Delta t \\ z(t_0 + \Delta t) \approx z(t_0) + v_z(t_0)\,\Delta t \end{cases} \tag{6.4}$$

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 $\boldsymbol{r}(t)$ and can therefore keep track of and predict the motion of an object.

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

look like. Suppose that the initial position $\boldsymbol{r}$ and momentum $\boldsymbol{P}$ are given, and that the influx $\boldsymbol{F}$ and supply $\boldsymbol{G}$ are known at all times:

$$
\begin{aligned}
\boldsymbol{v}(t_0) &= \boldsymbol{P}(t_0)/m \\
\boldsymbol{r}(t_0 + \Delta t) &\approx \boldsymbol{r}(t_0) + \boldsymbol{v}(t_0)\,\Delta t \\
\boldsymbol{P}(t_0 + \Delta t) &\approx \boldsymbol{P}(t_0) + [\boldsymbol{F}(t_0) + \boldsymbol{G}(t_0)]\,\Delta t \\
\\
\boldsymbol{v}(t_0 + \Delta t) &= \boldsymbol{P}(t_0 + \Delta t)/m \\
\boldsymbol{r}(t_0 + 2\Delta t) &\approx \boldsymbol{r}(t_0 + \Delta t) + \boldsymbol{v}(t_0 + \Delta t)\,\Delta t \\
\boldsymbol{P}(t_0 + 2\Delta t) &\approx \boldsymbol{P}(t_0 + \Delta t) + [\boldsymbol{F}(t_0 + \Delta t) + \boldsymbol{G}(t_0 + \Delta t)]\,\Delta t \\
\\
\boldsymbol{v}(t_0 + 2\Delta t) &= \boldsymbol{P}(t_0 + 2\Delta t)/m \\
\boldsymbol{r}(t_0 + 3\Delta t) &\approx \boldsymbol{r}(t_0 + 2\Delta t) + \boldsymbol{v}(t_0 + 2\Delta t)\,\Delta t \\
\boldsymbol{P}(t_0 + 3\Delta t) &\approx \boldsymbol{P}(t_0 + 2\Delta t) + [\boldsymbol{F}(t_0 + 2\Delta t) + \boldsymbol{G}(t_0 + 2\Delta t)]\,\Delta t \\
\\
&\ldots
\end{aligned}
\tag{6.5}
$$

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 $\boldsymbol{F}$ or supply $\boldsymbol{G}$ may also depend on the position $\boldsymbol{r}$. The scheme above works also in these cases.

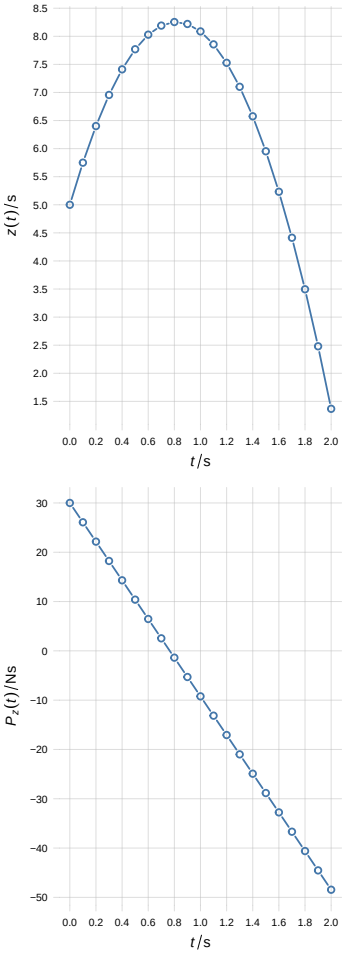
Numerical time integration of position: example

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

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

Then numerical integration up to $t = 0.2\,\mathrm{s}$ looks as follows:

$$\begin{aligned} v_z(0\,\mathrm{s}) &= P_z(0\,\mathrm{s})/m \\ &= 30\,\mathrm{N\,s}/4\,\mathrm{kg} = 7.50\,\mathrm{m/s} \\ z(0.1\,\mathrm{s}) &\approx z(0\,\mathrm{s}) + v_z(0\,\mathrm{s})\,\Delta t \\ &\approx 5\,\mathrm{m} + 7.50\,\mathrm{m/s}\cdot 0.1\,\mathrm{s} = 5.75\,\mathrm{m} \\ P_z(0.1\,\mathrm{s}) &\approx P_z(0\,\mathrm{s}) + [F_z(0\,\mathrm{s}) + G_z(0\,\mathrm{s})]\,\Delta t \\ &\approx 30\,\mathrm{N\,s} - 39.23\,\mathrm{N}\cdot 0.1\,\mathrm{s} = 26.08\,\mathrm{N\,s} \\ \\ v_z(0.1\,\mathrm{s}) &= P_z(0.1\,\mathrm{s})/m \\ &= 26.08\,\mathrm{N\,s}/4\,\mathrm{kg} = 6.52\,\mathrm{m/s} \\ z(0.2\,\mathrm{s}) &\approx z(0.1\,\mathrm{s}) + v_z(0.1\,\mathrm{s})\,\Delta t \\ &\approx 5.75\,\mathrm{m} + 6.52\,\mathrm{m/s}\cdot 0.1\,\mathrm{s} = 6.40\,\mathrm{m} \\ P_z(0.2\,\mathrm{s}) &\approx P_z(0.1\,\mathrm{s}) + [F_z(0.1\,\mathrm{s}) + G_z(0.1\,\mathrm{s})]\,\Delta t \\ &\approx 26.08\,\mathrm{N\,s} - 39.23\,\mathrm{N}\cdot 0.1\,\mathrm{s} = 22.15\,\mathrm{N\,s} \\ \\ &\dots \end{aligned}$$



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

Exercise 6.4

Modify the script you made for part **1.** of Exercise 6.3, so that it also evolves the position $\boldsymbol{r}(t)$ 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 $\boldsymbol{r}(t_0) = [0, 0, 2]\,\mathrm{m}$.
Plot $z(t)$ and $P_z(t)$ against the time t, similarly to the plots in the margin.

Applicability of numerical time integration

We have illustrated the numerical-evolution procedure, formulae (6.3) and (6.5), 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` webpage[6] of *Physics Simulation in Visual Computing*[7]. 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.

Example script for numerical time integration

We shall come back to numerical time integration in a later chapter, and will approach the problem of writing a simulation script in a more systematic way.

For the moment, here is an example script, written in Octave[8] (should also work in MATLAB), that is a solution for Exercise 6.4 p. 153. Blue lines are strictly related to numerical time integration; grey lines take care of saving and plotting the results. You can use this script as a starting point for exercises that require scripting in the next chapters.

Download `tennisball.m`[9]

```
%%% tennisball.m
%% Numerical simulation of object motion in 2D with gravity
%% (base SI units used throughout)
%% Coordinates (y,z)
%% Parameters
m = 0.059; % tennis ball's mass
%%
%% Initial values
t0 = 0; % initial time
y0 = 0; z0 = 2; % initial position
Py0 = 3; Pz0 = 0.75; % initial momentum
%%
t1 = t0 + 2; % final time
dt = 0.01; % time step
%%
%% Initialize values for loop
t = t0;
y = y0; z = z0;
Py = Py0; Pz = Pz0;
%%
Fy = 0; Fz = 0; % momentum influx (constant)
Gy = 0; Gz = -0.579; % momentum supply (constant)
%%
%% Plot & saving
%% adjust final time if not multiple of timestep
t1 = t1 + mod(t1-t0,dt);
%% Save values of all quantities at some steps during the simulation,
%% for subsequente analysis or plotting
%% (saving at all timesteps could be too costly)
Nsaves = 200; % number of timepoints to save during the simulation
```

```
%% Calculate time interval for saving
dsave = (t1-t0)/(Nsaves-1);
if abs(dsave) < abs(dt)
  error('time interval between saves is smaller than timestep')
end
%% Initialize vectors to contain saved values
tSave = nan(Nsaves,1);
ySave = nan(Nsaves,1); zSave = nan(Nsaves,1);
PySave = nan(Nsaves,1); PzSave = nan(Nsaves,1);
%% Save initial values
i = 1; % index that keeps count of savepoints
tSave(i) = t;
ySave(i) = y; zSave(i) = z;
PySave(i) = Py; PzSave(i) = Pz;
%% Initialize plot
cols = get(0, 'DefaultAxesColorOrder');
plot(ySave(1), zSave(1), '.', 'Color', cols(1,:)); axis('tight');
xlabel('y/m'); ylabel('z/m'); hold on;
%%
%% Numerical time integration
%% loop
while t < t1
  %% We need Py,Px,y,z,vy,vz
  %% we have y,z,Py,Pz
  %% find vy,vz using constitutive relations
  vy = Py/m; vz = Pz/m;
  %%
  %% Drive forward in time
  %% update momentum
  Py = Py + (Fy + Gy)*dt;
  Pz = Pz + (Fz + Gz)*dt;
  %% update position
  y = y + vy*dt;
  z = z + vz*dt;
  %% update time
  t = t + dt;
  %%
  %% Check whether to save & plot at this step
  if min(abs([0 dsave] - mod(t-t0, dsave))) <= abs(dt)/2
    i = i+1;
    tSave(i) = t;
    ySave(i) = y; zSave(i) = z;
    vySave(i) = vy; vzSave(i) = vz;
    PySave(i) = Py; PzSave(i) = Pz;
    plot(y, z, '.', 'Color', cols(1,:));
    pause(0.001);
  end
```

```
78 end
79 %% Plot full trajectory
80 plot(ySave, zSave, 'Color', cols(1,:)); axis('tight');
```

URLs for chapter 6

1. https://mars.nasa.gov/mars2020/
2. https://www.jpl.nasa.gov/news/nasas-perseverance-rover-is-midway-to-mars
3. https://encyclopediaofmath.org/wiki/Euler_method
4. https://github.com/InteractiveComputerGraphics/physics-simulation/blob/5bfb1022ac8b055bf52b197da22b0c2e4122d5f8/examples/particle_system.html#L181
5. https://interactivecomputergraphics.github.io/physics-simulation/examples/particle_system.html
6. https://interactivecomputergraphics.github.io/physics-simulation/examples/spring_plot.html
7. https://interactivecomputergraphics.github.io/physics-simulation/
8. https://octave.org/
9. https://pglpm.github.io/7wonders/code/tennisball.m

7 Conservation & balance of matter

> It is shown that the inertia of energy does not obviate the necessity for assuming the conservation of matter. *Matter* is to be interpreted as number of molecules, therefore, and not as inertia.
>
> C. Eckart 1940

7.1 Formulation and generalities

Balance and conservation of matter

Volume content: N Flux: J Supply: A

$$N(t_1) = N(t_0) + \underbrace{\int_{t_0}^{t_1} J(t)\,\mathrm{d}t + \int_{t_0}^{t_1} A(t)\,\mathrm{d}t}_{\text{integral form}} \qquad \underbrace{\frac{\mathrm{d}N(t)}{\mathrm{d}t} = J(t) + A(t)}_{\text{differential form}} \tag{7.1}$$

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.

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*. 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.

"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…"
Wilczek & Zee 1979

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.

› § 3.3 page 55

7.2 Examples of constitutive relations

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 $2\times10^{-5}\ \text{m}^3$ for liquid water) contains approximately $m = 1.62\times10^{15}\ \text{J}$ 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 –

› § 3.6 page 61

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

Mass-energy of matter

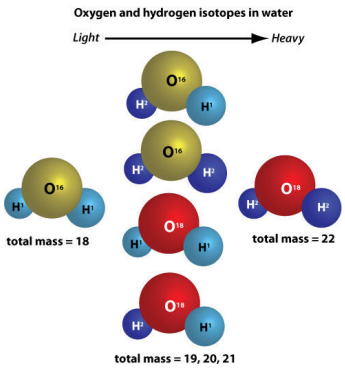
$$m = \rho N$$

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 constant[1] of approximately $\rho_{H_2O} = 0.0180\,\text{kg/mol}$. If a volume contains $N = 20\,\text{mol}$ of water, we usually attribute to it a mass-energy of

$$\begin{aligned} m &= \rho_{H_2O}\, N \\ &= 0.0180\,\text{kg/mol} \times 20\,\text{mol} \\ &= 0.36\,\text{kg}\,. \end{aligned}$$

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 isotopes[3] (molecules differing in the number of neutrons), and again the mixture proportions may vary with physical conditions.



There are several different isotopes of water, with different masses (image from U.S. Geological Survey[2])

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

$$\rho_{H_2O}\, J = 0.0180\,\text{kg/mol} \times 30\,\text{mol/s} = 0.54\,\text{kg/s}\,.$$

Conservation of mass: 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 Helium[4] $\rho_{\mathrm{He}} = 4 \times 10^{-3}$ 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?

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:

When we see this symbol we know that there's matter that's only balanced, not conserved – which involves danger

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[5]. If there is no influx of matter, $J(t) = 0$, the balance law for the substance then takes the form

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\lambda\, N(t)$$

called the *law of radioactive decay*[6].

7.3 Examples of applications

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.

› § 5.6 page 121

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 (6.5): 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.

› § 6.1 page 152

› § 7.2 page 160

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]

$$3\,P_4 + 10\,KClO_3 \rightarrow 3\,P_4O_{10} + 10\,KCl$$

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[9], the 10 moles of potassium (K) atoms[10], 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: $\mathrm{d}N(t)/\mathrm{d}t = J = 0$.

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.

Climate

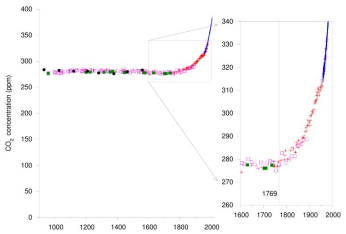
The laws of conservation and balances of matter turn out to be very useful also in problems related to climate.

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, like 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 like ice or wood.

› § 7.2 page 161

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].



Volume content of CO_2 in the atmosphere in the years 1000–2000. By the law of matter conservation, there must have been a net **in**flux of CO_2 into the atmosphere in these years. A vertical line is drawn at year 1769, when James Watt patented his steam engine (from MacKay 2008 p. 6)

Exercise 7.2

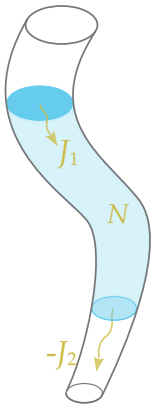
Measuring the relative amounts of carbon within an air pocket near the surface of a block of ice, you find that one mole of air contains 10^{-12} mol of the radioactive isotope ^{14}C. Making an analogous measurement for an air pocket deeper in the ice, you find instead a relative abundance of 2×10^{-13} 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

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\lambda\, N(t)$$

with $\lambda = 0.000\,122\ \mathrm{yr}^{-1}$, until you reach the amount $N(t_{\mathrm{ice}}) = 2 \times 10^{-13}$ mol. Use the procedure of eq. (6.1), starting from time $t_0 = 0$ yr and $N(t_0) = 10^{-12}$ mol, and assuming that the flow $J(t)$ of ^{14}C is zero. You can use a timestep $\Delta t = (1/365)$ yr.

› § 6.1 page 144

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.

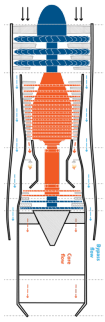
Consider a control volume, for instance the one indicated in light blue in the side picture. The amount of fluid N in this volume is constant in time: $\frac{\mathrm{d}N(t)}{\mathrm{d}t} = 0$. By the law of conservation of matter, the total influx must then be zero:

$$0 = \frac{\mathrm{d}N(t)}{\mathrm{d}t} = J(t)$$

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

$$0 = J(t) = J_1(t) + J_2(t) \qquad \Longrightarrow \qquad J_1(t) = -J_2(t)$$

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 at different sections of a jet engine, even if the flow of the fluid is turbulent, and in even more general situations.



(image from JetX[15])

If the surfaces are small enough, we can also use the connection between flux and velocity:

› § 4.15 page 105

$$v_1 = \frac{J_1/A_1}{N/V} \qquad v_2 = \frac{-J_2/A_2}{N/V}$$

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 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).

Sections on chemical reactions and stoichiometry to be added.

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/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://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

8 Conservation of electric charge

8.1 Formulation and generalities

Conservation of electric charge

Volume content: Q Flux: $\mathcal{I}$

$$Q(t_1) = Q(t_0) + \int_{t_0}^{t_1} \mathcal{I}(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \mathcal{I}(t) \tag{8.1}$$

integral form — differential form

To be written in a later version

9 Conservation of magnetic flux

9.1 Formulation and generalities

Conservation of magnetic flux

Flux: $\mathcal{B}$ Circuitation: $-\mathcal{E}$

$$\mathcal{B}(t_1) = \mathcal{B}(t_0) + \int_{t_0}^{t_1} \mathcal{E}(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}\mathcal{B}(t)}{\mathrm{d}t} = \mathcal{E}(t) \tag{9.1}$$

integral form — differential form

To be written in a later version

10 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."

10.1 Formulation and generalities

Balance of momentum

Volume content: $\boldsymbol{P}$ Flux: $\boldsymbol{F}$ Supply: $\boldsymbol{G}$

$$\boldsymbol{P}(t_1) = \boldsymbol{P}(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{G}(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t} = \boldsymbol{F}(t) + \boldsymbol{G}(t) \tag{10.1}$$

integral form — differential form

Among the seven universal balances, the balance of momentum is probably the 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.

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 $\boldsymbol{P}$ 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 $\boldsymbol{F}$ of momentum with many properties of matter, like 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.

› § 6.1 page 152

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 $\boldsymbol{F}$ and supply $\boldsymbol{G}$ are necessary to ensure that the volume content $\boldsymbol{P}$ of momentum is constantly zero.

Momentum balance keeps buildings from collapsing (photo: *Bryggen*, Bergen, from UNESCO World Heritage Centre[1])

Amounts of momentum depend on the coordinate system

Remember that the amount of momentum $\boldsymbol{P}$ in a volume, the flux $\boldsymbol{F}$ through a surface, and the supply $\boldsymbol{G}$ 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!

10.2 Examples of constitutive relations

Newtonian 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.

› § 6.2 page 148

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 $\boldsymbol{v}$, possibly zero, related to its flux. The **Newtonian formula for the momentum of matter** then says that this control volume also contains an amount of momentum $\boldsymbol{P}$ that is proportional to the mass-energy and the velocity:

› § 7.2 page 159

› § 4.15 page 105

Newtonian constitutive relation for momentum

If a control volume contains an amount of matter N having mass-energy m and velocity $\boldsymbol{v}$, then it also contains an amount of momentum

$$\boldsymbol{P} = m\boldsymbol{v} \qquad \text{or in components} \qquad \begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = m \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \tag{10.2}$$

This relation is only valid if the matter's velocity $\boldsymbol{v}$ is the same throughout the volume. This is usually the case if the volume is small enough.

You are probably familiar with the first equality above; old textbook present it as the *definition* of momentum. Today we know today that this equality only holds for matter but not, for instance, for the electromagnetic field. And even for matter it is only approximate: it's only valid when the velocity's magnitude v is small compared to the speed of light, only in weak gravitational fields, and only when changes in energy-mass are small compared to the total mass-energy.

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* $\boldsymbol{F}$.

› § 4.12 page 99

We can approximately treat these objects as control volumes for which the following constitutive properties apply:

Hookean spring

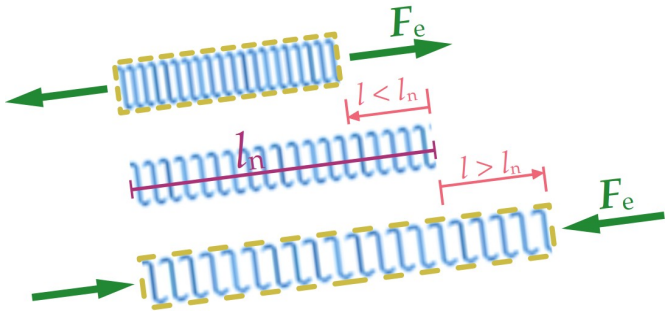
A **Hookean spring** is a control volume with the following properties:

- 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 $\boldsymbol{P}$ within the control volume are negligible.
- Across the (small) surfaces orthogonal to main dimension there is an **ef**flux of momentum equal to

$$\boldsymbol{F}_\mathrm{e} = -k\,\Delta\boldsymbol{l} \tag{10.3}$$

  this formula is called **Hooke's law**.

k is a constant, called the **elastic constant** of the spring; $\Delta\boldsymbol{l}$ is a vector along the main dimension, with magnitude $\Delta l = |l - l_\mathrm{n}|$ equal to difference between the present length l and the natural length l_n; it points *away* from the control volume if $l > l_\mathrm{n}$, and *toward* the control volume if $l < l_\mathrm{n}$.



Note the minus sign in Hooke's law. It means that if the present length is larger than the natural one, $l > l_\mathrm{n}$, then there flux of momentum getting out of the spring is oriented towards the spring; that is, the spring exerts a *tensile* force towards itself. And if the present length is smaller than the natural one, $l < l_\mathrm{n}$, then the flux of momentum getting out of the spring is oriented away from the spring; that is, the spring exerts a *compressive* force away from itself.

In modelling some physical problems we can put $l_\mathrm{n} = 0$ 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 10.1

What are the physical dimensions of the elastic constant k? in which units could we measure it?

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 like the following:

Constitutive equation for rubber bands and strings

Suppose that the two extremities of a rubber band have position vectors $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$, and that the natural length of the rubber band is l_n. Then the momentum **ef**fluxes at the extremities $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$ are approximately given by

$$\boldsymbol{F}_{as} = \begin{cases} 0\,, & \text{if } |\boldsymbol{r}_a - \boldsymbol{r}_b| \le l_\mathrm{n} \\ -k\left(|\boldsymbol{r}_a - \boldsymbol{r}_b| - l_\mathrm{n}\right) \frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|}\,, & \text{if } |\boldsymbol{r}_a - \boldsymbol{r}_b| > l_\mathrm{n} \end{cases} \tag{10.4}$$

Note that $\frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|}$ is just a *unit vector* directed from the position $\boldsymbol{r}_b$ to the position $\boldsymbol{r}_a$.

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

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 can be neglected (just like the mass and momentum of a real spring)

Under these conditions we have a constitutive relation of the general form

$$\boldsymbol{F}_{as} = f\left(|\boldsymbol{r}_a - \boldsymbol{r}_b|\right) \frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|}$$

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**:

$$\boldsymbol{F}_{as} = -G \frac{m_a\, m_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|^2} \frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|} \tag{10.5}$$

Another example is the force of the so-called *Lennard-Jones potential*:

$$\boldsymbol{F}_{as} = \frac{\epsilon}{|\boldsymbol{r}_a - \boldsymbol{r}_b|} \left[12 \left(\frac{\sigma}{|\boldsymbol{r}_a - \boldsymbol{r}_b|} \right)^{12} - 6 \left(\frac{\sigma}{|\boldsymbol{r}_a - \boldsymbol{r}_b|} \right)^{6} \right] \frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|} \tag{10.6}$$

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.

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 $\boldsymbol{G}$. 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 $\boldsymbol{G}$ 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.

The supply of momentum appearing in a slowed-down car can be deadly real

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:

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

$$\boldsymbol{G} = m\,\boldsymbol{g}\,, \qquad \text{with} \qquad \boldsymbol{g} = -g \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\,, \tag{10.7}$$

where g, the *gravitational acceleration*, depends on the planet. On Earth it is approximately

$$g \approx 9.8\,\mathrm{N/kg} \equiv 9.8\,\mathrm{m/s^2}\,. \tag{10.8}$$

This supply of momentum is approximately constant in time.

Contact forces

Many physical phenomena involve contact between two or more bulks of matter of different kinds or in different states of motion. Think of a book resting on a table, or a wooden crate pushed across the floor, or a layer of vegetable oil floating above water. The fluxes of momentum that occur through contact surfaces of this kind are called **contact forces** and have several peculiar features. In some cases they are mathematically quite difficult to describe.

We here discuss some particular constitutive relations for contact forces between rigid bodies, with the following simplifications:

- one of the bodies is at rest in the chosen coordinate system; typically this 'body' is the ground, floor, or the surface of a table;
- the contact surface is horizontal, orthogonal to the force of gravity.

A contact force in this kind of situations typically has both a component orthogonal to the surface (vertical component), and a component parallel to the surface (horizontal component), so its direction is oblique with respect to the surface. The surface-orthogonal component of a contact force is called **normal force**, and the surface-parallel component is called **friction**.

If we simplify the problem to two dimensions with coordinates (x, z), where x is horizontal and z is vertical, upward, then the z-component of the contact force is the normal force, and the x-component is the friction. Informally we can write

$$\textbf{contact force} = \begin{bmatrix} \text{contact force}_x \\ \text{contact force}_z \end{bmatrix} = \begin{bmatrix} \text{friction} \\ \text{normal force} \end{bmatrix}$$

In many cases, a contact force $\boldsymbol{F}_\text{c}$ has a constitutive equation of this type:

$$\boldsymbol{F}_\text{c} = \begin{bmatrix} \pm \mu \, F_\text{n} \\ F_\text{n} \end{bmatrix} \tag{10.9}$$

where F_n is the normal force and μ is approximately a constant, called **friction coefficient**. The normal force F_n has in turn a special expression. Let us examine it first.

Normal force The normal components of contact forces are peculiar: their mathematical expression can be said to be determined by the balance of momentum, rather than vice versa.

Consider an object such a book lying on a table, possibly pushed along the table. We *observe* that the vertical position of the book does not change: the book doesn't suddenly sink into the table, nor does it suddenly levitate upward. Its vertical velocity component is therefore always zero: $v_z = 0\,\mathrm{m/s}$. Using Newton's constitutive relation for momentum $\boldsymbol{P} = m\boldsymbol{v}$, where m is the mass of the book, we see that the vertical momentum component is also always zero: $P_z = mv_z = 0\,\mathrm{N\,s}$. Its time derivative is therefore also zero.

Now consider the sum of all momentum fluxes through the whole surface of the book *except the contact surface*: top surface, side surfaces; call this sum $\boldsymbol{F}_{\text{other}}$. Consider also the gravitational supply of momentum $\boldsymbol{G} = -m\,g\,[0,-1]$. Taking the vertical, z-component of the balance of momentum we have

$$0\,\mathrm{N} = \frac{\mathrm{d}P_z}{\mathrm{d}t} = F_{\mathrm{n}} + F_{\text{other},z} + G_z \qquad \Longrightarrow \qquad F_{\mathrm{n}} = -F_{\text{other},z} - G_z$$

That is, the normal force is equal to *minus* the sum of all z-components of the other momentum fluxes and of the momentum supply:

Normal force (horizontal case)

In many situations of horizontal contact between two solid bodies, the **normal force** F_{n} is given by

$$F_{\mathrm{n}} = -\begin{pmatrix}\text{sum of } z\text{-components of momentum fluxes through all other surfaces}\\ \text{and of momentum supply}\end{pmatrix}$$

The normal force is therefore never needed to predict the behaviour of the vertical momentum of a body. Rather, its mathematical expression comes from the fact that we already know that such momentum is zero and constant. But the value of this force is needed to determine the *horizontal component* of the contact force: the friction, as formula (10.9) shows.

Friction When we want to push or drag an object such as a table or large box across the floor, it's a common experience that exerting a small force (say, pushing with a finger) won't move the object. We need to exert a minimal amount of force to set it into motion; and usually this minimal force is the larger, the heavier is the object. It's also a common experience that once we manage to set the object into motion, the force we need to exert to keep it moving with a constant speed can be smaller than the force initially needed to set it into motion.

In both these experiences the total momentum content of the object is not changing (and it is zero in the first experience). In particular, the fact that the *horizontal* component of momentum is not changing, even if we are providing a horizontal momentum flux, means that there is a horizontal momentum flux coming from somewhere else. Obviously it is the horizontal component of the contact force exerted on the object by the floor: the *friction* between object and floor.

Our two experiences illustrate that this friction may be different depending on whether the object is at rest or in motion with respect to the floor. We call it **static friction** when the object's velocity is zero, and **kinetic** or **sliding friction** when the object's velocity is non-zero. Let's see their mathematical expressions.

Static friction In our first experience, with the object at rest on the floor, we noticed that the (zero) horizontal momentum does not change if we exert any amount of force smaller than a particular threshold. Reasoning as we did for the normal force, consider again the sum $\boldsymbol{F}_{\text{other}}$ of all momentum fluxes through the whole surface of the object except the contact surface. Even if not necessary in this specific case, consider also the gravitational supply of momentum $\boldsymbol{G} = -m\,g\,[0,-1]$. Finally, call F_{s} the static friction. Taking the horizontal, x-component of the balance of momentum we find

$$0\,\text{N} = \frac{\mathrm{d}P_x}{\mathrm{d}t} = F_{\text{s}} + F_{\text{other},x} + G_x \qquad \Longrightarrow \qquad F_{\text{s}} = -F_{\text{other},x} - G_x$$

Note how it is the balance of momentum that determines the amount of static friction, rather than vice versa, just like it happened for the normal force.

We also said that this friction occurs only as long as the total horizontal force is less that a particular threshold. This threshold turns out to be, approximately:

- independent of the area of the contact surface
- dependent on the nature of the two materials in contact
- proportional to the normal force exerted on the object

We can express this threshold with the equation

$$(F_{\text{s}})_{\text{threshold}} = \mu_{\text{s}}\,|F_{\text{n}}|\ .$$

Static friction (horizontal case)

In many situations of horizontal contact between two solid bodies, the **static friction** F_s is given by

$$F_\mathrm{s} = \begin{cases} -\boldsymbol{F}_\mathrm{other} & \text{if } |\boldsymbol{F}_\mathrm{other}| \le \mu_\mathrm{s} |F_\mathrm{n}| \\ -\boldsymbol{e}_{\boldsymbol{F}_\mathrm{other}} \mu_\mathrm{s} |F_\mathrm{n}| & \text{if } |\boldsymbol{F}_\mathrm{other}| \ge \mu_\mathrm{s} |F_\mathrm{n}| \end{cases} \tag{10.10}$$

where

$$\boldsymbol{F}_\mathrm{other} \coloneqq \begin{pmatrix} \text{sum of horizontal components of momentum fluxes} \\ \text{through all other surfaces} \end{pmatrix},$$

F_n is the normal force, μ_s is called the **coefficient of static friction**, and $\boldsymbol{e}_{\boldsymbol{F}_\mathrm{other}}$ is a *unit vector* having the same direction as $\boldsymbol{F}_\mathrm{other}$.

Kinetic friction In our second experience, with the object in motion, we noticed that the kinetic friction $\boldsymbol{F}_\mathrm{k}$ exerted by the floor on the object is, approximately:

- constant
- independent of the area of the contact surface
- dependent on the nature of the two materials in contact
- proportional to the normal force exerted on the object
- opposite to the velocity $\boldsymbol{v}$ of the object

Kinetic friction (horizontal case)

In many situations of horizontal contact between two solid bodies, the **kinetic friction** $\boldsymbol{F}_\mathrm{k}$ is given by

$$\boldsymbol{F}_\mathrm{k} = -\boldsymbol{e}_{\boldsymbol{v}} \mu_\mathrm{k} |F_\mathrm{n}| \tag{10.11}$$

where μ_k is the **coefficient of kinetic friction**, and $\boldsymbol{e}_{\boldsymbol{v}}$ is a *unit vector* having the same direction as the velocity $\boldsymbol{v}$; the minus sign indicates that the kinetic friction $\boldsymbol{F}_\mathrm{k}$ has opposite direction.

The formula above is valid only as long as the velocity of the object is not zero.

Note that the coefficients of static and kinetic friction, μ_s and μ_k, need not have the same value; in many cases the coefficient of kinetic friction is smaller than the other.

Exercise 10.2

1. The normal force has physical dimensions of force, SI units N. From formulae (10.10) and (10.11) find the physical dimensions and units of the coefficients of static and kinetic friction.
2. Two persons are pushing a parked car (side figure). The car rests on a thin layer of ice that formed on the tarmac. Each person is exerting a horizontal force of 600 N. The car has mass 1200 kg, and the acceleration of gravity is 9.8 N/kg. The coefficient of static friction between the wheel's rubber and ice is $\mu_s = 0.1$.

   Will the two persons manage to move the car?
3. Consider the following Octave/MATLAB function to calculate the friction in a numerical simulation in two dimensions with coordinates (x, z):

```
function F_fr = friction(Fother_x, Fother_z, v_x, mu_s, mu_k)
  if v_x == 0 % static friction
    threshold = mu_s * abs(Fother_z); % max magnitude
    if abs(Fother_x) <= threshold
      F_fr = -Fother_x;
    else
      F_fr = -sign(Fother_x) * threshold;
    end
  else % kinetic friction
    F_fr = -sign(v_x) * mu_k * abs(Fother_z);
  end
end
```

   Check that this function correctly covers and represents both formulae (10.10) and (10.11).

`sign()` is defined as

$$\texttt{sign}(x) \coloneqq \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

10.3 Examples of applications

Statics

As previously mentioned there are situations in which we must study or find which momentum fluxes $\boldsymbol{F}$ and supply $\boldsymbol{G}$ 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:

$$\boldsymbol{P}(t) = [0\,,0\,,0]\,\mathrm{N\,s} \;\text{ (constant)}.$$

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

$$\boldsymbol{G}(t) = [0\,,0\,,-2]\,\mathrm{N} \;\text{ (constant)}.$$

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* **in**flux of momentum must be

$$\boldsymbol{F}(t) = \underbrace{\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t}}_{\text{[0,0,0] because constant}} - \underbrace{\boldsymbol{G}(t)}_{[0,0,-2]\,\mathrm{N}}$$

$$= [0\,,0\,,2]\,\mathrm{N} \;\text{ (constant)}.$$

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.

› § 4.13 page 101

In the present case there's a constitutive relation about the force, through the *whole* surface, exerted by air on the book, which says that such total force is approximately zero:

$$\boldsymbol{F}_{\text{air-book}}(t) \approx [0\,,0\,,0]\,\mathrm{N}\,.$$

The only remaining force must be the one between the book and the table: a contact force. So the total influx $\boldsymbol{F}(t)$ that appears in the momentum balance must be the sum of these two:

› § 10.2 page 175

$$\boldsymbol{F}(t) = \boldsymbol{F}_{\text{air-book}}(t) + \boldsymbol{F}_{\text{table-book}}(t)\,.$$

From the last three equations we finally find

$$\boldsymbol{F}_{\text{table-book}}(t) \approx [0\,,0\,,2]\,\text{N} \quad \text{(constant)}.$$

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 is the **normal force** discussed in the previous section.

Exercise 10.3

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 $[0\,,0\,,-2]\,\text{N}$, 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.*

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 10^5 N (corresponding to a weight of about 10 000 kg!).
- 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.

› § 4.12 page 99

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 sometime 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 compensate 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 10.4

In this exercise we apply a reasoning very similar to that of Exercise 10.3. 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 $h\,A$

› § 10.3 page 179

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

$$G_z = -h \cdot A \cdot 11.8\,\mathrm{N/m^3}\,.$$

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 *divided by 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 $\boldsymbol{G}$ actually changes with altitude as well, and depends on other quantities such as temperature. But they do have the correct order of magnitude.

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 $\boldsymbol{G}$ 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 volume of the surrounding air (or atmosphere or other fluid), and the gravitational acceleration; this formula is called 'Archimedes's principle'[6]. 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.

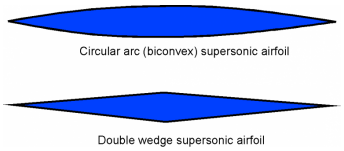
Airship *Airlander* from Hybrid Air Vehicles[5]

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[8] and is used by birds and gliders; it's also the mechanism that lifts a light object like a piece of paper or a feather.

A hang-glider (image from Jæren Luftsportsklubb[7])

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 flux 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.

› § 4.6 page 87



Sections of two kinds of

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 Center[9]

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.

› § 5.6 page 123

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.

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

Scenic Skyway[10] cable car, Australia

- There's a constant momentum supply because of gravity; let's say that in this case it's

$$\boldsymbol{G}(t) = [0\,,0\,,-8\times10^4]\,\mathrm{N}\ \text{(constant)}$$

- By the balance of momentum, there must then be an influx of upward-momentum of the same magnitude as the supply:

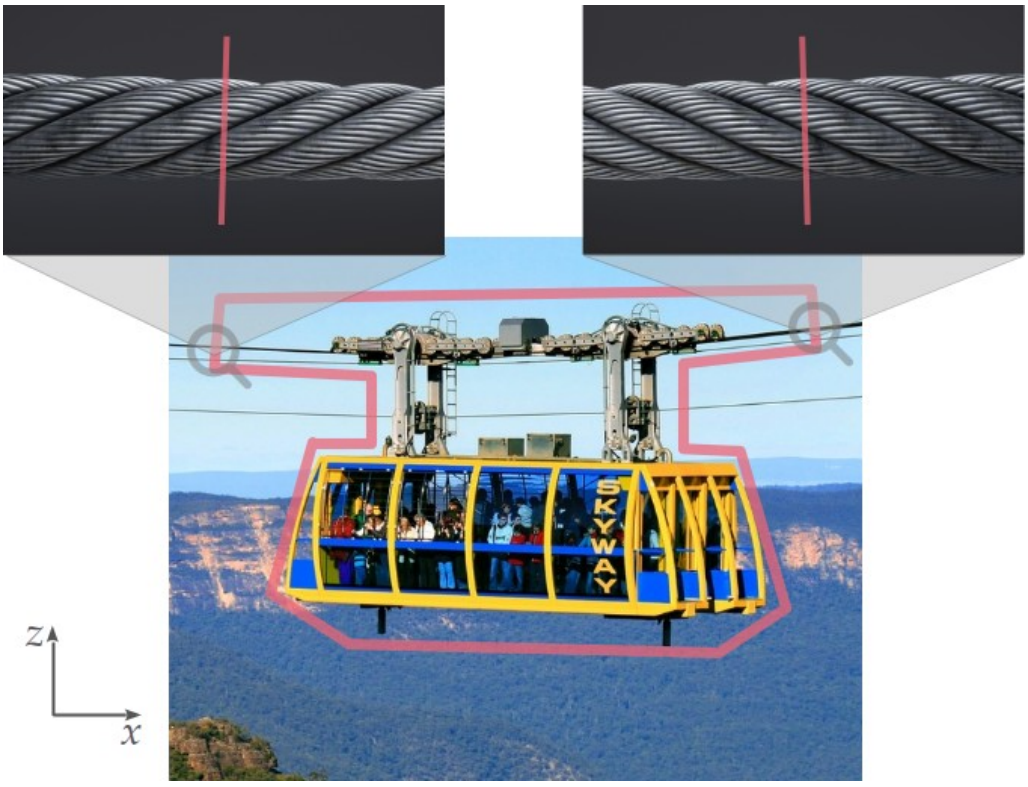
$$\boldsymbol{F}(t) = \underbrace{\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t}}_{[0,0,0]\text{ because constant}} - \underbrace{\boldsymbol{G}(t)}_{[0,0,-8\times10^4]\,\mathrm{N}}$$

$$= [0\,,0\,,8\times10^4]\,\mathrm{N}\ \text{(constant)}.$$

- Also in this case the influx through the parts of the control surface in contact with air is practically zero:

$$\boldsymbol{F}_{\text{air-car}}(t) \approx [0\,,0\,,0]\,\mathrm{N}\,.$$

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 depict, 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 $\boldsymbol{F}$ 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 $\boldsymbol{F}_{\text{left}}$ through surface cutting the cable on the left; and $\boldsymbol{F}_{\text{right}}$ 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

$$\boldsymbol{F}_{\text{left}} + \boldsymbol{F}_{\text{right}} = \boldsymbol{F} = \begin{bmatrix} 0 \\ 8\times10^4 \end{bmatrix} \text{N}$$

from this equation we find

$$(\boldsymbol{F}_{\text{left}})_x = -(\boldsymbol{F}_{\text{right}})_x \qquad (\boldsymbol{F}_{\text{left}})_z = 8\times10^4\,\text{N} - (\boldsymbol{F}_{\text{right}})_z$$

where '$(\dots)_x$' denotes the x-component, and similarly for the z-component. But this doesn't tell us how much $\boldsymbol{F}_{\text{left}}$ and $\boldsymbol{F}_{\text{right}}$ are individually. For example we could have

$$\boldsymbol{F}_{\text{left}} \stackrel{?}{=} \begin{bmatrix} 5 \\ -2\times10^4 \end{bmatrix} \text{N} \qquad \boldsymbol{F}_{\text{right}} \stackrel{?}{=} \begin{bmatrix} -5 \\ 20\times10^4 \end{bmatrix} \text{N}$$

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: $(\boldsymbol{F}_{\text{left}})_z = (\boldsymbol{F}_{\text{right}})_z$. So we can conclude that

$$(\boldsymbol{F}_{\text{left}})_z = (\boldsymbol{F}_{\text{right}})_z = 4\times10^4\,\text{N}$$

that is, each half of the cable is "taking half of the weight", as intuitively expected.

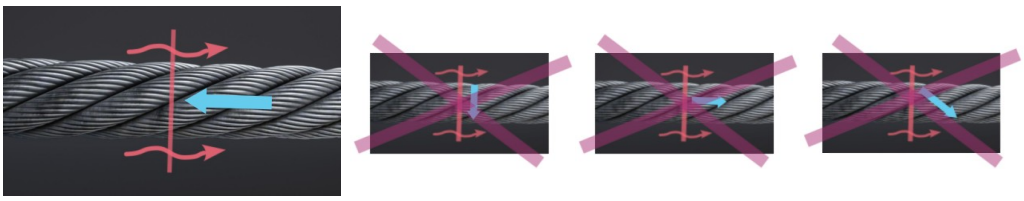
The last part of our mystery, about the x-components, can be solved thanks to an additional constitutive property:

Momentum flux allowed in cables and ropes

A cable, rope, or similar object can (approximately) only transmit *tensile momentum flux*, aligned along its axis.

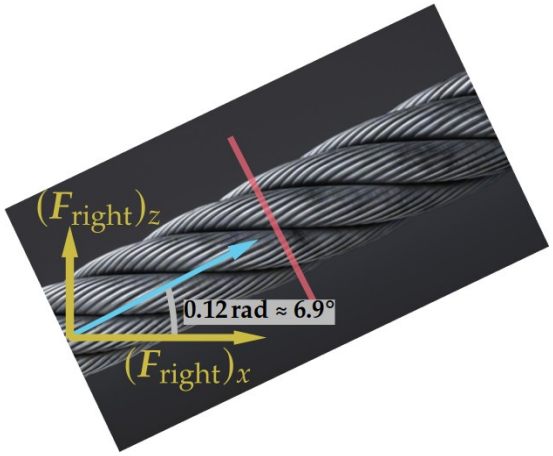
› § 4.12 page 100

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 rad $\approx 6.9°$, but in opposite directions. This must then also be the angle between the influx $\boldsymbol{F}_{\text{right}}$ and its horizontal component $(\boldsymbol{F}_{\text{right}})_x$, as in this picture (the angle has been exaggerated for clarity):



From trigonometry we find

$$(\boldsymbol{F}_{\text{right}})_x = (\boldsymbol{F}_{\text{right}})_z / \tan(0.12\ \text{rad}) = 4 \times 10^4\ \text{N}/0.12 \approx 3.3 \times 10^5\ \text{N}$$

and therefore $(\boldsymbol{F}_{\text{left}})_x \approx -3.3 \times 10^5$ N:

$$\boldsymbol{F}_{\text{left}} \approx \begin{bmatrix} -3.3{\times}10^5 \\ 4{\times}10^4 \end{bmatrix} \text{N} \qquad \boldsymbol{F}_{\text{right}} \approx \begin{bmatrix} 3.3{\times}10^5 \\ 4{\times}10^4 \end{bmatrix} \text{N}$$

The magnitude of these tensions is also approximately $|\boldsymbol{F}_{\text{left}}| = |\boldsymbol{F}_{\text{right}}| \approx 3 \times 10^5$ N or an equivalent weight of 35 tonnes (this result seems in the correct order of magnitude, comparing with the data in Brownjohn 1998).

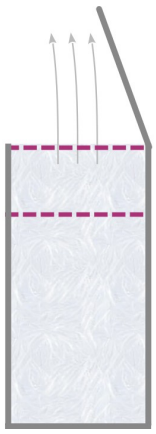
In solving the cable-car problem above we used the balance of momentum, but note and keep in mind how that balance alone wasn't enough. We also had to use:

- 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

Momentum fluxes in a gas

Imagine a rigid box at rest containing an amount of gas, and vacuum outside the box. Let gravity be negligible. It is common knowledge that if we open one side of the box, the gas within gets out. The gas's behaviour is to be contrasted with that of a solid: if the box contained, say, a brick, then the brick would simply stay where it is upon opening one side of the box.

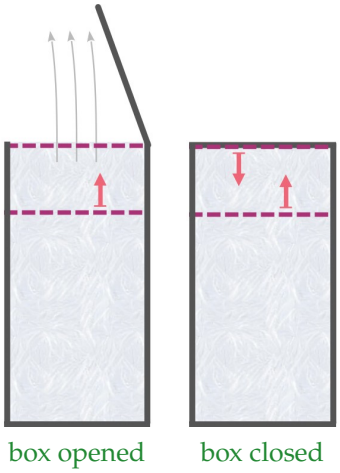
This peculiar behaviour of gases gives us an idea of what kind of momentum fluxes must occur at the boundaries of a control volume containing some gas.



Exercise 10.5

Before continuing to read, try to figure out by yourself how the momentum fluxes in a gas should be; formulate some hypotheses at least. Remember that you must first choose control surfaces. For instance, think of what kind of momentum flux there could be through control surfaces like the red dashed lines in the side figure: (a) at the opened side; (b) in between the opened side and the middle of the box, parallel to the opened side.

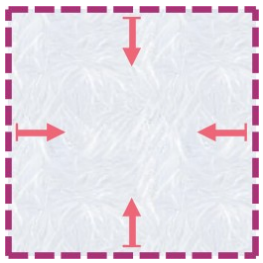
Let us try to infer what are the momentum fluxes through the two horizontal control surfaces depicted in the previous illustration.



- The portion of gas between the two surfaces gets out as soon as we open the upper side of the box. This portion of gas must therefore be receiving outward-oriented momentum. This momentum cannot come from the gravitational momentum supply, because the latter is oriented downward. It must therefore be the result of a momentum flux. This flux cannot be through the surface separating the gas from the vacuum, and it seems implausible that it is a flux through the contact surface with the box. Intuitively we understand that it is a flux coming from the gas further within the box (side picture, left).
- As long as the upper side of the box is closed, however, portion of gas between the two surfaces stays at rest. The flux of outward-oriented momentum coming from the gas further within the box must therefore be compensated by a flux of inward-oriented momentum that comes from the closed side of the box (side picture, right).

Repeating this kind of reasoning with horizontal control surfaces further within the box, and then with vertical control surfaces, we arrive

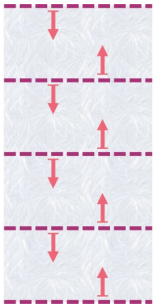
at the following conclusion: through any small control surface delimiting some gas *at rest* there is an influx of inward-oriented momentum. The total momentum within the control volume doesn't change because the influxes from opposite parts of the whole closed control surface cancel each other.



Internal pressure

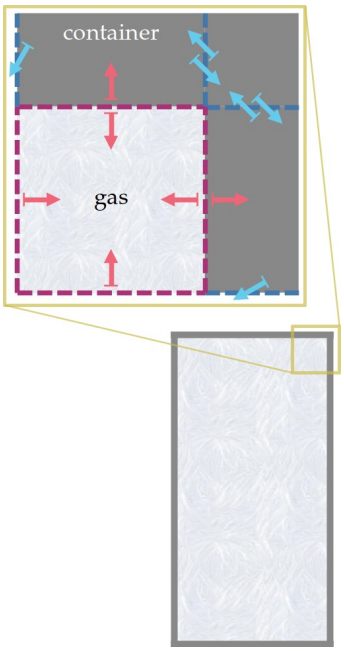
The compressive momentum flux that occurs through any control surface that we may choose within a gas is called the **internal pressure** of the gas. More precisely, the internal pressure is the flux divided by the area through which it occurs.

In Chapter 11 we shall discuss a constitutive relation for the internal pressure of particular kinds of gas.



If we consider a collection of parallel control surfaces within a gas, we see that there is a flux of momentum across all of them; this momentum has the same orientation as the chosen crossing direction. In the side illustration, for instance, if we cross the horizontal control surfaces (red dashed lines) from bottom to top, we'll measure a flux of upward momentum. By flux symmetry, if we choose the opposite crossing direction, from top to bottom, then we'll measure a flux of downward momentum.

From a molecular point of view, this momentum flux mainly comes from the momentum transported by the molecules that make up the gas. Some molecules are moving with an upward velocity, and therefore each of them has an upward momentum; this upward momentum travels upward, transported by the molecule. Other molecules are moving with a downward velocity, and therefore each of them has a downward momentum, which is also transported downward.



At the surface where the gas is in contact with its containing box, the momentum gets transferred to the box. An interesting question then arises. When the box is also at rest, the momentum contained in each portion of it is zero and remains zero. Where does the momentum that it receives from the gas go?

The answer is that an opposite circulation of momentum takes places through the walls of the box: the momentum that one side of the box receives from the gas is transported to the opposite side, where it is given back to the gas. By the symmetry of flux, this circulation can also be visualized in the opposite direction. The flux of momentum along the walls of the box is therefore tensile, or partially tensile and partially

shearing. The side picture gives a *qualitative* illustration of the momentum fluxes that occur through various control surfaces at the corner of a box that contains an amount of gas and is surrounded by a vacuum. From a molecular point of view, the flux of momentum through the box occurs not because of transport by molecules, but because of transport by the electromagnetic field that exists with the box's molecules.

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].

› § 10.2 page 170

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 10.6

Before continuing, try to do 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?

› § 6.1 page 149

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, $l_{\mathrm{n}} \approx 0\,\mathrm{m}$. We choose a coordinate system at rest with the Earth's surface; let's denote it (y, z), with z pointing upward. 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: one tightly enveloping body a, one body b, 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 and the elastic constant k, depend on the time t.

Body a: The control volume for body a is small and can be characterized simply by its position vector $\boldsymbol{r}_a$. The mass-energy in this control volume is m_a.

By construction, conservation of matter automatically holds for this control volume. It remains, hidden, in the relation between the position and velocity of the control volume: $\boldsymbol{v}_a(t) = \mathrm{d}\boldsymbol{r}_a(t)/\mathrm{d}t$.

› § 7.3 page 162

This control volume contains an amount of momentum $\boldsymbol{P}_a = m_a\,\boldsymbol{v}_a$, according to the Newtonian constitutive equation.

› § 10.2 page 169

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 **in**flux $\boldsymbol{F}_{as}$. We are using this notation for the fluxes: '$\boldsymbol{F}_{as}$' is the influx in body a coming from the spring; analogously for $\boldsymbol{F}_{bs}$ and body b.

If we consider gravity effects, this control volume also has a constant momentum supply $\boldsymbol{G}_a = -m_a\,g\,[0, 1]$, according to the constitutive equation for gravitational force.

› § 10.2 page 174

For this control volume we have therefore the following momentum balance and constitutive relations:

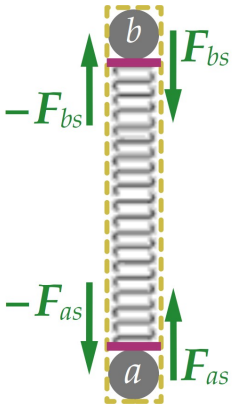
momentum balance	$\dfrac{\mathrm{d}\boldsymbol{P}_a(t)}{\mathrm{d}t} = \boldsymbol{F}_{as}(t) + \boldsymbol{G}_a(t)$
velocity	$\dfrac{\mathrm{d}\boldsymbol{r}_a(t)}{\mathrm{d}t} = \boldsymbol{v}_a(t)$
const. relation	$\boldsymbol{P}_a(t) = m_a \boldsymbol{v}_a(t)$
const. relation	$\boldsymbol{G}_a(t) = -m_a\, g\, [0,1]$

Body b: An analogous analysis can be made for the control volume of body b. It has position $\boldsymbol{r}_b$, total mass-energy m_b, total momentum $\boldsymbol{P}_b = m_b\, \boldsymbol{v}_b$.

The only momentum flux occurs across the portion of control surface in common with that of the spring; there we have an **in**flux $\boldsymbol{F}_{bs}$.

For this control volume we have the following momentum balance and constitutive relations:

momentum balance	$\dfrac{\mathrm{d}\boldsymbol{P}_b(t)}{\mathrm{d}t} = \boldsymbol{F}_{bs}(t) + \boldsymbol{G}_b(t)$
velocity	$\dfrac{\mathrm{d}\boldsymbol{r}_b(t)}{\mathrm{d}t} = \boldsymbol{v}_b(t)$
const. relation	$\boldsymbol{P}_b(t) = m_b \boldsymbol{v}_b(t)$
const. relation	$\boldsymbol{G}_b(t) = -m_b\, g\, [0,1]$



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 $\boldsymbol{F}_{as}$, $\boldsymbol{F}_{bs}$ happen across these surfaces, but their vectors are placed on the side in the illustration. The vectors representing the velocities $\boldsymbol{v}_a$, $\boldsymbol{v}_b$, and the momentum supplies $\boldsymbol{G}_a$, $\boldsymbol{G}_b$ are omitted for clarity.

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 $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$. 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 $\boldsymbol{P}$ and supply $\boldsymbol{G}$, also zero.

At the surface where the spring is in contact with body a (see side picture above), there is an **in**flux of momentum for the spring, equal to $-\boldsymbol{F}_{as}$. The minus sign comes from the symmetry of flux, because $\boldsymbol{F}_{as}$ is the influx for body a.

› § 4.6 page 87

Analogously, at the surface where the spring is in contact with body b, there is a momentum influx for the spring equal to $-\boldsymbol{F}_{bs}$.

The total momentum influx for the spring is therefore $\boldsymbol{F} = -\boldsymbol{F}_{as} - \boldsymbol{F}_{bs}$ by extensivity.

The Hookean constitutive relation (10.3) says that the momentum **ef**flux across one surface, say the one in contact with body a, must be

$$\boldsymbol{F}_{as} = -k\,(\boldsymbol{r}_a - \boldsymbol{r}_b)$$

because $\boldsymbol{r}_a - \boldsymbol{r}_b$ 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	$\dfrac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t} = \boldsymbol{F}(t) + \boldsymbol{G}(t)$
total influx by extensivity	$\boldsymbol{F}(t) = -\boldsymbol{F}_{as}(t) - \boldsymbol{F}_{bs}(t)$
assumption for spring	$\boldsymbol{P}(t) = 0$
assumption for spring	$\boldsymbol{G}(t) = 0$
Hooke const. relation	$\boldsymbol{F}_{as}(t) = -k\,[\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)]$

Something peculiar happens in the case of the spring: **the momentum balance for the spring reduces to a very simple equation:**

$$0 = \boldsymbol{F} = -\boldsymbol{F}_{as} - \boldsymbol{F}_{bs}$$

We thus find that the effluxes at the extremities of the spring must be opposite:

$$\boldsymbol{F}_{as} = -\boldsymbol{F}_{bs}$$

this result is therefore a consequence of momentum balance, *not* of the 'principle of action and reaction'.

This is of course an approximation. In a more detailed description of this physical system, the momentum balance for the spring would tell us how momentum flows from one end to the other of the spring, over time. In our simplified system we are essentially assuming that this flow is *instantaneous*. This is why we end up with an equation relating two momentum fluxes, $\boldsymbol{F}_{as}$ and $\boldsymbol{F}_{bs}$, *at the same time*.

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.

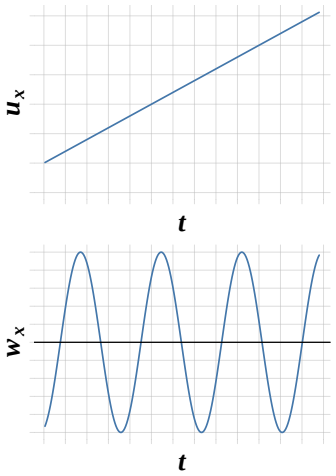
The set of equations above is usually solved analytically by introducing two alternative vector variables:

$$\boldsymbol{u} \coloneqq \frac{m_a\,\boldsymbol{r}_a + m_b\,\boldsymbol{r}_b}{m_a + m_b} \qquad \boldsymbol{w} \coloneqq \boldsymbol{r}_a - \boldsymbol{r}_b$$

the first, $\boldsymbol{u}$, 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):

$$\frac{\mathrm{d}^2\boldsymbol{u}(t)}{\mathrm{d}t^2} = 0 \qquad \frac{\mathrm{d}^2\boldsymbol{w}(t)}{\mathrm{d}t^2} + k\,\frac{m_a + m_b}{m_a\,m_b}\,\boldsymbol{w}(t) = 0$$

that can be solved analytically.



Every component of $\boldsymbol{u}(t)$ is a linear function of time, and every component of $\boldsymbol{w}(t)$ 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 $\boldsymbol{u}(t)$ and the amplitudes and phases of $\boldsymbol{w}(t)$ depend on the initial conditions.

Exercise 10.7

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 $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$.

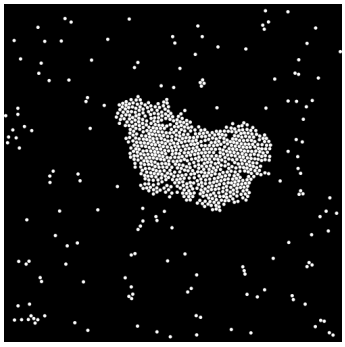
Many-body systems

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

› § 10.2 page 172

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.

This way we can numerically simulate complex physical phenomena like a planetary system, or a collection of molecules – thus entering the field of molecular-dynamics simulation[17]. You can see an example of the latter kind of simulation, using the Lennard-Jones pairwise force of formula (10.6), at the *Interactive Molecular Dynamics* webpage[18].



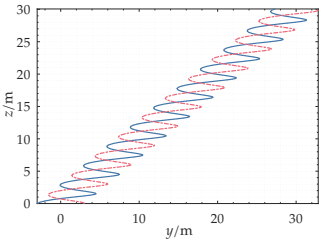
Snapshot of a molecular-dynamics simulation from *Interactive Molecular Dynamics*[16]

10.4 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, a particular set of initial values $\boldsymbol{r}_a(t_0), \boldsymbol{r}_b(t_0), \boldsymbol{v}_a(t_0), \boldsymbol{v}_b(t_0)$ for the tennis balls leads to the trajectories shown in the plot on the side. The distance between the two tennis balls 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, $\boldsymbol{r}_a \approx \boldsymbol{r}_b$; and we see that the system as a whole is essentially moving on a straight line. A zoomed-out plot of $\boldsymbol{r}_a(t)$ or $\boldsymbol{r}_b(t)$ 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 $\boldsymbol{P}_{\text{tot}}$, momentum flux $\boldsymbol{F}_{\text{tot}}$, and

› § 3.2 page 54

momentum supply $\boldsymbol{G}_{\text{tot}}$ for this closed control surface can be obtained by adding up those of the three original control surfaces, spring+a+b:

$$m_{\text{tot}} = 0 + m_a + m_b \qquad \boldsymbol{P}_{\text{tot}} = 0 + \boldsymbol{P}_a + \boldsymbol{P}_b$$
$$\boldsymbol{F}_{\text{tot}} = (-\boldsymbol{F}_{as} - \boldsymbol{F}_{bs}) + \boldsymbol{F}_{as} + \boldsymbol{F}_{bs} \qquad \boldsymbol{G}_{\text{tot}} = 0 + \boldsymbol{G}_a + \boldsymbol{G}_b$$

Note that the total momentum influx is zero: $\boldsymbol{F}_{\text{tot}} = 0$. 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:

$$\boldsymbol{P}_{\text{tot}}(t_1) = \boldsymbol{P}_{\text{tot}}(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}_{\text{tot}}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{G}_{\text{tot}}(t)\,\mathrm{d}t$$

If the total momentum influx and supplies are zero, then we find $\boldsymbol{P}_{\text{tot}}(t_1) = \boldsymbol{P}_{\text{tot}}(t_0)$, which explains why we found that the velocity for the whole system could be considered as constant.

10.5 Numerical time integration: a strategy

The physical system of two masses connected by a Hookean spring can be solved, and its behaviour predicted, analytically. But finding analytical solutions becomes more difficult or even impossible as soon as we consider non-Hookean springs and more general constitutive relations, or more masses. We then resort to numerical time-integration methods to predict the behaviour of such systems.

› § 10.2 page 172

The essential idea of such numerical methods was introduced in an earlier chapter. Let us now apply it to the masses-and-spring system. If we write an algorithm that numerically simulates a Hookean spring, it is then very easy to generalize it to a non-Hookean spring and to even more complex constitutive relations.

› § 6.1 page 144

We wrote our first time-integration algorithm by proceeding intuitively, without following any particular scheme. Our present physical system involves more quantities and more constitutive equations. It is therefore convenient to find a more systematic way to build a simulation algorithm. Let's see the main steps using the present physical system as an example.

› § 6.1 page 153

In trying to reach a systematic way we shall also realize again the importance of the universal balance laws.

Overview of the relevant equations

First we must have a clear list of the balance laws and constitutive relations that apply to the physical system. In our case they are

› § 10.3 page 191

$$\frac{\mathrm{d}\boldsymbol{P}_a(t)}{\mathrm{d}t} = \boldsymbol{F}_{as}(t) + \boldsymbol{G}_a(t) \qquad \text{momentum balance for body } a$$

$$\frac{\mathrm{d}\boldsymbol{P}_b(t)}{\mathrm{d}t} = \boldsymbol{F}_{bs}(t) + \boldsymbol{G}_b(t) \qquad \text{momentum balance for body } b$$

$$\frac{\mathrm{d}\boldsymbol{r}_a(t)}{\mathrm{d}t} = \boldsymbol{v}_a(t) \qquad \text{velocity of body } a$$

$$\frac{\mathrm{d}\boldsymbol{r}_b(t)}{\mathrm{d}t} = \boldsymbol{v}_b(t) \qquad \text{velocity of body } b$$

$$\boldsymbol{F}_{bs}(t) = -\boldsymbol{F}_{as}(t) \qquad \text{momentum balance for spring}$$

$$\boldsymbol{P}_a(t) = m_a \boldsymbol{v}_a(t) \qquad \text{Newtonian momentum body } a$$

$$\boldsymbol{P}_b(t) = m_b \boldsymbol{v}_b(t) \qquad \text{Newtonian momentum body } b$$

$$\boldsymbol{G}_a(t) = -m_a\, g\, [0, 1] \qquad \text{momentum supply body } a$$

$$\boldsymbol{G}_b(t) = -m_b\, g\, [0, 1] \qquad \text{momentum supply body } b$$

$$\boldsymbol{F}_{as}(t) = -k\, [\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)] \qquad \text{Hooke's law}$$

where each vector equation represents two equations: one for component y, one for z. We could have written the balance laws in integral, rather than differential, form. The latter form is simply more compact. Recall the peculiarity about the momentum balance for the spring: since the spring is assumed to always have zero momentum, the balance simplified to the equation $\boldsymbol{F}_{bs}(t) = -\boldsymbol{F}_{as}(t)$, where no time derivative appear anymore.

In order to do numerical time integration, we approximate the balance laws and the velocities with finite-difference approximations. Our set of

› § 6.1 page 146

equations is therefore rewritten as follows:

$$\boldsymbol{P}_a(t+\Delta t) \approx \boldsymbol{P}_a(t) + [\boldsymbol{F}_{as}(t) + \boldsymbol{G}_a(t)]\,\Delta t \qquad \text{momentum balance for body } a$$
$$\boldsymbol{P}_b(t+\Delta t) \approx \boldsymbol{P}_b(t) + [\boldsymbol{F}_{bs}(t) + \boldsymbol{G}_b(t)]\,\Delta t \qquad \text{momentum balance for body } b$$
$$\boldsymbol{r}_a(t+\Delta t) \approx \boldsymbol{r}_a(t) + \boldsymbol{v}_a(t)\,\Delta t \qquad \text{velocity of body } a$$
$$\boldsymbol{r}_b(t+\Delta t) \approx \boldsymbol{r}_b(t) + \boldsymbol{v}_b(t)\,\Delta t \qquad \text{velocity of body } b$$
$$\boldsymbol{F}_{bs}(t) = -\boldsymbol{F}_{as}(t) \qquad \text{momentum balance for spring}$$
$$\boldsymbol{P}_a(t) = m_a \boldsymbol{v}_a(t) \qquad \text{Newtonian momentum body } a$$
$$\boldsymbol{P}_b(t) = m_b \boldsymbol{v}_b(t) \qquad \text{Newtonian momentum body } b$$
$$\boldsymbol{G}_a(t) = -m_a\, g\, [0,1] \qquad \text{momentumsupply body } a$$
$$\boldsymbol{G}_b(t) = -m_b\, g\, [0,1] \qquad \text{momentum supply body } b$$
$$\boldsymbol{F}_{as}(t) = -k\,[\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)] \qquad \text{Hooke's law}$$

This is now our starting point. Let's see a reasoned set of steps to write a script that implements these equations.

A strategy for writing a numerical time-integration algorithm

Our strategy can be divided into six systematic steps, which we reason out now. The script can be written along, as we follow them. After each step, take a look at the example script of Table 10.1 on page 204, and locate the lines where that step was implemented.

0. Find any constants appearing in the equations. Constitutive relations typically contain constant quantities, that is, quantities that don't change in time. These constants must be known in order to simulate the system. They are therefore **declared at the beginning of the script**. In our case the constants are the masses m_a, m_b, the gravitational acceleration g, and the elastic constant k. Other examples could be fixed lengths, areas, volumes.

The discrete timestep Δt is also a constant in our scripts. In more advanced techniques, however, the timestep may be *adaptive*, that is, it may change at every iteration, depending on particular conditions. This can lead to smaller numerical errors.

Volume contents, fluxes, and supplies usually depend on time; but in special situations some of them may turn out to be constant as well. In such situations it is computationally efficient to declare them before the

time-stepping loop, rather than computing them anew (obtaining always the same value) at each time step. In our case we see that the gravity supplies of the two bodies, $\boldsymbol{G}_a(t) = -m_a\, g\, [0,1]$ and $\boldsymbol{G}_b(t) = -m_b\, g\, [0,1]$ are constant, because the masses, the gravitational acceleration, and the unit vector $[0,1]$ are constant.

page 204

1. Find which equations drive the system forward in time. Some equations must tell us the values of some quantities at the next time point $t + \Delta t$, given quantities at the present time point t. They must therefore be written in the **core of the time-stepping loop**, together with the update of the time variable t.

These driving equations easy to identify: '$t + \Delta t$' appears in them. In our case they are the balances for the momenta $\boldsymbol{P}_a$, $\boldsymbol{P}_b$, and the velocity equations for the two bodies, having two coordinate components each.

The fact that the forward-driving equations are fundamental balances is not a coincidence: as mentioned several times already, *the universal balances are important because they relate later times to earlier times*. This fact becomes especially evident when we write a simulation algorithm.

In our case there's also one more balance, the momentum balance for the spring, but owing to the special assumptions about the spring (being massless), it simplifies to a same-time equation.

page 204

2. Choose a state for the physical system. The driving equations in the core lines of the loop calculate later quantities from a particular set of present quantities. We must therefore make sure that the values of these quantities are defined before the core lines, and that they are known before the time-stepping loop begins as well. In our case we see that we need 8 quantities, underlined below, with two components each, for a total of 16 numbers:

$$\begin{aligned}
\boldsymbol{P}_a(t+\Delta t) &\approx \underline{\boldsymbol{P}_a(t)} + [\underline{\boldsymbol{F}_{as}(t)} + \boldsymbol{G}_a(t)]\,\Delta t \\
\boldsymbol{P}_b(t+\Delta t) &\approx \underline{\boldsymbol{P}_b(t)} + [\underline{\boldsymbol{F}_{bs}(t)} + \boldsymbol{G}_b(t)]\,\Delta t \\
\boldsymbol{r}_a(t+\Delta t) &\approx \underline{\boldsymbol{r}_a(t)} + \underline{\boldsymbol{v}_a(t)}\,\Delta t \\
\boldsymbol{r}_b(t+\Delta t) &\approx \underline{\boldsymbol{r}_b(t)} + \underline{\boldsymbol{v}_b(t)}\,\Delta t
\end{aligned}$$

The supplies $\boldsymbol{G}_a$, $\boldsymbol{G}_b$ are constant and already declared, so we don't need to worry about them.

The eight quantities above are not all independent, thanks to constitutive relations that relate some of them. For instance, if we assign the

velocity $\boldsymbol{v}_a$, then the momentum $\boldsymbol{P}_a$ is determined by the constitutive relation $\boldsymbol{P}_a = m_a \boldsymbol{v}_a$; and if we assign the positions $\boldsymbol{r}_a, \boldsymbol{r}_b$, then the force $\boldsymbol{F}_{as}$ is determined by the constitutive relation $\boldsymbol{F}_{as} = -k\,(\boldsymbol{r}_a - \boldsymbol{r}_b)$.

Our task now is to find a *minimum* set among these quantities, from which all others can be determined via constitutive relations. As a rule of thumb, the number of minimum quantities is given by

(number needed in driving equations) − (number of same-time relations)

considering all vector components. In our case the same-time relations are four, with two components each:

$$\boldsymbol{F}_{bs}(t) = -\boldsymbol{F}_{as}(t)$$
$$\boldsymbol{P}_a(t) = m_a \boldsymbol{v}_a(t) \qquad \boldsymbol{P}_b(t) = m_b \boldsymbol{v}_b(t)$$
$$\boldsymbol{F}_{as}(t) = -k\,[\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)]$$

so we should find a minimum set of $(2 \times 8) - (2 \times 4) = 8$ quantities.

The choice is not unique. Convince yourself that this minimal set could be used:

$$\boldsymbol{r}_a\,,\ \boldsymbol{v}_a\,,\ \ \boldsymbol{r}_b\,,\ \boldsymbol{v}_b$$

or this:

$$\boldsymbol{r}_a\,,\ \boldsymbol{P}_a\,,\ \ \boldsymbol{F}_{as}\,,\ \boldsymbol{v}_b$$

Usually we prefer a minimal set that consists of easily observable or measurable quantities, like positions, velocities, temperatures. In our case, let us agree to use $\boldsymbol{r}_a, \boldsymbol{v}_a, \boldsymbol{r}_b, \boldsymbol{r}_b$; note that each of these has two components: $y_a, z_a, y_b, \ldots$ and so on, for a total of four.

This minimum set is called the *state* of the physical system:

State of a physical system

The **state** of a physical system is the minimal amount of information needed to drive the system from one time point to the next.

It is usually encoded in a minimal set of time-dependent quantities, but the choice of quantities is often not unique. In this case we say that different sets of quantities *represent the same state*.

The values of a state at the beginning of a numerical time integration are called **initial conditions**.

The state of our system is $\boldsymbol{r}_a, \boldsymbol{v}_a, \boldsymbol{r}_b, \boldsymbol{r}_b$. The values of the state at the initial time, that is, the initial conditions, need to be **declared before the time-stepping loop**, together with the value of the initial time.

page 204

3. From the state, determine the quantities necessary for forward-driving. Recall that the forward-driving equations require, at each time step, the 2×8 quantities

$$\boldsymbol{P}_a(t)\,,\ \boldsymbol{F}_{as}(t)\,,\ \boldsymbol{P}_b(t)\,,\ \boldsymbol{F}_{bs}(t)\,,\ \boldsymbol{r}_a(t)\,,\ \boldsymbol{v}_a(t)\,,\ \boldsymbol{r}_b(t)\,,\ \boldsymbol{v}_b(t)$$

We must therefore find these from our state, at each new time step, by using any necessary constitutive relations. In our case:

$$\begin{aligned}
\boldsymbol{P}_a(t) &= m_a\,\boldsymbol{v}_a(t)\\
\boldsymbol{F}_{as}(t) &= -k\,[\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)]\\
\boldsymbol{P}_b(t) &= m_b\,\boldsymbol{v}_b(t)\\
\boldsymbol{F}_{bs}(t) &= -\boldsymbol{F}_{as}(t) = -\Big(-k\,[\boldsymbol{r}_a(t) - \boldsymbol{r}_b(t)]\Big)\\
\boldsymbol{r}_a(t) &\quad \text{given}\\
\boldsymbol{v}_a(t) &\quad \text{given}\\
\boldsymbol{r}_b(t) &\quad \text{given}\\
\boldsymbol{v}_b(t) &\quad \text{given}
\end{aligned}$$

Note that to find the force $\boldsymbol{F}_{bs}$ from the state we needed to use two equations.

In the code, the formulae that calculate the quantities necessary to the forward-driving lines from the state **need to be written within the time-stepping loop, right before the forward-driving lines**.

page 204

4. Find the new state from the time-updated quantities. The time-stepping loop of our code is now able to determine quantities at a later time $t + \Delta t$, given the state at the present time t. The quantities that we obtain at the later time are

$$\boldsymbol{P}_a(t+\Delta t)\,,\ \ \boldsymbol{P}_b(t+\Delta t)\,,\ \ \boldsymbol{r}_a(t+\Delta t)\,,\ \ \boldsymbol{r}_b(t+\Delta t)$$

We now want to be able to start the next loop iteration. The next iteration needs the new value of the state $\boldsymbol{r}_a, \boldsymbol{v}_a, \boldsymbol{r}_b, \boldsymbol{r}_b$; but we have $\boldsymbol{P}_a, \boldsymbol{P}_b, \boldsymbol{r}_a, \boldsymbol{r}_b$. We need therefore to **convert the updated quantities back to the quantities that constitute the state, within the time loop**. This is again done by using same-time equations. In our case we use

$$\begin{aligned}
\boldsymbol{r}_a(t+\Delta t) &\quad \text{given}\\
\boldsymbol{r}_b(t+\Delta t) &\quad \text{given}\\
\boldsymbol{v}_a(t+\Delta t) &= \boldsymbol{P}_a(t)/m_a\\
\boldsymbol{v}_b(t+\Delta t) &= \boldsymbol{P}_b(t)/m_b
\end{aligned}$$

Now are ready to go back to the beginning of the loop!

page 204

5. Decide the condition for stopping the time loop. For how long should our simulation run? This obviously depends on the application. We may want to run it for a given amount of time. Or we may want to stop it as soon as some condition is met; for instance, if the position or speed of a body reach particular values. The possibilities are endless.

We insert these conditions in the for- or while-loop, and initialize them as necessary.

page 204

After these six steps, the essential part of our numerical-time-integration script is ready. The script will probably need additional lines to perform tasks that depend on the specific problem. For example we may want to store the numerical values of some quantities at different times, for later analysis; or plot some of them as the simulation evolves. The ways these tasks are implemented are often heavily dependent on the programming language, and we therefore cannot discuss them in detail here.

Table 10.1 Example script for numerical time integration of the spring & bodies system. The constants, initial values, and the stop condition (stop at 10 s) can of course be different. Vector components are declared on one line.

```
%%%% 0.  Constants and timestep
ma = 1;
mb = 1;
g = 9.80665;
k = 4;
Gya = 0; Gza = -ma*g;
Gyb = 0; Gzb = -mb*g;
dt = 0.01;

%%%% 2.  State: ya,za,vya,vza,yb,zb,vyb,vzb; initial conditions
t = 0;
ya = 0; za = 0;
yb = 0.1; zb = 0.1;
vya = 0; vza = 0;
vyb = 0; vzb = 0.1;

%%%% 5.  Condition for stopping loop
while t < 10

    %%%% 3.  Calculate forward-driving quantities from state
    Pya = ma*vya;  Pza = ma*vza;
    Fyab = -k*(ya-yb);  Fzab = -k*(za-zb);
    Pyb = mb*vyb;  Pzb = mb*vzb;
    Fyba = -Fyab;  Fzba = -Fzab;

    %%%% 1.  Drive forward in time
    Pya = Pya + (Fyab + Gya)*dt;  Pza = Pza + (Fzab + Gza)*dt;
    Pyb = Pyb + (Fyba + Gyb)*dt;  Pzb = Pzb + (Fzba + Gzb)*dt;
    ya = ya + vya*dt;  za = za + vza*dt;
    yb = yb + vyb*dt;  zb = zb + vzb*dt;
    t = t + dt;

    %%%% 4.  Calculate new state from forward-driven quantities
    vya = Pya/ma;  vza = Pza/ma;
    vyb = Pyb/mb;  vzb = Pzb/mb;

end
```

Exercise 10.8

1. Examine our previous script `tennisball.m` for numerically integrating the motion of a tennis ball, and pinpoint where the six building steps above are implemented.

   Which quantities constitute the *state* in that script?

› § 6.1 page 154

2. Implement the pseudo-code of Table 10.1 on page 204 in your preferred programming language.

   Your script should save all or some values of the state $\boldsymbol{r}_a, \boldsymbol{r}_b, \boldsymbol{v}_a, \boldsymbol{v}_b$, and their corresponding t values, so that it's possible to plot their time evolution afterwards.

   (If you had difficulties writing your script, for the following exercises you can download and use the script `hooke_spring.m`[19], written for Octave[20]/MATLAB, as a starting point.)

3. Run your script with the following values:

$$
\begin{aligned}
t_0 = 0\,\mathrm{s} \qquad t_1 = 10\,\mathrm{s} \qquad \Delta t = 0.001\,\mathrm{s} \\
m_a = m_b = 2\,\mathrm{kg} \qquad k = 5\,\mathrm{N/m} \\
\boldsymbol{r}_a(t_0) = [-3,\, 0]\,\mathrm{m} \qquad \boldsymbol{r}_b(t_0) = [3,\, 0]\,\mathrm{m} \\
\boldsymbol{v}_a(t_0) = [0,\, 0]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [0,\, 0]\,\mathrm{m/s} \\
\boldsymbol{G}_a(t_0) = [0,\, 0]\,\mathrm{N} \qquad \boldsymbol{G}_b(t_0) = [0,\, 0]\,\mathrm{N}
\end{aligned}
$$

   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 $\boldsymbol{P}_a + \boldsymbol{P}_b$ for the system composed by the two bodies and the spring. How do these component change? Why?

4. Run the script with the same parameter as before but the following initial velocity values:

$$
\boldsymbol{v}_a(t_0) = [1,\, 1]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [-1,\, -1]\,\mathrm{m/s}
$$

   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?

5. Run the script with the following initial velocity values:

$$\boldsymbol{v}_a(t_0) = [2,\, 2]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [1,\, 1]\,\mathrm{m/s}$$

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.

6. Play with all the parameter values and initial conditions, and see what happens. Before simulating, try to intuitively predict what the behaviour of the system will be.

Non-Hookean spring: numerical time integration

The system of two masses connected by a *non-Hookean* spring with constitutive relation (10.4) is analytically challenging. Its numerical solution, however, only involves a change of a couple of lines to the code that you wrote for Exercise 10.8.

› § 10.2 page 172

A small mathematical change in a constitutive relation can lead to a much richer set of behaviours, as the examples in the side plots below show.

Exercise 10.9

Modify the script you wrote for Exercise 10.8 so that the spring is modelled by the non-Hookean constitutive relation (10.4). This constitutive relation can be implemented with an `if`-statement, for instance.
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:

$$t_0 = 0\,\mathrm{s} \qquad t_1 = 10\,\mathrm{s} \qquad \Delta t = 0.001\,\mathrm{s}$$
$$m_a = 0.1\,\mathrm{kg} \qquad m_b = 0.1\,\mathrm{kg} \qquad k = 0.5\,\mathrm{N/m} \qquad l_\mathrm{n} = 0.5\,\mathrm{m}$$
$$\boldsymbol{r}_a(t_0) = [-0.3,\, 0.3]\,\mathrm{m} \qquad \boldsymbol{r}_b(t_0) = [0,\, 0]\,\mathrm{m}$$
$$\boldsymbol{v}_a(t_0) = [1,\, 5]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [0,\, 5]\,\mathrm{m/s}$$
$$\boldsymbol{G}_a(t_0) = m_a \cdot [0,\, 0]\,\mathrm{N/kg} \qquad \boldsymbol{G}_b(t_0) = [0,\, 0]\,\mathrm{N}$$

Set 2:

$$t_0 = 0\,\mathrm{s} \qquad t_1 = 10\,\mathrm{s} \qquad \Delta t = 0.001\,\mathrm{s}$$
$$m_a = 2\,\mathrm{kg} \qquad m_b = 2\,\mathrm{kg} \qquad k = 5\,\mathrm{N/m} \qquad l_\mathrm{n} = 5\,\mathrm{m}$$
$$\boldsymbol{r}_a(t_0) = [-3,\, 0]\,\mathrm{m} \qquad \boldsymbol{r}_b(t_0) = [3,\, 0]\,\mathrm{m}$$
$$\boldsymbol{v}_a(t_0) = [0,\, 0]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [0,\, 0]\,\mathrm{m/s}$$
$$\boldsymbol{G}_a(t_0) = [0,\, 0]\,\mathrm{N} \qquad \boldsymbol{G}_b(t_0) = [0,\, 0]\,\mathrm{N}$$

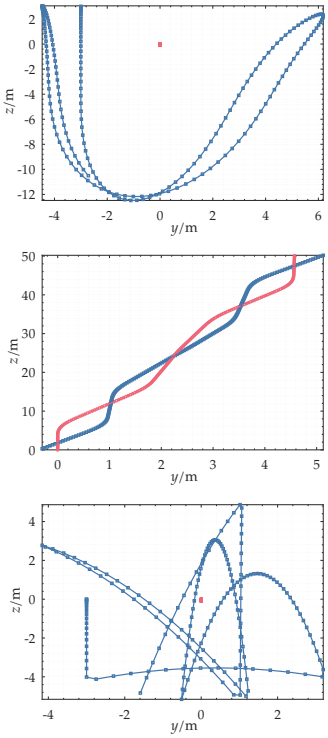
Set 3:

$$t_0 = 0\,\mathrm{s} \qquad t_1 = 10\,\mathrm{s} \qquad \Delta t = 0.001\,\mathrm{s}$$
$$m_a = 1\,\mathrm{kg} \qquad m_b = 5000\,\mathrm{kg} \qquad k = 5000\,\mathrm{N/m} \qquad l_\mathrm{n} = 5\,\mathrm{m}$$
$$\boldsymbol{r}_a(t_0) = [-3,\, 0]\,\mathrm{m} \qquad \boldsymbol{r}_b(t_0) = [0,\, 0]\,\mathrm{m}$$
$$\boldsymbol{v}_a(t_0) = [0,\, 0]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [0,\, 0]\,\mathrm{m/s}$$
$$\boldsymbol{G}_a(t_0) = -m_a\, g \cdot [0,\, 1] \qquad \boldsymbol{G}_b(t_0) = [0,\, 0]\,\mathrm{N}$$

Set 4:

$$t_0 = 0\,\mathrm{s} \qquad t_1 = 10\,\mathrm{s} \qquad \Delta t = 0.001\,\mathrm{s}$$
$$m_a = 1\,\mathrm{kg} \qquad m_b = 5000\,\mathrm{kg} \qquad k = 5\,\mathrm{N/m} \qquad l_\mathrm{n} = 5\,\mathrm{m}$$
$$\boldsymbol{r}_a(t_0) = [-3,\, 3]\,\mathrm{m} \qquad \boldsymbol{r}_b(t_0) = [0,\, 0]\,\mathrm{m}$$
$$\boldsymbol{v}_a(t_0) = [0,\, 0]\,\mathrm{m/s} \qquad \boldsymbol{v}_b(t_0) = [0,\, 0]\,\mathrm{m/s}$$
$$\boldsymbol{G}_a(t_0) = m_a\, g \cdot [0,\, 1] \qquad \boldsymbol{G}_b(t_0) = [0,\, 0]\,\mathrm{N}$$

Play with other parameters and initial values!



Examples of trajectories of two masses connected by a non-Hookean spring, for different values of parameters and initial conditions. Mass a in blue, Mass b in red

10.6 Example script for non-Hookean spring

Here is an example script that is a solution for Exercise 10.9 p. 206. 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.

› § 6.1 page 154

Download `rubberband2D.m`[21]

```
%%% rubberband2D.m
%% Simulation of two bodies connected by non-Hookean rubber band
%% SI units used throughout
%% Coordinates (y,z)
%%%% Constants
ma = 1; % mass of object a
mb = 5000; % mass of object b
g = 9.80665; % gravitational acceleration
k = 5; % spring constant
ln = 5; % natural length of rubber band
%%
Gya = 0; Gza = -ma*g; % gravity supply on object a
Gyb = 0; Gzb = -mb*g*0; % gravity supply on object b
%%
t1 = 10; % final time
dt = 0.001; % time step
%%%% STATE: y,z,vy,vz for a and b; initial conditions
t = 0; % initial time
ya = -3; za = 3; % initial position of object a
yb = 0; zb = 0; % initial position of object a
vya = 0; vza = 0; % initial velocity of object a
vyb = 0; vzb = 0; % initial velocity of object b
%%
%%%% Plot & saving
%% adjust final time if not multiple of timestep
t1 = t1 + mod(t1-t,dt);
%% Save values of all quantities at some steps during the simulation,
%% for subsequente analysis or plotting
%% (saving at all timesteps could be too costly)
Nsaves = 200; % number of timepoints to save during the simulation
%% Calculate time interval for saving
dsave = (t1-t)/(Nsaves-1);
if abs(dsave) < abs(dt)
  error('time interval between saves is smaller than timestep')
end
%% Initialize vectors to contain saved values
tSave = nan(Nsaves,1); % time
yaSave = nan(Nsaves,1); zaSave = nan(Nsaves,1); % position object a
ybSave = nan(Nsaves,1); zbSave = nan(Nsaves,1); % position object b
```

```
vyaSave = nan(Nsaves,1); vzaSave = nan(Nsaves,1); % velocity object a
vybSave = nan(Nsaves,1); vzbSave = nan(Nsaves,1); % velocity object b
%% Save initial values
i = 1; % index that keeps count of savepoints
t0 = t;
tSave(1) = t;
yaSave(1) = ya; zaSave(1) = za;
ybSave(1) = yb; zbSave(1) = zb;
vyaSave(1) = vya; vzaSave(1) = vza;
vybSave(1) = vyb; vzbSave(1) = vzb;
%% Initialize plot
close all;
cols = get(0, 'DefaultAxesColorOrder');
plot(yaSave(1), zaSave(1), 's','Color',cols(1,:)); axis('tight');
hold on;
plot(ybSave(1), zbSave(1), 'o','Color',cols(2,:));
xlabel('{\it y}/m'); ylabel('{\it z}/m');
%%
%%%% Numerical time integration
%% loop
while t < t1
  %% We need y,z,vy,vz,Py,Pz,Fy,Fz for a and b (G constant)
  %% we have y,z,vy,vz
  %% find Py,Pz,Fy,Fz using constitutive relations
  Pya = ma*vya; Pza = ma*vza;
  Pyb = mb*vyb; Pzb = mb*vzb;
  %% non-Hookean constitutive relation
  l = norm([ya-yb, za-zb]); % present length
  if l < ln
    Fyas = 0; % momentum flux from spring to a, y comp.
    Fzas = 0; % momentum flux from spring to a, z comp.
  else
    Fyas = -k*(ya-yb)*(l-ln)/l;
    Fzas = -k*(za-zb)*(l-ln)/l;
  end
  Fybs = -Fyas;
  Fzbs = -Fzas;
  %%
  %% Drive forward in time
  %% update momentum
  Pya = Pya + (Fyas + Gya)*dt;
  Pza = Pza + (Fzas + Gza)*dt;
  Pyb = Pyb + (Fybs + Gyb)*dt;
  Pzb = Pzb + (Fzbs + Gzb)*dt;
  %% update position
  ya = ya + vya*dt;
  za = za + vza*dt;
```

```
  yb = yb + vyb*dt;
  zb = zb + vzb*dt;
  %% update time
  t = t + dt;
  %%
  %% Find new state for next iteration
  %% We need y,z,vy,vz
  %% we have y,z,Py,Pz
  %% find vy,vz using constitutive relations
  vya = Pya/ma;  vza = Pza/ma;
  vyb = Pyb/mb;  vzb = Pzb/mb;
  %%
  %% Check whether to save & plot at this step
  if min(abs([0 dsave] - mod(t-t0, dsave))) <= abs(dt)/2
    i = i+1;
    tSave(i) = t;
    yaSave(i) = ya; zaSave(i) = za;
    ybSave(i) = yb; zbSave(i) = zb;
    vyaSave(i) = vya; vzaSave(i) = vza;
    vybSave(i) = vyb; vzbSave(i) = vzb;
    plot(ya, za, 's','Color',cols(1,:));
    plot(yb, zb, 'o','Color',cols(2,:));
    pause(0.001);
  end
end
%% Plot full trajectory
plot(yaSave,zaSave,'-','Color',cols(1,:));
plot(ybSave,zbSave,'-.','Color',cols(2,:));
```

URLs for chapter 10

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.hybridairvehicles.com/
6. https://www.britannica.com/science/Archimedes-principle
7. https://www.jaerenluftsport.no
8. https://www.britannica.com/topic/locomotion/Gravitational-gliding#ref497010
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-oscillations#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/hooke_spring.m
20. https://octave.org/
21. https://pglpm.github.io/7wonders/code/rubberband2D.m

11 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

11.1 Formulation and generalities

Balance of energy

Volume content: E Flux: Φ Supply: R

$$E(t_1) = E(t_0) + \int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t + \int_{t_0}^{t_1} R(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t) \tag{11.1}$$

integral form — differential form

The balance of energy is extremely important in physical phenomena that underlie many modern (post-industrial revolution[1]) technologies. It must often be explicitly accounted for, together with the balance of momentum; and it often is the main governing balance, when the balance of momentum can be neglected. It's the relevant balance 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 $\boldsymbol{P}$, and a plethora of constitutive relations for its flux $\boldsymbol{F}$. 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, auxiliary quantities like metric and temperature, and others.

› § 10.1 page 168

› § 3.11 page 71

The balance of energy is therefore also very important for numerical time integration. Recall that balance laws, if used alone, 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 particular "gluing" role.

› § 6.1 page 149

In physical phenomena where the balance of energy must be taken explicitly into account, an auxiliary quantity typically appears in our description of the phenomenon: **temperature**.

› § 3.11 page 71

Definitions of total energy

In the physics literature we can read about energies and energy fluxes having all sorts of names – 'internal', 'kinetic', 'potential', 'elastic', 'electromagnetic', and others. We can also read about different kinds of energy balances, which can be related to one another through particular sequences of mathematical steps.

One confusing aspect of such variety is that it becomes unclear what is, and what is not, to be counted as "energy content". One example is 'gravitational potential energy'. Some texts define it as "energy an object possesses because of its position in a gravitational field"[2], that is, as energy content. Other texts, especially in fluid dynamics or material science, do not include any gravitational terms in the energy content; related terms appear instead as energy fluxes or supplies. Another example is 'kinetic energy', which is sometimes included in the total energy content, and sometimes not, for instance in fluid dynamics.

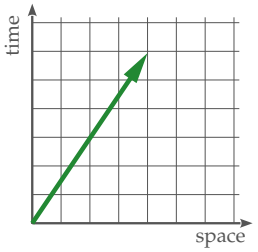
One can show that all these points of view have mathematically the same consequences. Is the definition of 'total energy content' arbitrary, then? Yes it is, and relativity theory gives a clear understanding of this arbitrariness.

Relativity shows that *the definition of "total energy" depends on the choice of a reference velocity and a reference clock*. This choice of velocity & clock

$$\underbrace{\frac{\partial}{\partial t}\left(\tfrac{1}{2}\rho v^2 + \rho\hat{U}\right)}_{\text{rate of increase of energy per unit volume}} \underbrace{+\,\rho(\mathbf{v}\cdot\mathbf{g})}_{\text{rate of work done on fluid per unit volume by external forces}}$$

Snippets from a formula on p. 330 of *Introductory Transport Phenomena* (Bird et al. 2015). This text does not consider gravitational potential energy to be part of the total energy. Compare the expression above on the right with eq. (11.3) on page 218 below.

can even be different at each spacetime point. Furthermore, this choice 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 clocks to define 'total energy'.



A reference 'velocity & clock' is simply a vector with four components in four-dimensional spacetime

Given a coordinate system (t, x, y, z), one of the following three choices is usually implicitly made, when electromagnetic fields are negligible:

- At each point in space and time we may choose a zero coordinate velocity – that is, we don't move with respect to the coordinates (x, y, z) – and we may measure time according to coordinate time t.

  In Newtonian approximation, the total energy content defined by this choice includes so-called internal-energy, kinetic-energy, and gravitational-potential-energy terms.

- If matter is present, we may choose the matter's velocity (which is related to the matter flux), and we may measure time according to the proper time of a clock moving together with that body of matter.

  › § 4.15 page 105

  In Newtonian approximation, the total energy content defined by this choice includes only an internal-energy term. 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 may choose a zero coordinate velocity, as in the first case, but measure time according to *proper time* rather than coordinate time t.

  › § 2.1 page 30

  In Newtonian approximation, the total energy content defined by this choice includes internal-energy and kinetic-energy terms, but no terms related to gravity (this is the choice underlying the formula snippet in the previous side picture).

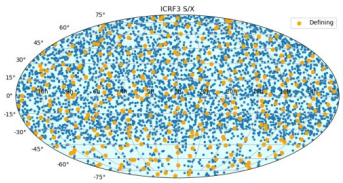
Each of these differently defined energies satisfies a balance law with different fluxes and supplies. An interesting property of the second definition above is that the gravitational field is completely absent in its volume content, flux, and supply.

Note that we do *not* need to use *all* of these differently defined energies at the same time; *one is enough*. Depending on the physical application or problem, one definition can be mathematically more convenient to use than another.

In these notes we shall use the 'total' or 'coordinate' energy, because it has the following interesting or convenient features:

- It is defined anywhere, even where no matter is present.

- Its supply is practically zero with coordinate systems and coordinate time typically used for physical phenomena on Earth (International Terrestrial Reference System, ITRS[3]) or in the solar system (International Celestial Reference System, ICRS[4]). In other words, this energy practically *satisfies a conservation law* in these common coordinate systems.
- Its volume content and flux typically contain terms that refer to the motion of matter and to the gravitational field.



Map of some distant astronomical objects used to define the International Celestial Reference Frame (from *The ICRF*[5])

An aspect still rarely discussed

Most physics textbooks today unfortunately do not mention the dependence of the energy definition on a reference velocity & clock; or worse they mix it up with the dependence on coordinate systems. This dependence, however, is as important as the fact that energy and mass are the same thing.

Forms of energy

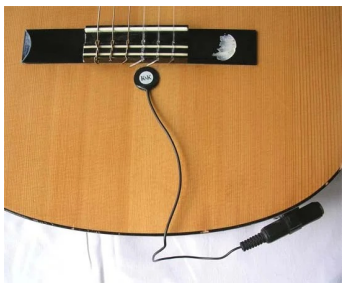
Once we have agreed on a definition of 'total energy content' by means of a reference velocity & clock, a distinction is usually further made between different "forms" of energy, for example 'elastic' or 'electromagnetic'. This distinction as a different origin: it comes from particular *constitutive relations* that connect the total energy to other quantities like matter, temperature, electromagnetic field. There is a wild variety of such constitutive relations, which depend on the physical phenomenon to be described.

In general it is not possible to separate the dependence of total energy on other quantities into a sum of neatly distinct "pieces" like:

$$\text{"mechanical energy"} + \text{"thermal energy"} + \text{"electromagnetic energy"} + \cdots$$

and say "this is the energy contributed by matter", "this is the energy contributed by the electromagnetic field", and so on. *Energy is a property of the other quantities taken as a whole*. The absence of such a separation is the origin of many interesting and useful physical phenomena like piezoelectricity[7] and magnetostriction[8].

In some situations, however, it is possible to *approximately* express energy as a sum of terms that depend on particular quantities only. In these cases we can speak for instance of 'elastic energy', 'radiation energy', and similar expressions denoting the terms in the approximate sum.



The functioning of a guitar pickup is based on the fact that energy cannot be clearly separated into 'mechanical' and 'electromagnetic' (image by Georg Feitscher[6])

Is energy conserved?

Energy is *balanced*, but not *conserved*, according to our definitions of these terms. This is not a controversial statement: recall that many texts use the term 'conservation' in the sense of 'balance', that is, they are not excluding the presence of a volume supply. In fluid mechanics, for instance, the internal-energy definition is typically used, and as previously mentioned this energy definition satisfies a balance law, not a conservation law.

› § 5.3 page 112

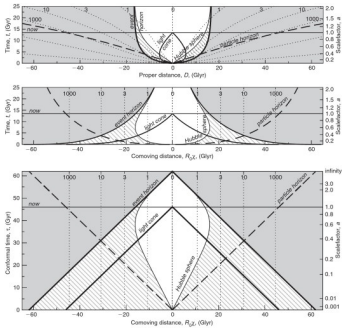
$$\frac{DE}{Dt} = \frac{\sigma_{ij}}{\rho}\frac{\partial u_i}{\partial x_j} + \frac{1}{\rho}\frac{\partial}{\partial x_i}\left(k\frac{\partial T}{\partial x_i}\right)$$

Balance of internal energy in a classic text on fluid dynamics (Batchelor 2000 eq. (3.4.3)). The first summand on the right is the energy supply.

The statement "energy is conserved", however, is often meant in another sense: if we take a control volume where nothing is flowing in or out – no energy, matter, electric charge, electromagnetic field, momentum – then isn't its energy content conserved, constant in time?

The answer to this question is also 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 the definition of energy. 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 universe[9] – and an energy that is strictly conserved can only be defined if the metric is constant along some time direction:

> cosmologists have not done a very good job of spreading the word about something that's been well-understood since at least the 1920's: energy is not conserved in General Relativity Carroll 2010[10]



Depiction of change of spacetime metric on cosmological scales, in three different coordinate systems (from Davis & Lineweaver 2004)

See Carroll's[11] and Baez's[12] interesting posts about this topic.

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 curious difference:

- On planetary scales the spacetime metric is not quite constant in space but 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 fact that energy is not conserved, however, does not mean that we could find ways for our cars or laptops to magically operate by themselves. First, the amount of energy supply 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, the human problem of "using energy" is not about energy *creation* but about energy *conversion* from one form to another. This will be the topic of a later chapter.

11.2 Constitutive relations for energy content

Energy probably has the greatest variety of constitutive relations for its content and its flux. Whole books are dedicated to the discussion of constitutive relations for energy.

In these notes we shall first restrict our focus on general constitutive relations that can be applied in physical phenomena that involve matter but not electric charge or the electromagnetic field. Then we shall restrict our focus even further, on constitutive relations for specific kinds of matter in particular states, for example ideal gases. Often we shall also simplify the discussion by considering physical phenomena where only one spatial dimension is relevant.

Internal, kinetic, gravitational potential energy

If a control volume contains matter, and any electric charge or electromagnetic field is negligible, then that control volume also contains an amount of total energy-mass that can be written in a very general form.

Energy associated with matter

Consider a control volume close to Earth's surface and containing an amount of matter N but no electric charges or electromagnetic fields. This control volume must also be such that the coordinate velocity $\boldsymbol{v}$, molar mass ρ of the matter within, and the z-coordinate are approximately the same throughout the volume (this is true, for instance, if the volume is small enough). Then the total energy-mass content in this control volume is given by

› § 7.2 page 160

$$mc^2 + U + \frac{1}{2}m\boldsymbol{v}^2 + mgz \qquad \text{with} \quad m = \rho N \tag{11.2}$$

In this expression, c is the speed of light, g is the gravitational acceleration, and we have assumed a coordinate system (x, y, z) with z pointing

› § 5.9 page 138

upward.

The term U is called **internal thermodynamic energy** and is given by some further constitutive relation that depends of the physical phenomenon. The terms $\frac{1}{2}m\boldsymbol{v}^2$ and mgz are called **kinetic energy** and **gravitational potential energy**.

Usually we change the "zero" of energy-mass content, removing the term mc^2 and defining the total energy as

$$E = U + \frac{1}{2}m\boldsymbol{v}^2 + mgz \tag{11.3}$$

The formulae above are valid in Newtonian approximation, for speeds smaller than the speed of light and weak gravitational fields.

As we discussed some chapters ago, energy and mass are the same thing. Formula (11.2) above is actually also the mass, multiplied by c^2, contained in the control volume. In these notes we often call only 'm' the mass contained in a control volume, but clearly this is an approximation: we are neglecting the terms $(U + \frac{1}{2}m\boldsymbol{v}^2 + mgz)/c^2$ because they are extremely small – there are in fact even smaller terms, coming from relativity theory, that are neglected in the expressions above.

› § 3.6 page 59

In the sum $mc^2 + U$ there actually isn't a clear-cut separation between mc^2 and U, because the constant value of the molar mass density ρ in $m = \rho N$ is arbitrary to some degree. We can for example remove some constant amount u (possibly negative) from the definition of U, while adding u/c^2 to the definition of m:

$$\underbrace{\left(m + \tfrac{1}{c^2}u\right)}_{\text{new } m} c^2 + \underbrace{(U - u)}_{\text{new } U}$$

The change in m is usually negligible, beyond its 9th decimal digit. But the change in U can be quite large. This means that we can arbitrarily redefine the "zero" of the thermodynamic internal energy U. For water and steam, for instance, U is decreed to be zero at a particular physical condition called triple point[13] (Wagner & Kretzschmar 2008).

'the baryon "mass" density ρ_0, *despite its name, and despite the fact it is sometimes even more misleadingly called "density of rest mass-energy," is actually a measure of the number density of baryons n, and nothing more. It is defined as the product of n with some standard figure for the mass per baryon, μ_0, in some well-defined standard state; thus, $\rho_0 \equiv n\mu_0$.'*
Misner et al. 2017 § 39.3

Exercise 11.1

Consider a control volume containing water, with $m = 1\,\text{kg}$ and internal energy $U = 100\,000\,\text{J}$. Suppose we want to redefine the internal energy so that it is zero instead. By how much should we redefine m, so that

the sum $mc^2 + U$ stays the same?

Zero-point of internal energy and conservation of matter

We must remember that $m = \rho N$, so the content of matter N and its balance should also enter the discussion. In phenomena where matter and antimatter are both present the zero-point of internal energy cannot be redefined arbitrarily. Usually electric charge and electromagnetic field become important in such phenomena, so the general constitutive relation (11.2) may not be appropriate anyway.
The discussion above about the zero-point of internal energy is somewhat imprecise, and meant only to give you an idea of why we can define the zero-point arbitrarily. A rigorous but more complicated analysis would involve the law of conservation of matter, and would show that in appropriate circumstances the zero-point of internal energy mathematically disappears from the balance of matter (Eckart 1940).

When the total mass-energy is required, for example in the Newtonian formula for momentum, we simply approximate it with m. But for physical phenomena where energy *exchanges* are important, we change the "zero" of our energy measurements, shifting it by mc^2, so that energy calculations are numerically more manageable.

› § 10.2 page 170

This is why we shall use the total energy E as defined in formula (11.3):

$$E = U + \frac{1}{2}m\boldsymbol{v}^2 + mgz$$

Note that in this definition we can still arbitrarily choose the zero of U.

This general constitutive relation applies to a large variety of physical phenomena. But it is of little use unless we specify a detailed constitutive relation for the internal energy U. The latter can be wildly different depending on the kind and state of matter, as we shall see later.

The separation between internal and kinetic energy depends on the observation scale

Given a control volume containing matter, at a particular time and in a coordinate system, the amount of *total* energy contained therein is uniquely determined and agreed upon by all researchers, even if one researcher

is making measurements on that volume with coarse instruments, and another researcher is making measurements on a molecular scale.

But the separation of this total energy into internal, kinetic, gravitational does *depend on the detail and scale of observation*. For instance, a researcher who describes the matter within the volume as a continuous gas may measure a total energy $E = 4000\,\mathrm{J}$, and describe this as a sum of $U = 4000\,\mathrm{J}$ internal energy, 0 J gravitational potential energy, and 0 J kinetic energy. Another researcher, who describes the matter within the same volume as a large number of atoms in motion, also measures a total energy $E = 4000\,\mathrm{J}$, but may describe this as a sum of $U = 0\,\mathrm{J}$ internal energy, 0 J gravitational potential energy, and 4000 J kinetic energy, obtained by summing up $\frac{1}{2}m\boldsymbol{v}^2$ for all the atoms in the volume.

The reason of this difference is that the first researcher does not measure any *visible* velocity in the gas: the measuring instruments average it out to zero. This researcher therefore attributes the total energy completely to internal energy, and indeed still detects the atomic motion indirectly, as a *temperature* possessed by the gas. The second researcher, on the other hand, can measure the velocities of the individual atoms, and therefore attributes all the total energy to kinetic energy. For this researcher the gas has no temperature and no internal energy.

Examples of the exchange between internal, kinetic, gravitational potential energies

Of importance in physics and engineering applications are not only changes in the total energy E, but also exchanges among its internal, kinetic, and gravitational terms, even when the total energy is constant. Here are some examples.

Bodies in motion. In the preceding chapters we have considered small bodies of matter such as a tennis ball. In many situations where the motion of a body is involved, its internal energy U is approximately constant, and therefore we only focus on its kinetic $\frac{1}{2}m\boldsymbol{v}^2$ and potential mgz energies. An exchange of energy can happen between them. For instance, in a tennis ball falling in a vacuum, the gravitational energy is decreasing while the kinetic energy is increasing at the same rate: the vertical coordinate of the ball is decreasing, and its downward velocity increasing.

In other situations the internal energy is not constant, and there are exchanges among all three energy components, as well as changes in the

total energy. A ball bouncing on the floor is an example. We can clearly see that its vertical position z, and therefore its gravitational energy mgz, alternately decreases and increases. The same is true for its velocity $\boldsymbol{v}$ and kinetic energy $\frac{1}{2}m\boldsymbol{v}^2$, which are zero at the highest and lowest (bouncing) points. But we also observe that the highest vertical position of the ball gets lower and lower, until the ball rests on the floor, which means that its kinetic and gravitational energies are both zero. They have been converted partly in internal energy U of the ball, and partly transferred, via energy flux, to the internal energy of the floor.

Another example where this kind of energy conversion and energy flux are important is skydiving. The gravitational energy of a skydiver is obviously decreasing. If it were completely converted to kinetic energy, the skydiver would acquire a dangerously excessive falling speed. Instead, this energy is luckily mainly transferred, via friction, to the internal energy of the surrounding air, and partly converted into internal energy of the skydiver. For this reason the kinetic energy, and therefore the falling velocity, of the skydiver remain constant after some time.

Exercise 11.2

A skydiver jumps with zero initial velocity at an altitude of 3000 m. What would be the skydiver's velocity at 1000 m, if there were no changes in internal energy and no energy fluxes?
Compare the velocity you found and the typical velocity of 200 km/h that a skydiver may have at 1000 m, when the parachute is deployed.

Springs and rubber bands. In studying Hookean and non-Hookean springs and rubber bands in Chapter 10, we said that they are usually modelled as objects without mass. They therefore have negligible kinetic and potential energies. All their energy is therefore internal energy.

Gases. Matter in a gaseous state is also often modelled as having negligible mass. All its energy is therefore modelled as internal energy. This is similar to the modelling of objects like springs and rubber bands. An important difference is that the internal energy of gases is typically highly dependent on temperature, whereas the the temperature dependence for springs can be neglected in many applications.

Solids and fluids. In modelling extended matter in a solid and fluid state, usually all three forms of energy must be taken into account. They can moreover be quite variable from one small control volume to another. This is why their modelling and simulation can be extremely complex.

11.3 Constitutive relations for energy flux

Comments about movement of matter at a surface

We shall now discuss constitutive relations for energy fluxes through a control surface. These relations are valid under two requirements:

- matter in contact with the control surface has velocity $\boldsymbol{v}$
- there is no flux of matter through the control surface.

These two requirements sometimes appear as contradictory at first sight: if there's no flux of matter, then how can matter have a non-zero velocity?

Both conditions can actually coexist, and there are two main ways in which they do. Let's discuss them briefly so that you can at least understand them intuitively.

Heat flux and power

Think of a person pushing a door or lifting a weight on the palm of their hand, a hammer hitting a piece of hot iron, or a bike-pump piston quickly pressed down. These and many other physical phenomena can be characterized as follows:

- there's a control surface in contact with matter on both sides
- the control surface may be moving
- there is no flux of matter through the control surface
- electric charges and electromagnetic fields are negligible on a macroscopic scale.

In situations like these there is a general constitutive formula for the flux of energy through the control surface.

Heat flux and power

Consider a control surface, possibly moving, satisfying the conditions above. This control surface must also be such that the velocity $\boldsymbol{v}$ of matter in contact with it, is the same at every point of the surface (this

is true, for instance, if the surface is small enough).

Choose a crossing direction in order to define fluxes. If $\boldsymbol{F}$ is the flux of momentum through this control surface, then the energy flux Φ through the surface is given by

$$\Phi = Q + \boldsymbol{F} \cdot \boldsymbol{v} \tag{11.4}$$

The term Q is called the **heat flux** or *heating*; the term $\boldsymbol{F} \cdot \boldsymbol{v}$ is called the **mechanical power** or *working* transmitted by the force $\boldsymbol{F}$.

If the heat flux vanishes, $Q = 0\,\mathrm{J/s}$, then the energy flux is called **adiabatic**.

We can also consider the integrated flux of energy, that is, the total energy that flows through the surface between times t_0 and t_1:

› § 4.14 page 103

$$\int_{t_0}^{t_1} \Phi(t)\,\mathrm{d}t = \int_{t_0}^{t_1} Q(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{F}(t) \cdot \boldsymbol{v}(t)\,\mathrm{d}t$$

The integrated heat flux is called the **heat** transferred through the surface, and the integrated mechanical power is called the **mechanical work** done by the force

Tricky points in applying the energy-flux formula

In applying the formula above, keep always in mind the following:

- The control surface can be purely imaginary, and there is no real physical separation between the matter on its sides.
- The heat flux could be negative and the mechanical power positive, or vice versa.
- When we deal with closed control surfaces, the conditions for applying the formula above often do **not** hold over the whole surface. For instance, the velocity $\boldsymbol{v}$ or the heat flux Q could be different on different parts the surface.

  To calculate the *total* energy flux in such cases, we must first divide the closed surface into parts for which the formula above can be applied, then add the results.

The heat flux Q can be controlled in some situations, and in that case it

is specified as a boundary condition in a physical problem or simulation. In other situations it is instead given by a constitutive relation. Analogously for the momentum flux $\boldsymbol{F}$: in some situations it's specified as a boundary condition; in others it's given by a constitutive relation, as we saw in many examples from the previous chapter.

› § 6.1 page 149

› § 10.2 page 169

The separation between heat and power depends on the observation scale

We have discussed the fact that the total energy content in a control volume doesn't depend on our manner of description and measuring instruments, but the separation into internal and kinetic energy does. A completely analogous situation, and for analogous reasons, occurs for the flux of total energy through a control surface, and its separation into heat flux and mechanical power.

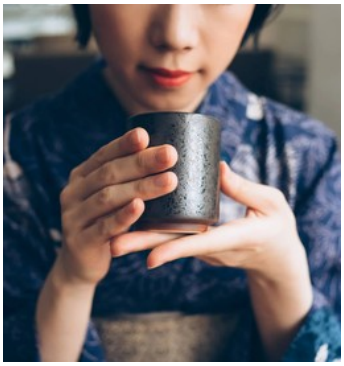
Consider two bodies of matter in contact. For one researcher, who observes these bodies macroscopically, there may be a flux of total energy $\Phi = 1\,\mathrm{J/s}$ through the contact surface, and no visible motion of matter. This researcher therefore consider this energy flux to consist completely in a heat flux $Q = 1\,\mathrm{J/s}$, and no mechanical power $\boldsymbol{F} \cdot \boldsymbol{v}$ since the velocity $\boldsymbol{v}$ is zero. For another researcher, who can keep track of molecular motions and velocities, the total-energy flux is still $\Phi = 1\,\mathrm{J/s}$; but this flux consists completely in $1\,\mathrm{J/s}$ of mechanical power, obtained by summing up $\boldsymbol{F} \cdot \boldsymbol{v}$ for all molecules at the contact surface; the heat flux Q is zero.

Examples of heat and mechanical-power fluxes

Let us see some examples in which energy flux appears as heat flux, flux of mechanical-power, or both.

Holding a cup of hot tea. When we hold a cup of hot tea or coffee, we can feel a flux of energy from the cup to our hands. How much of it is heat flux, and how much is mechanical power?

Let's consider a control surface between our hands and the cup, and let's choose a hands → cup crossing direction. There is obviously a momentum flux from our hands to the cup. The cup has a constant downward supply of momentum from gravity, but it isn't falling. This means that there must be a influx of upward momentum. We can also feel the pressure that the cup exerts on our hands; by the symmetry of flux this means that there



(Image from *Bunka Japan*[14])

must also be a flux of momentum to the cup that points towards its centre. So the total influx of momentum $\boldsymbol{F}$ is not zero. The cup, however, is not moving: its velocity $\boldsymbol{v}$ is zero. The mechanical power $\boldsymbol{F} \cdot \boldsymbol{v}$ is therefore zero as well.

The energy flux through our control surface must therefore be completely in the form of heat flux Q, and the heat **in**flux through our control surface is negative, because energy is flowing from the cup to our hands.

If we move the cup around, then its velocity is no longer zero, and some mechanical power $\boldsymbol{F} \cdot \boldsymbol{v}$ can be transmitted. It shows as an increase in the kinetic energy of the tea, which can even splash out of the cup.

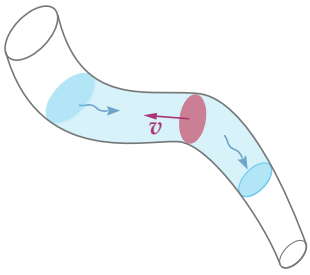
Cooking. When we cook something we create a heat **in**flux into the item being cooked. No mechanical power is transferred, since the velocity of matter is zero.

Spring and body. In studying Hookean and non-Hookean springs and rubber bands, we saw that there is a momentum flux from one end of the spring to the body attached at that end. If the momentum influx for the body is $\boldsymbol{F}$, and the body is moving with velocity $\boldsymbol{v}$, then there is an energy flux into the body, in the form of mechanical power $\boldsymbol{F} \cdot \boldsymbol{v}$. This energy flux increases the body's kinetic energy $\frac{1}{2}m\boldsymbol{v}^2$.

By the symmetry of flux, this same amount of energy flows out of the spring, reducing its internal energy. Any heat flux between the spring and the bodies attached to it is usually negligible.

Gases. In describing and using matter in a gaseous state, usually heat flux and mechanical power must both be taken into account. Many physical theories and technologies are indeed focused on the problem of generating *effluxes of power* as large as possible from a body of matter, by providing *influxes of heat* to it.

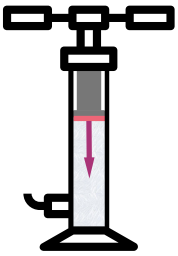
Exercise 11.3



1. Water is flowing downward in a pipe, as illustrated in the side picture by the blue squiggly arrows. Take a control surface moving with velocity $\boldsymbol{v}$ as depicted in red. Can we apply formula (11.4) for the energy flux through this control surface?
2. Suppose that there is no power $\boldsymbol{F} \cdot \boldsymbol{v}$ transmitted through a moving control surface. Does this mean that the momentum flux $\boldsymbol{F}$ is zero?

Or does this mean that the matter's velocity $\boldsymbol{v}$ is zero?

3. A control surface, with a given crossing direction, is moving with velocity $[0, 2, -3]\,\mathrm{m/s}$. Through this surface we have a heat flux of $-3\,\mathrm{J/s}$ and a momentum flux of $[1, -2, 1]\,\mathrm{N}$. How much is the energy flux through the surface?
4. Consider the control surface that separates air from the piston within a bike pump, schematized by the red line in the side illustration. The piston and the gas in contact with it, on the two sides of the control surface, are moving downward with a velocity $[0, 0, -0.5]\,\mathrm{m/s}$. The downward flux of momentum is $[0, 0, -20]\,\mathrm{N}$. The energy flux through this surface is adiabatic. How much is the downward flux of energy through this surface?
5. Consider again the control surface of question 3. above. We are now told that there is no matter flux through this surface, and the matter in contact with the surface on both sides is moving with the same velocity as the surface. How much is the energy flux through the surface?
6. Imagine a cylindrical *closed* control surface enveloping the air within the bike pump of question 4. above. The surface previously considered is part of this closed control surface. We are told that across the rest of the closed surface there is a total heat **ef**flux of $2\,\mathrm{J/s}$. How much is the total energy **in**flux through the closed control surface?



Summary: energy constitutive relations for matter

So far we have discussed two general constitutive relations: one for energy content E, formula (11.3); and one for energy flux Φ, formula (11.4). We also said that our definition of total energy approximately satisfies a conservation law, so the energy supply R is zero. Let us rewrite here these constitutive relations, explicitly indicating their time dependence:

› § 11.1 page 215

$$
\begin{aligned}
E(t) &= U(t) + \tfrac{1}{2} m \boldsymbol{v}(t)^2 + mgz(t) \\
\Phi(t) &= Q(t) + \boldsymbol{F}(t) \cdot \boldsymbol{v}(t) \\
R(t) &= 0
\end{aligned}
$$

Let us recall the physical conditions under which they are valid:

- matter present within and directly outside the control volume

- control volume and surface such that the matter velocity $\boldsymbol{v}$ and molar mass ρ, and therefore also m, are the same throughout
- no matter flux through the control surface
- no electric charges or electromagnetic fields
- Newtonian approximation, and close to the Earth's surface

Division of control volume or surface may be necessary

In some situations we must divide a control volume or closed control surface into several parts where these conditions are valid.

Under these conditions, the formulae above can be used together with the balance of energy (11.1) in order to describe many common physical situations, with all sorts of materials, and to make predictions, for instance by numerical simulations. Note how these constitutive formulae connect the fluxes and volume contents of energy, matter, momentum; and they even involve two new auxiliary quantities: the internal energy U and the heat flux Q.

› § 6.1 page 144

In order to concretely use these constitutive formulae, we need first more specific constitutive relations for the internal energy U, the heat flux Q, and the momentum flux $\boldsymbol{F}$. The variety of constitutive relations for these three quantities is the subject of materials science[15], and could fill numerous tomes. Some constitutive relations for these quantities are mathematically extremely complex, involving integrals and partial derivatives; this is no surprise, because they reflect the complexity of the materials they model.

We shall now focus on a set of simple constitutive relations for $\boldsymbol{F}, U, Q$ which can be used in a more restricted but still quite broad range of physical situations, involving gases.

What if there's matter flux? Transport terms

We have repeated many times that the constitutive relation for the energy flux is through a control surface

$$\Phi = Q + \boldsymbol{F} \cdot \boldsymbol{v}$$

is only valid when there isn't any flux of matter through that surface. What if there *is* a flux of matter? how does that constitutive relation change?
There is indeed a more general formula that can be used even when there is a flux of matter through the control surface. To understand it, however, we must:

- give the formulae for the fluxes of matter, momentum, energy at the same time
- take into account the contents of matter, momentum, energy in a thin control volume connected to the surface

- take into account the velocity of the control surface itself.

So the description is somewhat more complicated; you see why we choose to avoid this more general situation in the notes. But here are the fully general formulae, in case you're curious.

Choose a crossing direction for the surface, and

- $\boldsymbol{v}_\mathrm{s}$ is the velocity of the surface
- $\boldsymbol{n}_\mathrm{s}$ is a *unit* vector orthogonal to the surface, having the same direction as the crossing direction
- A is the area of the surface
- V is the volume of a thin control volume in contact with the surface
- N, $\boldsymbol{P}$, E are the matter, momentum, energy contents in that thin control volume
- $\boldsymbol{v}$ is the velocity of matter in contact with the surface

Then the fluxes of matter, momentum, energy are given by the following general formulae, if no electromagnetic fields or electric charges are present and in Newtonian approximation:

$$J = \frac{A}{V} N \boldsymbol{n}_\mathrm{s} \cdot (\boldsymbol{v} - \boldsymbol{v}_\mathrm{s})$$

$$\boldsymbol{F} = A \boldsymbol{n}_\mathrm{s} \cdot \boldsymbol{\sigma} + \frac{A}{V} \boldsymbol{P} \boldsymbol{n}_\mathrm{s} \cdot (\boldsymbol{v} - \boldsymbol{v}_\mathrm{s})$$

$$\Phi = Q + A \boldsymbol{n}_\mathrm{s} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{v} + \frac{A}{V} E \boldsymbol{n}_\mathrm{s} \cdot (\boldsymbol{v} - \boldsymbol{v}_\mathrm{s})$$

In these formulae, $\boldsymbol{\sigma}$ is a 3-by-3 matrix called the *stress tensor* or *pressure tensor*. The terms containing the expression $\boldsymbol{n}_\mathrm{s} \cdot (\boldsymbol{v} - \boldsymbol{v}_\mathrm{s})$ are called *transport terms*, because they originate from the transport of matter through the surface.

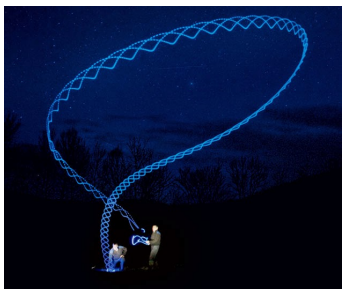
If the flux of matter is zero, $J = 0$, then the expression $\boldsymbol{n}_\mathrm{s} \cdot (\boldsymbol{v} - \boldsymbol{v}_\mathrm{s})$ must be zero too, and the analogous expressions in the fluxes of momentum and energy are zero as well; all fluxes become much simpler.

11.4 Rigid bodies

Rigid bodies have very peculiar properties in respect of the balance of energy. A **rigid body** is a body (of matter) that has constant volume and *shape*; or in other words a body that cannot deform. It can of course move, and in general its amount of matter, mass-energy, and velocity can be different in different parts of the body; think of a boomerang for instance.

The constitutive equations for the energy of a rigid body take on a peculiar form, because of its rigidity:

- **the heat flux contributes only to changes in internal energy**
- **the internal energy can only change through heat fluxes**
- **the mechanical power contributes only to changes in kinetic & potential gravitational energy**



Long exposure of the trajectory of a boomerang outfitted with LEDs[16]

- **kinetic & potential gravitational energy can only change through mechanical power**

Let's express these conditions mathematically. Consider a control volume containing all or part of a rigid body, and over which the velocity $\boldsymbol{v}$ is the same in different sub-volumes, even if it can change with time. The mass and internal energy contained in the control volume are m and U, the total momentum influx and heat flux across the closed control surface are $\boldsymbol{F}$ and Q. Use a vertical coordinate z. Then, in integral form:

$$U(t_1) = U(t_0) + \int_{t_0}^{t_1} Q(t)\,\mathrm{d}t \tag{11.5a}$$

$$\tfrac{1}{2}m\boldsymbol{v}(t_1) + mgz(t_1) = \tfrac{1}{2}m\boldsymbol{v}(t_0) + mgz(t_0) + \int_{t_0}^{t_1} \boldsymbol{F}(t)\cdot\boldsymbol{v}(t)\,\mathrm{d}t \tag{11.5b}$$

or equivalently in differential form:

$$\frac{\mathrm{d}U(t)}{\mathrm{d}t} = Q(t) \tag{11.6a}$$

$$\frac{\mathrm{d}\left[\tfrac{1}{2}m\boldsymbol{v}(t)^2 + mgz(t)\right]}{\mathrm{d}t} = \boldsymbol{F}(t)\cdot\boldsymbol{v}(t) \tag{11.6b}$$

We can say that mechanical and thermal phenomena gets completely separated in a rigid body. This separation does *not* occur for ordinary deformable bodies: in general, mechanical power can produce changes in internal energy.

Moreover, it turns out that **the equations relating kinetic & gravitational potential energy can be derived from the balance of momentum**. In other words, formulae (11.5b) and (11.6b) above can be derived from the balance of momentum applied to the same control volume: they do not represent some additional physical law.

The special constitutive property of rigid bodies has an important consequence. *If we are not interested in the changes of the internal energy of a rigid body, then we do not need to consider the balance of energy*. The balance of momentum is all we need.

This fact explain why we were able to solve many problems of motion in the previous chapters: many bodies we considered, such as tennis balls or blocks of materials, were practically rigid, and we were not investigating their internal energies. The balance of momentum was therefore sufficient to describe their motion. Springs and rubber bands are *not* rigid, but for

them we implicitly assumed that their internal energy was constant, and no heat fluxes occurred.

Rigidity is obviously only an approximation. In fact, General Relativity makes this notion strictly speaking impossible. But in situations where Newtonian approximations apply, the approximation of rigidity makes many physical phenomena much easier to describe.

11.5 Constitutive relations for ideal gases

There is a set of simple constitutive relations which can be used as approximations for many gases, especially when they are rarefied, that is, when their amount of matter per volume N/V is low. We say that these constitutive relations apply to **ideal gases**. When we say that some material can be modelled as an ideal gas – or simply say "it's an ideal gas", we mean that we can model it with good enough approximation using these relations.

The constitutive relations for ideal gases involve constants and a couple of functions that can be different depending on the kind of gas. So there isn't just one ideal gas, but a family of them. Some texts speak of *the* ideal gas, as if there was only one; but they do so because they discuss properties that are common to all ideal gases.

In discussing the constitutive relations below, we consider a closed control surface that encloses an amount of some ideal gas. We assume that there is no flux of the gas through the enclosing surface, and more generally that the conditions listed in the previous 11.3summary are satisfied. These conditions are approximately true for many practical cases, like air under compression or expansion in a bike pump.

If we want to model a body of gas accurately, however, we generally need to use a large number of such control volumes, each one enough small that the conditions above are approximately satisfied within it. This way we can try to simulate many interesting physical behaviours that ideal gases can have, such as turbulence.

Pressure (momentum flux) of an ideal gas

We previously discussed the peculiar fluxes of momentum that take place in a gas, which we generally call the internal pressure of the gas. Now we discuss a constitutive relation that connects:

› § 10.3 page 190

- the flux of momentum $\boldsymbol{F}$ through the closed control surface containing an ideal gas
- the amount of ideal gas within
- the temperature within
- the volume enclosed by the control surface

This constitutive relation is more easily expressed in terms of the momentum flux through a surface, *divided by the area* of the surface:

Pressure vector, stress vector, pressure

For a control surface of area A where the momentum flux $\boldsymbol{F}$ is uniform throughout, and through which *there is no matter flux*, we call **pressure vector** or **stress vector** $\boldsymbol{p}$ the momentum flux divided by the area: $\boldsymbol{p} \coloneqq \boldsymbol{F}/A$.

If the momentum flux $\boldsymbol{F}$, and therefore the pressure vector, are orthogonal to the surface and point *away from the surface*, then we call **pressure** the magnitude of the pressure vector, and usually denote it p:

$$p \coloneqq |\boldsymbol{p}| \equiv \frac{|\boldsymbol{F}|}{A} \tag{11.7}$$

The name *tension* is used instead of *pressure* if the momentum flux points towards the surface. Compare our discussion about compressive and tensile momentum fluxes.

› § 4.12 page 99

Pressure has physical dimensions of force per area, with units $\mathrm{N/m^2}$; this unit also called *pascal* (Pa).

'Pressure' has many different meanings

Be aware that the term 'pressure' is used in many different and sometimes even opposite ways in the physics literature. Some texts use 'pressure' to indicate not just the magnitude of $\boldsymbol{F}/A$, but the full vector; so for them 'pressure' is not just a number, but a set of three vector components. Some texts use 'pressure' to denote the pressure vector; so for them a pressure need not be orthogonal to a control surface. Some texts use 'pressure' also to indicate 'tension', or vice versa.
So when you read a text that uses or discusses 'pressure', make sure to get correctly in which sense the text is using this word.

In these notes we shall sometimes use 'pressure' in a slightly different or more general meaning, which should be clear from the context.

Now that the notion of pressure is introduced, we can formulate a particular constitutive relation for the momentum flux:

Ideal-gas law with viscosity

If a closed control surface encloses a volume V with an amount N of an ideal gas, at uniform temperature T, and the momentum flux is orthogonal to the surface and uniform throughout, and there is no matter flux through the surface, then the pressure p is given by

$$p = \frac{RNT}{V} - \mu \frac{1}{V} \frac{\mathrm{d}V}{\mathrm{d}t} \tag{11.8}$$

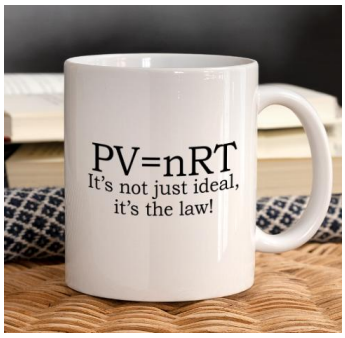
where $\frac{\mathrm{d}V}{\mathrm{d}t}$ is the rate of change of volume, μ is a **viscosity** coefficient, and R is the molar gas constant[17], having universal value

$$R = 8.314\,462\,618\,153\,24\ \mathrm{J/(mol \cdot K)} \quad \text{(exactly)}$$

The equation above is called the **ideal-gas law with viscosity**.

If the surface has area A, the magnitude of the momentum flux is therefore $|\boldsymbol{F}| = pA$.

Note that the formula above for pressure and momentum flux is only valid for surfaces through which no matter flux occurs.



In the expression above for the pressure you may recognize the famous "$pV = NRT$" formula, called the **ideal-gas law**. This famous formula is strictly speaking only valid when the ideal gas is at rest, which means that its volume is not changing, so that $\frac{\mathrm{d}V}{\mathrm{d}t} = 0$.

If the ideal gas is in motion, for instance expanding or contracting and its volume V changes with time, then additional terms must be added to the famous "$pV = NRT$" formula. This is what we see in the relation (11.8) above. In many cases these additional terms are extremely small and therefore neglected.

The fact that the viscosity coefficients must be positive is a consequence of the *balance of entropy* – the second law of thermodynamics – which we shall discuss later.

Internal energy of an ideal gas

Next we discuss a constitutive relation that connects the amount of internal energy in the control volume with the amount of ideal gas and its temperature:

Internal energy of an ideal gas

If a small control volume contains an amount N of ideal gas at uniform temperature T, then it also contains an amount of internal energy U given by

$$U = C\,N\,T \tag{11.9}$$

where the constant C depends on the kind of ideal gas and is called *molar heat capacity*.

This formula is very important: for an ideal gas, the absolute temperature is a direct measure of the internal energy, so it can often be used as a proxy for the latter.

Validity of the internal-energy formula

- Keep in mind that the particular constitutive relation (11.9) is valid for an ideal gas only. The internal energy of a generic material is related to other quantities besides amount of matter and temperature.
- Some books express the formula above saying "the internal energy of an ideal gas depends only on temperature". This statement is somewhat vague and easy to misunderstand. First, the internal energy clearly "depends" also on the amount of matter N. Second, if in some application we express the amount of matter or the temperature as functions of other quantities, such as volume, then the internal energy also becomes a function of those quantities. The formula above can be simply expressed as follow: if we know the amount (and kind) of matter in a volume, and the temperature in that volume, then we also know the amount of internal energy therein.

Heat flux between sides at different temperatures

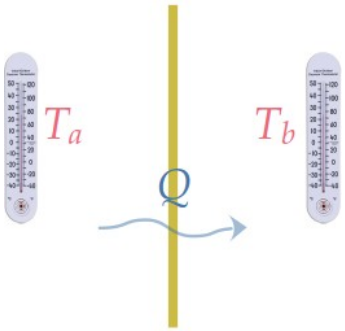
Lastly we discuss a constitutive relation that connects heat flux and temperature. This relation applies to many physical phenomena and materials: not only to ideal gases, but to many other fluids and solids as well.

Newton's law of cooling

Consider a control surface of area A and call its two sides a and b. If the matter on side a has approximately uniform temperature T_a, and the matter on side b has approximately uniform temperature T_b, then the heat flux Q in the $a \to b$ crossing direction is given by **Newton's 'law of cooling'**:

$$Q = Ah\,(T_a - T_b) \tag{11.10}$$

where h is called the *coefficient of heat transfer* and depends on the particular physical conditions of the matter on the two sides of the surface. This coefficient is usually positive, and may depend on the temperature.



The constitutive formula above with $h > 0$ says that if the temperature on side a is larger than the one on side b, then a positive heat flux occurs from a to b. In other words, positive heat flows from the hotter to the colder side of the surface.

Newton's law of cooling implies that we can approximately consider temperature as having a jump or discontinuity between the two sides of the surface. In situations where this approximation is too gross, another constitutive equation is often used: **Fourier's law of heat conduction**, which we can write as

$$Q = -Ak\frac{\partial T}{\partial x}$$

In this expression we imagine the surface to be orthogonal to the yz directions, and the crossing direction to be the positive x-direction. The derivative $\frac{\partial T}{\partial x}$ is the gradient of the temperature, expressing how much the temperature changes from one point to another very close one.

Let's make clear that Newton's law of cooling and Fourier's law *are not universal*. There are physical situations and materials for which the heat flux is connected to temperature in more complex ways; and not only to temperature, but also to momentum flux, matter flux, electromagnetic

quantities. Thermoelectric coolers[18] are an example application of these more complex constitutive relations for the heat flux.

Heat can flow from cold to hot

Some texts say that "heat cannot flow from cold to hot", and present this vague statement as a consequence of, or equivalent to, the second law of thermodynamics. This statement is actually *not* true, from several points of view.
Heat can flow from cold to hot. An everyday example is a refrigerator:

> It is obvious that, on a macroscopic scale, heat flows from a cold source to a hotter sink in a refrigerator: thus the idea that the second law requires heat to invariably flow from hot to cold is belied by the fact that refrigerators do work. (Astarita 1990)

Even more interesting examples, in which Fourier's law does not apply, occur for materials such as polymers add text and reference

Other common assumptions about ideal gases

Besides the particular constitutive relations for momentum flux, internal energy, and heat flux just discussed, it is typically assumed that a not-too-large volume of ideal gas has a negligible mass. This is usually a reasonable assumption.

Exercise 11.4

Calculate the mass of a litre, that is 10^{-3} m^3, of air. Use the following information:

- the molar mass ρ of air is around 0.03 kg/mol; this means that an amount N of air has mass $m \approx \rho N$ › § 7.2 page 160
- the pressure p of air is around 10^5 N/m^2
- take a thermodynamic temperature of air $T = 300\,\mathrm{K}$ (around 27 °C)
- pressure, volume, temperature, and amount of air are related by the ideal-gas law (11.8). › § 11.5 page 232

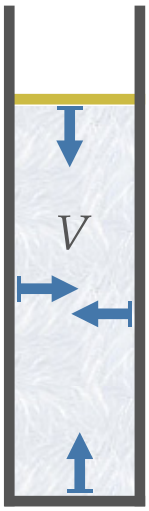
If the mass m of a volume of ideal gas is assumed to be zero, then four other quantities are zero as well in that volume, at all times:

- the total momentum content $\boldsymbol{P} = m\boldsymbol{v}$
- the momentum supply from gravity $\boldsymbol{G} = m\boldsymbol{g}$

- the kinetic energy $\frac{1}{2}m\boldsymbol{v}^2$
- the gravitational potential energy mgz

This assumption has an important consequence for the balance of momentum of a control volume containing an ideal gas: we have

$$\underset{=0}{\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t}} = \boldsymbol{F}(t) + \underset{=0}{\boldsymbol{G}(t)} \qquad \Longrightarrow \qquad \boldsymbol{F}(t) = 0$$



that is, the *total* momentum influx is always zero. But note that *this does not mean that the momentum influx is zero everywhere* across the closed control surface; it only means that the momentum influxes of opposite parts of the surface cancel out perfectly, as schematized in the side figure.

Remember that when we use the ideal-gas law (11.8), we are assuming that **any small surface of area A of the control volume has a momentum influx $\boldsymbol{F}_A$ of magnitude**

$$F_A = A\,p = A\,\frac{RNT}{V} - A\,\mu\frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}t}$$

and directed *inwards*, according to the ideal-gas law (11.8).

› § 11.5 page 232

Another important consequence of assuming that a volume of ideal gas has zero mass is that **its total energy is purely internal energy**:

$$E = \underset{=CNT}{U} + \underset{=0}{\tfrac{1}{2}m\boldsymbol{v}^2} + \underset{=0}{mgz} = CNT$$

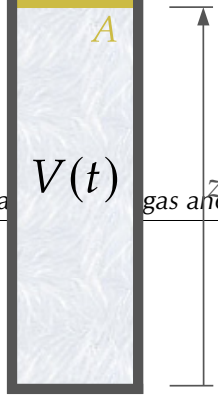
which also means that any energy influx or efflux only changes the internal energy of the gas.

The masslessness assumption is of course inadequate in some conditions like extremely fast compressions or expansions. In such conditions also the ideal-gas law (11.8) must be modified.

11.6 Example applications: ideal gas and piston

Setup

One of the simplest system involving an ideal gas consists of a chamber of variable volume, wherein an amount of ideal gas is enclosed, as illustrated in the side picture. To be more specific we consider a vertical tubular chamber containing an amount N of an ideal gas. The dark grey walls are rigid, the yellow piston can move vertically and has constant mass



m and constant surface area A. The base of the piston is at a height $z(t)$ from the bottom of the chamber. The height and therefore the volume $V(t) = A\,z(t)$ of the chamber can vary with time. We choose a one-dimensional coordinate system z; vectors are positive when directed upward.

Since the piston has mass m, it also has momentum $\boldsymbol{P} = m\,v$, according to the Newtonian constitutive relation for momentum, and it may have a vertical gravity momentum supply $\boldsymbol{G} = -m\,g$.

You may wonder why we need to consider a piston having mass. There are two main reasons. First, in a real situation a piston does have a mass that is non-negligible, compared with the mass of the gas. Second, the assumption that the gas has no mass makes the balance of momentum become singular, as discussed in the previous section, and any description and prediction of motion becomes therefore impossible – unless we add mass somewhere else. The piston is effectively also a proxy for the mass of the gas. We encountered an analogous situation with the spring-and-bodies system.

› § 11.5 page 235

Control volumes & surfaces

As usual we must choose a set of control volumes and surfaces to describe the physical system, so that we can define the contents, fluxes, and supplies of any relevant balance laws.

A natural choice is a control volume tightly wrapping the piston, and another control volume coinciding with the chamber's walls, containing the gas. The two control volumes have one horizontal surface in common, of area A, where the piston is in contact with the gas. Therefore there will be fluxes of quantities between the two control volumes. By the symmetry of flux, the flux X of any quantity from one volume to the other through this surface, corresponds to a flux $-X$ from the second volume to the first.

› § 4.6 page 87

The control volume of the piston is rigid but movable; its position is determined by the coordinate $z(t)$. The volume $V(t)$ of the chamber can instead change with time, and is related to $z(t)$ by

$$V(t) = A\,z(t) \tag{11.11}$$

Although this is a geometric relation, its deeper origin actually lies in the law of conservation of matter, because the shapes and positions of these control volumes are chosen so that matter is automatically conserved.

The vertical velocity $v(t)$ of the piston is given by

$$v(t) = \frac{\mathrm{d}z(t)}{\mathrm{d}t}$$

and is positive if the piston is moving upward. We also define the rate of change of the volume $V(t)$, which turns out to be related to the velocity $v(t)$:

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \frac{\mathrm{d}A\, z(t)}{\mathrm{d}t} = A\, v(t) \tag{11.12}$$

Let's now see how the main balances apply to the chosen control volumes of this physical system.

Conservation of matter

As mentioned in the previous section, the control volumes have been chosen so as to automatically satisfy this conservation law. But this law still appears very subtly through the geometric relations (11.11) and (11.12). We therefore don't need to worry about its balance, as long as we use those geometric relations.

The constant amount of matter in the piston is such that its mass is m, and the constant amount of ideal gas in the chamber is N.

Balances of momentum

Piston. Let's first apply this balance to the control volume of the piston.

The momentum content of the piston is $P(t) = mv(t)$, positive when directed upward.

The *total* momentum influx $F(t)$ is the sum of the fluxes through three main surfaces:

Side: There usually are shear forces between the side of the piston and the walls of the chamber, but they are sometimes made negligible, for instance by the use of lubricants. Here we assume that such forces are zero.

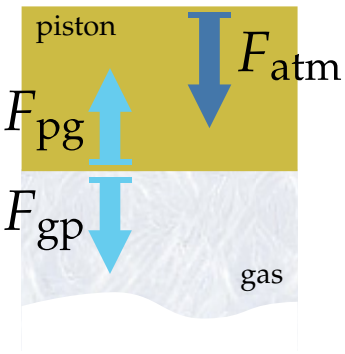
› § 4.12 page 100

Top: If there's atmosphere above the piston, it exerts a downward force F_{atm} approximately constant and equal to

$$F_{\text{atm}} = -A \cdot 10^5\ \mathrm{N/m^2} \tag{11.13}$$

Bottom: The momentum flux through this surface is particularly important, because it's momentum exchanged between the piston and the gas. Denote by $F_{pg}(t)$ the flux of momentum from the gas to the piston, and $F_{gp}(t)$ the flux of momentum from the piston into the gas. Again by the symmetry of flux we have

$$F_{pg}(t) = -F_{gp}(t)$$



The momentum influx $F_{gp}(t)$ for the gas, being a pressure, always points inwards; in our coordinates it is always negative, downward. The momentum influx $F_{pg}(t)$ for the piston is therefore always positive, upward.

For the moment we don't have any constitutive relation for the influx $F_{pg}(t)$, but we shall have one when we examine the control volume of the gas.

The total influx of momentum, or surface force, for the piston is therefore

$$F(t) = F_{atm} + F_{pg}(t) \tag{11.14}$$

The gravitational supply of momentum to the piston must be taken into account in our case, because the piston can move vertically. It would have been neglected if the piston had moved horizontally instead. It is given by

$$G = -m\,g \tag{11.15}$$

The equations governing the motion of the piston are therefore

$$\begin{gathered} \frac{\mathrm{d}P(t)}{\mathrm{d}t} = F(t) + G(t) \\ P(t) = mv(t) \qquad F(t) = F_{atm} + F_{pg}(t) \qquad G = -m\,g \\ F_{atm} = -A \cdot 10^5\ \mathrm{N/m^2} \end{gathered} \tag{11.16}$$

together with the relation between momentum influxes for piston and gas at the common surface:

$$F_{pg}(t) = -F_{gp}(t) \tag{11.17}$$

Ideal gas. We are considering the ideal gas to be practically massless; therefore its momentum content and gravitational supply are zero.

We saw that when the gas is considered massless, the balance of momentum just becomes a requirement that the *total* flux of momentum be zero as well. Also recall that the momentum flowing through a small surface, being a pressure, always has an inward orientation, orthogonal to the surface.

› § 11.5 page 235

The total momentum influx for the gas takes places through three main surfaces:

Side: The influxes on opposite parts of the side surface are equal in magnitude but opposite in orientation, and cancel each other out. The total force through the side surface is therefore zero, and we don't need to keep track of it.

Top: The gas is in contact with the piston at the top surface. The influx of momentum through the top, F_{gp}, is therefore equal to minus the momentum flux F_{pg} from the gas to the piston. We wrote this in formula (11.17):

$$F_{\mathrm{pg}}(t) = -F_{\mathrm{gp}}(t)$$

We have a constitutive relation for the pressure corresponding to this influx: the ideal-gas law:

› § 11.5 page 232

$$F_{\mathrm{gp}}(t) = -A\,p(t) \quad \text{with} \quad p(t) = \frac{RNT(t)}{V(t)} - \mu\frac{1}{V(t)}\frac{\mathrm{d}V(t)}{\mathrm{d}t}$$

The first equation has a minus sign because the influx always points inward, therefore downward in our case. Note how this constitutive relation brings into play the volume $V(t)$ and the thermodynamic temperature $T(t)$ of the ideal gas, both of which can change with time.

Bottom: It is usually assumed that the chamber containing the ideal gas rests on some support or on the ground. Through the corresponding surface there is therefore an efflux of downward-pointing momentum, as it happened in our analysis of the books on a table. The momentum influx at the bottom is opposite to the influx at the top surface. We therefore don't need to keep track of the momentum flux at the bottom.

› § 10.3 page 179

The only equations relevant to the momentum of the gas are therefore

$$F_{\mathrm{gp}}(t) = -A\,p(t) \qquad p(t) = \frac{RNT(t)}{V(t)} - \mu\frac{1}{V(t)}\frac{\mathrm{d}V(t)}{\mathrm{d}t} \tag{11.18}$$

together with the relation between momentum influxes for piston and gas at the common surface, equation (11.17).

Balances of energy

Piston. Let's first consider the control volume of the piston.

We are treating the piston as a *rigid* body. It motion is therefore completely determined by the balance of momentum. We would need to consider the balance of energy for the piston, if we were interested in how its internal energy changes. But in the present problem this is of no interest.

› § 11.4 page 228

You may object: "but we expect the gas to exchange heat, possibly with the piston as well". This is a fully valid and intelligent objection! The gas is indeed likely to exchange energy in the form of heat with the piston, unless we insulate the surface between them and make the energy flux through it adiabatic. We are silently making one or both of these assumptions:

› § 11.3 page 222

- the area A between piston and gas is small compared to the rest of the gas surface, and can therefore be neglected
- from the point of view of heat exchange (only), the piston can be considered just like the rigid walls and not treated in a special way.

Ideal gas. The energy content $E(t)$ of the control volume containing the ideal gas 11.5amounts only to its internal energy $U(t)$ – no kinetic or potential energies – owing to the zero-mass assumption. And we do have a constitutive relation for internal energy, which connects the latter to the amount N of gas and to its thermodynamic temperature $T(t)$. The energy content is therefore given by

› § 11.5 page 233

$$E(t) = U(t) = CNT(t) \tag{11.19}$$

Keep in mind that this constitutive relation can only be used when the temperature and the amount of matter per volume are approximately the same in every sub-volume.

Recall that the energy influx Φ through a surface of area a that satisfies the no-matter-flux and uniform-velocity conditions is given by the general constitutive equation for energy flux, together with Newton's law of cooling:

› § 11.3 page 232

$$\Phi(t) = Q(t) + \boldsymbol{F}(t) \cdot \boldsymbol{v}(t) \quad \text{with} \quad Q(t) = a\, h\, [T_{(ext)} - T(t)]$$

We are calling $T_{(ext)}$ the temperature of the walls and piston enclosing the ideal gas, and assuming it to be constant.

The *total* energy influx $\Phi(t)$ is the sum of fluxes through two main surfaces:

Side: There could be a heat flux through the side walls. Note that as the piston moves up or down, the area of the side walls changes, and this must be taken into account in Newton's law of cooling. For simplicity let's assume that the energy flux through the side walls is adiabatic.

The momentum flux $\boldsymbol{F}_{\text{side}}$ through the side walls is orthogonal to them, according to the ideal-gas law, and therefore horizontal. As the piston moves up or down, we assume that the velocity $\boldsymbol{v}$ of the gas is approximately always vertical instead. This means that, on the side walls, the mechanical power $\boldsymbol{F}_{\text{side}} \cdot \boldsymbol{v}$ is approximately zero, because $\boldsymbol{F}_{\text{side}}$ and $\boldsymbol{v}$ are orthogonal there.

We have therefore $\Phi_{\text{side}}(t) = 0$.

Bottom: Let's consider the possibility of a heat flux Q_{bot} through the bottom surface. Since the area A of this surface is constant, this flux is

$$Q_{\text{bot}}(t) = A\, h\, [T_{(ext)} - T(t)]$$

The momentum influx $\boldsymbol{F}_{\text{bottom}}$ through the bottom wall is directed upward. The velocity of the gas is approximately assumed to be vertical, and we also assume that it is zero at the bottom surface, because the bottom surface is not moving. Otherwise it would mean either that a vacuum is forming right above the surface, or that some gas is passing through it. Neither of these possibilities are contemplated in the present case; they would require different constitutive relations. If the gas velocity at this surface is zero, then the mechanical power is also zero.

The energy flux through the bottom surface is therefore

$$\Phi_{\text{bot}} = Q_{\text{bot}}(t) + 0$$

Top: We could consider a heat flux through the top surface, between gas and piston. Its formula would be similar to the one for the bottom surface with same area A but possibly different external temperature (the temperature of the piston). For simplicity let's assume that this heat flux is zero.

The momentum influx $F_{\text{gp}}(t)$ through the top surface is directed downward, and has magnitude given by the ideal-gas law. The velocity of the ideal gas at this surface must be equal to that of the surface itself, $v(t)$, otherwise there would either be a vacuum or gas would pass through the piston. At the top surface there is therefore a non-zero mechanical power $F_{\text{gp}} \cdot \boldsymbol{v}$. The energy flux across this surface is therefore

$$\Phi_{\text{top}} = 0 + F_{\text{gp}}(t) \cdot v(t)$$

where $F_{\text{gp}}(t)$ is given by formula (11.18).

The total energy influx into the control volume enclosing the ideal gas is therefore

$$\Phi(t) = Q_{\text{bot}}(t) + F_{\text{gp}}(t)\, v(t)$$

with

$$Q_{\text{bot}}(t) = A\, h\, [T_{(ext)} - T(t)] \tag{11.20}$$

$$F_{\text{gp}}(t) = -A\left[\frac{RNT(t)}{V(t)} - \mu\frac{1}{V(t)}\frac{\mathrm{d}V(t)}{\mathrm{d}t}\right]$$

We finally have all the equations to describe and numerically time-integrate our physical system.

Exercise 11.5

1. Use the basic script-writing strategy to write a script that simulates the system of ideal gas & piston. › § 10.5 page 197
2. What is the state of the system? › § 10.5 page 201

3. Simulate the system with the following numerical constants and initial conditions:

$$m = 10\,\mathrm{kg} \qquad N = 0.04\,\mathrm{mol} \qquad A = 0.01\,\mathrm{m}^2 \qquad T_{(ext)} = 296.15\,\mathrm{K}$$
$$C = 20\,\mathrm{J/(mol \cdot K)} \quad \mu = 0.000\,04\,\mathrm{N\,s/m^2} \quad h = 8000\,\mathrm{J/(K\,s\,m^2)}$$
$$g = 9.8\,\mathrm{N/kg} \qquad R = 8.314\,\mathrm{J/(mol\,K)}$$
$$t_0 = 0\,\mathrm{s} \qquad t_1 = 1\,\mathrm{s} \qquad \Delta t = 0.0001\,\mathrm{s}$$
$$z(t_0) = 0.1\,\mathrm{m} \qquad v(t_0) = 0\,\mathrm{m/s} \qquad T(t_0) = 296.15\,\mathrm{m/s}$$

   Plot the position $z(t)$ of the piston and the temperature $T(t)$ of the ideal gas as functions of time. What do you observe?

4. Simulate again with the same values as before but a coefficient of heat transfer $h = 0\,\mathrm{J/(K\,s\,m^2)}$, that is, assuming that all energy fluxes are adiabatic. Plot again position and temperature against time. What do you observe?

5. Keeping all energy fluxes adiabatic, simulate now by setting the viscosity coefficient $\mu = 400\,\mathrm{N\,s/m^2}$. What do you observe?

6. Play and simulate with other values!

Here is an example of Octave/MATLAB script for simulating the system of ideal gas and piston and plotting $z(t)$ and $T(t)$:

Download idealgas_piston.m[19]

```
%%% idealgas_piston.m
%% Simulation of ideal gas & mass piston in 1D
%% SI units used throughout
%% Coordinate z
%%%% Constants
m = 10; % mass of piston
N = 0.04; % amount of ideal gas
A = 0.1^2; % area of piston
g = 9.8; % gravitational acceleration
R = 8.31446261815; % molar gas constant
C = 20; % molar heat capacity
mu = 0.00004; % gas viscosity
h = 8000; % heat-transfer coefficient
Te = 273.15 + 23; % temperature of environment
Fatm = -100000*A; % force on piston by atmosphere
%%
G = -m*g; % gravity supply of momentum to piston
%%
t1 = 1; % final time
```

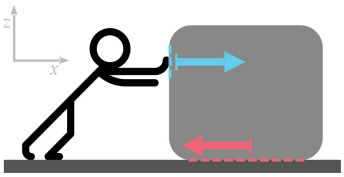
```
dt = 0.0001; % time step
%%%% STATE: z, v, T; initial conditions
t = 0; % initial time
z = 0.15; % initial position of piston
v = 0; % initial velocity of piston
T = 273.15 + 23; % initial temperature of gas
%%
%% Plot & saving
%% adjust final time if not multiple of timestep
t1 = t1 + mod(t1-t,dt);
%% Save values of all quantities at some steps during the simulation,
%% for subsequente analysis or plotting
%% (saving at all timesteps could be too costly)
Nsaves = 200; % number of timepoints to save during the simulation
%% Calculate time interval for saving
dsave = (t1-t)/(Nsaves-1);
if abs(dsave) < (dt)
  error('time interval between saves is smaller than timestep')
end
%% Initialize vectors to contain saved values
tSave = nan(Nsaves,1);
zSave = nan(Nsaves,1);
vSave = nan(Nsaves,1);
TSave = nan(Nsaves,1);
%% Save initial values
i = 1; % index that keeps count of savepoints
t0 = t;
tSave(1) = t;
zSave(1) = z;
vSave(1) = v;
TSave(1) = T;
%% Initialize plot
close all;
subplot(2,1,1)
cols = get(0, 'DefaultAxesColorOrder');
plot(tSave(1), zSave(1), 'o','Color',cols(1,:)); axis('tight');
xlabel('time {\it t}/s'); ylabel('position {\it z}/m'); hold on;
%%
%% Numerical time integration
%% loop
while t < t1
  %% We need P,Fpg,z,v,E,Fgp,Qbot (G constant)
  %% we have z,v,T
  %% find P,Fpg,E,Qbot,Fgp using constitutive relations
  P = m*v;
  Fgp = -(N*R*T/z - A*mu*v/z);
  Fpg = -Fgp;
```

```
  E = C*N*T;
  Qbot = A*h*(Te - T);
  %%
  %% Drive forward in time
  %% update momentum of piston
  P = P + (Fpg + Fatm + G)*dt;
  %% update position of piston
  z = z + v*dt;
  %% update internal energy of gas
  E = E + (Qbot + Fgp*v)*dt;
  %% update time
  t = t + dt;
  %%
  %% Find new state for next iteration
  %% We need z,v,T
  %% we have P,z,E
  %% find v,T using constitutive relations
  v = P/m;
  T = E/(C*N);
  %%
  %% Check whether to save & plot at this step
  if min(abs([0 dsave] - mod(t-t0, dsave))) <= dt/2
    i = i+1;
    tSave(i) = t;
    zSave(i) = z;
    vSave(i) = v;
    TSave(i) = T;
    plot(t, z, 'o','Color',cols(1,:));
    pause(0.001);
  end
end
%% Plot trajectory
plot(tSave,zSave,'-','Color',cols(1,:));
subplot(2,1,2)
plot(tSave,TSave-273.15,'-','Color',cols(2,:)); axis('tight');
xlabel('time {\it t}/s'); ylabel('temperature {\it T}/C');
```

11.7 Surfaces of discontinuity

From our discussion about friction, we know what happens from the point of view of momentum when a person is pushing a heavy object, say a crate, on the floor. The person is providing a constant influx of horizontal momentum F_p to the crate, but the floor is also providing a horizontal-momentum influx F_f, kinetic friction, having opposite orientation. Suppose

› § 10.2 page 175



that the influx by the person has same magnitude as the friction, $F_\mathrm{p} = -F_\mathrm{f}$. Then the crate will move with constant horizontal velocity v, because its time-rate change of horizontal momentum is zero:

$$\frac{\mathrm{d}P_x(t)}{\mathrm{d}t} = F_\mathrm{p} + F_\mathrm{f} = -F_\mathrm{f} + F_\mathrm{p} = 0\,\mathrm{N}\,.$$

Let us consider what happens from the point of view of energy. Through the control surface where the person pushes the crate there's an influx of mechanical power $F_\mathrm{p} \cdot \boldsymbol{v}$. Let's say that there's no heat flux, so the energy influx through that surface is

$$E_\mathrm{p} = F_\mathrm{p} \cdot v > 0\,.$$

This energy influx is positive because the force exerted by the person and the velocity have the same orientation.

What about the contact surface of crate and floor? There should be an energy flux through there too. In fact, there must be one if the internal energy U of the crate is constant in time: because the crate's kinetic energy $\frac{1}{2}mv^2$ and gravitational potential energy are also constant, so the total energy content is constant, and therefore the energy flowing in from the push must flow out from somewhere else in this case. Obviously its flowing through the surface between crate and floor. What kind of energy flux occurs there? is it heat? or is it mechanical power?

We don't have a constitutive relation for the energy flux Φ between crate and floor. The relation $\Phi = Q + \boldsymbol{F} \cdot v$ does *not* apply there, because the velocity of matter v is not the same in the proximity of the surface. On the upper side, the matter of the crate is moving with velocity v; but on the lower side, the matter of the floor is at rest, with velocity zero.

This contact surface is an example of *surface of discontinuity*

Surface of discontinuity

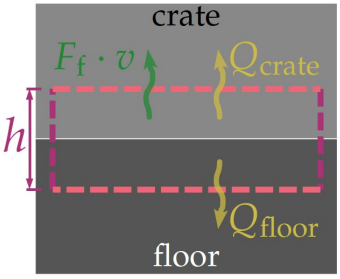
If the value of a physical quantity Q has two different values as we consider points closer and closer to the two sides of a small control surface, then the latter is called a **surface of discontinuity for the quantity** Q.

A control surface may be one of discontinuity for some quantity but not for another. In our present example, for instance, we might have the same temperature on the floor and at the bottom of the crate; the contact surface

between them is then a surface of discontinuity for the velocity, but not for the temperature.

We shall now learn a technique to deal with surfaces of discontinuity. This technique allows us to extend the application of some constitutive relations also to cases where they at first cannot be applied. Our discussion of the technique is largely intuitive, but it could be made mathematically rigorous.

Imagine to zoom in on the imaginary control surface that separates the crate and the floor. Replace this surface with a very thin imaginary control volume; see side picture. Two sides (light-red dashed lines) of this control volume have the same extension as the original surface and are parallel to it; but one of the lies completely within the crate, and the other completely within the floor. The lateral sides (dark-red dashed lines) of the control volume have a very small height h.



Consider the fluxes of momentum for this control volume. They can be separated (thanks to extensivity) into three contributions:

Surface within crate: Horizontal momentum, as friction (having leftward orientation in the illustration), is coming from the floor, and is passed on to the upper parts. Through this surface there is therefore an *ef* flux of momentum equal to F_f.

All matter close to this surface has velocity v, on both sides, because this surface is completely within the crate. For the energy *ef* flux Φ_{crate} through this surface we can therefore use the constitutive relation

$$\Phi_{crate} = Q_{crate} + F_f \cdot v$$

where Q_{crate} is a possible upward heat flux.

Surface within floor: Horizontal momentum, as friction (having rightward orientation in the illustration), is coming from the crate, and is passed on to the lower parts. Through this surface there is therefore an *ef* flux of momentum equal to $-F_f$.

All matter close to this surface has zero velocity, on both sides, because this surface is completely within the floor. For the energy *ef* flux Φ_{floor} through this surface we can therefore use the constitutive relation

$$\Phi_{floor} = Q_{floor}$$

where Q_{floor} is a possible downward heat flux, and there is no mechanical power owing to the zero velocity.

Side surfaces: We imagine to take the height h to be extremely small, so small that the area of the side control surfaces is negligible. All fluxes through these surfaces are, intuitively, negligible.

Now let's examine the balance of energy for this thin control volume. Owing to the very small height h, the volume is extremely small. Therefore the energy content E and its time-rate of change $\mathrm{d}E/\mathrm{d}t$ are, intuitively, negligible. The balance of energy then yields

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\Phi_{\text{crate}} - \Phi_{\text{floor}}$$

$$0\,\mathrm{J/s} = -(Q_{\text{crate}} + F_{\text{f}} \cdot v) - Q_{\text{floor}}$$

from which we find

$$\Phi_{\text{crate}} = -\Phi_{\text{floor}}$$

$$Q_{\text{crate}} + F_{\text{f}} \cdot v = -Q_{\text{floor}}$$

This is a very interesting result. Zoom out from the thin control volume, so that it looks like the initial control surface between crate and floor. What happens at this control surface is the following:

- The symmetry of flux is still valid: the energy flux from floor to crate, Φ_{crate}, and the energy flux from crate to floor, Φ_{floor}, are equal in magnitude but opposite: $\Phi_{\text{crate}} = -\Phi_{\text{floor}}$.
- However, the energy flux appears *on one side (crate) of the surface as heat flux plus mechanical power*, whereas *on the other side (floor) only as heat flux*.

› § 4.6 page 87

This makes sense from a molecular point of view: roughly speaking, the kinetic energy of the visible, coordinated motion of the molecules that make up the crate, is transformed into kinetic energy of the microscopic uncoordinated motion of the molecules that make up the floor. But note the amazing fact that we obtained this result by applying the balance laws, without invoking any molecular picture. In Chapter 14, about the balance of entropy, we'll also discover that at this kind of surfaces of discontinuity, positive influx of mechanical power on one side can be converted to positive outflux of heat on the other side – but the opposite conversion cannot happen.

URLs for chapter 11

1. https://www.britannica.com/event/Industrial-Revolution
2. http://hyperphysics.phy-astr.gsu.edu/hbase/gpot.html
3. https://itrf.ign.fr/en/background
4. https://aa.usno.navy.mil/faq/ICRS_doc
5. https://hpiers.obspm.fr/icrs-pc/newwww/icrf/index.php
6. https://commons.wikimedia.org/wiki/images/File:Piezoelectric_pickup1.jpg
7. http://hyperphysics.phy-astr.gsu.edu/hbase/Solids/piezo.html
8. http://hyperphysics.phy-astr.gsu.edu/hbase/Solids/magstrict.html
9. https://www.jpl.nasa.gov/edu/news/2023/7/24/exploring-the-mystery-of-our-expanding-universe/
10. https://www.preposterousuniverse.com/blog/2010/02/22/energy-is-not-conserved/
11. https://www.preposterousuniverse.com/blog/2010/02/22/energy-is-not-conserved/
12. https://math.ucr.edu/home/baez/physics/Relativity/GR/energy_gr.html
13. http://hyperphysics.phy-astr.gsu.edu/hbase/thermo/pvtsur.html#c1
14. https://bunkajapan.com/blogs/japanese-tea-culture/how-to-hold-a-tea-cup
15. https://www.britannica.com/technology/materials-science
16. https://imgur.com/BzcZOvl
17. https://doi.org/10.1351/goldbook.G02579
18. https://www.energy.gov/energysaver/thermoelectric-coolers
19. https://pglpm.github.io/7wonders/code/idealgas_piston.m

12 Balance of angular momentum

Around the world, around the world

Daft Punk 2005a

12.1 Formulation and generalities

Balance of angular momentum

Volume content: $\boldsymbol{L}$ Flux: $\boldsymbol{\tau}$ Supply: $\boldsymbol{M}$

$$\boldsymbol{L}(t_1) = \boldsymbol{L}(t_0) + \int_{t_0}^{t_1} \boldsymbol{\tau}(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{M}(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}\boldsymbol{L}(t)}{\mathrm{d}t} = \boldsymbol{\tau}(t) + \boldsymbol{M}(t)$$

integral form differential form

(12.1)

The balance of angular momentum – or rotational momentum or moment of momentum – is no less important and no less present in everyday phenomena than the balances of momentum or energy. Like the balance of momentum, it is at the core of applications where motion and stability are important. It is crucial in studying, predicting, and planning the motion of artificial satellites around Earth and of celestial bodies, from planets to galaxies.

Yet the balance of angular momentum is mentioned less often, and may *seem* less often, used than the balance of momentum. This happens because of several related reasons. The most important constitutive relations involving angular momentum require slightly more complicated mathematics, in particular they need explicit integration of functions and the use of vector cross-products or bivectors. Because of this, the consequences

of the balance of angular momentum are often pre-emptively baked-in into constitutive relations; we shall soon see an example with the Hookean spring.

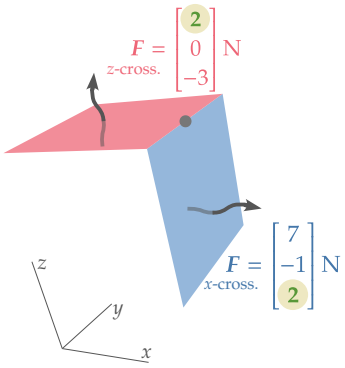
It moreover turns out that *the balance of angular momentum can be expressed without the time derivative of a volume content, or without the time-integrals of a flux and of a supply*. It can instead be expressed as a sort of symmetry condition, we can call this the "tensorial expression" of this balance:

Balance of angular momentum (tensorial form)

Take a small control surface passing through some point P, of area A, parallel to the zx directions, and with a positive-y crossing direction. Take another small control surface also passing through P and of area A, parallel to the xy directions, and with a positive-z crossing direction. Then the flux of z-momentum F_z through the zx control surface must be equal to the flux of y-momentum F_y through the xy control surface. Two more analogous laws hold by exchanging y & x or z & x in the statement above:

$$\underset{y\text{-crossing}}{F_z} = \underset{z\text{-crossing}}{F_y} \qquad \underset{x\text{-crossing}}{F_z} = \underset{z\text{-crossing}}{F_x} \qquad \underset{y\text{-crossing}}{F_x} = \underset{x\text{-crossing}}{F_y} \tag{12.2}$$

These equations only hold if the momentum fluxes are uniform throughout the surfaces.



The x-component of the momentum flux crossing in the z-direction must equal the z-component of the momentum flux crossing in the x-direction. In this picture each equals 2 N.

You notice that in this alternative formulation we don't even need to speak of a volume content, flux, or supply of angular momentum, and we don't use derivatives or integrals.

The two expressions of the balance of angular momentum are completely equivalent, as we shall prove later on. Each one has advantages and disadvantages. The integral and differential expressions (12.1) involve time-integration or time-differentiation, but they apply to control volumes of arbitrary shape, size, motion. The tensorial expression (12.2) does not involve integration or differentiation in time, but it can usually only be applied to very small control surfaces.

Each expression is used in particular applications and disciplines. The integral and differential expressions are typically used in applications with small bodies or approximately rigid bodies, or applications where forces are given explicitly, rather than via constitutive relations. The tensorial expression is typically used in material science and fluid dynamics, where

constitutive relations for the momentum flux $\boldsymbol{F}$ are given in such a way that the tensorial law is automatically satisfied. This is again one case in which the balance of angular momentum plays its role invisibly, automatically satisfied.

Definitions of angular momentum

Just like total energy, also angular momentum can be defined in different ways; in fact, the freedom we have in its definition is more visible than the one we have in defining total energy.

To be continued.

12.2 Examples of constitutive relations

Standard constitutive equation for angular momentum

If $\boldsymbol{P} = (P_x, P_y, P_z)$ is the momentum in a *small* volume, and $\boldsymbol{r} = (x, y, z)$ is the position vector of this small volume, then the angular momentum $\boldsymbol{L} = (L_x, L_y, L_z)$ *with respect to the origin of coordinates*, contained in that same volume, is given by the vector product

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{P} \qquad \text{or equivalently} \qquad \begin{cases} L_x = y\, P_z - z\, P_y \\ L_y = z\, P_x - x\, P_z \\ L_z = x\, P_y - y\, P_x \end{cases} \tag{12.3a}$$

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:

$$\begin{cases} L_{yz} = y\, P_z - z\, P_y \\ L_{zx} = z\, P_x - x\, P_z \\ L_{xy} = x\, P_y - y\, P_x \end{cases} \tag{12.3b}$$

Choose whichever you prefer.

If $\boldsymbol{F} = (F_x, F_y, F_z)$ is the flux of momentum through a *small* surface at a given time, and $\boldsymbol{r} = (x, y, z)$ is the position vector of that surface, 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

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

$$\text{or equivalently} \quad \begin{cases} \tau_x = y\,F_z - z\,F_y \\ \tau_y = z\,F_x - x\,F_z \\ \tau_z = x\,F_y - y\,F_x \end{cases} \quad \text{or} \quad \begin{cases} \tau_{yz} = y\,F_z - z\,F_y \\ \tau_{zx} = z\,F_x - x\,F_z \\ \tau_{xy} = x\,F_y - y\,F_x \end{cases} \tag{12.4}$$

Exercise 12.1

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

$$\boldsymbol{r} = \left[1.4 \times 10^{+7}\,,\ 1.6 \times 10^{+7}\,,\ 0\right] \text{m}$$
$$\boldsymbol{P} = \left[-3.1 \times 10^{+6}\,,\ 3.4 \times 10^{+6}\,,\ 0\right] \text{N} \cdot \text{s}$$

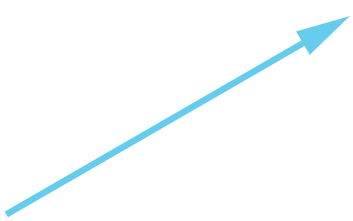
Assuming that the satellite's volume can be considered small enough for the present purpose, calculate the satellite's angular momentum.

To be continued.

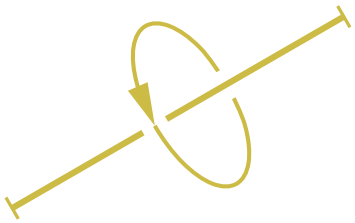
Add discussion: the fact that the forces exerted by a Hookean spring are parallel to the spring is a consequence of the balance of angular momentum.

12.3 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:



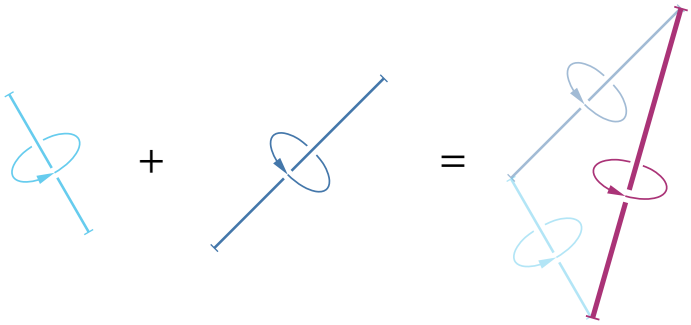
an ordinary vector



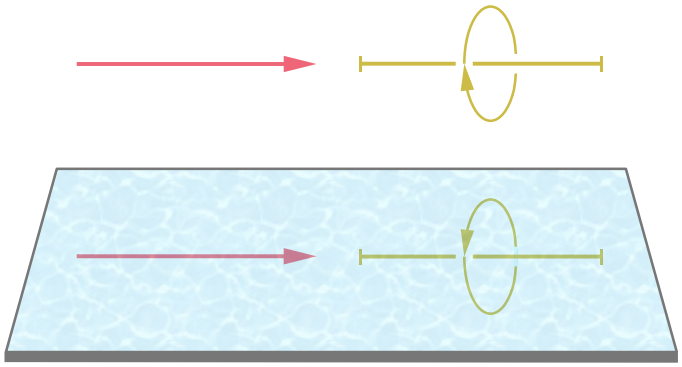
a twisted vector

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:

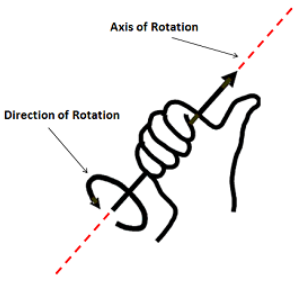


Ordinary vectors and twisted vectors behave very differently if we look at their images through a mirror parallel to their axis: the orientation of ordinary vectors appears unchanged, whereas the orientation of twisted vectors appears *reversed*:



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!

Remarks on momentum and energy 13

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

13.1 Common misunderstandings on momentum, energy, angular momentum

Momentum of what? Energy of what?

To be written in a later version

14 Balance of entropy

> Their various "second laws" sound more like warnings or threats than principles of a rational science.
>
> C. A. Truesdell, III 1984

14.1 Formulation and generalities

Balance of entropy

Volume content: S Flux: Π

$$S(t_1) \geq S(t_0) + \int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t \qquad \frac{\mathrm{d}S(t)}{\mathrm{d}t} \geq \Pi(t) \tag{14.1}$$

integral form — differential form

The balance of entropy expresses what's commonly called "second law of thermodynamics". Entropy and its balance are successfully used in many applications, but our understanding of them and of their physical foundation is still incomplete. This state of affairs is reflected in the many and wildly different presentations of entropy and its balance: different in wording, mathematical formulation, scope, and sometimes even in physical consequences.

Many textbooks only present limited and special cases of properties and uses of entropy and its balance, and unfortunately they often make these limited, special cases appear as more general, or of broader application, than they actually are. Such textbooks also typically restrict themselves to situations where there the contents of quantities in a control volume do not change with time, and the fluxes are zero. We call this a situation

of *equilibrium*. The discipline that studies equilibrium situations is called *thermostatics*.

In these lecture notes, entropy and its balance are presented from a point of view, actively developed and used since the 1960s, having the following features:

- It has been used for many years in concrete technological applications (some examples at NASA: Chang & Haddad 1971; Hughes et al. 1986; Turon et al. 2004; Diosady et al. 2018; Kato & Rose 2020), and in the study complex materials such as polymers and mixtures.
- It has led to new physical constitutive relations, or to the physical and mathematical foundation of existing ones, from first principles.
- It is formulated with the same mathematics, and at the same mathematical level, as the physics of matter, momentum, angular momentum, energy, and electromagnetism.
- It includes time-dependent phenomena and is fully connected with phenomena involving the other six basic quantities.

In the technical literature the entropy balance above, used in its full generality, goes under the name of **Clausius-Duhem inequality**[1].

Thermodynamic entropy and statistical entropy

One added difficulty is that entropy and its balance can also be approached from a completely different direction, especially when we study physical systems on small scales. It's the approach of *statistical mechanics*[2], which considers physical situations in where we lack information about initial conditions, or boundary conditions, or constitutive relations. In statistical mechanics, a conceptually different entropy appears, not as a physical quantity, but as a measure of our lack of information about the physical system, in the strict technical sense of Information Theory[3]. Also this entropy satisfies a balance law very similar to (14.1) above.

> One of the reasons for the bewilderment which is sometimes felt at an unheralded appearance of the term entropy is the superabundance of objects which bear this name. On the one hand, there is a large choice of macroscopic quantities (functions of state variables) called entropy, on the other hand, a variety of microscopic quantities, similarly named, associated with the logarithm of a probability or the mean value of the logarithm of a density. Each one of these concepts is suited for a specific purpose. More confusing, however, than the lack of imagination in terminology is the fact that several of these distinct concepts, different in meaning and in numerical value, may be significant in a single problem. (Grad 1961 § 1 p. 323)

We thus have a **thermodynamic entropy** and a **statistical entropy**. Their fascinating relation is only partly understood, and still the object of some debate. In the present notes we shall focus on *thermodynamic* entropy.

Entropy depends on the observation scale

The entropy content in a control volume and the entropy flux across a control surface are not uniquely defined; that is, several choices are possible, each one correct in a specific situation. This non-uniqueness is actually two-fold. In the present section we discuss a first sense in which entropy is non-unique; in a later section we'll discuss a second sense.

First let's make clear that the entropy content in a given control volume, and the entropy flux through a given control surface, at a given coordinate time, *do not depend on the coordinate system* chosen. In this regard they are like the content & flux of matter, electric charge, and magnetic flux; and unlike momentum, angular momentum, and energy.

On the other hand, entropy content & flux do *depend on the detail and scale of observation* of a physical phenomenon. In this regard they are unlike all other six fundamental quantities, whose *total* content and flux do not depend on the observation scale.

As an example, take a control volume containing air. This control volume could be studied, described, and measured in three different ways: (a) as containing a continuous, fluid amount of matter of one kind: 'air'; (b) as containing a continuous, fluid mixture of amounts of matter of different kinds: nitrogen, oxygen, and several others; (c) as containing a bunch of molecules in motion.

The *energy* content measured within this control volume, at a particular time instant, will be exactly the same in all three cases. What changes among them is the division of this total energy into internal and kinetic, but the total is the same.

› § 11.2 page 219

The *entropy* content assigned to this control volume, on the other hand, will be different in each case.

> A given object of study cannot always be assigned a unique value, its "entropy". It may have many different entropies, each one worthwhile. The proper choice will depend on the interests of the individual, the particular phenomena under study, the degree of precision available or arbitrarily decided upon, or the method of description which is employed; and each of these criteria is largely subject to the discretion of the individual. [. . .]
>
> For another example we turn to aerodynamics. The existence of diffusion between oxygen and nitrogen somewhere in a wind tunnel will usually be

> of no interest. Therefore the aerodynamicist uses an entropy which does not recognize the separate existence of the two elements but only that of "air". In other circumstances, the possibility of diffusion between elements with a much smaller mass ratio (e.g., 238/235) may be considered quite relevant. (Grad 1961 § 1 pp. 323, 325)

We shall now see that this peculiar dependence of entropy on the details and scale of observation actually makes a lot of sense when we understand how the entropy balance is used.

14.2 The physical role of the balance of entropy

Amazing variety of consequences

The entropy balance (14.1) has an apparent peculiarity, compared to the general form of a balance law: it is not an equality

› § 5.5 page 118

$$\ldots = \ldots$$

but an *inequality*

$$\ldots \geq \ldots$$

Why? and what are the consequences of this peculiarity?

Many texts and media try to summarize the meaning of the entropy balance and its inequality sign in simple terms. But the reality is that this inequality leads to an amazing variety of very different phenomena, and cannot be summarized in words.

This should not be surprising. Take the balance of momentum for example. It leads to all sorts of motions and deformations of objects, but also to the stability of objects: from the extremely complicated motion of the atmosphere to the stillness of a pen resting on a table. This balance law could not be summarized in some simple sentence, like "objects spontaneously fall downward". First, such a sentence would be false in many situations: just look at the motion of rocks expelled by a volcano, or at cosmological expansion. Second, it would be useless for precise predictions and numerical simulations.

The same remark holds true for the balance of entropy. This balance also leads to a wild variety of physical consequences. The mixing of two liquids can be said to be a consequence of the balance of entropy (together with the other balances). But also the appearance of life on Earth is a consequence of the balance of entropy; and all complex physical mechanisms underlying

The mixing of two liquids and the growing of life are both consequences of the second law of thermodynamics.

life are consequences of the balance of entropy, too (together with the other balances). Any simplistic summaries, like common ones mentioning "tendency to disorder" or "spontaneously doing this or that" or similar, are (a) simply *false* in many physical situations; (b) vague: what does 'to tend' mean? how is 'disorder' defined and quantified? to what initial values of position, velocity, and so on, does 'spontaneous' refer to? what's 'spontaneous' and what's not? (c) useless for quantitative predictions.

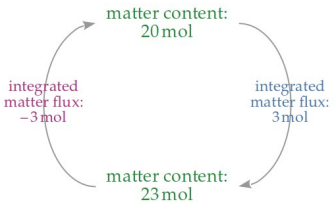
Irreversibility

From a qualitative point of view, the inequality sign in the balance of entropy expresses *irreversibility*; but we must understand this word in the right way.

Consider first a generic balance or conservation law; let's take conservation of matter for instance:

$$N(t_1) = N(t_0) + \int_{t_0}^{t_1} J(t)\,\mathrm{d}t$$

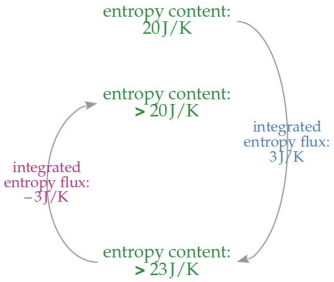
Suppose a control volume contains an amount 20 mol of matter. During an interval of time we provide a net flow of matter $\int_{t_0}^{t_1} J(t)\,\mathrm{d}t = 3$ mol into the control volume. By the conservation law, at the end we have an amount 23 mol of matter in the control volume. This also means that in principle we could revert to an amount of 20 mol by providing a *negative* amount of the same flow as before, −3 mol. We can, in principle, change the matter content in the volume between 20 mol and 23 mol by providing a net flow of +3 mol or its opposite −3 mol. If we consider a balance with a supply, the conclusion remains the same, if we invert also the sign of the supply.



Now consider the balance of entropy with a strict inequality:

$$S(t_1) > S(t_0) + \int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t$$

Suppose a control volume contains an amount 20 J/K of entropy. During an interval of time we provide a net flow of entropy $\int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t = 3$ J/K into the control volume. By the inequality above, at the end we *cannot* have an amount 23 J/K of entropy in the control volume: it will be larger, say 23.1 J/K. If we want to revert to an entropy content of 20 J/K, then simply reversing the total flow to −3 J/K won't be enough. In fact, even −3.1 J/K



won't be enough, because by the inequality above we must have

$$\text{new entropy amount} > 23.1\,\text{J/K} - 3.1\,\text{J/K}$$

and the new entropy content would be more than 20 J/K.

This is what we call **irreversibility**: we cannot revert to a given entropy content in the volume *by inverting the entropy flux*.

'Irreversibility' does not mean that we cannot revert to a given content

Note that the example above does *not* imply that we can never have again an entropy content of 20 J/K in the control volume. It is only showing that we cannot achieve this by simply providing an entropy flux opposite in sign to the flux we provided before. The original entropy content can still in principle be re-established; but in order to achieve this we need to providing a larger negative entropy flux than before. Analogously, the example does *not* imply that the entropy content of a control volume can only increase. The entropy content can very well decrease, provided an enough large negative entropy flow is provided.

In cases where the balance of entropy holds with an equal sign:

$$S(t_1) = S(t_0) + \int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t$$

then we can of course revert to a given entropy content by inverting the entropy flux.

Reversible and irreversible processes

We call a physical change, process, transformation, or phenomenon between two times t_0 and t_1 **reversible** if it satisfies the balance of entropy with an equal '=' sign, at least approximately.

We call it instead **irreversible** if it satisfies the balance of entropy with a strict inequality '<' sign.

There is no clear-cut division between 'reversible' and 'irreversible' processes. A process may lead to an entropy content that is *slightly* larger than the one we would have obtained with a perfectly reversible process; but the difference is so small, with respect to our approximations, that it can be neglected. In this case the process is called reversible for

all practical purposes. Arguably there are in fact no *exactly* reversible processes in nature.

Entropy balance as a meta-law

One important consequence of the inequality sign in the balance of entropy is that this balance cannot be used in a numerical-time-integration scheme. If we rewrite it in an approximate form for a short timestep Δt:

› § 6.1 page 144

$$S(t + \Delta t) \gtrsim S(t) + \Pi(t)\,\Delta t$$

this relation doesn't tell us the value of the entropy content at time $t + \Delta t$, but only an approximate minimum value that this content could have. As far as we know, it could be much larger than this minimum. For example, if a closed control surface at some time t contains an amount of entropy $S(t) = 10\,\mathrm{J/K}$, and the net entropy flux at that time is $\Pi(t) = 2\,\mathrm{J/(K\,s)}$, then we can say that a time $\Delta t = 0.1\,\mathrm{s}$ after the content should be

$$\begin{aligned} S(t + \Delta t) &\gtrsim S(t) + \Pi(t)\,\Delta t \\ &\gtrsim 10\,\mathrm{J/K} + 2\,\mathrm{J/(K\,s)} \cdot 0.1\,\mathrm{s} \\ &\gtrsim 10.2\,\mathrm{J/K} \end{aligned}$$

This mean that maybe $S(t + \Delta t) \approx 10.2\,\mathrm{J/K}$, or maybe $S(t + \Delta t) = 1000\,\mathrm{J/K}$, or maybe even larger values – the entropy balance doesn't really tell us. This law, therefore, apparently does not allow us to "drive forward" a physical system.

In fact, when we simulate or predict the behaviour of any physical system, it turns out that we can in principle always do without the balance of entropy. We only need to use the other six balances together with any relevant constitutive equations. The entropy content S itself may appear in these constitutive relations, and is often a useful quantity. But the balance of entropy is not used.

What is its physical role of this balance law, then?

The answer is that *the balance of entropy is a 'meta-law'*. It is not a law directly about physical phenomena, but a *law about laws* of physical phenomena. Roughly speaking, this meta-law determines which constitutive relations are physically admissible and which are not admissible.

> The subtlety arises, of course, because the second law is an inequality, and therefore only other inequalities can be deduced from it by a purely algebraic procedure. However, [. . .] one can in fact deduce from the second law, when coupled with appropriate constitutive assumptions, consequences which are equations, not inequalities; the procedure is more a logical than an algebraic one. (Astarita 1990 § 2.4 p. 46)

In very simple cases this meta-law leads to restrictions on the values of physical coefficients. For example, it's a consequence of the balance of entropy that the viscosity coefficient μ in the ideal-gas law and the coefficient of heat transfer h in Newton's law of cooling must be positive. We shall see a simple example of how this kind of restrictions come about.

› § 11.5 page 232
› § 11.5 page 234

But in more complex cases this meta-law leads to much more powerful results. For example, it can dictate that some constitutive relations cannot contain particular physical quantities as variables. The discussion of these complex and more fascinating cases unfortunately requires much more advanced mathematics, so we can only get a dim glimpse of them in these notes.

If you feel adventurous, check the simple examples discussed in Astarita's book, in the section cited in the quote, or in Chapter 2 of Samohýl & Pekař 2014.

Considering the role of the entropy balance as a meta-law that decides which constitutive relations are admissible and which aren't, and considering that constitutive relations heavily depend on the scales of time & space and on the measurement precision with which we describe a physical phenomenon, it then makes sense that the entropy content and the entropy flux should also depend on details and scale of observation, as discussed in a previous section.

› § 5.9 page 135
› § 14.1 page 259

Entropy depends on a reference state

Besides depending on the observation scale, entropy additionally depends on the choice of a *reference state* of a control volume; that is, on a particular set of values for the minimal number of quantities needed to predict what will happen in the control volume. The reference state is a state to which we assign zero entropy by convention.

› § 14.1 page 259
› § 10.5 page 201

For many physical phenomena and materials, a change of the reference state simply shifts all entropy values by a constant amount, so our entropy is defined except for a 'zero' of its measurement scale. But there are also important and quite common phenomena and materials for which a change of reference state leads to a *different* entropy function – not just the addition or subtraction of a constant. An ordinary paper clip is a simple

The entropies that can be defined for an ordinary paper clip do not differ simply by a constant.

example. The non-uniqueness of entropy is also a consequence of the peculiar inequality sign in the balance of entropy.

But this non-uniqueness is not a problem. Any one of the entropies, defined with respect to different reference states, can be used for instance as one of the quantities that define the state of the physical system. The prediction of the physical system's behaviour will be the same.

'First, there are bodies which have two entropies [...] whose difference is nonconstant [...]. Second, it can happen that the body does not have a smooth entropy. These two peculiarities are related to the fact that for the entropy we have only an inequality.' Šilhavý 1997 § 7.6

14.3 Examples of constitutive relations

We have emphasized that the balance of entropy (together with the other balances) imposes restrictions on the possible constitutive relations between the physical quantities used to model a physical phenomenon. Among such quantities are also the entropy content S and entropy flux Π which enter this very balance (if you think about it, it's extremely intriguing that a physical law manages to determine the expressions of the very terms it contains).

Entropy flux, heat, temperature

The most important constitutive relation for entropy is the one that relates its flux Π with the heat flux Q and the thermodynamic temperature T.

› § 11.3 page 222

Entropy flux

Consider a control surface, possibly moving, with an assigned crossing direction, and satisfying the following conditions:

- the surface is in contact with matter on both sides
- no flux of matter through the surface
- no chemical reactions (transformations between matter types) occur across the surface
- no electromagnetic phenomena involved
- the temperature T is the same on every part of the surface
- the heat flux through the surface is Q

Then across the control surface there is an entropy flux Π given by

$$\Pi = \frac{Q}{T} \tag{14.2}$$

This relation has wide applicability, but pay close attention to the conditions for its validity. In particular, it is not correct if there is a net flux of matter

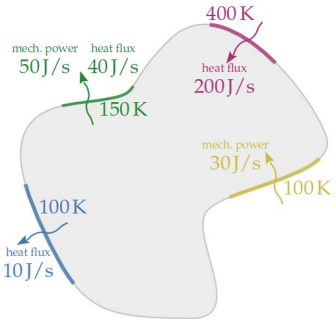
through the surface. Nor is it correct if there are opposite fluxes of *different kinds* of matter that cancel each one out – so the *net* flux is zero – but chemical reactions are occurring between these different matter kinds. It is also important that the temperature be well-defined and be uniform, that is, have the same value, on the surface and at least in a small spatial region on each side of the surface. If the temperature is not uniform, the surface is usually divided into smaller parts, so that the temperature can be considered uniform in each part separately, and then the total flux is obtained by summation, thanks to extensivity.

› § 3.2 page 54

Exercise 14.1

1. At a particular time instant, across a closed control surface there is a flux of energy as illustrated in the side picture:

- through one part of the surface there is a heat influx of 200 J/s and no flux of mechanical power; the temperature around that part is 400 K
- through one part there is a 40 J/s heat efflux and a 50 J/s efflux of mechanical power; the temperature around that part is 150 K
- through one part there is a 10 J/s heat efflux and no flux mechanical power; the temperature around that part is 100 K
- through one part there is a 30 J/s influx of mechanical power and no heat flux; the temperature around that part is 100 K
- through the rest of the surface there are no energy fluxes of any kind.



Each part of the surface satisfies the conditions of the constitutive relation (14.2).

How much is the *net influx* of entropy through the whole closed control surface?

2. Suppose that through a particular control surface, at a given time, there is zero *net* energy flux. The surface satisfies the conditions of the constitutive relation (14.2). Can we say that the entropy flux through that surface is also zero?

Entropy of an ideal gas

In Chapter 11 we discussed constitutive relations for the pressure p and the internal energy U of an control volume containing an ideal gas:

› § 11.5 page 230

$$p = \frac{RNT}{V} - \mu \frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}t}\,, \qquad U = C\,N\,T$$

where N is the amount of gas, T the temperature of the gas, assumed uniform, V is the volume, and R, C, μ are the molar gas constant, molar heat capacity, and viscosity coefficient.

Under the same conditions of validity for the constitutive relations above, we also have a constitutive relation for the entropy content S of the volume of ideal gas:

Entropy of an ideal gas

If a small control volume contains an amount N of ideal gas at uniform temperature T, then it also contains an amount of entropy S given by

$$S = CN \ln \frac{T}{T_0} - RN \ln \frac{N/V}{N_0/V_0} \tag{14.3}$$

where T_0, N_0, V_0 are arbitrary reference temperature, amount of matter, and volume.

14.4 Examples of applications

Thermal engines

Recall that we are completely free in our choice of a control volume: it can have any size and shape, and can move and deform in any way. This freedom is extremely powerful: we can for example imagine a control volume that wraps a very complex engine having moving parts. Through the surface of this control volume we can keep track of any exchanges of matter, momentum, energy between the engine and its exterior; in particular exchanges of heat and of mechanical power. And whatever happens within the engine, that is, within our imaginary control volume, must obey the seven universal balance laws.

› § 4.3 page 79

This powerful freedom in choosing a control volume, when combined with the balances of entropy and energy, can lead to amazingly general

physical results. Thanks to these results we can for example prevent waste of efforts in technological ideas that would eventually turn out to be unfeasible. Let us see a couple of examples.

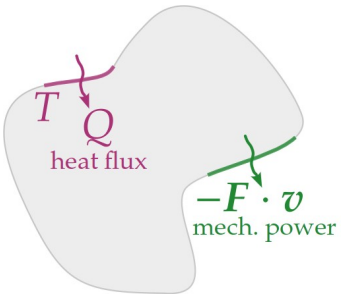
First of all let's define what we mean by 'thermal engine': a device that can absorb or emit both heat and mechanical work, and that can operate repeatedly, in principle forever. A device operated by an electric battery, for instance, is not an engine, because it will cease operating once the battery is exhausted. The ability to operate forever means that at recurring points in time the device must be find itself in the same state, so as to start over.

A thermal engine can also receive or release matter, momentum, angular momentum, and electromagnetic quantities.

An impossible thermal engine

Would it be possible to build a thermal engine that takes some energy in the form of heat from an inlet at constant temperature, and releases energy in the form of work, say by lifting an object?

The operation of such an engine is captured in the side picture. Imagine to wrap the engine, no matter how complex it could be, in a closed control surface, defining a control volume. A part (red) of the control surface delimits the inlet through which the engine receives a heat flux $Q(t)$, possibly variable in time. The temperature T at the inlet is *constant* in time. Another part (blue) of the control surface delimits the movable components through which the engine is releasing mechanical power $-\boldsymbol{F}(t) \cdot \boldsymbol{v}(t)$, where $\boldsymbol{F}(t)$ is the influx of momentum through that part, and $\boldsymbol{v}(t)$ is the velocity of the matter set into motion; both can vary with time. The expression for the mechanical power has a minus sign because it's the power *we* receive, so it's an *ef*flux for the engine. Through the rest (grey) of the control surface there are *no exchanges of heat*, but there may be fluxes of matter, momentum, angular momentum, and electromagnetic quantities; but we require that over an operation cycle the overall amount of each such flow be zero. Only the fluxes yQ and $-\boldsymbol{F} \cdot \boldsymbol{v}$ can have a non-zero net amount over an operation cycle.



The engine starts a cycle at time t_0 and operates until time t_1, at which time its state is exactly the same as at the initial one, the cycle is complete, and the engine is ready to start a new operation cycle. In a cycle, the total

amount of heat ΔH we provide to the engine and the total amount of work ΔW we *receive* from it are given, with a shorter notation, by

$$\Delta H \coloneqq \int_{t_0}^{t_1} Q(t)\,\mathrm{d}t \qquad\qquad \Delta W \coloneqq -\int_{t_0}^{t_1} \boldsymbol{F}(t)\cdot\boldsymbol{v}(t)\,\mathrm{d}t$$

The net amount of energy flowing into the control volume between t_0 and t_1 is therefore

$$\int_{t_0}^{t_1} \Phi_{\text{tot}}(t)\,\mathrm{d}t = \int_{t_0}^{t_1} Q(t)\,\mathrm{d}t + \int_{t_0}^{t_1} \boldsymbol{F}(t)\cdot\boldsymbol{v}(t)\,\mathrm{d}t \equiv \Delta H - \Delta W$$

Let's see what the balance of energy says about such an engine. We have

$$\begin{aligned} U(t_1) &= U(t_0) + \int_{t_0}^{t_1} \Phi_{\text{tot}}(t)\,\mathrm{d}t \\ &= U(t_0) + \Delta H - \Delta W \end{aligned}$$

But the state of the engine at t_1 is the same as at t_0, therefore the energy content at these two times must be the same: $U(t_1) = U(t_0)$. The equation above simplifies to

$$\Delta W = \Delta H \tag{14.4}$$

that is, the mechanical work produced in a cycle must be equal to the total amount of heat provided in a cycle, as expected. The balance of energy doesn't require more than this.

According to the balance of energy this engine is therefore admissible.

Let's see what the balance of entropy says. We have

$$\begin{aligned} S(t_1) &\geq S(t_0) + \int_{t_0}^{t_1} \Pi(t)\,\mathrm{d}t \\ &\geq S(t_0) + \int_{t_0}^{t_1} \frac{Q(t)}{T}\,\mathrm{d}t \\ &\geq S(t_0) + \frac{\Delta H}{T} \end{aligned}$$

The last step, where the temperature T goes out of the time integral, is possible because the temperature is constant, according to our engine

design. Also in this case the initial and final entropy content must be the same: $S(t_1) = S(t_0)$. The equation above simplifies to

$$\frac{\Delta H}{T} \leq 0 \qquad \underset{\text{because } T > 0}{\Longrightarrow} \qquad \Delta H \leq 0 \tag{14.5}$$

This is a remarkable result: *it's impossible to give a positive net amount of heat, in a cycle, to such an engine*; otherwise the entropy-balance law would be broken.

Together with the conclusion (14.4) from the balance of energy, we also obtain

$$\Delta W \leq 0 \tag{14.6}$$

that is, *it's impossible to receive a positive net amount of work, in a cycle, from such an engine*.

We conclude that an engine designed in this way is physically impossible. No matter which kind of ingenious technology or materials we tried to use, we would never be able to cyclically gain positive mechanical work from it.

Note that the opposite use, though, is physically possible: we can, cyclically, provide positive work to the engine and get heat out at a constant temperature. Indeed this is how most heating systems operate.

It's important to understand correctly what's possible and what's not. We *can* of course provide positive heat to the device, at a constant temperature, as much as we please. The result above says that *we won't be able to return the device to its initial state* as long as we do so. The incompatibility is between:

- positive net amount of heat
- constant temperature
- cyclic operation

But note that the converse is possible: we can *absorb* a net amount of heat from the device, at constant temperature, as much as we please, and return the device to its initial state.

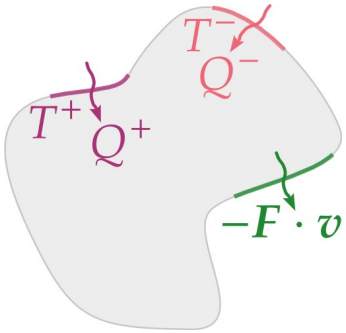
We can therefore try a different design, where one of the conditions above is dropped. The ability to operate cyclically is very convenient, so let's try to keep it. What happens if we let heat be exchanged at different temperatures?

A possible thermal engine, with limitations

Let modify our original design. Now we allow the exchange of heat to happen at two different temperatures. This could be done by changing the temperature of one part of the surface over time (within a cycle), or by allowing heat to be exchanged at two different inlets, having constant but different temperatures. We choose the second option as it's easier to analyse and lead to the same results as the first.

The new engine design is represented in the side picture. An influx of heat $Q^+(t)$ occurs through a part (dark red) of the closed control surface at constant temperature T^+; another influx of heat $Q^-(t)$ occurs through another part (light red) of the surface at constant temperature T^-. We assume

$$T^+ > T^-$$



but for the moment we are not making assumptions about $Q^+()$ and $Q^-(t)$; we only require that the net amount of heat provided to the engine in a cycle be positive. Through another, movable part (blue) of the control surface the engine is releasing mechanical power $-\boldsymbol{F}(t)\cdot\boldsymbol{v}(t)$. Through the rest (grey) of the surface there may be fluxes of matter and other quantities, except heat; but the net flow of such quantities is zero over an operation cycle.

We consider a cycle of the engine between times t_0, t_1. Employ again the shorter notation for the time-integrated fluxes:

$$\Delta H^+ \coloneqq \int_{t_0}^{t_1} Q^+(t)\,\mathrm{d}t \qquad \Delta H^- \coloneqq \int_{t_0}^{t_1} Q^-(t)\,\mathrm{d}t$$

$$\Delta W \coloneqq -\int_{t_0}^{t_1} \boldsymbol{F}(t)\cdot\boldsymbol{v}(t)\,\mathrm{d}t$$

so that the net amount of energy flowing into the control volume in this cycle is

$$\int_{t_0}^{t_1} \Phi_{\mathrm{tot}}(t)\,\mathrm{d}t = \Delta H^+ + \Delta H^- - \Delta W$$

The balance of energy applied to the engine's control volume requires that

$$U(t_1) = U(t_0) + \Delta H^+ + \Delta H^- - \Delta W$$

and since $U(t_1) = U(t_0)$ we find

$$\Delta W = \Delta H^+ + \Delta H^- \tag{14.7}$$

That is, the net work released in a cycle must be equal to the net heat provided, as expected. Note that we would like

The flux of entropy for the engine's control volume is

$$\int_{t_0}^{t_1} \Pi_{\text{tot}}(t)\,\mathrm{d}t = \int_{t_0}^{t_1} \frac{Q^+(t)}{T^+}\,\mathrm{d}t + \int_{t_0}^{t_1} \frac{Q^-(t)}{T^-}\,\mathrm{d}t \equiv \frac{\Delta H^+}{T^+} + \frac{\Delta H^-}{T^-}$$

The balance of entropy applied to the control volume therefore says

$$S(t_1) \geq S(t_0) + \frac{\Delta H^+}{T^+} + \frac{\Delta H^-}{T^-}$$

and since $S(t_1) = S(t_0)$ in this operation cycle we finally find

$$\frac{\Delta H^+}{T^+} + \frac{\Delta H^-}{T^-} \leq 0$$

For this new engine, the entropy balance is not saying that the net amount of heat provided in a cycle cannot be positive. It looks like the new engine design might work.

With a little algebra, and recalling that a thermodynamic temperature $T^- > 0$, we can rewrite the inequality above as follows:

$$\Delta H^- \leq -\frac{T^-}{T^+}\Delta H^+ \tag{14.8}$$

The fraction T^-/T^+ is positive; so if ΔH^+ is positive, ΔH^- must be negative, or vice versa. The entropy balance is therefore saying that *in a cycle, if the net amount of heat exchanged at one inlet is positive, then the net amount exchanged at the other inlet must be negative*.

We wished to have a positive net amount of heat provided to the engine in a cycle, that is,

$$\Delta H^+ + \Delta H^- > 0$$

This *is* indeed feasible! If ΔH^+ is positive, then ΔH^- is negative, but its absolute value is less than ΔH^+, because in formula (14.8) the fraction T^-/T^+ is less than 1. Suppose for instance that $\Delta H^+ = 8000\,\text{J}$, $T^+ = 400\,\text{K}$, $T^- = 100\,\text{K}$. Formula (14.8) then requires

$$\Delta H^- \leq -\frac{100\,\text{K}}{400\,\text{K}} \cdot 8000\,\text{J} = -2000\,\text{J}$$

and it would be possible to have, say, $\Delta H^- = -3000\,\text{J}$. The net heat amount provided to the engine in a cycle would then be $8000\,\text{J} - 3000 = +\,5000\,\text{J}$. According to the requirement (14.7) from the balance of energy, in a

cycle we would then gain a net work of 5000 J. The new engine design is successful!

An interesting question arises: how much work, in a cycle, can we squeeze out of our engine? We see that we are giving an energy amount ΔH^+ as heat to the engine, and the engine is returning to us ΔH^- as heat and ΔW as work. Can we somehow minimize ΔH^- and maximize ΔW?

It turns out that the balance of entropy also tells us what's the maximal amount of work that we can obtain from the engine. Combine together the requirements (14.7) and (14.8), by substituting ΔH^- from the latter into the former:

$$\Delta W = \Delta H^+ + \Delta H^- \text{ and } \Delta H^- \leq -\frac{T^-}{T^+}\Delta H^+ \quad \Longrightarrow \quad \Delta W \leq \Delta H^+ - \frac{T^-}{T^+}\Delta H^+$$

by using a little algebra we finally find the maximal work obtainable:

$$\Delta W \leq \left(1 - \frac{T^-}{T^+}\right)\Delta H^+ \tag{14.9}$$

The factor $1 - T^-/T^+$ is called the **efficiency** of the thermal engine. Since thermodynamic temperature is positive, the efficiency cannot be greater than 1.

In order to maximize the amount of work ΔW obtained and minimize the amount of heat ΔH^- received back from the engine, we must try to make the efficiency as close to 1 as possible. Looking at the fraction T^-/T^+ we see that there are two main ways, both of which can be pursued:

- Lower as much as possible, close to 0 K, the temperature T^- at which heat is received back from the engine
- Increase as much as possible the temperature T^+ at which heat is provided to the engine.

It is amazing that we can say, beforehand, how much work we can at most get from such an engine, without even knowing or needing to specify what kind of technology, materials, and way of operation it could be based upon. You see the strength of the consequences that the little "$\geq$" sign in the balance of entropy can have.

The thermal-engine example above also hints at the role of the balance of entropy as a meta-law about constitutive relations. In a real application and construction of an engine, the heat flux Q and momentum flux $\boldsymbol{F}$ will be concretely specified by constitutive relations; think for instance of

Newton's law of cooling for Q or the ideal-gas law for $\boldsymbol{F}$. But if a limitation such as the maximal work efficiency (14.9), which we can rewrite in full as

› § 11.5 page 230

$$-\int_{t_0}^{t_1} \boldsymbol{F}(t) \cdot \boldsymbol{v}(t)\, \mathrm{d}t \leq \left(1 - \frac{T^-}{T^+}\right) \int_{t_0}^{t_1} Q^+(t)\, \mathrm{d}t \ ,$$

is to be universally valid, then the specific mathematical formulae for Q and $\boldsymbol{F}$ cannot be whatever. In fact they turn out to have severe restrictions.

Constraints on constitutive relations for friction

To be written

URLs for chapter 14

1. https://encyclopediaofmath.org/wiki/Clausius-Duhem_inequality
2. https://plato.stanford.edu/entries/statphys-statmech/
3. https://www.britannica.com/science/information-theory

Postface to the teacher

事事無礙

The majority of everyday and forefront technologies is based on physical phenomena at the intersection of traditional categories such as "mechanics", "electromagnetics", "thermodynamics". In many cases it is not clear whether these phenomena should be labelled as belonging to one category rather than another. Sometimes such labelling is artificial, more misleading than helpful.

How to prepare students in physics or engineering for the myriad of possible physics specializations and applications lying ahead? The problem is not only that the students may not know yet which physics field they'll want to pursue, but also that they'll likely need some knowledge of all other fields anyway. I believe that the best approach is to teach them physical notions and physical laws that are common to as many physical phenomena as possible, and can be used in as many physical applications as possible. Notions and laws that the students will always be able to use afterwards, and upon which the students can gradually build more specialized knowledge.

As pointed out in the Preface, *we do actually have such notions and laws*, although some textbooks almost seem to forget their existence:

- what we can call 'matter' or 'substance' or 'particle number'
- energy
- momentum
- angular momentum (including the notion of centre of mass-energy)
- electric charge
- magnetic flux
- entropy

These quantities and their balances are usually recognized and emphasized in books on continuum physics, non-relativistic as well as relativistic. The classic treatise *The Classical Field Theories* (1960), for instance, states this programme very clearly:

> *Motion*, *stress*, *energy*, *entropy*, and *electromagnetism* are the concepts upon which field theories are constructed. Certain laws of *conservation* or *balance* are laid down as relating these quantities in all cases. These basic principles, which are in integral form, in regions where the variables change sufficiently smoothly are equivalent to differential *field equations* [. . .]. The field equations [. . .] form an underdetermined system, insufficient to yield specific answers unless further equations are supplied. Within the embracing concept of the balanced fields, it is possible to define *ideal materials* by certain further conditions. These defining conditions are called *constitutive equations*.

(As explained in Chapter 14, the balance of entropy is a *meta*law rather than a law.)

Besides being common to *all* physical phenomena, and to *all* our main physical theories, the notions and laws above have outstanding pedagogical features:

- They are few.
- They are easy to understand intuitively, because they express the idea of a "budget" or balance.
- They can be mathematically expressed by a common formula that is simple and yet can be directly applied, *as-is*, to all physical phenomena, theories, and disciplines.
- And this formula is directly and intuitively connected with the ideas of *prediction* and *simulation*.

Thanks to these amazing features, the students can be taught to recognize these notions and their mathematical "budgets" in the great variety of physical phenomena around them.

What will be left to learn – probably a lifelong learning – is the great variety of ways in which these seven basic notions can be mathematically connected to one another. It is this latter variety of *constitutive relations*, which depend on the physical phenomenon, that the students will thereafter specialize into. But note the advantage of having established

a foundation of a few simple universal notions: whenever our future engineers and physicists will need to deal with physical phenomena outside their specialization, they will only have to learn new ways to connect the physical quantities that *they already know*.

From a pedagogical point of view, one could hardly ask for a better internal simplicity out of a discipline. The present, very imperfect notes try to exploit this pedagogical potential.

Many current physics textbooks take a different approach. They introduce the students to a mixture of some notions and laws that are common to all physical phenomena, but together with others notions and laws that are specific to a particular physical field: the Newtonian mechanics of *point-particles*, with its intuition of momentum as "mass times velocity". In my opinion this approach has many grave drawbacks:

- The mixture of general and discipline-specific notions is often not clearly separated, and this lack of separation leads to confusion. Momentum as "mass times velocity" is not universal. "$\boldsymbol{F} = m\boldsymbol{a}$" breaks down in many physical phenomena. Then which law, if any, is still valid in its stead? The "law of action and reaction" is not universal either. Then can any similar law be universally extended? What happens to angular momentum and its balance then? If an electromagnetic field exerts a force on a mass, is the mass also exerting a force on the electromagnetic field? how can the latter sustain a force?
- This approach unnecessarily suppresses some useful intuitions that many students have from everyday life, such as the distinction between contact forces and body forces. The students will later have to recover these intuitions as they study more general phenomena.
- This particular mixture of notions and laws is today used almost only in textbooks; very little in physical and engineering applications. Students who will research and work in actual mechanics – buildings, bridges, aeroplanes, fluid flow, and generally extended bodies – will need to amend their *point-particle* intuitions to a *continuum* one. Students who will research and work in electromagnetics will need to amend their intuition of momentum as "mass times velocity". Students who will research in particle physics will also need to amend their intuition of momentum in several different ways. Students who will work in General Relativity will need all these amendments at once.

"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

The increasing importance of amending our Newtonian intuition of space and time was discussed in Chapter 2.

Newtonian point-mass mechanics is less and less used in astronomy as well. Ephemerides use post-Newtonian approximations of General Relativity (Park et al. 2021). NASA and the Jet Propulsion Laboratory[4] by default include relativistic effects (Moyer 2000) when they plan or calculate trajectories for Earth, Moon, and beyond. The same general-relativistic formulae are used to calculate and plan spacecraft dynamics, for navigating in cis-lunar[5] or geocentric space as well as for interplanetary missions: the same software is used for navigating in both regimes (Park & Chodas 2024).

- The students typically end up with the understanding that "every phenomenon is just a consequence of Newton's laws", of the second law in particular. This is clearly false: this law is just one out of six or seven that determine the evolution of physical phenomena.
- A fault, not of this approach per se, but of the way it is taught, is that students often remain confused about what this all-powerful second law *exactly* say, and about its precise mathematical expression. Is it "$\boldsymbol{F} = m\boldsymbol{a}$"? or is it "$\boldsymbol{F} = \mathrm{d}\boldsymbol{P}/\mathrm{d}t$"? Why do many extra added terms suddenly appear when this "second law" of Newton's is used in continuum mechanics? is it still the same law?
- Science and education have had the noble tradition of founding their teachings on the notions of the theories that proved to be most correct. That is how we got Newtonian mechanics and electromagnetics in our schools. Today we know that Newtonian mechanics and some of its intuitions are only approximate; whereas the predictions and explanations offered by General Relativity (and quantum theory) keep on being beautifully confirmed. It's time we continue our noble tradition and replace those Newtonian notions that are only approximate with more exact ones.

The student is, in other words, introduced to notions and a physical understanding that are fuzzy, are partially incorrect according to our present understanding of physics, and moreover cannot be used as-is, but will instead require revisions – some of which are quite drastic (I say this out of my own experience as a physics student, researcher, and teacher).

One might hear the argument that the teaching approach via Newtonian point-mass mechanics is closer to our "everyday intuition". But that's a

topsy-turvy argument. Our everyday intuition *comes from* that teaching approach. As an analogy, someone in the 16th century could have said that it's better to teach the geocentric model[6] of the solar system, than the heliocentric one, because the former is closer to everyday intuition. Yet our children today quickly develop a heliocentric intuition, simply because it's the one that enters our education from the very start. Similar arguments could be made for other concepts, such as energy or the electromagnetic field, that once were not part of everyday intuition, but today are.

Validity of the mathematical form of the balance laws in General Relativity

These notes state several times that the mathematical form of the equation for balance, for instance

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \Phi(t) + R(t) \tag{14.10}$$

or its integral form, is also valid in General Relativity. I would like to give a brief explanation, if not a proof, of this fact for those who are not familiar with General Relativity.

To be continued

The balance equations of the first six quantities are commonly used in numerical relativity for magneto-hydrodynamical problems. Here's an example reproduced from the textbook by Baumgarte & Shapiro 2010:

Box 5.1 The relativistic MHD equations

The coupled set of relativistic MHD equations can be written in conservative form as follows:

$$\partial_t \rho_* + \partial_j(\rho_* v^j) = 0, \tag{5.168}$$

$$\partial_t \tilde{S}_i + \partial_j(\alpha\sqrt{\gamma}\, T^j{}_i) = \frac{1}{2}\alpha\sqrt{\gamma}\, T^{ab} g_{ab,i}, \tag{5.169}$$

$$\partial_t \tilde{\tau} + \partial_i(\alpha^2\sqrt{\gamma}\, T^{0i} - \rho_* v^i) = s_{\tilde{\tau}}, \tag{5.170}$$

$$\partial_t \tilde{B}^i + \partial_j(v^j \tilde{B}^i - v^i \tilde{B}^j) = 0, \tag{5.171}$$

where we have the balances of matter, momentum, energy, and magnetic flux; the balance of angular momentum is satisfied implicitly by the symmetry of the four-stress tensor T^{ab}, and the balance of charge by the use of a specific constitutive relation. In another example, Komissarov 2005 merges together the balances of momentum and energy:

The evolution equations of ideal MHD include the continuity equation,

$$\partial_t(\alpha\sqrt{\gamma}\,\rho u^t) + \partial_i(\alpha\sqrt{\gamma}\,\rho u^i) = 0, \tag{2}$$

the energy–momentum equations,

$$\partial_t\left(\alpha\sqrt{\gamma}\,T^t_{\ \nu}\right) + \partial_i\left(\alpha\sqrt{\gamma}\,T^i_{\ \nu}\right) = \frac{1}{2}\partial_\nu(g_{\alpha\beta})T^{\alpha\beta}\alpha\sqrt{\gamma}, \tag{3}$$

and the induction equation,

$$\partial_t(B^i) + e^{ijk}\partial_j(E_k) = 0. \tag{4}$$

To be continued

URLs for chapter *Postface to the teacher*

4. https://www.jpl.nasa.gov
5. https://csps.aerospace.org/papers/cislunar-development-what-build-and-why
6. https://www.britannica.com/science/geocentric-model

Bibliography

> 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)

("de X" is listed under D, "van X" under V, and so on, regardless of national conventions.)

Oxford English Dictionary (2009), 2nd ed. Oxford University Press. First publ. 1857.

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.

Ashby, N. (2003): *Relativity in the global positioning system*. Living Rev. Relativity **6**, 1–42. DOI: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. First publ. 1989.

Batchelor, G. K. (2000): *An Introduction to Fluid Dynamics*, repr. (Cambridge University Press, Cambridge). First publ. 1967.

Baumgarte, T. W., Shapiro, S. L. (2010): *Numerical Relativity: Solving Einstein's Equations on the Computer*. (Cambridge University Press, Cambridge).

Bird, R. B., Stewart, W. E., Lightfoot, E. N., Klingenberg, D. J. (2015): *Introductory Transport Phenomena*. (Wiley, New York).

Biró, T. S. (2011): *Is There a Temperature?: Conceptual Challenges at High Energy, Acceleration and Complexity*. (Springer, New York). DOI:10.1007/978-1-4419-8041-0.

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.

Burke, W. L. (1987): *Applied Differential Geometry*, repr. (Cambridge University Press, Cambridge). DOI:10.1017/CBO9781139171786. First publ. 1985.

Burke, W. L. (1995): *Div, Grad, Curl Are Dead*. http://people.ucsc.edu/~rmont/papers/Burke_DivGradCurl.pdf, see also errata at https://www.ucolick.org/~burke/class/dgcaderr.html, and also http://www.ucolick.org/~burke.

Calvino, I. (1979): *Invisible Cities*. (Picador, London). Transl. by William Weaver. First publ. in Italian 1972.

Capitaine, N. (2010): *The astronomical reference systems in the framework of general relativity*. Lecture at the Institut d'Astrophysique de Paris (IAP), https://philippelefloch.org/wp-content/uploads/2010/12/2010-december-nicole-capitaine.pdf.

Chang, H. (2004): *Inventing Temperature: Measurement and Scientific Progress*. (Oxford University Press, New York). DOI:10.1093/0195171276.001.0001.

Chang, T. S., Haddad, G. N. (1971): *On dispersion and characteristic motions of temperature rate dependent materials*. Tech. rep. NASA-CR-1795. (NASA). https://ntrs.nasa.gov/citations/19710024301.

Chew, G. F. (1970): *Hadron bootstrap: triumph or frustration?* Phys. Today **23**[10], 23–28. DOI: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.

Daft Punk (2005a): *Around the World*. In: Daft Punk (2005b).

Daft Punk (2005b): *Human After All*. (Virgin, worldwide).

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.

Diosady, L., Murman, S., Carton de Wiart, C. (2018): *A higher-order space-time finite-element method for moving-body and fluid-structure interaction problems*. Tech. rep. ARC-E-DAA-TN58275. (NASA Ames Research Center, Moffett Field, USA). https://ntrs.nasa.gov/citations/20190030858, https://iccfd.org/iccfd10/proceedings.html.

Dirac, P. A. M. (1955): *Gauge-invariant formulation of quantum electrodynamics*. Can. J. Phys. **33**[11], 650–660. DOI:10.1139/p55-081.

Doyle, A. C. (1887): *A Study in Scarlet*. Repr. in Doyle (1998; 2014).

Doyle, A. C. (1998): *Camden House: The Complete Sherlock Holmes*. http://ignisart.com/camdenhouse/canon. First publ. 1887–1927.

Doyle, A. C. (2014): *The Complete Sherlock Holmes*. https://sherlock-holm.es. Version 3.1. First publ. 1887–1927.

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.

Einstein, A. (1905a): *Zur Elektrodynamik bewegter Körper*. Ann. Phys. (Berl.) **17**, 891–921. Transl. in Einstein (1989a) Doc. 23 pp. 140–171.

Einstein, A. (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.

Einstein, A. (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.

Einstein, A. (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).

Essmann, U., Träuble, H. (1971): *The magnetic structure of superconductors*. Sci. Am. **224**[3], 74–84. DOI:0.1038/scientificamerican0371-74.

Euler, L. (1761): *Principia motus fluidorum. Pars Prior*. Novi Comment. Acad. Sci. Petropolitanae **6**, 271–311, tab. IV. http://eulerarchive.maa.org, http://archive.org/details/novicommentarii14sssrgoog; repr. in *Opera Omnia* series 2, vol. 12, pp. 133–168. Probably written 1752. See also Euler (1770; 1771).

Euler, L. (1770): *Sectio secunda de principiis motus fluidorum*. Novi Comment. Acad. Sci. Petropolitanae **14/1**, 270–386, tab. V–VI. http://eulerarchive.maa.org, http://archive.org/details/novicommentarii05sssrgoog; repr. in *Opera Omnia* series 2, vol. 13, pp. 73–153. Written 1766. See also Euler (1761; 1771).

Euler, L. (1771): *Sectio tertia de motu fluidorum*. Novi Comment. Acad. Sci. Petropolitanae **15**, 219–360, tab. I–VI. http : / / eulerarchive . maa . org, urlhttp://archive.org/details/novicommentariac15impe; repr. in *Opera Omnia* series 2, vol. 13, pp. 154–261. Written 1766. See also Euler (1761; 1770).

Faraday, M. (1846): *Thoughts on ray-vibrations*. Philos. Mag. **28**[188], 345–350. DOI:10.1080/14786444608645431.

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.

Flügge, S., ed. (1960): *Handbuch der Physik: Band III/1: Prinzipien der klassischen Mechanik und Feldtheorie [Encyclopedia of Physics: Vol. III/1: Principles of Classical Mechanics and Field Theory]*. (Springer, Berlin). DOI:10.1007/978-3-642-45943-6.

Galilei, G. (1623): *Il Saggiatore*. http://www.liberliber.it/biblioteca/g/galilei/il_saggiatore/html/index.htm; parts transl. in Seeger (1966).

Gibbins, G., Haigh, J. D., Kato, S., Rose, F. G. (2021): *"Global and regional entropy production by radiation estimated from satellite observations": comments and reply*. J. Climate **34**[9], 3721–3731. DOI:10.1175/JCLI-D-20-0685.1, DOI:10.1175/JCLI-D-20-0950.1. See Kato, Rose (2020).

Grad, H. (1961): *The many faces of entropy*. Commun. Pure Appl. Math. **14**, 323–354. DOI:10.1002/cpa.3160140312.

Hughes, T. J. R., Franca, L. P., Mallet, M. (1986): *A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics*. Comput. Methods Appl. Mech. Eng. **54**[2], 223–234. DOI:10.1016/0045-7825(86)90127-1.

ISO (2009): *ISO 80000-1:2009: Quantities and units 1: General*. International Organization for Standardization.

ISO (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.

Kato, S., Rose, F. G. (2020): *Global and regional entropy production by radiation estimated from satellite observations*. J. Climate **3**[8], 2985–3000. DOI:10.1175/JCLI-D-19-0596.1. See also comments and reply in Gibbins, Haigh, Kato, Rose (2021).

Komissarov, S. S. (2005): *Observations of the Blandford-Znajek process and the magnetohydrodynamic Penrose process in computer simulations of black hole magnetospheres*. Mon. Not. R. Astron. Soc. **359**[3], 801–808. DOI:10.1111/j.1365-2966.2005.08974.x.

Kovalevsky, J., Seidelmann, P. K. (2004): *Fundamentals of Astrometry*. (Cambridge University Press, Cambridge). DOI: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. First publ. 2022.

Lorentz, H. A., Einstein, A., Minkowski, H., Weyl, H. (1952): *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, repr. (Dover, New York). https://archive.org/details/in.ernet.dli.2015.214561. With notes by A. Sommerfeld; transl. by W. Perrett and G. B. Jeffery, first publ. 1923.

MacKay, D. J. C. (2008): *Sustainable Energy – without the hot air*. (UIT, Cambridge). https://www.withouthotair.com.

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.

Maxwell (Clerk Maxwell), J. (1869): *Remarks on the mathematical classification of physical quantities*. Proc. London Math. Soc. **III**[34], 224–233. DOI:10.1112/plms/s1-3.1.224. Repr. in Maxwell (2010b) doc. XLVI pp. 257–266.

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.

Maxwell (Clerk Maxwell), J. (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.

Minkowski, H. (1908): *Space and time*. Address delivered at the 80th Assembly of German Natural Scientists and Physicians, Cologne. Transl. in Lorentz, Einstein, Minkowski, Weyl (1952), pp. 73–96.

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.

Moyer, T. D. (2000): *Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation*. (Jet Propulsion Lab, NASA). https://descanso.jpl.nasa.gov/monograph/series2_section.html, DOI:10.1002/0471728470.

Newton, I. (1726a): *Philosophiæ naturalis principia mathematica*, 1st ed. (Societatis Regiæ, London). https://archive.org/details/philosophiaenatu00newt_0, https://archive.org/details/philosophiaenat00newt. Transl. in Newton (1846; 1974).

Newton, I. (1726b): *Philosophiæ 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; 1974).

Newton, I. (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.

Newton, I. (1974): *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his system of the world. Vol. One: The Motion of Bodies; 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.

Park, R., Chodas, P. (2024). Personal communication. Feel free to enquire directly at https://www.jpl.nasa.gov.

Park, R. S., Folkner, W. M., Williams, J. G., Boggs, D. H. (2021): *The JPL planetary and lunar ephemerides DE440 and DE441*. Astron. J. **161**[3], 105. DOI:10.3847/1538-3881/abd414.

Parker, E. N. (1974a): *Hydraulic concentration of magnetic fields in the solar photosphere. II. Bernoulli effect*. Astrophys. J. **190**[2], 429–436. DOI:10.1086/152894. See also Parker (1974b).

Parker, E. N. (1974b): *Hydraulic concentration of magnetic fields in the solar photosphere. I. Turbulent pumping*. Astrophys. J. **189**[3], 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.

Samohýl, I. (1987): *Thermodynamics of Irreversible Processes in Fluid Mixtures: (Approached by Rational Thermodynamics)*. (Teubner, Leipzig).

Samohýl, I., Pekař, M. (2014): *The Thermodynamics of Linear Fluids and Fluid Mixtures*. (Springer, Cham). First published as Samohýl (1987). DOI:10.1007/978-3-319-02514-8.

Schieffer, G. (2013): *Aerodynamic and numerical flow modeling of elastic high lift configurations*. PhD thesis. (Aachen University, Aachen). https://publications.rwth-aachen.de/record/229479.

Seeger, R. J., ed. (1966): *Galileo Galilei, His life and His Works*. (Pergamon, Oxford).

Šilhavý, M. (1997): *The Mechanics and Thermodynamics of Continuous Media*. (Springer, Berlin). DOI:10.1007/978-3-662-03389-0.

Smith, G. (2024): *Newton's* Philosophiae Naturalis Principia Mathematica. In: Zalta, Nodelman (2024). https://plato.stanford.edu/archives/win2024/entries/newton-principia. First publ. 2007.

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.

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.

Truesdell III, C. A. (1966): *Six Lectures on Modern Natural Philosophy*. (Springer, Berlin). DOI:10.1007/978-3-662-29756-8.

Truesdell III, C. A. (1984): *Rational Thermodynamics*, 2nd ed. (Springer, New York). DOI:10.1007/978-1-4612-5206-1. First publ. 1969.

Truesdell III, C. A. (1987): *An Idiot's Fugitive Essays on Science: Methods, Criticism, Training, Circumstances*, 2nd pr., rev. and augmented. (Springer, New York). First publ. 1984.

Truesdell III, C. A., Toupin, R. A. (1960): *The Classical Field Theories*. In: Flügge (1960): I–VII, 226–902. With an appendix on invariants by Jerald LaVerne Ericksen. DOI:10.1007/978-3-642-45943-6_2.

Turon, A., Camanho, P. P., Costa, J., Davila, C. G. (2004): *An interface damage model for the simulation of delamination under variable-mode ratio in composite materials*. Tech. rep. NASA/TM-2004-213277. (NASA, Langley Research Center). https://ntrs.nasa.gov/citations/20040171493.

Wagner, W., Kretzschmar, H.-J. (2008): *International Steam Tables: Properties of Water and Steam Based on the Industrial Formulation IAPWS-IF97*, 2nd ed. (Springer, Berlin). First publ. 1998.

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.

Zalta, E. N., Nodelman, U., eds. (2024): *Stanford Encyclopedia of Philosophy*, continuously updated. (The Metaphysics Research Lab). https://plato.stanford.edu. First publ. 1995.