

Vector Calculus and Continuum Mechanics

Notes for a lecture course in the Department of Earth Sciences, University of Oxford

Professor Richard F. Katz

September 6, 2026

Contents

0	Background	6
0.1	Why Learn Vector Calculus & Continuum Mechanics?	6
0.2	What is Vector Calculus?	7
0.2.1	What is calculus?	7
0.2.2	What is a vector?	9
0.2.3	What is vector calculus?	12
1	Parameterised curves	14
1.1	Differentiating a parameterised curve	15
1.2	Arc length	17
1.3	Line integrals	18
1.3.1	Line integrals of scalar fields	18
1.3.2	Line integrals of vector fields	20
1.4	The gradient of a scalar	22
1.4.1	Line integrals of conservative vector fields	23
1.4.2	Application to mechanical energy conservation	25
2	Surfaces and Volumes	28
2.1	Area integrals	28
2.2	Aside on notation for multiple integrals	29
2.3	Evaluating area integrals	29
2.4	Changing coordinate systems	31
2.5	Aside on notation for scalars, vectors and tensors	33
2.6	2-D (plane) polar coordinates	33
2.7	Volume integrals	35
2.7.1	Volume integrals in spherical and cylindrical coordinates	36
2.7.2	Vector-valued volume integrals	39
2.8	Surface integrals	40
2.8.1	Parameterised surfaces	41
2.8.2	Integrals of scalar fields	42
2.8.3	Integrals of vector fields	43

3	Vector Operators	48
3.1	The gradient	48
3.2	The Nabla (or Del) operator	50
3.3	Aside on index notation, Einstein summation convention, dot products, and the Kronecker delta	50
3.4	The divergence	52
3.5	Aside on the Levi-Civita symbol, a notation for cross products	53
3.6	The curl	54
3.7	Mathematical properties of grad, div, and curl	56
3.8	The Laplacian operator	56
3.9	The Helmholtz decomposition	57
3.10	The streamfunction, streamlines and field lines	58
3.11	Orthogonal curvilinear coordinates	60
	3.11.1 Cartesian, cylindrical and spherical coordinates	61
	3.11.2 Grad	62
	3.11.3 Div, curl and Laplacian	63
4	Integral Theorems	67
4.1	Divergence Theorem	67
	4.1.1 The meaning of the divergence	68
	4.1.2 A proof(ish)	68
	4.1.3 An application of the Divergence Theorem	69
4.2	Stokes Theorem	71
	4.2.1 The meaning of the curl	71
	4.2.2 An application of Stokes Theorem	72
4.3	Gravity and Gauss' law	74
	4.3.1 Gauss' law	75
	4.3.2 Gravitational potential	77
5	Partial differential equations	78
5.1	The Laplace equation	79
5.2	The Poisson equation	79
5.3	Solution by separation of variables	83
6	Tensors	86
6.1	Why we need tensors: an example	86
6.2	Tensors are maps between vectors	87
6.3	Tensors are things that transform like tensors	88
	6.3.1 Transforming first-rank tensors (vectors)	89
	6.3.2 Transforming second-rank tensors (matrices)	90
	6.3.3 The rotation matrix	91
	6.3.4 Pardon me madame; is that a tensor?	93
6.4	Symmetric, antisymmetric, and isotropic tensors	93
6.5	Tensor fields	95

7	What is Continuum Mechanics?	99
7.1	What is mechanics?	99
7.2	What is a continuum?	101
7.3	What is rheology?	103
8	Newton's Second Law for a Continuum	104
8.1	Acceleration and the Lagrangian derivative	104
8.2	Inertial force per unit volume	106
8.3	Body force per unit volume	107
8.4	Surface force and the traction vector	107
8.5	The Cauchy stress tensor	109
8.6	Newton's second law for continua	112
8.7	Pressure and stress	113
9	Conservation of Mass	116
9.1	Conservation of mass for fluids	116
9.2	Conservation of mass for elastic solids	117
10	Rheological Models of Fluids and Solids	118
10.1	Elastic and viscous rheological models (scalar version)	118
10.1.1	Elastic deformation	118
10.1.2	Viscous deformation	119
10.1.3	Viscoelastic deformation	120
10.2	Strain and strain-rate tensors	123
10.2.1	Local analysis of the velocity field	124
10.2.2	Strain and rotation rates	125
10.3	Elastic and viscous rheological models (tensor version)	127
10.3.1	The Hookean elastic solid	128
10.3.2	The incompressible Newtonian fluid	129
10.3.3	The upper-convected Maxwell viscoelastic model	130
10.4	Principal components of stress and strain (rate)	130
11	Momentum Conservation	134
11.1	Elastic deformation	134
11.2	Viscous deformation	135
12	Incompressible Flows	137
12.1	Static and dynamic pressure	137
12.2	Scaling analysis of inertia and viscosity	138
12.3	Boundary conditions	140
12.3.1	Kinematic conditions	140
12.3.2	Dynamic conditions	140
12.4	Parallel viscous flows	141
12.4.1	Two-dimensional, steady, parallel flow	142
12.4.2	Flow between parallel plates	142
12.4.3	Flow down an inclined plane	144
12.5	Inviscid flows	146

12.6	Rotating flows	148
12.6.1	The vorticity equation	148
12.6.2	Rotating flows	151
12.6.3	Geostrophic balance	153
13	Elastodynamics and elastic flexure	157
13.1	(Seismic) wave equations	157
13.1.1	P-wave equation	157
13.1.2	S-wave equation	158
13.2	Plane waves	158
13.3	Spherical waves	159
13.4	Elastic stretching of a rod	160
13.5	Elastic flexure of plates	162
13.5.1	Plane stress	162
13.5.2	Cylindrical bending of a plate	162
13.5.3	Flexural–isostatic compensation of a floating plate	166

Acknowledgements & sources

A course in vector calculus and continuum mechanics is not written from scratch. These are classical subjects that have been taught by many people over many years. Whilst wishing to provide a course that is adapted to your needs (students of the Department of Earth Sciences, University of Oxford), I also wish to draw on the pedagogical materials of outstanding teachers. This has always been possible with the use of textbooks, but online course notes make it even easier.

The notes before you draw heavily from the following sources:

DT [Course notes by Professor David Tong on Vector Calculus](#) from the University of Cambridge Part IA Mathematical Tripos.

CT [Course notes by Professor Caroline Terquem on Fluid Mechanics](#) from the University of Oxford degree in Physics.

LM [Introduction to the Mechanics of a Continuous Medium](#) by Laurence Malvern. A comprehensive treatment of continua and a classic reference.

GW [Understanding Fluid Flow](#) by Grae Worster. An introduction based on physical examples and a didactic, minimalistic approach to the theory.

PK [Fluid Mechanics](#) by Pijush Kundu. A nice textbook.

PS [Introduction to Seismology](#) by Peter Shearer.

TS [Geodynamics](#) by Turcotte & Schubert. About applications of continuum mechanics to the solid Earth.

When a figure is re-used from these sources, credit is given in the figure caption via the relevant two-letter code. However, I also acknowledge these sources more broadly. The exposition of vector calculus in these notes closely follows that by Professor David Tong.

Thanks to Hannah Sanderson, Marin Mag and Justin Leung for proofreading. If you find (typographical) errors in these notes and/or if you find them useful, please email richard.katz@earth.ox.ac.uk

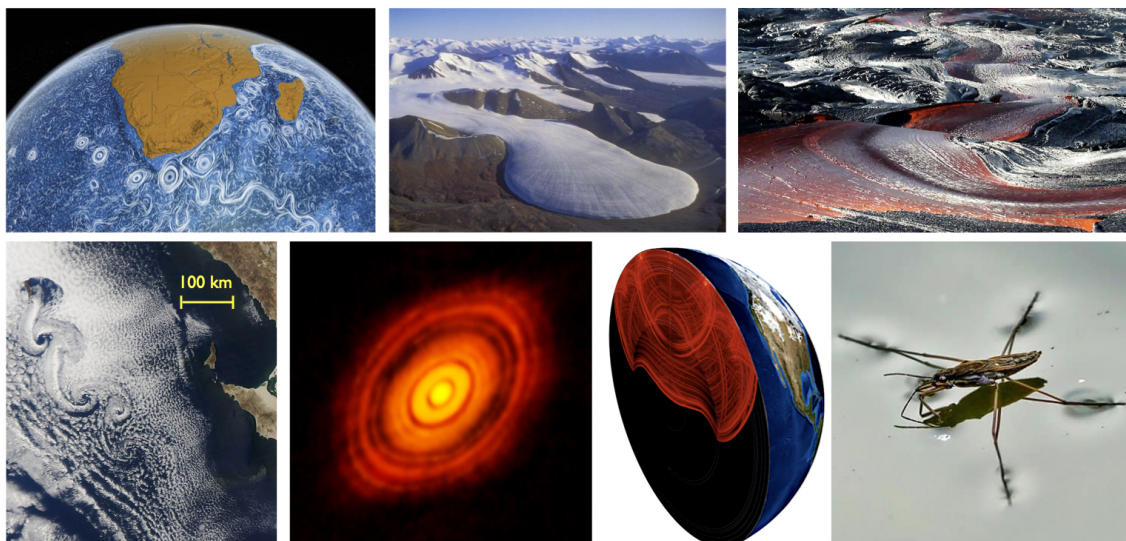


Figure 0.1: Geophysical continua. Clockwise from top left: simulated ocean currents (source: NASA), a piedmont glacier, lava flows, a water-strider, simulated seismic waves (source: AxisEM), a protoplanetary disk (source: ALMA), an atmospheric vortex street. Uncopyrighted images.

0 Background

0.1 Why Learn Vector Calculus & Continuum Mechanics?

How does the Universe work? How does Earth work? Why does nature do what it does and not something different? What will happen to our environment in the future? These questions motivate natural science, which is first and foremost an empirical venture. But observations lead to insight, and insight leads to conceptual models, which are qualitative answers to questions like those above.

Why not stop there, with qualitative answers? One reason is that competing conceptual models might arise from the same observations. To test which, if any, are correct, it is usually necessary to have quantitative predictions. These come from mathematical descriptions of the physical processes. *The mathematical descriptions that enable prediction require vector calculus as a basic grammatical framework.*

Three-dimensional space and time are evidently independent variables; forces have direction and magnitude; heat fluxes are driven by temperature gradients; magnets create fields of force; the ocean and atmosphere circulate; waves spread radially from their source. A quantitative physical theory needs to accommodate these and other characteristics of the physical world; it needs a language furnished with symbols and rules to describe them.

The great theoretical frameworks of physics are succinctly and elegantly expressed with vector calculus. What is succinct and elegant about these expressions? A huge diversity of geometry and physics, at scales from the atomic to the whole universe, can be described with short statements using a small set of grammatical rules (in comparison with written language, for example). In this course you will learn many of those rules, providing you with a foundation of mathematical knowledge to

study theories from electromagnetism to general relativity and many others.

A theory of great use to physical scientists (including geo- and planetary) concerns how the matter within extended regions moves and interacts as a consequence (and a cause) of applied forces. This theory is known as continuum mechanics. It allows us to make mathematical models (and quantitative predictions) of earthquakes and seismic waves, mantle convection and plate tectonics, ocean circulation, tides, atmospheric flow and weather, the geodynamo and magnetic field, flow of groundwater and magma, the motion of ice sheets, locomotion of organisms, volcanic eruptions and more. Continuum mechanics underpins climate models as the dominant basis for transport of heat. In this course you will learn the foundations of continuum mechanics, allowing you to specialise in any of the areas above.

0.2 What is Vector Calculus?

In this section we provide some informal definitions (and examples) of the most important mathematical objects that we'll be using in the course. These are not intended to be mathematically rigorous. Rather they are intended to be helpful means of connecting the abstract (mathematics) with the concrete (motion, physics).

0.2.1 What is calculus?

Calculus is a set of operations that are performed on *functions*. The simplest type of function is of the form

$$f : \mathbb{R} \rightarrow \mathbb{R}. \quad (0.1)$$

This statement means that the function f is a mapping from a *scalar* to another scalar ($\mathbb{R}$ symbolises a real number). A scalar is a quantity that can be represented with a single, real number expressing its magnitude. Mass, for example, is represented in terms of a scalar.

Example 0.1 of a function

As an example of (0.1), consider the function

$$f : t \rightarrow t^2.$$

Here the scalar t is mapped into the scalar t^2 . If the input to f is 0.5 the output is 0.25, but we can choose any real number as an input.

We will typically use a more familiar notation for functions, $f(t)$, but remember that they are maps from one space to another space.

The most basic calculus operation that we can do on a function f is *differentiation*. For a function of one scalar variable, say $f(t)$, the definition of the derivative is

$$\frac{df}{dt} = \lim_{\delta t \rightarrow 0} \frac{f(t + \delta t) - f(t)}{\delta t}. \quad (0.2)$$

This should be familiar, but it is worth bearing it in mind. It states that the derivative of a function with respect to its argument is the slope of the function in

the direction of a positive change in the argument. Most functions have derivatives that can be written analytically, but we will not review these specifics.

Example 0.2 of differentiation

As an example of (0.2), consider the function

$$g : t \rightarrow t^3/3.$$

The derivative of this function is

$$\frac{dg}{dt} : t \rightarrow t^2.$$

From this we can see that dg/dt is, itself, a function.

There are a few properties that we'll (usually) require of functions:

- That they are finite or zero. This means that they do not evaluate to infinity for any value of their input.
- That they are continuous. This means that an arbitrarily small change in their input maps to an arbitrarily small change in their output.
- That they have (at least) one continuous derivative. This means that they have no changes in slope that are discontinuous.

The other important operation in calculus is *integration*. We can think of integration in the geometric sense, as the area under a curve between two end-points, or we can think of it as the inverse of differentiation. In this latter case,

$$f(t) = \frac{d}{dt} \left(\int f dt \right). \tag{0.3}$$

This equation simply means that whatever $\int f dt$ does to the function $f(t)$, the derivative d/dt undoes it. Caution is needed here, because the derivative of any constant is zero! Therefore, when integrating, we must always add an arbitrary constant to the result. The value of this constant can be determined if we know the value of the integration at some point. Upon re-differentiation, this constant contributes zero to the result.

Example 0.3 of integration

As an example of (0.3), consider the function

$$f : t \rightarrow t^2.$$

The integral of this function is

$$\int f dt : t \rightarrow t^3/3 + c,$$

where c is the constant of integration. We can confirm our result if we notice that

$$\frac{d}{dt} \left(\int f dt \right) : t \rightarrow t^2,$$

which is the same as the mapping f itself.

As with differentiation, we will not review the integration of particular functions here.

In both differentiation and integration, we have worked with *infinitesimal* quantities, denoted as dt , for example. Discovery of the infinitesimal was the conceptual innovation that enabled the development of calculus. Yet it often goes unexplained and, hence, is not well understood. So please take careful note: *an infinitesimal is an arbitrarily small (but non-zero!) change in a quantity*. Infinitesimals can only appear in equations under certain circumstances¹:

- As a ratio, representing the derivative, e.g., $\frac{df}{dt}$;
- In combination with a $\int$ symbol, representing integration, e.g., $\int f dt$;
- In an equation where each term contains an infinitesimal quantity, e.g., $df = (3 + m)dt$. (This is called a *differential form*).

Note also that infinitesimals may be dimensional or dimensionless, just like other symbols. For example, if t is measured in seconds, then dt is also measured in seconds. We shall distinguish between the infinitesimal change dt and the finite-but-small change δt , though these concepts are obviously related and often connected in calculations.

0.2.2 What is a vector?

We now need to contrast the mathematical object called a *vector* with the scalar, that we met above. Whereas a scalar has only a magnitude, *a vector has magnitude and direction*. And when we think about direction, we immediately require that it be defined within a physical space. That space is at least one-dimensional, but more typically two- or three-dimensional. So whereas a scalar lives in the space $\mathbb{R}$, a vector lives in the space $\mathbb{R}^n$, where n is the dimension of the space.

¹Mathematicians have stricter rules about this than (geo)physicists.

Example 0.4 of vectors

Consider the space defined by a room with walls that form a cube (floor and ceiling are considered ‘walls’ here). A bee is flying around within the room. Define a coordinate system with an origin at a point where three walls come together, and axes directed from the origin along the lines where two walls come together. At an instant in time, the bee has a position vector $\mathbf{x}$ and a velocity vector $\mathbf{v}$ in that coordinate system. Both of these vectors are in a space $\mathbb{R}^3$.

From this example, we recognise two types of vectors. The position vector is an example of a vector that is rooted at the origin and points away from it. The velocity vector is an example of a vector that originates at the end of the position vector (i.e., at the position of the object that has a velocity — the bee in our example).

In a coordinate system, we can write a vector as a list of numbers. For a 3-D coordinate system as in the example above, we could write

$$\mathbf{x} = (x_1, x_2, x_3), \quad (0.4)$$

where x_i is the distance along each of the coordinate directions, measured in, say, metres. The velocity vector could be written in the same way,

$$\mathbf{v} = (v_1, v_2, v_3). \quad (0.5)$$

Each entry in the velocity vector v_i represents the component of velocity along the coordinate direction.

Review point about vector properties

The two properties of a vector are its magnitude and its direction. We can represent these mathematically as follows (taking the position vector as example),

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}, \quad \text{vector magnitude}, \quad (0.6)$$

$$\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|, \quad \text{vector direction}. \quad (0.7)$$

The first of these represents the length of the vector as the square root of the sum of the components squared; the proof of this is geometrical. The second represents the direction in terms of a unit vector: a vector with length 1 and no physical dimensions that points in the same direction as the original vector. The ‘hat’ on the unit vector means that it has unit length.

Exercise 0.1 on unit vectors

Construct unit vectors for the following vectors:

a. $\mathbf{x} = (2, 0, 0)$,

b. $\mathbf{v} = (3, 3, 0)$,

c. $\mathbf{u} = (1, 1, 1)$.

We can also think about the direction of a vector in terms of the angle that it makes with something else. An obvious set of reference directions are the coordinate axes. We can define *basis* vectors as the unit vectors that point along the coordinate axes. We refer to the basis vectors as $\hat{\mathbf{e}}_i$, where i is an index representing numbers from 1 to the spatial dimension n (we could have equivalently used a different symbol for the index such as j). In a Cartesian coordinate system, the basis vectors are

$$\begin{aligned}\hat{\mathbf{e}}_1 &= (1, 0, 0), \\ \hat{\mathbf{e}}_2 &= (0, 1, 0), \\ \hat{\mathbf{e}}_3 &= (0, 0, 1).\end{aligned}$$

Review point about basis vectors

Basis vectors are *orthonormal* (orthogonal and normalised) if

$$\delta_{ij} = \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \quad (0.8)$$

This is the definition of the Kronecker delta, δ_{ij} , an important symbol in this course.

Basis vectors are different in other coordinate systems but, for the purposes of this course, are always orthonormal².

Recall that the dot product of two vectors is defined in terms of the angle θ between them by

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta.$$

It follows that

$$\theta = \cos^{-1}(\hat{\mathbf{x}} \cdot \hat{\mathbf{e}}_i)$$

is the angle between the vector $\mathbf{x}$ and the coordinate axis with basis vector $\hat{\mathbf{e}}_i$.

Exercise 0.2 on vector direction

Use the cross product and the dot product to find an expression for the angle between the vector $\mathbf{x}$ and the plane formed by vectors $\mathbf{a}$ and $\mathbf{b}$. Note that the cross product satisfies

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin \theta,$$

where again θ is the angle between the two vectors.

For the vector $\mathbf{x} = (1, 1, 1)$, calculate

- The angle with each of the three Cartesian basis vectors, $\hat{\mathbf{e}}_i$;
- The angle with the x - y plane.

²There are coordinate systems where the axes are not orthogonal, but they are exotic and hard to work with. We will not consider them.

If you find the meaning (geometric and mathematical) of the dot product and the cross product hazy in your mind, it is essential that you review these concepts now!

0.2.3 What is vector calculus?

Vector calculus is the application of the operations of calculus (differentiation and integration) to functions that map from and to the space of vectors. Let's explore this definition by first examining such functions.

Recalling that a function is a map, let's consider a map of the form

$$f : \mathbb{R} \rightarrow \mathbb{R}^n. \quad (0.9)$$

This maps from a scalar to a vector in an n -dimensional space. For example, we can denote the points along a curve in $\mathbb{R}^n$ as $\mathbf{x}$ and we can denote a position in $\mathbb{R}$ as t . Then our map becomes

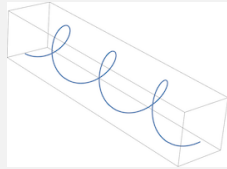
$$f : t \rightarrow \mathbf{x}(t) \quad (0.10)$$

If we regard t as time, this is a map from time to position. We can regard $\mathbf{x}(t)$ as the map itself.

Example 0.5 of a map of $\mathbb{R} \rightarrow \mathbb{R}^3$

Let's again consider a bee flying around a room. Over some time interval, the bee follows the trajectory

$$\mathbf{x}(t) = (\cos t, \sin t, t), \quad (0.11)$$



which is a helix.

What a clever bee.

Another important type of map is of the form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}. \quad (0.12)$$

This maps every point in an n -dimensional space to a scalar value. For example, we can denote the temperature within a room as T and the position in the room as $\mathbf{x}$. Then our map becomes

$$f : \mathbf{x} \rightarrow T(\mathbf{x}). \quad (0.13)$$

The temperature is known as a *scalar field*.

The most general mapping that we'll consider here is of a *vector field*. This mapping has the form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^n. \quad (0.14)$$

This maps every point in an n -dimensional space to an n -dimensional vector. An example is the velocity field $\mathbf{v}$ of the air as it circulates within a room

$$f : \mathbf{x} \rightarrow \mathbf{v}(\mathbf{x}). \quad (0.15)$$

Note that this velocity vector field is defined at all points in space whereas above, in the example of a flying bee, the velocity vector was defined only at the instantaneous position of the bee.

In much of the course we will be focused on vector fields and the physical equations that govern them. It is important to be able to visualise vector fields.

Exercise 0.3 on vector fields

Consider the two-dimensional field of two-dimensional vectors

$$\mathbf{V}(\mathbf{x}) = E(-x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y),$$

for some constant E , where $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y$ are basis vectors in the x and y directions. Within the domain $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$, sketch this field by hand or use plotting software to display it.

With this expanded conception of functions and mappings, we now return to our informal definition of vector calculus: Vector calculus is the application of the operations of calculus (differentiation and integration) to functions that map from and to the space of vectors. Let's dive in.

Summary of §0

- For our purposes, calculus means differentiation and integration of functions.
- Infinitesimal quantities like dt are arbitrarily small but not zero. Their use in equations is governed by certain rules.
- Functions are maps from one space to another. Examples are maps from scalar to scalar as in $f(t)$, from scalar to vector as in $\mathbf{x}(t)$, from vector to scalar as in $T(\mathbf{x})$, or from vector to vector as in $\mathbf{v}(\mathbf{x})$.
- Vector calculus means the application of differentiation and integration to functions on vector spaces.

1 Parameterised curves

In this section we consider mappings of the form

$$f : \mathbb{R} \rightarrow \mathbb{R}^3, \quad (1.1)$$

which correspond to *parameterised curves*. The curve is a set of points in $\mathbb{R}^n$ (where $n = 3$ for the usual 3 space dimensions) and the parameter is in $\mathbb{R}$. An instance of this is the position of a moving object as a function of time. We can represent the position as $\mathbf{x}(t)$, where t is the parameter. The curve can be expressed in an orthonormal basis as

$$\mathbf{x} = \sum_{i=1}^n x_i \hat{\mathbf{e}}_i, \quad (1.2)$$

$$= (x_1, \dots, x_n). \quad (1.3)$$

Usually we'll work with $n = 3$ — three-dimensional space.

Example 1.1 of a parameterised curve

Consider a curve (mapping from $\mathbb{R} \rightarrow \mathbb{R}^3$) of the form

$$\mathbf{x}(t) = (at, bt^2, 0).$$

This is a parabola in the $z = 0$ plane. We can eliminate the parameter t to obtain the usual form of the parabola, $y = (b/a^2)x^2$ with $z = 0$.

Sometimes we are given a curve that is not parameterised and we wish to introduce a parameterisation. There is no guaranteed method for doing this, nor is there a unique solution. A good approach is to look for a variable that does not pass through any values more than once. We say that the curve is *monotonic* in that variable. Then set that variable to be a function of time and use this function to eliminate the monotonic variable.

Example 1.2 of parameterising a curve

Consider a curve of the form $y = cx^2 + d$, where c and d are constants. Notice that x is monotonic while y is not (y takes the same value at x_0 and $-x_0$, for example). Now parameterise in terms of t by *choosing* $x = t$; substitute to obtain $y = ct^2 + d$. The curve can now be written as the map from $\mathbb{R} \rightarrow \mathbb{R}^2$ given by

$$\mathbf{x}(t) = (t, ct^2 + d).$$

Sometimes we need to parameterise a closed loop, where there is no monotonic variable. For example, exercise 1.1 provides the formula for a circle with unit radius, centred at the origin. The hint suggests the use of a periodic function of t to parameterise x . Think about why this is a good choice: starting from $(1, 0)$ and travelling around the circle anticlockwise, we move from positive x to negative x and then back to positive x and our starting point. For closed loops such as circles

and ellipses, it is always best to try to parameterise the curve in terms of sinusoidal functions of t . An ellipse is worked in example 1.3.

Exercise 1.1 on parameterising a curve

Parameterise the curve

$$x^2 + y^2 = 1.$$

Hint: try $x = \cos t$.

Example 1.3 of parameterising an ellipse

Consider the curve $2x^2 - 4x + 3y^2 = 4$. We can manipulate this algebraically to put it into the form of an ellipse,

$$\frac{(x-1)^2}{3} + \frac{y^2}{2} = 1.$$

Then, to parameterise, we can try $x-1 = \sqrt{3}\cos t$. Substituting this into the equation, we can determine that $y = \sqrt{2}\sin t$. Hence our parameterised ellipse is $\mathbf{x}(t) = (1 + \sqrt{3}\cos t, \sqrt{2}\sin t)$.

1.1 Differentiating a parameterised curve

It is straightforward to use Taylor expansion to investigate a parameterised curve over a small step in the parameter δt ,

$$\begin{aligned}\mathbf{x}(t + \delta t) &= \mathbf{x}(t) + \frac{d\mathbf{x}}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{x}}{dt^2}\delta t^2 + \dots, \\ &= \mathbf{x}(t) + \frac{d\mathbf{x}}{dt}\delta t + \mathcal{O}(\delta t^2),\end{aligned}\tag{1.4}$$

where $\mathcal{O}(\delta t^2)$ means that all additional terms are of a size proportional to δt^2 or smaller as $\delta t \rightarrow 0$.

Review point about the meaning of the derivative

Rearranging equation (1.4) gives us a definition for the derivative of a parameterised curve,

$$\frac{d\mathbf{x}}{dt} = \lim_{\delta t \rightarrow 0} \left[\frac{\mathbf{x}(t + \delta t) - \mathbf{x}(t)}{\delta t} - \mathcal{O}(\delta t) \right].\tag{1.5}$$

We will sometimes use the notation $\dot{\mathbf{x}}$ interchangeably with $d\mathbf{x}/dt$.

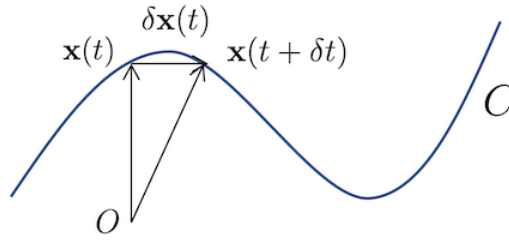


Figure 1.1: The tangent vector is the derivative of the curve C , obtained in the limit of $\delta t \rightarrow 0$. [Source: DT]

Example 1.4 of differentiating a parameterised curve

Recall the bee flying around a room that follows the parameterised curve

$$\mathbf{x}(t) = (\cos t, \sin t, t). \quad (1.6)$$

We can compute the bee's velocity as

$$\mathbf{v}(t) = \frac{d\mathbf{x}}{dt} = (-\sin t, \cos t, 1). \quad (1.7)$$

The derivative of a parameterised curve has an important meaning: it is the *tangent vector*. This is evident by examining figure 1.1 and imagining $\delta t \rightarrow 0$. Note, however, that the choice of parameterisation can affect the tangent vector. We can demonstrate this by considering the curve $y = x$ and choosing the parameterisation $x = t^3/3$. We obtain $\mathbf{x} = (t^3, t^3)/3$ and hence $\dot{\mathbf{x}} = (t^2, t^2)$. The problem is that this tangent vector vanishes at $t = 0$, despite the fact that the curve clearly has a tangent there.

Exercise 1.2 on parameterising a curve

Find a better parameterisation of the curve $y = x$ and show that its tangent is non-zero for all x .

Now the obligatory identities, which can be proven but probably won't provoke any doubts. Consider a scalar function $f(t)$ and two vector functions $\mathbf{h}(t)$ and $\mathbf{g}(t)$, all of which are continuous and differentiable

$$\frac{d}{dt}(f\mathbf{g}) = \frac{df}{dt}\mathbf{g} + f\frac{d\mathbf{g}}{dt}, \quad (1.8)$$

$$\frac{d}{dt}(\mathbf{g} \cdot \mathbf{h}) = \frac{d\mathbf{g}}{dt} \cdot \mathbf{h} + \mathbf{g} \cdot \frac{d\mathbf{h}}{dt}, \quad (1.9)$$

$$\frac{d}{dt}(\mathbf{g} \times \mathbf{h}) = \frac{d\mathbf{g}}{dt} \times \mathbf{h} + \mathbf{g} \times \frac{d\mathbf{h}}{dt}. \quad (1.10)$$

Care must be taken with the last of these because the order of entries in the cross product affects the result: $\mathbf{g} \times \mathbf{h} = -\mathbf{h} \times \mathbf{g}$.

1.2 Arc length

The tangent vector is useful for computing the arc length of a curve. We use s to represent arc length and δs to represent a finite increment of the same. By examination of figure 1.1, we can infer that

$$\begin{aligned}\delta s &= |\delta \mathbf{x}| + \mathcal{O}(|\delta \mathbf{x}|^2), \\ &= |\dot{\mathbf{x}} \delta t| + \mathcal{O}(\delta t^2),\end{aligned}\tag{1.11}$$

where we have used our Taylor expansion (1.4) to obtain the second line. Dividing through by δt and allowing this interval to become arbitrarily small we obtain

$$\frac{ds}{dt} = |\dot{\mathbf{x}}|.\tag{1.12}$$

This equation states that the speed at which we move along the arc length of a parameterised curve is given by the magnitude of the tangent vector.

Example 1.5 of speed along a curve

Recall the bee flying along a helix with a velocity

$$\dot{\mathbf{x}}(t) = (-\sin t, \cos t, 1).\tag{1.13}$$

This bee has a speed $|\dot{\mathbf{x}}| = \sqrt{\sin^2 t + \cos^2 t + 1^2} = \sqrt{2}$.

The equation (1.12) for the rate of arc length can be integrated to give the arc length. This calculation requires a starting point to be prescribed in terms of the curve parameter. For a curve parameterised in terms of t , we start our integration at some point t_0 and end it at a point that we can choose later,

$$s(t) = \int_{t_0}^t |\dot{\mathbf{x}}(t')| dt'.\tag{1.14}$$

In this equation, we have introduced a dummy variable of integration t' to distinguish it from the limit of integration t .

Example 1.6 of arc length along a curve

Coming back to our lovely little bee, we want to calculate the distance of its flight along the helix over the time interval $[0, 2\pi]$. (N.B. We are using dimensionless symbols here for convenience). Using $|\dot{\mathbf{x}}| = \sqrt{2}$, we obtain $s(t) = \sqrt{2}t$. Hence, after 2π , it has travelled a distance of $\sqrt{2}\pi$. It has gone from $\mathbf{x}(0) = (1, 0, 0)$ to $\mathbf{x}(2\pi) = (1, 0, 2\pi)$. The linear distance between these two points (“as the crow flies?!”) is 2π , which is obviously less than the arc length.

Now you might be wondering: do different parameterisations of a curve give different arc lengths? One should hope they don’t! But let’s find out. Suppose that a curve $y(x)$ is parameterised in two ways: $\mathbf{x}(t)$ and $\mathbf{x}(\tau)$. We note that t and τ are not equal, but can be related via x to give $t(\tau)$ or $\tau(t)$. To compute the arc length in terms of τ , we will use

$$dt = \frac{dt}{d\tau} d\tau \quad \text{and} \quad \dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} = \frac{d\mathbf{x}}{d\tau} \frac{d\tau}{dt}.\tag{1.15}$$

Starting with (1.14), substituting and adjusting the limits of integration (i.e., $\tau_0 = \tau(t_0)$), and assuming that $t(\tau)$ increases with τ , so that $dt/d\tau > 0$, we obtain

$$s = \int_{t_0}^t |\dot{\mathbf{x}}(t')| dt' = \int_{\tau_0}^{\tau} \left| \frac{d\mathbf{x}}{d\tau'} \frac{d\tau'}{dt} \right| \frac{dt}{d\tau'} d\tau', \quad (1.16)$$

$$= \int_{\tau_0}^{\tau} \left| \frac{d\mathbf{x}}{d\tau'} \right| d\tau'. \quad (1.17)$$

This shows us that arc length computed with τ is the same as the arc length computed with t ! In other words, *equation (1.14) for arc length is independent of the choice of parameterisation.*

Exercise 1.3 on arc length for different parameterisations

Consider the curve $y = x^2$ and compute the arc length from $(-1, 1)$ to $(1, 1)$ using two different parameterisations, $x = t - 1$ and $x = \tau^3 - 1$.

Because the arc length is independent of the parameterisation, we can say that it is inherent to the curve. Therefore it provides a natural parameterisation for the curve. We can think of a curve in terms of $\mathbf{x}(s)$ with tangent vector $d\mathbf{x}/ds$. Then, since $|d\mathbf{x}| = ds$, the tangent vector has the convenient property that it is a unit vector, $|d\mathbf{x}/ds| = 1$.

1.3 Line integrals

Lines are one-dimensional objects that are embedded in a higher dimensional space. A line on a piece of paper is a 1-D object (the line) embedded in a 2-D object (the paper). The trajectory of a bee flying through the air is a 1-D object (the trajectory) embedded in a 3-D object (the atmosphere).

As the bee flies through the air along $\mathbf{x}(t)$, it feels different temperatures $T(\mathbf{x})$ and it is affected by the flow field of the air $\mathbf{v}(\mathbf{x})$. We might like to calculate the cumulative effect of these changing conditions by integrating along the line. This is a *line integral*.

More formally: given a curve C in $\mathbb{R}^n$ and a field defined everywhere in $\mathbb{R}^n$, we can integrate the field along the curve. Our approach differs for scalar fields and vector fields, so we treat these in turn.

1.3.1 Line integrals of scalar fields

Recall that a scalar field is defined by a map

$$\phi : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (1.18)$$

that is, ϕ is a scalar function of $\mathbf{x}$; we write this as $\phi(\mathbf{x})$.

Given a curve C parameterised in terms of, say, t , we could consider $\phi(\mathbf{x}(t))$ and integrate this along C . However, our result would depend on the choice of parameterisation, which is awkward. We'd like to avoid this and so we instead

parameterise in terms of arc length s . Then we can write the line integral of ϕ along C from s_a to s_b (with $s_a < s_b$) as

$$\int_C \phi ds = \int_{s_a}^{s_b} \phi(\mathbf{x}(s)) ds. \quad (1.19)$$

In principle, this integral can be evaluated for suitably defined data. It is conventionally set up so that the arc length along the curve is positive — and hence the result is independent of the direction of integration.

The problem with (1.19) is that we usually have C in terms of a parameterisation that is not arc length — suppose it is time, t . There is some relationship between s and t and hence we can write $ds = (ds/dt)dt$. Using this in (1.19),

$$\int_C \phi ds = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) \frac{ds}{dt} dt,$$

where t_a and t_b are the times such that $s_a = s(t_a)$ and $s_b = s(t_b)$. Now recall that $ds/dt = |\dot{\mathbf{x}}|$, the magnitude of the tangent vector (the speed)! Substituting, we have

$$\int_C \phi ds = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) |\dot{\mathbf{x}}| dt. \quad (1.20)$$

This integral contains the length of the tangent vector within the integrand. It is derived from (1.19) and hence it ensures that the result is independent of the parameterisation of C .

In evaluating (1.20), we should be careful of the sign. If $t_b < t_a$, our parameterisation traverses the curve from s_b to s_a . To ensure that the result of integration is independent of direction, we should take the tangent to go in the opposite direction, $ds/dt = -|\dot{\mathbf{x}}|$.

Example 1.7 of a line integral of a scalar field

Let's return to our helix-flying bee and calculate its exposure to temperature as it moves through the room. We will take all symbols as dimensionless in this example. Recall that the bee follows a path

$$\mathbf{x}(t) = (\cos t, \sin t, t).$$

The temperature in the room varies in space as $T(\mathbf{x}) = ay + bz$, with a, b constant. The integrated temperature along the bee's path starting at $t = 0$ is

$$\begin{aligned} \int_0^t T(\mathbf{x}(t')) |\dot{\mathbf{x}}| dt' &= \int_0^t (a \sin t' + bt') |\dot{\mathbf{x}}| dt', \\ &= \int_0^t (a \sin t' + bt') \sqrt{2} dt', \end{aligned}$$

where we have used our equation for $T(\mathbf{x})$ and the value of $|\dot{\mathbf{x}}|$ determined above. We obtain

$$\sqrt{2} [a(1 - \cos t) + bt^2/2].$$

We can compute the temperature averaged over arc length $\bar{T}$ along the helix from $t = 0$ by finding the arc length using (1.14); we obtain $s(t) = \sqrt{2}t$ and hence, by dividing, $\bar{T}(t) = a(1 - \cos t)/t + bt/2$. Care must be taken in evaluating this at $t = 0$.

1.3.2 Line integrals of vector fields

Recall that a vector field is defined by a map

$$\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \tag{1.21}$$

that is, $\mathbf{F}$ is a vector function of $\mathbf{x}$; we write this as $\mathbf{F}(\mathbf{x})$.

In this context, a line integral of $\mathbf{F}$ can have (at least) two meanings. In the first meaning, we simply integrate the vector field along the arc length of the curve,

$$\int_C \mathbf{F} ds = \int_{s_a}^{s_b} \mathbf{F}(\mathbf{x}(s)) ds. \tag{1.22}$$

Note that the result of this integration is a *vector*, and that we would perform the integration component-wise. We could achieve this by using the change of coordinates from s to some other parameterisation t that we used in deriving (1.20). An integral of this type could be used to find, for example, the average wind velocity along the flight path of an airplane (or a bee).

The second meaning of the line integral of $\mathbf{F}$ is more commonly the relevant one. This is where we are integrating the projection of $\mathbf{F}$ onto the tangent of the curve,

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_{t_a}^{t_b} \mathbf{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}} dt. \tag{1.23}$$

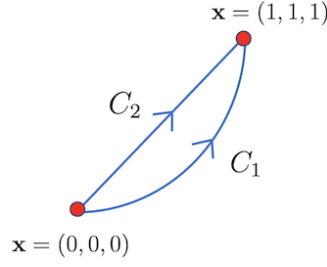


Figure 1.2: Two curves, sharing the same end-points, used for line integrals. Source: DT.

In this equation, we are evaluating the vector field $\mathbf{F}$ along the curve described by $\mathbf{x}(t)$ and dotting it with the tangent of that curve, $\dot{\mathbf{x}}$. Doing this produces a scalar, which we integrate along the parameterised curve from the starting point to the ending point. An integral of this type could be used, for example, to find the work done by a rocket to overcome gravity as it ascends along a path that is not necessarily vertical.

Example 1.8 of a line integral of a vector field

In this example we have a (somewhat contrived) vector field defined in $\mathbb{R}^3$ and given by

$$\mathbf{F} = (xe^y, z^2, xy). \quad (1.24)$$

We need to integrate this field in terms of its projection along a curve C_1 (Fig. 1.2). The curve C_1 is parameterised as $\mathbf{x}(t)$ and connects $\mathbf{x}(0) = (0, 0, 0)$ to $\mathbf{x}(1) = (1, 1, 1)$ according to

$$\mathbf{x}(t) = (t, t^2, t^3).$$

The problem is asking us to evaluate the integral given in equation (1.23). Hence we need expressions for $\mathbf{F}(\mathbf{x}(t))$ and $\dot{\mathbf{x}}(t)$. Evaluating $\mathbf{F}$ along $\mathbf{x}(t)$ we have

$$\mathbf{F}(t) = (te^{t^2}, t^6, t^3).$$

And taking the derivative of $\mathbf{x}$ with respect to t gives us

$$\dot{\mathbf{x}} = (1, 2t, 3t^2).$$

Putting this all together in equation (1.23), we have

$$\begin{aligned} \int_{t_a}^{t_b} \mathbf{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}} dt &= \int_0^1 (te^{t^2}, t^6, t^3) \cdot (1, 2t, 3t^2) dt, \\ &= \int_0^1 (te^{t^2} + 2t^7 + 3t^5) dt, \\ &= \frac{1}{4}(1 + 2e). \end{aligned} \quad (1.25)$$

This scalar is the result that we sought.

Exercise 1.4 on a line integral of a vector field

Integrate the vector field $\mathbf{F}$ given in equation (1.24) along the curve C_2 (Fig. 1.2) that connects $\mathbf{x}(0) = (0, 0, 0)$ to $\mathbf{x}(1) = (1, 1, 1)$ with the line given by

$$\mathbf{x}(t) = (t, t, t).$$

Show that the result is $5/3$. (Hint: you may need to use integration by parts.)

The example and the exercise above integrate the same vector field along different curves (that share their starting and ending points). The results of these two integrations are different. *In general, line integrals are path dependent.* Only under special circumstances are line integrals independent of the path.

It is often the case that we need to integrate along a closed curve (i.e., a loop). We denote such an integral as

$$\oint_C \mathbf{F} \cdot d\mathbf{x}. \quad (1.26)$$

Thinking of $\mathbf{F}$ as the velocity field of a fluid, this integral is called the circulation of $\mathbf{F}$ around C .

Example 1.9 of a circulation integral

We compute the circulation of the vector field given by (1.24) around the loop given by $C = C_1 - C_2$ (Fig. 1.2). We already have a result for C_1 , computed in the example above as $(1 + 2e)/4$. The result for $+C_2$ is $5/3$, but in this case we need the integral in the opposite direction, along $-C_2$. We can reverse the sign of the tangent vector and obtain the result $-5/3$. Hence we have

$$\oint_{C_1 - C_2} \mathbf{F} \cdot d\mathbf{x} = \frac{1}{4}(1 + 2e) - \frac{5}{3}.$$

There are some non-trivial vector fields $\mathbf{F}(\mathbf{x})$ that satisfy

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = 0 \quad (1.27)$$

for all closed curves C . These are known as *conservative* fields, and we discuss them after introducing an important vector operator, the gradient.

1.4 The gradient of a scalar

Consider a mapping ϕ that represents a scalar field

$$\phi : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (1.28)$$

which we can write as $\phi(\mathbf{x})$. The field depends on the n independent coordinates that span the space over which it is defined. For $n = 3$ (a 3-dimensional space), this can be written $\phi(x_1, x_2, x_3)$ or, in Cartesian coordinates, $\phi(x, y, z)$. Since ϕ depends

on three independent variables, we can take three different derivatives of ϕ . We denote these as partial derivatives

$$\frac{\partial \phi}{\partial x_i},$$

which are defined in the same way as (0.2), except that we vary x_i while holding the other coordinates constant. (Remember that x_i represents one entry in a set of coordinate directions. For example in Cartesian coordinates, x_i could be $x_2 = y$, in which case $x_1 = x$ and $x_3 = z$ are held constant in the partial derivative.)

Hence we see that since ϕ lives in $\mathbb{R}^n$, it has n partial derivatives in the n coordinate directions. Using a set of orthonormal basis vectors $\hat{\mathbf{e}}_i$ to span this space, we can define the *gradient* of a scalar as

$$\begin{aligned} \nabla \phi &= \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} \hat{\mathbf{e}}_i \\ &= \frac{\partial \phi}{\partial x_1} \hat{\mathbf{e}}_1 + \frac{\partial \phi}{\partial x_2} \hat{\mathbf{e}}_2 + \dots + \frac{\partial \phi}{\partial x_n} \hat{\mathbf{e}}_n. \end{aligned} \quad (1.29)$$

Here we have introduced the *gradient operator* ∇ , which appears in bold font because it is a vector.

For a scalar variable in Cartesian coordinates in $\mathbb{R}^3$ we have

$$\nabla \phi = \frac{\partial \phi}{\partial x} \hat{\mathbf{e}}_x + \frac{\partial \phi}{\partial y} \hat{\mathbf{e}}_y + \frac{\partial \phi}{\partial z} \hat{\mathbf{e}}_z. \quad (1.30)$$

It is very important to note that the gradient of a scalar is a vector!

We can use the gradient vector to calculate the derivative of ϕ in the direction $\hat{\mathbf{n}}$ (recall that $|\hat{\mathbf{n}}| = 1$). The slope of ϕ at the point $\mathbf{x}$ in the direction $\hat{\mathbf{n}}$ is simply given by

$$\hat{\mathbf{n}} \cdot \nabla \phi(\mathbf{x}).$$

This is known as the *directional derivative*. At any point, the directional derivative is maximal when $\hat{\mathbf{n}}$ points in the direction parallel to $\nabla \phi$. This tells us something important: that $\nabla \phi$ points in the direction in which ϕ is most rapidly increasing.

1.4.1 Line integrals of conservative vector fields

A vector field $\mathbf{F}$ is called *conservative* if it can be written as

$$\mathbf{F} = \nabla \phi \quad (1.31)$$

for some scalar field ϕ that is called a *potential* in this context. The name “conservative” derives from the association of such fields with conservation of energy, as we shall see below.

Given a vector field $\mathbf{F}$, how can we check whether it can be written as the gradient of a scalar potential? Since the components of $\mathbf{F}$ are

$$F_i = \frac{\partial \phi}{\partial x_i},$$

we can differentiate this with respect to x_j to obtain

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial^2 \phi}{\partial x_i \partial x_j} = \frac{\partial F_j}{\partial x_i}, \quad (1.32)$$

where the second equality follows from reversing the order of the derivatives. Hence we have the necessary condition

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial F_j}{\partial x_i} \quad (1.33)$$

for $\mathbf{F}$ to be conservative. On the simply connected domains considered here, this condition throughout the domain is also sufficient.

Example 1.10 of showing a vector field is conservative

Consider the vector field given by

$$\mathbf{F} = (3x^2y \sin z, x^3 \sin z, x^3y \cos z). \quad (1.34)$$

Is it conservative? Taking derivatives, we find that

$$\begin{aligned} \frac{\partial F_x}{\partial y} &= 3x^2 \sin z = \frac{\partial F_y}{\partial x}, \\ \frac{\partial F_x}{\partial z} &= 3x^2y \cos z = \frac{\partial F_z}{\partial x}, \\ \frac{\partial F_y}{\partial z} &= x^3 \cos z = \frac{\partial F_z}{\partial y}, \end{aligned}$$

so it passes the derivatives test.

Exercise 1.5 on scalar potentials and conservative fields

Show that the vector field in (1.34) is the gradient of the scalar potential $\phi = x^3y \sin z$.

We now show that for conservative fields, the line integral is independent of the path. Consider the integral of $\mathbf{F} \cdot d\mathbf{x}$ along a curve C parameterised by t from t_a to

t_b ,

$$\begin{aligned}
\int_C \mathbf{F} \cdot d\mathbf{x} &= \int_C \nabla \phi \cdot d\mathbf{x}, \quad (\text{substitution}) \\
&= \int_{t_a}^{t_b} \nabla \phi \cdot \dot{\mathbf{x}} \, dt, \quad (\text{using eq. (1.23)}) \\
&= \int_{t_a}^{t_b} \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} \frac{dx_i}{dt} \, dt, \quad (\text{expanding dot product}) \\
&= \int_{t_a}^{t_b} \frac{d}{dt} \phi(\mathbf{x}(t)) \, dt, \quad (\text{reversing chain rule}) \\
&= \int_{t_a}^{t_b} d\phi(\mathbf{x}(t)), \quad (\text{cancelling}) \\
&= [\phi(\mathbf{x}(t))]_{t_a}^{t_b} = \phi(\mathbf{x}_b) - \phi(\mathbf{x}_a). \tag{1.35}
\end{aligned}$$

These steps have proven that the line integral of a conservative field depends only on the end-points of the integral, not on the path.

Exercise 1.6 on conservative fields integrated on closed curves

Show that the conservative field $\mathbf{F} = \nabla \phi$ integrates to zero along any closed curve.

Notice that our proof of path independence relied on the use of *exact differentials*. A differential form is exact if it can be written as $\mathbf{F} \cdot d\mathbf{x} = d\phi$ for some function ϕ . This is true if (and only if) $\mathbf{F}$ is conservative, such that $\mathbf{F} = \nabla \phi$. Then we can write

$$\mathbf{F} \cdot d\mathbf{x} = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} dx_i = d\phi.$$

1.4.2 Application to mechanical energy conservation

These ideas of line integrals and conservative fields are related to Newtonian mechanics. Recall Newton's second law, written in vector form,

$$m\ddot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \tag{1.36}$$

where $\mathbf{F}(\mathbf{x})$ can be thought of as a force field (e.g., a gravitational field). Suppose that we're interested in the kinetic energy of an object and how it changes as the object moves through the force field. The kinetic energy and its rate of change are given by

$$\begin{aligned}
K &= \frac{1}{2} m |\dot{\mathbf{x}}|^2, \\
\frac{dK}{dt} &= m \dot{\mathbf{x}} \cdot \ddot{\mathbf{x}}.
\end{aligned}$$

(Note here how we have taken the derivative with respect to time!) Now we consider the change in kinetic energy between two times, t_1 and t_2 , as the particle moves and

is acted on by $\mathbf{F}$.

$$K(t_2) - K(t_1) = \int_{t_1}^{t_2} \frac{dK}{dt} dt = \int_{t_1}^{t_2} m\dot{\mathbf{x}} \cdot \ddot{\mathbf{x}} dt. \quad (1.37)$$

Then, substituting Newton's second law,

$$K(t_2) - K(t_1) = \int_{t_1}^{t_2} \dot{\mathbf{x}} \cdot \mathbf{F} dt = \int_C \mathbf{F} \cdot d\mathbf{x}. \quad (1.38)$$

This is an integral that we know how to evaluate — particularly if $\mathbf{F}$ is conservative. Let's assume that $\mathbf{F}$ is indeed conservative and write it in terms of the potential energy Φ ,

$$\mathbf{F} = -\nabla\Phi, \quad (1.39)$$

(Note that the minus sign is a convention but not mathematically necessary.) Substituting this into (1.38) we obtain

$$K(t_2) - K(t_1) = - \int_C \nabla\Phi \cdot d\mathbf{x} = \Phi(t_1) - \Phi(t_2). \quad (1.40)$$

Rearranging and generalising this result, we find that

$$K(t) + \Phi(t) = \text{constant}. \quad (1.41)$$

We have shown that a conservative force gives rise to a conserved energy $E = K + \Phi$ (in the absence of dissipative forces such as viscous drag). Indeed this is why we use the word “conservative” to describe such forces.

Summary of §1

- A curve is a continuous set of points embedded in $\mathbb{R}^n$ that can be parameterised by a scalar.
- Parameterised curves can be differentiated with respect to their parameter; doing so gives the tangent vector.
- The integral of the tangent vector magnitude with respect to the curve parameter gives the arc length; the same arc length is obtained for any parameterisation.
- The line integral of a scalar field ϕ along a curve C parameterised as $\mathbf{x}(t)$ is given by

$$\int_C \phi \, ds = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) |\dot{\mathbf{x}}| \, dt.$$

- The line integral of a vector field $\mathbf{F}$ along a curve C parameterised as $\mathbf{x}(t)$ is given by

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_{t_a}^{t_b} \mathbf{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}} \, dt.$$

- The gradient operator ∇ acting on a smooth, continuous scalar field $\phi(\mathbf{x})$ gives a vector $\nabla\phi$ that points in the direction of most rapid increase of ϕ .
- If a vector field can be expressed in terms of the gradient of a scalar potential, $\mathbf{F} = \nabla\phi$, then the vector field is conservative and its line integral is path independent,

$$\int_{\mathbf{x}_a}^{\mathbf{x}_b} \mathbf{F} \cdot d\mathbf{x} = \phi(\mathbf{x}_b) - \phi(\mathbf{x}_a).$$

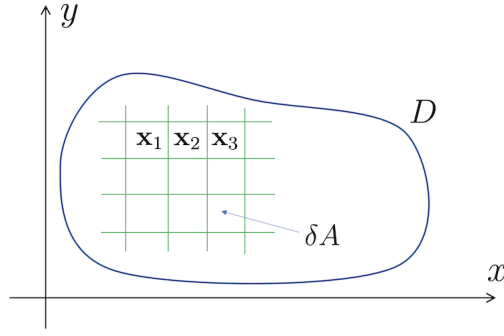


Figure 2.1: Schematic view of integration over a 2-D region D . The Cartesian space is divided into small-but-finite patches with area δA and centroids $\mathbf{x}_j$. Source: DT.

2 Surfaces and Volumes

In this section of the notes we generalise the one-dimensional integrals that we considered above to higher dimensions: surfaces and volumes.

2.1 Area integrals

Consider a two-dimensional region D in $\mathbb{R}^2$, as shown in figure 2.1. Now consider a function, $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ that maps from the position in a two-dimensional space to a scalar ϕ . (To fix ideas, we can imagine that ϕ represents topography.) We might like to integrate ϕ over the region D . We express this as

$$\int_D \phi(\mathbf{x}) dA, \quad (2.1)$$

where $\mathbf{x} = (x, y)$ is the position vector in a 2-D Cartesian coordinate system. The infinitesimal dA represents an arbitrarily small element of area; the integration is the sum over many such areas. This is the 2-D extension of the idea of integration in 1-D, $\int f dx$. In fact, we can think of $dA = dx dy$, the product of infinitesimal steps in each of the coordinate directions.

Figure 2.1 gives us a graphical way to think about the 2-D integral where we allow the infinitesimal dA to become a small-but-finite patch with area δA . Labelling each patch by its centroid $\mathbf{x}_j$, we can approximate the integral as a sum

$$\int_D \phi(\mathbf{x}) dA \approx \sum_j \phi(\mathbf{x}_j) \delta A, \quad (2.2)$$

where the sum is evaluated for all patches j with centroids inside of D . There are inaccuracies around the boundary where patches are cut by the boundary of D , but this is mitigated if we let $\delta A \rightarrow 0$ and increase the number of patches. Indeed, this is the great invention of calculus: the infinitesimal dA .

A special case of the two-dimensional area integral (2.1) is when $\phi(\mathbf{x}) = 1$. In this case the integral gives the area within D ,

$$A = \int_D dA. \quad (2.3)$$

Since A and dA both have dimensions of length-squared, the units are correct. In contrast, if we consider the case where ϕ represents topography and hence has dimensions of length, the product ϕdA has units of volume. In this case, we are integrating for the volume beneath the topography (and above the baseline from which elevation is measured).

2.2 Aside on notation for multiple integrals

To make it easier to work with surface and volume integrals, we introduce two changes to our usual notation for integrals.

First, we stop bothering to put the infinitesimal after the integrand. There's no real need for this, as you may have noticed. Instead we think of $\int dx$ as an *integration operator* that moves from left to right, integrating everything in its path (until the end of the integrand). So,

$$\begin{array}{ll} \text{old notation:} & \int f(x) dx, \\ \text{new notation:} & \int dx f(x). \end{array}$$

This makes multiple integrals more straightforward to write,

$$\begin{array}{ll} \text{old notation:} & \int_{y_a}^{y_b} \int_{x_a}^{x_b} \phi(\mathbf{x}) dx dy, \\ \text{new notation:} & \int_{y_a}^{y_b} dy \int_{x_a}^{x_b} dx \phi(\mathbf{x}). \end{array}$$

Now the $\int$ symbol is more closely associated with its infinitesimal element and the order of operations is also clearer.

Second, we change the number of $\int$ symbols to suit our needs without strict accounting. For example, when writing a surface integral,

$$\int_D dA \phi(x, y) = \int dy \int dx \phi(x, y).$$

Since we expanded the area element as $dA = dx dy$, we included an additional integral symbol so that there is a $\int$ with each infinitesimal. In some cases, however, it is clearer to match a single $\int$ with a product of two infinitesimals. The context will clarify the meaning (with some thought!).

2.3 Evaluating area integrals

A practical approach to evaluating area integrals is to reduce them to multiple ordinary integrals. There are many ways to do this, but it is usually a good idea to align things with the coordinate system. So in a Cartesian coordinate system, we would divide our region D into stripes parallel to one coordinate direction, integrate first in the direction of the stripes, then integrate over the set of parallel strips. However, even with this approach, there is still a choice of which way to align the stripes: in the x direction or the y direction. These choices are shown in figure 2.2.

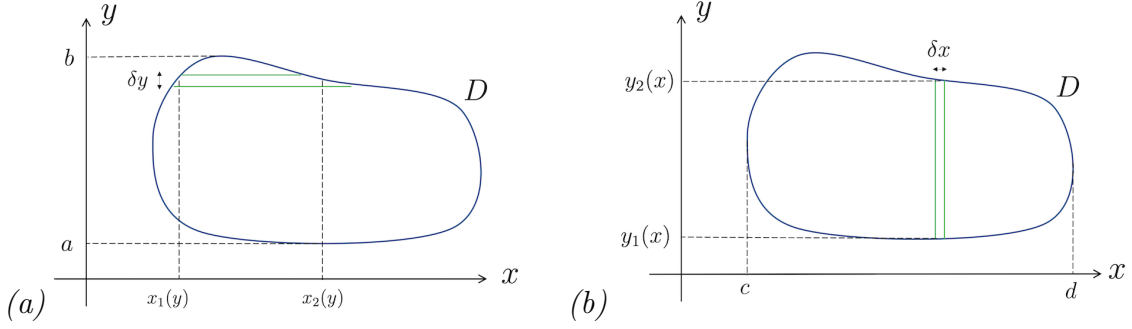


Figure 2.2: Two strategies to perform an area integral over the same region D . (a) Integrate first along stripes in the x direction and then integrate over stripes in the y direction, as written in equation (2.4). (b) Integrate first along stripes in the y direction and then integrate over stripes in the x direction. Source: DT.

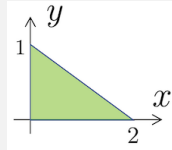
Drawing stripes of width δy in the x direction results in the schematic shown in figure 2.2a. The integration over elements dx is performed first, and it extends from x_1 to x_2 , where each of these limits is a function of y . This function is determined by the shape of the region D . This integration done, we then perform a second integration in the y direction over elements dy . This integration must span the full extent of D in the y direction, from $y = a$ to $y = b$. This process is written

$$\int_D dA \phi(x, y) = \int_a^b dy \int_{x_1(y)}^{x_2(y)} dx \phi(x, y). \quad (2.4)$$

To actually evaluate this, we need the following information: the shape of D in terms of $a, b, x_1(y), x_2(y)$ and the function $\phi(x, y)$.

Example 2.1 of an area integral with stripes in the x direction

We wish to integrate the function $\phi(x, y) = x^2 y$ over the triangular domain



We can do so using stripes in the x direction and equation (2.4). By the geometry of the triangle,

$$\begin{aligned} \int_D \phi dA &= \int_0^1 dy \int_0^{2(1-y)} dx x^2 y \\ &= \int_0^1 dy \left(y \left[\frac{x^3}{3} \right]_0^{2(1-y)} \right) = \frac{8}{3} \int_0^1 dy y (1-y)^3 = \frac{2}{15}. \end{aligned}$$

Note that we performed the integration along stripes (dx) before the integration over all stripes (dy).

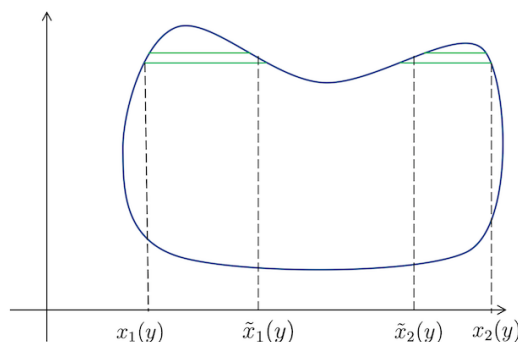


Figure 2.3: A situation where the shape of D and the direction of the first integration conspire to produce disjoint intervals. In this case, choosing stripes in the y direction for the first integral might make life easier. Source: DT.

Exercise 2.1 on an area integral with stripes in the y direction

Using the schematic diagram in figure 2.2b, derive the area integral formula equivalent to (2.4), but for stripes in the y direction. Then integrate $\phi = x^2y$ over the same triangular domain as in Example 2.1. You should get the same answer, $2/15$.

An obvious caution should apply here. Depending on the shape of D and the choice of stripe directions, it could happen that stripes are interrupted by boundary embayments, as in figure 2.3. It is also worth checking whether symmetries of the problem (that apply to both D and ϕ) make it possible to evaluate the integral with less work.

Exercise 2.2 on an area integral

Integrate the function

$$f(\mathbf{x}) = e^{-(x^2+y^2)/2}$$

over the whole first quadrant, i.e., in the region defined by $x \geq 0$, $y \geq 0$. Show that the result is $\pi/2$.

2.4 Changing coordinate systems

We have so far considered only Cartesian coordinates. However, many problems in Earth science are more conveniently represented in a spherical coordinate system, or even a cylindrical one. How can we compute integrals in those or other systems? This is our next topic. We focus first on $\mathbb{R}^2$, two-dimensional problems.

Consider a change of independent variables $(x, y) \rightarrow (u, v)$. To be a helpful change of coordinates, it must be smooth and invertible: each point in x, y maps uniquely to a point in u, v with neighbouring relationships preserved. Note, however, that this new coordinate system need not have orthogonal axes (though cylindrical and spherical coordinates are orthogonal). An example is shown in figure 2.4. How do we compute the area integral (2.3) in coordinates u, v ?

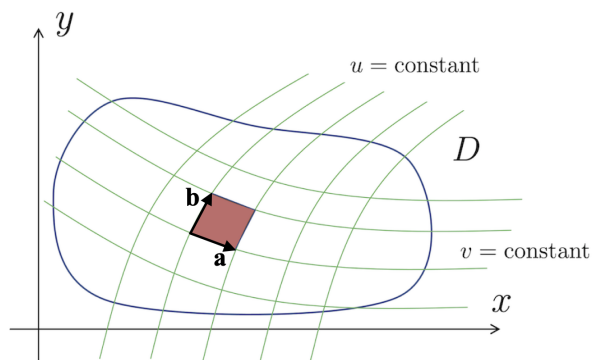


Figure 2.4: Region D of an area integration with Cartesian coordinates x, y and an alternate coordinate system u, v . The pink patch has coordinate increments δu and δv and an area δA . Source: DT.

For reasons that we will explain below, the area element can be expressed in terms of the *Jacobian matrix* of the coordinate transformation,

$$dA = |\mathbf{J}| \, du \, dv \quad (2.5)$$

where $|\mathbf{J}|$ is the determinant of the Jacobian. The Jacobian relates infinitesimal changes in a coordinate of one system to infinitesimal changes in the other,

$$\mathbf{J}(u, v) = \begin{pmatrix} \partial x / \partial u & \partial x / \partial v \\ \partial y / \partial u & \partial y / \partial v \end{pmatrix} = \frac{\partial(x, y)}{\partial(u, v)}. \quad (2.6)$$

The second equality defines a special notation for the Jacobian.

The area element (2.5) and the Jacobian (2.6) enable us to write the area integral as

$$\int_D f(\mathbf{x}) \, dA = \int_D du \, dv \, |\mathbf{J}| \, f(\mathbf{x}(u, v)), \quad (2.7)$$

where $f(\mathbf{x}(u, v))$ is a shorthand for $f(x(u, v), y(u, v))$. Notice how the determinant of the Jacobian in the scalar area integral (2.7) plays the same role as the magnitude of the tangent vector in the scalar line integral (1.20).

Why does this work? To understand we need to look at the geometric relationship between coordinate systems. Consider an arbitrary point in (u, v) and make some small change to its position $\delta u, \delta v$. As shown in figure 2.4, each of these increments can be represented as a vector,

$$\mathbf{a} = \frac{\partial \mathbf{x}}{\partial u} \delta u \quad \text{and} \quad \mathbf{b} = \frac{\partial \mathbf{x}}{\partial v} \delta v.$$

You should recall that the area of the parallelogram (pink in fig. 2.4) described by the two vectors is $|\mathbf{a} \times \mathbf{b}|$. This is the area δA and so we need to find a way to compute this cross product.

Since $x = x(u, v)$ and $y = y(u, v)$ are known functions, we can express $\delta x, \delta y$ in terms of $\delta u, \delta v$ by expanding in a Taylor series,

$$\begin{aligned} \delta x &= \frac{\partial x}{\partial u} \delta u + \frac{\partial x}{\partial v} \delta v + \mathcal{O}(\delta u^2, \delta v^2, \delta u \delta v), \\ \delta y &= \frac{\partial y}{\partial u} \delta u + \frac{\partial y}{\partial v} \delta v + \mathcal{O}(\delta u^2, \delta v^2, \delta u \delta v). \end{aligned}$$

Neglecting the second-order terms we then have

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \begin{pmatrix} \delta u \\ \delta v \end{pmatrix} \quad (2.8)$$

Using these relationships, we can write $\mathbf{a}$ and $\mathbf{b}$ in terms of the Cartesian basis vectors,

$$\mathbf{a} = \delta u \left(\frac{\partial x}{\partial u} \hat{\mathbf{e}}_x + \frac{\partial y}{\partial u} \hat{\mathbf{e}}_y \right) \quad \text{and} \quad \mathbf{b} = \delta v \left(\frac{\partial x}{\partial v} \hat{\mathbf{e}}_x + \frac{\partial y}{\partial v} \hat{\mathbf{e}}_y \right),$$

Then we can calculate the area of the parallelogram with our Cartesian basis vectors as

$$\delta A = |\mathbf{a} \times \mathbf{b}| = \delta u \delta v \begin{vmatrix} \partial x / \partial u & \partial x / \partial v \\ \partial y / \partial u & \partial y / \partial v \end{vmatrix} = |\mathbf{J}| \delta u \delta v, \quad (2.9)$$

where we have used the fact that the determinant of a matrix is equal to the determinant of its transpose. In the limit of $\delta u, \delta v \rightarrow 0$, this supports our assertion in equation (2.5).

2.5 Aside on notation for scalars, vectors and tensors

Let's acknowledge a notation that we've already been using. Vectors are indicated by bold, non-italic symbols,

vectors: $\mathbf{x}, \mathbf{v}, \mathbf{F}, \mathbf{a}, \mathbf{b}$.

We will also need to use tensors and matrices (such as the Jacobian matrix). We will discuss tensors in detail below. For now, suffice to say that such objects will be represented as bold italic or bold Greek letters,

matrices & tensors: $\mathbf{J}, \mathbf{E}, \mathbf{T}, \boldsymbol{\epsilon}$.

Scalars, of course, are not bold. Since the individual entries in vectors and tensors are scalars, these are also not bold,

scalars: $x, y, K, F_3, v_1, J_{21}$.

2.6 2-D (plane) polar coordinates

Plane-polar coordinates (figure 2.5) are a commonly used, two-dimensional coordinate system. They introduce the idea of angle as a coordinate. The angle ϕ is measured from the positive x direction to the position vector. The other coordinate is the radius, represented by ρ , which is the magnitude of the position vector.

The relationship between Cartesian and plane-polar coordinates is given by

$$x = \rho \cos \phi \quad \text{and} \quad y = \rho \sin \phi, \quad (2.10)$$

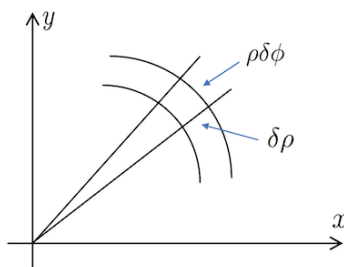


Figure 2.5: The area element δA in plane polar coordinates. Recall that arc length of a segment of circle is given by $\rho\delta\phi$. Source: DT.

where $\rho \geq 0$ and $0 \leq \phi < 2\pi$. *Important note: we previously used $\phi(x, y)$ to represent a two-dimensional, scalar field. In this section, we use ϕ as a coordinate. These two uses are not related!* We can form the Jacobian matrix,

$$\mathbf{J} = \frac{\partial(x, y)}{\partial(\rho, \phi)} = \begin{pmatrix} \cos \phi & -\rho \sin \phi \\ \sin \phi & \rho \cos \phi \end{pmatrix}, \quad (2.11)$$

and compute its determinant, $|\mathbf{J}| = \rho$. With this and (2.5) we can form the area element as

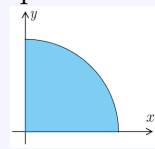
$$dA = \rho d\rho d\phi. \quad (2.12)$$

A small-but-finite version of the area element is shown in figure 2.5.

Exercise 2.3 on area integration in polar coordinates

Set up and solve an integral in polar coordinates to find the area of a quarter

circle, the region defined by the intervals $0 \leq \rho \leq R$ and $0 \leq \phi \leq \pi/2$,



Example 2.2 of area integration in polar coordinates

We seek to integrate

$$f = e^{-(x^2+y^2)/2} = e^{-\rho^2/2}$$

on the quarter-circle region defined by the intervals $0 \leq \rho \leq R$ and $0 \leq \phi \leq \pi/2$. Writing this in polar coordinates,

$$\int_D f dA = \int_0^{\pi/2} d\phi \int_0^R d\rho \rho e^{-\rho^2/2}.$$

The integrand is independent of ϕ and hence that integral gives a factor of $\pi/2$. The ρ integral can be carried out by making the substitution $u = \rho^2/2$ such that $du = \rho d\rho$. Carrying out both integrals we obtain

$$\int_D f dA = \frac{\pi}{2}(1 - e^{-R^2/2}).$$

Take the limit of this result for $R \rightarrow \infty$ and compare with the Cartesian integral in the exercise above.

2.7 Volume integrals

There are various reasons why you would need to do a volume integral. One is simply to compute the volume of a 3-D shape. Another is to compute the mass of a 3-D shape from an internal, non-uniform distribution of density.

Volume integration uses the same ideas and tools as area integration, but adds an additional dimension. The generic problem is to integrate a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ over a three-dimensional region V . We can think of the integral as a sum of many, small volume elements with volumes δV and centroids $\mathbf{x}_n$. Conceptually,

$$\int_V f(\mathbf{x}) dV = \lim_{\delta V \rightarrow 0} \sum_n f(\mathbf{x}_n) \delta V.$$

In practice, our approach is the same as the one we used for area integrals: to choose a direction for first integration and imagine a bundle of columns aligned along that direction. After integrating along a column, we integrate across all columns. We can think of this latter integration as an area integral.

For example, in Cartesian coordinates, choosing the z direction for first integration, we have

$$\int_V f(x, y, z) dV = \int dA \int_{z_1(x,y)}^{z_2(x,y)} dz f(x, y, z).$$

This approach is shown in the left-hand panel of figure 2.6. Alternatively, we could first do an integral over a thin slab of the region $D(z)$ and then integrate in dz ,

$$\int_V f(x, y, z) dV = \int dz \int_{D(z)} dx dy f(x, y, z).$$

This is illustrated in the right-hand panel of figure 2.6.

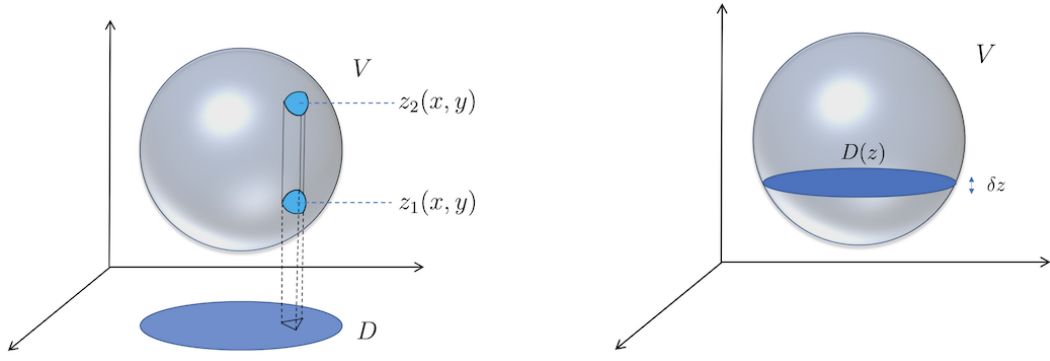


Figure 2.6: Two ways of approaching a volume integral, in this case taken over a spherical domain. At left, we first integrate along columns in z and then over the area D . At right, we first integrate over the area $D(z)$, then integrate over z . The result is identical. Source: DT.

2.7.1 Volume integrals in spherical and cylindrical coordinates

As with area integrals, it is often convenient to describe regions of integration (D or V) and functions to be integrated in terms of non-Cartesian coordinate systems. Spherical polar coordinates and cylindrical polar coordinate systems are the most common of these.

Again, the theory from 2-D area integrals translates directly into 3-D. Given a smooth and invertible coordinate mapping from $(u, v, w) \rightarrow (x, y, z)$, the volume element is given by

$$dV = dx dy dz = |\mathbf{J}| du dv dw \quad (2.13)$$

where $|\mathbf{J}|$ is the determinant of the Jacobian matrix. The latter is

$$\mathbf{J} = \frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{pmatrix}. \quad (2.14)$$

Hence the volume integral in the new coordinates is

$$\int_V f(\mathbf{x}) dV = \int_V du dv dw |\mathbf{J}| f(\mathbf{x}(u, v, w)). \quad (2.15)$$

We now apply this to cylindrical and spherical coordinates. These systems are considered further in section 3.11.

Cylindrical polar coordinates Shown in figure 2.7a, these are a simple extension of plane-polar coordinates, where the z axis is the same as in Cartesian coordinates. Hence the relationships between Cartesian and cylindrical are

$$\begin{aligned} x &= \rho \cos \phi, \\ y &= \rho \sin \phi, \\ z &= z, \end{aligned}$$

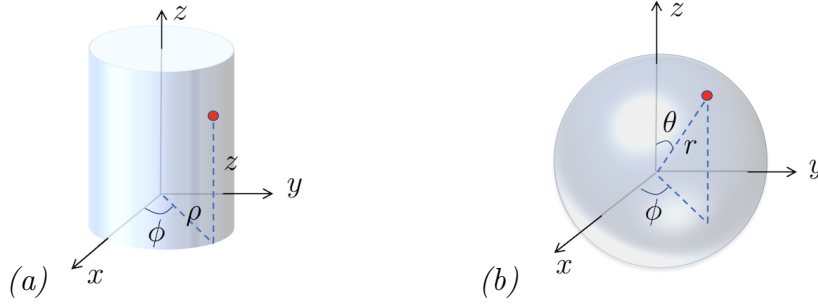


Figure 2.7: Two non-Cartesian coordinate systems: (a) cylindrical and (b) spherical coordinates. Both are defined with reference to a Cartesian system. Source: DT.

with radius $\rho \geq 0$, polar angle $0 \leq \phi < 2\pi$ and height $z \in (-\infty, \infty)$. It is an exercise to show that

$$|\mathbf{J}| = \rho \quad \Rightarrow \quad dV = \rho d\rho d\phi dz. \quad (2.16)$$

Exercise 2.4 on cylindrical coordinates

Show that the volume element for cylindrical coordinates is $\rho d\rho d\phi dz$.

Example 2.3 of volume of a cylinder

We compute the volume of a cylinder by integrating $f(\rho, \phi, z) = 1$ over the height Z and radius R of interest,

$$\int_V dV = \int_0^{2\pi} d\phi \int_0^Z dz \int_0^R d\rho \rho = \pi Z R^2$$

Spherical polar coordinates Shown in figure 2.7b, spherical coordinates use two angles: ϕ is the equivalent of longitude, measured counterclockwise from $\hat{\mathbf{e}}_x$, and θ is the equivalent of co-latitude, measured from $\hat{\mathbf{e}}_z$. These are related to Cartesians by

$$\begin{aligned} x &= r \sin \theta \cos \phi, \\ y &= r \sin \theta \sin \phi, \\ z &= r \cos \theta, \end{aligned}$$

where radius $r \geq 0$, co-latitude $0 \leq \theta \leq \pi$ and longitude $0 \leq \phi < 2\pi$. The volume element is obtained via the Jacobian as

$$|\mathbf{J}| = r^2 \sin \theta \quad \Rightarrow \quad dV = r^2 \sin \theta dr d\theta d\phi. \quad (2.17)$$

Exercise 2.5 on spherical coordinates

Show that the volume element for spherical coordinates is $r^2 \sin \theta dr d\theta d\phi$.

Example 2.4 of volume of a sphere

Consider a spherically symmetrical function $f(r)$. The volume integral of this function over a sphere of radius R is

$$\begin{aligned}\int_V f(r) dV &= \int_0^R dr \int_0^\pi d\theta \int_0^{2\pi} d\phi r^2 f(r) \sin \theta, \\ &= 2\pi [-\cos \theta]_0^\pi \int_0^R dr r^2 f(r) = 4\pi \int_0^R dr r^2 f(r).\end{aligned}$$

Taking $f(r) = 1$ we obtain the volume of a sphere, $\frac{4}{3}\pi R^3$.

Exercise 2.6 on volume integration over a sphere

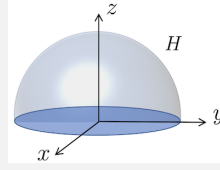
Consider the spherically symmetrical function

$$f(r) = \frac{1}{r} e^{-r^2/2}.$$

Integrate this function over the volume enclosed by $0 \leq r \leq R$, $0 \leq \theta \leq \pi/2$, $0 \leq \phi \leq \pi/2$ (this is 1/8 of a sphere). If this was easy, try the same integration of $g(r) = r e^{-r^2/2}$.

Example 2.5 of a dimensional volume integral over a sphere

Consider a density f that increases linearly in the z direction with $f(z) = f_0 z/R$



in a hemisphere H with outer radius R and $z \geq 0$. What is the total mass in H ?

Since we have $z \geq 0$, our region of integration is $0 \leq r \leq R$, $0 \leq \theta \leq \pi/2$, $0 \leq \phi < 2\pi$. Using $z = r \cos \theta$, we rewrite $f(r, \theta) = f_0 r \cos \theta / R$ and use $dV = r^2 \sin \theta dr d\theta d\phi$ and integrate

$$\begin{aligned} \int_H f dV &= \frac{f_0}{R} \int_0^{2\pi} d\phi \int_0^{\pi/2} d\theta \int_0^R dr r^3 \sin \theta \cos \theta, \\ &= \frac{2\pi f_0}{R} \left[\frac{1}{2} \sin^2 \theta \right]_0^{\pi/2} \left[\frac{r^4}{4} \right]_0^R, \\ &= \frac{1}{4} \pi R^3 f_0. \end{aligned}$$

Since this example is dimensional, we should check that the dimensions of our result are correct. The coefficient in the density function f_0 must have units of kg/m^3 to give density the correct units. Then $f_0 R^3$ has units of kg , which is what we wanted.

2.7.2 Vector-valued volume integrals

Above we have integrated scalar functions over volumes. There is no complication when integrating vector functions over volumes; we simply integrate each component independently. The final answer is a vector.

Suppose we wish to find the position $\mathbf{X}$ of the centre of mass of a 3-D object with variable density $f(\mathbf{x})$. The total mass of the object is given by the scalar integral

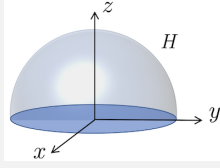
$$M = \int_V f dV.$$

The centre of mass is obtained by integrating $\mathbf{F} = \mathbf{x}f(\mathbf{x})$,

$$\mathbf{X} = \frac{1}{M} \int_V \mathbf{x}f(\mathbf{x}) dV.$$

Example 2.6 of a vector-valued integral

Find the centre of mass of the constant-density hemisphere H with outer radius



R and $z \geq 0$. Assume that the density is $f = f_0$ within H . Then we have $M = 2\pi f_0 R^3/3$ and

$$\mathbf{X} = \frac{f_0}{M} \int_H \mathbf{x} dV.$$

Before you start integrating, note two key facts:

- The vectors $\mathbf{X}$ and $\mathbf{x}$ have three independent components;
- By symmetry, we can infer that the x and y components of $\mathbf{X}$ are zero; the centre of mass lies on the z axis.

Be sure that you understand the above before continuing. Recalling that $z = r \cos \theta$ in spherical coordinates, the z component of the integral is

$$\mathbf{X} \cdot \hat{\mathbf{e}}_z = \frac{f_0}{M} \int_H z dV = \frac{f_0}{M} \int_0^{2\pi} d\phi \int_0^{\pi/2} d\theta \int_0^R dr (r \cos \theta) r^2 \sin \theta = \frac{3R}{8}.$$

As a quick check, determine the units of the result (you should work this integral yourself to check the details). The centre of mass is $\mathbf{X} = (0, 0, 3R/8)$, which is a point *inside* the hemisphere, as it must be.

2.8 Surface integrals

The last and most interesting category of integrals we'll consider are surface integrals. These are like area integrals, but over a surface that is not flat-lying. So now the surface on which we integrate is a mapping

$$\mathbf{x} : \mathbb{R}^2 \rightarrow \mathbb{R}^3,$$

which means a 2-D surface in a 3-D space. Note the similarity here to the curve on which we performed line integrals. For the curve, the mapping was $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^3$.

There is more than one way to represent two-dimensional surfaces³, but we will consider an approach that is an extension of the parameterised curve from section 1. Whereas for curves, we had a single parameter (usually called t or s), when we consider surfaces we require a set of parameters, which we'll generically refer to as u, v . If the surface corresponds to a simple shape such as a sphere or cylinder, it is straightforward to write $\mathbf{x}$ in terms of surface coordinates.

³An approach that we won't consider here is the level set, which is an isosurface of a 3-D scalar field. Calling this field $F(x, y, z)$, we could consider the surface $F(x, y, z) = 0$. Since this is one equation in three unknowns, the solution has two degrees of freedom — it is therefore a surface.

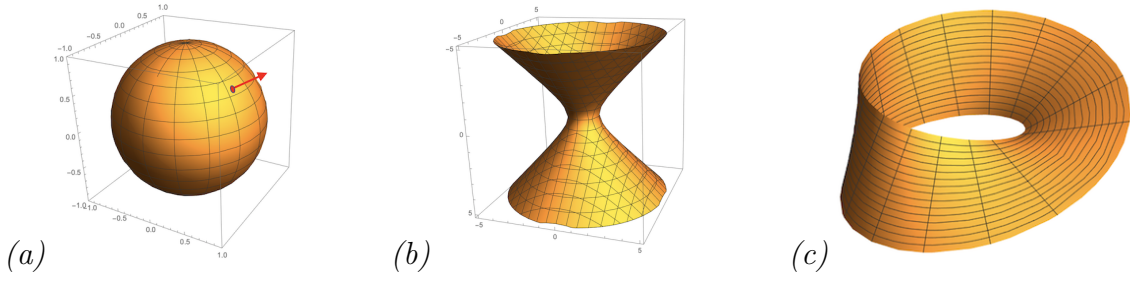


Figure 2.8: Examples of surfaces, $\mathbf{x} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$. (a) The sphere with outward-pointing normal vector. (b) The hyperboloid. (c) The Möbius strip. Note that this surface is unorientable, because there is no distinction between its two sides. Source: DT.

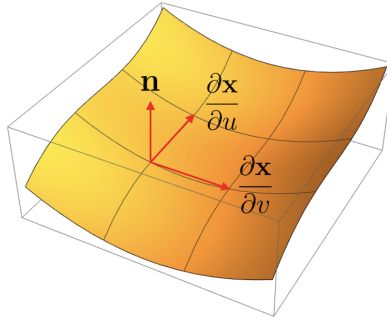


Figure 2.9: A surface $\mathbf{x}(u, v)$ with tangent vectors and a normal vector. Source: DT.

We will only consider orientable surfaces, which are surfaces where the sides of the surface are distinct. Two examples are shown in figure 2.8a,b. With an orientable surface, we can identify a side of the surface by the sign of the normal vector (as for the sphere in panel (a)). As a counter-example, consider the Möbius strip, shown in figure 2.8c.

2.8.1 Parameterised surfaces

A parameterised surface is like a parameterised curve. Two coordinates u, v give a unique position on the surface, corresponding to a unique position $\mathbf{x}$ in $\mathbb{R}^3$. This mapping from u, v to $\mathbf{x}$ allows us to obtain two independent tangent vectors at each point on the surface (for a curve, we had only one tangent vector at each point).

For a surface $\mathbf{x}(u, v)$, surface tangent vectors are given by

$$\frac{\partial \mathbf{x}}{\partial u} \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial v} \quad (\text{tangent vectors}).$$

The partial derivatives imply that the other independent surface coordinate is held constant. An example of these vectors is shown in figure 2.9.

Example 2.7 of parameterising a spherical surface

Consider the surface of a sphere with radius R . We wish to obtain a parameterisation of the form $\mathbf{x}(u, v)$. For the surface coordinates u, v we can use the spherical coordinates θ and ϕ . This gives us

$$\mathbf{x}(\theta, \phi) = R(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$$

for $\mathbf{x}$ in Cartesian coordinates. Now it is straightforward to calculate the tangents,

$$\frac{\partial \mathbf{x}}{\partial \theta} = R(\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta), \quad (2.18a)$$

$$\frac{\partial \mathbf{x}}{\partial \phi} = R(-\sin \theta \sin \phi, \sin \theta \cos \phi, 0). \quad (2.18b)$$

Exercise 2.7 on parameterising a cylindrical surface

Consider the surface of a cylinder with radius R that extends to $\pm\infty$ in the z direction. Write a surface parameterisation of the form $\mathbf{x}(\phi, z)$ and obtain the tangent vectors.

2.8.2 Integrals of scalar fields

We now wish to integrate the scalar field $f(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ on the surface S given by $\mathbf{x}(u, v)$. We need to write

$$\int_S f(\mathbf{x}) dS$$

in terms that allow us to actually evaluate it. The first thing we need is to obtain the area element dS in terms of du and dv . Here we make use of the idea introduced in section 2.4, equation (2.9), that the cross product of tangent vectors gives the area of a small-but-finite patch with sides of length δu and δv (shown in fig. 2.9),

$$\delta S = \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| \delta u \delta v.$$

In the limit of infinitesimal changes, we can write our surface integral,

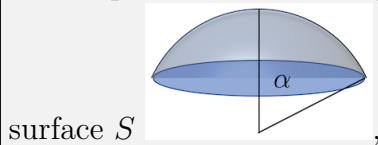
$$\int_S f(\mathbf{x}) dS = \int_D du dv \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| f(\mathbf{x}(u, v)), \quad (2.19)$$

where D is the two-dimensional region in u, v -space corresponding to S .

We can make two remarks about (2.19). First, the integral of $f = 1$ gives the surface area of S , which is always positive. Second, the result is independent of the parameterisation of S , as long as it is a valid mapping. Both of these properties are consistent with what we learned about the line integral of equation (1.20).

Example 2.8 of area of a spherical cap

To compute the area of a spherical cap of radius R we integrate $f = 1$ over the



$$\int_S dS = \int_0^{2\pi} d\phi \int_0^\alpha d\theta \left| \frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi} \right|,$$

where

$$\mathbf{x}(\theta, \phi) = R(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta).$$

We already have the tangent vectors from equations (2.18), but we need their cross product. Careful computation gives

$$\frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi} = (R \sin \theta) \mathbf{x}. \quad (2.20)$$

Noting that $|\mathbf{x}| = R$, the magnitude of this cross product is $R^2 \sin \theta$ and hence we have

$$\int_0^{2\pi} d\phi \int_0^\alpha d\theta \left| \frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi} \right| = R^2 \int_0^{2\pi} d\phi \int_0^\alpha d\theta \sin \theta = 2\pi R^2 (1 - \cos \alpha).$$

Now if we let $\alpha = \pi$ then we obtain the familiar area of a sphere, $4\pi R^2$.

2.8.3 Integrals of vector fields

When it comes to integrals of vector fields on surfaces, we are often motivated to know the total rate at which stuff is crossing the surface. Consider a vector field $\mathbf{F}(\mathbf{x})$ representing a *flux* of stuff. The word flux implies that $\mathbf{F}$ has units of (amount of stuff) $\times$ (area) $^{-1} \times$ (time) $^{-1}$ or, in words, “stuff per area per time.” If we integrate $\mathbf{F}$ over a surface, we obtain something with units of (amount of stuff)/(time), or “stuff per time,” a rate.

To fix ideas, think of $\mathbf{F}$ as the velocity field of a fluid flow through a long, straight, cylindrical pipe. Note that velocity is the volumetric flux per unit area; integrating it over a surface gives volumetric flow rate (check the units!). Now think of S as the circular opening at the end of the pipe. Then the integral of the flux over S gives the rate of volume leaving the pipe.

But we must exercise care with the *direction* of the vector field. In fact, what we are integrating is the component of the flux that is *normal to the surface*. This is the *flux vector dotted with the normal vector*. The normal vector $\hat{\mathbf{n}}$, shown in figures 2.8a and 2.9, is perpendicular to the surface. It is readily obtained from the cross product of tangent vectors,

$$d\mathbf{S} = \hat{\mathbf{n}} dS = \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right) du dv. \quad (2.21)$$

Here we have introduced the *vector area element* $d\mathbf{S}$, which has the magnitude dS and points in the direction of $\hat{\mathbf{n}}$.

Using the ideas and notation above, we write the surface integral of $\mathbf{F}$ crossing S as

$$\begin{aligned}\int_S \mathbf{F}(\mathbf{x}) \cdot d\mathbf{S} &= \int_S \mathbf{F}(\mathbf{x}) \cdot \hat{\mathbf{n}} dS, \\ &= \int_D du dv \mathbf{F}(\mathbf{x}(u, v)) \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right).\end{aligned}\tag{2.22}$$

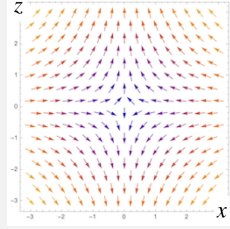
As with the scalar integral (2.19), the result of this integration is independent of the parameterisation of the surface.

Example 2.9 of the surface integral of a vector field

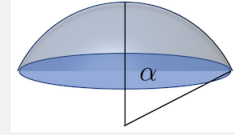
Consider the vector field given by

$$\mathbf{G}(\mathbf{x}) = (-x, 0, z).$$

We want to integrate this field over the spherical cap S subtended by an angle α .



The field $\mathbf{G}$:



The cap S :

Using (2.20), our vector area element for the spherical cap is

$$d\mathbf{S} = (R \sin \theta) \mathbf{x} \, d\theta \, d\phi.$$

with the parameterised surface $\mathbf{x}(\theta, \phi) = R(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$. Then the surface integral (2.22) becomes

$$\int_S \mathbf{G} \cdot d\mathbf{S} = R^3 \int_0^{2\pi} d\phi \int_0^\alpha d\theta \sin \theta (\cos^2 \theta - \sin^2 \theta \cos^2 \phi),$$

where we have used $x = R \sin \theta \cos \phi$ and $z = R \cos \theta$. Next we use $\int_0^{2\pi} \cos^2 \phi \, d\phi = \pi$ to obtain

$$\int_S \mathbf{G} \cdot d\mathbf{S} = \pi R^3 \int_0^\alpha d\theta \sin \theta (2 \cos^2 \theta - \sin^2 \theta).$$

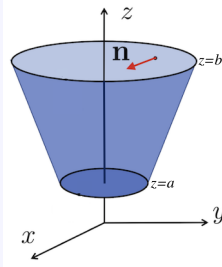
To integrate this, we use the trigonometric identity $\sin^2 \theta + \cos^2 \theta = 1$ to eliminate the $\sin^2 \theta$, then make the substitution $U = \cos \theta$ and $dU = -\sin \theta \, d\theta$. We can then integrate and restore θ to obtain

$$\int_S \mathbf{G} \cdot d\mathbf{S} = \pi R^3 [\cos \theta - \cos^3 \theta]_0^\alpha = \pi R^3 \cos \alpha \sin^2 \alpha,$$

where we have used the same trig identity. It is worth noting that this integral is zero when $\alpha = 0, \pi$, and understanding why.

Exercise 2.8 on the surface integral of a vector field

Consider the vector field $\mathbf{G}(\mathbf{x}) = (-x, 0, z)$ that was used in example 2.9. Now we



want to integrate this over the surface S defined by $z^2 = x^2 + y^2$ with z within the strictly positive interval $[a, b]$. Using cylindrical coordinates to parameterise the surface,

$$\mathbf{x}(\rho, \phi) = (\rho \cos \phi, \rho \sin \phi, \rho),$$

with $a \leq \rho \leq b$. Obtain the normal vector $\hat{\mathbf{n}}$ and show that the vector area element is $d\mathbf{S} = (-\cos \phi, -\sin \phi, 1) \rho d\rho d\phi$. Then perform the integral $\int_S \mathbf{G} \cdot d\mathbf{S}$. Show that the result is $\pi(b^3 - a^3)$.

Summary of §2

- The integral of a scalar field $f(\mathbf{x})$ over a 2-D, flat-lying surface D is given by

$$\int_D f(\mathbf{x}) dA.$$

Evaluating it in Cartesian coordinates usually requires choosing one coordinate direction for first integration and the other coordinate direction for second integration.

- Under a change of coordinates from $x, y \rightarrow u, v$, the area element becomes $dA = |\mathbf{J}| du dv$ where $\mathbf{J}$ is the Jacobian matrix $\partial(x, y)/\partial(u, v)$. Then the area integral of a scalar field f is

$$\int_D du dv |\mathbf{J}| f(\mathbf{x}(u, v)).$$

- Volume integrals follow directly from area integrals. The Cartesian and non-Cartesian formulae for a scalar field $f(\mathbf{x})$ are

$$\int_V f(\mathbf{x}) dV = \int_V du dv dw |\mathbf{J}| f(\mathbf{x}(u, v, w))$$

with a Jacobian matrix $\partial(x, y, z)/\partial(u, v, w)$. The volume integral of a vector field is done with the same approach, applied component-wise.

- In cylindrical and spherical coordinates we have

$$\begin{aligned} dA &= \rho d\rho d\phi && \text{(plane polar),} \\ dV &= \rho d\rho d\phi dz && \text{(cylindrical),} \\ dV &= r^2 \sin \theta dr d\theta d\phi && \text{(spherical).} \end{aligned}$$

- Surface integrals of a scalar field are performed in the coordinates u, v of the parameterised surface $\mathbf{x}(u, v)$ as,

$$\int_S f(\mathbf{x}) dS = \int_D du dv \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| f(\mathbf{x}(u, v)).$$

- Surface integrals of the normal component of a vector field $\mathbf{F}(\mathbf{x})$ are performed in the coordinates u, v of the parameterised surface as

$$\int_S \mathbf{F}(\mathbf{x}) \cdot d\mathbf{S} = \int_D du dv \mathbf{F}(\mathbf{x}(u, v)) \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right).$$

3 Vector Operators

The vector differential operators Div, Grad and Curl are perhaps the most important mathematical objects in vector calculus. They operate on fields and differentiate them in various ways. We'll be mostly concerned with how they operate on scalar and vector fields.

The crucial piece of notation is the symbol ∇ , which is called “nabla” or “del.” We will write this symbol in bold ∇ in these notes to remind you that it is a vector.

3.1 The gradient

We already introduced the gradient in section 1.4, where we looked at its action on a scalar field, $\phi(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$. In Cartesian coordinates (repeating equation (1.29)),

$$\begin{aligned}\nabla\phi &= \sum_{i=1}^n \frac{\partial\phi}{\partial x_i} \hat{\mathbf{e}}_i \\ &= \frac{\partial\phi}{\partial x_1} \hat{\mathbf{e}}_1 + \frac{\partial\phi}{\partial x_2} \hat{\mathbf{e}}_2 + \dots + \frac{\partial\phi}{\partial x_n} \hat{\mathbf{e}}_n.\end{aligned}\tag{3.1}$$

So the gradient of ϕ is a vector in an n -dimensional space; for our purposes, n is usually three.

We can also think of the gradient in a way that is independent of coordinate systems. Suppose that we want to use a Taylor series approximation to estimate the value of ϕ at a point close to $\mathbf{x}_0$. In fact, the point we wish to consider is $\mathbf{x}_0 + \mathbf{h}$. The Taylor series is

$$\phi(\mathbf{x}_0 + \mathbf{h}) = \phi(\mathbf{x}_0) + \mathbf{h} \cdot \nabla\phi|_{\mathbf{x}_0} + \mathcal{O}(|\mathbf{h}|^2),\tag{3.2}$$

where we have only explicitly written out the first two terms. Note that we have not chosen a coordinate system! But we can see that the linear change in ϕ is represented by the dot product of the displacement $\mathbf{h}$ and $\nabla\phi|_{\mathbf{x}_0}$ (the latter is the gradient evaluated at the point $\mathbf{x}_0$). Hence we understand that at the point $\mathbf{x}_0$, the largest slope of the field is in the direction given by its gradient. The slope of the field in the direction of $\mathbf{h}$ is the projection of $\nabla\phi$ onto $\hat{\mathbf{h}}$ (i.e., their dot product).

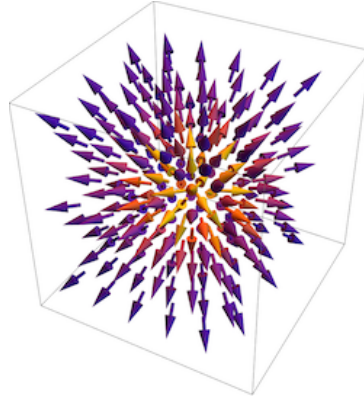


Figure 3.1: The vector field $\hat{\mathbf{e}}_r/r^2$. Vector magnitude decreases from yellow to dark purple. Source: DT.

Example 3.1 of gradient of a scalar field

Consider the scalar field

$$\phi(x, y, z) = -\frac{1}{\sqrt{x^2 + y^2 + z^2}} = -\frac{1}{r},$$

where $r = |\mathbf{x}|$ is the distance from the origin to $\mathbf{x}$. We seek to determine $\nabla\phi$.

Taking a derivative with respect to x ,

$$\frac{\partial\phi}{\partial x} = \frac{x}{(x^2 + y^2 + z^2)^{3/2}} = \frac{x}{r^3}.$$

The derivatives in the y and z directions are similar and hence we can write

$$\nabla\phi = \sum_{i=1}^3 \frac{\partial\phi}{\partial x_i} \hat{\mathbf{e}}_i = \frac{x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z}{r^3} = \frac{\hat{\mathbf{e}}_r}{r^2},$$

where we have introduced a radial basis vector,

$$\hat{\mathbf{e}}_r = \frac{x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z}{r}.$$

At any point $\mathbf{x}$, this unit vector points radially away from the origin. Figure 3.1 depicts the vector field, $\nabla\phi$, which points in the radial direction and has a magnitude that decreases as $1/r^2$.

We can use the gradient to compute the derivative of a field along a curve. Suppose that we have the curve $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^3$, parameterised by t , that moves through a scalar field $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$. We can consider the value of ϕ along the curve as a composite map, $\phi(\mathbf{x}(t)) : \mathbb{R} \rightarrow \mathbb{R}$. And we can consider the derivative of this

value along the curve, $d\phi(\mathbf{x}(t))/dt$. Using a version of the chain rule,

$$\begin{aligned}\frac{d\phi(\mathbf{x}(t))}{dt} &= \frac{\partial\phi}{\partial x} \frac{dx}{dt} + \frac{\partial\phi}{\partial y} \frac{dy}{dt} + \frac{\partial\phi}{\partial z} \frac{dz}{dt}, \\ &= \dot{\mathbf{x}} \cdot \nabla\phi.\end{aligned}\tag{3.3}$$

(You can check this by working backwards.) This says that the change in ϕ along the curve is the slope of the field projected onto the tangent of the curve.

Exercise 3.1 on change along a curve

Given the helical curve $\mathbf{x}(t) = (\cos t, \sin t, t)$ and the field $\phi(x, y, z) = xy - z$, show that

$$\frac{d\phi(\mathbf{x}(t))}{dt} = -2\sin^2 t.$$

Does this result depend on the parameterisation of the curve $\mathbf{x}$?

3.2 The Nabla (or Del) operator

It is very helpful to think of ∇ as an object with a life of its own, given by

$$\nabla = \sum_{i=1}^n \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i},\tag{3.4}$$

which we'll refer to as the Nabla or Del operator. We can notice two things about it straight away. First, it is a vector because it is a (weighted) sum over basis vectors, and second, it is an operator, ready to differentiate anything it encounters (on its right).

In (3.4), we have carefully placed the basis vector $\hat{\mathbf{e}}_i$ to the left of the differential operator because the differential should not operate on the basis vector.

3.3 Aside on index notation, Einstein summation convention, dot products, and the Kronecker delta

We have been writing vectors as sums over the components multiplied by basis vectors. As examples in Cartesian coordinates,

$$\begin{aligned}\mathbf{x} &= x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z = \sum_{i=1}^3 x_i \hat{\mathbf{e}}_i, \\ \nabla\phi &= \frac{\partial\phi}{\partial x} \hat{\mathbf{e}}_x + \frac{\partial\phi}{\partial y} \hat{\mathbf{e}}_y + \frac{\partial\phi}{\partial z} \hat{\mathbf{e}}_z = \sum_{i=1}^3 \frac{\partial\phi}{\partial x_i} \hat{\mathbf{e}}_i, \\ \nabla &= \hat{\mathbf{e}}_x \frac{\partial}{\partial x} + \hat{\mathbf{e}}_y \frac{\partial}{\partial y} + \hat{\mathbf{e}}_z \frac{\partial}{\partial z} = \sum_{i=1}^3 \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i}.\end{aligned}$$

This notation, using an integer index as a subscript (usually i , j or k), is called *index notation*. It is a very helpful notation, except that it is cluttered by a lot of

$\sum$ symbols. However, Einstein noticed that we don't need these. We can simply note that anytime there is a *repeated index*, it implies a sum. This is the *Einstein summation convention*. So, for example,

$$\begin{aligned} x_i \hat{\mathbf{e}}_i &= x_1 \hat{\mathbf{e}}_1 + x_2 \hat{\mathbf{e}}_2 + x_3 \hat{\mathbf{e}}_3, \\ &= x \hat{\mathbf{e}}_x + y \hat{\mathbf{e}}_y + z \hat{\mathbf{e}}_z. \end{aligned}$$

Let's try dotting two vectors $\mathbf{a}$ and $\mathbf{b}$, both in $\mathbb{R}^n$,

$$\mathbf{a} \cdot \mathbf{b} = (a_i \hat{\mathbf{e}}_i) \cdot (b_j \hat{\mathbf{e}}_j).$$

Importantly, we used two different indices, i and j for the two vectors, because we cannot have an index repeated four times (and because the indices are, at this point, independent). We can take the coefficients a_i and b_j out front, leaving

$$\mathbf{a} \cdot \mathbf{b} = a_i b_j (\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j).$$

Now recall something important we learned in equation (0.8) of section 0.2.2:

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

where δ_{ij} is the *Kronecker delta*. This equation is true because the basis vectors are mutually orthogonal. Therefore $\hat{\mathbf{e}}_x \cdot \hat{\mathbf{e}}_y = 0$ and similar for other combinations where $i \neq j$. So we only need to keep the entries where $i = j$! Hence we can write

$$\mathbf{a} \cdot \mathbf{b} = b_j a_i \delta_{ij} = b_i a_i = b_j a_j,$$

Now, using Einstein summation convention, we can see that we've got the usual definition of the dot product. The Kronecker delta tells us to replace i with j (or, equivalently, j with i).

With this new notation, we can write the Nabla operator in a cleaner form,

$$\nabla = \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i}, \tag{3.5}$$

where the summation is implied.

Exercise 3.2 on index notation and dot products

Convert the following from vector notation to index notation or vice versa.

- | | |
|---|---|
| 1. $\mathbf{c} \cdot \mathbf{d}$, | 5. $a_i b_k \delta_{ik} d_j \hat{\mathbf{e}}_j$ |
| 2. $(\mathbf{a} \cdot \mathbf{b}) \mathbf{c}$, | 6. $\nabla \psi$ |
| 3. $c_i \delta_{ij} \hat{\mathbf{e}}_j$, | 7. $(\nabla \psi) \cdot \mathbf{b}$ |
| 4. $a_i b_i c_j d_j$, | 8. $\mathbf{a} \cdot \nabla$ |

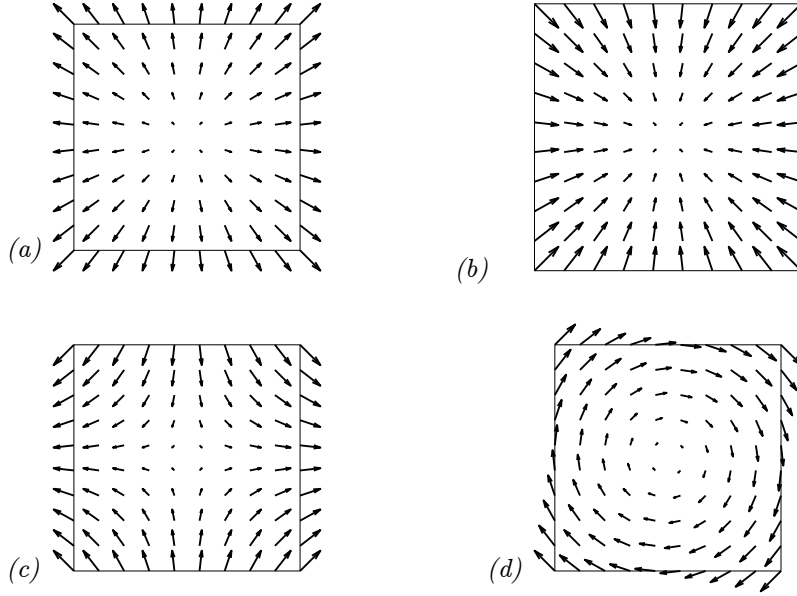


Figure 3.2: Vector fields plotted in the x - y plane at $z = 0$ to illustrate the concepts of divergence and curl. $E > 0$ is a constant. (a) Field $\mathbf{F}(\mathbf{x}) = E\mathbf{x}$. (b) Field $\mathbf{G}(\mathbf{x}) = -E\mathbf{x}$. (c) Field $\mathbf{H}(\mathbf{x}) = E(x, -y, 0)$. (d) Field $\mathbf{J}(\mathbf{x}) = E(y, -x, 0)$.

3.4 The divergence

We've got a nifty vector operator ∇ and we've seen how it operates on a scalar field to give the gradient. But obviously we are itching to dot it with a vector field. Let's use a Cartesian coordinate system and dot ∇ with $\mathbf{F}(\mathbf{x})$,

$$\begin{aligned}\nabla \cdot \mathbf{F} &= \left(\hat{\mathbf{e}}_i \frac{\partial}{\partial x_i} \right) \cdot (\hat{\mathbf{e}}_j F_j), \\ &= (\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j) \frac{\partial F_j}{\partial x_i} = \delta_{ij} \frac{\partial F_j}{\partial x_i}, \\ &= \frac{\partial F_i}{\partial x_i},\end{aligned}\tag{3.6}$$

where the repeated index implies a summation and, of course, $F_i(\mathbf{x})$ is the i^{th} component of $\mathbf{F}(\mathbf{x})$. Therefore, in Cartesian coordinates,

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.\tag{3.7}$$

This is known as *the divergence* of $\mathbf{F}$. We can get an intuitive feeling for what it means by considering some examples.

Example 3.2 of the divergence of a vector field

Compute the divergence of the vector field

$$\mathbf{F}(\mathbf{x}) = E\mathbf{x} \quad (\text{see fig. 3.2a}),$$

for some positive constant E .

We start by writing $\mathbf{F} = E(x, y, z)$. Then we use (3.7) to write

$$\nabla \cdot \mathbf{F} = E \left(\frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} \right) = 3E.$$

Note that the divergence of $\mathbf{F}$ is positive and uniform in space.

Exercise 3.3 on the divergence of a vector field

Compute the divergence of each of the following fields. E is a positive constant.

$$\mathbf{G}(\mathbf{x}) = -E\mathbf{x} \quad (\text{fig. 3.2b}),$$

$$\mathbf{H}(\mathbf{x}) = E(x, -y, 0) \quad (\text{fig. 3.2c}),$$

$$\mathbf{J}(\mathbf{x}) = E(y, -x, 0) \quad (\text{fig. 3.2d}),$$

$$\mathbf{K}(\mathbf{x}) = E(x^2, xy, 5\sqrt{x}).$$

If we think of a vector field as representing the flux of a quantity (e.g., mass), then we can give a simple physical meaning to divergence: at a point in space where the divergence is positive, there is a net flux of mass away from that point; at a point where the divergence is negative, there is a net flux toward that point. Thus a negative divergence is a convergence.

3.5 Aside on the Levi-Civita symbol, a notation for cross products

We are (or should be) familiar with computing cross products by the determinant method,

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = \hat{\mathbf{e}}_x(a_y b_z - a_z b_y) - \hat{\mathbf{e}}_y(a_x b_z - a_z b_x) + \hat{\mathbf{e}}_z(a_x b_y - a_y b_x).$$

Vector calculus uses a different notation for the cross product. Let's use indices (and the summation convention) to write

$$\mathbf{a} \times \mathbf{b} = (a_i \hat{\mathbf{e}}_i) \times (b_j \hat{\mathbf{e}}_j) = a_i b_j (\hat{\mathbf{e}}_i \times \hat{\mathbf{e}}_j).$$

This is a double sum over i and j , so it includes every i, j combination. Suppose we're using Cartesian coordinates⁴, so each index takes values x, y, z . Let's consider

⁴We require that the *coordinate system is right-handed*. This means that if I align the fingers of my right hand with the x axis then rotate them toward the y axis, my thumb points in the

all of the 9 possible combinations of $\hat{\mathbf{e}}_i \times \hat{\mathbf{e}}_j$,

$$\begin{aligned}\hat{\mathbf{e}}_x \times \hat{\mathbf{e}}_x &= \mathbf{0}, & \hat{\mathbf{e}}_x \times \hat{\mathbf{e}}_y &= \hat{\mathbf{e}}_z, & \hat{\mathbf{e}}_x \times \hat{\mathbf{e}}_z &= -\hat{\mathbf{e}}_y, \\ \hat{\mathbf{e}}_y \times \hat{\mathbf{e}}_x &= -\hat{\mathbf{e}}_z, & \hat{\mathbf{e}}_y \times \hat{\mathbf{e}}_y &= \mathbf{0}, & \hat{\mathbf{e}}_y \times \hat{\mathbf{e}}_z &= \hat{\mathbf{e}}_x, \\ \hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_x &= \hat{\mathbf{e}}_y, & \hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_y &= -\hat{\mathbf{e}}_x, & \hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_z &= \mathbf{0}.\end{aligned}$$

We can efficiently capture all of the patterns in this array with

$$\hat{\mathbf{e}}_i \times \hat{\mathbf{e}}_j = \epsilon_{ijk} \hat{\mathbf{e}}_k, \quad (3.8)$$

where ϵ_{ijk} is the *Levi-Civita symbol*,

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } (i, j, k) \text{ is } (1, 2, 3) \text{ or } (2, 3, 1) \text{ or } (3, 1, 2), \\ -1 & \text{if } (i, j, k) \text{ is } (1, 3, 2) \text{ or } (3, 2, 1) \text{ or } (2, 1, 3), \\ 0 & \text{if any two index values are equal.} \end{cases} \quad (3.9)$$

That is, even permutations of 1, 2, 3 give +1, odd permutations give -1, and anything else gives zero.

With the Levi-Civita symbol, we can write the cross product

$$\mathbf{a} \times \mathbf{b} = a_i b_j (\hat{\mathbf{e}}_i \times \hat{\mathbf{e}}_j) = \epsilon_{ijk} a_i b_j \hat{\mathbf{e}}_k.$$

This is now a triple sum over i , j , and k , so there are 27 terms in the explicit triple sum, of which 6 are non-zero.

Exercise 3.4 on index notation and cross products

Convert the following from vector notation to index notation or vice versa.

- | | |
|--|--|
| 1. $\mathbf{c} \times \mathbf{d}$, | 3. $\epsilon_{ijk} \hat{\mathbf{e}}_k \delta_{ij}$ (evaluate), |
| 2. $\mathbf{a} \cdot (\mathbf{c} \times \mathbf{d})$, | 4. $(\mathbf{a} \cdot \mathbf{b})(\mathbf{c} \times \mathbf{d})$. |

3.6 The curl

The next obvious thing to try is the cross product of the Nabla operator with a vector field. Let's use a Cartesian coordinate system and cross ∇ with $\mathbf{F}(\mathbf{x})$,

$$\begin{aligned}\nabla \times \mathbf{F} &= \left(\hat{\mathbf{e}}_i \frac{\partial}{\partial x_i} \right) \times (\hat{\mathbf{e}}_j F_j), \\ &= (\hat{\mathbf{e}}_i \times \hat{\mathbf{e}}_j) \frac{\partial F_j}{\partial x_i} = \epsilon_{ijk} \frac{\partial F_j}{\partial x_i} \hat{\mathbf{e}}_k, \\ &= \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}, \\ &= \hat{\mathbf{e}}_x \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) - \hat{\mathbf{e}}_y \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z} \right) + \hat{\mathbf{e}}_z \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right).\end{aligned} \quad (3.10)$$

direction of the z axis.

This is known as *the curl* of $\mathbf{F}$. Like the divergence, we need to look at the curl of some fields to get a feeling for the meaning.

Example 3.3 of the curl of a vector field

Compute the curl of the vector field

$$\mathbf{F}(\mathbf{x}) = E\mathbf{x} \quad (\text{see fig. 3.2a}),$$

for some positive constant E . We use the determinant,

$$\nabla \times \mathbf{F} = E \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x & y & z \end{vmatrix} = 0.$$

We can immediately find the result by noticing that the components of $\mathbf{F}$ depend only on their own coordinate (e.g., $F_x = F_x(x)$). This also applies to the vector fields in fig. 3.2b,c.

Example 3.4 of the curl of a vector field

Compute the curl of the vector field

$$\mathbf{J}(\mathbf{x}) = E(y, -x, 0) \quad (\text{see fig. 3.2d}),$$

for some positive constant E . We use the determinant,

$$\nabla \times \mathbf{J} = E \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y & -x & 0 \end{vmatrix} = E(0\hat{\mathbf{e}}_x - 0\hat{\mathbf{e}}_y + (-2)\hat{\mathbf{e}}_z) = -2E\hat{\mathbf{e}}_z.$$

It is the circulation of this vector field that gives a nonzero curl. We can use the right-hand rule to understand the sign of the resulting vector.

The vector fields of fig. 3.2a–c have zero curl, but the curl of the vector field in fig. 3.2d is nonzero. This illustrates the key aspect of the curl: it measures the *circulation* of the vector field. We'll see this in more detail below, when we relate the curl to the circulation integral.

Exercise 3.5 on curl of a vector field

Consider the vector field given by

$$\mathbf{F}(\mathbf{x}) = E(\sin y, -\sin x, 0).$$

Sketch or computer-plot the field. Compute $\nabla \times \mathbf{F}$ and comment on its form in relation to $\mathbf{F}$.

3.7 Mathematical properties of grad, div, and curl

Since ∇ is a linear operator, it has most of the properties that you'd expect. In particular,

$$\nabla(\alpha\phi + \psi) = \alpha\nabla\phi + \nabla\psi, \quad (3.11a)$$

$$\nabla \cdot (\alpha\mathbf{F} + \mathbf{G}) = \alpha\nabla \cdot \mathbf{F} + \nabla \cdot \mathbf{G}, \quad (3.11b)$$

$$\nabla \times (\alpha\mathbf{F} + \mathbf{G}) = \alpha\nabla \times \mathbf{F} + \nabla \times \mathbf{G}, \quad (3.11c)$$

for any scalar fields ϕ, ψ , vector fields $\mathbf{F}, \mathbf{G}$ and constant α . There are also the generalised product rules,

$$\nabla(\phi\psi) = \phi\nabla\psi + \psi\nabla\phi, \quad (3.12a)$$

$$\nabla \cdot (\phi\mathbf{F}) = \phi(\nabla \cdot \mathbf{F}) + \mathbf{F} \cdot (\nabla\phi), \quad (3.12b)$$

$$\nabla \times (\phi\mathbf{F}) = (\nabla\phi) \times \mathbf{F} + \phi(\nabla \times \mathbf{F}), \quad (3.12c)$$

These can be proved by writing them in index notation and using the standard product rule. As usual, care must be taken in the ordering of cross products, as this affects the sign of the result.

There are more complicated vector identities, all of which can be proved using index notation. A useful one is

$$\nabla \times \nabla \times \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}. \quad (3.13)$$

We will discuss others when they are needed. It is always easy to [look them up](#).

Since we are treating the Nabla operator as an individual object, we can also combine with a vector field from the left,

$$\mathbf{F} \cdot \nabla = F_i \frac{\partial}{\partial x_i} \quad \text{and} \quad \mathbf{F} \times \nabla = \epsilon_{ijk} \hat{\mathbf{e}}_k F_i \frac{\partial}{\partial x_j}. \quad (3.14)$$

Each of these is a differential operator, ready to differentiate any object it encounters on its right.

3.8 The Laplacian operator

What happens when we take the dot product of two Nabla operators? We get the Laplacian operator. Since dotting two vectors gives a scalar, we should expect that dotting two vector operators gives a scalar operator. Let's see:

$$\begin{aligned} \nabla \cdot \nabla &= \left(\hat{\mathbf{e}}_i \frac{\partial}{\partial x_i} \right) \cdot \left(\hat{\mathbf{e}}_j \frac{\partial}{\partial x_j} \right) = \delta_{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j}, \\ &= \frac{\partial^2}{\partial x_i \partial x_i} = \nabla^2, \end{aligned} \quad (3.15)$$

Indeed this is a scalar operator, but with a repeated index, and therefore an implied summation. In Cartesian coordinates, the Laplacian of a scalar is

$$\begin{aligned} \nabla^2 \phi &= \frac{\partial^2 \phi}{\partial x_i \partial x_i}, \\ &= \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2}. \end{aligned}$$

The Laplacian also operates on vector fields,

$$\nabla^2 \mathbf{F}(\mathbf{x}) = \frac{\partial^2 \mathbf{F}}{\partial x^2} + \frac{\partial^2 \mathbf{F}}{\partial y^2} + \frac{\partial^2 \mathbf{F}}{\partial z^2}. \quad (3.16)$$

The Laplacian of a vector field is a vector field.

Exercise 3.6 on the Laplacian of a vector field

Use index notation to show that $\nabla \cdot \nabla \mathbf{F}$ gives a result that is a vector field. Hint: think of this as $(\nabla \cdot \nabla) \mathbf{F}$.

Just as the second derivative of a curve $f''(x)$ represents its curvature, the Laplacian of a scalar field is a field that represents the *curvature of the field* everywhere in space.

3.9 The Helmholtz decomposition

There is a useful way to decompose any vector field into two distinct parts. It is called the Helmholtz decomposition. It uses two *potentials*⁵, $\mathbf{A}$ and ϕ ,

$$\mathbf{H}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}) + \nabla \phi(\mathbf{x}). \quad (3.17)$$

To understand the value of this decomposition, we need to consider two important properties of combinations of vector operators.

The first of these is the following. For any vector field $\mathbf{B}$ it is true that

$$\nabla \cdot (\nabla \times \mathbf{B}) = 0. \quad (3.18)$$

We can prove this by writing (3.18) in index notation,

$$\left(\hat{\mathbf{e}}_l \frac{\partial}{\partial x_l} \right) \cdot \left(\epsilon_{ijk} \hat{\mathbf{e}}_i \frac{\partial}{\partial x_j} B_k \right) = \epsilon_{ijk} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} B_k = 0.$$

This equals zero because, for any choice of i and j , we can make a conjugate term by exchanging the index values and flipping the order of the derivatives; these two terms have opposite sign because of ϵ_{ijk} and hence they sum to zero. For example,

$$\underbrace{\epsilon_{xyz}}_{\text{even, } 1} \frac{\partial}{\partial x} \frac{\partial}{\partial y} B_z + \underbrace{\epsilon_{yxz}}_{\text{odd, } -1} \frac{\partial}{\partial y} \frac{\partial}{\partial x} B_z = 0.$$

The second property is the following. For any scalar field ψ it is true that

$$\nabla \times (\nabla \psi) = 0. \quad (3.19)$$

It is an exercise to prove this theorem.

Exercise 3.7 on proving a property of vector operators

Use index notation and symmetry arguments to prove that equation (3.19) is true. Hint: this closely follows the proof of (3.18).

⁵In general, a potential is a function whose derivatives are used to describe other functions.

Now that we have the properties (3.18) and (3.19), we can understand the value of the Helmholtz decomposition. Consider the following operations on $\mathbf{H}$ as defined by (3.17),

$$\nabla \times \mathbf{H} = \nabla \times \nabla \times \mathbf{A}, \quad (3.20a)$$

$$\nabla \cdot \mathbf{H} = \nabla^2 \phi, \quad (3.20b)$$

where we have used (3.19) in the first one and (3.18) in the second one. Now it is evident that *the Helmholtz decomposition splits a vector field into a divergence-free part and a curl-free part.*

We can make the following generalisations.

- If a vector field $\mathbf{F}(\mathbf{x})$ has the property $\nabla \cdot \mathbf{F} = 0$ then we call it a *solenoidal* field and we can write it as the curl of a vector potential, $\mathbf{F} = \nabla \times \mathbf{A}$.
- If a vector field $\mathbf{G}(\mathbf{x})$ has the property $\nabla \times \mathbf{G} = 0$ then we call it *irrotational* and we can write it as the gradient of a scalar potential, $\mathbf{G} = \nabla \phi$.

You should recognise the second of these two generalisations as being related to conservative fields from section 1.4.1. Now we can provide a succinct way of checking for conservation: if $\nabla \times \mathbf{G} = 0$ then $\mathbf{G}(\mathbf{x})$ is a conservative field (on the simply connected domains considered here). Furthermore, in this case it is true that $\oint_C \mathbf{G} \cdot d\mathbf{x} = 0$.

Exercise 3.8 on potentials for irrotational (conservative) fields

Consider the irrotational field $\mathbf{F} = \nabla \phi$ with scalar potential $\phi(\mathbf{x})$. Show, for any constant α , that $\psi = \phi + \alpha$ is also a potential for $\mathbf{F}$, and hence that ϕ is not uniquely determined.

Exercise 3.9 on potentials for solenoidal fields

Consider the solenoidal field $\mathbf{F} = \nabla \times \mathbf{A}$ for vector potential $\mathbf{A}(\mathbf{x})$. Show, for any scalar field $g(\mathbf{x})$, that $\mathbf{B} = \mathbf{A} + \nabla g$ is also a potential for $\mathbf{F}$, and hence that $\mathbf{A}$ is not uniquely defined. This means we can, if desired, choose some ∇g such that the vector potential $\mathbf{A}$ is divergence-free.

3.10 The streamfunction, streamlines and field lines

There is a case when the vector potential is especially useful. This is for solenoidal vector fields that are also two-dimensional. This is often the case in fluid dynamics, where flows are (at least approximately) incompressible with translational invariance along one axis.

Consider the vector field $\mathbf{F}(\mathbf{x})$ that satisfies $\nabla \cdot \mathbf{F} = 0$ everywhere. Furthermore, assume that the vector field is of the form $\mathbf{F}(\mathbf{x}) = (F_x(x, y), F_y(x, y), 0)$, i.e., the z component is zero and there is no variation in the z direction. In this special case, the Helmholtz decomposition requires only the vector potential $\mathbf{A}$ and, furthermore, it

has the simple form $\mathbf{A} = \psi \hat{\mathbf{z}}$. The scalar field $\psi(x, y)$ is known as the *streamfunction*, for reasons we shall see shortly.

Using Helmholtz and the streamfunction, we can write

$$\mathbf{F} = \nabla \times (\psi \hat{\mathbf{z}}). \quad (3.21)$$

In terms of the components of $\mathbf{F}$ we have

$$F_x(x, y) = \frac{\partial \psi}{\partial y}, \quad F_y(x, y) = -\frac{\partial \psi}{\partial x}, \quad F_z = 0. \quad (3.22)$$

The streamfunction is convenient because it represents a vector field in terms of a scalar field.

Exercise 3.10 on the streamfunction

Show that the vector potential $\psi(x, y)\hat{\mathbf{z}}$ is sufficient to describe solenoidal vector fields of the form $\mathbf{F}(\mathbf{x}) = (F_x(x, y), F_y(x, y), 0)$.

Exercise 3.11 on the streamfunction

Consider the streamfunction

$$\psi = E_0 xy,$$

for a positive constant E_0 . Contours of this function are plotted in figure 3.3. Calculate the associated vector field.

Exercise 3.12 on the streamfunction

Consider the streamfunction

$$\psi = E_0 \sin(x) \sin(y),$$

for a positive constant E_0 . Calculate the associated vector field $\mathbf{F}(\mathbf{x})$. Using the domain $0 \leq x \leq \pi$, $0 \leq y \leq \pi$, plot contours of ψ and representative vectors of $\mathbf{F}$ on the same axes.

The streamfunction has another valuable property. *Its contours are everywhere tangent to the direction of the vector field that it represents.* This is evident by examination of the example shown in figure 3.3, but it can also be shown to be generally true. Because of this property, contours of the streamfunction are called *streamlines*. If the vector field represents a steady fluid velocity, then the streamlines are the paths that fluid mass takes as it is carried by the flow.

More generally, lines that are everywhere tangent to the direction of a vector field are called *field lines*. These can be computed for three-dimensional vector fields by means other than the streamfunction.

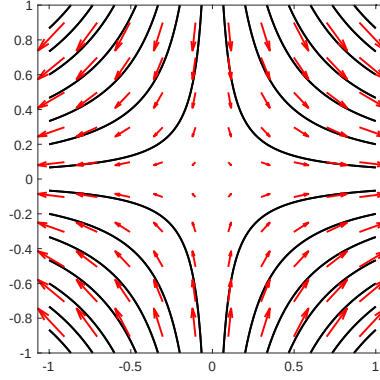


Figure 3.3: Contours of the streamfunction $\psi = E_0xy$ overlaid with vectors from the field $\mathbf{F} = \nabla \times \psi \hat{\mathbf{z}}$. The domain is $x \in [-1, 1]$, $y \in [-1, 1]$.

Exercise 3.13 on the meaning of streamlines

Consider a general 2-D vector field $\mathbf{F} = (F_x, F_y)$ associated with a streamfunction $\psi(x, y)$. Contours of the streamfunction are plotted at a constant interval in ψ .

1. Show that the contours of ψ are tangent to $\mathbf{F}$.
2. Noting that $|\mathbf{F}| = |\nabla \psi|$, show that the flow is faster where streamlines are closer together.

Hint: Recall that the gradient of ψ points in the direction in which ψ changes most rapidly and, therefore, $\nabla \psi$ is orthogonal to contours of ψ .

3.11 Orthogonal curvilinear coordinates

So far, our examination of the vector operators grad, div and curl has relied heavily on Cartesian coordinates. But like vectors, vector operators exist outside of any coordinate system, and hence it must be possible to express them within any valid coordinate system. So in this section we examine the general case in $\mathbb{R}^3$ and apply it to cylindrical and spherical coordinate systems.

The general case is of a coordinate system with coordinates u, v, w in which the position vector can be written as a function of those coordinates, $\mathbf{x} = \mathbf{x}(u, v, w)$. Then, a change in any one of u, v, w while holding the other two constant causes a change in $\mathbf{x}$. We express this idea as

$$d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial u} du + \frac{\partial \mathbf{x}}{\partial v} dv + \frac{\partial \mathbf{x}}{\partial w} dw. \quad (3.23)$$

Here $\partial \mathbf{x} / \partial u$ is a tangent vector along a line on which v and w are constant (and similarly for the other partial derivatives). These tangents represent the coordinate directions and must span $\mathbb{R}^n$, which they do if

$$\frac{\partial \mathbf{x}}{\partial u} \cdot \left(\frac{\partial \mathbf{x}}{\partial v} \times \frac{\partial \mathbf{x}}{\partial w} \right) \neq 0.$$

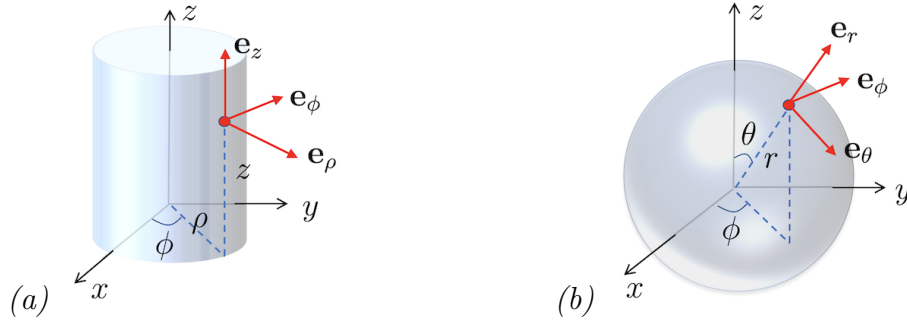


Figure 3.4: Curvilinear orthonormal coordinate systems. (a) Cylindrical. (b) Spherical. Source: DT.

Even better (and this is true of all the coordinate systems that we'll consider here) is a system for which the coordinate directions are mutually orthogonal. However, since the basis vectors (the tangent directions) are generally not constant in space, we will consider systems called *curvilinear orthogonal*.

Since the tangent vectors are parallel to the basis vectors, we can introduce proportionality constants and write

$$\frac{\partial \mathbf{x}}{\partial u} = h_u \hat{\mathbf{e}}_u, \quad \frac{\partial \mathbf{x}}{\partial v} = h_v \hat{\mathbf{e}}_v, \quad \frac{\partial \mathbf{x}}{\partial w} = h_w \hat{\mathbf{e}}_w.$$

We call $h_u, h_v, h_w > 0$ *scale factors*; the vectors $\hat{\mathbf{e}}_u, \hat{\mathbf{e}}_v, \hat{\mathbf{e}}_w$ form a right-handed orthonormal basis such that $\hat{\mathbf{e}}_u \times \hat{\mathbf{e}}_v = \hat{\mathbf{e}}_w$. We can use these to re-write (3.23) as

$$d\mathbf{x} = h_u \hat{\mathbf{e}}_u du + h_v \hat{\mathbf{e}}_v dv + h_w \hat{\mathbf{e}}_w dw. \quad (3.24)$$

It is instructive to consider the magnitude of $d\mathbf{x}$,

$$d\mathbf{x} \cdot d\mathbf{x} = h_u^2 du^2 + h_v^2 dv^2 + h_w^2 dw^2.$$

Recalling that each of the coordinates u, v, w may be changed while holding the others constant, it is clear that *the scale factors represent changes in length as we change the relevant coordinates*.

Let's look at the scale factors for three standard coordinate systems.

3.11.1 Cartesian, cylindrical and spherical coordinates

Cartesian coordinates Since this is our reference basis in which each of the coordinates already represents length, the scale factors are trivial,

$$\begin{aligned} \mathbf{x} = (x, y, z), \quad &\Rightarrow \quad h_x = h_y = h_z = 1, \\ &\Rightarrow \quad \hat{\mathbf{e}}_x = (1, 0, 0), \quad \hat{\mathbf{e}}_y = (0, 1, 0), \quad \hat{\mathbf{e}}_z = (0, 0, 1). \end{aligned}$$

Cylindrical coordinates See also section 2.4. The position vector is

$$\mathbf{x} = (\rho \cos \phi, \rho \sin \phi, z), \quad (3.25)$$

for $\rho \geq 0$ and $\phi \in [0, 2\pi)$ and $z \in \mathbb{R}$. In terms of Cartesians,

$$\rho = \sqrt{x^2 + y^2}, \quad \tan \phi = \frac{y}{x},$$

with the quadrant selected by the sign of the Cartesian coordinates. And, using the tangent vectors, it is straightforward to calculate

$$\begin{aligned}\hat{\mathbf{e}}_\rho &= (\cos \phi, \sin \phi, 0), \\ \hat{\mathbf{e}}_\phi &= (-\sin \phi, \cos \phi, 0), \\ \hat{\mathbf{e}}_z &= (0, 0, 1),\end{aligned}$$

with

$$h_\rho = h_z = 1, \quad h_\phi = \rho.$$

These vectors are shown in figure 3.4a in red. It is important to note that the vectors change direction with ϕ .

Spherical coordinates See also section 2.4. The position vector is

$$\mathbf{x} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta), \quad (3.26)$$

for $r \geq 0$ and $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. In terms of Cartesian coordinates,

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \tan \theta = \frac{\sqrt{x^2 + y^2}}{z}, \quad \tan \phi = \frac{y}{x},$$

(again, quadrant determined by sign of Cartesian components). And, using the tangent vectors, it is straightforward to calculate

$$\begin{aligned}\hat{\mathbf{e}}_r &= (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \\ \hat{\mathbf{e}}_\theta &= (\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta), \\ \hat{\mathbf{e}}_\phi &= (-\sin \phi, \cos \phi, 0),\end{aligned}$$

with

$$h_r = 1, \quad h_\theta = r, \quad h_\phi = r \sin \theta.$$

These vectors are shown in figure 3.4b in red. It is important to note that the vectors change direction with ϕ and θ .

3.11.2 Grad

Consider a scalar function $f(\mathbf{x}(u, v, w))$. An infinitesimal change in f can be written

$$df = \frac{\partial f}{\partial u} du + \frac{\partial f}{\partial v} dv + \frac{\partial f}{\partial w} dw.$$

A change in f can also be written

$$df = \nabla f \cdot d\mathbf{x} = \nabla f \cdot (h_u \hat{\mathbf{e}}_u du + h_v \hat{\mathbf{e}}_v dv + h_w \hat{\mathbf{e}}_w dw).$$

Since these two expressions of df must give the same result, we can compare terms on the right-hand sides, make use of orthogonality of the basis vectors, and infer that

$$\nabla = \frac{\hat{\mathbf{e}}_u}{h_u} \frac{\partial}{\partial u} + \frac{\hat{\mathbf{e}}_v}{h_v} \frac{\partial}{\partial v} + \frac{\hat{\mathbf{e}}_w}{h_w} \frac{\partial}{\partial w}, \quad (3.27)$$

where we have carefully placed the basis vectors to the left of the differential operators because they should not be differentiated. Hence we have

$$\begin{aligned} \nabla &= \hat{\mathbf{e}}_\rho \frac{\partial}{\partial \rho} + \frac{\hat{\mathbf{e}}_\phi}{\rho} \frac{\partial}{\partial \phi} + \hat{\mathbf{e}}_z \frac{\partial}{\partial z} \quad (\text{cylindrical}), \\ &= \hat{\mathbf{e}}_r \frac{\partial}{\partial r} + \frac{\hat{\mathbf{e}}_\theta}{r} \frac{\partial}{\partial \theta} + \frac{\hat{\mathbf{e}}_\phi}{r \sin \theta} \frac{\partial}{\partial \phi} \quad (\text{spherical}). \end{aligned}$$

It is important to note how the scale factors ensure that each term has the correct units, those of $1/\text{length}$.

3.11.3 Div, curl and Laplacian

Formulating the divergence, curl and Laplacian in curvilinear orthonormal coordinates is (almost) straightforward. We consider a vector field $\mathbf{F} = \mathbf{F}(u, v, w)$ written in terms of its components and basis vectors,

$$\mathbf{F} = F_u \hat{\mathbf{e}}_u + F_v \hat{\mathbf{e}}_v + F_w \hat{\mathbf{e}}_w. \quad (3.28)$$

(For the Laplacian, we consider a scalar field $f(u, v, w)$ but recall that it can be applied, component-wise, to a vector field.) Now we recall that the basis vectors are not necessarily constants, but can depend on the coordinates themselves. Instead of seeking a general formulation, we combine the Nabla operator from (3.27) with the general vector field (3.28) and substitute scale factors and basis vectors from section 3.11.1.

Cylindrical coordinates In cylindrical coordinates we have

$$\nabla \cdot \mathbf{F} = \frac{1}{\rho} \frac{\partial(\rho F_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}, \quad (3.29a)$$

$$\nabla \times \mathbf{F} = \left(\frac{1}{\rho} \frac{\partial F_z}{\partial \phi} - \frac{\partial F_\phi}{\partial z} \right) \hat{\mathbf{e}}_\rho + \left(\frac{\partial F_\rho}{\partial z} - \frac{\partial F_z}{\partial \rho} \right) \hat{\mathbf{e}}_\phi + \frac{1}{\rho} \left(\frac{\partial(\rho F_\phi)}{\partial \rho} - \frac{\partial F_\rho}{\partial \phi} \right) \hat{\mathbf{e}}_z, \quad (3.29b)$$

$$\nabla^2 f = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2}. \quad (3.29c)$$

Example 3.5 of the divergence in cylindrical coordinates

Verify the formula for divergence given in (3.29a).

Let's take this term by term. Starting with the radial term, we note that $h_\rho = 1$ and none of the basis vectors depends on ρ . Therefore,

$$\begin{aligned} \left(\frac{\hat{\mathbf{e}}_\rho}{h_\rho} \frac{\partial}{\partial \rho} \right) \cdot (F_\rho \hat{\mathbf{e}}_\rho + F_\phi \hat{\mathbf{e}}_\phi + F_z \hat{\mathbf{e}}_z) &= \hat{\mathbf{e}}_\rho \cdot \frac{\partial}{\partial \rho} (F_\rho \hat{\mathbf{e}}_\rho + F_\phi \hat{\mathbf{e}}_\phi + F_z \hat{\mathbf{e}}_z), \\ &= \hat{\mathbf{e}}_\rho \cdot \left(\hat{\mathbf{e}}_\rho \frac{\partial F_\rho}{\partial \rho} + \hat{\mathbf{e}}_\phi \frac{\partial F_\phi}{\partial \rho} + \hat{\mathbf{e}}_z \frac{\partial F_z}{\partial \rho} \right) = \frac{\partial F_\rho}{\partial \rho}. \end{aligned}$$

Next taking the z term, we note that similar to the radial term, $h_z = 1$ and none of the basis vectors depend on z . Hence, as for the radial term,

$$\left(\frac{\hat{\mathbf{e}}_z}{h_z} \frac{\partial}{\partial z} \right) \cdot (F_\rho \hat{\mathbf{e}}_\rho + F_\phi \hat{\mathbf{e}}_\phi + F_z \hat{\mathbf{e}}_z) = \frac{\partial F_z}{\partial z}.$$

The ϕ term is more difficult because two of the basis vectors, $\hat{\mathbf{e}}_\rho$ and $\hat{\mathbf{e}}_\phi$, are actually functions of ϕ . Moreover,

$$\begin{aligned} \frac{\partial \hat{\mathbf{e}}_\rho}{\partial \phi} &= \frac{\partial}{\partial \phi} (\cos \phi, \sin \phi, 0) = (-\sin \phi, \cos \phi, 0) = \hat{\mathbf{e}}_\phi, \\ \frac{\partial \hat{\mathbf{e}}_\phi}{\partial \phi} &= \frac{\partial}{\partial \phi} (-\sin \phi, \cos \phi, 0) = -(\cos \phi, \sin \phi, 0) = -\hat{\mathbf{e}}_\rho. \end{aligned}$$

Because of this and $h_\phi = \rho$ we obtain

$$\begin{aligned} \left(\frac{\hat{\mathbf{e}}_\phi}{h_\phi} \frac{\partial}{\partial \phi} \right) \cdot (F_\rho \hat{\mathbf{e}}_\rho + F_\phi \hat{\mathbf{e}}_\phi + F_z \hat{\mathbf{e}}_z) &= \frac{\hat{\mathbf{e}}_\phi}{\rho} \cdot \frac{\partial}{\partial \phi} (F_\rho \hat{\mathbf{e}}_\rho + F_\phi \hat{\mathbf{e}}_\phi + F_z \hat{\mathbf{e}}_z), \\ &= \frac{\hat{\mathbf{e}}_\phi}{\rho} \cdot \left(\hat{\mathbf{e}}_\rho \frac{\partial F_\rho}{\partial \phi} + \hat{\mathbf{e}}_\phi F_\rho + \hat{\mathbf{e}}_\phi \frac{\partial F_\phi}{\partial \phi} - \hat{\mathbf{e}}_\rho F_\phi + \hat{\mathbf{e}}_z \frac{\partial F_z}{\partial \phi} \right), \\ &= \frac{F_\rho}{\rho} + \frac{1}{\rho} \frac{\partial F_\phi}{\partial \phi}. \end{aligned}$$

Now, summing all three contributions we obtain

$$\begin{aligned} \nabla \cdot \mathbf{F} &= \frac{\partial F_\rho}{\partial \rho} + \frac{F_\rho}{\rho} + \frac{1}{\rho} \frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}, \\ &= \frac{1}{\rho} \frac{\partial (\rho F_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}, \end{aligned}$$

where we used the product rule (backwards) to obtain the last line. This verifies the result (3.29a).

Spherical coordinates In spherical coordinates we have

$$\nabla \cdot \mathbf{F} = \frac{1}{r^2} \frac{\partial(r^2 F_r)}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial(\sin \theta F_\theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial F_\phi}{\partial \phi}, \quad (3.30a)$$

$$\begin{aligned} \nabla \times \mathbf{F} = & \frac{1}{r \sin \theta} \left(\frac{\partial(\sin \theta F_\phi)}{\partial \theta} - \frac{\partial F_\theta}{\partial \phi} \right) \hat{\mathbf{e}}_r + \frac{1}{r} \left(\frac{1}{\sin \theta} \frac{\partial F_r}{\partial \phi} - \frac{\partial(r F_\phi)}{\partial r} \right) \hat{\mathbf{e}}_\theta \\ & + \frac{1}{r} \left(\frac{\partial(r F_\theta)}{\partial r} - \frac{\partial F_r}{\partial \theta} \right) \hat{\mathbf{e}}_\phi, \end{aligned} \quad (3.30b)$$

$$\nabla^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}. \quad (3.30c)$$

It is convenient to write the curl (3.30b) using the determinant as

$$\nabla \times \mathbf{F} = \begin{vmatrix} \frac{\hat{\mathbf{e}}_r}{r^2 \sin \theta} & \frac{\hat{\mathbf{e}}_\theta}{r \sin \theta} & \frac{\hat{\mathbf{e}}_\phi}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\ F_r & r F_\theta & r \sin \theta F_\phi \end{vmatrix}. \quad (3.31)$$

Exercise 3.14 on the divergence in spherical coordinates

Verify the formula for divergence given in (3.30a). Start by recalling that $\hat{\mathbf{e}}_r = \hat{\mathbf{e}}_r(\theta, \phi)$ and $\hat{\mathbf{e}}_\theta = \hat{\mathbf{e}}_\theta(\theta, \phi)$ and $\hat{\mathbf{e}}_\phi = \hat{\mathbf{e}}_\phi(\phi)$; compute the partial derivatives of these basis vectors.

General formulae for these vector operators are available for arbitrary curvilinear orthogonal coordinate systems, but we do not consider them here.

Summary of §3

- Index notation is a valuable tool for manipulating vector operators and vector equations. A repeated index within a term implies a summation over the values that the index can take.
- The bold Nabla symbol ∇ is a vector operator given in Cartesian coordinates by

$$\nabla = \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i},$$

where a summation over the repeated index is implied. It is used in the gradient, divergence, and curl operators.

- The gradient of a scalar field (e.g., $\nabla\phi(\mathbf{x})$) is formed by the Nabla operator acting on a scalar field. This gives a vector field expressing the maximum slope of the field at each point, and the direction in which that slope is realised.
- The divergence of a vector field (e.g., $\nabla \cdot \mathbf{F}(\mathbf{x})$) is formed by the dot product of the Nabla operator with the vector field. It is a scalar field expressing the net rate of transport out of an infinitesimal volume surrounding the point of evaluation.
- The curl of a vector field (e.g., $\nabla \times \mathbf{F}(\mathbf{x})$) is formed from the cross product of the Nabla operator with a vector field. It measures the circulation of the vector field and is a vector field itself.
- The Laplacian operator $\nabla \cdot \nabla = \nabla^2$ is a scalar operator. Applied to a scalar field, it gives the local curvature of that field.
- The Helmholtz decomposition is a way of expressing a general vector field in terms of two parts. The solenoidal part has zero divergence and is expressed as the curl of a vector potential. The irrotational part has zero circulation and is expressed as the gradient of a scalar potential.
- The gradient, divergence and curl have different expressions in different coordinate systems. These are obtained using the scale factors that relate changes in the curvilinear coordinate to changes in the Cartesian position.

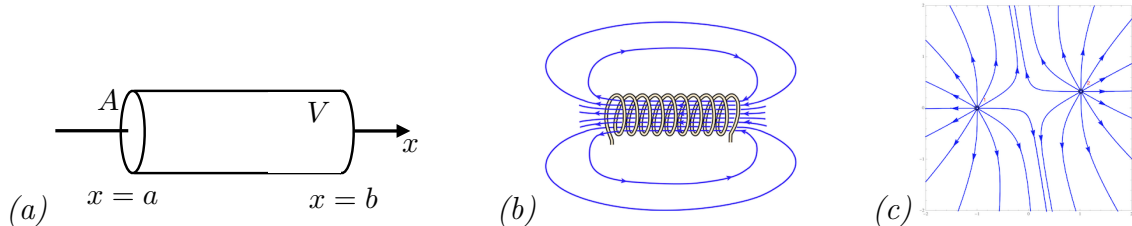


Figure 4.1: The divergence theorem. (a) A cylindrical volume aligned with the x direction. (b) Magnetic field lines, tangent to $\mathbf{B}$, around a solenoid coil. Source: hyperphysics web page. (c) Electric field lines, tangent to $\mathbf{E}$, surrounding two positive, electrical point-charges. Source: DT.

4 Integral Theorems

In this section we consider generalisations of the fundamental theorem of calculus

$$\int_a^b \frac{df}{dx} dx = f(b) - f(a)$$

for vector calculus. These are important because they are useful in deriving and solving the equations of continuum mechanics (and almost all other physical theories), but also because they give insight into the basic meaning of the mathematical operations of vector calculus.

4.1 Divergence Theorem

The Divergence Theorem, also known as Gauss' Theorem, is

$$\int_V \nabla \cdot \mathbf{F} dV = \int_S \mathbf{F} \cdot d\mathbf{S}. \quad (4.1)$$

The integral on the right is over a surface S that encloses the volume V . This surface must be piece-wise smooth and with an outward-pointing normal vector. It is common to see the bounding surface S represented as ∂V . For most purposes, this is simply a different but equivalent notation. However, there is a deeper mathematical connection between a boundary and a derivative that we won't go into here.

A simplified case We can see the connection to the Fundamental Theorem of Calculus if we consider a simplified case. Let's apply the Divergence Theorem to the vector field $\mathbf{F} = \phi(\mathbf{x})\mathbf{g}$, where $\mathbf{g}$ is a constant vector field and ϕ is a scalar field. Then we can take $\mathbf{g}$ outside the integral and write $\mathbf{g} \cdot \int_V \nabla \phi dV = \mathbf{g} \cdot \int_S \phi d\mathbf{S}$. Since this must be true for any $\mathbf{g}$,

$$\int_V \nabla \phi dV = \int_S \phi d\mathbf{S}. \quad (4.2)$$

This is the Divergence Theorem for scalar fields.

To fix ideas, let's suppose that $\phi = \phi(x)$, a function that is independent of y and z . Let's further consider that V is a section of cylindrical pipe, shown in figure 4.1a,

with cross sectional area A aligned with the x direction and spanning the distance from $x = a$ to $x = b$. By symmetry we can infer that the y and z components of the equation are trivially satisfied, leaving only the x component. Then, since we have $dV = A dx$, equation (4.2) becomes

$$A \int_a^b \frac{d\phi}{dx} dx = A [-\phi(a) + \phi(b)],$$

which is the Fundamental Theorem!

4.1.1 The meaning of the divergence

To better understand the meaning of the divergence theorem, we consider what happens when V becomes very small around a point $\mathbf{x}_0$. In that case,

$$\int_V \nabla \cdot \mathbf{F} dV \approx V (\nabla \cdot \mathbf{F})|_{\mathbf{x}_0} \quad (4.3)$$

because the integrand is approximately constant over the (vanishingly small) volume. Using this in the Divergence Theorem we can write

$$\nabla \cdot \mathbf{F} = \lim_{V \rightarrow 0} \frac{1}{V} \int_S \mathbf{F} \cdot d\mathbf{S}.$$

This equation gives us the meaning of the divergence, independent of the coordinate system. *The divergence is the net outward flow (when $\nabla \cdot \mathbf{F} > 0$) or inward flow (when $\nabla \cdot \mathbf{F} < 0$), per unit volume, of a small volume surrounding a point in space.*

When a vector field has zero divergence, such as for the magnetic field shown in figure 4.1b where $\nabla \cdot \mathbf{B} = 0$, it means that there is no net flux into/out of any small volume. Any volume you can draw has the same magnetic flux entering as exiting. The electric field $\mathbf{E}$ is different; it has its source in electrical charge. It is governed by $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, where $\rho(\mathbf{x})$ is the charge density in space and ϵ_0 is a constant. This equation states that where the charge density is zero, $\mathbf{E}$ is divergence free. It also states that field lines have their origin in places where there is electric charge. This situation is shown in figure 4.1c.

4.1.2 A proof(ish)

A simple proof of the Divergence Theorem (4.1) can be developed by considering a cubic volume V in $\mathbb{R}^3$. We can divide this volume into a grid of tiny cubes with volumes δV . On each tiny cube, using eqn. (4.3), the volume integral of the divergence is approximately $\delta V \nabla \cdot \mathbf{F}$; we can sum these over all tiny cubes to approximate the volume integral over V . On each tiny cube, we can also re-write the divergence integral in terms of a surface integral. However, when we sum the surface integrals over all the tiny cubes, most of them cancel because the cubes share faces with their neighbours; an outward flux (+ve) from one cube is an inward flux (−ve) into the neighbouring cube. The only surface contributions that don't cancel are those at the boundary of V , where there is no adjacent tiny cube. In the limit of $\delta V \rightarrow 0$, this is the Divergence Theorem.

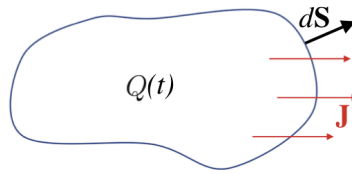
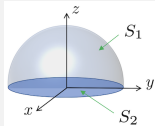


Figure 4.2: Why does the nightclub look like a potato? Source: DT.

Example 4.1 of the Divergence Theorem in action

Consider the hemispherical region with radius R bounded by $S = S_1 + S_2$



and the vector field $\mathbf{F} = (0, 0, z + R)$. Use the Divergence Theorem to compute the net flux out of S .

We first note that $\nabla \cdot \mathbf{F} = 1$. Therefore we have

$$\int_V \nabla \cdot \mathbf{F} dV = \int_V dV = \frac{1}{2} \left(\frac{4}{3} \pi R^3 \right).$$

To confirm that this is the same result as obtained by the surface integral, do exercise 4.1.

Exercise 4.1 on on the Divergence Theorem in action

Consider the hemispherical region with radius R bounded by $S = S_1 + S_2$ and the vector field $\mathbf{F} = (0, 0, z + R)$, as in example 4.1. Evaluate the integral $\int_S \mathbf{F} \cdot d\mathbf{S}$ and show that you get the same result. Hint: recall that $z = R \cos \theta$ and that $\hat{\mathbf{n}} = (x, y, z)/R$ on the hemispherical surface.

4.1.3 An application of the Divergence Theorem

The most important application of the Divergence Theorem is to conservation laws. These are physical statements about how a quantity can change locally when that quantity is *conserved*. Conserved means that it cannot be created or destroyed — but it can still move around.

Leaving aside birth and death, people are conserved. So let's think about a nightclub. If we want to know how many people are in the nightclub each hour from 5pm to 5am, we don't go in and count them. We count the people who are in the nightclub at 5pm and call this quantity Q . After this, we stand at the door and measure the number of people entering and leaving; we update Q accordingly, giving a running tally $Q(t)$. There's no need to worry about people vanishing entirely; people are conserved. Nor is there a need to worry about people de-materialising in the club and re-materialising in their homes. They must leave through the door; people are conserved *locally*.

Electric charge is also conserved, but we don't usually measure individual units of

charge. We measure charge density $\rho(\mathbf{x}, t)$, the amount of charge per unit volume at a particular place and time. The total amount of electric charge within the volume V of a nightclub at a particular time t is given by

$$Q(t) = \int_V \rho(\mathbf{x}, t) dV. \quad (4.4)$$

Now suppose that we want to know how Q changes with time. We need a running tally. But as shown in figure 4.2, charge can leave the nightclub through any wall or door, so we need to integrate the flux of ρ through the surface S bounding the nightclub. We can write,

$$\frac{dQ}{dt} = - \int_S \mathbf{J} \cdot d\mathbf{S}, \quad (4.5)$$

where the vector field $\mathbf{J}(\mathbf{x}, t)$ is the flux of charge (units of charge/area/time). The surface integral is the equivalent of counting people at the door; it is a count of the rate at which charge is going in the direction of the surface normal over the whole surface S . Since the surface normal points outward, the integral gives the rate at which charge is *leaving* — hence the negative sign.

We can substitute (4.4) into the left-hand side of (4.5) and use the Divergence Theorem on the right hand side,

$$\int_V \frac{\partial \rho}{\partial t} dV = - \int_V \nabla \cdot \mathbf{J} dV. \quad (4.6)$$

In this equation, the time derivative moved inside the integral because the nightclub is fixed (and so V is not changing in time).

Now, let's drop this nightclub silliness; equation (4.6) applies to any fixed volume V . So let's assume that V is small enough that things inside it are roughly uniform (e.g., eqn. (4.3)). Then we can replace the integration over V with a multiplication by V . Then, rearranging and dividing by V we obtain

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \quad (4.7)$$

This is the local conservation equation that we sought. It says that the rate of change of charge density local to a point in space is given by the negative divergence of the flux of charge density at that point.

But equation (4.7) is entirely general — it applies to any conserved quantity. Mass is a conserved quantity, so we can think of $\rho(\mathbf{x}, t)$ as the density field of a fluid. The flux of mass in a fluid with velocity field $\mathbf{v}(\mathbf{x}, t)$ is $\mathbf{J} = \rho \mathbf{v}$. Hence we have

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0.$$

In many fluid flows, the density is constant and uniform. If that is the case, then mass conservation simplifies to

$$\nabla \cdot \mathbf{v} = 0, \quad (\text{for } \rho = \text{const.})$$

which means that an incompressible flow is solenoidal. We'll say more about this equation when we get to continuum mechanics.

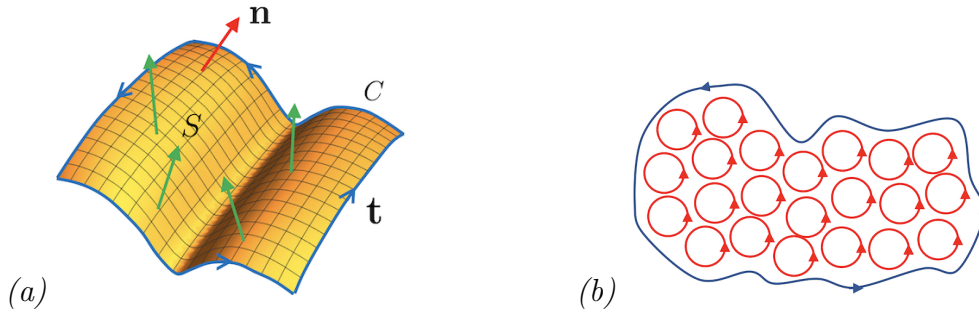


Figure 4.3: Stokes Theorem. (a) A surface S and bounding curve C for Stokes Theorem. The surface has normal vector $\hat{\mathbf{n}}$ shown in red. Green arrows represent $\nabla \times \mathbf{F}$. The tangent vector to the curve C is $\mathbf{t}$. (b) A sketched proof of Stokes Theorem. Red lines represent the normal-directed curl over tiny patches; the blue line represents the circulation as the sum of these patches. Source: DT.

4.2 Stokes Theorem

Stokes Theorem is analogous to the Divergence Theorem, but it applies to the curl. It is

$$\int_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S} = \oint_C \mathbf{F} \cdot d\mathbf{x}, \quad (4.8)$$

where C is a curve that encloses the surface S as shown in figure 4.3a.⁶ The orientations of S and C should be compatible. In particular, the direction of circulation on C given by the tangent $\mathbf{t}$ should be related to the outward-pointing normal $\hat{\mathbf{n}}$ such that $\mathbf{t} \times \hat{\mathbf{n}}$ points away from the surface (this becomes important in exercise 4.2 below).

Many surfaces can share the same bounding curve C . By Stokes Theorem, all of these surfaces must have the same integral of $(\nabla \times \mathbf{F}) \cdot d\mathbf{S}$.

A kinda-proof Something like a proof of Stokes theorem follows along the same lines as our proof(ish) of the Divergence Theorem. It is sketched in figure 4.3b. We consider a surface, flattened for simplicity, with normal vector pointing toward us. The surface is divided into many small, adjacent patches with areas δA . On each patch we calculate the curl and project it onto the surface normal. These curl-patches are represented by the red arrows. Summing over these patches leads to internal cancellation; only the flow along the boundary of A remains. This is the circulation on C .

4.2.1 The meaning of the curl

As we did for the Divergence Theorem, we can examine Stokes Theorem when the surface region becomes very small,

$$\int_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S} \approx A \hat{\mathbf{n}} \cdot (\nabla \times \mathbf{F})|_{\mathbf{x}_0},$$

⁶When the surface S lies in a plane, this is known as Green's Theorem.

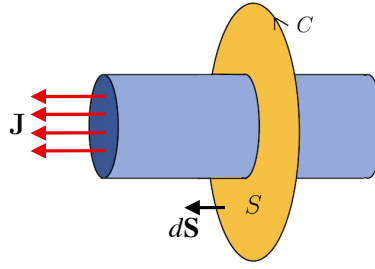


Figure 4.4: Electrical current density $\mathbf{J}$ through a wire. Disk-shaped surface S perpendicular to the wire, with vector area element $d\mathbf{S}$ parallel to the wire, and bounding curve C . Source: DT.

where A is the area of the surface S and $\hat{\mathbf{n}}$ is the unit normal vector. Using this in Stokes theorem,

$$\hat{\mathbf{n}} \cdot (\nabla \times \mathbf{F}) = \lim_{A \rightarrow 0} \frac{1}{A} \int_C \mathbf{F} \cdot d\mathbf{x}. \quad (4.9)$$

This equation tells us that the amount of curl in the direction of $\hat{\mathbf{n}}$ is a measure of the circulation in a plane normal to $\hat{\mathbf{n}}$.

4.2.2 An application of Stokes Theorem

We will use an example from electromagnetism to illustrate the use of Stokes Theorem in a real problem. We do not intend to make this a deep lesson in electromagnetism, but a few words are needed. Maxwell's equations of electromagnetism tell us that at steady state, the relationship between the magnetic field $\mathbf{B}$ and the electric current density $\mathbf{J}$ (i.e., the flux) is

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \quad (4.10)$$

where μ_0 is a constant.

We consider a straight and infinitely long wire carrying a steady electrical current I . A section of the wire is shown in figure 4.4. We wish to solve for the magnetic field $\mathbf{B}(\mathbf{x})$ associated with this current. We begin by drawing a surface, S , as shown in fig. 4.4 and a curve C on the boundary of the surface. The surface has a vector area element $d\mathbf{S}$; we dot this with both sides of equation (4.10) and integrate,

$$\int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{S} = \mu_0 \int_S \mathbf{J} \cdot d\mathbf{S}.$$

The right-hand side of this equation has the integral of electric flux along the wire, which is the current, I . The left-hand side can be treated with Stokes Theorem. We obtain

$$\int_C \mathbf{B} \cdot d\mathbf{x} = \mu_0 I. \quad (4.11)$$

Now we assume that S is a disk of radius ρ with a normal in the z direction, along the wire. Its bounding curve C is a circle parameterised as

$$\mathbf{x} = \rho(\cos \phi, \sin \phi, 0),$$

from which we calculate the tangent vector as

$$\mathbf{t} = \frac{d\mathbf{x}}{d\phi} = \rho(-\sin \phi, \cos \phi, 0).$$

Using what we understand about the curl (i.e., that it represents circulation), we can guess that $\mathbf{B}$ is in the direction of $\mathbf{t}$. But we don't know the magnitude of $\mathbf{B}$, or how that magnitude depends on ρ . So we write our guess as

$$\mathbf{B}(\mathbf{x}) = b(\rho)\hat{\mathbf{t}} = b(\rho)(-\sin \phi, \cos \phi, 0),$$

where $b(\rho)$ is an unknown function of radius. Using this and the tangent to C we can obtain

$$\mathbf{B} \cdot d\mathbf{x} = \mathbf{B} \cdot \mathbf{t} d\phi = \rho b(\rho) d\phi,$$

and substituting this into (4.11) gives

$$\mu_0 I = \int_C \mathbf{B} \cdot d\mathbf{x} = \int_0^{2\pi} d\phi \rho b(\rho) = 2\pi \rho b(\rho).$$

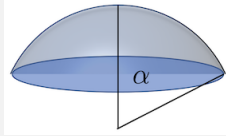
Rearranging this gives $b(\rho) = \mu_0 I / 2\pi \rho$, which we can use to write the solution for the magnetic field as

$$\mathbf{B} = \frac{\mu_0 I}{2\pi \rho} (-\sin \phi, \cos \phi, 0).$$

This solution tells us that the magnetic field circulates around the wire and that its magnitude drops off with $1/\rho$ away from the wire.

Example 4.2 of use of Stokes theorem

Use Stokes theorem to calculate $\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$ where S is the hemispherical cap



and $\mathbf{F}(\mathbf{x}) = (0, xz, 0)$. To verify the answer, note that defining $\mathbf{G} = \nabla \times \mathbf{F} = (-x, 0, z)$ we recognise the surface integral as what we computed in example 2.9, where we obtained $\pi R^3 \sin^2 \alpha \cos \alpha$.

We parameterise the Cartesian coordinates of the curve C bounding the cap in terms of spherical coordinates,

$$\begin{aligned} \mathbf{x} &= R(\sin \alpha \cos \phi, \sin \alpha \sin \phi, \cos \alpha), \\ d\mathbf{x} &= R(-\sin \alpha \sin \phi, \sin \alpha \cos \phi, 0) d\phi. \end{aligned}$$

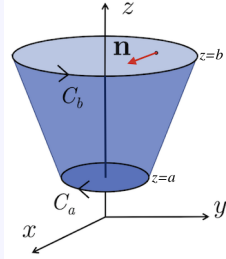
Using this tangent vector and the given field $\mathbf{F}$,

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{x} &= \int_0^{2\pi} d\phi (xz) R \sin \alpha \cos \phi, \\ &= R^3 \sin^2 \alpha \cos \alpha \int_0^{2\pi} d\phi \cos^2 \phi, \\ &= \pi R^3 \sin^2 \alpha \cos \alpha. \end{aligned}$$

This is in agreement with the previously obtained result.

Exercise 4.2 on use of Stokes Theorem

Use Stokes Theorem to calculate $\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$ where S is the cone segment



given by $\rho(z) = z$ and $\mathbf{F}(\mathbf{x}) = (0, xz, 0)$. To verify the answer, note that defining $\mathbf{G} = \nabla \times \mathbf{F} = (-x, 0, z)$ we recognise the surface integral as what we computed in exercise 2.8, where we obtained $\pi(b^3 - a^3)$. Use cylindrical coordinates to parameterise the Cartesian coordinates of C_a and C_b . Note that they circulate in opposite directions!

We have focused thus far on integration and differentiation of vector functions. But we can now extend our discussion to differential equations that are written in terms of vector calculus. The most basic of these are the Laplace and Poisson equations, which we consider below. But first we show how these equations arise by consideration of classical gravitation.

4.3 Gravity and Gauss' law

Gravity is a fundamental force and, at non-relativistic conditions, is described efficiently with vector fields. Here is Newton's law of gravitation for two point-masses,

$$\mathbf{F}(\mathbf{x}) = -\frac{GMm}{r^2} \hat{\mathbf{r}}, \quad (4.12)$$

where $G = 6.6743 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ is the Universal Constant of Gravitation, M, m are the attracting masses, and r is the distance separating them. We use a common notation for the radial basis vector, $\hat{\mathbf{r}} = \hat{\mathbf{e}}_r$. Note that we are taking a coordinate system centred on one mass (say M) and calculating the vector force on the other mass. ⁷

Why does the gravitational force decay as r^{-2} , just like the electrostatic force? Should this surprise us? No, the reason is not profound. First, both gravity and electrostatic forces are described by fields and second, space has three dimensions. We see in the next section how vector calculus can unravel this mystery.

⁷It is worth noting (but not essential for this course) that the Coulomb force between two particles with charges Q and q is

$$\mathbf{F}(\mathbf{x}) = \frac{Qq}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}}.$$

Note that this is identical in form to Newton's law of gravitation. There are different constants of proportionality (ϵ_0 is another fundamental constant of nature), but the dependence on r is the same. Everything we say about gravity thus applies to electrostatics, though there is no (known) physical connection between the two.

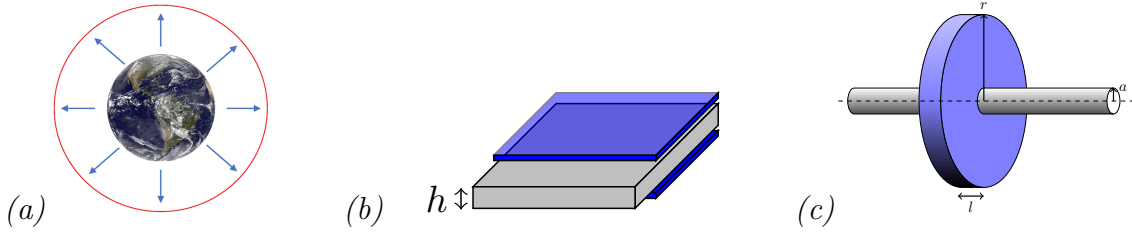


Figure 4.5: (a) Earth at the centre of a spherical surface of radius r . Source: DT. (b,c) Gravity sources (grey) and surfaces for Gauss-flux-method integrals (blue). Source: Connor & Connor, lecture notes on Potential Fields Geophysics, USF.

4.3.1 Gauss' law

In this section we'll assume that $M \gg m$, so that the acceleration of the body with mass M caused by the body with mass m is negligible (the converse is not true). We'll therefore refer to m as the *test mass*. We can define the vector force per unit mass on the test mass as $\mathbf{g}$, which has units of acceleration. This gives

$$\mathbf{g}(\mathbf{x}) = \frac{\mathbf{F}}{m} = -\frac{GM}{r^2} \hat{\mathbf{r}}. \quad (4.13)$$

Now we have the familiar form $\mathbf{F} = \mathbf{g}m$.

Suppose that we didn't know Newton's law but we knew Gauss' law (in integrated form), which is

$$\int_S \mathbf{g} \cdot d\mathbf{S} = -4\pi GM. \quad (4.14)$$

This equation states that the *total* gravitational force per unit mass through a surface S enclosing some mass M is proportional to M . Since $d\mathbf{S}$ is an outward-pointing vector area element, the minus sign means that the force is inward-pointing, i.e., attractive.

We can evaluate the left-hand side of (4.14) if we make one assumption – that $\mathbf{g}$ is spherically symmetric, i.e., if we assume $\mathbf{g} = g(r)\hat{\mathbf{r}}$. This will be true if the mass distribution within the body is spherically symmetric, which is a good approximation for Earth. Taking the Earth as our gravitating mass M , we enclose it in a spherical surface with radius r as in figure 4.5a, centred on the centre of mass. Then, recalling the vector area element $d\mathbf{S} = r^2 \sin \theta d\theta d\phi \hat{\mathbf{r}}$, we have

$$\begin{aligned} \int_S \mathbf{g} \cdot d\mathbf{S} &= \int_0^{2\pi} d\phi \int_0^\pi d\theta g(r) r^2 \sin \theta, \\ &= 4\pi r^2 g(r). \end{aligned}$$

Then, using Gauss' law (4.14) and rearranging, we find that

$$\mathbf{g}(r) = -\frac{GM}{r^2} \hat{\mathbf{r}},$$

which is exactly Newton's law of gravitation for a point mass M . Note, however, that we have derived Newton's law without assuming anything about M besides

that it is spherically symmetrical and contained within the radius r . Determining the gravity field in this way is called *Gauss' flux method*.

This result tells us why gravitational attraction decays as $1/r^2$. It is because the surface area of the sphere containing the attracting mass increases with r^2 . According to Gauss' law, the force per unit mass $\mathbf{g}$ is a flux (and hence gets integrated over an area).

Example 4.3 of Gauss' flux method

Determine the gravitational force per unit mass above a flat slab with areal density ϱ (kg/m²) and area A that is much larger than the typical height above it at which we might measure gravity.

We take a coordinate system where z increases normal to the slab. We define our vector area element using the normal to the slab $d\mathbf{S} = \hat{\mathbf{n}}dS$. Above the sheet, $\mathbf{g} = -g\hat{\mathbf{z}}$ and $\mathbf{n} = +\hat{\mathbf{z}}$; below it, $\mathbf{g} = +g\hat{\mathbf{z}}$ and $\mathbf{n} = -\hat{\mathbf{z}}$. Thus $\mathbf{g} \cdot \mathbf{n} = -g$ on both large faces. We set up the Gauss flux integral (4.14) as

$$-2 \int_A g(z) dA = -4\pi GM,$$

with slab and surface shown in figure 4.5b. It is important to note that the surface must enclose the slab, which means that it must be above *and below* the slab. The field has equal magnitude but opposite direction above and below the sheet.

We neglect the edges of the slab, assuming that they are very far away from the observation point (i.e., the slab is effectively infinite). Then the left-hand side integrates to $-2g(z)A$ and we rearrange to obtain

$$g(z) = 2\pi GM/A = 2\pi G\varrho.$$

Note that this attraction is independent of height (for heights much less than the lateral extent of the slab). We can write this in terms of a gravity anomaly if we assume the slab has a thickness h and replaces material with a different density. Then $\varrho = h\Delta\rho$ and $\Delta g(z) = 2\pi Gh\Delta\rho$. This is the Bouguer gravity anomaly.

Exercise 4.3 on Gauss' flux method

Determine the gravity anomaly surrounding an infinitely long cylindrical mass with radius a and density difference $\Delta\rho$ (kg/m³) relative to the embedding medium, as shown in figure 4.5c.

A different version of Gauss' law is more commonly used. To derive it, we write the mass M contained within the surface S as a volume integral of the density field $\rho(\mathbf{x})$,

$$M = \int_V \rho(\mathbf{x}) dV.$$

Then we use the Divergence theorem to express $\int_S \mathbf{g} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{g} dV$. Substituting

these into (4.14) and rearranging, we have

$$\int_V [\nabla \cdot \mathbf{g} + 4\pi G\rho(\mathbf{x})] dV = 0,$$

for any volume V . Since the volume (and enclosing surface) is arbitrary, we can choose one small enough that the integrand is roughly constant within. Then we have Gauss' law in differential form,

$$\nabla \cdot \mathbf{g} = -4\pi G\rho(\mathbf{x}). \quad (4.15)$$

This equation states that the gravitational field converges where there is mass or, in other words, mass creates gravitational attraction.

4.3.2 Gravitational potential

The above examples assumed symmetries that are not always present. In that case, we need to use another property of the gravity field: that it is conservative. Recall that a conservative field has $\nabla \times \mathbf{g} = 0$ and also $\oint_C \mathbf{g} \cdot d\mathbf{x} = 0$. You can check that both of these hold for the examples above.

Recall that for conservative fields, we can write the vector field in terms of a scalar potential,

$$\mathbf{g}(\mathbf{x}) = -\nabla\Phi(\mathbf{x}).$$

Here we have introduced the gravity potential Φ and related it to the gravitational force per unit mass $\mathbf{g}$. Note that the minus sign in this definition is purely a convention. Using the potential, we can re-write (4.15) as

$$\nabla^2\Phi = 4\pi G\rho(\mathbf{x}). \quad (4.16)$$

An equation of this form is known as a *Poisson equation*. When the right-hand side is zero (i.e., in free space, outside a gravitating body) we have

$$\nabla^2\Phi = 0, \quad (4.17)$$

which is known as the *Laplace equation*. These are fundamental differential equations in vector calculus and appear in many physical contexts.

Summary of §4

- The Divergence theorem (also known as Gauss' theorem) states that

$$\int_V \nabla \cdot \mathbf{F} dV = \int_S \mathbf{F} \cdot d\mathbf{S}.$$

The divergence of a vector field is the net outward flow from an infinitesimal volume surrounding a point.

- Stokes theorem states that

$$\int_S \nabla \times \mathbf{F} \cdot d\mathbf{S} = \oint_C \mathbf{F} \cdot d\mathbf{x},$$

where C is a curve that bounds the surface S . The curl of a vector field is a measure of the circulation of the field.

- Gauss' law of gravity in integral form is

$$\int_S \mathbf{g} \cdot d\mathbf{S} = -4\pi GM$$

and means that the total force (per unit mass) through a closed surface S is proportional to the total mass M within that surface.

- Gauss' law in differential form is

$$\nabla \cdot \mathbf{g} = -4\pi G\rho(\mathbf{x})$$

and means that mass at a point $\mathbf{x}$ causes a gravitational attraction to that point.

- The gravitational potential Φ is defined by $\mathbf{g} = -\nabla\Phi$ and turns Gauss' law in differential form into the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho(\mathbf{x}).$$

Where the right-hand side is zero, the Poisson equation becomes the Laplace equation $\nabla^2\Phi = 0$.

5 Partial differential equations

In this section we consider the Laplace and Poisson equations, which arise in our consideration of gravitation, above. These two partial differential equations (PDE) are extremely important across applied mathematics and physics. They provide a context for developing two methods for solving linear PDEs, Green's functions and separation of variables. What follows in this section is a very brief introduction to the topic of PDEs.

5.1 The Laplace equation

The Laplace equation is a scalar equation for a scalar field, e.g., $\psi(\mathbf{x})$,

$$\nabla^2 \psi = 0. \quad (5.1)$$

It is typically solved in two or three dimensions, but can be solved in an arbitrary number of dimensions. The Laplacian operator $\nabla^2 = \nabla \cdot \nabla$ is a scalar operator that takes a different form in different coordinate systems (see sec. 3.11.3).

Solutions to the Laplace equation differ depending on the geometry of the domain and the boundary conditions. However, all solutions to Laplace are known as *harmonic functions*. We see examples below.

Isotropic solutions Recall that isotropic solutions are the same in all directions. We can therefore seek isotropic solutions in spherical coordinates, but where the solution is a function of r only, i.e., $\psi(\mathbf{x}) = \psi(r)$ for some scalar function ψ . This is an important solution because it allows us to describe the field surrounding a point source (e.g., the potential field surrounding a point-mass).

Using the expression of the Laplace operator in spherical coordinates (3.30c),

$$\begin{aligned} \nabla^2 \psi = 0 & \Rightarrow \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial \psi}{\partial r} = 0, \\ & \Rightarrow \psi(r) = \frac{A}{r} + B, \quad (\text{spherical coords.}) \end{aligned} \quad (5.2)$$

where A and B are constants of integration (check this solution by direct substitution). The part of this solution A/r diverges as $r \rightarrow 0$, which is a clue that it doesn't work at $r = 0$.

We can also obtain an isotropic solution in cylindrical coordinates. It is independent of azimuth and so isotropic in $\mathbb{R}^2$, but not really isotropic in $\mathbb{R}^3$. Using the expression for the Laplace operator in cylindrical coordinates (3.29c) and denoting radius with r ,

$$\begin{aligned} \nabla^2 \psi = 0 & \Rightarrow \frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial \psi}{\partial r} = 0, \\ & \Rightarrow \psi(r) = C \ln r + D, \quad (\text{cylindrical coords.}) \end{aligned} \quad (5.3)$$

where C and D are constants of integration. Note that this diverges to $-\infty$ as $r \rightarrow 0$, along the full z axis.

5.2 The Poisson equation

The Poisson equation is also a scalar equation for a scalar field. The difference with the Laplace equation is only that there is a *source term* on the right-hand side of the equation. If, for example, we are solving for the gravitational potential, the source term describes a spatial distribution of mass. Mass is the source of gravitational potential.

In this section we will consider a tidier version of (4.16),

$$\nabla^2 \psi = -\rho(\mathbf{x}), \quad (5.4)$$

where $\psi = -\Phi/4\pi G$ is a rescaled gravity potential.

An isotropic solution to the Laplace equation $\nabla^2\psi = 0$ had the form $\psi(r) = A/r + B$. We noted that this solution is awkward at $r = 0$, because it is singular there. To understand what is happening at the origin, we interpret ψ as a scalar potential. Indeed, if ψ is the scaled gravitational potential, we might wonder what mass is creating this gravity? To answer this question, use Gauss' law in integral form,

$$\int_S \nabla\psi \cdot d\mathbf{S} = -\lambda,$$

where λ represents a mass within S that creates the gravity field associated with ψ .

We can take S to be a sphere of radius r . Using the Divergence theorem to rewrite as $\int_V \nabla^2\psi dV = -\lambda$, and substituting the Poisson equation (5.4), we can see that

$$\int_V \rho(\mathbf{x}) dV = \lambda,$$

which is the mass contained within V . But recall that our solution to Laplace, $\psi(r) = A/r + B$, is *only singular at the origin*. So we can take our volume V with surface S to be an *arbitrarily small ball around the origin*. This ball contains the point-mass!

What function gives us $\rho(\mathbf{x}) = 0$ for all $\mathbf{x} \neq \mathbf{0}$, but is non-zero at $\mathbf{x} = \mathbf{0}$? The answer is a function you've not previously met, the *3-D Dirac delta function* $\delta^3(\mathbf{x})$. The three-dimensional Dirac delta function can be thought of as an infinitely narrow spike of infinite height at $\mathbf{x} = \mathbf{x}'$, with unit volume under the spike. It is zero everywhere else, and

$$\int_V \delta^3(\mathbf{x} - \mathbf{x}') dV = \begin{cases} 1 & \text{if } V \text{ contains } \mathbf{x}', \\ 0 & \text{otherwise.} \end{cases} \quad (5.5)$$

So the Dirac delta function has strange properties:

- It is zero everywhere except at one point;
- It has a unit volume.

Since it has a finite volume but takes up zero space, it must have a value of ∞ at $\mathbf{x} = \mathbf{x}'$! Hence we learn that the isotropic solution $\psi = A/r + B$ corresponds to a Poisson equation with a very specific right-hand side, and its solution (obtained by Gauss' flux method)

$$\nabla^2\psi^* = -\lambda\delta^3(\mathbf{x}) \quad \Rightarrow \quad \psi^*(r) = \frac{\lambda}{4\pi r} + B, \quad (5.6)$$

where B is a constant of integration. Hence we see that our isotropic solution $A/r + B$ to the Laplace equation at $r > 0$ is, in fact, an isotropic solution for *all* r to the Poisson equation with a delta-function mass distribution.

Exercise 5.1 on the Dirac Delta function

Evaluate the integral

$$\int_V f(\mathbf{x}) dV$$

for the volume V with the surface defined by $|\mathbf{x}| = 10$ and with

$$f(\mathbf{x}) = \delta^3(\mathbf{x} - 2\hat{\mathbf{e}}_x) + \delta^3(\mathbf{x} - 5\hat{\mathbf{e}}_y) + \delta^3(\mathbf{x} - 30\hat{\mathbf{e}}_z).$$

Hint: the answer is 2.

Now, we started this section with the Poisson equation for a general distribution of mass $\rho(\mathbf{x})$ in equation (5.4), but in (5.6) we solved it for a very particular (and peculiar) distribution $\rho(\mathbf{x}) = \lambda\delta^3(\mathbf{x})$. But we did this for a good reason! *It turns out that we can use the latter to construct solutions to the former.* We do this using an approach called *superposition of solutions*, which just means summing multiple solutions to an equation.

Consider a case where we have a set of N point masses located at positions $\mathbf{x}_i$ where $i = 1, 2, \dots, N$ with masses λ_i . Each point mass generates a field ψ_i^* . The total field generated by the total set of N masses is obtained by superposition,

$$\psi(\mathbf{x}) = \sum_{i=1}^N \psi_i^*(\mathbf{x}). \quad (5.7)$$

So the Poisson equation for the total field is

$$\nabla^2 \left(\sum_{i=1}^N \psi_i^* \right) = - \sum_{i=1}^N \lambda_i \delta^3(\mathbf{x} - \mathbf{x}'_i). \quad (5.8)$$

But we already know the solution arising from each individual point of mass! It is given by equation (5.6) if the mass is located at $\mathbf{x}' = \mathbf{0}$. But the masses are *not* located at the origin. Each point-mass can define a spherical coordinate system with

$$r = |\mathbf{x} - \mathbf{x}'_i|.$$

Then we can write the individual solutions using (5.6) as

$$\psi_i^*(\mathbf{x}) = \frac{\lambda_i}{4\pi|\mathbf{x} - \mathbf{x}'_i|}, \quad (5.9)$$

where we have taken $B = 0$ by choosing ψ_i^* to be zero at infinity. And now we simply use (5.9) in (5.7) to obtain the full solution,

$$\begin{aligned} \psi(\mathbf{x}) &= \sum_{i=1}^N \psi_i^*(\mathbf{x}) \\ &= \sum_{i=1}^N \frac{\lambda_i}{4\pi|\mathbf{x} - \mathbf{x}'_i|}. \end{aligned} \quad (5.10)$$

The next step is to allow an infinite set of point-masses to become infinitely close together to form a continuous object — and hence to convert our sum into an integral!

This idea is perhaps clearer if we write the real density distribution $\rho(\mathbf{x})$ in a different way,

$$\rho(\mathbf{x}) = \int \rho(\mathbf{x}') \delta^3(\mathbf{x} - \mathbf{x}') dV', \quad (5.11)$$

where dV' is an infinitesimal volume element of the primed position $\mathbf{x}'$. This is like the right-hand side of (5.8) but with the sum as an integral and λ_i as $\rho(\mathbf{x}')$. The integral may be taken over all of space, or just the region containing non-zero density. The increment of mass at every point $\mathbf{x}'$ contributes to the total solution, which we obtain by integrating over all $\mathbf{x}'$. Then the total solution to (5.4) can be expressed as the integral of delta-function solutions (5.6),

$$\psi(\mathbf{x}) = \int_{V'} \frac{\rho(\mathbf{x}')}{4\pi|\mathbf{x} - \mathbf{x}'|} dV', \quad (5.12)$$

where the volume V' is parameterised by $\mathbf{x}'$. Note that this is an integral over $\mathbf{x}'$ so it treats $\mathbf{x}$ as a constant. Compare this with the discrete version in (5.10).

Green's functions The function

$$\mathcal{G}(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|}$$

is known as the *Green's function* solution of the equation

$$\nabla^2 \mathcal{G} = -\delta^3(\mathbf{x} - \mathbf{x}').$$

This is the Green's problem associated with the Poisson equation $\nabla^2 \psi = -\rho(\mathbf{x})$. With the Green's function solution, we can immediately write

$$\psi(\mathbf{x}) = \int_{V'} \rho(\mathbf{x}') \mathcal{G}(\mathbf{x}, \mathbf{x}') dV', \quad (5.13)$$

exactly as we did in equation (5.12).

There are Green's problems and solutions associated with many other linear partial differential equations. Linearity of the equation is essential because Green's functions work by superposition.

Example 5.1 of Green's function solution for gravity

Use equation (5.12) to write an expression for the gravitational potential $\Phi(\mathbf{x})$ around a cube-shaped planet with edges of length L and density ρ_0 .

Comparing equation (5.4) with equation (4.16), we see that $\Phi = -4\pi G\psi$ (recall that G is the gravitational constant). Therefore we can write our expression for the potential Φ as

$$\Phi(\mathbf{x}) = -G\rho_0 \int_{-L/2}^{L/2} dx' \int_{-L/2}^{L/2} dy' \int_{-L/2}^{L/2} dz' \frac{1}{|\mathbf{x} - \mathbf{x}'|}.$$

The mass of the cube is $M = \rho_0 L^3$.

5.3 Solution by separation of variables

When the boundaries of a problem align with the coordinate axes and the conditions along those boundaries are sufficiently simple, the method of separation of variables can be used to solve the Laplace equation.

To illustrate this method, we solve on a rectangular domain using Cartesian coordinates. The Laplacian is

$$\nabla^2 \psi = 0 \quad \Rightarrow \quad \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0. \quad (5.14)$$

Now the trick: we assume that the solution is of the form

$$\psi(\mathbf{x}) = X(x)Y(y)$$

for two independent functions X and Y . Notice that X is *not* a function of y and Y is *not* a function of x . Substituting this form in the Laplace equation we obtain

$$YX'' + XY'' = 0.$$

Here we have used primes to indicate a derivative with respect to the (only) argument of a function. Dividing by XY and rearranging gives

$$\frac{X''}{X} = -\frac{Y''}{Y}.$$

Now we notice something important. The left-hand side depends *only* on x and the right-hand side depends *only* on y . For this to be true for any x, y pair, it must be the case that both sides are equal to a constant⁸. Now at this stage the constant could be anything, so with foresight (for example 5.2) we will choose it to be k^2 . Then we have

$$X''/X = k^2 \quad \text{and} \quad Y''/Y = -k^2$$

which, for $k \neq 0$, has solutions⁹

$$X(x) = Ae^{kx} + Be^{-kx} \quad \text{and} \quad Y(y) = C \cos ky + D \sin ky,$$

and therefore a separated solution is

$$\psi(\mathbf{x}) = (Ae^{kx} + Be^{-kx})(C \cos ky + D \sin ky). \quad (5.16)$$

The constants A, B, C, D together with the allowable values of k are constrained by the boundary conditions. Sometimes a solution is constructed by summing over the allowable values of k . We return to this briefly below.

⁸To prove this, differentiate the whole equation with respect to either independent variable; in either case, one side of the equation vanishes.

⁹For certain, rare cases, the boundary conditions require that $k = 0$ such that we obtain solutions

$$X(x) = Ax + B \quad \text{and} \quad Y(y) = Cy + D. \quad (5.15)$$

Example 5.2 of separation of variables

Consider the Laplace equation (5.14) on a rectangular, semi-infinite domain with $x \in [0, \infty)$ and $y \in [0, 1]$. The boundary conditions are,

$$\begin{aligned}\psi &\rightarrow 0 & \text{as } x &\rightarrow \infty, \\ \psi &= 0 & \text{at } y &= 0, 1, \\ \psi &= \sin \pi y & \text{at } x &= 0.\end{aligned}$$

Find the solution that fits the boundary conditions.

Start with the general solution (5.16). The first boundary condition tells us that $A = 0$. The second boundary condition tells us that $C = 0$. What we have left, evaluated at $x = 0$, is $\psi = BD \sin ky$. The third boundary condition tells us that $BD = 1$ and constrains the eigenvalue to be $k = \pi$. Hence the solution with these boundary conditions is

$$\psi(\mathbf{x}) = e^{-\pi x} \sin \pi y.$$

Exercise 5.2 on separation of variables

Use separation of variables to obtain a general solution to the diffusion equation,

$$\kappa \frac{\partial^2 \psi}{\partial x^2} = \frac{\partial \psi}{\partial t}.$$

Your solution should be of the form $\psi(x, t) = X(x)T(t)$ for independent functions X and T , to be determined.

On a finite domain, where boundary conditions must be imposed on four walls, it is convenient to use the functions $\sinh kx$ and $\cosh kx$ as solutions instead of e^{kx} and e^{-kx} (noting that all of these solve the equation $X''/X = k^2$).

Summary of §5

- Isotropic (spherically symmetric) solutions to Laplace's equation have a singularity at $r = 0$. This singularity corresponds to a point-source at the origin in Poisson's equation.
- A point-source can be described with a Dirac Delta function, $\delta^3(\mathbf{x})$, an infinite spike with unit volume under it.
- Poisson's equation is a linear PDE, has a Green's function solution, and hence can be solved with a Green's function integral.
- Separation of variables is a useful technique for solving linear partial differential equations on simple domains. It assumes that the full solution $\psi(\mathbf{x}, t)$ is a product of functions, each of only one variable, e.g.,

$$\psi(\mathbf{x}, t) = X(x)Y(y)Z(z)T(t).$$

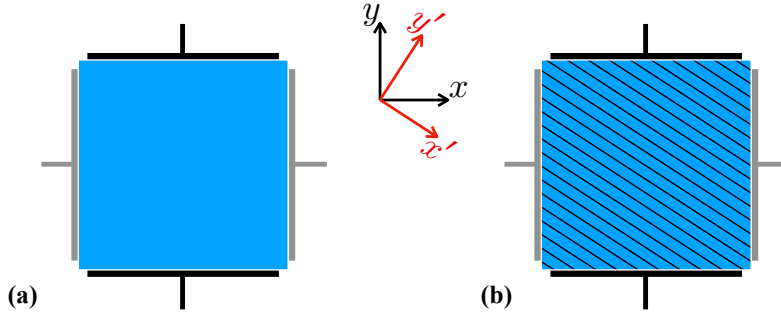


Figure 6.1: Blocks of conductive material between electrodes that can create an electric field. The grey electrodes create a field in the x direction; the black electrodes create a field in the y direction. **(a)** The block of material has uniform and isotropic conductivity σ_0 . **(b)** The block has a uniform background conductivity σ_0 plus an anisotropic conductivity σ_1 in the x' direction, coming from an array of nanowires.

6 Tensors

The word tensor is extremely general, which makes it annoyingly vague. A first-rank tensor has one index (e.g., v_j), which means it is a vector. A zeroth-rank tensor has zero indices (e.g., c) and is therefore a scalar. Third-rank tensors do not appear in this course and we will see fourth-rank tensors (with four indices) only in one situation.

In the context of continuum mechanics (and other field theories), most of the tensors are *second-rank tensors*, meaning that they have two free indices (e.g., T_{ij}) and can be written as square matrices. *So we'll use the word "tensor" to mean "second-rank tensor" unless otherwise noted.*

So if a (second-rank) tensor can be written as a matrix, are all matrices tensors? The answer is no. So we must ask, how can we tell when a matrix is a tensor? We return to this question in section 6.3.

6.1 Why we need tensors: an example

Consider the flow of electrical current through a block of material shown in Figure 6.1a. The material is *isotropic*, meaning that it is the same in all directions. Therefore, an electric field E_x in the x direction drives a current in the x direction and likewise for a field E_y in the y direction. We can write the relationship between the vector current density (flux of electric charge) $\mathbf{J}$ and the vector electric field $\mathbf{E}$ as

$$\mathbf{J} = \sigma_0 \mathbf{E}, \quad (6.1)$$

or

$$\begin{bmatrix} J_x \\ J_y \end{bmatrix} = \sigma_0 \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \quad (6.2)$$

This is Ohm's law for an isotropic medium. It states that $\mathbf{J}$ and $\mathbf{E}$ are in the same direction, but have different magnitudes. The conductivity, $\sigma_0 = |\mathbf{J}|/|\mathbf{E}|$, is independent of the direction of the imposed electric field.

Now consider the case of electrical current through the block shown in Figure 6.1b. In this case the isotropic background has been overprinted by an array of more conductive nanowires. The array of wires conducts only in the x' direction, at an angle to x . From the drawing, we can infer that an electric field in the y direction will drive some current in the x direction through the wires, and vice versa. We need a relationship that captures this, such as

$$\begin{bmatrix} J_x \\ J_y \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix}, \quad (6.3)$$

or,

$$\mathbf{J} = \boldsymbol{\sigma} \cdot \mathbf{E}, \quad (6.4)$$

where $\boldsymbol{\sigma}$ is the conductivity tensor. We need this tensor (matrix) representation to capture the *anisotropic* nature of the conductivity! But how to determine the entries in the conductivity matrix?

For the case shown in panel a, the conductivity tensor should be isotropic and should couple E_x to J_x and E_y to J_y . Therefore it is

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_0 & 0 \\ 0 & \sigma_0 \end{bmatrix} = \sigma_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \sigma_0 \mathbf{I}, \quad (6.5)$$

where $\mathbf{I}$ is the identity matrix (or identity tensor).

For the case in panel b, suppose that we know that the array of wires has a conductivity of σ_1 along the direction of the wires, which is the x' direction. Clearly the wires do not conduct in the y' direction. If we write Ohm's law in the primed coordinate system, we have

$$\mathbf{J}' = \boldsymbol{\sigma}' \cdot \mathbf{E}'. \quad (6.6)$$

Now the conductivity tensor has two contributions, the isotropic part $\sigma_0 \mathbf{I}$ and the anisotropic part,

$$\boldsymbol{\sigma}' = \sigma_0 \mathbf{I} + \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix}. \quad (6.7)$$

Note that the identity tensor is the same in any coordinate system! Also note that we've augmented the conductivity in the x' direction, but only due to the electric field in the x' direction! Now we have

$$\begin{bmatrix} J_{x'} \\ J_{y'} \end{bmatrix} = \begin{bmatrix} \sigma_0 + \sigma_1 & 0 \\ 0 & \sigma_0 \end{bmatrix} \begin{bmatrix} E_{x'} \\ E_{y'} \end{bmatrix}. \quad (6.8)$$

This is great, but what we really want is the conductivity tensor in the x - y coordinate system that is aligned with the electrodes that we can control. We'll come back to that below when we've learned about tensor rotations.

6.2 Tensors are maps between vectors

This is an important property of tensors, and the one that will be most useful in continuum mechanics. When we multiply a vector by a tensor, we get another vector,

$$\mathbf{u} = \mathbf{T} \cdot \mathbf{v} \quad \text{or} \quad u_i = T_{ij} v_j. \quad (6.9)$$

Recall that the properties of a vector are direction and magnitude. Therefore, in general, a tensor changes the direction and magnitude of a vector. This is not strictly the definition of a tensor (see below for that), but it is a useful way to think about tensors. (Note that we are referring to second-rank tensors here. A fourth-rank tensor maps between second-rank tensors.)

Example 6.1 of a map between vectors

Cauchy's stress theorem, which we will consider again later, under Continuum Mechanics, gives a meaning to the stress-tensor field $\mathbf{T}(\mathbf{x})$. The stress-tensor field gives us the state of stress throughout a continuous piece of matter. Suppose that we want to know, at a point $\mathbf{x}$ within that domain, what is the stress acting on an arbitrary, flat surface element? We call this the traction $\mathbf{t}$. Cauchy's stress theorem states that for a surface element at position $\mathbf{x}$ with a unit-normal vector $\hat{\mathbf{n}}$, the traction vector is given by

$$\mathbf{t}(\mathbf{x}, \hat{\mathbf{n}}) = \mathbf{T} \cdot \hat{\mathbf{n}} \quad \text{or} \quad t_i = T_{ij} \hat{n}_j.$$

(Do not confuse the traction vector with the tangent vector; they share a symbol but are distinct.)

Above in (6.9) we have included a $\cdot$ where we multiply a tensor with a vector. This is called *index contraction* and we use it to emphasise that there is a summation over one of the indices of the tensor. As usual, index notation provides an unambiguous statement of this. Summation over one index gives a result that has one less index; hence a second-rank tensor (2 free indices) contracts to a vector (1 free index). For this reason the dot product with a tensor is sometimes called a *tensor contraction*.

Often, a matrix–vector product is written without the $\cdot$, such as $\mathbf{u} = \mathbf{T} \mathbf{v}$. We use both notations below, but usually exclude the dot.

6.3 Tensors are things that transform like tensors

Annoyingly, the best strict definition of a tensor is recursive: *a tensor is a mathematical object that transforms like a tensor*. What does this mean? To clarify, let's consider a first-rank tensor, i.e., a vector. And we need to explain what it means to “transform” a tensor. We do this in the next paragraph.

A vector exists in the absence of any coordinate system. It is a direction and magnitude. To work with a vector, it is helpful to express it in a coordinate system. But which coordinate system to choose? Even if we just restrict ourselves to Cartesian systems in $\mathbb{R}^3$, we can rotate them arbitrarily in space. So a vector maintains its magnitude and direction when we transform its expression from one coordinate system to another. But a vector is a rank-one tensor, and so we can use this same definition for all tensors! Transforming a tensor means to convert its expression from one coordinate system to another.

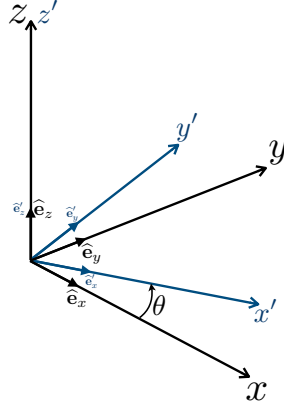


Figure 6.2: A coordinate rotation by θ around the z axis.

6.3.1 Transforming first-rank tensors (vectors)

Let's express a vector $\mathbf{x}$ in terms of a set of basis vectors, $\mathbf{x} = x_i \hat{\mathbf{e}}_i$. The basis vectors are mutually orthogonal such that $\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}$. But we could have equally expressed $\mathbf{x}$ in terms of a different set of basis vectors $\hat{\mathbf{e}}'_i$ that are related to $\hat{\mathbf{e}}_i$ by a rotation matrix,

$$\hat{\mathbf{e}}'_i = R_{ij} \hat{\mathbf{e}}_j. \quad (6.10)$$

This rotated set, shown in figure 6.2, is also orthonormal and so $\hat{\mathbf{e}}'_i \cdot \hat{\mathbf{e}}'_j = \delta_{ij}$. Combining this with (6.10) we can write,

$$\begin{aligned} \hat{\mathbf{e}}'_i \cdot \hat{\mathbf{e}}'_j &= \delta_{ij} = R_{ik} \hat{\mathbf{e}}_k \cdot R_{jl} \hat{\mathbf{e}}_l, \\ &= R_{ik} R_{jl} \delta_{kl}, \\ &= R_{ik} R_{jk}. \end{aligned}$$

Writing the last part in matrix notation,

$$\mathbf{R} \mathbf{R}^T = \mathbf{I} \quad \text{or} \quad R_{ik} R_{jk} = \delta_{ij}, \quad (6.11)$$

where $\mathbf{I} = \delta_{ij}$ is the identity matrix. This shows that the inverse of a rotation matrix is its transpose, which we will use below.

Since the vector $\mathbf{x}$ is the same in either basis, we can write

$$\mathbf{x} = x_i \hat{\mathbf{e}}_i = x'_i \hat{\mathbf{e}}'_i.$$

Then substituting equation (6.10), the last equality becomes

$$x_i \hat{\mathbf{e}}_i = x'_i R_{ij} \hat{\mathbf{e}}_j. \quad (6.12)$$

This equation shows that the *components* of the vector transform with the same rotation matrix that we used for the basis vectors, $x_j = R_{ij} x'_i$. Now, dotting both

sides of equation (6.12) with $\hat{\mathbf{e}}'_k = R_{kl}\hat{\mathbf{e}}_l$ we have

$$\begin{aligned} x_i \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}'_k &= x'_i R_{ij} \hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}'_k, \\ x_i R_{kl} \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_l &= x'_i R_{ij} R_{kl} \hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}_l, \\ x_i R_{ki} &= x'_i R_{ij} R_{kj}, \\ &= x'_i \delta_{ik}, \\ &= x'_k, \end{aligned}$$

where we used equation (6.11). The two results of this paragraph are re-written in the set

$$\mathbf{x}' = \mathbf{R}\mathbf{x} \quad \text{or} \quad x'_i = R_{ij}x_j, \quad (6.13a)$$

$$\mathbf{x} = \mathbf{R}^\top \mathbf{x}' \quad \text{or} \quad x_i = R_{ji}x'_j. \quad (6.13b)$$

Examine these two equations carefully and notice the difference: the first one sums over the second index of $\mathbf{R}$ while the second one sums over the first index.

6.3.2 Transforming second-rank tensors (matrices)

A vector is a first-rank tensor, so we have, in a sense, demonstrated how tensors transform. But we generally refer to second-rank tensors, so let's look at that case.

Consider the mapping $\mathbf{u} = \mathbf{T}\mathbf{v}$ from section 6.2. We can re-write this equation using (6.13) to put the vectors into the rotated basis,

$$\mathbf{R}^\top \mathbf{u}' = \mathbf{T} \mathbf{R}^\top \mathbf{v}' \quad \text{or} \quad R_{ik} u'_i = T_{kl} R_{jl} v'_j.$$

Note that we are contracting the vectors with the first index of $\mathbf{R}$. Now, multiplying both sides by $\mathbf{R}$ and recalling equation (6.11),

$$\mathbf{u}' = \mathbf{R} \mathbf{T} \mathbf{R}^\top \mathbf{v}' \quad \text{or} \quad u'_n = R_{nk} T_{kl} R_{jl} v'_j.$$

From this set we can infer the desired result that relates the tensor components in two coordinate systems that differ by a rotation,

$$\mathbf{T}' = \mathbf{R} \mathbf{T} \mathbf{R}^\top \quad \text{or} \quad T'_{ij} = R_{ik} R_{jl} T_{kl}. \quad (6.14)$$

We have shown that a second-rank tensor transforms by contracting each index with a rotation matrix.

It is important to emphasise that with index notation, the order that the entries are written in the product doesn't matter. What does matter is which of the indices is being summed over. Getting the indices right requires precise application of equation (6.13), and then careful bookkeeping.

Exercise 6.1 on the inverse rotation matrix

Prove that the inverse of the rotation matrix is its transpose. Hint: there are two separate ways to approach this:

1. Notice that the inverse of a rotation by θ is a rotation by $-\theta$.
2. Show that $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Example 6.2 of something that is not a tensor

The position vector $\mathbf{x}$ is obviously a rank-one tensor. However, consider this thing:

$$\mathbf{m} = \begin{bmatrix} x^2 \\ y^2 \\ z^2 \end{bmatrix}.$$

Is this a tensor or just an array? To find out, we rotate the coordinate system with a rotation matrix $\mathbf{R}$. In that case,

$$\mathbf{m}' = \begin{bmatrix} (R_{11}X + R_{21}Y + R_{31}Z)^2 \\ (R_{12}X + R_{22}Y + R_{32}Z)^2 \\ (R_{13}X + R_{23}Y + R_{33}Z)^2 \end{bmatrix}.$$

But this isn't the way that a rank-one tensor transforms! That is, $m'_i \neq R_{ij}m_j$. So it isn't a tensor, it is just an array. It looked like a rank-one tensor because it was written with the notation of a vector.

6.3.3 The rotation matrix

So far, we have avoided saying what is inside the rotation matrix, but the time has come. We consider two coordinate systems. The first system has coordinates x, y, z and the second system has coordinates X, Y, Z . They differ by a counterclockwise rotation θ around the z axis. This means that $\hat{\mathbf{e}}_z = \hat{\mathbf{e}}_Z$, but the other basis vectors are not equal between systems. In particular, we have

$$\hat{\mathbf{e}}_X = \hat{\mathbf{e}}_x \cos \theta + \hat{\mathbf{e}}_y \sin \theta, \quad (6.15a)$$

$$\hat{\mathbf{e}}_Y = \hat{\mathbf{e}}_y \cos \theta - \hat{\mathbf{e}}_x \sin \theta, \quad (6.15b)$$

$$\hat{\mathbf{e}}_Z = \hat{\mathbf{e}}_z. \quad (6.15c)$$

Suppose that we have a vector $\mathbf{a}$ expressed in terms of $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$ and we want to express it in terms of $\hat{\mathbf{e}}_X, \hat{\mathbf{e}}_Y, \hat{\mathbf{e}}_Z$. In other words, we want to transform it to the X, Y, Z system, where we will call it $\mathbf{a}'$ (note that these are the same vector, but expressed with different basis vectors). Working backwards and using equations (6.15),

$$\begin{aligned} \mathbf{a}' &= a_X \hat{\mathbf{e}}_X + a_Y \hat{\mathbf{e}}_Y + a_Z \hat{\mathbf{e}}_Z, \\ &= a_X (\hat{\mathbf{e}}_x \cos \theta + \hat{\mathbf{e}}_y \sin \theta) + a_Y (\hat{\mathbf{e}}_y \cos \theta - \hat{\mathbf{e}}_x \sin \theta) + a_Z \hat{\mathbf{e}}_Z, \\ &= \hat{\mathbf{e}}_x (a_X \cos \theta - a_Y \sin \theta) + \hat{\mathbf{e}}_y (a_Y \cos \theta + a_X \sin \theta) + \hat{\mathbf{e}}_z a_Z, \\ &= \hat{\mathbf{e}}_x a_x + \hat{\mathbf{e}}_y a_y + \hat{\mathbf{e}}_z a_z. \end{aligned}$$

This equation can be summarised succinctly by writing

$$\mathbf{a} = \mathbf{R}^T \mathbf{a}' = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} a_X \\ a_Y \\ a_Z \end{bmatrix}, \quad (6.16)$$

where $\mathbf{R}$ is a rotation matrix about the z axis. It appears in (6.16) with a transpose to be consistent with (6.13).

As we learned above in section 6.3.1, we can invert this equation; given a vector expressed in one basis $\mathbf{a}$, we can find the vector $\mathbf{a}'$ expressed in a rotated basis. Together, the forward and inverse rotations are

$$\mathbf{a}' = \mathbf{R}\mathbf{a} \quad \text{and} \quad \mathbf{a} = \mathbf{R}^\top \mathbf{a}', \quad (6.17)$$

which is identical to equation (6.13), but expressed in vector notation. Equation (6.17) is true for any rotation matrix, not just the one in (6.16).

Rotation matrices are also useful for rotating vectors within the same coordinate system. For example, $\mathbf{b} = \mathbf{R}\mathbf{a}$ with $\mathbf{R}^\top$ as in (6.16). In this case, $\mathbf{b}$ has the magnitude of $\mathbf{a}$ but a direction that has been rotated by θ in the *clockwise* direction. (Note that rotating a vector with a fixed coordinate system is opposite to rotating a coordinate system with a fixed vector).

Exercise 6.2 on rotation matrices

Construct a rotation matrix that rotates a vector anticlockwise around the y axis by an angle θ . Show that this works by rotating the vector $\mathbf{a} = \hat{\mathbf{e}}_x + \hat{\mathbf{e}}_z$ by 90° anticlockwise.

Example 6.3 of the conductivity tensor (6.7)

We now return to our conductivity tensor (6.7) armed with a means for transforming tensors to rotated coordinate systems. Recall that we had

$$\boldsymbol{\sigma}' = \sigma_0 \mathbf{I} + \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

in the primed system, aligned with the nanowires but not with the electrodes. We want a tensor aligned with the electrodes, in the x - y system. Using what we learned in sections 6.3.2 and 6.3.3, we can find that $\boldsymbol{\sigma} = \mathbf{R}^\top \boldsymbol{\sigma}' \mathbf{R}$ and hence

$$\begin{aligned} \boldsymbol{\sigma} &= \sigma_0 \mathbf{R}^\top \mathbf{I} \mathbf{R} + \sigma_1 \mathbf{R}^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{R}, \\ &= \sigma_0 \mathbf{I} + \sigma_1 \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \\ &= \sigma_0 \mathbf{I} + \sigma_1 \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix}. \end{aligned}$$

This is the conductivity tensor in the x - y coordinate system. It tells us that an electric field of, say, $\mathbf{E} = E_y \hat{\mathbf{e}}_y$ creates a current density in the x direction of $J_x = \sigma_1 E_y \sin \theta \cos \theta = \frac{\sigma_1}{2} E_y \sin 2\theta$. (It is left to the reader to find J_y .) And since the x' coordinate system was rotated to align with the nanowires, we can think of θ as being the angle of the wires measured *anticlockwise* from the x -direction. N.B. this means that θ is negative in Fig. 6.1b and hence J_x is negative in this example.

Example 6.4 of something that is not a tensor

The position vector $\mathbf{x}$ is obviously a rank-one tensor. However, consider this thing:

$$\mathbf{m} = \begin{bmatrix} x^2 \\ y^2 \\ z^2 \end{bmatrix}.$$

Is this a tensor or just an array? To find out, we rotate the coordinate system from x, y, z to X, Y, Z about the z axis. In that case,

$$\mathbf{m}' = \begin{bmatrix} (X \cos \theta + Y \sin \theta)^2 \\ (-X \sin \theta + Y \cos \theta)^2 \\ Z^2 \end{bmatrix}.$$

But this isn't the way that a rank-one tensor transforms! So it isn't a tensor, it is just an array. It looked like a rank-one tensor because it was written with the notation of a vector.

6.3.4 Pardon me madame; is that a tensor?

Suppose someone shows you a 3×3 matrix — this one for example

$$\begin{pmatrix} 2 & 3 & 0 \\ 0 & 4 & 4 \\ 46.8 & 11 & 2 \end{pmatrix}$$

— and asks you, “is that a tensor?” You should answer that it might be, but you'd need more information to know for certain. What would you need to know? You'd need to know what entries it has in a different coordinate system, and what the exact relationship between the two coordinate systems is. Only then could you say for certain.

However, this contrived situation is largely academic (ahem), and misses the point of section 6.3.2.

6.4 Symmetric, antisymmetric, and isotropic tensors

There are a few special types of tensor, and these play a particularly important role in mathematical theories of physics. These types of tensors will appear frequently in the Continuum Mechanics section of the course. Their definitions are relatively simple.

Symmetric tensors Symmetric tensors are those for which the matrix representation is symmetrical,

$$\mathbf{S} = \mathbf{S}^T \quad \text{or} \quad S_{ij} = S_{ji}. \quad (6.18)$$

This means that for a tensor in $\mathbb{R}^3$, instead of having 9 independent entries, it has only six independent entries. Each of the off-diagonal entries is repeated.

Antisymmetric tensors Antisymmetric tensors are those for which the matrix representation satisfies

$$\mathbf{A} = -\mathbf{A}^\top \quad \text{or} \quad A_{ij} = -A_{ji}. \quad (6.19)$$

You will notice that this can only be true if the diagonal entries are all zeros. Therefore, in $\mathbb{R}^3$, an antisymmetric tensor has only 3 independent entries.

Exercise 6.3 on (anti)symmetric tensors

Using the independent scalar variables $a, b, c, d, e, f \neq 0$, write down a symmetric tensor $\mathbf{S}$ and an antisymmetric tensor $\mathbf{A}$.

Isotropic tensors A tensor that is invariant under all rotations is *isotropic*, meaning that it is the same in all directions. It turns out that the only invariant, rank-2 tensor is the scaled identity matrix, $\alpha \mathbf{I}$ or $\alpha \delta_{ij}$, where α is a constant.

Exercise 6.4 on isotropic tensors

Show that the identity tensor is invariant to a rotation about the z axis by $\pi/2$.

Tensor decompositions It is a wonderful fact that any tensor can be decomposed into a symmetric part, an antisymmetric part, and an isotropic part. Let's see how.

Consider the tensor $\mathbf{T}$ in $\mathbb{R}^3$ that is not a special tensor. That is, it is not symmetric, antisymmetric or isotropic. It is written as a 3×3 matrix and has 9 independent components. We begin by defining two new matrices

$$S_{ij} = \frac{1}{2} (T_{ij} + T_{ji}) \quad \text{and} \quad A_{ij} = \frac{1}{2} (T_{ij} - T_{ji}). \quad (6.20)$$

You can show (exercise 6.5) that $\mathbf{S}$ is symmetric while $\mathbf{A}$ is antisymmetric. Importantly, it is also true that

$$S_{ij} + A_{ij} = T_{ij}. \quad (6.21)$$

Therefore, we can write any tensor as the sum of a symmetric part and an antisymmetric part. The number of independent entries sums to 9.

Exercise 6.5 on tensor decompositions

Show that $\mathbf{S}$ and $\mathbf{A}$ defined in equation (6.20) are symmetric and antisymmetric, respectively. Hint: transpose using index notation.

We can further decompose the symmetric tensor $\mathbf{S}$ to pull out an isotropic part. For this we need to define the *trace* of a matrix. The trace is the sum of the diagonal entries. For a 3×3 matrix $\mathbf{S}$,

$$\text{trace}(\mathbf{S}) = S_{kk} = S_{11} + S_{22} + S_{33}. \quad (6.22)$$

(We note that the repeated index implies summation.) Now we can use this to define,

$$\begin{aligned} Q &= S_{ii}/3, \\ M_{ij} &= S_{ij} - Q\delta_{ij}. \end{aligned}$$

Rearranging the latter, we can re-write (6.21) as

$$T_{ij} = M_{ij} + A_{ij} + Q\delta_{ij}. \quad (6.23)$$

In this equation, $\mathbf{M}$ has 5 independent entries, $\mathbf{A}$ has 3 independent entries, and Q has one independent entry. The total number of independent entries remains 9.

Exercise 6.6 on the trace of a matrix

Show that the trace of the right-hand side of equation (6.23) is $3Q$.

Exercise 6.7 on tensor decomposition

Find a decomposition of the form of (6.23) for the tensor

$$\mathbf{T} = \begin{pmatrix} 2 & 5 & 6 \\ 1 & 1 & 7 \\ 2 & 3 & 0 \end{pmatrix}.$$

6.5 Tensor fields

Tensor fields are tensors that vary in space (and possibly also over time). They are essential to the theory of continuum mechanics. Usually they are simply referred to as “tensors.” Two important examples are the stress tensor $\mathbf{T}(\mathbf{x})$ and the displacement gradient tensor $\mathbf{D}(\mathbf{x})$. We will study these in detail later, but let’s first see why they are tensor fields.

Consider some finite, simple domain that is filled with homogeneous solid. Suppose that at $t = 0$ the material is at rest. Force is applied to the boundaries of the domain and this causes motion, such that after a short time, the material has been displaced from its initial position. The displacement is specific to each infinitesimal parcel of matter (you can imagine tagging atoms and comparing their initial and final positions). We refer to this as the vector *displacement field* $\mathbf{u}(\mathbf{x})$. If $\mathbf{u}$ is uniform, the whole body has moved without distortion and there is no strain (we will go into this in more detail later). If the displacement is not uniform, there are gradients in the displacement. These gradients can be written as a tensor field,

$$\mathbf{D}(\mathbf{x}) = \nabla \mathbf{u}(\mathbf{x}) \quad \text{or} \quad D_{ij} = \frac{\partial u_i}{\partial x_j}. \quad (6.24)$$

From the expression in index notation, we can clearly see why this is a tensor: there are two free indices. One of these comes from the Nabla operator and one comes

from the displacement vector. Every combination of i and j must be considered, and so we have a table (or matrix) of numbers. In particular, in $\mathbb{R}^3$ with Cartesian coordinates,

$$\mathbf{D}(\mathbf{x}) = \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{pmatrix}, \quad (6.25)$$

where each entry in this matrix is itself a field. If we know the displacement field $\mathbf{u}(\mathbf{x})$, we can always calculate the displacement-gradient field $\mathbf{D}$ at any point in the domain.

Now, the stress at a point in an elastic solid depends on the displacement gradient at a point. We could express this as $\mathbf{T}(\mathbf{x}) = \mathbf{T}(\mathbf{D}(\mathbf{x}))$.¹⁰ And since the displacement gradient is a tensor field, so too must be the stress.

Suppose that $\mathbf{u}(\mathbf{x}, t)$ represents a displacement that is changing smoothly over time. Then we could define a vector *velocity field* as

$$\mathbf{v}(\mathbf{x}, t) = \lim_{\delta t \rightarrow 0} \frac{\mathbf{u}(\mathbf{x}, t + \delta t) - \mathbf{u}(\mathbf{x}, t)}{\delta t}.$$

We could therefore define a velocity-gradient tensor field as

$$\dot{\mathbf{D}}(\mathbf{x}, t) = \frac{\partial}{\partial t} \mathbf{D}(\mathbf{x}, t) = \nabla \mathbf{v}(\mathbf{x}, t) \quad \text{or} \quad \dot{D}_{ij} = \frac{\partial v_i}{\partial x_j}. \quad (6.26)$$

This tensor is essential in calculating the stress at a point in a fluid.

Exercise 6.8 on the displacement gradient tensor field

Sketch the following displacement fields and calculate the displacement-gradient tensor field. D_0 is a constant.

1. $\mathbf{u}(\mathbf{x}) = D_0(x, y, z)$;
2. $\mathbf{u}(\mathbf{x}) = D_0(-x, -y, 2z)$;
3. $\mathbf{u}(\mathbf{x}) = D_0(\sin x, 0, 0)$;
4. $\mathbf{u}(\mathbf{x}) = D_0(-y, x, 0)$.

Discuss the elastic stress that these displacements might create in terms of the directions of tension and compression.

¹⁰A large part of continuum mechanics is modelling the relationship between stress and material distortion (and its rate). We come back to discuss this in detail later.

Summary of §6

- The word tensor is very general, defined as any object that transforms like a tensor under coordinate-system rotation. Vectors are first-order tensors. Second-order tensors have two free indices and, in a particular coordinate system, are expressed as matrices. We will refer to second-order tensors (and second-order tensor fields) as “tensors.”
- A second-order tensor is a map between vectors. For example, here the tensor $\mathbf{T}$ maps $\mathbf{x}$ to $\mathbf{y}$.

$$\mathbf{y} = \mathbf{T} \cdot \mathbf{x} \quad \text{or} \quad y_i = T_{ij}x_j.$$

- Under coordinate rotation, tensors transform by contraction with a rotation matrix. First-order and second-order tensors transform as

$$x'_i = R_{ij}x_j \quad \text{and} \quad T'_{ij} = R_{ik}R_{jl}T_{kl}.$$

respectively. Note that in each case, the rotation matrix is contracted on its *second* index to move from the original coordinate system to the new one. To determine whether something is a tensor, it is necessary to know its matrix form in two coordinate systems that are related by a known rotation matrix.

- The rotation matrix *for a counterclockwise rotation around the z axis* by an angle θ is given by

$$\mathbf{R} = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The inverse of a rotation matrix is equal to its transpose, $\mathbf{R}^{-1} = \mathbf{R}^\top$.

- There are three special types of (second-order) tensor

$$\begin{aligned} \mathbf{S} &= \mathbf{S}^\top && \text{(symmetric),} \\ \mathbf{A} &= -\mathbf{A}^\top && \text{(antisymmetric),} \\ \alpha \mathbf{I} & \text{(or } \alpha \delta_{ij}) && \text{(isotropic).} \end{aligned}$$

- A general, (second-order) tensor $\mathbf{T}$ can be decomposed as follows,

$$\mathbf{T} = \mathbf{M} + \mathbf{A} + Q\mathbf{I},$$

where $\mathbf{M} = \frac{1}{2}(\mathbf{T} + \mathbf{T}^\top) - Q\mathbf{I}$ is symmetric and has zero trace, $\mathbf{A} = \frac{1}{2}(\mathbf{T} - \mathbf{T}^\top)$ is antisymmetric, and $Q = \text{trace}(\mathbf{T})/3$ is a scalar.

- Tensor fields are tensor functions of space (and, in some cases, time). For example, given a vector field $\mathbf{u}(\mathbf{x})$, a tensor field is produced by $\nabla \mathbf{u}$. We see this clearly by writing it in index notation, $\partial u_i / \partial x_j$.

7 What is Continuum Mechanics?

7.1 What is mechanics?

Mechanics is the study of how mass couples forces and motions. This statement is adequately general to be true but also vague. More specifically, mechanics is concerned with how forces cause mass to move and deform, and how motion and deformation create forces.

We will limit our study of mechanics to speeds that are much less than the speed of light, and hence for which relativistic effects are negligible. Moreover, we will limit our consideration to length scales that are large enough that quantum effects are negligible. This leaves us squarely in the realm of Newtonian mechanics, which can be concisely summarised with Newton's second law,

$$\sum \mathbf{F} = M\ddot{\mathbf{x}}. \quad (7.1)$$

This equation applies to a discrete object (e.g., a ball) and states that the sum of vector forces acting upon the ball is proportional to its vector acceleration $\ddot{\mathbf{x}}$. It turns out that the constant of proportionality is the mass M of the object.

Equation (7.1) describes physical things and hence is a dimensional equation. The left-hand side has units of force and, hence, so does the right-hand side. The SI metric unit of force is the Newton,

$$1 \text{ Newton} = 1 \text{ kg m/s}^2.$$

In this section of the course, it is crucial that we are careful with units in all equations. *An equal sign, beyond indicating numerical equality, implies that quantities on either side must have the same units.*

A few remarks about equation (7.1).

- The right-hand side is sometimes called the *inertial force*.
- The inertial force is the time-rate-of-change of the *momentum*, $\mathbf{m} = M\dot{\mathbf{x}}$.
- This equation is written in terms of a Cartesian basis for position. We can see this because if there are no forces on the object, we can integrate once to obtain $\dot{\mathbf{x}} = \text{const}$. This is Newton's first law, which states that an object moves with constant (vector!) velocity unless a force acts on it.
- Newton's second law gives a vector equation that is usually solved component-wise.
- We usually assume that the mass of the object is constant during the motion, i.e., $\dot{M} = 0$. We can think of this as a statement of *conservation of mass*.

Example 7.1 of Newtonian mechanics of an object

Write down Newton's second law for a ball of density ρ_b and radius R in a fluid of density ρ_f . Assume that the fluid exerts a drag force $\mathbf{F}_d = -\beta\dot{\mathbf{x}}$ on the ball.

We have three forces to consider: the body force of gravity on the ball, the buoyancy force, and the drag force. The sum of these forces balances the inertial force,

$$\mathbf{F}_g + \mathbf{F}_a + \mathbf{F}_d = M_b\ddot{\mathbf{x}}.$$

Assuming a constant gravitational acceleration $\mathbf{g}$ and combining the body force and Archimedes into one term we have

$$\frac{4}{3}\pi R^3(\rho_b - \rho_f)\mathbf{g} - \beta\dot{\mathbf{x}} = \frac{4}{3}\pi R^3\rho_b\ddot{\mathbf{x}}.$$

To clean this up, we divide through by $4\pi R^3\rho_b/3$ and rearrange to obtain

$$\ddot{\mathbf{x}} + B\dot{\mathbf{x}} - \Delta\rho'\mathbf{g} = \mathbf{0}, \quad (7.2)$$

where $B = 3\beta/4\pi R^3\rho_b$ is a scaled drag coefficient and $\Delta\rho' = (\rho_b - \rho_f)/\rho_b$ is a relative density difference.

Exercise 7.1 on Newtonian mechanics of an object

Referring to equation (7.2), consider a case where the ball (and the fluid) are initially stationary. Take a coordinate system in which $\hat{\mathbf{z}}$ points in the direction of gravity (i.e., downward).

1. Find the acceleration $\ddot{z}$ of the ball at $t = 0$.
2. Suppose that after some time, the ball stops accelerating and reaches its terminal falling speed $\dot{z}_{\text{tm}}$. Find this speed.
3. Show that

$$\dot{z}(t) = (\dot{z}_0 - \dot{z}_{\text{tm}})e^{-Bt} + \dot{z}_{\text{tm}}$$

is a solution with initial speed $\dot{z}_0$. Plot this and discuss its shape in terms of the shifting physical balances.

Exercise 7.2 on Newtonian mechanics of a damped oscillator

Consider a block of mass M on a spring with spring constant k . The block rests on a frictionless surface that is normal to gravity. The spring exerts a force $\mathbf{F} = -kx\hat{\mathbf{e}}_x$.

1. Assume that the block moves only in the x direction and write Newton's second law for this motion. Solve for the motion of the block at $t > 0$ for the initial condition $x(0) = x_0$, $\dot{x}(0) = 0$.
2. Extend Newton's second law for the block to model a frictional force of the form $\mathbf{F}_f = -\beta\dot{x}\hat{\mathbf{e}}_x$. Rewrite your governing equation in the form

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0$$

and give definitions and units for parameters γ and ω_0 .

3. Use the trial solution

$$x(t) = Ae^{(i\omega_1 + \sigma)t}$$

where A is a complex number. Find $\sigma(\gamma)$ and $\omega_1(\gamma)$, both pure reals. Hence write the full solution. Note the value of γ above which the underdamped solution must be replaced by its critically damped or overdamped form.

4. Plot several valid solutions for different values of γ and explain in physical terms.

7.2 What is a continuum?

A continuum is a mathematical concept that is used to describe a class of physical objects: regions of space that contain matter. The matter could be in any physical state or states (gas, liquid, solid, plasma), but it is assumed that the matter fills the space so that there are no sub-regions of the space that contain no matter.

The concept of a continuum clearly cannot be valid at the molecular scale, where there are discrete particles that have mass, with massless voids between them. Indeed, we require the continuum to have a minimum scale that is sufficiently large that we can define a mass-density $\rho(\mathbf{x})$ that is a smooth function of position, but no larger. This means we are considering scales larger than the molecular scale, as shown in figure 7.1a.

The properties that we define to describe the continuum (density, displacement, velocity, stress, etc.) are then all defined as averages over a volume that is as small as possible (in the sense of the previous paragraph). This volume is known as the representative elementary volume (REV) and is depicted (for a liquid) in figure 7.1b. At the level of the continuum model, the REV exists at a point with position $\mathbf{x}$. But at the molecular scale, the REV is a box containing many molecules — sufficiently many that volume averages are well defined and vary smoothly from one REV to its neighbour.

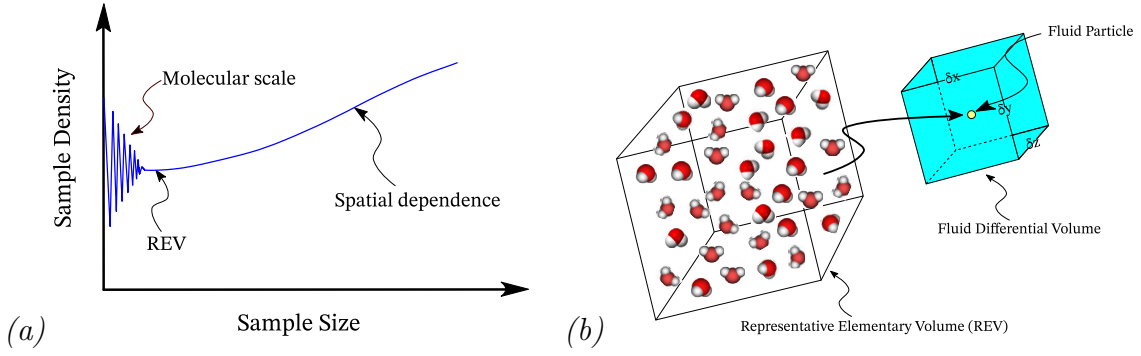


Figure 7.1: Basic concepts in continuum mechanics. (a) Density as a function of the spatial scale on which it is measured. The x axis should be considered to have a logarithmic scale. (b) The representative elementary volume (REV) of a liquid (for example) that contains many molecules but is small compared to the scale of spatial dependence of averaged properties. Source: [10.1016/C2016-0-04780-0](https://doi.org/10.1016/C2016-0-04780-0), Chapter 1.

The mechanics of continua are formulated in terms of these volume averages. Hence we can derive and analyse models of how density varies in space and time within a continuum, and the same applies to displacement, velocity, stress, etcetera. This is the purpose of continuum mechanics.

Figure 7.2 illustrates the meaning of averaged variables representing displacement for an elastic continuum in panel (a), and velocity for a fluid medium in panel (b). Both the displacement $\mathbf{u}(\mathbf{x}, t)$ and the velocity $\mathbf{v}(\mathbf{x}, t)$ are vector fields, and hence almost all of vector calculus is relevant in their analysis. The displacement field represents how REVs at positions $\mathbf{x}$ in the initial configuration of the elastic solid have been displaced to new positions $\mathbf{X}$ in the deformed configuration. Hence $\mathbf{u}$ is a vector pointing from $\mathbf{x}$ to $\mathbf{X}$. The fluid velocity has a different meaning, of course. It is the *rate* of change of position of REVs. In a fluid the displacements can be very large and complicated, and hence we don't usually record the 'initial' and 'final' positions of a REV.

There is, nonetheless, a relationship between the velocity $\mathbf{v}$ and the displacement $\mathbf{u}$ that has limited validity, as we will discuss further below. Consider an REV with velocity $\mathbf{v}(t)$. In a short time interval δt , it moves a distance that is approximately $\mathbf{v}(t)\delta t$. This distance is a displacement from its initial position $\mathbf{x}(t)$ to its new position $\mathbf{X}(t + \delta t)$. Hence we can write

$$\mathbf{v}(t) = \frac{\mathbf{u}(t)}{\delta t} = \frac{\mathbf{X}(t + \delta t) - \mathbf{x}(t)}{\delta t}.$$

Taking the limit of $\delta t \rightarrow 0$, we see that

$$\mathbf{v}(\mathbf{x}, t) = \dot{\mathbf{u}}(\mathbf{x}, t). \quad (7.3)$$

This relationship is certainly true at the instant when deformation begins and $\mathbf{u} = \mathbf{0}$. It is approximately true when the displacement is sufficiently small that $\mathbf{v}(\mathbf{x}, t)$ is still a good approximation of $\mathbf{v}(\mathbf{X}, t)$.

If displacements are sufficiently large that the velocity differs over the displacement, then the relationship (7.3) is no longer correct. Fortunately, for elastic con-

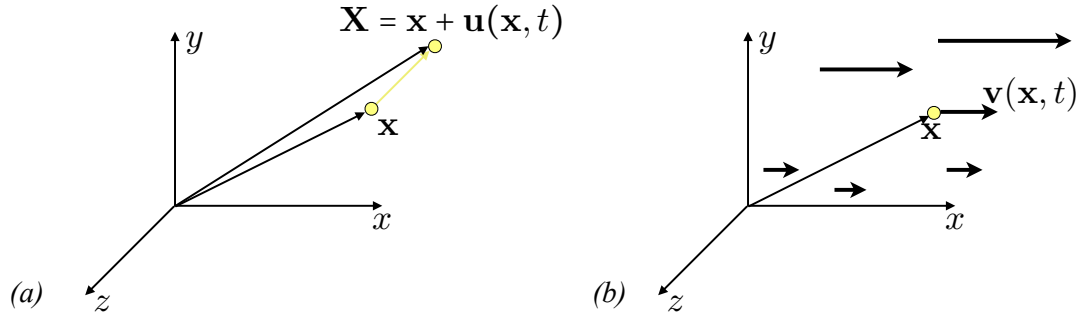


Figure 7.2: Variables describing motion in continua. (a) The displacement field $\mathbf{u}(\mathbf{x}, t)$ used to describe an elastic deformation. (b) The velocity field $\mathbf{v}(\mathbf{x}, t)$ used to describe a fluid flow.

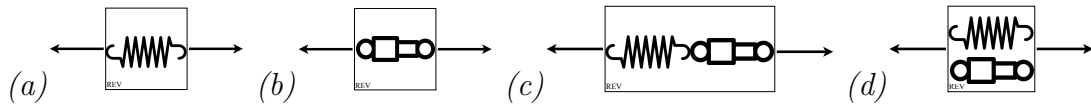


Figure 7.3: Rheological models. (a) Elastic model. (b) Viscous model. (c) Maxwell viscoelastic model. (d) Kelvin viscoelastic model.

tinua we will consider only small displacements, and for fluid continua we will not be concerned with displacements $\mathbf{u}$.

7.3 What is rheology?

The particles that move through space in basic Newtonian mechanics are generally assumed to be rigid. The forces on them cause acceleration (or at least motion), not deformation. But in a continuum, forces cause the matter to deform: metals bend, liquids pour, gases convect, and there are myriad other examples in geoscience alone. *The scientific study of the relationship between forces and deformations is called rheology.* (Informally, the relationship between forces and deformations itself is called rheology — as in, “what is the rheology of the mantle?”).

In this course we will consider two rheological laws, the law of elasticity and the law of viscosity. As Earth scientists, we know that materials (such as hot rock) behave differently on different time scales. The mantle transmits seismic waves but also flows. In the case of elasticity, force (actually stress, which is force per unit area) is proportional to strain (the gradient of displacement); in the case of viscosity, stress is proportional to strain rate (the gradient of displacement rate). Despite this difference, most of the continuum theory is shared between these two situations. You will often hear people describe these two rheological models as “solids and fluids” but keep in mind that liquids can have elastic properties and solids can have viscous properties. Fluids are anything that flows.

Rheological models are often represented in terms of physically familiar objects, as in figure 7.3. Elasticity is represented as a spring, because the elastic rheological law is essentially the same as Hooke’s law for a spring. Viscosity is represented as a dashpot because displacement of the piston occurs at a rate that is proportional to the applied force (at least over the range of intended forces).

8 Newton's Second Law for a Continuum

Newton's second law must apply to every parcel of mass in the universe (except for relativistic or quantum situations, of course). So it must apply to any REV within a continuum. Our work in this section is to convert equation (7.1) from a statement about a discrete and rigid particle to a statement that holds at every REV in a continuum. We note that our REV is conceptual rather than physical; it doesn't have an intrinsic volume and so we cannot specify its mass, M . Therefore we must work with density ρ , the mass per unit volume. *Our statement of Newton's second law for a continuum therefore balances forces per unit volume.* Note that our SI units are

$$1 \text{ N/m}^3 = 1 \text{ Pa/m} = 1 \text{ kg s}^{-2} \text{ m}^{-2}.$$

We now dive into the mathematics of a continuum (hoping that it is a liquid with the viscosity and density of water...)!

8.1 Acceleration and the Lagrangian derivative

Think back to the Newtonian mechanics exemplified in example 7.1. We were concerned with the motion of the ball and, in particular, with the ball's centre of mass, which was located at a position $\mathbf{x}(t)$. The acceleration $\ddot{\mathbf{x}}$ referred specifically to the ball as it moved through space. This was so obvious we didn't need to mention it. But suppose that we wanted to describe the motion of the fluid surrounding the ball. In that case, the acceleration is a vector field, not a single vector! We cannot simply write the acceleration as the second derivative of a single position vector, $\ddot{\mathbf{x}}$.

This is a notational problem, not a physical one. We simply need a better notation. A good start would be to use the partial derivative, rather than a dot. In a continuum, the velocity field is defined over all positions $\mathbf{x}$ and hence we could define an acceleration field at all positions as

$$\frac{\partial \mathbf{v}}{\partial t} \quad (\text{Eulerian contribution to acceleration}). \quad (8.1)$$

We call this the Eulerian contribution to the acceleration (or, more simply, the *Eulerian acceleration*) because it refers to positions $\mathbf{x}$ in a fixed reference frame. (A fixed reference frame is called an Eulerian reference frame). This reference frame is illustrated in figure 8.1 by the fox named Euler that sits in a fixed position on the riverbank, watching the flow. She looks at one position $\mathbf{x}$ and measures the acceleration there, as the water moves through it.

It is important to note that Newton's second law refers to a specific piece of matter as it moves through space (think of the ball). So the Eulerian contribution to the acceleration is nice, but it isn't exactly what we need because it doesn't track the matter as it moves. An acceleration that tracks the matter has a special notation,

$$\frac{D\mathbf{v}}{Dt} \quad (\text{Lagrangian acceleration}). \quad (8.2)$$

We call this the *Lagrangian acceleration* because it follows a material parcel along its trajectory $\mathbf{X}(t)$. Here $\mathbf{X}(t)$ is measured in the fixed spatial frame; it is not itself

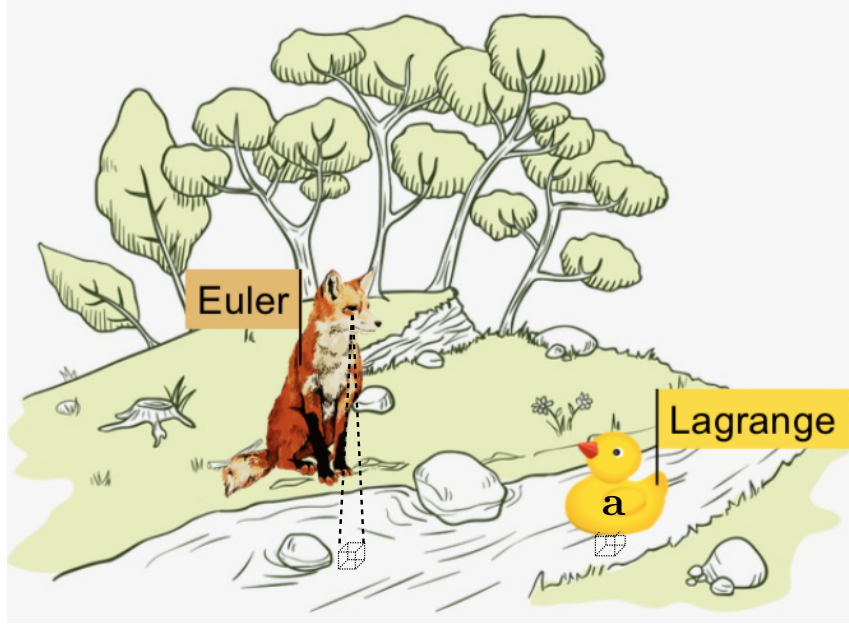


Figure 8.1: The Eulerian and Lagrangian perspectives on the flow of a river. Source: CT.

a fixed coordinate. The notation indicates that the field it describes is moving with the flow itself — a somewhat confusing concept at first encounter. It is illustrated in figure 8.1 by the (rubber) duck, which is stuck to the water beneath it. At a time t , the duck is at a position $\mathbf{X}(t)$ and measures the acceleration there. The acceleration it measures is always that of the same water molecules — the ones contained in the REV it is stuck to. (This also applies to elastic media, where displacements are much smaller.)

Now you may be thinking, I know how to evaluate the Eulerian derivative already, but how do I evaluate a Lagrangian derivative? An excellent question. It seems as if we'd need to know $\mathbf{X}(t)$. This is sorta true, but it isn't as bad as it sounds. To see how this works, let's first recognise that the Eulerian and Lagrangian derivatives are not exclusively related to acceleration — they are defined for any field where the quantity can be transported by material motion. This includes scalar fields, so let's focus our attention there for simplicity.

Consider a scalar field $Q(\mathbf{x}, t)$ (you can imagine it is the concentration of some tracer isotope, if you wish). Now this quantity moves with the material (say, water) that it is in. We can consider a specific REV of water at position $\mathbf{x} = \mathbf{X}(t)$ and ask, how does $Q(\mathbf{X}(t), t)$ change along its path. We learn this by taking a *total derivative* dQ , a derivative with respect to all of its arguments. Using the chain rule,

$$\begin{aligned} dQ &= \frac{\partial Q}{\partial t} dt + \frac{\partial Q}{\partial x} dX + \frac{\partial Q}{\partial y} dY + \frac{\partial Q}{\partial z} dZ, \\ \Rightarrow \frac{dQ}{dt} &= \frac{\partial Q}{\partial t} + \underbrace{\left(\frac{\partial Q}{\partial x}, \frac{\partial Q}{\partial y}, \frac{\partial Q}{\partial z} \right)}_{\nabla Q} \cdot \underbrace{\left(\frac{dX}{dt}, \frac{dY}{dt}, \frac{dZ}{dt} \right)}_{d\mathbf{X}/dt}. \end{aligned}$$

Now we can recognise that $d\mathbf{X}/dt = \mathbf{v}$ and that $dQ/dt = DQ/Dt$. Hence we have

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla Q. \quad (8.3)$$

This very important equation tells us how to evaluate the Lagrangian derivative DQ/Dt at a time t and position $\mathbf{x}$ given the fields $Q(\mathbf{x}, t)$ and $\mathbf{v}(\mathbf{x}, t)$. (Note, however, that it isn't necessary to differentiate the velocity field.) It says that the Lagrangian derivative has two contributions: the Eulerian derivative and the *advective derivative*.

The advective derivative sometimes appears when Q is in a steady state. For example, we will later encounter the equation

$$\mathbf{v} \cdot \nabla Q = 0.$$

This equation has an important meaning. We can see this by recalling that $\mathbf{v}$ is everywhere tangent to streamlines of a flow. So if the directional derivative of Q along streamlines is zero, it means that Q is *constant along streamlines*. This tells us that Q is transported with the flow without loss.

8.2 Inertial force per unit volume

Using equation (8.3), we can now write an equation for the Lagrangian acceleration of (8.2). It is

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}. \quad (8.4)$$

The Eulerian term is straightforward to evaluate by taking the derivative of each component of the vector $\mathbf{v}$. The advective derivative is more complicated because $\nabla \mathbf{v}$ is a tensor. We can clarify the situation with index notation. Writing an equation for the k^{th} component of $\mathbf{v}$,

$$\begin{aligned} \frac{Dv_k}{Dt} &= \frac{\partial v_k}{\partial t} + (v_i \hat{\mathbf{e}}_i) \cdot \left(\hat{\mathbf{e}}_j \frac{\partial}{\partial x_j} v_k \right), \\ &= \frac{\partial v_k}{\partial t} + v_j \frac{\partial v_k}{\partial x_j}. \end{aligned}$$

This perspective shows that we can group entries in the advective derivative as $(\mathbf{v} \cdot \nabla)\mathbf{v}$. Writing it this way, we see an expression where there is no tensor, but rather a scalar operator $(\mathbf{v} \cdot \nabla)$ acting on a vector $\mathbf{v}$. Since (8.4) is a vector equation, we should evaluate it component-wise. The x component is

$$\frac{Dv_x}{Dt} = \frac{\partial v_x}{\partial t} + \left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_z \frac{\partial}{\partial z} \right) v_x,$$

and similarly for the y and z components.

Multiplying the Lagrangian acceleration of equation (8.4) by the density gives the right-hand side of our continuum version of Newton's second law,

$$\rho \frac{D\mathbf{v}}{Dt} \quad (\text{Inertial force per unit volume}). \quad (8.5)$$

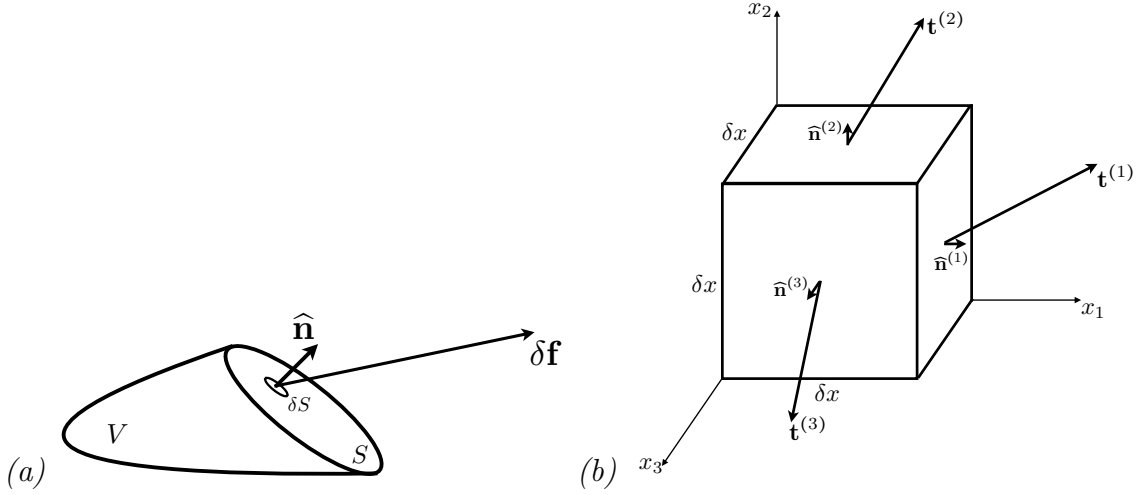


Figure 8.2: Traction. (a) A surface force $\delta \mathbf{f}$ distributed on a patch δS with normal vector $\hat{\mathbf{n}}$. The surface S is an arbitrary cut through a continuum that has volume V . (b) An REV with edges of length δx , aligned with a Cartesian coordinate system. The faces are numbered by the axis that is parallel to their normal. Three normal vectors and three traction vectors are shown.

In elastic deformations (e.g., seismic waves), the transport of momentum by displacements is typically small relative to the Eulerian acceleration. In that case,

$$\frac{D\mathbf{v}}{Dt} \approx \frac{\partial \mathbf{v}}{\partial t} \approx \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (\text{for small displacements}). \quad (8.6)$$

8.3 Body force per unit volume

Just as the ball of example 7.1 felt the force of gravity in proportion to its mass, so too can every REV of a continuum feel a gravitational force. However, as in the previous section, we work with the mass density within the REV, rather than the mass. So the gravitational force enters as a force per unit volume. It acts throughout the REV, on all of the mass it contains. Hence it is known as a *body force* (in contrast to surface forces, that we will meet below).

There are other body forces. These can, for example, arise from interaction of matter with electric and magnetic fields. The Lorentz force is a body force exerted by a magnetic field on a moving conductor. It plays an important role in the dynamics of the outer core. In this course we will be concerned only with the gravitational body force.

Recall that in equation (4.13) we defined the field $\mathbf{g}(\mathbf{x})$ as the gravitational force per unit mass. Multiplying this by the density ρ gives us the gravitational force per unit volume. This will slot into the continuum version of Newton's second law.

$$\rho \mathbf{g} \quad (\text{Gravitational body force per unit volume}). \quad (8.7)$$

8.4 Surface force and the traction vector

The same REV that feels the pull of gravity also feels forces on its surface. These are exerted by the surrounding material and need to be accounted for in the balance

of forces. However, these forces are not applied at a point on the exterior of the REV, but are rather distributed over its surface. Hence we need to consider the surface force per unit area, which is called the *traction*. We will refer to traction with the symbol $\mathbf{t}$ (not to be confused with the tangent vector, for which we used the same symbol earlier in the course).

We can think of the traction as the limit of the force on a finite patch of area within the continuum as the size of that patch goes to zero. Figure 8.2a shows a cut through a volume V containing a continuum. The cut makes a surface S and a small patch of that surface is called δS with unit normal vector $\hat{\mathbf{n}}$. The force on that surface is $\delta \mathbf{f}$. Then, allowing the patch to shrink in size, we define the traction as

$$\mathbf{t} = \lim_{\delta S \rightarrow 0} \frac{\delta \mathbf{f}}{\delta S}.$$

Because the force is understood to diminish in proportion to the area, this has a well-defined, finite limit: the traction at a point.

Here are a few important properties of the traction.

- The component of the traction normal to the surface is given by $\mathbf{t} \cdot \hat{\mathbf{n}}$. This is called the *normal traction*.
- Any surface that passes through the same point and has the same normal vector at that point also has the same traction vector. That is to say, *traction is a function of position and surface-normal only*; it is entirely local.
- Given a surface with normal $\hat{\mathbf{n}}$ and traction $\mathbf{t}^{(\hat{\mathbf{n}})}$, the opposite side of the same surface has a traction that is equal in magnitude and opposite in direction. Mathematically,

$$\mathbf{t}^{(\hat{\mathbf{n}})} = -\mathbf{t}^{(-\hat{\mathbf{n}})}.$$

This is Cauchy's stress lemma and is an expression of Newton's third law for continua.

Figure 8.2b shows the normal and traction vectors at two points on the surface of an REV. Assuming that we know the traction everywhere on the surface of this REV, we can write the net force on the REV due to traction as the integral

$$\frac{1}{\delta V} \int_S \mathbf{t} dS, \quad (\text{Net surface force per unit volume on REV}) \quad (8.8)$$

where S is the surface that encloses the REV (shown as a cube in fig. 8.2b) and $\delta V = \delta x^3$ is its volume.

Now we have to confront the fact that we don't actually know the traction vector everywhere, and an REV is a fictitious construct — what we really want is the net force per unit volume at a point. We shall see below that our expression (8.8) will prove useful in obtaining this.

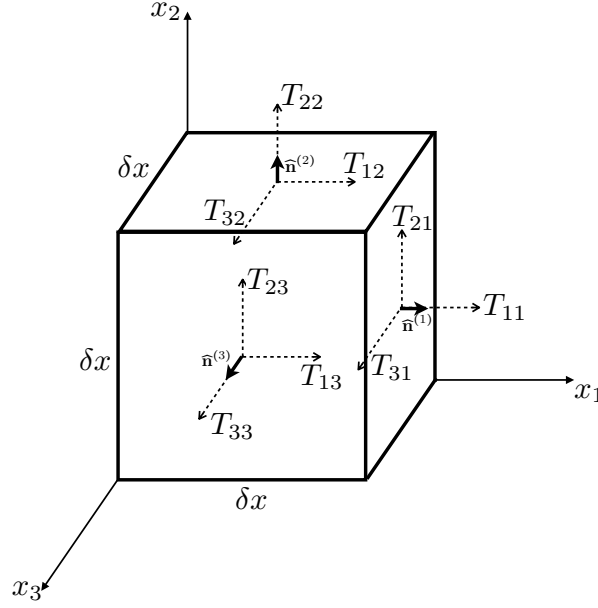


Figure 8.3: The REV for development of the stress tensor, with the positive faces numbered according to the direction of their normals. Positive directions of the stress-tensor components are shown on the positive faces. On the negative faces (not shown), the positive stress direction is opposite those shown. Source: LM (adapted).

8.5 The Cauchy stress tensor

We'd like to express the vector $\mathbf{t}$ as a function of both position and unit normal vector. This would enable us to evaluate tractions on any surface within a continuum (e.g., on the surface bounding our REV). What sort of mathematical object would be suitable for this? A tensor field! A tensor field can be evaluated at a position and can, at that point, be dotted with a vector to give another vector. The *stress tensor* field is exactly the field that we need. This will be demonstrated below.

Note that we define stress as force per unit area, which is the same as traction. But whereas traction is a vector that is defined with reference to a specified surface (normal), stress is a tensor that quantifies traction on a surface of any orientation. Hence the stress tensor allows us to compute the traction on *any* surface with orientation given by its normal vector $\hat{\mathbf{n}}$, at a position $\mathbf{x}$ within the domain. The divergence of the stress tensor gives the net force per unit volume—a vector—at any position within the continuum.

Below we discuss the formulation and properties of the stress tensor. Relevant proofs can be found in [LM¹¹](#). In developing the theory, we will refer to an REV that is cubic, as shown in figure 8.3. It has sides of length δx and faces aligned with the Cartesian axes x_i . The normal vectors are identified by the coordinate axis with which they are parallel.

¹¹There is also a very nice Wikipedia page about the Cauchy stress tensor [here](#).

Entries in the stress tensor The stress tensor in $\mathbb{R}^3$ is represented as a 3×3 matrix with entries T_{ij} ,

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix}. \quad (8.9)$$

The first index indicates the direction of the traction component. The second index indicates the direction of the normal to the REV face. Thus, for example,

T_{11} acts in the x_1 direction on a face with normal in the x_1 direction. It is a normal stress.

T_{12} acts in the x_1 direction on a face with normal in the x_2 direction. It is a shear stress.

T_{33} acts in the x_3 direction on a face with normal in the x_3 direction. It is a normal stress.

T_{23} acts in the x_2 direction on a face with normal in the x_3 direction. It is a shear stress.

It can be shown that the columns of the stress tensor are the three traction vectors $\mathbf{t}^{(j)}$ that act on faces with normals $\hat{\mathbf{n}}^{(j)}$. These three vectors describe the stress state at a point and allow us to find the traction on a surface with any normal. We shall see below that the entries in those three vectors (and hence in the stress tensor itself) are not entirely independent.

Sign convention(s) of the stress tensor There are two important sign conventions that are required to clarify the meaning of the stress tensor. These are indicated by the direction of the arrows labelled with the tensor components in figure 8.3.

First, *normal stresses are positive when they are directed outward from faces of the REV*. Hence if $T_{11} > 0$, then the REV is in a state of tension in the x_1 direction¹². This is true on faces with normal vectors that are in the negative direction. So the traction on the REV face with normal $-\hat{\mathbf{n}}^{(1)}$ points toward negative x_1 if T_{11} is positive.

Second, *shear stresses are positive when they point in a positive coordinate direction on a positive face*. A positive face has a normal in the positive coordinate direction. On a negative face, a shear stress is positive if it points in a negative coordinate direction. So, for example, $T_{12} > 0$ means that the traction on a face with normal $\hat{\mathbf{n}}^{(2)}$ acts in the positive x_1 direction, whereas the traction on the opposite face acts in the negative x_1 direction.

¹²In rock mechanics and soil mechanics, compressive stress is taken to be positive, which has implications for sign changes elsewhere in the theory.

Exercise 8.1 on sign conventions for 2-D stress

In $\mathbb{R}^2$, the stress tensor is 2×2 and the REV can be represented as a square aligned with the coordinate directions. Draw this square and indicate the positive directions of the normal and shear stresses. Do this on both the positive and negative faces.

Symmetry of the stress tensor The stress tensor is symmetrical, meaning that

$$\mathbf{T} = \mathbf{T}^\top \quad \text{or} \quad T_{ij} = T_{ji}. \quad (8.10)$$

Therefore, instead of 9 independent entries in the stress tensor, there are only 6 independent entries. For example, if we know T_{13} , then we know, by symmetry, T_{31} because it is equal to T_{13} .

This follows from conservation of angular momentum. Looking at figure 8.3, imagine that $T_{12} \neq T_{21}$. The stresses then apply a net torque to the REV about the x_3 direction. In an ordinary continuum with no distributed torques (body couples), local angular-momentum balance requires this torque to vanish (to avoid giving material elements an accelerating spin). Hence $T_{12} = T_{21}$, and similarly $T_{ij} = T_{ji}$.

Application of the stress tensor As previously noted, the stress tensor has the following wonderful property,

$$\mathbf{t} = \mathbf{T} \cdot \hat{\mathbf{n}} \quad \text{or} \quad t_i = T_{ij} \hat{n}_j, \quad (8.11)$$

which can be evaluated for any position $\mathbf{x}$ and any surface defined by normal $\hat{\mathbf{n}}$. Since the stress tensor is symmetrical, $T_{kj} \hat{n}_j = T_{ik} \hat{n}_i$.

Exercise 8.2 on stress and traction

Consider a stress tensor in two dimensions given by

$$\mathbf{T}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that this tensor field is uniform in space! Write down the outward normal vector $\hat{\mathbf{n}}$ on the unit square with faces at $x = 0, 1$ and $y = 0, 1$ in terms of its components. Use these to evaluate the traction vector $\mathbf{t}$ along the sides of the square.

Equation (8.11) is valuable because it can be substituted into (8.8), which can then be manipulated as follows,

$$\begin{aligned} \frac{1}{\delta V} \int_S \mathbf{t} \, dS &= \frac{1}{\delta V} \int_S \mathbf{T} \cdot \hat{\mathbf{n}} \, dS, \\ &= \frac{1}{\delta V} \int_S \mathbf{T} \cdot d\mathbf{S}, \\ &= \frac{1}{\delta V} \int_{\delta V} \nabla \cdot \mathbf{T} \, dV, \end{aligned}$$

where we have used the Divergence theorem. Then, allowing the size of the REV to shrink,

$$\lim_{\delta V \rightarrow 0} \frac{1}{\delta V} \int_{\delta V} \nabla \cdot \mathbf{T} dV = \nabla \cdot \mathbf{T},$$

where we have used equation (4.3) in taking the limit. Hence we obtain the quantity that we needed,

$$\nabla \cdot \mathbf{T} \quad (\text{Net surface force per unit volume}). \quad (8.12)$$

We next use this to write Newton's second law.

8.6 Newton's second law for continua

We are now in a position to write down Newton's second law for a point (hosting a vanishingly small REV) within a continuum. Our equation has units of force/volume. On the left-hand side, in the equivalent of $\sum \mathbf{F}$, is the net surface force from (8.12) and the body force from (8.7); on the right-hand side, in the equivalent of $M\ddot{\mathbf{x}}$, is the rate of change of momentum (the inertial force) from (8.5). Together these are

$$\nabla \cdot \mathbf{T} + \rho \mathbf{g} = \rho \frac{D\mathbf{v}}{Dt} \quad \text{or} \quad \frac{\partial}{\partial x_j} T_{ij} + \rho g_i = \rho \frac{Dv_i}{Dt}. \quad (8.13)$$

This extremely important equation is the foundation for (almost) all continuum mechanics. It is sometimes known as Cauchy's stress equation.

This equation appears to be suitable for fluids, where motion is described by a velocity field. But it is equally suitable for elastic media. In fact, since the displacements are generally very small in elastic deformation, it is typically assumed that the momentum they transport is negligible and hence, using (8.6),

$$\nabla \cdot \mathbf{T} + \rho \mathbf{g} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (\text{for small displacements}). \quad (8.14)$$

This equation is used in elastodynamics and governs, for example, the propagation of seismic waves.

Good news. We have derived a continuum version of Newton's second law, which is a statement of conservation of momentum. It should, in combination with suitable boundary conditions, allow us to solve for the motion of a continuum subject to external forces.

Bad news. To solve for the velocity or displacement, we need to know the stress tensor throughout the continuum at all times. How can we possibly know this? The answer is that we need to use additional physics linking the stress to the motion, $\mathbf{T}(\mathbf{v}(\mathbf{x}))$ or $\mathbf{T}(\mathbf{u}(\mathbf{x}))$. This relationship is known as a *constitutive law*.

More good news. We cover elastic and viscous (and viscoelastic) constitutive laws in chapter 10.

8.7 Pressure and stress

Pressure is the isotropic component of the stress — meaning that it is the same in all coordinate directions. Recall from section 6.4 that symmetric tensors can be decomposed into a deviatoric part (with zero trace) and an isotropic part. For the case of the stress tensor, we recall that we have chosen to define tension with a positive sign. However a positive pressure implies compression, not tension. So we must reverse the sign of pressure to ensure it makes sense in the decomposition of equation (6.23). The decomposition then reads

$$\mathbf{T} = \boldsymbol{\tau} + (-P)\mathbf{I} \quad \text{or} \quad T_{ij} = \tau_{ij} - P\delta_{ij}, \quad (8.15)$$

where

$$P = -\frac{1}{3}\text{trace}(\mathbf{T}).$$

Here we have introduced a tensor field $\boldsymbol{\tau}$ called the deviatoric stress. Although it may have non-zero entries on the diagonal, by its definition it has $\text{trace}(\boldsymbol{\tau}) = 0$.

Example 8.1 of stress and static pressure

Consider a static fluid (i.e., zero velocity, zero acceleration). Assign a coordinate system with $z = 0$ at the surface of the fluid and $\hat{\mathbf{z}}$ pointing downward. By symmetry, there are no gradients in the x or y directions within the fluid, so all $\partial/\partial x$ and $\partial/\partial y$ are zero. We consider the z direction and define the magnitude of gravity as $g = \mathbf{g} \cdot \hat{\mathbf{z}}$. Equation (8.13) reduces to

$$\frac{\partial}{\partial x_i} T_{iz} + \rho g = 0 \quad \Rightarrow \quad \frac{\partial T_{zz}}{\partial z} = -\rho g.$$

The sum over i has only the z entry because of the symmetry noted above. This equation states that the tensile normal stress decreases with depth (or, in other words, becomes more compressive with depth). Following (8.15), we write the tensile stress as equal to minus the pressure, a compressive stress per unit area: $T_{zz} = -P$. Then we have

$$\frac{\partial P}{\partial z} = \rho g \quad \Rightarrow \quad P(z) = \rho g z + c,$$

where c is a constant of integration. If the pressure at $z = 0$, the surface of the fluid, is P_0 , then $c = P_0$ and we have

$$P(z) = P_0 + \rho g z.$$

This is the equation of static pressure in a gravitational field. The pressure is called *hydrostatic* in water and *lithostatic* in rock, and has other names in other contexts.

Exercise 8.3 on static pressure

Consider the example 8.1 above, where we obtained $P(z)$. Provide an equation for the full solution $P(x, y, z, t)$. Then, using physical intuition and the relationship between pressure and stress, write down the stress-tensor entries T_{xx} and T_{yy} . Hence, write the full (static) stress tensor $\mathbf{T}$ as an array. What is the deviatoric stress in this case?

Summary of §8

- Continuum mechanics is an extension of Newton's second law, $\sum \mathbf{F} = M\ddot{\mathbf{x}}$, to apply to points in a continuum domain containing mass. The theory is derived by considering forces per unit volume acting on representative elementary volumes (REVs).
- The derivative of a quantity Q in a reference frame that is moving with a materially fixed REV is expressed with a Lagrangian derivative,

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla Q.$$

The terms are the Eulerian and advective derivatives, respectively.

- The acceleration of a continuum is expressed by the Lagrangian derivative,

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}.$$

- The traction on an REV surface, whether that surface is external or internal to a continuum, is written as $\mathbf{t}$. It is the force per unit area on an infinitesimal patch of the surface. The normal component of traction is positive if it imposes a tension on the REV.
- The Cauchy stress-tensor field $\mathbf{T}(\mathbf{x}, t)$ represents the state of stress at any point $\mathbf{x}$ within a continuum for an (internal) surface with any unit normal vector $\hat{\mathbf{n}}$. The traction on that surface at that point is given by $\mathbf{t} = \mathbf{T} \cdot \hat{\mathbf{n}}$.
- The stress tensor has 9 entries but is symmetric. This means that $\mathbf{T} = \mathbf{T}^\top$ and hence there are only 6 independent entries in the tensor. The stress tensor has two sign conventions. The first is that tension is positive. The second is that shear-stress is positive if it is in a positive axis-direction on a positive face of the REV.
- Conservation of momentum equation for all continua (Newton's second law),

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \mathbf{T} + \rho \mathbf{g}, \quad \text{or} \quad \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \nabla \cdot \mathbf{T} + \rho \mathbf{g}.$$

The second equation applies if displacements are small (elastic solids).

- The pressure is defined as minus the isotropic part of the stress tensor, $P(\mathbf{x}) = -\text{trace}(\mathbf{T}(\mathbf{x})) / 3$. With this we can write the stress tensor in terms of its isotropic and deviatoric parts, $\mathbf{T} = -P\mathbf{I} + \boldsymbol{\tau}$.

9 Conservation of Mass

In the example of a ball falling through a fluid, the ball was rigid and had mass M . We took for granted that the mass was constant during the motion of the ball. But in continuum mechanics, we consider not the total mass but rather the mass density $\rho(\mathbf{x})$. At sufficiently large (or small) pressures, this density can change locally. Sound waves are an example of local changes in density.

But even with density changes, mass must be conserved. That is, it is possible to move mass around, but it cannot be created or destroyed (under non-relativistic conditions). So we can make a strict accounting of mass. That is what we do in this chapter. The most important part of this was covered as an application of the Divergence theorem in section 4.1.3.

9.1 Conservation of mass for fluids

The basic idea here is to treat an REV as a mass storage unit and monitor the mass it stores,

(rate of change of mass in REV) = (rate of transport of mass into REV through its boundary).

The total mass within an Eulerian REV (that is, an REV that is fixed in space) with surface S and volume V is given by $\int_V \rho dV$. The mass flux is $\rho \mathbf{v}$. Using these, we can rewrite the word-equation above as

$$\frac{d}{dt} \int_V \rho dV = - \int_S \rho \mathbf{v} \cdot d\mathbf{S}.$$

Bringing the time derivative into the integral and using the Divergence theorem on the right-hand side gives

$$\int_V \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right) dV = 0.$$

Since the REV is arbitrary, we can allow it to be sufficiently small that the quantities within it are constant to a good approximation. Then we have

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (9.1)$$

Applying the product rule to the divergence term, we can re-write this equation as

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (9.2)$$

which states that relative density changes in a frame that is moving with a fluid parcel are proportional to the local divergence of velocity.

Exercise 9.1 on the mass conservation equation

Derive (9.2) from (9.1), showing all steps and assumptions.

For many flows, changes in density are negligible. These flows are incompressible in the sense that $D\rho/Dt = 0$, even if the density is not necessarily uniform (though it is usually also uniform). Applying this condition to (9.2) we obtain the *Continuity equation*,

$$\nabla \cdot \mathbf{v} = 0. \quad (9.3)$$

This is also known as the incompressibility condition. In this course we will only consider incompressible flows.

9.2 Conservation of mass for elastic solids

When modelling an elastic solid, it is usually more convenient to work in terms of displacement $\mathbf{u}(\mathbf{x})$. Assuming that displacement gradients and hence changes in density from a uniform state are small, we can integrate equation (9.2) in time and rearrange to give

$$\rho(\mathbf{x}) = \rho_0 (1 - \nabla \cdot \mathbf{u}). \quad (9.4)$$

This equation states that differences in density from the reference state are due to divergence of the displacement field.

Exercise 9.2 on density changes in elastic solids

Derive (9.4) from (9.2), showing all steps and assumptions.

Summary of §9

- Mass conservation for a compressible flow is written as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0.$$

- If the flow is incompressible (constant density), then mass conservation becomes the continuity equation

$$\nabla \cdot \mathbf{v} = 0.$$

- For elastic media with small displacements, mass conservation integrates to

$$\rho = \rho_0 (1 - \nabla \cdot \mathbf{u}).$$

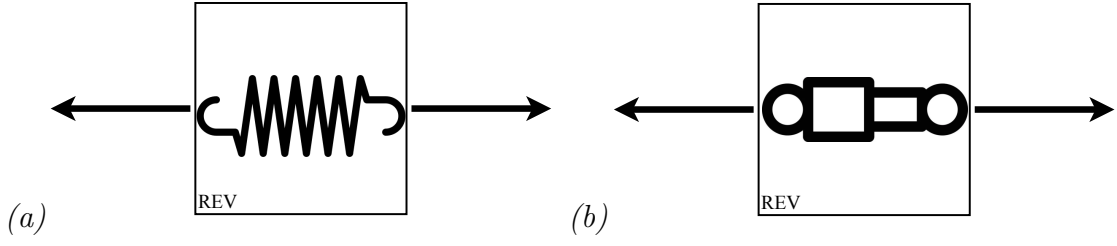


Figure 10.1: Representative elementary volumes depicting rheological models. Arrows indicate the state of stress, which is unidirectional and positive (tensile). (a) The elastic rheological model is represented as a spring. (b) The viscous rheological model is represented as a dashpot.

10 Rheological Models of Fluids and Solids

In this section we develop the constitutive relationships required to write the stress tensor $\mathbf{T}$ in terms of the velocity $\mathbf{v}$, in the case of a viscous constitutive law, or the displacement $\mathbf{u}$, in the case of an elastic one. We also consider a simple version of viscoelasticity, that combines these two approaches.

To build intuition, we first look at the simplest case where deformation occurs at a point and only in one direction. This avoids the use of vectors and tensors, and focuses on the physical model. We then build up the tensors required to provide a full description of the deformation. Finally, we derive the full tensor constitutive laws.

10.1 Elastic and viscous rheological models (scalar version)

The concept of a zero-dimensional model of deformation is awkward because to deform something means to change its shape. But a 0-D object is a point, and hence cannot have a shape. So what we investigate in this section is a one-dimensional deformation of material at a point. It is useful to think of this as a 1-D deformation of an REV, as shown in figure 10.1.

The stress that is depicted is a tensile stress, which we shall call T . A negative value of T indicates a compressive stress. The same stress T is applied to both faces of the REV, which means that there is no stress gradient, and hence there is no acceleration. We also ignore the body force for this section.

10.1.1 Elastic deformation

Elasticity, depicted schematically in Fig. 10.1a, can be thought of as the force associated with elongation or compression of a spring. It therefore relates a stress with an elongation. Let's assume that the spring in the REV has a neutral length of ℓ_0 that is a positive but small quantity. The elongation, which we'll refer to as the *strain*, is then defined to be the change in length relative to the neutral length,

$$E = \frac{\ell - \ell_0}{\ell_0}. \quad (10.1)$$

It is important to note that strain is dimensionless!

Now we can write an elastic rheological law in general terms as $T = f(E)$, where f is some scalar function. We would expect that f is an odd function, such that negative strains (shortening) correspond to negative stresses (compression). We would also expect that f is monotonically increasing with E . *The simplest rheological law that conforms to these basic expectations is linear,*

$$T = GE, \quad (10.2)$$

where G is a positive constant called the elastic modulus that has the same units as the stress. We can think of this as the first entry in a Taylor expansion around zero strain, valid at small strains. Or we can think of it as being empirical because, for most elastic materials, the measured relationship between stress and strain is linear for small strains. A linear relationship between stress and strain is known as Hooke's law.

Exercise 10.1 on elastic deformation

Plot the strain as a function of time $E(t)$ using (10.2) for each of the following stress curves.

1. $T = 0$ for $t < 0$ and then $T = \frac{G}{100}$ for $t \geq 0$.
2. $T = \max\left(0, \frac{G}{1000\text{ s}}t\right)$ for $t < 1$ s followed by $T = G/1000$.
3. $T = \frac{G}{100} \sin(2\pi t/t_p)$ for all time, with $t_p = 5$ s.

10.1.2 Viscous deformation

Viscosity, as depicted schematically in figure 10.1b, can be thought of as the force associated with the elongation-rate of a dashpot. The elongation rate is the strain rate $\dot{E}$, which we can compute as the derivative of (10.1) with respect to time. Since ℓ_0 is a constant,

$$\dot{E} = \frac{\partial E}{\partial t} = \frac{1}{\ell_0} \frac{\partial \ell}{\partial t}. \quad (10.3)$$

Note that the strain rate has dimensions of inverse time! These come from taking a derivative with respect to time of the dimensionless strain.

Now we can write a rheological law in general terms as $T = g(\dot{E})$, where g is some scalar function. As with elasticity, we can choose the simplest law that satisfies basic expectations,

$$T = \eta \dot{E}, \quad (10.4)$$

where η is a positive constant called the dynamic viscosity that has the dimensions of stress $\times$ time. A linear relationship between stress and strain rate is known as a Newtonian viscous law.

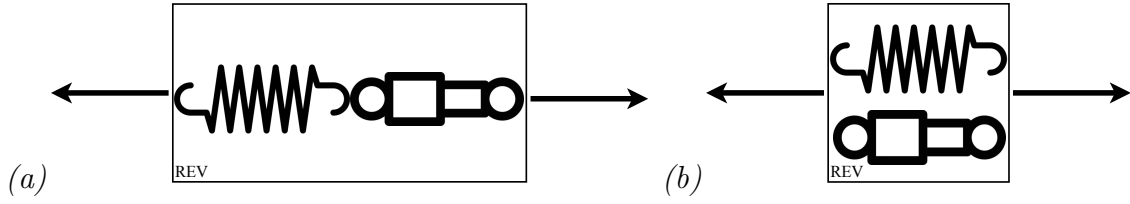


Figure 10.2: Representative elementary volumes depicting rheological models of viscoelasticity. (a) A Maxwell viscoelastic element. The spring and dashpot have identical stress. (b) A Kelvin viscoelastic element. The spring and the dashpot have identical strain.

Exercise 10.2 on viscous deformation

For each of the following stress curves, plot the strain rate as a function of time $\dot{E}(t)$ using the Newtonian viscous law (10.4). Assume a viscosity $\eta = 1$ Pa-s and a constant $t_p = 5$ s.

1. $T = 0$ for $t < 0$ and then $T = 10$ Pa for $t \geq 0$.
2. $T = 10 \max(0, t/t_p)$ Pa for $t/t_p < 1$ followed by $T = 10$ Pa.
3. $T = 10 \sin(2\pi t/t_p)$ Pa for all time.

Also, calculate and plot the strain. What assumption do you need to make to do this?

10.1.3 Viscoelastic deformation

Nearly all materials can respond both viscously and elastically when stresses are applied on different timescales. Silicate rocks, for example, can transmit elastic waves and also undergo irreversible viscous creep.

Maxwell viscoelasticity The simplest rheological models of viscoelastic deformation are illustrated in figure 10.2. We'll only consider the Maxwell model, in panel (a). In the Maxwell model, both elements feel the same stress but undergo different amounts of strain. The strain of the REV is the sum of that of the two elements. This is written

$$E = E_{\text{el}} + E_{\text{vi}}, \quad (\text{Maxwell viscoelasticity}), \quad (10.5)$$

where the subscripts indicate viscous and elastic parts of the deformation. Hence, the strain rate is the sum of the individual strain rates,

$$\dot{E} = \dot{E}_{\text{el}} + \dot{E}_{\text{vi}}. \quad (10.6)$$

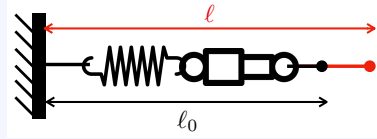
Using equations (10.2) and (10.4) we can re-write equation (10.5) as

$$\dot{E} = \frac{\dot{T}}{G} + \frac{T}{\eta}. \quad (10.7)$$

Note the dot indicating the time derivative on the stress. This is the Maxwell viscoelastic constitutive law.

Exercise 10.3 on the Maxwell time

Consider a Maxwell body that, at $t = 0$, has just undergone a change of length from ℓ_0 to ℓ , incurring a strain of E_0 , as shown below.



For $t \geq 0$, the strain rate is zero but the stress evolves. Use the Maxwell viscoelastic constitutive law (10.7) to show that

$$T(t) = T_0 e^{-t/t_{\text{maxwell}}},$$

and determine the values of T_0 and t_{maxwell} in terms of E_0 , G , η . Concisely explain the physical meaning of t_{maxwell} .

Kelvin viscoelasticity There is a crucial difference between the Kelvin viscoelastic formulation and the Maxwell one. In the Kelvin case, there is only one strain (rate); the stress is the sum of the two stresses on the two elements,

$$T = T_{\text{el}} + T_{\text{vi}}. \quad (\text{Kelvin viscoelasticity}) \quad (10.8)$$

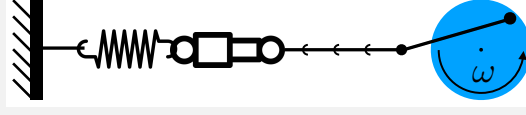
From this equation, following the procedure used above for Maxwell viscoelasticity, you can derive the Kelvin viscoelastic constitutive law.

Exercise 10.4 on Kelvin viscoelasticity

1. Derive the relationship between strain (rate) and stress that is appropriate for a Kelvin viscoelastic law. This is the Kelvin viscoelastic constitutive law.
2. Solve for the evolution of strain immediately following the instantaneous application of a constant stress, T_0 , at $t = 0$. *Hint:* First find the solution for $t \rightarrow \infty$ when the system has stopped extending. Then find the solution at $t = 0$ when the strain is zero. Then solve for the time-dependence.

Example 10.1 of the Deborah number

Consider a Maxwell body subject to a periodic strain of $E = E_0 \cos \omega t$, as shown below.



Find the stress and plot its amplitude as a function of frequency.

Solving this problem is facilitated by complex numbers. We write the strain as $E(t) = E_0 e^{i\omega t}$, with the understanding that only the real part represents the physical strain. Since the Maxwell law is linear, we seek a stress of the same form, $T(t) = T_0 e^{i\omega t}$. Substitution into equation (10.7) gives

$$i\omega E_0 = \left(\frac{i\omega}{G} + \frac{1}{\eta} \right) T_0.$$

Defining

$$t_{\text{maxwell}} = \frac{\eta}{G} \quad \text{and} \quad \text{De} = \omega t_{\text{maxwell}} = \frac{t_{\text{maxwell}}}{t_{\text{forcing}}},$$

where $t_{\text{forcing}} = 1/\omega$, we obtain

$$T_0 = GE_0 \frac{i\text{De}}{1 + i\text{De}}.$$

Taking the real part gives the physical stress

$$T(t) = \frac{GE_0 \text{De}}{\sqrt{1 + \text{De}^2}} \cos(\omega t + \Phi), \quad (10.9)$$

where

$$\Phi = \tan^{-1} \left(\frac{1}{\text{De}} \right).$$

The Deborah number compares the Maxwell time with the forcing timescale. Small De gives mainly viscous behaviour, while large De gives mainly elastic behaviour.

“The mountains flowed before the Lord” –Prophetess Deborah, Judges 5:5.

Exercise 10.5 on the Deborah number

Consider the example 10.1 with oscillatory strain and, consequently, oscillatory stress. Make a log–log plot of the scaled amplitude of the stress oscillation $|T|/GE_0$ as a function of the Deborah number. Explain the two regimes evident in the graph and the transition between them in terms of the forcing frequency ω . To support this, obtain asymptotic expressions for the stress amplitude when $\text{De} \gg 1$ and when $\text{De} \ll 1$.

Exercise 10.6 on the Deborah number in geoscience

For the following processes, infer a reasonable timescale of stress changes and, using relevant material properties, estimate the Deborah number.

1. Seismic waves in the mantle;
2. Tidal forces in the mantle;
3. Post-glacial rebound of the crust;
4. Post-glacial rebound of the mantle;
5. Mantle convection.

Given your results, state which processes are dominantly elastic, which are dominantly viscous, and which should be considered viscoelastic.

10.2 Strain and strain-rate tensors

We have already discussed, in section 8.5, the tensorial nature of the stress within a continuum. And yet the constitutive laws that we discussed in section 10.1 applied to scalars. To derive tensorial constitutive laws, we need a full description of the deformation of a continuum. Obtaining that is the aim of this section.

Consider a brick-shaped object resting, perfectly still, in outer space. Imagine a Cartesian coordinate system with origin at a corner of the brick, aligned with the edges. Now consider the vector displacement field $\mathbf{u}$ that, in the current resting state of the brick, is zero everywhere in the brick.

An astronaut approaches the brick-shaped object and wishes to change the displacement field away from zero. What are the ways she could do this? One way would simply be to displace the whole brick by a constant distance. This would give a uniform displacement field within the brick. A second way would be to rotate the brick around its central line by a constant angle. In this case, the displacement field would be non-uniform, but would be given by a simple, solid-body rotation. A third way would be to stretch the brick along its central line, and a fourth way would be to shear it across this line. These latter two would create non-uniform displacement fields and overall shape change. Importantly, they would require the astronaut to use both hands. Why? Because approaches 3 and 4 generate internal stresses within the brick.

Notice that we haven't specified the material (or rheological model) of the brick-like object. It could be elastic, viscous, or something else. In any case, shape change and gradients in the displacement (rate) field (that are not associated with solid-body rotation) are associated with internal stresses.

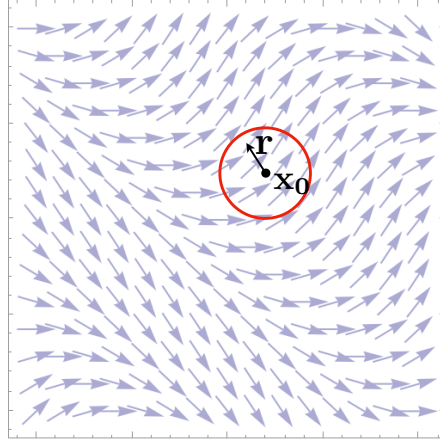


Figure 10.3: A velocity field containing the position $\mathbf{x}_0$. Around $\mathbf{x}_0$ we identify a small neighbourhood in which the position vector $\mathbf{r} = \mathbf{x} - \mathbf{x}_0$ is used to identify points nearby to $\mathbf{x}_0$.

10.2.1 Local analysis of the velocity field

Recognising that stresses arise from gradients in displacement (rate), we now investigate the *kinematics*¹³ of a continuum in the neighbourhood of a point. Below we will focus on the kinematics of the incompressible velocity field $\mathbf{v}$, but the same applies to an incompressible displacement field $\mathbf{u}$, which becomes the velocity by a time derivative. (Our approach follows GW.)

The word ‘neighbourhood’ is a hint about our technique: Taylor expansion. As shown in figure 10.3, we envision a continuum with some complex velocity field $\mathbf{v}(\mathbf{x})$ and select an arbitrary point in that continuum $\mathbf{x}_0$. The neighbourhood of $\mathbf{x}_0$ is the small region within the red circle. We define a local position vector $\mathbf{r}$ that is related to the global coordinate system as $\mathbf{r} = \mathbf{x} - \mathbf{x}_0$. Now we Taylor expand the velocity field,

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}(\mathbf{x}_0) + \nabla \mathbf{v}|_{\mathbf{x}_0} \cdot \mathbf{r} + \mathcal{O}(|\mathbf{r}|^2), \quad (10.10)$$

where we haven’t bothered to write out the higher-order terms. The reason we haven’t bothered is that we’ll immediately neglect them, under the assumption that $|\mathbf{r}|$ is very small relative to the scale of variation of the velocity field. Then, defining $\mathbf{v}_0 = \mathbf{v}(\mathbf{x}_0)$, we can write

$$\mathbf{v}(\mathbf{x}) \approx \mathbf{v}_0 + \dot{\mathbf{D}}|_{\mathbf{x}_0} \cdot \mathbf{r} \quad (\text{for positions } \mathbf{x} \text{ close to } \mathbf{x}_0), \quad (10.11)$$

where

$$\dot{\mathbf{D}} = \nabla \mathbf{v} \quad \text{or} \quad \dot{D}_{ik} = \frac{\partial v_i}{\partial x_k} \quad (10.12)$$

is the *velocity-gradient tensor*. We previously discussed this tensor field in section 6.5.

¹³Kinematics is the study of motions without forces (in contrast to dynamics).

Displacement It is important to recognise that all of the analysis above in section 10.2.1 concerning the velocity field can apply equally to the displacement field. Indeed, it is straightforward to show that we can expand the displacement $\mathbf{u}(\mathbf{x})$ as

$$\mathbf{u}(\mathbf{x}) \approx \mathbf{u}_0 + \mathbf{D}|_{\mathbf{x}_0} \cdot \mathbf{r} \quad (\text{for positions } \mathbf{x} \text{ close to } \mathbf{x}_0), \quad (10.13)$$

where we have defined the *displacement gradient tensor* as

$$\mathbf{D} = \nabla \mathbf{u} \quad \text{or} \quad D_{ik} = \frac{\partial u_i}{\partial x_k}. \quad (10.14)$$

In what follows, we focus on the velocity vector and velocity gradient tensor. Please note, however, that all of our results would hold for the displacement.

10.2.2 Strain and rotation rates

Now, based on our discussion above, we might anticipate that the velocity gradient tensor $\dot{\mathbf{D}}$ is related to the stress by a constitutive law. This stress would then explain why there is a difference in velocity between neighbouring points $\mathbf{x}_0$ and $\mathbf{x}_0 + \mathbf{r}$. This is almost true! But what is missing so far is the possibility of rotation. We next decompose $\dot{\mathbf{D}}$ to tease out the contribution from rotation.

It is evident from (10.12) that $\dot{\mathbf{D}}$ is a tensor. We can therefore decompose it into symmetric and antisymmetric tensors, as we did in section 6.4. We do this in index notation for convenience,

$$\begin{aligned} \dot{D}_{jk} &= \frac{\partial v_j}{\partial x_k}, \\ &= \underbrace{\frac{1}{2} \left(\frac{\partial v_j}{\partial x_k} + \frac{\partial v_k}{\partial x_j} \right)}_{\dot{\mathbf{E}}} + \underbrace{\frac{1}{2} \left(\frac{\partial v_j}{\partial x_k} - \frac{\partial v_k}{\partial x_j} \right)}_{\dot{\mathbf{\Theta}}}, \\ &= \frac{1}{2} [\nabla \mathbf{v} + (\nabla \mathbf{v})^\top] + \frac{1}{2} [\nabla \mathbf{v} - (\nabla \mathbf{v})^\top], \\ &= \dot{\mathbf{E}} + \dot{\mathbf{\Theta}}. \end{aligned} \quad (10.15)$$

And therefore equation (10.11) becomes

$$\mathbf{v}(\mathbf{x}) \approx \mathbf{v}_0 + \dot{\mathbf{E}}|_{\mathbf{x}_0} \cdot \mathbf{r} + \dot{\mathbf{\Theta}}|_{\mathbf{x}_0} \cdot \mathbf{r} \quad (\text{for positions } \mathbf{x} \text{ close to } \mathbf{x}_0), \quad (10.16)$$

Now, our choice of tensor symbols in the decomposition of $\dot{\mathbf{D}}$ indicates that the first part is the strain rate, which suggests that $\dot{\mathbf{\Theta}}$ represents rotation rate. But let's look at this in more detail.

Rotation rate Our purpose here is to demonstrate that $\dot{\mathbf{\Theta}}$ represents rotation. We're going to need an identity called the vector triple product. In vector notation this reads $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}$. It is more convenient, however, to use index notation. Considering the j^{th} component,

$$\begin{aligned} [(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}]_j &= a_k c_k b_j - b_k c_k a_j, \\ &= c_k (a_k b_j - a_j b_k). \end{aligned}$$

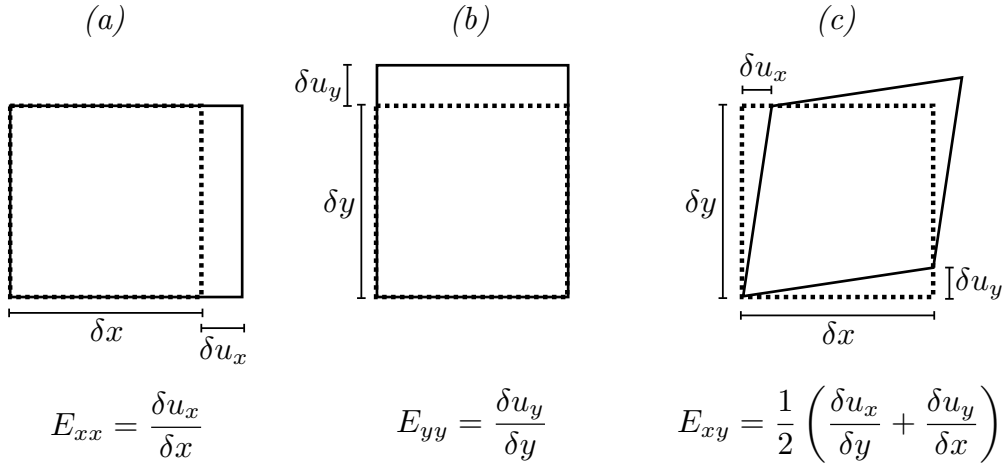


Figure 10.4: Geometrical meaning of the strain tensor. The dotted square is in the reference state. (a) Normal strain in the x direction. (b) Normal strain in the y direction. (c) Shear strain without rotation. All strains are correctly defined in the limit of the square shrinking to a point. Strain rates are given by replacing the displacement with the velocity. Source: LM (adapted).

This identity becomes useful when we apply it to a special triple product that is chosen with foresight. This is

$$[(\nabla \times \mathbf{v}) \times \mathbf{r}]_j = r_k \left(\frac{\partial v_j}{\partial x_k} - \frac{\partial v_k}{\partial x_j} \right).$$

Examining the right-hand side, we can see that this is equal to $2r_k \dot{\Theta}_{jk}$. Therefore we can write

$$\dot{\Theta} \cdot \mathbf{r} = \frac{1}{2} \boldsymbol{\omega} \times \mathbf{r},$$

where $\boldsymbol{\omega} = \nabla \times \mathbf{v}$ is the *vorticity* of the flow. This shows us that a velocity field given by $\dot{\Theta} \cdot \mathbf{r}$ is a solid-body rotation with angular velocity $\boldsymbol{\omega}/2$. With this understanding, we can infer that $\dot{\mathbf{E}}$ represents the part of $\dot{\mathbf{D}}$ that describes deformation of the continuum.

Strain rate Let's look carefully at $\dot{\mathbf{E}}$ to better understand its meaning. For simplicity, consider the matrix components in a 2-D system,

$$\dot{\mathbf{E}} = \frac{1}{2} \begin{pmatrix} \frac{\partial v_j}{\partial x_k} + \frac{\partial v_k}{\partial x_j} \\ \frac{1}{2} \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \end{pmatrix} = \begin{pmatrix} \frac{\partial v_x}{\partial x} & \frac{1}{2} \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \\ \frac{1}{2} \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) & \frac{\partial v_y}{\partial y} \end{pmatrix}.$$

Each of the entries in the tensor has a geometric meaning. Figure 10.4 plots these in terms of the strain E_{ij} rather than the strain rate, but the geometric interpretation is the same. Now notice that

$$\text{trace}(\dot{\mathbf{E}}) = \nabla \cdot \mathbf{v}$$

and, by our earlier assumption, the divergence of the velocity is zero.

Recall from linear algebra that *any real, symmetric matrix has a set of orthogonal eigenvectors and corresponding eigenvalues*. In a coordinate system aligned with those eigenvectors, the matrix is diagonalised. Also recall that *the trace of a tensor (matrix) is invariant to changes in coordinates*, so we still have $\text{trace}(\dot{\mathbf{E}}) = \nabla \cdot \mathbf{v} = 0$. This means our 2-D matrix must be

$$\dot{\mathbf{E}} = \begin{pmatrix} \dot{E}_0 & 0 \\ 0 & -\dot{E}_0 \end{pmatrix}.$$

Then, in our rotated coordinate system, we have

$$\frac{\partial v_{x'}}{\partial x'} = \dot{E}_0 \quad \text{and} \quad \frac{\partial v_{y'}}{\partial y'} = -\dot{E}_0$$

and taking the velocity to be zero at the origin fixes the integration constants. Integrating and using this condition we find that the local velocity field is $\dot{E}_0(x', -y')$. This local pattern of flow is shown in figures 3.3 and 10.5; it is called *pure shear*. Pure shear is a fundamental ‘ingredient’ of all flows, even those of much greater complexity.

Imagine injecting a square-shaped region of ink surrounding the origin in a pure-shear flow and watching it as the ink is carried by the flow. What happens? Does it translate, rotate, or deform (i.e., change its shape)? The square deforms. This informs us that the tensor $\dot{\mathbf{E}}$ describes the part of $\nabla \mathbf{v}$ that is associated with deformation or strain. Hence we call $\dot{\mathbf{E}}$ the *strain-rate tensor*.

Components of the local flow Using the above analysis to interpret equation (10.16)

$$\mathbf{v}(\mathbf{x}) \approx \underbrace{\mathbf{v}_0}_{\text{translation}} + \underbrace{\dot{\mathbf{E}}|_{\mathbf{x}_0} \cdot \mathbf{r}}_{\text{deformation}} + \underbrace{\frac{1}{2}\boldsymbol{\omega}|_{\mathbf{x}_0} \times \mathbf{r}}_{\text{rotation}} \quad (\text{for positions } \mathbf{x} \text{ close to } \mathbf{x}_0), \quad (10.17)$$

These three contributions are illustrated in figure 10.5.

Now we can come to a fundamental conclusion! The only part of the velocity field that is associated with stress is the deformation, and this part is given by $\dot{\mathbf{E}}$. Therefore, the tensor form of the viscous constitutive law must be of the form $\mathbf{T} = g(\dot{\mathbf{E}})$. Following the same logic, the elastic constitutive law must be of the form $\mathbf{T} = f(\mathbf{E})$. These relations exclude the translational and rotational parts of the flow!

10.3 Elastic and viscous rheological models (tensor version)

In section 10.2.2, we obtained an expression for the deformational part of the velocity gradient tensor $\nabla \mathbf{v}$, which we called the strain-rate tensor $\dot{\mathbf{E}}$. In the case of elasticity, the corresponding tensors are the displacement-gradient tensor $\nabla \mathbf{u}$ and the strain tensor $\mathbf{E}$. We now relate the strain (rate) tensor to the stress tensor that we examined in section 8.5. We follow closely the assumptions and form of the scalar laws in section 10.1.

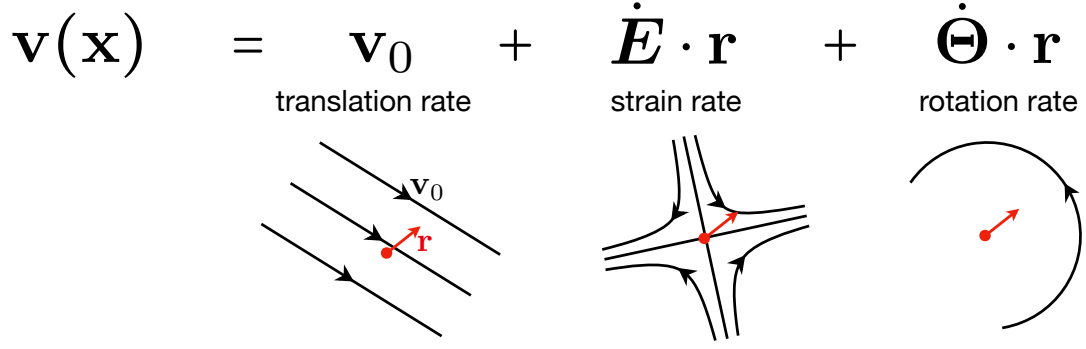


Figure 10.5: Three contributions to the velocity field in the neighbourhood of a point $\mathbf{x}_0$: translation $\mathbf{v}_0$, strain rate $\dot{\mathbf{E}}$, and rotation rate $\dot{\boldsymbol{\Theta}}$. Shown in two dimensions and without density change for simplicity. Source: GW (adapted).

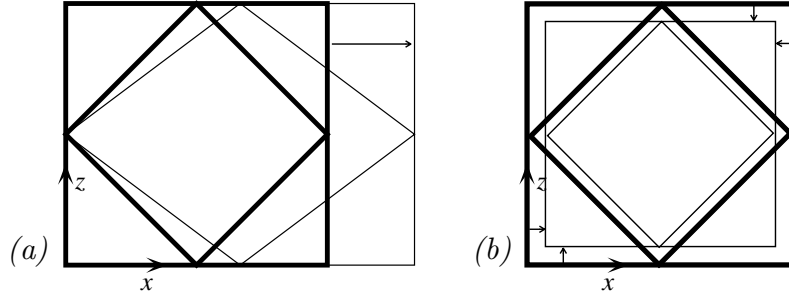


Figure 10.6: Uniaxial versus isotropic strain. (a) Uniaxial strain with $E_{11} > 0$ and all other entries in $\mathbf{E}$ equal to zero. (b) Isotropic strain with zero deviatoric part and $\boldsymbol{\nabla} \cdot \mathbf{u} < 0$. Source: PS (adapted).

10.3.1 The Hookean elastic solid

This is the most commonly used model of elastic deformation. It is also known as the ideal solid model.

An elastic medium has, in general, two modes of deformation. In the first, compression or dilation of the material leads to small changes in density (allowing, for example, sound waves). In the second mode, the material shears without a change in density.

It is important to recognise that a general strain separates into deviatoric and volumetric parts; even a diagonal uniaxial strain contains both parts. Figure 10.6 illustrates a uniaxial and an isotropic strain associated with the tensors

$$\mathbf{E}_{\text{uniaxial}} = \begin{pmatrix} E_a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{E}_{\text{isotropic}} = \begin{pmatrix} -E_b & 0 & 0 \\ 0 & -E_b & 0 \\ 0 & 0 & -E_b \end{pmatrix}.$$

We see from the embedded diamond shape in panel (a) that a uniaxial extension causes internal angles to change, which indicates shear has occurred. In contrast, panel (b) shows an isotropic (compressive) strain. This isotropic strain is the part that is measured by $\boldsymbol{\nabla} \cdot \mathbf{u}$, and hence doesn't enter into the deviatoric tensor.

The resistance to each of these two modes may be different and so it is necessary

to separate them. To this end, we decompose the strain tensor into an incompressible part (with zero trace) and a diagonal part. This decomposition was introduced in section 6.4. We define the *deviatoric strain tensor* as

$$\boldsymbol{\epsilon} = \boldsymbol{E} - \frac{1}{3}\text{trace}(\boldsymbol{E})\boldsymbol{I}.$$

We call this the deviatoric strain because it has zero trace.

Exercise 10.7 on the deviatoric strain tensor

Show that $\text{trace}(\boldsymbol{E}) = \nabla \cdot \mathbf{u}$. Then show that $\text{trace}(\boldsymbol{\epsilon}) = 0$.

Now the two modes of elastic deformation are separated and represented by $\boldsymbol{\epsilon}$ (shear) and $\nabla \cdot \mathbf{u}$ (expansion, equal to $\text{trace}(\boldsymbol{E})$). Using the same reasoning as in section 10.1 on scalar rheological models, we can independently relate these to the deviatoric and isotropic stress (recall that $\boldsymbol{T} = -p\boldsymbol{I} + \boldsymbol{\tau}$) using proportionality constants,

$$\boldsymbol{\tau} = 2G\boldsymbol{\epsilon} \quad \text{and} \quad p = -K\nabla \cdot \mathbf{u}.$$

Multiplying the isotropic equation by $\boldsymbol{I}$ and adding them together gives

$$\boldsymbol{T} = 2G\boldsymbol{\epsilon} + K(\nabla \cdot \mathbf{u})\boldsymbol{I} \quad \text{or} \quad T_{ij} = 2G\epsilon_{ij} + K\frac{\partial u_k}{\partial x_k}\delta_{ij}. \quad (10.18)$$

In this constitutive law we have introduced the shear modulus G and the bulk modulus K . These typically are expressed with units of Pascals.

This simple form with only two elastic constants is valid for *isotropic* materials. These are materials that have the same elastic properties independent of the direction of the applied stress. For anisotropic materials the stress and strain are related by a fourth-order tensor with more than two distinct entries,

$$T_{ij} = C_{ijkl}E_{kl}. \quad (10.19)$$

Although this tensor has $3^4 = 81$ entries, only 21 of them are distinct under fully general anisotropy. We will not consider elastic anisotropy here.

10.3.2 The incompressible Newtonian fluid

This is the most common rheological model in fluid dynamics, and most widely relevant to materials. Since an incompressible flow has $\text{trace}(\dot{\boldsymbol{E}}) = \nabla \cdot \mathbf{v} = 0$, it is true that $\dot{\boldsymbol{E}} = \dot{\boldsymbol{\epsilon}}$. This means that there is only one mode of deformation for incompressible flows: shear.

For a Newtonian fluid, the deviatoric stress $\boldsymbol{\tau}$ is linearly related to the strain rate by the viscosity, and hence we assert the proportionality

$$\boldsymbol{\tau} = 2\eta\dot{\boldsymbol{\epsilon}}.$$

Writing this in terms of the full stress tensor gives

$$\boldsymbol{T} = -p\boldsymbol{I} + 2\eta\dot{\boldsymbol{\epsilon}} \quad \text{or} \quad T_{ij} = -p\delta_{ij} + 2\eta\dot{\epsilon}_{ij}. \quad (10.20)$$

In this viscous constitutive law the constant of proportionality is the viscosity η , typically expressed with units of Pascal-seconds.

10.3.3 The upper-convected Maxwell viscoelastic model

For (a feeling of) completeness, we introduce a rather advanced rheological model. This constitutive law extends the Maxwell viscoelastic formulation of equation (10.7) to a tensor form, assuming incompressibility of the material. It reads

$$\frac{\boldsymbol{\tau}}{2\eta} + \frac{\overset{\nabla}{\boldsymbol{\tau}}}{2G} = \dot{\boldsymbol{\epsilon}}. \quad (10.21)$$

This looks straightforward, but there's a twist. The $\overset{\nabla}{\boldsymbol{\tau}}$ represents an upper-convected time derivative, also called an Oldroyd derivative, which is essentially a Lagrangian derivative for tensors. It represents the translation, rotation and deformation of material elements within the flow. You can learn more about it [elsewhere](#).

10.4 Principal components of stress and strain (rate)

We have learned that the strain (rate), and hence the stress, are real and symmetric tensors. From linear algebra we know that real, symmetric tensors have two important properties. First, that their eigenvalues are always real, and second, that their eigenvectors are mutually perpendicular. A set of mutually perpendicular vectors forms an orthogonal basis for a coordinate system. Therefore, any real, symmetric tensor has an associated coordinate system built from its eigenvectors. These are the *principal axes* of the tensor.

Consider the tensor field $\boldsymbol{T}$ representing the stress throughout a continuum. Consider also a vector $\mathbf{n}$ that represents the outward normal to a specified surface at a point $\mathbf{x}$ in the continuum. We say that $\mathbf{n}$ is an eigenvector if

$$\boldsymbol{T} \cdot \mathbf{n} = \lambda \mathbf{n}, \quad (10.22)$$

where λ is a scalar. This equation states that for the special vector $\mathbf{n}$, the effect of contracting with the tensor $\boldsymbol{T}$ is equivalent to multiplying by a scalar (in fact, *eigen* means special in German). Recall that $\mathbf{t} = \boldsymbol{T} \cdot \mathbf{n}$ is the traction on the surface specified by $\mathbf{n}$. If $\mathbf{n}$ is an eigenvector, then this traction is entirely in the direction of $\mathbf{n}$; it is exactly normal to the surface. For this special direction, the action of tensor $\boldsymbol{T}$ is simply to scale the eigenvector by a factor λ , which is the eigenvalue.

It is instructive to write equation (10.22) in matrix–vector form, assuming a two-dimensional system for simplicity,

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \lambda \quad (10.23)$$

Now recall that our 2nd-rank tensor, being real and symmetric, has two eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ and two eigenvectors $\mathbf{n}^{(1)}$ and $\mathbf{n}^{(2)}$. We can write a single equation representing both of these cases as

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} n_1^{(1)} & n_1^{(2)} \\ n_2^{(1)} & n_2^{(2)} \end{bmatrix} = \begin{bmatrix} n_1^{(1)} & n_1^{(2)} \\ n_2^{(1)} & n_2^{(2)} \end{bmatrix} \begin{bmatrix} \lambda^{(1)} & 0 \\ 0 & \lambda^{(2)} \end{bmatrix}. \quad (10.24)$$

Going back to tensor form, we write this equation as $\mathbf{T}\mathbf{N} = \mathbf{N}\mathbf{\Lambda}$ where the definitions of tensors $\mathbf{N}$ and $\mathbf{\Lambda}$ can be inferred from a comparison with (10.24). Multiplying both sides by $\mathbf{N}^{-1}$ we have

$$\mathbf{\Lambda} = \mathbf{N}^{-1}\mathbf{T}\mathbf{N}. \quad (10.25)$$

This equation states that the stress tensor $\mathbf{T}$ takes a diagonal form $\mathbf{\Lambda}$ when it is transformed into the coordinate system defined by the eigenvectors. The entries in $\mathbf{\Lambda}$ are the eigenvalues $\lambda^{(j)}$.

For a three-dimensional, symmetric stress tensor, we can choose three mutually perpendicular, unit-length eigenvectors; these define the principal directions. Finding these vectors requires first solving for the eigenvalues¹⁴. When forming the diagonal matrix $\mathbf{\Lambda}$ of eigenvalues, the entries are listed in descending order. For the stress tensor, the entries are called the *principal stresses*. Since we have chosen a tension-positive sign convention for stress, the first principal stress is the maximum tension and the third is the minimum tension (maximum compression). These principal stresses act as normal stresses along the principal axes of the tensor. This is illustrated for a simple shear flow in example 10.2.

¹⁴If there are three distinct eigenvalues, then it is straightforward to find the orthogonal basis of eigenvectors, as in example 10.2. If one or more of the eigenvalues is repeated, the solution is more subtle. See any introductory textbook about Linear Algebra.

Example 10.2 of principal axes for simple shear

Consider the flow known as simple shear,

$$\mathbf{v}(\mathbf{x}) = \dot{\gamma}y\hat{\mathbf{x}}, \quad (10.26)$$

where $\dot{\gamma}$ is a constant representing the strain rate. This is a parallel flow with strain-rate tensor

$$\dot{\mathbf{E}} = \begin{bmatrix} 0 & \dot{\gamma}/2 \\ \dot{\gamma}/2 & 0 \end{bmatrix}.$$

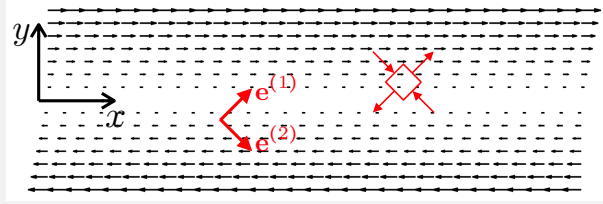
Note that these are independent of position (why?). We find the eigenvalues by forming the characteristic equation, $\det(\dot{\mathbf{E}} - \lambda\mathbf{I}) = 0$, which leads to $\lambda^2 - (\dot{\gamma}/2)^2 = 0$ and hence to the eigenvalues

$$\lambda^{(1)} = \dot{\gamma}/2, \quad \lambda^{(2)} = -\dot{\gamma}/2.$$

To find the eigenvectors we solve $(\dot{\mathbf{E}} - \lambda^{(j)}\mathbf{I})\mathbf{e}^{(j)} = 0$. For $j = 1$ and $j = 2$ we find

$$\mathbf{e}^{(1)} = (1, 1)/\sqrt{2}, \quad \mathbf{e}^{(2)} = (1, -1)/\sqrt{2},$$

where the factor of $\sqrt{2}$ has been included to give the vectors unit length. The flow and eigenvectors look like this:



In a coordinate system where $\hat{\mathbf{x}}' = \mathbf{e}^{(1)}$ and $\hat{\mathbf{y}}' = \mathbf{e}^{(2)}$ we have a diagonal strain rate tensor

$$\dot{\mathbf{E}}' = \begin{bmatrix} \dot{\gamma}/2 & 0 \\ 0 & -\dot{\gamma}/2 \end{bmatrix},$$

so $\mathbf{e}^{(1)}$ is the stretching direction and $\mathbf{e}^{(2)}$ is the squeezing direction. The stress tensor in this coordinate system is $\mathbf{T}' = 2\eta\dot{\mathbf{E}}'$. The principal stresses are shown as arrows around the red box, above, which has faces aligned with normals along principal axes.

Exercise 10.8 on rotation in simple shear

Consider the simple-shear flow given by equation (10.26) and compute the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{v}$. Comment on the rate of rotation in the flow. How does a material ring (a circle of red dye, for example), evolve in the flow over time?

Summary of §10

- Normal strain is the change in length relative to the neutral length. It is dimensionless. Normal strain rate is the rate of change of strain. It has units of inverse time.
- In elasticity, strain and stress are linearly proportional. In viscous flow, strain rate and stress are linearly proportional.
- Viscoelasticity combines the viscous and elastic rheological laws. For the Maxwell model of viscoelasticity, strain rate is the sum of the viscous and elastic rates, in response to the same stress.
- The Maxwell time is the timescale on which elastic stress is viscously relaxed after a sudden strain is imposed. The ratio of the Maxwell time to the timescale of forced changes in stress is the Deborah number.
- The displacement-gradient tensor is $\nabla \mathbf{u}$ and the velocity-gradient tensor is $\nabla \mathbf{v}$. These tensors carry information about the deformation and the rotation of the continuum.
- Deformation is described by the symmetric part of the displacement-gradient or velocity-gradient tensor. The strain and strain-rate tensors are

$$\mathbf{E} = \frac{1}{2} [\nabla \mathbf{u} + (\nabla \mathbf{u})^T] \quad \text{and} \quad \dot{\mathbf{E}} = \frac{1}{2} [\nabla \mathbf{v} + (\nabla \mathbf{v})^T].$$

- Solid-body rotation is described by the antisymmetric part of the displacement (rate)-gradient tensor. The rotation and rotation-rate tensors are

$$\mathbf{\Theta} = \frac{1}{2} [\nabla \mathbf{u} - (\nabla \mathbf{u})^T] \quad \text{and} \quad \dot{\mathbf{\Theta}} = \frac{1}{2} [\nabla \mathbf{v} - (\nabla \mathbf{v})^T].$$

- Deviatoric strain (rate) is defined as

$$\boldsymbol{\epsilon} = \mathbf{E} - \frac{1}{3} \text{trace}(\mathbf{E}) \mathbf{I} \quad \text{and} \quad \dot{\boldsymbol{\epsilon}} = \dot{\mathbf{E}} - \frac{1}{3} \text{trace}(\dot{\mathbf{E}}) \mathbf{I}.$$

For an incompressible continuum, the deviatoric strain (rate) tensor is equal to the strain (rate) tensor.

- The tensorial rheological model of a Hookean elastic solid is

$$\mathbf{T} = 2G\boldsymbol{\epsilon} + K(\nabla \cdot \mathbf{u})\mathbf{I}.$$

- The tensorial rheological model of a Newtonian viscous fluid is

$$\mathbf{T} = -p\mathbf{I} + 2\eta\dot{\boldsymbol{\epsilon}}.$$

- Stress and strain (rate) tensors can be diagonalised by finding their eigenvalues and eigenvectors.

11 Momentum Conservation

Before moving ahead, let's review where we are. In chapter 8, we derived a continuum version of Newton's second law. In the continuum version, the equation applies to an element (in fact, each element) of the continuum. The relevant forces are surface forces (from neighbouring elements), the body force (from gravity), and the inertial force, which is really the mass $\times$ acceleration part. Each of these appears as force per unit volume. Surface forces are expressed as the divergence of the stress tensor.

Also in chapter 8 we investigated the meaning of the stress tensor in terms of forces per unit area on orthogonal faces of a fictitious representative elementary volume of the fluid. We recognised normal and shear stresses. We noted a sign convention of the stress tensor: that tension is positive, independent of the coordinate directions. And we discussed the decomposition of the stress tensor into deviatoric and isotropic parts.

However, at the end of section 8.6, we recognised that to solve for the motion of the continuum (displacement or velocity), we need to be able to write the stress $\mathbf{T}$ in terms of the motion of the continuum. After a necessary aside in chapter 9 to derive equations for conservation of mass, we set out in chapter 10 to derive constitutive laws relating stress to motion. What we found was that stress is related to strain for an elastic medium or to strain rate for a viscous medium. We derived tensors that isolate the strain (rate). Then, in section 10.3, we finally obtained the needed equations relating the stress $\mathbf{T}$ to the motion.

Now it is time to combine the Cauchy stress equation (8.13) with the elastic and viscous constitutive laws, (10.18) and (10.20). For convenience, we repeat the Cauchy stress equation,

$$\nabla \cdot \mathbf{T} + \rho \mathbf{g} = \rho \frac{D\mathbf{v}}{Dt}, \quad (11.1)$$

noting that for elasticity, assuming small strains, we can replace the Lagrangian derivative with $\partial^2 \mathbf{u} / \partial t^2$.

11.1 Elastic deformation

We recall the tensor version of Hooke's law of linear elasticity,

$$\mathbf{T} = 2G\boldsymbol{\epsilon} + K(\nabla \cdot \mathbf{u})\mathbf{I},$$

with bulk modulus K and shear modulus G , which we shall take to be constant and uniform. Substituting this into (11.1) gives

$$2G\nabla \cdot \boldsymbol{\epsilon} + K\nabla(\nabla \cdot \mathbf{u}) + \rho \mathbf{g} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2}. \quad (11.2)$$

Exercise 11.1 on divergence of a diagonal tensor

Use index notation to show that $\nabla \cdot (Q\mathbf{I}) = \nabla Q$, for some scalar field $Q(\mathbf{x})$.

We can simplify the first term of this equation by re-writing $\nabla \cdot \boldsymbol{\epsilon} = \frac{1}{6}\nabla(\nabla \cdot \mathbf{u}) + \frac{1}{2}\nabla^2 \mathbf{u}$. Then (11.2) becomes

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \rho \mathbf{g} + G \nabla^2 \mathbf{u} + \left(K + \frac{1}{3}G\right) \nabla(\nabla \cdot \mathbf{u}), \quad (11.3)$$

where we have also rearranged the order of the terms to put the equation into a standard form. The vector equation (11.3) is known as the *Navier-Cauchy* equation of linear, isotropic *elastodynamics*, with extremely broad uses. It can be solved numerically or analytically, but the latter is facilitated by making further manipulations.

Exercise 11.2 on divergence of the deviatoric strain

Show that

$$\nabla \cdot \boldsymbol{\epsilon} = \frac{1}{6}\nabla(\nabla \cdot \mathbf{u}) + \frac{1}{2}\nabla^2 \mathbf{u}$$

by combining the definition of the deviatoric strain tensor $\boldsymbol{\epsilon} = \mathbf{E} - \frac{1}{3}\text{trace}(\mathbf{E})\mathbf{I}$ with the definition of the strain tensor $\mathbf{E} = \frac{1}{2}[\nabla \mathbf{u} + (\nabla \mathbf{u})^\top]$. Hint: use index notation.

For modelling the propagation of seismic waves, Equation (11.3) is sometimes written

$$\ddot{\mathbf{u}} = \frac{\lambda + 2\mu}{\rho} \nabla(\nabla \cdot \mathbf{u}) - \frac{\mu}{\rho} \nabla \times \nabla \times \mathbf{u}, \quad (11.4)$$

where $\mu = G$ and $\lambda = K - 2G/3$ are the Lamé elastic constants. The nice feature of this formulation is that it cleanly separates the volumetric strain from the shear strain. The body force has been dropped from this equation, a standard assumption in seismology.

As we will see below in section 13.1, equation (11.4) is a wave equation; actually it is two wave equations.

11.2 Viscous deformation

We recall the tensor version of Newtonian viscosity,

$$\mathbf{T} = -p\mathbf{I} + 2\eta\dot{\boldsymbol{\epsilon}},$$

where η is the shear viscosity, which we will (for now) take to be constant and uniform. Substituting the constitutive law into the Cauchy stress equation (11.1) we obtain

$$-\nabla p + 2\eta \nabla \cdot \dot{\boldsymbol{\epsilon}} + \rho \mathbf{g} = \rho \frac{D\mathbf{v}}{Dt}. \quad (11.5)$$

Simplifying the term in $\nabla \cdot \dot{\boldsymbol{\epsilon}}$ using the condition of incompressibility, $\nabla \cdot \mathbf{v} = 0$, we obtain

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho \mathbf{g}. \quad (11.6)$$

This is the *Navier-Stokes equation of fluid dynamics*. Combined with the continuity equation $\nabla \cdot \mathbf{v} = 0$, this system is the most commonly used framework for fluid mechanics. We should note, however, the assumptions that limit its applicability:

- We assumed uniform viscosity, but viscosity can vary with temperature, pressure, and the strain-rate itself. This is important in the mantle.
- We assumed that gravity provides the sole body force. There can be others, such as magnetism. This is important in the outer core.
- We assumed an incompressible flow, but flows near the speed of sound or across large pressure gradients can be compressible. This is important in the atmosphere.

Exercise 11.3 on divergence of the deviatoric strain rate

Show that $\nabla \cdot \dot{\epsilon} = \frac{1}{2} \nabla^2 \mathbf{v}$ in the case of an incompressible flow. Hint: use index notation.

Summary of §11

- The Navier-Cauchy equation of linear, isotropic elastodynamics is

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \rho \mathbf{g} + G \nabla^2 \mathbf{u} + \left(K + \frac{1}{3}G\right) \nabla (\nabla \cdot \mathbf{u}),$$

where G is the shear modulus and K is the bulk modulus; both have SI units of Pascals. This equation is frequently written in terms of the Lamé constants, $\mu = G$ and $\lambda = K - 2G/3$.

- The Navier-Stokes equations of incompressible, Newtonian, isoviscous flow are

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho \mathbf{g},$$

$$\nabla \cdot \mathbf{v} = 0,$$

where η is the viscosity with SI units of Pascal-seconds.

12 Incompressible Flows

In this chapter we look at *solutions* $\mathbf{v}(\mathbf{x}, t)$, $p(\mathbf{x}, t)$ to the incompressible Navier-Stokes (N-S) equations,

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho \mathbf{g}, \quad (12.1a)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (12.1b)$$

Here we have expanded the Lagrangian derivative as a reminder of its contents. In particular, we emphasise that this equation is *nonlinear* due to the advection of momentum $\rho \mathbf{v} \cdot \nabla \mathbf{v}$. This makes N-S very challenging to solve, in general. Our focus will be on problems where we can neglect this term.

Exercise 12.1 on expanding the Navier-Stokes momentum equation

Note that equation (12.1a) is a vector equation, meaning that it governs each of the components of the vector $\mathbf{v}(\mathbf{x}, t)$ by a separate scalar equation. Assuming a Cartesian coordinate system, write the x , y and z components of equation (12.1a), governing v_x , v_y and v_z , respectively. *Hint:* be sure to use the fact that $\mathbf{v} \cdot \nabla$ and ∇^2 are scalar operators!

Neglecting terms is a great way to make problems easier to solve. But solutions are only relevant if they represent the physical effects that are important. How do we discover and justify if/when we can neglect a term in the equations? Scaling analysis is one way, and we discuss this in section 12.2.

But when it comes to solving partial differential equations, we must always respect the boundaries and the conditions that apply there. So in section 12.3 we consider the types of boundary conditions that are typically encountered, and in section 12.4 we examine problems that place an emphasis on the boundary conditions.

Numerical methods and codes for computational fluid dynamics (CFD) are readily accessible these days, and the standard laptop now has enough computing power to do model 3-D flows. But without an understanding of the physics and mathematics (i.e., the continuum mechanics), numerical solutions can be misleading or worse. So it is always important to keep an eye on what can be learned from simplified problems.

12.1 Static and dynamic pressure

We all know that in a gravitational field, even static fluids have stress. So there's at least a part of the pressure that is independent of the flow; this is called the static pressure. We find the static pressure by taking equation (12.1a) and setting $\mathbf{v} = \mathbf{0}$ to obtain

$$\nabla p = \rho \mathbf{g} \quad (\text{when } \mathbf{v} = \mathbf{0}). \quad (12.2)$$

This tells us that pressure increases downward through a static fluid, in the direction of gravity. If we take a coordinate system with $\hat{\mathbf{z}}$ pointing upward such that $\mathbf{g} = -g\hat{\mathbf{z}}$

and place the origin at the upper surface of the static fluid, we can integrate in z (12.2) to obtain

$$p_S(z) = p_0 - \rho g z \quad \text{for } z \leq 0,$$

where we use p_S to indicate that this is the *static pressure*.

In many non-static problems, it is useful to write the full pressure as $p = p_S + P$, where P is called the *dynamic pressure* to indicate that it is associated with motion.

Exercise 12.2 on static and dynamic pressure

Substitute $p = p_S + P$ into equation (12.1a) and use the meaning of p_S to simplify the equation.

12.2 Scaling analysis of inertia and viscosity

Scaling analysis has different meanings and uses. To many students it doesn't feel very mathematical, and there is some truth in this. It is a semi-rigorous way to let knowledge of the physics (empirical, intuitive, etc) guide and simplify the mathematics.

Scaling analysis usually begins by choosing “characteristic scales” of variables. This means that, for a particular problem that we wish to mathematically model, we identify typical values of independent and dependent variables. These should be approximate! For example, a typical length-scale of mantle convection is the depth of the mantle, about 3000 km. But as with all mathematics, when we don't know something, we just create a symbol to represent that thing in the hope that later we'll know it and can replace the symbol with a number. This is what we do next.

Consider the Navier-Stokes equation (12.1). It is easy to choose scales for things that are considered to be constant (such as density, gravity and viscosity in this context); they are their own scales. It is less easy to choose scales for other quantities: length, time, velocity, pressure. But let's suppose that we have a (yet unknown) length scale that we denote $[x]$ and a velocity scale that we denote $[v]$. We can combine these to form a time scale $[t] = [x]/[v]$. Pressure is less straightforward, so for now we just call the pressure scale $[p]$.

The next step is to define a set of dimensionless variables, denoted with primes ($'$). We use the characteristic scales to make these variables dimensionless. Hence we write

$$\mathbf{x} = [x]\mathbf{x}', \quad t = \frac{[x]}{[v]}t', \quad \mathbf{v} = [v]\mathbf{v}', \quad p = [p]p'. \quad (12.3)$$

On this basis, we also have $\nabla = [x]^{-1}\nabla'$ and $\partial/\partial t = ([v]/[x])\partial/\partial t'$. Substituting these into equations (12.1), we obtain

$$\frac{\rho[v]^2}{[x]} \left(\frac{\partial \mathbf{v}'}{\partial t'} + \mathbf{v}' \cdot \nabla' \mathbf{v}' \right) = -\frac{[p]}{[x]} \nabla' p' + \frac{\eta[v]}{[x]^2} \nabla'^2 \mathbf{v}' + \rho g \hat{\mathbf{g}}.$$

Now, multiplying through by $[x]^2/\eta[v]$ we obtain

$$\text{Re} \left(\frac{\partial \mathbf{v}'}{\partial t'} + \mathbf{v}' \cdot \nabla' \mathbf{v}' \right) = -\frac{[p][x]}{\eta[v]} \nabla' p' + \nabla'^2 \mathbf{v}' + \frac{\rho g [x]^2}{\eta[v]} \hat{\mathbf{g}},$$

where we have defined the *Reynolds number*

$$\text{Re} = \frac{\text{Inertial force}}{\text{Viscous force}} = \frac{\rho[v][x]}{\eta}. \quad (12.4)$$

This is sometimes written as $\text{Re} = [v][x]/\nu$, where the symbol ‘nu’ $\nu = \eta/\rho$ represents the *kinematic viscosity* (for reasons that are obscure at this point in the course). Now if we think that pressure variations are a consequence of viscous stresses, we could choose $[p] = \eta[v]/[x]$. Then we have

$$\text{Re} \left(\frac{\partial \mathbf{v}'}{\partial t'} + \mathbf{v}' \cdot \nabla' \mathbf{v}' \right) = -\nabla' p' + \nabla'^2 \mathbf{v}' + \frac{\rho g [x]^2}{\eta [v]} \hat{\mathbf{g}}.$$

Let’s step back and think about this equation. First, it is dimensionless; each term has no physical dimensions. The Reynolds number is a dimensionless number, as is the ratio in the body-force term. Moreover, each of the primed symbols has been rescaled by its characteristic scale. So now, all of the variables (position, time, velocity) except pressure are of $\mathcal{O}(1)$. The viscous force per unit volume term is of $\mathcal{O}(1)$ because we have rescaled the other terms with a ratio to its magnitude.

How big is the inertial force relative to the viscous force? To the extent that our choice of characteristic scales is appropriate, the Reynolds number gives us an estimate of this. In the case that $\text{Re} \ll 1$, then viscous forces are dominant and inertial forces are unimportant. We can neglect the inertial term and, returning to dimensional variables, seek solutions of

$$\nabla p = \eta \nabla^2 \mathbf{v} + \rho \mathbf{g}, \quad (12.5a)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (12.5b)$$

These are the *Stokes equations* for viscous, incompressible flow.

In the opposite case that $\text{Re} \gg 1$, then inertia is very important and viscous stresses are unimportant. We can neglect the viscous term and, returning to dimensional variables, seek solutions of

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \rho \mathbf{g}, \quad (12.6a)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (12.6b)$$

These are the *Euler equations* for inviscid, incompressible, unsteady flow. It is important to recognise here that neglecting the viscous term has reduced the spatial order of the momentum equation from second order to first order. Consequently, solutions of the Euler equations satisfy fewer boundary conditions than solutions of the Navier–Stokes equations. We discuss boundary conditions below in section 12.3.

It is crucial to note that we can *never assume that the pressure gradient is negligibly small on the basis of this scaling analysis* because pressure is needed to enforce the incompressibility constraint.

Exercise 12.3 on the Reynolds number

For each of the following flows, provide an estimate of the Reynolds number.

1. Mantle convection.
2. Flow down the river Thames near Oxford.
3. Wind as it blows through city streets.
4. Ocean circulation.
5. Flow around a single-celled, swimming organism.
6. Flow around a swimming tuna fish.

12.3 Boundary conditions

The Navier–Stokes momentum equation is second order in space because of the ∇^2 term and therefore requires boundary conditions. What are the standard types of boundary conditions that can be imposed? There are only a few.

12.3.1 Kinematic conditions

Fluid is often contained within or bounded by rigid walls; think of a tea cup, the ocean floor, or an airplane wing as examples. The wall can be described as a parameterised surface in 3-D and, at every point, it has a normal vector $\hat{\mathbf{n}}$ (pointing into the fluid) and two orthogonal tangent vectors $\hat{\mathbf{t}}_1$ and $\hat{\mathbf{t}}_2$. There are two types of conditions we can apply:

No-penetration condition This condition states the fluid cannot flow *through* the wall, and hence that

$$\mathbf{v} \cdot \hat{\mathbf{n}} = 0.$$

No-slip condition This condition states that exactly at the wall, the local fluid molecules are adhered to the wall and cannot slip along the wall. This is written as

$$\mathbf{v} \cdot \hat{\mathbf{t}}_{1,2} = 0.$$

Together, these give us the boundary condition $\mathbf{v} = \mathbf{0}$.

Above, we have assumed that the wall is not moving. If the wall is moving, we would have $\mathbf{v} = \mathbf{v}_{\text{wall}}$, where the latter is the velocity of the wall.

12.3.2 Dynamic conditions

Liquids are often bounded by air; think of tea in a cup or water in the ocean. Both water/tea and air are fluids. These boundaries are more complicated than rigid walls because they can move with no geometrical constraint (hence breaking waves, for

example). The conditions are described in terms of the *traction* on the boundary. It is simplest (and usually adequate) to think of the boundaries as having a fixed position (e.g., sea level) and use the normal and tangential vectors defined above (the normal points into the liquid of interest).

The first step is to obtain the traction vector on the boundary. We get this by contracting the stress tensor (on the boundary) with the normal vector of the boundary,

$$\mathbf{f} = \mathbf{T} \cdot \hat{\mathbf{n}},$$

where we use $\mathbf{f}$ for traction to avoid ambiguity with the tangent vector. Now we can impose conditions on the components of $\mathbf{f}$. Again, there are two categories:

Normal-traction conditions Here we are considering the force (per unit area) normal to the surface. This is usually a pressure, in which case we have

$$\hat{\mathbf{n}} \cdot \mathbf{f} = \hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}} = -p_0. \quad (12.7)$$

We need a minus sign because tension has a positive sign.

Shear-traction conditions Here we impose the force (per unit area) tangential to the surface. This is a shear stress. In 3-D, the surface has two tangent directions, but let's just consider one of them:

$$\hat{\mathbf{t}}_1 \cdot \mathbf{f} = \hat{\mathbf{t}}_1 \cdot \mathbf{T} \cdot \hat{\mathbf{n}} = -\tau_1, \quad (12.8)$$

where τ_1 is the stress imposed on the boundary (by the wind, for example). It has a minus sign because the normal vector is pointing into the fluid.

Dynamic boundary conditions are easier to understand in the context of an actual problem. We'll look at these below.

12.4 Parallel viscous flows

In this section we consider a simple class of flows that are parallel, in the sense that the velocity vectors all point in the same direction. Let's orient our coordinate system such that this is the x direction. Then, our parallel flow can be represented as

$$\mathbf{v}(\mathbf{x}, t) = v_x(\mathbf{x}, t) \hat{\mathbf{e}}_x.$$

However, we can simplify this using the continuity equation $\nabla \cdot \mathbf{v} = 0$. Since $v_y = v_z = 0$,

$$\nabla \cdot \mathbf{v} = 0 \quad \Rightarrow \quad \frac{\partial v_x}{\partial x} = 0. \quad (12.9)$$

This means that the flow is unchanging in the x direction and we can write

$$\mathbf{v}(\mathbf{x}, t) = v_x(y, z, t) \hat{\mathbf{e}}_x. \quad (\text{no } x \text{ dependence!})$$

The continuity equation says nothing, however, about how v_x changes in the other coordinate directions! For that, we need the momentum equation.

12.4.1 Two-dimensional, steady, parallel flow

To simplify further, we consider a flow that is two dimensional (i.e., no variation in the z direction). Hence we require that our velocity solution is of the form

$$\mathbf{v}(\mathbf{x}) = (v_x(y), 0, 0).$$

Since the flow is steady (the velocity field is independent of time), we can immediately see that $\partial \mathbf{v} / \partial t = 0$.

Hence we can simplify the governing equations. To see this we write them in component form,

$$\begin{aligned} x \text{ direction: } & \rho \left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} \right) v_x = -\frac{\partial p}{\partial x} + \eta \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) v_x + \rho g_x, \\ y \text{ direction: } & \rho \left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} \right) v_y = -\frac{\partial p}{\partial y} + \eta \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) v_y + \rho g_y, \end{aligned}$$

where g_j are the components of the gravity vector in the x, y coordinate system. To simplify further we use $v_y = 0$ and $\partial v_x / \partial x = 0$ to obtain

$$x \text{ direction: } 0 = -\frac{\partial p}{\partial x} + \eta \frac{\partial^2 v_x}{\partial y^2} + \rho g_x, \quad (12.10a)$$

$$y \text{ direction: } 0 = -\frac{\partial p}{\partial y} + \rho g_y, \quad (12.10b)$$

These are the equations of steady, 2-D, parallel flow and can always be used when the coordinates are chosen such that the flow is in the x direction. It is usually best to solve the y equation first.

12.4.2 Flow between parallel plates

Consider a fluid that is sandwiched between two plates that are parallel to each other, as shown in figure 12.1. The gap between the plates has a height H . The plates are very large compared to H and hence we consider them to be infinite. The bottom plate is stationary and the top plate is moving at a speed v_0 in a direction parallel to the gap. A constant pressure gradient $-G$ is *imposed* on the fluid in the same direction as the boundary motion. Gravity pulls the fluid in a direction normal to the plates.

We consider a two-dimensional slice, with the x direction parallel to the direction of motion and the y direction opposite to gravity. The parallel flow equations (12.10) become

$$\begin{aligned} x \text{ direction: } & \frac{\partial p}{\partial x} = \eta \frac{\partial^2 v_x}{\partial y^2}, \\ y \text{ direction: } & \frac{\partial p}{\partial y} = -\rho g. \end{aligned}$$

The second equation can be immediately integrated to give $p = -\rho g y + C(x)$ where C is an arbitrary function of x only. The first term is the static pressure, which does

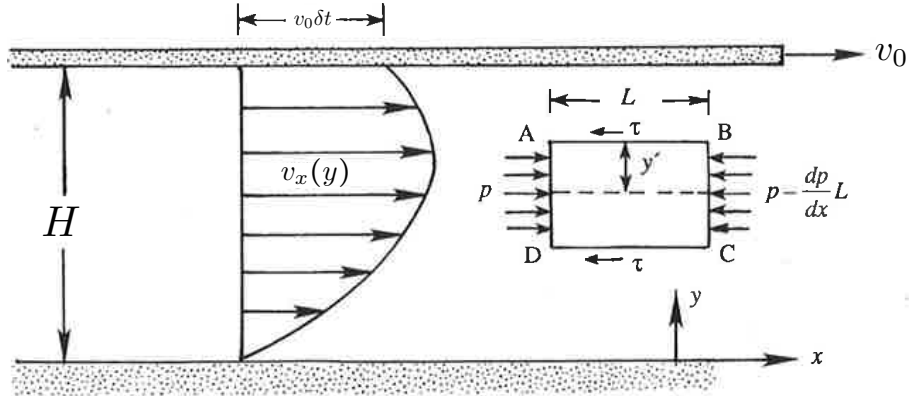


Figure 12.1: Schematic diagram for flow between parallel plates. Inset shows stresses acting on a REV within the flow. The pressure gradient is $\partial p / \partial x = -G$. Source: PK (adapted).

not cause any flow. The second term enables us to write $\partial p / \partial x = C'(x) = -G$. Then we have

$$-\frac{G}{\eta} = \frac{\partial^2 v_x}{\partial y^2}.$$

Integrating this twice with respect to y we obtain

$$v_x(y) = -\frac{G}{2\eta}y^2 + Ay + B,$$

with constants of integration A, B .

Given the physical description above, we impose the *no-slip* condition on $v_x(y)$,

$$v_x(0) = 0, \quad v_x(H) = v_0.$$

The first of these tells us that $B = 0$ and the second one tells us that $A = v_0/H + GH/2\eta$. Using this in the solution we obtain

$$v_x(y) = \frac{G}{2\eta}y(H - y) + \frac{v_0}{H}y.$$

It is simple to verify that this solution satisfies the no-slip boundary conditions. The units also check out: all terms on the right-hand side have units of length/time.

The solution to the parallel flow problem has two control parameters, the pressure gradient G and the plate speed v_0 . Plots for different parameter combinations are shown in figure 12.2. It is helpful to consider the end-member flow regimes associated with these parameters.

Couette flow Setting $G = 0$ (no pressure gradient) we have the linear velocity profile $v_x = v_0 y / H$, which goes from zero at the bottom to v_0 at the top. This is known as Couette flow.

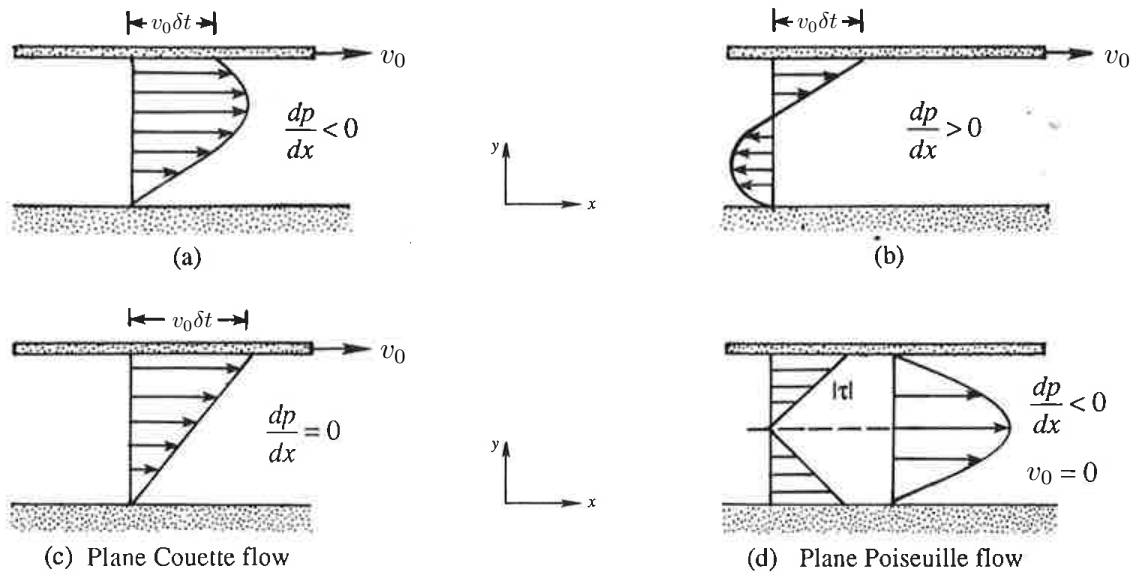


Figure 12.2: Representative solutions for parallel plate flow. The imposed pressure gradient is $-G$, so cases with $G > 0$ have $\partial p / \partial x < 0$.

Poiseuille flow Setting $v_0 = 0$ (stationary upper plate), we have a parabolic flow profile. Since $y \leq H$, the flow is always down the pressure gradient in the positive x direction (for $G > 0$).

Exercise 12.4 on Poiseuille flow through a pipe

Consider an infinitely long, horizontal pipe of radius R containing a fluid of viscosity η . The flow is driven by a pressure gradient $-G$ along the pipe, with $G > 0$.

1. Make a sketch of the problem including only the relevant details. Be sure to include a coordinate system.
2. Write down a parallel-flow equation for axisymmetric flow in a cylinder (in cylindrical coordinates!). You only need to consider the equation governing flow in the direction of the pipe.
3. Solve the equation for the non-zero velocity component as a function of radius.
4. Obtain the flow rate Q as the total volume per time moving through the pipe, $Q = \int_A v dA$, where A is the circular cross-section.

12.4.3 Flow down an inclined plane

Consider an infinite plane that makes an angle θ to the horizontal. A uniform layer of fluid of thickness H coats the surface and flows steadily downhill. We align our coordinate system with the x axis parallel to the flow, along the base, and y

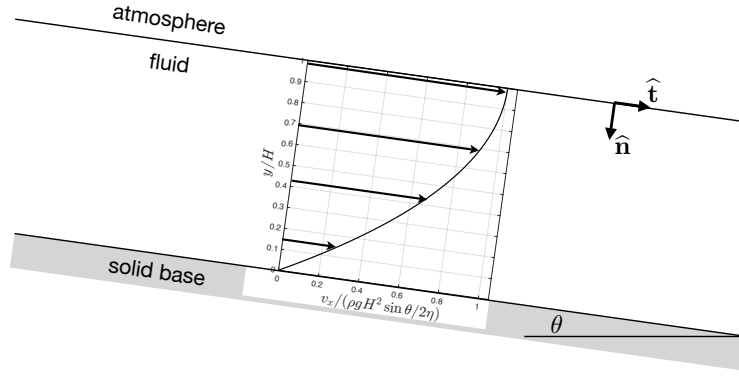


Figure 12.3: Schematic diagram for flow down an inclined plane and plot of the solution for the flow.

perpendicular. The parallel flow equations are

$$\begin{aligned} x \text{ direction: } \frac{\partial p}{\partial x} &= \eta \frac{\partial^2 v_x}{\partial y^2} + \rho g \sin \theta, \\ y \text{ direction: } \frac{\partial p}{\partial y} &= -\rho g \cos \theta, \end{aligned}$$

where $g = |\mathbf{g}|$. We can immediately integrate the second equation to give

$$p = -\rho g y \cos \theta + C(x),$$

where the constant of integration C may be a function of x .

What boundary conditions should we use? Clearly we want no-slip at the base, $v_x(0) = 0$, but we need another condition on v_x at $y = H$, and we need a condition on the pressure to determine C . Starting with the latter, we expect atmospheric pressure outside of the flow. We take this pressure to be p_0 , a constant, and so solving for C leads to $p = p_0 + (H - y)\rho g \cos \theta$.

Now we can simplify the x component because $\partial p / \partial x = 0$. We have

$$\eta \frac{\partial^2 v_x}{\partial y^2} = -\rho g \sin \theta.$$

Integrating once gives

$$\eta \frac{\partial v_x}{\partial y} = -\rho g y \sin \theta + A.$$

Another boundary condition on v_x is needed to solve for A . Here we recognise that the air above offers negligible resistance to the flow — that is, it supplies *zero shear traction at $y = H$* . This is our boundary condition! Using $\hat{\mathbf{n}} = (0, -1)$ and $\hat{\mathbf{t}} = (1, 0)$ in equation (12.8), we obtain the condition

$$T_{xy}(y = H) = \eta \left. \frac{\partial v_x}{\partial y} \right|_{y=H} = 0.$$

Using this to solve gives $A = \rho g H \sin \theta$ and hence

$$\eta \frac{\partial v_x}{\partial y} = \rho g (H - y) \sin \theta.$$

Integrating again and using the no-slip condition gives the solution

$$v_x(y) = \frac{\rho g \sin \theta}{2\eta} y(2H - y).$$

This is plotted in figure 12.3. The axes are scaled to avoid having to specify parameter values.

Examine the shape of $v_x(y)$ in the figure. Note that it goes to zero at the base, satisfying the no-slip condition there. Note also that it has zero slope as it approaches $y/H = 1$, a consequence of the zero-shear traction boundary condition.

12.5 Inviscid flows

When the Reynolds number is much larger than 1, it means that viscosity is unimportant in the bulk of the flow and inertia is dominant. Typically such flows are turbulent, but there are cases where a steady, laminar flow can occur. The momentum equation is then

$$\rho \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \rho \mathbf{g}. \quad (12.11)$$

It is convenient to write the body force in terms of a gravitational potential $\Phi = -\rho \mathbf{g} \cdot \mathbf{x}$. We then use the vector identity

$$\mathbf{v} \times (\nabla \times \mathbf{v}) = \nabla \left(\frac{1}{2} |\mathbf{v}|^2 \right) - \mathbf{v} \cdot \nabla \mathbf{v}. \quad (12.12)$$

to obtain

$$\rho \left[\nabla \left(\frac{1}{2} |\mathbf{v}|^2 \right) - \mathbf{v} \times (\nabla \times \mathbf{v}) \right] = -\nabla p - \nabla \Phi.$$

Dotting both sides of this equation with $\mathbf{v}$ gives

$$\mathbf{v} \cdot \nabla \left[\frac{1}{2} \rho |\mathbf{v}|^2 + \Phi + p \right] = 0. \quad (12.13)$$

This result motivates us to define H as the *Bernoulli energy function*,

$$H = \frac{1}{2} \rho |\mathbf{v}|^2 + \Phi + p, \quad (12.14)$$

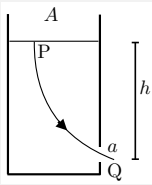
which is a sum of the kinetic and potential energy per unit volume (pressure is a potential to do mechanical work). *Bernoulli's theorem*, $\mathbf{v} \cdot \nabla H = 0$, states that for the steady, inviscid, constant-density flow considered here, H is constant along streamlines of the flow. Put differently, $H = H(\psi)$, where ψ is the two-dimensional streamfunction. *Under these assumptions, the mechanical energy is conserved along each streamline.* This theorem is very useful in solving problems.

Exercise 12.5 on vector calculus

Explain why the term $\mathbf{v} \cdot [\mathbf{v} \times (\nabla \times \mathbf{v})]$ has vanished from (12.13). Geometric logic should be sufficient.

One of the interesting consequences of Bernoulli's theorem is that when a fluid parcel accelerates (at constant height), its kinetic energy increases and therefore its pressure decreases. This helps airplanes to fly (though the fluid dynamics of lift is actually more subtle and interesting than this; see GW or PK).

Example 12.1 of Bernoulli's theorem



Consider a bucket with cross-sectional area A , filled with water, that is draining through a hole with cross-sectional area a . The water level above the hole is h . Find the speed of the flow as it leaves the hole.

The first step of a Bernoulli problem is to identify a relevant streamline. In this case, a good choice is the streamline that goes from P to Q. (Any streamline that exits the hole must connect to the free surface. Why?) We then apply Bernoulli's theorem,

$$\rho v_P^2/2 + p_P + \Phi_P = \rho v_Q^2/2 + p_Q + \Phi_Q.$$

Now, the pressure all around the bucket is atmospheric, so the difference in pressure between the two points is negligible. But the GPE differs by an amount ρgh , where ρ is the density of water. Using these two points we have

$$v_Q^2 - v_P^2 = 2gh.$$

This is one equation in two unknowns... so we need another equation. We can obtain this by thinking of conservation of mass. In particular, mass conservation tells us that the volume rate at the free surface must equal the volume rate at the outflow. This gives us

$$Av_P = av_Q.$$

Using this to eliminate v_P gives

$$v_Q = \sqrt{\frac{2gh}{1 - a^2/A^2}}.$$

Note that for $a \ll A$, this result is asymptotically equal to $\sqrt{2gh}$. Checking the units gives us confidence in our algebra.

Exercise 12.6 on applying Bernoulli's theorem

Use Bernoulli to find the speed of steady water flow out of the nozzle of a fire hose into the atmosphere. The hose and the nozzle have cross-sectional areas A and a , respectively. The pressure in the hose is P and all parts of the hose are at the same elevation.

There are many applications of inviscid flow theory to geoscience, including waves on the surface of the ocean (both standard and tsunami). Unfortunately we will not cover this topic here, but there is an accessible treatment in GW, chapter 7.

A word of caution While treating large-Re flows as inviscid can be convenient and allow significant mathematical insight, it can also lead to predictions that are

physically nonsensical¹⁵. The reason can be seen if we look again at the Reynolds number, $[v][x]/\nu$. Suppose that we have a flow of speed $[v]$ and kinematic viscosity ν over a flat, rigid surface. If we measure $[x]$ as the distance above the surface, there is always some distance at which the Reynolds number is $O(1)$ or smaller. In this region, *viscosity is important*. This region is known as the *viscous boundary layer*. When the kinematic viscosity ν is small and/or the flow speed $[v]$ is large, the boundary layer thickness can be very small. But size isn't everything! The viscous boundary layer enables airplanes to fly, for example. It creates turbulence and dissipates energy, as we'll see in the next section.

12.6 Rotating flows

Earth is rotating (if you aren't sure about this, watch sunrise or sunset). This means that all fluids on Earth are rotating too. In some cases, this matters a lot and in some cases not at all. But even when planetary rotation is unimportant, we still see that fluids develop internal rotations. This is called *vorticity*, and it is a crucial and interesting aspect of fluid mechanics — especially in oceanography and atmospheric science. Here we discuss the basics.

12.6.1 The vorticity equation

The incompressible Navier-Stokes equations capture the physics of rotating flow and vorticity, but it is helpful to write them in a different form to bring this to the fore. Here it is helpful to define the vorticity as the curl of the velocity,

$$\boldsymbol{\omega} = \nabla \times \mathbf{v}. \quad (12.15)$$

(We already saw this above in section 10.2.2 when we analysed the rotation-rate tensor.) To derive an equation governing the rotation, we use the vector identity (12.12) to rewrite the Navier-Stokes momentum equation (12.1a) as

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \nabla \left(\frac{1}{2} |\mathbf{v}|^2 \right) - \mathbf{v} \times \boldsymbol{\omega} \right) = -\nabla p + \eta \nabla^2 \mathbf{v} - \nabla \Phi, \quad (12.16)$$

where we have again introduced the gravitational potential Φ . Taking the curl of this whole equation, using (12.15), and using another vector identity (see exercise 12.7), we obtain the *vorticity equation*,

$$\frac{D\boldsymbol{\omega}}{Dt} = \boldsymbol{\omega} \cdot \nabla \mathbf{v} + \nu \nabla^2 \boldsymbol{\omega}, \quad (12.17)$$

where the Lagrangian derivative on the left-hand side represents transport of vorticity by the flow itself.

¹⁵For example, inviscid flow theory elegantly predicts that flow around a sphere produces zero drag. This is obviously wrong, and the reasons are fairly clear. The solution is still quite useful.

Exercise 12.7 on vector calculus

Use the vector identity

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = (\nabla \cdot \mathbf{b})\mathbf{a} + (\mathbf{b} \cdot \nabla)\mathbf{a} - (\nabla \cdot \mathbf{a})\mathbf{b} - (\mathbf{a} \cdot \nabla)\mathbf{b}$$

to derive equation (12.17) from (12.16), assuming an incompressible flow.

Equation (12.17) states that as parcels of fluid transport their vorticity, there are two processes that modify it. The first term on the right-hand side represents *vortex stretching* while the second term represents viscous diffusion, which can include both dissipation and generation of vorticity. We next consider each of these terms separately.

Vortex stretching If we neglect the effect of viscosity then the vorticity equation becomes

$$\frac{D\boldsymbol{\omega}}{Dt} = \boldsymbol{\omega} \cdot \nabla \mathbf{v} = \boldsymbol{\omega} \cdot (\dot{\mathbf{E}} + \dot{\boldsymbol{\Theta}}),$$

where we have written the velocity gradient tensor in terms of its symmetric (strain rate) and antisymmetric (rotation rate) parts. Now, we dot both sides with the vorticity vector, making it a scalar equation,

$$\frac{D}{Dt} \left(\frac{1}{2} \omega^2 \right) = \boldsymbol{\omega} \cdot (\dot{\mathbf{E}} + \dot{\boldsymbol{\Theta}}) \cdot \boldsymbol{\omega} = \boldsymbol{\omega} \cdot \dot{\mathbf{E}} \cdot \boldsymbol{\omega}.$$

Here the rotation rate contribution vanishes because the tensor is antisymmetric (exercise 12.8). What remains is the strain-rate tensor, which is symmetrical, and thus it can be diagonalised. Then, in a coordinate system aligned with the principal axes of the flow (eigenvectors of the tensor),

$$\boldsymbol{\omega} \cdot \dot{\mathbf{E}} \cdot \boldsymbol{\omega} = \dot{E}_1 \omega_1^2 + \dot{E}_2 \omega_2^2 + \dot{E}_3 \omega_3^2,$$

where $\dot{E}_j$ are the eigenvalues of the strain-rate tensor and ω_j are the components of the vorticity vector.

This result tells us that the vorticity can be amplified, $D(\omega^2/2)/Dt > 0$, when the vorticity vector has a component in a direction in which $\dot{E}_j > 0$. Recall that $\dot{E}_j$ is the normal strain rate in the j^{th} direction; for example, $\partial v_1 / \partial x_1$. So $\dot{E}_j > 0$ means that the j^{th} direction is a *stretching direction*.

What we learn is that *stretching a vortex along its axis of rotation leads to an amplification of vorticity*. This is also true of a ballerina or figure skater (though don't stretch them too much). As you may have experienced when you were a ballerina, drawing the arms in and extending them above the head while spinning leads to amplification of angular velocity.

Exercise 12.8 on vector calculus

Show that for any antisymmetric tensor $\mathbf{A}$ and any vector $\mathbf{b}$,

$$\mathbf{b} \cdot \mathbf{A} \cdot \mathbf{b} = 0.$$

It is sufficient to show this for a 2-D system (i.e. a 2×2 matrix).

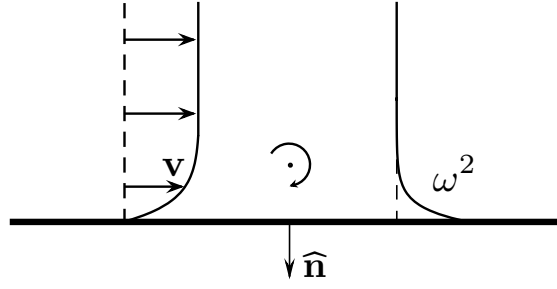


Figure 12.4: Vorticity generation by shear over a no-slip boundary. Source: GW (adapted).

Exercise 12.9 on vorticity and vortex stretching

Consider the flow $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 = E(-x, -y, 2z)$ and $\mathbf{v}_2 = \frac{1}{2}\omega(t)(-y, x, 0)$. Show that this flow has vorticity $\nabla \times \mathbf{v} = (0, 0, \omega(t))$ and that it satisfies the vorticity equation for a particular function $\omega(t)$ that you should find. Sketch each of the flow fields $\mathbf{v}_1$ and $\mathbf{v}_2$ and discuss how this problem illustrates the amplification of vorticity by vortex stretching. (Source: GW)

Vorticity generation & dissipation If we again dot the vorticity equation with the vorticity vector and neglect the vortex stretching term we obtain

$$\begin{aligned} \frac{D}{Dt} \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) &= \nu \omega_i \frac{\partial^2 \omega_i}{\partial x_j \partial x_j}, \\ &= \nu \frac{\partial}{\partial x_j} \left(\omega_i \frac{\partial \omega_i}{\partial x_j} \right) - \nu \left(\frac{\partial \omega_i}{\partial x_j} \right) \left(\frac{\partial \omega_i}{\partial x_j} \right), \\ &= \nu \nabla \cdot \nabla \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) - \nu |\nabla \boldsymbol{\omega}|^2, \end{aligned}$$

where we have reversed the chain rule to obtain the second line. Integrating this equation over some volume V and using the divergence theorem,

$$\int_V \frac{D}{Dt} \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) dV = \nu \int_S \hat{\mathbf{n}} \cdot \nabla \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) dS - \nu \int_V |\nabla \boldsymbol{\omega}|^2 dV. \quad (12.18)$$

The second term on the right-hand side is always negative and hence always causes the vorticity to decrease. It therefore represents *viscous dissipation*. The first term on the right-hand side can be best understood if we take one side of the volume to be a rigid, no-slip boundary, as shown in figure 12.4. Shear is generated by the boundary, within a layer where the no-slip condition affects the flow. Vorticity increases toward the boundary, in the same direction as the outward-pointing normal vector. Hence the first term on the right-hand side represents vorticity generation.

We have shown in this section that the Navier-Stokes momentum equation contains a description of the vorticity of the flow, and that this description can be isolated by deriving a vorticity equation (12.17). We have further shown that the

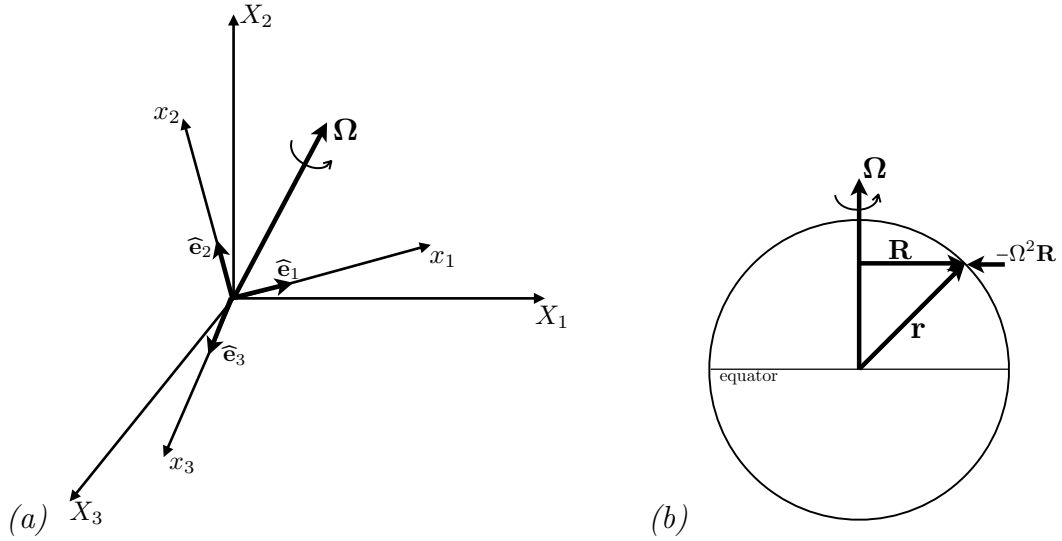


Figure 12.5: Diagrams for calculation of acceleration in rotating frame. (a) Frame with coordinates x_i is rotating with angular velocity Ω with respect to fixed frame with coordinates X_i . (b) Relationship between position vector $\mathbf{r}$ (at Earth surface) and perpendicular position vector $\mathbf{R}$. Centripetal acceleration vector with magnitude $\Omega^2|\mathbf{R}|$ is shown. Source: PK (adapted).

terms in the vorticity equation represent vorticity transport, amplification by vortex stretching (which can be negative), vorticity generation at no-slip boundaries, and vorticity dissipation in the interior of the flow.

There are several further remarks that we should make here.

- In models of *inviscid* flows governed by Euler's equations (12.6), the mechanisms of vorticity generation and dissipation are absent. Therefore, if the flow is initially irrotational, it remains irrotational for all time.
- No natural flows are inviscid, so real flows always have vorticity generation and dissipation.
- Dissipation of vorticity by viscosity is a conversion of energy from kinetic to thermal. The flow slows down but the average random motion of molecules speeds up.

12.6.2 Rotating flows

Here, on the surface of Earth, we are in a rotating frame of reference. Objects in a rotating frame of reference should not, strictly speaking, obey Newton's laws of motion as derived for a non-rotating frame. Obviously we get away with this in some cases, but oceanographers and atmospheric scientists are not fooled. They take rotation into account. You can too; here's how.

We need to derive the Navier-Stokes equations in a rotating reference frame. This will affect only the term representing acceleration (the inertial force). To recognise how this changes, we consider the following simple scenario.

Consider a vector $\mathbf{P}$ that (like all vectors) exists independent of reference frames. It can be expressed in a rotating frame of reference with coordinates x_1, x_2, x_3 . In this rotating frame, shown in figure 12.5a,

$$\mathbf{P} = P_1 \hat{\mathbf{e}}_1 + P_2 \hat{\mathbf{e}}_2 + P_3 \hat{\mathbf{e}}_3 = P_j \hat{\mathbf{e}}_j,$$

$$\left(\frac{d\mathbf{P}}{dt} \right)_R = \hat{\mathbf{e}}_j \frac{dP_j}{dt},$$

where the subscript R indicates the rotating frame. This frame is rotating with respect to a fixed reference frame that has coordinates X_1, X_2, X_3 . The rotation is described by a fixed angular velocity vector $\boldsymbol{\Omega}$. In the fixed reference frame, the motion of the vector $\mathbf{P}$ is obtained in terms of the time derivative as,

$$\begin{aligned} \left(\frac{d\mathbf{P}}{dt} \right)_F &= \frac{d}{dt} (P_j \hat{\mathbf{e}}_j), \\ &= \hat{\mathbf{e}}_j \frac{dP_j}{dt} + P_j \frac{d\hat{\mathbf{e}}_j}{dt}, \\ &= \left(\frac{d\mathbf{P}}{dt} \right)_R + P_j \frac{d\hat{\mathbf{e}}_j}{dt}, \end{aligned}$$

where subscript F represents the fixed frame and we have substituted the time derivative of $\mathbf{P}$ in the rotating frame. We can express the final term as a cross product with $\boldsymbol{\Omega}$,

$$\left(\frac{d\mathbf{P}}{dt} \right)_F = \left(\frac{d\mathbf{P}}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{P}. \quad (12.19)$$

This equation holds true for *any* vector $\mathbf{P}$.

Let's apply (12.19) to the position vector $\mathbf{r}$. We obtain

$$\mathbf{v}_F = \mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r}, \quad (12.20)$$

where we have substituted velocities in place of the time-derivative of position. Now, since we are interested in the acceleration in a fixed frame, let's apply (12.19) to the velocity vector $\mathbf{v}_F$. We obtain

$$\left(\frac{d\mathbf{v}_F}{dt} \right)_F = \left(\frac{d\mathbf{v}_F}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{v}_F.$$

Using (12.20) in this equation we obtain

$$\begin{aligned} \left(\frac{d\mathbf{v}_F}{dt} \right)_F &= \frac{d}{dt} (\mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r})_R + \boldsymbol{\Omega} \times (\mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r}), \\ &= \left(\frac{d\mathbf{v}_R}{dt} \right)_R + \boldsymbol{\Omega} \times \left(\frac{d\mathbf{r}}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{v}_R + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}). \end{aligned}$$

And writing this in terms of acceleration vectors we have

$$\mathbf{a}_F = \mathbf{a}_R + 2\boldsymbol{\Omega} \times \mathbf{v}_R + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}). \quad (12.21)$$

This is what we needed! An equation relating the acceleration in the fixed reference frame (where Newton's laws of motion apply) to that in the rotating reference frame (where we make measurements and models).

However, it is standard to make a further mathematical simplification. We can write $\boldsymbol{\Omega} \times \mathbf{r} = \boldsymbol{\Omega} \times \mathbf{R}$, where $\mathbf{R}$ is the component of position that is perpendicular to $\boldsymbol{\Omega}$, as shown in figure 12.5b. Then, using the vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$ we obtain $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) = -\Omega^2 \mathbf{R}$. Equation (12.21) becomes

$$\mathbf{a}_F = \mathbf{a} + \underbrace{2\boldsymbol{\Omega} \times \mathbf{v}}_{\text{Coriolis}} + \underbrace{(-\Omega^2 \mathbf{R})}_{\text{centripetal}}, \quad (12.22)$$

where we have dropped the subscripts indicating the rotating reference frame but they are implied. Now we see that there are three contributions to the acceleration in a fixed reference frame: acceleration in the rotating frame, the Coriolis acceleration, and the centripetal acceleration. Substituting this into the Navier-Stokes momentum equation (12.1a) we obtain

$$\rho \left(\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} \right) = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho (\mathbf{g} + \Omega^2 \mathbf{R}). \quad (12.23)$$

This equation now rigorously accounts for rotation of the reference frame, and is hence appropriate for modelling flows on Earth. In writing this equation, we have made an arbitrary choice to treat the Coriolis term as an acceleration and keep it on the left-hand side, while moving the centripetal acceleration to the right-hand side (and hence considering it as the centrifugal force).

Exercise 12.10 on centrifugal force

Calculate the magnitude of the centrifugal force due to Earth rotation at the mean radius of Earth as a function of latitude θ . Ratio this quantity to the magnitude of the gravitational force. Using this ratio, explain the existence of the equatorial bulge in Earth's profile. Going further, assume that the equatorial bulge relative to the mean radius scales with the centrifugal force relative to the gravitational force. Use this to predict the height of the equatorial bulge. (The measured difference in radius between the poles and the equator is about 20 km).

12.6.3 Geostrophic balance

In the atmosphere and oceans, it is broadly true (away from the equator and at large scales) that horizontal motions are governed by an approximate balance between the Coriolis force and the pressure-gradient force. To study this we assume a Cartesian coordinate system at some latitude θ , with x increasing to the east and y increasing to the north, as shown in figure 12.6a. In this coordinate system,

$$\Omega_x = 0, \quad \Omega_y = \Omega \cos \theta, \quad \Omega_z = \Omega \sin \theta,$$

where $\Omega = |\mathbf{\Omega}|$ is the magnitude of the angular velocity vector. The Coriolis acceleration is therefore

$$2\mathbf{\Omega} \times \mathbf{v} = \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ 0 & 2\Omega \cos \theta & 2\Omega \sin \theta \\ v_x & v_y & 0 \end{vmatrix}.$$

For our present purposes, we have assumed that the vertical component of the velocity is negligible (in most cases, especially at large scales, the horizontal components are much larger than the vertical). This is a 3-D vector, but we care about the horizontal components,

$$\begin{aligned} (2\mathbf{\Omega} \times \mathbf{v})_x &= -(2\Omega \sin \theta)v_y = -fv_y, \\ (2\mathbf{\Omega} \times \mathbf{v})_y &= (2\Omega \sin \theta)v_x = fv_x, \end{aligned}$$

where we have defined

$$f = 2\Omega \sin \theta.$$

The quantity f is known as the *planetary vorticity* or the *Coriolis parameter* or the *Coriolis frequency*. It has units of inverse time. It is positive in the northern hemisphere ($\theta > 0$) and negative in the southern hemisphere.

If we assume a (horizontal) balance between Coriolis and pressure-gradient forces and divide through by f , we obtain

$$(y \text{ direction:}) \quad v_x = -\frac{1}{f\rho} \frac{\partial p}{\partial y}, \quad (12.24a)$$

$$(x \text{ direction:}) \quad v_y = \frac{1}{f\rho} \frac{\partial p}{\partial x}. \quad (12.24b)$$

This is known as *geostrophic balance*.

Equations (12.24) might look familiar! They are similar to our definition of the streamfunction ψ . In fact, if we take f and ρ to be constant and define $\psi = -p/(f\rho)$ then we have $v_x = \partial\psi/\partial y$ and $v_y = -\partial\psi/\partial x$. This is important because it means that *contours of $p/(f\rho)$ are (approximately) streamlines of the flow*. It is why the wind is usually tangent to isobars on a weather map!

Figure 12.6b shows the flow around low and high pressure centres in the northern hemisphere. The Coriolis force acts perpendicular to the velocity and to its right (for $\theta > 0$ and hence $f > 0$). Because of this, low-pressure storm systems rotate anti-clockwise in the northern hemisphere and clockwise in the southern hemisphere (viewed from above).

Geostrophic balance occurs when the Coriolis acceleration is more important than the Lagrangian acceleration. The Coriolis acceleration scales as $|f|[v]$ whereas the Lagrangian acceleration scales as $[v]^2/[x]$. Their ratio is the Rossby number,

$$\text{Ro} = \frac{\text{Lagrangian acceleration}}{\text{Coriolis acceleration}} = \frac{[v]}{|f|[x]}. \quad (12.25)$$

The Coriolis force is important at small Rossby number ($\text{Ro} \ll 1$), which occurs at sufficiently large length scales $[x]$, away from the equator.

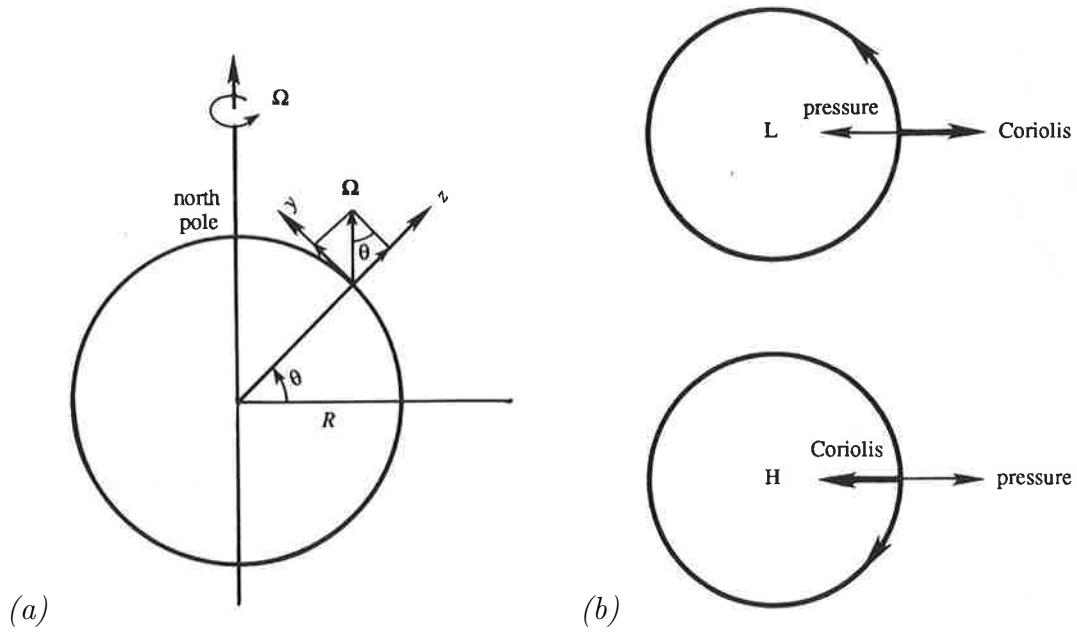


Figure 12.6: Diagrams for geostrophic balance model. (a) Rotating Cartesian coordinate system at latitude θ . (b) Geostrophic balance in the northern hemisphere surrounding a low-pressure centre (top) and a high-pressure centre (bottom), viewed from above. The pressure force $-\nabla p$ is shown by the thin arrow and the Coriolis force is the thick arrow. Flow follows the circular streamline. Source: PK.

Exercise 12.11 on bathtub vortices

People say that planetary vorticity controls the direction of the plug-hole vortex in your bathtub. Estimate the bathtub Rossby number and comment on their hypothesis.

Summary of §12

- A gradient of static pressure balances the body force when the velocity is zero and the Navier-Stokes momentum equation becomes $\nabla p = \rho \mathbf{g}$. With z increasing opposite to gravity, the static pressure within the fluid is $p_S(z) = p_0 - \rho g z$. This balance can sometimes be subtracted from the momentum equation to leave the gradient of the dynamic pressure.
- Scaling analysis of the N-S momentum equation leads to the dimensionless Reynolds number, $\text{Re} = \rho[v][x]/\eta$, which is a measure of the importance of inertia to the importance of viscosity. Large-Re flows are sometimes approximated as being inviscid (with Euler's equations), while small-Re flows are modelled as quasi-static (with Stokes equations).
- Boundary conditions on flows can be kinematic (on the velocity) or dynamic (on the stress and pressure). The no-penetration and no-slip conditions are the most basic and most important.
- Parallel viscous flows have a velocity field with only one component that varies only in the perpendicular directions. Typically we take $\mathbf{v} = v_x(y)\hat{\mathbf{e}}_x$ and solve a reduced system of equations.
- Inviscid flows conserve the Bernoulli energy density $H = \frac{1}{2}\rho|\mathbf{v}|^2 + \Phi + p$, which is found to be constant along streamlines.
- The vorticity equation

$$\frac{D\boldsymbol{\omega}}{Dt} = \boldsymbol{\omega} \cdot \nabla \mathbf{v} + \nu \nabla^2 \boldsymbol{\omega}$$

is derived from N-S and models the evolution of vorticity due to vortex stretching, vorticity generation at no-slip boundaries, and viscous diffusion of vorticity.

- Flows in a rotating reference frame satisfy a modified version of N-S momentum,

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{v} + (\mathbf{g} + \Omega^2 \mathbf{R})$$

that includes a Coriolis acceleration (or inertial force) and a centripetal acceleration (or centrifugal force), where $\boldsymbol{\Omega}$ is the angular velocity of the reference frame and $|\mathbf{R}|$ is the perpendicular distance to the axis of rotation.

- In geostrophic balance, the Coriolis force balances the pressure gradient. In regions with approximately constant planetary vorticity f , this leads to flows that are approximately tangent to contours of pressure.

13 Elastodynamics and elastic flexure

We return, in this chapter, to the topic of elasticity and elastodynamics. For convenience, the governing equations are

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \rho \mathbf{g} + G \nabla^2 \mathbf{u} + \left(K + \frac{1}{3}G\right) \nabla(\nabla \cdot \mathbf{u}), \quad (13.1a)$$

$$\rho(\mathbf{x}) = \rho_0(1 - \nabla \cdot \mathbf{u}). \quad (13.1b)$$

The first of these is conservation of momentum and the second is conservation of mass. The variations in density associated with elastic (de)compression are small and usually inconsequential, so we will treat the density as a constant.

13.1 (Seismic) wave equations

We know that elastic media transmit two types of waves, both pressure waves and shear waves. So these must be solutions of equation (13.1a). To facilitate an analysis of these waves, we make two changes. First, we recall that elastic waves are associated with the exchange of kinetic and elastic energy and that gravity is not of leading-order importance. *Hence we neglect the body force term.* Second, we use a Helmholtz decomposition to split the displacement field into its solenoidal (shear) and irrotational (compression) parts,

$$\mathbf{u} = \nabla \times \Psi + \nabla U, \quad (13.2)$$

where Ψ and U are the vector and scalar potentials, respectively. Substituting this decomposition into (13.1a) and recalling that $\nabla \cdot \nabla \times \mathbf{a} = 0$ for all $\mathbf{a}$ and that $\nabla \cdot \nabla = \nabla^2$, we have

$$\rho \frac{\partial^2}{\partial t^2} (\nabla \times \Psi + \nabla U) = G \nabla^2 (\nabla \times \Psi + \nabla U) + \left(K + \frac{1}{3}G\right) \nabla(\nabla^2 U). \quad (13.3)$$

This looks messier but in fact we've nearly done something remarkable.

13.1.1 P-wave equation

To remove the shear part and leave the compression part, we take the divergence of (13.3). Again using $\nabla \cdot \nabla \times \mathbf{a} = 0$ and $\nabla \cdot \nabla = \nabla^2$, we obtain

$$\rho \frac{\partial^2}{\partial t^2} \nabla^2 U = \left(K + \frac{4}{3}G\right) \nabla^2 \nabla^2 U. \quad (13.4)$$

Now we can reverse the order of the operators on the left-hand side and 'integrate' to remove a ∇^2 operator leaving

$$\frac{\partial^2 U}{\partial t^2} = c_P^2 \nabla^2 U. \quad (13.5)$$

This is a wave equation for the scalar potential U , where $c_P = \sqrt{(K + 4G/3)/\rho}$ is the P-wave speed.

It is important to note that this type of wave doesn't require the continuum to have an elastic shear modulus. Most fluids transmit sound waves because they respond elastically to compression.

13.1.2 S-wave equation

To remove the compression part and leave only the shear part, we take the curl of (13.3). Here we need to use the following vector identity, $\nabla \times \nabla \times \Psi = \nabla(\nabla \cdot \Psi) - \nabla^2 \Psi$. However, since we can construct the vector potential Ψ to be divergence free (see exercise 3.9), we can simplify the identity to $\nabla \times \nabla \times \Psi = -\nabla^2 \Psi$. Then we have

$$-\rho \frac{\partial^2}{\partial t^2} \nabla^2 \Psi = -G \nabla^2 \nabla^2 \Psi.$$

Again we can reverse the order of the operators on the left-hand side and ‘integrate’ to remove a ∇^2 operator leaving

$$\frac{\partial^2 \Psi}{\partial t^2} = c_s^2 \nabla^2 \Psi. \quad (13.6)$$

This is a wave equation for the vector potential Ψ , where $c_s = \sqrt{G/\rho}$ is the S-wave speed.

Exercise 13.1 on one-dimensional wave solutions

Consider the one-dimensional wave equation

$$\frac{\partial^2 q}{\partial t^2} = c^2 \frac{\partial^2 q}{\partial x^2}$$

on an infinite domain (you can think of this as a very long stretched string with perpendicular displacement $q(x, t)$). Show that for any twice-differentiable function $f(X)$, a solution to the wave equation is $q = f(x \pm ct)$. Explain the physical meaning of this solution using the string analogy.

13.2 Plane waves

Plane waves are waves where the wavefront is uncurved (i.e., a plane). Elastic waves in Earth typically radiate radially from something like a point source, so they are not plane waves. However, far from the source, where the radius of the wavefront is very large, the wave can be approximated as a plane wave. Plane P- and S-waves are illustrated in figure 13.1.

As noted in exercise 13.1, the general solution to the one-dimensional wave equation is $f(x \pm ct)$ for a wavespeed c . Seismologists typically express plane-wave solutions (here for the scalar potential) as

$$U(\mathbf{x}, t) = f(t - \mathbf{s} \cdot \mathbf{x}),$$

where $\mathbf{s} = \hat{\mathbf{s}}/c_P$ is the *slowness vector* and $\hat{\mathbf{s}}$ points in the direction of wave propagation. It is also usual to represent more complex wave solutions in terms of a sum over Fourier modes, where each mode is of the form

$$U(\mathbf{x}, t) = A(\omega) e^{-i\omega(t - \mathbf{s} \cdot \mathbf{x})}, \quad (13.7)$$

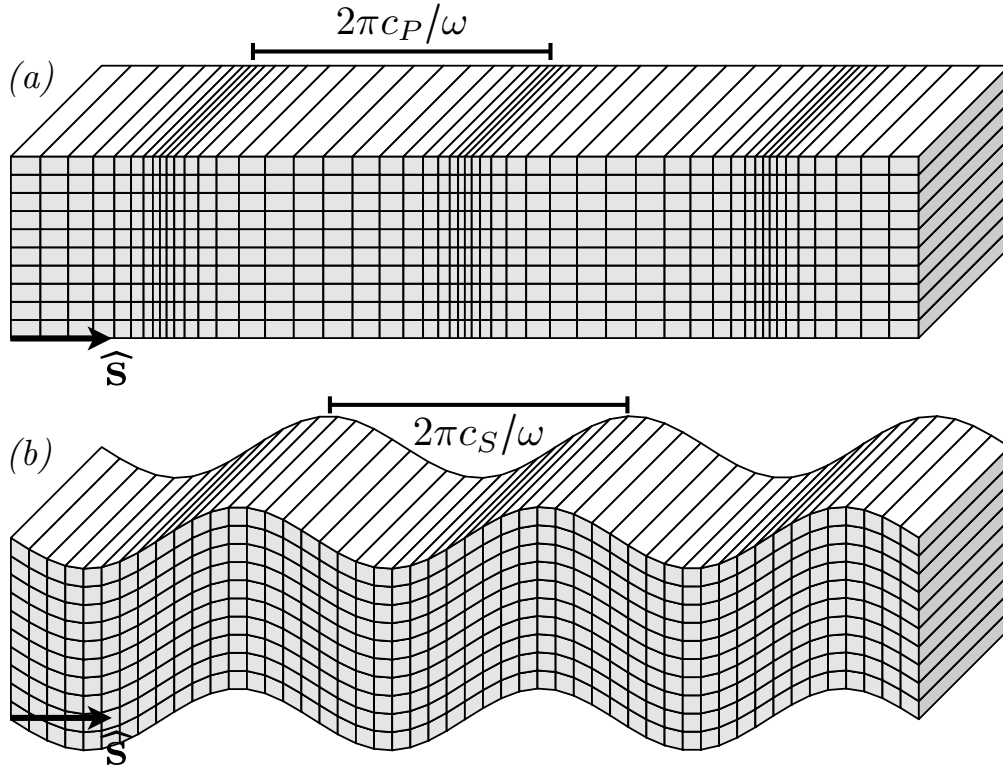


Figure 13.1: Seismic plane waves. (a) A monochromatic P-wave. (b) A monochromatic S-wave.

where ω is the angular frequency and A is a frequency-dependent amplitude. Of course the physical wave is represented by the real part,

$$\Re [A(\omega)e^{-i\omega(t-\mathbf{s}\cdot\mathbf{x})}] = A(\omega) \cos[(\mathbf{k} \cdot \mathbf{x} - \omega t)],$$

where we have taken $A(\omega)$ to be pure real and introduced the *vector wavenumber* $\mathbf{k} = \hat{\mathbf{s}}\omega/c_P$. A plane wave of this type is shown in figure 13.1a.

Exercise 13.2 on plane waves

Verify that (13.7) is a solution to equation (13.5) for a wave propagating with slowness $\mathbf{s} = \hat{\mathbf{e}}_x/c_P$.

13.3 Spherical waves

Imagine a mechanical disturbance (e.g., an explosion) embedded in an infinite, 3-D, elastic medium. The elastic waves created by the disturbance propagate outward, away from the source. If the medium is isotropic, the waves are spherically symmetrical. In that case, the Laplacian operator in the scalar-potential wave equation (13.5) takes the spherical form we obtained in (3.30c), but with only the radial derivatives non-zero,

$$\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial U}{\partial r} = \frac{1}{c_P^2} \frac{\partial^2 U}{\partial t^2}. \quad (13.8)$$

You can verify (exercise 13.3) that the general solution to this equation for outwardly propagating waves is

$$U(r, t) = \frac{f(t - r/c_P)}{r}. \quad (13.9)$$

Note that the solution decays as $1/r$ for any waveform f . Also note that it is singular at $r = 0$, and hence cannot be a (physical) solution there.

Exercise 13.3 on spherical waves

Verify that (13.9) is a general solution to the spherically symmetrical wave equation (13.8).

Equation (13.9) is very close to the general solution of Laplace's equation with spherical symmetry $\psi(r) = A/r + B$. If we return to section 5.2 and review the Poisson equation associated with this solution, we can infer that our solution (13.9) is actually the solution of

$$\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial U}{\partial r} - \frac{1}{c_P^2} \frac{\partial^2 U}{\partial t^2} = -4\pi \delta^3(\mathbf{x}) f(t), \quad (13.10)$$

where $\delta^3(\mathbf{x})$ is the 3-D Dirac delta function centred at the origin, and $f(t)$ is the same function as in (13.9), except now only a function of time. The right-hand side represents an energy source, such as an earthquake or underground explosion, that leads to the outward radiation of waves. In fact, $f(t)$ is sometimes called the *source-time function*.

Exercise 13.4 on spherical waves

Consider the spherically symmetrical, isotropic wave equation (13.10). For the source-time function $f(t) = \cos t$ (which means that the source has been oscillating for all past time and will continue for all future time). Find the solution $U(r, t)$ that satisfies (13.10).

Our discussion of plane and spherical waves has focused exclusively on P-waves, which can be described by a scalar potential. But seismology requires solutions to the full elastodynamic equation (13.1a) that describe both S- and P-waves. In that case, a vector potential is also required. This is a more advanced topic.

13.4 Elastic stretching of a rod

A simple but important example of static elastic deformation is the case of an elastic rod (e.g., a steel wire) being extended by a principal stress T_1 along its length. The wire is surrounded by fluid (e.g., air), and so the stress in each of the perpendicular directions is zero, $T_2 = T_3 = 0$. These stresses cause a set of principal strains in the same directions E_1, E_2, E_3 . Under uniaxial stress, none of these strains are zero! The rod extends in the 1 direction and contracts in the 2 and 3 directions. In this section we model that elastic deformation.

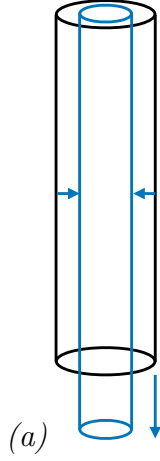


Figure 13.2: Schematic diagram showing an elastic rod being stretched by a uniaxial stress. The undeformed rod has a radius R and a length L . The ratio of strains is given by Poisson's ratio, $\Delta R/R = -\nu \Delta L/L$.

To do so, we use the linear elastic rheological model (10.18). Using the Lamé constant $\lambda = K - 2G/3$, introduced with equation (11.4), this model can also be written

$$\mathbf{T} = 2G\mathbf{E} + \lambda(\nabla \cdot \mathbf{u})\mathbf{I}.$$

Since our coordinates are aligned with the principal components, the tensors are diagonal and we have

$$T_1 = (\lambda + 2G)E_1 + \lambda E_2 + \lambda E_3, \quad (13.11a)$$

$$T_2 = \lambda E_1 + (\lambda + 2G)E_2 + \lambda E_3, \quad (13.11b)$$

$$T_3 = \lambda E_1 + \lambda E_2 + (\lambda + 2G)E_3. \quad (13.11c)$$

Now, setting $T_2 = T_3 = 0$ we can use the second two equations to determine that

$$E_2 = E_3 = \frac{-\lambda}{2(\lambda + G)}E_1 \quad (\text{uniaxial stress}). \quad (13.12)$$

To clean up this equation we can define

$$\nu = \frac{\lambda}{2(\lambda + G)},$$

which is *Poisson's ratio* (not to be confused with kinematic viscosity, that conventionally shares the same symbol). Writing

$$E_{2,3} = -\nu E_1,$$

we see that if $E_1 > 0$ representing extension, then $E_{2,3} < 0$ representing contraction. So as the rod is pulled and extended, it becomes more slender as a consequence.

Exercise 13.5 on Poisson's ratio

Find the value that Poisson's ratio must take for an incompressible material.

Using the definition of Poisson's ratio and equation (13.11a), we can obtain the relationship between axial stress and axial strain as

$$T_1 = \mathcal{E}E_1 \quad (\text{uniaxial stress}), \quad (13.13)$$

where we have introduced Young's modulus, $\mathcal{E} = G(3\lambda + 2G)/(\lambda + G)$ to simplify the relationship between the uniaxial stress and the co-linear strain.

Young's modulus and Poisson's ratio form a set of elastic parameters that are determined by λ, G , or equivalently by K, G . Indeed, $\mathcal{E}$ and ν can be obtained by measuring the strains in a uniaxial loading test and used to obtain λ and G (and hence also K and the other Lamé parameter $\mu = G$).

13.5 Elastic flexure of plates

Plate tectonics on Earth requires that plates bend as they descend into subduction zones. Plates also bend when they are loaded by a superimposed volcano or accumulation of sediment. This bending is associated with an elastic stress that we analyse in this section.

13.5.1 Plane stress

From uniaxial stress (where $T_2 = T_3 = 0$), it is straightforward to consider a situation of stress where there are two non-zero principal stresses in a plane. This is the situation we encounter with a thin elastic plate (e.g., a tectonic plate). In this case $T_3 = 0$, but as with uniaxial stress, all of the strains are potentially non-zero. In particular, we can use equations (13.11), Young's modulus and Poisson's ratio to obtain the elastic constitutive equations under plane stress

$$E_1 = \frac{1}{\mathcal{E}}(T_1 - \nu T_2), \quad (13.14a)$$

$$E_2 = \frac{1}{\mathcal{E}}(T_2 - \nu T_1), \quad (13.14b)$$

$$E_3 = \frac{-\nu}{\mathcal{E}}(T_1 + T_2). \quad (13.14c)$$

We will use these relationships in the next section.

13.5.2 Cylindrical bending of a plate

A flat plate with Cartesian coordinate system is shown in figure 13.3a. In the neutral, reference state of no flexure, there are no elastic stresses. In general, elastic bending of the plate can cause displacements that vary with both x and y . However, in this course we will consider only *cylindrical flexure*, as shown in figure 13.3b. In that case, the plate bends in the x direction and is perfectly straight in the y direction.

The flexure shown in fig. 13.3b is circular, with a radius of curvature R that is much larger than the thickness of the plate H . Although flexure means that the plate is no longer parallel to the x axis, we will assume that the deviation is small and hence distance along the plate can still be measured by x .

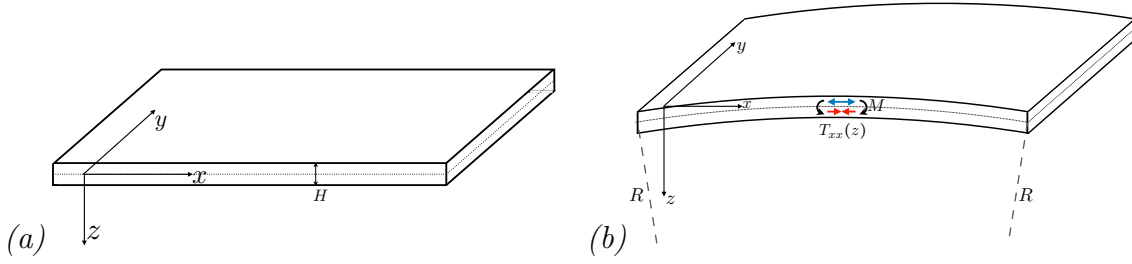


Figure 13.3: The Euler-Bernoulli elastic plate and coordinate system. Note that z increases downward. (a) The plate in its flat, neutral configuration. The origin of the coordinates is on the mid-plane of the plate, so the plate occupies $-H/2 \leq z \leq H/2$. (b) The plate with a cylindrical bend having radius of curvature R . The fibre stresses T_{xx} and their resultant bending moment M are shown schematically. As shown, the bending moment is negative.

The bending generates normal stresses T_{xx} within the plate. These stresses are also known as *fibre stresses* because they remain parallel to material lines (i.e., fibres) within the plate. With the flexure shown in fig. 13.3b, these stresses are positive (tensile) in the upper half of the plate ($z < 0$) and negative (compressive) in the lower half of the plate ($z > 0$). We can conveniently describe them in terms of a *bending moment*,

$$M(x) = \int_{-H/2}^{H/2} T_{xx} z \, dz. \quad (13.15)$$

This moment measures the torque (per unit length in y) on the beam at a point in x . Torque usually has units force-distance, but torque per y -length has units of force.

To derive an equation governing the vertical displacement of the plate $w(x)$, we consider vertical force balance and torque balance applied to a short segment of the plate, shown in figure 13.4. In both cases, we assume the plate is in static equilibrium, i.e., it is not accelerating. We consider a segment of the plate length δx , which we will assume to be very small. Finally, we ignore any net compression in the x direction.

Vertical force balance Referring to figure 13.4a, we now derive an equation for the balance of forces. There is a force per unit area $q(x)$ pushing downward or upward on the plate, giving a total force $q(x)\delta x$ (per unit length in y). This could be due to the weight of accumulated sediments, pressure in the mantle beneath, etc. We also must account for the shear stresses T_{xz} at the ends of the segment. Force balance (per unit length in y) gives us

$$q(x)\delta x + H[T_{xz}(x + \delta x) - T_{xz}(x)] = 0. \quad (13.16)$$

We can simplify by Taylor expanding

$$T_{xz}(x + \delta x) \approx T_{xz}(x) + \left. \frac{dT_{xz}}{dx} \right|_x \delta x.$$

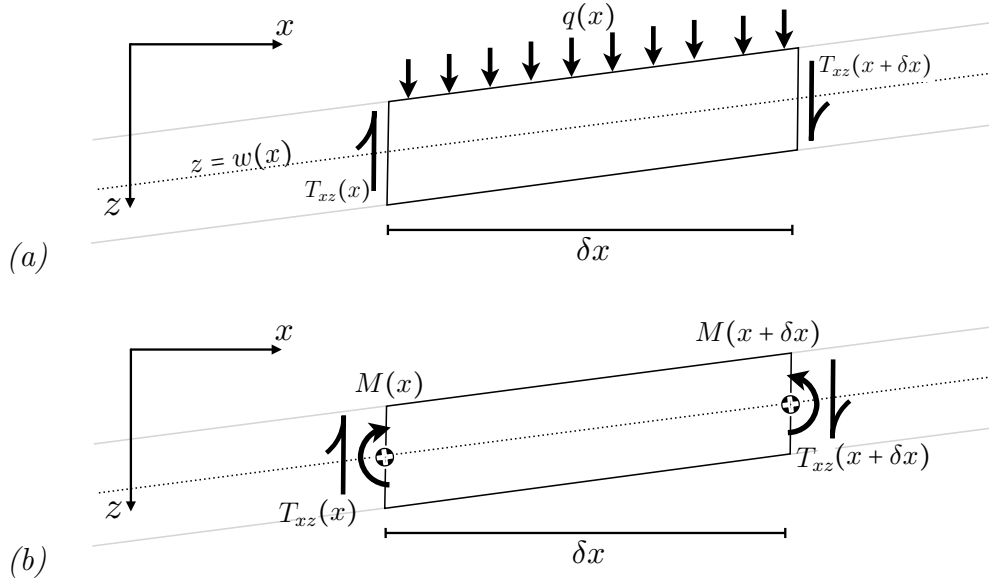


Figure 13.4: Mechanical balances in an x - z section of the plate. All forces and moments are shown as being positive. (a) Entries in the vertical force balance. (b) Entries in the torque balance. Source: adapted from TS.

Using this in (13.16) gives

$$H \frac{dT_{xz}}{dx} = -q. \quad (13.17)$$

This balance must hold at all x in the plate.

Torque balance Referring to figure 13.4b, we next consider the balance of torques (per unit length in y) on the segment of plate. Counterclockwise torques are positive. The bending moment is a torque and enters directly. The shear stress at $x + \delta x$ exerts a torque in proportion to its moment arm length, δx . Hence the torque balance is

$$M(x + \delta x) - M(x) - HT_{xz}(x + \delta x)\delta x = 0. \quad (13.18)$$

We Taylor expand the moment

$$M(x + \delta x) \approx M(x) + \left. \frac{dM}{dx} \right|_x \delta x,$$

and use the Taylor expansion of the shear stress from above. Then equation (13.18) becomes

$$\frac{dM}{dx} \delta x - H \left(T_{xz} + \frac{dT_{xz}}{dx} \delta x \right) \delta x = 0.$$

Since we are considering the limit of δx becoming very small, we can neglect the term in $(\delta x)^2$. We obtain

$$\frac{dM}{dx} = HT_{xz}. \quad (13.19)$$

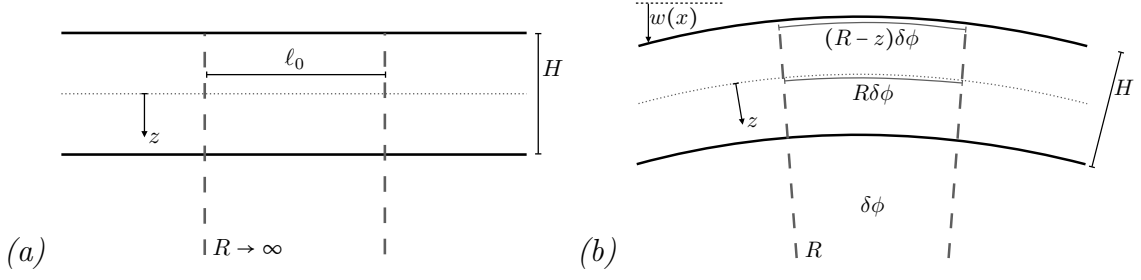


Figure 13.5: Relationship between flexure and fibre strain in an Euler-Bernoulli plate in cylindrical flexure. (a) The plate is flat (the radius of curvature is infinite) and all fibre strains are zero. A material segment has width ℓ_0 . (b) The plate is flexed with radius of curvature R . The lines bounding the segment (dashed) have an angle $\delta\phi$ between them. The mid-plane fibre has segment length $\ell_0 = R\delta\phi$. Fibre segment length increases in the $-z$ direction, with a maximum of $(R + H/2)\delta\phi$ at the upper surface of the plate.

Equation governing the vertical displacement Now we take an x derivative of equation (13.19) and use equation (13.17) to eliminate $H dT_{xz}/dx$ to give

$$\frac{d^2 M}{dx^2} = -q. \quad (13.20)$$

This is the equation governing the bending moment $M(x)$.

To make progress, we must write M and q in terms of the vertical displacement $w(x)$. We focus on M first, and return to equation (13.15), where we aim to replace T_{xx} with the fibre strain E_{xx} . We note that the plate is in a state of plane stress ($T_{zz} = 0$), and that the strain in the y direction is zero ($E_{yy} = 0$). Returning to equations (13.14), we identify x as the 1-direction and y as the 2-direction. Hence from equation (13.14b) we have $T_{yy} = \nu T_{xx}$, which can be used in (13.14a) to give

$$T_{xx} = \frac{\mathcal{E}}{1 - \nu^2} E_{xx}. \quad (13.21)$$

This result gets us closer, but we still need to relate the fibre strain E_{xx} to the z displacement of the mid-plane of the plate, $w(x)$.

Consider a flexed plate as shown in figure 13.5b, where the local shape is approximated by a circle with radius R . This is known as the radius of curvature. The local curvature of the plate is then $1/R$. In the upper half of the plate, at $z < 0$, fibres are stretched with $E_{xx} > 0$ and in the lower half, at $z > 0$, the strain is compressive $E_{xx} < 0$; the fibre strain at $z = 0$ is zero. Recall that the strain is $\delta\ell/\ell_0$. Then, using the geometry of segment lengths in figure 13.5b, the strain is

$$E_{xx}(z) = \frac{(R - z)\delta\phi - R\delta\phi}{R\delta\phi} = -\frac{z}{R}.$$

This doesn't appear helpful because it isn't in terms of $w(x)$ — but we are nearly there! The factor of $1/R$ is the curvature of the cylindrical bend. For small curvature ($R \gg H$) and small vertical displacements w , the curvature is closely approximated by the second derivative of w ,

$$\frac{1}{R} \sim \frac{d^2 w}{dx^2}.$$

Combining this with (13.21) and substituting into the definition of the bending moment (13.15) we obtain

$$\begin{aligned} M(x) &= - \int_{-H/2}^{H/2} \mathcal{E}' z^2 \frac{d^2 w}{dx^2} dz, \\ &= -\mathcal{E}' I \frac{d^2 w}{dx^2}, \end{aligned} \quad (13.22)$$

where $\mathcal{E}' = \mathcal{E}/(1 - \nu^2)$ is a modified Young's modulus and $I = H^3/12$ is called the second moment of area (per unit length in y). The product $B = \mathcal{E}' I$ is referred to as the bending stiffness.

Substituting (13.22) into the force-and-torque balance (13.20) gives us the equation governing the vertical displacement due to bending,

$$B \frac{d^4 w}{dx^4} = q. \quad (13.23)$$

This is a fourth-order equation for the vertical displacement of the beam away from its neutral position. To solve it we need an expression for q and four boundary conditions.

13.5.3 Flexural–isostatic compensation of a floating plate

One interesting application of the Euler-Bernoulli plate model for cylindrical bending avoids the need to impose boundary conditions by considering periodic solutions on an infinite domain. We envision a floating elastic plate of mean thickness H and density ρ_p on a fluid of density ρ , at its level of neutral buoyancy (the weight of the layer is supported by the pressure in the fluid below). We can think of this as less-dense crust floating on the mantle, or an ice shelf floating on sea water.

As illustrated in figure 13.6a, suppose that the plate has a periodic variation in thickness, representing an additional load that is distributed as $\rho_p g h_0 \sin(2\pi x/\lambda)$, where g is gravity, h_0 is the amplitude of the variation in thickness, and λ is the wavelength of the thickness variation. How much topography will this variation generate? This depends on the amount of isostatic compensation! The vertical load causes an unknown vertical displacement $w(x)$. The displacement pushes the plate downward into the fluid, but the plate is less dense and wants to rise. The force at a displacement w beneath the neutral level is associated with the weight of the plate pushing downward with $\rho_p g w$ and the pressure of the fluid pushing upward with $-\rho g w$. The net downward force is thus $(\rho_p - \rho) g w$, where we note that $\rho_p < \rho$. The total downward force per unit area is then

$$q(x) = \rho_p g h_0 \sin(2\pi x/\lambda) - \Delta \rho g w(x),$$

where $\Delta \rho = \rho - \rho_p$. This is substituted into equation (13.23) to obtain

$$B \frac{d^4 w}{dx^4} + \Delta \rho g w = \rho_p g h_0 \sin(2\pi x/\lambda). \quad (13.24)$$

We seek a solution $w(x)$ to this equation.

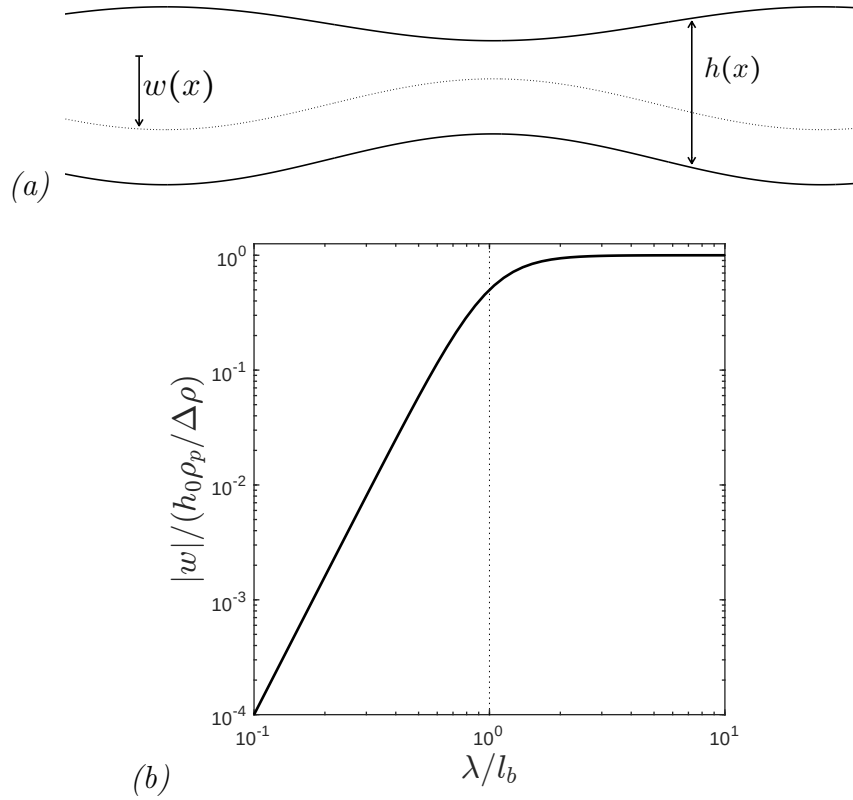


Figure 13.6: Flexural-isostatic compensation. (a) Schematic diagram of a plate with thickness variation $h(x) = H + h_0 \sin(2\pi x/\lambda)$ that has caused some vertical displacement $w(x)$. (b) Scaled amplitude of vertical displacement as a function of scaled wavelength. The wavelength is scaled by the natural bending wavelength and the amplitude is scaled by the displacement for complete isostatic compensation.

It is reasonable to expect that $w(x) \propto \sin(2\pi x/\lambda)$, because this is the distribution of the applied load on the plate. So we guess that $w(x) = w_0 \sin(2\pi x/\lambda)$ and substitute this into equation (13.24) to obtain

$$w_0[B(2\pi/\lambda)^4 + \Delta\rho g] = \rho_p g h_0.$$

Solving this for w_0 and putting this into our trial solution gives

$$w(x) = \frac{\rho_p h_0 / \Delta\rho}{B(2\pi/\lambda)^4 / \Delta\rho g + 1} \sin(2\pi x/\lambda). \quad (13.25)$$

This indicates that the amount of displacement (and bending) depends on the wavelength λ .

To be more specific, we consider the denominator, noting that the first term is negligible in the long-wavelength limit $B(2\pi/\lambda)^4 / \Delta\rho g \ll 1$. In this case, $|w| = h_0(\rho_p / \Delta\rho)$, which represents full isostatic compensation. At the other limit, the second term is negligible for small wavelengths because $B(2\pi/\lambda)^4 / \Delta\rho g \gg 1$. In this case, $|w| \rightarrow 0$. There is no displacement of the plate when the loading wavelength is short; flexural rigidity prevents it. So we see that at sufficiently short wavelengths, the resistance to bending prevents isostatic compensation. Where is the transition between these two regimes? The transition is where $B(2\pi/\lambda)^4 / \Delta\rho g \sim 1$ or, in terms of wavelength, when

$$\lambda \sim l_b = 2\pi \left(\frac{B}{\Delta\rho g} \right)^{1/4}.$$

This is the natural bending length scale of the plate.

The amplitude of vertical displacement of the plate is plotted in figure 13.6b as a function of wavelength. The wavelength is normalised by l_b , the bending length scale, and the displacement amplitude is normalised by $h_0 \rho_p / \Delta\rho$, the displacement with total isostatic compensation. We see the transition from vanishing flexure at small wavelength to complete isostatic compensation at large wavelength.

Summary of §13

- The equations of elastodynamics govern the displacement $\mathbf{u}$ and elastic variations of density. However, the density variations are typically small and inconsequential, so they are often neglected.
- A Helmholtz decomposition of the displacement separates it into solenoidal (shear) and irrotational (compression) parts. Each of these is associated with a potential that satisfies a wave equation.
- The irrotational part gives the equation for P-waves while the solenoidal part gives the equation for S-waves.
- Two simple solutions of the wave equations are plane waves and spherical waves. Scalar potentials for these solutions are, respectively

$$U(\mathbf{x}, t) = f(t - \mathbf{s} \cdot \mathbf{x}) \quad \text{and} \quad U(r, t) = \frac{f(t - r/c_P)}{r},$$

where $\mathbf{s} = \hat{\mathbf{s}}/c_P$ is the slowness vector and $\hat{\mathbf{s}}$ is the propagation direction.

- Elastic bodies such as rods, when stretched uniaxially, deform in the perpendicular direction by an amount given by Poisson's ratio, ν . For a uniaxial stretch of $\Delta L/L = E_1 > 0$, the perpendicular strain is $E_2 = -\nu E_1$. Poisson's ratio is positive for all conventional materials so E_2 is a contraction.
- The elastic bending of thin plates is modelled by balancing the second derivative of the bending moment with external loads (typically arising from gravity). This leads to a fourth-order equation for vertical displacement.