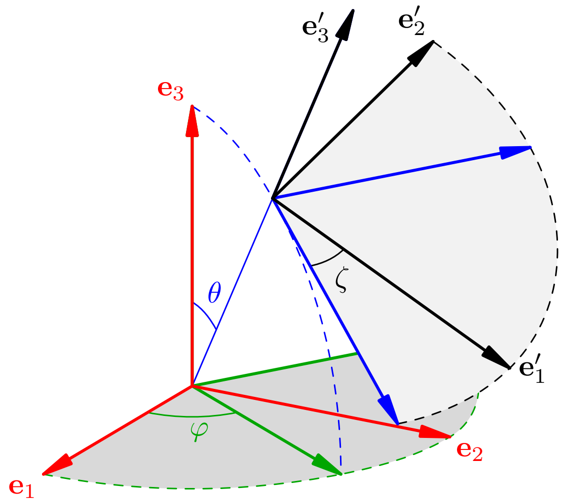


MATH 321

VECTOR AND COMPLEX CALCULUS FOR THE PHYSICAL SCIENCES

©Fabian Waleffe
Department of Mathematics &
Department of Engineering Physics

University of Wisconsin, Madison

December 15, 2023

Preface

These notes have been developed to fill the gap between the mathematics covered in standard multivariable calculus courses and the vector and complex calculus needed for intermediate and advanced undergraduate courses in the physical and engineering sciences. Emphasis is placed on geometric understanding of the fundamental concepts and connections with physical world experiences and applications. Vector, index and matrix notations are discussed since each have their own merits and limitations and all three are used interchangeably in applications.

The mathematical concepts and tools covered here are fundamental to sciences such as *mechanics, electromagnetism, fluid dynamics, aerodynamics, transport phenomena, flight dynamics, astrodynamics, continuum mechanics, elasticity, plasma physics, geophysical and astrophysical fluid dynamics*, for example, as well as *computer graphics* that requires a deeper understanding of the mathematical modeling of curves, surfaces and volumes. It is common for instructors of courses in those areas to spend 1/3 or more of their time quickly 'reviewing' the requisite vector and complex calculus that is needed for adequate presentation and understanding of the material. This creates excessive redundancy at an introductory level that is frustrating and inefficient for students and instructors alike.

The material is presented at an intermediate to advanced undergraduate level. Students are expected to have mastered basic geometry, trigonometry, algebra and calculus including multi-variable calculus (partial derivatives, multiple integrals). Familiarity with differential equations and linear algebra is recommended but not required. Knowledge of introductory physics is assumed. The concepts are motivated, derived and justified carefully but the approach is intended to be more 'intuitive' than 'rigorous,' although those terms are subjective. The derivations and proofs should be studied and understood as they lead to deeper understanding of the concepts. Solved examples are provided but not too many as to drown out the fundamental concepts. The concepts and examples should be studied thoroughly before attempting the exercises. Exercises appear after most subsections to encourage the productive cycle of studying a particular concept then practicing its applications in non-trivial problems and going back to studying the concept more deeply. Learning is an iterative process.

Acknowledgements: Thanks to Prof. Jean-Luc Thiffeault and Drs. Anakewit 'Tete' Boonkasame and Alfredo Wetzel for using and commenting on these notes. Skull surface image courtesy of `https://doc.cgal.org/latest/Surface_mesher/index.html`.

Contents

Chapter 1

Vector Geometry and Algebra

1 Find your bearings

What is a vector? In calculus, you may have defined *vectors* as lists of numbers such as (a_1, a_2) or (a_1, a_2, a_3). Mathematicians favor that approach nowadays because it readily generalizes to n dimensional space $(a_1, a_2, \ldots, a_n)$ and reduces geometry to arithmetic. But in physics, and in geometry before that, you encountered vectors as quantities with both *magnitude* and *direction* such as *displacements*, *velocities* and *forces*. The geometric point of view emphasizes the invariance of these quantities with respect to changes of the system of coordinates.

1.1 Magnitude and Direction

The prototypical vectors are the displacements in two or three dimensional space such as the displacement of a boat on the surface of a lake, or a drone flying over a prairie. We write

$$\boldsymbol{a} = a\,\hat{\boldsymbol{a}} \tag{1}$$

for a vector $\boldsymbol{a}$ of magnitude a in the direction $\hat{\boldsymbol{a}}$.

In navigation, the direction $\hat{\boldsymbol{a}}$ is specified by the *heading* (or *azimuth*) α which is the clockwise angle from the Northern direction. The magnitude a is specified in appropriate physical units such as meters for a displacement, or meters/second for a velocity. Magnitude does not have to come before direction, '1km Northeast' is the same as 'Northeast 1km,'

$$a\,\hat{\boldsymbol{a}} = \hat{\boldsymbol{a}}\,a.$$

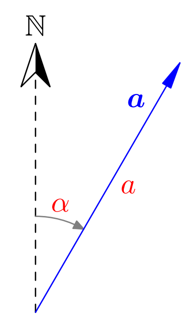


Fig. 1.1: Vector $\boldsymbol{a}$ with magnitude a, heading α

Vectors are denoted in boldface type such as $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{u}$, $\boldsymbol{v}$, …, usually lower-case but often upper-case also in Physics, for example for the magnetic field $\boldsymbol{B}$, electric field $\boldsymbol{E}$ or resultant force $\boldsymbol{F}$. The magnitude of vector $\boldsymbol{a}$ is written in regular type a or with vertical bars $|\boldsymbol{a}|$

$$|\boldsymbol{a}| = |a\hat{\boldsymbol{a}}| = a \geq 0. \tag{2}$$

The magnitude $|\boldsymbol{a}| = a \geq 0$ is a positive real number with appropriate physical units (meters, Newtons, ...). The direction of $\boldsymbol{a}$ is denoted in boldface with a hat, $\hat{\boldsymbol{a}}$. Geometrically speaking, a direction such as *North*, *East* or *heading* α, has no magnitude, however vector algebra operations such as (2) and (7) below lead to

$$|\hat{\boldsymbol{a}}| = 1. \tag{3}$$

For that reason, *direction vectors* are often called *unit* vectors, although these 'unit' vectors do not have physical units.[1]

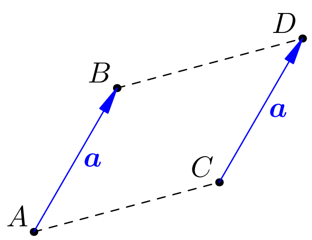


Fig. 1.2: $\overrightarrow{AB} = \overrightarrow{CD} = \boldsymbol{a}$

Two points A and B specify a displacement vector $\boldsymbol{a} = \overrightarrow{AB}$. The same *displacement* $\boldsymbol{a}$ starting from point C leads to point D, with

$$\overrightarrow{AB} = \boldsymbol{a} = \overrightarrow{CD}.$$

Conversely, we can specify points by specifying displacements from a reference point, thus point B is located displacement $\boldsymbol{a}$ from point A, and D is $\boldsymbol{a}$ from C. Two vectors are equal when their magnitudes and directions match

$$\boldsymbol{a} = \boldsymbol{b} \Leftrightarrow a = b \text{ and } \hat{\boldsymbol{a}} = \hat{\boldsymbol{b}}. \tag{4}$$

Notation. Some authors write $\vec{a}$ for $\boldsymbol{a}$ and we use that notation to denote the displacement from point A to point B as $\overrightarrow{AB}$. Writing by hand, we use the typographical notation for boldface which is

$$\underset{\sim}{a} \equiv \boldsymbol{a}, \quad \underset{\sim}{\hat{a}} \equiv \hat{\boldsymbol{a}}, \text{ etc.}$$

This notation allows us to distinguish between a collection of unit vectors

$$\{\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2, \hat{\boldsymbol{a}}_3\} \equiv \{\underset{\sim}{\hat{a}}_1, \underset{\sim}{\hat{a}}_2, \underset{\sim}{\hat{a}}_3\}$$

and the cartesian components of a unit vector

$$\hat{\boldsymbol{a}} \equiv \underset{\sim}{\hat{a}} \equiv (\hat{a}_1, \hat{a}_2, \hat{a}_3).$$

That notational ability to distinguish between those two concepts, unit vector $\hat{\boldsymbol{a}}_i$ and component $\hat{a}_i$ of unit vector $\hat{\boldsymbol{a}}$, will be useful in some instances.

1.2 Vectors on a plane

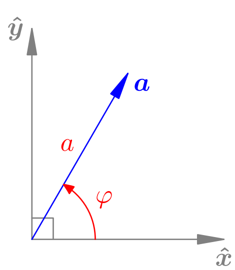


Fig. 1.3: Polar form of $\boldsymbol{a}$: magnitude a, azimuth φ

In mathematical physics, we specify a direction in a plane (say a horizontal plane) using an angle φ counterclockwise from a reference direction $\hat{\boldsymbol{x}}$. A 2D vector $\boldsymbol{a}$ can then be specified by the pair (a, φ) for vector $\boldsymbol{a}$ of magnitude $|\boldsymbol{a}| = a$ and direction $\hat{\boldsymbol{a}}$ specified by azimuth φ, the angle from $\hat{\boldsymbol{x}}$ to $\hat{\boldsymbol{a}}$, toward $\hat{\boldsymbol{y}}$. That is the *polar representation* of $\boldsymbol{a}$ illustrated in fig. 1.3.

[1] This is the usual convention that we follow in these notes. We could attach the physical units to the 'unit' vectors with the vector magnitudes then being unit-less real numbers. That may be useful in computer graphics for instance, but less so in physics where quantities with different physical units may have the same direction. For instance, $\boldsymbol{F} = m\boldsymbol{a}$ tells us that the force $\boldsymbol{F}$ and the acceleration $\boldsymbol{a}$ have the same direction, that is $\hat{\boldsymbol{F}} = \hat{\boldsymbol{a}}$ if our unit vectors have no physical units.

We can also specify the vector $\boldsymbol{a}$ as a sum of displacements in two reference perpendicular directions, a_x in direction $\hat{\boldsymbol{x}}$ and a_y in direction $\hat{\boldsymbol{y}}$ perpendicular to $\hat{\boldsymbol{x}}$. Now the pair (a_x, a_y) specifies $\boldsymbol{a}$. That is the *cartesian representation* of $\boldsymbol{a}$ (fig. 1.4). The cartesian reference directions $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are independent of $\boldsymbol{a}$ and thus the cartesian components a_x and a_y can be negative, contrary to magnitude $a = |\boldsymbol{a}| \geq 0$.

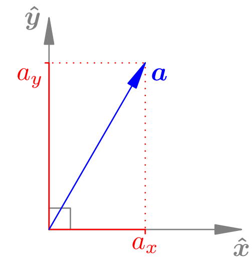


Fig. 1.4: Cartesian form of $\boldsymbol{a}$: components a_x and a_y

The relationships between the polar (a, φ) and cartesian (a_x, a_y) representations follow directly from basic trigonometry,

$$\begin{cases} a_x = a\cos\varphi \\ a_y = a\sin\varphi \end{cases} \Leftrightarrow \begin{cases} a = \sqrt{a_x^2 + a_y^2}, \\ \varphi = \text{atan2}(a_y, a_x), \end{cases} \tag{5}$$

where atan2 is the arctangent function with range in $(-\pi, \pi]$. The classic arctangent function $\text{atan}(a_y/a_x)$ has the range $[-\pi/2, \pi/2]$ and determines φ only up to a multiple of π, that is $\varphi = \arctan(a_y/a_x) + k\pi$ where k is an integer.

It is clear from (5) that

$$(a, \varphi) \neq (a_x, a_y),$$

in general, yet these two pairs of numbers (a, φ) and (a_x, a_y) represent the *same* 2D vector $\boldsymbol{a}$. To express equality of the representations, we need the direction vectors to write

$$\boldsymbol{a} = a\,\hat{\boldsymbol{a}} = a_x\,\hat{\boldsymbol{x}} + a_y\,\hat{\boldsymbol{y}}, \tag{6}$$

yielding

$$\hat{\boldsymbol{a}} = \frac{a_x}{a}\hat{\boldsymbol{x}} + \frac{a_y}{a}\hat{\boldsymbol{y}} = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}} \tag{7}$$

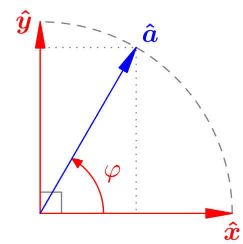


Fig. 1.5: $\hat{\boldsymbol{a}} = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}}$

as illustrated in fig. 1.5. Thus $\hat{\boldsymbol{a}}$ is a vector function of φ but independent of magnitude a, while the cartesian direction vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are independent of both a and φ, and thus independent of (a_x, a_y) as well. The representation $\boldsymbol{a} = a\,\hat{\boldsymbol{a}}$ with $\hat{\boldsymbol{a}}$ specified by the angle φ is the *polar* form of the 2D vector $\boldsymbol{a}$. The representation $\boldsymbol{a} = a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}}$ is the *cartesian* form, with $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ two arbitrary but fixed perpendicular directions.

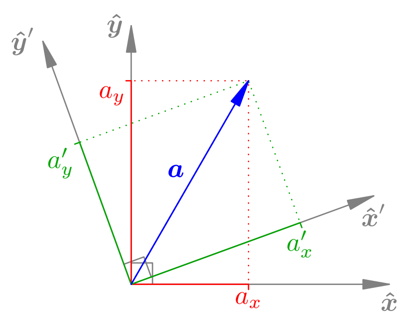


Fig. 1.6: Two cartesian decompositions of the same $\boldsymbol{a}$ but $(a_x, a_y) \neq (a_x', a_y')$.

The vector $\boldsymbol{a}$ is specified by the pair (a, φ) or (a_x, a_y), but the *same* physical 2D vector can be represented by *any* pair of numbers depending on how we choose our reference magnitudes and directions. This is the case even if we restrict to cartesian representations as shown in fig. 1.6 where the same vector $\boldsymbol{a}$ has distinct cartesian components $(a_x, a_y) \neq (a'_x, a'_y)$, since the reference directions $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ and $\{\hat{\boldsymbol{x}}', \hat{\boldsymbol{y}}'\}$ are distinct. To express equality, we need the direction vectors to write

$$\boldsymbol{a} = a_x\,\hat{\boldsymbol{x}} + a_y\,\hat{\boldsymbol{y}} = a'_x\,\hat{\boldsymbol{x}}' + a'_y\,\hat{\boldsymbol{y}}'. \tag{8}$$

Components are useless without the directions.

Notation. We use $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ to denote the x, y, z directions, respectively, instead of the old $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ notation used in many elementary calculus and physics books. The $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ notation originates from *quaternions*, a generalization of complex numbers that preceded the concept of vector, but it clashes with the very useful index notation that we use later in these notes. It also clashes with the complex plane where $\boldsymbol{i}$ is the y direction, not x!

1.3 Vectors in 3D space

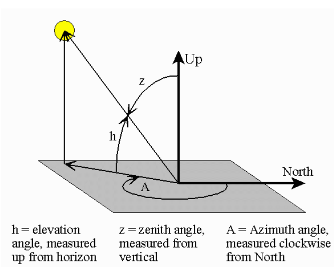


Fig. 1.7: Azimuth and Elevation angles in Astronomy

Direction in 3D space. How do you specify a direction $\hat{\boldsymbol{a}}$ in 3D space? Astronomers use an *azimuth* angle measured clockwise from North in the horizontal plane and an *inclination* (or *zenith*) angle measured from the vertical. An *elevation* angle measured up from the horizontal plane may be used instead of the inclination. Azimuth and elevation angles are also used in ballistics and computer graphics. Figure 1.7 is from NOAA, the National Ocean and Atmosphere Administration.

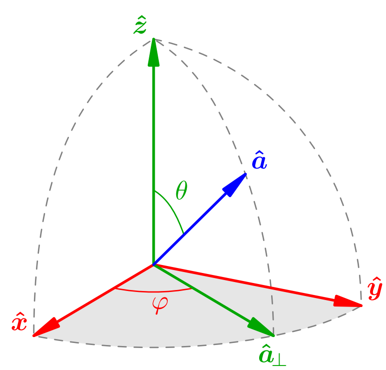


Fig. 1.8: Direction $\boldsymbol{a}$ in 3D specified by polar angle θ and azimuth φ.

The mathematical physics convention, illustrated in fig. 1.8, is to use the angle θ between a reference direction $\hat{\boldsymbol{z}}$ and the arbitrary direction $\hat{\boldsymbol{a}}$, as well as the angle φ *around* $\hat{\boldsymbol{z}}$, that is, the angle between the vertical planes $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{x}})$ and $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{a}})$. See for example Fig. 10-10 in Volume III of *The Feynman Lectures on Physics.*[2]

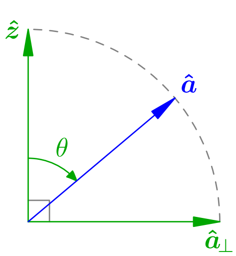


Fig. 1.9: Side view: $\hat{\boldsymbol{z}}$, $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{a}}_\perp$. Polar angle θ.

A little trigonometry in the vertical plane $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{a}})$ shown in fig. 1.9 yields the direction vector $\hat{\boldsymbol{a}}$ in terms of the vertical direction $\hat{\boldsymbol{z}}$ and a horizontal direction $\hat{\boldsymbol{a}}_\perp$

$$\hat{\boldsymbol{a}} = \cos\theta\;\hat{\boldsymbol{z}} + \sin\theta\;\hat{\boldsymbol{a}}_\perp. \tag{9}$$

[2] http://www.feynmanlectures.caltech.edu/III_10.html#Ch10-F10

Analysis of the horizontal plane $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ shown in fig. 1.10 then gives

$$\hat{\boldsymbol{a}}_\perp = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}}. \tag{10}$$

Substituting expression (10) for $\hat{\boldsymbol{a}}_\perp$ into (9) gives

$$\hat{\boldsymbol{a}} = \sin\theta\,\cos\varphi\,\hat{\boldsymbol{x}} + \sin\theta\,\sin\varphi\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}}, \tag{11}$$

that expresses an arbitrary direction $\hat{\boldsymbol{a}}$ in 3D space in terms of the mutually orthogonal cartesian directions $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ using two angles: a polar angle θ and an azimuth φ.

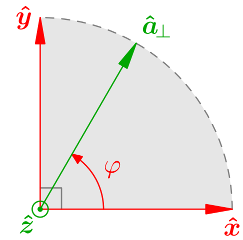


Fig. 1.10: Top view: $\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}}_\perp, \hat{\boldsymbol{y}}$. Azimuth φ.

Vectors in 3D space. An arbitrary vector $\boldsymbol{a} = a\hat{\boldsymbol{a}}$ in 3D can thus be specified by its magnitude a with its direction $\hat{\boldsymbol{a}}$ specified by the polar and azimuthal angles (θ, φ). That is the *spherical representation* $\boldsymbol{a} \equiv (a, \theta, \varphi)$. The *cylindrical* representation $(a_\perp, \varphi, a_z)$ specifies $\boldsymbol{a}$ by its horizontal magnitude $a_\perp = a\sin\theta$, azimuth φ and vertical component $a_z = a\cos\theta$. The *cartesian* representation consists of the familiar (a_x, a_y, a_z). A 3D vector $\boldsymbol{a}$ can thus be represented by 3 real numbers, for instance (a, θ, φ) or $(a_\perp, \varphi, a_z)$ or (a_x, a_y, a_z). Each of these triplets represent the same 3D vector $\boldsymbol{a}$ yet they are not equal to each other, in general,

$$(a, \theta, \varphi) \neq (a_\perp, \varphi, a_z) \neq (a_x, a_y, a_z). \tag{12}$$

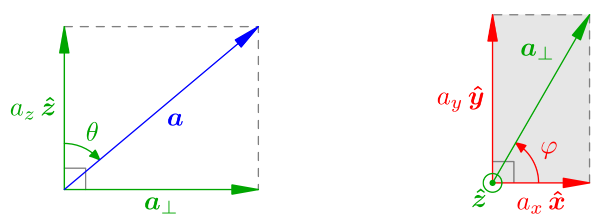


Fig. 1.11: Vertical and horizontal decompositions of vector $\boldsymbol{a} = a\hat{\boldsymbol{a}}$ in 3D.

To express equality, we need the direction vectors to write (fig. 1.11)

$$\boxed{\boldsymbol{a} = a\,\hat{\boldsymbol{a}} = a_\perp\hat{\boldsymbol{a}}_\perp + a_z\hat{\boldsymbol{z}} = a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}} + a_z\hat{\boldsymbol{z}}.} \tag{13}$$

From this vector equation (13), and with a little help from Pythagoras and basic trigonometry, we deduce (fig. 1.11)

$$a^2 = a_\perp^2 + a_z^2, \quad a_\perp^2 = a_x^2 + a_y^2, \tag{14}$$

and

$$\hat{\boldsymbol{a}} = \frac{a_\perp}{a}\,\hat{\boldsymbol{a}}_\perp + \frac{a_z}{a}\,\hat{\boldsymbol{z}} = \sin\theta\,\hat{\boldsymbol{a}}_\perp + \cos\theta\,\hat{\boldsymbol{z}}, \tag{15}$$

$$\hat{\boldsymbol{a}}_\perp = \frac{a_x}{a_\perp}\,\hat{\boldsymbol{x}} + \frac{a_y}{a_\perp}\,\hat{\boldsymbol{y}} = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}}. \tag{16}$$

Unique and positive angles can be specified by restricting them to the ranges $0 \le \varphi < 2\pi$ and $0 \le \theta \le \pi$, similar to the navigation convention where headings are specified as angles between 0 and 359° from North (in degrees). In mathematical physics, it is more common to use the definition $-\pi < \varphi \leq \pi$ together with $0 \leq \theta \leq \pi$, then

$$\varphi = \operatorname{atan2}(a_y, a_x), \qquad \theta = \operatorname{acos}(a_z/a) \tag{17}$$

where `atan2` is the arctangent function with range in $(-\pi, \pi]$ and `acos` is the arccosine function whose range is $[0, \pi]$.

1.4 Addition and scaling of vectors

Geometric vectors such as displacements and forces can be *added* and *scaled* to compose or decompose new vectors, as in eqn. (13), for instance.

Geometric vectors add according to the 'parallelogram rule' illustrated in Fig. 1.12, and addition is *commutative*, $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$. To any vector $\boldsymbol{b}$ we can associate an opposite vector denoted $-\boldsymbol{b}$ that is the vector with the same magnitude but reverse direction to $\boldsymbol{b}$. Vector subtraction $\boldsymbol{a} - \boldsymbol{b}$ is then defined as the addition of $\boldsymbol{a}$ and $-\boldsymbol{b}$. In particular $\boldsymbol{a} + (-\boldsymbol{a}) = \boldsymbol{0}$ corresponds to no net displacement. This is an important difference between points and displacements, there is no special point in our space, but there is one special displacement: the zero vector $\boldsymbol{0}$ such that $\boldsymbol{a} + (-\boldsymbol{a}) = \boldsymbol{0} = (-\boldsymbol{a}) + \boldsymbol{a}$ and $\boldsymbol{a} + \boldsymbol{0} = \boldsymbol{a}$, for any $\boldsymbol{a}$. Vector addition is also *associative*, $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$, as illustrated in Fig. 1.13.

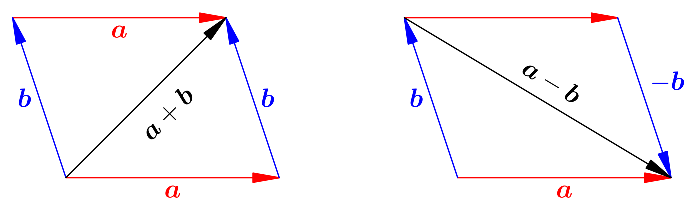


Fig. 1.12: Vector addition $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$ and subtraction $\boldsymbol{a} - \boldsymbol{b} = -\boldsymbol{b} + \boldsymbol{a}$. The vectors $\boldsymbol{a} + \boldsymbol{b}$ and $\boldsymbol{a} - \boldsymbol{b}$ are the diagonals of the parallelogram with sides $\boldsymbol{a}$ and $\boldsymbol{b}$.

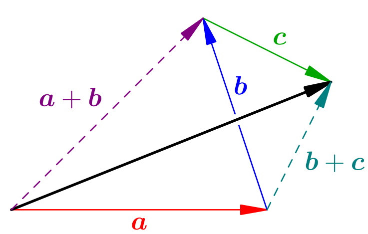


Fig. 1.13: Associativity $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$. Note that $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are *not* in the same plane, in general.

Geometric vectors can also be *scaled,* that is, multiplied by a real number $\alpha \in \mathbb{R}$, called a *scalar*. Geometrically, $\boldsymbol{v} = \alpha \boldsymbol{a}$ is a new vector parallel to $\boldsymbol{a}$ but of length $|\boldsymbol{v}| = |\alpha||\boldsymbol{a}|$. The direction of $\boldsymbol{v}$ is the same as $\boldsymbol{a}$ if $\alpha > 0$ and opposite to $\boldsymbol{a}$ if $\alpha < 0$. Obviously $(-1)\boldsymbol{a} = (-\boldsymbol{a})$, multiplying $\boldsymbol{a}$ by (-1) yields the previously defined opposite of $\boldsymbol{a}$. Other geometrically immediate properties are *distributivity with respect to addition of real factors*: $(\alpha+\beta)\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}$, and *with respect to multiplication of real factors:* $(\alpha\beta)\boldsymbol{a} = \alpha(\beta\boldsymbol{a})$. Slightly less trivial is *distributivity with respect to vector addition:* $\alpha(\boldsymbol{a}+\boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b}$, which geometrically corresponds to *similarity of triangles* (fig. 1.15).

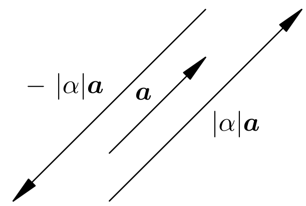


Fig. 1.14: Scaling $\boldsymbol{a}$ by $\alpha = \pm|\alpha|$

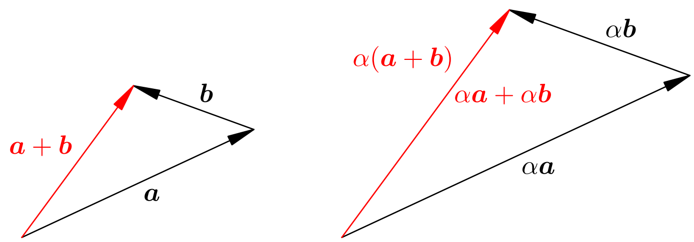


Fig. 1.15: Scaling is distributive with respect to addition: $\alpha(\boldsymbol{a}+\boldsymbol{b}) = (\alpha\boldsymbol{a}) + (\alpha\boldsymbol{b})$.

Exercises

1. What is the magnitude of a direction? the direction of a magnitude? an azimuth?

2. What are $|\boldsymbol{v}|$ and $\hat{\boldsymbol{v}}$ for the vector $\boldsymbol{v} \equiv -5$ miles per hour heading Northeast?

3. What is φ for the vector whose cartesian components are $(-1, \sqrt{3})$?

4. If you move 2 miles heading 030° then 3 miles heading 290°, how far are you from your original position? what heading from your original position? Sketch. Headings are clockwise from North.

5. If you move distance a heading α then distance b heading β, how far are you and in what heading from your original position? Draw a sketch and explain your *algorithm*. Headings are clockwise from North.

6. An airplane travels at airspeed V, heading α (clockwise from north). The wind has speed W heading β. Make a clean sketch. Show/explain the algorithm to calculate the airplane's ground speed and heading.

7. Write `cart2polar` and `polar2cart` codes (in Matlab or Python, for example, or pseudo-code) that transform Cartesian to Polar representations and Polar to Cartesian, respectively, for 2D vectors.

8. True or False: in SI units $|\hat{\boldsymbol{a}}|$ = 1 meter while in CGS units $|\hat{\boldsymbol{a}}| = 1$ cm.

9. True or False: $\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}$ is a unit vector in the northeast direction.

10. In astronomy, meteorology and computer graphics: what is an azimuth? What is an elevation? What is an inclination? Sketch, explain.

11. What are longitude and latitude? Sketch, explain. What are the longitude and latitude at the North pole? in Madison, WI? in London, UK? What are meridians and parallels?

12. Find φ and θ (fig. 1.8) for the vector $\boldsymbol{a}$ whose cartesian components are $(1, 1, 1)$. What is the angle between $\hat{\boldsymbol{x}}$ and $\boldsymbol{a}$?

13. Find the azimuthal and polar angles φ and θ (fig. 1.8) for the vector $\boldsymbol{a}$ given in cartesian form as $(-1, -1, -1)$.

14. A vector $\boldsymbol{v}$ is specified in cartesian components as $(-3, 2, 1)$. Find $\hat{\boldsymbol{v}}_\perp$ for that vector and express $\hat{\boldsymbol{v}}_\perp$ in terms of the cartesian direction vectors. Write $\boldsymbol{v}$ in cylindrical representation.

15. A vector is specified in cartesian coordinates as $(3, 2, 1)$. Find its magnitude and direction. Express its direction in cartesian, cylindrical and spherical representations using the *cartesian* direction vectors.

16. Write `cart2spher` and `spher2cart` codes (in Matlab or Python, for example, or pseudo-code) that transform Cartesian to Spherical representations and Spherical to Cartesian, respectively, for 3D vectors.

17. Sketch $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{a} + \boldsymbol{b}/2$ and $\boldsymbol{a} - \boldsymbol{b}/3$ for arbitrary $\boldsymbol{a}$ and $\boldsymbol{b}$.

18. If $\overrightarrow{AB} = \overrightarrow{CD}$, prove that $\overrightarrow{AC} = \overrightarrow{BD}$.

2 Vector Spaces

2.1 Abstract Vector Spaces

The concept of vector has evolved beyond geometric vectors. In its most general form, a **Vector Space** is defined as any set of mathematical objects $\boldsymbol{a}, \boldsymbol{b}, \ldots$ (such as displacements or ordered lists of numbers) for which addition $\boldsymbol{a}+\boldsymbol{b}$ and scalar multiplication $\alpha\boldsymbol{a}$ are defined and satisfy the following 8 properties or *axioms*. It is assumed that addition $\boldsymbol{a}+\boldsymbol{b}$ and scaling $\alpha\boldsymbol{a}$ yield objects in the same set as that of the objects $\boldsymbol{a}, \boldsymbol{b}$ that are added or scaled.

Vector addition must satisfy:

$$\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}, \tag{18}$$
$$\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})=(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}, \tag{19}$$
$$\boldsymbol{a}+\mathbf{0}=\boldsymbol{a}, \tag{20}$$
$$\boldsymbol{a}+(-\boldsymbol{a})=\mathbf{0}. \tag{21}$$

Scalar multiplication must satisfy:

$$(\alpha+\beta)\boldsymbol{a}=\alpha\boldsymbol{a}+\beta\boldsymbol{a}, \tag{22}$$
$$(\alpha\beta)\boldsymbol{a}=\alpha(\beta\boldsymbol{a}), \tag{23}$$
$$\alpha(\boldsymbol{a}+\boldsymbol{b})=\alpha\boldsymbol{a}+\alpha\boldsymbol{b}, \tag{24}$$
$$1\ \boldsymbol{a}=\boldsymbol{a}. \tag{25}$$

2.2 The vector space $\mathbb{R}^n$

Consider the set of ordered n-tuplets of real numbers $\mathbf{x} \equiv (x_1, x_2, \ldots, x_n)$. These could correspond to student grades on a particular exam, for instance. What kind of operations would we want to do on these lists of student grades? We'll probably want to *add* several grades for *each student* and we'll probably want to *rescale* the grades. So the natural operations on these n-tuplets are *addition* defined[3] by adding the respective components:

$$\mathbf{x}+\mathbf{y} \triangleq (x_1+y_1, x_2+y_2, \ldots, x_n+y_n) = \mathbf{y}+\mathbf{x}. \tag{26}$$

and multiplication by a real number[4] $\alpha \in \mathbb{R}$ defined as

$$\alpha\mathbf{x} \triangleq (\alpha x_1, \alpha x_2, \ldots, \alpha x_n). \tag{27}$$

The set of n-tuplets of real numbers equipped with addition and multiplication by a real number as just defined is a fundamental vector space called $\mathbb{R}^n$. The vector

[3]The symbol $\triangleq$ means "equal by definition".
[4]The symbol $\in$ means *in* or *belonging to*.

spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ are particularly important to us as they will soon correspond to the components of our physical vectors. But we also use $\mathbb{R}^n$ for very large n when studying systems of equations, for instance. The vector space $\mathbb{C}^n$, that consists of ordered lists of n complex numbers, is also useful in applications.

2.3 Bases and Components

Addition and scaling of vectors allow us to define the concepts of *linear combination*, *linear (in)dependence*, *dimension*, *basis* and *components*. These concepts apply to *any* vector space.

A *linear combination* of k vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k\}$ is an expression of the form

$$\chi_1 \boldsymbol{a}_1 + \chi_2 \boldsymbol{a}_2 + \cdots + \chi_k \boldsymbol{a}_k$$

where χ_1, χ_2, ..., χ_k are arbitrary real numbers. The k vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k\}$ are *linearly independent* if

$$\chi_1 \boldsymbol{a}_1 + \chi_2 \boldsymbol{a}_2 + \cdots + \chi_k \boldsymbol{a}_k = 0 \quad \Leftrightarrow \quad \chi_1 = \chi_2 = \cdots = \chi_k = 0.$$

Otherwise, the vectors are linearly dependent. For instance if $3\boldsymbol{a}_1 + 2\boldsymbol{a}_2 + \boldsymbol{a}_3 = 0$, then $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are linearly *dependent*. The *dimension* of a vector space is the largest number of linearly independent vectors, n say, in that space.

A *basis* for an n dimensional vector space is *any* collection of n linearly independent vectors. If $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\}$ is a basis for an n dimensional vector space, then any vector $\boldsymbol{v}$ in that space can be *expanded* (or *decomposed*) as

$$\boldsymbol{v} = v_1 \boldsymbol{a}_1 + v_2 \boldsymbol{a}_3 + \cdots + v_n \boldsymbol{a}_n.$$

The n real numbers $(v_1, v_2, \dots, v_n)$ are the *(scalar) components* of $\boldsymbol{v}$ in the basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\}$. The expansion of $\boldsymbol{v}$ is unique for that basis as the reader is asked to show in the exercises.

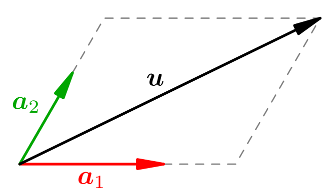


Fig. 1.16: Decomposing $\boldsymbol{u}$ into $\boldsymbol{u} = u_1 \boldsymbol{a}_1 + u_2 \boldsymbol{a}_2$ by projections parallel to $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in 2D.

Basis in a plane. Any two *non-parallel* vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in a plane (a horizontal plane, a vertical plane, or any slanted plane) are linearly independent and form a **basis** for vectors (*e.g.* displacements or forces) in that plane. Any given vector $\boldsymbol{u}$ in the plane can be expanded as $\boldsymbol{u} = u_1 \boldsymbol{a}_1 + u_2 \boldsymbol{a}_2$, for a unique pair (u_1, u_2). Geometrically, the components are obtained by *projections parallel* to each of the basis vectors (fig. 1.16). Three or more vectors in a plane are necessarily *linearly dependent*. Vectors in a plane form a 2-dimensional vector space.

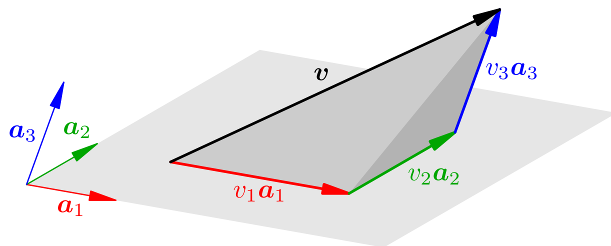


Fig. 1.17: Expanding $\boldsymbol{v}$ in terms of an arbitrary basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ in 3D.

Basis in 3D space. Any three *non-coplanar* vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ in 3D space are linearly independent and those vectors form a basis for 3D space. Any given vector $\boldsymbol{v}$ can be expanded as $\boldsymbol{v} = v_1\boldsymbol{a}_1 + v_2\boldsymbol{a}_2 + v_3\boldsymbol{a}_3$, for a unique triplet of real numbers (v_1, v_2, v_3). Geometrically, the components are obtained by *projections parallel* to each of the basis vectors as illustrated in fig. 1.17. Thus, any four or more vectors in 3D are necessarily *linearly dependent.* If the vectors are known in terms of cartesian components, we can calculate the components using *dot* and *cross* products without the need for geometrical constructions, as discussed in later sections.

The 8 properties of addition and scalar multiplication imply that if two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are expanded with respect to the *same* basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$, that is if

$$\begin{aligned} \boldsymbol{u} &= u_1\boldsymbol{a}_1 + u_2\boldsymbol{a}_2 + u_3\boldsymbol{a}_3, \\ \boldsymbol{v} &= v_1\boldsymbol{a}_1 + v_2\boldsymbol{a}_2 + v_3\boldsymbol{a}_3, \end{aligned}$$

then

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1)\boldsymbol{a}_1 + (u_2 + v_2)\boldsymbol{a}_2 + (u_3 + v_3)\boldsymbol{a}_3,$$

$$\alpha\boldsymbol{v} = (\alpha v_1)\boldsymbol{a}_1 + (\alpha v_2)\boldsymbol{a}_2 + (\alpha v_3)\boldsymbol{a}_3,$$

thus addition and scalar multiplication are performed component by component and the triplets of real components (u_1, u_2, u_3) and (v_1, v_2, v_3) are elements of the vector space $\mathbb{R}^3$. A basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ in 3D (Euclidean) space, call it $\mathbb{E}^3$, provides a one-to-one correspondence (mapping) between geometric vectors $\boldsymbol{v}$ in $\mathbb{E}^3$ and algebraic vectors (v_1, v_2, v_3) in $\mathbb{R}^3$

$$\text{Basis} \in \mathbb{E}^3 \Rightarrow \boldsymbol{v} \in \mathbb{E}^3 \longleftrightarrow (v_1, v_2, v_3) \in \mathbb{R}^3.$$

Exercises

1. Verify that addition and scalar multiplication of n-tuplets as defined in (26) and (27) satisfy the 8 required vector properties.

2. Show that the set of polynomials $P(x)$ of degree at most n is a vector space, where $x \in \mathbb{R}$. What is the dimension of that vector space? What is a basis for that vector space?

3. Suppose you define addition of n-tuplets $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ as usual but define scalar multiplication according to $\alpha\mathbf{x} = (\alpha x_1, x_2, \cdots, x_n)$, that is, only the first component is multiplied by α. Which property is violated? What if you defined $\alpha\mathbf{x} = (\alpha x_1, 0, \cdots, 0)$, which property would be violated?

4. From the 8 properties, show that $(0)\boldsymbol{a} = \mathbf{0}$ and $(-1)\boldsymbol{a} = (-\boldsymbol{a})$, $\forall\boldsymbol{a}$, *i.e.* show that multiplication by the scalar 0 yields the neutral element for addition, and multiplication by -1 yields the additive inverse.

5. Given vectors $\boldsymbol{a}$, $\boldsymbol{b}$ in $\mathbb{E}^3$, show that the set of all $\boldsymbol{v} = \alpha\boldsymbol{a} + \beta\boldsymbol{b}$, $\forall\alpha, \beta \in \mathbb{R}$ is a vector space. What is the dimension of that vector space?

$\forall$ means *for all* or *for any*.

6. Show that the set of all vectors $\boldsymbol{v} = \alpha\boldsymbol{a} + \boldsymbol{b}$, $\forall\alpha \in \mathbb{R}$ and fixed $\boldsymbol{a}$, $\boldsymbol{b}$ is *not* a vector space, in general.

7. Consider the set of all unit vectors in $\mathbb{E}^3$. Is that a vector space? Explain.

8. Prove that if $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ and $\{\mathbf{b}_1, \ldots, \mathbf{b}_l\}$ are two distinct bases for the same vector space, then $k = l$. [Hint: start with $k = 2$, $l = 3$ and express $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ as linear combinations of $\mathbf{a}_1, \mathbf{a}_2$. Show that this leads to a contradiction since $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are assumed to be linearly independent.]

9. Prove that the components of any $\mathbf{v}$ with respect to a basis $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ are unique. [Hint: assume that $\mathbf{v}$ can be expanded in two distinct ways, then use what you know about $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$]

10. Find the components of *North* in the bases (a) *East*, *West*, (b) *East*, *Northeast*. Sketch and briefly explain your *geometrical* construction.

11. Given three points P_1, P_2, P_3 in Euclidean 3D space, let M be the midpoint of segment P_1P_2, what are the components of $\overrightarrow{P_2P_3}$ and $\overrightarrow{P_3M}$ in the basis $\overrightarrow{P_1P_2}$, $\overrightarrow{P_1P_3}$? Sketch.

12. Show that $\mathbf{e}_1 = (1, 0, \cdots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \cdots, 0)$, etc. is a basis for $\mathbb{R}^n$. It is called the *natural basis* for $\mathbb{R}^n$.

13. True or False: $\hat{\boldsymbol{x}} = (1, 0, 0)$. Discuss.

14. Show that the set of smooth functions $f(x)$ periodic of period 2π is a vector space. What is the dimension of that space? What is a basis for that space?

3 Points and Coordinates

3.1 Position vector

In elementary calculus and linear algebra it is easy to confuse points and vectors. In $\mathbb{R}^3$ for instance, a point P is defined as a triplet (x_1, x_2, x_3) but a vector $\boldsymbol{a}$ is also defined by a real number triplet (a_1, a_2, a_3), yet in physical space, points and displacements are two clearly different things.

The confusion arises from the fundamental way to locate points by specifying displacements from a *reference point* called the *origin* and denoted O. An arbitrary point P is then specified by providing the displacement vector $\boldsymbol{r} = \overrightarrow{OP}$. That vector is called the *position vector* of P and denoted $\boldsymbol{r}$ for *radial* vector from the origin O.

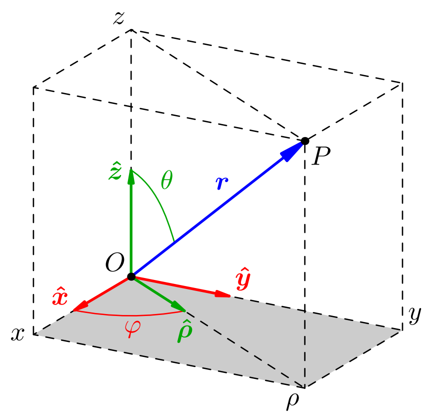


Fig. 1.18: Position vector $\boldsymbol{r} = \overrightarrow{OP}$ in spherical, cylindrical and cartesian coordinates. Mathematical physics convention, θ is the angle between $\hat{\boldsymbol{z}}$ and $\boldsymbol{r}$.

A *Cartesian system of coordinates* consists of a reference point O and three mutually orthogonal directions $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ that provide a basis for displacements in 3D Euclidean space $\mathbb{E}^3$. The position vector $\boldsymbol{r} = \overrightarrow{OP}$ of point P can then be specified in spherical, cylindrical or cartesian form as in sect. 1.3 now for the position vector $\boldsymbol{r}$ instead of the arbitrary vector $\boldsymbol{a}$,

$$\boxed{\overrightarrow{OP} = \boldsymbol{r} = r\,\hat{\boldsymbol{r}} = \rho\,\hat{\boldsymbol{\rho}} + z\hat{\boldsymbol{z}} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}.} \tag{28}$$

It follows that

$$\rho\hat{\boldsymbol{\rho}} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} \tag{29}$$

and

$$\rho = \sqrt{x^2 + y^2} \quad \Leftrightarrow \quad \begin{cases} x = \rho\cos\varphi \\ y = \rho\sin\varphi \end{cases} \tag{30}$$

$$r = \sqrt{\rho^2 + z^2} \quad \Leftrightarrow \quad \begin{cases} \rho = r\sin\theta \\ z = r\cos\theta \end{cases} \tag{31}$$

and we can eliminate ρ to obtain

$$r = \sqrt{x^2 + y^2 + z^2} \quad \Leftrightarrow \quad \begin{cases} x = r\sin\theta\cos\varphi, \\ y = r\sin\theta\sin\varphi, \\ z = r\cos\theta. \end{cases} \tag{32}$$

We can also deduce expressions for the direction vectors $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{r}}$ in terms of the cartesian directions $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$

$$\hat{\boldsymbol{\rho}} = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}}, \qquad \hat{\boldsymbol{r}} = \sin\theta\,\hat{\boldsymbol{\rho}} + \cos\theta\,\hat{\boldsymbol{z}}. \tag{33}$$

We refer to these as *hybrid* representations using spherical coordinates (θ, φ) with cartesian direction vectors, to distinguish them from the full cartesian expressions

$$\hat{\boldsymbol{\rho}} = \frac{x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}}{\sqrt{x^2+y^2}}, \qquad \hat{\boldsymbol{r}} = \frac{x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\,\hat{\boldsymbol{z}}}{\sqrt{x^2+y^2+z^2}}. \tag{34}$$

Note that θ is the angle between $\hat{\boldsymbol{z}}$ and $\boldsymbol{r}$ (called the *polar* or *zenith* or *inclination* angle, depending on the context) while φ is the azimuthal angle *about* the z axis, the angle between the $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{x}})$ and the $(\hat{\boldsymbol{z}}, \boldsymbol{r})$ planes. The distance to the origin is r while ρ is the distance to the z-axis. This is the physics convention in ISO 80000-2 (International Standards Organization) that has been used for many decades in mathematical physics. See for example Fig. 10-10 in Volume III of *The Feynman Lectures on Physics*.[5] American calculus teachers often reverse the definitions of θ and φ, confusing applied mathematics, engineering and physics students.

The unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ form a basis for 3D Euclidean vector space, but $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{z}}$ do not, and $\hat{\boldsymbol{r}}$ does not either. The cartesian basis $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ consists of three fixed and mutually orthogonal directions independent of P, but $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{r}}$ depend on P, each point has its own $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{r}}$. We will construct and use full cylindrical and spherical orthogonal bases, $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{z}})$ and $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varphi}})$ later in vector calculus. These cylindrical and spherical basis vectors vary with P, or more precisely with φ and θ.

Once a Cartesian system of coordinates, $O, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$, has been chosen, the cartesian *coordinates* (x, y, z) of P are the cartesian *components* of $\boldsymbol{r} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$. The cylindrical coordinates of P are (ρ, φ, z) and its *spherical coordinates* are (r, θ, φ), but cylindrical and spherical coordinates are *not* vector components, they are the cylindrical and spherical representations of $\boldsymbol{r}$.

3.2 Curvilinear coordinates

Coordinates can be specified in many other ways that do not correspond to a displacement vector. In 2D for instance, point P can be located by specifying the angles θ_1 and θ_2 between the vector $\overrightarrow{F_1F_2}$ and $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$, respectively, where F_1 and F_2 are two reference points (the foci). In navigation, F_1 and F_2 would be lighthouses or radio beacons. Alternatively, one could specify P by specifying the *distances* $r_1 = |F_1P|$ and $r_2 = |F_2P|$ (as in the global positioning system (GPS) that measures distance from satellites in 3D space). *Bipolar coordinates* specify P through the angle α between $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$ and the natural log of the distance ratio $\ln(r_1/r_2)$. Bipolar coordinates arise in various areas of physics that lead to Laplace's equation, including electrostatics and aerodynamics. Thus, in general, *coordinates* of points are not necessarily the *components* of vectors. We will study 'generalized' or 'curvilinear' coordinates in vector calculus.

[5] http://www.feynmanlectures.caltech.edu/III_10.html#Ch10-F10

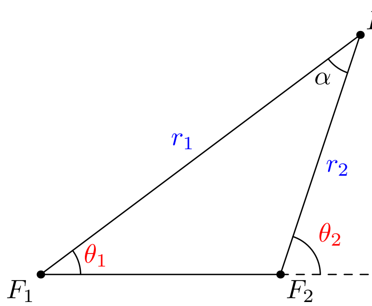


Fig. 1.19: *Beyond* (x, y)*:* A point P in a plane can be specified using *biangular* coordinates (θ_1, θ_2), or *biradial* coordinates (r_1, r_2) or *bipolar* coordinates $(\ln(r_1/r_2), \alpha)$. The latter bipolar coordinates occur in electrostatics and aerodynamics in the definition of source panels.

3.3 Lines and Planes

Lines. The line through point A that is parallel to the vector $\boldsymbol{a}$ consists of all points P such that

$$\overrightarrow{AP} = t\,\boldsymbol{a}, \quad \forall t \in \mathbb{R}. \tag{35}$$

This vector equation expresses that the vector $\overrightarrow{AP}$ is parallel to $\boldsymbol{a}$. In terms of an origin O we have $\overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{AP}$, that is

$$\boldsymbol{r} = \boldsymbol{r}_A + t\,\boldsymbol{a}, \tag{36}$$

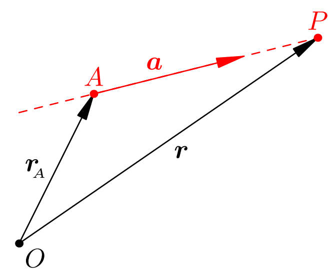


Fig. 1.20: Points on a line passing through A, parallel to $\boldsymbol{a}$

where $\boldsymbol{r} = \overrightarrow{OP}$ and $\boldsymbol{r}_A = \overrightarrow{OA}$ are the position vectors of P and A with respect to O, respectively. The real number t is the parameter of the line, it is the coordinate of P in the system of coordinates $(A, \boldsymbol{a})$ specified by the reference point A and the reference direction $\boldsymbol{a}$. In cartesian coordinates, $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$, $\boldsymbol{r}_A = x_A\hat{\boldsymbol{x}} + y_A\hat{\boldsymbol{y}} + z_A\hat{\boldsymbol{z}}$ and $\boldsymbol{a} = a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}} + a_z\hat{\boldsymbol{z}}$, thus the vector equation (36) corresponds to the 3 scalar equations

$$x = x_A + t\,a_x, \quad y = y_A + t\,a_y, \quad z = z_A + t\,a_z,$$

that provide a map from $t \in \mathbb{R} \to (x, y, z) \in \mathbb{R}^3$.

Planes. Likewise the equation of a plane passing through A and parallel to the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ consists of all points P such that

$$\overrightarrow{AP} = s\,\boldsymbol{a} + t\,\boldsymbol{b}, \quad \forall s, t \in \mathbb{R} \tag{37}$$

or with respect to the origin O:

$$\boldsymbol{r} = \boldsymbol{r}_A + s\,\boldsymbol{a} + t\,\boldsymbol{b}. \tag{38}$$

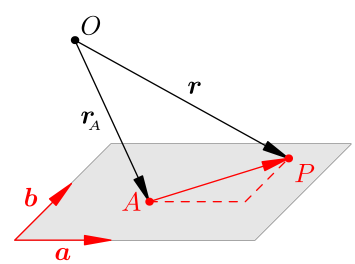


Fig. 1.21: Points on a plane passing through A, parallel to $\boldsymbol{a}$ and $\boldsymbol{b}$

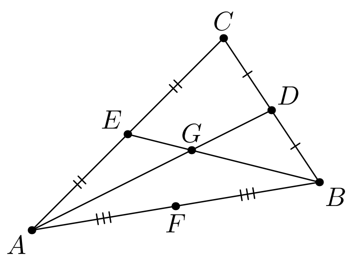
This is the parametric vector equation of that plane with parameters s, t, that are the coordinates of P in the system of coordinates specified by A, $\boldsymbol{a}$, $\boldsymbol{b}$. In cartesian coordinates, the vector equation (38) yields the 3 scalar equations

$$x = x_A + s\,a_x + t\,b_x, \quad y = y_A + s\,a_y + t\,b_y, \quad z = z_A + s\,a_z + t\,b_z,$$

that provide a map from $(s,t) \in \mathbb{R}^2 \to (x,y,z) \in \mathbb{R}^3$.

Notation. We use $\boldsymbol{r}$ for the position vector of an arbitrary *point* P with cartesian coordinates (x,y,z). We use $\boldsymbol{r}_A$ for the position vector of a specific point A with cartesian coordinates (x_A, y_A, z_A) and (a_x, a_y, a_z) for the cartesian *components* of vector $\boldsymbol{a}$, to help distinguish between points and vectors.

3.4 Medians of a triangle



As an example of these vector concepts and methods we solve the following problem: *Prove that the medians of any triangle are concurrent.*

A median of a triangle is a line that passes through a vertex and the midpoint of the opposite side. Thus if D is the midpoint of segment BC then the line through A and D is a median. Any point P on that median is such that

$$\overrightarrow{AP} = t_1\,\overrightarrow{AD} = t_1\left(\overrightarrow{AB} + \frac{1}{2}\overrightarrow{BC}\right), \tag{39}$$

for some real number t_1. This is the vector equation of the line through A and D. Likewise, if E is the midpoint of segment AC then any point P on the median through B is such that

$$\overrightarrow{BP} = t_2\,\overrightarrow{BE} = t_2\left(\overrightarrow{BA} + \frac{1}{2}\overrightarrow{AC}\right), \tag{40}$$

for some real number t_2. That is the vector equation of the line through B and E.

It is *self-evident* that those two medians intersect at a point that we call G. It is not self-evident however that G lies on the third median through C and F, the midpoint of AB. We better prove that.

To avoid going around in triangles and reusing the same information over and over in different forms, we select a unique *intrinsic system of coordinates.* That is, we select a reference point and a vector basis that *belong* to the problem at hand. Since the problem is inherently two-dimensional even if A, B, C are points in 3D space (or even in $\mathbb{R}^n$), we can specify any point in the plane A, B, C in terms of a reference point in that plane, A say, and *two* basis vectors, $\overrightarrow{AB}$ and $\overrightarrow{BC}$ say.

Median (39) is already expressed in terms of point A with $\overrightarrow{AB}$ and $\overrightarrow{BC}$ as basis vectors. Median (40) uses B as reference point and $\overrightarrow{BA}$ and $\overrightarrow{AC}$ as basis vectors, but it is easy to rewrite (40) in terms of A, $\overrightarrow{AB}$ and $\overrightarrow{BC}$ as

$$\overrightarrow{AP} = \overrightarrow{AB} + \overrightarrow{BP} = (1 - \frac{t_2}{2})\overrightarrow{AB} + \frac{t_2}{2}\overrightarrow{BC}. \tag{41}$$

Now G is the point that is on both medians (39) and (41), therefore

$$\overrightarrow{AG} = t_1\left(\overrightarrow{AB} + \frac{1}{2}\overrightarrow{BC}\right) = (1 - \frac{t_2}{2})\overrightarrow{AB} + \frac{t_2}{2}\overrightarrow{BC}, \tag{42}$$

for some t_1 and t_2. This is a vector equation for t_1 and t_2. Since $\overrightarrow{AB}$ and $\overrightarrow{BC}$ are *linearly independent*, the coefficients of $\overrightarrow{AB}$ must match on both sides of the equation, requiring $t_1 = 1 - t_2/2$. The coefficients of $\overrightarrow{BC}$ must match also, giving $t_1 = t_2$. This yields the two equations

$$t_1 = 1 - \frac{t_2}{2} = t_2$$

whose solutions are

$$t_1 = t_2 = \frac{2}{3}. \tag{43}$$

Thus $\overrightarrow{AG} = \frac{2}{3}\overrightarrow{AD}$ and $\overrightarrow{BG} = \frac{2}{3}\overrightarrow{BE}$, that is, G is 2/3 down the medians from the vertices.

If we follow the same reasoning starting from the medians AD and CF we will find the same intersection point G, with $\overrightarrow{AG} = \frac{2}{3}\overrightarrow{AD}$, and the three medians intersect at G. Another way to verify that result is to show that $\overrightarrow{CG}$ is parallel to $\overrightarrow{CF}$. In fact

$$\begin{aligned}\overrightarrow{CG} &= \overrightarrow{CB} + \overrightarrow{BA} + \overrightarrow{AG} = \overrightarrow{CB} + \overrightarrow{BA} + \frac{2}{3}\left(\overrightarrow{AB} + \frac{1}{2}\overrightarrow{BC}\right)\\ &= \frac{2}{3}\left(\overrightarrow{CB} + \frac{1}{2}\overrightarrow{BA}\right) = \frac{2}{3}\,\overrightarrow{CF}.\end{aligned}$$

Exercises

1. Pick two generic vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and some arbitrary point A in the plane of your sheet of paper. If $\overrightarrow{AB} = \alpha\boldsymbol{a} + \beta\boldsymbol{b}$, sketch the region where B can be if: (i) α and β are both between 0 and 1, (ii) $|\beta| \le |\alpha| \le 1$.

2. Let $\boldsymbol{r}_A$ and $\boldsymbol{r}_B$ be the position vectors of points A and B, respectively and consider all points $\boldsymbol{r} = (\alpha\boldsymbol{r}_A + \beta\boldsymbol{r}_B)/(\alpha+\beta)$ for all real α and β with $\alpha+\beta \neq 0$. Do these points lie along a line or a plane?

3. Given three points P_1, P_2, P_3 in $\mathbb{E}^3$, what are the coordinates of the midpoints of the triangle sides with respect to P_1 and the basis $\overrightarrow{P_1P_2}$, $\overrightarrow{P_1P_3}$?

4. Sketch the *coordinate curves* — the curves along which the coordinates are constant — in 2D for (1) cartesian coordinates (x, y), (2) polar coordinates (ρ, φ) (often denoted (r, θ) in 2D), (3) Biangular coordinates (θ_1, θ_2), (4) Biradial coordinates (r_1, r_2), (5) Bipolar coordinates $(\alpha, \ln(r_1/r_2))$.

5. Show that the line segment connecting the midpoints of two sides of a triangle is parallel to and equal to half of the third side using (1) high school geometry, (2) vectors.

6. Show that the diagonals of a parallelogram intersect at their midpoints using (1) high school geometry, (2) vectors.

7. Show that the medians of a triangle intersect at the same point, the *centroid* G, which is 2/3 of the way down from the vertices along each median (a median is a line that connects a vertex to the middle of the opposite side). Do this in two ways: (1) using high school geometry and (2) using vector methods.

8. Consider the point A whose position vector $\boldsymbol{r}_A = \overrightarrow{OA}$ is the average of the position vectors of three arbitrary points P_1, P_2, P_3 in 2D or 3D Euclidean space, $\boldsymbol{r}_A \triangleq (\boldsymbol{r}_1 + \boldsymbol{r}_2 + \boldsymbol{r}_3)/3$. Show that A is independent of the origin O and that A lies on each of the medians of the triangle P_1, P_2, P_3.

9. Find a point X such that $\overrightarrow{XA} + \overrightarrow{XB} + \overrightarrow{XC} = 0$, where A, B, C are given but arbitrary points in Euclidean space. If A, B, C are not co-linear, show that the line through A and X cuts BC at its mid-point. Deduce similar results for the other sides of the triangle ABC and therefore that X is the point of intersection of the medians. Sketch. [Hint: since X is unknown, the vector equation has three unknown vectors, however, any two of those can be expressed in terms of the third.]

10. From Martin Isaacs' *Geometry for College Students*: Given an arbitrary triangle with vertices A, B, C, consider A' at $1/3$ from A along edge AB, and likewise for B' along BC, and C' along CA. Similarly, consider A'' at $1/3$ from A' along $A'B'$, and likewise for B'' along $B'C'$, and C'' along $C'A'$. Sketch. Use vectors to prove that the triangle $A''B''C''$ is similar to the original triangle ABC.

11. Find a point X such that $\overrightarrow{XA} + \overrightarrow{XB} + \overrightarrow{XC} + \overrightarrow{XD} = 0$, where A, B, C, D are given but arbitrary points in Euclidean space. If A, B, C, D are not co-planar, show that the line through A and X intersects the triangle BCD at its centroid. Deduce similar results for the other faces and therefore that the medians of the tetrahedron $ABCD$, defined as the lines joining each vertex to the centroid of the opposite triangle, all intersect at the same point X which is 3/4 of the way down from the vertices along the medians. Visualize. [Hint: solve # 9 first.]

12. Given four points A, B, C, D not co-planar, let G be the intersection of the medians of triangle A, B, C. Find the point X that is the intersection of the plane passing through A and the midpoints of BD and CD and the line through D and G. Sketch. How far down DG is X? Show your work.

4 Dot Product

4.1 Geometry and algebra of dot product

The geometric definition of the dot product of vectors in 3D Euclidean space is

$$\boxed{\boldsymbol{a} \cdot \boldsymbol{b} \triangleq |\boldsymbol{a}|\,|\boldsymbol{b}|\,\cos\theta,} \tag{44}$$

where θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, with $0 \le \theta \le \pi$. The dot product is a real number such that $\boldsymbol{a} \cdot \boldsymbol{b} = 0$ iff $\boldsymbol{a}$ and $\boldsymbol{b}$ are *orthogonal*, that is when $\theta = \pi/2$ if $|\boldsymbol{a}|$ and $|\boldsymbol{b}|$ are not zero. The $\mathbf{0}$ vector is considered orthogonal to any vector. The dot product of any vector with itself is the square of its magnitude

$$\boldsymbol{a} \cdot \boldsymbol{a} = |\boldsymbol{a}|^2. \tag{45}$$

The dot product is also called the *scalar product* since its result is a scalar, or the *inner product* in linear algebra (where the fundamental data structure is the 'matrix' and the dot product is done over the 'inner indices').

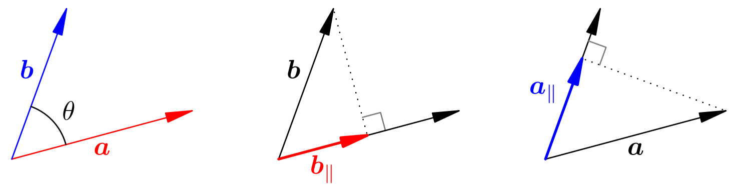


Fig. 1.22: Orthogonal projections $\boldsymbol{b}_\parallel = (\boldsymbol{b}\cdot\hat{\boldsymbol{a}})\hat{\boldsymbol{a}}$ and $\boldsymbol{a}_\parallel = (\boldsymbol{a}\cdot\hat{\boldsymbol{b}})\hat{\boldsymbol{b}}$.

The dot product is directly related to the *orthogonal projections* of $\boldsymbol{b}$ onto $\boldsymbol{a}$ and $\boldsymbol{a}$ onto $\boldsymbol{b}$. The latter are, respectively,

$$\begin{aligned} \boldsymbol{b}_\parallel &= b\cos\theta\,\hat{\boldsymbol{a}} = (\boldsymbol{b}\cdot\hat{\boldsymbol{a}})\,\hat{\boldsymbol{a}}, \\ \boldsymbol{a}_\parallel &= a\cos\theta\,\hat{\boldsymbol{b}} = (\boldsymbol{a}\cdot\hat{\boldsymbol{b}})\,\hat{\boldsymbol{b}}, \end{aligned} \tag{46}$$

where $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ are the *unit* vectors in the $\boldsymbol{a}$ and $\boldsymbol{b}$ directions, respectively. The orthogonal projections are not equal in general $\boldsymbol{a}_\parallel \neq \boldsymbol{b}_\parallel$, they have different magnitudes and different directions, but the dot product has the fundamental property that

$$\boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{a}\cdot\boldsymbol{b}_\parallel = \boldsymbol{a}_\parallel\cdot\boldsymbol{b}. \tag{47}$$

In physics, the work W done by a force $\boldsymbol{F}$ on a particle undergoing the displacement $\boldsymbol{\ell}$ is equal to distance ℓ times $F_\parallel = \boldsymbol{F}\cdot\hat{\boldsymbol{\ell}}$, but that is equal to the total force F times $\ell_\parallel = \boldsymbol{\ell}\cdot\hat{\boldsymbol{F}}$,

$$W = F_\parallel \ell = F\ell_\parallel = \boldsymbol{F}\cdot\boldsymbol{\ell}.$$

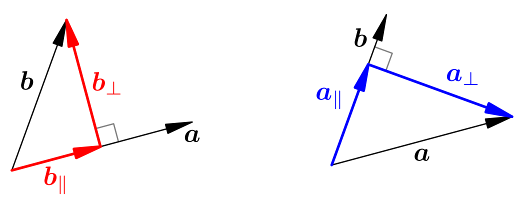


Fig. 1.23: Parallel and perpendicular components

It is often useful to decompose a vector $\boldsymbol{b}$ into *vector* components, $\boldsymbol{b}_\parallel$ and $\boldsymbol{b}_\perp$, parallel and perpendicular to a vector $\boldsymbol{a}$, respectively, such that $\boldsymbol{b} = \boldsymbol{b}_\parallel + \boldsymbol{b}_\perp$ with

$$\boldsymbol{b}_\parallel = (\boldsymbol{b}\cdot\hat{\boldsymbol{a}})\,\hat{\boldsymbol{a}}, \qquad \boldsymbol{b}_\perp = \boldsymbol{b} - \boldsymbol{b}_\parallel. \tag{48}$$

Likewise

$$\boldsymbol{a}_\parallel = (\boldsymbol{a}\cdot\hat{\boldsymbol{b}})\,\hat{\boldsymbol{b}}, \qquad \boldsymbol{a}_\perp = \boldsymbol{a} - \boldsymbol{a}_\parallel, \tag{49}$$

such that $\boldsymbol{a} = \boldsymbol{a}_\parallel + \boldsymbol{a}_\perp$ is the decomposition of $\boldsymbol{a}$ into vector components parallel and perpendicular to $\boldsymbol{b}$, respectively.

Properties of the dot product

The dot product has the following properties:

1. $\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{b}\cdot\boldsymbol{a}$, *(commutativity)*
2. $\boldsymbol{a}\cdot\boldsymbol{a}\geq 0, \quad \boldsymbol{a}\cdot\boldsymbol{a}=0 \Leftrightarrow \boldsymbol{a}=\boldsymbol{0}$, *(positive definiteness)*
3. $(\boldsymbol{a}\cdot\boldsymbol{b})^2\leq(\boldsymbol{a}\cdot\boldsymbol{a})\,(\boldsymbol{b}\cdot\boldsymbol{b})$ *(Cauchy-Schwarz)*
4. $(\alpha\boldsymbol{a}+\beta\boldsymbol{b})\cdot\boldsymbol{c}=\alpha(\boldsymbol{a}\cdot\boldsymbol{c})+\beta(\boldsymbol{b}\cdot\boldsymbol{c})$ *(distributivity)*

The first three properties follow directly from the geometric definition (44).

Proof of distributivity. Distributivity with respect to scalar multiplication, that is $(\alpha\boldsymbol{a})\cdot\boldsymbol{c}=\alpha(\boldsymbol{a}\cdot\boldsymbol{c})$, is left as an exercise. Let $\boldsymbol{v}\triangleq\alpha\boldsymbol{a}+\beta\boldsymbol{b}$ and decompose all three vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{v}$ into components parallel and perpendicular to $\boldsymbol{c}$, that is

$$\boldsymbol{v}=\boldsymbol{v}_{\parallel}+\boldsymbol{v}_{\perp}=\alpha(\boldsymbol{a}_{\parallel}+\boldsymbol{a}_{\perp})+\beta(\boldsymbol{b}_{\parallel}+\boldsymbol{b}_{\perp}) \tag{50}$$

where $\boldsymbol{a}_{\parallel}=(\boldsymbol{a}\cdot\hat{\boldsymbol{c}})\hat{\boldsymbol{c}}$, $\boldsymbol{b}_{\parallel}=(\boldsymbol{b}\cdot\hat{\boldsymbol{c}})\hat{\boldsymbol{c}}$, $\boldsymbol{v}_{\parallel}=(\boldsymbol{v}\cdot\hat{\boldsymbol{c}})\hat{\boldsymbol{c}}$. Now parallel and perpendicular components to $\boldsymbol{c}$ are linearly independent, so (50) implies that parallel and perpendicular components must match separately, thus $\boldsymbol{v}_{\parallel}=\alpha\boldsymbol{a}_{\parallel}+\beta\boldsymbol{b}_{\parallel}$, therefore $(\alpha\boldsymbol{a}+\beta\boldsymbol{b})\cdot\hat{\boldsymbol{c}}=\alpha(\boldsymbol{a}\cdot\hat{\boldsymbol{c}})+\beta(\boldsymbol{b}\cdot\hat{\boldsymbol{c}})$, multiplying by $|\boldsymbol{c}|$ yields the desired distributivity property. ☐

The distributivity property is a fundamental algebraic property of the dot product. It allows us to deduce that

$$\begin{aligned}\boldsymbol{a}\cdot\boldsymbol{b}&=(a_x\hat{\boldsymbol{x}}+a_y\hat{\boldsymbol{y}}+a_z\hat{\boldsymbol{z}})\cdot(b_x\hat{\boldsymbol{x}}+b_y\hat{\boldsymbol{y}}+b_z\hat{\boldsymbol{z}})\\&=a_xb_x\,\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{x}}+a_xb_y\,\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}+\cdots\\&=a_xb_x+a_yb_y+a_zb_z\end{aligned} \tag{51}$$

in terms of cartesian components for $\boldsymbol{a}$ and $\boldsymbol{b}$, since $\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{y}}=0$, $\hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{z}}=0$, $\hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{x}}=0$. Thus computing a dot product is easy when the vectors are known in cartesian form. In particular, the angle between two cartesian vectors can be obtained from

$$\cos\theta=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}=\frac{a_xb_x+a_yb_y+a_zb_z}{\sqrt{a_x^2+a_y^2+a_z^2}\,\sqrt{b_x^2+b_y^2+b_z^2}}.$$

That result (51) is the standard definition of dot product in $\mathbb{R}^3$ but we deduced it from the geometric definition (44). The latter is more general. For example, if vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are expanded in terms of an arbitrary basis $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3\}$, the distributivity and commutativity of the dot product yield

$$\begin{aligned}\boldsymbol{u}\cdot\boldsymbol{v}&=(u_1\boldsymbol{a}_1+u_2\boldsymbol{a}_2+u_3\boldsymbol{a}_3)\cdot(v_1\boldsymbol{a}_1+v_2\boldsymbol{a}_2+v_3\boldsymbol{a}_3)\\&=u_1v_1\,(\boldsymbol{a}_1\cdot\boldsymbol{a}_1)+u_2v_2\,(\boldsymbol{a}_2\cdot\boldsymbol{a}_2)+u_3v_3\,(\boldsymbol{a}_3\cdot\boldsymbol{a}_3)\\&\quad+(u_1v_2+u_2v_1)\,(\boldsymbol{a}_1\cdot\boldsymbol{a}_2)+(u_1v_3+u_3v_1)\,(\boldsymbol{a}_1\cdot\boldsymbol{a}_3)\\&\quad+(u_2v_3+u_3v_2)\,(\boldsymbol{a}_2\cdot\boldsymbol{a}_3)\\&\neq u_1v_1+u_2v_2+u_3v_3.\end{aligned} \tag{52}$$

In matrix notation, this reads

$$\boldsymbol{u}\cdot\boldsymbol{v} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1\cdot\boldsymbol{a}_1 & \boldsymbol{a}_1\cdot\boldsymbol{a}_2 & \boldsymbol{a}_1\cdot\boldsymbol{a}_3 \\ \boldsymbol{a}_2\cdot\boldsymbol{a}_1 & \boldsymbol{a}_2\cdot\boldsymbol{a}_2 & \boldsymbol{a}_2\cdot\boldsymbol{a}_3 \\ \boldsymbol{a}_3\cdot\boldsymbol{a}_1 & \boldsymbol{a}_3\cdot\boldsymbol{a}_2 & \boldsymbol{a}_3\cdot\boldsymbol{a}_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{53}$$

while in index notation, this is

$$\boldsymbol{u}\cdot\boldsymbol{v} = u_i v_j \, g_{ij} \tag{54}$$

with implicit sums over the repeated i and j indices, where $g_{ij} \triangleq \boldsymbol{a}_i\cdot\boldsymbol{a}_j = g_{ji}$ is the *metric* of the basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$. Matrix and index notations are discussed later in these notes.

Exercises

1. A skier slides down an inclined plane with a total vertical drop of h, show that the work done by gravity is independent of the slope. Use $\boldsymbol{F}$ and $\boldsymbol{\ell}$'s and sketch the geometry of this result.
2. Discuss and sketch the solutions $\boldsymbol{u}$ of $\boldsymbol{a}\cdot\boldsymbol{u} = \alpha$, where $\boldsymbol{a}$ and α are known.
3. Prove that $(\alpha\boldsymbol{a})\cdot\boldsymbol{b} = \alpha(\boldsymbol{a}\cdot\boldsymbol{b})$ from the geometric definition (44). Discuss both $\alpha \geq 0$ and $\alpha < 0$.
4. Sketch $\boldsymbol{c} = \boldsymbol{a}+\boldsymbol{b}$ then calculate $\boldsymbol{c}\cdot\boldsymbol{c} = (\boldsymbol{a}+\boldsymbol{b})\cdot(\boldsymbol{a}+\boldsymbol{b})$ and deduce the 'law of cosines'.
5. Use vector algebra to show that $\boldsymbol{a}_\perp \triangleq \boldsymbol{a} - (\boldsymbol{a}\cdot\hat{\boldsymbol{n}})\hat{\boldsymbol{n}}$ is orthogonal to $\hat{\boldsymbol{n}}$, for any $\boldsymbol{a}$, $\hat{\boldsymbol{n}}$. Sketch.
6. $\boldsymbol{B}$ is a magnetic field and $\boldsymbol{v}$ is the velocity of a particle. We want to decompose $\boldsymbol{v} = \boldsymbol{v}_\perp + \boldsymbol{v}_\parallel$ where $\boldsymbol{v}_\perp$ is perpendicular to the magnetic field and $\boldsymbol{v}_\parallel$ is parallel to it. Derive vector expressions for $\boldsymbol{v}_\perp$ and $\boldsymbol{v}_\parallel$ in terms of $\boldsymbol{v}$ and $\boldsymbol{B}$.
7. True or false?: $\boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{a}\cdot\boldsymbol{b}_\parallel = \boldsymbol{a}_\parallel\cdot\boldsymbol{b} = a_\parallel b_\parallel$.
8. If $\boldsymbol{a}$ and $\boldsymbol{b}$ have components $(1,2)$ and $(3,4)$ in the basis {*North*, *Northwest*}, respectively, what is their dot product?
9. Show that the diagonals of the parallelogram spanned by $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal to each other if and only if $|\boldsymbol{a}| = |\boldsymbol{b}|$.
10. Consider $\boldsymbol{v}(t) = \boldsymbol{a} + t\boldsymbol{b}$ where $t \in \mathbb{R}$. What is the minimum $|\boldsymbol{v}|$ and for what t? Solve two ways: (1) geometrically and (2) using calculus.
11. Given three arbitrary points A, B, C in 3D space, specified by their cartesian coordinates, what is the equation of the line through A that intersects BC orthogonally? Give a vector equation and specify the algorithm to compute all vectors from the cartesian data. For example, what is the line through point $(2,3,4)$ that intersects the line through points $(3,1,5)$ and $(4,2,3)$ at 90 degrees?

12. Use vectors to show that the three heights (a.k.a. *altitudes* or *normals*) dropped from the vertices of a triangle perpendicular to their opposite sides intersect at the same point, the *orthocenter* H. [Hint: this is similar to the median problem in section 3.4 but now $\overrightarrow{AD}$ and $\overrightarrow{BE}$ are defined by $\overrightarrow{AD} \cdot \overrightarrow{CB} = 0$ and $\overrightarrow{BE} \cdot \overrightarrow{CA} = 0$ and the goal is to show that $\overrightarrow{CH} \cdot \overrightarrow{AB} = 0$. Do you need to find H to prove the result?].

13. Three points A, B, C in 3D space are specified by their cartesian coordinates. Show that the three equations $\overrightarrow{AB} \cdot \overrightarrow{CH} = 0$, $\overrightarrow{BC} \cdot \overrightarrow{AH} = 0$, $\overrightarrow{CA} \cdot \overrightarrow{BH} = 0$, are not sufficient to find the coordinates of H. Explain.

14. Three points A, B, C in 3D space are specified by their cartesian coordinates. Derive an algorithm to compute the coordinates of the point H that is the intersection of the heights using linear combinations of known vectors and 3 dot products.

15. A and B are two points on a sphere of radius R specified by their longitude and latitude. What are longitude and latitude? Draw clean sketches and explain. Specify the algorithm to compute the shortest distance between A and B, traveling on the sphere. What is the shortest distance between Morey Field in Middleton, Wisconsin and Nadi airport in the Fiji Islands?

16. Use vectors to show that if A, B and C lie on a circle, with AC as a diameter, then $\widehat{ABC}$ is a right angle.

17. Use vectors to prove that any point P on the perpendicular bisector of AB is equidistant from A and B. Prove that the perpendicular bisectors of the sides of a triangle are concurrent and that their intersection O is the center of a circle that passes through all three vertices of the triangle. That point O is called the *circumcenter*.

18. Prove that any point P on an angle bisector is equidistant from the sides of the angle. Prove that the angle bisectors of a triangle are concurrent and that their intersection I is the center of a circle tangent to the three sides (the *incircle*).

19. Prove that $\hat{\boldsymbol{a}} + \hat{\boldsymbol{b}}$ bisects the angle between arbitrary vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Sketch.

20. Given three points P_1, P_2, P_3 in 3D space, specified by their cartesian coordinates ($P_1 \equiv (x_1, y_1, z_1)$ etc.), derive a vector algorithm to find the center C of the circle through the 3 points (the *circumcircle*). Write a code in Matlab or Python (or other language). Test your code.

21. Given three points P_1, P_2, P_3 in 3D space, specified by their cartesian coordinates ($P_1 \equiv (x_1, y_1, z_1)$ etc.), derive a vector algorithm to find the center I of the circle tangent to the three sides (the *inner circle*). Write a code in Matlab or Python (or other language). Test your code.

22. In an arbitrary triangle, let O be the circumcenter and G the centroid. Consider the point P such that $\overrightarrow{GP} = 2\overrightarrow{OG}$. Show that P is the orthocenter H, hence

O, G and H are on the same line called the *Euler line*. Sketch. [Easy to prove with the right similar triangles. With vectors, consider any vertex V and let V' be the midpoint of the opposite side. Relate $\overrightarrow{VP}$ to $\overrightarrow{OV'}$. Analyze, conclude.]

23. In an arbitrary triangle, let H be the orthocenter and G the centroid. Consider the point P such that $\overrightarrow{HG} = 2\overrightarrow{GP}$. Show that P is the circumcenter O, hence O, G and H are on the same line called the *Euler line*. [Easy to prove with the right similar triangles. With vectors, consider any vertex V and let V' be the midpoint of the opposite side. Relate $\overrightarrow{VH}$ to $\overrightarrow{PV'}$. Analyze, conclude.]

4.2 Orthonormal bases

Arbitrary basis. Given any vector $\boldsymbol{v}$ and three non co-planar vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ in $\mathbb{E}^3$, you can in principle find the three (signed) scalars (v_1, v_2, v_3) such that

$$\boldsymbol{v} = v_1\,\boldsymbol{a}_1 + v_2\,\boldsymbol{a}_2 + v_3\,\boldsymbol{a}_3 \tag{55}$$

by *parallel* projections (figs. 1.16 & 1.17). The scalars (v_1, v_2, v_3) are the *components* of $\boldsymbol{v}$ in the basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$. For an arbitrary basis, we cannot find v_1 without knowing the directions of $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$, and likewise for v_2 and v_3.

Orthogonal basis. If the basis vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ are mutually orthogonal, that is if $\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 = \boldsymbol{a}_2 \cdot \boldsymbol{a}_3 = \boldsymbol{a}_3 \cdot \boldsymbol{a}_1 = 0$ then we can find each component independently of the others. Indeed, dotting (55) with $\boldsymbol{a}_1$ yields

$$\boldsymbol{v} \cdot \boldsymbol{a}_1 = v_1\,\boldsymbol{a}_1 \cdot \boldsymbol{a}_1 + v_2\,\boldsymbol{a}_2 \cdot \boldsymbol{a}_1 + v_3\,\boldsymbol{a}_3 \cdot \boldsymbol{a}_1 \tag{56}$$

and

$$v_1 = \frac{\boldsymbol{v} \cdot \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1},$$

independent of $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ as long as $\boldsymbol{a}_2 \cdot \boldsymbol{a}_1 = 0 = \boldsymbol{a}_3 \cdot \boldsymbol{a}_1$. The other components are obtained likewise by projecting the vector $\boldsymbol{v}$ onto the respective basis vector and using orthogonality $\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 = \boldsymbol{a}_2 \cdot \boldsymbol{a}_3 = \boldsymbol{a}_3 \cdot \boldsymbol{a}_1 = 0$ to obtain

$$v_2 = \frac{\boldsymbol{a}_2 \cdot \boldsymbol{v}}{\boldsymbol{a}_2 \cdot \boldsymbol{a}_2}, \quad v_3 = \frac{\boldsymbol{a}_3 \cdot \boldsymbol{v}}{\boldsymbol{a}_3 \cdot \boldsymbol{a}_3}.$$

Orthonormal basis. An orthonormal basis is such that the basis vectors are mutually orthogonal *and* of unit norm. Our usual cartesian basis $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ is an orthonormal basis and, conversely, any orthonormal basis can serve as a cartesian basis. Such a basis is often denoted[6] $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Its compact definition is

$$\boxed{\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}} \tag{57}$$

[6] Forget about the notation $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ for cartesian unit vectors. This is 19th century notation, it is unfortunately still very common in elementary courses but that old notation will get in the way if you stick to it. We will NEVER use $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$, instead we use $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ or $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ or $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ to denote a set of three orthonormal vectors in 3D euclidean space. We use **indices** i, j and k to denote *anyone* of those 3 directions. Those indices are positive integers that can take all the values from 1 to n, the dimension of the space. We spend most of our time in 3D space, so most of the time our indices i, j and k can *each* be 1, 2 or 3. They should not be confused with those old orthonormal vectors $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ from elementary calculus.

where i and j can each be any of $1, 2, 3$ and δ_{ij} is the **Kronecker symbol** defined as

$$\delta_{ij} \triangleq \begin{cases} 1 & \text{if } \ i = j, \\ 0 & \text{if } \ i \neq j. \end{cases} \tag{58}$$

Orthogonal Expansion and Projections. The components of a vector $\boldsymbol{v}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are the scalars v_1, v_2, v_3 such that

$$\boldsymbol{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3. \tag{59}$$

Equations (59) is the *expansion* of $\boldsymbol{v}$ in terms of the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Dotting (59) with each of the basis vectors yields

$$v_1 = \boldsymbol{v} \cdot \mathbf{e}_1, \quad v_2 = \boldsymbol{v} \cdot \mathbf{e}_2, \quad v_3 = \boldsymbol{v} \cdot \mathbf{e}_3 \tag{60}$$

since $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are orthonormal (57). Equations (60) are the *orthogonal projections* of $\boldsymbol{v}$ onto each of the unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. The orthogonal projections of a vector $\boldsymbol{v}$ onto an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are the components of $\boldsymbol{v}$ in that basis.

The expansion (59) and projection (60) equations for $\boldsymbol{v}$ can be written more compactly in index notation with sigma notation as

$$\boldsymbol{v} = \sum_{j=1}^{3} v_j\mathbf{e}_j \quad \Leftrightarrow \quad v_i = \mathbf{e}_i \cdot \boldsymbol{v}, \quad \forall i = 1, 2, 3. \tag{61}$$

Invariance of cartesian dot product. If two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are expanded in terms of the orthonormal $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, that is

$$\boldsymbol{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3, \qquad \boldsymbol{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3,$$

then the distributivity properties of the dot product and the orthonormality of the basis yield

$$\boxed{\boldsymbol{a} \cdot \boldsymbol{b} = a_1b_1 + a_2b_2 + a_3b_3.} \tag{62}$$

One remarkable property of this formula is that its value is independent of the orthonormal basis. The dot product is a geometric property of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, independent of the basis. This is obvious from the geometric definition (44) but not from its expression in terms of components (62). If $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ are two distinct orthonormal bases then

$$\boldsymbol{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 = a'_1\mathbf{e}'_1 + a'_2\mathbf{e}'_2 + a'_3\mathbf{e}'_3$$

but, in general, the components in the two bases are distinct: $a_1 \neq a'_1$, $a_2 \neq a'_2$, $a_3 \neq a'_3$, and likewise for another vector $\boldsymbol{b}$, yet

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1b_1 + a_2b_2 + a_3b_3 = a'_1b'_1 + a'_2b'_2 + a'_3b'_3. \tag{63}$$

The simple algebraic form of the dot product is *invariant under a change of orthonormal basis*, although each component may be different.

Exercises

1. Given the orthonormal (cartesian) basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, consider $\boldsymbol{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$, $\boldsymbol{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2$, $\boldsymbol{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$. What are the components of $\boldsymbol{v}$ (i) in terms of $\boldsymbol{a}$ and $\boldsymbol{b}$? (ii) in terms of $\boldsymbol{a}$ and $\boldsymbol{b}_\perp$ where $\boldsymbol{a} \cdot \boldsymbol{b}_\perp = 0$?

2. If $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ are two distinct orthogonal bases and $\boldsymbol{a}$ and $\boldsymbol{b}$ are arbitrary vectors, prove (63) but construct an example that $a_1b_1 + 2a_2b_2 + 3a_3b_3 \neq a'_1b'_1 + 2a'_2b'_2 + 3a'_3b'_3$ in general.

3. If $\boldsymbol{w} = \sum_{i=1}^3 w_i\mathbf{e}_i$, calculate $\mathbf{e}_j \cdot \boldsymbol{w}$ using $\sum$ notation and (57).

4. Why is it not true that $\mathbf{e}_i \cdot \sum_{i=1}^3 w_i\mathbf{e}_i = \sum_{i=1}^3 w_i(\mathbf{e}_i \cdot \mathbf{e}_i) = \sum_{i=1}^3 w_i\delta_{ii} = w_1 + w_2 + w_3$? How can you fix this calculation?

5. If $\boldsymbol{v} = \sum_{i=1}^3 v_i\mathbf{e}_i$ and $\boldsymbol{w} = \sum_{i=1}^3 w_i\mathbf{e}_i$, calculate $\boldsymbol{v} \cdot \boldsymbol{w}$ using $\sum$ notation and (57).

6. If $\boldsymbol{v} = \sum_{i=1}^3 v_i\boldsymbol{a}_i$ and $\boldsymbol{w} = \sum_{i=1}^3 w_i\boldsymbol{a}_i$, where the basis $\boldsymbol{a}_i$, $i = 1, 2, 3$, is *not* orthonormal, calculate $\boldsymbol{v} \cdot \boldsymbol{w}$ using $\sum$ notation.

7. Calculate (i) $\sum_{j=1}^3 \delta_{ij}a_j$, (ii) $\sum_{i=1}^3 \sum_{j=1}^3 \delta_{ij}a_jb_i$, (iii) $\sum_{j=1}^3 \delta_{jj}$.

4.3 Dot product and norm in $\mathbb{R}^n$ (Optional)

Dot product in $\mathbb{R}^n$

The geometric definition of the dot product (44) is great for oriented line segments as it emphasizes the geometric aspects, but the algebraic formula (62) is very useful for calculations. It's also a way to define the dot product for other vector spaces where the concept of 'angle' between vectors may not be obvious *e.g.* what is the angle between the vectors (1,2,3,4) and (4,3,2,1) in $\mathbb{R}^4$?! The dot product (*a.k.a. scalar* product or *inner* product) of the vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ is defined as suggested by (62):

$$\mathbf{x} \cdot \mathbf{y} \triangleq x_1y_1 + x_2y_2 + \cdots + x_ny_n. \tag{64}$$

The reader will verify that this definition satisfies the fundamental properties of the dot product (sect. 4.1) (commutativity $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$, positive definiteness $\mathbf{x} \cdot \mathbf{x} \geq 0$ and multi-linearity (or distributivity) $(\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2) \cdot \mathbf{y} = \alpha_1(\mathbf{x}_1 \cdot \boldsymbol{y}) + \alpha_2(\mathbf{x}_2 \cdot \mathbf{y})$.

To show the *Cauchy-Schwarz* property, you need a bit of Calculus and a classical trick: consider $\mathbf{v} = \mathbf{x} + \lambda\mathbf{y}$, then

$$F(\lambda) \triangleq \mathbf{v} \cdot \mathbf{v} = \lambda^2\mathbf{y} \cdot \mathbf{y} + 2\lambda\mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{x} \geq 0.$$

For given, but arbitrary, $\mathbf{x}$ and $\mathbf{y}$, this is a quadratic polynomial in λ. That polynomial $F(\lambda)$ has a single minimum at $\lambda_* = -(\mathbf{x} \cdot \mathbf{y})/(\mathbf{y} \cdot \mathbf{y})$. That minimum value is

$$F(\lambda_*) = (\mathbf{x} \cdot \mathbf{x}) - \frac{(\mathbf{x} \cdot \mathbf{y})^2}{(\mathbf{y} \cdot \mathbf{y})} \geq 0$$

which must still be positive since $F \geq 0, \forall \lambda$, hence the Cauchy-Schwarz inequality.

Once we know that the definition (64) satisfies Cauchy-Schwarz, $(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})\,(\mathbf{y} \cdot \mathbf{y})$, we can define the length of a vector by $|\mathbf{x}| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$ (this is called the *Euclidean* length since it corresponds to length in Euclidean geometry by Pythagoras's theorem) and the angle θ between two vectors in $\mathbb{R}^n$ by $\cos\theta = (\mathbf{x} \cdot \mathbf{y})/(|\mathbf{x}|\;|\mathbf{y}|)$. A vector space for which a dot (or inner) product is defined is called a *Hilbert space* (or an *inner product space*).

The bottom line is that for more complex vector spaces, the dot (or scalar or inner) product is a key mathematical construct that allows us to generalize the concept of 'angle' between vectors and, most importantly, to define 'orthogonal vectors'.

Norm of a vector

The *norm* of a vector, denoted $\|\mathbf{a}\|$, is a positive real number that defines its size or 'length' (but not in the sense of the number of its components). For displacement vectors in Euclidean spaces, the norm is the length of the displacement, $\|\boldsymbol{a}\| = |\boldsymbol{a}|$, that is the distance between point A and B if $\overrightarrow{AB} = \boldsymbol{a}$. The following properties are geometrically straightforward for length of displacement vectors:

1. $\|\mathbf{a}\| \geq 0$ and $\|\mathbf{a}\| = 0 \Leftrightarrow \mathbf{a} = \mathbf{0}$,
2. $\|\alpha\mathbf{a}\| = |\alpha|\;\|\mathbf{a}\|$,
3. $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$. (triangle inequality)

Draw the triangle formed by $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{a} + \boldsymbol{b}$ to see why the latter is called the *triangle inequality*. For more general vector spaces, these properties become the *defining properties* (*axioms*) that a norm must satisfy. A vector space for which a norm is defined is called a *Banach space*.

Norms for $\mathbb{R}^n$

For other types of vector space, there are many possible definitions for the norm of a vector as long as those definitions satisfy the 3 norm properties. In $\mathbb{R}^n$, the p-norm of vector $\boldsymbol{x}$ is defined by the positive number

$$\|\mathbf{x}\|_p \triangleq \Big(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\Big)^{1/p}, \tag{65}$$

where $p \geq 1$ is a real number. Commonly used norms are the 2-norm $\|\mathbf{x}\|_2$ which is the square root of the sum of the squares, the 1-norm $\|\mathbf{x}\|_1$ (sum of absolute values) and the infinity norm, $\|\mathbf{x}\|_\infty$, defined as the limit as $p \to \infty$ of the above expression.

Note that the 2-norm $\|\mathbf{x}\|_2 = (\mathbf{x} \cdot \mathbf{x})^{1/2}$ and for that reason is also called the *Euclidean norm*. In fact, if a dot product is defined, then a norm can always be defined as the square root of the dot product. In other words, *every Hilbert space is a Banach space,* but the converse is not true.

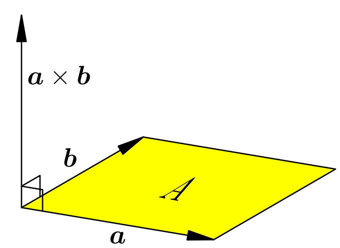
Exercises

1. So what is the angle between $(1, 2, 3, 4)$ and $(4, 3, 2, 1)$?
2. Can you define a dot product for the vector space of real functions $f(x)$?
3. Find a vector orthogonal to $(1, 2, 3, 4)$. Find all the vectors orthogonal to $(1, 2, 3, 4)$.
4. Decompose (4,2,1,7) into the sum of two vectors one of which is parallel and the other perpendicular to $(1, 2, 3, 4)$.
5. Show that $\cos nx$ with n integer, is a set of orthogonal functions on $(0, \pi)$. Find formulas for the components of a function $f(x)$ in terms of that orthogonal basis. In particular, find the components of $\sin x$ in terms of the cosine basis in that $(0, \pi)$ interval.
6. Show that the infinity norm $\|\mathbf{x}\|_\infty = \max_i |x_i|$.
7. Show that the p-norm satisfies the three norm properties for $p = 1, 2, \infty$.
8. Define a norm for $\mathbb{C}^n$.
9. Define the 2-norm for real functions $f(x)$ in $0 < x < 1$.

5 Cross Product

5.1 Geometry and algebra of cross product

The cross product, also called the *vector product* or the *area product*, is defined as



$$\boxed{\boldsymbol{a} \times \boldsymbol{b} \triangleq |\boldsymbol{a}|\,|\boldsymbol{b}|\,\sin\theta\,\hat{\boldsymbol{n}} = A\,\hat{\boldsymbol{n}}} \tag{66}$$

where θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$, with $0 \le \theta \le \pi$. The cross product is a vector whose *magnitude* is the area A of the parallelogram with sides $\boldsymbol{a}$ and $\boldsymbol{b}$, and *direction* $\hat{\boldsymbol{n}}$ is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$, with $(\boldsymbol{a}, \boldsymbol{b}, \hat{\boldsymbol{n}})$ right handed.

The right hand-rule implies that the cross product *anti-commutes*

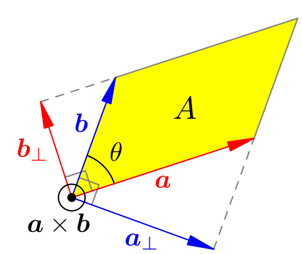
$$\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}, \tag{67}$$

and thus the cross product of any vector with itself vanishes

$$\boldsymbol{a} \times \boldsymbol{a} = 0. \tag{68}$$

Since the area of a parallelogram is *base* $\times$ *height* and there are two ways to pick a base and a height, $(\boldsymbol{a}, \boldsymbol{b}_\perp)$ and $(\boldsymbol{a}_\perp, \boldsymbol{b})$, the geometric definition yields the fundamental cross product identity

$$\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{a} \times \boldsymbol{b}_\perp = \boldsymbol{a}_\perp \times \boldsymbol{b} \tag{69}$$



where $\boldsymbol{b}_\perp$ is the vector component of $\boldsymbol{b}$ perpendicular to $\boldsymbol{a}$, and $\boldsymbol{a}_\perp$ is the vector component of $\boldsymbol{a}$ perpendicular to $\boldsymbol{b}$ (eqn. (48)).

The cross product satisfies the distributivity property:

$$\boldsymbol{\omega}\times(\alpha\boldsymbol{a}+\beta\boldsymbol{b})=\alpha(\boldsymbol{\omega}\times\boldsymbol{a})+\beta(\boldsymbol{\omega}\times\boldsymbol{b}), \tag{70}$$

for any vectors $\boldsymbol{a},\boldsymbol{b},\boldsymbol{\omega}$ and scalars α,β.

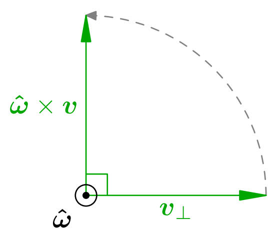


Fig. 1.24: $\hat{\boldsymbol{\omega}}\times\boldsymbol{v}$ rotates $\boldsymbol{v}_\perp$ by $\pi/2$ counterclockwise about $\hat{\boldsymbol{\omega}}$

Proof of Distributivity. Let $\boldsymbol{\omega}=|\boldsymbol{\omega}|\,\hat{\boldsymbol{\omega}}$ and eliminate $|\boldsymbol{\omega}|$ by property 2, then it suffices to show that $\hat{\boldsymbol{\omega}}\times(\boldsymbol{a}+\boldsymbol{b})=\hat{\boldsymbol{\omega}}\times\boldsymbol{a}+\hat{\boldsymbol{\omega}}\times\boldsymbol{b}$. Now $\hat{\boldsymbol{\omega}}\times\boldsymbol{v}=\hat{\boldsymbol{\omega}}\times\boldsymbol{v}_\perp$ is the right hand rotation of $\boldsymbol{v}_\perp$ by $\pi/2$ about $\hat{\boldsymbol{\omega}}$ (fig. 1.24) for any $\boldsymbol{v}$, as implied by (66) and (69), hence $\hat{\boldsymbol{\omega}}\times\boldsymbol{a}$, $\hat{\boldsymbol{\omega}}\times\boldsymbol{b}$, $\hat{\boldsymbol{\omega}}\times(\boldsymbol{a}+\boldsymbol{b})$ are the $\pi/2$ rotation of $\boldsymbol{a}_\perp$, $\boldsymbol{b}_\perp$, $(\boldsymbol{a}+\boldsymbol{b})_\perp$, respectively (fig. 1.25). Since $(\boldsymbol{a}+\boldsymbol{b})_\perp=\boldsymbol{a}_\perp+\boldsymbol{b}_\perp$ as shown in (50), it follows that $\hat{\boldsymbol{\omega}}\times(\boldsymbol{a}+\boldsymbol{b})=\hat{\boldsymbol{\omega}}\times\boldsymbol{a}+\hat{\boldsymbol{\omega}}\times\boldsymbol{b}$. □

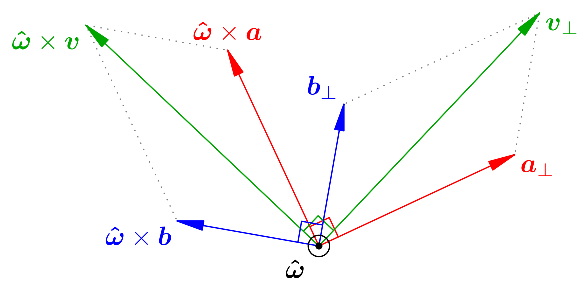
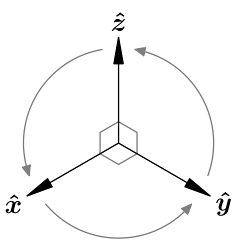


Fig. 1.25: $\hat{\boldsymbol{\omega}}\times(\boldsymbol{a}+\boldsymbol{b})=\hat{\boldsymbol{\omega}}\times\boldsymbol{a}+\hat{\boldsymbol{\omega}}\times\boldsymbol{b}$ is the $\frac{\pi}{2}$ rotation of $(\boldsymbol{a}+\boldsymbol{b})_\perp=\boldsymbol{a}_\perp+\boldsymbol{b}_\perp$.

The distributivity property yields the cartesian formula for the cross product

$$\begin{aligned}\boldsymbol{a}\times\boldsymbol{b}&=(a_x\hat{\boldsymbol{x}}+a_y\hat{\boldsymbol{y}}+a_z\hat{\boldsymbol{z}})\times(b_x\hat{\boldsymbol{x}}+b_y\hat{\boldsymbol{y}}+b_z\hat{\boldsymbol{z}})\\&=a_xb_x\,\hat{\boldsymbol{x}}\times\hat{\boldsymbol{x}}+a_xb_y\,\hat{\boldsymbol{x}}\times\hat{\boldsymbol{y}}+a_xb_z\,\hat{\boldsymbol{x}}\times\hat{\boldsymbol{z}}+\cdots\\&=\hat{\boldsymbol{x}}(a_yb_z-a_zb_y)+\hat{\boldsymbol{y}}(a_zb_x-a_xb_z)+\hat{\boldsymbol{z}}(a_xb_y-a_yb_x)\end{aligned} \tag{71}$$



since

$$\hat{\boldsymbol{x}}\times\hat{\boldsymbol{y}}=\hat{\boldsymbol{z}},\quad \hat{\boldsymbol{y}}\times\hat{\boldsymbol{z}}=\hat{\boldsymbol{x}},\quad \hat{\boldsymbol{z}}\times\hat{\boldsymbol{x}}=\hat{\boldsymbol{y}} \tag{72}$$

and

$$\hat{\boldsymbol{x}}=\hat{\boldsymbol{y}}\times\hat{\boldsymbol{z}},\quad \hat{\boldsymbol{y}}=\hat{\boldsymbol{z}}\times\hat{\boldsymbol{x}},\quad \hat{\boldsymbol{z}}=\hat{\boldsymbol{x}}\times\hat{\boldsymbol{y}}. \tag{73}$$

Note that each of these expressions is a cyclic permutation of the previous one

$$(x,y,z)\to(y,z,x)\to(z,x,y)$$

and this enables us to easily reconstruct formula (71) – or any one of its components, without having to use the right hand rule explicitly. We can figure them out simply with that cyclic (even) or acyclic (odd) permutation rule. The cartesian expansion (71) of the cross product is often remembered using the formal ‘determinants’

$$\begin{aligned}\begin{vmatrix}\hat{\boldsymbol{x}}&\hat{\boldsymbol{y}}&\hat{\boldsymbol{z}}\\a_x&a_y&a_z\\b_x&b_y&b_z\end{vmatrix}&=\begin{vmatrix}\hat{\boldsymbol{x}}&a_x&b_x\\\hat{\boldsymbol{y}}&a_y&b_y\\\hat{\boldsymbol{z}}&a_z&b_z\end{vmatrix}\\&=\hat{\boldsymbol{x}}(a_yb_z-a_zb_y)+\hat{\boldsymbol{y}}(a_zb_x-a_xb_z)+\hat{\boldsymbol{z}}(a_xb_y-a_yb_x),\end{aligned} \tag{74}$$

although the cyclic permutation rules (72), (73), enable reconstruction of the cartesian expansion formula just as easily.

5.2 Double cross product ('Triple vector product')

Double cross products[7] occur frequently in applications (*e.g.* angular momentum of a rotating body) directly or indirectly (see the discussion below about mirror reflection and cross-products in physics). An important special case of double cross product is

$$(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{a} = (\boldsymbol{a} \cdot \boldsymbol{a})\, \boldsymbol{b}_\perp = \boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{a}) \tag{75}$$

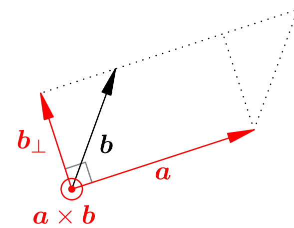


Fig. 1.26: $(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{a} = a^2 \boldsymbol{b}_\perp$.

where $\boldsymbol{b}_\perp$ is the vector component of $\boldsymbol{b}$ perpendicular to $\boldsymbol{a}$. The identity (75) follows from the geometric definition of the cross product since $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{a} \times \boldsymbol{b}_\perp$, that demonstration is left as an exercise. The general double cross products obey the following identity

$$\boxed{\begin{aligned} (\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} &= (\boldsymbol{a} \cdot \boldsymbol{c})\, \boldsymbol{b} - (\boldsymbol{b} \cdot \boldsymbol{c})\, \boldsymbol{a}, \\ \boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) &= (\boldsymbol{a} \cdot \boldsymbol{c})\, \boldsymbol{b} - (\boldsymbol{a} \cdot \boldsymbol{b}) \boldsymbol{c}. \end{aligned}} \tag{76}$$

thus in general

$$(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} \neq \boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}),$$

except when $\boldsymbol{c} = \alpha\, \boldsymbol{a}$ as in (75). The identities (76) follow from each other after some manipulations and renaming of vectors, but we can remember both at once as:

> *middle vector times dot product of the other two, minus*
> *other vector in parentheses times dot product of the other two.*[8]

We give two proofs of the double cross product identity (76) that serve as exercises on the vector geometry and algebra of the cross product.

Vector Geometric Proof of (76). Let $\boldsymbol{X} = (\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c}$ for short. By definition of the cross product, $\boldsymbol{X}$ must be orthogonal to $(\boldsymbol{a} \times \boldsymbol{b})$ and to $\boldsymbol{c}$. Orthogonality to $\boldsymbol{a} \times \boldsymbol{b}$ means that $\boldsymbol{X}$ is a linear combination of $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\boldsymbol{X} = x_1\, \boldsymbol{a} + x_2\, \boldsymbol{b}.$$

Orthogonality to $\boldsymbol{c}$ then means $\boldsymbol{X} \cdot \boldsymbol{c} = x_1\, \boldsymbol{a} \cdot \boldsymbol{c} + x_2\, \boldsymbol{b} \cdot \boldsymbol{c} = 0$, hence $x_1 = \mu(\boldsymbol{b} \cdot \boldsymbol{c})$, $x_2 = -\mu(\boldsymbol{a} \cdot \boldsymbol{c})$ for some scalar μ, that is

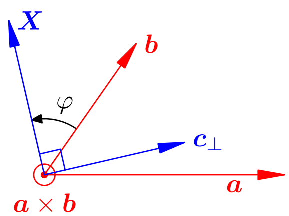
$$\boldsymbol{X} = \mu(\boldsymbol{b} \cdot \boldsymbol{c})\boldsymbol{a} - \mu(\boldsymbol{a} \cdot \boldsymbol{c})\boldsymbol{b},$$

This almost gives the formula (76) but we still need to find μ. Let $\boldsymbol{a} \times \boldsymbol{b} = A\, \hat{\boldsymbol{n}}$ and $\boldsymbol{c} = \boldsymbol{c}_\parallel + \boldsymbol{c}_\perp$, where $\parallel$ and $\perp$ here refer to the $\boldsymbol{a} \times \boldsymbol{b}$ direction. The vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{X}$ and $\boldsymbol{c}_\perp$ are all in the plane orthogonal to $\hat{\boldsymbol{n}}$, and furthermore $\boldsymbol{X} \perp \boldsymbol{c}_\perp$, as illustrated in the figure. Then

$$\boldsymbol{X} = (\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} = A\, |\boldsymbol{c}_\perp|\, \hat{\boldsymbol{X}}$$

[7] The double cross product is often called 'triple vector product', there are 3 vectors but only 2 products!

[8] This is more useful than the confusing 'BAC-CAB' rule for remembering the 2nd. Try applying the BAC-CAB mnemonic to $(\boldsymbol{b} \times \boldsymbol{c}) \times \boldsymbol{a}$ for confusing fun!



but also

$$\boldsymbol{X} = \mu(\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a} - \mu(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} = \mu(\boldsymbol{b}\cdot\boldsymbol{c}_\perp)\boldsymbol{a} - \mu(\boldsymbol{a}\cdot\boldsymbol{c}_\perp)\boldsymbol{b}.$$

Crossing both of these expressions with $\boldsymbol{b}$ yields

$$\boldsymbol{b}\times\boldsymbol{X} = A\,|\boldsymbol{c}_\perp|\,\boldsymbol{b}\times\hat{\boldsymbol{X}} = A\,|\boldsymbol{c}_\perp||\boldsymbol{b}|\,\sin\varphi\,\hat{\boldsymbol{n}}$$

and

$$\boldsymbol{b}\times\boldsymbol{X} = \mu(\boldsymbol{b}\cdot\boldsymbol{c}_\perp)\,\boldsymbol{b}\times\boldsymbol{a} = \mu\,(|\boldsymbol{b}|\,|\boldsymbol{c}_\perp|\,\sin\varphi)\,A\,(-\hat{\boldsymbol{n}}),$$

where φ is the angle from $\boldsymbol{b}$ to $\boldsymbol{X}$, positive for right-hand rotation about $\boldsymbol{a}\times\boldsymbol{b}$. Comparing the two expressions for $\boldsymbol{b}\times\boldsymbol{X}$ gives $\mu=-1$ and formula (76).

□

Vector Algebra Proof of (76). Consider the intrinsic orthogonal basis $\boldsymbol{a}, \boldsymbol{b}_\perp, (\boldsymbol{a}\times\boldsymbol{b})$. In that basis

$$\boldsymbol{a}=\boldsymbol{a}, \quad \boldsymbol{b} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a} + \boldsymbol{b}_\perp, \quad \boldsymbol{c} = \frac{\boldsymbol{a}\cdot\boldsymbol{c}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a} + \frac{\boldsymbol{b}_\perp\cdot\boldsymbol{c}}{\boldsymbol{b}_\perp\cdot\boldsymbol{b}_\perp}\boldsymbol{b}_\perp + \gamma(\boldsymbol{a}\times\boldsymbol{b}) \tag{77}$$

for some[9] γ that does not contribute to $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$. Now $\boldsymbol{a}\times\boldsymbol{b} = \boldsymbol{a}\times\boldsymbol{b}_\perp$ thus

$$(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c} = (\boldsymbol{a}\times\boldsymbol{b}_\perp)\times\boldsymbol{c}.$$

Substituting (77) for $\boldsymbol{c}$ and using (75) twice, once for $(\boldsymbol{a}\times\boldsymbol{b}_\perp)\times\boldsymbol{a} = (\boldsymbol{a}\cdot\boldsymbol{a})\boldsymbol{b}_\perp$ and another for $(\boldsymbol{a}\times\boldsymbol{b}_\perp)\times\boldsymbol{b}_\perp = -(\boldsymbol{b}_\perp\cdot\boldsymbol{b}_\perp)\boldsymbol{a}$, yields

$$(\boldsymbol{a}\times\boldsymbol{b}_\perp)\times\boldsymbol{c} = (\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}_\perp - (\boldsymbol{b}_\perp\cdot\boldsymbol{c})\boldsymbol{a},$$

but

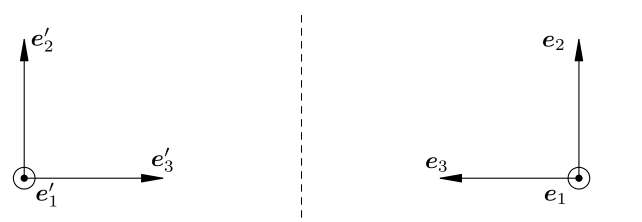
$$(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}_\perp - (\boldsymbol{b}_\perp\cdot\boldsymbol{c})\boldsymbol{a} = (\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}$$

since $\boldsymbol{b} = \alpha\boldsymbol{a} + \boldsymbol{b}_\perp$ and $(\boldsymbol{a}\cdot\boldsymbol{c})(\alpha\boldsymbol{a}) - (\alpha\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{a} = 0$. Thus (76) is true for all $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$.

□

5.3 Orientation of Bases

If we pick an arbitrary unit vector $\mathbf{e}_1$, then a unit vector $\mathbf{e}_2$ orthogonal to $\mathbf{e}_1$, there are two opposite unit vectors $\mathbf{e}_3$ orthogonal to both $\mathbf{e}_1$ and $\mathbf{e}_2$. One choice gives a *right-handed basis* (*i.e.* $\mathbf{e}_1$ in right thumb direction, $\mathbf{e}_2$ in right index direction and $\mathbf{e}_3$ in right major direction). The other choice gives a *left-handed basis*. These two types of bases are *mirror images* of each other as illustrated in the following figure, where $\mathbf{e}_1{}' = \mathbf{e}_1$ point straight out of the paper (or screen).

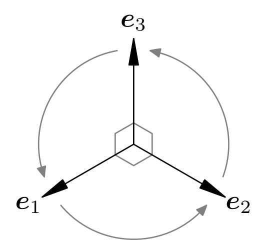


Left handed *mirror* *Right handed*

[9] $\gamma = (\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}/|\boldsymbol{a}\times\boldsymbol{b}|^2$

This figure reveals an interesting subtlety of the cross product. For this particular choice of left and right handed bases (other arrangements are possible of course), $\mathbf{e}_1{}' = \mathbf{e}_1$ and $\mathbf{e}_2{}' = \mathbf{e}_2$ but $\mathbf{e}_3{}' = -\mathbf{e}_3$ so $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$ and $\mathbf{e}_1{}' \times \mathbf{e}_2{}' = \mathbf{e}_3 = -\mathbf{e}_3{}'$. This indicates that the mirror image of the cross-product is *not* the cross-product of the mirror images. On the opposite, the mirror image of the cross-product $\mathbf{e}_3{}'$ is *minus* the cross-product of the images $\mathbf{e}_1{}' \times \mathbf{e}_2{}'$. We showed this for a special case, but this is general, the cross-product is not invariant under reflection, it changes sign. Physical laws should not depend on the choice of basis, so this implies that they should not be expressed in terms of an *odd* number of cross products. When we write that the velocity of a particle is $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$, $\boldsymbol{v}$ and $\boldsymbol{r}$ are 'good' vectors (reflecting as they should under mirror symmetry) but $\boldsymbol{\omega}$ is not quite a true vector, it is a *pseudo-vector*. It changes sign under reflection. That is because rotation vectors are themselves defined according to the right-hand rule, so an expression such as $\boldsymbol{\omega} \times \boldsymbol{r}$ actually contains two applications of the right hand rule. Likewise in the Lorentz force $\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$, $\boldsymbol{F}$ and $\boldsymbol{v}$ are good vectors, but since the definition involves a cross-product, it must be that $\boldsymbol{B}$ is a pseudo-vector. Indeed $\boldsymbol{B}$ is itself a cross-product so the definition of $\boldsymbol{F}$ actually contains two cross-products.

The orientation (right-handed or left-handed) did not matter to us before but, now that we've defined the cross-product with the right-hand rule, we'll typically choose right-handed bases. We don't have to, geometrically speaking, but we need to from an algebraic point of view otherwise we'd need two sets of algebraic formula, one for right-handed bases and one for left-handed bases. In terms of our right-handed cross product definition, we can define a right-handed basis by



$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3 \quad \Rightarrow \quad \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1, \qquad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2, \tag{78}$$

$$\Rightarrow \mathbf{e}_2 \times \mathbf{e}_1 = -\mathbf{e}_3, \quad \mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2, \quad \mathbf{e}_3 \times \mathbf{e}_2 = -\mathbf{e}_1. \tag{79}$$

Note that (78) are *cyclic* rotations of the basis vectors $(1,2,3) \to (2,3,1) \to (3,1,2)$. The orderings of the basis vectors in (79) correspond to a-cyclic rotations of $(1,2,3)$. For 3 elements, a *cyclic* rotation corresponds to an *even* number of permutations. For instance we can go from $(1,2,3)$ to $(2,3,1)$ in 2 permutations $(1,2,3) \to (2,1,3) \to (2,3,1)$. The concept of *even* and *odd* number of permutations is more general. But for three elements it is useful to think in terms of cyclic and acyclic permutations.

If we expand $\boldsymbol{a}$ and $\boldsymbol{b}$ in terms of the right-handed $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, then apply the 3 properties of the cross-product *i.e.* in compact summation form

$$\boldsymbol{a} = \sum_{i=1}^{3} a_i \mathbf{e}_i, \quad \boldsymbol{b} = \sum_{j=1}^{3} b_j \mathbf{e}_j, \quad \Rightarrow \boldsymbol{a} \times \boldsymbol{b} = \sum_{i=1}^{3} \sum_{j=1}^{3} \; a_i b_j \; (\mathbf{e}_i \times \mathbf{e}_j),$$

we obtain

$$\boldsymbol{a} \times \boldsymbol{b} = \mathbf{e}_1(a_2 b_3 - a_3 b_2) + \mathbf{e}_2(a_3 b_1 - a_1 b_3) + \mathbf{e}_3(a_1 b_2 - a_2 b_1), \tag{80}$$

which, again, we can reconstruct using the cyclic permutations of

$$\begin{aligned} + &: (1,2,3) \to (2,3,1) \to (3,1,2), \\ - &: (1,3,2) \to (2,1,3) \to (3,2,1). \end{aligned} \tag{81}$$

Exercises

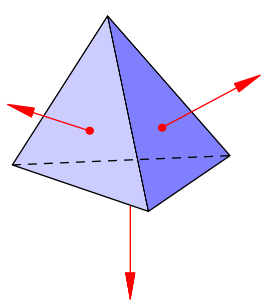
1. Show that $|\boldsymbol{a}\times\boldsymbol{b}|^2+(\boldsymbol{a}\cdot\boldsymbol{b})^2=|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2,\ \forall\boldsymbol{a},\boldsymbol{b}$.

2. A particle of charge q moving at velocity $\boldsymbol{v}$ in a magnetic field $\boldsymbol{B}$ experiences the Lorentz force $\boldsymbol{F}=q\boldsymbol{v}\times\boldsymbol{B}$. Show that there is no force in the direction of the magnetic field and that the Lorentz force does no work on the particle. Does it have any effect on the particle?

3. Sketch three vectors such that $\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}=0$, show that $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{b}\times\boldsymbol{c}=\boldsymbol{c}\times\boldsymbol{a}$ in two ways (1) from the geometric definition of the cross product and (2) from the algebraic properties of the cross product. Deduce the 'law of sines' relating the sines of the angles of a triangle and the lengths of its sides
$$\frac{\sin\alpha}{a}=\frac{\sin\beta}{b}=\frac{\sin\gamma}{c}.$$

4. Consider any three points P_1, P_2, P_3 in 3D Euclidean space, with position vectors $\boldsymbol{r}_1=\overrightarrow{OP_1}$, etc. Show that
$$A\hat{\boldsymbol{n}}\triangleq\frac{1}{2}(\boldsymbol{r}_1\times\boldsymbol{r}_2+\boldsymbol{r}_2\times\boldsymbol{r}_3+\boldsymbol{r}_3\times\boldsymbol{r}_1)$$
is a vector whose magnitude is the area of the triangle and direction is perpendicular to the triangle in a direction determined by the ordering P_1, P_2, P_3. Show that the result is independent of O. Sketch for four cases: (1) O is a vertex, (2) O is inside triangle, (3) O is outside triangle but in same plane, (4) O is not in the triangle plane.

5. Consider an arbitrary non-self intersecting quadrilateral in the (x,y) plane with vertices P_1,P_2,P_3,P_4. Show that
$$A\hat{\boldsymbol{n}}\triangleq\frac{1}{2}(\boldsymbol{r}_1\times\boldsymbol{r}_2+\boldsymbol{r}_2\times\boldsymbol{r}_3+\boldsymbol{r}_3\times\boldsymbol{r}_4+\boldsymbol{r}_4\times\boldsymbol{r}_1)$$
is a vector whose magnitude is the area of the quadrilateral and points in the $\pm\hat{\boldsymbol{z}}$ direction depending on whether P_1,P_2,P_3,P_4 are oriented counterclockwise, or clockwise, respectively, where $\boldsymbol{r}_n=\overrightarrow{OP_n}$ is the position vector of point P_n. What is the compact formula for the area of the quadrilateral in terms of the (x,y) coordinates of each point? Is the vector formula still valid if the points are in a plane but not the (x,y) plane?



6. Four arbitrary points in 3D Euclidean space $\mathbb{E}^3$ are the vertices of a tetrahedron. Consider the *area vectors* perpendicular pointing outward for each of the four faces of the tetrahedron, with magnitudes equal to the area of the corresponding triangular face. Show that the sum of these area vectors is zero.

7. Imagine 'gluing' together two tetrahedra with a common interior face that is removed. The resulting polyhedron has 6 triangular faces (hexahedron). Show that the sum of the outward pointing area vectors is zero. This result and procedure extend to more complex polyhedra and even to curved surfaces as can be shown with the *'gradient' theorem* in vector calculus.

8. Let V be the number of vertices, F the number of faces and E the number of edges of the polyhedron obtained by gluing together tetrahedra, as in the previous problem. Deduce Euler's formula that $V + F - E = 2$ for all such polyhedra.

9. Point P rotates at angular velocity ω about the axis parallel to $\boldsymbol{a}$ anchored at point A. Derive a vector formula for the *velocity* of P. Make sketches to illustrate your derivation.

10. Vector $\boldsymbol{c}$ is the (right hand) rotation of vector $\boldsymbol{b}$ about $\mathbf{e}_3$ by angle φ. Find the cartesian components of $\boldsymbol{c}$ in terms of the components of $\boldsymbol{b}$.

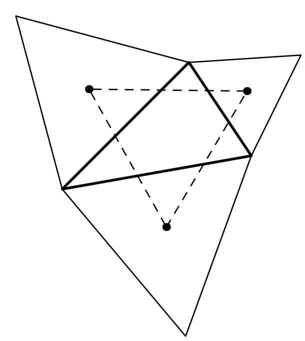
11. Find the cartesian components of the vector $\boldsymbol{c}$ obtained by rotating $\boldsymbol{b} = (b_1, b_2, b_3)$ about $\boldsymbol{a} = (a_1, a_2, a_3)$ by an angle α. What is $\boldsymbol{c}$ if $\boldsymbol{a} \equiv (3, 2, 1)$, $\boldsymbol{b} = (2, 3, 4)$ and $\alpha = \pi/3$? [Hint: consider the intrinsic orthogonal basis $\hat{\boldsymbol{a}}$, $\boldsymbol{b}_\perp$ and $\hat{\boldsymbol{a}} \times \boldsymbol{b}_\perp$ and obtain the vector solution first, *then* translate into cartesian components.]

12. Vector $\boldsymbol{b}'$ is the (right hand) rotation of vector $\boldsymbol{b}$ about $\boldsymbol{a}$ by angle φ. Show that

$$\boxed{\boldsymbol{b}' = \boldsymbol{b}_\parallel + \boldsymbol{b}_\perp \, \cos\varphi + (\hat{\boldsymbol{a}} \times \boldsymbol{b}_\perp) \, \sin\varphi.} \tag{82}$$

Write a computer code (in Matlab or Python, for instance) that finds $\boldsymbol{b}'$ given $\boldsymbol{a}$, $\boldsymbol{b}$, φ, assuming that the vectors are provided in cartesian form.

13. Test the rotation code (82) to make sure it is correct. Describe your tests. Find $\boldsymbol{b}'$ if $\boldsymbol{a} \equiv (3, 2, 1)$, $\boldsymbol{b} = (2, 3, 4)$ and $\alpha = \pi/3$.



14. For an arbitrary triangle, construct the outer equilateral triangles attached to each of the sides, in the plane of the original triangle. Show that the triangle connecting the centroids of those equilateral triangles is itself equilateral. [Hint: Use good cyclic notation, for instance $\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c} = 0$ for the original sides. Show that the vector connecting two centroids is the rotation by $2\pi/3$ of the previous such vector.]

15. Point P is rotated by angle φ about the axis parallel to $\boldsymbol{a}$ that passes through point A. Derive a vector formula for the new position of P. Make sketches to illustrate your derivation. Modify code (82) to handle rotations of points about axes.

16. True or false: $\boldsymbol{v} \perp (\boldsymbol{a} \times \boldsymbol{b}) \Leftrightarrow \boldsymbol{v} = x_1\boldsymbol{a} + x_2\boldsymbol{b}$ for some real x_1 and x_2. Explain.

17. Show by (1) cross product geometry and (2) cross product algebra that all the vectors $\boldsymbol{u}$ such that $\boldsymbol{a} \times \boldsymbol{u} = \boldsymbol{b}$ have the form

$$\boldsymbol{u} = \alpha\boldsymbol{a} + \frac{\boldsymbol{b} \times \boldsymbol{a}}{|\boldsymbol{a}|^2}, \quad \forall \alpha \in \mathbb{R}$$

18. Show the *Jacobi identity:* $\quad \boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) + \boldsymbol{b} \times (\boldsymbol{c} \times \boldsymbol{a}) + \boldsymbol{c} \times (\boldsymbol{a} \times \boldsymbol{b}) = 0.$

19. If $\hat{\boldsymbol{n}}$ is any unit vector, show algebraically *and* geometrically that any vector $\boldsymbol{a}$ can be decomposed as

$$\boxed{\boldsymbol{a} = (\hat{\boldsymbol{n}} \cdot \boldsymbol{a})\hat{\boldsymbol{n}} + \hat{\boldsymbol{n}} \times (\boldsymbol{a} \times \hat{\boldsymbol{n}}) \equiv \boldsymbol{a}_{\parallel} + \boldsymbol{a}_{\perp}.} \tag{83}$$

The first component is parallel to $\hat{\boldsymbol{n}}$, the second is perpendicular to $\hat{\boldsymbol{n}}$.

20. A left-handed basis $\mathbf{e}_1{}', \mathbf{e}_2{}', \mathbf{e}_3{}'$, is defined by $\mathbf{e}_i{}' \cdot \mathbf{e}_j{}' = \delta_{ij}$ and $\mathbf{e}_1{}' \times \mathbf{e}_2{}' = -\mathbf{e}_3{}'$. Show that $(\mathbf{e}_i{}' \times \mathbf{e}_j{}') \cdot \mathbf{e}_k{}'$ has the opposite sign to the corresponding expression for a right-handed basis, $\forall i, j, k$ (the definition of the cross-product remaining its right-hand rule self). Thus deduce that the formula for the components of the cross-product in the left handed basis would all change sign.

21. Show (75) from the geometric definition of the cross-product (66).

22. Starting from the identity for $(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c}$ derive the identity for $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c})$. Briefly explain your steps.

23. Explain the expansions in (77).

24. Prove (76) by brute force cartesian calculations.

25. Prove (76) using the intrinsic right-handed orthonormal basis $\mathbf{e}_1 = \boldsymbol{a}/|\boldsymbol{a}|$, $\mathbf{e}_3 = (\boldsymbol{a} \times \boldsymbol{b})/|\boldsymbol{a} \times \boldsymbol{b}|$ and $\mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_1$. Then $\boldsymbol{a} = a_1\mathbf{e}_1$, $\boldsymbol{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2$, $\boldsymbol{c} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$. Visualize and explain why this is a general result and therefore a proof of the double cross product identity.

26. Magnetic fields $\boldsymbol{B}$ are created by electric currents according to the Ampère and Biot-Savart laws. The simplest current is a moving charge. Consider two electric charges q_1 and q_2 moving at velocity $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively. Assume along the lines of the Biot-Savart law that the magnetic field induced by q_1 at q_2 is

$$\boldsymbol{B}_2 = \frac{\mu_0}{4\pi} \frac{q_1 \boldsymbol{v}_1 \times (\boldsymbol{r}_2 - \boldsymbol{r}_1)}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|^3} \tag{84}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the positions of q_1 and q_2, respectively, and μ_0 is the magnetic constant. The Lorentz force experienced by q_2 is $\boldsymbol{F}_2 = q_2\boldsymbol{v}_2 \times \boldsymbol{B}_2$. What is the corresponding magnetic field and Lorentz force induced by q_2 at q_1? Do the forces satisfy Newton's action-reaction law?

6 Cartesian index notation

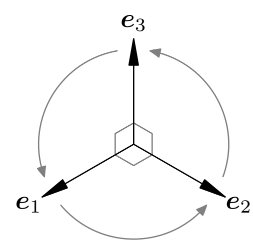
6.1 Levi-Civita symbol

We have used the Kronecker delta δ_{ij} (58) to express orthonormality of the cartesian basis vectors as $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. There is a similar symbol, ϵ_{ijk}, the *Levi-Civita* symbol (also known as the *alternating* or *permutation* symbol), defined as

$$\epsilon_{ijk} \triangleq \begin{cases} 1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3), \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3), \\ 0 & \text{otherwise}, \end{cases} \tag{85}$$

or, explicitly: $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$ and $\epsilon_{213} = \epsilon_{132} = \epsilon_{321} = -1$, all other $\epsilon_{ijk} = 0$. For 3 distinct elements, (a,b,c) say, an *even* permutation is the same as a *cyclic* permutation – for example, the cyclic permutation $(a,b,c) \to (b,c,a)$ is equivalent to the two permutations $(a,b,c) \to (b,a,c) \to (b,c,a)$. Thus the even permutations of $(1,2,3)$ are the cyclic permutations $(1,2,3)$, $(2,3,1)$, $(3,1,2)$ and the odd permutations are the acyclic permutations $(2,1,3)$, $(3,2,1)$, $(1,3,2)$. This implies that

$$\epsilon_{ijk} = \epsilon_{jki} = \epsilon_{kij}, \qquad \forall i,j,k \tag{86}$$



(why?). The ϵ_{ijk} symbol provides a compact expression for the components of the cross-product of right-handed cartesian basis vectors:

$$\boxed{(\mathbf{e}_i \times \mathbf{e}_j) \cdot \mathbf{e}_k = \epsilon_{ijk}.} \tag{87}$$

but since this is the k-component of $(\mathbf{e}_i \times \mathbf{e}_j)$ we can also write

$$\boxed{(\mathbf{e}_i \times \mathbf{e}_j) = \sum_{k=1}^{3} \epsilon_{ijk} \mathbf{e}_k.} \tag{88}$$

Note that there is only one non-zero term in the latter sum (but then, why can't we drop the sum?). Verify this result for yourself.

Sigma notation, free and dummy indices

The *expansion* of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in terms of basis $\mathbf{e}_1$,$\mathbf{e}_2$,$\mathbf{e}_3$, $\boldsymbol{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ and $\boldsymbol{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3$, can be written compactly using the sigma (Σ) notation

$$\boldsymbol{a} = \sum_{i=1}^{3} a_i \mathbf{e}_i, \qquad \boldsymbol{b} = \sum_{i=1}^{3} b_i \mathbf{e}_i. \tag{89}$$

We have introduced the **Kronecker** symbol δ_{ij} and the **Levi-Civita** symbol ϵ_{ijk} in order to write and perform our basic vector operations such as dot and cross products in compact forms, *when the basis is orthonormal and right-handed.* For example, using (58) and (88)

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{3}\sum_{j=1}^{3} a_i b_j \, \mathbf{e}_i \cdot \mathbf{e}_j = \sum_{i=1}^{3}\sum_{j=1}^{3} a_i b_j \delta_{ij} = \sum_{i=1}^{3} a_i b_i, \tag{90}$$

$$\boldsymbol{a} \times \boldsymbol{b} = \sum_{i=1}^{3}\sum_{j=1}^{3} a_i b_j \, \mathbf{e}_i \times \mathbf{e}_j = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} a_i b_j \epsilon_{ijk} \, \mathbf{e}_k. \tag{91}$$

Note that i and j are *dummy* or *summation* indices in the sums (89) and (90), they do not have a specific value, they have *all* the possible values in their range. It is their *place* in the particular expression and their *range* that matters, not their name

$$\boldsymbol{a} = \sum_{i=1}^{3} a_i \mathbf{e}_i = \sum_{j=1}^{3} a_j \mathbf{e}_j = \sum_{k=1}^{3} a_k \mathbf{e}_k = \cdots \neq \sum_{k=1}^{3} a_k \mathbf{e}_i \tag{92}$$

Indices come in two kinds, the *dummies* and the *free*. Here's an example

$$\mathbf{e}_i \cdot (\boldsymbol{a} \cdot \boldsymbol{b})\boldsymbol{c} = \left(\sum_{j=1}^{3} a_j b_j \right) c_i, \tag{93}$$

here j is a dummy summation index, but i is *free*, we can pick for it any value $1, 2, 3$. Freedom comes with constraints. If we use i on the left-hand side of the equation, then we have no choice, we must use i for c_i on the right hand side. By convention we try to use i, j, k, l, m, n, to denote indices, which are positive integers. Greek letters are sometimes used for indices.

Mathematical operations impose some naming constraints however. Although, we can use the same index name, i, in the expansions of $\boldsymbol{a}$ and $\boldsymbol{b}$, when they appear separately as in (89), we *cannot use the same index name if we multiply them* as in (90) and (91). Bad things will happen if you do, for instance

$$\boldsymbol{a} \times \boldsymbol{b} = \left(\sum_{i=1}^{3} a_i \mathbf{e}_i \right) \times \left(\sum_{i=1}^{3} b_i \mathbf{e}_i \right) = \sum_{i=1}^{3} a_i b_i \, \mathbf{e}_i \times \mathbf{e}_i = 0 \quad \text{(WRONG!)} \tag{94}$$

6.2 Einstein summation convention

While he was developing the theory of general relativity, Einstein noticed that many of the sums that occur in calculations involve terms where the summation index appears twice. For example, i appears twice in the single sums in (89), i and j appear twice in the double sum in (90) and i, j and k each appear twice in the triple sum in (91). To facilitate such manipulations he dropped the Σ signs and adopted the **summation convention** that **a repeated index implicitly denotes a sum over all values of that index**. In a letter to a friend he wrote *"I have made a great discovery in mathematics; I have suppressed the summation sign every time that the summation must be made over an index which occurs twice"*. If it had been an email, he would have punctuated it with ;-).

Thus with Einstein's summation convention we write the sum $a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ simply as $a_i\mathbf{e}_i$ since the index i is repeated

$$\boldsymbol{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 \equiv \; a_i\mathbf{e}_i. \tag{95}$$

The name of the index does not matter if it is repeated – it is a *dummy* or *summation* index, thus

$$a_i\mathbf{e}_i = a_j\mathbf{e}_j = a_k\mathbf{e}_k = a_l\mathbf{e}_l \ \equiv a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$$

and any repeated index i, j, k, l, $\ldots$ implies a sum over all values of that index. With the summation convention, the sum in (90) is written simply as

$$\boldsymbol{a}\cdot\boldsymbol{b} = a_ib_i \tag{96}$$

where $a_ib_i = a_1b_1 + a_2b_2 + a_3b_3$ is a sum over all values of i, while the triple sum in (91) reduces to the very compact form

$$\boldsymbol{a}\times\boldsymbol{b} = \epsilon_{ijk}a_ib_j\mathbf{e}_k, \tag{97}$$

which is a sum over all values of i, j and k and has $3^3 = 27$ terms. However $\epsilon_{ijk} = 0$ for $3^3 - 3! = 27 - 6 = 21$ of those terms, whenever an index value is repeated in the triplet (i, j, k). Note that these two index expressions for $\boldsymbol{a}\cdot\boldsymbol{b}$ and $\boldsymbol{a}\times\boldsymbol{b}$ assume that the underlying basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, is a right handed orthonormal basis.

The summation convention is a very useful and widely used notation but you have to use it with care – not write or read an i for a j or a 1 for an l, for examples – and there are cases where it cannot be used. Some basic rules facilitate manipulations.

Dummy repetition rule: Indices can never appear more than twice *in the same term*, if they are, that's probably a mistake as in (94),

$$\boldsymbol{a}\times\boldsymbol{b} = (a_i\mathbf{e}_i)\times(b_i\mathbf{e}_i) = a_ib_i\ \mathbf{e}_i\times\mathbf{e}_i = 0\ ??! \quad \text{(WRONG!)}$$

where i appears 4 times in the same term.[10] However

$$\begin{aligned} a_i + b_i + c_i &\equiv (a_1 + b_1 + c_1,\ a_2 + b_2 + c_2,\ a_3 + b_3 + c_3) \\ &\equiv \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} \end{aligned}$$

is OK since the index i appears in different terms and is in fact the index form for the vector sum $\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}$. In contrast, the expression

$$a_i + b_j + c_k$$

does not make vector sense since i, j, k are free indices here and there is no vector operation that adds components corresponding to different basis vectors — free indices in different terms must match. Going back to that dot product $\boldsymbol{a}\cdot\boldsymbol{b}$ in index notation, we need to change the name of one of the dummies, for example $i \to j$ in the $\boldsymbol{b}$ expansion, and

$$\boldsymbol{a}\cdot\boldsymbol{b} = (a_i\mathbf{e}_i)\cdot(b_j\mathbf{e}_j) = a_ib_j\ \mathbf{e}_i\cdot\mathbf{e}_j = a_ib_j\delta_{ij} = a_ib_i = a_jb_j.$$

Substitution rule: if one of the indices of δ_{ij} is involved in a sum, we substitute the summation index for the other index in δ_{ij} and drop δ_{ij}, for example

$$a_i\delta_{ij} \equiv a_1\delta_{1j} + a_2\delta_{2j} + a_3\delta_{3j} = a_j, \tag{98}$$

[10] *Terms* are elements of a sum, *factors* are elements of a product.

since i is a dummy in this example and δ_{ij} eliminates all terms in the sum except that corresponding to index j, whatever its value, thus $a_i\delta_{ij} = a_j$. If both indices of δ_{ij} are summed over as in the double sum $a_ib_j\delta_{ij}$, it does not matter which index we substitute for, thus

$$a_ib_j\delta_{ij} = a_ib_i = a_jb_j$$

and likewise

$$\delta_{kl}\delta_{kl} = \delta_{kk} = \delta_{ll} = 3, \qquad \delta_{ij}\epsilon_{ijk} = \epsilon_{iik} = 0. \tag{99}$$

Note the result $\delta_{kk} = 3$ because k is repeated, so there is a sum over all values of k and $\delta_{kk} \equiv \delta_{11} + \delta_{22} + \delta_{33} = 1 + 1 + 1 = 3$, not 1. The last result is because ϵ_{ijk} vanishes whenever two indices have the same value.

Let's compute the l component of $\boldsymbol{a}\times\boldsymbol{b}$ from (97) as an exercise. We pick l because i, j and k are already taken. The l component is

$$\mathbf{e}_l \cdot (\boldsymbol{a}\times\boldsymbol{b}) = \epsilon_{ijk}a_ib_j\mathbf{e}_k\cdot\mathbf{e}_l = \epsilon_{ijk}a_ib_j\delta_{kl} = \epsilon_{ijl}a_ib_j = \epsilon_{lmn}a_mb_n \tag{100}$$

what happened on that last step? first, $\epsilon_{ijk} = \epsilon_{kij}$ because (i, j, k) to (k, i, j) is a cyclic permutation which is an even permutation in 3D space and the value of ϵ_{ijk} does not change under even permutations. Then i and j are dummies and we renamed them m and n respectively being careful to keep the place of the indices. The final result is worth memorizing: if $\boldsymbol{v} = \boldsymbol{a}\times\boldsymbol{b}$, the l component of $\boldsymbol{v}$ is $v_l = \epsilon_{lmn}a_mb_n$, or switching indices to i, j, k

$$\boldsymbol{v} = \boldsymbol{a}\times\boldsymbol{b} \iff v_i = \epsilon_{ijk}a_jb_k \iff \boldsymbol{v} = \mathbf{e}_i\,\epsilon_{ijk}a_jb_k. \tag{101}$$

As an another example that will lead us to a fundamental identity, let's write the double cross product identity $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ in index notation. Let $\boldsymbol{v} = \boldsymbol{a}\times\boldsymbol{b}$ then the i component of the double cross product $\boldsymbol{v}\times\boldsymbol{c}$ is $\epsilon_{ijk}v_jc_k$. Now we need the j component of $\boldsymbol{v} = \boldsymbol{a}\times\boldsymbol{b}$. Since i and k are taken we use l, m as new dummy indices, and we have $v_j = \epsilon_{jlm}a_lb_m$. So the i component of the double cross product $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$ is

$$(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c} \equiv\; \epsilon_{ijk}\epsilon_{jlm}\,a_lb_mc_k. \tag{102}$$

Note that j, k, l and m are repeated, so this expression is a quadruple sum! According to our double cross product identity it should be equal to the i component of $(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}$ *for any* $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$. We want the i component of the latter expression since i is a free index in (102), that i component is

$$(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a} \equiv\; (a_jc_j)\,b_i - (b_jc_j)\,a_i \tag{103}$$

wait! isn't j repeated 4 times? no, it's not. It's repeated twice in *separate* terms so this is a difference of two sums over j, j is a dummy but i is free and must match in both terms. Since (102) and (103) are equal to each other for any $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, this should be telling us something about ϵ_{ijk}, but to extract that out we need to rewrite (103) in the form $a_lb_mc_k$. How? by making use of our ability to rename dummy variables and adding variables using δ_{ij} and the substitution rule. Let's look at the first term in (103), $(a_jc_j)b_i$, here's how to write it in the form $a_lb_mc_k$ as in (102):

$$(a_jc_j)b_i = (a_kc_k)b_i = (\delta_{lk}a_lc_k)(\delta_{im}b_m) = \delta_{lk}\delta_{im}a_lc_kb_m. \tag{104}$$

Do similar manipulations to the second term in (103) to obtain $(b_j c_j)a_i = \delta_{il}\delta_{km} a_l c_k b_m$ and

$$\epsilon_{ijk}\epsilon_{jlm} a_l b_m c_k = (\delta_{lk}\delta_{im} - \delta_{il}\delta_{km}) a_l c_k b_m. \tag{105}$$

Since this equality holds *for any* a_l, c_k, b_m, we must have $\epsilon_{ijk}\epsilon_{jlm} = (\delta_{lk}\delta_{im} - \delta_{il}\delta_{km})$. That's true but it's not written in a nice way so let's clean it up to a form that's easier to reconstruct. First note that $\epsilon_{ijk} = \epsilon_{jki}$ since ϵ_{ijk} is invariant under a cyclic permutation of its indices. So our identity becomes $\epsilon_{jki}\epsilon_{jlm} = (\delta_{lk}\delta_{im} - \delta_{il}\delta_{km})$. We've done that flipping so the summation index j is in first place in both ϵ factors. Now we prefer the lexicographic order (i, j, k) to (j, k, i) so let's rename all the indices being careful to rename the correct indices on both sides. This yields

$$\boxed{\epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}} \tag{106}$$

This takes some digesting but it is an excellent exercise and example of index notation manipulations.

The identity (106) is actually pretty easy to remember and verify. First, $\epsilon_{ijk}\epsilon_{ilm}$ is a sum over i but there is never more than one non-zero term (why?). Second, the only possible values for that expression are $+1$, 0 and -1 (why?). The only way to get 1 is to have $(j, k) = (l, m)$ with $j = l \neq k = m$ (why?), but in that case the right hand side of (106) is also 1 (why?). The only way to get -1 is to have $(j, k) = (m, l)$ with $j = m \neq k = l$ (why?), but in that case the right hand side is -1 also (why?). Finally, to get 0 we need $j = k$ or $l = m$ and the right-hand side again vanishes in either case. For instance, if $j = k$ then we can switch j and k in one of the terms and $\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} = \delta_{jl}\delta_{km} - \delta_{km}\delta_{jl} = 0$.

Formula (106) has a generalization that does not include summation over one index

$$\begin{aligned}\epsilon_{ijk}\epsilon_{lmn} =& \delta_{il}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{kl} + \delta_{in}\delta_{jl}\delta_{km} \\ & -\delta_{im}\delta_{jl}\delta_{kn} - \delta_{in}\delta_{jm}\delta_{kl} - \delta_{il}\delta_{jn}\delta_{km}\end{aligned} \tag{107}$$

note that the first line correspond to (i, j, k) and (l, m, n) matching up to cyclic rotations, while the second line corresponds to (i, j, k) matching with an odd (acyclic) rotation of (l, m, n).

Exercises

1. Explain why $\epsilon_{ijk} = \epsilon_{jki} = -\epsilon_{jik}$ for any integer i, j, k.
2. Using (87) and Einstein's summation convention to show that $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \mathbf{e}_k = \epsilon_{ijk} a_i b_j = \epsilon_{kij} a_i b_j$ and $(\boldsymbol{a} \times \boldsymbol{b}) = \epsilon_{ijk} a_i b_j\, \mathbf{e}_k = \mathbf{e}_i\, \epsilon_{ijk} a_j b_k$.
3. Show that $\epsilon_{ijk}\epsilon_{ijk} = 6$ by direct deduction and by application of (106).
4. Show that $\epsilon_{ijk}\epsilon_{ljk} = 2\delta_{il}$ by direct deduction and by application of (106).
5. Deduce (106) from (107).
6. True or False?: $\epsilon_{ijk} a_i b_j c_k = \epsilon_{lmn} a_m b_n c_l$. Explain.

7. Consider $\Omega_{ij} = -\Omega_{ji}$ (that is Ω_{ij} is an *antisymmetric tensor*). If i and j take independent values in $\{1, 2, 3\}$ so that there are 9 Ω_{ij}'s, *a priori*, show that Ω_{ij} has in fact only three independent components. Show that those three components can be defined as $\omega_k = \frac{1}{2}\epsilon_{kij}\Omega_{ij}$ and then $\Omega_{ij} = \epsilon_{ijk}\omega_k$.

8. Show that $\boldsymbol{\omega} \times \mathbf{e}_i = \epsilon_{ijk}\omega_k\mathbf{e}_j$ where $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a right handed orthonormal basis and $\omega_k = \boldsymbol{\omega} \cdot \mathbf{e}_k$.

9. If $\Omega_{ij} = -\Omega_{ji}$, show that $\Omega_{ij}\mathbf{e}_j = \boldsymbol{\omega} \times \mathbf{e}_i$ where $\omega_k = \mathbf{e}_k \cdot \boldsymbol{\omega} = \frac{1}{2}\epsilon_{kij}\Omega_{ij}$.

10. If $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{a}$, show that $v_i = \Omega_{ij}a_j$ where $\Omega_{ij} = \epsilon_{ikj}\omega_k = -\Omega_{ji}$.

11. Let $\boldsymbol{v} = (\boldsymbol{a} \cdot \boldsymbol{b})\boldsymbol{a}$ yielding $v_i = a_{ij}b_j$ in a cartesian basis. Find a_{ij} and show that $a_{ij} = a_{ji}$ so that a_{ij} is symmetric.

12. Let $\boldsymbol{v} = \boldsymbol{b} - 2(\hat{\boldsymbol{a}} \cdot \boldsymbol{b})\hat{\boldsymbol{a}}$ yielding $v_i = a_{ij}b_j$ in a cartesian basis. Find a_{ij}. What is the geometric relation between $\boldsymbol{v}$ and $\boldsymbol{b}$?

13. Let $\boldsymbol{v} = \boldsymbol{a} \times \boldsymbol{b}$ yielding $v_i = a_{ij}b_j$ in a cartesian basis. Find a_{ij} and show that $a_{ij} = -a_{ji}$ so that a_{ij} is anti-symmetric.

14. Let $\boldsymbol{v} = (\hat{\boldsymbol{a}} \times \boldsymbol{b}) \times \hat{\boldsymbol{a}}$ yielding $v_i = a_{ij}b_j$ in a cartesian basis. Find a_{ij}. What is the geometric relation between $\boldsymbol{v}$ and $\boldsymbol{b}$?

7 Mixed product and Determinant

A mixed product, also called the *box product* or the *'triple scalar product'*, is a combination of a cross and a dot product, $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$, the result is a *scalar*. We have already encountered mixed products (*e.g.* eqn. (87)) but their geometric and algebraic properties are so important that they merit their own subsection.

$$\boxed{\begin{array}{c}(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = (\boldsymbol{b} \times \boldsymbol{c}) \cdot \boldsymbol{a} = (\boldsymbol{c} \times \boldsymbol{a}) \cdot \boldsymbol{b} = \\ \textit{signed} \text{ volume of the parallelepiped } \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\end{array}} \tag{108}$$

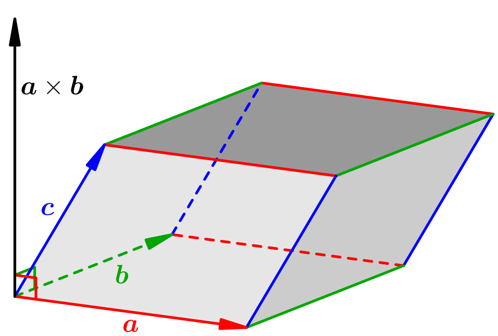


Fig. 1.27: $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$ is the signed volume of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$.

To derive the fundamental identity (108), take $\boldsymbol{a}$ and $\boldsymbol{b}$ as the base of the parallelepiped then $\boldsymbol{a} \times \boldsymbol{b} = A\hat{\boldsymbol{n}}$ has magnitude A equal to area of the base parallelogram and direction $\hat{\boldsymbol{n}}$ perpendicular to that base. The height h of the parallelepiped is simply $h = \hat{\boldsymbol{n}} \cdot \boldsymbol{c}$, thus the volume is indeed

$$A\,h = A\hat{\boldsymbol{n}} \cdot \boldsymbol{c} = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}.$$

Signwise, $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} > 0$ if $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, in that order, form a right-handed basis (not orthogonal in general), and $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} < 0$ if the triplet is left-handed. Taking $\boldsymbol{b}$ and $\boldsymbol{c}$, or $\boldsymbol{c}$ and $\boldsymbol{a}$, as the base, yields the same volume and sign. The dot product commutes, so $(\boldsymbol{b} \times \boldsymbol{c}) \cdot \boldsymbol{a} = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$, yielding the identity

$$(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}). \tag{109}$$

That is nice and easy! we can switch the dot and the cross without changing the result. In summary, the mixed product is invariant to a *cyclic permutation* of the vectors $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = (\boldsymbol{b} \times \boldsymbol{c}) \cdot \boldsymbol{a}$, and to a swap of the dot and cross $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$.

We have shown (108) and (109) geometrically. The properties of the dot and cross products yield many other results such as $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = -(\boldsymbol{b} \times \boldsymbol{a}) \cdot \boldsymbol{c}$, etc. There are 12 different ways to write a mixed product of the vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ but they are all equal to $\pm$ the volume of the parallelepiped; 6 are positive with right-handed ordering, 6 are negative with left-handed ordering. We can collect all these algebraic properties as follows.

A mixed product is a scalar function of three vectors called the *determinant*

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}, \tag{110}$$

whose value is the *signed volume* of the parallelepiped with sides $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$. The determinant has three fundamental properties

1. it changes sign if *any* two vectors are permuted, *e.g.*

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = -\det(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}) = \det(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}), \tag{111}$$

2. it is linear in *any* of its vectors *e.g.* $\forall\ \alpha$, $\boldsymbol{d}$,

$$\det(\alpha\boldsymbol{a} + \boldsymbol{d}, \boldsymbol{b}, \boldsymbol{c}) = \alpha \det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) + \det(\boldsymbol{d}, \boldsymbol{b}, \boldsymbol{c}), \tag{112}$$

3. if the triplet $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is right-handed and orthonormal then

$$\det(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = 1. \tag{113}$$

You can deduce these from the properties of the dot and cross products as well as geometrically. Property (112) is a combination of the distributivity properties of the dot and cross products with respect to vector addition and multiplication by a scalar. For example,

$$\begin{aligned}\det(\alpha\boldsymbol{a} + \boldsymbol{d}, \boldsymbol{b}, \boldsymbol{c}) = (\alpha\boldsymbol{a} + \boldsymbol{d}) \cdot (\boldsymbol{b} \times \boldsymbol{c}) &= \alpha\big(\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})\big) + \boldsymbol{d} \cdot (\boldsymbol{b} \times \boldsymbol{c}) \\ &= \alpha \det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) + \det(\boldsymbol{d}, \boldsymbol{b}, \boldsymbol{c}).\end{aligned}$$

From these three properties, you deduce easily that the determinant is zero if any two vectors are identical (from prop 1), or if any vector is zero (from prop 2 with $\alpha = 1$ and $\boldsymbol{d} = \boldsymbol{0}$), and that the determinant does not change if we add a multiple of one vector to another, for example

$$\begin{aligned} \det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}) &= 0, \\ \det(\boldsymbol{a}, \boldsymbol{0}, \boldsymbol{c}) &= 0, \\ \det(\boldsymbol{a} + \beta\boldsymbol{b}, \boldsymbol{b}, \boldsymbol{c}) &= \det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}). \end{aligned} \tag{114}$$

Geometrically, this last one corresponds to a *shearing* of the parallelepiped, with no change in volume or orientation.

The *determinant* determines whether three vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are linearly independent and can be used as a basis for the vector space

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \neq 0 \;\Leftrightarrow\; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \text{ form a basis.} \tag{115}$$

If $\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = 0$ then either one of the vectors is zero or they are co-planar and $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ cannot provide a basis for vectors in $\mathbb{E}^3$. This is how the determinant is introduced in elementary linear algebra, it determines whether a system of linear equations has a solution or not. But the determinant is much more than a number that may or may not be zero, it 'determines' the volume of the parallelepiped and its orientation!

The 3 fundamental properties fully specify the determinant as explored in exercises 5, 6 below. If the vectors are expanded in terms of a right-handed orthonormal basis, *i.e.* $\boldsymbol{a} = a_i\mathbf{e}_i$, $\boldsymbol{b} = b_j\mathbf{e}_j$, $\boldsymbol{c} = c_k\mathbf{e}_k$ (summation convention), then we obtain the following formula for the determinant in terms of the vector components

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = a_i b_j c_k (\mathbf{e}_i \times \mathbf{e}_j) \cdot \mathbf{e}_k = \epsilon_{ijk}\, a_i b_j c_k. \tag{116}$$

Expanding that expression

$$\epsilon_{ijk}\, a_i b_j c_k = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_2b_1c_3 - a_3b_2c_1 - a_1b_3c_2, \tag{117}$$

we recover the familiar algebraic determinants

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \epsilon_{ijk}\, a_i b_j c_k \equiv \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \tag{118}$$

Note that it does not matter whether we put the vector components along rows or columns. This is a non-trivial and important property of determinants, that the determinant of a matrix is the determinant of its transpose (see section on matrices).

This familiar determinant has the same three fundamental properties (111), (112), (113) of course

1. it changes sign if *any* two columns (or rows) are permuted, *e.g.*

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} b_1 & a_1 & c_1 \\ b_2 & a_2 & c_2 \\ b_3 & a_3 & c_3 \end{vmatrix}, \tag{119}$$

2. it is linear in *any* of its columns (or rows) *e.g.* $\forall\ \alpha$, (d_1, d_2, d_3),

$$\begin{vmatrix} \alpha a_1 + d_1 & b_1 & c_1 \\ \alpha a_2 + d_2 & b_2 & c_2 \\ \alpha a_3 + d_3 & b_3 & c_3 \end{vmatrix} = \alpha \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}, \tag{120}$$

3. finally, the determinant of the *natural basis* is

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1. \tag{121}$$

You can deduce from these three properties that the determinant vanishes if any column (or row) is zero or if any two columns (or rows) is a multiple of another, and that the determinant does not change if we add to one column (row) a linear combination of the other columns (rows). These properties allow us to calculate determinants by successive shearings and column-swapping.

There is another explicit formula for determinants, in addition to the $\epsilon_{ijk}a_ib_jc_k$ formula, it is the *Laplace (or Cofactor) expansion* in terms of 2-by-2 determinants, *e.g.*

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \tag{122}$$

where the 2-by-2 determinants are

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1. \tag{123}$$

This formula is nothing but $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$ expressed with respect to a right handed basis. To verify that, compute the components of $(\boldsymbol{b} \times \boldsymbol{c})$ first, then dot with the components of $\boldsymbol{a}$. This cofactor expansion formula can be applied to any column or row, however there are ± 1 factors that appear. We won't go into the straightforward details, but all that follows directly from the column swapping property (119). That's essentially the identities $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{c} \times \boldsymbol{a}) = \cdots$.

Exercises

1. Show that the 2-by-2 determinant $\left|\begin{smallmatrix} a_1 & b_1 \\ a_2 & b_2 \end{smallmatrix}\right| = a_1 b_2 - a_2 b_1$, is the *signed area* of the parallelogram with sides $\boldsymbol{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$ and $\boldsymbol{b} = b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2$. It is positive if $\boldsymbol{a}, \boldsymbol{b}, -\boldsymbol{a}, -\boldsymbol{b}$ is a counterclockwise cycle, negative if the cycle is clockwise. Sketch.

2. The determinant $\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ of three oriented line segments $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ is a geometric quantity. Show that $\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = |\boldsymbol{a}|\,|\boldsymbol{b}|\,|\boldsymbol{c}| \sin\phi \cos\theta$. Specify ϕ and θ. Sketch.

3. Show that $-|\boldsymbol{a}|\,|\boldsymbol{b}|\,|\boldsymbol{c}| \le \det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \le |\boldsymbol{a}|\,|\boldsymbol{b}|\,|\boldsymbol{c}|$. When do the equalities apply? Sketch.

4. Use properties (111) and (112) to show that

$$\det(\alpha\boldsymbol{a}+\lambda\boldsymbol{d},\beta\boldsymbol{b}+\mu\mathbf{e},\boldsymbol{c}) =$$

$$\alpha\beta\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})+\alpha\mu\det(\boldsymbol{a},\mathbf{e},\boldsymbol{c})+\beta\lambda\det(\boldsymbol{d},\boldsymbol{b},\boldsymbol{c})+\lambda\mu\det(\boldsymbol{d},\mathbf{e},\boldsymbol{c}).$$

5. Use properties (111) and (113) to show that $\det(\mathbf{e}_i,\mathbf{e}_j,\mathbf{e}_k)=\epsilon_{ijk}$.

6. Use property (112) and exercise 5 above to show that if $\boldsymbol{a}=a_i\mathbf{e}_i$, $\boldsymbol{b}=b_i\mathbf{e}_i$, $\boldsymbol{c}=c_i\mathbf{e}_i$ *(summation convention)* then $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\epsilon_{ijk}a_ib_jc_k$.

7. Prove the identity $(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{c}\times\boldsymbol{d})=(\boldsymbol{a}\cdot\boldsymbol{c})(\boldsymbol{b}\cdot\boldsymbol{d})-(\boldsymbol{a}\cdot\boldsymbol{d})(\boldsymbol{b}\cdot\boldsymbol{c})$ using both vector identities and indicial notation.

8. Express $(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{a}\times\boldsymbol{b})$ in terms of dot products of $\boldsymbol{a}$ and $\boldsymbol{b}$.

9. Show that $(\boldsymbol{a}\cdot\boldsymbol{a})(\boldsymbol{b}\cdot\boldsymbol{b})-(\boldsymbol{a}\cdot\boldsymbol{b})^2$ is the square of the area of the parallelogram spanned by $\boldsymbol{a}$ and $\boldsymbol{b}$.

10. If A is the area the parallelogram with sides $\boldsymbol{a}$ and $\boldsymbol{b}$, show that

$$A^2=\begin{vmatrix}\boldsymbol{a}\cdot\boldsymbol{a} & \boldsymbol{a}\cdot\boldsymbol{b}\\ \boldsymbol{a}\cdot\boldsymbol{b} & \boldsymbol{b}\cdot\boldsymbol{b}\end{vmatrix}.$$

11. If $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\neq 0$, then any vector $\boldsymbol{v}$ can be expanded as $\boldsymbol{v}=\alpha\boldsymbol{a}+\beta\boldsymbol{b}+\gamma\boldsymbol{c}$. Find explicit expressions for the components α,β,γ in terms of $\boldsymbol{v}$ and the basis vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ in the general case when the latter are *not* orthogonal. [Hint: project on cross products of the basis vectors, then collect the mixed products into determinants and deduce *Cramer's rule*].

12. Given three vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ such that $D=\boldsymbol{a}_1\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)\neq 0$, define

$$\boldsymbol{a}_1'=\frac{1}{D}\,\boldsymbol{a}_2\times\boldsymbol{a}_3,\quad \boldsymbol{a}_2'=\frac{1}{D}\,\boldsymbol{a}_3\times\boldsymbol{a}_1,\quad \boldsymbol{a}_3'=\frac{1}{D}\,\boldsymbol{a}_1\times\boldsymbol{a}_2. \tag{124}$$

This is the **reciprocal basis** of the basis $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$.

(i) Show that $\boldsymbol{a}_i\cdot\boldsymbol{a}_j'=\delta_{ij},\ \forall\, i,j=1,2,3$.

(ii) Show that if $\boldsymbol{v}=v_i\,\boldsymbol{a}_i$ and $\boldsymbol{v}=v_i'\,\boldsymbol{a}_i'$ (summation convention), then $v_i=\boldsymbol{v}\cdot\boldsymbol{a}_i'$ and $v_i'=\boldsymbol{v}\cdot\boldsymbol{a}_i$. So the components in one basis are obtained by projecting onto the other basis.

13. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly independent and $\boldsymbol{c}$ is any arbitrary vector, find α, β and γ such that $\boldsymbol{c}=\alpha\boldsymbol{a}+\beta\boldsymbol{b}+\gamma(\boldsymbol{a}\times\boldsymbol{b})$. Express α, β and γ in terms of dot products only. [Hint: find α and β first, then use $\boldsymbol{c}_\parallel=\boldsymbol{c}-\boldsymbol{c}_\perp$.]

14. Express $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ in terms of dot products of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ only. [Hint: solve problem 13 first.]

15. Provide an algorithm to compute the volume of the parallelepiped ($\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$) by taking only dot products. [Hint: 'rectify' the parallelepiped $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) \to (\boldsymbol{a},\boldsymbol{b}_\perp,\boldsymbol{c}_\perp) \to (\boldsymbol{a},\boldsymbol{b}_\perp,\boldsymbol{c}_{\perp\perp})$ where $\boldsymbol{b}_\perp$ and $\boldsymbol{c}_\perp$ are perpendicular to $\boldsymbol{a}$, and $\boldsymbol{c}_{\perp\perp}$ is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}_\perp$. Explain geometrically why these transformations do not change the volume. Explain why these transformations do not change the determinant by using the properties of determinants.]

16. (*) If V is the volume of the parallelepiped with sides $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ show that

$$V^2 = \begin{vmatrix} \boldsymbol{a}\cdot\boldsymbol{a} & \boldsymbol{a}\cdot\boldsymbol{b} & \boldsymbol{a}\cdot\boldsymbol{c} \\ \boldsymbol{b}\cdot\boldsymbol{a} & \boldsymbol{b}\cdot\boldsymbol{b} & \boldsymbol{b}\cdot\boldsymbol{c} \\ \boldsymbol{c}\cdot\boldsymbol{a} & \boldsymbol{c}\cdot\boldsymbol{b} & \boldsymbol{c}\cdot\boldsymbol{c} \end{vmatrix}.$$

Do this in several ways: (i) from problem 13, (ii) using indicial notation and the formula (107).

8 Points, Lines, Planes, etc.

We discussed points, lines and planes in section 3 and reviewed the concepts of position vector

$$\overrightarrow{OP} = \boldsymbol{r} = r\hat{\boldsymbol{r}} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$$

and parametric equations of lines and planes. Here, we briefly review *implicit* equations of lines and planes and some applications of dot and cross products to points, lines and planes.

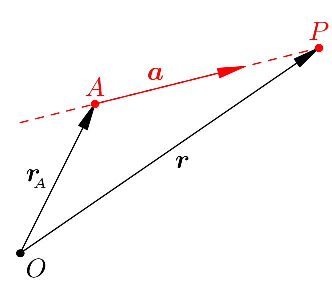
Center of mass. The center of mass, $\boldsymbol{r}_c$, of a system of N particles of mass m_i located at position $\boldsymbol{r}_i$, $i = 1,\dots,N$, is the *mass averaged position* defined by

$$M\,\boldsymbol{r}_c \triangleq \sum_{i=1}^{N} m_i \boldsymbol{r}_i \tag{125}$$

where $M = \sum_{i=1}^{N} m_i$ is the total mass. In particular, if all the masses are equal then for $N=2$, $\boldsymbol{r}_c = (\boldsymbol{r}_1+\boldsymbol{r}_2)/2$, for $N=3$, $\boldsymbol{r}_c = (\boldsymbol{r}_1+\boldsymbol{r}_2+\boldsymbol{r}_3)/3$. Note that we do not use the summation convention for the sum over the N particles. This N is not the dimension of the space in which the particles are located.

Equations of lines. The vector equation of a line parallel to $\boldsymbol{a}$ passing through a point A is

$$\overrightarrow{AP} = t\,\boldsymbol{a}, \quad \forall t \in \mathbb{R} \tag{126}$$



This vector equation expresses that the vector $\overrightarrow{AP}$ is parallel to $\boldsymbol{a}$, where t is a real parameter. In terms of an origin O we have $\overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{AP}$ and we can write the vector (parametric) equation of that same line as

$$\boldsymbol{r} = \boldsymbol{r}_A + t\,\boldsymbol{a}, \tag{127}$$

where $\boldsymbol{r} = \overrightarrow{OP}$ and $\boldsymbol{r}_A = \overrightarrow{OA}$ are the position vectors of P and A with respect to O, respectively.

The real number t is the parameter of the line, it is the coordinate of P in the system A, $\boldsymbol{a}$. We can eliminate that parameter by crossing the parametric equation with $\boldsymbol{a}$:

$$\boxed{\boldsymbol{r} = \boldsymbol{r}_A + t\boldsymbol{a}, \quad \forall t \in \mathbb{R}} \Leftrightarrow \boxed{(\boldsymbol{r} - \boldsymbol{r}_A) \times \boldsymbol{a} = 0.} \tag{128}$$

This is the *(explicit) parametric equation* $\boldsymbol{r} = \boldsymbol{r}_A + t\boldsymbol{a}$, with parameter t, and the *implicit* equation $(\boldsymbol{r} - \boldsymbol{r}_A) \times \boldsymbol{a} = 0$ of a line.

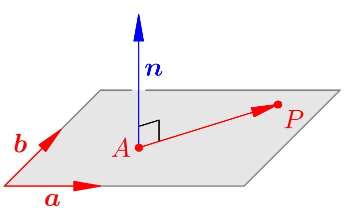
In cartesian coordinates, $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$, $\boldsymbol{r}_A = x_A\hat{\boldsymbol{x}} + y_A\hat{\boldsymbol{y}} + z_A\hat{\boldsymbol{z}}$ and $\boldsymbol{a} = a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}} + a_z\hat{\boldsymbol{z}}$, thus the parametric vector equation $\boldsymbol{r} = \boldsymbol{r}_A + t\boldsymbol{a}$ corresponds to component equations

$$\begin{cases} x = x_A + t\, a_x, \\ y = y_A + t\, a_y, \\ z = z_A + t\, a_z, \end{cases}$$

with one parameter t, while the implicit equation $(\boldsymbol{r} - \boldsymbol{r}_A) \times \boldsymbol{a} = 0$ corresponds to the 2 equations

$$\frac{x - x_A}{a_x} = \frac{y - y_A}{a_y} = \frac{z - z_A}{a_z}.$$

Equations of planes. The equation of a plane passing through point A, parallel to $\boldsymbol{a}$ and $\boldsymbol{b}$ (with $\boldsymbol{a} \times \boldsymbol{b} \neq 0$) is



$$\overrightarrow{AP} = t_1\boldsymbol{a} + t_2\boldsymbol{b}, \quad \forall t_1, t_2 \in \mathbb{R}$$

or $\boldsymbol{r} = \boldsymbol{r}_A + t_1\boldsymbol{a} + t_2\boldsymbol{b}$ since $\boldsymbol{r} = \overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{AP}$. The parameters t_1 and t_2 can be eliminated by dotting the parametric equation with $\boldsymbol{n} = \boldsymbol{a} \times \boldsymbol{b}$:

$$\boxed{\boldsymbol{r} = \boldsymbol{r}_A + t_1\boldsymbol{a} + t_2\boldsymbol{b}, \quad \forall t_1, t_2 \in \mathbb{R}} \Leftrightarrow \boxed{(\boldsymbol{r} - \boldsymbol{r}_A) \cdot \boldsymbol{n} = 0.} \tag{129}$$

This is the parametric equation of the plane with parameters t_1 and t_2, and the implicit equation of the plane passing through A and perpendicular to $\boldsymbol{n}$.

In cartesian coordinates, the vector equation $\boldsymbol{r} = \boldsymbol{r}_A + t_1\boldsymbol{a} + t_2\boldsymbol{b}$ corresponds to component equations

$$\begin{cases} x = x_A + t_1 a_x + t_2 b_x, \\ y = y_A + t_1 a_y + t_2 b_y, \\ z = z_A + t_1 a_z + t_2 b_z, \end{cases}$$

with 2 parameters (t_1, t_2), while the implicit equation $(\boldsymbol{r} - \boldsymbol{r}_A) \cdot \boldsymbol{n} = 0$ corresponds to 1 equation

$$n_x(x - x_A) + n_y(y - y_A) + n_z(z - z_A) = 0,$$

where $n_x = (a_y b_z - a_z b_y)$, $n_y = (a_z b_x - a_x b_z)$, $n_z = (a_x b_y - a_y b_x)$.

Equations of spheres. The equation of a sphere of center $\boldsymbol{r}_c$ and radius R is

$$\boxed{|\boldsymbol{r} - \boldsymbol{r}_c| = R} \Leftrightarrow \boxed{\boldsymbol{r} = \boldsymbol{r}_c + R\,\hat{\boldsymbol{a}},} \quad \forall \hat{\boldsymbol{a}} \text{ s.t. } |\hat{\boldsymbol{a}}| = 1, \tag{130}$$

where $\hat{\boldsymbol{a}}$ is any direction in 3D space. We have seen in eqn. (11) that such a direction can be expressed as

$$\hat{\boldsymbol{a}}(\theta, \varphi) = \cos\varphi \sin\theta \, \hat{\boldsymbol{x}} + \sin\varphi \sin\theta \, \hat{\boldsymbol{y}} + \cos\theta \, \hat{\boldsymbol{z}},$$

where $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ are any set of mutually orthogonal unit vectors. The angles θ and φ are the 2 parameters appearing in this parametrization of a sphere.

[ADD: one or two solved examples, intersection between line and plane, intersections between 2 planes, distance from point to plane,...]

Exercises

1. Show that the center of gravity of three points of equal mass is at the point of intersection of the medians of the triangle formed by the three points.

2. Show that the center of mass G is independent of the origin O.

3. Find vector equations for the line passing through the two points $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and the plane through the three points $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$.

4. What is the distance between the point $\boldsymbol{r}_1$ and the plane through $\boldsymbol{r}_0$ perpendicular to $\boldsymbol{a}$?

5. What is the distance between the point $\boldsymbol{r}_1$ and the plane through $\boldsymbol{r}_0$ parallel to $\boldsymbol{a}$ and $\boldsymbol{b}$?

6. What is the distance between the line parallel to $\boldsymbol{a}$ that passes through point A and the line parallel to $\boldsymbol{b}$ that passes through point B?

7. A particle was at point P_1 at time t_1 and is moving at the constant velocity $\boldsymbol{v}_1$. Another particle was at P_2 at t_2 and is moving at the constant velocity $\boldsymbol{v}_2$. How close did the particles get to each other and at what time? What conditions are needed for a collision?

8. Point C is obtained by rotating point B about the axis passing through point A, with direction $\boldsymbol{a}$, by angle α (right hand rotation by α about $\boldsymbol{a}$). Find an explicit vector expression for $\overrightarrow{OC}$ in terms of $\overrightarrow{OB}$, $\overrightarrow{OA}$, $\boldsymbol{a}$ and α. Make clean sketches. Express your vector result in cartesian form.

9 Orthogonal Transformations and Matrices

9.1 Change of cartesian basis in 2D

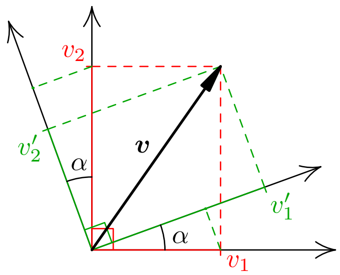


Fig. 1.28: Two cartesian representations of $\boldsymbol{v}$.

A vector $\boldsymbol{v}$ is represented by two distinct sets of cartesian components (v_1, v_2) and (v_1', v_2') as illustrated in fig. 1.28. In 2D, we can find the relations between the components (v_1, v_2) and (v_1', v_2') using trigonometry,

$$\begin{cases} v_1' = v_1 \cos\alpha + v_2 \sin\alpha, \\ v_2' = -v_1 \sin\alpha + v_2 \cos\alpha. \end{cases} \tag{131}$$

These transformation equations can be written in *matrix-vector* form as

$$\begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{132}$$

where the matrix-vector product is performed *row by column* as given in (131).

However, it is easier to obtain the transformation rules by using the *basis vectors* explicitly and dot products. The starting point is the vector identity

$$\boldsymbol{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 = v_1'\mathbf{e}_1' + v_2'\mathbf{e}_2', \tag{133}$$

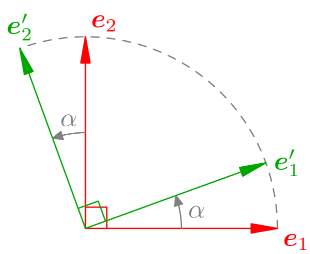


Fig. 1.29: Change of cartesian bases in 2D.

then simple dot products with $\mathbf{e}_1'$ and $\mathbf{e}_2'$ yield

$$\begin{cases} v_1' = \mathbf{e}_1' \cdot \boldsymbol{v} = \mathbf{e}_1' \cdot \mathbf{e}_1\ v_1 + \mathbf{e}_1' \cdot \mathbf{e}_2\ v_2, \\ v_2' = \mathbf{e}_2' \cdot \boldsymbol{v} = \mathbf{e}_2' \cdot \mathbf{e}_1\ v_1 + \mathbf{e}_2' \cdot \mathbf{e}_2\ v_2, \end{cases} \tag{134}$$

and these are identical to (131) since

$$\begin{cases} \mathbf{e}_1' \cdot \mathbf{e}_1 = \cos\alpha, & \mathbf{e}_1' \cdot \mathbf{e}_2 = \sin\alpha, \\ \mathbf{e}_2' \cdot \mathbf{e}_1 = -\sin\alpha, & \mathbf{e}_2' \cdot \mathbf{e}_2 = \cos\alpha. \end{cases}$$

The inverse transformation is obtained just as easily by dotting (133) by $\mathbf{e}_1$ and $\mathbf{e}_2$ yielding

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1' & \mathbf{e}_1 \cdot \mathbf{e}_2' \\ \mathbf{e}_2 \cdot \mathbf{e}_1' & \mathbf{e}_2 \cdot \mathbf{e}_2' \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \end{pmatrix}. \tag{135}$$

A transformation $(v_1, v_2) \to (v_1', v_2')$ followed by its inverse $(v_1', v_2') \to (v_1, v_2)$ should return the original components. Indeed, substituting (132) for (v_1', v_2') into (135) gives

$$\begin{aligned} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \end{aligned} \tag{136}$$

where the *matrix-matrix* product is performed *row by column*.

9.2 Change of cartesian basis in 3D

Consider two orthonormal bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ in 3D euclidean space. The corresponding cartesian components (v_1, v_2, v_3) and (v'_1, v'_2, v'_3) of an arbitrary vector $\boldsymbol{v}$ will be distinct in general, $(v_1, v_2, v_3) \neq (v'_1, v'_2, v'_3)$, but equality of the representations can be expressed with the direction vectors

$$\boldsymbol{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 = v'_1\mathbf{e}'_1 + v'_2\mathbf{e}'_2 + v'_3\mathbf{e}'_3,$$

that is

$$\boldsymbol{v} = v_i\mathbf{e}_i = v'_i\mathbf{e}'_i,$$

using the convention of summation over repeated indices. The relationships between the two sets of coordinates are then easily obtained by simple dot products

$$v'_i = \mathbf{e}'_i \cdot \boldsymbol{v} = (\mathbf{e}'_i \cdot \mathbf{e}_j)\, v_j \triangleq q_{ij}\, v_j, \tag{137}$$

and

$$v_i = \mathbf{e}_i \cdot \boldsymbol{v} = (\mathbf{e}_i \cdot \mathbf{e}'_j)\, v'_j = q_{ji}\, v'_j, \tag{138}$$

where

$$\boxed{q_{ij} \triangleq \mathbf{e}'_i \cdot \mathbf{e}_j} \tag{139}$$

It is useful to view the 9 coefficients q_{ij} as the elements of a 3-by-3 *matrix* $\mathbf{Q}$, that is, a 3-by-3 table with the first index i corresponding to the row index and the second index j to the column index

$$\mathbf{Q} \equiv \left[q_{ij}\right] = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{e}'_1 \cdot \mathbf{e}_1 & \mathbf{e}'_1 \cdot \mathbf{e}_2 & \mathbf{e}'_1 \cdot \mathbf{e}_3 \\ \mathbf{e}'_2 \cdot \mathbf{e}_1 & \mathbf{e}'_2 \cdot \mathbf{e}_2 & \mathbf{e}'_2 \cdot \mathbf{e}_3 \\ \mathbf{e}'_3 \cdot \mathbf{e}_1 & \mathbf{e}'_3 \cdot \mathbf{e}_2 & \mathbf{e}'_3 \cdot \mathbf{e}_3 \end{bmatrix} \tag{140}$$

The transformation rule (137) can then be viewed as a matrix vector product

$$v'_i = q_{ij}v_j \;\Leftrightarrow\; \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \;\Leftrightarrow\; \mathbf{v}' = \mathbf{Q}\mathbf{v} \tag{141}$$

where $\mathbf{v}'$ and $\mathbf{v}$ are the *columns* of cartesian components (v'_1, v'_2, v'_3) and (v_1, v_2, v_3), respectively. The inverse transformation (138) is

$$v_i = q_{ji}v'_j \;\Leftrightarrow\; \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{21} & q_{31} \\ q_{12} & q_{22} & q_{32} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix} \;\Leftrightarrow\; \mathbf{v} = \mathbf{Q}^T\mathbf{v}' \tag{142}$$

where $\mathbf{Q}^T$ is the *transpose* of $\mathbf{Q}$, the matrix whose *rows* are the corresponding *columns* of $\mathbf{Q}$ (and thus the *columns* of $\mathbf{Q}^T$ are the *rows* of $\mathbf{Q}$),

$$q^T_{ij} = q_{ji}.$$

Notation. We write $\boldsymbol{v}$ for a geometric vector, a quantity with magnitude and direction independent of the system of coordinates. The cartesian components of that vector with respect to any orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are written compactly as v_i in index notation, with $i = 1, 2, 3$. Those components (v_1, v_2, v_3) form a vector in $\mathbb{R}^3$ that is considered to be a *column* vector $\mathbf{v}$ in *matrix notation.* These are equivalent but distinct representations for $\boldsymbol{v}$

$$\boldsymbol{v} \equiv v_i \equiv \mathbf{v}.$$

The index notation, as in v_i and q_{ij}, does not in itself have the concept of 'rows' and 'columns,' but it distinguishes between first and second index, thus

$$q_{ij}\, v_j \neq q_{ji}\, v_j.$$

Seeing q_{ij} as the (row i, column j) element of matrix $\mathbf{Q}$, the sum $q_{ij}v_j$ is three dot products between the *rows* of $\mathbf{Q}$ and the column $\mathbf{v}$

$$q_{ij}\, v_j \;\equiv\; \mathbf{Q}\mathbf{v},$$

but the sum $q_{ji}v_j$ is three dot products between the *columns* of $\mathbf{Q}$ and the column $\mathbf{v}$. Thus in matrix notation where matrices and vectors are multiplied *rows by columns*,

$$q_{ji}\, v_j \;\equiv\; \mathbf{v}^T\mathbf{Q} \equiv \mathbf{Q}^T\mathbf{v},$$

where $\mathbf{v}^T$ and $\mathbf{v}^T\mathbf{Q}$ are *row vectors* while $\mathbf{Q}^T\mathbf{v}$ and $\mathbf{v}$ are *column vectors*

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

In matrix notation, a row cannot equal a column, unless they are mere 1-by-1, thus $\mathbf{v}^T \neq \mathbf{v}$, although both are equivalent to $v_i \equiv (v_1, v_2, v_3)$.

Direction Cosine Matrix. The quantities q_{ij} defined in (139) are

$$q_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j = \cos\theta_{ij}$$

where θ_{ij} is the angle between the unit vectors $\mathbf{e}'_i$ and $\mathbf{e}_j$. Thus the $q_{ij} = \cos\theta_{ij}$ are *direction cosines* and the 9 of them form the *Direction Cosine Matrix*

$$\mathbf{Q} \equiv \begin{bmatrix} q_{ij} \end{bmatrix} = \begin{bmatrix} \cos\theta_{11} & \cos\theta_{12} & \cos\theta_{13} \\ \cos\theta_{21} & \cos\theta_{22} & \cos\theta_{23} \\ \cos\theta_{31} & \cos\theta_{32} & \cos\theta_{33} \end{bmatrix}. \tag{143}$$

There are 9 angles θ_{ij}, but those angles are not independent.

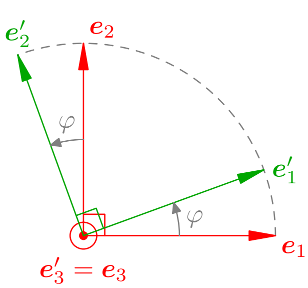


Fig. 1.30: Rotation about $\mathbf{e}_3$.

Example. If $\mathbf{e}'_1, \mathbf{e}'_2$ are rotated from $\mathbf{e}_1, \mathbf{e}_2$ about $\mathbf{e}'_3 = \mathbf{e}_3$ by φ,we find

$$\mathbf{Q} \equiv \begin{bmatrix} q_{ij} \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{144}$$

with $\theta_{11} = \varphi$, $\theta_{12} = \pi/2 - \varphi$, $\theta_{21} = \varphi + \pi/2$ and $\theta_{22} = \varphi$. $\square$

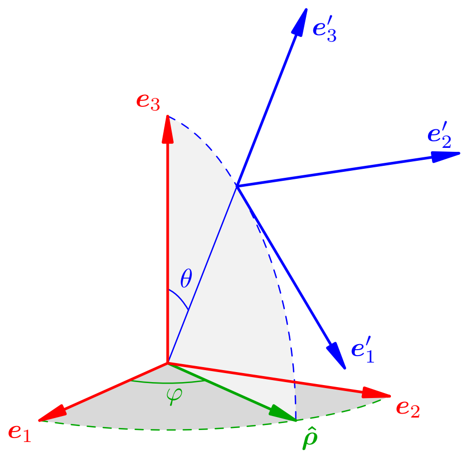


Fig. 1.31: Earth basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and local basis $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ at longitude φ, polar angle θ. $\mathbf{e}_1'$ is South, $\mathbf{e}_2'$ East, and $\mathbf{e}_3'$ is Up.

Example. Earth to local basis. Consider an Earth basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ with $\mathbf{e}_3$ in the *polar axis* direction, $\mathbf{e}_1, \mathbf{e}_3$ in the plane of the *prime meridian* and a local basis $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ at longitude φ, polar angle θ (latitude $\lambda = \pi/2 - \theta$) with $\mathbf{e}_1'$ south, $\mathbf{e}_2'$ east and $\mathbf{e}_3'$ up, as illustrated in fig. 1.31.
Most of the required vector analysis has already been done in section 1.3 with the help of the horizontal radial direction vector $\hat{\boldsymbol{\rho}} = \cos\varphi\,\mathbf{e}_1 + \sin\varphi\,\mathbf{e}_2$. We find (exercise 5)

$$\left\{\begin{array}{rrcrcr} \mathbf{e}_1' = & \cos\varphi\cos\theta\,\mathbf{e}_1 & + & \sin\varphi\cos\theta\,\mathbf{e}_2 & - & \sin\theta\,\mathbf{e}_3, \\ \mathbf{e}_2' = & -\sin\varphi\,\mathbf{e}_1 & + & \cos\varphi\,\mathbf{e}_2, & & \\ \mathbf{e}_3' = & \cos\varphi\sin\theta\,\mathbf{e}_1 & + & \sin\varphi\sin\theta\,\mathbf{e}_2 & + & \cos\theta\,\mathbf{e}_3, \end{array}\right. \tag{145}$$

yielding the transformation matrix

$$\mathbf{Q} = \begin{bmatrix} \cos\varphi\cos\theta & \sin\varphi\cos\theta & -\sin\theta \\ -\sin\varphi & \cos\varphi & 0 \\ \cos\varphi\sin\theta & \sin\varphi\sin\theta & \cos\theta \end{bmatrix}. \tag{146}$$

The two angles φ and θ specify this direction cosine matrix. In general, three angles are needed, the third angle corresponding to an arbitrary rotation of $\mathbf{e}_1'$ and $\mathbf{e}_2'$ about $\mathbf{e}_3'$. This is discussed below in the Euler angles section.

□

Orthonormality. There are 9 direction cosines $\cos\theta_{ij}$ (in 3D space) but orthonormality of both bases imply several constraints, so these 9 angles are not independent and the $\mathbf{Q}$ matrices have very special characteristics.

These constraints follow from (137), (138) which must hold for any (v_1, v_2, v_3) and (v_1', v_2', v_3'). Substituting (138) into (137) yields

$$v_i' = q_{ik}q_{jk}\,v_j', \quad \forall v_i' \;\Rightarrow\; q_{ik}q_{jk} = \delta_{ij}. \tag{147}$$

Viewing q_{ij} as the (i, j) element of matrix $\mathbf{Q}$, the sums $q_{ik}q_{jk}$ are dot products between the *rows* of $\mathbf{Q}$. In *row by column* matrix notation this is

$$q_{ik}q_{jk} = q_{ik}q^T_{kj} = \delta_{ij} \;\Leftrightarrow\; \mathbf{Q}\mathbf{Q}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{148}$$

This means that the *rows* of matrix $\mathbf{Q}$ (140) are *orthogonal* to each other and of unit magnitude.

Likewise, substituting (137) into (138) gives

$$v_i = q_{ki}q_{kj}\, v_j, \quad \forall v_i \;\Rightarrow\; q_{ki}q_{kj} = \delta_{ij}.$$

Now, the sums $q_{ki}q_{kj}$ are dot products between the *columns* of $\mathbf{Q}$. In *row by column* matrix notation this is

$$q_{ki}q_{kj} = q^T_{ik}q_{kj} = \delta_{ij} \;\Leftrightarrow\; \mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{149}$$

and the *columns* of matrix $\mathbf{Q}$ (140) are also *orthogonal* to each other and of unit magnitude.

Geometric interpretation of orthogonal matrix. The two sets of orthogonal relationships (148), (149) can be derived more geometrically as follows. The coefficient $q_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j$ is both the j component of $\mathbf{e}'_i$ in the $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ basis, and the i component of $\mathbf{e}_j$ in the $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ basis. Therefore we can write

$$\mathbf{e}'_i = (\mathbf{e}'_i \cdot \mathbf{e}_j)\, \mathbf{e}_j = q_{ij}\, \mathbf{e}_j. \tag{150}$$

In other words,

$$\mathbf{e}'_i \equiv \begin{bmatrix} q_{i1} & q_{i2} & q_{i3} \end{bmatrix}$$

in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, and this is the i-th *row* of matrix $\mathbf{Q}$ (140). Now $\mathbf{e}'_i \cdot \mathbf{e}'_j = \delta_{ij}$ thus

$$\mathbf{e}'_i \cdot \mathbf{e}'_j = q_{ik}\mathbf{e}_k \cdot q_{jl}\mathbf{e}_l = q_{ik}q_{jl}\mathbf{e}_k \cdot \mathbf{e}_l = q_{ik}q_{jl}\delta_{kl} = q_{ik}q_{jk}. \tag{151}$$

hence

$$\mathbf{e}'_i \cdot \mathbf{e}'_j = \boxed{q_{ik}q_{jk} = \delta_{ij}.} \tag{152}$$

So the *rows* of $\mathbf{Q}$ are orthonormal because they are the components of $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

Likewise,

$$\mathbf{e}_j = (\mathbf{e}_j \cdot \mathbf{e}'_k)\, \mathbf{e}'_k = q_{kj}\, \mathbf{e}'_k, \tag{153}$$

and

$$\mathbf{e}_j \equiv \begin{bmatrix} q_{1j} \\ q_{2j} \\ q_{3j} \end{bmatrix}$$

in the basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$, this is the j-th *column* of matrix $\mathbf{Q}$ in (140). Then

$$\mathbf{e}_i \cdot \mathbf{e}_j = \boxed{q_{ki}q_{kj} = \delta_{ij},} \tag{154}$$

and the columns of $\mathbf{Q}$ are orthonormal because they are the components of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ in the basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$.

A *square* matrix whose rows are mutually orthonormal to each other will have mutually orthonormal columns, and *vice-versa.* Such a matrix is called an *orthogonal matrix.*

Exercises

1. Find $\mathbf{Q}$ if $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ is the right hand rotation of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ about $\mathbf{e}_3$ by angle φ. Verify (152) and (154) for your Q. If $\boldsymbol{v} = v_i\mathbf{e}_i$, find the components of $\boldsymbol{v}$ in the basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ in terms of (v_1, v_2, v_3) and φ.

2. Find $\mathbf{Q}$ if $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ is the right hand rotation of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ about $\mathbf{e}_2$ by angle θ. Verify (152) and (154) for your Q.

3. Find $\mathbf{Q}$ if $\mathbf{e}'_1 = -\mathbf{e}_1$, $\mathbf{e}'_2 = \mathbf{e}_3$, $\mathbf{e}'_3 = \mathbf{e}_2$ and verify (152) and (154) for it.

4. Verify that the rows of (144) and (146) are orthonormal, and likewise for the columns.

5. Derive (146) (i) using meridional $(\hat{\boldsymbol{\rho}}, \mathbf{e}_3)$ and equatorial $(\mathbf{e}_1, \mathbf{e}_2)$ projections of $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ as in section 1.3, (ii) finding $\mathbf{e}'_3$ by projections onto $\mathbf{e}_3$ and $\hat{\boldsymbol{\rho}}$ then $\mathbf{e}_1$ and $\mathbf{e}_2$, then calculating $\mathbf{e}'_2 = (\mathbf{e}_3 \times \mathbf{e}'_3)/|\mathbf{e}_3 \times \mathbf{e}'_3|$ and $\mathbf{e}'_1 = \mathbf{e}'_2 \times \mathbf{e}'_3$.

6. Find $\mathbf{Q}$ if $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is an Earth basis as defined in the text and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ is a local basis at longitude φ and latitude λ with $\mathbf{e}'_1$ east, $\mathbf{e}'_2$ north and $\mathbf{e}'_3$ up. Write $\mathbf{Q}$ in terms of φ and λ.

7. *Orthogonal projection of a 3D scene.* One way to make a 2D picture of a 3D scene is to plot the orthogonal projection of the 3D data specified in a $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ basis onto a plane perpendicular to the viewpoint in direction $\mathbf{e}'_3$ at azimuth φ and elevation λ. Find the relevant $\mathbf{Q}$ in terms of φ and λ and specify how to obtain the 2D plotting data from the 3D data.

8. The velocity of a satellite is v_1 east, v_2 north, v_3 up as measured from a cartesian basis located at longitude φ, latitude λ. What are the corresponding velocity components with respect to the Earth basis?

9. The velocity of a satellite is v_1 east, v_2 north, v_3 up as measured from a cartesian basis located at longitude φ_1, latitude λ_1. What are the corresponding velocity components with respect to a local basis at longitude φ_2, latitude λ_2? Derive and explain an algorithm to compute those components.

10. In relation to (147), prove that if $v_i = a_{ij}v_j$ for all v_i then $a_{ij} = \delta_{ij}$.

11. Explain why the determinant of any orthogonal matrix $\mathbf{Q}$ is ± 1. When is $\det(\mathbf{Q}) = +1$ and when is it -1, in general? Give explicit examples.

12. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are two arbitrary vectors and $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ and $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are two distinct orthonormal bases, we have shown that $a_i b_i = a'_i b'_i$ (eqn. (63), here with summation convention). Verify this invariance directly from the transformation rule (137), $v'_i = q_{ij} v_j$, showing your mastery of index notation.

13. If $\boldsymbol{v}'$ is the rotation of $\boldsymbol{v}$ about $\boldsymbol{a}$ by α (exercise 12 in cross product section) then $v'_i = p_{ij} v_j$ in a cartesian basis. Find p_{ij} in terms of α and the cartesian components of $\boldsymbol{a}$. Show that the matrix $[p_{ij}]$ is orthogonal. Note that this is *not* a change of basis, it is a rotation of vectors, however rotation of vectors and rotation of bases are closely connected.

14. Let $\boldsymbol{v}' = \boldsymbol{v} - 2(\hat{\boldsymbol{a}} \cdot \boldsymbol{v})\hat{\boldsymbol{a}}$ then $v'_i = p_{ij} v_j$ in a cartesian basis. Find p_{ij}. Show that $[p_{ij}]$ is an orthogonal matrix. Note that this is *not* a change of basis.

9.3 Matrices

That $\mathbf{Q}$ was a very special matrix, an *orthogonal* matrix. More generally a 3-by-3 real matrix $\mathbf{A}$ is a table of 9 real numbers

$$\mathbf{A} \equiv [a_{ij}] \equiv \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{155}$$

Matrices are denoted by a capital letter, $\mathbf{A}$ and $\mathbf{Q}$, and by square brackets [ ]. By convention, vectors in $\mathbb{R}^3$ are defined as 3-by-1 matrices

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_1 \\ x_3 \end{bmatrix},$$

although for typographical reasons we often write $\mathbf{x} = (x_1, x_2, x_3)$ but not $[x_1, x_2, x_3]$ that denotes a 1-by-3 matrix, or *row* vector. The term *matrix* is similar to *vectors* in that it implies precise rules for manipulations of these objects (for vectors these are the two fundamental addition and scalar multiplication operations with specific properties, see Sect. 1.4).

Matrix-vector multiply

Equation (137) shows how matrix-vector multiplication should be defined. The matrix vector product $\mathbf{A}\mathbf{x}$ ($\mathbf{A}$ 3-by-3, $\mathbf{x}$ 3-by-1) is a 3-by-1 vector $\mathbf{b}$ in $\mathbb{R}^3$ whose i-th component is the dot product of row i of matrix $\mathbf{A}$ with the column $\mathbf{x}$,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad b_i = a_{ij} x_j \tag{156}$$

where $a_{ij} x_j \equiv a_{i1} x_1 + a_{i2} x_2 + a_{i3} x_3$ in the summation convention. The product of a matrix with a (column) vector is performed *row-by-column.* This product is defined only if the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{x}$. A 2-by-1 vector cannot be multiplied by a 3-by-3 matrix.

Identity Matrix

There is a unique matrix such that $\mathbf{Ix} = \mathbf{x}$, $\forall \mathbf{x}$. For $\mathbf{x} \in \mathbb{R}^3$, that is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{157}$$

Matrix-Matrix multiply

Two successive linear transformation of coordinates, that is,

$$x_i' = a_{ij}\, x_j, \quad \text{then} \quad x_i'' = b_{ij}\, x_j'$$

(summation over repeated indices) can be combined into one transformation from x_j to x_i''

$$x_i'' = b_{ik} a_{kj}\ x_j = c_{ij}\, x_j$$

where

$$c_{ij} = b_{ik} a_{kj} = (\mathbf{BA})_{ij}. \tag{158}$$

This defines matrix multiplication. The product of two matrices $\mathbf{BA}$ is a matrix, $\mathbf{C}$ say, whose (i, j) element c_{ij} is the dot product of row i of $\mathbf{B}$ with column j of $\mathbf{A}$. As for matrix-vector multiplication, the product of two matrices is done *row-by-column.* This requires that the *number of columns of the first matrix* in the product ($\mathbf{B}$) equals the *number of rows of the second matrix* ($\mathbf{A}$). Thus, the product of a 3-by-3 matrix and a 2-by-2 matrix is not defined, for instance. We can only multiply M-by-N by an N-by-P, that is *'inner dimensions must match'*. In general, $\mathbf{BA} \neq \mathbf{AB}$, matrix multiplication does not commute. You can visualize this by considering two successive rotation of axes, one by angle α about $\mathbf{e}_3$, followed by one by β about $\mathbf{e}_2'$. This is not the same as rotating by β about $\mathbf{e}_2$, then by α about $\mathbf{e}_3'$. You can also see it algebraically

$$(\mathbf{BA})_{ij} = b_{ik} a_{kj} \neq a_{ik} b_{kj} = (\mathbf{AB})_{ij}.$$

Matrix transpose

The transformation (138) involves the sum $a_{ji} x_j'$ that is similar to the matrix vector multiply except that the multiplication is column-by-column! To write this as a matrix-vector multiply, we define the *transpose matrix* $\mathbf{A}^T$ whose row i correspond to column i of A. If the (i, j) element of A is a_{ij} then the (i, j) element of $\mathbf{A}^T$ is a_{ji}

$$(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}.$$

Then

$$x_i = a_{ji} x_j' \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{A}^T \mathbf{x}'. \tag{159}$$

A *symmetric matrix* $\mathbf{A}$ is such that $\mathbf{A} = \mathbf{A}^T$, but an *anti-symmetric matrix* $\mathbf{A}$ is such that $\mathbf{A} = -\mathbf{A}^T$.

It is left as an exercise in index notation to show that *the transpose of a product is equal to the product of the transposes in reverse order* $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$.

9.4 Euler Angles

Arbitrary matrices are typically denoted $\mathbf{A}$, while orthogonal matrices are typically denoted $\mathbf{Q}$ in the literature. In matrix notation, the orthogonality conditions (152), (154) read

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}. \tag{160}$$

Such a matrix is called an *orthogonal matrix.* A *proper* orthogonal matrix has determinant equal to 1 and corresponds geometrically to a pure rotation. An *improper* orthogonal matrix has determinant -1. It corresponds geometrically to a combination of rotations and an *odd* number of reflections. The *product* of orthogonal matrices is an orthogonal matrix but the sum (performed element by element) is not.

As we have seen at the beginning of this section, the elements q_{ij} of an orthogonal matrix can be interpreted as the dot products of unit vectors of two distinct orthonormal bases, $q_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j = \cos\theta_{ij}$, where θ_{ij} is the angle between $\mathbf{e}'_i$ and $\mathbf{e}_j$. In 3D, there are 9 such angles but these angles are not independent since both bases $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ and $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ consist of mutually orthogonal unit vectors. If both bases are right handed (or both are left handed), each can be transformed into the other through only *three* elementary rigid body rotations, in general. Therefore, any 3 by 3 proper orthogonal matrix can be decomposed into the product of three elementary orthogonal matrices. The 3 angles corresponding to those 3 elementary rotations are called *Euler angles*, in general.

That only 3 angles are needed should not be a surprise since we learned early in this chapter that an arbitrary direction $\hat{\boldsymbol{a}}$ in 3D space can be specified by 2 angles. Thus, 2 angles φ and θ are sufficient to specify $\mathbf{e}'_3$, say, in terms of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and we only need one extra angle, call it ζ, to specify the orientation of $(\mathbf{e}'_1, \mathbf{e}'_2)$ about the direction $\mathbf{e}'_3$. That's spherical coordinates φ, θ plus a twist ζ. Thus we can construct, or represent, the matrix $\mathbf{Q}$ corresponding to the transformation from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ in the following 3 elementary rotations.

1. Rotation about $\mathbf{e}_3$ by φ to obtain the intermediate basis (fig. 1.30)

$$\begin{cases} \mathbf{e}_1^{(1)} = & \cos\varphi\,\mathbf{e}_1 & + & \sin\varphi\,\mathbf{e}_2 & \\ \mathbf{e}_2^{(1)} = & -\sin\varphi\,\mathbf{e}_1 & + & \cos\varphi\,\mathbf{e}_2 & \\ \mathbf{e}_3^{(1)} = & & & & \mathbf{e}_3 \end{cases} \tag{161}$$

defining $q_{ij}^{(1)} = \mathbf{e}_i^{(1)} \cdot \mathbf{e}_j$, yields the rotation matrix

$$\mathbf{Q}^{(1)} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{162}$$

for the rotation from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \mathbf{e}_3^{(1)}\}$. The goal of this rotation about $\mathbf{e}_3$ by φ is is to align $\mathbf{e}_2^{(1)}$ with $\mathbf{e}_3 \times \mathbf{e}'_3$, that is to put $\mathbf{e}_1^{(1)}$ in the plane of $\mathbf{e}_3$ and the target $\mathbf{e}'_3$.

2. Rotation about $\mathbf{e}_2^{(1)}$ by angle θ to obtain the basis

$$\begin{cases} \mathbf{e}_1^{(2)} = \cos\theta\,\mathbf{e}_1^{(1)} - \sin\theta\,\mathbf{e}_3^{(1)} \\ \mathbf{e}_2^{(2)} = \mathbf{e}_2^{(1)} \\ \mathbf{e}_3^{(2)} = \sin\theta\,\mathbf{e}_1^{(1)} + \cos\theta\,\mathbf{e}_3^{(1)} \end{cases} \tag{163}$$

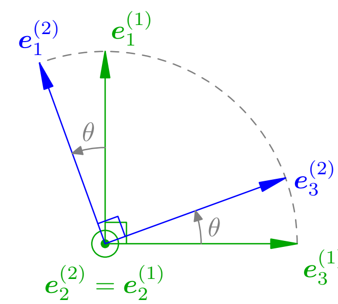


Fig. 1.32: Rotation about $\mathbf{e}_2$.

with $q_{ij}^{(2)} = \mathbf{e}_i^{(2)} \cdot \mathbf{e}_j^{(1)}$, this corresponds to the rotation matrix

$$\mathbf{Q}^{(2)} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{164}$$

for the rotation from $\{\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \mathbf{e}_3^{(1)}\}$ to $\{\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \mathbf{e}_3^{(2)}\}$. These 2 rotations achieve $\mathbf{e}_3^{(2)} = \mathbf{e}_3'$, the target $\mathbf{e}_3'$ but the vectors $(\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)})$ do not necessarily match the target $(\mathbf{e}_1', \mathbf{e}_2')$.

3. To match those vectors in general requires another elementary rotation about $\mathbf{e}_3^{(2)} = \mathbf{e}_3'$ to align $(\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)})$ with the target $(\mathbf{e}_1', \mathbf{e}_2')$

$$\begin{cases} \mathbf{e}_1^{(3)} = \cos\zeta\,\mathbf{e}_1^{(2)} + \sin\zeta\,\mathbf{e}_2^{(2)} \\ \mathbf{e}_2^{(3)} = -\sin\zeta\,\mathbf{e}_1^{(2)} + \cos\zeta\,\mathbf{e}_2^{(2)} \\ \mathbf{e}_3^{(3)} = \mathbf{e}_3^{(2)} \end{cases} \tag{165}$$

defining $q_{ij}^{(3)} = \mathbf{e}_i^{(3)} \cdot \mathbf{e}_j^{(2)}$, this corresponds to the rotation matrix

$$\mathbf{Q}^{(3)} = \begin{bmatrix} \cos\zeta & \sin\zeta & 0 \\ -\sin\zeta & \cos\zeta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{166}$$

The orthogonal transformation $\mathbf{Q}$ from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \mathbf{e}_3^{(3)}\}$ is then obtained by taking the matrix product

$$\begin{aligned} \mathbf{Q} &= \mathbf{Q}^{(3)}\mathbf{Q}^{(2)}\mathbf{Q}^{(1)} \\ &= \begin{bmatrix} \cos\zeta & \sin\zeta & 0 \\ -\sin\zeta & \cos\zeta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{167}$$

Watch out for the order! The first transformation is the rightmost matrix $\mathbf{Q}^{(1)}$.

Any 3 by 3 proper orthogonal matrix $\mathbf{Q}$ can thus be represented with only three angles, (φ, θ, ζ), for example. There are many ways to choose those angles however. The choice made above is consistent with spherical coordinates and would be labelled a ZYZ representation in the literature since we rotated about the original z axis, then the new y, then the new z again. A ZXZ definition would perform a rotation about the original z so that $\mathbf{e}_1'$ is in the $\mathbf{e}_3 \times \mathbf{e}_3'$ direction, instead of $\mathbf{e}_2'$ as we chose above. Then a rotation about $\mathbf{e}_1'$ would be performed to align $\mathbf{e}_3'' = \mathbf{e}_3$, followed by a rotation about $\mathbf{e}_3''$ to align $(\mathbf{e}_1'', \mathbf{e}_2'')$. There are 4 other possible choices of Euler angles: XYX, XZX, YXY, YZY. In all cases, the 1st and 3rd rotation are about the same intrinsic direction. Such choices of Euler angles are sometimes called 'proper' Euler angles.

Proper Euler angles inherit a singularity from spherical coordinates where the azimuth φ is undetermined when the polar angle $\theta = 0$ or π. Likewise in (167) when $\theta = 0$ or π, the first and 3rd rotation are about the same actual direction, hence φ and ζ are not uniquely determined. When θ is near 0 or π, the decomposition (167) may lead to large angles φ, ζ that almost cancel each other to yield a small net effect. This may lead to computational inaccuracies, or wild motions if the angles are used to control a 3D body. This singularity issue is known as 'Gimbal lock' in mechanical engineering.

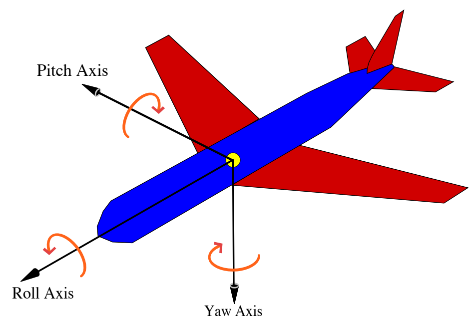


Fig. 1.33: Airplane attitude, yaw, pitch and roll, from `wikipedia`.

In aircraft dynamics and control, 3 angles about 3 distinct axes are used as illustrated in fig. 1.33. Imagine a frame $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ attached to an airplane, with $\hat{\boldsymbol{x}}$ pointing from tail to nose, $\hat{\boldsymbol{z}}$ perpendicular to the plane of the airplane and $\hat{\boldsymbol{y}}$ pointing from one wingtip to the other. The orientation of the airplane with respect to a fixed reference frame can be specified by the *heading* (or *yaw*) — the angle around $\hat{\boldsymbol{z}}$ to align $\hat{\boldsymbol{x}}$ with the desired horizontal direction, the *elevation* (or *pitch*) — the angle about $\hat{\boldsymbol{y}}$ to pitch the nose up or down to align $\hat{\boldsymbol{x}}$ with the desired direction in the vertical plane, and the *bank* (or *roll*) – the angle about $\hat{\boldsymbol{x}}$ to rotate the wings around the axis of the airplane to achieve the desired *bank* angle. The yaw-pitch-roll decomposition would be a ZYX decomposition (or *factorization*) of the orthogonal matrix $\mathbf{Q}$. Such choices of Euler angles are often called *Bryan* or *Tait-Bryan* angles, after G.H. Bryan, a pioneer in airplane stability and dynamics. A ZYX factorization of an orthogonal $\mathbf{Q}$ reads

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{168}$$

where φ is the *heading*, θ the *pitch* and ψ the *bank*.

Gram-Schmidt

To define an arbitrary orthogonal matrix, we can then simply pick any three arbitrary (Euler) angles φ, θ, ζ and construct an orthonormal matrix using (167). Another important procedure to do this is the **Gram-Schmidt** procedure: pick any three $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ and orthonormalize them, that is

(1) First, let $\boldsymbol{q}_1 = \boldsymbol{a}_1/|\boldsymbol{a}_1|$ and $\boldsymbol{a}'_2 = \boldsymbol{a}_2 - (\boldsymbol{a}_2 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1$, $\boldsymbol{a}'_3 = \boldsymbol{a}_3 - (\boldsymbol{a}_3 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1$,

(2) next, let $\boldsymbol{q}_2 = \boldsymbol{a}'_2/|\boldsymbol{a}'_2|$ and $\boldsymbol{a}''_3 = \boldsymbol{a}'_3 - (\boldsymbol{a}'_3 \cdot \boldsymbol{q}_2)\boldsymbol{q}_2$,

(3) finally, let $\boldsymbol{q}_3 = \boldsymbol{a}''_3/|\boldsymbol{a}''_3|$.

The vectors $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$ form an orthonormal basis. This procedure generalizes not only to any dimension but also to other vector spaces, *e.g.* to construct orthogonal polynomials.

Exercises

1. Give explicit examples of 2-by-2 and 3-by-3 symmetric and antisymmetric matrices.

2. If $\mathbf{x}^T = [x_1, x_2, x_3]$, calculate $\mathbf{x}^T\mathbf{x}$ and $\mathbf{x}\mathbf{x}^T$.

3. For $\mathbf{x} \in \mathbb{R}^n$, show that $\mathbf{x}^T\mathbf{x}$ and $\mathbf{x}\mathbf{x}^T$ are symmetric, (i) explicitly using indices, (ii) by matrix manipulations.

4. Let $\boldsymbol{v} = \boldsymbol{a} \times \boldsymbol{b}$ in $\mathbb{E}^3$ yielding $v_i = a_{ij}b_j$ in a cartesian basis. Find a_{ij}. Show that $[a_{ij}]$ is anti-symmetric but not orthogonal.

5. Prove that the transpose of a product $(\mathbf{AB})^T$ is the product of the transposes in *reverse* order $\mathbf{B}^T\mathbf{A}^T$.

6. Prove that the product of two orthogonal matrices is an orthogonal matrix. Interpret geometrically.

7. What is the most general form of a 2-by-2 orthogonal matrix?

8. Give a non-trivial example of a 3-by-3 orthogonal *and* symmetric matrix.

9. Assume that $\mathbf{Q}$ is a proper orthogonal and symmetric matrix. What is the geometric meaning of such matrices? Give a non-trivial example.

10. Joe claims that
$$\mathbf{Q} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$
is an orthogonal matrix. How do you verify Joe's data? Assuming the data is valid, what are $\mathbf{e}'_2 \cdot \mathbf{e}_3$ and $q_{1k}q_{k2}$ for that matrix?

11. What is the orthogonal matrix corresponding to a reflection about the (x, z) plane? What is its determinant?

12. We want to rotate an object (*i.e.* a set of points) by an angle γ about an axis passing through the origin. Provide an algorithm (or a Matlab or Python code) to calculate the cartesian coordinates of the rotated points. Compare (1) the vector approach of exercise 12 in the cross-product section and (2) the elementary rotation matrix approach. The latter performs two elementary rotation of bases to obtain the coordinates of the points in a basis whose $\mathbf{e}_3'$ is the desired rotation direction, rotates in that basis, then returns to the original basis. How many elementary rotation matrices are involved? Compare the computational complexity of both approaches.

13. What are Euler angles? Are they unique (modulo 2π)?

14. Verify that (167) yields (146) when $\zeta = 0$.

15. Let $\mathbf{Q}$ be any orthogonal matrix. What is the form of its ZYZ Euler angle decomposition? What is the form of its ZXZ Euler angle decomposition? What is the form of its ZYX decomposition?

16. Find the ZYZ, ZXZ and ZYX factorizations of (146).

17. The orthonormal basis $\{\mathbf{e}_1'', \mathbf{e}_2'', \mathbf{e}_3''\}$ is the rotation of $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ by ζ about $\mathbf{e}_3'$. The transformation from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ is given in (146). What is the matrix corresponding to the transformation from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}_1'', \mathbf{e}_2'', \mathbf{e}_3''\}$? Find an elementary rotation factorization of that matrix.

18. Prove that the product of two orthogonal matrices is an orthogonal matrix but that their sum is not, in general.

19. The basis $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ is the rotation of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ by α about $\mathbf{e}_1$. What is the matrix $\mathbf{Q}$ corresponding to the transformation $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} \to \{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$? Find its explicit ZYZ factorization. Discuss and visualize the elementary rotations when $\alpha \ll 1$.

20. Construct an algorithm to find the ZYZ Euler angle representation of a given arbitrary orthogonal matrix $\mathbf{Q}$.

21. What are the Euler angles and the transformation matrix $\mathbf{Q}$ from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ when the latter is the (right-handed) rotation of the former by angle α about the direction $\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$?

22. The basis $\{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$ is the rotation of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ by α about $\hat{\boldsymbol{a}}$. What is the ZYZ Euler factorization of the matrix $\mathbf{Q}$ corresponding to the transformation $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} \to \{\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'\}$? Invert this problem and construct an algorithm to determine α, $\hat{\boldsymbol{a}}$ given an arbitrary $\mathbf{Q}$.

23. Pick three non-trivial but arbitrary vectors in $\mathbb{R}^3$ (using Matlab's `randn(3)` for instance), then construct an orthonormal basis $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ using the Gram-Schmidt procedure. Verify that the matrix $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]$ is orthogonal. Note in particular that the rows are orthogonal eventhough you orthogonalized the columns only.

24. Pick *two* arbitrary vectors $\mathbf{a}_1$, $\mathbf{a}_2$ in $\mathbb{R}^3$ and orthogonalize them to construct $\mathbf{q}_1$, $\mathbf{q}_2$. Consider the 3-by-2 matrix $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2]$ and compute $\mathbf{Q}\mathbf{Q}^T$ and $\mathbf{Q}^T\mathbf{Q}$. Explain.

9.5 Determinant of a matrix (Optional)

See earlier discussion of determinants (section on mixed product). The determinant of a matrix has the explicit formula $\det(\mathbf{A}) = \epsilon_{ijk} a_{i1} a_{j2} a_{k3}$, the only non-zero terms are for (i, j, k) equal to a permutation of $(1, 2, 3)$. We can deduce several fundamental properties of determinants from that formula. We can reorder $a_{i1} a_{j2} a_{k3}$ into $a_{1l} a_{2m} a_{3n}$ using an even number of permutations if (i, j, k) is an even perm of (1,2,3) and an odd number for odd permutations. So

$$\det(\mathbf{A}) = \epsilon_{ijk} a_{i1} a_{j2} a_{k3} = \epsilon_{lmn} a_{1l} a_{2m} a_{3n} = \det(\mathbf{A}^T). \tag{169}$$

Another useful result is that

$$\epsilon_{ijk} a_{il} a_{jm} a_{kn} = \epsilon_{ijk} \epsilon_{lmn} a_{i1} a_{j2} a_{k3}. \tag{170}$$

We can then prove that $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ by a direct calculation in compact index notation:

$$\begin{aligned} \det(\mathbf{AB}) &= \epsilon_{ijk} a_{il} b_{l1} a_{jm} b_{m2} a_{kn} b_{n3} = \epsilon_{ijk} \epsilon_{lmn} a_{i1} a_{j2} a_{k3} b_{l1} b_{m2} b_{n3} \\ &= \det(\mathbf{A}) \det(\mathbf{B}) \end{aligned} \tag{171}$$

These results and manipulations generalize straightforwardly to any dimension.

9.6 Three views of $\mathbf{A}\mathbf{x} = \mathbf{b}$ (Optional)

Column View

▶ View $\mathbf{b}$ as a linear combination of the columns of $\mathbf{A}$.

Write $\mathbf{A}$ as a row of columns, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$, where $\mathbf{a}_1^T = [a_{11}, a_{21}, a_{31}]$ etc., then

$$\mathbf{b} = \mathbf{A}\mathbf{x} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + x_3 \mathbf{a}_3$$

and $\mathbf{b}$ is a linear combination of the columns $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$. If $\mathbf{x}$ is unknown, the linear system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ will have a solution for any $\mathbf{b}$ if and only if the columns form a basis, *i.e.* iff $\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \equiv \det(\mathbf{A}) \neq 0$. If the determinant is zero, then the 3 columns are in the same plane and the system will have a solution only if $\mathbf{b}$ is also in that plane.

As seen in earlier exercises, we can find the components (x_1, x_2, x_3) by thinking geometrically and projecting on the *reciprocal basis e.g.*

$$x_1 = \frac{\mathbf{b} \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \equiv \frac{\det(\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3)}{\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}. \tag{172}$$

Likewise

$$x_2 = \frac{\det(\mathbf{a}_1, \mathbf{b}, \mathbf{a}_3)}{\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}, \quad x_3 = \frac{\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{b})}{\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}.$$

This is a nifty formula. Component x_i equals the determinant where vector i is replaced by $\mathbf{b}$ divided by the determinant of the basis vectors. You can deduce this directly from the algebraic properties of determinants, for example,

$$\det(\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3) = \det(x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3, \mathbf{a}_2, \mathbf{a}_3) = x_1 \det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3).$$

This is **Cramer's rule** and it generalizes to any dimension, however computing determinants in higher dimensions can be very costly and the next approach is computationally much more efficient.

Row View:

▶ View $\mathbf{x}$ as the intersection of planes perpendicular to the rows of $\mathbf{A}$.

View $\mathbf{A}$ as a column of rows, $\mathbf{A} = [\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3]^T$, where $\mathbf{n}_1^T = [a_{11}, a_{12}, a_{13}]$ is the first *row* of $\mathbf{A}$, etc., then

$$\mathbf{b} = \mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{n}_1^T \\ \mathbf{n}_2^T \\ \mathbf{n}_3^T \end{bmatrix} \mathbf{x} \quad \Leftrightarrow \begin{cases} \mathbf{n}_1 \cdot \mathbf{x} = b_1 \\ \mathbf{n}_2 \cdot \mathbf{x} = b_2 \\ \mathbf{n}_3 \cdot \mathbf{x} = b_3 \end{cases}$$

and $\mathbf{x}$ is seen as the position vector of the intersection of three planes. Recall that $\mathbf{n} \cdot \mathbf{x} = C$ is the equation of a plane perpendicular to $\mathbf{n}$ and passing through a point $\mathbf{x}_0$ such that $\mathbf{n} \cdot \mathbf{x}_0 = C$, for instance the point $\mathbf{x}_0 = C\mathbf{n}/|\mathbf{n}|$.

To find $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$, for given $\mathbf{b}$ and A, we can combine the equations in order to eliminate unknowns, *i.e.*

$$\begin{cases} \mathbf{n}_1 \cdot \mathbf{x} = b_1 \\ \mathbf{n}_2 \cdot \mathbf{x} = b_2 \\ \mathbf{n}_3 \cdot \mathbf{x} = b_3 \end{cases} \Leftrightarrow \begin{cases} \mathbf{n}_1 \cdot \mathbf{x} & = b_1 \\ (\mathbf{n}_2 - \alpha_2 \mathbf{n}_1) \cdot \mathbf{x} & = b_2 - \alpha_2 b_1 \\ (\mathbf{n}_3 - \alpha_3 \mathbf{n}_1) \cdot \mathbf{x} & = b_3 - \alpha_3 b_1 \end{cases}$$

where we pick α_2 and α_3 such that the new normal vectors $\mathbf{n}_2' = \mathbf{n}_2 - \alpha_2\mathbf{n}_1$ and $\mathbf{n}_3' = \mathbf{n}_3 - \alpha_3\mathbf{n}_1$ have a zero 1st component *i.e.* $\mathbf{n}_2' = (0, a_{22}', a_{23}')$, $\mathbf{n}_3' = (0, a_{32}', a_{33}')$. At the next step, one defines a $\mathbf{n}_3'' = \mathbf{n}_3' - \beta_3\mathbf{n}_2'$ picking β_3 so that the 1st and 2nd components of $\mathbf{n}_3''$ are zero, *i.e.* $\mathbf{n}_3'' = (0, 0, a_{33}'')$. And the resulting system of equations is then easy to solve by *backward substitution.* This is **Gaussian Elimination** which in general requires swapping of equations to avoid dividing by small numbers. We could also pick the α's and β's to orthogonalize the $\mathbf{n}$'s, just as in the Gram-Schmidt procedure. That is better in terms of roundoff error and does not require equation swapping but is computationally twice as expensive as Gaussian elimination.

Linear Transformation of vectors into vectors

▶ View $\mathbf{b}$ as a linear transformation of $\mathbf{x}$.

Here A is a 'black box' that transforms the vector input $\mathbf{x}$ into the vector output $\mathbf{b}$. This is the most general view of $A\mathbf{x} = \mathbf{b}$. The transformation is linear, this means that

$$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha(A\mathbf{x}) + \beta(A\mathbf{y}), \quad \forall\, \alpha, \beta \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{173}$$

This can be checked directly from the explicit definition of matrix-vector multiply:

$$\sum_k a_{ik}(\alpha x_k + \beta y_k) = \sum_k \alpha a_{ik} x_k + \sum_k \beta a_{ik} y_k.$$

This linearity property is a key property because if A is really a black box (*e.g.* the "matrix" is not actually known, it's just a machine that takes a vector and spits out another vector) we can figure out the effect of A onto any vector $\mathbf{x}$ once we know $A\mathbf{e}_1$, $A\mathbf{e}_2, \ldots, A\mathbf{e}_n$.

This transformation view of matrices leads to the following extra rules of matrix manipulations.

Matrix-Matrix addition

$$A\mathbf{x} + B\mathbf{x} = (A+B)\mathbf{x} \;\Leftrightarrow\; \sum_k a_{ik}x_k + \sum_k B_{ik}x_k = \sum_k (a_{ik} + B_{ik})x_k, \;\forall x_k \tag{174}$$

so matrices are added *components by components* and $A+B = B+A$, $(A+B)+C = A+(B+C)$. The zero matrix is the matrix whose entries are all zero.

Matrix-scalar multiply

$$A(\alpha\mathbf{x}) = (\alpha A)\mathbf{x} \Leftrightarrow \sum_k a_{ik}(\alpha x_k) = \sum_k (\alpha a_{ik})x_k, \;\forall \alpha, x_k \tag{175}$$

so multiplication by a scalar is also done component by component and $\alpha(\beta A) = (\alpha\beta)A = \beta(\alpha A)$.

In other words, matrices can be seen as elements of a vector space! This point of view is also useful in some instances (in fact, computer languages like C and Fortran typically store matrices as long vectors. Fortran stores it column by column, and C row by row). The set of *orthogonal* matrices does NOT form a vector space because the sum of two orthogonal matrices is not, in general, an orthogonal matrix. The set of orthogonal matrices is a *group*, the *orthogonal group* $O(3)$ (for 3-by-3 matrices). The *special orthogonal* group $SO(3)$ is the set of all 3-by-3 proper orthogonal matrices, *i.e.* orthogonal matrices with determinant =+1 that correspond to pure rotation, not reflections. The motion of a rigid body about its center of inertia is a motion in $SO(3)$, not $\mathbb{R}^3$. $SO(3)$ is the *configuration space* of a rigid body.

Exercises

▷ Pick a random 3-by-3 matrix A and a vector $\mathbf{b}$, ideally in matlab using its

```
» A=randn(3),b=randn(3,1).
```

Solve $Ax = b$ using Cramer's rule and Gaussian Elimination. Ideally again in Matlab. Matlab knows all about matrices and vectors. To compute $\det(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) = \det(\mathbf{A})$ and $\det(\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3)$ in matlab, simply use

» `det(A)`, `det(b,A(:,2),A(:,3))`.

Type `help matfun`, or `help elmat`, and or `demos` for a peek at Matlab capabilities.

Chapter 2

Vector Calculus

1 Vector functions and their derivatives

Vector Calculus deals with vector functions. We begin with vector functions $\boldsymbol{a}(t)$ of one real variable t. The function $\boldsymbol{a}(t)$ could represent the position vector of a particle at time t, or its velocity or its acceleration, or a time-dependent force, for example. The tip of vector $\boldsymbol{a}(t)$ traces a curve $\mathcal{C}$ with respect to its tail. The difference $\Delta\boldsymbol{a}(t) = \boldsymbol{a}(t') - \boldsymbol{a}(t)$ is a secant vector for that curve.

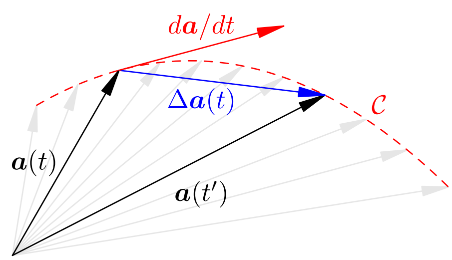


Fig. 2.1: A vector function $\boldsymbol{a}(t)$ at t and $t' > t$ with secant vector $\Delta\boldsymbol{a}(t) = \boldsymbol{a}(t') - \boldsymbol{a}(t)$ and $d\boldsymbol{a}/dt$ at t.

We assume that the vector function $\boldsymbol{a}(t)$ is *continuous* at t, that is

$$\lim_{t'\to t} \boldsymbol{a}(t') = \boldsymbol{a}(t), \tag{1}$$

then the magnitude of the secant vector $\Delta\boldsymbol{a}(t) = \boldsymbol{a}(t') - \boldsymbol{a}(t)$ goes to zero as $\Delta t = t' - t \to 0$. The limit of the ratio $\Delta\boldsymbol{a}(t)/\Delta t$, if it exists, is the derivative of $\boldsymbol{a}(t)$ at t:

$$\frac{d\boldsymbol{a}}{dt} \triangleq \lim_{t'\to t} \frac{\boldsymbol{a}(t') - \boldsymbol{a}(t)}{t' - t} = \lim_{\Delta t\to 0} \frac{\Delta\boldsymbol{a}(t)}{\Delta t}. \tag{2}$$

It is intuitively clear that the vector derivative $d\boldsymbol{a}/dt$ is tangent to the curve at $\boldsymbol{a}(t)$ since it is the scaled limit of secant vectors. We use Newton's compact dot notation for time derivatives: $\dot{\boldsymbol{a}} = d\boldsymbol{a}/dt$, $\ddot{\boldsymbol{a}} = d^2\boldsymbol{a}/dt^2$, etc. when there is no risk of confusion.

Differentiation rules. Differentiation rules for vector functions are similar to those of simple functions. The derivative of a sum of vector functions is the sum of the derivatives,

$$\frac{d}{dt}(\boldsymbol{a}+\boldsymbol{b}) = \frac{d\boldsymbol{a}}{dt} + \frac{d\boldsymbol{b}}{dt}. \tag{3}$$

There are several kinds of products involving vectors, but the product rules are all similar

$$\frac{d}{dt}(\alpha\boldsymbol{a}) = \frac{d\alpha}{dt}\boldsymbol{a} + \alpha\frac{d\boldsymbol{a}}{dt}, \tag{4}$$

$$\frac{d}{dt}(\boldsymbol{a}\cdot\boldsymbol{b}) = \frac{d\boldsymbol{a}}{dt}\cdot\boldsymbol{b} + \boldsymbol{a}\cdot\frac{d\boldsymbol{b}}{dt}, \tag{5}$$

$$\frac{d}{dt}(\boldsymbol{a}\times\boldsymbol{b}) = \frac{d\boldsymbol{a}}{dt}\times\boldsymbol{b} + \boldsymbol{a}\times\frac{d\boldsymbol{b}}{dt}, \tag{6}$$

$$\frac{d}{dt}[(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}] = (\frac{d\boldsymbol{a}}{dt}\times\boldsymbol{b})\cdot\boldsymbol{c} + (\boldsymbol{a}\times\frac{d\boldsymbol{b}}{dt})\cdot\boldsymbol{c} + (\boldsymbol{a}\times\boldsymbol{b})\cdot\frac{d\boldsymbol{c}}{dt}, \tag{7}$$

therefore

$$\frac{d}{dt}\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = \det(\frac{d\boldsymbol{a}}{dt},\boldsymbol{b},\boldsymbol{c}) + \det(\boldsymbol{a},\frac{d\boldsymbol{b}}{dt},\boldsymbol{c}) + \det(\boldsymbol{a},\boldsymbol{b},\frac{d\boldsymbol{c}}{dt}). \tag{8}$$

All of these are as expected but the formula for the derivative of a determinant is worth noting because it generalizes to any dimension and the reader used to deeply ingrained visual recipes for evaluating simple determinants may not realize that a determinant is a sum of products.[1] The proofs of these product rules from the limit definition are similar to those for functions of one variable. Some important consequences of theses differentiation rules are highlighted below.

Derivative of a magnitude. The derivative of a magnitude $|\boldsymbol{a}|$ can be computed from the vector derivative $d\boldsymbol{a}/dt$ using the chain rule and the product rule applied to $|\boldsymbol{a}| = \sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}$,

$$\frac{d|\boldsymbol{a}|}{dt} = \frac{d\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}}{dt} = \frac{1}{2\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}}2\boldsymbol{a}\cdot\frac{d\boldsymbol{a}}{dt} = \hat{\boldsymbol{a}}\cdot\frac{d\boldsymbol{a}}{dt}. \tag{9}$$

Thus the rate of change of the magnitude $|\boldsymbol{a}|$ is the orthogonal projection of $d\boldsymbol{a}/dt$ onto the direction $\hat{\boldsymbol{a}}$. Consequently, if $|\boldsymbol{a}|$ is constant then

$$\boldsymbol{a}\cdot\frac{d\boldsymbol{a}}{dt} = 0 \;\Leftrightarrow\; \frac{d\boldsymbol{a}}{dt} = \boldsymbol{\omega}(t)\times\boldsymbol{a} \tag{10}$$

[1] For determinants in $\mathbb{R}^3$ it reads

$$\frac{d}{dt}\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} \dot{a}_1 & b_1 & c_1 \\ \dot{a}_2 & b_2 & c_2 \\ \dot{a}_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} a_1 & \dot{b}_1 & c_1 \\ a_2 & \dot{b}_2 & c_2 \\ a_3 & \dot{b}_3 & c_3 \end{vmatrix} + \begin{vmatrix} a_1 & b_1 & \dot{c}_1 \\ a_2 & b_2 & \dot{c}_2 \\ a_3 & b_3 & \dot{c}_3 \end{vmatrix}$$

and of course we could also take the derivatives along rows instead of columns.

for some angular rotation rate vector $\boldsymbol{\omega}(t)$, that depends on t in general. The vector derivative $d\boldsymbol{a}/dt$ is orthogonal to $\boldsymbol{a}$ if the magnitude $|\boldsymbol{a}|$ is constant.

Derivative of a direction. The derivative $d\hat{\boldsymbol{a}}/dt$ of *any* unit vector function $\hat{\boldsymbol{a}}(t)$ is always orthogonal to $\hat{\boldsymbol{a}}$ since $|\hat{\boldsymbol{a}}|^2 = \hat{\boldsymbol{a}} \cdot \hat{\boldsymbol{a}} = 1$ for all t. This follows from (9) but is worth deriving again directly from the product rule,

$$\hat{\boldsymbol{a}} \cdot \frac{d\hat{\boldsymbol{a}}}{dt} = \frac{1}{2}\frac{d}{dt}\left(\hat{\boldsymbol{a}} \cdot \hat{\boldsymbol{a}}\right) = 0. \tag{11}$$

Derivative of magnitude and direction. The rate of change of an arbitrary vector $\boldsymbol{a}(t)$ arises from changes in both its magnitude $a = |\boldsymbol{a}|$ and its direction $\hat{\boldsymbol{a}}$. The product rule gives

$$\frac{d\boldsymbol{a}}{dt} = \frac{d}{dt}\big(a\,\hat{\boldsymbol{a}}\big) = \frac{da}{dt}\hat{\boldsymbol{a}} + a\frac{d\hat{\boldsymbol{a}}}{dt} \tag{12}$$

and it follows from (11) that these two vector components are orthogonal to each other (fig. 2.2). Since there is always a vector $\boldsymbol{\omega}(t)$ such that $d\hat{\boldsymbol{a}}/dt = \boldsymbol{\omega} \times \hat{\boldsymbol{a}}$, the general vector derivative (12) can always be written

$$\frac{d\boldsymbol{a}}{dt} = \frac{d}{dt}\big(a\,\hat{\boldsymbol{a}}\big) = \frac{da}{dt}\,\hat{\boldsymbol{a}} + \boldsymbol{\omega}(t) \times \boldsymbol{a}, \tag{13}$$

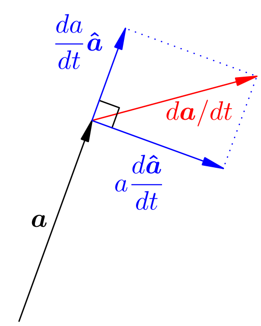


Fig. 2.2: Orthogonal components of $d\boldsymbol{a}/dt$.

highlighting that the rate of change of $\boldsymbol{a}$ arises from two orthogonal contributions: one in the direction of $\boldsymbol{a}$ due to a change in its magnitude $a = |\boldsymbol{a}|$, and one orthogonal to $\boldsymbol{a}$ due to an instantaneous rotation about $\boldsymbol{\omega}(t)$.

Derivatives of cylindrical and spherical directions. The position vector $\boldsymbol{r} = \overrightarrow{OP}$ is

$$\boldsymbol{r} = r\hat{\boldsymbol{r}} = \rho\hat{\boldsymbol{\rho}} + z\hat{\boldsymbol{z}} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}} \tag{14}$$

in spherical, cylindrical and cartesian coordinates, respectively. For a moving particle, $\boldsymbol{r} = \boldsymbol{r}(t)$, but the cartesian direction vectors $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ are fixed independent of the particle motion, however the radial directions $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\rho}}$ depend on the particle position and thus vary with t in general. Hence the time derivative of (14) gives

$$\frac{d\boldsymbol{r}}{dt} = \dot{r}\hat{\boldsymbol{r}} + r\frac{d\hat{\boldsymbol{r}}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\frac{d\hat{\boldsymbol{\rho}}}{dt} + \dot{z}\hat{\boldsymbol{z}} = \dot{x}\hat{\boldsymbol{x}} + \dot{y}\hat{\boldsymbol{y}} + \dot{z}\hat{\boldsymbol{z}}.$$

The horizontal radial direction $\hat{\boldsymbol{\rho}}(\varphi)$ depends only on the azimuth φ while the radial direction $\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}(\theta, \varphi)$ is a function of the polar angle θ and the azimuth φ. The changes in $\hat{\boldsymbol{\rho}}(t)$ and $\hat{\boldsymbol{r}}(t)$ thus arise from θ and φ varying with t.

Cylindrical direction vectors. Let $\hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{z}} \times \hat{\boldsymbol{\rho}}$ thus $\{\hat{\boldsymbol{z}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}\}$ is a right handed orthonormal basis (fig. 2.4). If $\varphi = \varphi(t)$ then $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\varphi}}$ rotate about $\hat{\boldsymbol{z}}$ at angular rotation rate $\dot{\varphi} = d\varphi/dt$, that is their rotation vector is $\boldsymbol{\omega} = \dot{\varphi}\hat{\boldsymbol{z}}$ and

$$\boxed{\begin{aligned} \frac{d\hat{\boldsymbol{\rho}}}{dt} &= \dot{\varphi}\,\hat{\boldsymbol{z}} \times \hat{\boldsymbol{\rho}} = \dot{\varphi}\,\hat{\boldsymbol{\varphi}}, \\ \frac{d\hat{\boldsymbol{\varphi}}}{dt} &= \dot{\varphi}\,\hat{\boldsymbol{z}} \times \hat{\boldsymbol{\varphi}} = -\dot{\varphi}\,\hat{\boldsymbol{\rho}}. \end{aligned}} \tag{15}$$

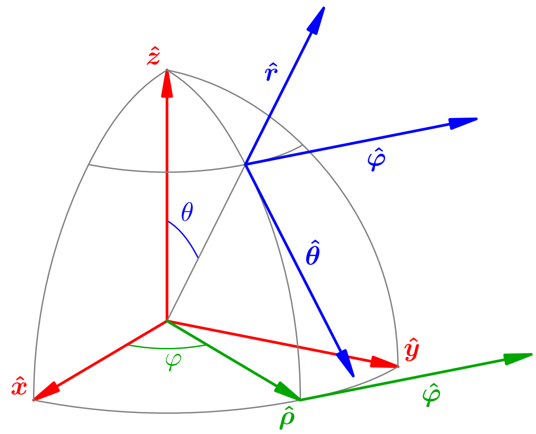


Fig. 2.3: 3D orthogonal projection of cartesian basis $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$, cylindrical basis $\{\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{z}}\}$ and spherical basis $\{\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varphi}}\}$. The vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ do not depend on position, in particular they do not depend on θ, φ. The vectors $\{\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}\}$ are orthogonal to $\hat{\boldsymbol{z}}$ and depend only on φ.

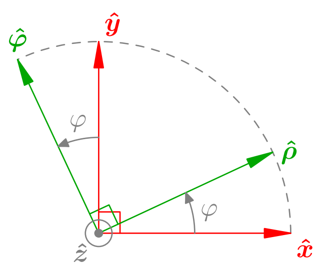


Fig. 2.4: Cartesian and cylindrical directions

These results (15) can also be derived from the hybrid representations

$$\hat{\boldsymbol{\rho}} = \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}}, \qquad \hat{\boldsymbol{\varphi}} = -\sin\varphi\,\hat{\boldsymbol{x}} + \cos\varphi\,\hat{\boldsymbol{y}}, \tag{16}$$

and the chain rule

$$\frac{d\hat{\boldsymbol{\rho}}}{dt} = \frac{d\hat{\boldsymbol{\rho}}}{d\varphi}\frac{d\varphi}{dt} = \dot{\varphi}\,(-\sin\varphi\,\hat{\boldsymbol{x}} + \cos\varphi\,\hat{\boldsymbol{y}}) = \dot{\varphi}\,\hat{\boldsymbol{\varphi}},$$

$$\frac{d\hat{\boldsymbol{\varphi}}}{dt} = \frac{d\hat{\boldsymbol{\varphi}}}{d\varphi}\frac{d\varphi}{dt} = \dot{\varphi}\,(-\cos\varphi\,\hat{\boldsymbol{x}} - \sin\varphi\,\hat{\boldsymbol{y}}) = -\dot{\varphi}\,\hat{\boldsymbol{\rho}}.$$

Spherical direction vectors. Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\varphi}} \times \hat{\boldsymbol{r}}$ with $\hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{z}} \times \hat{\boldsymbol{\rho}}$ as before. The relationship between all those direction vectors is illustrated in figs. 2.3 and 2.5. The radial $\hat{\boldsymbol{r}}$, polar $\hat{\boldsymbol{\theta}}$ and azimuthal $\hat{\boldsymbol{\varphi}}$ direction vectors rotate rigidly since their magnitudes are fixed and the angles between them are fixed. There are two sources of rotation now, $\dot{\varphi}\,\hat{\boldsymbol{z}}$ and $\dot{\theta}\,\hat{\boldsymbol{\varphi}}$, and the spherical rotation vector is $\boldsymbol{\omega} = \dot{\varphi}\,\hat{\boldsymbol{z}} + \dot{\theta}\,\hat{\boldsymbol{\varphi}}$, each of the basis vectors evolve according to the rotation equation

$$\frac{d\boldsymbol{a}}{dt} = \boldsymbol{\omega} \times \boldsymbol{a} = \left(\dot{\varphi}\,\hat{\boldsymbol{z}} + \dot{\theta}\,\hat{\boldsymbol{\varphi}}\right) \times \boldsymbol{a}.$$

That is

$$\boxed{\begin{aligned}
\frac{d\hat{\boldsymbol{r}}}{dt} &= \left(\dot{\varphi}\,\hat{\boldsymbol{z}} + \dot{\theta}\,\hat{\boldsymbol{\varphi}}\right) \times \hat{\boldsymbol{r}} = \dot{\varphi}\sin\theta\,\hat{\boldsymbol{\varphi}} + \dot{\theta}\,\hat{\boldsymbol{\theta}},\\
\frac{d\hat{\boldsymbol{\theta}}}{dt} &= \left(\dot{\varphi}\,\hat{\boldsymbol{z}} + \dot{\theta}\,\hat{\boldsymbol{\varphi}}\right) \times \hat{\boldsymbol{\theta}} = \dot{\varphi}\cos\theta\,\hat{\boldsymbol{\varphi}} - \dot{\theta}\,\hat{\boldsymbol{r}},\\
\frac{d\hat{\boldsymbol{\varphi}}}{dt} &= \left(\dot{\varphi}\,\hat{\boldsymbol{z}} + \dot{\theta}\,\hat{\boldsymbol{\varphi}}\right) \times \hat{\boldsymbol{\varphi}} = -\dot{\varphi}\,\hat{\boldsymbol{\rho}} = -\dot{\varphi}\,\sin\theta\,\hat{\boldsymbol{r}} - \dot{\varphi}\,\cos\theta\,\hat{\boldsymbol{\theta}}.
\end{aligned}} \tag{17}$$

These results can also be obtained from the chain rule

$$\frac{d\hat{\boldsymbol{r}}}{dt} = \frac{\partial \hat{\boldsymbol{r}}}{\partial \theta}\,\dot{\theta} + \frac{\partial \hat{\boldsymbol{r}}}{\partial \varphi}\,\dot{\varphi}, \qquad \frac{d\hat{\boldsymbol{\theta}}}{dt} = \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta}\,\dot{\theta} + \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \varphi}\,\dot{\varphi}, \qquad \frac{d\hat{\boldsymbol{\varphi}}}{dt} = \frac{d\hat{\boldsymbol{\varphi}}}{d\varphi}\,\dot{\varphi}.$$

applied to

$$\hat{\boldsymbol{r}} = \cos\theta\,\hat{\boldsymbol{z}} + \sin\theta\,\hat{\boldsymbol{\rho}}, \qquad \hat{\boldsymbol{\theta}} = -\sin\theta\,\hat{\boldsymbol{z}} + \cos\theta\,\hat{\boldsymbol{\rho}}, \tag{18}$$

with $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\varphi}}$ defined in (16).

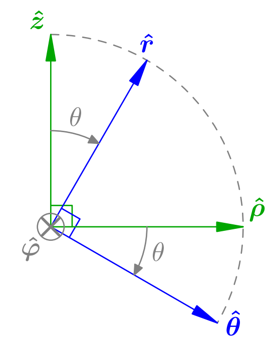


Fig. 2.5: Cylindrical and spherical directions

Velocity. The velocity in cartesian coordinates reads

$$\frac{d\boldsymbol{r}}{dt} = \dot{x}\hat{\boldsymbol{x}} + \dot{y}\hat{\boldsymbol{y}} + \dot{z}\hat{\boldsymbol{z}} \tag{19}$$

but in cylindrical coordinates

$$\frac{d\boldsymbol{r}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\frac{d\hat{\boldsymbol{\rho}}}{dt} + \dot{z}\hat{\boldsymbol{z}} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\dot{\varphi}\hat{\boldsymbol{\varphi}} + \dot{z}\hat{\boldsymbol{z}} \tag{20}$$

and in spherical

$$\frac{d\boldsymbol{r}}{dt} = \dot{r}\hat{\boldsymbol{r}} + r\frac{d\hat{\boldsymbol{r}}}{dt} = \dot{r}\hat{\boldsymbol{r}} + r\dot{\varphi}\sin\theta\,\hat{\boldsymbol{\varphi}} + r\dot{\theta}\,\hat{\boldsymbol{\theta}}. \tag{21}$$

Acceleration. The acceleration has the cartesian expression

$$\frac{d^2\boldsymbol{r}}{dt^2} = \ddot{x}\hat{\boldsymbol{x}} + \ddot{y}\hat{\boldsymbol{y}} + \ddot{z}\hat{\boldsymbol{z}} \tag{22}$$

but in cylindrical coordinates, the derivative of (20) using (15), yields

$$\begin{aligned}\frac{d^2\boldsymbol{r}}{dt^2} &= \ddot{\rho}\hat{\boldsymbol{\rho}} + 2\dot{\rho}\frac{d\hat{\boldsymbol{\rho}}}{dt} + \rho\frac{d^2\hat{\boldsymbol{\rho}}}{dt^2} + \ddot{z}\hat{\boldsymbol{z}} \\ &= \left(\ddot{\rho} - \rho\dot{\varphi}^2\right)\hat{\boldsymbol{\rho}} + \left(\rho\ddot{\varphi} + 2\dot{\rho}\dot{\varphi}\right)\hat{\boldsymbol{\varphi}} + \ddot{z}\hat{\boldsymbol{z}}.\end{aligned} \tag{23}$$

Likewise in spherical coordinates, repeated use of (17) yields

$$\begin{aligned}\frac{d^2\boldsymbol{r}}{dt^2} &= \ddot{r}\hat{\boldsymbol{r}} + 2\dot{r}\frac{d\hat{\boldsymbol{r}}}{dt} + r\frac{d^2\hat{\boldsymbol{r}}}{dt^2} = \\ &= \left(\ddot{r} - r\dot{\theta}^2 - r\dot{\varphi}^2\sin^2\theta\right)\hat{\boldsymbol{r}} + \left(r\ddot{\theta} + 2\dot{r}\dot{\theta} - r\dot{\varphi}^2\sin\theta\cos\theta\right)\hat{\boldsymbol{\theta}} \\ &\quad + \left(r\ddot{\varphi}\sin\theta + 2r\dot{\theta}\dot{\varphi}\cos\theta + 2\dot{r}\dot{\varphi}\sin\theta\right)\hat{\boldsymbol{\varphi}}.\end{aligned} \tag{24}$$

Exercises

1. Show that for any vector function $\boldsymbol{u}(t)$

   (a)
   $$\boldsymbol{u}\cdot\frac{d\boldsymbol{u}}{dt} = \frac{d}{dt}\left(\frac{\boldsymbol{u}\cdot\boldsymbol{u}}{2}\right).$$

(b)
$$\boldsymbol{u} \times \frac{d^2\boldsymbol{u}}{dt^2} = \frac{d}{dt}\left(\boldsymbol{u} \times \frac{d\boldsymbol{u}}{dt}\right).$$

(c)
$$\boldsymbol{u} \times \frac{d\boldsymbol{u}}{dt} \cdot \frac{d^3\boldsymbol{u}}{dt^3} = \frac{d}{dt}\left(\boldsymbol{u} \times \frac{d\boldsymbol{u}}{dt} \cdot \frac{d^2\boldsymbol{u}}{dt^2}\right).$$

2. Show that if $\boldsymbol{u}(t)$ is any vector with constant magnitude then
$$\boldsymbol{u} \cdot \frac{d\boldsymbol{u}}{dt} = 0, \quad \forall t.$$

3. If $\boldsymbol{r} = r\hat{\boldsymbol{r}}$, show that
$$\hat{\boldsymbol{r}} \cdot \frac{d\boldsymbol{r}}{dt} = \frac{dr}{dt} \quad \text{and} \quad \boldsymbol{r} \cdot \frac{d\boldsymbol{r}}{dt} = r\frac{dr}{dt}.$$

4. Show that if $\boldsymbol{a}(t)$, $\boldsymbol{b}(t)$, $\boldsymbol{c}(t)$ are three distinct solutions $\boldsymbol{u}(t)$ of
$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{\omega}(t) \times \boldsymbol{u}$$
for the same $\boldsymbol{\omega}(t)$, then $\boldsymbol{a} \times \boldsymbol{b} \cdot \boldsymbol{c}$ is constant. Interpret geometrically.

5. If $\boldsymbol{v}(t) = d\boldsymbol{r}/dt$ and $\boldsymbol{r}_c$ is any constant vector, show that
$$\frac{d}{dt}\big((\boldsymbol{r} - \boldsymbol{r}_c) \times m\boldsymbol{v}\big) = (\boldsymbol{r} - \boldsymbol{r}_c) \times \frac{d}{dt}(m\boldsymbol{v}).$$
In mechanics, $\boldsymbol{L} \triangleq (\boldsymbol{r} - \boldsymbol{r}_c) \times m\boldsymbol{v}$ is the *angular momentum* with respect to the point $\boldsymbol{r}_c$ of the particle of mass m with velocity $\boldsymbol{v}$ at position $\boldsymbol{r}$, and $(\boldsymbol{r} - \boldsymbol{r}_c) \times d(m\boldsymbol{v})/dt = (\boldsymbol{r} - \boldsymbol{r}_c) \times \boldsymbol{F}$ by Newton's law. The cross product $(\boldsymbol{r} - \boldsymbol{r}_c) \times \boldsymbol{F} = \boldsymbol{T}$ is the *torque* about the point $\boldsymbol{r}_c$.

6. Consider $\boldsymbol{r}(t) = \boldsymbol{r}_C + \boldsymbol{a}\cos t + \boldsymbol{b}\sin t$ where $\boldsymbol{r}_C$, $\boldsymbol{a}$, $\boldsymbol{b}$ are arbitrary constant vectors in 3D space. Sketch $\boldsymbol{r}(t)$ and indicate all points where $\boldsymbol{r} \cdot d\boldsymbol{r}/dt = 0$ for both $\boldsymbol{r}_C$ zero or non-zero.

7. The position of a particle at time t is given by $\boldsymbol{r}(t) = \boldsymbol{a}\cos\theta(t) + \boldsymbol{b}\sin\theta(t)$, with $\theta(t) = 2\cos t$ and $\boldsymbol{a}$, $\boldsymbol{b}$ arbitrary constants. What are the velocity and the acceleration of the particle? Sketch the particle motion and a few representative position, velocity and acceleration vectors.

8. Verify (17) for the derivatives of the direction vectors using the chain rule.

9. Verify (23) and (24) for the acceleration in cylindrical and spherical coordinates.

10. A bead is rotating at constant angular velocity ω about a circular hoop. The hoop rotates about one of its diameters at constant angular velocity Ω. What are the bead velocity and acceleration? If the particle has mass m, what is the torque on the hoop about the center of the hoop? (see exercise 5 for the definition of torque.)

2 Newtonian mechanics

2.1 Motion of a single particle

In classical mechanics, the motion of a particle of constant mass m is governed by Newton's law

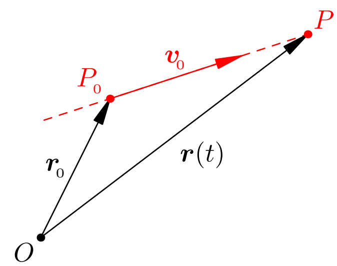
$$\boldsymbol{F} = m\boldsymbol{a}, \tag{25}$$

where $\boldsymbol{F}$ is the resultant of the forces acting on the particle and $\boldsymbol{a}(t) = d\boldsymbol{v}/dt = d^2\boldsymbol{r}/dt^2$ is its acceleration, with $\boldsymbol{r}(t)$ its position vector. Newton's law is a vector differential equation for the particle position $\boldsymbol{r}(t)$ at time t given the mass m and the force $\boldsymbol{F}(t)$. We cover a few fundamental examples.

Constant velocity

If $\boldsymbol{F} = 0$ then $\boldsymbol{a} = d\boldsymbol{v}/dt = 0$ so the velocity of the particle is constant, $\boldsymbol{v}(t) = \boldsymbol{v}_0$ say, and its position is given by the vector differential equation $d\boldsymbol{r}/dt = \boldsymbol{v}_0$ whose solution is

$$\boldsymbol{r}(t) = \boldsymbol{r}_0 + t\boldsymbol{v}_0 \tag{26}$$



where $\boldsymbol{r}_0$ is a constant of integration whose meaning is clear, it is the position of the particle at time $t = 0$. If the particle is at P_0 at time t_0 then

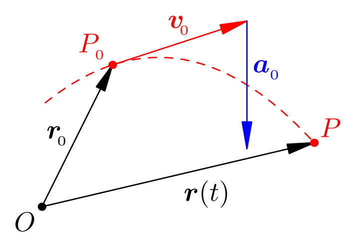
$$\boldsymbol{r}(t) = \boldsymbol{r}_0 + (t - t_0)\boldsymbol{v}_0. \tag{27}$$

The trajectory of the particle is a straight line parallel to $\boldsymbol{v}_0$.

Constant acceleration

If $\boldsymbol{F} = \boldsymbol{F}_0$ is a non-zero vector constant then the acceleration $\boldsymbol{a} = \boldsymbol{a}_0 = \boldsymbol{F}_0/m$ is a constant and

$$\frac{d^2\boldsymbol{r}}{dt^2} = \frac{d\boldsymbol{v}}{dt} = \boldsymbol{a}(t) = \boldsymbol{a}_0 \tag{28}$$



where $\boldsymbol{a}_0$ is a time-independent vector. Integrating once we find $\dot{\boldsymbol{r}} = \boldsymbol{v} = \boldsymbol{v}_0 + t\boldsymbol{a}_0$ and another integration yields

$$\boldsymbol{r}(t) = \boldsymbol{r}_0 + t\boldsymbol{v}_0 + \frac{t^2}{2}\boldsymbol{a}_0 \tag{29}$$

where $\boldsymbol{v}_0$ and $\boldsymbol{r}_0$ are *vector* constants of integration. They are easily interpreted as the velocity and position at $t = 0$, respectively. If the initial data $\boldsymbol{r}_0$ and $\boldsymbol{v}_0$ are given at t_0, it suffices to replace t by $t - t_0$ in (29).

The trajectory is a parabola passing through $\boldsymbol{r}_0$ parallel to $\boldsymbol{v}_0$ at $t = 0$. The parabolic motion is in the plane through $\boldsymbol{r}_0$ that is parallel to $\boldsymbol{v}_0$ and $\boldsymbol{a}_0$ but the origin O may not be in that plane. We can write this parabola in standard form by selecting cartesian axes such that $\boldsymbol{a}_0 = -g\hat{\boldsymbol{y}}$, $\boldsymbol{v}_0 = u_0\hat{\boldsymbol{x}} + v_0\hat{\boldsymbol{y}}$ and $\boldsymbol{r}_0 = 0$ then

$$\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}} = -g\frac{t^2}{2}\hat{\boldsymbol{y}} + (u_0\hat{\boldsymbol{x}} + v_0\hat{\boldsymbol{y}})t$$

yielding $x = u_0 t$, $y = v_0 t - gt^2/2$. Eliminating t when $u_0 \neq 0$ yields

$$y = \frac{v_0}{u_0} x - \frac{g}{2u_0^2} x^2.$$

Uniform rotation

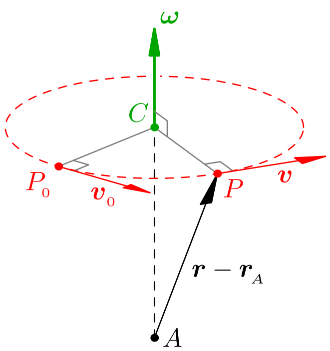


Fig. 2.6: Rotation about axis $(A, \boldsymbol{\omega})$. 3D view.

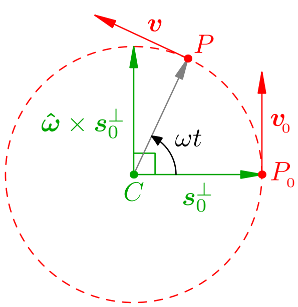


Fig. 2.7: Rotation about axis $(A, \boldsymbol{\omega})$. 2D view $\perp \boldsymbol{\omega}$.

If a particle rotates with angular frequency ω about an *axis* $(A, \hat{\boldsymbol{\omega}})$ that passes through point A and is parallel to unit vector $\hat{\boldsymbol{\omega}}$, then its velocity is

$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} = \boldsymbol{\omega} \times (\boldsymbol{r} - \boldsymbol{r}_A) \tag{30}$$

where $\boldsymbol{\omega} = \omega \hat{\boldsymbol{\omega}}$ is the *angular velocity* and $\boldsymbol{r}_A = \overrightarrow{OA}$ is the position vector of A. If $\boldsymbol{\omega}$ and A are constants, this is circular motion about the axis $(A, \hat{\boldsymbol{\omega}})$.

Let $\boldsymbol{s}(t) = \overrightarrow{AP} = \boldsymbol{r}(t) - \boldsymbol{r}_A$ be the position vector of P with respect to A. The motion is best described in terms of the intrinsic system of coordinates with the center of the circle C as reference point and the orthogonal vector basis $\{\boldsymbol{s}_0^\parallel, \boldsymbol{s}_0^\perp, \hat{\boldsymbol{\omega}} \times \boldsymbol{s}_0^\perp\}$ where

$$\boldsymbol{s}_0 \triangleq \overrightarrow{AP_0}, \quad \boldsymbol{s}_0^\parallel = (\boldsymbol{s}_0 \cdot \hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}}, \quad \boldsymbol{s}_0^\perp = \boldsymbol{s}_0 - \boldsymbol{s}_0^\parallel, \tag{31}$$

that is derived from the initial position P_0 with respect to A and the rotation direction $\hat{\boldsymbol{\omega}}$. The position of the circle center C with respect to an arbitrary origin O is then

$$\overrightarrow{OC} = \overrightarrow{OA} + \boldsymbol{s}_0^\parallel$$

and the vector $\overrightarrow{CP}$ simply rotates at angular velocity ω in the plane $\{\boldsymbol{s}_0^\perp, \hat{\boldsymbol{\omega}} \times \boldsymbol{s}_0^\perp\}$ orthogonal to $\boldsymbol{\omega}$, that is (fig. 2.7)

$$\overrightarrow{CP} = \boldsymbol{s}_0^\perp \cos\omega t + (\hat{\boldsymbol{\omega}} \times \boldsymbol{s}_0^\perp) \sin\omega t,$$

such that the angle between $\overrightarrow{CP_0}$ and $\overrightarrow{CP}$ is ωt. The position vector $\boldsymbol{r}(t) = \overrightarrow{OP} = \overrightarrow{OC} + \overrightarrow{CP}$ from an arbitrary origin O is then

$$\boldsymbol{r}(t) = \boldsymbol{r}_C + \boldsymbol{s}_0^\perp \cos\omega t + (\hat{\boldsymbol{\omega}} \times \boldsymbol{s}_0^\perp) \sin\omega t, \tag{32}$$

where $\boldsymbol{r}_C = \boldsymbol{r}_A + \boldsymbol{s}_0^\parallel$ is the position vector of the circle center C determined from the initial position P_0 and the rotation direction $\hat{\boldsymbol{\omega}}$, with $\boldsymbol{s}_0^\parallel$ and $\boldsymbol{s}_0^\perp$ defined from the initial position $\overrightarrow{AP_0}$ in (31).

Motion under a central force

A force $\boldsymbol{F} = F(r)\,\hat{\boldsymbol{r}}$ where $r = |\boldsymbol{r}|$ that always points toward the origin (if $F(r) < 0$, away if $F(r) > 0$) and depends only on the distance to the origin is called a *central force*. The gravitational force for planetary motion and the Coulomb force in electromagnetism are of that kind. Newton's law for a particle submitted to such a force is

$$m\frac{d\boldsymbol{v}}{dt} = F(r)\,\hat{\boldsymbol{r}} \tag{33}$$

where $\boldsymbol{v} = d\boldsymbol{r}/dt$ and $\boldsymbol{r}(t) = r\hat{\boldsymbol{r}}$ is the position vector of the particle, hence both r and $\hat{\boldsymbol{r}}$ are functions of time t, in general. Motion due to such a force has two conserved quantities, *angular momentum* and *energy*.

Conservation of angular momentum. The cross product of (33) with $\boldsymbol{r}$ yields

$$\boldsymbol{r} \times \frac{d\boldsymbol{v}}{dt} = 0 \Leftrightarrow \frac{d}{dt}(\boldsymbol{r} \times \boldsymbol{v}) = 0 \Leftrightarrow \boldsymbol{r} \times \boldsymbol{v} = \boldsymbol{r}_0 \times \boldsymbol{v}_0 \triangleq \frac{\boldsymbol{L}_0}{m} \tag{34}$$

where $\boldsymbol{L}_0 = L_0\,\hat{\boldsymbol{L}}_0$ is a constant vector (exercise 5 in the previous section). The vector $\boldsymbol{L} = \boldsymbol{r} \times m\boldsymbol{v}$ is called *angular momentum* in physics. The fact that $\boldsymbol{r} \times \boldsymbol{v}$ is constant implies that the motion is in the plane that passes through the origin O and is orthogonal to $\boldsymbol{L}_0$ (why?) and that *'the radius vector sweeps equal areas in equal times'*. Indeed $\boldsymbol{v}dt = d\boldsymbol{r}$ is the displacement during the infinitesimal time span dt thus

$$\boldsymbol{r} \times \boldsymbol{v}dt = \boldsymbol{r} \times d\boldsymbol{r} = \frac{\boldsymbol{L}_0}{m}\,dt$$

but

$$dA(t) = \frac{1}{2}\,|\boldsymbol{r}(t) \times d\boldsymbol{r}(t)| = \frac{L_0}{2m}\,dt$$

is the infinitesimal triangular area swept by $\boldsymbol{r}(t)$ in time dt. This yields Kepler's law that the area swept by $\boldsymbol{r}(t)$ in time T is independent of the start time t_1

$$A = \int_{t_1}^{t_1+T} dA(t) = \int_{t_1}^{t_1+T} \frac{L_0}{2m}\,dt = \frac{L_0}{2m}\,T.$$

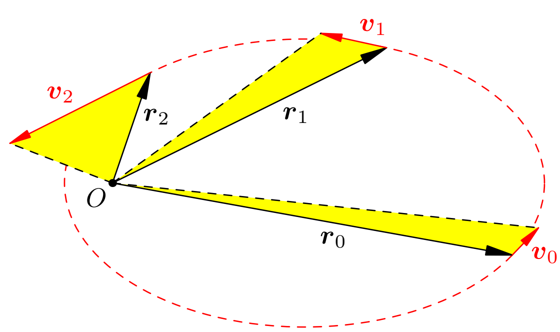


Fig. 2.8: Kepler's law: *The radius vector sweeps equal areas in equal times.* Here for the classic $F(r) = -1/r^2$ in which case the trajectories are ellipses with the origin as a focus.

Conservation of energy. The dot product of (33) with $\boldsymbol{v}$ yields

$$m\frac{d\boldsymbol{v}}{dt} \cdot \boldsymbol{v} - F(r)\,\hat{\boldsymbol{r}} \cdot \boldsymbol{v} = 0 \;\Leftrightarrow\; \frac{d}{dt}\left(m\frac{\boldsymbol{v} \cdot \boldsymbol{v}}{2} + V(r)\right) = 0, \tag{35}$$

where $V(r)$ is an antiderivative of $-F(r)$, $dV(r)/dr \triangleq -F(r)$. The function $V(r)$ is called the *potential* and $V(r) = -1/r$ for the classic inverse square law $F(r) =$

$-1/r^2$. This follows from the product rule $\boldsymbol{v} \cdot d\boldsymbol{v}/dt = d/dt(\boldsymbol{v} \cdot \boldsymbol{v})/2$ and the chain rule

$$\frac{dV(r)}{dt} = \frac{dV}{dr}\frac{dr}{dt} = \frac{dV}{dr}\hat{\boldsymbol{r}} \cdot \boldsymbol{v} = -F(r)\hat{\boldsymbol{r}} \cdot \boldsymbol{v}$$

since

$$\hat{\boldsymbol{r}} \cdot \boldsymbol{v} = \hat{\boldsymbol{r}} \cdot \frac{d\boldsymbol{r}}{dt} = \frac{d|\boldsymbol{r}|}{dt} = \frac{dr}{dt},$$

which is equation (9) applied to $\boldsymbol{r}$. Thus

$$\left(m\frac{|\boldsymbol{v}|^2}{2} + V(r)\right) = E_0 \tag{36}$$

is a constant. The first term, $m|\boldsymbol{v}|^2/2$, is the kinetic energy and the second term, $V(r)$, is the potential energy which is defined up to an arbitrary constant. The constant E_0 is the total conserved energy. Note that $V(r)$ and E_0 can be negative but $m|\boldsymbol{v}|^2/2 \geq 0$, so the physically admissible r domain is that were $V(r) \leq E_0$.

Exercises:

1. Show that $|\boldsymbol{r}(t) - \boldsymbol{r}_A|$ remains constant if $\boldsymbol{r}(t)$ evolves according to (30) with $\boldsymbol{r}_A$ constant but non-constant $\boldsymbol{\omega}(t)$. If $\hat{\boldsymbol{\omega}}$ and $\boldsymbol{r}_A$ are constants, show that $\hat{\boldsymbol{\omega}} \cdot (\boldsymbol{r} - \boldsymbol{r}_A)$ is also constant. What kind of $\boldsymbol{r}(t)$ are possible in such cases? Finally, if $\boldsymbol{\omega}$ and $\boldsymbol{r}_A$ are both constants, show that $|\boldsymbol{v}|$ is also constant. Given all those constants of motion in the latter case, what type of particle motion is left? Find the force $\boldsymbol{F}$ required to sustain the latter motion for a particle of mass m according to Newton's law.

2. Verify that (32) yields the correct initial condition at $t = 0$ and satisfies the vector differential equation (30).

3. If $d\boldsymbol{r}/dt = \boldsymbol{\omega} \times \boldsymbol{r}$ with $\boldsymbol{\omega}$ constant, show that $d^2\boldsymbol{r}_\perp/dt^2 = -\omega^2\boldsymbol{r}_\perp$ where $\boldsymbol{r}_\perp$ is the component of $\boldsymbol{r}$ perpendicular to $\boldsymbol{\omega}$ and $\omega = |\boldsymbol{\omega}|$. What happens to $\boldsymbol{r}_\parallel$? Describe all the possible $\boldsymbol{r}(t)$.

4. Find the general solution $\boldsymbol{u}(t)$ to $d\boldsymbol{u}/dt = \boldsymbol{\omega} \times \boldsymbol{u}$ for arbitrary constant $\boldsymbol{\omega}$.

5. A particle of mass m and electric charge q moving at velocity $\boldsymbol{v}$ in a magnetic field $\boldsymbol{B}$ experiences the Lorentz force $\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$. If m, q and $\boldsymbol{B}$ are constants, use Newton's law to show that the velocity $\boldsymbol{v}(t)$ rotates about $\boldsymbol{B}$ at angular rotation rate
$$\omega = -q\frac{|\boldsymbol{B}|}{m}.$$
Find $\boldsymbol{r}(t)$ and show that the particle trajectories are helices. Find the radius of rotation about the helix in terms of the position $\boldsymbol{r}_0$ and velocity $\boldsymbol{v}_0$ at some observation time t_0.

6. Find the general solution $\boldsymbol{r}(t)$ to $d^2\boldsymbol{r}/dt^2 = -\omega^2\boldsymbol{r}$ with ω real and constant. How does this differ from exercise 3?

7. Find the general solution $\boldsymbol{r}(t)$ to $d^2\boldsymbol{r}/dt^2 = \omega^2\boldsymbol{r}$ with ω real and constant.

8. Investigate $\boldsymbol{r}(t)$ if $m\,d^2\boldsymbol{r}/dt^2 = -k(r-\ell_0)\hat{\boldsymbol{r}}$ for constant m and arbitrary initial conditions $\boldsymbol{r}(0) = \boldsymbol{r}_0$ and $\boldsymbol{v}(0) = \boldsymbol{v}_0$, where $\boldsymbol{v} = d\boldsymbol{r}/dt$, $\boldsymbol{r} = r\hat{\boldsymbol{r}}$, with k, ℓ_0 constant (modeling a spring of stiffness k and rest length ℓ_0). Can you find conserved quantities? Can you define a potential $V(r)$ for this problem? What type of motion do you expect? Are there special initial conditions?

9. Investigate $\boldsymbol{r}(t)$ if $d^2\boldsymbol{r}/dt^2 = -\boldsymbol{r}/r^3$ with $\boldsymbol{r}(0) = \boldsymbol{r}_0$ and $d\boldsymbol{r}/dt(0) = \boldsymbol{v}_0$. Are there constants of motions, and if yes, what are they explicitly? What types of motion are possible?

10. A bead of constant mass m is sliding without friction along a circular hoop of radius R that rotates about the vertical diameter at constant angular velocity Ω. The only external force acting on the particle is gravity $-mg\hat{\boldsymbol{z}}$. If $\theta(t)$ specifies the angular position of the particle around the hoop, use Newton's law to derive the 2nd order differential equation for $\theta(t)$. What is the force from the hoop on the particle? Find all the possible equilibrium positions for which θ remains constant.

2.2 Motion of a system of particles (optional)

Consider N particles of mass m_i at positions $\boldsymbol{r}_i$, $i = 1,\ldots,N$. The net force acting on particle number i is $\boldsymbol{F}_i$ and Newton's law for each particle reads $m_i\ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i$. Summing over all i's yields

$$\sum_{i=1}^{N} m_i\ddot{\boldsymbol{r}}_i = \sum_{i=1}^{N} \boldsymbol{F}_i.$$

Great cancellations occur on both sides. On the left side, let $\boldsymbol{r}_i = \boldsymbol{r}_c + \boldsymbol{s}_i$, where $\boldsymbol{r}_c$ is the center of mass and $\boldsymbol{s}_i$ is the position vector of particle i with respect to the center of mass, then

$$\sum_i m_i\boldsymbol{r}_i = \sum_i m_i\left(\boldsymbol{r}_c + \boldsymbol{s}_i\right) = M\boldsymbol{r}_c + \sum_i m_i\boldsymbol{s}_i \;\Rightarrow \sum_i m_i\boldsymbol{s}_i = 0,$$

as, by definition of the center of mass $\sum_i m_i\boldsymbol{r}_i = M\boldsymbol{r}_c$, where $M = \sum_i m_i$ is the total mass. If the masses m_i are constants then $\sum_i m_i\boldsymbol{s}_i = 0 \Rightarrow \sum_i m_i\dot{\boldsymbol{s}}_i = 0$ $\Rightarrow \sum_i m_i\ddot{\boldsymbol{s}}_i = 0$. In that case, $\sum_i m_i\ddot{\boldsymbol{r}}_i = \sum_i m_i\left(\ddot{\boldsymbol{r}}_c + \ddot{\boldsymbol{s}}_i\right) = \sum_i m_i\ddot{\boldsymbol{r}}_c = M\ddot{\boldsymbol{r}}_c$. On the right-hand side, by action-reaction, all internal forces cancel out and the resultant is therefore the sum of all external forces only $\sum_i \boldsymbol{F}_i = \sum_i \boldsymbol{F}_i^{(e)} = \boldsymbol{F}^{(e)}$.

Therefore,

$$M\ddot{\boldsymbol{r}}_c = \boldsymbol{F}^{(e)} \tag{37}$$

where M is the total mass and $\boldsymbol{F}^{(e)}$ is the resultant of all external forces acting on all the particles. The motion of the center of mass of a system of particles is that of a single particle of mass M with position vector $\boldsymbol{r}_c$ under the action of the sum of all external forces. This is a fundamental theorem of mechanics.

There are also nice cancellations occurring for the motion about the center of mass. This involves considering angular momentum and torques about the center of mass. Taking the cross-product of Newton's law, $m_i\ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i$, with $\boldsymbol{s}_i$ for each particle and summing over all particles gives

$$\sum_i \boldsymbol{s}_i \times m_i\ddot{\boldsymbol{r}}_i = \sum_i \boldsymbol{s}_i \times \boldsymbol{F}_i.$$

On the left hand side, $\boldsymbol{r}_i \equiv \boldsymbol{r}_c + \boldsymbol{s}_i$ and the definition of center of mass implies $\sum_i m_i\boldsymbol{s}_i = 0$. Therefore

$$\sum_i \boldsymbol{s}_i \times m_i\ddot{\boldsymbol{r}}_i = \sum_i \boldsymbol{s}_i \times m_i(\ddot{\boldsymbol{r}}_c + \ddot{\boldsymbol{s}}_i) = \sum_i \boldsymbol{s}_i \times m_i\ddot{\boldsymbol{s}}_i = \frac{d}{dt}\left(\sum_i \boldsymbol{s}_i \times m_i\dot{\boldsymbol{s}}_i\right).$$

This last expression is the rate of change of the total angular momentum about the center of mass

$$\boldsymbol{L}_c \equiv \sum_{i=1}^{N} (\boldsymbol{s}_i \times m_i\dot{\boldsymbol{s}}_i).$$

On the right hand side, one can argue that the (internal) force exerted by particle j on particle i is in the direction of the relative position of j with respect to i, $\boldsymbol{f}_{ij} \equiv \alpha_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)$. By action-reaction the force from i onto j is $\boldsymbol{f}_{ji} = -\boldsymbol{f}_{ij} = -\alpha_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)$, and the net contribution to the torque from the internal forces will cancel out: $\boldsymbol{r}_i \times \boldsymbol{f}_{ij} + \boldsymbol{r}_j \times \boldsymbol{f}_{ji} = 0$. This is true with respect to any point and in particular, with respect to the center of mass $\boldsymbol{s}_i \times \boldsymbol{f}_{ij} + \boldsymbol{s}_j \times \boldsymbol{f}_{ji} = 0$. Hence, for the motion about the center of mass we have

$$\frac{d\boldsymbol{L}_c}{dt} = \boldsymbol{T}_c^{(e)} \tag{38}$$

where $\boldsymbol{T}^{(e)} = \sum_i \boldsymbol{s}_i \times \boldsymbol{F}_i$ is the net torque about the center of mass due to external forces only. This is another fundamental theorem, that the rate of change of the total angular momentum about the center of mass is equal to the total torque due to the external forces only.

Exercise: If $\boldsymbol{f}_{ij} = \alpha_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)$ where $\alpha_{ij} = \alpha_{ji}$ can be a function of the distance $|\boldsymbol{r}_i - \boldsymbol{r}_j|$, show algebraically and geometrically that $\boldsymbol{s}_i \times \boldsymbol{f}_{ij} + \boldsymbol{s}_j \times \boldsymbol{f}_{ji} = 0$, where $\boldsymbol{s}$ is the position vector from the center of mass such that $\boldsymbol{r} = \boldsymbol{r}_c + \boldsymbol{s}$.

2.3 Motion of a rigid body (optional)

The two vector differential equations for motion of the center of mass and evolution of the angular momentum about the center of mass are sufficient to fully determine the motion of a rigid body.

A rigid body is such that all lengths and angles are preserved within the rigid body. If A, B and C are any three points of the rigid body, then $\overrightarrow{AB} \cdot \overrightarrow{AC}$ is constant. The evolution of any vector $\overrightarrow{AB}$ is thus a rotation and

$$\frac{d}{dt}\overrightarrow{AB} = \boldsymbol{\omega}(t) \times \overrightarrow{AB}.$$

Here we show that $\boldsymbol{\omega}(t)$ is the same for any points A and B within the same rigid body. That unique angular rotation vector $\boldsymbol{\omega}(t)$ is sometimes called the *Poisson vector*.

Kinematics of a rigid body

Consider a right-handed orthonormal basis, $\{\mathbf{e}_1(t), \mathbf{e}_2(t), \mathbf{e}_3(t)\}$ tied to the body. These vectors are functions of time t because they are frozen into the body so they rotate with it. However the basis remains orthonormal as all lengths and angles are preserved in a rigid body, hence $\mathbf{e}_i(t) \cdot \mathbf{e}_j(t) = \delta_{ij}$ and

$$\frac{d}{dt}(\mathbf{e}_i \cdot \mathbf{e}_j) = \dot{\boldsymbol{e}}_i \cdot \mathbf{e}_j + \mathbf{e}_i \cdot \dot{\boldsymbol{e}}_j = 0, \tag{39}$$

for all i, j and t. In particular, $\mathbf{e}_1 \cdot \dot{\boldsymbol{e}}_1 = 0$ and

$$\dot{\boldsymbol{e}}_1 = \boldsymbol{\omega}(t) \times \mathbf{e}_1 = \omega_3\, \mathbf{e}_2 - \omega_2\, \mathbf{e}_3 \tag{40}$$

for some $\boldsymbol{\omega} = \omega_1\mathbf{e}_1 + \omega_2\mathbf{e}_2 + \omega_3\mathbf{e}_3$. The unit vector $\mathbf{e}_1$ and its derivative $\dot{\boldsymbol{e}}_1$ determine ω_2 and ω_3 but not ω_1. Applying (39) for $(i, j) = (1, 2)$ and $(1, 3)$, then substituting for $\dot{\boldsymbol{e}}_1$ from (40) yields

$$\begin{aligned} \mathbf{e}_1 \cdot \dot{\boldsymbol{e}}_2 &= -\dot{\boldsymbol{e}}_1 \cdot \mathbf{e}_2 = -\omega_3, \\ \mathbf{e}_1 \cdot \dot{\boldsymbol{e}}_3 &= -\dot{\boldsymbol{e}}_1 \cdot \mathbf{e}_3 = \omega_2. \end{aligned} \tag{41}$$

Hence $\dot{\boldsymbol{e}}_2 = -\omega_3\mathbf{e}_1 + \alpha\mathbf{e}_3$ and $\dot{\boldsymbol{e}}_3 = \omega_2\mathbf{e}_1 + \beta\mathbf{e}_2$ for some α and β, since $\mathbf{e}_2 \cdot \dot{\boldsymbol{e}}_2 = 0 = \mathbf{e}_3 \cdot \dot{\boldsymbol{e}}_3$, but $\dot{\boldsymbol{e}}_2 \cdot \mathbf{e}_3 + \mathbf{e}_2 \cdot \dot{\boldsymbol{e}}_3 = 0$ requires that $\alpha = -\beta \triangleq \omega_1$. In summary

$$\begin{aligned} \dot{\boldsymbol{e}}_2 &= -\omega_3\mathbf{e}_1 + \omega_1\mathbf{e}_3 = \boldsymbol{\omega} \times \mathbf{e}_2, \\ \dot{\boldsymbol{e}}_3 &= \omega_2\mathbf{e}_1 - \omega_1\mathbf{e}_2 = \boldsymbol{\omega} \times \mathbf{e}_3, \end{aligned} \tag{42}$$

and all three basis vectors have the same angular rotation vector $\boldsymbol{\omega}(t)$.

This derivation can be performed more quickly using index notation. Indeed if we define $\dot{\boldsymbol{e}}_i = \Omega_{ij}\mathbf{e}_j$ then (39) implies that $\Omega_{ij} = -\Omega_{ji}$. As seen in earlier exercise, the antisymmetric tensor Ω_{ij} thus defines a vector $\boldsymbol{\omega} = \frac{1}{2}\epsilon_{kij}\Omega_{ij}\mathbf{e}_k$ such that $\dot{\boldsymbol{e}}_i = \Omega_{ij}\mathbf{e}_j = \boldsymbol{\omega} \times \mathbf{e}_i$, and all three basis vectors have the same rotation vector $\boldsymbol{\omega}$.

The rotation vector $\boldsymbol{\omega}(t)$ gives the rate of change of *any* vector tied to the body. Indeed, if A and B are any two points of the body then the vector $\boldsymbol{c} \equiv \overrightarrow{AB}$ can be expanded with respect to the body basis $\mathbf{e}_1(t)$, $\mathbf{e}_2(t)$, $\mathbf{e}_3(t)$

$$\boldsymbol{c}(t) = c_1\mathbf{e}_1(t) + c_2\mathbf{e}_2(t) + c_3\mathbf{e}_3(t),$$

but the components $c_i = \boldsymbol{c}(t) \cdot \mathbf{e}_i(t)$ are constants because all lengths and angles, and therefore all dot products, are time-invariant. Thus

$$\frac{d\boldsymbol{c}}{dt} = \sum_{i=1}^{3} c_i \frac{d\mathbf{e}_i}{dt} = \sum_{i=1}^{3} c_i\,(\boldsymbol{\omega} \times \mathbf{e}_i) = \boldsymbol{\omega} \times \boldsymbol{c}.$$

This is true for any vector tied to the body (*material vectors*), implying that the rotation vector is unique for the body.

Dynamics of a rigid body

The center of mass of a rigid body moves according to the sum of the external forces as for a system of particles. A continuous rigid body can be considered as a continuous distribution of 'infinitesimal' masses dm

$$\sum_{i=1}^{N} m_i \boldsymbol{s}_i \longrightarrow \int_V \boldsymbol{s}\, dm(\boldsymbol{s})$$

where the three-dimensional integral is over all points $\boldsymbol{s}$ in the domain V of the body ($dm(\boldsymbol{s})$ is the mass 'measure' of the infinitesimal volume element dV at point $\boldsymbol{s}$, or in other words $dm = \rho dV$, where $\rho(\boldsymbol{s})$ is the mass density at point $\boldsymbol{s}$).

For the motion about the center of mass, the position vectors $\boldsymbol{s}_i$ are frozen into the body hence $\dot{\boldsymbol{s}}_i = \boldsymbol{\omega} \times \boldsymbol{s}_i$ for any point of the body. The total angular momentum for a rigid system of particles then reads

$$\boldsymbol{L} = \sum_i m_i \boldsymbol{s}_i \times \dot{\boldsymbol{s}}_i = \sum_i m_i \boldsymbol{s}_i \times (\boldsymbol{\omega} \times \boldsymbol{s}_i) = \sum_i m_i \left(|\boldsymbol{s}_i|^2 \boldsymbol{\omega} - \boldsymbol{s}_i \left(\boldsymbol{s}_i \cdot \boldsymbol{\omega} \right) \right). \quad (43)$$

and for a continuous rigid body

$$\boldsymbol{L} = \int_V \left(|\boldsymbol{s}|^2 \boldsymbol{\omega} - \boldsymbol{s} \left(\boldsymbol{s} \cdot \boldsymbol{\omega} \right) \right) \, dm. \quad (44)$$

The Poisson vector is unique for the body, so it does not depend on $\boldsymbol{s}$ and we should be able to take it out of the sum, or integral. That's easy for the $\|\boldsymbol{s}\|^2 \boldsymbol{\omega}$ term, but how can we get $\boldsymbol{\omega}$ out of the $\int \boldsymbol{s} \left(\boldsymbol{s} \cdot \boldsymbol{\omega} \right) dm$ term?! We need to introduce the concepts of *tensor product* and *tensors* to do this, but we can give a hint by switching to index notation $\boldsymbol{L} \to L_i$, $\boldsymbol{s} \to s_i$, $|\boldsymbol{s}| = s$, $\boldsymbol{\omega} \to \omega_i$, with $i = 1, 2, 3$ (this i is now a cartesian coordinate counter, not the particle counter $i = 1, \ldots, N$ used above) and writing

$$L_i = \int_V \left(s^2 \omega_i - s_i \left(s_j \omega_j \right) \right) \, dm = \left(\int_V \left(s^2 \delta_{ij} - s_i s_j \right) \, dm \right) \omega_j \triangleq \mathcal{J}_{ij} \, \omega_j \quad (45)$$

where $\mathcal{J}$ is the *tensor of inertia* of the rigid body, independent of the rotation vector $\boldsymbol{\omega}$. In vector notation, this is

$$\boldsymbol{L} = \mathcal{J} \cdot \boldsymbol{\omega}. \quad (46)$$

The tensor $\mathcal{J}$ is symmetric positive definite, thus it has real positive eigenvalues and orthogonal eigenvectors. Let I_1, I_2, I_3 be the eigenvalues with $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ the corresponding eigenvectors. Then

$$\mathcal{J} = I_1 \mathbf{e}_1 \mathbf{e}_1 + I_2 \mathbf{e}_2 \mathbf{e}_2 + I_3 \mathbf{e}_3 \mathbf{e}_3. \quad (47)$$

If no torque acts on the body, then $\dot{\boldsymbol{L}} = 0$, that is

$$\frac{d}{dt} \left(\mathcal{J} \cdot \boldsymbol{\omega} \right) = \frac{d}{dt} \left(I_1 \omega_1 \mathbf{e}_1 + I_2 \omega_2 \mathbf{e}_2 + I_3 \omega_3 \mathbf{e}_3 \right) = 0, \quad (48)$$

where $\boldsymbol{\omega} = \omega_1 \mathbf{e}_1 + \omega_2 \mathbf{e}_2 + \omega_3 \mathbf{e}_3$ is the rotation vector expressed in the body frame. As a rigid body moves, the *principal moments of inertia* I_1, I_2, I_3 are constants since they describe the distribution of mass about axes fixed in the body. The *principal axes of inertia* $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are also fixed in the body but this means that they are functions of time t, with respect to an inertial frame of reference. The rotation vector $\boldsymbol{\omega}$ of the rigid body is also a function of time, in general, and its rate of change arises from the changes of the components $(\omega_1, \omega_2, \omega_3)$ in the body frame, as well as the rotation of the body frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. The rates of change of the latter are simply $\dot{\mathbf{e}}_i = \boldsymbol{\omega} \times \mathbf{e}_i$. Equation (48) thus expands to

$$I_1(\dot{\omega}_1\mathbf{e}_1 + \omega_1\dot{\mathbf{e}}_1) + I_2(\dot{\omega}_2\mathbf{e}_2 + \omega_2\dot{\mathbf{e}}_2) + I_3(\dot{\omega}_3\mathbf{e}_3 + \omega_3\dot{\mathbf{e}}_3) = 0.$$

Projecting onto $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ using $\dot{\mathbf{e}}_i \cdot \mathbf{e}_j = (\boldsymbol{\omega} \times \mathbf{e}_i) \cdot \mathbf{e}_j = \epsilon_{ijk}\omega_k$, yields

$$\begin{cases} I_1\dot{\omega}_1 = (I_2 - I_3)\,\omega_2\omega_3, \\ I_2\dot{\omega}_2 = (I_3 - I_1)\,\omega_3\omega_1, \\ I_3\dot{\omega}_3 = (I_1 - I_2)\,\omega_1\omega_2. \end{cases} \tag{49}$$

These are *Euler's equations* for the torque-less motion of a rigid body. The dynamics conserves energy, $E = \frac{1}{2}\boldsymbol{\omega} \cdot \mathcal{J} \cdot \boldsymbol{\omega}$,

$$\frac{1}{2}\frac{d}{dt}\left(I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2\right) = 0 \tag{50}$$

and angular momentum, $\frac{1}{2}\boldsymbol{L} \cdot \boldsymbol{L} = \frac{1}{2}\left(\boldsymbol{\omega} \cdot \mathcal{J}\right) \cdot \left(\mathcal{J} \cdot \boldsymbol{\omega}\right)$,

$$\frac{1}{2}\frac{d}{dt}\left(I_1^2\omega_1^2 + I_2^2\omega_2^2 + I_3^2\omega_3^2\right) = 0. \tag{51}$$

Euler's equations show that rotation about any principal axis is a fixed point, that is $(\omega_1, \omega_2, \omega_3) = (\Omega, 0, 0)$ or $(0, \Omega, 0)$ or $(0, 0, \Omega)$ are steady solutions of (49), for any constant Ω. However rotation about the axis corresponding to the middle moment of inertia is unstable. Indeed, if we label the principal moments of inertia such that $I_1 \le I_2 \le I_3$, then linearization of (49) about $(0, \Omega, 0)$ yields

$$\begin{cases} I_1\dot{\omega}_1 = (I_2 - I_3)\,\Omega\omega_3, \\ I_3\dot{\omega}_3 = (I_1 - I_2)\,\omega_1\Omega, \end{cases} \tag{52}$$

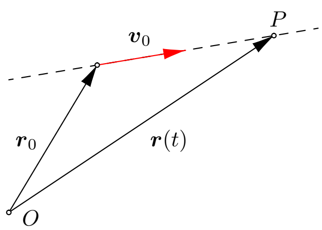
that reduces to the 2nd order ODE

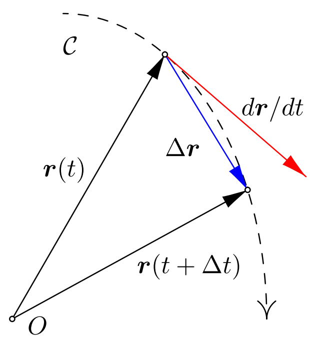
$$I_1I_3\ddot{\omega}_1 = (I_1 - I_2)\,(I_2 - I_3)\,\Omega^2\omega_1 \tag{53}$$

that has exponentially growing solutions whenever $(I_1 - I_2)\,(I_2 - I_3) > 0$ which is the case if I_2 is the middle principal moment. The linearization (52) and the exponential growth of ω_1 and ω_3 are only valid as long as $(I_3 - I_1)\int_0^t \omega_1(t')\omega_3(t')dt'$ is small compared to $I_2\Omega$. Otherwise, ω_2 cannot be assumed to be fixed at Ω and the full nonlinear system (49) needs to be considered.

3 Curves

3.1 Elementary curves





Recall the parametric equation of a line: $\boldsymbol{r}(t) = \boldsymbol{r}_0 + t\boldsymbol{v}_0$, where $\boldsymbol{r}(t) = \overrightarrow{OP}$ is the position vector of a point P on the line with respect to some 'origin' O, $\boldsymbol{r}_0$ is the position vector of a reference point on the line and $\boldsymbol{v}_0$ is a vector parallel to the line. Note that this can be interpreted as the linear motion of a particle with constant velocity $\boldsymbol{v}_0$ that was at the point $\boldsymbol{r}_0$ at time $t=0$ and $\boldsymbol{r}(t)$ is the position at time t.

More generally, a vector function $\boldsymbol{r}(t)$ of a real variable t defines a curve $\mathcal{C}$. The vector function $\boldsymbol{r}(t)$ is the *parametric representation* of that curve and t is the parameter. It is useful to think of t as time and $\boldsymbol{r}(t)$ as the position of a particle at time t. The collection of all the positions for a range of t is the particle trajectory. The vector $\Delta\boldsymbol{r} = \boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t)$ is a *secant vector* connecting two points on the curve, if we divide $\Delta\boldsymbol{r}$ by Δt and take the limit as $\Delta t \to 0$ we obtain the vector $d\boldsymbol{r}/dt$ which is *tangent* to the curve at $\boldsymbol{r}(t)$. If t is time, then $d\boldsymbol{r}/dt = \boldsymbol{v}$ is the velocity.

Circle. The parameter can have any name and does not need to correspond to time. For instance the circle of radius a centered at O can be parameterized by

$$\boldsymbol{r}(\theta) = a\cos\theta\,\hat{\boldsymbol{x}} + a\sin\theta\,\hat{\boldsymbol{y}}, \tag{54}$$

where θ is a real parameter that is in fact the angle between $\boldsymbol{r}$ and $\hat{\boldsymbol{x}}$.

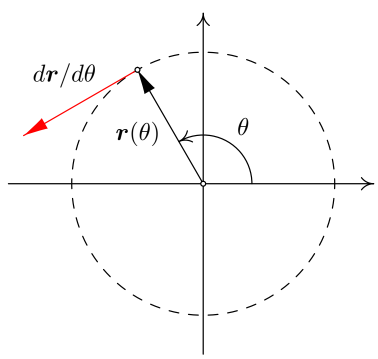


Fig. 2.9: A circle of radius a centered at O is $\boldsymbol{r}(\theta) = a\cos\theta\,\hat{\boldsymbol{x}} + a\sin\theta\,\hat{\boldsymbol{y}}$ and $d\boldsymbol{r}/d\theta = -a\sin\theta\,\hat{\boldsymbol{x}} + a\cos\theta\,\hat{\boldsymbol{y}}$ is the tangent vector to the circle at $\boldsymbol{r}(\theta)$, with $\boldsymbol{r}\cdot d\boldsymbol{r}/d\theta = 0$ for any θ.

Ellipse. The circle parameterization is easily extended to an arbitrary ellipse centered at C with major radius vector $\boldsymbol{a}$ and minor radius vector $\boldsymbol{b} \perp \boldsymbol{a}$

$$\boldsymbol{r}(\theta) = \boldsymbol{r}_C + \boldsymbol{a}\,\cos\theta + \boldsymbol{b}\,\sin\theta. \tag{55}$$

The parameter θ is an angle, but it is *not* the angle between $\boldsymbol{r} - \boldsymbol{r}_C$ and $\boldsymbol{a}$, unless $|\boldsymbol{a}| = |\boldsymbol{b}|$. If we pick cartesian coordinates with C as the origin and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{a}}$, $\hat{\boldsymbol{y}} = \hat{\boldsymbol{b}}$, then $\boldsymbol{r}_C = 0$, $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}$ and (55) becomes

$$\begin{cases} x = a\cos\theta \\ y = b\sin\theta \end{cases} \quad \Leftrightarrow \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{56}$$

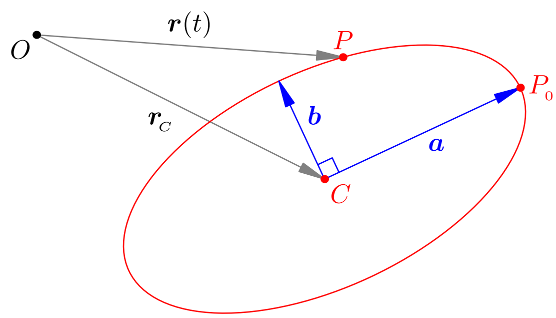


Fig. 2.10: An arbitrary ellipse: $\boldsymbol{r}(t) = \boldsymbol{r}_C + \boldsymbol{a}\cos t + \boldsymbol{b}\sin t$, centered at C with major radius $\boldsymbol{a}$ and minor radius $\boldsymbol{b}$ if $\boldsymbol{a}\cdot\boldsymbol{b} = 0$.

which is the standard implicit equation of an ellipse. However, the vector equation (55) is coordinate independent and if $\boldsymbol{r}_C = x_c\hat{\boldsymbol{x}} + y_c\hat{\boldsymbol{y}}$ and $\boldsymbol{a} = a\cos\alpha\,\hat{\boldsymbol{x}} + a\sin\alpha\,\hat{\boldsymbol{y}}$, thus $\boldsymbol{b} = -b\sin\alpha\,\hat{\boldsymbol{x}} + b\cos\alpha\,\hat{\boldsymbol{y}}$ then (55) becomes

$$\begin{cases} x = x_c + a\cos\alpha\cos\theta - b\sin\alpha\sin\theta, \\ y = y_c + a\sin\alpha\cos\theta + b\cos\alpha\sin\theta, \end{cases} \tag{57}$$

that can be written in matrix form as

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} a\cos\theta \\ b\sin\theta \end{bmatrix} = \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix}. \tag{58}$$

We recognize the matrix as a rotation matrix whose inverse is its transpose. Solving for $(\cos\theta, \sin\theta)$ then eliminating θ yields the general implicit equation for an ellipse in the (x, y) plane

$$(x - x_c)^2\left(\frac{\cos^2\alpha}{a^2} + \frac{\sin^2\alpha}{b^2}\right) + (y - y_c)^2\left(\frac{\sin^2\alpha}{a^2} + \frac{\cos^2\alpha}{b^2}\right) \\ + (x - x_c)(y - y_c)\sin 2\alpha\left(\frac{1}{a^2} - \frac{1}{b^2}\right) = 1. \tag{59}$$

Actually, the vector equation (55) is even more general since it applies to ellipses in 3D (and even in $\mathbb{R}^n$), and $\boldsymbol{a}$ and $\boldsymbol{b}$ do not have to be orthogonal for the curve to be an ellipse (exercise 8).

Hyperbola. Likewise,

$$\boldsymbol{r}(t) = \boldsymbol{r}_C + \boldsymbol{a}\,\cosh t + \boldsymbol{b}\,\sinh t \tag{60}$$

is a hyperbolic branch in the $C, \boldsymbol{a}, \boldsymbol{b}$ plane, where the hyperbolic cosine and sine

$$\cosh t \triangleq \frac{e^t + e^{-t}}{2}, \qquad \sinh t \triangleq \frac{e^t - e^{-t}}{2}$$

satisfy

$$\cosh^2 t - \sinh^2 t = 1$$

for any real t.

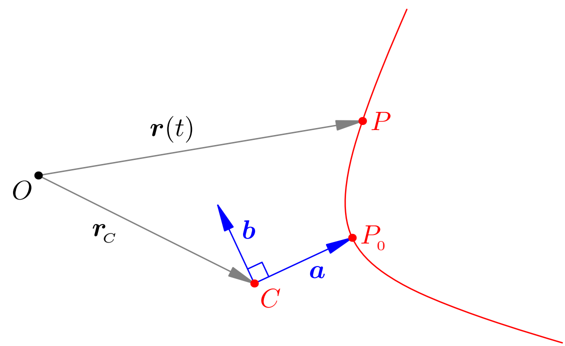


Fig. 2.11: $\boldsymbol{r}(t) = \boldsymbol{r}_C + \boldsymbol{a}\cosh t + \boldsymbol{b}\sinh t$, hyperbolic branch centered at C.

3.2 Speeding through Curves

Consider a point P (or a Particle, or a Plane, or a Planet) at position or radius vector $\boldsymbol{r} = \overrightarrow{OP}$ at time t, thus $\boldsymbol{r} = \boldsymbol{r}(t)$. In cartesian coordinates

$$\boldsymbol{r}(t) = x(t)\hat{\boldsymbol{x}} + y(t)\hat{\boldsymbol{y}} + z(t)\hat{\boldsymbol{z}} \tag{61}$$

with fixed basis vector $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$, while in spherical coordinates

$$\boldsymbol{r}(t) = r(t)\hat{\boldsymbol{r}}(t) \tag{62}$$

where the magnitude $r(t)$ and the direction $\hat{\boldsymbol{r}}(t)$ are functions of t in general. The radial direction $\hat{\boldsymbol{r}}$ is a function of the azimuthal angle φ and the polar angle θ, $\hat{\boldsymbol{r}}(\varphi, \theta)$ as seen in Chapter 1, section 5. Thus the time dependence of $\hat{\boldsymbol{r}}$ would specified by $\hat{\boldsymbol{r}}(\varphi(t), \theta(t))$. Whatever those functions, we know well that $\hat{\boldsymbol{r}} \cdot d\hat{\boldsymbol{r}}/dt = 0$.

The velocity $\boldsymbol{v} = d\boldsymbol{r}/dt$ contains geometric information about the *curve* $\mathcal{C}$ traced by $P(t)$. Imagine for instance a car on highway 90, or a car on a rollercoaster track. The position of the car defines the curve (road, track) but also contains speed information about how fast the car is going along that track. The concept of arclength – the distance along the curve – allows us to separate speed information from trajectory information.

Arclength s is specified in differential form as

$$ds = |d\boldsymbol{r}| = \sqrt{d\boldsymbol{r} \cdot d\boldsymbol{r}} = \sqrt{dx^2 + dy^2 + dz^2}, \tag{63}$$

it should not be confused with the differential of distance to the origin r which is

$$dr = d|\boldsymbol{r}| = d(\sqrt{x^2 + y^2 + z^2}) = \frac{x dx + y dy + z dz}{\sqrt{x^2 + y^2 + z^2}} = \hat{\boldsymbol{r}} \cdot d\boldsymbol{r} \tag{64}$$

The arclength s can be defined in terms of the speed $v(t) = |\boldsymbol{v}(t)|$ by the scalar differential equation

$$\frac{ds}{dt} = v = \left|\frac{d\boldsymbol{r}}{dt}\right| \geq 0 \tag{65}$$

the latter definition picks the direction of increasing s as the direction of travel as t increases. Thus arclength s is a monotonically increasing function of t and there is a one-to-one correspondence between s and t (but watch out if the particle stops and backtracks.[2]).

Velocity $\boldsymbol{v}$ is

$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} = \frac{d\boldsymbol{r}}{ds}\frac{ds}{dt} = v\,\hat{\boldsymbol{t}} \tag{66}$$

where

$$\hat{\boldsymbol{t}} = \frac{d\boldsymbol{r}}{ds} = \frac{d\boldsymbol{r}/dt}{|d\boldsymbol{r}/dt|} = \frac{\boldsymbol{v}}{v} \tag{67}$$

is the *unit tangent vector* to the curve at point $\boldsymbol{r}$. Here we think of $\boldsymbol{r}$ as $\boldsymbol{r}(s(t))$, that is, position $\boldsymbol{r}$ is given as a function of distance along a known track $\boldsymbol{r}(s)$ (for instance, 42 miles from Madison westbound on highway 90 specifies a point in 3D space) and $s(t)$ as a function of time t obtained by integrating $ds/dt = v(t)$ with respect to time (for instance, given the speed $v(t)$ of a car traveling westbound on highway 90). In general, we write $\boldsymbol{v} = v\,\hat{\boldsymbol{v}}$ for a velocity vector $\boldsymbol{v}$ of magnitude v and direction $\hat{\boldsymbol{v}}$, however here we are separating geometric track information $\hat{\boldsymbol{t}}$ from speed v along that track and the velocity direction $\hat{\boldsymbol{v}}$ is the same as the track direction, $\hat{\boldsymbol{v}} = \hat{\boldsymbol{t}}$, with the direction of $\hat{\boldsymbol{t}}$ given as the direction of increasing time. So $\hat{\boldsymbol{t}}$ is the *unit tangent* to the curve at $\boldsymbol{r}$ and also the direction of motion as t increases, $\hat{\boldsymbol{v}}$.

Acceleration $\boldsymbol{a}$ is

$$\begin{aligned}\boldsymbol{a} = \frac{d^2\boldsymbol{r}}{dt^2} = \frac{d\boldsymbol{v}}{dt} &= \frac{d\boldsymbol{v}}{ds}\frac{ds}{dt} = \frac{d(v\,\hat{\boldsymbol{t}})}{ds}v = v\frac{dv}{ds}\hat{\boldsymbol{t}} + v^2\frac{d\hat{\boldsymbol{t}}}{ds}\\ &= \frac{d}{ds}\left(\frac{1}{2}v^2\right)\hat{\boldsymbol{t}} + \kappa v^2\,\hat{\boldsymbol{n}}\end{aligned} \tag{68}$$

where

$$\frac{d\hat{\boldsymbol{t}}}{ds} \triangleq \kappa\,\hat{\boldsymbol{n}} = \frac{1}{R}\,\hat{\boldsymbol{n}} \tag{69}$$

is the rate of change of the curve direction $\hat{\boldsymbol{t}}$ with respect to distance along the curve s. Since $\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} = 1$, we have $\hat{\boldsymbol{t}} \cdot d\hat{\boldsymbol{t}}/ds = 0$ and $\hat{\boldsymbol{n}}$ is perpendicular to $\hat{\boldsymbol{t}}$. Indeed $\hat{\boldsymbol{n}}$ points in the direction of the turn and is a unit vector that is *normal* (that is orthogonal or perpendicular) to the curve at $\boldsymbol{r}$. The *curvature*

$$\kappa = \left|\frac{d\hat{\boldsymbol{t}}}{ds}\right| = \frac{1}{R} \tag{70}$$

has units of inverse length and can thus be written as $1/R$ where $R = R(s)$ is the local *radius of curvature*. Thus (68) yields a decomposition of the acceleration $\boldsymbol{a}$ in terms of a component in the curve direction $\hat{\boldsymbol{t}}$ and a component in the turn direction $\hat{\boldsymbol{n}}$. The $\hat{\boldsymbol{n}}$ component $\kappa v^2 = v^2/R$ is the *centripetal acceleration* of the particle. This

[2]That is dangerous and illegal when traveling westbound on Highway 90, but that might happen on a roller coaster where the point is to jerk you around!

result is completely general, holding for any curve, not just for circles, and in general $\kappa = \kappa(s)$, $v = v(s)$ vary along the curve.

Jerk $\boldsymbol{j}$ is

$$\begin{aligned}
\boldsymbol{j} = \frac{d^3\boldsymbol{r}}{dt^3} = \frac{d\boldsymbol{a}}{dt} &= \frac{d\boldsymbol{a}}{ds}\frac{ds}{dt} \\
&= v\frac{d^2}{ds^2}\left(\frac{1}{2}v^2\right)\hat{\boldsymbol{t}} + v^2\frac{dv}{ds}\frac{d\hat{\boldsymbol{t}}}{ds} + v\frac{d}{ds}(\kappa v^2)\hat{\boldsymbol{n}} + \kappa v^3\frac{d\hat{\boldsymbol{n}}}{ds} \\
&= v\frac{d^2}{ds^2}\left(\frac{1}{2}v^2\right)\hat{\boldsymbol{t}} + \frac{d}{ds}\left(\kappa v^3\right)\hat{\boldsymbol{n}} + \kappa v^3\frac{d\hat{\boldsymbol{n}}}{ds} \\
&= \left(v\frac{d^2}{ds^2}\left(\frac{1}{2}v^2\right) - \kappa^2 v^3\right)\hat{\boldsymbol{t}} + \frac{d}{ds}\left(\kappa v^3\right)\hat{\boldsymbol{n}} + \kappa\tau v^3\,\hat{\boldsymbol{b}}
\end{aligned} \tag{71}$$

where

$$\frac{d\hat{\boldsymbol{n}}}{ds} = \tau\,\hat{\boldsymbol{b}} - \kappa\,\hat{\boldsymbol{t}} \tag{72}$$

in terms of the *binormal*

$$\hat{\boldsymbol{b}} \triangleq \hat{\boldsymbol{t}} \times \hat{\boldsymbol{n}} \tag{73}$$

and the *torsion*

$$\tau(s) \triangleq \hat{\boldsymbol{b}} \cdot \frac{d\hat{\boldsymbol{n}}}{ds} = -\hat{\boldsymbol{n}} \cdot \frac{d\hat{\boldsymbol{b}}}{ds} \tag{74}$$

since $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{b}} = 0$ and $\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{n}} = 0$ with $d\hat{\boldsymbol{t}}/ds \triangleq \kappa\,\hat{\boldsymbol{n}}$. The binormal $\hat{\boldsymbol{b}}$ will be constant and the torsion τ will be 0 *for a planar curve.*

The cross product of (66) with (68)

$$\boldsymbol{v} \times \boldsymbol{a} = \dot{\boldsymbol{r}} \times \ddot{\boldsymbol{r}} = \kappa\, v^3\,\hat{\boldsymbol{b}} \tag{75}$$

provides another way to obtain the curvature κ and binormal $\hat{\boldsymbol{b}}$ directly from the velocity and acceleration without passing through arclength, where $\boldsymbol{v} = \dot{\boldsymbol{r}} = d\boldsymbol{r}/dt$, $\boldsymbol{a} = \ddot{\boldsymbol{r}} = d^2\boldsymbol{r}/dt^2$ and $v = |\dot{\boldsymbol{r}}|$. The dot product of (75) with (71),

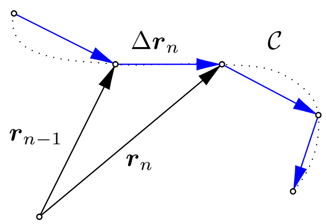
$$(\boldsymbol{v} \times \boldsymbol{a}) \cdot \boldsymbol{j} = (\dot{\boldsymbol{r}} \times \ddot{\boldsymbol{r}}) \cdot \dddot{\boldsymbol{r}} = (\kappa v^3)^2\,\tau \tag{76}$$

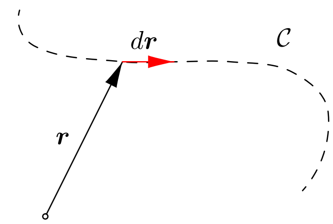
yields an expression for the torsion τ once v and κ are known.

3.3 Integrals over curves

Line element: Given a curve $\mathcal{C}$, the *line element* denoted $d\boldsymbol{r}$ is an ‘infinitesimal’ secant vector. This is a useful shortcut for the procedure of approximating the curve by a succession of secant vectors $\Delta\boldsymbol{r}_n = \boldsymbol{r}_n - \boldsymbol{r}_{n-1}$ where $\boldsymbol{r}_{n-1}$ and $\boldsymbol{r}_n$ are two consecutive points on the curve, with $n = 1, 2, \ldots, N$ integer, then taking the limit $\max|\Delta\boldsymbol{r}_n| \to 0$ (so $N \to \infty$). In that limit, the direction of the secant vector $\Delta\boldsymbol{r}_n$ becomes identical with that of the tangent vector at that point. If an explicit parametric representation $\boldsymbol{r}(t)$ is known for the curve then

$$d\boldsymbol{r} = \frac{d\boldsymbol{r}(t)}{dt}\,dt \tag{77}$$





A *line integral* is an integral over a curve $\mathcal{C}$. The basic *line integral* along a curve $\mathcal{C}$ is $\int_{\mathcal{C}} \boldsymbol{F} \cdot d\boldsymbol{r}$ where $\boldsymbol{F}(\boldsymbol{r})$ is a vector *field*, that is a vector function of position $\boldsymbol{r}$. If $\boldsymbol{F}(\boldsymbol{r})$ is a force, this integral represent the net work done by the force on a particle as the latter moves along the curve. We can make sense of this integral as the *limit of a sum*, namely partitioning the curve into a chain of N secant vectors $\Delta \boldsymbol{r}_n$ as above, then

$$\boxed{\int_{\mathcal{C}} \boldsymbol{F} \cdot d\boldsymbol{r} = \lim_{\max |\Delta \boldsymbol{r}_n| \to 0} \sum_{n=1}^{N} \boldsymbol{F}_n \cdot \Delta \boldsymbol{r}_n} \tag{78}$$

where $\boldsymbol{F}_n$ is an estimate of the average value of $\boldsymbol{F}$ along the line segment $\boldsymbol{r}_{n-1} \to \boldsymbol{r}_n$. A simple choice is $\boldsymbol{F}_n = \boldsymbol{F}(\boldsymbol{r}_{n-1})$ or $\boldsymbol{F}(\boldsymbol{r}_n)$ but better choices are the *trapezoidal rule*

$$\boldsymbol{F}_n \triangleq \frac{1}{2}\big(\boldsymbol{F}(\boldsymbol{r}_{n-1}) + \boldsymbol{F}(\boldsymbol{r}_n)\big),$$

or the *midpoint rule*

$$\boldsymbol{F}_n \triangleq \boldsymbol{F}\left(\frac{\boldsymbol{r}_n + \boldsymbol{r}_{n-1}}{2}\right).$$

These different choices for $\boldsymbol{F}_n$ give finite sums that converge to the same limit, $\int_{\mathcal{C}} \boldsymbol{F} \cdot d\boldsymbol{r}$, but the trapezoidal and midpoint rules will converge faster for smooth functions, and give more accurate finite sum approximations. If an explicit representation $\boldsymbol{r}(t)$ is known then we can reduce the line integral to a regular integral over t:

$$\int_{\mathcal{C}} \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{t_a}^{t_b} \left(\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \frac{d\boldsymbol{r}(t)}{dt}\right) dt, \tag{79}$$

where $\boldsymbol{r}(t_a)$ is the starting point of curve $\mathcal{C}$ and $\boldsymbol{r}(t_b)$ is its end point. These may be the same point even if $t_a \neq t_b$ (for example, an integral once around a circle from $\theta = 0$ to $\theta = 2\pi$).

Example. For the near simplest case of $\boldsymbol{F}(\boldsymbol{r}) = \boldsymbol{r}$ and the straight line from point A to point B

$$\mathcal{C} : \boldsymbol{r}(t) = \boldsymbol{r}_{A} + t\,(\boldsymbol{r}_{B} - \boldsymbol{r}_{A})$$

with $t = 0 \to 1$ then

$$\int_{\mathcal{C}} \boldsymbol{r} \cdot d\boldsymbol{r} = \int_0^1 \Big(\boldsymbol{r}_A + t\,(\boldsymbol{r}_B - \boldsymbol{r}_A)\Big) \cdot (\boldsymbol{r}_B - \boldsymbol{r}_A)\, dt$$
$$= \frac{1}{2}\,(\boldsymbol{r}_B \cdot \boldsymbol{r}_B - \boldsymbol{r}_A \cdot \boldsymbol{r}_A)\,. \qquad (80)$$

Easy enough, but this is a very special $\boldsymbol{F}(\boldsymbol{r}) = \boldsymbol{r}$, so special that this line integral is actually *independent of the path* from A to B. Indeed for *any* curve $\mathcal{C}$ from A to B

$$\int_{\mathcal{C}} \boldsymbol{r} \cdot d\boldsymbol{r} = \frac{1}{2}\int_{\mathcal{C}} d(\boldsymbol{r}\cdot\boldsymbol{r}) = \frac{1}{2}\,(\boldsymbol{r}_B \cdot \boldsymbol{r}_B - \boldsymbol{r}_A \cdot \boldsymbol{r}_A) \qquad (81)$$

using differentials, or with whatever $\boldsymbol{r}(t)$ specifies the curve:

$$\int_{\mathcal{C}} \boldsymbol{r} \cdot d\boldsymbol{r} = \int_{t_A}^{t_B} \boldsymbol{r}(t) \cdot \frac{d\boldsymbol{r}(t)}{dt}\, dt = \frac{1}{2}\int_{t_A}^{t_B} \frac{d(\boldsymbol{r}\cdot\boldsymbol{r})}{dt} dt = \frac{1}{2}\,(\boldsymbol{r}_B \cdot \boldsymbol{r}_B - \boldsymbol{r}_A \cdot \boldsymbol{r}_A)\,. \qquad (82)$$

Either way, we recover (80) *whatever the curve from A to B*. This is only true for very special vector fields $\boldsymbol{F}(\boldsymbol{r})$ called *conservative vector fields* (or conservative *force* if $\boldsymbol{F}$ is a force). The vector field $\boldsymbol{v} = \boldsymbol{\Omega} \times \boldsymbol{r}$ with $\boldsymbol{\Omega}$ constant is a simple example of a non-conservative vector field discussed later below. □

There are many other types of line integrals such as

$$\int_{\mathcal{C}} f(\boldsymbol{r})\,|d\boldsymbol{r}|, \quad \int_{\mathcal{C}} \boldsymbol{F}\,|d\boldsymbol{r}|, \quad \int_{\mathcal{C}} f(\boldsymbol{r})\,d\boldsymbol{r}, \quad \int_{\mathcal{C}} \boldsymbol{F} \times d\boldsymbol{r}.$$

The first one gives a scalar result and the latter three give *vector* results. Recall from the previous section that $ds = |d\boldsymbol{r}|$ is differential arclength, not to be confused with $dr = d|\boldsymbol{r}|$, the differential of distance to the origin. We can make sense of these integrals from the *limit of a sum* definition.

Example: The length of curve $\mathcal{C}$ from point A to point B is determined by the integral

$$\int_{\mathcal{C}} ds = \int_{\mathcal{C}} |d\boldsymbol{r}| = \lim_{|\Delta\boldsymbol{r}_n|\to 0} \sum_{n=1}^{N} |\Delta\boldsymbol{r}_n| \qquad (83)$$

with $\boldsymbol{r}_0 = \boldsymbol{r}_A$ and $\boldsymbol{r}_N = \boldsymbol{r}_B$. If a parametrization $\boldsymbol{r} = \boldsymbol{r}(t)$ is known then

$$\int_{\mathcal{C}} |d\boldsymbol{r}| = \int_{t_A}^{t_B} \left|\frac{d\boldsymbol{r}(t)}{dt}\right| |dt| \qquad (84)$$

where $\boldsymbol{r}_A = \boldsymbol{r}(t_A)$ and $\boldsymbol{r}_B = \boldsymbol{r}(t_B)$. That's *almost* a Calc I integral, except for that $|dt|$, what does that mean?! Again you can understand that from the limit-of-a-sum definition with $t_0 = t_A$, $t_N = t_B$ and $\Delta t_n = t_n - t_{n-1}$. If $t_A < t_B$ then $\Delta t_n > 0$ and $dt > 0$, so $|dt| = dt$ and we're blissfully happy. But if $t_B < t_A$ then $\Delta t_n < 0$ and $dt < 0$, so $|dt| = -dt$ and

$$\int_{t_A}^{t_B} (\cdots)|dt| = \int_{t_B}^{t_A} (\cdots)dt, \qquad \text{if} \quad t_A > t_B. \qquad (85)$$

□

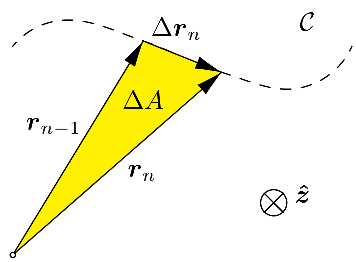
Example: A special example of a $\int_{\mathcal{C}} \boldsymbol{F} \times d\boldsymbol{r}$ integral is

$$\int_{\mathcal{C}} \boldsymbol{r} \times d\boldsymbol{r} = \lim_{|\Delta \boldsymbol{r}_n| \to 0} \sum_{n=1}^{N} \boldsymbol{r}_n \times \Delta \boldsymbol{r}_n = \int_{t_A}^{t_B} \left(\boldsymbol{r}(t) \times \frac{d\boldsymbol{r}(t)}{dt} \right) dt \qquad (86)$$

This integral yields a *vector* $2A\hat{\boldsymbol{z}}$ whose magnitude is *twice* the area A swept by the position vector $\boldsymbol{r}(t)$ when the curve $\mathcal{C}$ lies in a plane perpendicular to $\hat{\boldsymbol{z}}$ and O is in that plane (recall Kepler's law that *'the radius vector sweeps equal areas in equal times'*). This follows from the fact that

$$\frac{1}{2} \boldsymbol{r}_n \times \Delta \boldsymbol{r}_n = \Delta A\, \hat{\boldsymbol{z}}$$

is the area ΔA of the triangle with sides $\boldsymbol{r}_n$ and $\Delta \boldsymbol{r}_n$.



If $\mathcal{C}$ and O are not coplanar then the vectors $\boldsymbol{r}_n \times \Delta \boldsymbol{r}_n$ are not necessarily in the same direction and their *vector* sum is *not* the area swept by $\boldsymbol{r}$. In that more general case, the surface is conical and to calculate its area S we would need to calculate $S = \frac{1}{2} \int_{\mathcal{C}} |\boldsymbol{r} \times d\boldsymbol{r}|$. □

Exercises:

1. What is the geometric interpretation for the angle θ in (55)? Is it the angle between $\boldsymbol{r} - \boldsymbol{r}_c$ and $\boldsymbol{a}$? [Hint: consider the circles of radius a and b]

2. Show that the points $P \equiv (x, y)$ on the ellipse $x^2/a^2 + y^2/b^2 = 1$ are such that the sum of their distances to the *foci* $F_1 \equiv (-c, 0)$ and $F_2 \equiv (c, 0)$ is equal to $2a$, where $c = \sqrt{a^2 - b^2}$, for $a \geq b$. This is the geometric definition of an ellipse.

3. Given two arbitrary fixed points F_1 and F_2 and a constant $2a \geq |F_1F_2|$, consider the curve traced by the point $P(t)$ such that the sum of the distances $|F_1P| + |F_2P| = 2a$ in an arbitrary plane containing F_1 and F_2. Show that the tangent to that curve makes equal angles with the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$. This is a fundamental geometric property of the ellipse. Make a nice sketch to show your comprehension of the problem and result.

4. Show that the points $P \equiv (x, y)$ on the hyperbola $x^2/a^2 - y^2/b^2 = 1$ are such that the *difference* of their distances to the *foci* $F_1 \equiv (-c, 0)$ and $F_2 \equiv (c, 0)$ is equal to $\pm 2a$, where $c = \sqrt{a^2 + b^2}$. This is the geometric definition of a hyperbola.

5. Given two arbitrary fixed points F_1 and F_2 and a constant $2a \leq |F_1F_2|$, consider the curve traced by the point $P(t)$ such that the difference of the distances $|F_1P| - |F_2P| = \pm 2a$ in an arbitrary plane containing F_1 and F_2. Show that the tangent to that curve makes equal angles with the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$. This is a fundamental geometric property of the hyperbola. Make a nice sketch to show your comprehension of the problem and result.

6. Show that (58) yields $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$ that expands out to (59), where

$$\mathbf{A} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$$

and $\mathbf{x}^T = (x - x_c, y - y_c)$. Hence $\mathbf{A}$ is a symmetric positive definite matrix with eigenvalues $1/a^2$ and $1/b^2$ corresponding to eigenvectors $(\cos\alpha, \sin\alpha)^T$ and $(-\sin\alpha, \cos\alpha)^T$, respectively.

7. Show that $\cosh t$ is even, $\sinh t$ is odd and $e^t = \cosh t + \sinh t$, hence deduce that $e^{-t} = \cosh t - \sinh t$ and $1 = \cosh^2 t - \sinh^2 t$ for any real t. Sketch e^t, e^{-t}, $\cosh t$ and $\sinh t$.

8. Show that $\boldsymbol{r}(t) = \boldsymbol{a}\cos t + \boldsymbol{b}\sin t$ is an ellipse even if $\boldsymbol{a}$ and $\boldsymbol{b}$ are not orthogonal. How do you find the major and minor radii vectors? [Hint: find t_0 such that $\boldsymbol{r}\cdot\dot{\boldsymbol{r}} = 0$ and define $\boldsymbol{a}' = \boldsymbol{r}(t_0)$ and $\boldsymbol{b}' = \dot{\boldsymbol{r}}(t_0)$. Express $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{r}(t)$ in terms of $\boldsymbol{a}'$, $\boldsymbol{b}'$.]

9. Show that $\boldsymbol{r}(t) = \boldsymbol{a}\cosh t + \boldsymbol{b}\sinh t$ is a hyperbola even if $\boldsymbol{a}$ and $\boldsymbol{b}$ are not orthogonal. How do you find the major and minor radii vectors? [Hint: find t_0 such that $\boldsymbol{r}\cdot\dot{\boldsymbol{r}} = 0$ and define $\boldsymbol{a}' = \boldsymbol{r}(t_0)$ and $\boldsymbol{b}' = \dot{\boldsymbol{r}}(t_0)$. Express $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{r}(t)$ in terms of $\boldsymbol{a}'$, $\boldsymbol{b}'$.]

10. What is the most general $\boldsymbol{r}(t)$ for a curve with constant acceleration? What is the most general $\boldsymbol{r}(s)$ for a curve with constant $d^2\boldsymbol{r}/ds^2$ where s is arclength along the curve? Are these the same curves?

11. Give an example of a curve with constant curvature. Are such curves necessarily planar?

12. Given $\boldsymbol{r}(t)$, how do you compute its curvature and its torsion?

13. What is a curve with constant jerk? Find $\boldsymbol{r}(t)$ for such a curve.

14. Consider the curve $\boldsymbol{r}(t) = (1-t)^2\boldsymbol{r}_0 + 2t(1-t)\boldsymbol{r}_C + t^2\boldsymbol{r}_1$. Show that the curve is planar. Show that the curve passes through the points $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$. What are $\boldsymbol{v}(0)$ and $\boldsymbol{v}(1)$ where $\boldsymbol{v} = d\boldsymbol{r}/dt$? Show that $\boldsymbol{v}(0.5)$ is parallel to $\boldsymbol{r}_1 - \boldsymbol{r}_0$. Sketch the curve in a generic case.

15. What is the curve described by $\boldsymbol{r}(t) = a\cos\omega t\,\hat{\boldsymbol{x}} + a\sin\omega t\,\hat{\boldsymbol{y}} + bt\,\hat{\boldsymbol{z}}$, where a, b and ω are constant real numbers and $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ are a cartesian basis? Sketch the curve. What are the velocity, acceleration and jerk for this curve? Find $\hat{\boldsymbol{t}}$, $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{b}}$ for this curve. What are the curvature and torsion for this curve?

16. (i) Show that the tangent, normal and binormal unit vectors each satisfy the vector differential equation

$$\frac{d\boldsymbol{v}}{ds} = \boldsymbol{\omega}(s) \times \boldsymbol{v}$$

with $\boldsymbol{\omega} = \tau\hat{\boldsymbol{t}} + \kappa\hat{\boldsymbol{b}}$. Interpret geometrically. (ii) Write each equation in the intrinsic (Frenet) frame $\hat{\boldsymbol{t}}$, $\hat{\boldsymbol{n}}$, $\hat{\boldsymbol{b}}$. What are the units of $\boldsymbol{\omega}(s)$?

17. Consider the vector function $\boldsymbol{r}(\theta) = \boldsymbol{r}_c + a\cos\theta\,\mathbf{e}_1 + b\sin\theta\,\mathbf{e}_2$, where $\boldsymbol{r}_c$, $\mathbf{e}_1$, $\mathbf{e}_2$, a and b are constants, with $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. What kind of curve is this? Next, assume that $\boldsymbol{r}_c$, $\mathbf{e}_1$ and $\mathbf{e}_2$ are in the same plane. Consider cartesian coordinates (x, y) in that plane such that $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}$. Assume that the angle between $\mathbf{e}_1$ and $\hat{\boldsymbol{x}}$ is α. Derive the equation of the curve in terms of the cartesian coordinates (x, y) (i) in parametric form, (ii) in implicit form $f(x, y) = 0$. Simplify your equations as much as possible.

18. Generalize the previous exercise to the case where $\boldsymbol{r}_c$ is not in the same plane as $\mathbf{e}_1$ and $\mathbf{e}_2$. Consider general cartesian coordinates (x, y, z) such that $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$. Assume that all the angles between $\mathbf{e}_1$ and $\mathbf{e}_2$ and the basis vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ are known. How many independent angles is that? Specify those angles. Derive the parametric equations of the curve for the cartesian coordinates (x, y, z) in terms of the parameter θ.

19. Derive integrals for the length and area of the planar curve in the previous exercise. Clean up your integrals and compute them – if possible (one is trivial, the other is not).

20. Calculate $\int_{\mathcal{C}} d\boldsymbol{r}$ and $\int_{\mathcal{C}} \boldsymbol{r}\cdot d\boldsymbol{r}$ along the curve of the preceding exercise from $\boldsymbol{r}(0)$ to $\boldsymbol{r}(-3\pi/2)$. What are these integrals if $\mathcal{C}$ is an arbitrary curve from point A to point B?

21. Calculate $\int_{\mathcal{C}} \boldsymbol{B} \cdot d\boldsymbol{r}$ and $\int_{\mathcal{C}} \boldsymbol{B} \times d\boldsymbol{r}$ with $\boldsymbol{B} = (\hat{\boldsymbol{z}} \times \boldsymbol{r})/|\hat{\boldsymbol{z}} \times \boldsymbol{r}|^2$ when $\mathcal{C}$ is the circle of radius R in the x, y plane centered at the origin. How do the integrals depend on R? [Hint: visualize $\boldsymbol{B}$ and the curve]

22. If $\boldsymbol{F}(\boldsymbol{r}(t)) = m\,d\boldsymbol{v}/dt$ with $\boldsymbol{v} = d\boldsymbol{r}/dt$, show that $\int_{\mathcal{C}} \boldsymbol{F} \cdot d\boldsymbol{r} = \frac{1}{2}m\left(v_B^2 - v_A^2\right)$ for any curve $\mathcal{C}$ from A to B parameterized by $\boldsymbol{r}(t)$, where m is a constant and $v_A = |\boldsymbol{v}|$ at A, $v_B = |\boldsymbol{v}|$ at B.

23. Show that $\oint_{\mathcal{C}} \boldsymbol{r} \times d\boldsymbol{r} = 2A\,\hat{\boldsymbol{n}}$ where $\mathcal{C}$ is a closed curve in a plane perpendicular to $\hat{\boldsymbol{n}}$ and A is the area enclosed by the curve, even if O is *not* in the plane of the curve.

4 Surfaces

4.1 Parameterizations

Surfaces can be specified in *implicit* form as

$$f(\boldsymbol{r}) = 0 \tag{87}$$

where f is a scalar function of position, $f : \mathbb{E}^3 \to \mathbb{R}$, with $\boldsymbol{r}$ the position vector in 3D Euclidean space $\mathbb{E}^3$. For example, a sphere of radius R centered at the origin:

$$f(\boldsymbol{r}) = \boldsymbol{r} \cdot \boldsymbol{r} - R^2 = 0 = x^2 + y^2 + z^2 - R^2,$$

or a plane perpendicular to $\boldsymbol{a} = a_1\hat{\boldsymbol{x}} + a_2\hat{\boldsymbol{y}} + a_3\hat{\boldsymbol{z}}$,

$$f(\boldsymbol{r}) = \boldsymbol{a} \cdot \boldsymbol{r} - \alpha = 0 = a_1 x + a_2 y + a_3 z - \alpha.$$

Surfaces can also be specified in explicit or *parametric* form as

$$\boldsymbol{r} = \boldsymbol{r}(u, v) \tag{88}$$

where $\boldsymbol{r}(u, v)$ is a vector function of two real parameters u and v, $\boldsymbol{r} : \mathbb{R}^2 \to \mathbb{E}^3$. Those parameters (u, v) are coordinates for points $\boldsymbol{r}$ on the surface S.

Example: If $\boldsymbol{b}$ and $\boldsymbol{c}$ are any two non-parallel vectors that are both perpendicular to $\boldsymbol{a}$, then any point on the plane $\boldsymbol{a} \cdot \boldsymbol{r} - \alpha = 0$ can be specified explicitly by

$$\boldsymbol{r} = \frac{\alpha}{a}\,\hat{\boldsymbol{a}} + u\,\boldsymbol{b} + v\,\boldsymbol{c} = \boldsymbol{r}(u, v)$$

for any real u and v.

$\square$

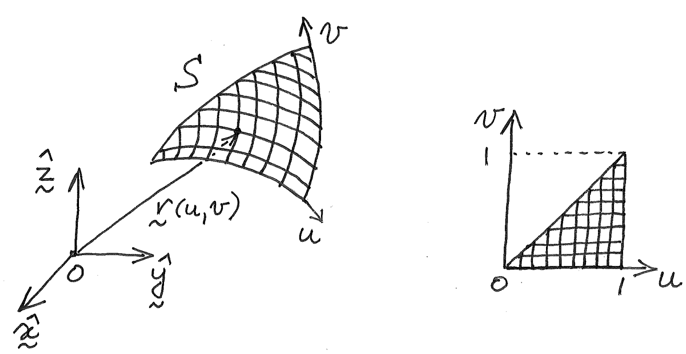


Fig. 2.12: Conceptual example of a vector function $\boldsymbol{r}(u, v)$ providing a mapping from the triangle with vertices $(0, 0), (1, 0), (1, 1)$ in the (u, v) plane to the curvy triangular surface S in the 3D (x, y, z) Euclidean space.

Example: If A, B, C are the three vertices of any triangle then any point P in that triangle is such that its position vector $\overrightarrow{OP} = \boldsymbol{r}$ satisfies

$$\boldsymbol{r} = \overrightarrow{OA} + u\,\overrightarrow{AB} + v\,\overrightarrow{BC} = \boldsymbol{r}(u, v)$$

for any (u, v) in the domain $0 \le u \le 1, 0 \le v \le u$.

$\square$

Example: For the sphere of radius R centered at O, we can solve $x^2+y^2+z^2-R^2 = 0$ for z as a function of x, y to obtain the parameterization $z = \pm\sqrt{R^2 - x^2 - y^2}$ and

$$\boldsymbol{r}(x, y) = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} \pm \sqrt{R^2 - x^2 - y^2}\,\hat{\boldsymbol{z}}, \tag{89}$$

which is defined for all (x, y) in the disk $x^2 + y^2 \le R^2$. There are two vector functions $\boldsymbol{r}(x, y)$ needed to describe the spherical surface, one for the upper or northern hemisphere with $z = \sqrt{R^2 - x^2 - y^2}$ and one for the lower or southern hemisphere with $z = -\sqrt{R^2 - x^2 - y^2}$.

$\square$

Example: Spherical coordinates. The same sphere surface $x^2 + y^2 + z^2 = R^2$ can be parametrized as

$$\boldsymbol{r}(\theta, \varphi) = R\cos\varphi\sin\theta\,\hat{\boldsymbol{x}} + R\sin\varphi\sin\theta\,\hat{\boldsymbol{y}} + R\cos\theta\,\hat{\boldsymbol{z}}, \tag{90}$$

where $\boldsymbol{r} \cdot \hat{\boldsymbol{z}} = R\cos\theta$ so θ is the polar angle, while φ is the azimuthal (or longitude) angle, the angle between $\hat{\boldsymbol{x}}$ and $\boldsymbol{\rho} = \boldsymbol{r} - z\hat{\boldsymbol{z}} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}$, the projection of $\boldsymbol{r}$ in the (x, y) plane.

$\square$

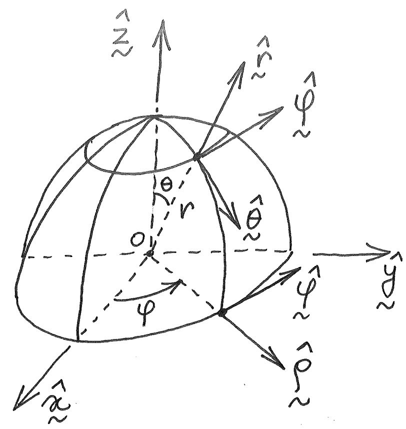


Fig. 2.13: Spherical coordinates r, θ, φ and direction vectors $\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varphi}}$.

Coordinate curves and Tangent vectors. If one of the parameters is held fixed, $v = v_0$ say, we obtain a curve $\boldsymbol{r}(u, v_0)$. There is one such curve for every value of v. For the sphere parameterized as in (90), $\boldsymbol{r}(\theta, \varphi_0)$ is the φ_0 *meridian*, the circle passing through the north and south poles and whose plane is an angle φ_0 from the *prime meridian* plane. Likewise $\boldsymbol{r}(u_0, v)$ describes another curve. For the sphere, $\boldsymbol{r}(\theta_0, \varphi)$ would be the θ_0 *parallel*, a circle parallel to the equatorial circle. The set of all such curves *generates* the surface. These two families of curves are *parametric* or *coordinate curves* for the surface. The vectors $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$ are tangent to their respective parametric curves and hence to the surface. These two vectors taken at the same point $\boldsymbol{r}(u,v)$ define the tangent plane at that point.

The coordinates are *orthogonal* if the tangent vectors $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$ at each point $\boldsymbol{r}(u,v)$ are orthogonal to each other (except perhaps at some singular points where one or both tangent vectors vanish),

$$\frac{\partial\boldsymbol{r}}{\partial u}\cdot\frac{\partial\boldsymbol{r}}{\partial v}=0. \tag{91}$$

The coordinates are *conformal* if they are orthogonal and

$$\left|\frac{\partial\boldsymbol{r}}{\partial u}\right|=\left|\frac{\partial\boldsymbol{r}}{\partial v}\right|\neq 0. \tag{92}$$

Conformal coordinates preserve angles between curves in the (u,v) and their images on the surface $\boldsymbol{r}(u,v)$. This is a very important property for navigation, for example, since angles measured on the map ((u,v) plane) are the same as angles 'in the real world' (the earth's spherical surface say).

Normal to the surface at a point. At any point $\boldsymbol{r}(u,v)$ on a surface, there is an infinity of tangent directions but there is only one normal direction. The normal to the surface at a point $\boldsymbol{r}(u,v)$ is given by

$$\boldsymbol{N}=\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}. \tag{93}$$

Note that the ordering (u,v) specifies an orientation for the surface, *i.e.* an 'up' and 'down' side, and that $\boldsymbol{N}$ is not a unit vector, in general.

Surface element. The surface element $d\boldsymbol{S}=\hat{\boldsymbol{n}}\,dS$ at a point $\boldsymbol{r}$ on a surface S is a vector of infinitesimal magnitude dS and direction $\hat{\boldsymbol{n}}$ which is the unit normal to the surface at that point. Imagine one of the small triangles that make up the surface of the skull. At different points on the skull we have a little triangular patch with small surface element $\Delta\boldsymbol{S}$ equal to 1/2 of the cross product of two small edges. If a parametric representation $\boldsymbol{r}(u,v)$ for the surface is known then

$$d\boldsymbol{S}=d\boldsymbol{r}_u\times d\boldsymbol{r}_v=\left(\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}\right)dudv, \tag{94}$$

since the right hand side represents the area of the parallelogram formed by the line elements

$$d\boldsymbol{r}_u=\frac{\partial\boldsymbol{r}}{\partial u}\,du,\qquad d\boldsymbol{r}_v=\frac{\partial\boldsymbol{r}}{\partial v}\,dv,$$

along the u and v coordinate curves, respectively. Note that $\hat{\boldsymbol{n}}\neq\boldsymbol{N}$ but $\hat{\boldsymbol{n}}=\boldsymbol{N}/|\boldsymbol{N}|$ since $\hat{\boldsymbol{n}}$ is a unit vector. Although we often need to refer to the unit normal $\hat{\boldsymbol{n}}$, it is usually not needed to compute it explicitly since in practice it is the area element $d\boldsymbol{S}$ that is needed.

Example: For the triangle $\boldsymbol{r}(u,v)=\boldsymbol{r}_A+u\,\overrightarrow{AB}+v\,\overrightarrow{BC}$ we have

$$d\boldsymbol{S}=(\overrightarrow{AB}\times\overrightarrow{BC})\,du\,dv.$$

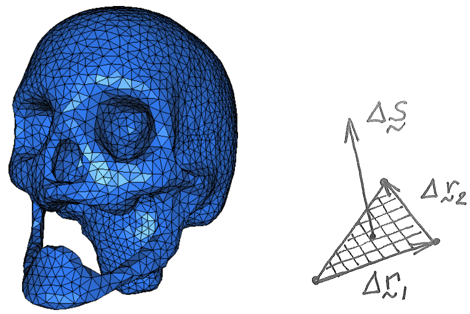


Fig. 2.14: The skull surface is modeled by a large collection of small triangles. Every small triangular patch provides a small but finite approximation $\Delta\boldsymbol{S} \triangleq \frac{1}{2}(\Delta\boldsymbol{r}_1 \times \Delta\boldsymbol{r}_2)$ to the vector surface element $d\boldsymbol{S}$, where $\Delta\boldsymbol{r}_1$ and $\Delta\boldsymbol{r}_2$ are two vector edges of the triangle oriented such that the cross-product points everywhere outward.

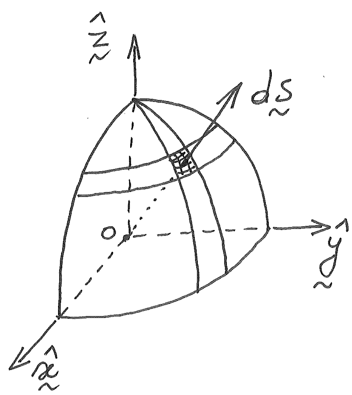


Fig. 2.15: Spherical surface element $d\boldsymbol{S}(\theta,\varphi) = \dfrac{\partial\boldsymbol{r}}{\partial\theta} \times \dfrac{\partial\boldsymbol{r}}{\partial\varphi}\, d\theta d\varphi$.

$\square$

Example: For spherical coordinates, we have

$$d\boldsymbol{S} = \frac{\partial\boldsymbol{r}}{\partial\theta}d\theta \times \frac{\partial\boldsymbol{r}}{\partial\varphi}d\varphi.$$

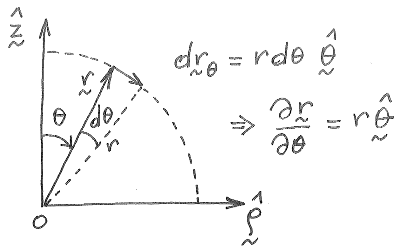
We can calculate $\partial\boldsymbol{r}/\partial\theta$ and $\partial\boldsymbol{r}/\partial\varphi$ from the hybrid representation (90)

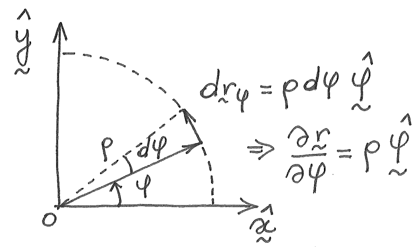
$$\boldsymbol{r}(\theta,\varphi) = r\cos\varphi\sin\theta\,\hat{\boldsymbol{x}} + r\sin\varphi\sin\theta\,\hat{\boldsymbol{y}} + r\cos\theta\,\hat{\boldsymbol{z}},$$

where r is fixed but arbitrary, yielding

$$\begin{aligned}\frac{\partial\boldsymbol{r}}{\partial\theta} &= \quad r\cos\varphi\cos\theta\,\hat{\boldsymbol{x}} + r\sin\varphi\cos\theta\,\hat{\boldsymbol{y}} - r\sin\theta\,\hat{\boldsymbol{z}},\\ \frac{\partial\boldsymbol{r}}{\partial\varphi} &= -\,r\sin\varphi\sin\theta\,\hat{\boldsymbol{x}} + r\cos\varphi\sin\theta\,\hat{\boldsymbol{y}},\end{aligned} \tag{95}$$

then compute their cross product to obtain $d\boldsymbol{S}(\theta,\varphi)$. We can also obtain those partials from direct geometric reasoning as illustrated in the following figures





where

$$\hat{\boldsymbol{r}} = \frac{\boldsymbol{r}}{r}, \qquad \hat{\boldsymbol{\varphi}} = \frac{\hat{\boldsymbol{z}}\times\boldsymbol{r}}{|\hat{\boldsymbol{z}}\times\boldsymbol{r}|}, \qquad \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\varphi}}\times\hat{\boldsymbol{r}},$$

and $\rho = r\sin\theta = \sqrt{x^2+y^2}$. Either approach yields

$$\frac{\partial\boldsymbol{r}}{\partial\theta} = r\,\hat{\boldsymbol{\theta}}, \qquad \frac{\partial\boldsymbol{r}}{\partial\varphi} = r\sin\theta\,\hat{\boldsymbol{\varphi}}, \tag{96}$$

and the surface element for a sphere of radius r as

$$d\boldsymbol{S} = \frac{\partial\boldsymbol{r}}{\partial\theta}\times\frac{\partial\boldsymbol{r}}{\partial\varphi}\,d\theta d\varphi = r\hat{\boldsymbol{\theta}}\times r\sin\theta\,\hat{\boldsymbol{\varphi}}\,d\theta d\varphi = \hat{\boldsymbol{r}}\,r^2\sin\theta\,d\theta d\varphi, \tag{97}$$

for a sphere of radius r fixed.

□

Exercises:

1. Consider the sphere $x^2+y^2+z^2=R^2$. Parametrize the northern hemisphere $z\geq 0$ using both (89) and (90). Make 3D perspective sketches of the coordinate curves for both parameterizations. What happens at the pole and at the equator? Calculate $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$ for both parameterizations. Are the coordinates orthogonal?

2. Compute tangent vectors and the normal to the surface $z=h(x,y)$. Show that $\partial\boldsymbol{r}/\partial x$ and $\partial\boldsymbol{r}/\partial y$ are not orthogonal to each other in general. Determine the class of functions $h(x,y)$ for which (x,y) are orthogonal coordinates on the surface $z=h(x,y)$ and interpret geometrically. Derive an explicit formula for the area element $dS=|d\boldsymbol{S}|$ in terms of $h(x,y)$.

3. Deduce from the implicit equation $|\boldsymbol{r}-\boldsymbol{r}_c|=R$ for a sphere of radius R centered at $\boldsymbol{r}_c$ that $(\boldsymbol{r}-\boldsymbol{r}_c)\cdot\partial\boldsymbol{r}/\partial u = (\boldsymbol{r}-\boldsymbol{r}_c)\cdot\partial\boldsymbol{r}/\partial v = 0$ for any u and v, where $\boldsymbol{r}(u,v)$ is *any* parameterization of a point on that sphere. Find a $\boldsymbol{r}(u,v)$ for that sphere and compute $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$. Are your coordinates orthogonal? Compute the surface element $d\boldsymbol{S}$ for that sphere. Do your surface elements point toward or away from the center of the sphere?

4. Show that spherical coordinates (90) are orthogonal but not conformal. What happens on the polar axis?

5. *Stereographic coordinates* map a point (x, y, z) on the sphere $x^2+y^2+z^2 = R^2$ to the point $(u, v, 0)$ that is the intersection of the $z = 0$ plane and the line passing through the south pole $(0, 0, -R)$ and the point (x, y, z). Sketch. What is the parametric equation of that line? Find $\boldsymbol{r}(u, v)$ and show that these (u, v) coordinates are *conformal.*

6. Explain why the surface of a torus (*i.e.* 'donut' or tire) can be parameterized as $x = (R + a\cos v)\cos u$, $y = (R + a\cos v)\sin u$, $z = a\sin v$. Interpret the geometric meaning of the constants R, a, and the variables (u, v). What are the ranges of u and v needed to cover the entire torus? Do these parameters provide orthogonal coordinates for the torus? Calculate the surface element $\mathrm{d}\boldsymbol{S}$.

7. Describe the surface $\boldsymbol{r}(u, v) = R\cos u\ \hat{\boldsymbol{x}} + R\sin u\ \hat{\boldsymbol{y}} + v\ \hat{\boldsymbol{z}}$ where $0 \le u < 2\pi$ and $0 \le v \le H$, and R and H are positive constants. What is the surface element and what is the total surface area? Show that $\partial\boldsymbol{r}/\partial u$, $\partial\boldsymbol{r}/\partial v$ are continuous across the angle cut $u = 2\pi$ and unique for u modulo 2π.

8. The *Möbius* strip can be parameterized as
$$\boldsymbol{r}(u, v) = \left(R + v\sin\frac{u}{2}\right)\ \cos u\,\hat{\boldsymbol{x}} + \left(R + v\sin\frac{u}{2}\right)\ \sin u\,\hat{\boldsymbol{y}} + v\cos\frac{u}{2}\ \hat{\boldsymbol{z}}$$
where $0 \le u < 2\pi$ and $-h \le v \le h$. Sketch this surface. Show that $\partial\boldsymbol{r}/\partial u$ is continuous across the angle cut $u = 2\pi$, $v = 0$, but $\partial\boldsymbol{r}/\partial v$ is not. Show that $\partial\boldsymbol{r}/\partial u$ is unique but $\partial\boldsymbol{r}/\partial v$ is double-valued at any u modulo 2π with $v = 0$.

9. Explain why the surface described by
$$\boldsymbol{r}(u, v) = \hat{\boldsymbol{x}}\,a\cos u\cos v + \hat{\boldsymbol{y}}\,b\sin u\cos v + \hat{\boldsymbol{z}}\,c\sin v$$
where a, b and c are real constants is the surface of an *ellipsoid.* Are u and v orthogonal coordinates for that surface? Consider cartesian coordinates (x, y, z) such that $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$. Derive the implicit equation $f(x, y, z) = 0$ satisfied by all such $\boldsymbol{r}(u, v)$'s.

10. Consider the vector function $\boldsymbol{r}(u, v) = \boldsymbol{r}_c + \mathbf{e}_1\,a\cos u\cos v + \mathbf{e}_2\,b\sin u\cos v + \mathbf{e}_3\,c\sin v$, where $\boldsymbol{r}_c, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, a, b and c are constants with $\mathbf{e}_i\cdot\mathbf{e}_j = \delta_{ij}$. What does the set of all such $\boldsymbol{r}(u, v)$'s represent? Consider cartesian coordinates (x, y, z) such that $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$. Assume that all the angles between $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and the basis vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ are known. How many independent angles is that? Specify which angles. Can you assume $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is right handed? Explain. Derive the implicit equation $f(x, y, z) = 0$ satisfied by all such $\boldsymbol{r}(u, v)$'s. Express your answer in terms of the minimum independent angles that you specified earlier.

4.2 Surface integrals

The typical surface integral is of the form $\int_S \boldsymbol{v} \cdot d\boldsymbol{S}$. This represents the *flux* of $\boldsymbol{v}$ through the surface S. If $\boldsymbol{v}(\boldsymbol{r})$ is the velocity of a fluid, water or air, at point $\boldsymbol{r}$, then $\int_S \boldsymbol{v} \cdot d\boldsymbol{S}$ is the time-rate at which volume of fluid is flowing through that surface per unit time. Indeed, $\boldsymbol{v} \cdot d\boldsymbol{S} = (\boldsymbol{v} \cdot \hat{\boldsymbol{n}})\, dS$ where $\hat{\boldsymbol{n}}$ is the unit normal to the surface and $\boldsymbol{v} \cdot \hat{\boldsymbol{n}}$ is the component of fluid velocity that is perpendicular to the surface. If that component is zero, the fluid moves tangentially to the surface, not through the surface. Speed $\times$ area = volume per unit time, so $\boldsymbol{v} \cdot d\boldsymbol{S}$ is the volume of fluid passing through the surface element $d\boldsymbol{S}$ per unit time at that point at that time. The total volume passing through the entire surface S per unit time is $\int_S \boldsymbol{v} \cdot d\boldsymbol{S}$. Such integrals are often called *flux* integrals.

We can make sense of such integrals as *limits of sums*, for instance,

$$\int_S \boldsymbol{v} \cdot d\boldsymbol{S} = \lim_{\Delta S_n \to 0} \sum_{n=1}^{N} \boldsymbol{v}_n \cdot \Delta \boldsymbol{S}_n \tag{98}$$

where the sum is over, say, a triangular partition of the surface (such as the skull surface) with N triangles and $\boldsymbol{v}_n$ is for instance $\boldsymbol{v}$ at the center of area of triangle n whose area vector $\Delta\boldsymbol{S}_n$. Another good choice is to define $\boldsymbol{v}_n$ as the average of $\boldsymbol{v}$ at the vertices of triangle n. Here $\Delta S_n \to 0$ means that the partition is uniformly refined with $N \to \infty$. This is conceptually intuitive although that limit is a bit tricky to define precisely mathematically. Roughly speaking, each triangle should be shrinking to a point, not to a line, and of course we need more and more of smaller and smaller triangles to cover the whole surface.

If a parametric representation $\boldsymbol{r}(u, v)$ is known for the surface then we can also write

$$\int_S \boldsymbol{v} \cdot d\boldsymbol{S} = \int_A \left(\boldsymbol{v} \cdot \left(\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} \right) \right) dudv, \tag{99}$$

where A is the domain in the u, v parameter plane that corresponds to S.

As for line integrals, we can make sense of many other types of surface integrals such as

$$\int_S p\, d\boldsymbol{S},$$

which would represent the net pressure force on S if $p = p(\boldsymbol{r})$ is the pressure at point $\boldsymbol{r}$. Other surface integrals could have the form $\int_S \boldsymbol{v} \times d\boldsymbol{S}$, etc. In particular the total area of surface S is $\int_S |d\boldsymbol{S}|$.

Examples: For the triangle ABC parameterized as $T : \boldsymbol{r}(u, v) = \boldsymbol{r}_A + u\,\overrightarrow{AB} + v\,\overrightarrow{BC}$, $0 \le u \le 1$, $0 \le v \le u$, we can directly evaluate the surface integral

$$\int_T d\boldsymbol{S} = \frac{1}{2}\,\overrightarrow{AB} \times \overrightarrow{BC}$$

since it is the vector sum of all surface elements $d\boldsymbol{S}$, and that is the total triangle area times the unit normal to the triangle. We can also compute that integral explicitly as

$$\int_T d\boldsymbol{S} = \int_0^1 du \int_0^u \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\, dv = \overrightarrow{AB} \times \overrightarrow{BC} \int_0^1 du \int_0^u dv = \frac{1}{2}\overrightarrow{AB} \times \overrightarrow{BC}.$$

The surface integral $\int_T dS$ is the sum of the magnitudes of the surface elements, therefore it is the total triangle area. It can also be computed as

$$\int_T |d\boldsymbol{S}| = \int_0^1 du \int_0^u dv \left| \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} \right| = \frac{1}{2} \left| \overrightarrow{AB} \times \overrightarrow{BC} \right|.$$

Finally, the integral

$$\int_T \boldsymbol{r} \cdot d\boldsymbol{S} = \frac{1}{2}(\boldsymbol{r}_A \cdot \hat{\boldsymbol{n}}) \left| \overrightarrow{AB} \times \overrightarrow{BC} \right|$$

since $d\boldsymbol{S} = \hat{\boldsymbol{n}}\, dS$ with $\hat{\boldsymbol{n}}$ is constant for the triangle and $\boldsymbol{r} \cdot \hat{\boldsymbol{n}} = \boldsymbol{r}_A \cdot \hat{\boldsymbol{n}}$ for all points in that plane, and that constant $(\boldsymbol{r}_A \cdot \hat{\boldsymbol{n}})$ is the distance from the origin to the plane of the triangle. □

Exercises:

1. Compute the percentage of surface area that lies north of the arctic circle on the earth (assume it is a perfect sphere). Show your work, don't just google it.

2. Provide an explicit integral for the total surface area of the torus parametrized as in §4, problem 6. Compute the total surface area.

3. Calculate $\int_S \boldsymbol{r} \cdot d\boldsymbol{S}$ where S is (i) the square $0 \le x \le a$, $0 \le y \le a$ at $z = b$, (ii) the surface of the sphere of radius R centered at $(0,0,0)$, (iii) the surface of the sphere of radius R centered at $x = x_0$, $y = z = 0$.

4. Calculate $\int_S (\boldsymbol{r}/r^3) \cdot d\boldsymbol{S}$ where S is the surface of the sphere of radius R centered at the origin. How does the result depend on R?

5. The pressure *outside* the sphere of radius R centered at $\boldsymbol{r}_c$ is $p = p_0 + A\boldsymbol{r} \cdot \hat{\boldsymbol{a}}$ where $\hat{\boldsymbol{a}}$ is an arbitrary but fixed unit vector and p_0 and A are constants. The pressure inside the sphere is the constant $p_1 > p_0$. Calculate the net force on the sphere. Calculate the net torque on the sphere about its center $\boldsymbol{r}_c$ and about the origin.

4.3 Curves on surfaces

[Future version of these notes will discuss curves on surface, geodesics, fundamental forms, surface curvature]

The line element on a surface $\boldsymbol{r}(u, v)$ is

$$d\boldsymbol{r} = \frac{\partial \boldsymbol{r}}{\partial u}\, du + \frac{\partial \boldsymbol{r}}{\partial v}\, dv \tag{100}$$

thus the arclength element $ds = |d\boldsymbol{r}| = \sqrt{(d\boldsymbol{r}) \cdot (d\boldsymbol{r})}$ is

$$ds^2 = \frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial u}\, du^2 + 2\frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial v}\, du\, dv + \frac{\partial \boldsymbol{r}}{\partial v} \cdot \frac{\partial \boldsymbol{r}}{\partial v}\, dv^2. \tag{101}$$

where $ds^2 \triangleq (ds)^2$ not $d(s^2) = 2s\,ds$, and likewise for du^2 and dv^2. In differential geometry, this is called the *first fundamental form* of the surface usually written

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2. \tag{102}$$

The area element $dS = |d\boldsymbol{S}|$ is

$$\begin{aligned} dS^2 =& \left(\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right) \cdot \left(\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right)\, du^2\, dv^2 \\ =& (EG - F^2)\, du^2 dv^2 \end{aligned} \tag{103}$$

using the vector identity $(\boldsymbol{a} \times \boldsymbol{b}) \cdot (\boldsymbol{a} \times \boldsymbol{b}) = (\boldsymbol{a} \cdot \boldsymbol{a})(\boldsymbol{b} \cdot \boldsymbol{b}) - (\boldsymbol{a} \cdot \boldsymbol{b})^2$.

5 Volumes

We have seen that $\boldsymbol{r}(t)$ is the parametric equation of a curve, $\boldsymbol{r}(u, v)$ represents a surface, now we discuss $\boldsymbol{r}(u, v, w)$ which is the parametric representation of a volume. Curves $\boldsymbol{r}(t)$ are one dimensional objects so they have only one parameter t or each point on the known curve is determined by a single *coordinate*. Surfaces are two-dimensional and require two parameters u and v, which are coordinates for points on that surface. Volumes are three-dimensional objects that require three parameters u, v, w say. Each point is specified by three *coordinates*.

Example: Any point P inside the tetrahedron with vertices A, B, C, D has

$$\boldsymbol{r} = \boldsymbol{r}_A + u\,\overrightarrow{AB} + v\,\overrightarrow{BC} + w\,\overrightarrow{CD} = \boldsymbol{r}(u, v, w). \tag{104}$$

with

$$0 \le u \le 1, \qquad 0 \le v \le u, \qquad 0 \le w \le v.$$

$\square$

Example: A sphere of radius R centered at $\boldsymbol{r}_c$ has the implicit equation $|\boldsymbol{r} - \boldsymbol{r}_c| \le R$, or $(\boldsymbol{r} - \boldsymbol{r}_c) \cdot (\boldsymbol{r} - \boldsymbol{r}_c) \le R^2$ to avoid square roots. In cartesian coordinates this translates into the implicit equation

$$(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2 \le R^2. \tag{105}$$

An explicit parametrization for that sphere is

$$\boldsymbol{r}(r, \theta, \varphi) = \boldsymbol{r}_c + \hat{\boldsymbol{x}}\, r \sin\theta \cos\varphi + \hat{\boldsymbol{y}}\, r \sin\theta \sin\varphi + \hat{\boldsymbol{z}}\, r \cos\theta, \tag{106}$$

where r is distance to the origin, θ the polar angle and φ the azimuthal (or longitude) angle. To describe the sphere fully but uniquely we need

$$0 \le r \le R, \qquad 0 \le \theta \le \pi, \qquad 0 \le \varphi < 2\pi. \tag{107}$$

$\square$

Coordinate curves

For a curve $\boldsymbol{r}(t)$, all we needed to worry about was the tangent $d\boldsymbol{r}/dt$ and the line element $d\boldsymbol{r} = (d\boldsymbol{r}/dt)dt$. For surfaces, $\boldsymbol{r}(u, v)$ we have two sets of coordinates curves with tangents $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$, a normal $\boldsymbol{N} = (\partial\boldsymbol{r}/\partial u) \times (\partial\boldsymbol{r}/\partial v)$ and a surface element $d\boldsymbol{S} = \boldsymbol{N}dudv$. Now for volumes $\boldsymbol{r}(u, v, w)$, we have three sets of coordinates curves with tangents $\partial\boldsymbol{r}/\partial u$, $\partial\boldsymbol{r}/\partial v$ and $\partial\boldsymbol{r}/\partial w$. A u-coordinate curve for instance, corresponds to $\boldsymbol{r}(u, v, w)$ with v and w fixed. There is a double infinity of such one dimensional curves, one for each v, w pair. For the parametrization (106), the φ-coordinate curves correspond to *parallels*, *i.e.* circles of fixed radius r at fixed polar angle θ. The θ-coordinate curves are *meridians*, *i.e.* circles of fixed radius r through the poles. The r-coordinate curves are radial lines out of the origin.

Coordinate surfaces

For volumes $\boldsymbol{r}(u, v, w)$, we also have three sets of *coordinate surfaces* corresponding to one parameter fixed and the other two free. A w-isosurface for instance corresponds to $\boldsymbol{r}(u, v, w)$ for a fixed w. There is a single infinity of such two dimensional (u, v) surfaces. For the parametrization (106) such surfaces correspond to spherical surfaces of radius r centered at $\boldsymbol{r}_c$. Likewise, if we fix u but let v and w free, we get another surface, and v fixed with u and w free is another coordinate surface.

Line Elements

Thus given a volume parametrization $\boldsymbol{r}(u, v, w)$ we can define four types of line elements, one for each of the coordinate directions $(\partial\boldsymbol{r}/\partial u)du$, $(\partial\boldsymbol{r}/\partial v)dv$, $(\partial\boldsymbol{r}/\partial w)dw$ and a general line element corresponding to the infinitesimal displacement from coordinates (u, v, w) to the coordinates $(u + du, v + dv, w + dw)$. That general line element $d\boldsymbol{r}$ is given by (chain rule):

$$d\boldsymbol{r} = \frac{\partial\boldsymbol{r}}{\partial u}du + \frac{\partial\boldsymbol{r}}{\partial v}dv + \frac{\partial\boldsymbol{r}}{\partial w}dw. \tag{108}$$

The arclength follows from

$$\begin{aligned} ds^2 = d\boldsymbol{r}\cdot d\boldsymbol{r} = &\frac{\partial\boldsymbol{r}}{\partial u}\cdot\frac{\partial\boldsymbol{r}}{\partial u}du^2 + \frac{\partial\boldsymbol{r}}{\partial v}\cdot\frac{\partial\boldsymbol{r}}{\partial v}dv^2 + \frac{\partial\boldsymbol{r}}{\partial w}\cdot\frac{\partial\boldsymbol{r}}{\partial w}dw^2 \\ &+ 2\frac{\partial\boldsymbol{r}}{\partial u}\cdot\frac{\partial\boldsymbol{r}}{\partial v}du\,dv + 2\frac{\partial\boldsymbol{r}}{\partial v}\cdot\frac{\partial\boldsymbol{r}}{\partial w}dv\,dw + 2\frac{\partial\boldsymbol{r}}{\partial w}\cdot\frac{\partial\boldsymbol{r}}{\partial u}du\,dw \end{aligned} \tag{109}$$

and contains 6 terms if the coordinates are not orthogonal. Note that ds^2 must be interpreted as $(ds)^2$ not $d(s^2) = 2s\,ds$, and likewise for du^2, dv^2, dw^2. Switching to index notation with $(u, v, w) \to (u_1, u_2, u_3)$ and summation convention, the arclength element is

$$ds^2 = g_{ij}\; du_i\, du_j \tag{110}$$

where

$$g_{ij} \triangleq \frac{\partial\boldsymbol{r}}{\partial u_i}\cdot\frac{\partial\boldsymbol{r}}{\partial u_j} = g_{ji} \tag{111}$$

is the *metric* tensor.

Surface Elements

Likewise, there are three basic types of *surface elements*, one for each coordinate surface. The surface element on a w-isosurface, for example, is given by

$$d\boldsymbol{S}_w = \left(\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}\right)dudv, \tag{112}$$

while the surface elements on a u-isosurface and a v-isosurface are respectively

$$d\boldsymbol{S}_u = \left(\frac{\partial \boldsymbol{r}}{\partial v} \times \frac{\partial \boldsymbol{r}}{\partial w}\right) dv dw, \qquad d\boldsymbol{S}_v = \left(\frac{\partial \boldsymbol{r}}{\partial w} \times \frac{\partial \boldsymbol{r}}{\partial u}\right) du dw. \tag{113}$$

Note that surface orientations are built into the order of the coordinates.

Volume Element

Last but not least, a parametrization $\boldsymbol{r}(u, v, w)$ defines a *volume element* given by the mixed (i.e. triple scalar) product

$$dV = \left(\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right) \cdot \frac{\partial \boldsymbol{r}}{\partial w}\, du\, dv\, dw \equiv \det\left(\frac{\partial \boldsymbol{r}}{\partial u}, \frac{\partial \boldsymbol{r}}{\partial v}, \frac{\partial \boldsymbol{r}}{\partial w}\right) du\, dv\, dw. \tag{114}$$

The definition of volume integrals as limit-of-a-sum should be obvious by now. If an explicit parametrization $\boldsymbol{r}(u, v, w)$ for the volume is known, we can use the volume element (114) and write the volume integral in $\boldsymbol{r}$ space as an iterated triple integral over u, v, w. Be careful that there is an orientation implicitly built into the ordering of the parameters, as should be obvious from the definition of the mixed product and determinants. The volume element dV is usually meant to be positive so the sign of the mixed product and the bounds of integrations for the parameters u, v and w must be chosen to respect that. (Recall the definition of $|dt|$ in the line integrals section).

Exercises

1. Calculate the line, surface and volume elements for the coordinates (106). You need to calculate 4 line elements and 3 surfaces elements. One line element for each coordinate curve and the general line element. Verify that these coordinates are orthogonal.

2. Formulate integral expressions in terms of the coordinates (106) for the surface and volume of a sphere of radius R. Calculate those integrals.

3. A curve $\boldsymbol{r}(t)$ is given in terms of the (u, v, w) coordinates, *i.e.* $\boldsymbol{r}(t) = \boldsymbol{r}(u, v, w)$ with $(u(t), v(t), w(t))$ for $t = t_a$ to $t = t_b$. Find an explicit expression in terms of $(u(t), v(t), w(t))$ as a t-integral for the length of that curve.

4. Find suitable coordinates for a torus. Are your coordinates orthogonal? Compute the volume of that torus.

6 Maps, curvilinear coordinates

Parameterizations of curves, surfaces and volumes is essentially equivalent to the concepts of 'maps' and curvilinear coordinates.

Maps

The parametrization (90) for the surface of a sphere of radius R provides a *map* between that surface and the $0 \le \varphi < 2\pi$, $0 \le \theta \le \pi$ rectangle in the (φ, θ) plane. In a mapping $\boldsymbol{r}(u, v)$ a small rectangle of sides du, dv at a point (u, v) in the (u, v) plane is mapped to a small parallelogram of sides $(\partial \boldsymbol{r}/\partial u)du$, $(\partial \boldsymbol{r}/\partial v)dv$ at point $\boldsymbol{r}(u, v)$ in the Euclidean space.

The parametrization (106) for the sphere of radius R centered at $\boldsymbol{r}_c$ provides a *map* between the sphere of radius R in Euclidean space and the box $0 \le r \le R$, $0 \le \theta \le \pi$, $0 \le \varphi < 2\pi$, in the (r, θ, φ) space. In a mapping $\boldsymbol{r}(u, v, w)$, the infinitesimal box of sides du, dv, dw located at point (u, v, w) in the (u, v, w) space is mapped to a parallelepiped of sides $(\partial \boldsymbol{r}/\partial u)du$, $(\partial \boldsymbol{r}/\partial v)dv$, $(\partial \boldsymbol{r}/\partial w)dw$ at the point $\boldsymbol{r}(u, v, w)$ in the Euclidean space.

Curvilinear coordinates, orthogonal coordinates

The parameterizations $\boldsymbol{r}(u, v)$ and $\boldsymbol{r}(u, v, w)$ define *coordinates* for a surface or a volume, respectively. If the coordinate curves are not straight lines one talks of *curvilinear coordinates.* These maps define good coordinates if the coordinate curves intersect *transversally*, that is if the coordinate curves are not tangent to each other, for a surface that is

$$\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} \neq 0$$

and for a volume

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

If the coordinate curves intersect transversally at a point then the coordinates tangent vectors at that point provide linearly independent directions in the space of $\boldsymbol{r}$. Tangent intersections at a point $\boldsymbol{r}$ would imply that the tangent vectors are linearly dependent at that point. This is acceptable on a *set of measure zero*, for instance at the equator for the parametrization (89) or on the polar axis for (106).

The coordinates (u, v, w) are *orthogonal* if the coordinate curves in $\boldsymbol{r}$-space intersect at right angles. This is the best kind of 'transversal' intersection and these are the most desirable type of coordinates, however non-orthogonal coordinates are sometimes more convenient for some problems. Two fundamental examples of orthogonal curvilinear coordinates are

▶ *Cylindrical (or polar) coordinates*

$$x = \rho \cos \theta, \quad y = \rho \sin \theta, \quad z = z. \tag{115}$$

▶ *Spherical coordinates*

$$x = r \sin \theta \cos \varphi, \quad y = r \sin \theta \sin \varphi, \quad z = r \cos \theta. \tag{116}$$

Changing notation from (x, y, z) to (x_1, x_2, x_3) and from (u, v, w) to (q_1, q_2, q_3) a general change of coordinates from cartesian $(x, y, z) \equiv (x_1, x_2, x_3)$ to curvilinear

(q_1, q_2, q_3) coordinates is expressed succinctly by

$$x_i = x_i(q_1, q_2, q_3), \quad i = 1, 2, 3. \tag{117}$$

The position vector $\boldsymbol{r}$ can be expressed in terms of the q_j's through the cartesian expression:

$$\boldsymbol{r}(q_1, q_2, q_3) = \hat{\boldsymbol{x}}\, x(q_1, q_2, q_3) + \hat{\boldsymbol{y}}\, y(q_1, q_2, q_3) + \hat{\boldsymbol{z}}\, z(q_1, q_2, q_3) = \sum_{i=1}^{3} \mathbf{e}_i\ x_i(q_1, q_2, q_3), \tag{118}$$

where $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. The q_i coordinate curve is the curve $\boldsymbol{r}(q_1, q_2, q_3)$ where q_i is free but the other two variables are fixed. The q_i isosurface is the surface $\boldsymbol{r}(q_1, q_2, q_3)$ where q_i is fixed and the other two parameters are free.

The coordinate tangent vectors $\partial \boldsymbol{r}/\partial q_i$ are key to the coordinates. They provide a natural vector basis for those coordinates. The coordinates are orthogonal if these tangent vectors are orthogonal to each other at each point. In that case it is useful to define the unit vector $\hat{\boldsymbol{q}}_i$ in the q_i coordinate direction by

$$h_i = \left| \frac{\partial \boldsymbol{r}}{\partial q_i} \right|, \qquad \frac{\partial \boldsymbol{r}}{\partial q_i} = h_i \hat{\boldsymbol{q}}_i, \quad \text{(no sum)} \tag{119}$$

where h_i is the the magnitude of the tangent vector in the q_i direction, $\partial \boldsymbol{r}/\partial q_i$, and $\hat{\boldsymbol{q}}_i \cdot \hat{\boldsymbol{q}}_j = \delta_{ij}$ for orthogonal coordinates. These h_i's are called the *scale* factors. Note that this decomposition of $\partial \boldsymbol{r}/\partial q_i$ into a scale factor h_i and a direction $\hat{\boldsymbol{q}}_i$ clashes with the summation convention. The distance traveled in x-space when changing q_i by dq_i, keeping the other q's fixed, is $|d\boldsymbol{r}| = h_i dq_i$ (no summation). The distance ds travelled in x-space when the orthogonal curvilinear coordinates change from (q_1, q_2, q_3) to $(q_1 + dq_1, q_2 + dq_2, q_3 + dq_3)$ is

$$ds^2 = d\boldsymbol{r} \cdot d\boldsymbol{r} = h_1^2\, dq_1^2 + h_2^2\, dq_2^2 + h_3^2\, dq_3^2. \tag{120}$$

Although the cartesian unit vectors $\mathbf{e}_i$ are independent of the coordinates, the curvilinear unit vectors $\hat{\boldsymbol{q}}_i$ in general *are* functions of the coordinates, even if the latter are orthogonal. Hence $\partial \hat{\boldsymbol{q}}_i/\partial q_j$ is in general non-zero. For orthogonal coordinates, those derivatives $\partial \hat{\boldsymbol{q}}_i/\partial q_j$ can be expressed in terms of the scale factors and the unit vectors.

For instance, for spherical coordinates $(q_1, q_2, q_3) \equiv (r, \varphi, \theta)$, the unit vector in the $q_1 \equiv r$ direction is the vector

$$\frac{\partial \boldsymbol{r}(r, \varphi, \theta)}{\partial r} = \hat{\boldsymbol{x}}\ \sin\theta \cos\varphi + \hat{\boldsymbol{y}}\ \sin\theta \sin\varphi + \hat{\boldsymbol{z}}\ \cos\theta, \tag{121}$$

so the scale coefficient $h_1 \equiv h_r = 1$ and the unit vector

$$\hat{\boldsymbol{q}}_1 \equiv \hat{\boldsymbol{r}} \equiv \mathbf{e}_r = \hat{\boldsymbol{x}}\ \sin\theta \cos\varphi + \hat{\boldsymbol{y}}\ \sin\theta \sin\varphi + \hat{\boldsymbol{z}}\ \cos\theta. \tag{122}$$

The position vector $\boldsymbol{r}$ can be expressed as

$$\boldsymbol{r} = r\, \hat{\boldsymbol{r}} = x\, \hat{\boldsymbol{x}} + y\, \hat{\boldsymbol{y}} + z \hat{\boldsymbol{z}} \ \equiv\ \sum_{i=1}^{3} x_i\, \mathbf{e}_i. \tag{123}$$

So its expression in terms of spherical coordinates and their unit vectors, $\boldsymbol{r} = r\hat{\boldsymbol{r}}$, is simpler than in cartesian coordinates, $\boldsymbol{r} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$, but there is a catch! The radial unit vector $\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}(\varphi, \theta)$ varies in the azimuthal and polar angle directions, while the cartesian unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$ are independent of the coordinates!

For orthogonal coordinates, the scale factors h_i's determine everything. In particular, the surface and volume elements can be expressed in terms of the h_i's. For instance, the surface element for a q_3-isosurface is

$$d\boldsymbol{S}_3 = \hat{\boldsymbol{q}}_3\, h_1 h_2\, dq_1 dq_2, \tag{124}$$

and the volume element

$$dV = h_1 h_2 h_3\, dq_1 dq_2 dq_3, \tag{125}$$

assuming that q_1, q_2, q_3 is right-handed. These follow directly from (112) and (114) and (119) when the coordinates are orthogonal.

Exercises

1. Find the scale factors h_i and the unit vectors $\hat{\boldsymbol{q}}_i$ for cylindrical and spherical coordinates. Express the 3 surface elements and the volume element in terms of those scale factors and unit vectors. Sketch the unit vector $\hat{\boldsymbol{q}}_i$ in the (x, y, z) space (use several 'views' rather than trying to make an ugly 3D sketch!). Express the position vector $\boldsymbol{r}$ in terms of the unit vectors $\hat{\boldsymbol{q}}_i$. Calculate the derivatives $\partial\hat{\boldsymbol{q}}_i/\partial q_j$ for all i, j and express these derivatives in terms of the scale factors h_k and the unit vectors $\hat{\boldsymbol{q}}_k$, $k = 1, 2, 3$.

2. A curve in the (x, y) plane is given in terms of polar coordinates as $\rho = \rho(\theta)$. Deduce θ-integral expressions for the length of the curve and for the area swept by the radial vector.

3. Consider *elliptical coordinates* (u, v, w) defined by $x = \alpha \cosh u \cos v$, $y = \alpha \sinh u \sin v$, $z = w$ for some $\alpha > 0$, where x, y and z are standard cartesian coordinates in 3D Euclidean space. What do the coordinate curves correspond to in the (x, y, z) space? Are these orthogonal coordinates? What is the volume element in terms of elliptical coordinates?

4. For general curvilinear coordinates, not necessarily orthogonal, is the q_i-isosurface perpendicular to $\partial\boldsymbol{r}/\partial q_i$? is it orthogonal to $(\partial\boldsymbol{r}/\partial q_j) \times (\partial\boldsymbol{r}/\partial q_k)$ where i, j, k are all distinct? What about for orthogonal coordinates?

7 Change of variables

Parameterizations of surfaces and volumes and curvilinear coordinates are geometric examples of a *change of variables*. These changes of variables and the associated formula and geometric concepts can occur in non-geometric contexts. The fundamental relationship is the formula for a volume element (114). In the context of a general change of variables from (x_1, x_2, x_3) to (q_1, q_2, q_3) that formula (114) reads

$$dx_1 dx_2 dx_3 = dV_x = J\, dq_1 dq_2 dq_3 = J\, dV_q \tag{126}$$

where

$$J = \begin{vmatrix} \partial x_1/\partial q_1 & \partial x_1/\partial q_2 & \partial x_1/\partial q_3 \\ \partial x_2/\partial q_1 & \partial x_2/\partial q_2 & \partial x_2/\partial q_3 \\ \partial x_3/\partial q_1 & \partial x_3/\partial q_2 & \partial x_3/\partial q_3 \end{vmatrix} = \det\left(\frac{\partial x_i}{\partial q_j}\right) \tag{127}$$

is the *Jacobian determinant* and dV_x is a volume element in the x-space while dV_q is the corresponding volume element in q-space. The Jacobian is the determinant of the Jacobian *matrix*

$$\boldsymbol{J} = \begin{bmatrix} \partial x_1/\partial q_1 & \partial x_1/\partial q_2 & \partial x_1/\partial q_3 \\ \partial x_2/\partial q_1 & \partial x_2/\partial q_2 & \partial x_2/\partial q_3 \\ \partial x_3/\partial q_1 & \partial x_3/\partial q_2 & \partial x_3/\partial q_3 \end{bmatrix} \quad \Leftrightarrow \quad J_{ij} = \frac{\partial x_i}{\partial q_j}. \tag{128}$$

The vectors $(dq_1, 0, 0)$, $(0, dq_2, 0)$ and $(0, 0, dq_3)$ at point (q_1, q_2, q_3) in q-space are mapped to the vectors $(\partial \boldsymbol{r}/\partial q_1) dq_1$, $(\partial \boldsymbol{r}/\partial q_2) dq_2$, $(\partial \boldsymbol{r}/\partial q_3) dq_3$. In component form this is

$$\begin{bmatrix} dx_1^{(1)} \\ dx_2^{(1)} \\ dx_3^{(1)} \end{bmatrix} = \begin{bmatrix} \partial x_1/\partial q_1 & \partial x_1/\partial q_2 & \partial x_1/\partial q_3 \\ \partial x_2/\partial q_1 & \partial x_2/\partial q_2 & \partial x_2/\partial q_3 \\ \partial x_3/\partial q_1 & \partial x_3/\partial q_2 & \partial x_3/\partial q_3 \end{bmatrix} \begin{bmatrix} dq_1 \\ 0 \\ 0 \end{bmatrix}, \tag{129}$$

where $(dx_1^{(1)}, dx_2^{(1)}, dx_3^{(1)})$ are the x-components of the vector $(\partial \boldsymbol{r}/\partial q_1) dq_1$. Similar relations hold for the other basis vectors. Note that the rectangular box in q-space is in general mapped to a non-rectangular parallelepiped in x-space so the notation $dx_1 dx_2 dx_3$ for the volume element in (126) is a (common) abuse of notation.

The formulas (126), (127) tells us how to change variables in multiple integrals. This formula generalizes to higher dimension, and also to lower dimension. In the 2 variable case, we have a 2-by-2 determinant that can also be understood as a special case of the surface element formula (94) for a mapping $\boldsymbol{r}(q_1, q_2)$ from a 2D space (q_1, q_2) to another 2D-space (x_1, x_2). In that case $\boldsymbol{r}(q_1, q_2) = \mathbf{e}_1 x_1(q_1, q_2) + \mathbf{e}_2 x_2(q_1, q_2)$ so $(\partial \boldsymbol{r}/\partial q_1) \times (\partial \boldsymbol{r}/\partial q_2) dq_1 dq_2 = \mathbf{e}_3\, dA_x$ and

$$dx_1 dx_2 = dA_x = J\, dq_1 dq_2 = J\, dA_q \tag{130}$$

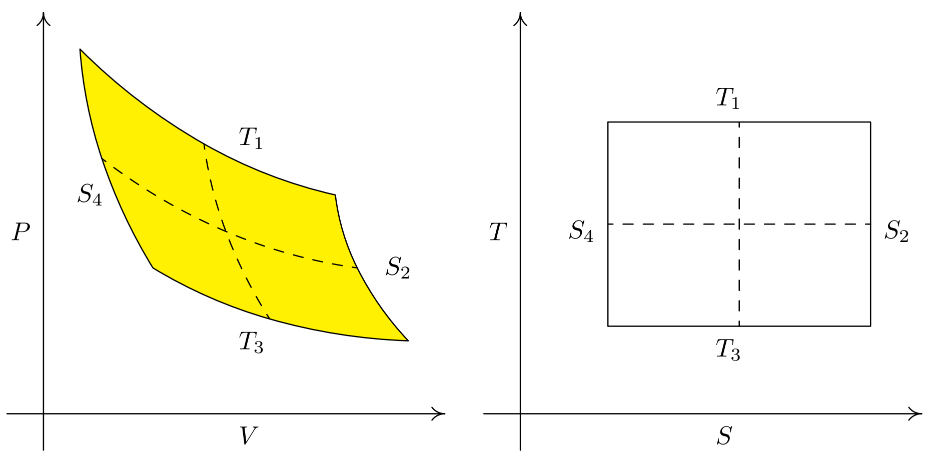
where the *Jacobian determinant* is now

$$J = \begin{vmatrix} \partial x_1/\partial q_1 & \partial x_1/\partial q_2 \\ \partial x_2/\partial q_1 & \partial x_2/\partial q_2 \end{vmatrix}. \tag{131}$$

Change of variables example: Carnot cycle

Consider a Carnot cycle for a perfect gas. The equation of state is $PV = nRT = NkT$, where P is pressure, V is the volume of gas and T is the temperature in Kelvins.

The volume V contains n moles of gas corresponding to N molecules, R is the gas constant and k is Boltzmann's constant with $nR = Nk$. A *Carnot cycle* is an idealized thermodynamic cycle in which a gas goes through (1) a heated isothermal expansion at temperature T_1, (2) an adiabatic expansion at constant entropy S_2, (3) an isothermal compression releasing heat at temperature $T_3 < T_1$ and (4) an adiabatic compression at constant entropy $S_4 < S_2$. For a perfect monoatomic gas, constant entropy means constant PV^γ where $\gamma = C_P/C_V = 5/3$ with C_P and C_V the heat capacity at constant pressure P or constant volume V, respectively. Thus let $S = PV^\gamma$ (this S is not the physical entropy but it is constant whenever entropy is constant, we can call S a 'pseudo-entropy').



Now the work done by the gas when its volume changes from V_a to V_b is $\int_{V_a}^{V_b} P dV$ (since work = Force × displacement, P = force/area and V=area × displacement). Thus the (yellow) area inside the cycle in the (P, V) plane is the net work performed by the gas during one cycle. Although we can calculate that area by working in the (P, V) plane, it is easier to calculate it by using a change of variables from P, V to T, S. The area inside the cycle in the (P, V) plane is not the same as the area inside the cycle in the (S, T) plane. There is a distortion. An element of area in the (P, V) plane aligned with the T and S coordinates (*i.e.* with the dashed curves in the (P, V) plane) is

$$dA = \left| \left(\frac{\partial P}{\partial S} \mathbf{e}_{_P} + \frac{\partial V}{\partial S} \mathbf{e}_{_V} \right) dS \times \left(\frac{\partial P}{\partial T} \mathbf{e}_{_P} + \frac{\partial V}{\partial T} \mathbf{e}_{_V} \right) dT \right| \tag{132}$$

where $\mathbf{e}_{_P}$ and $\mathbf{e}_{_V}$ are the unit vectors in the P and V directions in the (P, V) plane, respectively. This is entirely similar to the area element of surface $\boldsymbol{r}(u, v)$ being equal to $d\boldsymbol{S} = \frac{\partial \boldsymbol{r}}{\partial u} du \times \frac{\partial \boldsymbol{r}}{\partial v} dv$ (but don't confuse the surface element $d\boldsymbol{S}$ with the pseudo-entropy differential dS used in the present example!). Calculating out the cross product, we obtain

$$dA = \left| \left(\frac{\partial P}{\partial S} \frac{\partial V}{\partial T} - \frac{\partial V}{\partial S} \frac{\partial P}{\partial T} \right) dS dT \right| = \pm\, J_{_{S,T}}^{^{P,V}}\; dS dT \tag{133}$$

where the $\pm$ sign will be chosen to get a positive area (this depends on the bounds of integrations) and

$$J_{S,T}^{P,V} = \det(\boldsymbol{J}_{S,T}^{P,V}) = \begin{vmatrix} \dfrac{\partial P}{\partial S} & \dfrac{\partial P}{\partial T} \\[2ex] \dfrac{\partial V}{\partial S} & \dfrac{\partial V}{\partial T} \end{vmatrix} \tag{134}$$

is the *Jacobian* determinant (here the vertical bars are the common notation for determinants). It is the determinant of the Jacobian *matrix* $\boldsymbol{J}_{S,T}^{P,V}$ that corresponds to the mapping from (S,T) to (P,V). The cycle area $A_{P,V}$ in the (P,V) plane is thus

$$A_{P,V} = \int_{T_3}^{T_1} \int_{S_4}^{S_2} \left| J_{S,T}^{P,V} \right| dSdT. \tag{135}$$

Note that the vertical bars in this formula are for absolute value of $J_{S,T}^{P,V}$ and the bounds have been selected so that $dS > 0$ and $dT > 0$ (in the limit-of-a-sum sense). This expression for the (P,V) area expressed in terms of (S,T) coordinates is simpler than if we used (P,V) coordinates, except for that $\left| J_{S,T}^{P,V} \right|$ since we do not have explicit expression for $P(S,T)$ and $V(S,T)$. What we have in fact are the inverse functions $T = PV/(Nk)$ and $S = PV^{\gamma}$. To find the partial derivatives that we need we could (1) find the inverse functions by solving for P and V in terms of T and S then compute the partial derivatives and the Jacobian, or (2) use implicit differentiation *e.g.* $Nk\,\partial T/\partial T = Nk = V\partial P/\partial T + P\partial V/\partial T$ and $\partial S/\partial T = 0 = V^{\gamma}\partial P/\partial T + P\gamma V^{\gamma-1}\partial V/\partial T$ etc. and solve for the partial derivatives we need. But there is a simpler way that makes use of an important property of Jacobians.

Geometric meaning of the Jacobian determinant and its inverse

The Jacobian $J_{S,T}^{P,V}$ represents the stretching factor of area elements when moving from the (S,T) plane to the (P,V) plane. If $dA_{S,T}$ is an area element centered at point (S,T) in the (S,T) plane then that area element gets mapped to an area element $dA_{P,V}$ centered at the corresponding point in the (P,V) plane. That's what equation (133) represents. In that equation we have in mind the mapping of a rectangular element of area $dSdT$ in the (S,T) plane to a parallelogram element in the (P,V) plane. The stretching factor is $|J_{S,T}^{P,V}|$ (as we saw earlier, the meaning of the sign is related to orientation, but here we are worrying only about areas, so we take absolute values). That relationship is valid for area elements of any shape, not just rectangles to parallelogram since the differential relationships implies an implicit limit-of-a-sum and in that limit, the 'pointwise' area stretching is independent of the shape of the area elements. A disk element in the (S,T) plane would be mapped to an ellipse element in the (P,V) plane but the pointwise area stretching would be the same as for a rectangular element (this is not true for *finite* size areas). So equation (133) can be written in the more general form

$$dA_{P,V} = |J_{S,T}^{P,V}|\, dA_{S,T} \tag{136}$$

which is locally valid for area elements of any shape. The key point is that if we consider the inverse map, back from (P, V) to (S, T) then there is an are stretching give by the Jacobian $J^{S,T}_{P,V} = (\partial S/\partial P)(\partial T/\partial V) - (\partial S/\partial V)(\partial T/\partial P)$ such that

$$dA_{S,T} = |J^{S,T}_{P,V}|\, dA_{P,V} \tag{137}$$

but since we are coming back to the original $dA_{S,T}$ element we must have

$$\boxed{J^{P,V}_{S,T}\, J^{S,T}_{P,V} = 1,} \tag{138}$$

so the Jacobian determinant are inverses of one another. This inverse relationship actually holds for the Jacobian *matrices* also

$$\boxed{\boldsymbol{J}^{P,V}_{S,T}\, \boldsymbol{J}^{S,T}_{P,V} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.} \tag{139}$$

The latter can be derived from the chain rule since (note the consistent ordering of the partials)

$$\boldsymbol{J}^{P,V}_{S,T} = \begin{pmatrix} \partial P/\partial S & \partial P/\partial T \\ \partial V/\partial S & \partial V/\partial T \end{pmatrix}, \qquad \boldsymbol{J}^{S,T}_{P,V} = \begin{pmatrix} \partial S/\partial P & \partial S/\partial V \\ \partial T/\partial P & \partial T/\partial V \end{pmatrix} \tag{140}$$

and the matrix product of those two Jacobian matrices yields the identity matrix. For instance, the first row times the first column gives

$$\left(\frac{\partial P}{\partial S}\right)_T \left(\frac{\partial S}{\partial P}\right)_V + \left(\frac{\partial P}{\partial T}\right)_S \left(\frac{\partial T}{\partial P}\right)_V = \left(\frac{\partial P}{\partial P}\right)_V = 1.$$

A subscript has been added to remind which other variable is held fixed during the partial differentiation. The inverse relationship between the Jacobian determinants (138) then follows from the inverse relationship between the Jacobian matrices (139) since the determinant of a product is the product of the determinants. This important property of determinants can be verified directly by explicit calculation for this 2-by-2 case.

So what is the work done by the gas during one Carnot cycle? Well,

$$\begin{aligned} J^{P,V}_{S,T} = \left(J^{S,T}_{P,V}\right)^{-1} &= \left(\frac{\partial S}{\partial P}\frac{\partial T}{\partial V} - \frac{\partial S}{\partial V}\frac{\partial T}{\partial P}\right)^{-1} \\ &= Nk\left(V^{\gamma}P - \gamma V^{\gamma-1}PV\right)^{-1} = \frac{Nk}{(1-\gamma)S} \end{aligned} \tag{141}$$

so

$$\begin{aligned} A_{P,V} &= \int_{T_3}^{T_1}\int_{S_4}^{S_2} \left|J^{P,V}_{S,T}\right| dSdT \\ &= \int_{T_3}^{T_1}\int_{S_4}^{S_2} \frac{Nk}{(\gamma-1)S} dSdT = \frac{Nk(T_1-T_3)}{(\gamma-1)} \ln\frac{S_2}{S_4}, \end{aligned} \tag{142}$$

since $\gamma > 1$ and other quantities are positive.

Exercises

1. Calculate the area between the curves $xy = \alpha_1$, $xy = \alpha_2$ and $y = \beta_1 x$, $y = \beta_2 x$ in the (x, y) plane. Sketch the area. (α_1, α_2, β_1, $\beta_2 > 0$.)

2. Calculate the area between the curves $xy = \alpha_1$, $xy = \alpha_2$ and $y^2 = 2\beta_1 x$, $y^2 = 2\beta_2 x$ in the (x, y) plane. Sketch the area. (α_1, α_2, β_1, $\beta_2 > 0$.)

3. Calculate the area between the curves $x^2 + y^2 = 2\alpha_1 x$, $x^2 + y^2 = 2\alpha_2 x$ and $x^2 + y^2 = 2\beta_1 y$, $x^2 + y^2 = 2\beta_2 y$. Sketch the area. (α_1, α_2, β_1, $\beta_2 > 0$.)

4. Calculate the area of the ellipse $x^2/a^2 + y^2/b^2 = 1$ and the volume of the ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ by transforming them to a disk and a sphere, respectively, using a change of variables. [Hint: consider the change of variables $x = au$, $y = bv$, $z = cw$.]

5. Calculate the integral $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$. Deduce the value of the Poisson integral $\int_{-\infty}^{\infty} e^{-x^2} dx$. [Hint: switch to polar coordinates].

6. Calculate $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a^2 + x^2 + y^2)^{\alpha} dxdy$. Where $a \neq 0$ and α is real. Discuss the values of α for which the integral exists.

8 Grad, div, curl

8.1 Scalar fields and iso-contours

Consider a scalar function of a vector variable: $f(\boldsymbol{r})$, for instance the pressure $p(\boldsymbol{r})$ as a function of position, or the temperature $T(\boldsymbol{r})$ at point $\boldsymbol{r}$, etc. One way to visualize such functions is to consider *isosurfaces* or *level sets*, these are the sets of all $\boldsymbol{r}$'s for which $f(\boldsymbol{r}) = f_0$, for some constant f_0. In cartesian coordinates $\boldsymbol{r} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} + z\,\hat{\boldsymbol{z}}$ and the scalar function of position, that is the *scalar field*, is a function of the three coordinates $f(\boldsymbol{r}) \equiv f(x, y, z)$.

Example: The isosurfaces of $f(\boldsymbol{r}) = \boldsymbol{r}\cdot\boldsymbol{r} = x^2 + y^2 + z^2$ are determined by the equation $f(\boldsymbol{r}) = f_0$. These are spheres or radius $\sqrt{f_0}$, if $f_0 \geq 0$.

Example: The isosurfaces of $f(\boldsymbol{r}) = \boldsymbol{a}\cdot\boldsymbol{r} = a_1x + a_2y + a_3z$ where $\boldsymbol{a} = a_1\hat{\boldsymbol{x}} + a_2\hat{\boldsymbol{y}} + a_3\hat{\boldsymbol{z}}$ is constant, are determined by the equation $f(\boldsymbol{r}) = f_0 = a_1x + a_2y + a_3z$. These are planes perpendicular to $\boldsymbol{a}$ at a signed distance $f_0/|\boldsymbol{a}|$ from the origin.

8.2 Geometric concept of the Gradient

The gradient of a scalar field $f(\boldsymbol{r})$ denoted $\boldsymbol{\nabla} f$ is a *vector* that points in the direction of *greatest increase of* f and whose magnitude equals the rate of change of f with distance in that direction. Let $\hat{\boldsymbol{S}}$ be the direction of *steepest* (or greatest) increase of f at point $\boldsymbol{r}$ and s the distance in that $\hat{\boldsymbol{S}}$ direction, then

$$\boldsymbol{\nabla} f = \hat{\boldsymbol{S}}\,\frac{\partial f}{\partial s}. \tag{143}$$

The gradient of a scalar field is a *vector field*, it is a function of position and varies from point to point, in general.

Fundamental Examples:

1. The gradient of $f(\boldsymbol{r}) = |\boldsymbol{r}| = r$, the distance to the origin, is

   $$\boldsymbol{\nabla} r = \hat{\boldsymbol{r}} \tag{144}$$

   since the direction of steepest increase of r is $\hat{\boldsymbol{r}}$ and the rate of increase of r with distance in the $\hat{\boldsymbol{r}}$ direction is $dr/dr = 1$.

2. The gradient of $f(\boldsymbol{r}) = |\boldsymbol{r} - \boldsymbol{r}_A| = |\overrightarrow{AP}|$, the distance from point A to point P, is

   $$\boldsymbol{\nabla}|\boldsymbol{r} - \boldsymbol{r}_A| = \frac{\boldsymbol{r} - \boldsymbol{r}_A}{|\boldsymbol{r} - \boldsymbol{r}_A|} \tag{145}$$

   since the direction of greatest increase of $|\boldsymbol{r} - \boldsymbol{r}_A|$ is in the direction of $\boldsymbol{r} - \boldsymbol{r}_A$, that is the unit vector $(\boldsymbol{r} - \boldsymbol{r}_A)/|\boldsymbol{r} - \boldsymbol{r}_A|$, and the rate of change of that distance $|\boldsymbol{r} - \boldsymbol{r}_A|$ in that direction is obviously 1.

3. The gradient of $f(\boldsymbol{r}) = \boldsymbol{a}\cdot\boldsymbol{r}$ is

   $$\boldsymbol{\nabla}(\boldsymbol{a}\cdot\boldsymbol{r}) = \boldsymbol{a} \tag{146}$$

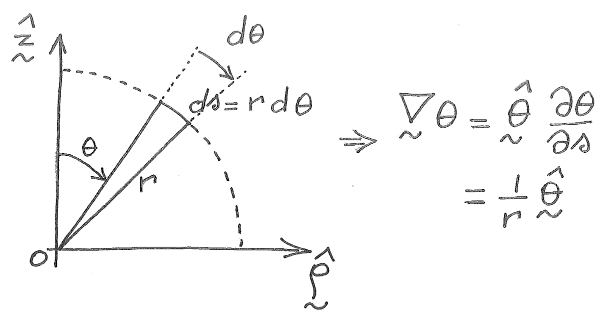
since $\boldsymbol{a}\cdot\boldsymbol{r} = |\boldsymbol{a}|s$ where $s = \hat{\boldsymbol{a}}\cdot\boldsymbol{r}$ is the (signed) distance from the origin in the direction $\hat{\boldsymbol{a}}$, thus $\hat{\boldsymbol{S}} = \hat{\boldsymbol{a}}$ for this scalar field and the rate of change of that distance s in that direction is again 1. In particular

$$\begin{aligned} \boldsymbol{\nabla}(\hat{\boldsymbol{x}}\cdot\boldsymbol{r}) &= \boldsymbol{\nabla}x = \hat{\boldsymbol{x}}, \\ \boldsymbol{\nabla}(\hat{\boldsymbol{y}}\cdot\boldsymbol{r}) &= \boldsymbol{\nabla}y = \hat{\boldsymbol{y}}, \\ \boldsymbol{\nabla}(\hat{\boldsymbol{z}}\cdot\boldsymbol{r}) &= \boldsymbol{\nabla}z = \hat{\boldsymbol{z}}. \end{aligned} \tag{147}$$

4. The gradient of $f(\boldsymbol{r}) = \theta$ where θ is the polar angle between the fixed direction $\hat{\boldsymbol{z}}$ and $\boldsymbol{r}$ is

$$\boldsymbol{\nabla}\theta = \hat{\boldsymbol{\theta}}\,\frac{\partial\theta}{\partial s} = \frac{1}{r}\hat{\boldsymbol{\theta}} \tag{148}$$

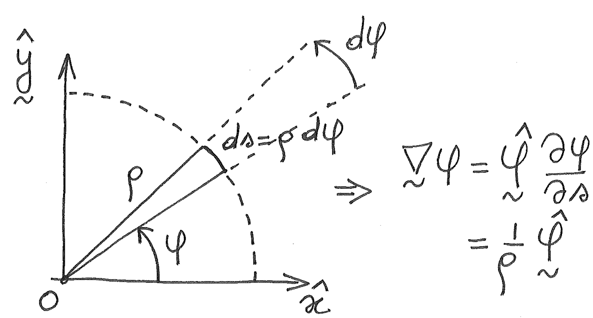
since $\hat{\boldsymbol{\theta}}$ is the direction (south) of greatest increase of θ (the angle from the North pole) and an infinitesimal step ds in that direction yields an increment $d\theta$ such that $ds = r d\theta$.



5. The gradient of $f(\boldsymbol{r}) = \varphi$ where φ is the azimuthal angle between the planes $\hat{\boldsymbol{z}},\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}},\boldsymbol{r}$

$$\boldsymbol{\nabla}\varphi = \hat{\boldsymbol{\varphi}}\,\frac{\partial\varphi}{\partial s} = \frac{1}{r\sin\theta}\hat{\boldsymbol{\varphi}} \tag{149}$$

since $\hat{\boldsymbol{\varphi}}$ is the direction (east) of greatest increase of φ (longitude) and an infinitesimal step ds in that direction yields an increment $d\varphi$ such that $ds = \rho\,d\varphi = r\sin\theta\,d\varphi$ since $\rho = r\sin\theta$ is the distance to the $\hat{\boldsymbol{z}}$ axis.



6. The gradient of $f(r)$, a function that depends only on distance to the origin $r = |\boldsymbol{r}|$ is
$$\boldsymbol{\nabla} f(r) = \frac{df}{dr}\,\hat{\boldsymbol{r}}. \tag{150}$$
since the direction of greatest increase is $\pm\hat{\boldsymbol{r}}$ and the rate of change of $f(r)$ in that direction is $\pm df/dr$. In particular
$$\begin{aligned}\boldsymbol{\nabla}\left(r^2\right) &= 2r\hat{\boldsymbol{r}} = 2\boldsymbol{r},\\ \boldsymbol{\nabla}\left(\frac{1}{r}\right) &= -\frac{\hat{\boldsymbol{r}}}{r^2} = -\frac{\boldsymbol{r}}{r^3}.\end{aligned} \tag{151}$$

7. More generally, if $f = f(s)$ where s is an arbitrary scalar field $s(\boldsymbol{r})$
$$\boldsymbol{\nabla} f = \frac{df}{ds}\,\boldsymbol{\nabla} s. \tag{152}$$
For example
$$\boldsymbol{\nabla} f(\theta) = \frac{\partial f}{\partial\theta}\boldsymbol{\nabla}\theta = \frac{\hat{\boldsymbol{\theta}}}{r}\frac{\partial f}{\partial\theta}.$$

Note that the gradient $\boldsymbol{\nabla} f$ is perpendicular to level sets of $f(\boldsymbol{r})$.

8.3 Directional derivative, gradient and the $\boldsymbol{\nabla}$ operator

The rate of change of $f(\boldsymbol{r})$ with respect to t along the curve $\boldsymbol{r}(t)$ is
$$\frac{df}{dt} = \lim_{\Delta t\to 0}\frac{f(\boldsymbol{r}+\Delta\boldsymbol{r}) - f(\boldsymbol{r})}{\Delta t} \tag{153}$$
where $\Delta\boldsymbol{r} = \boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t)$ and $\boldsymbol{r} = \boldsymbol{r}(t)$.

In Cartesian coordinates, $f(\boldsymbol{r}) \equiv f(x,y,z)$ and $f(\boldsymbol{r}(t)) \equiv f\left(x(t),y(t),z(t)\right)$ then, by the chain rule, we obtain
$$\begin{aligned}\frac{df}{dt} =& \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dy}{dt}\frac{\partial f}{\partial y} + \frac{dz}{dt}\frac{\partial f}{\partial z},\\ =& \frac{d\boldsymbol{r}}{dt}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial f}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial f}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial f}{\partial z}\right).\end{aligned} \tag{154}$$

If t is time then $d\boldsymbol{r}/dt = \boldsymbol{v}$ is the local velocity. If t is arclength s along the curve, then $d\boldsymbol{r}/ds = \hat{\boldsymbol{t}} = \hat{\boldsymbol{s}}$ is the unit tangent to the curve at that point, that is the direction of increasing arclength s. Then (154) reads
$$\frac{\partial f}{\partial s} = \hat{\boldsymbol{s}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial f}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial f}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial f}{\partial z}\right) \tag{155}$$
and this is the *directional derivative*, the rate of change of $f(\boldsymbol{r})$ with respect to distance s in some arbitrary direction $\hat{\boldsymbol{s}}$, whose fundamental definition is
$$\frac{\partial f}{\partial s} = \lim_{\Delta s\to 0}\frac{f(\boldsymbol{r}+\hat{\boldsymbol{s}}\,\Delta s) - f(\boldsymbol{r})}{\Delta s}, \tag{156}$$

now written as a partial derivative since the function varies in all directions but we are looking at its rate of change in a particular direction $\hat{\boldsymbol{s}}$. This fits with our earlier notions of partial derivatives $\partial f/\partial x$, $\partial f/\partial y$ and $\partial f/\partial z$ that indeed corresponds to rates of change in directions $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$, respectively.

Inspection of (155) shows that the directional derivative $\partial f/\partial s$ will be largest when $\hat{\boldsymbol{s}}$ is in the direction of the vector $\hat{\boldsymbol{x}}\partial f/\partial x + \hat{\boldsymbol{y}}\partial f/\partial y + \hat{\boldsymbol{z}}\partial f/\partial z$, thus that vector must be the gradient of f

$$\boldsymbol{\nabla} f = \hat{\boldsymbol{x}}\frac{\partial f}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial f}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial f}{\partial z}, \tag{157}$$

in cartesian coordinates x, y, z. The directional derivative (156) in arbitrary direction $\hat{\boldsymbol{s}}$ can then be written as

$$\frac{\partial f}{\partial s} = \hat{\boldsymbol{s}} \cdot \boldsymbol{\nabla} f. \tag{158}$$

This provides the cartesian coordinates expression for the *del operator*

$$\begin{aligned} \boldsymbol{\nabla} &= \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}, && (159)\\ &= \mathbf{e}_1\frac{\partial}{\partial x_1} + \mathbf{e}_2\frac{\partial}{\partial x_2} + \mathbf{e}_3\frac{\partial}{\partial x_3} && (160)\\ &\equiv \mathbf{e}_i\partial_i && (161)\end{aligned}$$

where ∂_i is short for $\partial/\partial x_i$ and we have used the convention of summation over all values of the repeated index i.

Del is a vector differential operator, it yields a vector field $\boldsymbol{\nabla} f$ when we operate with it on a scalar field $f(\boldsymbol{r})$. From the rules of multivariable calculus we obtain the following fundamental properties of $\boldsymbol{\nabla}$, where $f = f(\boldsymbol{r})$ and $g = g(\boldsymbol{r})$ are arbitrary scalar fields.

Gradient of a sum:

$$\boldsymbol{\nabla}(f+g) = \boldsymbol{\nabla} f + \boldsymbol{\nabla} g, \tag{162}$$

Product rule for $\boldsymbol{\nabla}$:

$$\boldsymbol{\nabla}(fg) = (\boldsymbol{\nabla} f)g + f(\boldsymbol{\nabla} g), \tag{163}$$

Chain rule for $\boldsymbol{\nabla}$:

$$\boldsymbol{\nabla} f(u_1, u_2, u_3, \cdots) = \frac{\partial f}{\partial u_1}\boldsymbol{\nabla} u_1 + \frac{\partial f}{\partial u_2}\boldsymbol{\nabla} u_2 + \frac{\partial f}{\partial u_3}\boldsymbol{\nabla} u_3 + \cdots \tag{164}$$

In particular, the chain rule yields expressions for $\boldsymbol{\nabla}$ in other coordinates systems, for instance

$$\boldsymbol{\nabla} f(r,\theta,\varphi) = \frac{\partial f}{\partial r}\boldsymbol{\nabla} r + \frac{\partial f}{\partial \theta}\boldsymbol{\nabla}\theta + \frac{\partial f}{\partial \varphi}\boldsymbol{\nabla}\varphi,$$

which combined with our earlier direct geometric results that $\boldsymbol{\nabla} r = \hat{\boldsymbol{r}}$, $\boldsymbol{\nabla}\theta = \hat{\boldsymbol{\theta}}/r$, $\boldsymbol{\nabla}\varphi = \hat{\boldsymbol{\varphi}}/(r\sin\theta)$ gives

$$\boldsymbol{\nabla} = \hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \frac{\hat{\boldsymbol{\theta}}}{r}\frac{\partial}{\partial\theta} + \frac{\hat{\boldsymbol{\varphi}}}{r\sin\theta}\frac{\partial}{\partial\varphi}. \tag{165}$$

8.4 Div and Curl

We define divergence and curl of a vector field $\boldsymbol{V}(\boldsymbol{r})$ based on their cartesian representations. Geometric interpretations of div and curl will be discussed later. *Vector fields* $\boldsymbol{V}(\boldsymbol{r})$ are vector-valued functions of position $\boldsymbol{r}$, for example the wind velocity $\boldsymbol{V}(\boldsymbol{r})$ of air or water at point $\boldsymbol{r}$, or the electric field $\boldsymbol{E}(\boldsymbol{r})$ at point $\boldsymbol{r}$, or the magnetic field $\boldsymbol{B}(\boldsymbol{r})$.

Divergence

The *divergence* of a vector field $\boldsymbol{V}(\boldsymbol{r})$ is the 'dot product' $\boldsymbol{\nabla}\cdot\boldsymbol{V}$, of the vector differential operator $\boldsymbol{\nabla}$ with the vector field $\boldsymbol{V}(\boldsymbol{r})$. In cartesian coordinates, $\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\partial_x + \hat{\boldsymbol{y}}\partial_y + \hat{\boldsymbol{z}}\partial_z$ and $\boldsymbol{V} = v_x\hat{\boldsymbol{x}} + v_y\hat{\boldsymbol{y}} + v_z\hat{\boldsymbol{z}}$, and

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{V} &= (\hat{\boldsymbol{x}}\partial_x + \hat{\boldsymbol{y}}\partial_y + \hat{\boldsymbol{z}}\partial_z)\cdot\boldsymbol{V}\\ &= \hat{\boldsymbol{x}}\cdot\partial_x\boldsymbol{V} + \hat{\boldsymbol{y}}\cdot\partial_y\boldsymbol{V} + \hat{\boldsymbol{z}}\cdot\partial_z\boldsymbol{V}\\ &= \partial_x v_x + \partial_y v_y + \partial_z v_z,\end{aligned} \tag{166}$$

where $\partial_x v_x = (\partial v_x/\partial x)$, etc. since the direction vector $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ do not vary with position. In cartesian index notation, $\boldsymbol{\nabla} = \mathbf{e}_i\partial_i$ and $\boldsymbol{V} = v_j\mathbf{e}_j$, thus $\boldsymbol{\nabla}\cdot\boldsymbol{V} = \big(\mathbf{e}_i\partial_i\big)\cdot\big(\mathbf{e}_j v_j\big) = (\mathbf{e}_i\cdot\mathbf{e}_j)\,\partial_i v_j = \delta_{ij}\partial_i v_j$ and

$$\boxed{\boldsymbol{\nabla}\cdot\boldsymbol{V} = \partial_i v_i.} \tag{167}$$

Examples:

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{\rho} &= 2,\\ \boldsymbol{\nabla}\cdot\boldsymbol{r} &= 3,\\ \boldsymbol{\nabla}\cdot(\hat{\boldsymbol{z}}\times\boldsymbol{r}) &= 0,\\ \boldsymbol{\nabla}\cdot(x^2\hat{\boldsymbol{x}} + xyz\,\hat{\boldsymbol{y}}) &= 2x + xz,\end{aligned} \tag{168}$$

where $\boldsymbol{\rho} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}$, $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}\times\boldsymbol{r} = x\,\hat{\boldsymbol{y}} - y\,\hat{\boldsymbol{x}} = \hat{\boldsymbol{z}}\times\boldsymbol{\rho}$.

Commutativity of the dot product $\boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{b}\cdot\boldsymbol{a}$ which holds for regular vectors, does *not* hold for the vector *differential operator* $\boldsymbol{\nabla}$,

$$\boldsymbol{\nabla}\cdot\boldsymbol{V} \neq \boldsymbol{V}\cdot\boldsymbol{\nabla} \tag{169}$$

since $\boldsymbol{\nabla}\cdot\boldsymbol{V} = \partial_i v_i$ is a scalar *field* but $\boldsymbol{V}\cdot\boldsymbol{\nabla} = v_i\partial_i$ is a scalar *operator*. For example,

$$\boldsymbol{\nabla}\cdot\boldsymbol{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3 \tag{170}$$

but

$$\boldsymbol{r}\cdot\boldsymbol{\nabla} = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z} = r\hat{\boldsymbol{r}}\cdot\boldsymbol{\nabla} = r\frac{\partial}{\partial r}. \tag{171}$$

The last result follows from the fact that $\boldsymbol{V}\cdot\boldsymbol{\nabla} = |\boldsymbol{V}|\hat{\boldsymbol{V}}\cdot\boldsymbol{\nabla}$ but $\hat{\boldsymbol{V}}\cdot\boldsymbol{\nabla}$ is the directional derivative in the direction of $\boldsymbol{V}$. In fluid dynamics where $\boldsymbol{V}$ is a fluid velocity, the operator $\boldsymbol{V}\cdot\boldsymbol{\nabla}$ is the advective (or convective) derivative.

Curl

The *curl* of a vector field $\boldsymbol{V}(\boldsymbol{r})$ is the 'cross product' $\boldsymbol{\nabla}\times\boldsymbol{V}$. In cartesian coordinates, that is

$$\begin{aligned}\boldsymbol{\nabla}\times\boldsymbol{V} &= (\hat{\boldsymbol{x}}\partial_x+\hat{\boldsymbol{y}}\partial_y+\hat{\boldsymbol{z}}\partial_z)\times(v_x\hat{\boldsymbol{x}}+v_y\hat{\boldsymbol{y}}+v_z\hat{\boldsymbol{z}})\\ &= \hat{\boldsymbol{x}}\,(\partial_y v_z-\partial_z v_y)+\hat{\boldsymbol{y}}\,(\partial_z v_x-\partial_x v_z)+\hat{\boldsymbol{z}}\,(\partial_x v_y-\partial_y v_x)\,.\end{aligned} \tag{172}$$

or in cartesian index notation $\boldsymbol{\nabla}\times\boldsymbol{V}=(\mathbf{e}_j\partial_j)\times(\mathbf{e}_k v_k)=(\mathbf{e}_j\times\mathbf{e}_k)\,\partial_j v_k$. Recall that $\epsilon_{ijk}=\mathbf{e}_i\cdot(\mathbf{e}_j\times\mathbf{e}_k)$, thus $\mathbf{e}_j\times\mathbf{e}_k=\mathbf{e}_i\,\epsilon_{ijk}$ and

$$\boxed{\boldsymbol{\nabla}\times\boldsymbol{V}=\mathbf{e}_i\,\epsilon_{ijk}\,\partial_j v_k.} \tag{173}$$

The right hand side is a *triple sum* over all values of the repeated indices i, j and k. That triple sum has only 6 non-zero terms since $\epsilon_{ijk}=\pm 1$ depending when (i,j,k) is a cyclic (=even) permutation or an acyclic (=odd) permutation of $(1,2,3)$ and vanishes in all other instances. Thus (173) expands to

$$\boldsymbol{\nabla}\times\boldsymbol{V}=\mathbf{e}_1\,(\partial_2 v_3-\partial_3 v_2)+\mathbf{e}_2\,(\partial_3 v_1-\partial_1 v_3)+\mathbf{e}_3\,(\partial_1 v_2-\partial_2 v_1) \tag{174}$$

which is the index equivalent of (172). We can also write that the i-th cartesian component of the curl $\boldsymbol{C}=\boldsymbol{\nabla}\times\boldsymbol{V}$ directly as

$$\boxed{\mathbf{e}_i\cdot\big(\boldsymbol{\nabla}\times\boldsymbol{V}\big)=C_i=\epsilon_{ijk}\,\partial_j v_k.} \tag{175}$$

Examples:

$$\begin{aligned}\boldsymbol{\nabla}\times\boldsymbol{\rho}&=0,\\ \boldsymbol{\nabla}\times\boldsymbol{r}&=0,\\ \boldsymbol{\nabla}\times(\hat{\boldsymbol{z}}\times\boldsymbol{r})&=2\hat{\boldsymbol{z}},\\ \boldsymbol{\nabla}\times(x^2\hat{\boldsymbol{x}}+xyz\,\hat{\boldsymbol{y}})&=-xy\,\hat{\boldsymbol{x}}+yz\,\hat{\boldsymbol{z}},\end{aligned} \tag{176}$$

where $\boldsymbol{\rho}=x\hat{\boldsymbol{x}}+y\hat{\boldsymbol{y}}$ and $\boldsymbol{r}=x\hat{\boldsymbol{x}}+y\hat{\boldsymbol{y}}+z\hat{\boldsymbol{z}}$. Note again that $\boldsymbol{a}\times\boldsymbol{b}=-\boldsymbol{b}\times\boldsymbol{a}$ for regular vectors, but this does not hold with the vector differential operator $\boldsymbol{\nabla}$

$$\boldsymbol{\nabla}\times\boldsymbol{V}\neq-\boldsymbol{V}\times\boldsymbol{\nabla} \tag{177}$$

since $\boldsymbol{\nabla}\times\boldsymbol{V}$ is a vector *field* while $\boldsymbol{V}\times\boldsymbol{\nabla}$ is a vector *operator.*

8.5 Vector identities

In the following $f=f(\boldsymbol{r})$ is an arbitrary scalar field while $\boldsymbol{V}(\boldsymbol{r})$ and $\boldsymbol{W}(\boldsymbol{r})$ are vector fields. Two fundamental identities that can be remembered from $\boldsymbol{a}\cdot(\boldsymbol{a}\times\boldsymbol{b})=0$ and $\boldsymbol{a}\times(\alpha\boldsymbol{a})=0$ are

$$\boxed{\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{V})=0,} \tag{178}$$

and

$$\boxed{\boldsymbol{\nabla}\times(\boldsymbol{\nabla}f)=0.} \tag{179}$$

The divergence of a curl and the curl of a gradient vanish identically (assuming all those derivatives exist). These can be proved by expanding out in cartesian coordinates, for instance

$$\begin{aligned}&\boldsymbol{\nabla}\times\boldsymbol{\nabla}f=\\&\hat{\boldsymbol{x}}\,(\partial_y\partial_z f-\partial_z\partial_y f)+\hat{\boldsymbol{y}}\,(\partial_z\partial_x f-\partial_x\partial_z f)+\hat{\boldsymbol{z}}\,(\partial_x\partial_y f-\partial_y\partial_x f)=0\end{aligned}$$

by equality of the mixed partials $\partial_y\partial_z f = \partial_z\partial_y f$, etc. In cartesian index notation, $\boldsymbol{\nabla}\times\boldsymbol{\nabla}f = \mathbf{e}_i\,\epsilon_{ijk}\partial_j\partial_k f = -\boldsymbol{\nabla}\times\boldsymbol{\nabla}f$ and thus $\boldsymbol{\nabla}\times\boldsymbol{\nabla}f = 0$, since

$$\epsilon_{ijk}\partial_j\partial_k f = -\epsilon_{ikj}\partial_j\partial_k f = -\epsilon_{ijk}\partial_k\partial_j f = -\epsilon_{ijk}\partial_j\partial_k f. \tag{180}$$

Other fundamental identities are the *product rules* for divergence and curl:

$$\boxed{\begin{aligned}\boldsymbol{\nabla}\cdot(f\boldsymbol{V}) &= (\boldsymbol{\nabla}f)\cdot\boldsymbol{V}+f(\boldsymbol{\nabla}\cdot\boldsymbol{V}),\\ \boldsymbol{\nabla}\times(f\boldsymbol{V}) &= (\boldsymbol{\nabla}f)\times\boldsymbol{V}+f(\boldsymbol{\nabla}\times\boldsymbol{V}).\end{aligned}} \tag{181}$$

Both are easily proved using index notation,

$$\boldsymbol{\nabla}\cdot(f\boldsymbol{V}) = \partial_i(fv_i) = (\partial_i f)v_i + f(\partial_i v_i) = (\boldsymbol{\nabla}f)\cdot\boldsymbol{V}+f(\boldsymbol{\nabla}\cdot\boldsymbol{V}).$$

and

$$\begin{aligned}\boldsymbol{\nabla}\times(f\boldsymbol{V}) = \mathbf{e}_i\epsilon_{ijk}\partial_j(fv_k) &= \mathbf{e}_i\epsilon_{ijk}(\partial_j f)v_k + f\mathbf{e}_i\epsilon_{ijk}(\partial_j v_k)\\ &= (\boldsymbol{\nabla}f)\times\boldsymbol{V}+f(\boldsymbol{\nabla}\times\boldsymbol{V}).\end{aligned}$$

The following identity is fundamental in fluid mechanics and electromagnetism

$$\boxed{\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{V}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{V}) - \nabla^2\boldsymbol{V},} \tag{182}$$

where

$$\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \partial_1^2+\partial_2^2+\partial_3^2 \tag{183}$$

is the *Laplacian* operator. The identity (182) is easily reconstructed from the double cross product formula $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c}) = \boldsymbol{b}(\boldsymbol{a}\cdot\boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a}\cdot\boldsymbol{b})$ but note that the $\boldsymbol{\nabla}$ in the first term *must* appear first since $\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{V}) \neq (\boldsymbol{\nabla}\cdot\boldsymbol{V})\boldsymbol{\nabla}$. That identity can be verified using index notation if one recalls the double cross product identity in terms of the permutation tensor, namely

$$\epsilon_{ijk}\epsilon_{klm} = \epsilon_{kij}\epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}. \tag{184}$$

A slightly trickier identity is

$$\boxed{\boldsymbol{\nabla}\times(\boldsymbol{V}\times\boldsymbol{W}) = (\boldsymbol{W}\cdot\boldsymbol{\nabla})\boldsymbol{V} - (\boldsymbol{\nabla}\cdot\boldsymbol{V})\boldsymbol{W} + (\boldsymbol{\nabla}\cdot\boldsymbol{W})\boldsymbol{V} - (\boldsymbol{V}\cdot\boldsymbol{\nabla})\boldsymbol{W},} \tag{185}$$

where

$$(\boldsymbol{V}\cdot\boldsymbol{\nabla})\boldsymbol{W} = (v_j\partial_j)\boldsymbol{W} \equiv (v_1\partial_1+v_2\partial_2+v_3\partial_3)\boldsymbol{W}$$

and similarly for $\boldsymbol{W}\cdot\boldsymbol{\nabla}\boldsymbol{V}$. This identity can be verified using (184) and can be reconstructed from the double cross-product identity remembering that $\boldsymbol{\nabla}$ is a vector *operator*, not a regular vector, hence $\boldsymbol{\nabla}\times(\boldsymbol{V}\times\boldsymbol{W})$ represents derivatives of a product and this doubles the number of terms of the resulting expression. The first two terms are the double cross product $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})=(\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}-(\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}$ for derivatives of $\boldsymbol{V}$ while the last two terms are the double cross product for derivatives of $\boldsymbol{W}$.

Another similar identity is

$$\boxed{\boldsymbol{V}\times(\boldsymbol{\nabla}\times\boldsymbol{W})=(\boldsymbol{\nabla}\boldsymbol{W})\cdot\boldsymbol{V}-(\boldsymbol{V}\cdot\boldsymbol{\nabla})\boldsymbol{W}.}\tag{186}$$

In index notation this reads

$$\epsilon_{ijk}v_j\left(\epsilon_{klm}\partial_l w_m\right)=(\partial_i w_j)v_j-(v_j\partial_j)w_i.\tag{187}$$

Note that this last identity involves the gradient of a vector field $\boldsymbol{\nabla}\boldsymbol{W}$. This makes sense and is a *tensor*, *i.e.* a geometric object whose components with respect to a basis form a *matrix*. In indicial notation, the components of $\boldsymbol{\nabla}\boldsymbol{W}$ are $\partial_i w_j$ and there are 9 of them. This is very different from $\boldsymbol{\nabla}\cdot\boldsymbol{W}=\partial_i w_i$ which is a scalar.

Another fundamental identity is

$$\boxed{\boldsymbol{\nabla}\cdot(\boldsymbol{V}\times\boldsymbol{W})=\boldsymbol{W}\cdot(\boldsymbol{\nabla}\times\boldsymbol{V})-\boldsymbol{V}\cdot(\boldsymbol{\nabla}\times\boldsymbol{W}),}\tag{188}$$

that can again be reconstructed from the product rule together with the vector identity $\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})=(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c})$, and can be proved using cartesian coordinates.

The bottom line is that these identities can be reconstructed relatively easily from our knowledge of regular vector identities for dot, cross and double cross products, however $\boldsymbol{\nabla}$ is a vector of derivatives and one needs to watch out more carefully for the order and the product rules. If in doubt, jump to cartesian coordinates with index notation.

The combination of these fundamental vector identities with the fundamental examples of gradients often leads to shortcuts to computing divergence and curl. For example, computing the divergence and curl for the fundamental vector field of gravity and electrostatics,

$$\boldsymbol{F}=\frac{\hat{\boldsymbol{r}}}{r^2}=\frac{x\hat{\boldsymbol{x}}+y\hat{\boldsymbol{y}}+z\hat{\boldsymbol{z}}}{(x^2+y^2+z^2)^{3/2}},\tag{189}$$

can be handled relatively easily using cartesian coordinates, but much more elegantly using vector identities and fundamental gradients. Indeed, using (181) and (150) gives

$$\boldsymbol{\nabla}\cdot\frac{\hat{\boldsymbol{r}}}{r^2}=\boldsymbol{\nabla}\cdot\frac{\boldsymbol{r}}{r^3}=r^{-3}\left(\boldsymbol{\nabla}\cdot\boldsymbol{r}\right)+\left(\boldsymbol{\nabla}r^{-3}\right)\cdot\boldsymbol{r}=\frac{3}{r^3}-\frac{3}{r^4}\hat{\boldsymbol{r}}\cdot\boldsymbol{r}=0,\tag{190}$$

and (150) and (179) yield

$$\boldsymbol{\nabla}\times\frac{\hat{\boldsymbol{r}}}{r^2}=-\boldsymbol{\nabla}\times\boldsymbol{\nabla}r^{-1}=0.\tag{191}$$

This is a very special vector field indeed, whose divergence and curl vanish everywhere *except* at $r=0$ where the field is singular.

8.6 Grad, Div, Curl in cylindrical and spherical coordinates

Although the del operator is $\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\partial_x + \hat{\boldsymbol{y}}\partial_y + \hat{\boldsymbol{z}}\partial_z$ in cartesian coordinates it is *not* simply $\hat{\boldsymbol{r}}\partial_r + \hat{\boldsymbol{\theta}}\partial_\theta + \hat{\boldsymbol{\varphi}}\partial_\varphi$ in spherical coordinates! In fact that expression is not even dimensionally correct since ∂_r has units of inverse length but ∂_θ and ∂_φ have no units! To obtain the correct expression for $\boldsymbol{\nabla}$ in more general *curvilinear* coordinates u, v, w, we can start from the chain rule

$$\boldsymbol{\nabla} f(u,v,w) = \frac{\partial f}{\partial u}\boldsymbol{\nabla} u + \frac{\partial f}{\partial v}\boldsymbol{\nabla} v + \frac{\partial f}{\partial w}\boldsymbol{\nabla} w \tag{192}$$

and in particular for cylindrical coordinates

$$\begin{aligned}\boldsymbol{\nabla} f(\rho,\varphi,z) &= \frac{\partial f}{\partial \rho}\boldsymbol{\nabla}\rho + \frac{\partial f}{\partial \varphi}\boldsymbol{\nabla}\varphi + \frac{\partial f}{\partial z}\boldsymbol{\nabla} z \\ &= \hat{\boldsymbol{\rho}}\frac{\partial f}{\partial \rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial f}{\partial \varphi} + \hat{\boldsymbol{z}}\frac{\partial f}{\partial z}\end{aligned} \tag{193}$$

since we figured out earlier that

$$\boldsymbol{\nabla}\rho = \hat{\boldsymbol{\rho}}, \quad \boldsymbol{\nabla}\varphi = \frac{1}{\rho}\hat{\boldsymbol{\varphi}}, \quad \boldsymbol{\nabla} z = \hat{\boldsymbol{z}}. \tag{194}$$

Likewise for spherical coordinates,

$$\begin{aligned}\boldsymbol{\nabla} f(r,\theta,\varphi) &= \frac{\partial f}{\partial r}\boldsymbol{\nabla} r + \frac{\partial f}{\partial \theta}\boldsymbol{\nabla}\theta + \frac{\partial f}{\partial \varphi}\boldsymbol{\nabla}\varphi \\ &= \hat{\boldsymbol{r}}\frac{\partial f}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial f}{\partial \theta} + \hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial f}{\partial \varphi}\end{aligned} \tag{195}$$

since we figured out earlier that

$$\boldsymbol{\nabla} r = \hat{\boldsymbol{r}}, \quad \boldsymbol{\nabla}\theta = \frac{1}{r}\hat{\boldsymbol{\theta}}, \quad \boldsymbol{\nabla}\varphi = \frac{1}{r\sin\theta}\hat{\boldsymbol{\varphi}}. \tag{196}$$

These yield the following expressions for the del operator $\boldsymbol{\nabla}$ in cartesian, cylindrical and spherical coordinates, respectively,

$$\begin{aligned}\boldsymbol{\nabla} &= \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z} \\ &= \hat{\boldsymbol{\rho}}\frac{\partial}{\partial \rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial}{\partial \varphi} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z} \\ &= \hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \varphi}\end{aligned} \tag{197}$$

where

- $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ are fixed mutually orthogonal cartesian unit vectors,
- $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{z}})$ are mutually orthogonal cylindrical unit vectors but $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\varphi}}$ depend on φ,

- $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varphi}})$ are mutually orthogonal spherical unit vectors but $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ depend on both θ and φ.

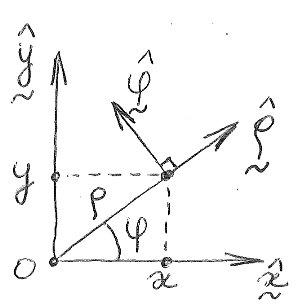
Note also that this is why we write the direction vectors $\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}$ in front of the partials in the del operator (197) since they vary with the coordinates. The cylindrical and spherical direction vectors obey the following relations

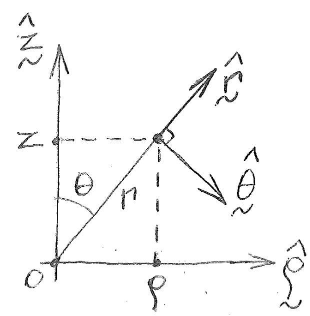
$$
\begin{aligned}
&\frac{\partial}{\partial \varphi}\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\varphi}}, \qquad && \frac{\partial}{\partial \varphi}\hat{\boldsymbol{\varphi}} = -\hat{\boldsymbol{\rho}} \\
&\frac{\partial}{\partial \varphi}\hat{\boldsymbol{r}} = \sin\theta\ \hat{\boldsymbol{\varphi}}, \qquad && \frac{\partial}{\partial \varphi}\hat{\boldsymbol{\theta}} = \cos\theta\ \hat{\boldsymbol{\varphi}} \\
&\frac{\partial}{\partial \theta}\hat{\boldsymbol{r}} = \hat{\boldsymbol{\theta}}, \qquad && \frac{\partial}{\partial \theta}\hat{\boldsymbol{\theta}} = -\hat{\boldsymbol{r}}
\end{aligned}
\tag{198}
$$

See Chapter 1, section 1.3 to verify these using the hybrid formulation, for example

$$
\left\{
\begin{array}{rcr}
\hat{\boldsymbol{\rho}} & = & \cos\varphi\,\hat{\boldsymbol{x}} + \sin\varphi\,\hat{\boldsymbol{y}} \\
\hat{\boldsymbol{\varphi}} & = & -\sin\varphi\,\hat{\boldsymbol{x}} + \cos\varphi\,\hat{\boldsymbol{y}}
\end{array}
\right.
\Rightarrow \ \frac{\partial \hat{\boldsymbol{\rho}}}{\partial \varphi} = \hat{\boldsymbol{\varphi}},
$$

but learn also to re-derive these relationships geometrically as all mathematical physicists know how to do from the following figures:





With these relations for the rates of change of the direction vectors with the coordinates, one can derive the expression for div, curl, Laplacian, etc in cylindrical and spherical coordinates.

Example: $\boldsymbol{\nabla} \cdot (v\hat{\boldsymbol{\theta}})$ using (197) and (198)

$$
\begin{aligned}
\boldsymbol{\nabla} \cdot (v\hat{\boldsymbol{\theta}}) &= (\boldsymbol{\nabla} v) \cdot \hat{\boldsymbol{\theta}} + v\ (\boldsymbol{\nabla} \cdot \hat{\boldsymbol{\theta}}) \\
&= \frac{1}{r}\frac{\partial v}{\partial \theta} + v\left(\hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \varphi}\right) \cdot \hat{\boldsymbol{\theta}} \\
&= \frac{1}{r}\frac{\partial v}{\partial \theta} + \frac{v}{r}\,\hat{\boldsymbol{\theta}} \cdot \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta} + \frac{v}{r\sin\theta}\,\hat{\boldsymbol{\varphi}} \cdot \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \varphi} \\
&= \frac{1}{r}\frac{\partial v}{\partial \theta} + \frac{v}{r\tan\theta} = \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(v\sin\theta)
\end{aligned}
\tag{199}
$$

□

Example: $\boldsymbol{\nabla}\cdot(v\hat{\boldsymbol{\theta}})$ using $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\varphi}}\times\hat{\boldsymbol{r}}=r\sin\theta\,\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r$ and vector identies:

$$\begin{aligned}\boldsymbol{\nabla}\cdot(v\hat{\boldsymbol{\theta}})&=\boldsymbol{\nabla}\cdot(rv\sin\theta\,\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r)\\&=\boldsymbol{\nabla}(rv\sin\theta)\cdot(\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r)+rv\sin\theta\,\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r)\\&=\boldsymbol{\nabla}(rv\sin\theta)\cdot\frac{1}{r\sin\theta}\hat{\boldsymbol{\theta}}=\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(v\sin\theta)\end{aligned}\tag{200}$$

since $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r)=\boldsymbol{\nabla}r\cdot(\boldsymbol{\nabla}\times\boldsymbol{\nabla}\varphi)-\boldsymbol{\nabla}\varphi\cdot(\boldsymbol{\nabla}\times\boldsymbol{\nabla}r)=0-0.$ □

Example: $\boldsymbol{\nabla}\times(v\hat{\boldsymbol{\theta}})$ using $\hat{\boldsymbol{\theta}}=r\boldsymbol{\nabla}\theta$ and vector identities:

$$\begin{aligned}\boldsymbol{\nabla}\times(v\hat{\boldsymbol{\theta}})&=\boldsymbol{\nabla}\times(rv\boldsymbol{\nabla}\theta)\\&=\boldsymbol{\nabla}(rv)\times\boldsymbol{\nabla}\theta+rv\boldsymbol{\nabla}\times\boldsymbol{\nabla}\theta\\&=\frac{1}{r}\frac{\partial(rv)}{\partial r}(\hat{\boldsymbol{r}}\times\hat{\boldsymbol{\theta}})+\frac{1}{r^2\sin\theta}\frac{\partial(rv)}{\partial\varphi}\hat{\boldsymbol{\varphi}}\times\hat{\boldsymbol{\theta}}\\&=\hat{\boldsymbol{\varphi}}\frac{1}{r}\frac{\partial}{\partial r}(rv)-\hat{\boldsymbol{r}}\frac{1}{r\sin\theta}\frac{\partial v}{\partial\varphi}.\end{aligned}\tag{201}$$

□

In that second approach, we use $\hat{\boldsymbol{\theta}}=r\sin\theta\boldsymbol{\nabla}\varphi\times\boldsymbol{\nabla}r$ when computing a divergence, but $\hat{\boldsymbol{\theta}}=r\boldsymbol{\nabla}\theta$ when computing a curl since that allows useful vector identities.

Proceeding with either method leads to the general formula for divergence in cylindrical coordinates

$$\boxed{\boldsymbol{\nabla}\cdot(u\hat{\boldsymbol{\rho}}+v\hat{\boldsymbol{\varphi}}+w\hat{\boldsymbol{z}})=\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho u)+\frac{1}{\rho}\frac{\partial}{\partial\varphi}v+\frac{\partial}{\partial z}w}\tag{202}$$

and in spherical coordinates

$$\boxed{\boldsymbol{\nabla}\cdot(u\hat{\boldsymbol{r}}+v\hat{\boldsymbol{\theta}}+w\hat{\boldsymbol{\varphi}})=\frac{1}{r^2}\frac{\partial}{\partial r}(r^2u)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(v\sin\theta)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}w,}\tag{203}$$

and similarly for the curl. Note that the meaning of (u,v,w) changes with the coordinates. We could use the notation $\boldsymbol{V}=v_\rho\hat{\boldsymbol{\rho}}+v_\varphi\hat{\boldsymbol{\varphi}}+v_z\hat{\boldsymbol{z}}$ and $\boldsymbol{V}=v_r\hat{\boldsymbol{r}}+v_\theta\hat{\boldsymbol{\theta}}+v_\varphi\hat{\boldsymbol{\varphi}}$ but using (u,v,w) for the components in the corresponding directions is lighter if there is no confusion.

The divergence of a gradient is the Laplacian, $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}=\nabla^2$, applying (202) to the gradient in cylindrical coordinates (197) yields the Laplacian in cylindrical coordinates

$$\nabla^2 f=\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}f=\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho\frac{\partial f}{\partial\rho})+\frac{1}{\rho^2}\frac{\partial^2 f}{\partial\varphi^2}+\frac{\partial^2 f}{\partial z^2}.\tag{204}$$

Likewise, applying (203) to the gradient in spherical coordinates (197) yields the Laplacian in spherical coordinates

$$\begin{aligned}&\nabla^2 f=\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}f=\\&\qquad\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial f}{\partial r})+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\frac{\sin\theta}{r}\frac{\partial f}{\partial\theta}\right)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}\left(\frac{1}{r\sin\theta}\frac{\partial f}{\partial\varphi}\right)\end{aligned}\tag{205}$$

For further formulae, the following wikipedia page is quite complete:
`https://en.wikipedia.org/wiki/Del_in_cylindrical_and_spherical_coordinates`

Exercises

1. $\boldsymbol{\nabla} f$ is the gradient of f at point P with $\boldsymbol{r} = \overrightarrow{OP}$, calculate $\boldsymbol{\nabla} f$ briefly justifying your work, for (i) $f(P) = r \sin\theta$ where r is distance to the origin and θ is the angle between $\boldsymbol{r}$ and $\hat{\boldsymbol{z}}$ (think geometrically); (ii) $f(P)$ is the sum of the distances from point P to two fixed points F_1 and F_2. Prove that the angle between $\boldsymbol{\nabla} f$ and $\overrightarrow{F_1P}$ is the same as the angle between $\boldsymbol{\nabla} f$ and $\overrightarrow{F_2P}$, for any P. What are the isosurfaces of f?

2. Calculate $\boldsymbol{\nabla} f$ for $f(\boldsymbol{r}) = \dfrac{A}{|\boldsymbol{r}-\boldsymbol{a}|} + \dfrac{B}{|\boldsymbol{r}-\boldsymbol{b}|}$ where $A, B, \boldsymbol{a}, \boldsymbol{b}$ are constants. (i) Using vector identities and the geometric concept of gradient, (ii) in cartesian x, y, z notation.

3. For the region $\mathcal{R}$ in a plane that is defined in cartesian coordinates (x, y) by $0 < u_1 \leq xy \leq u_2$ and $0 < v_1 \leq y/x \leq v_2$, consider the coordinates $u = xy$ and $v = y/x$. (i) Sketch the u and v coordinate curves. Highlight $\mathcal{R}$. (ii) Find and sketch $\partial\boldsymbol{r}/\partial u$, $\partial\boldsymbol{r}/\partial v$, $\boldsymbol{\nabla} u$ and $\boldsymbol{\nabla} v$. Are these orthogonal coordinates? (iii) Calculate
$$\frac{\partial\boldsymbol{r}}{\partial u}\cdot\boldsymbol{\nabla} u, \quad \frac{\partial\boldsymbol{r}}{\partial u}\cdot\boldsymbol{\nabla} v, \quad \frac{\partial\boldsymbol{r}}{\partial v}\cdot\boldsymbol{\nabla} u, \quad \frac{\partial\boldsymbol{r}}{\partial v}\cdot\boldsymbol{\nabla} v.$$

4. Let θ be the polar angle. Calculate $\partial_x\theta$. (In general, we would need to specify what is fixed when we compute the partial derivative, but we'll assume here naturally that ∂_x means (y, z) fixed)

5. Consider the bipolar coordinates $u = \ln(r_1/r_2)$, $v = \theta_2 - \theta_1 = \alpha$ (see fig. in Chap. 1). Calculate $\boldsymbol{\nabla} u$, $\boldsymbol{\nabla} v$ and $\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v$.

6. Prove that $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{V}) = 0$ and $\boldsymbol{\nabla}\times\boldsymbol{\nabla} f = 0$ for any sufficiently differentiable f and $\boldsymbol{V}$.

7. Show that $\boldsymbol{\nabla} f\cdot(\boldsymbol{\nabla}\times\boldsymbol{V}) \neq 0$ in general. How does this differ from (178)?

8. Show that $\boldsymbol{\nabla} f\times\boldsymbol{\nabla} g \neq 0$ in general. How does this differ from (179)?

9. Digest and verify the identity (184) *ab initio*.

10. Verify (185) and (186) using index notation and (184).

11. Use (184) to derive vector identities for $(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{c}\times\boldsymbol{d})$ and $(\boldsymbol{\nabla}\times\boldsymbol{V})\cdot(\boldsymbol{\nabla}\times\boldsymbol{W})$.

12. Show that $\boldsymbol{\nabla}\cdot(\boldsymbol{V}\times\boldsymbol{W}) = \boldsymbol{W}\cdot(\boldsymbol{\nabla}\times\boldsymbol{V}) - \boldsymbol{V}\cdot(\boldsymbol{\nabla}\times\boldsymbol{W})$ and explain how to reconstruct this from the rules for the mixed (or box) product $\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})$ of three regular vectors.

13. Find the fastest way to show that $\boldsymbol{\nabla}\cdot\boldsymbol{r} = 3$ and $\boldsymbol{\nabla}\times\boldsymbol{r} = 0$.

14. Show that $\boldsymbol{\nabla}\cdot(\hat{\boldsymbol{r}}/r^2)=0$ and $\boldsymbol{\nabla}\times(\hat{\boldsymbol{r}}/r^2)=0$ for all $\boldsymbol{r}\neq 0$ using (1) vector identities and a product rule, (2) cartesian coordinates, (3) spherical coordinates.

15. Show that $\boldsymbol{\nabla}\cdot\hat{\boldsymbol{r}}=2/r$ using (1) vector identities and the product rule, (2) cartesian coordinates, (3) spherical coordinates.

16. Show that $\boldsymbol{\nabla}\cdot\hat{\boldsymbol{\rho}}=1/\rho$ using (1) vector identities and the product rule, (2) cartesian coordinates, (3) cylindrical coordinates.

17. Consider $\Phi=|\boldsymbol{r}-\boldsymbol{r}_{_A}|^{-1}$ where $\boldsymbol{r}_{_A}$ is constant. Calculate $\boldsymbol{\nabla}\Phi$ and show that $\nabla^2\Phi=0$. This is the potential of an electric charge in electrostatics, or of a *source* in aerodynamics.

18. Consider $\Phi=\boldsymbol{a}\cdot\boldsymbol{r}/r^3$ with $\boldsymbol{a}$ constant. Calculate $\boldsymbol{\nabla}\Phi$ and show that $\nabla^2\Phi=0$ for $r\neq 0$. This is the potential of an electric *dipole* in electrostatics, or of a *doublet* in aerodynamics.

19. Consider $\boldsymbol{V}=\boldsymbol{a}\times\boldsymbol{r}$ with $\boldsymbol{a}$ constant. Show that $\boldsymbol{\nabla}\cdot\boldsymbol{V}=0$ and $\boldsymbol{\nabla}\times\boldsymbol{V}=2\boldsymbol{a}$ using (1) vector identities, (2) cartesian coordinates.

20. Consider $\boldsymbol{A}=\boldsymbol{a}\times\boldsymbol{r}/r^3$ with $\boldsymbol{a}$ constant. Calculate $\boldsymbol{\nabla}\cdot\boldsymbol{A}$ and $\boldsymbol{\nabla}\times\boldsymbol{A}$ and show that $\nabla^2\boldsymbol{A}=0$ for $r\neq 0$. This is the magnetic potential of a magnetic *dipole* in magnetostatics.

21. Show that $\hat{\boldsymbol{z}}\times\boldsymbol{r}=\hat{\boldsymbol{z}}\times\boldsymbol{\rho}=-y\hat{\boldsymbol{x}}+x\hat{\boldsymbol{y}}=\rho\hat{\boldsymbol{\varphi}}=\rho^2\boldsymbol{\nabla}\varphi$.

22. Show that $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}u\times\boldsymbol{\nabla}v)=0$ using vector identities.

23. Calculate $\boldsymbol{\nabla}\cdot(v\hat{\boldsymbol{\varphi}})$ and $\boldsymbol{\nabla}\times(v\hat{\boldsymbol{\varphi}})$ in cylindrical and spherical coordinates using methods illustrated in sect. 8.6.

24. Consider $\boldsymbol{B}=(\hat{\boldsymbol{z}}\times\boldsymbol{r})/|\hat{\boldsymbol{z}}\times\boldsymbol{r}|^2$, show that $\boldsymbol{\nabla}\cdot\boldsymbol{B}=0$ and $\boldsymbol{\nabla}\times\boldsymbol{B}=0$ everywhere *except* on the z-axis where $\rho=0$ (1) using vector identities and fundamental gradients, (2) cylindrical coordinates.

25. Calculate $\boldsymbol{\nabla}\cdot\boldsymbol{V}$ and $\boldsymbol{\nabla}\times\boldsymbol{V}$ for $\boldsymbol{V}=A\,(\boldsymbol{r}-\boldsymbol{a})/|\boldsymbol{r}-\boldsymbol{a}|^3+B\,(\boldsymbol{r}-\boldsymbol{b})/|\boldsymbol{r}-\boldsymbol{b}|^3$ where $A,B,\boldsymbol{a},\boldsymbol{b}$ are constants using vector identities and fundamental gradients.

26. Write the following fields in cartesian, cylindrical and spherical representations. Sketch the vector fields and compute their divergence $\boldsymbol{\nabla}\cdot\boldsymbol{V}$ and curl $\boldsymbol{\nabla}\times\boldsymbol{V}$.

    (a) $\boldsymbol{V}=\alpha\,\boldsymbol{r}+\boldsymbol{\omega}\times\boldsymbol{r}$, with α and $\boldsymbol{\omega}$ constants.
    (b) $\boldsymbol{V}=Sy\,\hat{\boldsymbol{x}}$, where S is constant (shear flow)
    (c) $\boldsymbol{V}=\alpha(x\hat{\boldsymbol{x}}-y\hat{\boldsymbol{y}})$ with α constant (stagnation point flow)
    (d) $\boldsymbol{E}=\hat{\boldsymbol{r}}/r^2$, (inverse square law in gravity and electrostatics)
    (e) $\boldsymbol{B}=(\hat{\boldsymbol{z}}\times\boldsymbol{r})/|\hat{\boldsymbol{z}}\times\boldsymbol{r}|^2$, (Magnetic field of a line current)

27. Calculate the divergence and curl of $\boldsymbol{V}=\boldsymbol{A}\cdot\boldsymbol{r}=a_{ij}x_j\mathbf{e}_i$ where $\boldsymbol{A}\equiv[a_{ij}]$ are constants (general linear flow).

28. Calculate $\boldsymbol{\nabla}f$ and $\nabla^2 f$ for $f=\boldsymbol{r}\cdot\boldsymbol{A}\cdot\boldsymbol{r}=a_{ij}x_ix_j$ where a_{ij} are constants.

9 Inverting Grad, Div, Curl (optional)

Scalar and Vector Potentials

The following three problems reduce to Poisson's equation.

1. Find a scalar potential Φ such that $\boldsymbol{\nabla}\Phi = \boldsymbol{F}$ given the vector field $\boldsymbol{F}(\boldsymbol{r})$.
2. Find a vector potential $\boldsymbol{U}$ such that $\boldsymbol{\nabla} \cdot \boldsymbol{U} = f$ given the scalar field $f(\boldsymbol{r})$.
3. Find a vector potential $\boldsymbol{U}$ such that $\boldsymbol{\nabla} \times \boldsymbol{U} = \boldsymbol{F}$ given the vector field $\boldsymbol{F}(\boldsymbol{r})$.

Problem 1 has a solution Φ only if $\boldsymbol{\nabla} \times \boldsymbol{F} = 0$ since $\boldsymbol{\nabla} \times \boldsymbol{\nabla}\Phi = 0$. Then Φ is obtained from Poisson's equation

$$\boldsymbol{\nabla} \cdot \boldsymbol{\nabla}\Phi = \nabla^2\Phi = \boldsymbol{\nabla} \cdot \boldsymbol{F}. \tag{206}$$

Problem 2 has an infinite number of solutions. If $\boldsymbol{U}$ is a solution, then so is $\boldsymbol{U} + \boldsymbol{\nabla} \times \boldsymbol{F}$ for any $\boldsymbol{F}$ since $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \boldsymbol{F} = 0$. Thus the curl of $\boldsymbol{U}$ is arbitrary. If we set $\boldsymbol{\nabla} \times \boldsymbol{U} = 0$, then $\boldsymbol{U} = \boldsymbol{\nabla}\Phi$ and

$$\boldsymbol{\nabla} \cdot \boldsymbol{U} = \nabla^2\Phi = f. \tag{207}$$

So Φ, then $\boldsymbol{U}$, also follows from Poissson's equation.

Problem 3 also has an infinite number of solutions, but a solution $\boldsymbol{U}$ exists only if $\boldsymbol{\nabla} \cdot \boldsymbol{F} = 0$ since $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \boldsymbol{U} = 0$. If $\boldsymbol{U}$ is a solution, then so is $\boldsymbol{U} + \boldsymbol{\nabla}\Psi$ since $\boldsymbol{\nabla} \times \boldsymbol{\nabla}\Psi = 0$ for any scalar function Ψ. Thus the divergence of $\boldsymbol{U}$ is arbitrary and we can pick $\boldsymbol{\nabla} \cdot \boldsymbol{U} = 0$ since given a $\boldsymbol{U}$ we can always find a Ψ such that $\boldsymbol{\nabla} \cdot (\boldsymbol{U} + \boldsymbol{\nabla}\Psi) = 0$ by solving the Poisson equation $\nabla^2\Psi = -\boldsymbol{\nabla} \cdot \boldsymbol{U}$.

Taking the curl of the equation $\boldsymbol{\nabla} \times \boldsymbol{U} = \boldsymbol{F}$ yields

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{U}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{U}) - \nabla^2\boldsymbol{U} = \boldsymbol{\nabla} \times \boldsymbol{F}.$$

For $\boldsymbol{\nabla} \cdot \boldsymbol{U} = 0$ this reduces to

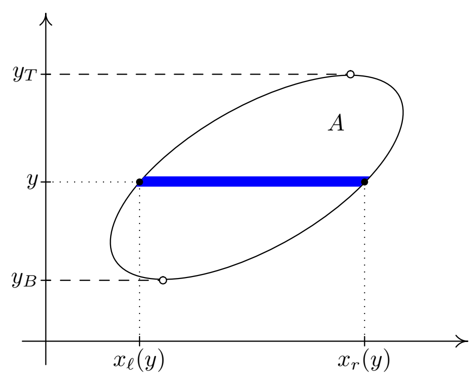
$$-\nabla^2\boldsymbol{U} = \boldsymbol{\nabla} \times \boldsymbol{F}. \tag{208}$$

This is a vector Poisson equation for $\boldsymbol{U}$.

10 Fundamental theorems of vector calculus

10.1 Integration in $\mathbb{R}^2$ and $\mathbb{R}^3$

The integral of a function of two variables $f(x,y)$ over a domain A of $\mathbb{R}^2$ denoted $\int_A f(x,y)dA$ can be defined as a the limit of a $\sum_n f(x_n,y_n)\Delta A_n$ where the A_n's, $n=1,\ldots,N$ provide an (approximate) partition of A that breaks up A into a set of small area elements, squares or triangles for instance. ΔA_n is the area of those element n and (x_n,y_n) is a point inside that element, for instance the center of area of the triangle. The integral would be the limit of such sums when the area of the triangles goes to zero and their number N must then go to infinity. This limit should be such that the aspect ratios of the triangles remain bounded away from 0 so we get a finer and finer sampling of A. This definition also provides a way to approximate the integral by such a finite sum.



If we imagine breaking up A into small squares aligned with the x and y axes then the sum over all squares inside A can be performed row by row. Each row-sum then tends to an integral in the x-direction, this leads to the conclusion that the integral can be calculated as *iterated* integrals

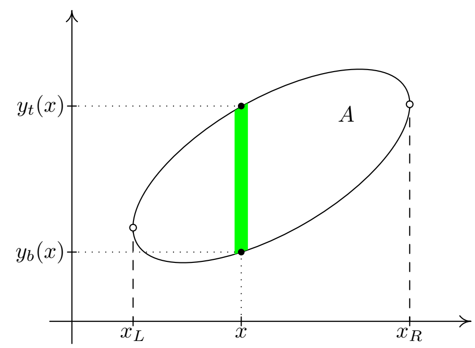
$$\int_A f(x,y)dA = \int_{y_B}^{y_T} dy \int_{x_\ell(y)}^{x_r(y)} f(x,y)dx \tag{209}$$

We can also imagine summing up column by column instead and each column-sum then tends to an integral in the y-direction, this leads to the *iterated* integrals

$$\int_A f(x,y)dA = \int_{x_L}^{x_R} dx \int_{y_b(x)}^{y_t(x)} f(x,y)dy. \tag{210}$$

Note of course that the limits of integrations differ from those of the previous iterated integrals.

This iterated integral approach readily extends to integrals over three-dimensional domains in $\mathbb{R}^3$ and more generally to integrals in $\mathbb{R}^n$.



10.2 Fundamental theorem of Calculus

The fundamental theorem of calculus can be written

$$\boxed{\int_a^b \frac{dF}{dx}dx = F(b) - F(a).} \tag{211}$$

Once again we can interpret this in terms of limits of finite differences. The derivative is defined as

$$\frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} \tag{212}$$

where $\Delta F = F(x + \Delta x) - F(x)$, while the integral

$$\int_a^b f(x)dx = \lim_{\Delta x_n \to 0} \sum_{n=1}^{N} f(\tilde{x}_n)\Delta x_n \tag{213}$$

where $\Delta x_n = x_n - x_{n-1}$ and $x_{n-1} \le \tilde{x}_n \le x_n$, with $n = 1, \ldots, N$ and $x_0 = a$, $x_N = b$, so the set of x_n's provides a partition of the interval $[a, b]$. The best choice for $\tilde{x}_n$ is the midpoint $\tilde{x}_n = (x_n + x_{n-1})/2$. This is the *midpoint* scheme in numerical integration methods. Putting these two limits together and setting $\Delta F_n = F(x_n) - F(x_{n-1})$ we can write

$$\int_a^b \frac{dF}{dx}dx = \lim_{\Delta x_n \to 0} \sum_{n=1}^{N} \frac{\Delta F_n}{\Delta x_n}\,\Delta x_n = \lim_{\Delta x_n \to 0} \sum_{n=1}^{N} \Delta F_n = F(b) - F(a). \tag{214}$$

We can also write this in the integral form

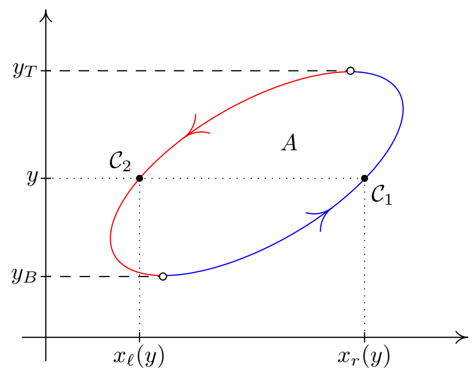
$$\int_a^b \frac{dF}{dx}dx = \int_{F(a)}^{F(b)} dF = F(b) - F(a). \tag{215}$$

10.3 Fundamental theorem in $\mathbb{R}^2$

From the fundamental theorem of calculus and the reduction of integrals on a domain A of $\mathbb{R}^2$ to iterated integrals on intervals in $\mathbb{R}$ we obtain for a function $G(x, y)$

$$\int_A \frac{\partial G}{\partial x}dA = \int_{y_B}^{y_T} dy \int_{x_\ell(y)}^{x_r(y)} \frac{\partial G}{\partial x}dx = \int_{y_B}^{y_T} \big[G(x_r(y), y) - G(x_\ell(y), y)\big]dy. \tag{216}$$

This looks nice enough but we can rewrite the integral on the right-hand side as a line integral over the boundary of A. The boundary of A is a closed curve $\mathcal{C}$ often denoted ∂A (not to be confused with a partial derivative). The boundary $\mathcal{C}$ has two parts $\mathcal{C}_1$ and $\mathcal{C}_2$.



The curve $\mathcal{C}_1$ can be parametrized in terms of y as $\boldsymbol{r}(y) = \hat{\boldsymbol{x}} x_r(y) + \hat{\boldsymbol{y}} y$ with $y = y_B \to y_T$, hence

$$\int_{\mathcal{C}_1} G(x,y)\hat{\boldsymbol{y}} \cdot d\boldsymbol{r} = \int_{y_B}^{y_T} G(x_r(y), y) dy.$$

Likewise, the curve $\mathcal{C}_2$ can be parametrized using y as $\boldsymbol{r}(y) = \hat{\boldsymbol{x}} x_\ell(y) + \hat{\boldsymbol{y}} y$ with $y = y_T \to y_B$, hence

$$\int_{\mathcal{C}_2} G(x,y)\hat{\boldsymbol{y}} \cdot d\boldsymbol{r} = \int_{y_T}^{y_B} G(x_\ell(y), y) dy.$$

Putting these two results together the right hand side of (216) becomes

$$\int_{y_B}^{y_T} \big[G(x_r(y), y) - G(x_\ell(y), y)\big] dy = \int_{\mathcal{C}_1} G\,\hat{\boldsymbol{y}} \cdot d\boldsymbol{r} + \int_{\mathcal{C}_2} G\,\hat{\boldsymbol{y}} \cdot d\boldsymbol{r} = \oint_{\mathcal{C}} G\,\hat{\boldsymbol{y}} \cdot d\boldsymbol{r},$$

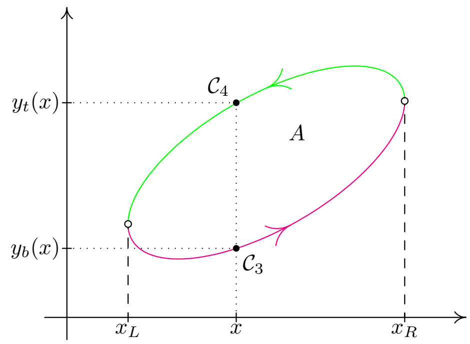
where $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$ is the *closed* curve bounding A. Then (216) becomes

$$\int_A \frac{\partial G}{\partial x}\, dA = \oint_{\mathcal{C}} G\,\hat{\boldsymbol{y}} \cdot d\boldsymbol{r}. \tag{217}$$

The symbol $\oint$ is used to emphasize that the integral is over a closed curve. Note that the curve $\mathcal{C}$ has been oriented *counter-clockwise* such that the interior is to the left of the curve.

Similarly the fundamental theorem of calculus and iterated integrals lead to the result that

$$\int_A \frac{\partial F}{\partial y} dA = \int_{x_L}^{x_R} dx \int_{y_b(x)}^{y_t(x)} \frac{\partial F}{\partial y} dy = \int_{x_L}^{x_R} \big[F(x, y_t(x)) - F(x, y_b(x))\big] dx, \tag{218}$$



and the integral on the right hand side can be rewritten as a line integral around the boundary curve $\mathcal{C} = \mathcal{C}_3 + \mathcal{C}_4$.

The curve $\mathcal{C}_3$ can be parametrized in terms of x as $\boldsymbol{r}(x) = \hat{\boldsymbol{x}}x + \hat{\boldsymbol{y}}\, y_b(x)$ with $x = x_L \to x_R$, hence

$$\int_{\mathcal{C}_3} F(x,y)\hat{\boldsymbol{x}} \cdot d\boldsymbol{r} = \int_{x_L}^{x_R} F(x, y_b(x))dx.$$

Likewise, the curve $\mathcal{C}_4$ can be parametrized using x as $\boldsymbol{r}(x) = \hat{\boldsymbol{x}}x + \hat{\boldsymbol{y}}\, y_t(x)$ with $x = x_R \to x_L$, hence

$$\int_{\mathcal{C}_4} F(x,y)\hat{\boldsymbol{x}} \cdot d\boldsymbol{r} = \int_{x_R}^{x_L} F(x, y_t(x))dx.$$

The right hand side of (218) becomes

$$\int_{x_L}^{x_R} \big[F(x,y_t(x)) - F(x,y_b(x))\big]dx = -\int_{\mathcal{C}_4} F\hat{\boldsymbol{x}} \cdot d\boldsymbol{r} - \int_{\mathcal{C}_3} F\hat{\boldsymbol{x}} \cdot d\boldsymbol{r} = -\oint_{\mathcal{C}} F\hat{\boldsymbol{x}} \cdot d\boldsymbol{r},$$

where $\mathcal{C} = \mathcal{C}_3 + \mathcal{C}_4$ is the *closed* curve bounding A oriented *counter-clockwise* as before. Then (218) becomes

$$\int_A \frac{\partial F}{\partial y}\, dA = -\oint_{\mathcal{C}} F\hat{\boldsymbol{x}} \cdot d\boldsymbol{r}. \tag{219}$$

10.4 Green and Stokes' theorems

The two results (217) and (219) can be combined into a single important formula. Subtract (219) from (217) to deduce the *curl form of Green's theorem*

$$\boxed{\int_A \left(\frac{\partial G}{\partial x} - \frac{\partial F}{\partial y}\right) dA = \oint_{\mathcal{C}} (F\hat{\boldsymbol{x}} + G\hat{\boldsymbol{y}}) \cdot d\boldsymbol{r} = \oint_{\mathcal{C}} (F dx + G dy).} \tag{220}$$

Note that dx and dy in the last integral are *not* independent quantities, they are the projection of the line element $d\boldsymbol{r}$ onto the basis vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ as written in the middle line integral. If $(x(t), y(t))$ is a parametrization for the curve then $dx = (dx/dt)dt$

and $dy = (dy/dt)dt$ and the t-bounds of integration should be picked to correspond to counter-clockwise orientation. Note also that Green's theorem (220) is the formula to remember since it includes both (217) when $F = 0$ and (219) when $G = 0$.

Green's theorem can be written in several equivalent forms. Define the vector field $\boldsymbol{V} = F(x,y)\hat{\boldsymbol{x}} + G(x,y)\hat{\boldsymbol{y}}$. A simple calculation verifies that its curl is purely in the $\hat{\boldsymbol{z}}$ direction indeed (174) gives $\boldsymbol{\nabla}\times\boldsymbol{V} = \hat{\boldsymbol{z}}\,(\partial G/\partial x - \partial F/\partial y)$ thus (220) can be rewritten in the form

$$\int_A (\boldsymbol{\nabla}\times\boldsymbol{V})\cdot\hat{\boldsymbol{z}}\,dA = \oint_{\mathcal{C}} \boldsymbol{V}\cdot d\boldsymbol{r}, \tag{221}$$

This result also applies to any 3D vector field $\boldsymbol{V}(x,y,z) = F(x,y,z)\hat{\boldsymbol{x}} + G(x,y,z)\hat{\boldsymbol{y}} + H(x,y,z)\hat{\boldsymbol{z}}$ and any planar surface A perpendicular to $\hat{\boldsymbol{z}}$ since $\hat{\boldsymbol{z}}\cdot(\boldsymbol{\nabla}\times\boldsymbol{V})$ still equals $\partial G/\partial x - \partial F/\partial y$ for such 3D vector fields and the line element $d\boldsymbol{r}$ of the boundary curve $\mathcal{C}$ of such planar area is perpendicular to $\hat{\boldsymbol{z}}$ so $\boldsymbol{V}\cdot d\boldsymbol{r}$ is still equal to $F(x,y,z)dx + G(x,y,z)dy$. The extra z coordinate is a mere parameter for the integrals and (221) applies equally well to 3D vector field $\boldsymbol{V}(x,y,z)$ provided A is a planar area perpendicular to $\hat{\boldsymbol{z}}$.

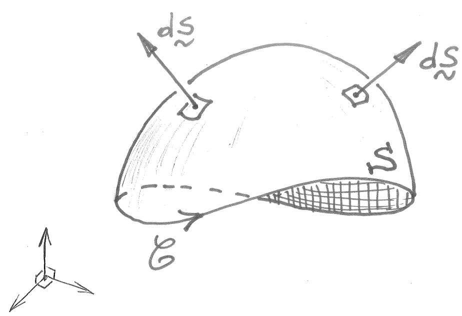


Fig. 2.16: Surface S with boundary $\mathcal{C}$ and right hand rule orientations for Stokes' Theorem.

In fact that last restriction on A itself can be removed. This is *Stokes' theorem* which reads

$$\boxed{\int_S (\boldsymbol{\nabla}\times\boldsymbol{V})\cdot d\boldsymbol{S} = \oint_{\mathcal{C}} \boldsymbol{V}\cdot d\boldsymbol{r},} \tag{222}$$

where S is a bounded orientable surface in 3D space, not necessarily planar, and $\mathcal{C}$ is its closed curve boundary. The orientation of the surface as determined by the direction of its normal $\hat{\boldsymbol{n}}$, where $d\boldsymbol{S} = \hat{\boldsymbol{n}}dS$, and the orientation of the boundary curve $\mathcal{C}$ must obey the right-hand rule. Thus a corkscrew turning in the direction of $\mathcal{C}$ would go through S in the direction of its normal $\hat{\boldsymbol{n}}$. That restriction is a direct consequence of the fact that the right hand rule enters the definition of the curl as the cross product $\boldsymbol{\nabla}\times\boldsymbol{V}$.

Stokes' theorem (222) provides a geometric interpretation for the curl. At any point $\boldsymbol{r}$, consider a small disk of area A perpendicular to an arbitrary unit vector $\hat{\boldsymbol{n}}$, then Stokes' theorem states that

$$(\boldsymbol{\nabla}\times\boldsymbol{V})\cdot\hat{\boldsymbol{n}} = \lim_{A\to 0}\frac{1}{A}\oint_{\mathcal{C}}\boldsymbol{V}\cdot d\boldsymbol{r} \tag{223}$$

where $\mathcal{C}$ is the circle bounding the disk A oriented with $\hat{\boldsymbol{n}}$. The line integral $\oint \boldsymbol{V}\cdot d\boldsymbol{r}$ is called the *circulation* of the vector field $\boldsymbol{v}$ around the closed curve $\mathcal{C}$.

Proof of Stokes' theorem

Index notation enables a fairly straightforward proof of Stokes' theorem for the more general case of a surface S in 3D space that can be parametrized by a 'good' function $\boldsymbol{r}(u,v)$ (differentiable and integrability as needed). Such a surface S can fold and twist (it could even intersect itself!) and is therefore of a more general kind than those that can be parametrized by the cartesian coordinates x and y. The main restriction on S is that it must be 'orientable'. This means that it must have an 'up' and a 'down' as defined by the direction of the normal $\partial\boldsymbol{r}/\partial u\times\partial\boldsymbol{r}/\partial v$. The famous Möbius strip only has one side and is the classical example of a non-orientable surface.

Let $x_i(u,v)$ represent the i cartesian component of the position vector $\boldsymbol{r}(u,v)$ on the surface S, with $i=1,2,3$. Then

$$\begin{aligned}
(\boldsymbol{\nabla}\times\boldsymbol{V})\cdot\left(\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}\right) &= \epsilon_{ijk}\frac{\partial V_k}{\partial x_j}\,\epsilon_{ilm}\frac{\partial x_l}{\partial u}\frac{\partial x_m}{\partial v} = \epsilon_{ijk}\epsilon_{ilm}\frac{\partial V_k}{\partial x_j}\frac{\partial x_l}{\partial u}\frac{\partial x_m}{\partial v}\\
&= (\delta_{jl}\delta_{km}-\delta_{jm}\delta_{kl})\frac{\partial V_k}{\partial x_j}\frac{\partial x_l}{\partial u}\frac{\partial x_m}{\partial v}\\
&= \frac{\partial V_k}{\partial x_j}\frac{\partial x_j}{\partial u}\frac{\partial x_k}{\partial v}-\frac{\partial V_k}{\partial x_j}\frac{\partial x_k}{\partial u}\frac{\partial x_j}{\partial v}\\
&= \frac{\partial \bar{V}_k}{\partial u}\frac{\partial x_k}{\partial v}-\frac{\partial \bar{V}_k}{\partial v}\frac{\partial x_k}{\partial u}\\
&= \frac{\partial}{\partial u}\left(\bar{V}_k\frac{\partial x_k}{\partial v}\right)-\frac{\partial}{\partial v}\left(\bar{V}_k\frac{\partial x_k}{\partial u}\right)\\
&= \frac{\partial G(u,v)}{\partial u} - \frac{\partial F(u,v)}{\partial v},
\end{aligned} \tag{224}$$

where we have used (184), then the chain rule

$$\frac{\partial V_k}{\partial x_j}\frac{\partial x_j}{\partial u} = \frac{\partial V_k}{\partial x_1}\frac{\partial x_1}{\partial u}+\frac{\partial V_k}{\partial x_2}\frac{\partial x_2}{\partial u}+\frac{\partial V_k}{\partial x_3}\frac{\partial x_3}{\partial u} = \frac{\partial\bar{V}_k}{\partial u},$$

with

$$\bar{V}_k(u,v)\triangleq V_k\big(x_1(u,v),x_2(u,v),x_3(u,v)\big),$$

and equality of mixed partials

$$\frac{\partial^2 x_k}{\partial u\partial v} = \frac{\partial^2 x_k}{\partial v\partial u}.$$

This proof demonstrates the power of index notation with the summation convention. The first line of (224) is a *quintuple* sum over all values of the indices i, j, k, l and m! This would be unmanageable without the compact notation.

Now for any point on the surface S with position vector $\boldsymbol{r}(u,v)$

$$\begin{aligned}\boldsymbol{V}\cdot d\boldsymbol{r} &= \bar{V}_i\left(\frac{\partial x_i}{\partial u}du+\frac{\partial x_i}{\partial v}dv\right)=\left(\bar{V}_i\frac{\partial x_i}{\partial u}\right)du+\left(\bar{V}_i\frac{\partial x_i}{\partial v}\right)dv\\ &= F(u,v)du+G(u,v)dv,\end{aligned}\tag{225}$$

and we have again reduced Stokes' theorem to Green's theorem (220) but expressed in terms of u and v instead of x and y. In details we have shown that

$$\begin{aligned}\int_S(\boldsymbol{\nabla}\times\boldsymbol{V})\cdot \mathrm{d}\boldsymbol{S} &= \int_A(\boldsymbol{\nabla}\times\boldsymbol{V})\cdot\left(\frac{\partial \boldsymbol{r}}{\partial u}\times\frac{\partial \boldsymbol{r}}{\partial v}\right)dudv\\ &= \int_A\left(\frac{\partial G(u,v)}{\partial u}-\frac{\partial F(u,v)}{\partial v}\right)dudv,\end{aligned}\tag{226}$$

$$\oint_{\mathcal{C}}\boldsymbol{V}\cdot d\boldsymbol{r}=\oint_{\mathcal{C}_A}F(u,v)du+G(u,v)dv.\tag{227}$$

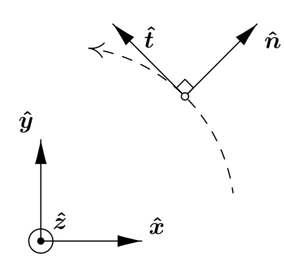
The right hand sides of (226) and (227) are equal by Green's theorem (220). ■

10.5 Divergence form of Green's theorem

The fundamental theorems (217) and (219) in $\mathbb{R}^2$ can be rewritten in a more palatable form.

The line element $d\boldsymbol{r}$ at a point on the curve $\mathcal{C}$ is in the direction of the unit tangent $\hat{\boldsymbol{t}}$ at that point, so $d\boldsymbol{r}=\hat{\boldsymbol{t}}\,ds$, where $\hat{\boldsymbol{t}}$ points in the counterclockwise direction of the curve. Then $\hat{\boldsymbol{t}}\times\hat{\boldsymbol{z}}=\hat{\boldsymbol{n}}$ is the unit *outward* normal $\hat{\boldsymbol{n}}$ to the curve at that point and $\hat{\boldsymbol{z}}\times\hat{\boldsymbol{n}}=\hat{\boldsymbol{t}}$ so

$$\begin{aligned}\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{n}}&=\hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{t}},\\ \hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{t}}&=-\hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{n}}.\end{aligned}\tag{228}$$



Hence since $d\boldsymbol{r}=\hat{\boldsymbol{t}}\,ds$, the fundamental theorems (217) and (219) can be rewritten

$$\int_A\frac{\partial F}{\partial x}\,dA=\oint_{\mathcal{C}}F\hat{\boldsymbol{y}}\cdot d\boldsymbol{r}=\oint_{\mathcal{C}}F\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{n}}\,ds,\tag{229}$$

$$\int_A\frac{\partial F}{\partial y}\,dA=-\oint_{\mathcal{C}}F\hat{\boldsymbol{x}}\cdot d\boldsymbol{r}=\oint_{\mathcal{C}}F\hat{\boldsymbol{y}}\cdot\hat{\boldsymbol{n}}\,ds.\tag{230}$$

The right hand side of these equations is easier to remember since they have $\hat{\boldsymbol{x}}$ going with $\partial/\partial x$ and $\hat{\boldsymbol{y}}$ with $\partial/\partial y$ and both equations have positive signs. But there are hidden subtleties. The arclength element $ds=|d\boldsymbol{r}|$ is positive by definition and $\hat{\boldsymbol{n}}$ must be the unit *outward* normal to the boundary, so if an explicit parametrization of the boundary curve is known, the bounds of integration should be picked so that $\oint_{\mathcal{C}} ds = \oint_{\mathcal{C}} |d\boldsymbol{r}| > 0$ would be the length of the curve with a positive sign. For the $d\boldsymbol{r}$ line integrals, the bounds of integration must correspond to counter-clockwise orientation of $\mathcal{C}$.

Formulas (229) and (230) can be combined in more useful forms. First, if u is the signed distance in the direction of the unit vector $\hat{\boldsymbol{u}}$, then the (directional) derivative in the direction $\hat{\boldsymbol{u}}$ is $\partial F/\partial u = \hat{\boldsymbol{u}}\cdot\boldsymbol{\nabla} F \equiv u_x\partial F/\partial x + u_y\partial F/\partial y$, where $\hat{\boldsymbol{u}} = u_x\hat{\boldsymbol{x}} + u_y\hat{\boldsymbol{y}}$, therefore combining (229) and (230) accordingly we obtain

$$\boxed{\int_A \frac{\partial F}{\partial u}\, dA = \oint_{\mathcal{C}} F\hat{\boldsymbol{u}}\cdot\hat{\boldsymbol{n}}\, ds.} \tag{231}$$

This result is written in a coordinate-free form. It applies to *any* direction $\hat{\boldsymbol{u}}$ in the x,y plane.

Another useful combination is to add (229) to (230), the latter written for an arbitrary function $G(x,y)$ in place of $F(x,y)$, yielding the *divergence-form of Green's theorem*

$$\boxed{\int_A \boldsymbol{\nabla}\cdot\boldsymbol{V}\; dA = \oint_{\mathcal{C}} \boldsymbol{V}\cdot\hat{\boldsymbol{n}}\; ds.} \tag{232}$$

for the arbitrary vector field $\boldsymbol{V} = F\hat{\boldsymbol{x}} + G\hat{\boldsymbol{y}}$ and arbitrary area A with boundary $\mathcal{C}$, where $\hat{\boldsymbol{n}}$ is the local unit outward normal and $ds = |d\boldsymbol{r}|$ is the arclength element along the curve.

10.6 Gauss' theorem

Gauss' theorem is the 3D version of the divergence form of Green's theorem. It is proved by first extending the fundamental theorem of calculus to 3D.

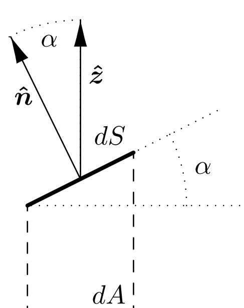
If V is a bounded volume in 3D space and $F(x,y,z)$ is a scalar function of the cartesian coordinates (x,y,z), then we have

$$\int_V \frac{\partial F}{\partial z}\, dV = \oint_S F\,\hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{n}} dS, \tag{233}$$

where S is the *closed* surface enclosing V and $\hat{\boldsymbol{n}}$ is the unit *outward* normal to S.

The proof of this result is similar to that for (217). Assume that the surface can be parametrized using x and y in two pieces: an upper 'hemisphere' at $z = z_u(x,y)$ and a lower 'hemisphere' at $z = z_l(x,y)$ with x,y in a domain A, the projection of S onto the x,y plane, that is the *same domain* for both the upper and lower surfaces. The closed surface S is not a sphere in general but we used the word 'hemisphere' to help visualize the problem. For a sphere, A is the equatorial disk, $z = z_u(x,y)$ is the northern hemisphere and $z = z_l(x,y)$ is the southern hemisphere.

Iterated integrals with $dV = dA\,dz$ and the fundamental theorem of calculus give



$$\int_V \frac{\partial F}{\partial z}\, dV = \int_A dA \int_{z_l(x,y)}^{z_u(x,y)} \frac{\partial F}{\partial z}\, dz = \int_A \big[F(x,y,z_u(x,y)) - F(x,y,z_l(x,y))\big]\, dA. \tag{234}$$

We can interpret that integral over A as an integral over the entire closed surface S that bounds V. All we need for that is a bit of geometry. If dA is the projection of the surface element $d\boldsymbol{S} = \hat{\boldsymbol{n}} dS$ onto the x, y plane then we have $dA = \cos\alpha\, dS = \pm\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{n}}\, dS$. The + sign applies to the upper surface for which $\hat{\boldsymbol{n}}$ is pointing up (*i.e.* in the direction of $\hat{\boldsymbol{z}}$) and the - sign for the bottom surface where $\hat{\boldsymbol{n}}$ points down (and would be opposite to the $\hat{\boldsymbol{n}}$ on the side figure). Thus we obtain

$$\int_V \frac{\partial F}{\partial z}\, dV = \oint_S F\, \hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{n}}\, dS, \tag{235}$$

where $\hat{\boldsymbol{n}}$ is the unit *outward* normal to S.

We can obtain similar results for the volume integrals of $\partial F/\partial x$ and $\partial F/\partial y$ and combine those to obtain

$$\boxed{\int_V \frac{\partial F}{\partial u}\, dV = \oint_S F\, \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{n}}\, dS,} \tag{236}$$

for arbitrary but fixed direction $\hat{\boldsymbol{u}}$. This is the 3D version of (231) and of the fundamental theorem of calculus.

We can combine this theorem into many useful forms. Writing it for $F(x, y, z)$ in the $\hat{\boldsymbol{x}}$ direction, with the $\hat{\boldsymbol{y}}$ version for a function $G(x, y, z)$ and the $\hat{\boldsymbol{z}}$ version for a function $H(x, y, z)$ we obtain *Gauss's theorem*

$$\boxed{\int_V \boldsymbol{\nabla} \cdot \boldsymbol{v}\, dV = \oint_S \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, dS,} \tag{237}$$

where $\boldsymbol{v} = F\hat{\boldsymbol{x}} + G\hat{\boldsymbol{y}} + H\hat{\boldsymbol{z}}$. This is the 3D version of (232). Note that both (236) and (237) are expressed in coordinate-free forms. These are general results. The integral $\oint_S \boldsymbol{v} \cdot \hat{\boldsymbol{n}} dS$ is the *flux* of $\boldsymbol{v}$ through the surface S. If $\boldsymbol{v}(\boldsymbol{r})$ is the velocity of a fluid at point $\boldsymbol{r}$ then that integral represents the time-rate at which volume of fluid flows through the surface S.

Gauss' theorem provides a coordinate-free interpretation for the divergence. Consider a small sphere of volume V and surface S centered at a point $\boldsymbol{r}$, then Gauss' theorem states that

$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = \lim_{V \to 0} \frac{1}{V} \oint_S \boldsymbol{v} \cdot \hat{\boldsymbol{n}}\, dS. \tag{238}$$

Note that (236) and (237) are equivalent. We deduced (237) from (236), but we can also deduce (236) from (237) by considering the special $\boldsymbol{v} = F\hat{\boldsymbol{u}}$ where $\hat{\boldsymbol{u}}$ is a unit vector independent of $\boldsymbol{r}$. Then from our vector identities $\boldsymbol{\nabla} \cdot (F\hat{\boldsymbol{u}}) = \hat{\boldsymbol{u}} \cdot \boldsymbol{\nabla} F = \partial F/\partial u$.

10.7 Other forms of the fundamental theorem in 3D

A useful form of (236) is to write it in indicial form as

$$\boxed{\int_V \frac{\partial F}{\partial x_j}\, dV = \oint_S n_j F\, dS.} \tag{239}$$

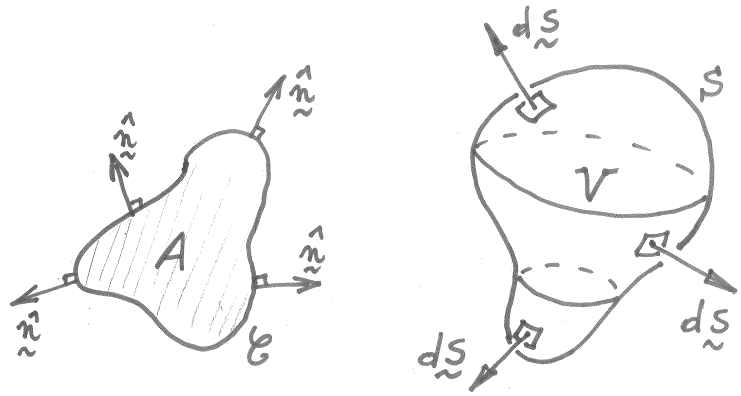


Fig. 2.17: Gauss' theorem, eqns. (232) and (237), applies to a flat area A with boundary curve $\mathcal{C}$ and to a volume V with boundary surface S and surface element $d\boldsymbol{S} = \hat{\boldsymbol{n}}\, dS$, where $\hat{\boldsymbol{n}}$ is the local unit outward normal.

Then with $f(\boldsymbol{r})$ in place of $F(\boldsymbol{r})$ we deduce that

$$\int_V \mathbf{e}_j \frac{\partial f}{\partial x_j}\, dV = \oint_S \mathbf{e}_j n_j f\, dS \tag{240}$$

since the cartesian unit vectors $\mathbf{e}_j$ are independent of position. With the summation convention, this is a sum of all three partial derivatives in (239), multiplied by their respective direction vector. This result can be written in coordinate-free form as

$$\int_V \boldsymbol{\nabla} f\, dV = \oint_S f\, \hat{\boldsymbol{n}}\, dS. \tag{241}$$

One application of this form of the fundamental theorem is to prove *Archimedes' principle.*

Next, writing (239) for v_k in place of F yields

$$\int_V \frac{\partial v_k}{\partial x_j}\, dV = \oint_S n_j\, v_k\, dS, \tag{242}$$

that represents 9 different integrals since j and k are free indices. Multiplying (242) by the position-independent ϵ_{ijk} and summing over all j and k gives

$$\int_V \epsilon_{ijk} \frac{\partial v_k}{\partial x_j}\, dV = \oint_S \epsilon_{ijk} n_j\, v_k\, dS. \tag{243}$$

The coordinate-free form of this is

$$\int_V \boldsymbol{\nabla} \times \boldsymbol{v}\, dV = \oint_S \hat{\boldsymbol{n}} \times \boldsymbol{v}\, dS. \tag{244}$$

The integral theorems (241) and (244) provide yet other geometric interpretations for the gradient

$$\boldsymbol{\nabla} f = \lim_{V \to 0} \frac{1}{V} \oint_S f\, \hat{\boldsymbol{n}}\, dS, \tag{245}$$

and the curl

$$\boldsymbol{\nabla} \times \boldsymbol{v} = \lim_{V \to 0} \frac{1}{V} \oint_S \hat{\boldsymbol{n}} \times \boldsymbol{v} \, dS. \tag{246}$$

These are similar to the result (238) for the divergence. Recall the other geometric interpretations for the gradient and the curl – the gradient as the vector pointing in the direction of greatest rate of change (sect. 8.2), and of the $\hat{\boldsymbol{n}}$ component of the curl, as the limit of the local circulation per unit area as given by Stokes' theorem (223).

In applications, we use the fundamental theorem as we do in 1D, namely to reduce a 3D integral to a 2D integral, for instance. However we also use them the other way, to evaluate a complicated surface integral as a simpler volume integral, for instance when $\boldsymbol{\nabla} f$ or $\boldsymbol{\nabla} \cdot \boldsymbol{v}$ or $\boldsymbol{\nabla} \times \boldsymbol{v}$ are constants and the volume integral is then trivial.

Exercises

1. Calculate $\oint_{\mathcal{C}} y \, dx$ where $\mathcal{C} : \boldsymbol{r}(\theta) = \boldsymbol{r}_C + \boldsymbol{a} \cos\theta + \boldsymbol{b} \sin\theta$ with $\boldsymbol{r}_C, \boldsymbol{a}, \boldsymbol{b}$ constants.

2. Let A be the area of the triangle with vertices $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$, $P_3 \equiv (x_3, y_3)$ in the cartesian x, y plane and $\mathcal{C} \equiv \partial A$ denotes the boundary of that area, oriented counterclockwise. (i) Calculate $\int_A x^2 \, dA$. Show/explain your work. (ii) Calculate $\oint_{\partial A} \boldsymbol{v} \cdot d\boldsymbol{r}$ where (1) $\boldsymbol{v} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}}$, (2) $\boldsymbol{v} = y\hat{\boldsymbol{x}} - x\hat{\boldsymbol{y}}$. Sketch $\boldsymbol{v}$.

3. If $\mathcal{C}$ is any closed curve in 3D space (i) calculate $\oint_{\mathcal{C}} \boldsymbol{r} \cdot d\boldsymbol{r}$ in two ways, (ii) calculate $\oint_{\mathcal{C}} \boldsymbol{\nabla} f \cdot d\boldsymbol{r}$ in two ways, where $f(\boldsymbol{r})$ is a scalar function. [Hint: by direct calculation and by Stokes theorem].

4. If $\mathcal{C}$ is any closed curve in 3D space not passing through the origin calculate $\oint_{\mathcal{C}} r^{-3} \boldsymbol{r} \cdot d\boldsymbol{r}$ in two ways.

5. Calculate the circulation of the vector field $\boldsymbol{B} = (\hat{\boldsymbol{z}} \times \boldsymbol{r})/|\hat{\boldsymbol{z}} \times \boldsymbol{r}|^2$ (i) about a circle of radius R centered at the origin in a plane perpendicular to $\hat{\boldsymbol{z}}$; (ii) about any closed curve $\mathcal{C}$ in 3D that does not go around the z-axis; (ii) about any closed curve $\mathcal{C}_0$ that does go around the $\hat{\boldsymbol{z}}$ axis. What's wrong with the z-axis anyway?

6. Consider $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$ where $\boldsymbol{\omega}$ is a constant vector, independent of $\boldsymbol{r}$. (i) Evaluate the circulation of $\boldsymbol{v}$ about the circle of radius R centered at the origin in the plane perpendicular to the direction $\hat{\boldsymbol{n}}$ by direct calculation of the line integral; (ii) Calculate the curl of $\boldsymbol{v}$ using vector identities; (iii) calculate the circulation of $\boldsymbol{v}$ about a circle of radius R centered at $\boldsymbol{r}_0$ in the plane perpendicular to $\hat{\boldsymbol{n}}$.

7. If $\mathcal{C}$ is any closed curve in 2D, calculate $\oint_{\mathcal{C}} \hat{\boldsymbol{n}} \, ds$ where $\hat{\boldsymbol{n}}(\boldsymbol{r})$ is the unit outside normal to $\mathcal{C}$ at the point $\boldsymbol{r}$ of $\mathcal{C}$ and $ds = |d\boldsymbol{r}|$.

8. If S is any closed surface in 3D, calculate $\oint \hat{\boldsymbol{n}} dS$ where $\hat{\boldsymbol{n}}(\boldsymbol{r})$ is the unit outside normal to S at a point $\boldsymbol{r}$ on S.

9. If S is any closed surface in 3D, calculate $\oint p\,\hat{\boldsymbol{n}}\,dS$ where $\hat{\boldsymbol{n}}$ is the unit outside normal to S and $p(\boldsymbol{r}) = (p_0 - \rho g \hat{\boldsymbol{z}} \cdot \boldsymbol{r})$ where p_0, ρ and g are constants (This is *Archimedes' principle* with ρ as fluid density and g as the acceleration of gravity.) Calculate the torque, $\oint_S \boldsymbol{r} \times (-p\hat{\boldsymbol{n}})\,dS$ for the same pressure field p.

10. Calculate the flux of $\boldsymbol{r}$ through (i) the surface of a sphere of radius R centered at the origin in two ways; (ii) through the surface of the sphere of radius R centered at $\boldsymbol{r}_0$; (iii) through the surface of a cube of side L with one corner at the origin in two ways.

11. (i) Calculate the flux of $\boldsymbol{v} = \boldsymbol{r}/r^3$ through the surface of a sphere of radius ϵ centered at the origin. (ii) Calculate the flux of that vector field through a closed surface that does not enclose the origin [Hint: use the divergence theorem] (iii) Calculate the flux through an arbitrary closed surface that encloses the origin [Hint: use divergence theorem and (i) to isolate the origin. What's wrong with the origin anyway?]

12. Calculate $\boldsymbol{\nabla}|\boldsymbol{r} - \boldsymbol{r}_0|^{-1}$ and $\boldsymbol{\nabla} \cdot \boldsymbol{v}$ with $\boldsymbol{v} = (\boldsymbol{r} - \boldsymbol{r}_0)/|\boldsymbol{r} - \boldsymbol{r}_0|^3$ where $\boldsymbol{r}_0$ is a constant vector.

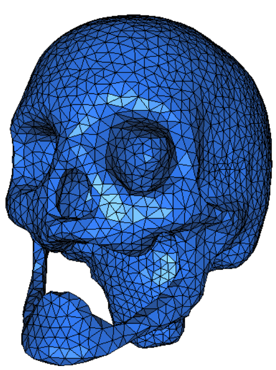
13. What are all the possible values of

$$\oint_S \boldsymbol{v} \cdot d\boldsymbol{S} \quad \text{for} \quad \boldsymbol{v} = A\,\frac{\boldsymbol{r} - \boldsymbol{a}}{|\boldsymbol{r} - \boldsymbol{a}|^3} + B\,\frac{\boldsymbol{r} - \boldsymbol{b}}{|\boldsymbol{r} - \boldsymbol{b}|^3},$$

where $A, B, \boldsymbol{a}, \boldsymbol{b}$ constants and S is any closed surface that does not pass through $\boldsymbol{a}$ or $\boldsymbol{b}$? Explain/justify carefully.

14. Calculate the flux of $\boldsymbol{v}$ through the surface of a sphere of radius R centered at the origin when $\boldsymbol{v} = \alpha_1(\boldsymbol{r} - \boldsymbol{r}_1)/|\boldsymbol{r} - \boldsymbol{r}_1|^3 + \alpha_2(\boldsymbol{r} - \boldsymbol{r}_2)/|\boldsymbol{r} - \boldsymbol{r}_2|^3$ where α_1 and α_2 are scalar constants and (i) $|\boldsymbol{r}_1|$ and $|\boldsymbol{r}_2|$ are both less than R; (ii) $|\boldsymbol{r}_1| < R < |\boldsymbol{r}_2|$. Generalize to $\boldsymbol{v} = \sum_{i=1}^{N} \alpha_i(\boldsymbol{r} - \boldsymbol{r}_i)/|\boldsymbol{r} - \boldsymbol{r}_i|^3$.

15. Calculate $\boldsymbol{F}(\boldsymbol{r}) = \int_{V_0} \boldsymbol{f}\,dV_0$ where V_0 is the inside of a sphere of radius R centered at O, $\boldsymbol{f} = (\boldsymbol{r} - \boldsymbol{r}_0)/|\boldsymbol{r} - \boldsymbol{r}_0|^3$ with $|\boldsymbol{r}| > R$ and *the integral is over* $\boldsymbol{r}_0$. This is essentially the gravity field or force at $\boldsymbol{r}$ due to a sphere of uniform mass density. The integral can be cranked out if you're good at analytic integration. But the smart solution is to realize that the integral over $\boldsymbol{r}_0$ is essentially a sum over $\boldsymbol{r}_i$ as in the previous exercise so we can figure out the flux of $\boldsymbol{F}$ through any closed surface enclosing all of V_0. Now by symmetry $\boldsymbol{F}(\boldsymbol{r}) = F(r)\hat{\boldsymbol{r}}$, so knowing the flux is enough to figure out $F(r)$.



16. Suppose you have all the data to plot the skull surface, how do you compute its volume? Provide an explicit formula or algorithm and specify what data is needed and in what form.

Chapter 3

Complex Calculus

1 Complex Numbers and Elementary functions

1.1 Complex Algebra and Geometry

Imaginary and complex numbers were introduced by Italian mathematicians in the early 1500's as a result of algorithms to find the solutions of the cubic equation $ax^3 + bx^2 + cx + d = 0$. Surprisingly, the formula they derived involved square roots of negative numbers when the cubic was known to have three real roots. If they carried through their calculations, *imagining* that $\sqrt{-1}$ existed, they obtained the correct real roots. Following Euler, we define

$$i \triangleq \sqrt{-1} \quad \text{such that} \quad i^2 = -1, \tag{1}$$

and a complex number z

$$z = x + i\,y \tag{2}$$

consists of a *real part* $\mathfrak{Re}(z) = x$ and an *imaginary part* $\mathfrak{Im}(z) = y$, both of which are real numbers. For example

$$z = 2 + 3i \;\Leftrightarrow\; \mathfrak{Re}(z) = 2,\; \mathfrak{Im}(z) = 3.$$

Note that the *imaginary part* of a complex number is *real*, $\mathfrak{Im}(2+3i) = 3$, not $3i$ and more generally $\mathfrak{Im}(z) = y$ not $i\,y$.

In the early 1800's, the French amateur mathematician Jean-Robert Argand introduced the geometric interpretation of complex numbers as *vectors* in the (x, y) plane with multiplication by i corresponding to counterclockwise rotation by $\pi/2$

$$z = x + iy \equiv (x, y) \Rightarrow iz = i(x + iy) = -y + ix \equiv (-y, x).$$

A second multiplication by i corresponds to another $\pi/2$ rotation $(-y, x) \to (-x, -y)$ resulting in a sign change after 2 successive rotations by $\pi/2$, $(x, y) \to (-x, -y) = -(x, y)$ and indeed $i^2 = -1$. That (x, y) plane is called the *complex plane*, historically known also as the Argand plane and illustrated in figure 3.1. Argand introduced the term *vector* as well as the geometric concepts of *modulus* $|z|$ and *argument* $\arg(z)$.

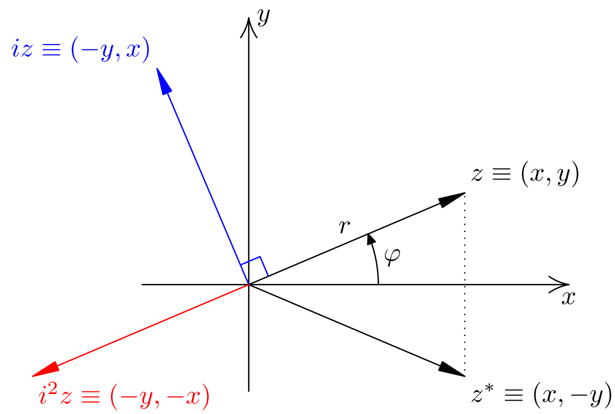


Fig. 3.1: The complex plane $\mathbb{C}$, also known as the Argand plane, interprets complex numbers $z = x + iy$ as vectors (x, y) in $\mathbb{R}^2$ with additional algebraic/geometric operations. The magnitude of z is $|z| = r = \sqrt{x^2 + y^2}$, its *argument* $\arg(z) = \varphi$ with $x = r\cos\varphi$, $y = r\sin\varphi$.

Modulus (or Norm or Magnitude)

The modulus of $z = x + iy$ is

$$|z| \triangleq \sqrt{x^2 + y^2} = r. \tag{3}$$

This modulus is equivalent to the euclidean norm of the 2D vector (x, y).

Argument (or angle)

The 'argument' of $z = x + iy$ is

$$\arg(z) \equiv \text{angle}(z) = \varphi \tag{4}$$

such that $(x, y) = (r\cos\varphi, r\sin\varphi)$. It is the angle between the vector (x, y) and the real direction $(1, 0)$. The notation $\arg(z)$ is classic but angle(z) is clearer. The standard unique definition specifies the angle in $(-\pi, \pi]$

$$-\pi < \text{angle}(z) \leq \pi \tag{5}$$

thus $\angle(z) = \text{atan2}(y, x)$. The function atan2$(y, x)$ was introduced in the Fortran computer language and is now common in all computer languages. The atan2(y, x) function is the arctangent function but returns an angle in $(-\pi, \pi]$, in contrast to $\arctan(y/x)$ that loses sign information and returns an angle in $[-\pi/2, \pi/2]$.

Complex conjugate

The reflection of $z = x + iy$ about the real axis is called the *complex conjugate* of z

$$z^* \triangleq x - iy, \tag{6}$$

written $\bar{z}$ by some authors.

Addition

$$z_1 + z_2 \triangleq (x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2).$$

That is $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$ so complex numbers add like 2D cartesian vectors, by adding the corresponding real and imaginary components. Thus complex addition satisfies the *triangle inequality*

$$|z_1 + z_2| \leq |z_1| + |z_2|. \tag{7}$$

Multiplication

From usual algebra but with the additional rule $i^2 = -1$, we obtain

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i\,(x_1 y_2 + x_2 y_1).$$

It is a straightforward exercise to show that the magnitude of a product is the product of the magnitudes

$$|z_1 z_2| = |z_1|\,|z_2| \tag{8}$$

and the angle of a product is the *sum* of the angles *modulo* 2π

$$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) \quad \text{modulo } 2\pi. \tag{9}$$

For instance, $\arg((-2)(-3)) = \arg(6) = 0$ but $\arg(-2) = \pi$ and $\arg(-3) = \pi$ so $\arg(-2) + \arg(-3) = 2\pi$ which is equal to 0 up to a multiple of 2π, that is *modulo* 2π.

The complex product is an operation that does not exist for 2D real cartesian vectors. However, note that

$$z_1^* z_2 = (x_1 - iy_1)(x_2 + iy_2) = (x_1 x_2 + y_1 y_2) + i(x_1 y_2 - x_2 y_1)$$

has a real part that equals the dot product of the real vectors

$$\boldsymbol{r}_1 \cdot \boldsymbol{r}_2 = (x_1, y_1) \cdot (x_2, y_2) = x_1 x_2 + y_1 y_2 \equiv \mathfrak{Re}(z_1^* z_2)$$

and an imaginary part that equals the $\mathbf{e}_3$ component of the cross product

$$(x_1, y_1, 0) \times (x_2, y_2, 0) = (0, 0, x_1 y_2 - x_2 y_1).$$

Note in particular that

$$z^* z = z z^* = x^2 + y^2 = r^2 = |z|^2, \tag{10}$$

and this relation is often useful to compute complex magnitudes just as the dot product is useful to compute vector magnitudes. For instance

$$|a + b| = \sqrt{(a + b)(a + b)^*} = \sqrt{aa^* + ab^* + ba^* + bb^*}$$

is the complex equivalent of

$$|\boldsymbol{a} + \boldsymbol{b}| = \sqrt{(\boldsymbol{a} + \boldsymbol{b}) \cdot (\boldsymbol{a} + \boldsymbol{b})} = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a} + 2\boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{b} \cdot \boldsymbol{b}}$$

for real vectors.

However, while the dot product of two vectors is not a vector, and the cross product of horizontal vectors is not a horizontal vector, the product of complex numbers is a complex number. This allows definition of complex division as the opposite of multiplication.

Division

Given z_1 and $z_2 \neq 0$ we can find a unique z such that $zz_2 = z_1$. That z is denoted z_1/z_2 and equal to

$$\frac{z_1}{z_2} = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{(x_1 + iy_1)(x_2 - iy_2)}{x_2^2 + y_2^2} = \left(\frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2}\right) + i\left(\frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}\right). \quad (11)$$

Thus while division of vectors by vectors is not defined (what is North divided by Northwest?!), we can divide complex numbers by complex numbers and the complex interpretation of 'North/Northwest' is

$$\frac{i}{\frac{-1+i}{\sqrt{2}}} = \frac{\sqrt{2}\, i(-1-i)}{(-1+i)(-1-i)} = \frac{1-i}{\sqrt{2}},$$

which is "Southeast"! But we won't use that confusing language and will not divide directions by directions. However we will divide unit complex numbers by unit complex numbers, especially after we have (re)viewed Euler's fabulous formula $e^{i\theta} = \cos\theta + i\,\sin\theta$.

It is left as an exercise to show that the magnitude of a ratio is the ratio of the magnitudes

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}$$

and the angle of a ratio is the *difference* of the angles, modulo 2π

$$\arg\left(\frac{z_1}{z_2}\right) = \arg(z_1) - \arg(z_2) \quad \text{modulo } 2\pi.$$

For example, $\arg(-2/i) = \arg(2i) = \pi/2$ but $\arg(-2) = -\pi$ and $\arg(i) = \pi/2$ so $\arg(-2) - \arg(i) = -3\pi/2$ which is $\pi/2 - (2\pi)$, so $\arg(z_1/z_2) = \arg(z_1) - \arg(z_2)$ up to a multiple of 2π.

All the algebraic formula that should be well known for real numbers hold for complex numbers as well, for instance

$$z^2 - a^2 = (z-a)(z+a)$$

$$z^3 - a^3 = (z-a)(z^2 + az + a^2)$$

and for any positive integer n

$$z^{n+1} - a^{n+1} = (z-a)\sum_{k=0}^{n} z^{n-k} a^k. \quad (12)$$

Also,

$$(z+a)^2 = z^2 + 2za + a^2$$
$$(z+a)^3 = z^3 + 3z^2 a + 3za^2 + a^3$$

and for any positive integer n we have the **binomial formula** (defining $0! = 1$)

$$(z+a)^n = \sum_{k=0}^{n} \binom{n}{k} z^{n-k} a^k = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} z^{n-k} a^k. \tag{13}$$

These classic algebraic formula will be useful to prove that $dz^n/dz = nz^{n-1}$ and that $\exp(z+a) = \exp(z)\exp(a)$.

Exercises:

1. Express the following expressions in the form $a + ib$ with a and b real, and $i^2 = -1$:
   $(2+3i)^*$, $|2+3i|$, $\mathfrak{Im}(2+3i)$, $\dfrac{1}{2+3i}$, $\arg(2+3i)$. Show your work.

2. Calculate $(1+i)/(2+3i)$. Find its magnitude and its angle.

3. Consider two arbitrary complex numbers z_1 and z_2. Sketch z_1, z_2, z_1+z_2, z_1-z_2, $z_1+z_1^*$ and $z_1-z_1^*$ in the complex plane.

4. Consider $z_1 = -1 + i\,10^{-17}$ and $z_2 = -1 - i\,10^{-17}$. Calculate $|z_1 - z_2|$ and $\arg(z_1) - \arg(z_2)$. Are z_1 and z_2 close to each other? Are $\arg(z_1)$ and $\arg(z_2)$ close to each other? Show your work.

5. Prove that $(z_1+z_2)^* = z_1^* + z_2^*$ and $(z_1 z_2)^* = z_1^* z_2^*$ for any complex numbers z_1 and z_2.

6. Prove that $|z_1 z_2| = |z_1||z_2|$ but $|z_1+z_2| \le |z_1| + |z_2|$ for any complex numbers z_1 and z_2. When is $|z_1+z_2| = |z_1| + |z_2|$?

7. Show that $|z_1 z_2| = |z_1||z_2|$ and $\angle(z_1 z_2) = \angle(z_1) + \angle(z_2)$ modulo 2π, (1) using trigonometry with $z = r(\cos\varphi + i\sin\varphi)$ and angle sum identities, and (2) geometrically by distributing one of the factors, for instance $z_1 z_2 = x_1 z_2 + y_1(iz_2)$ and interpreting the result as the linear combination of two orthogonal vectors of same magnitude $z_2 = (x_2, y_2) \equiv \boldsymbol{r}_2$ and $iz_2 = (-y_2, x_2) \equiv \boldsymbol{r}_2^{\perp}$.

8. Show that $|z_1/z_2| = |z_1|/|z_2|$ and $\angle(z_1/z_2) = \angle(z_1) - \angle(z_2)$ modulo 2π. (Hint: use the corresponding results for product and $z_1/z_2 = z_1 z_2^*/|z_2|^2$.)

9. If a and b are arbitrary complex numbers, prove that $ab^* + a^*b$ is real and $ab^* - a^*b$ is imaginary.

10. True or false: $(iz)^* z = 0$ since $i\,z$ is perpendicular to z. Explain.

11. For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the dot product $\boldsymbol{a}\cdot\boldsymbol{b} = 0$ when the vectors are perpendicular and the cross product $\boldsymbol{a}\times\boldsymbol{b} = 0$ when they are parallel. Show that two complex numbers a and b are parallel when $ab^* = a^*b$ and perpendicular when $ab^* = -a^*b$.

12. Show that the final formula (11) for division follows from the definition of multiplication (as it should), that is, if $z = z_1/z_2$ then $zz_2 = z_1$, solve for $\mathfrak{Re}(z)$ and $\mathfrak{Im}(z)$ to find z.

13. Prove (12) and the binomial formula (13) as you (should have) learned in high school algebra.

1.2 Limits and Derivatives

The modulus allows the definition of distance and limit. The **distance** between two complex numbers z_1 and z_2 is the modulus of their difference $|z_1 - z_2|$. A complex number z_1 tends to a complex number z if $|z_1 - z| \to 0$, where $|z_1 - z|$ is the euclidean distance between the complex numbers z_1 and z in the complex plane.

Continuity

A function $f(z)$ is continuous at z if

$$\lim_{z_1 \to z} f(z_1) = f(z), \tag{14}$$

which means that $f(z_1)$ can be as close as we want to $f(z)$ by taking z_1 close enough to z. In mathematical notation, that is

$$\forall \epsilon > 0\ \exists \delta > 0 \quad \text{s.t.} \quad |z_1 - z| < \delta \Rightarrow |f(z_1) - f(z)| < \epsilon.$$

For example $f(z) = a_2 z^2 + a_1 z + a_0$ is continuous everywhere since

$$\begin{aligned} |f(z_1) - f(z)| =& |a_2(z_1^2 - z^2) + a_1(z_1 - z)| \\ & \leq |a_2||z_1^2 - z^2| + |a_1||z_1 - z| = |z_1 - z| \left(|a_2||z_1 + z| + |a_1|\right) \end{aligned}$$

thus $|f(z_1) - f(z)| \to 0$ as $|z_1 - z| \to 0$. See the exercises for (ϵ, δ) for this example.

Differentiability

The *derivative* of a function $f(z)$ at z is

$$\frac{df(z)}{dz} = \lim_{a \to 0} \frac{f(z+a) - f(z)}{a} \tag{15}$$

where a is a complex number and $a \to 0$ means $|a| \to 0$. This limit must be the same no matter *how* $a \to 0$.

We can use the binomial formula (13) as done in Calc I to deduce that

$$\frac{dz^n}{dz} = n z^{n-1} \tag{16}$$

for any integer $n = 0, \pm 1, \pm 2, \ldots$, and we can define the *anti-derivative* of z^n as $z^{n+1}/(n+1) + C$ for all integer $n \neq -1$. All the usual rules of differentiation:

product rule, quotient rule, chain rule,..., still apply for complex differentiation and we will not bother to prove those here, the proofs are just like in Calc I.

So there is nothing special about complex derivatives, or is there? Consider the function $f(z) = \mathfrak{Re}(z) = x$, the real part of z. What is its derivative? Hmm..., none of the rules of differentiation help us here, so let's go back to first principles:

$$\frac{d\mathfrak{Re}(z)}{dz} = \lim_{a\to 0} \frac{\mathfrak{Re}(z+a) - \mathfrak{Re}(z)}{a} = \lim_{a\to 0} \frac{\mathfrak{Re}(a)}{a} = ?! \tag{17}$$

What is that limit? If a is real, then $a = \mathfrak{Re}(a)$ so the limit is 1, but if a is imaginary then $\mathfrak{Re}(a) = 0$ and the limit is 0. So there is no limit that holds for all $a \to 0$. The limit depends on *how* $a \to 0$, and we cannot define the z-derivative of $\mathfrak{Re}(z)$. $\mathfrak{Re}(z)$ is continuous everywhere, but nowhere z-differentiable!

Exercises:

1. Prove that the functions $f(z) = \mathfrak{Re}(z)$ and $f(z) = z^*$ are continuous everywhere.

2. For $f(z) = a_2 z^2 + a_1 z + a_0$ show that $|f(z_1) - f(z)| < \epsilon$ when $|z_1 - z| < \delta = \frac{-|a_1| + \sqrt{|a_1|^2 + 4\epsilon|a_2|}}{2|a_2|}$. Does this hold also when $a_2 \to 0$? What $\delta(\epsilon)$ do you expect for $a_2 = 0$?

3. Prove formula (16) from the limit definition of the derivative (a) using (12), (b) using (13).

4. Prove that (16) also applies to negative integer powers $z^{-n} = 1/z^n$ from the limit definition of the derivative.

5. Investigate the existence of df/dz from the limit definition for $f(z) = \mathfrak{Im}(z)$ and $f(z) = |z|$.

1.3 Geometric sums and series

For any complex number $q \neq 1$, the *geometric sum*

$$1 + q + q^2 + \cdots + q^n = \frac{1 - q^{n+1}}{1 - q}. \tag{18}$$

To prove this, let $S_n = 1 + q + \cdots + q^n$ and note that $qS_n = S_n + q^{n+1} - 1$, solving for S_n yields $S_n = (1 - q^{n+1})/(1 - q)$.

The *geometric series* is the limit of the sum as $n \to \infty$. It follows from (18), that the geometric series converges to $1/(1-q)$ if $|q| < 1$, and diverges if $|q| > 1$,

$$\boxed{\sum_{n=0}^{\infty} q^n = 1 + q + q^2 + \cdots = \frac{1}{1-q}, \quad \text{iff} \quad |q| < 1.} \tag{19}$$

Note that we have two different functions of q: (1) the series $\sum_{n=0}^{\infty} q^n$ which only exists when $|q| < 1$, (2) the function $1/(1-q)$ which is defined and smooth everywhere except at $q = 1$. These two expressions, the geometric series and the function $1/(1-q)$ are identical in the disk $|q| < 1$, but they are not at all identical outside of that disk since the series does not make any sense (*i.e.* it diverges) outside of it. What happens on the unit circle $|q| = 1$? (consider for example $q = 1$, $q = -1$, $q = i$, ...)

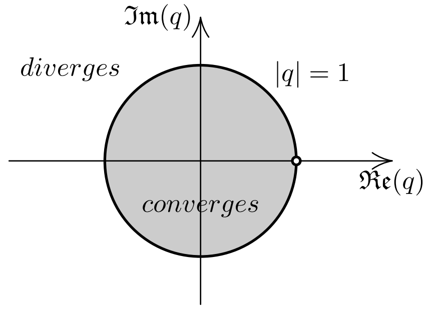


Fig. 3.2: The geometric series (18) $G(q) = \sum_{n=0}^{\infty} q^n$ converges in the disk $|q| < 1$ where it equals $1/(1-q)$. The series diverges outside the unit disk although the function $1/(1-q)$ exists for any $q \neq 1$.

Exercises:

1. Prove formula (18). What is the sum when $q = 1$?
2. What is $\sum_{n=0}^{\infty} q^n$ when $q = 1$? $q = -1$? $q = i$?
3. Calculate $1 + z + z^2 + \cdots + z^{321}$ for $z = i$ and $z = 1 + i$. Explain your work.
4. Calculate $1 - z + z^2 - z^3 + \cdots - z^{321}$ for $z = i$ and $z = 1 + i$. Explain your work.
5. Calculate $q^N + q^{N+2} + q^{N+4} + q^{N+6} + \cdots$ with $|q| < 1$.
6. What is the connection between the geometric sum (18) and the classic algebraic formula (12)?

1.4 Ratio test

The geometric series leads to a useful test for convergence of the general series

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots \tag{20}$$

We can make sense of this series again as the limit of the partial sums $S_n = a_0 + a_1 + \cdots + a_n$ as $n \to \infty$. Any one of these finite partial sums exists but the infinite sum does not necessarily converge. Example: take $a_n = 1 \; \forall n$, then $S_n = n + 1$ and $S_n \to \infty$ as $n \to \infty$.

A *necessary* condition for convergence is that $a_n \to 0$ as $n \to \infty$ as you learned in Calculus, but that is not sufficient. A *sufficient* condition for convergence is obtained by comparison to a geometric series. This leads to the *Ratio Test*: the series (20) converges if

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L < 1. \tag{21}$$

Why does the ratio test work? If $L < 1$, then pick *any* q such that $L < q < 1$ and one can find a (sufficiently large) N such that $|a_{n+1}|/|a_n|$ is within $\epsilon = q - L$ of L, in other words such that $|a_{n+1}|/|a_n| < q$ for all $n \geq N$, so we can write

$$\begin{aligned} &|a_N| + |a_{N+1}| + |a_{N+2}| + |a_{N+3}| + \cdots = \\ &\qquad |a_N| \left(1 + \frac{|a_{N+1}|}{|a_N|} + \frac{|a_{N+2}|}{|a_{N+1}|}\frac{|a_{N+1}|}{|a_N|} + \cdots \right) \\ &\qquad < |a_N| \left(1 + q + q^2 + \cdots \right) = \frac{|a_N|}{1-q} < \infty. \end{aligned} \tag{22}$$

If $L > 1$, then we can reverse the proof (*i.e.* pick q with $1 < q < L$ and N such that $|a_{n+1}|/|a_n| > q \; \forall n \geq N$) to show that the series *diverges*. If $L = 1$, the ratio test does not determine convergence.

1.5 Taylor series

A *power series* has the form

$$\sum_{n=0}^{\infty} c_n (z-a)^n = c_0 + c_1(z-a) + c_2(z-a)^2 + \cdots \tag{23}$$

where the c_n's are complex coefficients and z and a are complex numbers. It is a series in powers of $(z - a)$. By the *ratio test*, the power series converges if

$$\lim_{n\to\infty} \left| \frac{c_{n+1}(z-a)^{n+1}}{c_n(z-a)^n} \right| = |z-a| \lim_{n\to\infty} \left| \frac{c_{n+1}}{c_n} \right| \equiv \frac{|z-a|}{R} < 1, \tag{24}$$

where we have defined

$$\lim_{n\to\infty} \left| \frac{c_{n+1}}{c_n} \right| = \frac{1}{R}. \tag{25}$$

The power series converges if $|z-a| < R$. It diverges if $|z-a| > R$. Since $|z-a| = R$ is a circle of radius R centered at a, R is called the *radius of convergence* of the power series. R can be 0, ∞ or anything in between.

This geometric convergence inside a disk implies that power series can be differentiated (and integrated) term-by-term inside their disk of convergence (why?). The disk of convergence of the derivative or integral series is the same as that of the original series. For instance, the geometric series $\sum_{n=0}^{\infty} z^n$ converges in $|z| < 1$ and its term-by-term derivative $\sum_{n=0}^{\infty} n z^{n-1}$ does also, as you can verify by the ratio test.

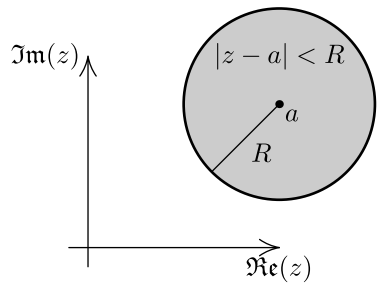


Fig. 3.3: A power series (23) converges in a disk $|z-a|<R$ and diverges outside of that disk. The radius of convergence R can be 0 or ∞.

If a power series converges to $f(z)$ say, then

$$f(z)=\sum_{n=0}^{\infty} c_n (z-a)^n,$$
$$f'(z)=\sum_{n=1}^{\infty} n c_n (z-a)^{n-1},$$
$$f''(z)=\sum_{n=2}^{\infty} n(n-1) c_n (z-a)^{n-2}, \ldots$$

thus

$$f(a)=c_0, \quad f'(a)=c_1, \quad f''(a)=2c_2, \ \ldots$$

and it is straightforward to verify by repeated differentiation of the series and evaluation at $z=a$ that

$$c_n=\frac{1}{n!}\frac{d^n f}{dz^n}(a),$$

in other words, the convergent power series (23) can be written

$$\boxed{\begin{aligned} f(z) &= f(a)+f'(a)(z-a)+\frac{f''(a)}{2}(z-a)^2+\cdots \\ &= \sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n, \end{aligned}} \tag{26}$$

where $f^{(n)}(a)=d^n f/dz^n(a)$ is the nth derivative of $f(z)$ at a and $n!=n(n-1)\cdots 1$ is the factorial of n, with $0!=1$ by convenient definition. The power series (26) is the *Taylor series* of $f(z)$ about $z=a$.

The equality between $f(z)$ and its Taylor series is only valid if the series converges. The geometric series

$$\frac{1}{1-z}=1+z+z^2+\cdots=\sum_{n=0}^{\infty} z^n \tag{27}$$

is the Taylor series of $f(z) = 1/(1-z)$ about $z = 0$, but the function $1/(1-z)$ exists and is infinitely differentiable everywhere except at $z = 1$, while the series $\sum_{n=0}^{\infty} z^n$ only exists in the unit circle $|z| < 1$. The convergence of the series about $z = 0$ is limited by the singularity of $1/(1-z)$ at $z = 1$.

Several useful Taylor series are more easily derived from the geometric series (19), (27) than from the general formula (26) (even if you really like calculating lots of derivatives!). For instance

$$\frac{1}{1-z^2} = 1 + z^2 + z^4 + \cdots = \sum_{n=0}^{\infty} z^{2n} \tag{28}$$

$$\frac{1}{1+z} = 1 - z + z^2 - \cdots = \sum_{n=0}^{\infty} (-z)^n \tag{29}$$

$$\ln(1+z) = z - \frac{z^2}{2} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n z^{n+1}}{n+1} \tag{30}$$

The last series is obtained by integrating both sides of the previous equation and matching at $z = 0$ to determine the constant of integration. These series converge only in $|z| < 1$ while the functions on the left hand side exist for (much) larger domains of z.

Exercises:

1. Explain why the domain of convergence of a power series is always a disk (possibly infinitely large), not an ellipse or a square or any other shape [Hint: read the notes carefully]. (Anything can happen on the boundary of the disk: weak (algebraic) divergence or convergence, perpetual oscillations, etc., recall the geometric series).

2. Show that if a function $f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n$ for all z's within the (non-zero) disk of convergence of the power series, then the c_n's must have the form provided by formula (26).

3. What is the Taylor series of $1/(1-z)$ about $z = 0$? what is its radius of convergence? does the series converge at $z = -2$? why not?

4. What is the Taylor series of the function $1/(1+z^2)$ about $z = 0$? what is its radius of convergence? Use a computer or calculator to test the convergence of the series inside and outside its disk of convergence.

5. What is the Taylor series of $1/z$ about $z = 2$? what is its radius of convergence? [Hint: $z = a + (z-a)$]

6. What is the Taylor series of $1/(1+z)^2$ about $z = 0$?

1.6 Complex transcendentals

We have defined $i^2 = -1$ and made geometric sense of it, now what is 2^i? We can make sense of such complex powers as $\exp(i \ln 2)$ but we first need to define the exponential function $\exp z$ and its inverse the (natural) logarithm, $\ln z$.

The complex versions of the Taylor series definition for the exponential, cosine and sine functions

$$\exp z = 1 + z + \frac{z^2}{2} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{31}$$

$$\cos z = 1 - \frac{z^2}{2} + \frac{z^4}{4!} \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} \tag{32}$$

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} \tag{33}$$

converge in the *entire* complex plane for any z with $|z| < \infty$ as is readily checked from the ratio test. These convergent series serve as the *definition* of these functions for complex arguments.

We can verify the usual properties of these functions from the series expansions. In general, we can integrate and differentiate series term by term inside the disk of convergence of the power series. Doing so for $\exp z$ shows that it is equal to its derivative

$$\frac{d \exp z}{dz} = \frac{d}{dz}\left(\sum_{n=0}^{\infty} \frac{z^n}{n!}\right) = \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = \exp z, \tag{34}$$

meaning that $\exp z$ is the solution of the complex differential equation

$$\frac{df}{dz} = f \quad \text{with} \quad f(0) = 1.$$

Likewise the series (32) for $\cos z$ and (33) for $\sin z$ imply

$$\frac{d}{dz} \cos z = -\sin z, \qquad \frac{d}{dz} \sin z = \cos z. \tag{35}$$

Taking another derivative of both sides shows that $f_1(z) = \cos z$ and $f_2(z) = \sin z$ are solutions of the 2nd order differential equation

$$\frac{d^2 f}{dz^2} = -f,$$

with $(f_1(0), f_1'(0)) = (1, 0)$ and $(f_2(0), f_2'(0)) = (0, 1)$.

Another slight *tour de force* with the series for $\exp(z)$ is to use the binomial formula (13) to obtain

$$\exp(z + a) = \sum_{n=0}^{\infty} \frac{(z+a)^n}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} \frac{z^k a^{n-k}}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k a^{n-k}}{k!(n-k)!}. \tag{36}$$

The double sum is over the triangular region $0 \le n \le \infty$, $0 \le k \le n$ in n, k space. If we interchange the order of summation, we'd have to sum over $k = 0 \to \infty$ and $n = k \to \infty$ (sketch it!). Changing variables to k, $m = n - k$ the range of m is 0 to ∞ as that of k and the double sum reads

$$\exp(z+a) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \frac{z^k a^m}{k!m!} = \left(\sum_{k=0}^{\infty} \frac{z^k}{k!}\right)\left(\sum_{m=0}^{\infty} \frac{a^m}{m!}\right) = \exp(z)\exp(a). \tag{37}$$

This is a major property of the exponential function and we verified it from its series expansion (31) for general complex arguments z and a. It implies that if we define as before

$$e = \exp(1) = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots = 2.71828... \tag{38}$$

then $\exp(n) = [\exp(1)]^n = e^n$ and $\exp(1) = [\exp(1/2)]^2$ thus $\exp(1/2) = e^{1/2}$ etc. so we can still identify $\exp(z)$ as the number e to the *complex power* z and (37) is the regular algebraic rule for exponents: $e^{z+a} = e^z e^a$. In particular

$$\exp z = e^z = e^{x+iy} = e^x e^{iy}, \tag{39}$$

where e^x is the regular exponential function of a real variable x but e^{iy} is the exponential of a pure imaginary number. We can make sense of this from the series (31), (32) and (33) to obtain the very useful formula

$$\begin{aligned} e^{iz} &= \cos z + i \sin z, \\ e^{-iz} &= \cos z - i \sin z, \end{aligned} \tag{40}$$

that yield

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z = \frac{e^{iz} - e^{-iz}}{2i} \tag{41}$$

that could serve as the complex definitions of $\cos z$ and $\sin z$. These hold for any complex number z. For z real, this is *Euler's formula* usually written in terms of a real angle θ

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta.} \tag{42}$$

This is arguably one of the most important formula in all of mathematics! It reduces all of trigonometry to algebra among other things. For instance $e^{i(\alpha+\beta)} = e^{i\alpha}e^{i\beta}$ implies

$$\begin{aligned} \cos(\alpha+\beta) + i\sin(\alpha+\beta) &= (\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) \\ &= (\cos\alpha\cos\beta - \sin\alpha\sin\beta) \\ &\quad + i(\sin\alpha\cos\beta + \sin\beta\cos\alpha) \end{aligned} \tag{43}$$

Equating real and imaginary parts[1] yields two trigonometric identities in one swoop, the angle sum formula from high school trigonometry.

[1] when α, β are real. More generally, equating parts even and odd in (α, β)

Exercises:

1. Use series to compute the number e to 4 digits. How many terms do you need?

2. Use series to compute $\exp(i)$, $\cos(i)$ and $\sin(i)$ to 4 digits.

3. Express $\cos(2+3i)$ in terms of cos, sin and exp of real numbers. Same question for $\sin(2+3i)$.

4. The *hyperbolic* cosine and sine are defined as

$$\begin{aligned} \cosh z &= 1 + \frac{z^2}{2} + \frac{z^4}{4!} \cdots = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!} \\ \sinh z &= z + \frac{z^3}{3!} + \frac{z^5}{5!} \cdots = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!} \end{aligned} \tag{44}$$

   (a) Use the series definition to show that

$$\frac{d\cosh z}{dz} = \sinh z, \qquad \frac{d\sinh z}{dz} = \cosh z, \tag{45}$$

   and that $\cosh z$ and $\sinh z$ are both solutions of the differential equation

$$\frac{d^2 f}{dz^2} = f$$

   with initial conditions $(f(0), f'(0)) = (1, 0)$ for $\cosh z$, and $(f(0), f'(0)) = (0, 1)$ for $\sinh z$ (here $f' = df/dz$).

   (b) Show that

$$\begin{aligned} e^z &= \cosh z + \sinh z, \\ e^{-z} &= \cosh z - \sinh z, \\ \cosh z &= \frac{e^z + e^{-z}}{2} = \cos(iz), \\ \sinh z &= \frac{e^z - e^{-z}}{2} = \frac{\sin(iz)}{i} \end{aligned} \tag{46}$$

   (c) Show that

$$\begin{aligned} \cos z &= \cos(x+iy) = \cos x \cosh y - i \sin x \sinh y, \\ \sin z &= \sin(x+iy) = \sin x \cosh y + i \sinh y \cos x. \end{aligned} \tag{47}$$

   (d) Sketch e^x, e^{-x}, $\cosh x$ and $\sinh x$ for real x.

5. Prove that $\cos^2 z + \sin^2 z = 1$ for any complex number z.

6. Prove that $\cosh^2 z - \sinh^2 z = 1$ for any complex number z.

7. Prove that the following well-known identities hold for complex a and b also:

$$\begin{aligned}
\cos(a+b) &= \cos a \cos b - \sin a \sin b,\\
\sin(a+b) &= \sin a \cos b + \sin b \cos a,\\
\cosh(a+b) &= \cosh a \cosh b + \sinh a \sinh b,\\
\sinh(a+b) &= \sinh a \cosh b + \sinh b \cosh a.
\end{aligned}$$

8. Is $e^{-iz} = \left(e^{iz}\right)^*$?

9. Show that

$$\begin{aligned}
f_1(z) &= 1 + \frac{z^3}{2\cdot 3} + \frac{z^6}{(2\cdot 3)\cdot(5\cdot 6)} + \cdots = \sum_{k=0}^{\infty} 3^k \left(\frac{1}{3}\right)_k \frac{z^{3k}}{(3k)!}\\
f_2(z) &= z + \frac{z^4}{3\cdot 4} + \frac{z^7}{(3\cdot 4)\cdot(6\cdot 7)} + \cdots = \sum_{k=0}^{\infty} 3^k \left(\frac{2}{3}\right)_k \frac{z^{3k+1}}{(3k+1)!}
\end{aligned} \tag{48}$$

are two linearly independent solutions of *Airy's equation*

$$\frac{d^2 f}{dz^2} = z\, f, \tag{49}$$

where $(z)_k = z(z+1)(z+2)\cdots(z+k-1)$ with $(z)_0 = 1$, is the *Pochhammer symbol.* Prove that the series converge in the *entire* complex plane.

10. Prove the identity $\dfrac{1}{1-e^{ix}} = \dfrac{ie^{-ix/2}}{2\sin(x/2)}$.

11. Use Euler's formula and geometric sums to derive the following identities

$$\begin{aligned}
\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx &= \frac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x},\\
\sin x + \sin 2x + \cdots + \sin nx &= \frac{\cos\frac{1}{2}x - \cos(n+\frac{1}{2})x}{2\sin\frac{1}{2}x}.
\end{aligned} \tag{50}$$

These identities are important in the study of waves and Fourier series. Use computer graphing software to plot the sums for a few increasing n. What happens when $n \to \infty$? [Hint: derive both identities at once as in (43).]

12. Show that if p is real with $|p| < 1$ then

$$1 + 2p\cos x + 2p^2\cos 2x + 2p^3\cos 3x + \cdots = \frac{1-p^2}{1+p^2-2p\cos x} \tag{51}$$

otherwise the series diverges if $|p| > 1$. What happens when $p = 1$?

13. The formula (43) leads to the well-known *double* and *triple* angle formula

$$\begin{aligned} \cos 2\theta &= 2\cos^2\theta - 1, & \sin 2\theta &= 2\sin\theta\cos\theta, \\ \cos 3\theta &= 4\cos^3\theta - 3\cos\theta, & \sin 3\theta &= \sin\theta(4\cos^2\theta - 1). \end{aligned} \tag{52}$$

These formula suggests that $\cos n\theta$ is a polynomial of degree n in $\cos\theta$ and that $\sin n\theta$ is $\sin\theta$ times a polynomial of degree $n-1$ in $\cos\theta$. The polynomial for $\cos n\theta$ in powers of $\cos\theta$ is the *Chebyshev* polynomial $T_n(x)$ of degree n in x such that

$$T_n(\cos\theta) = \cos n\theta,$$

thus

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x, \ldots$$

Derive explicit formulas for those polynomials for any n. These Chebyshev polynomials are very important in numerical calculations, your calculator uses them to evaluate all the 'elementary' functions such as e^x, $\cos x$, $\sin x$, etc. [Hint: use Euler's formula $e^{in\theta} = \cos n\theta + i\sin n\theta = (e^{i\theta})^n = (\cos\theta + i\sin\theta)^n$ and the binomial formula].

1.7 Polar representation

Introducing polar coordinates in the complex plane such that $x = r\cos\theta$ and $y = r\sin\theta$, then using Euler's formula (42), any complex number can be written

$$\boxed{z = x + iy = re^{i\theta} = |z|e^{i\arg(z)}.} \tag{53}$$

This is the *polar form* of the complex number z. Its modulus is $|z| = r$ and the angle $\theta = \arg(z) + 2k\pi$ is called the *phase* of z, where $k = 0, \pm 1, \pm 2, \ldots$ is an integer. A key issue is that for a given z, its phase θ is only defined up to an arbitrary multiple of 2π since replacing θ by $\theta \pm 2\pi$ does not change z. However the argument $\arg(z)$ is a function of z and therefore we want it to be uniquely defined for every z. For instance we can define $0 \le \arg(z) < 2\pi$, or $-\pi < \arg(z) \le \pi$. These are just two among an infinite number of possible definitions. Although computer functions (Fortran, C, Matlab, ...) make a specific choice (typically the 2nd one), that choice may not be suitable in some cases. The proper choice is problem dependent. This is because while θ is continuous, $\arg(z)$ is necessarily discontinuous. For example, if we define $0 \le \arg(z) < 2\pi$, then a point moving about the unit circle at angular velocity ω will have a phase $\theta = \omega t$ but $\arg(z) = \omega t \mod 2\pi$ which is discontinuous at $\omega t = 2k\pi$.

The cartesian representation $x + iy$ of a complex number z is perfect for addition/subtraction but the polar representation $re^{i\theta}$ is more convenient for multiplication and division since

$$z_1 z_2 = r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)}, \tag{54}$$

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}. \tag{55}$$

1.8 Logs and complex powers

The power series expansion of functions is remarkably powerful and closely tied to the theory of functions of a complex variable. Many complex functions beyond $\exp z$ can be defined by series. *A priori*, it doesn't seem very general, how, for instance, could we expand $f(z) = 1/z$ into a series in *positive* powers of z

$$\frac{1}{z} = a_0 + a_1 z + a_2 z^2 + \cdots \quad ??$$

We cannot expand $1/z$ in powers of z but we can expand in powers of $z - a$ for any $a \neq 0$. That Taylor series is obtained easily using the geometric series, again,

$$\frac{1}{z} = \frac{1}{a + (z-a)} = \frac{1}{a}\,\frac{1}{1 + \left(\dfrac{z-a}{a}\right)} = \sum_{n=0}^{\infty} (-1)^n \frac{(z-a)^n}{a^{n+1}}. \tag{56}$$

Thus we can expand $1/z$ in powers of $z - a$ for any $a \neq 0$. That (geometric) series converges in the disk $|z - a| < |a|$. This is the disk of radius $|a|$ centered at a. By taking a sufficiently far away from 0, that disk where the series converges can be made as big as one wants but it can never include the origin which of course is the sole *singular point* of the function $1/z$. Integrating (56) for $a = 1$ term by term yields

$$\ln z = \sum_{n=0}^{\infty} (-1)^n \frac{(z-1)^{n+1}}{n+1} \tag{57}$$

as the antiderivative of $1/z$ that vanishes at $z = 1$. This looks nice, however that series only converges for $|z - 1| < 1$. We need a better definition that works for a larger domain in the z-plane.

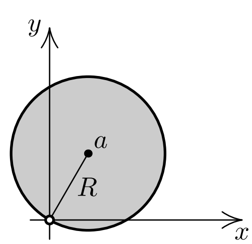


Fig. 3.4: The disk of convergence of the Taylor series (56) for $1/z$ is limited by the singularity at $z = 0$.

The Taylor series definition of the exponential $\exp(z) = \sum_{n=0}^{\infty} z^n/n!$ is very good. It converges for all z's, it led us to Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$ and it allowed us to verify the key property of the exponential, namely $\exp(a+b) = \exp(a)\exp(b)$ (where a and b are any *complex* numbers), from which we deduced other goodies: $\exp(z) \equiv e^z$ with $e = \exp(1) = 2.71828\ldots$, and $e^z = e^{x+iy} = e^x e^{iy}$.

What about $\ln z$? As for functions of a single real variable we can introduce $\ln z$ as the inverse of e^z or as the integral of $1/z$ that vanishes at $z = 1$.

$\ln z$ as the inverse of e^z

Given z we want to define the function $\ln z$ as the inverse of the exponential. This means we want to find a complex number $w = \ln z$ such that $e^w = z$. We can solve this equation for w as a function of z by using the polar representation for z, $z = |z|e^{i\arg(z)}$, together with the cartesian form for w, $w = u+iv$, where $u = \mathfrak{Re}(w)$ and $v = \mathfrak{Im}(w)$ are real. We obtain

$$\begin{aligned} e^w = z \;&\Leftrightarrow\; e^{u+iv} = |z|e^{i\arg(z)}, \\ &\Leftrightarrow\; e^u = |z|, \quad e^{iv} = e^{i\arg(z)}, \quad \text{(why?)} \\ &\Leftrightarrow\; u = \ln|z|, \quad v = \arg(z) + 2k\pi, \end{aligned} \tag{58}$$

where $k = 0, \pm 1, \pm 2, \cdots$ Note that $|z| \geq 0$ is a *positive real number* so $\ln|z|$ is our good old natural log of a positive real number. We have managed to find the inverse of the exponential

$$e^w = z \;\Leftrightarrow\; w = \ln|z| + i\arg(z) + 2ik\pi. \tag{59}$$

The equation $e^w = z$ for w, given z, has an infinite number of solutions. This make sense since $e^w = e^u e^{iv} = e^u(\cos v + i\sin v)$ is periodic of period 2π in v, so if $w = u + iv$ is a solution, so is $u + i(v + 2k\pi)$ for any integer k. We can take any *one* of those solutions as our definition of $\ln z$, in particular

$$\boxed{\ln z = \ln\left(|z|e^{i\arg(z)}\right) = \ln|z| + i\arg(z).} \tag{60}$$

This definition is unique since we assume that $\arg z$ is uniquely defined in terms of z. However different definitions of $\arg z$ lead to different definitions of $\ln z$.

Example: If $\arg(z)$ is defined by $0 \leq \arg(z) < 2\pi$ then $\ln(-3) = \ln 3 + i\pi$, but if we define instead $-\pi \leq \arg(z) < \pi$ then $\ln(-3) = \ln 3 - i\pi$.

Note that you can now take logs of negative numbers! Note also that the $\ln z$ definition fits with our usual manipulative rules for logs. In particular since $\ln(ab) = \ln a + \ln b$ then $\ln z = \ln(re^{i\theta}) = \ln r + i\theta$. This is the easy way to remember what $\ln z$ is.

Complex powers

As for functions of real variables, we can now define general complex powers such as 2^i in terms of the complex log and the complex exponential

$$a^b = e^{b\ln a} = e^{b\ln|a|}e^{ib\arg(a)}, \tag{61}$$

for example: $2^i = e^{i\ln 2} = \cos(\ln 2) + i\sin(\ln 2)$, is that what your calculator gives you? Be careful that b is complex in general, so $e^{b\ln|a|}$ is not necessarily real. Once again we need to define $\arg(a)$ and different definitions can actually lead to different values for a^b.

In particular, we have the complex power functions

$$z^a = e^{a\ln z} = e^{a\ln|z|}e^{ia\arg(z)} \tag{62}$$

and the complex exponential functions

$$a^z = e^{z \ln a} = e^{z \ln |a|} e^{iz \arg(a)}. \tag{63}$$

These functions are well-defined once we have defined a range for $\arg(z)$ in the case of z^a and for $\arg(a)$ in the case of a^z.

Note that different definitions for the $\arg(a)$ imply definitions for a^b that do *not* simply differ by an *additive multiple* of $2\pi i$ as was the case for $\ln z$. For example

$$(-1)^i = e^{i \ln(-1)} = e^{-\arg(-1)} = e^{-\pi - 2k\pi}$$

for some k, so the various possible definitions of $(-1)^i$ will differ by a *multiplicative integer power* of $e^{-2\pi}$.

Roots

The **fundamental theorem of algebra** states that any nth order polynomial equation of the form $c_n z^n + c_{n-1} z^{n-1} + \cdots + c_1 z + c_0 = 0$ with $c_n \neq 0$ always has n roots in the complex plane, where n is a positive integer. This means that there exists n complex numbers $z_1, \ldots, z_n$ such that

$$c_n z^n + c_{n-1} z^{n-1} + \cdots + c_0 = c_n (z - z_1) \cdots (z - z_n). \tag{64}$$

The numbers $z_1, \ldots, z_n$ are the *roots* or *zeros* of the polynomial. These roots can be repeated as for the polynomial $2z^2 - 4z + 2 = 2(z-1)^2$. This expansion is called *factoring* the polynomial.

Examples: The equation $2z^2 - 2 = 0$ has two real roots $z = \pm 1$ and

$$2z^2 - 2 = 2(z-1)(z+1).$$

The equation $3z^2 + 3 = 0$ has no real roots, however it has two imaginary roots $z = \pm i$ and

$$3z^2 + 3 = 3(z-i)(z+i).$$

$\square$

Roots of a. The equation

$$z^n = a,$$

with a complex and integer $n > 0$, therefore has n roots. We might be tempted to write the solution as

$$z = a^{1/n} = e^{(\ln a)/n} = e^{(\ln |a|)/n} e^{i \arg(a)/n},$$

but this is only *one root* whose value depends on the definition of $\arg(a)$. The n-th root function, $a^{1/n}$, must have a unique value for a given a but here we are looking for *all* the z's such that $z^n = a$. Using the polar representations $z = |z| e^{i \arg(z)}$ and $a = |a| e^{i \arg(a)}$ yields

$$z^n = a \;\Leftrightarrow\; |z|^n e^{i\, n \arg(z)} = |a| e^{i \arg(a)}.$$

Now $a = b$ implies

$$\begin{cases} |a| = |b|, \\ \arg(a) = \arg(b) + 2k\pi, \end{cases}$$

the magnitudes are equal $|a| = |b|$ but the angles are equal only up to a multiple of 2π. Indeed $e^{i\pi/2} = i = e^{-i3\pi/2}$ for example, and $\pi/2 \neq -3\pi/2$ but $\pi/2 = -3\pi/2 + 2\pi$. The equation $z^n = a$ thus implies that

$$|z^n| = |z|^n = |a|, \qquad \arg(z^n) = n\arg(z) = \arg(a) + 2k\pi$$

where $k = 0, \pm 1, \pm 2, \ldots$ is any integer. Solving for the real $|z| \geq 0$ and $\arg(z)$ yields

$$|z| = |a|^{1/n}, \qquad \arg(z) = \frac{\arg(a)}{n} + k\frac{2\pi}{n}. \tag{65}$$

When n is a positive integer, this yields n distinct values for $\arg(z)$ modulo 2π, yielding n distinct values for z. The roots are equispaced by angle $2\pi/n$ on the circle of radius $|a|^{1/n}$ in the complex plane.

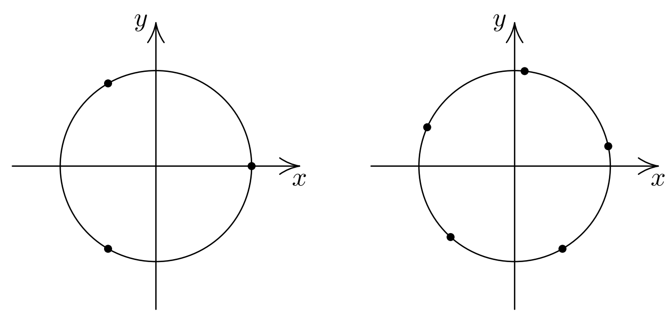


Fig. 3.5: The 3 roots of $z^3 = 1$ and the 5 roots of $z^5 = e^{i\pi/3}$.

For example

$$z^3 = 1 \;\Leftrightarrow\; \begin{cases} z = 1, \\ z = e^{i2\pi/3} = (-1 + i\sqrt{3})/2, \\ z = e^{i4\pi/3} = (-1 - i\sqrt{3})/2 = e^{-i2\pi/3}, \end{cases}$$

the 3 roots are on the unit circle, equispaced by $2\pi/3$, and

$$z^3 - 1 = (z - 1)(z - e^{i2\pi/3})(z - e^{-i2\pi/3}).$$

Likewise, the 5 roots of $z^5 = e^{i\pi/3}$ are $z = e^{i\pi/15}e^{i2k\pi/5}$, for $k = 0, \pm 1, \pm 2, \ldots$, they lie on the unit circle equispaced by $2\pi/5$.

Exercises:

1. Evaluate $\ln(-1)$ and $\ln i$ *without* a calculator.

2. Analyze and evaluate e^i and i^e.

3. Find all the roots of $z^5 = e^{i\pi/3}$. Do your roots match those in fig. 3.5?

4. Find all the roots, visualize and locate them in the complex plane and factor the corresponding polynomial (i) $z^2 + 1 = 0$, (ii) $z^3 + 1 = 0$, (iii) $z^4 + 1 = 0$, (iv) $z^2 = i$, (v) $2z^2 + 5z + 2 = 0$, (vi) $2\,z^5 + 1 + i = 0$.

5. Investigate the solutions of the equation $z^b = 1$ when (i) b is a rational number, that is $b = p/q$ with p, q integers, (ii) when b is irrational *e.g.* $b = \pi$, (iii) when b is complex, *e.g.* $b = 1 + i$. Make a sketch showing the solutions in the complex plane.

6. If a and b are complex numbers, what is wrong with saying that if $w = e^{a+ib} = e^a e^{ib}$ then $|w| = e^a$ and $\arg(w) = b + 2k\pi$? Isn't that what we did in (58)?

7. Consider $w = \sqrt{z}$ for $z = re^{i\varphi}$ with $r \geq 0$ fixed and $\varphi = 0 \to 4\pi$. Sketch $u = \mathfrak{Re}(w)$ and $v = \mathfrak{Im}(w)$ as functions of φ from $\varphi = 0 \to 4\pi$ for

$$\text{(i)}\ \sqrt{z} \triangleq \sqrt{r}\, e^{i\varphi/2}, \qquad \text{(ii)}\ \sqrt{z} \triangleq \sqrt{r}\, e^{i\ \arg(z)/2}.$$

   Why is (i) not a valid definition of $\sqrt{z}$?

8. Use Matlab, Python or your favorite complex calculator to evaluate $\sqrt{z^2-1}$ and $\sqrt{z-1}\sqrt{z+1}$ for $z = -1 + i$. Since $z^2 - 1 = (z-1)(z+1)$, shouldn't these square roots equal each other? Explain.

9. Show that the inverse of $w = \cosh z$, that is *one* solution w to $\cosh w = z$ is

$$w = \ln(z + \sqrt{z^2 - 1}), \tag{66}$$

   where $\sqrt{z^2-1}$ is defined with the branch cut $z \in [-1, 1]$, that is $\varphi_{1,2} = \arg(z \pm 1)$ with $-\pi < \arg(\cdot) \leq \pi$ and

$$\sqrt{z^2 - 1} \triangleq \sqrt{|z^2 - 1|}\, e^{i\varphi_1/2} e^{i\varphi_2/2}.$$

10. Show that the inverse of $w = \sinh z$, that is *one* solution w to $\sinh w = z$ is

$$w = \ln(z + \sqrt{z^2 + 1}), \tag{67}$$

   with branch cuts $z \in i(-\infty, -1) \cup i(1, \infty)$, that is $\sqrt{z^2+1} = \sqrt{|z^2+1|}e^{i\theta/2}$ with $\theta = \arg(z^2 + 1)$.

2 Functions of a complex variable

2.1 Visualization of complex functions

A function $w = f(z)$ of a complex variable $z = x+iy$ has complex values $w = u+iv$, where u, v are real. The real and imaginary parts of $w = f(z)$ are functions of the real variables x and y

$$f(z) = u(x, y) + i\, v(x, y). \tag{68}$$

For example, $w = z^2 = (x + iy)^2$ is

$$w = z^2 = (x^2 - y^2) + i\, 2xy \tag{69}$$

with a real part $u = x^2 - y^2$ and an imaginary part $v = 2xy$.

How do we visualize complex functions? In calculus I, for real functions of one real variable, $y = f(x)$, we made an xy plot. In complex calculus, x and y are independent variables and $w = f(z)$ corresponds to *two* real functions of *two* real variables $\mathfrak{Re}(f(z)) = u(x, y)$ and $\mathfrak{Im}(f(z)) = v(x, y)$. One way to visualize $f(z)$ is to make a 3D plot with u as the *height* above the (x, y) plane. We could do the same for $v(x, y)$, however a prettier idea is to *color* the surface $u = u(x, y)$ in the 3D space (x, y, u) by the value of $v(x, y)$ as in fig. 3.6.

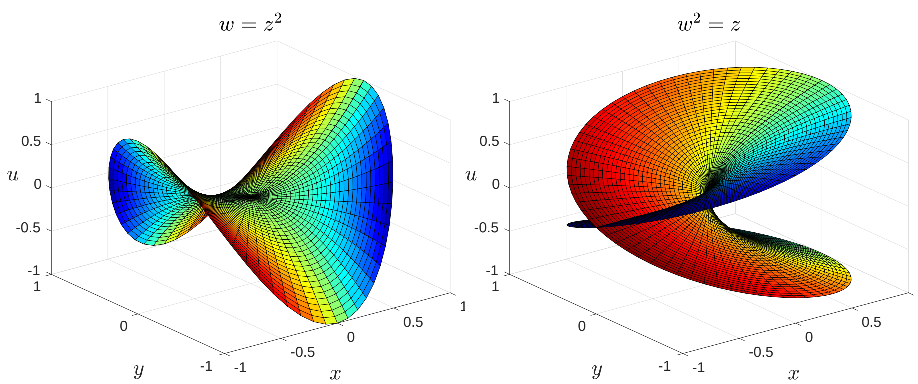


Fig. 3.6: 3D visualization of $w = z^2$ and $w^2 = z$ with $u = \mathfrak{Re}(w)$ as the height over the (x, y) plane and $v = \mathfrak{Im}(w)$ as the color.

The left-side graph in fig. 3.6 shows $\mathfrak{Re}(w) = u = x^2 - y^2 = \mathfrak{Re}(z^2)$. It is the upright parabola $u = x^2$ along the real axis $y = 0$, but $u = -y^2$ along the imaginary axis, $x = 0$. The right-side graph of $w^2 = z$ shows the two roots of that quadratic equation, that is $w = \pm\sqrt{z}$ computed as $z = re^{i\varphi}$ and $w = \sqrt{r}\, e^{i\varphi/2}$ but for a double cover of the z plane, that is $\varphi = 0 \to 4\pi$, not just 2π. This shows that those two roots are smoothly connected as one circles around the *branch point* $z = 0$ where both roots collide. The standard $\sqrt{z}$ function is $\sqrt{z} = \sqrt{|z|}e^{i \arg(z)/2}$ with $-\pi < \arg(z) \leq \pi$. This selects the root with positive real part $u \geq 0$, that is the upper surface on the $w^2 = z$ graph.

This type of 3D visualization yields beautiful pictures – and illustrates the concept of a *Riemann surface* – but they quickly become too complicated, in part because of the 2D projection on the page. It is usually more useful to make 2D *contour plots* of $u(x,y)$ and $v(x,y)$ as in fig. 3.7.

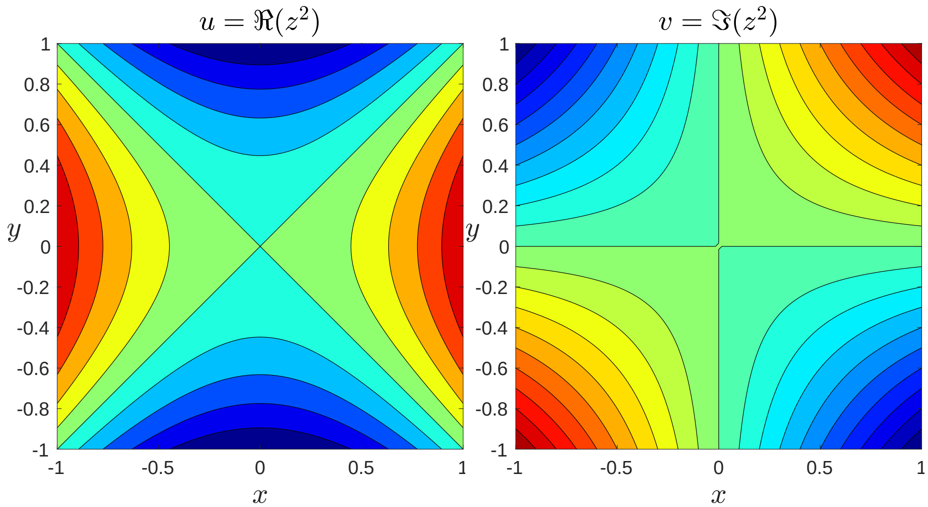


Fig. 3.7: 2D visualization of $w = z^2$ with contour plots of $u = \mathfrak{Re}(w) = x^2 - y^2$ on the left and $v = \mathfrak{Im}(w) = 2xy$ on the right. Note the saddle structures in both u and v.

2.2 Cauchy-Riemann equations

A function $f(z)$ of a complex variable $z = x + iy$ is a special function $f(x+iy)$ of two real variables (x,y), and consists of two real functions $u(x,y)$ and $v(x,y)$ of two real variables

$$f(z) = f(x+iy) = u(x,y) + i\,v(x,y). \tag{70}$$

The example $z^2 = (x^2 - y^2) + i\,(2xy)$ was discussed and illustrated above. Other examples are

$$e^z = e^{x+iy} = e^x \cos y + i\,e^x \sin y$$

for which $u = e^x \cos y$ and $v = e^x \sin y$, and

$$\frac{1}{z} = \frac{1}{x+iy} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}$$

that has $u = x/(x^2+y^2)$ and $v = -y/(x^2+y^2)$.

Now, if $f(z)$ is z-differentiable, then

$$\frac{\mathrm{d}f(z)}{\mathrm{d}z} = \lim_{\Delta z\to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z}$$
$$= \lim_{\Delta z\to 0} \frac{f\left(x+\Delta x + i(y+\Delta y)\right) - f(x+iy)}{\Delta x + i\Delta y}$$

has the same value no matter how $\Delta z = \Delta x + i\Delta y \to 0$. Picking $\Delta z = \Delta x$ with $\Delta y = 0$ gives

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial f}{\partial x},$$

while picking $\Delta z = i\Delta y$ with $\Delta x = 0$ yields

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{1}{i}\frac{\partial f}{\partial y},$$

thus

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial f}{\partial x} = -i\,\frac{\partial f}{\partial y}, \tag{71}$$

and since $f = u + iv$ this is

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{72}$$

Equating the real and imaginary part of that last equation yields the *Cauchy-Riemann* equations

$$\boxed{\begin{aligned}\frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y},\\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}\end{aligned}} \tag{73}$$

relating the partial derivatives of the real and imaginary part of a differentiable function of a complex variable $f(z) = u(x,y) + i\,v(x,y)$. This derivation shows that the Cauchy-Riemann equations are *necessary* conditions on $u(x,y)$ and $v(x,y)$ if $f(z)$ is differentiable in a neighborhood of z. If $\mathrm{d}f/\mathrm{d}z$ exists then the Cauchy-Riemann equations (73) *necessarily* hold. They are also *sufficient* as shown later below.

Example 1: The function $f(z) = z^2$ has $u = x^2 - y^2$ and $v = 2xy$. Its z-derivative $\mathrm{d}z^2/\mathrm{d}z = 2z$ exists everywhere and the Cauchy-Riemann equations (73) are satisfied everywhere since

$$\begin{cases} \dfrac{\partial u}{\partial x} = 2x = \dfrac{\partial v}{\partial y}, \\ \dfrac{\partial u}{\partial y} = -2y = -\dfrac{\partial v}{\partial x}. \end{cases}$$

□

Example 2: The function $f(z) = \ln z = \ln|z| + i\arg z$ has

$$u = \ln\sqrt{x^2+y^2} \quad \text{and} \quad v = \operatorname{atan2}(y,x).$$

Its z-derivative $\mathrm{d}\ln z/\mathrm{d}z = 1/z$ appears to exist everywhere except at $z = 0$, however $v = \arg z = \operatorname{atan2}(y, x)$ jumps by $\pm 2\pi$ as z crosses the negative real axis $x < 0$, $y = 0$, with the standard definition $-\pi < \arg z \le \pi$. Thus $f'(z)$ exists everywhere and the Cauchy-Riemann equations apply *except along the negative real axis* which is a *branch cut* for $\arg z$ and $\ln z$,

$$\begin{cases} \dfrac{\partial u}{\partial x} = \dfrac{x}{x^2+y^2} = \dfrac{\partial v}{\partial y}, \\ \dfrac{\partial u}{\partial y} = \dfrac{y}{x^2+y^2} = -\dfrac{\partial v}{\partial x}. \end{cases}$$

□

Example 3: The function $f(z) = z^* = x - iy$ has $u = x$, $v = -y$. The Cauchy-Riemann equations (73) do *not* hold anywhere for $f(z) = z^* = x - iy$ since

$$\frac{\partial u}{\partial x} = 1 \ne \frac{\partial v}{\partial y} = -1.$$

Its z-derivative $\mathrm{d}z^*/\mathrm{d}z$ does *not* exist anywhere. Indeed, from the limit definition of the derivative,

$$\frac{\mathrm{d}z^*}{\mathrm{d}z} = \lim_{a\to 0} \frac{(z^*+a^*) - z^*}{a} = \lim_{a\to 0} \frac{a^*}{a} = e^{-2i\alpha} \tag{74}$$

where $\Delta z = a = |a|e^{i\alpha}$, so the limit is different for every α. If a is real, then $\alpha = 0$ and the limit is 1, but if a is imaginary then $\alpha = \pi/2$ and the limit is -1. If $|a| = e^{-\alpha}$ then $a \to 0$ in a *logarithmic* spiral as $\alpha \to \infty$, but there is no limit in that case since $e^{-2i\alpha}$ keeps spinning around the unit circle without ever converging to anything. We cannot define a unique limit as $a \to 0$, so z^* is not differentiable with respect to z. □

The Cauchy-Riemann equations (73) are also *sufficient* for $f(x, y) = u(x, y) + iv(x, y)$ to be differentiable with respect to $z = x + iy$. Functions $u(x, y)$ and $v(x, y)$ that satisfy the Cauchy-Riemann equations are called *conjugate harmonic functions*. They are the real and imaginary part of a differentiable function $f(x + iy)$.

To prove this we need to show that the z-derivative of the function $f(x, y) \triangleq u(x, y) + iv(x, y)$, that is

$$\frac{\mathrm{d}f}{\mathrm{d}z} \triangleq \lim_{\Delta z\to 0} \frac{f(x+\Delta x, y+\Delta y) - f(x, y)}{\Delta z}$$

exists independently of how the limit $\Delta z = \Delta x + i\Delta y \to 0$ is taken. Using differential notation to simplify the writing, we have $\mathrm{d}f = \mathrm{d}u + i\mathrm{d}v$ with $\mathrm{d}z = \mathrm{d}x + i\mathrm{d}y$ and

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\mathrm{d}u + i\mathrm{d}v}{\mathrm{d}x + i\mathrm{d}y} \tag{75}$$

where $\mathrm{d}u = u(x + \mathrm{d}x, y + \mathrm{d}y) - u(x, y)$ and $\mathrm{d}v = v(x + \mathrm{d}x, y + \mathrm{d}y) - v(x, y)$. By the chain rule,

$$\mathrm{d}u = \frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y, \qquad \mathrm{d}v = \frac{\partial v}{\partial x}\mathrm{d}x + \frac{\partial v}{\partial y}\mathrm{d}y,$$

substituting those differentials in (75) and rearranging terms yields

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right)\frac{\mathrm{d}x}{\mathrm{d}x + i\mathrm{d}y} + \left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right)\frac{\mathrm{d}y}{\mathrm{d}x + i\mathrm{d}y} \tag{76a}$$

in the sense of limits, that is

$$\frac{\mathrm{d}x}{\mathrm{d}x + i\mathrm{d}y} \equiv \lim_{\Delta z \to 0} \frac{\Delta x}{\Delta x + i\Delta y}, \qquad \frac{\mathrm{d}y}{\mathrm{d}x + i\mathrm{d}y} \equiv \lim_{\Delta z \to 0} \frac{\Delta y}{\Delta x + i\Delta y}.$$

Picking $\Delta z = \Delta x$ with $\Delta y = 0$, (76a) yields

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \tag{76b}$$

while $\Delta z = i\Delta y$ with $\Delta x = 0$ gives

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{1}{i}\left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right) = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}. \tag{76c}$$

These $\mathrm{d}f/\mathrm{d}z$ expressions (76b) and (76c) are different, for general $u(x, y)$ and $v(x, y)$, *unless* u and v satisfy the Cauchy-Riemann equations (73). In that case, (76a) becomes

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}z} &= \left(\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}\right)\frac{\mathrm{d}x}{\mathrm{d}x + i\mathrm{d}y} + \left(\frac{\partial u}{\partial y} + i\frac{\partial u}{\partial x}\right)\frac{\mathrm{d}y}{\mathrm{d}x + i\mathrm{d}y} \\ &= \left(\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}\right)\frac{\mathrm{d}x + i\mathrm{d}y}{\mathrm{d}x + i\mathrm{d}y} = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y},\end{aligned} \tag{77}$$

and the value of $\mathrm{d}f/\mathrm{d}z$ is independent of how $\Delta z \to 0$. $\square$

Another important consequence of z-differentiability and the Cauchy-Riemann equations (73) is that the real and imaginary parts of a differentiable function $f(z) = u(x, y) + i\,v(x, y)$ both satisfy *Laplace's equation* (problem 1)

$$\boxed{\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}.} \tag{78}$$

Since both the real and imaginary parts of $f(z) = u(x, y) + iv(x, y)$ satisfy Laplace's equation, this is also true for $f(z)$ seen as a function of (x, y) and

$$\nabla^2 f = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f = 0. \tag{79}$$

Complex differentiability of a function $f(z)$ of a complex variable z implies the Cauchy-Riemann and Laplace's equations. In fact, z-differentiability of $f(z)$ in a neighborhood of z implies that $f(z)$ is infinitely differentiable in the neighborhood of z and that its Taylor series converges in a disk in that neighborhood. This follows from Cauchy's formula as shown later below. A function whose Taylor series converges in the neighborhood of a point is called *analytic* in that neighborhood. Since z-differentiability implies analyticity, the word *analytic* tends to be used interchangeably with complex differentiable.

Exercises:

1. Deduce (78) from the Cauchy-Riemann equations (73). Find $u(x,y)$ and $v(x,y)$ and verify (73) and (78) for (i) $f(z) = z^2$, (ii) z^3, (iii) e^z and (iv) $1/z$.

2. Given $u(x,y)$ find its conjugate harmonic function $v(x,y)$, if possible, such that $u(x,y) + iv(x,y) \equiv f(z)$ is z-differentiable, for (i) $u = y$, (ii) $u = x + y$, (iii) $u = \cos x \cosh y$, (iv) $v = \ln \sqrt{x^2 + y^2}$.

3. Is the function $|z|$ differentiable with respect to z? Why? What about the functions $\mathfrak{Re}(z)$?

4. Find $u(x,y)$ and $v(x,y)$ for $f(z) = \sqrt{z} = \sqrt{|z|} e^{i \arg(z)/2}$. Show that $f(z)$ is differentiable everywhere except across the negative real axis $x \le 0$, $y = 0$ with the standard definition $-\pi < \arg(z) \le \pi$.

5. Consider
$$f(z) = \sqrt{z^2 - 1} \triangleq \sqrt{|z^2 - 1|}\, e^{\frac{i}{2} \arg(z^2-1)}.$$
For the standard definition $-\pi < \arg(z^2 - 1) \le \pi$, show that $\arg(z^2 - 1)$ jumps by $\pm 2\pi$ when z crosses the imaginary axis *or* the real segment $[-1, 1]$. Show that $f(z)$ is not differentiable for such z but is continuous and differentiable for all other z.

6. Consider
$$f(z) = \sqrt{z^2 - 1} \triangleq \sqrt{z - 1}\sqrt{z + 1} = \sqrt{r_1 r_2}\, e^{i(\theta_1 + \theta_2)/2},$$
where $r_{1,2} = |z \pm 1|$ and $\theta_{1,2} = \arg(z \pm 1)$. For the standard definition $-\pi < \arg(\cdot) \le \pi$, show that $\theta_1 + \theta_2$ jumps by $\pm 2\pi$ across $(-1, 1)$ but by $\pm 4\pi$ across $(-\infty, -1)$. Deduce that $f(z)$ is not differentiable on $[-1, 1]$ but is continuous and differentiable for all other z, including $x < -1$, $y = 0$.

7. Consider
$$f(z) = \ln \frac{z + a}{z - a} = \ln \left| \frac{z + a}{z - a} \right| + i \left(\arg(z + a) - \arg(z - a) \right)$$
where a is an arbitrary *positive real* number and $-\pi < \arg(z \pm a) \le \pi$. Show that $f(z)$ is continuous across the semi-infinite real line $(-\infty, -a)$ but jumps by $\pm 2\pi i$ when z crosses the real segment $(-a, a)$. Conclude that $f(z)$ is differentiable everywhere except along the *branch cut* $[-a, a]$. Show that the real and imaginary part of $f(z)$ correspond to the *bipolar coordinates* of chapter 1 and therefore that those functions satisfy Laplace's equation everywhere, except for $z = \pm a$ for the real part, and the *branch cut* $[-a, a]$ for the imaginary part.

8. Complex numbers $z = x + iy$ correspond to vectors $\mathbf{r} = (x, y)$ in $\mathbb{R}^2$. The complex form of the gradients of the real functions $u(x,y)$ and $v(x,y)$ are thus
$$\boldsymbol{\nabla} u \equiv \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \right) \equiv \frac{\partial u}{\partial x} + i \frac{\partial u}{\partial y}, \qquad \boldsymbol{\nabla} v \equiv \frac{\partial v}{\partial x} + i \frac{\partial v}{\partial y}.$$

Show that if $w(z) = u(x,y) + iv(x,y)$ is z-differentiable then

$$\boldsymbol{\nabla} u \equiv \left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)^*, \qquad \boldsymbol{\nabla} v \equiv i\left(\frac{\mathrm{d}w}{\mathrm{d}z}\right)^*. \tag{80}$$

This is useful in 2D fluid dynamics where $w(z)$ represents a complex potential whose real part u is a velocity potential and imaginary part v is a streamfunction. The relationships (80) yield the velocity directly from the complex potential $w(z)$. For example if $w(z) = z^3$ then the velocity is $(3z^2)^*$, that is $\boldsymbol{v} = 3(x^2 - y^2)\,\hat{\boldsymbol{x}} - 6xy\,\hat{\boldsymbol{y}}$.

9. Show that $u = \cos(kx)e^{-ky}$ and $v = \sin(kx)e^{-ky}$ are solutions of Laplace's equation for any real k (i) by direct calculation and (ii) by finding a z differentiable function $f(z) = u(x,y) + iv(x,y)$. These solutions occur in a variety of applications, *e.g.* surface gravity waves with the surface at $y = 0$ and $ky < 0$.

10. Show that $w = z^3$ provides two solutions of Laplace's equation $\nabla^2\Phi = 0$ in the polar wedge $r \geq 0$, $0 < \varphi < \pi/3$, one solution to the *Dirichlet problem* with $\Phi = 0$ on the boundaries $\varphi = 0$ and $\varphi = \pi/3$, and another solution to the *Neumann problem* with $\partial\Phi/\partial\varphi = 0$ on $\varphi = 0, \pi/3$. Specify Φ for both problems in both cartesian coordinates (x,y) and polar coordinates (r,φ).

11. Generalizing the previous problem, show that $w = z^{3n}$ with n a non-zero integer, provide two solutions to Laplace's equation in the same polar wedge $r \geq 0$, $0 < \varphi < \pi/3$, one solution Φ_n with $\partial\Phi_n/\partial\varphi = 0$ for $\varphi = 0, \pi/3$ and one solution Ψ_n with $\Psi_n = 0$ for $\varphi = 0, \pi/3$. Specify Φ_n and Ψ_n in both cartesian coordinates (x,y) and polar coordinates (r,φ).

12. Generalize the previous problems by finding a solution Φ of Laplace's equation in the wedge $r \geq 0$, $0 < \varphi < \alpha$ that vanishes on the boundaries of the wedge at $\varphi = 0$ and $\varphi = \alpha$, where α is a constant such that $0 < \alpha \leq 2\pi$. What is the behavior of the gradient of Φ as $r \to 0$ as a function of α? Can you find other non-zero solutions to the same problem?

2.3 Geometry of Cauchy-Riemann, Conformal Mapping

The Cauchy-Riemann equations connecting the real and imaginary part of a complex differentiable function $f(z) = u(x,y) + i\,v(x,y)$ have great geometric implications illustrated in fig. 3.8 and figures below.

Orthogonality

The Cauchy-Riemann equations (73) imply *orthogonality of the contours* of $u(x,y) = \mathfrak{Re}(f(z))$ and $v(x,y) = \mathfrak{Im}(f(z))$ wherever $\mathrm{d}f(z)/\mathrm{d}z$ exists but does not vanish. Indeed, consider the gradients $\boldsymbol{\nabla} u \equiv (\partial u/\partial x, \partial u/\partial y)$ and $\boldsymbol{\nabla} v \equiv (\partial v/\partial x, \partial v/\partial y)$ at a point (x,y). The Cauchy-Riemann equations (73) imply that those two gradients are perpendicular to each other

$$\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v = \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} = \frac{\partial v}{\partial y}\frac{\partial v}{\partial x} - \frac{\partial v}{\partial x}\frac{\partial v}{\partial y} = 0. \tag{81}$$

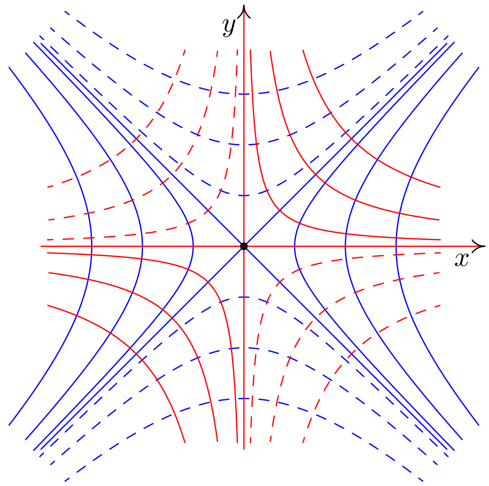


Fig. 3.8: For $z^2 = (x^2 - y^2) + i\,2xy$, the contours of $u(x, y) = x^2 - y^2 = 0, \pm 1, \pm 4, \pm 9$ (blue) are *hyperbolas* with asymptotes $y = \pm x$. The contours of $v(x, y) = 2xy = 0, \pm 1, \pm 4, \pm 9$ (red) are also *hyperbolas* but now with asymptotes $x = 0$ and $y = 0$. Solid is positive, dashed is negative. The u and v contours intersect everywhere at 90 degrees, except at $z = 0$ where $dz^2/dz = 2z = 0$.

Since gradients are always perpendicular to their respective isocontours, $\boldsymbol{\nabla} u$ is perpendicular to the u-contour and $\boldsymbol{\nabla} v$ to the v-contour through that point, the orthogonality of $\boldsymbol{\nabla} u$ and $\boldsymbol{\nabla} v$ imply orthogonality of the contours (level curves) of u and v.

Orthogonality of the u and v contours for $u = \mathfrak{Re}(w)$ and $v = \mathfrak{Im}(w)$ with $w = f(z)$ holds wherever $\mathrm{d}w/\mathrm{d}z$ exists except at critical points where $\mathrm{d}w/\mathrm{d}z = 0$ and $\boldsymbol{\nabla} u = \boldsymbol{\nabla} v = 0$. For the example $w = z^2$ in fig. 3.8, the contours are orthogonal everywhere except at $z = 0$ where $\mathrm{d}z^2/\mathrm{d}z = 2z = 0$. The gradients vanish at that point which is a saddle point for both functions u and v.

The real and imaginary parts, $u(x, y)$ and $v(x, y)$, of any z-differentiable complex function $f(z)$ therefore provide orthogonal coordinates in the (x, y) plane. But the Cauchy-Riemann equations also imply that the gradients of u and v are not only orthogonal, but also have equal magnitudes. This implies that such (u, v) coordinates are not only orthogonal but also *conformal*, as discussed hereafter.

Conformal Mapping

We can visualize the function $w = f(z) = u(x, y) + iv(x, y)$ as a *map* from the complex plane $z = x + iy$ to the complex plane $w = u + iv$. For example, the map $z \to w = z^2$ is illustrated in figs. 3.9 and 3.12.

In cartesian form, that map is $z = x + iy \to w = u + iv = (x^2 - y^2) + i(2xy)$. The vertical line $u = u_0$ in the w-plane is the image of the hyperbola $x^2 - y^2 = u_0$ in the z-plane. The horizontal line $v = v_0$ in the w-plane is the image of the hyperbola $2xy = v_0$ in the z-plane. Every point z in the z-plane has a single image w in the

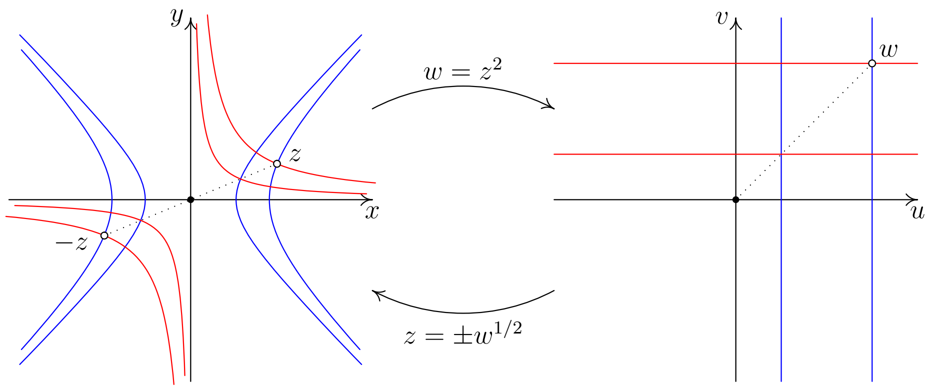


Fig. 3.9: $w = z^2$ as a mapping from the z-plane to the w-plane. There are two possible inverse maps $z = \pm\sqrt{w}$ since there are two z's for every w. The angles between corresponding curves are preserved. For example, the dotted line in the w-plane intersects the vertical and horizontal lines at $\pi/4$ and likewise for its pre-image in the z-plane intersecting the hyperbolas corresponding to vertical and horizontal lines in the w plane.

w-plane, however the latter has two *pre-images* z and $-z$ in the z-plane, indeed the inverse functions are $z = \pm w^{1/2}$, as illustrated in fig. 3.9.

The curves corresponding to constant u and constant v intersect at 90 degrees in *both* planes, except at $z = w = 0$. That is the orthogonality of u and v contours, but the dotted radial line intersects the blue and red curves at 45 degrees in *both* planes, for example. In fact *any angle* between any two curves in the z-plane is preserved in the w-plane *except at* $z = w = 0$ where they are *doubled* from the z to the w-plane.

In polar form $z = re^{i\theta} \to w = r^2 e^{i2\theta}$. This means that every radial line from the origin with angle θ from the x-axis in the z-plane is mapped to a radial line from the origin with angle 2θ from the u-axis in the w-plane.

Angle preservation

That is the conformal map property of z-differentiable complex functions $f(z)$. If $f(z)$ is z-differentiable (analytic) then the mapping $w = f(z)$ *preserves all angles* at all z's where $\mathrm{d}f(z)/\mathrm{d}z \neq 0$ exists *and* does not vanish.

To show preservation of angles in general, consider three neighboring points in the z-plane: z, $z + \Delta z_1$ and $z + \Delta z_2$. We are interested in seeing what happens to the angle between the two vectors Δz_1 and Δz_2. If $\Delta z_1 = |\Delta z_1| e^{i\theta_1}$ and $\Delta z_2 = |\Delta z_2| e^{i\theta_2}$ then the angle between those two vectors is $\alpha = \theta_2 - \theta_1$ and this is the angle of the ratio

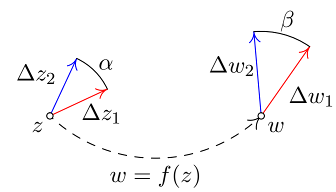
$$\frac{\Delta z_2}{\Delta z_1} = \frac{|\Delta z_2|}{|\Delta z_1|}\, e^{i(\theta_2 - \theta_1)} = \frac{|\Delta z_2|}{|\Delta z_1|}\, e^{i\alpha}.$$

The point z is mapped to the point $w = f(z)$. The points $z + \Delta z_1$ and $z + \Delta z_2$ are mapped to

$$\begin{aligned} w_1 &= f(z + \Delta z_1) \simeq f(z) + f'(z)\Delta z_1 + \frac{1}{2} f''(z)\Delta z_1^2 + \cdots \\ w_2 &= f(z + \Delta z_2) \simeq f(z) + f'(z)\Delta z_2 + \frac{1}{2} f''(z)\Delta z_2^2 + \cdots \end{aligned} \tag{82}$$

respectively, where $f' = \mathrm{d}f/\mathrm{d}z$, $f'' = \mathrm{d}^2 f/\mathrm{d}z^2$, etc.

In general, a line segment $t \in [0,1] \in \mathbb{R} \to z(t) = z + t\,\Delta z_1 \in \mathbb{C}$ in the z plane is mapped to a curve $w(t) = f(z(t))$ in the w plane. The angle between those curves is the angle between the tangent to those curves at their intersection point w. Thus we are interested in the limit $\Delta z_1 \to 0$ and $\Delta z_2 \to 0$, with fixed ratio $|\Delta z_2|/|\Delta z_1|$ and fixed angles θ_1, θ_2.



The angle between the tangents to the mapped curves at point w in the w plane is that limit of the angle between the secant vectors $\Delta w_1 = f(z + \Delta z_1) - f(z)$ and $\Delta w_2 = f(z + \Delta z_2) - f(z)$, which is the angle β of the ratio

$$\frac{\Delta w_2}{\Delta w_1} = \frac{|\Delta w_2|}{|\Delta w_1|}\, e^{i\beta}.$$

From the Taylor series expansions about z (82), this ratio

$$\frac{\Delta w_2}{\Delta w_1} = \frac{f'(z)\Delta z_2 + \frac{1}{2} f''(z)\Delta z_2^2 + \cdots}{f'(z)\Delta z_1 + \frac{1}{2} f''(z)\Delta z_1^2 + \cdots} \longrightarrow \frac{\Delta z_2}{\Delta z_1} \tag{83}$$

as Δz_1 and $\Delta z_2 \to 0$, with $|\Delta z_2|/|\Delta z_1|$ fixed, provided $f'(z) \neq 0$. Hence, in that limit,

$$\frac{\Delta w_2}{\Delta w_1} = \frac{|\Delta w_2|}{|\Delta w_1|}\, e^{i\beta} \;\to\; \frac{\Delta z_2}{\Delta z_1} = \frac{|\Delta z_2|}{|\Delta z_1|}\, e^{i\alpha} \tag{84}$$

implying $|\Delta w_2|/|\Delta w_1| \to |\Delta z_2|/|\Delta z_1|$ and $\beta \to \alpha$, so angles are preserved.

Singular points

The analysis above requires that the Taylor approximations be valid in a neighborhood of point z, and that $f'(z)\Delta z$ dominates as $\Delta z \to 0$, hence $f'(z)$ must exist and not vanish. If $f'(z)$ exists but $f'(z) = 0$, then angles are *not* preserved. At such points z, we need to go to the next order in the Taylor expansions to figure out what happens to angles. That is determined by the first nonzero term in the Taylor expansions.

If $f'(z) = 0$ but $f''(z) \neq 0$ then in the limit Δz_1,$\Delta z_2 \to 0$, with $|\Delta z_2|/|\Delta z_1|$ fixed, (83) yields

$$\frac{\Delta w_2}{\Delta w_1} = \frac{|\Delta w_2|}{|\Delta w_1|} e^{i\beta} \to \left(\frac{\Delta z_2}{\Delta z_1}\right)^2 = \frac{|\Delta z_2|^2}{|\Delta z_1|^2} e^{i2\alpha}$$

that is $\beta \to 2\alpha$ so the angles are *doubled* at those z's. For example, $f(z) = z^2$ has $f'(z) = 2z$ that exists everywhere but vanishes at $z = 0$ but $f'' = 2 \neq 0$ for all z.

Thus the mapping $w = z^2$ preserves all angles – angles between any two curves in the z-plane will be the same as the angles between the image curves in the w-plane, *except* at $z = 0$ where the angles will be *doubled* in the w-plane.

If $f'(z) = f''(z) = 0$ but $f'''(z) \neq 0$ then in the same limit (83) gives

$$\frac{\Delta w_2}{\Delta w_1} = \frac{|\Delta w_2|}{|\Delta w_1|}\, e^{i\beta} \to \left(\frac{\Delta z_2}{\Delta z_1}\right)^3 = \frac{|\Delta z_2|^3}{|\Delta z_1|^3}\, e^{i3\alpha}$$

so $\beta \to 3\alpha$ and the angles are *tripled* at those z's. For instance, $w = z^3$ has $w' = 3z^2$, $w' = 6z$ and $w''' = 6$, its derivative exists everywhere but vanishes together with its 2nd derivative at $z = 0$. Thus $w = z^3$ preserves all angles except at $z = 0$ where the angles will be tripled in the w-plane since the 3rd derivative does not vanish.

In general, if $f'(z) = f''(z) = \cdots = f^{(n-1)}(z) = 0$ but $f^{(n)}(z) \neq 0$, then angles at z will be multiplied by n at point $w = f(z)$.

A mapping that preserves all angles is called **conformal**. Analytic functions $f(z)$ provide conformal mappings between z and $w = f(z)$ at all points where $f'(z) \neq 0$. Analytic functions $z = f(w) = x(u, v) + iy(u, v)$ thus provide conformal coordinates (u, v) for domains in the (x, y) plane. Figure 3.10 shows conformal coordinates for a Joukowski airfoil obtained by mapping the domain $(u, v) = [0, 1] \times [0, 2\pi)$ using a shifted and rotated exponential map, $z_1(u, v) = z_c + (1 - z_c)\exp(u + iv)$ with $z_c = -0.1 + 0.1i$, followed by a Joukowski map $z = z_1 + 1/z_1$.

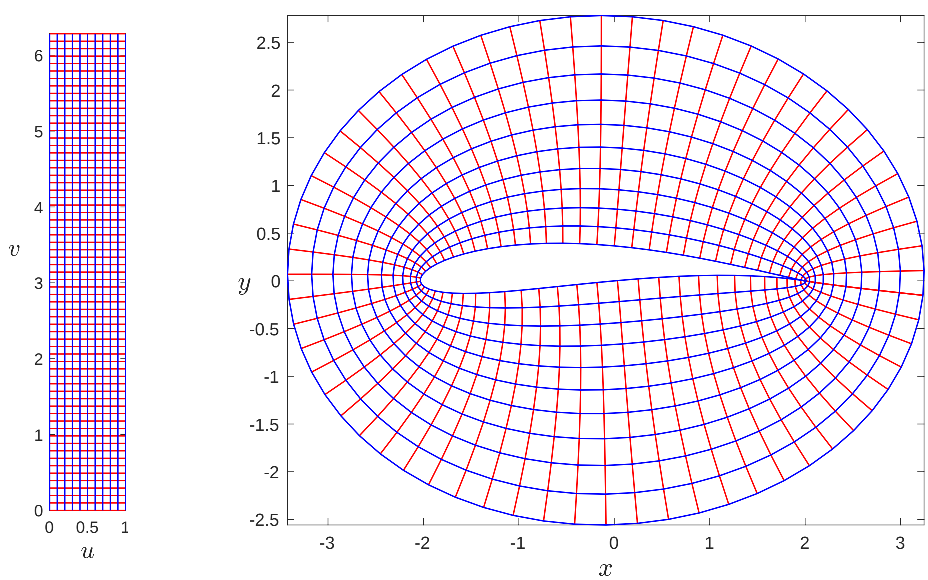


Fig. 3.10: Conformal coordinates (u, v) for a Joukowski airfoil $z = x + iy = z_1 + 1/z_1$ with $z_1(u, v) = z_c + (1 - z_c)e^{u+iv}$ for $z_c = -0.1 + 0.1i$. Red: v contours. Blue: u contours.

Solving Laplace's equation by conformal mapping

Another application of conformal mapping is to solve Laplace's equation in a given z domain by conformally mapping that domain to another domain $z \to Z = f(z)$ where the corresponding solution of Laplace's equation is known, say $F(Z)$. The solution(s) of Laplace's equation in the original domain is then simply $F(f(z))$.

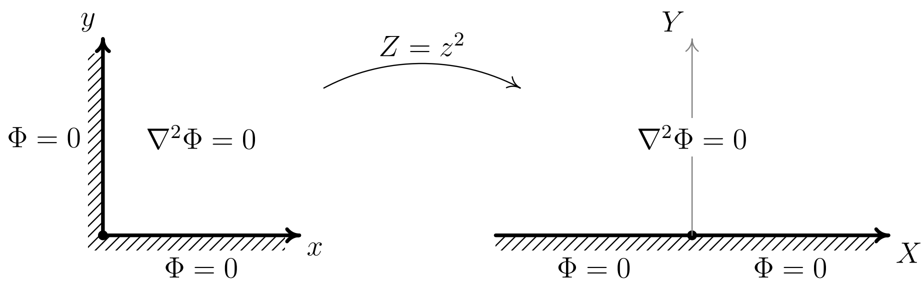


Fig. 3.11: Solving Laplace's equation in the first quadrant by mapping it to the upper half plane where the corresponding solution is trivial.

For example, to find a solution $\Phi(x, y)$ of Laplace's equation in the first quadrant $x \geq 0$, $y \geq 0$ with $\Phi = 0$ on $x = 0$ and on $y = 0$, we can map the $z = x + iy$ quadrant to the $Z = X + iY$ upper half plane with $Z = z^2$. The corresponding solution $\Phi(X, Y)$ in that upper half plane, that is a solution of

$$\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2}\right)\Phi = 0$$

with $\Phi = 0$ along $Y = 0$ is simply $\Phi = Y = \mathfrak{Im}(Z)$. The solution in the original quadrant is then $\Phi = Y(x, y) = \mathfrak{Im}(z^2) = 2xy$. See problems 9 – 12 in §2.2 and fig. 3.16 for other examples.

2.4 Conformal mapping examples

Figures such as figs. 3.8, 3.9 and 3.10, that show contours of $u(x, y)$ and $v(x, y)$ in the $z = x + iy$ plane visualize the inverse map from $(u, v) \to (x, y)$, with contours of u and v corresponding to vertical lines and horizontal lines in the (u, v) plane, respectively.

Figures such as fig. 3.12, visualize the forward map from $z \to w = f(z)$. Vertical lines in the z plane have $x = x_0$ fixed and are mapped to the curves $(u(x_0, y), v(x_0, y))$ in the w-plane. Horizontal lines in the z plane have $y = y_0$ fixed and are mapped to the curves $(u(x, y_0), v(x, y_0))$ in the w plane. Depending on the specific map $f(z)$, a polar point of view with $z = re^{i\theta}$ and/or $w = \rho e^{i\varphi}$ may be more revealing than the cartesian forms $z = x + iy$ and $w = u + iv$. We analyze a few forward maps $w = f(z)$ below. In applications, the roles of z and w maybe switched, with $z = f(w)$ providing maps and conformal coordinates (u, v) for (x, y) domains of interest.

Quadratic map

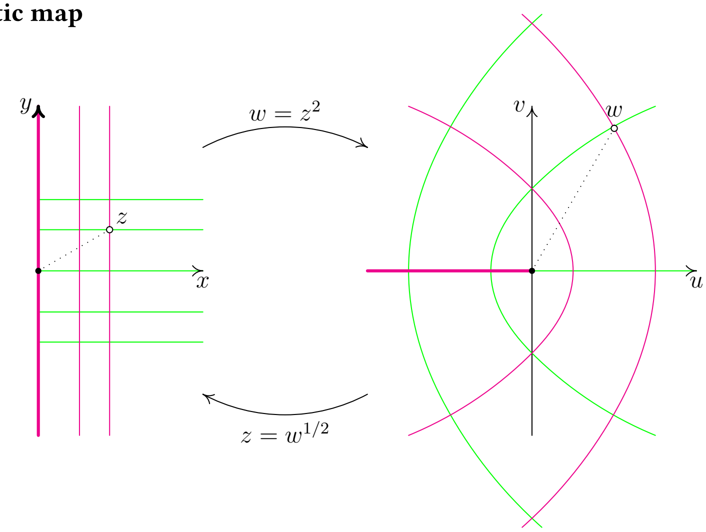


Fig. 3.12: $w = z^2$ and $z = \sqrt{w}$ with $-\pi < \arg(w) \le \pi$.

The map

$$w = z^2 = (x^2 - y^2) + i(2xy),$$

illustrated in fig. 3.12, maps the right half z-plane to the entire w-plane. This map provides *conformal, confocal parabolic coordinates* (x, y) for the (u, v) plane

$$u = x^2 - y^2, \qquad v = 2xy.$$

A (magenta) vertical line with $x = x_0$ is mapped to the *parabola* $u = x_0^2 - y^2$, $v = 2x_0 y$, that is

$$u = x_0^2 - \frac{v^2}{4x_0^2}$$

in the $w = (u, v)$ plane. A (green) horizontal line $y = y_0$ is mapped to the (green) parabola $u = x^2 - y_0^2$, $v = 2xy_0$

$$u = \frac{v^2}{4y_0^2} - y_0^2$$

in the (u, v) plane. All of those parabolas have $(u, v) = (0, 0)$ as their focus. The green and magenta curves intersect at 90 degrees in *both* planes.

Radial lines from the origin $z = re^{i\theta}$ in the z-plane are mapped to radials from the origin $w = r^2 e^{i2\theta}$ in the w-plane. The angles between radials at the origin are doubled in the w-plane compared to angles in the z-plane.

Angles between the radial dotted lines and the green and magenta curves are the same in *both* planes, *except at* $z = w = 0$ where the angles are doubled. The definition of the inverse function that was selected here is $w^{1/2} = |w|^{1/2} e^{i \arg(w)/2}$ with $-\pi <$ $\arg(w) \le \pi$ and correspond to a one-to-one map between the right-half z-plane to the entire w-plane. The inverse function $w^{1/2}$ has a *branch cut* along the negative real axis, $u < 0$, $v = 0$. The $\arg(w)$ jumps by 2π across the cut and $w^{1/2}$ jumps from positive to negative imaginary, or vice versa. The definition $0 \leq \arg(w) < 2\pi$ would have a branch cut along the positive real axis in the w-plane and map the upper half z-plane to the entire w-plane, instead of the right half z-plane.

Exponential map

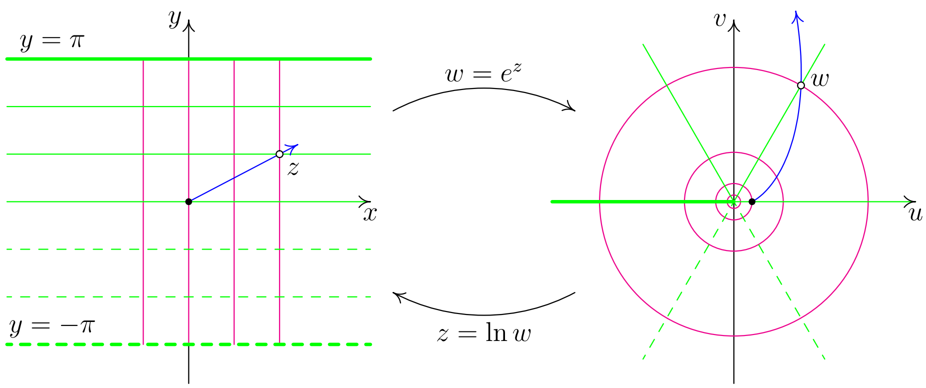


Fig. 3.13: $w = e^z$ and $z = \ln w = \ln |w| + i \arg(w)$ with $-\pi < \arg(w) \leq \pi$.

The map

$$w = e^z = e^x e^{iy}$$

maps the *strip* $-\infty < x < \infty$, $-\pi < y \leq \pi$ to the *entire* w-plane. Indeed $e^{z+2i\pi} = e^z$ is periodic of complex period $2i\pi$ in z, so e^z maps an infinite number of z's to the same w. This map provides conformal *log-polar coordinates* (x, y) for the (u, v) plane

$$u = e^x \cos y, \qquad v = e^x \sin y.$$

The vertical lines $z = x_0 + iy \to w = e^{x_0} e^{iy}$ are mapped to circles of radius $|w| = e^{x_0}$ in the w-plane (magenta). The horizontal lines $z = x + iy_0 \to w = e^x e^{iy_0}$ are mapped to radial lines with polar angle $\arg(w) = y_0$ in w-plane (green).

The radial from the origin $z = x + iax \to w = e^x e^{iax}$, with a real and fixed, is mapped to a *logarithmic spiral* in the w-plane since $|w| = e^x = e^{(\arg(w))/a}$ (blue). The inverse function $z = \ln w = \ln |w| + i \arg(w)$ showed in this picture corresponds to the definition $-\pi < \arg(w) \leq \pi$. All angles are preserved *e.g.* the angles between green and magenta curves, as well as between blue and colored curves, except at $w = 0$.

Hyperbolic cosine map

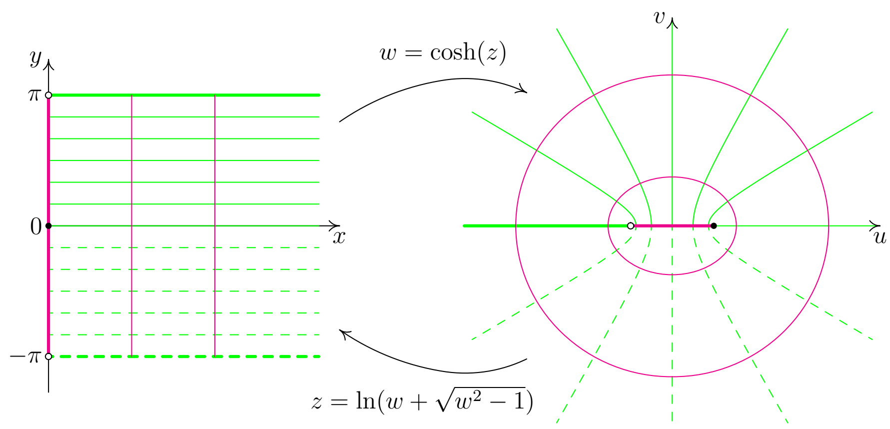


Fig. 3.14: $w = \cosh z$ and $z = \ln(w + \sqrt{w^2-1})$.

The map

$$w = \cosh(z) = \frac{e^z + e^{-z}}{2} = \frac{e^x e^{iy} + e^{-x} e^{-iy}}{2} = \cosh x \cos y + i \sinh x \sin y$$

maps the *semi-infinite strip* $(x, y) \in [0, \infty) \times (-\pi, \pi]$ to the *entire* w-plane. Indeed, $\cosh(z) = \cosh(-z)$ and $\cosh(z+2\pi i) = \cosh(z)$ so $\cosh z$ is even in z and periodic of period $2\pi i$. This map gives *conformal, confocal elliptic coordinates* (x, y) for the (u, v) plane

$$u = \cosh x \cos y, \qquad v = \sinh x \sin y$$

that are the right coordinates to solve Laplace's equation in an elliptical geometry (in the u, v plane). The vertical line segments $z = x_0 + iy$ with $x = x_0 \geq 0$ and $-\pi < y \leq \pi$ are mapped to *confocal ellipses* in the w-plane (magenta) with

$$\frac{u^2}{\cosh^2 x_0} + \frac{v^2}{\sinh^2 x_0} = 1.$$

The semi-infinite horizontal lines $z = x + iy_0$ are mapped to *confocal hyperbolic arcs* in the w-plane (green) with

$$\frac{u^2}{\cos^2 y_0} - \frac{v^2}{\sin^2 y_0} = 1.$$

All of those ellipses and hyperbolas have $(u, v) = (\pm 1, 0)$ as their foci.

The inverse map is $z = \ln(w + \sqrt{w^2-1})$ defined with $\varphi_{1,2} = \arg(z \pm 1)$ and

$$\sqrt{w^2-1} \triangleq \sqrt{|w^2-1|}\, e^{i\varphi_1/2} e^{i\varphi_2/2},$$

such that $\sqrt{w^2-1} = \pm i\sqrt{|w^2-1|}$ on the top side or bottom side of $(-1, 1)$, respectively, and $\ln(w + \sqrt{w^2-1})$ has the branch cut $w \in (-\infty, 1)$.

Joukowski Map

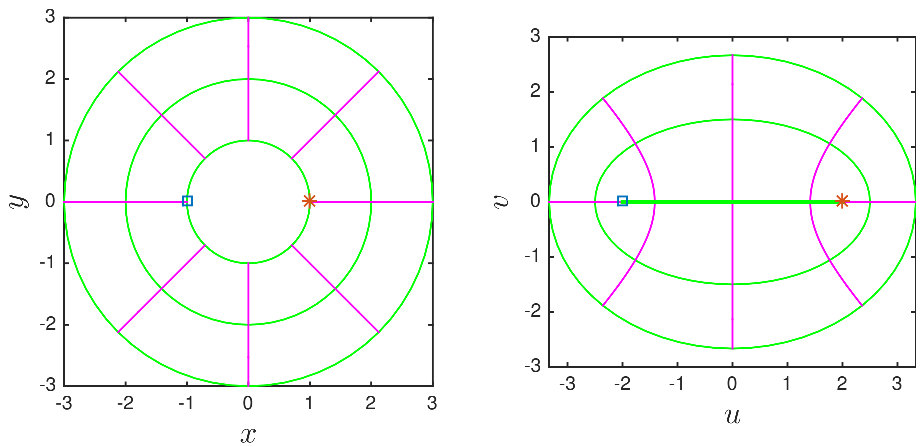


Fig. 3.15: The Joukowski map $w = z + 1/z$ maps circles centered at the origin in the z-plane to ellipses in the w plane. Radials from the origin in z are mapped to hyperbolic arcs in w.

The Joukowski map

$$w = z + \frac{1}{z}, \tag{85}$$

sends circles $|z| = r_0$ fixed to ellipses in the w plane with

$$\frac{u^2}{\left(r_0 + \frac{1}{r_0}\right)^2} + \frac{v^2}{\left(r_0 - \frac{1}{r_0}\right)^2} = 1.$$

In particular, it maps the unit circle $z = e^{i\theta}$ to the line segment $w = 2\cos\theta$. Radials $z = re^{i\theta_0}$ with θ_0 fixed are mapped to hyperbolic arcs with

$$\frac{u^2}{\cos^2\theta_0} - \frac{v^2}{\sin^2\theta_0} = 1.$$

The Joukowski map is useful in 2D electrostatics and fundamental in 2D aerodynamics. Its real and imaginary parts

$$\begin{aligned} u &= x + \frac{x}{x^2+y^2} = \left(r + \frac{1}{r}\right)\cos\theta, \\ v &= y - \frac{y}{x^2+y^2} = \left(r - \frac{1}{r}\right)\sin\theta, \end{aligned} \tag{86}$$

provide solutions to

$$\nabla^2 u = 0 \quad \text{for } r > 1 \quad \text{with} \quad \frac{\partial u}{\partial r} = 0 \quad \text{on } r = 1$$

and

$$\nabla^2 v = 0 \quad \text{for } r > 1 \quad \text{with} \quad v = 0 \quad \text{on } r = 1.$$

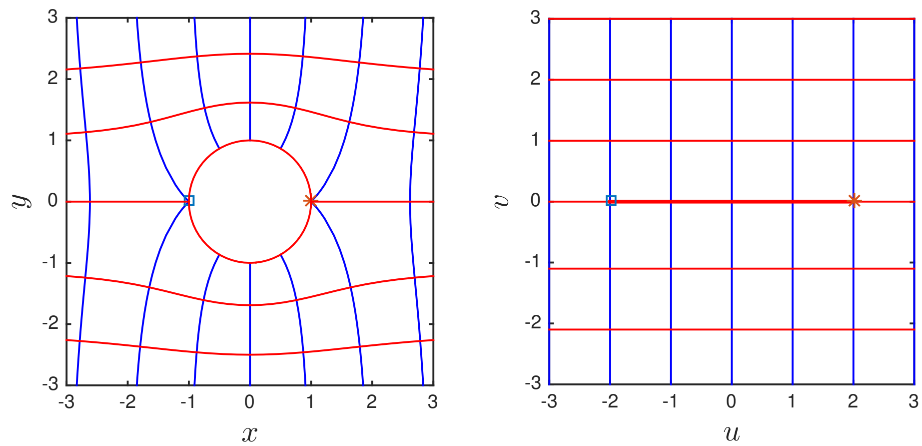


Fig. 3.16: The Joukowski map $w = z + 1/z$ maps the outside of the unit circle $|z| \geq 1$ to the entire w plane. The contours of $u(x, y) = \mathfrak{Re}(w)$ are the electric field lines (blue, vertical) when a perfect cylindrical conductor of radius 1 is placed in a uniform electric field. The contours of $v(x, y) = \mathfrak{Im}(w)$ are the streamlines (red, horizontal) for potential flow around the cylinder of radius 1.

Exercises:

1. Consider the map $w = z^2$. Determine where the triangle with vertices (i) $(1, 0)$, $(1, 1)$, $(0, 1)$ in the z-plane is mapped in the $w = u + iv$ plane; (ii) same but for triangle $(0, 0)$, $(1, 0)$, $(1, 1)$. Do not simply map the vertices, parametrize and map each edge and determine what happens to each angle.

2. Analyze $w = 1/z$. Show that contours of u and v are circles in the z-plane. Determine the maps of constant (i) $x = \mathfrak{Re}(z)$, (ii) $y = \mathfrak{Im}(z)$, (iii) $\varphi = \arg(z)$ and (iv) $r = |z|$. Show that arbitrary lines $z(t) = a + ita$, for any complex constant a with t real, are mapped to circles.

3. If (u, v) are curvilinear coordinates for the Euclidean plane with cartesian coordinates $(x, y) \in \mathbb{R}^2$, then the del operator is (§8.6, eqn. (192))

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} = (\boldsymbol{\nabla} u)\,\frac{\partial}{\partial u} + (\boldsymbol{\nabla} v)\,\frac{\partial}{\partial v}.$$

Show that the 2D Laplacian operator $\nabla^2 = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla}$ reads

$$\begin{aligned}\nabla^2 =& |\boldsymbol{\nabla} u|^2 \frac{\partial^2}{\partial u^2} + 2(\boldsymbol{\nabla} u) \cdot (\boldsymbol{\nabla} v)\frac{\partial^2}{\partial u \partial v} + |\boldsymbol{\nabla} v|^2 \frac{\partial^2}{\partial v^2} \\ &+ (\nabla^2 u)\,\frac{\partial}{\partial u} + (\nabla^2 v)\,\frac{\partial}{\partial v}.\end{aligned} \tag{87}$$

Deduce that if the coordinates correspond to a conformal map, $z = f(w)$ with

$z = x + iy$ and $w = u + iv$, then

$$\nabla^2 = \left|\frac{\mathrm{d}z}{\mathrm{d}w}\right|^{-2} \left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right). \tag{88}$$

4. Consider the change of variables from cartesian (x, y) to (u, v)

$$x = \cosh u \cos v, \qquad y = \sinh u \sin v.$$

Show that the coordinates curves with u fixed are confocal ellipses. Show that the coordinates curves with v fixed are confocal hyperbolas with the same foci as the ellipses. Identify the foci. Show that these (u, v) coordinates are conformal and that the Laplacian reads

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{1}{\sinh^2 u + \sin^2 v} \left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right).$$

5. The Laplacian in cylindrical coordinates (ρ, φ, z) is derived in §8.6, eqn. (204). Translating those results to polar coordinates $x = r\cos\theta$, $y = r\sin\theta$, the 2D Laplacian in polar coordinates is

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}.$$

Show that the Laplacian in *log-polar coordinates* $x = e^u \cos v$, $y = e^u \sin v$ is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = e^{-2u} \left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right).$$

Show that polar coordinates (r, θ) are orthogonal but not conformal, but log-polar coordinates (u, v) are orthogonal and conformal, with $r = e^u$ and $\theta = v$.

3 Integration of Complex Functions

3.1 Complex integrals are path integrals

What do we mean by $\int_a^b f(z)\mathrm{d}z$ when $f(z)$ is a complex function of the complex variable z and the bounds a and b are complex numbers in the z-plane? In general, we need to specify the *path* $\mathcal{C}$ in the complex plane to go from a to b and we need to write the integral as $\int_{\mathcal{C}} f(z)\mathrm{d}z$. Then if $z_0 = a, z_1, z_2, \ldots, z_N = b$ are successive points on the path from a to b we can define the integral as usual as

$$\int_{\mathcal{C}} f(z)\,\mathrm{d}z \triangleq \lim_{\Delta z_n \to 0} \sum_{n=1}^{N} f_n\,\Delta z_n \tag{89}$$

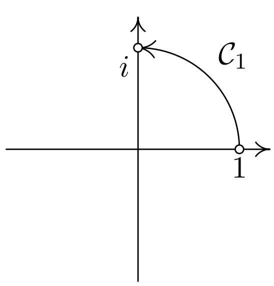
where $\Delta z_n = z_n - z_{n-1}$ and f_n is an approximation of $f(z)$ on the segment (z_{n-1}, z_n). The first order Euler approximation selects $f_n = f(z_{n-1})$ or $f_n = f(z_n)$ while the second order trapezoidal rule picks the average between those two values, that is $f_n = \frac{1}{2}(f(z_{n-1}) + f(z_n))$. The Riemann sum definition also provides a practical way to estimate the integral as a sum of finite differences.

Bound. If $|f(z)| \le M$ along the curve then

$$\left|\int_{\mathcal{C}} f(z)\mathrm{d}z\right| \le M \int_{\mathcal{C}} |\mathrm{d}z| = ML \tag{90}$$

where $L = \int_{\mathcal{C}} |\mathrm{d}z| \ge 0$ is the length of the curve $\mathcal{C}$ from a to b. Note also that the integral from a to b along $\mathcal{C}$ is minus that from b to a along the same curve since all the Δz_n change sign for that 'return' curve. If $\mathcal{C}$ is from a to b, we sometimes use $-\mathcal{C}$ to denote the same path but with the opposite orientation, from b to a. If we have a parametrization for the curve, say $z(t)$ with t real and $z(t_a) = a$, $z(t_b) = b$ then the partition $z_0, \ldots, z_N$ can be obtained from a partition of the real t interval $t_0 = t_a, \ldots, t_N = t_b$, and in the limit the integral can be expressed as

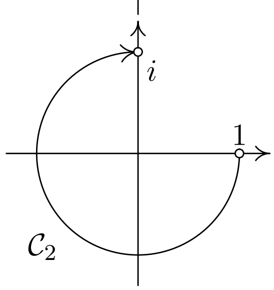
$$\int_{\mathcal{C}} f(z)\mathrm{d}z = \int_{t_a}^{t_b} f\big(z(t)\big)\,\frac{\mathrm{d}z}{\mathrm{d}t}\mathrm{d}t \tag{91}$$



that can be separated into real and imaginary parts to end up with a complex combination of real integrals over the real variable t.

Examples: To compute the integral of $1/z$ along the path $\mathcal{C}_1$ that consists of the unit circle counterclockwise from $a = 1$ to $b = i$, we can parametrize the circle as $z(\theta) = e^{i\theta}$ with $\mathrm{d}z = ie^{i\theta}\mathrm{d}\theta$ for $\theta = 0 \to \pi/2$, then

$$\int_{\mathcal{C}_1} \frac{1}{z}\mathrm{d}z = \int_0^{\pi/2} \frac{1}{e^{i\theta}} ie^{i\theta}\mathrm{d}\theta = i\frac{\pi}{2}.$$



For the path $\mathcal{C}_2$ which consists of the portion of the unit circle *clockwise* from $a = 1$ to $b = i$, the parametrization is again $z = e^{i\theta}$ however now $\theta = 0 \to -3\pi/2$ and

$$\int_{\mathcal{C}_2} \frac{1}{z}\mathrm{d}z = \int_0^{-3\pi/2} \frac{1}{e^{i\theta}} ie^{i\theta}\mathrm{d}\theta = -i\frac{3\pi}{2} \ne i\frac{\pi}{2}.$$

Clearly the integral of $1/z$ from $a = 1$ to $b = i$ depends on the path. Note that these two integrals differ by $2\pi i$.

However for the function z^2 over the same two paths with $z = e^{i\theta}$, $z^2 = e^{i2\theta}$ and $\mathrm{d}z = ie^{i\theta}\mathrm{d}\theta$, we find

$$\int_{\mathcal{C}_1} z^2 \mathrm{d}z = \int_0^{\pi/2} ie^{i3\theta}\mathrm{d}\theta = \frac{1}{3}\left(e^{i3\pi/2} - 1\right) = \frac{-i-1}{3} = \frac{b^3 - a^3}{3},$$

and

$$\int_{\mathcal{C}_2} z^2 \mathrm{d}z = \int_0^{-3\pi/2} ie^{i3\theta}\mathrm{d}\theta = \frac{1}{3}\left(e^{-i9\pi/2} - 1\right) = \frac{-i-1}{3} = \frac{b^3 - a^3}{3}.$$

Thus for z^2 it appears that we obtain the expected result $\int_a^b z^2 \mathrm{d}z = (b^3 - a^3)/3$, independently of the path. We've only checked two special paths, so we do not know for sure but, clearly, a key issue is to determine when an integral depends on the path of integration or not.

3.2 Cauchy's theorem

The integral of a complex function is independent of the path of integration if and only if the integral over a *closed* contour always vanishes. Indeed if $\mathcal{C}_1$ and $\mathcal{C}_2$ are two distinct paths from a to b then the curve $\mathcal{C} = \mathcal{C}_1 - \mathcal{C}_2$ which goes from a to b along $\mathcal{C}_1$ then back from b to a along $-\mathcal{C}_2$, is closed. The integral along that closed curve is zero if and only if the integral along $\mathcal{C}_1$ and $\mathcal{C}_2$ are equal.

Writing $z = x + iy$ and $f(z) = u(x, y) + iv(x, y)$ the complex integral around a closed curve $\mathcal{C}$ can be written as

$$\oint_{\mathcal{C}} f(z)\mathrm{d}z = \oint_{\mathcal{C}} (u + iv)(\mathrm{d}x + i\mathrm{d}y) = \oint_{\mathcal{C}} (u\mathrm{d}x - v\mathrm{d}y) + i \oint_{\mathcal{C}} (v\mathrm{d}x + u\mathrm{d}y) \tag{92}$$

hence the real and imaginary parts of the integral are real *line integrals*. These line integrals can be turned into area integrals using the curl form of Green's theorem:

$$\begin{aligned}\oint_{\mathcal{C}} f(z)\mathrm{d}z = & \oint_{\mathcal{C}} (u\mathrm{d}x - v\mathrm{d}y) \quad + \ i \oint_{\mathcal{C}} (v\mathrm{d}x + u\mathrm{d}y) \\ = & \int_A \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \mathrm{d}A + i \int_A \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \mathrm{d}A,\end{aligned} \tag{93}$$

where A is the interior domain bounded by the closed curve $\mathcal{C}$. But the Cauchy-Riemann equations (73) give

$$\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = 0, \quad \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = 0 \tag{94}$$

whenever the function $f(z)$ is differentiable in the neighborhood of the point $z = x+iy$. Thus both integrals vanish if $f(z)$ is analytic at *all* points of A. This is *Cauchy's theorem*,

$$\boxed{\oint_{\mathcal{C}} f(z)\mathrm{d}z = 0} \tag{95}$$

if $\mathrm{d}f/\mathrm{d}z$ *exists everywhere inside (and on) the closed curve* $\mathcal{C}$.

Functions like e^z, $\cos z$, $\sin z$ and z^n with $n \geq 0$ are *entire* functions — functions that are complex differentiable in the *entire* complex plane, hence the integral of such functions around *any* closed contour $\mathcal{C}$ vanishes.

3.3 Poles and Residues

Functions such as

$$f(z) = \frac{1}{z-a}$$

are differentiable everywhere except at the isolated singular point $z = a$. A function $f(z)$ has an *isolated singularity* at $z = a$ if it is differentiable in a neighborhood of $z = a$ but $f(z)$ is singular at $z = a$. The singularity is called a *pole of order* $n > 0$ if

$$\lim_{z \to a} (z-a)^n f(z) = C \neq 0,$$

that is if

$$f(z) \sim \frac{C}{(z-a)^n}$$

near $z = a$. For examples, $f(z) = (z-1)^{-1}$ and $f(z) = e^z/(z-1)$ have poles of order 1 (*simple* poles) at $z = 1$, while $f(z) = (z-i)^{-2}$ and $f(z) = (\cos z)/(z-i)^2$ have poles of order 2 (*double* poles) at $z = i$.

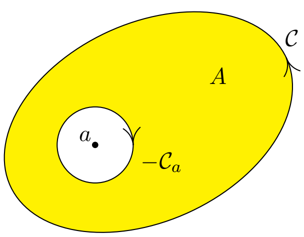


Fig. 3.17: Isolating the singularity at a.

The integral of $1/(z-a)^n$ around any closed contour that does *not* include $z = a$ vanishes, by Cauchy's theorem. We can figure out the integral of $(z-a)^{-n}$, with n a positive integer, about any closed curve $\mathcal{C}$ enclosing the pole a by considering a small circle $\mathcal{C}_a$ centered at a of radius $\epsilon > 0$ as small as needed to be inside the outer closed curve $\mathcal{C}$. Now, the function $(z-a)^{-n}$ is analytic everywhere inside the domain A bounded by the *counter-clockwise* outer boundary $\mathcal{C}$ and the *clockwise* inner circle boundary $-\mathcal{C}_a$ (emphasized here by the minus sign), so the interior A is always to the *left* when traveling on the boundary.

By Cauchy's theorem, this implies that the integral over the closed contour of A, which consists of the sum $\mathcal{C} + (-\mathcal{C}_a)$ of the outer *counter-clockwise* curve $\mathcal{C}$ and the inner *clockwise* circle $-\mathcal{C}_a$, vanishes

$$\oint_{\mathcal{C}+(-\mathcal{C}_a)} \frac{1}{(z-a)^n} \mathrm{d}z = 0 = \oint_{\mathcal{C}} \frac{1}{(z-a)^n} \mathrm{d}z + \oint_{-\mathcal{C}_a} \frac{1}{(z-a)^n} \mathrm{d}z$$

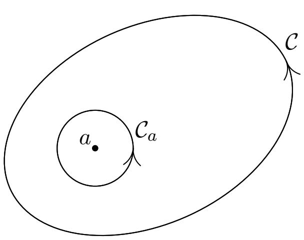


Fig. 3.18: Deforming the contour.

therefore

$$\oint_{\mathcal{C}} \frac{1}{(z-a)^n} \mathrm{d}z = \oint_{\mathcal{C}_a} \frac{1}{(z-a)^n} \mathrm{d}z,$$

since $\oint_{-\mathcal{C}_a} = -\oint_{\mathcal{C}_a}$, in other words the integral about the closed contour $\mathcal{C}$ equals the integral about the closed inner circle $\mathcal{C}_a$, now with the *same* orientation as the outer contour.[2] Thus the integral is invariant under deformation of the loop $\mathcal{C}$ as long as

[2] Recall that we used this singularity isolation technique in conjunction with the divergence theorem to evaluate the flux of $\hat{\boldsymbol{r}}/r^2 = \boldsymbol{r}/r^3$ (the inverse square law of gravity and electrostatics) through any closed surface enclosing the origin at $r = 0$, as well as in conjunction with Stokes' theorem for the circulation of a line current $\boldsymbol{B} = (\hat{\boldsymbol{z}} \times \boldsymbol{r})/|\hat{\boldsymbol{z}} \times \boldsymbol{r}|^2 = \hat{\boldsymbol{\varphi}}/\rho = \boldsymbol{\nabla}\varphi$ around a loop enclosing the z-axis at $\rho = 0$.

such deformation does not cross the *pole* at a. The loop can be shrunk to an arbitrary small circle surrounding the pole without changing the value of the integral. That remaining integral about the arbitrarily small circle is the *residue* integral.

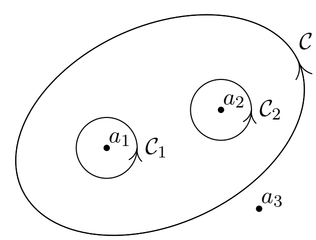
That *residue* integral can be calculated explicitly with the circle parameterization $z = a + \epsilon e^{i\theta}$ and $dz = i\epsilon e^{i\theta} d\theta$ yielding the simple result that

$$\oint_{\mathcal{C}_a} \frac{\mathrm{d}z}{(z-a)^n} = \int_0^{2\pi} \frac{i\epsilon e^{i\theta}}{\epsilon^n e^{in\theta}} \mathrm{d}\theta = i\epsilon^{1-n} \int_0^{2\pi} e^{i(1-n)\theta} \mathrm{d}\theta = \begin{cases} 2\pi i & \text{if } n = 1, \\ 0 & \text{if } n \neq 1. \end{cases}$$

We can collect all these various results in the following useful formula that for any integer $n = 0, \pm 1, \pm 2, \ldots$ and a closed contour $\mathcal{C}$ oriented *counter-clockwise*

$$\boxed{\oint_{\mathcal{C}} \frac{\mathrm{d}z}{(z-a)^n} = \begin{cases} 2\pi i & \text{if } n = 1 \text{ and } \mathcal{C} \text{ encloses } a, \\ 0 & \text{otherwise.} \end{cases}} \tag{96}$$

For $n \leq 0$, that is for $n = -|n|$, this follows directly from Cauchy's theorem since $(z-a)^{-n} = (z-a)^{|n|}$ is complex differentiable for all z. For $n > 0$ this also follows from Cauchy's theorem that allows the deformation of the contour $\mathcal{C}$ to a small circle $\mathcal{C}_a$ around the pole, the integral *residue*, that can be calculated explicitly, as done above. If n is *not* an integer, then $z = a$ is a *branch point* from which a *branch cut* emanates to define the function $(z-a)^\alpha$ uniquely. The result (96) does not apply for non-integer exponent n, the integral depends on the choice of branch cut and the exponent.



If the function $f(z)$ has several poles, at a_1, a_2 and a_3 for example, we can use the same procedure to isolate all the poles inside the contour, say a_1 and a_2 for example. The poles outside the contour $\mathcal{C}$ (a_3 on the side figure) do not contribute to the integral, *the contour integral is driven by what's inside* and Cauchy's theorem yields

$$\oint_{\mathcal{C}} f(z)\mathrm{d}z = \oint_{\mathcal{C}_1} f(z)\mathrm{d}z + \oint_{\mathcal{C}_2} f(z)\mathrm{d}z, \tag{97}$$

the integral is then the sum of the residue integrals, where the sum is over all isolated singularities inside the contour. Each of these residue integrals can be calculated as above. The procedure is called the *'calculus of residues'*. In general, if $f(z)$ is differentiable everywhere inside and on the simple closed curve $\mathcal{C}$, except at the isolated singularities at $z = a_1, \ldots, a_N$ then

$$\oint_{\mathcal{C}} f(z)\mathrm{d}z = \sum_{k=1}^{N} \oint_{\mathcal{C}_k} f(z)\mathrm{d}z, \tag{98}$$

where $\mathcal{C}_k$ is a sufficiently small simple closed loop that encloses only a_k. The residue integrals $\oint_{\mathcal{C}_k} f(z)\mathrm{d}z$ can be computed explicitly as in (96).

Example. Simple poles. Consider

$$\oint_{\mathcal{C}} \frac{1}{z^2+1}\,\mathrm{d}z.$$

The function $f(z) = (z^2+1)^{-1}$ has simple poles at $z = \pm i$ since $z^2+1 = (z-i)(z+i)$. The integral can be calculated using partial fractions since

$$\frac{1}{z^2+1} = \frac{1}{2i}\left(\frac{1}{z-i} - \frac{1}{z+i}\right)$$

and

$$\oint_{\mathcal{C}} \frac{\mathrm{d}z}{z^2+1} = \frac{1}{2i}\oint_{\mathcal{C}} \frac{\mathrm{d}z}{z-i} - \frac{1}{2i}\oint_{\mathcal{C}} \frac{\mathrm{d}z}{z+i}$$

The integral is reduced to a sum of simple integrals such as (96), here with $a = \pm i$ and $n = 1$. The integrals are $2\pi i$ or 0 depending on whether or not the poles i and $-i$ are inside $\mathcal{C}$. If $\mathcal{C}_1$ is a circle centered at i sufficiently small to not include $-i$, the *residue* at i is

$$\oint_{\mathcal{C}_1} \frac{1}{z^2+1}\,\mathrm{d}z = \frac{1}{2i}\oint_{\mathcal{C}_1} \frac{1}{z-i}\mathrm{d}z = \pi$$

and likewise if $\mathcal{C}_2$ is a circle centered at $-i$ sufficiently small to not include i, the *residue* at $-i$ is

$$\oint_{\mathcal{C}_2} \frac{1}{z^2+1}\,\mathrm{d}z = -\frac{1}{2i}\oint_{\mathcal{C}_2} \frac{1}{z+i}\mathrm{d}z = -\pi.$$

In summary,

$$\oint_{\mathcal{C}} \frac{1}{z^2+1}\,\mathrm{d}z = \begin{cases} \pi & \text{if } \mathcal{C} \text{ encloses } i \text{ but not } -i, \\ -\pi & \text{if } \mathcal{C} \text{ encloses } -i \text{ but not } i, \\ 0 & \text{if } \mathcal{C} \text{ encloses both } \pm i \text{ or neither.} \end{cases} \tag{99}$$

This assumes that $\mathcal{C}$ is a *simple closed curve* (no self-intersections) and is oriented counter-clockwise (otherwise all the signs would be reversed). If the contour passes through a singularity then the integral is not defined. It may have to be defined as a suitable limit (for instance as the *Cauchy Principal Value*) that may or may not exist. A more general approach is to use Taylor series expansions about the respective pole, instead of continued fraction expansions. For $\mathcal{C}_1$ around pole i for instance, we have

$$\frac{1}{z^2+1} = \frac{1}{(z-i)}\frac{1}{(z+i)} = \frac{1}{(z-i)}\,\frac{1}{2i+(z-i)} = \frac{1}{2i}\frac{1}{z-i}\sum_{k=0}^{\infty}\frac{(z-i)^k}{(2i)^k},$$

where the geometric series (19) has been used as a shortcut to the requisite Taylor series. Using (96) term by term then yields $\oint_{\mathcal{C}_1} f(z)\,dz = \pi$. A similar Taylor series expansion about $z = -i$ yields the other residue. For simple poles, only the first term of the Taylor series contributes, but that is not the case for higher order poles, as illustrated in the next examples. □

Example. Double pole. Consider $f(z) = z^2/(z-1)^2$. That function is differentiable for all z except at $z = 1$ where there is a 2nd order pole since $f(z) \sim 1/(z-1)^2$, near $z = 1$. This asymptotic behavior as $z \to 1$ might suggests that its residue around that pole vanishes if we apply (96) too quickly, but we have to be a bit more careful about integrating functions that diverge around vanishing circles, that's the story of calculus, $\infty \times 0$ requires further analysis. The Taylor series of z^2 about $z = 1$ is easily obtained as $z^2 = (z-1+1)^2 = (z-1)^2 + 2(z-1) + 1$, then the residue integral is, using (96) term-by-term,

$$\oint_{\mathcal{C}_1} \frac{z^2}{(z-1)^2}\,\mathrm{d}z = \oint_{\mathcal{C}_1} \left(1 + \frac{2}{z-1} + \frac{1}{(z-1)^2}\right)\mathrm{d}z = 4\pi i.$$

□

Example. Essential singularity. Consider

$$\oint_{\mathcal{C}} e^{1/z}\mathrm{d}z, \tag{100}$$

the function $e^{1/z}$ has an *infinite order pole* – an *essential singularity* – at $z = 0$ but Taylor series for e^t with $t = 1/z$ and the simple (96) makes this calculation straightforward

$$\oint_{\mathcal{C}} e^{1/z}\mathrm{d}z = \oint_{\mathcal{C}} \sum_{n=0}^{\infty} \frac{1}{n!\,z^n}\mathrm{d}z = \sum_{n=0}^{\infty} \frac{1}{n!} \oint_{\mathcal{C}} \frac{1}{z^n}\mathrm{d}z = 2\pi i \tag{101}$$

since all integrals vanish except the $n = 1$ term from (96), if $\mathcal{C}$ encloses $z = 0$, otherwise the integral is zero by Cauchy's theorem. Note that this 'Taylor' series for $e^{1/z}$ is in powers of $t = 1/z$ expanding about $t = 0$ that corresponds to $z = \infty$! In terms of z, this is an example of a *Laurent series*, a power series that includes negative powers to capture poles and essential singularities. □

Connection with $\ln z$

The integral of $1/z$ is of course directly related to $\ln z$, the natural log of z which can be defined as the antiderivative of $1/z$ that vanishes at $z = 1$, that is

$$\ln z \equiv \int_1^z \frac{1}{\zeta} d\zeta.$$

We use ζ as the *dummy* variable of integration since z is the upper limit of integration.

But along what path from 1 to z? Here's the $2\pi i$ multiplicity again. We saw earlier that the integral of $1/z$ from a to b depends on how we go around the origin. If we get one result along one path, we can get the same result + $2\pi i$ if we use a path that loops around the origin one more time counterclockwise than the original path. Or $-2\pi i$ if it loops clockwise, etc. Look back at exercise (7) in section (3.2). If we define a range for $\arg(z)$, *e.g.* $0 \le \arg(z) < 2\pi$, we find

$$\int_1^z \frac{1}{\zeta} d\zeta = \ln|z| + i\arg(z) + 2ik\pi \tag{102}$$

for some *specific* k that depends on the actual path taken from 1 to z *and* our definition of $\arg(z)$. The notation $\int_1^z$ is not complete for this integral. The integral is *path-dependent* and it is necessary to specify that path in more details, however all possible paths give the same answer *modulo* $2\pi i$.

Exercises:

Closed paths are oriented *counterclockwise* unless specified otherwise.

1. Calculate the integral of $f(z) = z + 2/z$ along the path $\mathcal{C}$ that goes once around the circle $|z| = R > 0$. How does your result depend on R?

2. Calculate the integral of $f(z) = az + b/z + c/(z+1)$, where a, b and c are complex constants, around (i) the circle of radius $R > 0$ centered at $z = 0$, (ii) the circle of radius 2 centered at $z = 0$, (iii) the triangle $-1/2$, $-2+i$, $-1-2i$.

3. Calculate the integral of $f(z) = 1/(z^2 - 4)$ around (i) the unit circle, (ii) the parallelogram 0, $2-i$, 4, $2+i$. [Hint: use partial fractions]

4. Calculate the integral of $f(z) = 1/(z^4 - 1)$ along the circle of radius 1 centered at i.

5. Calculate the integral of $\sin(1/(3z))$ over the square 1, i, -1, $-i$. [Hint: use the Taylor series for $\sin z$].

6. Calculate the integral of $1/z$ from $z = 1$ to $z = 2e^{i\pi/4}$ along (i) the path $1 \to 2$ along the real line then $2 \to 2e^{i\pi/4}$ along the circle of radius 2, (ii) along $1 \to 2$ on the real line, followed by $2 \to 2e^{i\pi/4}$ along the circle of radius 2, *clockwise*.

7. If a is an arbitrary complex number, show that the integral of $1/z$ along the straight line from 1 to a is equal to the integral of $1/z$ from 1 to $|a|$ along the real line + the integral of $1/z$ along a circular path of radius $|a|$ from $|a|$ to a. *Draw a sketch of the 3 paths from* 1 *to* a. Calculate the integral using both a clockwise and a counterclockwise circular arc from $z = |a|$ to $z = a$. What happens if a is real but negative?

8. Does the integral of $1/z^2$ from $z = a$ to $z = b$ (with a and b complex) depend on the path? Explain.

9. Does the integral of $z^* = \text{conj}(z)$ depend on the path? Calculate $\int_a^b z^* \mathrm{d}z$ along (i) the straight line from a to b, (ii) along the real direction from $\mathfrak{Re}(a)$ to $\mathfrak{Re}(b)$ then up in the imaginary direction from $\mathfrak{Im}(a)$ to $\mathfrak{Im}(b)$. Sketch the paths and compare the answers.

10. The expansion (56) with $a = 1$ gives $1/z = \sum_{n=0}^{\infty}(1-z)^n$. Using this expansion together with (96) we find that

$$\oint_{|z|=1} \frac{1}{z}\,\mathrm{d}z = \sum_{n=0}^{\infty} \oint_{|z|=1} (1-z)^n \mathrm{d}z = 0.$$

But this does not match with our explicit calculation that $\oint_{|z|=1} \frac{\mathrm{d}z}{z} = 2\pi i$. Why not?

3.4 Cauchy's formula

The combination of (95) with (96) and partial fraction and/or Taylor series expansions is quite powerful as we have already seen in the exercises, but there is another fundamental result that can be derived from them. This is Cauchy's formula

$$\boxed{\oint_{\mathcal{C}} \frac{f(z)}{z-a} \mathrm{d}z = 2\pi i f(a)} \tag{103}$$

which holds for *any* closed counterclockwise contour $\mathcal{C}$ that encloses a provided $f(z)$ is analytic (differentiable) everywhere inside and on $\mathcal{C}$.

The proof of this result follows the approach we used to calculate $\oint_{\mathcal{C}} dz/(z-a)$ in section 3.3. Using Cauchy's theorem (95), the integral over $\mathcal{C}$ is equal to the integral over a small counterclockwise circle $\mathcal{C}_a$ of radius ϵ centered at a. That's because the function $f(z)/(z-a)$ is analytic in the domain between $\mathcal{C}$ and the circle $\mathcal{C}_a : z = a + \epsilon e^{i\theta}$ with $\theta = 0 \to 2\pi$, so

$$\oint_{\mathcal{C}} \frac{f(z)}{z-a} \mathrm{d}z = \oint_{\mathcal{C}_a} \frac{f(z)}{z-a} \mathrm{d}z = \int_0^{2\pi} f(a + \epsilon e^{i\theta})\, i \mathrm{d}\theta = 2\pi i f(a). \tag{104}$$

The final step follows from the fact that the integral has the same value no matter what $\epsilon > 0$ we pick. Then taking the limit $\epsilon \to 0^+$, the function $f(a + \epsilon e^{i\theta}) \to f(a)$ because $f(z)$ is a nice continuous and differentiable function everywhere inside $\mathcal{C}$, and in particular at $z = a$.

Cauchy's formula has major consequences that follow from the fact that it applies to *any* a inside $\mathcal{C}$. To emphasize that, let us rewrite it with z in place of a, using ζ as the dummy variable of integration

$$2\pi i f(z) = \oint_{\mathcal{C}} \frac{f(\zeta)}{\zeta - z} d\zeta. \tag{105}$$

This provides an integral formula for $f(z)$ at *any* z inside $\mathcal{C}$ in terms of its values on $\mathcal{C}$. Thus knowing $f(z)$ on $\mathcal{C}$ completely determines $f(z)$ everywhere inside the contour!

Mean Value Theorem

Since (105) holds for any closed contour $\mathcal{C}$ as long as $f(z)$ is continuous and differentiable inside and on that contour, we can write it for a circle of radius r centered at z, $\zeta = z + re^{i\theta}$ where $d\zeta = ire^{i\theta}\mathrm{d}\theta$ and (105) yields

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} f(z + re^{i\theta}) \mathrm{d}\theta \tag{106}$$

which states that $f(z)$ is equal to its average over a circle centered at z. This is true as long as $f(z)$ is differentiable at all points inside the circle of radius r. This *mean value theorem* also applies to the real and imaginary parts of $f(z) = u(x, y) + iv(x, y)$. It implies that $u(x, y)$, $v(x, y)$ and $|f(z)|$ do not have extrema inside a domain where $f(z)$ is differentiable. Points where $f'(z) = 0$ and therefore $\partial u/\partial x = \partial u/\partial y = \partial v/\partial x = \partial v/\partial y = 0$ are *saddle* points, not local maxima or minima.

Generalized Cauchy formula and Taylor Series

Cauchy's formula also implies that if $f'(z)$ exists in the neighborhood of a point a then $f(z)$ is *infinitely differentiable in that neighborhood!* Furthermore, $f(z)$ can be expanded in a Taylor series about a that converges inside a disk whose radius is equal to the distance between a and the nearest singularity of $f(z)$. That is why we use the special word *analytic* instead of simply 'differentiable'. For a function of a complex variable being differentiable in a neighborhood is a really big deal!

$\triangleright$ To show that $f(z)$ is infinitely differentiable, we can show that the derivative of the right-hand side of (105) with respect to z exists by using the limit definition of the derivative and being careful to justify existence of the integrals and the limit. The final result is the same as that obtained by differentiating with respect to z under the integral sign, yielding

$$2\pi i f'(z) = \oint_{\mathcal{C}} \frac{f(\zeta)}{(\zeta - z)^2} d\zeta. \tag{107}$$

Doing this repeatedly we obtain

$$2\pi i f^{(n)}(z) = n! \oint_{\mathcal{C}} \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta. \tag{108}$$

where $f^{(n)}(z)$ is the nth derivative of $f(z)$ and $n! = n(n-1)\cdots 1$ is the factorial of n. Since all the integrals exist, all the derivatives exist. Formula (108) is the *generalized Cauchy formula* which we can rewrite in the form

$$\boxed{\oint_{\mathcal{C}} \frac{f(z)}{(z-a)^{n+1}}\, \mathrm{d}z = 2\pi i\, \frac{f^{(n)}(a)}{n!}} \tag{109}$$

$\square$

$\triangleright$ Another derivation of these results that establishes convergence of the Taylor series expansion at the same time is to use the geometric series (19) and the slick trick that we used in (56) to write

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta - a) - (z - a)} = \frac{1}{\zeta - a} \frac{1}{1 - \dfrac{z-a}{\zeta - a}} = \sum_{n=0}^{\infty} \frac{(z-a)^n}{(\zeta - a)^{n+1}} \tag{110}$$

where the geometric series converges provided $|z - a| < |\zeta - a|$. Cauchy's formula (105) then becomes

$$\begin{aligned} 2\pi i f(z) = \oint_{\mathcal{C}_a} \frac{f(\zeta)}{\zeta - z} d\zeta &= \oint_{\mathcal{C}_a} \sum_{n=0}^{\infty} f(\zeta) \frac{(z-a)^n}{(\zeta - a)^{n+1}}\, d\zeta \\ &= \sum_{n=0}^{\infty} (z-a)^n \oint_{\mathcal{C}_a} \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta \end{aligned} \tag{111}$$

where $\mathcal{C}_a$ is a circle centered at a whose radius is as large as desired provided $f(z)$ is differentiable inside and on that circle. For instance if $f(z) = 1/z$ then the radius of the circle must be less then $|a|$ since $f(z)$ has a singularity at $z = 0$ but is nice everwhere else. If $f(z) = 1/(z+i)$ then the radius must be less than $|a+i|$ which is the distance between a and $-i$ since $f(z)$ has a singularity at $-i$. In general, the radius of the circle must be less than the distance between a and the nearest singularity of $f(z)$. To justify interchanging the integral and the series we need to show that each integral exists and that the series of the integrals converges. If $|f(\zeta)| \le M$ on $\mathcal{C}_a$ and $|z-a|/|\zeta - a| \le q < 1$ since $\mathcal{C}_a$ is a circle of radius r centered at a and z is inside that circle while ζ is on the circle so $\zeta - a = re^{i\theta}$, $d\zeta = ire^{i\theta} d\theta$ and

$$\left| \oint_{\mathcal{C}_a} \frac{(z-a)^n f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta \right| \le 2\pi M q^n \tag{112}$$

showing that all integrals converge and the series of integrals also converges since $q < 1$.

The series (111) provides a power series expansion for $f(z)$

$$2\pi i f(z) = \sum_{n=0}^{\infty} (z-a)^n \oint_{\mathcal{C}_a} \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta = \sum_{n=0}^{\infty} c_n(a)(z-a)^n \tag{113}$$

that converges inside a disk centered at a with radius equal to the distance between a and the nearest singularity of $f(z)$. The series can be differentiated term-by-term and the derivative series also converges in the same disk. Hence all derivatives of $f(z)$ exist in that disk. In particular we find that

$$c_n(a) = 2\pi i\, \frac{f^{(n)}(a)}{n!} = \oint_{\mathcal{C}_a} \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta. \tag{114}$$

which is the generalized Cauchy formula (108), (109), and the series (111) is none other than the familiar Taylor Series

$$f(z) = f(a) + f'(a)(z-a) + \frac{f''(a)}{2}(z-a)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n. \tag{115}$$

□

Finally, Cauchy's theorem tells us that the integral on the right of (114) has the same value on any closed contour (counterclockwise) enclosing a but no other singularities of $f(z)$, so the formula holds for any such closed contour as written in (108). However convergence of the Taylor series only occurs inside a disk centered at a and of radius equal to the distance between a and the nearest singularity of $f(z)$.

Exercises:

1. Why can we take $(z-a)^n$ outside of the integrals in (111)?
2. Verify the estimate (112). Why does that estimate implies that the series of integrals converges?

3. Consider the integral of $f(z)/(z-a)^2$ about a small circle $\mathcal{C}_a$ of radius ϵ centered at a: $z = a + \epsilon e^{i\theta}$, $0 \le \theta < 2\pi$. Study the limit of the θ-integral as $\epsilon \to 0^+$. Does your limit agree with the generalized Cauchy formula (108), (114)?

4. Derive (109) by deforming the contour $\mathcal{C}$ to a small circle $\mathcal{C}_a$ about a, expanding $f(z)$ in a Taylor series about a (assuming that it exists) and using (96) term by term. In applications, this is a practical and more general approach that works for essential singularities as well. The approach in this section actually proves that the Taylor series converges, a major result of complex analysis.

5. Find the Taylor series of $1/(1+x^2)$ and show that its radius of convergence is $|x| < 1$ [Hint: use the geometric series]. Explain why the radius of convergence is one in terms of the singularities of $1/(1+z^2)$. Would the Taylor series of $1/(1+x^2)$ about $a = 1$ have a smaller or larger radius of convergence than that about $a = 0$?

6. Calculate the integrals of $\cos(z)/z^n$ and $\sin(z)/z^n$ over the unit circle, where n is a positive integer.

4 Real examples of complex integration

One application of complex (a.k.a. 'contour') integration is to turn difficult real integrals into simple complex integrals.

Example 1: What is the average of the function $f(\theta) = 3/(5 + 4\cos\theta)$? The function is periodic of period 2π so its average is

$$\frac{1}{2\pi}\int_0^{2\pi} f(\theta)\mathrm{d}\theta.$$

To compute that integral we think *integral over the unit circle in the complex plane!* Indeed the unit circle with $|z| = 1$ has the simple parametrization

$$z = e^{i\theta} \to \mathrm{d}z = ie^{i\theta}\mathrm{d}\theta \;\Leftrightarrow\; \mathrm{d}\theta = \frac{\mathrm{d}z}{iz}. \tag{116}$$

Furthermore

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{z + 1/z}{2},$$

so we obtain

$$\begin{aligned}\int_0^{2\pi} \frac{3}{5+4\cos\theta}\mathrm{d}\theta = &\oint_{|z|=1} \frac{3}{5+2(z+1/z)}\frac{\mathrm{d}z}{iz} \\ =&\frac{3}{2i}\oint_{|z|=1}\frac{\mathrm{d}z}{(z+\frac{1}{2})(z+2)} = \frac{3}{2i}\left(\frac{2\pi i}{z+2}\right)_{z=-\frac{1}{2}} = 2\pi.\end{aligned} \tag{117}$$

We turned the integral of a real periodic function $f(\theta)$ over its period from $\theta = 0$ to 2π into a complex z integral over the unit circle $z = e^{i\theta}$. This is a general idea that

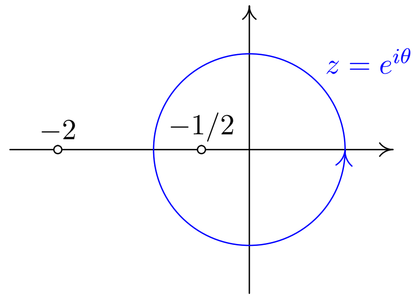


Fig. 3.19: Unit circle $z = e^{i\theta}$, poles and residues for (117).

applies for the integral of any periodic function over its period. That led us to the integral over a closed curve of a simple rational function (that is not always the case, it depends on the actual $f(\theta)$).

In this example, our complex function $(2z^2 + 5z + 2)^{-1}$ has two simple poles, at $-1/2$ and -2, and the denominator factors as $2z^2 + 5z + 2 = 2(z + 1/2)(z + 2)$. Since -2 is outside the unit circle, it does not contribute to the integral, but the simple pole at $-1/2$ does. So the integrand has the form $g(z)/(z - a)$ with $a = -1/2$ inside our domain and $g(z) = 1/(z + 2)$, is a good analytic function inside the unit circle. So one application of Cauchy's formula, *et voilà*. The function $3/(5 + 4\cos\theta)$ which oscillates between $1/3$ and 3 has an average of 1.

Related exercises: calculate

$$\int_0^\pi \frac{3\cos n\theta}{5 + 4\cos\theta} \mathrm{d}\theta \tag{118}$$

where n is an integer. [Hint: use symmetries to write the integral in $[0, 2\pi]$, do not use $2\cos n\theta = e^{in\theta} + e^{-in\theta}$ (why not? try it out to find out the problem), use instead $\cos n\theta = \mathfrak{Re}(e^{in\theta})$ with $n \geq 0$.] An even function $f(\theta) = f(-\theta)$ periodic of period 2π can be expanded in a *Fourier cosine series* $f(\theta) = a_0 + a_1\cos\theta + a_2\cos 2\theta + a_3\cos 3\theta + \cdots$. This expansion is useful in all sorts of applications: numerical calculations, signal processing, etc. The coefficient a_0 is the average of $f(\theta)$. The other coefficients are given by $a_n = 2\pi^{-1}\int_0^\pi f(\theta)\cos n\theta \, \mathrm{d}\theta$, *i.e.* the integrals (118). So what is the Fourier (cosine) series of $3/(5 + 4\cos\theta)$? Can you say something about its convergence?

□

Example 2:

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1 + x^2} = \pi \tag{119}$$

This integral is easily done with the fundamental theorem of calculus since $1/(1 + x^2) = \mathrm{d}(\arctan x)/\mathrm{d}x$, but we use contour integration to demonstrate the method. The integral is equal to the integral of $1/(1 + z^2)$ over the real line $z = x$ with $x = -\infty \to \infty$. That complex function has two simple poles at $z = \pm i$ since $z^2 + 1 = (z + i)(z - i)$.

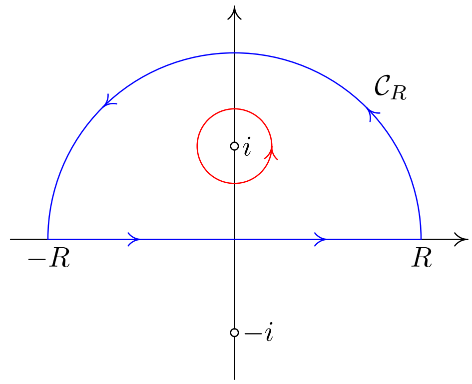


Fig. 3.20: Closing the contour and calculating the 'residue' for (119).

So we turn this into a contour integration by considering the closed path $\mathcal{C}$ consisting of the real interval $z = x$ with $x = -R \to R$ together with the semi-circle $\mathcal{C}_R : z = Re^{i\theta}$ with $\theta = 0 \to \pi$. Since $z = i$ is the only simple pole inside our closed contour $\mathcal{C}$, Cauchy's formula gives

$$\oint_{\mathcal{C}} \frac{\mathrm{d}z}{z^2+1} = \oint_{\mathcal{C}} \frac{(z+i)^{-1}}{z-i}\mathrm{d}z = 2\pi i \left(\frac{1}{z+i}\right)_{z=i} = \pi.$$

To get the integral we want, we need to take $R \to \infty$ and figure out the $\mathcal{C}_R$ integral. The latter goes to zero as $R \to \infty$ since $|dz| = Rd\theta$ and $R^2 - 1 \le |z^2+1| \le R^2+1$ on $z = Re^{i\theta}$, thus

$$\left|\int_{\mathcal{C}_R} \frac{\mathrm{d}z}{z^2+1}\right| \le \int_{\mathcal{C}_R} \frac{|\mathrm{d}z|}{|z^2+1|} < \int_0^\pi \frac{R\mathrm{d}\theta}{R^2-1} = \frac{\pi R}{R^2-1} \to 0.$$

□

Example 3: We use the same technique for

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^4} = \int_{\mathbb{R}} \frac{\mathrm{d}z}{1+z^4} \tag{120}$$

This integrand has 4 simple poles at the simple roots of $z^4 + 1 = 0$. Those roots are $z_k = e^{i(2k-1)\pi/4}$ for $k = 1, 2, 3, 4$, they are on the unit circle, equispaced by angle $\pi/2$.

We use the same closed contour $\mathcal{C}$ consisting of the real interval $[-R, R]$ and the semi-circle $\mathcal{C}_R$, as in the previous example, but now there are *two* simple poles inside that contour. We need to isolate both singularities leading to

$$\oint_{\mathcal{C}} = \oint_{\mathcal{C}_1} + \oint_{\mathcal{C}_2}.$$

Note that $z^4 + 1 = (z^2 - i)(z^2 + i) = (z^2 - z_1^2)(z^2 - z_2^2)$ where $z_1^2 = i$ and $z_2^2 = -i$. Then for $\mathcal{C}_1$ that includes only z_1, we write $z^4 + 1 = (z - z_1)(z + z_1)(z^2 + i)$

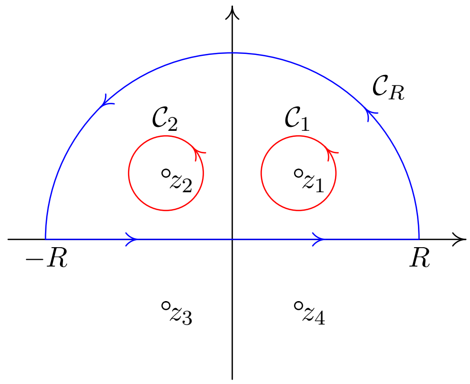


Fig. 3.21: Contour and 'residues' for (120) with $z_1 = e^{i\pi/4}$ and $z_2 = -e^{-i\pi/4}$.

and Cauchy's formula gives

$$\oint_{\mathcal{C}_1} \frac{\mathrm{d}z}{z^4+1} = 2\pi i \left(\frac{1}{2z_1(z_1^2 - z_2^2)} \right) = \frac{\pi}{2z_1}.$$

Likewise for $\mathcal{C}_2$ that includes only z_2, we write $z^4 + 1 = (z - z_2)(z + z_2)(z^2 - z_1^2)$ and Cauchy's formula yields

$$\oint_{\mathcal{C}_2} \frac{dz}{z^4+1} = 2\pi i \left(\frac{1}{2z_2(z_2^2 - z_1^2)} \right) = \frac{\pi}{2(-z_2)}.$$

Adding both residues and recalling $z_1 = e^{i\pi/4}$ and $z_2 = -e^{-i\pi/4}$ gives

$$\oint_{\mathcal{C}} \frac{\mathrm{d}z}{z^4+1} = \pi \frac{e^{-i\pi/4} + e^{i\pi/4}}{2} = \pi \cos\frac{\pi}{4} = \frac{\pi}{\sqrt{2}}.$$

As before we need to take $R \to \infty$ and figure out the $\mathcal{C}_R$ part. That part goes to zero as $R \to \infty$ since

$$\left| \int_{\mathcal{C}_R} \frac{\mathrm{d}z}{z^4+1} \right| \le \int_{\mathcal{C}_R} \frac{|\mathrm{d}z|}{|z^4+1|} < \int_0^\pi \frac{R\mathrm{d}\theta}{R^4-1} = \frac{\pi R}{R^4-1}.$$

□

We could extend the same method to

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^8} \mathrm{d}x. \tag{121}$$

We would use the same closed contour again, but now there would be 4 simple poles inside it and therefore 4 separate contributions.

□

Example 4:

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{(1+x^2)^2} = \int_{\mathbb{R}} \frac{\mathrm{d}z}{(z^2+1)^2} = \int_{\mathbb{R}} \frac{\mathrm{d}z}{(z-i)^2(z+i)^2}. \tag{122}$$

We use the same closed contour once more, but now we have a *double pole* inside the contour at $z = i$. We can figure out the contribution from that double pole by using the generalized form of Cauchy's formula (108). The integral over $\mathcal{C}_R$ can be shown to vanish as $R \to \infty$ and

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{(1+x^2)^2} = 2\pi i \left(\frac{\mathrm{d}}{\mathrm{d}z}(z+i)^{-2} \right)_{z=i} = \frac{\pi}{2}. \tag{123}$$

A $(z-a)^n$ in the denominator, with n a positive integer, is called an *n-th order pole.*

□

Example 5:

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} \mathrm{d}x = \pi. \tag{124}$$

This is a trickier problem. Our impulse is to consider $\int_{\mathbb{R}} (\sin z)/z \, dz$ but that integrand is a super good function! Indeed $(\sin z)/z = 1 - z^2/3! + z^4/5! - \cdots$ is analytic in the entire plane, its Taylor series converges in the entire plane. For obvious reasons such functions are called *entire* functions. But we love singularities now since they actually make our life easier. So we write

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} \mathrm{d}x = \mathfrak{Im} \int_{-\infty}^{\infty} \frac{e^{ix}}{x} \mathrm{d}x = \mathfrak{Im} \int_{\mathbb{R}} \frac{e^{iz}}{z} \mathrm{d}z, \tag{125}$$

where $\mathfrak{Im}$ stands for "imaginary part of". Now we have a nice simple pole at $z = 0$. But that's another problem since the pole is *on* the contour! We have to modify our favorite contour a little bit to avoid the pole by going over or below it. If we go below and close along $\mathcal{C}_\theta$ as before, then we'll have a pole inside our contour. If we go over it, we won't have any pole inside the closed contour. We get the same result either way (luckily!), but the algebra is a tad simpler if we leave the pole out.

So we consider the closed contour $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3 + \mathcal{C}_4$ where $\mathcal{C}_1$ is the real axis

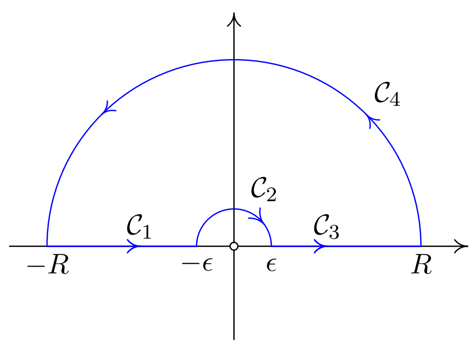


Fig. 3.22: Closing the contour for (125) and calculating a 'half-residue'.

from $-R$ to $-\epsilon$, $\mathcal{C}_2$ is the semi-circle from $-\epsilon$ to ϵ in the top half-plane, $\mathcal{C}_3$ is the real axis from ϵ to R and $\mathcal{C}_4$ is our good old semi-circle of radius R. The integrand e^{iz}/z

is analytic everywhere except at $z = 0$ where it has a simple pole, but since that pole is outside our closed contour, Cauchy's theorem gives $\oint_{\mathcal{C}} = 0$ or

$$\int_{\mathcal{C}_1+\mathcal{C}_3} = -\int_{\mathcal{C}_2} - \int_{\mathcal{C}_4}$$

The integral over the semi-circle $\mathcal{C}_2 : z = \epsilon e^{i\theta}$, $dz = i\epsilon e^{i\theta}\mathrm{d}\theta$, is

$$-\int_{\mathcal{C}_2} \frac{e^{iz}}{z}\mathrm{d}z = i\int_0^{\pi} e^{i\epsilon e^{i\theta}}\mathrm{d}\theta \to \pi i \quad \text{as} \quad \epsilon \to 0.$$

As before we'd like to show that the $\int_{\mathcal{C}_4} \to 0$ as $R \to \infty$. This is trickier than the previous cases we've encountered. On the semi-circle $z = Re^{i\theta}$ and $\mathrm{d}z = iRe^{i\theta}d\theta$, as we've seen so many times, we don't even need to think about it anymore (do you?), so

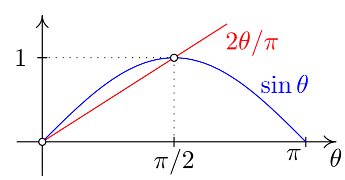
$$\int_{\mathcal{C}_4} \frac{e^{iz}}{z}\mathrm{d}z = i\int_0^{\pi} e^{iRe^{i\theta}}\mathrm{d}\theta = i\int_0^{\pi} e^{iR\cos\theta}e^{-R\sin\theta}\mathrm{d}\theta. \tag{126}$$

This is a pretty scary integral. But with a bit of courage and intelligence it's not as bad as it looks. The integrand has two factors, $e^{iR\cos\theta}$ whose norm is always 1 and $e^{-R\sin\theta}$ which is real and exponentially small for all θ in $0 < \theta < \pi$, except at 0 and π where it is exactly 1. Sketch $e^{-R\sin\theta}$ in $0 \le \theta \le \pi$ and it's pretty clear the integral should go to zero as $R \to \infty$. To show this rigorously, let's consider its modulus (norm) as we did in the previous cases. Then since (i) the modulus of a sum is less or equal to the sum of the moduli (triangle inequality), (ii) the modulus of a product is the product of the moduli and (iii) $|e^{iR\cos\theta}| = 1$ when R and θ are real (which they are)

$$0 \le \left|\int_0^{\pi} e^{iR\cos\theta}e^{-R\sin\theta}\mathrm{d}\theta\right| < \int_0^{\pi} e^{-R\sin\theta}\mathrm{d}\theta \tag{127}$$

we still cannot calculate that last integral but we don't need to. We just need to show that it is smaller than something that goes to zero as $R \to \infty$, so our integral will be *squeezed* to zero.

The plot of $\sin\theta$ for $0 \le \theta \le \pi$ illustrates that it is symmetric with respect to $\pi/2$ and that $2\theta/\pi \le \sin\theta$ when $0 \le \theta \le \pi/2$, or changing the signs $-2\theta/\pi \ge -\sin\theta$ and since e^x increases monotonically with x,



$$e^{-R\sin\theta} < e^{-2R\theta/\pi}$$

in $0 \le \theta \le \pi/2$. This yields **Jordan's Lemma**

$$\left|\int_0^{\pi} e^{iRe^{i\theta}}\mathrm{d}\theta\right| < \int_0^{\pi} e^{-R\sin\theta}\mathrm{d}\theta = 2\int_0^{\pi/2} e^{-R\sin\theta}\mathrm{d}\theta$$
$$< 2\int_0^{\pi/2} e^{-2R\theta/\pi}\mathrm{d}\theta = \pi\frac{1-e^{-R}}{R} \tag{128}$$

so $\int_{\mathcal{C}_4} \to 0$ as $R \to \infty$ and collecting our results we obtain (124).

□

Exercises

1. Study all the solved examples in this section.

2. Sketch $f(x) = 1/(a + b\sin x)$. Calculate $\int_0^{2\pi} f(x)\mathrm{d}x$ where a and b are real numbers with $|a| > |b|$. Why does $|a| > |b|$ matter?

3. Sketch $f(x) = 1/(1+\cos^2 x)$ and $g(x) = 1/(1+\sin^2 x)$. Calculate $\int_0^{2\pi} f(x)\mathrm{d}x$. Why is it equal to $\int_0^{2\pi} g(x)\mathrm{d}x$?

4. Sketch $f(\theta) = e^{R\cos\theta}$ and $g(\theta) = e^{R\sin\theta}$ where R is a positive constant. Calculate $\int_0^{2\pi} e^{R\cos\theta}\mathrm{d}\theta$.

5. Calculate $c_n \triangleq \int_0^{2\pi} e^{R\cos\theta}\, \cos n\theta\, \mathrm{d}\theta$, where n is an integer.

6. Make up and solve an exam question which is basically the same as $\int_{-\infty}^{\infty} \mathrm{d}x/(1+x^2)$ in terms of the logic and difficulty, but is different in the details.

7. Calculate $\int_{-\infty}^{\infty} \mathrm{d}x/(1 + x^2 + x^4)$. Can you provide an *a priori* upper bound for this integral based on integrals calculated earlier?

8. Consider $F(k) = \displaystyle\int_{-\infty}^{\infty} \frac{e^{ikx}}{a^2 + x^2}\,\mathrm{d}x$, the Fourier transform of the *Lorentzian*, where a and k are real. Show that $F(k) = F(-k)$. Calculate $F(k)$.

9. Given the Poisson integral $\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x = \sqrt{\pi}$, what is $\int_{-\infty}^{\infty} e^{-x^2/a^2} dx$ where a is real? (that should be easy!). Next, calculate
$$\int_{-\infty}^{\infty} e^{-x^2/a^2} e^{ikx}\mathrm{d}x$$
where a and k are arbitrary *real* numbers. This is the *Fourier transform* of the Gaussian e^{-x^2/a^2}. [Complete the square. Note that you can pick $a,\ k > 0$ (why?), then integrate over the rectangle $-R \to R \to R + ika^2/2 \to -R + ika^2/2 \to -R$. (why?justify).]

10. Calculate
$$\int_{-\infty}^{\infty} \frac{e^{ax}}{1 + e^x}\mathrm{d}x$$
for $0 < a < 1$ using the rectangular contour $-R \to R \to R + 2\pi i \to -R + 2\pi i \to -R$.

11. Show that
$$\int_0^{\infty} \frac{\mathrm{d}x}{1 + x^3} = \frac{2\pi}{3\sqrt{3}}$$
using the straight line $z = 0 \to R$, the circular arc $z = R \to Re^{i2\pi/3}$ and the straight line $z = Re^{2i\pi/3} \to 0$.

12. Calculate
$$\int_0^\infty \frac{\mathrm{d}x}{1+x^5}.$$
Explain what contour to take and why.

13. The *Fresnel* integrals come up in optics and quantum mechanics. They are
$$\int_{-\infty}^{\infty} \cos x^2 \,\mathrm{d}x, \quad \text{and} \quad \int_{-\infty}^{\infty} \sin x^2 \,\mathrm{d}x.$$
Calculate them both by considering $\int_0^\infty e^{ix^2}\mathrm{d}x$. The goal is to reduce this to a Poisson integral. This would be the case if $x^2 \to (e^{i\pi/4}x)^2$. So consider the closed path that goes from 0 to R on the real axis, then on the circle of radius R to $Re^{i\pi/4}$ (or the vertical line $R \to R+iR$), then back along the diagonal $z = se^{i\pi/4}$ with s real.

14. Show that the *Glauert integrals* of Aerodynamics
$$\int_0^{2\pi} \frac{\cos n\theta}{\cos\theta - \cos\theta_0}\mathrm{d}\theta = \pi\frac{\sin n\theta_0}{\sin\theta_0},$$
where the integral is interpreted as the *Cauchy Principal value* (that is, with a $\pm\epsilon$ interval around the singularity, as in fig. 3.22), where n is an integer and θ_0 is real.

Peeking at branch cuts

You may have noticed that we've only dealt with integer powers. What about fractional powers? First let's take a look at the integral of $\sqrt{z}$ over the unit circle $z = e^{i\theta}$ from $\theta = \theta_0$ to $\theta_0 + 2\pi$
$$\oint_{|z|=1} \sqrt{z}\,\mathrm{d}z = \int_{\theta_0}^{\theta_0+2\pi} e^{i\theta/2} ie^{i\theta}\mathrm{d}\theta = \frac{2i}{3}e^{3i\theta_0/2}(e^{i3\pi}-1) = \frac{-4i}{3}e^{3i\theta_0/2} \qquad (129)$$
The answer depends on θ_0! The integral over the closed circle depends on where we start on the circle?! This is weird, what's going on? The problem is with the definition of $\sqrt{z}$. We have implicitly defined $\sqrt{z} = |z|^{1/2}e^{i\arg(z)/2}$ with $\theta_0 \le \arg(z) < \theta_0 + 2\pi$ or $\theta_0 < \arg(z) \le \theta_0 + 2\pi$. But each θ_0 corresponds to a different definition for $\sqrt{z}$.

For real variables the equation $y^2 = x \ge 0$ had two solutions $y = \pm\sqrt{x}$ and we defined $\sqrt{x} \ge 0$. Can't we define $\sqrt{z}$ in a similar way? The equation $w^2 = z$ in the complex plane always has two solutions. We can say $\sqrt{z}$ and $-\sqrt{z}$ but we still need to define $\sqrt{z}$ since z is complex. Could we define $\sqrt{z}$ to be such that its *real* part is always positive? yes, and that's equivalent to defining $\sqrt{z} = |z|^{1/2}e^{i\arg(z)/2}$ with $-\pi < \arg(z) < \pi$ (check it). But that's not complete because the sqrt of a negative real number is pure imaginary, so what do we do about those numbers? We can define $-\pi < \arg(z) \leq \pi$, so real negative numbers have $\arg(z) = \pi$, not $-\pi$, by definition. This is indeed the definition that Matlab chooses. But it may not be appropriate for our problem *because it introduces a discontinuity in* $\sqrt{z}$ *as we cross the negative real axis.* If that is not desirable for our problem then we could define $0 \leq \arg(z) < 2\pi$. Now $\sqrt{z}$ is continuous across the negative real axis but there is a jump across the positive real axis. No matter what definition we pick, there will always be a discontinuity somewhere. We cannot go around $z = 0$ without encountering such a jump, $z = 0$ is called a *branch point* and the semi-infinite curve emanating from $z = 0$ across which $\arg(z)$ jumps is called a *branch cut.* Functions like $z^{3/2}$, $\ln z$, $\sqrt{z^2 - 1}$, etc. have similar issues. The functions are not differentiable across the cut and therefore Cauchy's theorem does apply directly. It can still be used as long as we do not cross branch cuts.

Here's a simple example that illustrates the extra subtleties and techniques.

$$\int_0^\infty \frac{\sqrt{x}}{1+x^2} \mathrm{d}x$$

First note that this integral does indeed exist since $\sqrt{x}/(1 + x^2) \sim x^{-3/2}$ as $x \to \infty$ and therefore goes to zero fast enough to be integrable. Our first impulse is to see this as an integral over the real axis from 0 to ∞ of the complex function $\sqrt{z}/(z^2 + 1)$. That function has simple poles at $z = \pm i$ as we know well. But there's a problem: $\sqrt{z}$ is not analytic at $z = 0$ which is on our contour again. No big deal, we can avoid it as we saw in the $(\sin x)/x$ example. So let's take the same 4-piece closed contour as in that problem. But we're not all set yet because we have a $\sqrt{z}$, what do we mean by that when z is complex? We need to define that function so that it is *analytic everywhere inside and on our contour.* Writing $z = |z|e^{i \arg(z)}$ then we can define $\sqrt{z} = |z|^{1/2} e^{i \arg(z)/2}$. We need to define $\arg(z)$ so $\sqrt{z}$ is analytic inside and on our contour. The definitions $-\pi \leq \arg(z) < \pi$ would not work with our decision to close in the upper half place. Why? because $\arg(z)$ and thus $\sqrt{z}$ would not be continuous at the junction between $\mathcal{C}_4$ and $\mathcal{C}_1$. We could close in the lower half plane, or we can pick another branch cut for $\arg(z)$. The standard definition $-\pi < \arg(z) \leq \pi$ works. Try it!