MATH 241

Calculus III

Semester: Fall 2025

Professor: Scott Ahlgren

Contents

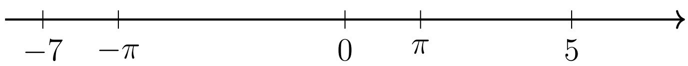
1 Vectors and the Geometry of Space

1.1 An introduction to multivariable calculus

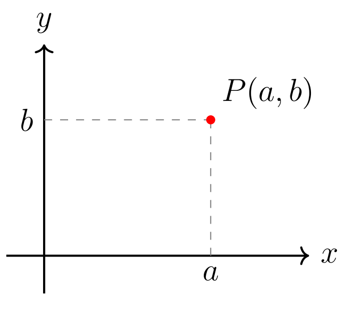
Single-variable calculus deals with functions of one input. The territory is the real line, and a function maps $\mathbb{R}$ into itself; differentiation gives a slope, integration gives an area, and most geometric phenomena reduce to drawings on a sheet of paper. Multivariable calculus opens up the same machinery in higher dimensions. Functions can take several inputs, produce several outputs, or do both. Their graphs are surfaces, their derivatives are vectors of partial derivatives, and their integrals measure volumes, masses, or flow. Before any of this works, we need a vocabulary for points, vectors, and operations in higher-dimensional space, which is the content of this first chapter.

1.2 $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$

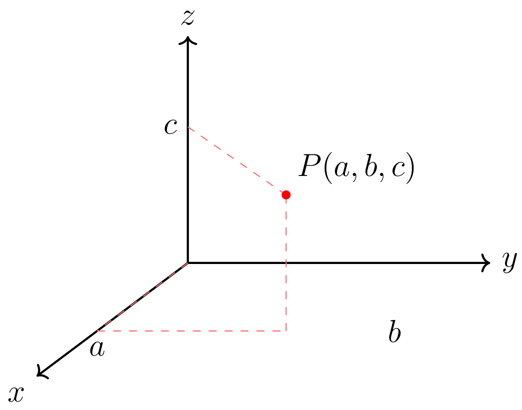
$\mathbb{R}$ denotes the set of real numbers. A real number can be thought of as a point on a number line.



$\mathbb{R}^2$ refers to the plane: a point in $\mathbb{R}^2$ requires two coordinates (x, y) to be specified.



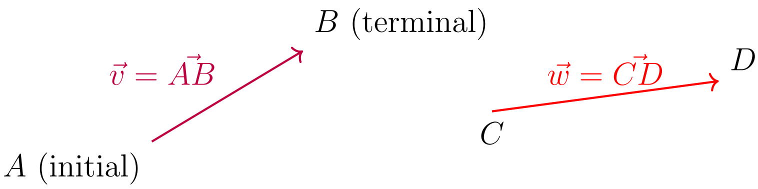
$\mathbb{R}^3$ refers to three-dimensional space: a point requires three coordinates (x, y, z). We will work mostly in $\mathbb{R}^3$ for the rest of the course, although many results generalise without much effort to $\mathbb{R}^n$ for any positive integer n.



1.3 An introduction to vectors

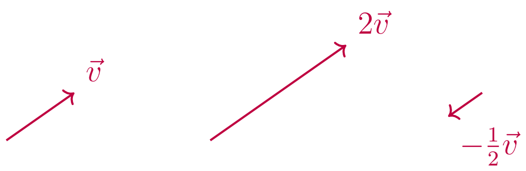
A vector is a quantity carrying both magnitude and direction. Velocity, momentum, and force are all naturally vector quantities: knowing how fast a car is moving is not enough information — we also need to know which way. By contrast a scalar is a quantity with magnitude alone. Speed, temperature, and time are scalars: a single number suffices to describe each.

Vectors can be drawn as arrows in a coordinate system. The length of the arrow gives the magnitude, and the direction the arrow points gives the direction.

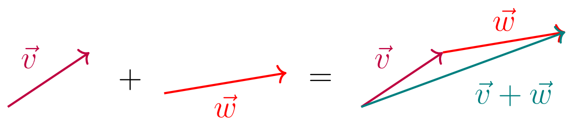


Two operations are fundamental. We can scale a vector by a real number and we can add two vectors together.

Multiplying a vector by a scalar produces a parallel vector with a possibly different length, and possibly the opposite direction (if the scalar is negative). Multiplying by a positive scalar preserves direction; multiplying by zero produces the zero vector; multiplying by a negative scalar reverses direction. Scalar multiples of a single nonzero vector are all parallel to one another.



Vector addition is performed geometrically by the head-to-tail rule: place the tail of the second vector at the head of the first, and draw a new arrow from the tail of the first to the head of the second. The resulting arrow is the sum.



Vector addition is commutative: $\vec{v}+\vec{w}=\vec{w}+\vec{v}$. Geometrically, this is the parallelogram law: completing the parallelogram on $\vec{v}$ and $\vec{w}$ gives the same diagonal regardless of the order in which we add. Vector subtraction $\vec{v}-\vec{w}$ is defined as $\vec{v}+(-\vec{w})$ and is not commutative.

A position vector is a vector $\vec{v}$ whose tail is fixed at the origin and whose head sits at some point $P(v_1, v_2)$. We write

$$\vec{v} = \langle v_1, v_2 \rangle\,, \tag{1.1}$$

where the angled brackets are intentionally distinct from the parentheses we used for points: $\langle v_1, v_2\rangle$ is a vector, (v_1, v_2) is a point. The components v_1 and v_2 are the x- and y-components of $\vec{v}$. Any vector in the plane is equal to a unique position vector — given $\overrightarrow{P_1P_2}$ with $P_1=(x_1,y_1)$ and $P_2=(x_2,y_2)$, we translate so the tail sits at the origin, and the corresponding position vector is

$$\vec{v} = \langle x_2 - x_1,\ y_2 - y_1 \rangle\,. \tag{1.2}$$

The magnitude of $\vec{v}=\langle v_1, v_2\rangle$ is the length of the corresponding arrow, given by the

Pythagorean theorem,

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}. \tag{1.3}$$

If we denote by θ the angle the vector makes with the positive x-axis, simple right-triangle trigonometry gives the components in terms of magnitude and angle,

$$v_1 = \|\vec{v}\| \cos\theta, \tag{1.4}$$

$$v_2 = \|\vec{v}\| \sin\theta. \tag{1.5}$$

Two vectors are equal if and only if they have the same magnitude and the same direction; equivalently, in component form, if and only if their components agree slot by slot. Operations on position vectors are particularly clean to work with in components: addition is componentwise, and scalar multiplication scales each component by the scalar:

$$\vec{v} + \vec{w} = \langle v_1 + w_1,\ v_2 + w_2 \rangle\,, \tag{1.6}$$

$$k\vec{v} = \langle kv_1,\ kv_2 \rangle\,. \tag{1.7}$$

1.3.1 Unit vectors

A unit vector is a vector of magnitude exactly 1. We use a hat to denote unit vectors: $\hat{\boldsymbol{v}}$. Any nonzero vector $\vec{v}$ can be turned into a unit vector pointing in the same direction by dividing by its magnitude,

$$\hat{\boldsymbol{v}} = \frac{\vec{v}}{\|\vec{v}\|}. \tag{1.8}$$

Direct calculation confirms that this formula does produce a unit vector, since $\|\vec{v}/\,\|\vec{v}\|\| = \|\vec{v}\|\,/\,\|\vec{v}\| = 1$. Multiplying back through gives a useful decomposition of any vector into magnitude times direction,

$$\vec{v} = \|\vec{v}\| \cdot \hat{\boldsymbol{v}}. \tag{1.9}$$

Two parallel vectors share the same unit vector — this is one way to define "same direction" precisely.

In $\mathbb{R}^2$ there are two especially important unit vectors that point along the coordinate axes,

$$\hat{\boldsymbol{\imath}} = \langle 1, 0 \rangle , \qquad \hat{\boldsymbol{\jmath}} = \langle 0, 1 \rangle , \tag{1.10}$$

called the standard basis vectors. Any planar vector can be written as a linear combination of these two: a vector $\vec{v} = \langle a, b \rangle$ may be expressed as $\vec{v} = a\hat{\boldsymbol{\imath}} + b\hat{\boldsymbol{\jmath}}$, where the scalars a and b specify how far the vector reaches in the x and y directions respectively.

Proposition 1.1. For any vector $\vec{v} = \langle a, b \rangle$ in $\mathbb{R}^2$, $\vec{v} = a\hat{\boldsymbol{\imath}} + b\hat{\boldsymbol{\jmath}}$.

Proof. Computing the right-hand side using the rules for scalar multiplication and addition,

$$a\hat{\boldsymbol{\imath}} + b\hat{\boldsymbol{\jmath}} = a\langle 1, 0 \rangle + b\langle 0, 1 \rangle = \langle a \cdot 1, a \cdot 0 \rangle + \langle b \cdot 0, b \cdot 1 \rangle = \langle a, 0 \rangle + \langle 0, b \rangle = \langle a, b \rangle = \vec{v}.$$

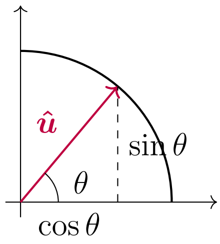
Each step is one of: applying scalar multiplication componentwise, or applying addition componentwise. ■

A unit vector at a known angle θ from the x-axis can also be expressed in terms of trigonometric functions, which is often a more useful form when angles rather than components are given.

Proposition 1.2. If $\hat{\boldsymbol{u}}$ is a unit vector in $\mathbb{R}^2$ making angle θ with the positive x-axis, then

$$\hat{\boldsymbol{u}} = \langle \cos\theta, \sin\theta \rangle = \cos(\theta)\hat{\boldsymbol{\imath}} + \sin(\theta)\hat{\boldsymbol{\jmath}}. \tag{1.11}$$

Proof. Place the tail of $\hat{\boldsymbol{u}}$ at the origin. Because $\hat{\boldsymbol{u}}$ has length 1, its head lies on the unit circle. The geometric definition of cosine and sine on the unit circle is precisely that the horizontal coordinate of a point on the unit circle at angle θ from the positive x-axis is $\cos\theta$, and the vertical coordinate is $\sin\theta$. Therefore the head of $\hat{\boldsymbol{u}}$ is at the point $(\cos\theta, \sin\theta)$, and the corresponding position vector is $\langle \cos\theta, \sin\theta \rangle$.



Rewriting the position vector in standard basis form gives $\hat{\boldsymbol{u}} = \cos(\theta)\hat{\boldsymbol{\imath}} + \sin(\theta)\hat{\boldsymbol{\jmath}}$. ■

1.3.2 Vectors in $\mathbb{R}^3$

A position vector in $\mathbb{R}^3$ has three components,

$$\vec{v} = \langle v_1, v_2, v_3 \rangle \,, \tag{1.12}$$

with three corresponding standard basis vectors $\hat{\boldsymbol{\imath}} = \langle 1, 0, 0 \rangle$, $\hat{\boldsymbol{\jmath}} = \langle 0, 1, 0 \rangle$, $\hat{\boldsymbol{k}} = \langle 0, 0, 1 \rangle$. Any 3D vector can be written $\vec{v} = v_1\hat{\boldsymbol{\imath}} + v_2\hat{\boldsymbol{\jmath}} + v_3\hat{\boldsymbol{k}}$. The magnitude formula extends from $\mathbb{R}^2$ in the natural way,

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}, \tag{1.13}$$

obtained by applying the Pythagorean theorem twice (once in the xy-plane to get the projected length, then again to combine that projection with the z-component). Two vectors in $\mathbb{R}^3$ are parallel if one is a scalar multiple of the other; vector addition and scalar multiplication are again componentwise.

1.4 The dot product

The dot product takes two vectors and returns a scalar. It is the simplest way to combine two vectors algebraically, and from it we can extract information about angles, lengths, and orthogonality. For $\vec{v} = \langle v_1, v_2, v_3 \rangle$ and $\vec{w} = \langle w_1, w_2, w_3 \rangle$, define

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + v_3 w_3. \tag{1.14}$$

The output is a single scalar, even though both inputs are vectors. The same formula applies in $\mathbb{R}^2$ with the third term omitted.

The geometric content of the dot product comes from a second formula equivalent to the algebraic definition above,

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \, \|\vec{w}\| \cos\theta, \qquad 0 \le \theta \le \pi, \tag{1.15}$$

where θ is the angle between $\vec{v}$ and $\vec{w}$ when their tails are placed at a common point.

Proposition 1.3 (Geometric formula for the dot product)**.** For nonzero vectors $\vec{v}$ and $\vec{w}$ separated by angle θ,

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \, \|\vec{w}\| \cos\theta.$$

Proof. Place $\vec{v}$ and $\vec{w}$ tail to tail. The triangle formed by their two heads and their common tail has sides of length $\|\vec{v}\|$, $\|\vec{w}\|$, and $\|\vec{v} - \vec{w}\|$, where $\vec{v} - \vec{w}$ is the vector connecting the two heads. The angle opposite the third side is θ. The law of cosines applied to this triangle says

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2 \, \|\vec{v}\| \, \|\vec{w}\| \cos\theta.$$

On the other hand, the squared magnitude of any vector $\vec{u}$ equals $\vec{u} \cdot \vec{u}$ by the algebraic formula. Applying this to $\vec{v} - \vec{w}$ and expanding by linearity of the dot product (which follows directly from the algebraic definition),

$$\|\vec{v} - \vec{w}\|^2 = (\vec{v} - \vec{w}) \cdot (\vec{v} - \vec{w}) = \vec{v} \cdot \vec{v} - 2\vec{v} \cdot \vec{w} + \vec{w} \cdot \vec{w} = \|\vec{v}\|^2 - 2\vec{v} \cdot \vec{w} + \|\vec{w}\|^2 .$$

Setting the two expressions for $\|\vec{v} - \vec{w}\|^2$ equal,

$$\|\vec{v}\|^2 - 2\vec{v} \cdot \vec{w} + \|\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2 \, \|\vec{v}\| \, \|\vec{w}\| \cos\theta.$$

The $\|\vec{v}\|^2$ and $\|\vec{w}\|^2$ terms cancel from both sides, leaving $-2\vec{v} \cdot \vec{w} = -2 \, \|\vec{v}\| \, \|\vec{w}\| \cos\theta$. Dividing by -2 produces the claim. ■

Solving for the angle gives a way to compute it from the components alone,

$$\theta = \arccos\left(\frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\,\|\vec{w}\|}\right). \tag{1.16}$$

Since cos is a one-to-one function on $[0,\pi]$, this formula determines θ uniquely in that range.

Two special cases of this formula deserve to be highlighted. The angle is 0 or π exactly when the vectors are parallel (in the same or opposite directions). The angle is $\pi/2$ exactly when the vectors are perpendicular, in which case $\cos\theta = 0$ and the dot product vanishes. Conversely, if the dot product of two nonzero vectors is zero, the angle between them must be $\pi/2$. We summarise:

$$\vec{v}\cdot\vec{w} = 0 \iff \vec{v}\perp\vec{w}. \tag{1.17}$$

This is by far the most common use of the dot product in practice: a quick algebraic test for orthogonality.

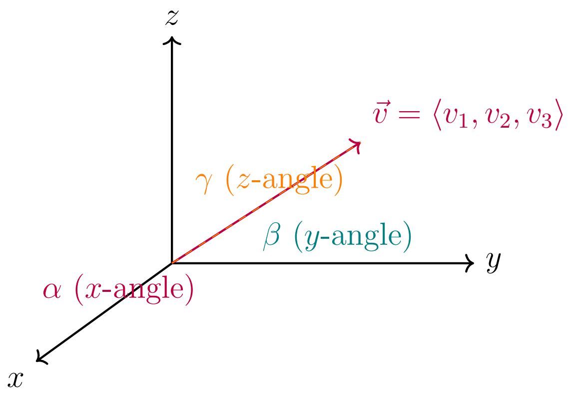
1.4.1 Work and direction cosines

Work is the prototypical physical example of a dot product. When a constant force $\vec{F}$ acts on an object that moves through a displacement $\vec{d}$, the work done by the force is

$$W = \vec{F}\cdot\vec{d} = \left\|\vec{F}\right\|\left\|\vec{d}\right\|\cos\theta. \tag{1.18}$$

Geometrically, only the component of $\vec{F}$ along $\vec{d}$ contributes to the work; the perpendicular component does no work. The dot product picks this out automatically.

A second application is the direction cosines of a vector, which describe its direction by giving the angles it makes with each coordinate axis. Let $\vec{v}$ be a nonzero vector in $\mathbb{R}^3$ and let α, β, γ denote the angles between $\vec{v}$ and the positive x-, y-, z-axes respectively.



By the dot-product formula applied to $\vec{v}$ and the standard basis vectors,

$$\cos\alpha = \frac{\vec{v}\cdot\hat{\boldsymbol{\imath}}}{\|\vec{v}\|} = \frac{v_1}{\|\vec{v}\|}, \tag{1.19}$$

$$\cos\beta = \frac{v_2}{\|\vec{v}\|}, \tag{1.20}$$

$$\cos\gamma = \frac{v_3}{\|\vec{v}\|}. \tag{1.21}$$

The three direction cosines are not independent — they satisfy a single Pythagorean-type identity.

Proposition 1.4. $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$

Proof. Substituting the formulas for the direction cosines and combining over a common denominator,

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = \frac{v_1^2}{\|\vec{v}\|^2} + \frac{v_2^2}{\|\vec{v}\|^2} + \frac{v_3^2}{\|\vec{v}\|^2} = \frac{v_1^2 + v_2^2 + v_3^2}{\|\vec{v}\|^2}.$$

The numerator $v_1^2 + v_2^2 + v_3^2$ is exactly $\|\vec{v}\|^2$ by the magnitude formula. Hence the right-hand side is $\|\vec{v}\|^2 / \|\vec{v}\|^2 = 1$. ∎

1.4.2 Projections

Given two nonzero vectors, we often want to extract the component of one in the direction of the other. There are two flavours of this: the scalar projection (a number giving the signed length of the component), and the vector projection (a vector pointing along

$\vec{w}$ with the appropriate length).

The scalar projection of $\vec{v}$ onto $\vec{w}$, denoted $\text{comp}_{\vec{w}}\vec{v}$, is the signed length of the shadow that $\vec{v}$ would cast on the line spanned by $\vec{w}$,

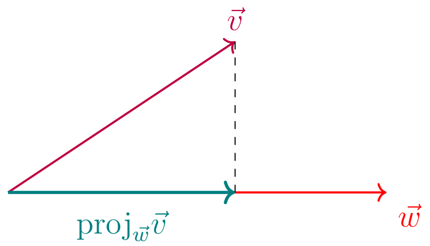
$$\text{comp}_{\vec{w}}\vec{v} = \frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|} = \|\vec{v}\|\cos\theta. \tag{1.22}$$

The two expressions are equivalent by the geometric formula for the dot product. The first expression is convenient for computation, while the second makes the geometry transparent.

The vector projection of $\vec{v}$ onto $\vec{w}$ is the scalar projection times the unit vector in the direction of $\vec{w}$,

$$\text{proj}_{\vec{w}}\vec{v} = \left(\frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|^2}\right)\vec{w} = \left(\frac{\vec{v}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\right)\vec{w}. \tag{1.23}$$

The geometry: drop a perpendicular from the head of $\vec{v}$ onto the line through $\vec{w}$; the resulting foot of the perpendicular, viewed as the head of a vector based at the tail, is $\text{proj}_{\vec{w}}\vec{v}$.



1.5 The cross product

Where the dot product takes two vectors and returns a scalar, the cross product takes two vectors in $\mathbb{R}^3$ and returns a third vector. It exists only in $\mathbb{R}^3$ (and, with modifications, in $\mathbb{R}^7$); there is no direct analogue in $\mathbb{R}^2$. The cross product is engineered to satisfy two properties simultaneously: the result is a vector orthogonal to both inputs, and its magnitude equals the area of the parallelogram they span.

For $\vec{v} = \langle v_1, v_2, v_3 \rangle$ and $\vec{w} = \langle w_1, w_2, w_3 \rangle$, the cross product is most compactly written as a determinant,

$$\vec{v} \times \vec{w} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \hat{\boldsymbol{\imath}} \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - \hat{\boldsymbol{\jmath}} \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + \hat{\boldsymbol{k}} \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}. \tag{1.24}$$

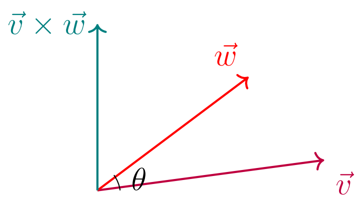
The first "row" of the determinant is symbolic: we treat $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$ as if they were scalars during expansion, then collect like terms.

The magnitude of the cross product has a clean geometric interpretation,

$$\|\vec{v} \times \vec{w}\| = \|\vec{v}\| \, \|\vec{w}\| \sin\theta, \tag{1.25}$$

where θ is the angle between $\vec{v}$ and $\vec{w}$. The right-hand side is exactly the area of the parallelogram spanned by $\vec{v}$ and $\vec{w}$: $\|\vec{v}\|$ is the base, and $\|\vec{w}\| \sin\theta$ is the perpendicular height. Half this quantity is the area of the triangle with vertices at the common tail and the two heads,

$$A_{\triangle} = \tfrac{1}{2} \|\vec{v} \times \vec{w}\|. \tag{1.26}$$



The direction of the cross product is determined by the right-hand rule: point the fingers of your right hand in the direction of $\vec{v}$, curl them toward $\vec{w}$, and your thumb points in the direction of $\vec{v} \times \vec{w}$. This convention makes the cross product orientation-sensitive: swapping the inputs flips the sign.

1.5.1 Properties of the cross product

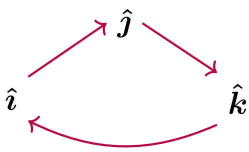
The cross product has several algebraic properties that follow from the determinant definition. Some of these mirror the properties of multiplication, but with crucial differences.

1. Anticommutativity: $\vec{v} \times \vec{w} = -(\vec{w} \times \vec{v})$. Swapping the two rows of a determinant changes its sign.
2. Self-cross product: $\vec{v} \times \vec{v} = \mathbf{0}$. A determinant with two equal rows vanishes.
3. Cross product with zero: $\vec{v} \times \mathbf{0} = \mathbf{0}$.
4. Distributivity: $\mathbf{u} \times (\vec{+}\vec{w}) = \mathbf{u} \times \vec{+}\mathbf{u} \times \vec{w}$.
5. Scalar triple product symmetry: $\mathbf{u} \cdot (\vec{\times}\vec{w}) = \vec{w} \cdot (\mathbf{u} \times \vec{)}$.

The standard basis vectors satisfy the cyclic identities

$$\hat{\boldsymbol{\imath}} \times \hat{\boldsymbol{\jmath}} = \hat{\boldsymbol{k}}, \qquad \hat{\boldsymbol{\jmath}} \times \hat{\boldsymbol{k}} = \hat{\boldsymbol{\imath}}, \qquad \hat{\boldsymbol{k}} \times \hat{\boldsymbol{\imath}} = \hat{\boldsymbol{\jmath}}, \tag{1.27}$$

with the reverse products giving the negatives. The mnemonic is to imagine the three letters arranged in a cycle; advancing in the forward direction yields the next basis vector, while reversing produces the negative.



1.5.2 Torque

One direct physical application of the cross product is the torque exerted by a force applied at a position. Where force is the linear analogue of acceleration, torque is its rotational analogue. If a force $\vec{F}$ is applied at the head of a position vector $\mathbf{r}$ from a pivot, the torque is

$$\vec{\tau} = \mathbf{r} \times \vec{F}, \qquad \|\vec{\tau}\| = \|\mathbf{r}\| \left\|\vec{F}\right\| \sin\theta. \tag{1.28}$$

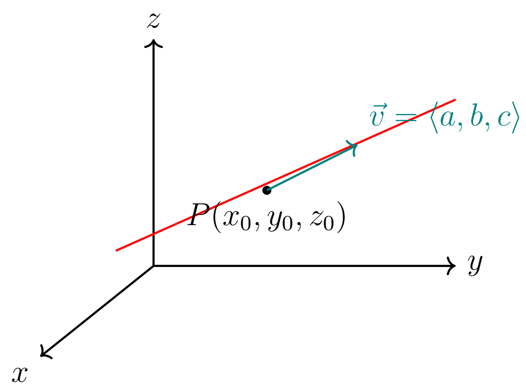
Only the component of $\vec{F}$ perpendicular to $\mathbf{r}$ contributes to torque (otherwise we would simply be pushing or pulling along the lever arm). The cross product packages this geometric fact into a single algebraic operation, with the result vector pointing along the axis of rotation.

1.6 Lines and planes

1.6.1 Lines in space

A line in $\mathbb{R}^2$ is determined by a slope-intercept pair, but in $\mathbb{R}^3$ slope is a richer concept — there is no single "slope" for a line in space. Instead, we describe a line by a point through which it passes and a direction vector that gives its orientation.

Let $P = (x_0, y_0, z_0)$ be a fixed point on the line and let $\vec{v} = \langle a, b, c\rangle$ be a nonzero direction vector parallel to the line. Any other point on the line is reached by starting at P and travelling some scalar multiple of $\vec{v}$.



The vector equation of the line is

$$L : \mathbf{r} = \mathbf{r}_0 + \lambda\vec{v}, \tag{1.29}$$

where $\mathbf{r}_0 = \langle x_0, y_0, z_0\rangle$ is the position vector of P and the parameter λ ranges over $\mathbb{R}$. As λ varies, the head of $\mathbf{r}$ traces out every point of the line. Writing this equation componentwise gives the parametric equations,

$$x = x_0 + \lambda a, \tag{1.30}$$

$$y = y_0 + \lambda b, \tag{1.31}$$

$$z = z_0 + \lambda c. \tag{1.32}$$

Solving each for λ (assuming a, b, c are all nonzero) and equating the three expressions yields the symmetric equations,

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{1.33}$$

The numbers a, b, c are the direction numbers of the line; they are unique up to a common nonzero scalar. If one of the direction numbers is zero, the line lies in a plane perpendicular to that axis (we cannot divide by zero in the symmetric form, so we instead state that coordinate as constant). Two lines are parallel exactly when their direction vectors are scalar multiples of each other.

1.6.2 Planes

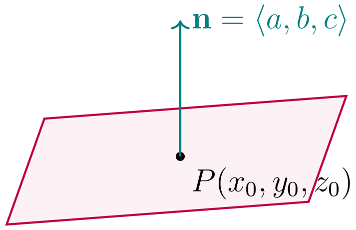
A plane in $\mathbb{R}^3$ is determined by an equation of the form

$$ax + by + cz = d. \tag{1.34}$$

Geometrically, a plane is fixed by a single point on it together with a vector perpendicular to it; we call the perpendicular a normal vector. If the plane passes through $P = (x_0, y_0, z_0)$ with normal $\mathbf{n} = \langle a, b, c \rangle$, every other point $Q = (x, y, z)$ on the plane satisfies the condition that the vector $\vec{PQ}$ lies entirely within the plane, hence is perpendicular to $\mathbf{n}$.

Translating “perpendicular” into a vanishing dot product gives

$$\mathbf{n} \cdot \vec{PQ} = 0, \tag{1.35}$$



which on expansion becomes

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \tag{1.36}$$

Distributing and collecting the constant terms recovers the standard form $ax+by+cz = d$ with $d = ax_0 + by_0 + cz_0$.

One useful corollary is the formula for the distance from a point to a plane. Given a plane $Ax + By + Cz = D$ and a point (x_1, y_1, z_1), the perpendicular distance from the point to the plane is

$$\text{dist} = \frac{|Ax_1 + By_1 + Cz_1 - D|}{\sqrt{A^2 + B^2 + C^2}}. \tag{1.37}$$

The numerator measures by how much the point fails to satisfy the plane equation; the denominator normalises by the magnitude of the normal vector $\langle A, B, C\rangle$ so that the result is a true distance rather than a multiple of it.

Example 0

Find the equation of the plane through $P(1, 2, 3)$ with normal $\mathbf{n} = \langle 2, -1, 4\rangle$.

Solution

Substituting the point and the normal into $a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$,

$$2(x - 1) - (y - 2) + 4(z - 3) = 0.$$

Distributing and collecting the constants on the right-hand side,

$$2x - y + 4z = 12.$$

1.7 Cylinders and quadric surfaces

A cylinder in $\mathbb{R}^3$ is the surface generated by all lines parallel to a given line that pass through a fixed planar curve. A familiar example is the right circular cylinder $x^2+y^2 = 1$, which extends indefinitely along the z-axis. Whenever an equation in $\mathbb{R}^3$ omits one of the variables, the resulting surface is a cylinder along the missing axis: every value of that variable is allowed, so the cross-section is the same in every plane perpendicular to the axis.

A quadric surface is the graph of a degree-two polynomial equation in x, y, z,

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0. \tag{1.38}$$

Quadric surfaces include several geometrically important examples that we should know by sight.

The ellipsoid is the locus

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1,$$

a generalisation of the sphere; when $a = b = c$ it is a sphere of that common radius.

The elliptic paraboloid satisfies

$$\frac{z}{c} = \frac{x^2}{a^2} + \frac{y^2}{b^2}.$$

Its level sets at constant z are ellipses (the “paraboloid” part is that the cross-sections in vertical planes are parabolas).

The hyperbolic paraboloid (the saddle surface) satisfies

$$\frac{z}{c} = \frac{x^2}{a^2} - \frac{y^2}{b^2}.$$

It curves upward in one direction and downward in the perpendicular one — the prototypical saddle.

The cone has the equation

$$\frac{z^2}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2},$$

a quadric whose level sets at constant nonzero z are ellipses, growing linearly with $|z|$.

The hyperboloid of one sheet,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$

looks like a single connected piece pinched at the waist; level sets are ellipses.

The hyperboloid of two sheets,

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1,$$

consists of two disjoint bowl-shaped pieces, one above and one below the xy-plane.

The standard tool for identifying a quadric surface from its equation is to take traces: substitute $z = k$ for various constants k and observe what curve the equation produces in the resulting plane. Combining the traces from all three coordinate planes (and from various heights) usually pins the surface down.

1.8 Cylindrical and spherical coordinates

Cartesian coordinates work well for many problems, but for problems with rotational symmetry around an axis (or around a point) they often hide the symmetry behind cumbersome algebra. Cylindrical and spherical coordinates are alternative coordinate systems that make rotational symmetry explicit.

1.8.1 Cylindrical coordinates

Cylindrical coordinates (r, θ, z) are obtained by replacing the xy-plane Cartesian coordinates with polar coordinates (r, θ) and leaving the z-coordinate unchanged. The conversion formulas are

$$x = r\cos\theta, \qquad r = \sqrt{x^2 + y^2}, \tag{1.39}$$

$$y = r\sin\theta, \qquad \theta = \arctan(y/x), \tag{1.40}$$

$$z = z. \tag{1.41}$$

Cylindrical coordinates are well-suited to problems with axial symmetry around the z-axis: a circular cylinder around the z-axis becomes simply $r = \text{const}$, and a horizontal plane is $z = \text{const}$.

1.8.2 Spherical coordinates

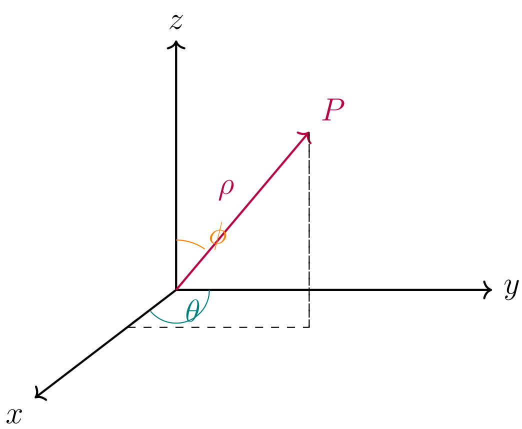
Spherical coordinates (ρ, θ, ϕ) describe a point by its distance ρ from the origin, the azimuthal angle θ (the angle around the z-axis, measured from the positive x-axis in the xy-plane), and the polar angle ϕ (measured from the positive z-axis down to the position vector). The conversion formulas are

$$x = \rho\sin\phi\cos\theta, \tag{1.42}$$

$$y = \rho\sin\phi\sin\theta, \tag{1.43}$$

$$z = \rho\cos\phi. \tag{1.44}$$

By convention $\rho \geq 0$, $0 \leq \phi \leq \pi$, and $0 \leq \theta < 2\pi$, with $\rho = \sqrt{x^2 + y^2 + z^2}$. Spherical coordinates are ideal for problems with spherical symmetry around the origin: a sphere becomes simply $\rho = \text{const}$, and a half-cone becomes $\phi = \text{const}$.



z
P
ρ
ϕ
y
θ
x

2 Vector Functions

Single-variable calculus dealt with scalar-valued functions of one variable. We now consider vector-valued functions of one variable: the input is still a single real number t, but the output is a vector. Geometrically, such functions trace out curves in space, and they provide the natural language for describing the motion of a particle, the path of a roller coaster, or the geometry of a smooth curve. Most of the chapter is devoted to extending the operations of differentiation and integration to vector functions and to studying the geometric quantities — tangent vectors, curvature, torsion — that result.

2.1 Vector functions and space curves

A vector function is a function whose input is a real parameter t and whose output is a vector. We write

$$\mathbf{r}(t) = x(t)\hat{\boldsymbol{\imath}} + y(t)\hat{\boldsymbol{\jmath}} + z(t)\hat{\boldsymbol{k}} = \langle x(t),\ y(t),\ z(t)\rangle\,, \tag{2.1}$$

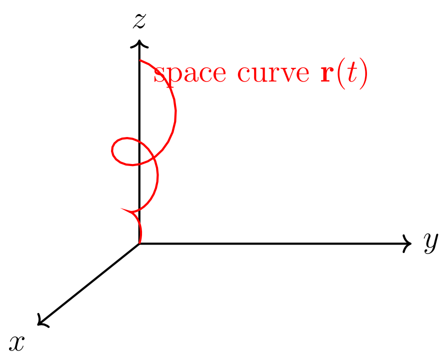
where the three components $x(t), y(t), z(t)$ are ordinary scalar functions of t, called the component functions of $\mathbf{r}$. The parameter t ranges over some common interval I on which all three component functions are defined; this interval is the domain of the vector function.

For each value of t in the domain, $\mathbf{r}(t)$ is a position vector — a vector based at the origin with its head at the point $(x(t), y(t), z(t))$. As t varies, the head of $\mathbf{r}(t)$ moves through space, tracing out a curve called a space curve. The vector function gives an oriented description of the curve, with the orientation determined by the direction of increasing t.

To sketch a space curve from a vector function, we plot the heads $(x(t), y(t), z(t))$ for many values of t in the domain and connect them in order. The resulting curve carries direction (an arrow that follows increasing t), even though we plot only the heads of the vectors and not the vectors themselves.

Example 1

For $\mathbf{r}(t) = t^2\hat{\boldsymbol{\imath}} + t\hat{\boldsymbol{\jmath}} + t^3\hat{\boldsymbol{k}}$, evaluating at $t = 2$ gives $\mathbf{r}(2) = 4\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} + 8\hat{\boldsymbol{k}}$, so the point $(4, 2, 8)$ lies on the curve. Plotting heads for many values of t produces a smooth space curve.



2.1.1 Limits and continuity

The limit of a vector function is taken componentwise. If each component function has a limit as $t \to a$, the limit of the vector function exists and is the vector of those component limits,

$$\lim_{t\to a} \mathbf{r}(t) = \left\langle \lim_{t\to a} x(t),\ \lim_{t\to a} y(t),\ \lim_{t\to a} z(t) \right\rangle, \tag{2.2}$$

provided each of the three scalar limits exists. If any single component limit fails to exist, the vector limit fails to exist as well. Continuity follows the analogous componentwise pattern: $\mathbf{r}$ is continuous at $t = a$ if each component function is continuous at a, equivalently if $\lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a)$.

2.2 Derivatives and integrals of vector functions

The derivative of a vector function is defined by the same difference-quotient limit as the derivative of a scalar function. The only change is that we are now subtracting and dividing vectors,

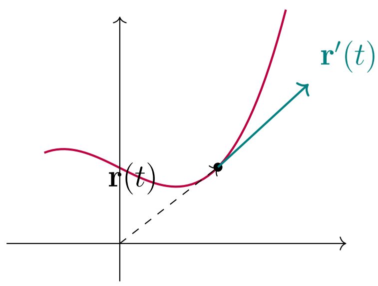
$$\mathbf{r}'(t) = \lim_{h\to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}. \tag{2.3}$$

Carrying the limit through componentwise — which we may do because limits of vector functions are componentwise — gives the practical formula

$$\mathbf{r}'(t) = \langle x'(t),\ y'(t),\ z'(t)\rangle\,. \tag{2.4}$$

To differentiate a vector function we differentiate each component separately.

Geometrically, $\mathbf{r}'(t)$ is a tangent vector to the curve traced by $\mathbf{r}$ at the point $\mathbf{r}(t)$, pointing in the direction of increasing t. If $\mathbf{r}(t)$ describes the position of a moving particle as a function of time, then $\mathbf{r}'(t)$ is the velocity vector and $\|\mathbf{r}'(t)\|$ is the speed.



2.2.1 Differentiation rules

The familiar differentiation rules from single-variable calculus all extend to vector functions. The two operations on vectors that did not exist in the scalar setting — the dot product and the cross product — get their own product rules. For differentiable vector functions $\mathbf{u}(t)$ and $\vec{v}(t)$, scalar function $f(t)$, and constant vector $\mathbf{c}$,

$$\frac{d}{dt}[\mathbf{u}+\vec{v}] = \mathbf{u}' + \vec{v}', \tag{2.5}$$

$$\frac{d}{dt}[f\mathbf{u}] = f'\mathbf{u} + f\mathbf{u}', \tag{2.6}$$

$$\frac{d}{dt}[\mathbf{u}\cdot\vec{v}] = \mathbf{u}'\cdot\vec{v} + \mathbf{u}\cdot\vec{v}', \tag{2.7}$$

$$\frac{d}{dt}[\mathbf{u}\times\vec{v}] = \mathbf{u}'\times\vec{v} + \mathbf{u}\times\vec{v}', \tag{2.8}$$

$$\frac{d}{dt}[\mathbf{u}(f(t))] = f'(t)\,\mathbf{u}'(f(t)). \tag{2.9}$$

Each rule follows from differentiating componentwise and applying the corresponding scalar rule. The order in the cross-product rule matters because the cross product is not commutative; reversing the factors negates the result.

2.2.2 Integrals

Integration is also componentwise,

$$\int_a^b \mathbf{r}(t)\,dt = \left\langle \int_a^b x(t)\,dt,\ \int_a^b y(t)\,dt,\ \int_a^b z(t)\,dt \right\rangle. \tag{2.10}$$

The result is a vector whose components are the ordinary scalar integrals of the component functions over the same interval. Antiderivatives have the same componentwise form, with a constant vector $\mathbf{c}$ playing the role of the integration constant.

2.3 Arc length

The total length of a smooth curve is computed by integrating its speed. If $\mathbf{r}(t) = \langle x(t),\ y(t),\ z(t)\rangle$ traces out a smooth curve as t varies over $[a,b]$, the arc length of the resulting curve is

$$L = \int_a^b \|\mathbf{r}'(t)\|\ dt = \int_a^b \sqrt{[x'(t)]^2+[y'(t)]^2+[z'(t)]^2}\,dt. \tag{2.11}$$

The integrand is the speed; we are summing infinitesimal speeds times infinitesimal time, which gives infinitesimal distance, then integrating to get the total distance traversed by the head of $\mathbf{r}$.

It is sometimes useful to track the running length from the starting point as a function of the parameter. The arc length function measured from $\mathbf{r}(a)$ is

$$s(t) = \int_a^t \|\mathbf{r}'(u)\| \, du, \tag{2.12}$$

and by the fundamental theorem of calculus,

$$\frac{ds}{dt} = \|\mathbf{r}'(t)\|. \tag{2.13}$$

For some applications it is convenient to switch to arc length itself as the parameter; a curve is said to be parameterised by arc length when $\|\mathbf{r}'(s)\| = 1$ for every s, in which case the parameter measures distance directly. Geometric quantities like curvature and torsion take their cleanest form when the curve is parameterised this way.

2.4 The TNB frame, curvature, and torsion

At every point on a smooth curve we can attach a triple of mutually perpendicular unit vectors — the unit tangent, the principal unit normal, and the binormal — forming a moving orthonormal frame that travels with the curve. This is the TNB (or Frenet-Serret) frame, and from it we extract two scalar quantities, curvature and torsion, that describe how the curve bends and twists in space.

2.4.1 Unit tangent, normal, and binormal

The unit tangent vector at parameter value t is the unit vector pointing in the direction of motion,

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}. \tag{2.14}$$

Because we have divided by the magnitude, $\mathbf{T}$ has length 1 at every point of the curve, regardless of how fast the parameter sweeps along.

Although $\mathbf{T}$ has constant length, its direction changes as the curve bends. The rate of change $\mathbf{T}'(t)$ measures this turning. A useful observation: because $\mathbf{T} \cdot \mathbf{T} = 1$ is constant, differentiating both sides gives $\mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 0$, so $2\mathbf{T} \cdot \mathbf{T}' = 0$, which means $\mathbf{T}' \perp \mathbf{T}$. The principal unit normal vector is the unit vector in the direction of $\mathbf{T}'$,

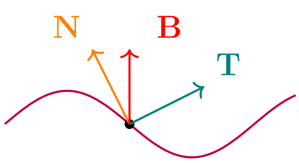
$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|}. \tag{2.15}$$

$\mathbf{N}$ points in the direction the curve is turning at each instant, into the concave side of the bend.

The third member of the frame, the binormal vector, completes a right-handed orthonormal triple,

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t). \tag{2.16}$$

$\mathbf{B}$ is automatically a unit vector (the cross product of two perpendicular unit vectors has magnitude $\|\mathbf{T}\|\,\|\mathbf{N}\|\sin(\pi/2) = 1$) and is perpendicular to both $\mathbf{T}$ and $\mathbf{N}$. The plane spanned by $\mathbf{T}$ and $\mathbf{N}$ at a given point is called the osculating plane; it is the plane in which the curve is locally turning, and $\mathbf{B}$ is normal to it.



2.4.2 Curvature

Curvature, denoted κ, measures how sharply a curve is bending at a point. The defining idea is to ask how quickly the unit tangent vector is rotating per unit arc length. Formally,

$$\kappa = \left\| \frac{d\mathbf{T}}{ds} \right\|, \tag{2.17}$$

the magnitude of the rate of change of $\mathbf{T}$ with respect to arc length. A perfectly straight line has $\mathbf{T}$ constant, hence $\kappa = 0$ identically; a tightly bent curve has $\mathbf{T}$ rotating rapidly per unit arc length, hence large κ.

The arc length parameterisation is rarely available in practice, so we typically convert to a derivative with respect to t via the chain rule. Since $ds/dt = \|\mathbf{r}'(t)\|$,

$$\kappa = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|}. \tag{2.18}$$

Both formulas are equivalent and either can be used. There is also a third, computationally convenient formula that avoids constructing $\mathbf{T}$ at all,

$$\kappa = \frac{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3}. \tag{2.19}$$

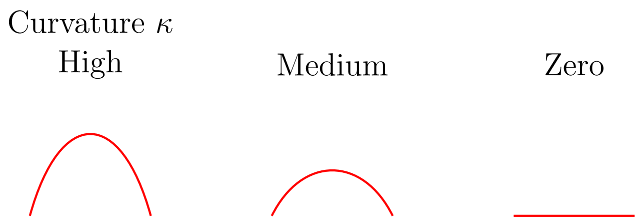
This form is the one most often used in calculation, especially for curves with polynomial or trigonometric components.

2.4.3 Torsion

Curvature alone does not tell us everything about the shape of a curve in $\mathbb{R}^3$. Two curves can have identical curvature at every point and still differ: a circle and a helix both have constant curvature, but a circle lies in a plane while a helix climbs out of that plane as it sweeps. The quantity that captures this out-of-plane twisting is the torsion τ,

$$\tau = \frac{(\mathbf{r}'(t) \times \mathbf{r}''(t)) \cdot \mathbf{r}'''(t)}{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|^2} = -\frac{d\mathbf{B}}{ds} \cdot \mathbf{N}. \tag{2.20}$$

Torsion is the rate at which the binormal vector $\mathbf{B}$ tilts away from a fixed plane as we move along the curve. A planar curve has $\mathbf{B}$ constant (perpendicular to the plane it lives in) and therefore zero torsion; a helix has $\mathbf{B}$ rotating uniformly and therefore constant nonzero torsion.

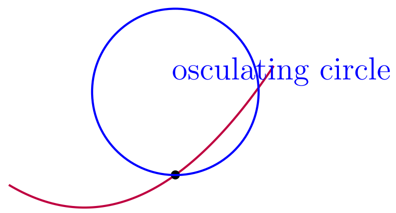


2.4.4 The osculating circle

At any point of a smooth curve where the curvature is nonzero, there is a unique circle that "best fits" the curve, in the sense of having the same tangent direction and the same curvature at the point of contact. This is the osculating circle. Its radius, called the radius of curvature, is the reciprocal of the curvature,

$$R = \frac{1}{\kappa}, \tag{2.21}$$

and its centre lies in the osculating plane (spanned by $\mathbf{T}$ and $\mathbf{N}$) at distance R from the curve in the direction of $\mathbf{N}$. Explicitly the centre is at $\mathbf{r}(t) + (1/\kappa)\mathbf{N}$. Where the curve bends sharply, the osculating circle is small; where the curve straightens out, the osculating circle becomes very large.



2.5 Motion in space

Vector functions provide the natural language for the motion of a particle. If $\mathbf{r}(t)$ denotes the position of a particle at time t, then differentiation gives velocity and acceleration,

$$\mathbf{v}(t) = \mathbf{r}'(t), \qquad \text{velocity,} \tag{2.22}$$

$$\text{speed} = \|\mathbf{v}(t)\| = \|\mathbf{r}'(t)\|, \tag{2.23}$$

$$\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t), \qquad \text{acceleration.} \tag{2.24}$$

Velocity is a vector that points in the direction of motion and has magnitude equal to the speed. Acceleration is the rate of change of velocity, and unlike in one-dimensional motion, acceleration in $\mathbb{R}^3$ has two distinct geometric roles.

Decomposing acceleration along the directions of $\mathbf{T}$ and $\mathbf{N}$ (the two members of the TNB frame in the osculating plane) gives

$$\mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N}, \tag{2.25}$$

where the tangential component a_T is the rate of change of speed and the normal component a_N is responsible for changing direction. Explicitly,

$$a_T = \frac{d}{dt}\,\|\mathbf{r}'(t)\| = \frac{\mathbf{r}'(t)\cdot\mathbf{r}''(t)}{\|\mathbf{r}'(t)\|}, \qquad a_N = \kappa\,\|\mathbf{r}'(t)\|^2 = \frac{\|\mathbf{r}'(t)\times\mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|}. \tag{2.26}$$

The tangential component changes how fast the particle moves; the normal component changes which way it points. A particle moving in a straight line at varying speed has $a_N = 0$ and $a_T \neq 0$; a particle in uniform circular motion has $a_T = 0$ and $a_N \neq 0$. In general both contribute, and the geometry of the trajectory determines their relative size.

3 Partial Derivatives

Up to this point, the functions we have considered have either had vector input and scalar output (paths) or scalar input and vector output (vector functions). We now turn to functions of several scalar inputs producing a scalar output — functions of the form $z = f(x, y)$ or $w = f(x, y, z)$. These describe surfaces in space, scalar fields, and most of the quantities of interest in physics: temperature, pressure, density, and so on. Differentiation needs to be reconsidered, because there is no longer a unique direction in which to take a derivative; the partial derivative is the natural extension, fixing all but one variable and differentiating in that direction.

3.1 Multivariable functions

A multivariable function takes several real-valued inputs. A single-variable function $y = f(x)$ has one input; a two-variable function

$$z = f(x, y) \tag{3.1}$$

has two inputs, and a three-variable function has three. Beyond three inputs, geometric visualisation becomes difficult, although the analytic machinery extends without trouble.

To graph a function we use one dimension for the output and one dimension for each input. A two-variable function therefore graphs as a surface in $\mathbb{R}^3$: the x- and y-axes carry the inputs, and the z-axis carries the output. The graph of $z = x^2 + y^2$, for instance, is a paraboloid opening upward.

The domain of a multivariable function is plotted in the input space, which has the same dimension as the number of inputs. For $f(x, y)$, the domain is a subset of $\mathbb{R}^2$. Finding the domain proceeds by applying the same rules as in single-variable calculus — denominators must be nonzero, square roots must have non-negative arguments, logarithms must have positive arguments — and intersecting the resulting constraints.

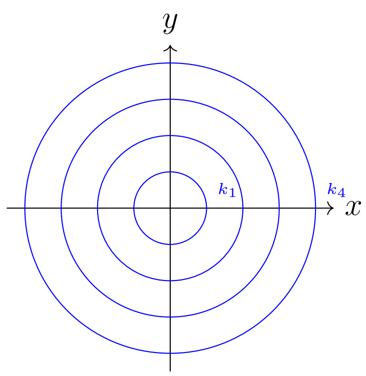
Example 2

For $f(x, y) = \dfrac{xy}{x - y}$, the only constraint comes from the denominator: $x - y$ must be nonzero. The domain is therefore $\{(x, y) \in \mathbb{R}^2 \mid x \neq y\}$, the plane with the line $y = x$ removed. The range is all of $\mathbb{R}$.

3.1.1 Level curves and contour plots

A useful tool for visualising a two-variable function without leaving the plane is to fix a value k for the output and ask which inputs produce it. The set of inputs satisfying $f(x, y) = k$ is the level curve of f at height k. Stacking many level curves at different heights produces a contour plot, the same kind of diagram one sees on a topographic map: each curve marks a constant elevation. Where the level curves are close together, the surface is steep; where they are far apart, the surface is gentle. Closed level curves around a point usually indicate a peak or a valley.

Geometrically, the level curve at height k is the projection onto the xy-plane of the cross-section of the surface $z = f(x, y)$ by the horizontal plane $z = k$. So drawing level curves is the same as slicing the surface by horizontal planes and looking from directly above.



3.2 Limits and continuity

The notion of a limit extends to multivariable functions, but with a subtle and crucial difference. In one variable, a point can only be approached from the left or the right — a one-dimensional kind of approach. In two or more variables, a point can be approached from infinitely many directions, along infinitely many paths. The limit

$$\lim_{(x,y)\to(a,b)} f(x,y) = L \tag{3.2}$$

exists if and only if $f(x,y) \to L$ along *every* path approaching (a,b). The strictness of this requirement is what makes multivariable limits qualitatively harder than single-variable ones: to show a limit exists, we must somehow control the function on every path; to show it fails to exist, it suffices to find two paths giving different limit values.

Example 3

Determine whether

$$\lim_{(x,y)\to(0,0)} \frac{xy^3 \cos x}{2x^2 + y^6}$$

exists.

Solution

We probe the limit along two different paths approaching the origin.

Along the path $x = y^3$ (substitute and let $y \to 0$),

$$\lim_{y\to 0} \frac{y^3 \cdot y^3 \cos(y^3)}{2y^6 + y^6} = \lim_{y\to 0} \frac{y^6 \cos(y^3)}{3y^6} = \frac{1}{3}.$$

Along the path $x = 0$ (substitute and let $y \to 0$),

$$\lim_{y\to 0} \frac{0}{y^6} = 0.$$

Two paths approaching the same point yield two different limits ($1/3$ and 0). Since not all paths agree, the limit does not exist.

Continuity carries over from the single-variable definition: f is continuous at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b). \tag{3.3}$$

Polynomials in x and y are continuous everywhere on $\mathbb{R}^2$, and rational functions (ratios of polynomials) are continuous wherever the denominator does not vanish. Sums, products, quotients, and compositions of continuous functions are continuous on the appropriate domains, mirroring the situation for single-variable functions.

3.3 Partial derivatives

Differentiating a multivariable function means asking how the output changes as one input changes, while the other inputs are held fixed. The partial derivative of $f(x, y)$ with respect to x at the point (x, y) is

$$f_x(x, y) = \frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(x + h, y) - f(x, y)}{h}. \tag{3.4}$$

Operationally we treat y as a constant and apply ordinary single-variable differentiation rules to x. The partial derivative with respect to y is defined symmetrically.

Geometrically, $f_x(a, b)$ is the slope of the tangent line to the graph of $z = f(x, y)$ at the point $(a, b, f(a, b))$, taken along the curve where y is held constant at b. Imagine slicing the surface with the vertical plane $y = b$; the resulting cross-section is a curve in the xz-plane, and $f_x(a, b)$ is the slope of that curve at $x = a$. Similarly $f_y(a, b)$ is the slope along the curve where x is held fixed at a.

3.3.1 Higher-order partial derivatives

Each of the partial derivatives f_x and f_y is itself a function of x and y, so we can differentiate again. The four second-order partial derivatives are

$$f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{yy} = \frac{\partial^2 f}{\partial y^2}, \quad f_{xy} = \frac{\partial^2 f}{\partial y \partial x}, \quad f_{yx} = \frac{\partial^2 f}{\partial x \partial y}. \tag{3.5}$$

The notation f_{xy} means: differentiate with respect to x first, then with respect to y. The order of subscripts in f_{xy} matches the order in which we differentiate, but the order in $\partial^2 f/(\partial y\, \partial x)$ reverses (the rightmost differentiation operator is applied first when reading the symbol left-to-right). The two conventions are unfortunate, but each is widespread.

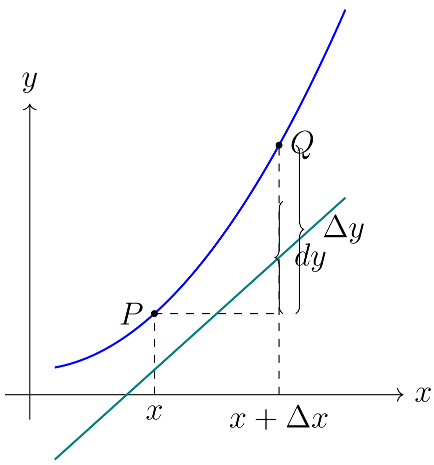
The mixed partials f_{xy} and f_{yx} involve differentiating in different orders, but for sufficiently smooth functions they always agree.

Theorem 3.1 (Clairaut's theorem)**.** If f_{xy} and f_{yx} are both continuous on an open set, then $f_{xy} = f_{yx}$ throughout that set.

In practice, this means we can usually differentiate in whichever order is computationally easier. Functions appearing in this course are smooth enough that Clairaut's theorem applies.

3.4 Differentials and linear approximation

Recall that for a single-variable function $y = f(x)$, the differential dy is the change in y along the tangent line as x changes by dx; it is the linear approximation to the actual change Δy in y along the curve.



The increment Δy is the change in height from x to $x + \Delta x$ measured along the curve, and the differential dy is the corresponding change measured along the tangent line at x. As $\Delta x \to 0$, the tangent line and the curve agree to higher and higher order, and

$\Delta y \to dy$. Since the slope of the tangent line is $f'(x) = dy/dx$, we have

$$dy = f'(x)\,dx. \tag{3.6}$$

For a two-variable function $z = f(x, y)$, the analogous object is the total differential, which captures the change in z along the tangent plane as both inputs change,

$$dz = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = f_x\,dx + f_y\,dy. \tag{3.7}$$

The interpretation: when x changes by dx and y changes by dy, the linear approximation to the change in z is $f_x\,dx + f_y\,dy$ — one contribution from each input, weighted by the corresponding partial derivative.

We say f is differentiable at (a, b) if its partial derivatives f_x, f_y exist and are continuous near (a, b). In that case dz provides a good linear approximation to the actual increment Δz, in the sense that the error vanishes faster than the size of the input change. Differentiability is a stronger condition than the mere existence of partial derivatives, but the gap matters mostly in pathological examples.

3.5 The chain rule

The chain rule extends to multivariable functions, and the form it takes depends on the dependency structure of the variables. Consider first the case of a function $z = f(x, y)$ where x and y are themselves both functions of a single parameter t. A change in t produces changes in both x and y, each of which propagates to a change in z through the partial derivatives. Adding the two contributions gives

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \tag{3.8}$$

The left-hand side is a total derivative because, after substituting $x(t)$ and $y(t)$, z is a function of the single variable t. The partial derivatives on the right are those of the original function $f(x, y)$, evaluated at the current point $(x(t), y(t))$.

Example 4

Given $z = x^2 - y^2$ with $x = t^2 + 1$ and $y = t^3 + t$, compute dz/dt.

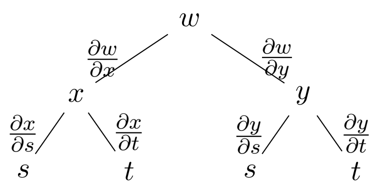
Solution

The partial derivatives are $\partial z/\partial x = 2x$ and $\partial z/\partial y = -2y$. The ordinary derivatives are $dx/dt = 2t$ and $dy/dt = 3t^2 + 1$. Assembling these into the chain rule,

$$\frac{dz}{dt} = (2x)(2t) + (-2y)(3t^2 + 1) = 4t(t^2 + 1) - 2(t^3 + t)(3t^2 + 1).$$

We could simplify further but the form above shows the chain-rule structure clearly.

When the intermediate variables themselves depend on more than one parameter, the situation is more elaborate but the rule is the same: each output partial derivative is a sum over all paths from the parameter to the output in the dependency graph. A tree diagram is helpful for keeping track of the structure. Suppose $w = f(x, y)$ with $x = x(s, t)$ and $y = y(s, t)$.



Each branch of the tree is labelled by a partial derivative. To compute $\partial w/\partial s$, multiply the partial derivatives along each path from w down to s, then sum over all such paths. There are two paths in the tree above (one through x, one through y), so

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s}, \tag{3.9}$$

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t}. \tag{3.10}$$

The same scheme generalises to any number of intermediate and independent variables.

3.6 Directional derivatives and the gradient

The partial derivatives f_x and f_y measure the slope of f in the directions of the two coordinate axes. Many problems demand the slope in some other direction — for example, "how fast does temperature change as we walk northeast?". The directional derivative gives this slope in any direction.

3.6.1 Directional derivatives

For a function $f(x, y)$ and a unit vector $\hat{\boldsymbol{u}} = \langle u_1, u_2 \rangle$, the directional derivative of f in the direction $\hat{\boldsymbol{u}}$ is

$$D_{\hat{\boldsymbol{u}}} f(x, y) = \frac{\partial f}{\partial x} u_1 + \frac{\partial f}{\partial y} u_2. \tag{3.11}$$

This is the slope of the tangent line to the surface $z = f(x, y)$ at the point above (x, y), taken along the vertical plane that contains $\hat{\boldsymbol{u}}$.

It is essential that $\hat{\boldsymbol{u}}$ be a unit vector. If we used a non-unit vector instead, the result would be scaled by its magnitude and would no longer represent a true rate of change per unit distance. For any nonzero direction vector, divide by its magnitude before substituting.

3.6.2 The gradient

The components of the directional-derivative formula are exactly the dot product of two vectors. The first vector is the gradient of f,

$$\nabla f = \langle f_x, f_y \rangle = \frac{\partial f}{\partial x} \hat{\boldsymbol{\imath}} + \frac{\partial f}{\partial y} \hat{\boldsymbol{\jmath}}, \tag{3.12}$$

and analogously $\nabla f = \langle f_x, f_y, f_z \rangle$ in three variables. The directional derivative is then simply

$$D_{\hat{\boldsymbol{u}}} f = \nabla f \cdot \hat{\boldsymbol{u}}. \tag{3.13}$$

Using the geometric formula for the dot product, since $\hat{\boldsymbol{u}}$ is a unit vector,

$$D_{\hat{\boldsymbol{u}}} f = \|\nabla f\| \cos\theta, \tag{3.14}$$

where θ is the angle between ∇f and $\hat{\boldsymbol{u}}$. This identity unlocks three important facts about the gradient.

First, the directional derivative is maximised when $\hat{\boldsymbol{u}}$ points in the direction of ∇f, in which case $\cos\theta = 1$ and the maximum value is $\|\nabla f\|$. The gradient therefore points in the direction of steepest ascent, and its magnitude is the slope along that direction.

Second, the directional derivative is minimised when $\hat{\boldsymbol{u}}$ points opposite to ∇f, with minimum value $-\|\nabla f\|$. This is the direction of steepest descent — the direction water would flow on the surface.

Third, the directional derivative is zero when $\hat{\boldsymbol{u}}$ is perpendicular to ∇f. Since the directional derivative is the rate of change of f in the direction $\hat{\boldsymbol{u}}$, a zero directional derivative means f does not change instantaneously in that direction; this is exactly the direction tangent to the level curve through (x, y). Therefore ∇f is everywhere perpendicular to the level curves of f.

Example 5

Find the slope of the tangent line to $f(x, y) = x^3 - 2x^2 + y^3$ at $(1, 2)$ in the direction making angle $\theta = \pi/6$ with the positive x-axis.

Solution

The unit direction vector at angle $\pi/6$ is $\hat{\boldsymbol{u}} = \langle \cos(\pi/6), \sin(\pi/6) \rangle = \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle$. Computing the partials, $f_x = 3x^2 - 4x$ and $f_y = 3y^2$. Evaluating at $(1, 2)$,

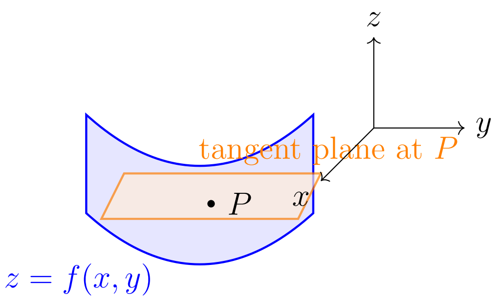
$$f_x(1, 2) = 3 - 4 = -1, \qquad f_y(1, 2) = 12.$$

The directional derivative is therefore

$$D_{\hat{\boldsymbol{u}}} f(1, 2) = (-1) \cdot \tfrac{\sqrt{3}}{2} + 12 \cdot \tfrac{1}{2} = 6 - \tfrac{\sqrt{3}}{2}.$$

3.7 Tangent planes and normal lines

A surface in $\mathbb{R}^3$ has a tangent plane at each smooth point, just as a curve in $\mathbb{R}^2$ has a tangent line at each smooth point. The tangent plane is the plane that best approximates the surface near the point of contact, and like the tangent line, it is determined by the local first-order behaviour of the function.



For a surface given explicitly by $z = f(x, y)$, we obtain the tangent plane by viewing the graph as the level surface $F(x, y, z) = 0$ where $F = f(x, y) - z$. The gradient of F in $\mathbb{R}^3$ is normal to this level surface, and a quick calculation gives

$$\nabla F = \langle f_x, f_y, -1 \rangle . \tag{3.15}$$

Using this normal vector and the point (x_0, y_0, z_0) on the surface, the equation of the tangent plane at this point is

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) - (z - z_0) = 0. \tag{3.16}$$

Solving for z gives the more familiar form,

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \tag{3.17}$$

This is the multivariable analogue of the tangent line $y - y_0 = f'(x_0)(x - x_0)$ from single-variable calculus. The normal line at (x_0, y_0, z_0) is the line through this point in the direction ∇F,

$$\frac{x - x_0}{f_x} = \frac{y - y_0}{f_y} = \frac{z - z_0}{-1}. \tag{3.18}$$

For an implicitly defined surface $F(x, y, z) = k$, the gradient ∇F is again normal to the surface, and the tangent plane at (x_0, y_0, z_0) is

$$F_x(x - x_0) + F_y(y - y_0) + F_z(z - z_0) = 0, \tag{3.19}$$

where the partials are evaluated at the point. This implicit form is more general; the explicit form $z = f(x, y)$ is a special case.

3.8 Extrema of multivariable functions

We now develop the multivariable analogue of the first and second derivative tests for finding maxima and minima. The first-derivative test generalises directly: at a relative extremum, the gradient must vanish. The second-derivative test is more subtle, because there are now several second-order partials to compare, and the geometry includes a new possibility absent in single-variable calculus — the saddle point, where a function increases in some directions and decreases in others.

A point (x_0, y_0) is a critical point of f if both partial derivatives vanish,

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0, \tag{3.20}$$

or if one of the partials fails to exist. Relative extrema (and saddle points) can occur only at critical points. As in single-variable calculus, this is necessary but not sufficient: not every critical point is an extremum.

Geometrically, a critical point is a point where the tangent plane is horizontal: the surface neither rises nor falls in any of the coordinate directions instantaneously. From there the surface might curve up in all directions (a relative minimum), curve down in all directions (a relative maximum), or curve up in some directions and down in others (a saddle point).

3.8.1 The second-derivative test

To classify a critical point we look at the second-order partials. The discriminant or Hessian determinant is

$$D(x, y) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2, \tag{3.21}$$

using Clairaut's theorem to set $f_{xy} = f_{yx}$. At a critical point (x_0, y_0), the test classifies the point according to the signs of D and f_{xx}:

1. $D(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) > 0$: relative minimum (equivalently $f_{yy} > 0$, since when $D > 0$ the two diagonal entries must have the same sign).

2. $D(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) < 0$: relative maximum.

3. $D(x_0, y_0) < 0$: saddle point.

4. $D(x_0, y_0) = 0$: the test is inconclusive and we must use other means.

The conceptual content: $D > 0$ means the surface bends in the same way in both diagonal directions, so we have a definite extremum; the sign of f_{xx} then tells us whether the bend is upward (minimum) or downward (maximum). $D < 0$ means the surface bends one way in one direction and the other way in the perpendicular direction, producing a saddle.

The procedure for finding all relative extrema of a function $z = f(x, y)$ on an open region therefore reduces to four steps. Compute f_x and f_y and find all critical points by solving $f_x = f_y = 0$. For each critical point, compute the second-order partials and the discriminant D. Apply the test above to classify each point.

3.8.2 Absolute extrema on closed regions

On a closed and bounded region, a continuous function is guaranteed to attain both an absolute maximum and an absolute minimum. These extrema occur either at critical points in the interior of the region or somewhere on the boundary.

The procedure mirrors single-variable optimization on a closed interval: find candidates from the interior, find candidates from the boundary, evaluate the function at each candidate, and compare. Specifically, find interior critical points by setting $\nabla f = \mathbf{0}$ and discarding any solutions that lie outside the region. For each piece of the boundary, parameterise it by a single variable and reduce f to a one-variable function on that piece, then find the extrema of that single-variable function on the parameter interval. Finally evaluate f at every candidate point and pick the largest and smallest values.

Example 6

Find the absolute maximum and minimum of $f(x, y) = x^2 + xy + y^2$ on the rectangle $\{(x, y) \mid -2 \leq x \leq 2,\ -1 \leq y \leq 1\}$.

Solution

Interior critical points. The partials are $f_x = 2x + y$ and $f_y = x + 2y$. Setting both to zero gives the system $2x + y = 0$ and $x + 2y = 0$, whose only solution is $x = y = 0$. The point $(0, 0)$ lies inside the rectangle, so it is a candidate. We evaluate $f(0, 0) = 0$.

Boundary. The rectangle has four edges; we examine each in turn.

On the edge $x = -2$, f reduces to a function of y alone, $f(-2, y) = 4 - 2y + y^2$. Differentiating, $f'(y) = -2 + 2y$, which vanishes at $y = 1$. Combined with the endpoints $y = \pm 1$, the candidates on this edge are at $y = -1$ and $y = 1$, giving $f(-2, -1) = 4 + 2 + 1 = 7$ and $f(-2, 1) = 4 - 2 + 1 = 3$.

On the edge $x = 2$, by the symmetry $f(2, y) = f(-2, -y)$ we get $f(2, 1) = 7$ and $f(2, -1) = 3$.

On the edge $y = -1$, $f(x, -1) = x^2 - x + 1$, with $f'(x) = 2x - 1 = 0$ at $x = 1/2$. The values at $x = 1/2$ and the endpoints $x = \pm 2$ give $f(1/2, -1) = 3/4$ (the new candidate) and the corner values already computed.

On the edge $y = 1$, $f(x, 1) = x^2 + x + 1$, with $f'(x) = 2x + 1 = 0$ at $x = -1/2$. This gives $f(-1/2, 1) = 3/4$, plus corner values.

Comparison. Among all candidate values $\{0, 3/4, 3, 7\}$, the absolute minimum is $f(0, 0) = 0$ and the absolute maximum is $f(-2, -1) = f(2, 1) = 7$.

3.9 Lagrange multipliers

The methods above find unconstrained extrema. Many real problems have a constraint: optimise the volume of a box subject to a fixed surface area, minimise the cost of a route subject to a fixed length, or maximise utility subject to a budget. The method of Lagrange multipliers handles such constraints elegantly.

Suppose we want to find the extrema of $f(x, y, z)$ subject to a constraint $g(x, y, z) = k$. The constraint defines a surface, and we are looking for the points on this surface where

f achieves its maximum or minimum. At any such constrained extremum, walking along the constraint surface produces no instantaneous change in f to first order, which means ∇f has no component tangent to the constraint surface; equivalently, ∇f is perpendicular to the constraint surface. But ∇g is itself perpendicular to the constraint surface (since the constraint surface is a level surface of g). Therefore ∇f and ∇g are parallel, meaning there exists a scalar λ — the Lagrange multiplier — such that

$$\nabla f = \lambda \nabla g, \qquad g(x, y, z) = k. \tag{3.22}$$

The first equation gives three scalar equations (one per component), and the constraint gives a fourth, producing a system of four equations in four unknowns (x, y, z, λ). Solve, then evaluate f at each solution to identify maxima and minima.

Example 7

Find the dimensions of an open rectangular box of volume 108 in^3 with minimum surface area.

Solution

Let the dimensions be x, y, z (with z the height of the open top). The constraint is $V = xyz = 108$. The surface area of an open box (no top) is $A = xy + 2yz + 2xz$. Rather than running the full Lagrange system, we substitute $z = 108/(xy)$ from the constraint into A, reducing to an unconstrained problem in two variables,

$$A(x, y) = xy + 2y \cdot \frac{108}{xy} + 2x \cdot \frac{108}{xy} = xy + \frac{216}{x} + \frac{216}{y}.$$

The first-order conditions are $A_x = y - 216/x^2 = 0$ and $A_y = x - 216/y^2 = 0$. From the first, $y = 216/x^2$; substituting into the second,

$$x - \frac{216}{(216/x^2)^2} = x - \frac{x^4}{216} = 0 \implies x(1 - x^3/216) = 0 \implies x^3 = 216 \implies x = 6.$$

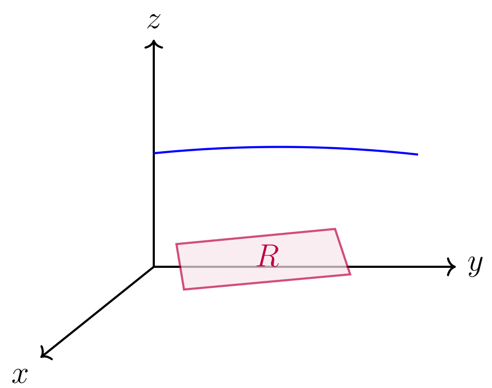
Hence $y = 216/36 = 6$ and $z = 108/36 = 3$. The minimum surface area is $A = 36 + 36 + 36 = 108$ in^2.

4 Multiple Integrals

Single-variable integration computes the area under a curve. Multiple integration extends this idea to higher dimensions: a double integral computes the volume under a surface, a triple integral computes a measurement of a four-dimensional region (or, more usefully, the mass of a three-dimensional solid with a given density). All of these constructions descend from a single common idea: partition the domain into infinitesimally small pieces, evaluate the integrand on each piece, multiply by the piece's volume, and sum. Different coordinate systems lead to different volume elements, but the underlying construction is the same.

4.1 Double integrals over rectangles

Consider a non-negative continuous function $z = f(x, y)$ on a region R in the xy-plane. The graph of f is a surface above R, and the volume of the solid lying above R and below the surface is what the double integral will compute.



We construct the integral as a Riemann sum. Cut R into mn subrectangles using m partitions along the x-direction and n along the y-direction. In the (i, j)th subrectangle pick a sample point (x_i^*, y_j^*). The volume of the rectangular column with base equal to that subrectangle and height equal to $f(x_i^*, y_j^*)$ is approximately

$$f(x_i^*, y_j^*)\,\Delta A_{ij},$$

where ΔA_{ij} is the area of the subrectangle. The total volume is approximated by summing over all subrectangles,

$$V \approx \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i^*, y_j^*)\, \Delta A_{ij}. \tag{4.1}$$

As we refine the partition by letting $m, n \to \infty$ (and consequently the size of each subrectangle shrinks to zero), this Riemann sum converges to the double integral

$$\iint_R f(x,y)\, dA = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i^*, y_j^*)\, \Delta A_{ij}. \tag{4.2}$$

4.1.1 Properties of the double integral

From the limit definition, the double integral inherits the standard linearity and monotonicity properties of integration:

1. $\iint_R cf\, dA = c \iint_R f\, dA$ for any constant c.
2. $\iint_R (f+g)\, dA = \iint_R f\, dA + \iint_R g\, dA$.
3. If $R = R_1 \cup R_2$ where R_1 and R_2 overlap only on a set of zero area, then $\iint_R f\, dA = \iint_{R_1} f\, dA + \iint_{R_2} f\, dA$.
4. If $f \geq g$ on R, then $\iint_R f\, dA \geq \iint_R g\, dA$.
5. $\iint_R 1\, dA = \text{area}(R)$, since the volume of a column of height 1 over R is just the area of the base.

These properties follow directly from the corresponding facts about Riemann sums and survive the passage to the limit.

4.2 Fubini's theorem and iterated integrals

The Riemann-sum definition is conceptually clean but useless for actual calculation. Fubini's theorem reduces a double integral to two ordinary single-variable integrations, performed one after the other. Over a rectangle $R = [a, b] \times [c, d]$, if f is continuous,

$$\iint_R f(x, y)\, dA = \int_a^b \int_c^d f(x, y)\, dy\, dx = \int_c^d \int_a^b f(x, y)\, dx\, dy. \tag{4.3}$$

Either order of integration produces the same answer, so we are free to choose the order that simplifies the calculation. The geometric content: integrating with respect to y first gives the area of a vertical cross-section at fixed x, and integrating that area over x stacks the cross-sections to produce the total volume.

4.2.1 General regions

For non-rectangular regions, the inner limits depend on the outer variable. We classify regions by which variable is allowed to vary freely.

A region is of Type I if it can be described by

$$R = \{(x, y) \mid a \leq x \leq b,\ g_1(x) \leq y \leq g_2(x)\},$$

with x ranging over a fixed interval and y trapped between two functions of x. A region is of Type II if instead

$$R = \{(x, y) \mid c \leq y \leq d,\ h_1(y) \leq x \leq h_2(y)\},$$

with the roles reversed. Many regions are simultaneously of both types, and we are free to choose whichever description leads to easier integrals. Fubini's theorem on these general regions takes the form

$$\iint_R f(x, y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\, dx = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, dx\, dy. \tag{4.4}$$

Example 8

Evaluate $\iint_R \frac{x}{1+xy}\, dA$ over the unit square $R = [0,1] \times [0,1]$.

Solution

Set up the iterated integral with y as the inner variable,

$$\int_0^1 \int_0^1 \frac{x}{1+xy}\, dy\, dx.$$

For the inner integral, substitute $u = 1 + xy$ with $du = x\, dy$. When $y = 0$, $u = 1$; when $y = 1$, $u = 1 + x$. Then

$$\int_0^1 \frac{x}{1+xy}\, dy = \int_1^{1+x} \frac{x}{u} \cdot \frac{du}{x} = \int_1^{1+x} \frac{1}{u}\, du = \ln(1+x).$$

The outer integral is then $\int_0^1 \ln(1+x)\, dx$, which can be evaluated by integration by parts ($u = \ln(1+x)$, $dv = dx$),

$$\int_0^1 \ln(1+x)\, dx = [(1+x)\ln(1+x) - (1+x)]_0^1 = (2\ln 2 - 2) - (0 - 1) = 2\ln 2 - 1.$$

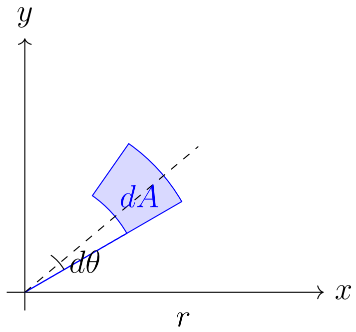
4.2.2 Reversing the order of integration

Sometimes an iterated integral is intractable in one order but straightforward in the other. To reverse the order, sketch the region described by the bounds of the original iterated integral and re-express the same region with the variables swapped. The integrand stays the same; only the bounds and the order of dx and dy change. This is purely an algebraic-geometric exercise on the region.

4.3 Double integrals in polar coordinates

For regions with circular symmetry — disks, annuli, sectors — Cartesian coordinates are awkward and polar coordinates are natural. The substitution $x = r\cos\theta$ and $y = r\sin\theta$ converts the integrand, but the area element also changes. In Cartesian coordinates we have $dA = dx\,dy$, but in polar coordinates the corresponding infinitesimal area element is

$$dA = r\,dr\,d\theta. \tag{4.5}$$



The geometric reason for the extra factor of r: an infinitesimal polar rectangle with radial extent dr and angular extent $d\theta$ at radius r has dimensions dr in the radial direction and $r\,d\theta$ in the angular direction (because the arc length subtended by an angle $d\theta$ at radius r is $r\,d\theta$). Multiplying gives the area $r\,dr\,d\theta$. Without the factor of r, regions far from the origin would be undercounted; the r factor compensates for the larger area swept out at larger radii.

The change-of-variable formula is therefore

$$\iint_R f(x,y)\,dA = \iint_R f(r\cos\theta,\ r\sin\theta)\,r\,dr\,d\theta. \tag{4.6}$$

Polar coordinates are the right choice when the region of integration is a disk, annulus, or sector, or when the integrand simplifies under the substitution $x^2 + y^2 = r^2$.

Example 9

Find the volume below $z = y^2/(x^2+y^2)$ above the xy-plane and between the cylinders $x^2 + y^2 = 1$ and $x^2 + y^2 = 2$.

Solution

The region of integration is the annulus $1 \leq r^2 \leq 2$, i.e. $1 \leq r \leq \sqrt{2}$, with θ ranging over a full revolution.

In polar coordinates the integrand simplifies dramatically: $y^2 = r^2 \sin^2\theta$ and $x^2 + y^2 = r^2$, so $y^2/(x^2+y^2) = \sin^2\theta$. Including the factor of r from dA,

$$\begin{aligned} V &= \int_0^{2\pi} \int_1^{\sqrt{2}} \sin^2\theta \cdot r\,dr\,d\theta \\ &= \int_0^{2\pi} \sin^2\theta \left[\frac{r^2}{2}\right]_1^{\sqrt{2}} d\theta = \int_0^{2\pi} \sin^2\theta \cdot \frac{1}{2}\,d\theta. \end{aligned}$$

Using the identity $\sin^2\theta = (1-\cos 2\theta)/2$,

$$V = \int_0^{2\pi} \frac{1-\cos 2\theta}{4}\,d\theta = \frac{1}{4}\left[\theta - \frac{\sin 2\theta}{2}\right]_0^{2\pi} = \frac{2\pi}{4} = \frac{\pi}{2}.$$

4.4 Applications to mass and centre of mass

Suppose a thin plate occupies a region R in the plane, with mass density $\rho(x,y)$ varying from point to point. The total mass is the integral of density over the region,

$$m = \iint_R \rho(x,y)\,dA. \tag{4.7}$$

This is a direct generalisation of the single-variable formula for total mass given a linear mass density.

The moments of the plate about the coordinate axes are integrals of distance times density,

$$M_y = \iint_R x\,\rho(x,y)\,dA \qquad \text{(moment about the } y\text{-axis)}, \tag{4.8}$$

$$M_x = \iint_R y\,\rho(x,y)\,dA \qquad \text{(moment about the } x\text{-axis)}. \tag{4.9}$$

The naming convention can be confusing: the moment "about the y-axis" weighs each mass element by its x-coordinate (its distance from the y-axis). The centre of mass of the plate is the point $(\bar{x}, \bar{y})$ where the moments equal those of a point mass m located at the centre,

$$\bar{x} = \frac{M_y}{m}, \qquad \bar{y} = \frac{M_x}{m}. \tag{4.10}$$

For a plate of uniform density, the centre of mass coincides with the geometric centroid of the region.

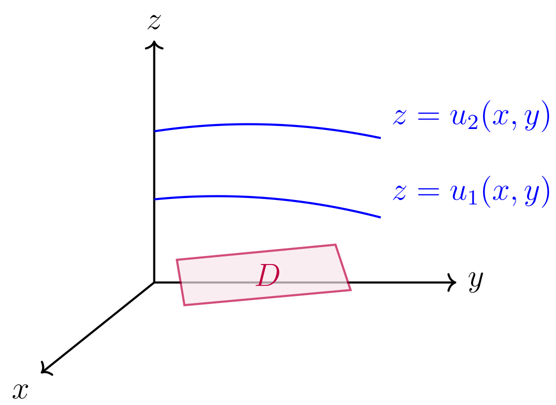
4.5 Triple integrals

Triple integrals integrate over three-dimensional regions. The construction parallels the double-integral case: partition the region into small rectangular boxes, evaluate the integrand on each box, multiply by the box's volume ΔV_{ijk}, and take the limit as the partition refines. The result is written

$$\iiint_T f(x, y, z)\, dV, \tag{4.11}$$

with the volume element $dV = dx\, dy\, dz$ in Cartesian coordinates.

The geometric meaning depends on the integrand. Where the double integral computes the volume under a surface, the triple integral computes a measurement of a four-dimensional region (which we cannot directly visualise) or, for a physically meaningful integrand, the mass of the solid: if ρ is a mass-density function, then $\iiint_T \rho\, dV$ is the total mass of the solid T.



Triple integrals are evaluated as iterated integrals. For a region of the form

$$T = \{(x, y, z) \mid (x, y) \in D,\ u_1(x, y) \le z \le u_2(x, y)\},$$

the inner z-integral evaluates the integrand on a vertical line above each point of the planar region D, and the outer double integral over D assembles these into the total,

$$\iiint_T f(x, y, z)\, dV = \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)\, dz \right] dA. \tag{4.12}$$

Example 10

Compute $\iiint_T z\, dV$ where T is bounded by $x^2 + z^2 = 4$, $x = 2y$, $y = 0$, $z = 0$, and $x \ge 0$.

Solution

The constraint $z \geq 0$ together with $x^2 + z^2 = 4$ gives $0 \leq z \leq \sqrt{4 - x^2}$ as the z-range. At $z = 0$, the boundary is $x^2 = 4$, so $x \leq 2$. The constraints $y = 0$, $x = 2y$, and $x \geq 0$ confine the planar region to $0 \leq y \leq x/2$ for $0 \leq x \leq 2$.

The triple integral becomes

$$\int_0^2 \int_0^{x/2} \int_0^{\sqrt{4-x^2}} z \, dz \, dy \, dx.$$

The innermost integral is straightforward,

$$\int_0^{\sqrt{4-x^2}} z \, dz = \left[\frac{z^2}{2} \right]_0^{\sqrt{4-x^2}} = \frac{4 - x^2}{2}.$$

Substituting,

$$\int_0^2 \int_0^{x/2} \frac{4 - x^2}{2} \, dy \, dx = \int_0^2 \frac{4 - x^2}{2} \cdot \frac{x}{2} \, dx = \int_0^2 \frac{4x - x^3}{4} \, dx.$$

Evaluating the final integral,

$$\frac{1}{4} \left[2x^2 - \frac{x^4}{4} \right]_0^2 = \frac{1}{4}(8 - 4) = 1.$$

4.6 Triple integrals in cylindrical coordinates

For solids with axial symmetry around the z-axis — cylinders, cones, paraboloids — cylindrical coordinates are usually easier than Cartesian. The volume element in cylindrical coordinates is

$$dV = r \, dz \, dr \, d\theta, \tag{4.13}$$

with the same factor of r that appeared in the planar polar transformation, for the same geometric reason: an infinitesimal polar rectangle has area $r \, dr \, d\theta$, and stacking these vertically with thickness dz gives a volume element of $r \, dz \, dr \, d\theta$.

Cylindrical coordinates are the right choice when the bounding surfaces are cylinders, planes parallel to the z-axis, paraboloids, or cones whose axis is the z-axis.

Example 11

Set up the volume of the region bounded by the sphere $x^2 + y^2 + z^2 = 9$ and the paraboloid $8z = x^2 + y^2$.

Solution

The two surfaces meet where $z^2 + 8z = 9$, i.e. $(z-1)(z+9) = 0$. Taking the relevant root $z = 1$, this corresponds to $x^2 + y^2 = 8$.

In cylindrical coordinates the paraboloid becomes $z = r^2/8$ and the sphere $z = \sqrt{9 - r^2}$. The volume is then

$$V = \int_0^{2\pi} \int_0^{\sqrt{8}} \int_{r^2/8}^{\sqrt{9-r^2}} r\,dz\,dr\,d\theta = \int_0^{2\pi} \int_0^{\sqrt{8}} r\left(\sqrt{9-r^2} - \frac{r^2}{8}\right) dr\,d\theta.$$

The remaining integral can be evaluated explicitly but the setup illustrates the cylindrical-coordinate procedure clearly.

4.7 Triple integrals in spherical coordinates

For solids with central symmetry around a single point — balls, cones with apex at the origin, and surfaces bounded by both — spherical coordinates are the natural choice. With $x = \rho \sin\phi \cos\theta$, $y = \rho \sin\phi \sin\theta$, $z = \rho\cos\phi$, the volume element is

$$dV = \rho^2 \sin\phi \, d\rho \, d\phi \, d\theta. \tag{4.14}$$

The factor $\rho^2 \sin\phi$ has a geometric explanation: an infinitesimal spherical box has radial extent $d\rho$, polar-angle extent $\rho\,d\phi$, and azimuthal extent $\rho \sin\phi\,d\theta$ (the radius of the small circle at polar angle ϕ is $\rho\sin\phi$). Multiplying these three lengths gives $\rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$.

Spherical coordinates work best when the region is bounded by spheres centred at the origin or by cones with apex at the origin.

Example 12

Evaluate $\iiint_T xz\,dV$ where T is the region bounded above by the sphere $x^2+y^2+z^2=4$ and below by the cone $z=\sqrt{x^2+y^2}$.

Solution

The sphere is $\rho=2$ in spherical coordinates. The cone $z=\sqrt{x^2+y^2}$ corresponds to $\phi=\pi/4$ (the angle where the polar angle from the z-axis equals $\pi/4$). The integrand xz in spherical coordinates is $(\rho\sin\phi\cos\theta)(\rho\cos\phi)=\rho^2\sin\phi\cos\phi\cos\theta$. Including the volume element,

$$\begin{aligned}\iiint_T xz\,dV &= \int_0^{2\pi}\int_0^{\pi/4}\int_0^2 \rho^2\sin\phi\cos\phi\cos\theta\cdot\rho^2\sin\phi\,d\rho\,d\phi\,d\theta\\ &= \int_0^{2\pi}\int_0^{\pi/4}\int_0^2 \rho^4\sin^2\phi\cos\phi\cos\theta\,d\rho\,d\phi\,d\theta.\end{aligned}$$

Because the integrand factors as a product of separate functions of ρ, ϕ, and θ, the triple integral factors as a product of three single integrals,

$$\left[\int_0^{2\pi}\cos\theta\,d\theta\right]\cdot\left[\int_0^{\pi/4}\sin^2\phi\cos\phi\,d\phi\right]\cdot\left[\int_0^2\rho^4\,d\rho\right]=0,$$

because $\int_0^{2\pi}\cos\theta\,d\theta=0$. The full integral therefore vanishes regardless of the values of the other two factors.

4.8 Change of variables and the Jacobian

Polar, cylindrical, and spherical coordinates are special cases of a more general procedure: change of variables. Given any smooth invertible transformation $T:(u,v)\mapsto(x,y)$ with $x=g(u,v)$ and $y=h(u,v)$, we can pull back an integral over the (x,y)-region to an integral over the (u,v)-region. The factor that appears under the change of variables is the Jacobian determinant,

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}. \tag{4.15}$$

The Jacobian measures, locally, how much the transformation stretches or shrinks area. Where $|J|$ is large, a small (u, v)-rectangle maps to a large (x, y)-region; where $|J|$ is small, the same (u, v)-rectangle maps to a small (x, y)-region. The change-of-variables formula in $\mathbb{R}^2$ is

$$\iint_R f(x,y)\,dA = \iint_S f(g(u,v), h(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv, \tag{4.16}$$

where $R = T(S)$ is the image of the source region S. The absolute value of the Jacobian appears because area is non-negative; the determinant itself can be negative if the transformation reverses orientation. In $\mathbb{R}^3$ the analogous formula is

$$\iiint_R f(x,y,z)\,dV = \iiint_S f \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| du\,dv\,dw. \tag{4.17}$$

Remark. The factors that appeared earlier in polar, cylindrical, and spherical coordinates are special cases of Jacobian determinants. For polar coordinates, $|\partial(x, y)/\partial(r, \theta)| = r$. For cylindrical coordinates, $|\partial(x, y, z)/\partial(r, \theta, z)| = r$. For spherical coordinates, $|\partial(x, y, z)/\partial(\rho, \phi, \theta)| = \rho^2 \sin\phi$. These can be derived directly from the determinant formula.

Example 13

Transform the region R bounded by the lines $y = -2x$, $y = -2x+10$, $y = x/2-15/2$, $y = x/2$ using the change of variables $u = x + 2y$, $v = y - 2x$, and compute the Jacobian.

Solution

Inverting the transformation, we solve for x and y in terms of u and v. From $u = x + 2y$ and $v = y - 2x$, multiplying the first by 1 and the second by -2 and adding gives $u - 2v = 5x$, so $x = (u - 2v)/5$. From the first equation $y = (u - x)/2$; substituting and simplifying gives $y = (v + 2u)/5$.

The Jacobian is therefore

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} 1/5 & -2/5 \\ 2/5 & 1/5 \end{vmatrix} = \frac{1}{25} - \left(-\frac{4}{25}\right) = \frac{1}{5}.$$

Mapping the boundaries: substituting into the original boundary equations and simplifying converts each of the four lines into a coordinate line in the (u, v)-plane. The line $y = -2x$ becomes $v = 0$; similar substitutions give $u = 3$, $u = 0$, and $v = 2$ for the other three lines. The transformed region is the rectangle $\{0 \le u \le 3,\ 0 \le v \le 2\}$.

The change-of-variables formula gives

$$\iint_R f\, dA = \int_0^2 \int_0^3 f(x(u, v), y(u, v)) \cdot \frac{1}{5}\, du\, dv.$$

5 Vector Calculus

Vector calculus is the calculus of vector fields: functions that assign a vector to each point of space. The fundamental objects are the divergence and curl, which describe how a vector field flows out of a region or rotates around a point, and the line and surface integrals, which integrate a vector field along a curve or across a surface. The chapter culminates in three structural theorems — Green's theorem, Stokes' theorem, and the divergence theorem — which all generalise the fundamental theorem of calculus by relating an integral over a region to an integral over its boundary.

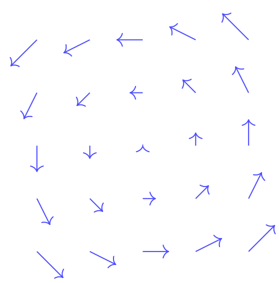
5.1 Vector fields

A vector field is a function that assigns to each point in space a vector. We will work mostly in $\mathbb{R}^2$ and $\mathbb{R}^3$. In components,

$$\vec{F}(x, y) = f(x, y)\hat{\boldsymbol{\imath}} + g(x, y)\hat{\boldsymbol{\jmath}}, \tag{5.1}$$

$$\vec{F}(x, y, z) = f(x, y, z)\hat{\boldsymbol{\imath}} + g(x, y, z)\hat{\boldsymbol{\jmath}} + h(x, y, z)\hat{\boldsymbol{k}}. \tag{5.2}$$

When we draw a vector field, we plot many sample arrows; the tail of each arrow sits at the point at which the field is being evaluated. Examples are everywhere in physics: the velocity field of a fluid attaches a velocity vector to each point of the fluid; a gravitational field attaches a force-per-unit-mass vector to each point in space; an electric field assigns an electric force to each point.



5.1.1 Gradient fields

A particularly important class of vector fields is the gradient fields. Given a scalar function f, the gradient

$$\nabla f = \frac{\partial f}{\partial x}\hat{\boldsymbol{\imath}} + \frac{\partial f}{\partial y}\hat{\boldsymbol{\jmath}} \quad (\text{or } + \frac{\partial f}{\partial z}\hat{\boldsymbol{k}}) \tag{5.3}$$

is itself a vector field, attaching a gradient vector to each point of the domain. A vector field $\vec{F}$ is called conservative if there exists some scalar function f such that

$$\vec{F} = \nabla f. \tag{5.4}$$

The function f is the potential function for $\vec{F}$. Conservative fields are exactly the gradient fields, and they are special: as we will see, line integrals of conservative fields depend only on the endpoints of the path, not on the path itself.

Example 14

Determine whether $\vec{F}(x, y) = 2xy\hat{\boldsymbol{\imath}} + (x^2 - 3y^2)\hat{\boldsymbol{\jmath}}$ is conservative, and if so, find a potential function.

Solution

We seek f such that $f_x = 2xy$ and $f_y = x^2 - 3y^2$. Integrating the first equation with respect to x,

$$f(x, y) = x^2 y + g(y),$$

where the "constant of integration" $g(y)$ may depend on y since y was held fixed during the x-integration. Differentiating with respect to y and matching against f_y,

$$f_y = x^2 + g'(y) = x^2 - 3y^2 \implies g'(y) = -3y^2 \implies g(y) = -y^3 + C.$$

Therefore $f(x, y) = x^2 y - y^3 + C$, and $\vec{F}$ is conservative with this potential.

5.2 Divergence and curl

Two scalar/vector quantities derived from a vector field encode information about its local behaviour: the divergence describes the net outflow of the field at a point, and the curl describes its tendency to rotate. Both can be motivated by imagining the vector field as the velocity of a fluid.

5.2.1 Divergence

The divergence of a vector field at a point measures the net rate at which fluid leaves a small volume around that point. If we imagine an infinitesimal sphere centred at the point and ask how much more fluid flows out across its surface than flows in, the answer per unit volume is the divergence.

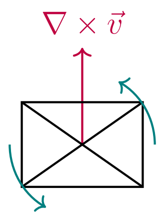
For $\vec{F} = P\hat{\boldsymbol{\imath}} + Q\hat{\boldsymbol{\jmath}} + R\hat{\boldsymbol{k}}$, the divergence is the scalar

$$\operatorname{div} \vec{F} = \nabla \cdot \vec{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}. \tag{5.5}$$

The notation $\nabla \cdot \vec{F}$ is suggestive: we "dot" the formal vector operator $\nabla = \langle \partial/\partial x, \partial/\partial y, \partial/\partial z \rangle$ with the vector field $\vec{F}$. The signs: a positive divergence at a point indicates a source (more flowing out than in); a negative divergence indicates a sink (more flowing in than out); a zero divergence indicates an incompressible field at that point (mass is locally conserved).

5.2.2 Curl

The curl of a vector field at a point measures the tendency of the field to rotate a small "paddle wheel" placed at the point. The wheel spins counter-clockwise (positive curl), clockwise (negative curl), or not at all (zero curl) when placed in the field. The curl is itself a vector, with magnitude giving the rotation speed and direction giving the rotation axis (by the right-hand rule).



For $\vec{F} = P\hat{\boldsymbol{\imath}} + Q\hat{\boldsymbol{\jmath}} + R\hat{\boldsymbol{k}}$ in $\mathbb{R}^3$, the curl is the symbolic determinant

$$\nabla \times \vec{F} = \nabla \times \vec{F} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}. \tag{5.6}$$

Expanding the determinant produces three components, each a difference of two partial derivatives.

The curl-of-a-gradient is always zero; intuitively, the field of steepest ascents from a scalar potential cannot “spin”.

Theorem 5.1. If f has continuous second partial derivatives, then $\nabla \times (\nabla f) = \mathbf{0}$.

Proof. Each component of $\nabla \times (\nabla f)$ is the difference of two mixed second partial derivatives of f. For example, the $\hat{\boldsymbol{\imath}}$-component is

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial y}\right) = f_{zy} - f_{yz}.$$

By Clairaut’s theorem, $f_{zy} = f_{yz}$ when these mixed partials are continuous, so the difference is zero. The other two components vanish by the same argument applied to different pairs of mixed partials. Hence every component of $\nabla \times (\nabla f)$ is zero, and $\nabla \times (\nabla f) = \mathbf{0}$. ∎

The contrapositive gives a useful diagnostic: a field with nonzero curl cannot be conservative.

Corollary 5.1.1. If $\vec{F}$ is conservative and its components have continuous partial derivatives, then $\nabla \times \vec{F} = \mathbf{0}$.

On simply connected regions of $\mathbb{R}^3$, the converse also holds.

Theorem 5.2. If $\vec{F}$ is defined on all of $\mathbb{R}^3$ with continuous component functions and $\nabla \times \vec{F} = \mathbf{0}$, then $\vec{F}$ is conservative.

On more general regions the converse can fail (a famous example is the field $\vec{F}(x, y) = (-y, x)/(x^2 + y^2)$ on $\mathbb{R}^2 \setminus \{0\}$, which has zero curl but is not conservative because of the hole at the origin).

5.3 Line integrals

We now extend integration from intervals to curves. There are two flavours: line integrals of scalar functions, which generalise arc-length integrals, and line integrals of vector fields, which compute work and circulation.

5.3.1 Line integrals of scalar functions

For a curve C in $\mathbb{R}^3$ with smooth parameterisation $\mathbf{r}(t)$, $a \leq t \leq b$, the line integral of a scalar function f along C is

$$\int_C f\, ds = \int_a^b f(\mathbf{r}(t))\ \|\mathbf{r}'(t)\|\ dt. \tag{5.7}$$

The element $ds = \|\mathbf{r}'(t)\|\ dt$ is the infinitesimal arc length, so the integral sums the values of f along the curve, weighted by infinitesimal arc length. Geometrically, if we visualise f as a height function above C, the line integral is the area of the curtain hanging above the curve. If f is a linear mass density along a wire shaped like C, the line integral gives the total mass of the wire.

Example 15

Compute $\int_C y\, ds$ where C is parameterised by $\mathbf{r}(t) = 2t\hat{\boldsymbol{\imath}} + t^3\hat{\boldsymbol{\jmath}}$ for $0 \leq t \leq 1$.

Solution

The derivative is $\mathbf{r}'(t) = 2\hat{\imath} + 3t^2\hat{\jmath}$ and its magnitude is $\|\mathbf{r}'(t)\| = \sqrt{4+9t^4}$. The integrand $y = t^3$ on the curve, so

$$\int_C y\,ds = \int_0^1 t^3\sqrt{4+9t^4}\,dt.$$

Substitute $u = 9t^4 + 4$, so $du = 36t^3\,dt$ and $t^3\,dt = du/36$. The bounds transform to $u(0) = 4$ and $u(1) = 13$,

$$\int_4^{13} \sqrt{u}\cdot\frac{du}{36} = \frac{1}{36}\cdot\frac{2}{3}\left[u^{3/2}\right]_4^{13} = \frac{1}{54}(13^{3/2} - 8).$$

5.3.2 Line integrals through vector fields

Given a vector field $\vec{F}$ and a curve C, the line integral of $\vec{F}$ along C is

$$\int_C \vec{F}\cdot d\mathbf{r} = \int_a^b \vec{F}(\mathbf{r}(t))\cdot\mathbf{r}'(t)\,dt, \tag{5.8}$$

where $d\mathbf{r} = \mathbf{r}'(t)\,dt$ is the infinitesimal displacement vector along the curve. Geometrically, the integrand at each point of the curve is the component of $\vec{F}$ along the direction of motion, weighted by the infinitesimal arc length; the sign matters because the dot product picks up the sign of the angle between $\vec{F}$ and the direction of motion. This integral is sometimes called the circulation of $\vec{F}$ along C.

If $\vec{F}$ is a force field and C is the path of a particle, then the line integral gives the work done by $\vec{F}$ on the particle.

Example 16

Find the work done by $\vec{F}(x,y) = xe^y\hat{\imath} + y\hat{\jmath}$ on a particle moving along the parabola $y = x^2$ from $(-1,1)$ to $(2,4)$.

Solution

Parameterise the path with $x = t$ and $y = t^2$, so $\mathbf{r}(t) = \langle t, t^2 \rangle$ and $\mathbf{r}'(t) = \langle 1, 2t \rangle$ for $-1 \le t \le 2$. Substituting into $\vec{F}$ along the path,

$$\vec{F}(\mathbf{r}(t)) = \left\langle te^{t^2},\ t^2 \right\rangle.$$

The dot product with $\mathbf{r}'(t)$ is

$$\vec{F} \cdot \mathbf{r}' = te^{t^2} \cdot 1 + t^2 \cdot 2t = te^{t^2} + 2t^3.$$

Therefore

$$W = \int_{-1}^{2} (te^{t^2} + 2t^3)\, dt.$$

The first piece integrates by the substitution $u = t^2$, giving $\frac{1}{2}e^{t^2}$. The second piece integrates to $t^4/2$. Combining,

$$W = \left[\frac{1}{2}e^{t^2} + \frac{t^4}{2}\right]_{-1}^{2} = \frac{1}{2}(e^4 - e) + \frac{1}{2}(16 - 1) = \frac{1}{2}(e^4 - e) + \frac{15}{2}.$$

5.3.3 Orientation

A curve has two possible orientations, corresponding to traversing it in two opposite directions. Reversing the orientation negates the line integral of a vector field,

$$\int_{-C} \vec{F} \cdot d\mathbf{r} = -\int_{C} \vec{F} \cdot d\mathbf{r}, \tag{5.9}$$

because the displacement vector $d\mathbf{r}$ flips sign. The scalar line integral, however, is unchanged by reversal,

$$\int_{-C} f\, ds = \int_{C} f\, ds,$$

since arc length is a non-negative quantity that does not depend on direction of travel.

5.4 The fundamental theorem of line integrals

Conservative fields admit a clean evaluation rule for their line integrals, in direct analogy to the fundamental theorem of calculus.

Theorem 5.3 (Fundamental theorem of line integrals)**.** If $\vec{F} = \nabla f$ on an open region D and C is a piecewise smooth curve in D from point A to point B, then

$$\int_C \vec{F} \cdot d\mathbf{r} = \int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A). \tag{5.10}$$

The line integral of a gradient field depends only on the endpoints, not on the path taken between them. This is the central feature of conservative fields, and it lets us evaluate work integrals by simply subtracting potential values rather than parameterising the path.

5.4.1 Path independence

A line integral $\int_C \vec{F} \cdot d\mathbf{r}$ is path-independent on a region D if its value depends only on the endpoints of C, not on the choice of path between them in D. This is equivalent to the condition $\oint_C \vec{F} \cdot d\mathbf{r} = 0$ for every closed curve C in D (one direction is immediate; for the other, observe that two paths between the same two points combined with reversing one of them gives a closed curve).

For sufficiently nice regions, path independence is equivalent to the field being conservative: if $\vec{F}$ has continuous components on an open connected region D and the line integral is path-independent there, then $\vec{F} = \nabla f$ for some scalar potential f on D.

5.4.2 Region vocabulary

To state precise hypotheses about the regions on which various theorems apply, we need a short vocabulary. A region $D \subseteq \mathbb{R}^n$ is:

- open if it does not contain any of its boundary points;

- connected if any two points in D can be joined by a path lying entirely in D;
- simply connected if it is connected and contains no holes (more precisely, every closed curve in D can be continuously contracted to a point within D).

On simply connected open regions, the local condition $\partial P/\partial y = \partial Q/\partial x$ is enough to conclude that a field is conservative.

Theorem 5.4. Let $\vec{F} = P\hat{\boldsymbol{\imath}} + Q\hat{\boldsymbol{\jmath}}$ be a vector field on an open simply connected region D. If P and Q have continuous partial derivatives and

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \tag{5.11}$$

throughout D, then $\vec{F}$ is conservative on D.

Example 17

Find the potential function of $\vec{F}(x, y) = (x + \arctan y)\hat{\boldsymbol{\imath}} + \dfrac{x + y}{1 + y^2}\hat{\boldsymbol{\jmath}}$.

Solution

We seek f with $f_x = x + \arctan y$ and $f_y = (x+y)/(1+y^2)$.

Integrating the first equation with respect to x,

$$f(x,y) = \frac{x^2}{2} + x \arctan y + g(y),$$

where $g(y)$ is an unknown function of y alone. Differentiating with respect to y,

$$f_y = \frac{x}{1+y^2} + g'(y).$$

Setting this equal to the given $f_y = (x+y)/(1+y^2)$ and solving for g',

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

Integrating, $g(y) = \frac{1}{2}\ln(1+y^2) + C$. Therefore

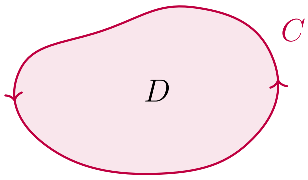
$$f(x,y) = \frac{x^2}{2} + x \arctan y + \frac{1}{2}\ln(1+y^2) + C.$$

5.5 Green's theorem

Green's theorem connects a line integral around a closed planar curve to a double integral over the region the curve bounds. In doing so it provides one of the simplest examples of the deeper principle that pervades vector calculus: an integral over a boundary equals an integral of a related quantity over the interior.

Let C be a positively oriented (counter-clockwise) piecewise smooth simple closed curve in $\mathbb{R}^2$, and let D be the region enclosed by C. If P and Q have continuous partial derivatives on an open region containing D, then

$$\oint_C P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA. \tag{5.12}$$



Special cases of Green's theorem yield several useful identities. Setting $P = -y/2$ and $Q = x/2$ makes the integrand $Q_x - P_y = 1$, so the double integral computes the area of D, and we obtain a formula for area as a line integral around the boundary,

$$\text{Area}(D) = \frac{1}{2}\oint_C x\,dy - y\,dx. \tag{5.13}$$

This is the kind of formula a planimeter (a mechanical device for measuring areas of irregular shapes) implements physically: the line integral over the boundary computes the area of the interior.

Example 18

Evaluate $\oint_C (x^2y + x^3)\,dx + 2xy\,dy$, where C is the positively oriented boundary of the region between $y = x$ and $y = x^2$.

Solution

Identifying $P = x^2y + x^3$ and $Q = 2xy$, the partial derivatives are $Q_x = 2y$ and $P_y = x^2$. The integrand of the corresponding double integral is $Q_x - P_y = 2y - x^2$. The region D is bounded above by $y = x$ and below by $y = x^2$, with x ranging from 0 to 1 (the curves meet at $(0,0)$ and $(1,1)$). Therefore

$$\begin{aligned}
\oint_C P\,dx + Q\,dy &= \int_0^1 \int_{x^2}^{x} (2y - x^2)\,dy\,dx \\
&= \int_0^1 \left[y^2 - x^2y\right]_{x^2}^{x} dx \\
&= \int_0^1 (x^2 - x^3) - (x^4 - x^4)\,dx \\
&= \int_0^1 (x^2 - x^3)\,dx = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.
\end{aligned}$$

5.5.1 Vector forms of Green's theorem

Green's theorem can be reformulated in two vector forms that foreshadow the higher-dimensional theorems coming later.

The circulation form expresses the line integral of $\vec{F}$ along C in terms of an integral over D of the third component of the curl,

$$\oint_C \vec{F} \cdot d\mathbf{r} = \iint_D (\nabla \times \vec{F}) \cdot \hat{\boldsymbol{k}} \, dA. \tag{5.14}$$

This is exactly the planar version of Stokes' theorem, which we will see in full generality below.

The flux form expresses the outward flux of $\vec{F}$ across C in terms of the divergence integrated over D,

$$\oint_C \vec{F} \cdot \mathbf{n} \, ds = \iint_D \nabla \cdot \vec{F} \, dA, \tag{5.15}$$

where $\mathbf{n}$ is the outward unit normal to C. This is the planar version of the divergence theorem.

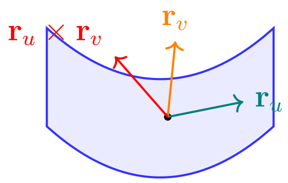
5.6 Parametric surfaces and surface area

A parametric surface is the image of a smooth function $\mathbf{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$,

$$\mathbf{r}(u, v) = x(u, v)\hat{\boldsymbol{\imath}} + y(u, v)\hat{\boldsymbol{\jmath}} + z(u, v)\hat{\boldsymbol{k}}. \tag{5.16}$$

As the parameter point (u, v) varies over the domain D, the head of $\mathbf{r}(u, v)$ sweeps out the surface $S = \mathbf{r}(D)$ in space. Each fixed value of one parameter (with the other free) traces out a grid curve on the surface, and the partial derivatives $\mathbf{r}_u$ and $\mathbf{r}_v$ are tangent to the surface along these grid curves at each point.

Because $\mathbf{r}_u$ and $\mathbf{r}_v$ are tangent to the surface and (for a smooth surface) linearly independent, their cross product $\mathbf{r}_u \times \mathbf{r}_v$ is normal to the surface. Its magnitude $\|\mathbf{r}_u \times \mathbf{r}_v\|$



is the area of the infinitesimal parallelogram in space spanned by $\mathbf{r}_u\,du$ and $\mathbf{r}_v\,dv$, so the surface area of S is

$$A = \iint_D \|\mathbf{r}_u \times \mathbf{r}_v\| \, du\, dv. \tag{5.17}$$

For the special case of a graph $z = g(x, y)$ over a region D in the xy-plane, parameterising by $\mathbf{r}(x, y) = \langle x, y, g(x, y)\rangle$ and computing $\mathbf{r}_x \times \mathbf{r}_y$ gives the formula

$$A = \iint_D \sqrt{1 + g_x^2 + g_y^2}\, dA, \tag{5.18}$$

the multivariable analogue of the arc-length formula $\int \sqrt{1 + (f'(x))^2}\, dx$.

5.7 Surface integrals

Surface integrals integrate a function over a surface; like line integrals, they come in scalar and vector flavours.

5.7.1 Scalar surface integrals

For a scalar function f on a parametric surface $S = \mathbf{r}(D)$,

$$\iint_S f\, dS = \iint_D f(\mathbf{r}(u, v))\, \|\mathbf{r}_u \times \mathbf{r}_v\| \, du\, dv. \tag{5.19}$$

The factor $\|\mathbf{r}_u \times \mathbf{r}_v\|$ is the infinitesimal surface area element dS, just as before. If f is a surface mass density (mass per unit area), this gives the total mass of a thin shell shaped like S.

5.7.2 Flux integrals

For a vector field $\vec{F}$ and an oriented surface S with unit normal $\mathbf{n}$, the flux of $\vec{F}$ across S is

$$\iint_S \vec{F} \cdot d\vec{S} = \iint_S \vec{F} \cdot \mathbf{n}\, dS = \iint_D \vec{F}(\mathbf{r}(u,v)) \cdot (\mathbf{r}_u \times \mathbf{r}_v)\, du\, dv. \tag{5.20}$$

The integrand at each point is the component of $\vec{F}$ along the normal direction; integrating this gives the net amount of flow through the surface in the direction of the chosen normal. The cross product $\mathbf{r}_u \times \mathbf{r}_v$ packages both the normal direction and the area element into a single vector, which is convenient for computation.

The orientation of the surface determines the sign of the flux: swapping u and v in the parameterisation reverses the cross product and therefore the sign. For closed surfaces, the standard convention is to use the outward-pointing normal.

5.8 Stokes' theorem

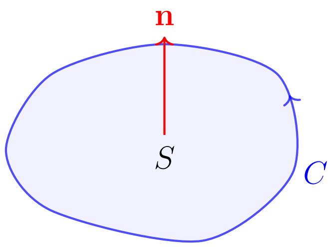
Stokes' theorem is the higher-dimensional generalisation of the circulation form of Green's theorem. It relates the circulation of a vector field around a closed curve in $\mathbb{R}^3$ to the flux of its curl through any surface that has the curve as its boundary.

Let S be an oriented piecewise smooth surface in $\mathbb{R}^3$ bounded by a simple closed piecewise smooth curve C, with C oriented positively relative to S in the sense of the right-hand rule (curl the fingers along C, and the thumb points in the direction of the chosen normal $\mathbf{n}$ on S). If $\vec{F}$ has continuous partial derivatives on an open region containing S, then

$$\oint_C \vec{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \vec{F}) \cdot d\vec{S}. \tag{5.21}$$

The two integrals describe the same quantity from two different viewpoints: the circulation around the boundary equals the total "rotational density" of the field passing through any surface that spans the boundary.

Green's theorem is the special case in which S is a planar region in the xy-plane and the normal is $\hat{\boldsymbol{k}}$. Many physics applications — Faraday's law of induction, Ampere's law — are statements of Stokes' theorem in disguise.



Example 19

Verify Stokes' theorem for $\vec{F} = -y^2\hat{\boldsymbol{\imath}} + x\hat{\boldsymbol{\jmath}} + z^2\hat{\boldsymbol{k}}$ with S the part of the paraboloid $z = x^2 + y^2$ below $z = 1$, oriented with upward-pointing normal.

Solution

The boundary C is the circle $x^2 + y^2 = 1$ at $z = 1$, oriented counter-clockwise viewed from above. Parameterise C by $x = \cos t$, $y = \sin t$, $z = 1$ for $0 \le t \le 2\pi$. For the line integral, compute $\vec{F}$ along C as $\left\langle -\sin^2 t,\ \cos t,\ 1 \right\rangle$ and the velocity as $\mathbf{r}' = \langle -\sin t,\ \cos t,\ 0\rangle$. Their dot product is

$$\vec{F} \cdot \mathbf{r}' = (-\sin^2 t)(-\sin t) + (\cos t)(\cos t) + 1 \cdot 0 = \sin^3 t + \cos^2 t.$$

Integrating over $0 \le t \le 2\pi$, the term $\sin^3 t$ integrates to zero (odd function over a full period), and $\cos^2 t$ integrates to π. So $\oint_C \vec{F} \cdot d\mathbf{r} = \pi$.
For the surface integral, compute $\nabla \times \vec{F} = \langle 0, 0, 1 + 2y\rangle$. Parameterise S by $\mathbf{r}(u, v) = \langle u \cos v,\ u \sin v,\ u^2\rangle$ for $0 \le u \le 1$, $0 \le v \le 2\pi$. The cross product is

$$\mathbf{r}_u \times \mathbf{r}_v = \left\langle -2u^2 \cos v,\ -2u^2 \sin v,\ u \right\rangle.$$

The dot product of $\nabla \times \vec{F}$ with $\mathbf{r}_u \times \mathbf{r}_v$ is $(1 + 2u \sin v) \cdot u$, since the curl has only a $\hat{\boldsymbol{k}}$-component. Integrating,

$$\begin{aligned} \iint_S (\nabla \times \vec{F}) \cdot d\vec{S} &= \int_0^{2\pi} \int_0^1 (u + 2u^2 \sin v)\, du\, dv \\ &= \int_0^{2\pi} \left(\frac{1}{2} + \frac{2}{3} \sin v \right) dv = \pi + 0 = \pi. \end{aligned}$$

Both integrals equal π, confirming Stokes' theorem.

5.9 The divergence theorem

The divergence theorem (also known as Gauss's theorem) is the higher-dimensional generalisation of the flux form of Green's theorem. It relates the outward flux of a vector field across a closed surface to the integral of its divergence over the enclosed solid.

Let E be a simple solid region in $\mathbb{R}^3$ with closed boundary surface S oriented with outward-pointing normal. If $\vec{F}$ has continuous partial derivatives on an open region containing E, then

$$\iint_S \vec{F} \cdot d\vec{S} = \iiint_E \nabla \cdot \vec{F}\, dV. \tag{5.22}$$

The conceptual content: the net outward flow of $\vec{F}$ through the enclosing surface equals the total amount of "source minus sink" inside the region. This is a conservation law connecting the surface flux to volumetric sources and sinks. In physics, it underlies Gauss's law in electromagnetism (the flux of the electric field through a closed surface equals the enclosed charge divided by ϵ_0) and many similar conservation statements.

Example 20

Compute the outward flux of $\vec{F}(x, y, z) = x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}$ across the unit sphere.

Solution

The divergence is $\nabla \cdot \vec{F} = 1 + 1 + 1 = 3$, a constant. The volume of the unit ball is $4\pi/3$. By the divergence theorem,

$$\iint_S \vec{F} \cdot d\vec{S} = \iiint_E 3\, dV = 3 \cdot \frac{4\pi}{3} = 4\pi.$$

Computing the surface integral directly would require parameterising the sphere and computing $\mathbf{r}_u \times \mathbf{r}_v$ explicitly, a much more painful calculation.

5.9.1 Relations between the theorems

The three "fundamental theorems" of vector calculus in $\mathbb{R}^3$ are all generalisations of the same underlying principle, the fundamental theorem of calculus, which says that integrating a derivative over a region reduces to evaluating an antiderivative at the boundary. In each case, an integral of a derivative over a region is equal to an integral of the original quantity over the boundary of that region. Schematically,

$$\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A) \qquad \text{(fundamental theorem of line integrals),} \tag{5.23}$$

$$\oint_{\partial S} \vec{F} \cdot d\mathbf{r} = \iint_S \nabla \times \vec{F} \cdot d\vec{S} \qquad \text{(Stokes' theorem),} \tag{5.24}$$

$$\iint_{\partial E} \vec{F} \cdot d\vec{S} = \iiint_E \nabla \cdot \vec{F}\, dV \qquad \text{(divergence theorem).} \tag{5.25}$$

In each line, the operator on the left side of the right-hand integral (gradient, curl, or divergence) is in some sense a "derivative", and the equation says this derivative integrated over a region equals the underlying quantity integrated over the boundary. The unifying generalisation is the modern Stokes theorem on differential forms, which collects all three of these into a single statement that applies in any dimension.