MathCastle

Multivariable Calculus

Complete course booklet • 11 topics

Contents

1 Vectors in Space

Dot, cross, and the geometry they encode

1.1 Dot and Cross Products

Dot product = angle detector

$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta$, computed coordinate-wise as $u_1v_1 + u_2v_2 + u_3v_3$. Zero dot product means perpendicular; the sign tells acute vs. obtuse.

Cross product = area machine

$\mathbf{u} \times \mathbf{v}$ is perpendicular to both inputs (right-hand rule) with magnitude $|\mathbf{u}||\mathbf{v}| \sin\theta$ — the AREA of the parallelogram they span. For plane vectors $(a, b, 0)$ and $(c, d, 0)$ it collapses to the determinant $|ad - bc|$ on the z-axis.

A 60° surprise

The cube-edge vectors $(1, 1, 0)$ and $(0, 1, 1)$: dot product 1, lengths $\sqrt{2}$ each, so $\cos\theta = \frac{1}{2}$ and $\theta = 60^\circ$ — connecting face diagonals of a cube always form equilateral triangles.

Why vectors run all of physics

Force, velocity, momentum, electric field — anything with a magnitude AND a direction is a vector, and the dot/cross pair are the only two ways to multiply them that respect rotations. Work done is a dot product ($W = \mathbf{F} \cdot \mathbf{d}$); torque and magnetic force are cross products. Learn these two operations and you have learned the grammar of mechanics.

Try it: Try it: perpendicular or not?

Are $(2, -3, 1)$ and $(4, 3, 1)$ perpendicular? Work: dot product $= 8 - 9 + 1 = 0$ — yes, exactly perpendicular, no angle computation needed. Now check $(1, 2, 2)$ against itself: $\mathbf{v} \cdot \mathbf{v} = 1 + 4 + 4 = 9 = |\mathbf{v}|^2$, so $|\mathbf{v}| = 3$ — the dot product also measures length.

1.2 Proofs & Why It Matters

Proof: u·v = |u||v|cos

Apply the law of cosines to the triangle with sides $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$: $|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}| \cos\theta$.

Expand the left side coordinate-wise: $|\mathbf{u} - \mathbf{v}|^2 = \sum(u_i - v_i)^2 = |\mathbf{u}|^2 - 2\sum u_iv_i + |\mathbf{v}|^2$. Comparing the two expansions, the $|\mathbf{u}|^2$ and $|\mathbf{v}|^2$ terms cancel and $\sum u_iv_i = |\mathbf{u}||\mathbf{v}| \cos\theta$. ■

Proof: $|\mathbf{u}\times\mathbf{v}|$ is the parallelogram area

The parallelogram on $\mathbf{u}, \mathbf{v}$ has base $|\mathbf{u}|$ and height $|\mathbf{v}|\sin\theta$, so its area is $|\mathbf{u}||\mathbf{v}|\sin\theta$.

Now compute $|\mathbf{u}\times\mathbf{v}|^2$ from the component formula; the Lagrange identity gives $|\mathbf{u}\times\mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u}\cdot\mathbf{v})^2 = |\mathbf{u}|^2|\mathbf{v}|^2(1-\cos^2\theta) = (|\mathbf{u}||\mathbf{v}|\sin\theta)^2$. Taking square roots matches the area exactly. ■

Proof: Cauchy–Schwarz $|\mathbf{u}\cdot\mathbf{v}|$ $|\mathbf{u}||\mathbf{v}|$

For every real t, $0 \le |\mathbf{u}+t\mathbf{v}|^2 = |\mathbf{u}|^2 + 2t(\mathbf{u}\cdot\mathbf{v}) + t^2|\mathbf{v}|^2$ — a quadratic in t that is never negative, so its discriminant is ≤ 0: $4(\mathbf{u}\cdot\mathbf{v})^2 - 4|\mathbf{u}|^2|\mathbf{v}|^2 \le 0$.

Rearranged, $|\mathbf{u}\cdot\mathbf{v}| \le |\mathbf{u}||\mathbf{v}|$, with equality exactly when $\mathbf{u}+t\mathbf{v} = \mathbf{0}$ for some t — parallel vectors. ■

1.3 Seeing It: Vectors in 2D & 3D

Vector addition is a parallelogram

Placing $\mathbf{v}$'s tail at $\mathbf{u}$'s head (or vice versa — the parallelogram shows both orders at once) lands at the same corner: $\mathbf{u}+\mathbf{v}$. Coordinates just add: $(3,1)+(1,2)=(4,3)$, and the picture explains why — the horizontal runs add, the vertical rises add, independently.

(diagram in the online version)

The diagonal of the parallelogram IS the sum; the other diagonal, from head of $\mathbf{v}$ to head of $\mathbf{u}$, is the difference $\mathbf{u}-\mathbf{v}$. Every force diagram, velocity composition, and displacement chain in physics is this one picture reused.

The angle lives in the dot product

Draw both vectors from one point and the angle θ between them is captured algebraically by $\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}||\mathbf{v}|\cos\theta$.

(diagram in the online version)

Reading the picture: when θ is small the vectors reinforce and the dot product is large and positive; as θ passes $90°$ the cosine — and the dot product — crosses zero and turns negative. So perpendicularity is the single cleanest test in the subject: compute one number, check if it is 0. No angle ever needs to be found unless the problem asks for it.

Projection: the shadow and the perpendicular remainder

The green vector is the SHADOW of $\mathbf{u}$ on the line through $\mathbf{v}$ — the projection $\mathrm{proj}_{\mathbf{v}}\mathbf{u} = \dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$.

(diagram in the online version)

The dashed red segment is what remains, $\mathbf{u} - \mathrm{proj}_{\mathbf{v}}\mathbf{u}$, and the small square marks the right angle that DEFINES the projection: the remainder is exactly perpendicular to $\mathbf{v}$. The remainder's length is the distance from $\mathbf{u}$'s head to the line — which is why this one picture powers point-to-line distances, least squares, and Gram–Schmidt alike.

The cross product in 3D: perpendicular, with area for length

In three dimensions $\mathbf{u}\times\mathbf{v}$ points PERPENDICULAR to the shaded parallelogram that $\mathbf{u}$ and $\mathbf{v}$ span — choose which of the two perpendicular directions by the right-hand rule: curl the right hand's fingers from $\mathbf{u}$ toward $\mathbf{v}$ and the thumb gives $\mathbf{u}\times\mathbf{v}$ (so $\mathbf{v}\times\mathbf{u}$ points the OPPOSITE way: the cross product is anti-commutative). Its length equals the parallelogram's area, $|\mathbf{u}||\mathbf{v}|\sin\theta$: parallel vectors span no area and give the zero vector, perpendicular ones give the maximum $|\mathbf{u}||\mathbf{v}|$. Normal vectors to planes, torque, and area computations all read straight off this picture.

(diagram in the online version)

1.4 Going Deeper: Explanations & Worked Problems

What the dot product really is, from scratch

Start with lengths: in 3D, $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + u_3^2}$ — Pythagoras applied twice, first in the floor plane to get the horizontal run $\sqrt{u_1^2 + u_2^2}$, then in the vertical plane with the height u_3.

Now ask: how much do two vectors AGREE in direction? Multiply matching coordinates and add: $\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$. Why is that a good measure? Because it is LINEAR in each slot ($(\mathbf{a}+\mathbf{b})\cdot\mathbf{v} = \mathbf{a}\cdot\mathbf{v} + \mathbf{b}\cdot\mathbf{v}$, straight from distributing the products), it gives $\mathbf{u}\cdot\mathbf{u} = |\mathbf{u}|^2$ (agreement with yourself is total), and — the theorem proved on the proof slide — it equals $|\mathbf{u}||\mathbf{v}|\cos\theta$. So the sign of the dot product is the sign of $\cos\theta$: positive means the angle is acute (the vectors lean the same way), zero means perpendicular, negative means obtuse. Every fact in this topic flows from those three properties.

Worked: the angle between (3, 4, 0) and (5, 12, 0), every step

Step 1 — dot product: $3\cdot5+4\cdot12+0\cdot0 = 15+48 = 63$. Step 2 — lengths: $|\mathbf{u}| = \sqrt{9+16} = 5$ and $|\mathbf{v}| = \sqrt{25+144} = 13$ (two Pythagorean triples). Step 3 — solve the angle formula for $\cos\theta$: $\cos\theta = \dfrac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \dfrac{63}{65}$.

Step 4 — interpret: $\frac{63}{65}$ is close to 1, so the angle is small ($\theta \approx 14.25°$); both vectors point into the first quadrant, so a small positive angle is exactly what geometry predicts. Always end with this sanity check — a negative cosine for two first-quadrant vectors would signal an arithmetic slip.

Worked: a full cross product, and checking it two ways

Compute $(2,1,0)\times(1,3,0)$. Step 1 — the determinant recipe: $\mathbf{u}\times\mathbf{v} = (u_2v_3 - u_3v_2, ; u_3v_1 - u_1v_3, ; u_1v_2 - u_2v_1) = (1\cdot0 - 0\cdot3, ; 0\cdot1 - 2\cdot0, ; 2\cdot3 - 1\cdot1) = (0,0,5)$.

Step 2 — perpendicularity check: $(0,0,5)\cdot(2,1,0) = 0$ ✓ and $(0,0,5)\cdot(1,3,0) = 0$ ✓, as the cross product must be orthogonal to both inputs. Step 3 — magnitude check by the area formula:

$|\mathbf{u}| = \sqrt{5}$, $|\mathbf{v}| = \sqrt{10}$, and $\cos\theta = \frac{2+3}{\sqrt{5}\sqrt{10}} = \frac{5}{\sqrt{50}} = \frac{1}{\sqrt{2}}$, so $\sin\theta = \frac{1}{\sqrt{2}}$ and $|\mathbf{u}||\mathbf{v}|\sin\theta = \sqrt{50}\cdot\frac{1}{\sqrt{2}} = 5$ — matching $|(0,0,5)| = 5$ exactly. Two independent routes to the same number is how you KNOW the computation is right.

2 Lines & Planes in Space

Normals, distances, and intersections

2.1 Equations of Lines and Planes

A plane is a point and a normal

Every plane in space is captured by ONE linear equation $ax + by + cz = d$, and the coefficient vector $\mathbf{n} = (a, b, c)$ is no accident: it is the NORMAL, perpendicular to the plane.

Why: if P_0 lies on the plane and P is any other plane point, the displacement $P - P_0$ lies in the plane, and the equation says exactly $\mathbf{n} \cdot (P - P_0) = 0$ — the normal is perpendicular to every in-plane direction. So building planes is mechanical: given a point and a normal, dot them; given three points, form two edge vectors and take their CROSS PRODUCT for the normal. Given the equation, read the normal straight off the coefficients. Parallel planes share a normal and differ only in d; perpendicular planes have perpendicular normals — every geometric relationship between planes translates into a dot- or cross-product statement about their normals.

Why lines need a parameter but planes need an equation

In 3D a single linear equation removes ONE degree of freedom, leaving a 2-dimensional plane — so a line (1-dimensional) needs TWO equations, or more usefully a parametric form $\mathbf{r}(t) = \mathbf{p} + t\mathbf{v}$: start at a point, walk along a direction. The dimension count ($3 -$ equations $=$ dimension) is rank–nullity from linear algebra wearing geometric clothes — one of many places these two courses are secretly the same course.

All distances are projections onto the normal

Distance from a point $P_0 = (x_0, y_0, z_0)$ to the plane $ax + by + cz = d$: $\dfrac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}}$. The logic: take any plane point Q, and project the displacement $P_0 - Q$ onto the unit normal — only the normal component carries you away from the plane.

The numerator is the (signed) un-normalized projection; dividing by $|\mathbf{n}|$ finishes it. The same idea handles parallel planes ($\frac{|d_1 - d_2|}{|\mathbf{n}|}$) and, with a cross product replacing the dot, point-to-line distances. One projection, every distance formula in the chapter.

Try it: Try it: a plane through three points

Find the plane through $(1, 0, 0)$, $(0, 2, 0)$, $(0, 0, 3)$. Work: edge vectors $(-1, 2, 0)$ and $(-1, 0, 3)$;

cross product $= (6,3,2)$; plane $6x+3y+2z=d$ with $d=6$ from the first point. Check the other two: $3\cdot 2=6$ ✓, $2\cdot 3=6$ ✓.

Answer. Shortcut worth knowing: intercepts p,q,r give $\frac{x}{p}+\frac{y}{q}+\frac{z}{r}=1$ instantly.

2.2 Proofs & Why It Matters

Proof: the point-to-plane distance formula

Let Q be any point on the plane ($\mathbf{n}\cdot Q=d$) and P_0 the outside point. The distance is the length of the projection of P_0-Q onto the unit normal: $\left|\frac{\mathbf{n}\cdot(P_0-Q)}{|\mathbf{n}|}\right| = \frac{|\mathbf{n}\cdot P_0-\mathbf{n}\cdot Q|}{|\mathbf{n}|} = \frac{|ax_0+by_0+cz_0-d|}{\sqrt{a^2+b^2+c^2}}$ — the arbitrary Q vanished into d, which is why the formula needs no particular plane point. ■ SIGNIFICANCE: this projection pattern is the template for every "distance to a flat object" problem — and in machine learning it is literally the margin of a linear classifier.

Where this leads

Planes are the first example of the general idea "linear equation = flat object + normal vector." In linear algebra the same picture becomes hyperplanes and the row space; in optimization, constraint surfaces and Lagrange multipliers; in graphics, every triangle of every 3D model carries its normal for lighting. The dot-product test you just learned is computed billions of times per frame in a video game.

2.3 Going Deeper: Worked Problems

Worked: the line where two planes meet

Find parametric equations for the intersection of $x+y+z=6$ and $x-y+2z=3$.

Step 1 — the direction of the line is perpendicular to BOTH normals, so take the cross product $(1,1,1)\times(1,-1,2)=(1\cdot 2-1\cdot(-1),;1\cdot 1-1\cdot 2,;1\cdot(-1)-1\cdot 1)=(3,-1,-2)$. Step 2 — find one point on both planes: set $z=0$ and solve $x+y=6$, $x-y=3$: $x=4.5$, $y=1.5$. Step 3 — assemble: $\mathbf{r}(t)=(4.5,1.5,0)+t(3,-1,-2)$. Step 4 — verify at $t=1$: the point $(7.5,0.5,-2)$ gives $7.5+0.5-2=6$ ✓ and $7.5-0.5-4=3$ ✓. Two planes, one cross product, one small system.

Worked: distance from a point to a line in space

Find the distance from $P=(1,2,3)$ to the line through $A=(0,0,1)$ with direction $\mathbf{v}=(1,1,0)$.

Step 1 — displacement $\mathbf{w}=P-A=(1,2,2)$. Step 2 — the perpendicular distance is $\frac{|\mathbf{w}\times\mathbf{v}|}{|\mathbf{v}|}$ (the parallelogram area divided by its base). Compute $\mathbf{w}\times\mathbf{v}=(2\cdot 0-2\cdot 1,;2\cdot 1-1\cdot 0,;1\cdot 1-2\cdot 1)=(-2,2,-1)$, magnitude 3. Step 3 — divide by $|\mathbf{v}|=\sqrt{2}$: distance $\frac{3}{\sqrt{2}}=\frac{3\sqrt{2}}{2}$. Step 4 — cross-check by projection: the component of $\mathbf{w}$ along $\mathbf{v}$ is $\frac{3}{\sqrt{2}}$, and $|\mathbf{w}|^2 - \left(\frac{3}{\sqrt{2}}\right)^2 = 9 - 4.5 = 4.5 = \left(\frac{3\sqrt{2}}{2}\right)^2$ ✓ — Pythagoras agrees with the cross product.

3 Curves & Motion

Vector-valued functions, speed, arc length, curvature

3.1 Vector-Valued Functions

Differentiate coordinate by coordinate

A path $\mathbf{r}(t) = (x(t), y(t), z(t))$ has velocity $\mathbf{r}'(t)$ (tangent to the path) and speed $|\mathbf{r}'(t)|$. Arc length integrates the speed: $L = \int_a^b |\mathbf{r}'(t)|\, dt$. The helix $(\cos t, \sin t, t)$ has CONSTANT speed $\sqrt{2}$ — its length is just $\sqrt{2}$ times the time elapsed.

Curvature measures how hard you turn

For a graph $y = f(x)$: $\kappa = \dfrac{|y''|}{(1 + y'^2)^{3/2}}$. A circle of radius R has constant curvature $\frac{1}{R}$ — so $\frac{1}{\kappa}$ is the radius of the circle that best hugs the curve. The parabola $y = x^2$ turns hardest at its vertex, where $\kappa = 2$.

Projectile motion, decomposed

For $\mathbf{r}(t) = (3t,\ 4t - 5t^2)$: velocity $(3, 4 - 10t)$, acceleration $(0, -10)$. The top of the arc is where velocity is horizontal — perpendicular to gravity — at $t = \frac{2}{5}$. Physics questions become dot-product questions.

Significance: kinematics is vector calculus

GPS arc lengths, roller-coaster g-forces (curvature times speed squared), and orbital mechanics are all $\mathbf{r}(t)$ computations. When engineers design a highway transition curve they are literally prescribing a curvature function — the "clothoid" ramps curvature linearly so the steering wheel turns at constant speed.

Try it: Try it: speed vs. velocity

A particle follows $\mathbf{r}(t) = (3\cos t, 3\sin t)$. Work: velocity $(-3\sin t, 3\cos t)$ — always changing direction; speed $\sqrt{9\sin^2 t + 9\cos^2 t} = 3$ — never changing at all. Uniform circular motion in one line: constant speed, perpetually turning velocity, acceleration pointing inward with magnitude $\frac{v^2}{r} = 3$.

3.2 Proofs & Why It Matters

Proof: arc length = |r(t)| dt

Chop $[a, b]$ into small steps Δt. Each step moves the point by $\mathbf{r}(t + \Delta t) - \mathbf{r}(t) \approx \mathbf{r}'(t)\,\Delta t$, a straight segment of length $|\mathbf{r}'(t)|\Delta t$ (Pythagoras applied to the coordinate increments).

Summing the segments gives a Riemann sum $\sum |\mathbf{r}'(t_i)|\,\Delta t$, and the limit as $\Delta t \to 0$ is by definition $\int_a^b |\mathbf{r}'(t)|\,dt$. ■

Proof: = |y|/(1 + y)$^{3/2}$ for graphs

Curvature is the turning rate of the tangent angle per unit of ARC length: $\kappa = \left|\frac{d\phi}{ds}\right|$ where $\tan\phi = y'$. Differentiate in x: $\sec^2\phi\,\frac{d\phi}{dx} = y''$, so $\frac{d\phi}{dx} = \frac{y''}{1+y'^2}$. Meanwhile arc length grows at $\frac{ds}{dx} = \sqrt{1 + y'^2}$.

Divide: $\kappa = \left|\frac{d\phi/dx}{ds/dx}\right| = \dfrac{|y''|}{(1 + y'^2)^{3/2}}$. ■

3.3 Going Deeper: Explanations & Worked Problems

Velocity, speed, and why arc length is an integral

A moving point $\mathbf{r}(t)$ has two different "how fast" questions hiding in it. The VELOCITY $\mathbf{r}'(t)$ is a vector: it points along the instantaneous direction of travel (tangent to the path) and its components are just the one-variable derivatives of each coordinate.

The SPEED $|\mathbf{r}'(t)|$ is a number — the length of that vector. Distance traveled is speed accumulated over time, which is why arc length is $L = \int_a^b |\mathbf{r}'(t)|\,dt$ and not anything simpler: over a tiny interval Δt the point moves approximately $\mathbf{r}'(t)\Delta t$, a straight step of length $|\mathbf{r}'(t)|\Delta t$, and adding infinitely many tiny straight steps IS the integral. Note what this is not: it is not $|\mathbf{r}(b) - \mathbf{r}(a)|$, the straight-line displacement — a runner on a circular track covers a full lap of distance with zero displacement.

Worked: the helix length, with the speed computed honestly

Find the length of $\mathbf{r}(t) = (\cos t, \sin t, t)$ for $0 \le t \le 2\pi$.

Step 1 — differentiate coordinate-wise: $\mathbf{r}'(t) = (-\sin t, \cos t, 1)$. Step 2 — speed: $|\mathbf{r}'(t)| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{1+1} = \sqrt{2}$, using the Pythagorean identity to collapse the first two squares. Step 3 — notice the speed is CONSTANT: the integral needs no antiderivative work, $L = \int_0^{2\pi} \sqrt{2}\,dt = 2\sqrt{2}\,\pi$. Step 4 — sanity check by unrolling: the helix wraps around a cylinder of circumference 2π while climbing 2π; unrolled flat it is the hypotenuse of a right triangle with both legs 2π, length $\sqrt{(2\pi)^2 + (2\pi)^2} = 2\sqrt{2}\,\pi$ ✓ — the same answer with no calculus at all.

Worked: curvature of y = x away from the vertex

Find κ for $y = x^2$ at $x = 1$. Step 1 — derivatives: $y' = 2x = 2$ and $y'' = 2$ at $x = 1$. Step 2 — plug into the formula: $\kappa = \dfrac{|y''|}{(1 + y'^2)^{3/2}} = \dfrac{2}{(1 + 4)^{3/2}} = \dfrac{2}{5\sqrt{5}}$.

Step 3 — compare with the vertex, where $y' = 0$ gives $\kappa = 2$: at $x = 1$ the curvature is $\frac{2}{5\sqrt{5}} \approx 0.18$, roughly a tenth of the vertex value. Step 4 — interpret via the osculating circle: the best-fitting circle at $x = 1$ has radius $\frac{1}{\kappa} = \frac{5\sqrt{5}}{2} \approx 5.6$, versus radius $\frac{1}{2}$ at the vertex — the parabola straightens out fast as you leave the bottom, and the formula quantifies exactly how fast.

4 Partial Derivatives & the Gradient

Slopes in every direction at once

4.1 Partial Derivatives

Freeze everything but one variable

For $f(x, y)$, the partial f_x differentiates in x while treating y as a CONSTANT (and vice versa for f_y). All the one-variable rules apply unchanged. Example: $f = x^3y^2$ gives $f_x = 3x^2y^2$ and $f_y = 2x^3y$.

The gradient points uphill

The vector $\nabla f = (f_x, f_y)$ points in the direction of STEEPEST increase, and its length is that steepest slope.

The directional derivative along a unit vector $\mathbf{u}$ is the dot product $D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$ — so no direction can beat $|\nabla f|$, and the level curves are always perpendicular to the gradient.

Tangent planes

The tangent plane to $z = f(x, y)$ at (a, b) is $z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ — the two partials are exactly the two tilts of the plane. For $z = x^2 + y^2$ at $(1, 2, 5)$: $z = 5 + 2(x - 1) + 4(y - 2)$.

Significance: the gradient is how machines learn

Gradient DESCENT — step opposite ∇f to reduce a loss — trains essentially every neural network. The fact that $-\nabla f$ is the steepest way down, proved in this topic, is the reason a billion-parameter model can be trained at all: you never need the whole landscape, only the local gradient.

Try it: Try it: a two-line gradient

For $f(x,y) = x^2 + xy$, compute $\nabla f(2,1)$. Work: $f_x = 2x + y = 5$, $f_y = x = 2$: $\nabla f = (5,2)$. Steepest slope: $|\nabla f| = \sqrt{29}$. And the slope heading straight along $(0,1)$? $\nabla f \cdot (0,1) = 2$ — less than $\sqrt{29} \approx 5.4$, as it must be.

4.2 Proofs & Why It Matters

Proof: the gradient is the steepest direction

For a unit vector $\mathbf{u}$, the slope in direction $\mathbf{u}$ is $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u} = |\nabla f| \cos\theta$, where θ is the angle between $\mathbf{u}$ and ∇f (this dot-product form follows from the chain rule applied to $t \mapsto f(\mathbf{x}+t\mathbf{u})$).

Since $\cos\theta \leq 1$ with equality only at $\theta = 0$, the slope is maximized precisely when $\mathbf{u}$ points along ∇f, and the maximal slope is $|\nabla f|$. ∎

Proof: level curves gradient

Walk along a level curve $\mathbf{r}(t)$ with $f(\mathbf{r}(t)) = c$ constant. Differentiate both sides in t (chain rule): $\nabla f \cdot \mathbf{r}'(t) = 0$. The tangent to the level curve is always orthogonal to the gradient — which is why contour maps and steepest-descent paths cross at right angles. ∎

Proof: the tangent plane formula

The plane through $(a, b, f(a,b))$ with slopes f_x in the x-direction and f_y in the y-direction is $z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$ by construction.

Differentiability of f at (a,b) says exactly that $f(x,y)$ minus this plane is $o\left(\sqrt{(x-a)^2 + (y-b)^2}\right)$ — the plane matches the surface to first order, which is what makes it THE tangent plane and the basis of linear approximation. ∎

4.3 Going Deeper: Explanations & Worked Problems

Partial derivatives from the limit up

The definition mirrors one-variable calculus exactly: $f_x(a,b) = \lim_{h\to 0} \dfrac{f(a+h,b) - f(a,b)}{h}$ — wiggle x only, watch the output, divide, take the limit. Geometrically, slice the surface $z = f(x,y)$ with the vertical plane $y = b$; the slice is an ordinary curve, and f_x is its slope.

That is why every one-variable rule (product, quotient, chain) transfers verbatim: during the computation, y is a constant like any other. Two subtleties are worth knowing early. First, the mixed partials usually agree: $f_{xy} = f_{yx}$ for smooth functions (Clairaut's theorem), so the order of differentiation is a free choice. Second, having both partials is weaker than true differentiability — the honest statement is that when the partials exist and are continuous, the tangent-plane approximation $f(a+h,b+k) \approx f(a,b) + f_x h + f_y k$ has error shrinking faster than $\sqrt{h^2+k^2}$, which is what every application (linearization, error propagation, Newton's method) actually uses.

Worked: gradient, steepest slope, and a directional derivative together

Let $f(x,y) = x^2 y$ at the point $(1,2)$.

Step 1 — partials: $f_x = 2xy$ (treat y constant) and $f_y = x^2$ (treat x constant). Step 2 — evaluate: $\nabla f(1,2) = (2 \cdot 1 \cdot 2, ; 1^2) = (4,1)$. Step 3 — steepest ascent: the maximum slope at this point is $|\nabla f| = \sqrt{16+1} = \sqrt{17} \approx 4.12$, achieved walking in the direction of $(4,1)$. Step 4 — slope in a GIVEN direction $\mathbf{u} = \left(\frac{3}{5}, \frac{4}{5}\right)$ (note $|\mathbf{u}| = 1$, required!): $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u} = 4 \cdot \frac{3}{5} + 1 \cdot \frac{4}{5} = \frac{12+4}{5} = \frac{16}{5} = 3.2$. Step 5 — coherence check: $3.2 < 4.12$ ✓, as no direction can beat the gradient direction; and walking perpendicular to $(4,1)$, e.g. along $(1,-4)/\sqrt{17}$, gives slope $\frac{4-4}{\sqrt{17}} = 0$ — that is the direction the level curve runs.

Worked: a tangent plane and what it predicts

Find the tangent plane to $z = x^2 + y^2$ at $(1,2,5)$, then use it to estimate $f(1.1, 1.9)$.

Step 1 — partials: $f_x = 2x = 2$ and $f_y = 2y = 4$ at the point. Step 2 — assemble the plane: $z = 5 + 2(x-1) + 4(y-2)$. Step 3 — estimate: $z \approx 5 + 2(0.1) + 4(-0.1) = 5 + 0.2 - 0.4 = 4.8$. Step 4 — compare with the truth: $f(1.1, 1.9) = 1.21 + 3.61 = 4.82$. The linear estimate missed by 0.02 — and the miss is exactly the curvature the plane cannot see: $(0.1)^2 + (-0.1)^2 = 0.02$. For this surface the tangent-plane error is ALWAYS $h^2 + k^2$, a perfect illustration of "error shrinks like the square of the step."

5 The Chain Rule & Implicit Differentiation

Dependency trees, implicit curves, and Euler's theorem

5.1 Differentiating Through Dependencies

The chain rule: one term per path

When $z = f(x,y)$ and both x and y depend on t, the total rate of change adds one contribution per route of dependence: $\dfrac{dz}{dt} = \dfrac{\partial z}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial z}{\partial y}\dfrac{dy}{dt}$.

Draw the dependency tree (z on top, x and y below, t at the bottom): multiply derivatives ALONG each path, add ACROSS paths. The rule extends to any number of variables and any depth of nesting — partials with respect to several inputs just select which bottom variable you follow. Intuition: a small change dt perturbs x by $x'dt$ and y by $y'dt$; the tangent-plane approximation converts each into a z-change, and to first order the two effects simply add.

The same rule runs backpropagation

Training a neural network is nothing but this chain rule applied through thousands of nested dependencies: the loss depends on outputs, which depend on weights through layer after layer. "Multiply along paths, add across paths" is precisely the backpropagation algorithm — the multivariable chain rule is arguably the most economically important formula in modern computing.

Implicit differentiation, industrialized

A curve given implicitly by $F(x, y) = 0$ has slope $\dfrac{dy}{dx} = -\dfrac{F_x}{F_y}$ wherever $F_y \neq 0$ — no solving for y, ever. Derivation in one line: differentiate $F(x, y(x)) = 0$ by the chain rule: $F_x + F_y\, y' = 0$.

This retires the ad-hoc implicit differentiation of first-year calculus: compute two partials, form the ratio, attach a minus sign. In three variables, a surface $F(x, y, z) = 0$ has $\dfrac{\partial z}{\partial x} = -\dfrac{F_x}{F_z}$ by the same argument, and ∇F is the surface normal — tangent planes to implicit surfaces come free.

Try it: Try it: the folium at (2, 4)

For $x^3 + y^3 = 9xy$ find y' at $(2, 4)$. Work: $F = x^3 + y^3 - 9xy$; $F_x = 3x^2 - 9y = -24$ and $F_y = 3y^2 - 9x = 30$ at the point; slope $= -\frac{-24}{30} = \frac{4}{5}$. Confirm the point first ($8 + 64 = 72 = 9 \cdot 8$ ✓) — implicit-differentiation errors are usually really point-not-on-curve errors.

5.2 Proofs & Why It Matters

Proof: the chain rule from the tangent plane

Differentiability means $\Delta z = f_x \Delta x + f_y \Delta y + \varepsilon$, where ε shrinks faster than $\sqrt{\Delta x^2 + \Delta y^2}$. Divide by Δt: $\dfrac{\Delta z}{\Delta t} = f_x \dfrac{\Delta x}{\Delta t} + f_y \dfrac{\Delta y}{\Delta t} + \dfrac{\varepsilon}{\Delta t}$.

As $\Delta t \to 0$ the difference quotients converge to $x'(t), y'(t)$, and $\frac{\varepsilon}{\Delta t} \to 0$ because ε is sub-linear in the displacement. What survives is exactly the chain rule. ■ SIGNIFICANCE: every related-rates problem, every implicit derivative, every Jacobian change of variables, and every gradient-descent step is this limit argument reused.

Proof: Euler's homogeneous-function theorem

Suppose $f(tx, ty) = t^n f(x, y)$ for all t (homogeneity of degree n). Differentiate BOTH sides with respect to t — the left by the chain rule: $x\, f_x(tx, ty) + y\, f_y(tx, ty)$; the right: $nt^{n-1} f(x, y)$. Set $t = 1$: $xf_x + yf_y = nf$.

■ SIGNIFICANCE: in economics this says total output equals the sum of inputs paid their marginal products (constant returns to scale); in thermodynamics it generates the Gibbs–Duhem relation — one chain-rule differentiation, two sciences.

5.3 Going Deeper: Worked Problems

Worked: a related-rates problem the multivariable way

The radius of a cylinder grows at 2 cm/s while its height shrinks at 1 cm/s. How fast is the volume changing when $r = 3$ and $h = 5$?

Step 1 — $V = \pi r^2 h$ depends on two moving inputs, so $\frac{dV}{dt} = V_r \frac{dr}{dt} + V_h \frac{dh}{dt}$. Step 2 — partials: $V_r = 2\pi rh = 30\pi$ and $V_h = \pi r^2 = 9\pi$ at the given moment. Step 3 — plug in the rates: $\frac{dV}{dt} = 30\pi(2) + 9\pi(-1) = 51\pi$ cm^3/s. Step 4 — interpret: the radius term dominates because volume is quadratic in r but only linear in h — the chain rule made that visible as the sizes of the two partials.

Worked: an implicit surface and its tangent plane

Find the tangent plane to the sphere $x^2 + y^2 + z^2 = 14$ at $(1, 2, 3)$.

Step 1 — write the surface as $F(x, y, z) = x^2 + y^2 + z^2 - 14 = 0$; the normal to a level surface is the gradient $\nabla F = (2x, 2y, 2z)$. Step 2 — evaluate: $\nabla F(1, 2, 3) = (2, 4, 6)$, or just $(1, 2, 3)$ after scaling. Step 3 — the plane through the point with that normal: $1(x-1)+2(y-2)+3(z-3) = 0$, i.e. $x + 2y + 3z = 14$. Step 4 — sanity: for a sphere centered at the origin the normal at a point IS the point's position vector, and 14 on the right equals $|(1, 2, 3)|^2$ — both facts confirm the plane without redoing anything.

6 Optimization & Lagrange Multipliers

Critical points, the second-derivative test, constrained maxima

6.1 Free and Constrained Optimization

Critical points and the D-test

Interior extrema need $\nabla f = \mathbf{0}$. Classify with $D = f_{xx}f_{yy} - f_{xy}^2$: $D > 0$ with $f_{xx} > 0$ is a MIN, $D > 0$ with $f_{xx} < 0$ a MAX, and $D < 0$ a SADDLE — the genuinely new 2-variable phenomenon, downhill one way and uphill another.

Lagrange: gradients align at constrained extrema

To optimize f subject to $g = c$, solve $\nabla f = \lambda \nabla g$: at the best point the level curve of f is TANGENT to the constraint, so the gradients are parallel. Maximizing xy on $x+y = 10$ gives $x = y = 5$ — AM–GM, rediscovered by calculus.

A linear function on a circle

Maximize $x+2y$ on $x^2+y^2=5$: Lagrange forces $(x,y) \parallel (1,2)$, giving $(1,2)$ and the maximum 5. In general $ax+by$ on a radius-r circle peaks at $r\sqrt{a^2+b^2}$ — Cauchy–Schwarz with a geometric proof.

Significance: everything optimal is a critical point

Prices in economics, shapes of soap films, maximum-likelihood estimates in statistics, minimum-energy configurations in chemistry — all sit where a gradient vanishes, usually subject to constraints. Lagrange multipliers even carry meaning: λ is the "shadow price," how much the optimum improves per unit of loosened constraint.

Try it: Try it: a fenced field

Maximize the area xy of a rectangle with $2x+y=40$ (a river covers one side). Work: Lagrange or substitute: $A = x(40-2x)$, $A' = 40-4x = 0$, $x = 10$, $y = 20$, area 200. Note the pattern: the constrained side gets half its "budget" — Lagrange conditions encode it automatically.

6.2 Proofs & Why It Matters

Proof sketch: the second-derivative D-test

Near a critical point, Taylor gives $f \approx f(a,b) + \frac{1}{2}\left(f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2\right)$ — a quadratic form $Q(h,k)$. Complete the square: $Q = f_{xx}\left(h + \frac{f_{xy}}{f_{xx}}k\right)^2 + \frac{D}{f_{xx}}k^2$ with $D = f_{xx}f_{yy} - f_{xy}^2$.

If $D > 0$ both terms share the sign of f_{xx} (a definite bowl: min or max); if $D < 0$ the two terms disagree, so Q takes both signs — a saddle. ■

Proof: why f = g at a constrained extremum

Let $\mathbf{r}(t)$ be any curve inside the constraint surface $g = c$ through the extremum at $t = 0$. Then $t \mapsto f(\mathbf{r}(t))$ has an extremum at 0, so $0 = \frac{d}{dt} f(\mathbf{r}(t))\big|_0 = \nabla f \cdot \mathbf{r}'(0)$.

Thus ∇f is orthogonal to EVERY tangent direction of the constraint — but the vectors orthogonal to that whole tangent plane are exactly the multiples of ∇g. Hence $\nabla f = \lambda \nabla g$. ■

6.3 Going Deeper: Explanations & Worked Problems

Why critical points, and what can go wrong

At an interior maximum or minimum, every slice through the point is a one-variable function with an extremum there, so every directional derivative vanishes — hence $\nabla f = \mathbf{0}$.

But the converse fails in a way one variable never shows: at a SADDLE, the surface rises along one line and falls along another, with $\nabla f = \mathbf{0}$ all the same ($f = x^2 - y^2$ at the origin is the model: uphill along the x-axis, downhill along the y-axis). That is why the D-test exists: near a critical point, f is governed by its quadratic part $\frac{1}{2}(f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2)$, and $D = f_{xx}f_{yy} - f_{xy}^2$ decides whether that quadratic keeps one sign (a genuine bowl or dome) or changes sign (a saddle). And remember the boundary: on a closed region, extrema can sit on the edge with $\nabla f \neq \mathbf{0}$ — that is precisely the situation Lagrange multipliers handle.

Worked: classify every critical point of f = x 3x + y

Step 1 — set the gradient to zero: $f_x = 3x^2 - 3 = 0$ gives $x = \pm 1$; $f_y = 2y = 0$ gives $y = 0$. Critical points: $(1, 0)$ and $(-1, 0)$. Step 2 — second partials: $f_{xx} = 6x$, $f_{yy} = 2$, $f_{xy} = 0$, so $D = 12x$.

Step 3 — test $(1, 0)$: $D = 12 > 0$ and $f_{xx} = 6 > 0$ — both slice curvatures upward: a LOCAL MIN, value $f(1, 0) = 1 - 3 = -2$. Step 4 — test $(-1, 0)$: $D = -12 < 0$ — a SADDLE, no extremum, even though the gradient vanishes. Step 5 — global view: as $x \to -\infty$, x^3 drags $f \to -\infty$, so the local min at $(1, 0)$ is NOT global — a reminder that the D-test speaks only about a neighborhood.

Worked: Lagrange multipliers on x + 2y with x + y = 5, no steps skipped

Step 1 — set up $\nabla f = \lambda \nabla g$: $f = x + 2y$, $g = x^2 + y^2$, so $(1, 2) = \lambda(2x, 2y)$, giving $1 = 2\lambda x$ and $2 = 2\lambda y$. Step 2 — eliminate λ: divide the equations: $\frac{1}{2} = \frac{x}{y}$, so $y = 2x$.

Step 3 — enforce the constraint: $x^2 + (2x)^2 = 5$ gives $5x^2 = 5$, $x = \pm 1$, hence the candidates $(1, 2)$ and $(-1, -2)$. Step 4 — evaluate f at each: $f(1, 2) = 1 + 4 = 5$ and $f(-1, -2) = -5$. The constraint circle is closed and bounded, so these are the global max and min: maximum 5. Step 5 — see it geometrically: the level lines $x + 2y = c$ are parallel lines sliding outward as c grows; the largest c still touching the circle touches it TANGENTLY, at the point where the radius is parallel to $(1, 2)$ — which is exactly what Step 1 said.

7 Double Integrals

Volume by slicing twice

7.1 Iterated Integration

Integrate one variable at a time

$\iint_R f\, dA$ over a rectangle is two ordinary integrals nested: inner in y (holding x), then outer in x — and Fubini says the order does not matter for continuous f. Over non-rectangles, the INNER limits depend on the outer variable: $\int_0^1 \int_0^x \cdots\, dy\, dx$ covers the triangle under $y = x$.

Polar coordinates carry an extra r

On disks and rings, switch to $x = r\cos\theta$, $y = r\sin\theta$ — but the area element becomes $dA = r\,dr\,d\theta$. That extra r (a small polar rectangle is $r\,\Delta\theta$ wide, not $\Delta\theta$) is the single most-forgotten factor in the course. It is also the magic that evaluates $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$.

A separable rectangle

$\iint_{[0,2]\times[0,3]} xy\,dA = \left(\int_0^2 x\,dx\right)\left(\int_0^3 y\,dy\right) = 2\cdot\frac{9}{2} = 9$ — when the integrand factors and the region is a rectangle, the double integral is a product of singles.

Significance: averages, mass, and probability

A double integral is how you total anything spread over a region: mass from density, rainfall from intensity maps, probability from joint densities. The average value of f over a region is $\iint f\,dA$ divided by the area — centroids and centers of mass are just coordinate-weighted versions of that same computation.

Try it: Try it: an average the fast way

Find the average of $f(x,y) = x$ over the rectangle $[0,4]\times[0,2]$. Work: $\iint x\,dA = \left(\int_0^4 x\,dx\right)(2) = 8\cdot 2 = 16$; divide by area 8: average 2. Sanity: x ranges 0 to 4 uniformly, so its average must be the midpoint 2 ✓.

7.2 Proofs & Why It Matters

Proof: dA = r dr d in polar coordinates

A small polar cell between radii r and $r+\Delta r$ and angles θ and $\theta+\Delta\theta$ is nearly a rectangle: its radial side has length Δr, and its circular side is an arc of radius r subtending $\Delta\theta$, of length $r\,\Delta\theta$.

So its area is $\approx r\,\Delta r\,\Delta\theta$ (exactly $\frac{1}{2}\left[(r+\Delta r)^2 - r^2\right]\Delta\theta = r\,\Delta r\,\Delta\theta + \frac{1}{2}\Delta r^2\Delta\theta$, and the second term vanishes in the limit). Summing cells gives $dA = r\,dr\,d\theta$. ■

Proof: the Gaussian integral equals

Let $I = \int_{-\infty}^{\infty} e^{-x^2}\,dx$. Then $I^2 = \iint e^{-(x^2+y^2)}\,dA$ over the plane — switch to polar: $\int_0^{2\pi}\int_0^{\infty} e^{-r^2}\,r\,dr\,d\theta$. The extra r makes the inner integral elementary ($u = r^2$): it equals $\frac{1}{2}$, so $I^2 = 2\pi\cdot\frac{1}{2} = \pi$ and $I = \sqrt{\pi}$. No one-variable method ever evaluates this integral — the plane does. ■

7.3 Going Deeper: Explanations & Worked Problems

Setting up limits: the one skill that decides everything

A double integral over a non-rectangle succeeds or fails at the limits. The procedure: (1) sketch the region; (2) choose an order, say inner y, outer x; (3) for the OUTER variable record the full interval of x values the region occupies; (4) for the INNER variable, freeze x, draw the vertical line through the region, and record where it enters and exits — those (usually x-dependent) heights are the inner limits. For the triangle with vertices $(0,0),(1,0),(1,1)$ (below the line $y=x$): x runs 0 to 1; the vertical line at x enters at $y=0$ and exits at $y=x$, so $\int_0^1\int_0^x \cdots\,dy\,dx$. Reversing the order re-derives, never guesses: a horizontal line at height y enters at $x=y$ and exits at $x=1$, so the same integral is $\int_0^1\int_y^1 \cdots\,dx\,dy$. Fubini guarantees the two agree — and when one order is hard (or impossible, as with $\int e^{y^2}$), swapping is the standard rescue.

Worked: a triangle integral both ways

Compute $\iint_T 2y\,dA$ over the triangle below $y=x$ on $[0,1]$.

Order 1 (inner y): Step 1 — inner: $\int_0^x 2y\,dy = \left[y^2\right]_0^x = x^2$. Step 2 — outer: $\int_0^1 x^2 dx = \frac{1}{3}$. Order 2 (inner x): Step 1 — inner: $\int_y^1 2y\,dx = 2y(1-y)$ (the integrand is constant in x, so multiply by the interval length). Step 2 — outer: $\int_0^1(2y-2y^2)dy = 1-\frac{2}{3}=\frac{1}{3}$. Both orders give $\frac{1}{3}$ — Fubini in action, and a complete template: inner antiderivative, evaluate at the x-dependent limits, then an ordinary outer integral.

Worked: a polar integral with the extra r, start to finish

Compute $\iint_D (x^2+y^2)\,dA$ over the disk of radius 2.

Step 1 — translate to polar: $x^2+y^2=r^2$ and $dA = r\,dr\,d\theta$, so the integrand becomes $r^2\cdot r = r^3$. Step 2 — limits: the full disk is $0\le r\le 2$, $0\le\theta\le 2\pi$. Step 3 — inner integral: $\int_0^2 r^3 dr = \left[\frac{r^4}{4}\right]_0^2 = 4$. Step 4 — outer: $\int_0^{2\pi} 4\,d\theta = 8\pi$. Step 5 — the cautionary rerun: forgetting the extra r gives $\int_0^{2\pi}\int_0^2 r^2 dr\,d\theta = \frac{16\pi}{3}\approx 16.8$ instead of $8\pi\approx 25.1$ — a wrong answer with no warning sign attached. Writing "$dA = r\,dr\,d\theta$" before computing anything is the habit that prevents it.

8 Triple Integrals & Coordinate Systems

Cylindrical and spherical coordinates

8.1 Integrating in Three Dimensions

Three nests, three coordinate systems

Boxes: integrate $dz\,dy\,dx$ straight through. Solids with circular symmetry: CYLINDRICAL coordinates, $dV = r\,dz\,dr\,d\theta$. Balls and cones: SPHERICAL, $dV=\rho^2\sin\phi\,d\rho\,d\phi\,d\theta$. Choosing coordinates to match the region is most of the work.

The sphere, verified

The ball $\rho \le R$ in spherical: $\int_0^{2\pi}\int_0^{\pi}\int_0^{R} \rho^2 \sin\phi \, d\rho \, d\phi \, d\theta = 2\pi \cdot 2 \cdot \frac{R^3}{3} = \frac{4}{3}\pi R^3$ — the middle-school formula, finally proved.

A paraboloid dome

Under $z = 4 - x^2 - y^2$: cylindrical gives $\int_0^{2\pi}\int_0^{2}(4 - r^2)\, r \, dr \, d\theta = 8\pi$. Forgetting the extra r gives a wrong answer that LOOKS plausible — always write the volume element first.

Significance: why three coordinate systems exist

Nature keeps producing round things — planets, pipes, atoms — and Cartesian boxes fit them badly. Cylindrical and spherical coordinates exist because the volume elements r and $\rho^2 \sin\phi$, once understood, turn week-long integrals into three lines. The hydrogen atom of quantum mechanics is solved in spherical coordinates for exactly this reason.

Try it: Try it: a cone by the slice

Volume of the cone z from r up to 1 over the unit disk (i.e. between $z = r$ and $z = 1$). Work: $\int_0^{2\pi}\int_0^{1}(1 - r)\, r \, dr \, d\theta = 2\pi\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{\pi}{3}$ — one third of the enclosing cylinder, the classical cone rule, derived rather than remembered.

8.2 Proofs & Why It Matters

Proof: dV = sin d d d

A small spherical cell has three nearly-perpendicular edges: radial, of length $\Delta\rho$; along a meridian, an arc of radius ρ subtending $\Delta\phi$, length $\rho\,\Delta\phi$; and along a parallel, an arc of radius $\rho \sin\phi$ (the distance to the z-axis) subtending $\Delta\theta$, length $\rho \sin\phi \, \Delta\theta$. The cell volume is the product $\rho^2 \sin\phi \, \Delta\rho \, \Delta\phi \, \Delta\theta$. ■

Proof: the sphere volume formula

Integrate the element over the ball of radius R: $\int_0^{2\pi}\int_0^{\pi}\int_0^{R} \rho^2 \sin\phi \, d\rho \, d\phi \, d\theta$. The three integrals separate: $2\pi \cdot [-\cos\phi]_0^{\pi} \cdot \frac{R^3}{3} = 2\pi \cdot 2 \cdot \frac{R^3}{3} = \frac{4}{3}\pi R^3$. ■

8.3 Going Deeper: Explanations & Worked Problems

Choosing coordinates: read the region, then the integrand

The decision tree: does the region have an axis of circular symmetry? If the boundary involves $x^2 + y^2$ (cylinders, paraboloids, cones over the z-axis), use CYLINDRICAL — polar in the floor plane with z kept, $dV = r\,dz\,dr\,d\theta$.

If distances from a single point rule (balls, spherical shells, cones from the origin), use SPHERICAL: ρ (distance from origin), ϕ (angle down from the north pole, 0 to π), θ (longitude), with $dV = \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$. Then check the integrand cooperates: $x^2+y^2 = r^2$ in cylindrical; $x^2+y^2+z^2 = \rho^2$ and $z = \rho\cos\phi$ in spherical. A sphere in Cartesian coordinates costs three nested square-root limits; the SAME sphere in spherical coordinates is a box $[0, R] \times [0, \pi] \times [0, 2\pi]$ — coordinates are chosen to make the region a box, because boxes separate.

Worked: the paraboloid volume in cylindrical, every step

Volume under $z = 4 - x^2 - y^2$ above the xy-plane.

Step 1 — find the floor region: $z = 0$ forces $x^2 + y^2 = 4$, the disk of radius 2. Step 2 — write in cylindrical: z runs from 0 up to $4 - r^2$; the volume element contributes the extra r. Step 3 — innermost (z): $\int_0^{4-r^2} dz = 4 - r^2$. Step 4 — middle ($r$): $\int_0^2 (4 - r^2)\,r\,dr = \int_0^2 (4r - r^3)dr = \left[2r^2 - \frac{r^4}{4}\right]_0^2 = 8 - 4 = 4$. Step 5 — outer ($\theta$): $\int_0^{2\pi} 4\,d\theta = 8\pi$. Step 6 — sanity: the enclosing cylinder (radius 2, height 4) has volume 16π; the paraboloid fills exactly HALF of it — a clean known fact that confirms the arithmetic.

Worked: an off-center integrand in spherical

Compute $\iiint_B z^2\,dV$ over the ball $\rho \le 1$.

Step 1 — translate: $z = \rho\cos\phi$, so $z^2 = \rho^2\cos^2\phi$, and with $dV = \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$ the integrand is $\rho^4\cos^2\phi\sin\phi$. Step 2 — separate the three integrals: $\left(\int_0^1 \rho^4 d\rho\right)\left(\int_0^\pi \cos^2\phi\sin\phi\,d\phi\right)\left(\int_0^{2\pi} d\theta\right)$. Step 3 — evaluate each: $\int_0^1 \rho^4 d\rho = \frac{1}{5}$; for the middle, substitute $u = \cos\phi$, $du = -\sin\phi\,d\phi$: $\int_{-1}^1 u^2 du = \frac{2}{3}$; the last is 2π. Step 4 — multiply: $\frac{1}{5} \cdot \frac{2}{3} \cdot 2\pi = \frac{4\pi}{15}$. Step 5 — symmetry check: by symmetry $\iiint x^2 = \iiint y^2 = \iiint z^2$, and the three must add to $\iiint \rho^2\,dV = \int_0^1 \rho^4 d\rho \cdot 4\pi = \frac{4\pi}{5}$; indeed $3 \cdot \frac{4\pi}{15} = \frac{4\pi}{5}$ ✓.

9 Change of Variables & the Jacobian

How substitutions stretch area and volume

9.1 The Jacobian Determinant

Substitution in several variables

One-variable substitution carries a factor $\frac{dx}{du}$; in several variables the factor is a DETERMI-

NANT: for a map $(u,v) \mapsto (x(u,v), y(u,v))$, $dx\,dy = \left|\det \dfrac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$, where the Jacobian matrix collects all four partials.

The reason is geometric: the map sends a tiny $du \times dv$ rectangle to a tiny parallelogram spanned by the column vectors (x_u, y_u) and (x_v, y_v), and a parallelogram's area is exactly the absolute determinant. Every coordinate system you have used is a special case: polar's factor r, cylindrical's r, spherical's $\rho^2 \sin\phi$ — all Jacobians, all computable in two minutes from the definition rather than memorized.

Which direction is the determinant?

The classic trap: if you define the NEW variables from the old ($u = x+y$, $v = x-y$), you naturally compute det $\frac{\partial(u,v)}{\partial(x,y)}$ — but the integral needs the INVERSE factor $\left|\det \frac{\partial(x,y)}{\partial(u,v)}\right|$, its reciprocal. Rule of thumb: the Jacobian you multiply by always has the OLD variables on top. When in doubt, test on a unit square.

Try it: Try it: the area of an ellipse in one line

Find the area of $\frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1$. Work: the map $x = au$, $y = bv$ turns it into the unit disk $u^2 + v^2 \le 1$; the Jacobian is det $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = ab$; area $= ab \times \pi = \pi ab$. Setting $a = b = r$ recovers πr^2 — the circle formula was the special case all along.

9.2 Proofs & Why It Matters

Proof: why the determinant appears

Near a point, the substitution map is approximated by its derivative — the linear map with the Jacobian matrix J (this is the tangent-plane idea applied to a MAP rather than a function).

A linear map multiplies all areas by $|\det J|$: the unit square spanned by $\mathbf{e}_1, \mathbf{e}_2$ lands on the parallelogram spanned by the columns of J, whose area is $|\det J|$, and general regions are limits of little squares. Summing $f \times$ (image area) over a fine grid gives Riemann sums of $\iint f\,|\det J|\,du\,dv$. ■ SIGNIFICANCE: this is the bridge between linear algebra ("determinant = volume factor") and calculus ("substitution rule") — and in probability it is exactly how densities transform, the change-of-variables formula behind every normalizing flow in machine learning.

Proof: polar coordinates, this time with no hand-waving

$x = r\cos\theta$, $y = r\sin\theta$: $J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$, $\det J = r\cos^2\theta + r\sin^2\theta = r$.

The "curvy rectangle" argument from the double-integrals topic was the geometric shadow of this two-line computation — and now spherical's $\rho^2 \sin\phi$ is just a bigger determinant, not a new idea to memorize. ■

9.3 Going Deeper: Worked Problems

Worked: an integral that only the right substitution can do

Compute $\iint_R (x+y)^2\,dA$ over the square R with vertices $(1,0),(0,1),(-1,0),(0,-1)$.

Step 1 — the region is a diamond, awkward in x,y; substitute $u=x+y$, $v=x-y$, which maps it to the square $|u|\le 1$, $|v|\le 1$. Step 2 — Jacobian: $\dfrac{\partial(u,v)}{\partial(x,y)} = \det\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = -2$, so $dx\,dy = \frac{1}{2}\,du\,dv$. Step 3 — the integrand is u^2: $\displaystyle\int_{-1}^{1}\int_{-1}^{1} u^2\cdot\tfrac{1}{2}\,du\,dv = \tfrac{1}{2}\cdot\tfrac{2}{3}\cdot 2 = \tfrac{2}{3}$. Step 4 — check plausibility: the diamond has area 2 and $(x+y)^2$ averages less than 1 on it, so an answer below 2 is right-sized; the exact value $\frac{2}{3}$ took three lines because the substitution matched BOTH region and integrand.

Worked: deriving the cylindrical volume element

Show that $dV = r\,dr\,d\theta\,dz$ for $x=r\cos\theta$, $y=r\sin\theta$, $z=z$.

Step 1 — the Jacobian matrix of $(r,\theta,z)\mapsto(x,y,z)$ is $\begin{pmatrix}\cos\theta & -r\sin\theta & 0\\ \sin\theta & r\cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$. Step 2 — expand along the last row: the determinant is $1\cdot(r\cos^2\theta + r\sin^2\theta) = r$. Step 3 — so $dV = |r|\,dr\,d\theta\,dz = r\,dr\,d\theta\,dz$ since $r\ge 0$. Step 4 — the geometric reading: the z-direction is untouched, so the 3D element is just the polar area element $r\,dr\,d\theta$ times dz — the same determinant computed in the Jacobian topic, with one extra dimension riding along for free.

10 Vector Calculus

Line integrals, Green's theorem, div and curl

10.1 Fields, Circulation, and Flux

Conservative fields have potentials

If $\mathbf{F} = \nabla f$ then $\int_C \mathbf{F}\cdot d\mathbf{r} = f(\text{end}) - f(\text{start})$ — the fundamental theorem for line integrals: path-independent, and zero around every loop. Test: in the plane, $\mathbf{F} = (P,Q)$ is conservative when $Q_x = P_y$.

Green's theorem trades a loop for a region

$\oint_C P\,dx + Q\,dy = \iint_R (Q_x - P_y)\,dA$ for a counterclockwise loop. The famous special case $\frac{1}{2}\oint(-y\,dx + x\,dy) = $ area powers planimeters and the shoelace formula alike.

Divergence and curl, in one breath

$\nabla \cdot \mathbf{F}$ measures how much the field SPREADS from a point (source strength); $\nabla \times \mathbf{F}$ measures how much it ROTATES. Conservative fields are curl-free; incompressible flows are divergence-free — and the Stokes/divergence theorems extend Green to surfaces and solids.

Significance: conservation laws live here

Green's theorem and its 3D siblings are the mathematical form of 'what is created inside crosses the boundary.' Fluid flow, heat, electric charge — every conservation law in physics is one of these theorems applied to the right field. The planimeter (a 19th-century gadget that measures area by tracing a boundary) is Green's theorem built in brass.

Try it: Try it: spot the conservative field

Is $\mathbf{F} = (2xy, x^2)$ conservative? Work: $P_y = 2x$ and $Q_x = 2x$ — equal, so yes; potential $f = x^2 y$ (check $\nabla f = (2xy, x^2)$ ✓). So $\int \mathbf{F} \cdot d\mathbf{r}$ from $(0,0)$ to $(3,2)$ along ANY path is $f(3,2) = 18$.

10.2 Proofs & Why It Matters

Proof: the fundamental theorem for line integrals

If $\mathbf{F} = \nabla f$ and C is parametrized by $\mathbf{r}(t)$, $a \le t \le b$, then $\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\, dt$.

By the chain rule the integrand is exactly $\frac{d}{dt} f(\mathbf{r}(t))$, so the integral is $f(\mathbf{r}(b)) - f(\mathbf{r}(a))$ by the ordinary fundamental theorem of calculus. Only the endpoints survive. ■

Proof sketch: Green's theorem on a rectangle

On $[a,b] \times [c,d]$: $\iint Q_x\, dA = \int_c^d [Q(b,y) - Q(a,y)]\, dy$, which is the line integral of $Q\, dy$ up the right side and down the left. Similarly $-\iint P_y\, dA$ produces $P\, dx$ along the bottom and top. Adding gives the counterclockwise loop integral.

A general region is tiled by small rectangles: interior edges are traversed twice in opposite directions and cancel, leaving only the outer boundary. ■

Proof: the (x dy y dx) area formula

Apply Green's theorem with $P = -\frac{y}{2}$, $Q = \frac{x}{2}$: then $Q_x - P_y = \frac{1}{2} + \frac{1}{2} = 1$, so the loop integral equals $\iint 1\, dA$ — the enclosed area. Choosing polygon vertices as the path gives the shoelace formula as a corollary. ■

10.3 Going Deeper: Explanations & Worked Problems

Conservative or not: the full workflow

Given $\mathbf{F} = (P, Q)$, the question "does a potential exist?" has a mechanical answer. Test: compute Q_x and P_y; if they differ, no potential, stop.

If they agree (on a region without holes), CONSTRUCT the potential: integrate P in x to get $f = \int P\,dx + g(y)$ with an unknown function of y alone; differentiate this f in y, set it equal to Q, and solve for $g'(y)$; integrate once more. Example: $\mathbf{F} = (2xy, ; x^2 + 3y^2)$. Check: $P_y = 2x = Q_x$ ✓. Integrate: $f = x^2y + g(y)$. Match: $f_y = x^2 + g'(y) \stackrel{!}{=} x^2 + 3y^2$, so $g'(y) = 3y^2$ and $g = y^3$. Potential: $f = x^2y + y^3$, and every line integral of $\mathbf{F}$ is now just f at the endpoints. The hole caveat is real: $\left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$ passes the derivative test everywhere it is defined yet has integral 2π around the origin — the missing point matters.

Worked: verifying Green's theorem on the unit circle

Take $\mathbf{F} = (-y, x)$ around the unit circle, counterclockwise. LEFT SIDE (the line integral, done honestly):

Step 1 — parametrize: $\mathbf{r}(t) = (\cos t, \sin t)$, $0 \le t \le 2\pi$, so $dx = -\sin t\,dt$, $dy = \cos t\,dt$. Step 2 — substitute: $\oint(-y\,dx + x\,dy) = \int_0^{2\pi} [(-\sin t)(-\sin t) + (\cos t)(\cos t)]\,dt = \int_0^{2\pi} (\sin^2 t + \cos^2 t)\,dt$. Step 3 — the identity collapses the integrand to 1: the integral is 2π. RIGHT SIDE (the double integral): $Q_x - P_y = 1 - (-1) = 2$, and $\iint_{\text{disk}} 2\,dA = 2\pi$. The two sides agree — and dividing by 2 gives the area formula $\frac{1}{2}\oint(x\,dy - y\,dx) = \pi$, the area of the unit disk, computed purely from its boundary.

Worked: divergence and curl side by side

Let $\mathbf{F} = (x^2, ; xy, ; z^2)$. DIVERGENCE: differentiate each component by ITS OWN variable and add: $\partial_x(x^2) + \partial_y(xy) + \partial_z(z^2) = 2x + x + 2z = 3x + 2z$ — at $(1, 2, 3)$ this is $9 > 0$, so the field is expanding there (a net source).

CURL: the determinant recipe gives $\nabla \times \mathbf{F} = \big(\partial_y(z^2) - \partial_z(xy), ; \partial_z(x^2) - \partial_x(z^2), ; \partial_x(xy) - \partial_y(x^2)\big) = (0 - 0, ; 0 - 0, ; y - 0) = (0, 0, y)$. Nonzero curl means $\mathbf{F}$ is NOT conservative — no potential exists, and loop integrals of $\mathbf{F}$ can be nonzero — even though two of the three components look innocent. One derivative test settles what no amount of staring at the formula would.

11 Surface Integrals, Stokes & Divergence

Flux through surfaces and the two capstone theorems

11.1 The Capstone Theorems

Flux: how much field crosses a surface

The surface integral $\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_S \mathbf{F} \cdot \mathbf{n}\,dS$ measures net flow through S: only the component of $\mathbf{F}$ along the unit normal $\mathbf{n}$ counts.

For a FLAT surface with constant field this collapses to (normal component) $\times$ (area) — the flux of $(0, 0, 5)$ up through a radius-3 disk is $5 \cdot 9\pi = 45\pi$, no machinery required. For curved surfaces, parametrize and integrate; for CLOSED surfaces there is usually a shortcut, and it is the next box.

The divergence theorem and Stokes' theorem

DIVERGENCE THEOREM: outward flux through a closed surface equals $\iiint \nabla \cdot \mathbf{F}\, dV$ over the enclosed solid — total outflow equals the sum of all the little sources inside.

STOKES: circulation $\oint \mathbf{F} \cdot d\mathbf{r}$ around a closed curve equals the flux of $\nabla \times \mathbf{F}$ through ANY surface spanning it — total rotation around the rim equals the sum of the little whirlpools on the membrane. Green's theorem is Stokes flattened into the plane; the fundamental theorem of calculus is the 1-dimensional case of everything. One slogan covers the whole family: the integral of a derivative over a region equals the original quantity summed over the boundary.

Why physicists write Maxwell's equations twice

Each of Maxwell's equations has a differential form (about $\nabla\cdot$ and $\nabla\times$ at a point) and an integral form (about flux and circulation over regions) — and the divergence and Stokes theorems are precisely the dictionary between them. Gauss's law "flux of $\mathbf{E}$ = enclosed charge" IS the divergence theorem applied to $\nabla \cdot \mathbf{E} = \rho/\varepsilon_0$. Learning these two theorems is learning to read electromagnetism.

Try it: Try it: flux through a sphere without parametrizing anything

Find the outward flux of $\mathbf{F} = (x, y, z)$ through the sphere of radius 2. Work: $\nabla \cdot \mathbf{F} = 3$, so flux $= 3 \times \text{volume} = 3 \cdot \frac{4}{3}\pi \cdot 8 = 32\pi$.

Answer. Cross-check on the unit sphere: there $\mathbf{F} = \mathbf{n}$ on the surface, so flux = area = 4π — and the theorem gives $3 \cdot \frac{4}{3}\pi = 4\pi$ ✓. When the divergence is constant, flux problems are volume problems.

11.2 Proofs & Why It Matters

Proof sketch: the divergence theorem on a box

On a box, split the flux into three pairs of opposite faces. For the z-pair: $\iint [R(x, y, z_{\text{top}}) - R(x, y, z_{\text{bot}})]\, dA = \iiint \partial_z R\, dz\, dA$ by the one-variable fundamental theorem — the flux difference across the pair equals the integral of $\partial R/\partial z$ through the inside. Adding the three pairs assembles $\iiint (\partial_x P + \partial_y Q + \partial_z R)\, dV$. A general solid is tiled by small boxes: fluxes across interior walls cancel in pairs (shared walls, opposite normals), leaving the outer boundary. ■ SIGNIFICANCE: "interior contributions cancel, only the boundary survives" is the single deepest pattern in integration — it is why conservation laws exist in physics.

The grand unification

Line up the family: FTC ($\int_a^b f' = f(b) - f(a)$), FTL for gradients, Green, Stokes, divergence. Each says: derivative integrated over the inside = function summed over the boundary. In advanced mathematics all five become ONE statement about differential forms ($\int_\Omega d\omega = \int_{\partial\Omega} \omega$, the generalized Stokes theorem) — this course has been climbing the rungs of a single ladder the whole time.

11.3 Going Deeper: Worked Problems

Worked: flux through a closed box with the divergence theorem

Find the outward flux of $\mathbf{F} = (x^2, y, z)$ through the surface of the box $[0, 2] \times [0, 1] \times [0, 3]$.

Step 1 — divergence: $\nabla \cdot \mathbf{F} = 2x + 1 + 1 = 2x + 2$. Step 2 — integrate over the box: $\displaystyle\int_0^3 \int_0^1 \int_0^2 (2x + 2)\, dx\, dy\, dz$. The inner integral is $\left[x^2 + 2x\right]_0^2 = 8$. Step 3 — the outer two multiply by the remaining side lengths: $8 \cdot 1 \cdot 3 = 24$. Step 4 — direct check of one face pair: on $x = 2$ the flux is $\iint 4\, dA = 4 \cdot 3 = 12$ outward, on $x = 0$ it is 0; the y-faces contribute $1 \cdot (2 \cdot 3) = 6$ net and the z-faces $3 \cdot 2 - 0 = 6$: total $12 + 6 + 6 = 24$ ✓. Six face integrals or one volume integral — the theorem lets you choose.

Worked: Stokes' theorem picking the easy surface

Compute $\displaystyle\oint_C \mathbf{F} \cdot d\mathbf{r}$ for $\mathbf{F} = (-y, x, z^2)$ around the circle $x^2 + y^2 = 4$ in the plane $z = 1$, counterclockwise from above.

Step 1 — curl: $\nabla \times \mathbf{F} = \big(\partial_y(z^2) - \partial_z(x), ; \partial_z(-y) - \partial_x(z^2), ; \partial_x(x) - \partial_y(-y)\big) = (0, 0, 2)$. Step 2 — cap the circle with the flat disk $z = 1$, upward normal $(0, 0, 1)$: the flux of the curl is $\iint 2\, dA = 2 \cdot \pi \cdot 4 = 8\pi$. Step 3 — by Stokes, the circulation is 8π. Step 4 — verify directly: parametrize $\mathbf{r} = (2\cos t, 2\sin t, 1)$; then $\mathbf{F} \cdot \mathbf{r}' = (-2\sin t)(-2\sin t) + (2\cos t)(2\cos t) + 1 \cdot 0 = 4$, and $\int_0^{2\pi} 4\, dt = 8\pi$ ✓. The z^2 component never mattered — its curl contribution was zero, and Stokes said so before any parametrizing.