

Derivatives in Calculus

[Read this note online](#)

The **derivative** in calculus is the instantaneous rate of change of a function. Geometrically, it's the slope of the tangent line at a point. Physically, it's velocity if the function is position, or acceleration if the function is velocity. Economically, it's marginal cost or marginal revenue. The derivative shows up anywhere a quantity changes and you want to know how fast.

Most of single-variable calculus is built on three ideas: the limit definition, a small set of rules for differentiating common functions, and the chain rule. Memorize those, and 90% of derivative problems become routine. The interesting work happens at the edges, non-differentiable points, optimization at constraint boundaries, and the derivative-driven core of modern machine learning and physics.

A derivative records the local change in a function. The sign, magnitude and units are all part of interpreting it.

DERIVATIVES IN CALCULUS

THE DERIVATIVE measures the instantaneous rate of change of a function.

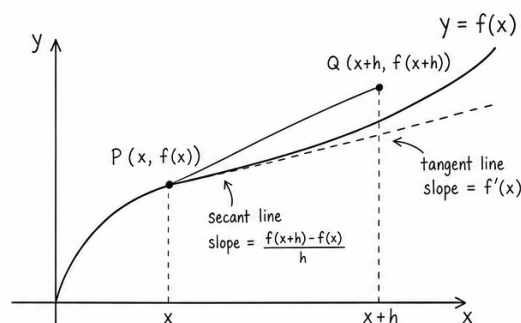
$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

COMMON NOTATIONS

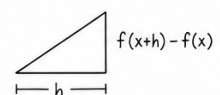
- $\frac{dy}{dx}$ → Leibniz notation
- $f'(x)$ → Prime notation
- $Df(x)$ → Operator notation ("D" stands for derivative)

INTERPRETATION

$f'(x)$ is the slope of the tangent line to $y = f(x)$ at the point $(x, f(x))$. It gives the instantaneous rate of change at x .



SLOPE TRIANGLE



$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{f(x+h) - f(x)}{h}$$

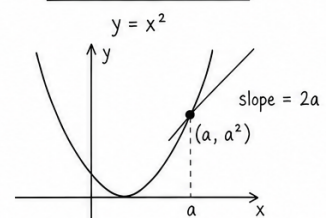
EXAMPLE: $f(x) = x^2$

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \\ &= \lim_{h \rightarrow 0} (2x + h) \\ &= 2x \end{aligned}$$

Therefore,

$$\frac{d}{dx} x^2 = 2x$$

TANGENT EXAMPLE



The tangent line at $x = a$ is $y = 2a(x-a) + a^2$

The Limit Definition

The derivative of f at x is:

The fraction is the average rate of change between x and $x + h$, the slope of the secant line through the two points. As h shrinks toward zero, the secant becomes the tangent, and the average rate becomes the instantaneous rate.

The derivative exists at x only when this limit exists. Sharp corners, vertical tangents, and discontinuities all kill differentiability. The function $f(x) = |x|$ is continuous everywhere but not differentiable at $x = 0$, where the left and right slopes disagree (-1 and +1, respectively).

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Notation Conventions

The same idea has many costumes:

Pick whichever the textbook uses; they all mean the same thing. Different fields developed different conventions because they emphasized different uses. Leibniz notation makes the **chain rule** feel like fraction cancellation. Newton's dot notation keeps physics equations compact. Lagrange's prime notation is the cleanest for general algebraic work.

Lagrange: $f'(x)$, $f''(x)$, $f^{(n)}(x)$, compact for general use **Leibniz:** $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$, clear about what's being differentiated with respect to what; mandatory for the chain rule **Newton:** $\dot{y}$, $\ddot{y}$, mostly physics, dots mean derivatives with respect to time **Operator:** $D_x f$ or $\frac{d}{dx}f$, useful in differential equations

The Rules You'll Actually Use

From the limit definition you derive a small toolkit:

Plus the standard library: $\frac{d}{dx} \sin x = \cos x$, $\frac{d}{dx} \cos x = -\sin x$, $\frac{d}{dx} e^x = e^x$, $\frac{d}{dx} \ln x = \frac{1}{x}$, $\frac{d}{dx} \tan x = \sec^2 x$. Memorize these and you can differentiate nearly anything algebraically.

Power rule: $\frac{d}{dx} x^n = nx^{n-1}$ for any real n **Constant rule:** $\frac{d}{dx} c = 0$ **Sum rule:** $(f+g)' = f' + g'$ **Product rule:** $(fg)' = f'g + fg'$ **Quotient rule:** $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ **Chain rule:** $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$

Worked Examples

Polynomial: Differentiate $f(x) = 3x^4 - 2x^2 + 5x - 7$:

Power rule term by term, constants drop out. Bread-and-butter polynomial differentiation.

Product: Differentiate $g(x) = x^2 \sin x$:

Product rule with $f = x^2$ and $g = \sin x$.

Chain: Differentiate $h(x) = \sin(x^2)$:

Outer derivative $\cos(\cdot)$ evaluated at the inner function, multiplied by inner derivative $2x$.

$$f'(x) = 12x^3 - 4x + 5$$

$$g'(x) = 2x \sin x + x^2 \cos x$$

$$h'(x) = \cos(x^2) \cdot 2x = 2x \cos(x^2)$$

What the Derivative Tells You

This is the toolkit behind optimization: find where $f' = 0$, check the second derivative, classify each critical point. It's also how machine learning gradient descent works under the hood, read more in [best machine learning courses](#) and [best data science courses](#).

Sign of f' : positive means f is increasing, negative means decreasing. $f' = 0$: a critical point, possible local max, local min, or saddle. **Sign of f'' :** positive means f is concave up, negative means concave down. $f'' = 0$ **with sign change:** inflection point.

The Second Derivative Test

For a critical point c where $f'(c) = 0$:

The test is fast for well-behaved functions. For $f(x) = x^3$, the only critical point is at $x = 0$; $f''(0) = 0$, so the second-derivative test is inconclusive. Inspecting the sign of f' reveals $x = 0$ is neither a max nor a min, it's an inflection point where the tangent line crosses the curve.

For multivariable functions, the second-derivative test generalizes to the Hessian matrix. The **eigenvalues** of the Hessian classify critical points: all positive means local min, all negative means local max, mixed signs mean saddle. Quasi-Newton optimization methods (BFGS, L-BFGS) approximate the inverse Hessian to choose efficient search directions.

If $f''(c) > 0$, then c is a local minimum. If $f''(c) < 0$, then c is a local maximum. If $f''(c) = 0$, the test is inconclusive, use the first-derivative sign chart or a higher-order test.

Implicit Differentiation

When a relationship between x and y isn't explicitly solved for y , differentiate both sides with respect to x , treating y as a function of x and applying the chain rule wherever y appears.

Example: differentiate $x^2 + y^2 = 25$ (the equation of a circle).

Implicit differentiation handles equations that can't be solved cleanly for y , like $\sin(xy) + e^{x+y} = x^2$. It also produces derivatives of inverse functions efficiently, the derivative of $\arcsin x$ is derived this way from $\sin y = x$.

$$2x + 2y \frac{dy}{dx} = 0 \implies \frac{dy}{dx} = -\frac{x}{y}$$

Higher-Order Derivatives

The second derivative is the derivative of the derivative. The third derivative is the derivative of the second. And so on.

Higher-order derivatives appear in Taylor series, where they encode the local behavior of a function around a point. They also matter in physics: position is the original quantity, velocity is its derivative, acceleration is the second derivative, jerk is the third, snap is the fourth (yes, those are real names).

If $f(x) = \sin x$, then $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x$, $f^{(4)}(x) = \sin x$. The pattern repeats with period 4. For $f(x) = e^x$, all derivatives equal e^x . For polynomials, the n -th derivative of x^n is $n!$, and all higher derivatives are zero.

Differentiability vs Continuity

Differentiability is strictly stronger than continuity. Every differentiable function is continuous, but not every continuous function is differentiable.

Counterexamples:

For practical work, almost every function you'll meet is differentiable except at isolated points. Software-defined functions (neural network activations, splines, decision tree predictions) have known non-differentiable points that engineers handle deliberately.

Sharp corners: $|x|$ is continuous everywhere but not differentiable at 0. **Vertical tangents:** $x^{1/3}$ has an infinite-slope tangent at 0, so the derivative doesn't exist there. **Cusps:** functions like $x^{2/3}$ have a cusp at 0, also non-differentiable. **Pathological cases:** the Weierstrass function is continuous everywhere but differentiable nowhere, a deeply counterintuitive example that motivated 19th-century rigor.

Where Derivatives Live in Practice

Physics: velocity, acceleration, force, related rates, Lagrangian and Hamiltonian mechanics. Most differential equations describe how physical quantities change over time or space. **Economics:** marginal cost, marginal revenue, marginal utility, elasticity. Almost every microeconomic optimization is a derivative-based problem. **Engineering:** stress-strain curves, signal processing (the derivative of position gives velocity, of voltage gives current through capacitors), control systems. **Machine learning:** gradient descent, backpropagation, automatic differentiation. Modern deep learning would not exist without efficient derivative computation across millions of parameters. **Probability and statistics:** probability density functions are derivatives of cumulative distribution functions; maximum likelihood estimation uses derivatives to find optimal parameters. **Finance:** option Greeks (delta, gamma, theta, vega) are derivatives of option prices with respect to underlying parameters.

Optimization in One Dimension

To find local maxima and minima of $f(x)$ on an interval:

This is the standard recipe in calculus textbooks and the foundation for nearly all classical optimization. Real-world problems usually have multiple local optima; finding the global optimum requires either exhaustive search of critical points or specialized algorithms (gradient descent with random restarts, simulated annealing, branch-and-bound for discrete cases).

Compute $f'(x)$. Find critical points where $f'(x) = 0$ or f' is undefined. Apply the second-derivative test or first-derivative sign chart to classify each critical point. Evaluate f at the critical points and the interval endpoints. The largest value is the absolute max; the smallest is the absolute min.

Numerical Differentiation

When you can't differentiate a function symbolically (no closed form, defined by data, or built from simulation), numerical differentiation approximates derivatives from function values.

Forward difference: $f'(x) \approx (f(x+h) - f(x))/h$. Simple but only first-order accurate.

Central difference: $f'(x) \approx (f(x+h) - f(x-h))/(2h)$. Second-order accurate, almost always preferred when both $x+h$ and $x-h$ are accessible.

Higher-order finite-difference schemes use more sample points and achieve fourth- or sixth-order accuracy. Automatic differentiation, used in modern ML frameworks, computes exact derivatives by tracking computational graphs and applying the chain rule symbolically, neither numerical nor purely symbolic, but a third approach that's faster and more accurate than both.

A Brief History of Derivatives

Newton developed his version of calculus around 1665 to 1666, calling derivatives "fluxions" and using dot notation that survives in physics. Leibniz developed his version independently in the 1670s, using dy/dx notation that survives in nearly every other field. The bitter priority dispute that followed slowed mathematical progress in England for nearly a century, while continental Europe (using Leibniz's notation) raced ahead.

The 19th century made derivatives rigorous. Cauchy and Weierstrass replaced infinitesimal arguments with limit-based definitions. By the 20th century, the limit-based derivative was the standard in textbooks worldwide. Nonstandard analysis in the 1960s rehabilitated infinitesimals on rigorous grounds, but the standard $\varepsilon - \delta$ approach kept its dominance in education.

The modern era of derivatives is computational. Automatic differentiation transformed scientific computing in the 1980s-2000s and is now the foundation of every deep learning framework. Symbolic differentiation in computer algebra systems (Mathematica, SymPy) handles textbook problems instantly. The pencil-and-paper era is largely over for routine work; the conceptual understanding still matters because every technique builds on it.

Mean Value Theorem

If f is continuous on $[a, b]$ and differentiable on (a, b) , there exists $c \in (a, b)$ such that:

Geometrically, somewhere between a and b the tangent line is parallel to the secant line. This is the workhorse theorem behind most existence results in single-variable calculus and the foundation for the Fundamental Theorem of Calculus.

Practical applications: if a car travels 100 miles in 2 hours, the MVT guarantees that at some moment the speedometer read exactly 50 mph. Speed cameras and average-speed enforcement use this principle: if your average speed exceeded the limit, MVT guarantees you exceeded the limit at some instant.

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Derivatives of Inverse Functions

If f has an inverse f^{-1} and $f'(f^{-1}(x)) \neq 0$, then:

This formula derives the standard inverse function derivatives. $\frac{d}{dx} \arcsin x = 1/\sqrt{1-x^2}$, $\frac{d}{dx} \arctan x = 1/(1+x^2)$, and $\frac{d}{dx} \ln x = 1/x$ all come from applying this rule to standard functions.

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

Linear Approximation and Differentials

Near a point a , a function is approximately linear: $f(x) \approx f(a) + f'(a)(x - a)$. This is the linear approximation, and the difference $dy = f'(a)dx$ is called the differential.

Linear approximation is the basis for Newton's method (root-finding), error propagation in measurements, and first-order numerical analysis. The same idea generalizes to multivariable Taylor expansion and is the foundation for differential geometry.

Symbolic vs Numerical Differentiation

Symbolic differentiation manipulates algebraic expressions to produce closed-form derivatives. Computer algebra systems (Mathematica, SymPy, Maple) do this automatically. The output is exact and human-readable but can grow very large for complex expressions.

Numerical differentiation estimates derivatives from function evaluations using finite-difference formulas. It's faster for one-off evaluations but accumulates floating-point error and only delivers approximate results.

Automatic differentiation, used in modern ML frameworks, is a third approach: it tracks the computational graph and applies the chain rule symbolically to each node, producing exact derivatives without ever simplifying the algebraic expression. AD is faster than symbolic for complex models and more accurate than numerical differentiation.

Why Derivatives Underpin Modern Machine Learning

Every neural network training step involves computing gradients of a loss function with respect to model parameters. With models running into hundreds of billions of parameters, the engineering challenge is computing these derivatives efficiently. Backpropagation, automatic differentiation, and frameworks like PyTorch and JAX exist precisely to solve this problem at scale.

The same derivative-driven optimization powers reinforcement learning policy updates, generative model training, and probabilistic programming inference. Without derivatives, none of modern AI exists. The classical calculus you learn in college is the conceptual core of trillion-dollar industries.

Partial Derivatives Quick Look

For multivariable functions $f(x, y)$, the partial derivative $\partial f/\partial x$ treats y as a constant and differentiates with respect to x . The other partial $\partial f/\partial y$ does the opposite. Together they generalize the single-variable derivative to surfaces and higher-dimensional functions, and they're the building blocks for gradients, divergence, curl, and the Hessian matrix in multivariable calculus.

Partial derivatives are essential in optimization, gradient descent (each parameter gets its own partial derivative for the update step), thermodynamics (state variables change with respect to each independent input), and economics (marginal effects of one input while others are held fixed).

The related [OpenStax mathematics textbooks](#) provide additional explanations and exercises for the mathematical definitions used here.

A Stationary Point

Let $s(t) = t^3 - 6t^2 + 9t$ describe position in meters, with t in seconds. Its velocity is $v(t) = 3t^2 - 12t + 9 = 3(t - 1)(t - 3)$. The velocity is zero at 1 s and 3 s.

Checking signs shows positive velocity before 1 s, negative velocity between 1 s and 3 s, and positive velocity afterward. The object changes direction at both stationary times. A zero derivative alone does not guarantee a direction change: $s(t) = t^3$ has zero derivative at zero while increasing on both sides.

The points to check in this example are:

- Differentiating position gives velocity, with units m/s.
- Differentiating velocity gives acceleration, with units m/s².
- A stationary point requires classification using nearby behavior or another valid test.

Check Your Understanding

Try the questions before reading the answers.

Question 1. What is the acceleration at 2 s in the worked example?

Answer. $6t - 12$ gives 0 m/s².

Question 2. Must every differentiable local extremum have zero derivative at an interior point?

Answer. Yes, but the converse does not hold.

Practice Questions

Work each question before reading its solution. The set runs from direct recall and substitution to the applied questions that exams actually use to separate grades. The downloadable PDF includes these questions and their solutions.

Question 1. Differentiate $f(x) = x^5$.

Solution. Power rule: bring the exponent down, reduce it by 1. $f'(x) = 5x^4$.

Question 2. Differentiate $f(x) = 3x^4 - 2x^2 + 7x - 5$.

Solution. Term by term: $12x^3 - 4x + 7$. Constants vanish, and the derivative of $7x$ is just 7.

Question 3. Differentiate $f(x) = x^2 \sin x$.

Solution. Product rule, first times derivative of second plus second times derivative of first: $f'(x) = 2x \sin x + x^2 \cos x$.

Question 4. Differentiate $f(x) = \frac{x}{x^2 + 1}$.

Solution. Quotient rule: $f'(x) = \frac{(x^2 + 1)(1) - x(2x)}{(x^2 + 1)^2} = \frac{1 - x^2}{(x^2 + 1)^2}$. The derivative is zero at $x = \pm 1$, exactly where this function peaks and bottoms.

Question 5. Differentiate $f(x) = (3x^2 + 1)^5$.

Solution. Chain rule, outside first, then inside: $f'(x) = 5(3x^2 + 1)^4 \cdot 6x = 30x(3x^2 + 1)^4$. Forgetting the inner factor $6x$ is the most common chain-rule error.

Question 6. Find the tangent line to $y = x^2$ at $x = 3$.

Solution. Slope: $y' = 2x = 6$ at $x = 3$; point: $(3, 9)$. Line: $y - 9 = 6(x - 3)$, so $y = 6x - 9$. The tangent is the straight line the curve momentarily agrees with.

Question 7. A dropped object falls $s(t) = 4.9t^2$ meters in t seconds. Find its speed at $t = 2$.

Solution. $v(t) = s'(t) = 9.8t$, so $v(2) = 19.6$ m/s. The derivative of position is velocity; differentiating again gives the constant acceleration 9.8 m/s^2 , gravity itself.

Question 8. Find and classify the critical points of $f(x) = x^3 - 3x$.

Solution. $f'(x) = 3x^2 - 3 = 0$ gives $x = \pm 1$. Second derivative $f''(x) = 6x$: at $x = 1$, positive, a local minimum, $f(1) = -2$; at $x = -1$, negative, a local maximum, $f(-1) = 2$. Sign of the second derivative reads the curvature.

Question 9. Find dy/dx on the circle $x^2 + y^2 = 25$ at the point $(3, 4)$.

Solution. Differentiate both sides treating y as a function of x : $2x + 2y y' = 0$, so $y' = -x/y$. At $(3, 4)$: $-3/4$. Implicit differentiation handles curves that are not graphs of single functions, and the result matches geometry: the tangent to a circle is perpendicular to its radius.

Question 10. Differentiate $f(x) = e^{2x} \ln x$.

Solution. Product plus chain: $f'(x) = 2e^{2x} \ln x + e^{2x} \cdot \frac{1}{x} = e^{2x} \left(2 \ln x + \frac{1}{x} \right)$. The exponential survives differentiation with only the factor 2 from its inner function; the logarithm contributes its reciprocal.

FAQs

What does a derivative actually represent?

It's the instantaneous rate of change of a function. Geometrically, it's the slope of the tangent line at a point. In physics, the derivative of position is velocity; the derivative of velocity is acceleration. In economics, it's marginal anything (cost, revenue, utility).

When is a function not differentiable?

At sharp corners (like $|x|$ at $x = 0$), at vertical tangents (like $x^{1/3}$ at $x = 0$), at discontinuities, and at endpoints if you require a two-sided limit. Differentiability is strictly stronger than continuity: every differentiable function is continuous, but not every continuous function is differentiable.

What's the difference between dy/dx and $\frac{dy}{dx}$?

Notation only. dy/dx is Leibniz notation that visually emphasizes 'change in y per change in x.' It's not literally division, but it can be manipulated as if it were in many situations (chain rule, separable differential equations), which is why physicists love it.

How do I know when to use the chain rule?

Whenever you see one function nested inside another: $f(g(x))$. Examples: $\sin(x^2)$, $e^{(3x)}$, $\sqrt{(x^2 + 1)}$, $\ln(\cos x)$. The outer function gets differentiated normally; multiply by the derivative of the inner function. If you don't see nesting, you don't need the chain rule.

What is the derivative used for in real life?

Anywhere instantaneous rates matter: velocity and acceleration in physics, growth rates in biology, marginal cost and elasticity in economics, gradient descent in machine learning, sensitivity analysis in engineering, hazard rates in actuarial science. If something changes, derivatives describe how fast.

What is implicit differentiation?

A technique for finding dy/dx when y is defined implicitly by an equation involving both x and y . You differentiate both sides with respect to x , treating y as a function of x and applying the chain rule wherever y appears. The result expresses dy/dx in terms of both x and y .

How is the derivative used in machine learning?

Gradient descent uses derivatives of the loss function with respect to model parameters to update parameters in the direction that reduces loss. Backpropagation in neural networks is the chain rule applied across many layers to compute these derivatives efficiently. Without derivatives, modern machine learning is impossible.

What is the second derivative test?

If $f'(c) = 0$ and $f''(c) > 0$, then c is a local minimum. If $f''(c) < 0$, then c is a local maximum. If $f''(c) = 0$, the test is inconclusive. The test extends to multivariable functions via the Hessian matrix and its eigenvalues.

What's the difference between average rate of change and instantaneous rate of change?

Average rate of change is the slope of the secant line over an interval: $(f(b) - f(a))/(b - a)$. Instantaneous rate of change is the derivative, the slope of the tangent line at a single point. As the interval shrinks to zero, the average rate approaches the instantaneous rate.

Can a function have an infinite derivative?

Yes, the derivative can be infinite at points where the tangent line is vertical. Examples: $x^{1/3}$ has an infinite derivative at $x = 0$ because the tangent line is vertical. The function is continuous but not differentiable there, depending on how you define differentiability for infinite slopes.

How do automatic differentiation and symbolic differentiation differ?

Symbolic differentiation manipulates expressions algebraically and produces a closed-form derivative. Automatic differentiation tracks the computational graph of a function and applies the chain rule numerically at each node, producing exact derivatives without ever simplifying the algebraic expression. Both are exact; AD scales better to complex models like neural networks.

Why do we need both Leibniz and Lagrange notation?

Different contexts favor different notation. Leibniz notation (dy/dx) makes the chain rule feel intuitive and is essential in differential equations and physics. Lagrange notation ($f'(x)$) is cleaner for general algebraic work and theoretical statements. Most mathematicians use both fluently and switch based on what makes the problem easier to read.

What is the Mean Value Theorem?

If f is continuous on $[a, b]$ and differentiable on (a, b) , there exists c in (a, b) where $f'(c)$ equals the average rate of change $(f(b) - f(a))/(b - a)$. Geometrically, some tangent line in the interval is parallel to the secant joining the endpoints.

How is the derivative of an inverse function computed?

Using the inverse function theorem: $(f^{-1})'(x) = 1 / f'(f^{-1}(x))$. This is how standard derivatives of $\arcsin$, $\arctan$, and $\ln$ are derived from the derivatives of $\sin$, $\tan$, and $\exp$.