CALCULUS

Derivatives in Calculus

What the derivative measures, every core rule with worked examples, optimization, and a 10-question practice set with full solutions.

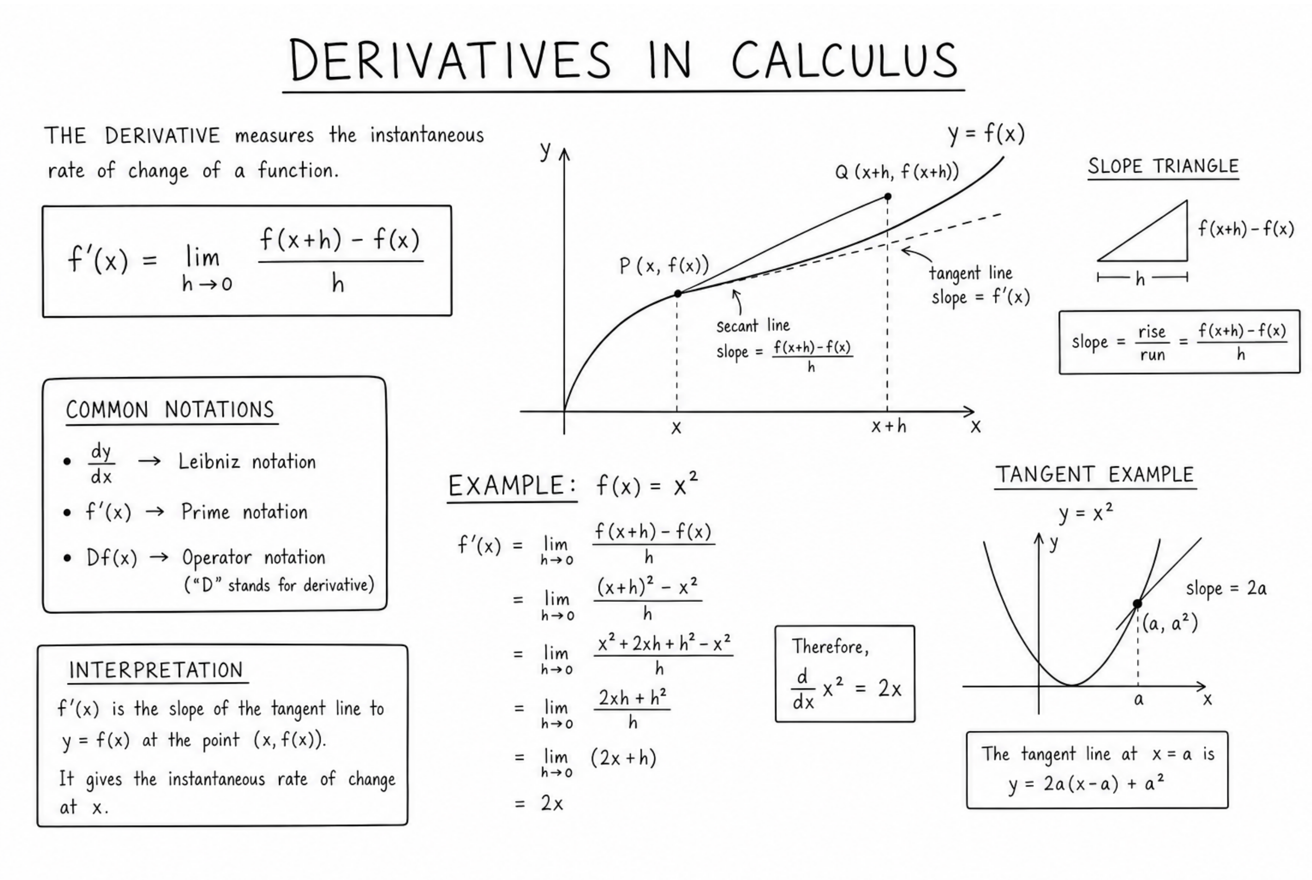
SUBJECT	LEVEL	INCLUDES
Mathematics	High school and early college	Practice set, answer key, revision sheet

Read this note online, with updates: gauravtiwari.org/study-notes/derivatives-in-calculus

INSIDE

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.

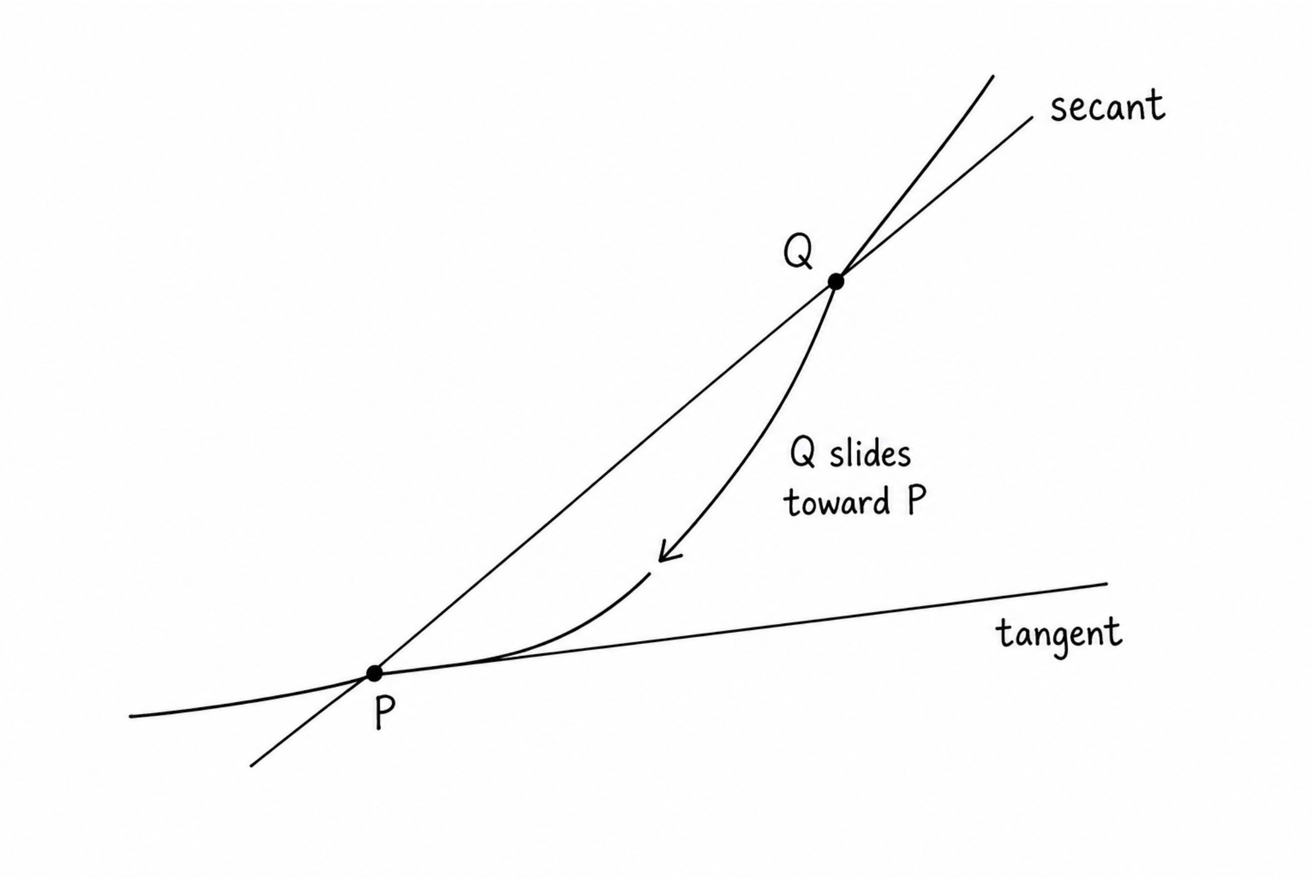
Most of single-variable calculus is built on three ideas: the limit definition, a small set of rules for differentiating common functions, and the chain rule. Memorize those, and 90% of derivative problems become routine. The interesting work happens at the edges, non-differentiable points, optimization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



Derivatives in calculus tangent line and secant line modern textbook illustration

The Limit Definition

The derivative of f at x is:



The derivative as a limit: as Q slides toward P, the secant line tips into the tangent, and its slope becomes the derivative at P.

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

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).

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.

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.

Worked Examples

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

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

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

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

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

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

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

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

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

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.

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.

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).

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

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 $\sin y = x$.

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).

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.

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).

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:

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

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.

Derivatives of Inverse Functions

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

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

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.

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).

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.

Q1. Differentiate $f(x) = x^5$.

SOLUTION

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

Q2. 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.

Q3. Differentiate $f(x) = x^2 \sin x$.

SOLUTION

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

Q4. Differentiate $f(x) = \dfrac{x}{x^2 + 1}$.

SOLUTION

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

Q5. 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.

Q6. 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.

Q7. 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 $v(2) = 19.6\text{ m/s}$. The derivative of position is velocity; differentiating again gives the constant acceleration 9.8 m/s^2, gravity itself.

Q8. 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.

Q9. 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.

Q10. 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.

Answer Key

QUICK ANSWERS

Q1 $5x^4$ **Q2** $12x^3 - 4x + 7$ **Q3** $2x \sin x + x^2 \cos x$ **Q4** $\dfrac{1 - x^2}{(x^2 + 1)^2}$ **Q5** $30x(3x^2 + 1)^4$

Q6 $y = 6x - 9$ **Q7** $19.6\,\mathrm{m/s}$ **Q8** $\max(-1, 2), \min(1, -2)$ **Q9** $-3/4$ **Q10** $e^{2x}(2 \ln x + 1/x)$

Revision Sheet

The definition

$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$: instantaneous rate of change, slope of the tangent.

Special derivatives

$(e^x)' = e^x$, $(\ln x)' = \frac{1}{x}$, $(\sin x)' = \cos x$, $(\cos x)' = -\sin x$.

Core rules

Power: $(x^n)' = nx^{n-1}$. Product: $(uv)' = u'v + uv'$. Quotient: $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$.

Optimization recipe

Solve $f' = 0$, then classify with f'': positive = min, negative = max.

Chain rule

$[f(g(x))]' = f'(g(x)) \cdot g'(x)$. Outside first, then multiply by the inside's derivative.

Motion

$s' = v$, $v' = a$. Position, velocity, acceleration: each the derivative of the one before.

About These Notes

This note is part of a free study-notes series on gauravtiwari.org covering mathematics, physics, chemistry, and biology: the concepts that keep deciding grades in high school, board, and entrance exams.

2008
Writing and teaching online since

2,000+
Articles published on gauravtiwari.org

125+
Free study notes in this series

M.Sc.
Mathematics, with physics and chemistry training

Gaurav Tiwari is a developer, educator, and founder of Gatilab. The study-notes series exists because most exam prep material is either a full textbook or a thin definition page, and revision needs something in between: one concept, complete, with the traps named and the practice attached.

GET THE FULL SERIES

Every note in the series is free to read online and free to download as a PDF like this one. Browse the complete collection at **gauravtiwari.org/free-study-notes-exam-prep** and pick the topics your next exam actually covers.

Gaurav Tiwari

gauravtiwari.org