Differentiation

alexander

December 1, 2025

the derivative of a function f at $x = a$ is the limit of the difference quotients (if it exists):

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

when the limit exists, we say that f is differentiable at $x = a$. an equivalent definition of the derivative is

$$f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a}$$

tangent lines:
assume that $f(x)$ is differentiable at $x = a$. the tangent line to the graph of $y = f(x)$ at $P = (a, f(a))$ is the line through P of slope $f'(a)$. the equation of the tangent line in point-slope form is

$$y - f(a) = f'(a)(x - a)$$

derivative of linear and constant functions:

- if $f(x) = mx + b$ is a linear function, then $f'(a) = m$ for all a

- if $f(x) = b$ is a constant function, then $f'(a) = 0$ for all a

the power rule (1)

1. $\frac{d}{dx}(x^n) = \lim_{h \to 0} \frac{(x+h)^n - x^n}{h}$

2. $(x+h)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} h^k$

3. $(x+h)^n - x^n = \sum_{k=1}^{n} \binom{n}{k} x^{n-k} h^k$

4. $\frac{(x+h)^n - x^n}{h} = \sum_{k=1}^{n} \binom{n}{k} x^{n-k} h^{k-1}$

5. take the limit as $h \to 0$. all terms with h^{k-1} for $k \geq 2$ go to zero. only the $k = 1$ term remains.

6. $\binom{n}{1} x^{n-1} = n x^{n-1}$

7. $\frac{d}{dx}(x^n) = n x^{n-1}$

the power rule (2)

1. $f'(a) = \lim_{x \to a} \frac{x^n - a^n}{x - a}$

2. $x^n - a^n = (x - a)(x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1})$

3. $\frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1}$

4. $f'(a) = na^{n-1}$

linearity rules assume that f and g are differentiable functions

- sum rule: the function $f + g$ is differentiable and $(f + g)' = f' + g'$

- constant multiple rule: for any constant c, cf is differentiable and $(cf)' = cf'$

differentiability implies continuity

1. theorem: if f is differentiable at c, then f is continuous at c

2. since f is differentiable at c, the limit $f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$ exists

3. to expose the structure behind this limit, we rewrite the difference quotient as the sum of the tangent slope and a small error term measuring the difference between the secant slope and the tangent slope: $f'(c) + \epsilon(x) = \frac{f(x) - f(c)}{x - c}$

4. $\epsilon(x) = \frac{f(x) - f(c)}{x - c} - f'(c)$ because the difference quotient tends to $f'(c)$ as $x \to c$, this error term must satisfy $\lim_{x \to c} \epsilon(x) = 0$

5. this decomposition allows us to express $f(x)$ in a linearized form: $y - y_1 = m(x - x_1) \Rightarrow L(x) = f(c) + f'(c)(x - c) \Rightarrow f(x) = f(c) + f'(c)(x - c) + \epsilon(x)(x - c)$

6. this expression shows that near c, the function behaves like its tangent line plus an error term. and because $\epsilon(x) \to 0$ and $x - c \to 0$, the product $\epsilon(x)(x - c)$ also goes to 0. thus, $\lim_{x \to c} f(x) = f(c) + 0 + 0 = f(c)$

proof of the product rule

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

2. add and subtract $f(x)g(x + h)$

3. $\lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h}$

4. $\lim_{h \to 0} \left[\frac{f(x+h) - f(x)}{h} g(x + h) + f(x) \frac{g(x+h) - g(x)}{h} \right]$

5. $\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \cdot \lim_{h \to 0} g(x + h) + f(x) \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}$

6. $(fg)'x = f'(x)g(x) + f(x)g'(x)$

proof of the quotient rule

1. $\left(\frac{f}{g}(x)\right)' = \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h}$

2. $\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)} = \frac{f(x+h)g(x) - f(x)g(x+h)}{g(x+h)g(x)}$

3. $\lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} = \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x+h)g(x)}$

4. add and subtract $f(x)g(x)$

5. $\lim_{h\to 0} \frac{g(x)(f(x+h)-f(x)) - f(x)(g(x+h)-g(x))}{hg(x+h)g(x)}$

6. $\lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h} \cdot \frac{g(x)}{g(x+h)g(x)} - \frac{g(x+h)-g(x)}{h} \cdot \frac{f(x)}{g(x+h)g(x)}\right]$

7. $f'(x) \cdot \frac{1}{g(x)} - g'(x)\frac{f(x)}{g(x)g(x)}$

8. $\frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{[g(x)]^2} = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$

proof of the derivative of sin

1. $\frac{d}{dx}\sin(x) = \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h}$

2. $\sin(x+h) = \sin(x)\cos(h) + \cos(x)\sin(h)$

3. $\lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h} = \lim_{h\to 0} \frac{\sin(x)\cos(h)-\sin(x)+\cos(x)\sin(h)}{h}$

4. $\lim_{h\to 0} \frac{\sin(x)(\cos(h)-1)+\cos(x)\sin(h)}{h}$

5. $\lim_{h\to 0}\left[\sin(x) \cdot \frac{\cos(h)-1}{h} + \cos(x) \cdot \frac{\sin(h)}{h}\right]$

6. $\sin(x) \cdot (0) + \cos(x) \cdot (1)$

7. $\frac{d}{dx}\sin(x) = \cos(x)$

proof of the chain rule

1. $h(x) = f(g(x))$

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

3. notice that the numerator is Δf and the denominator is Δx not the change in Δg

4. $\frac{f(g(x+h))-f(g(x))}{1} \cdot \frac{1}{g(x+h)-g(x)} \cdot \frac{1}{h} \cdot \frac{g(x+h)-g(x)}{1}$

5. $\lim_{h\to 0} \frac{f(g(x+h))-f(g(x))}{h} = \lim_{h\to 0} \frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)} \cdot \lim_{h\to 0} \frac{g(x+h)-g(x)}{h}$

6. $h'(x) = f'(g(x)) \cdot g'(x)$

3

shifting and scaling rule if $f(x)$ is differentiable, then for any constants k and b, $\frac{d}{dx}f(kx+b) = kf'(kx+b)$

implicit differentiation:
to differentiate using the methods covered thus far, we must have a formula for y in terms of x, for instance, $y = x^3 + 1$. but suppose that y is determined instead by an equation such as $y^4 + xy = x^3 - x + 2$. in this case, we say that y is defined impplicitly. how can we find the slope of the tangent line at a point on the graph? it may be inconvenient or even impossible to solve for y explicitly as a function of x, but we can find $\frac{dy}{dx}$ using the method of implicit differentiation. to illustrate, consider the equation of the unit circle:

1. $x^2 + y^2 = 1$

2. to compute $\frac{dy}{dx}$, first take the derivative of both sides of the equation and evaluate:
   $\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(1)$

3. $\frac{d}{dx}(x^2) + \frac{d}{dx}(y^2) = 0$

4. $2x + \frac{d}{dx}(y^2) = 0$

5. what should we do with the term $\frac{d}{dx}(y^2)$? we use the chain rule. if we think of y as a function $y = f(x)$, then $y^2 = f(x)^2$, and by the chain rule,
   $\frac{d}{dx}y^2 = \frac{d}{dx}f(x)^2 = 2f(x)\frac{df}{dx} = 2y\frac{dy}{dx}$

6. $2x + \frac{d}{dx}(y^2) \to 2x + 2y\frac{dy}{dx}$, and we may solve for $\frac{dy}{dx}$ if $y \neq 0$:

7. $\frac{dy}{dx} = -\frac{x}{y}$

related rates:
in related rate problems, the goal is to calculate an unknown rate of change in terms of other rates of change that are known. one typical problem involves a ladder learning against a wall. the question is: how fast does the ladder move if the bottom of the ladder is pulled away from the wall at constant speed? what is interesting and perhaps surprising is that the top and bottom travel at different speeds.

linear approximation:
we can use the derivative to estimate Δf without computing it exactly. by definition, the derivative is the limit
$f'(a) = \lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$
so when Δx is small, we have $\frac{\Delta f}{\Delta x} \approx f'(a)$, and thus, $\Delta f \approx f'(a)\Delta x$

the linear approximation if often called the tangent line approximation because of the following interpretation. the quantitiy Δf is the vertical change from $x = a$ to $a = a + \Delta x$ in the graph of $f(x)$. recall that for a nonvertical straight line, the vertical change is equal to the slope times the horizontal change. since

the tangent line has slope $f'(a)$. the vertical change in the tangent line is $f'(a)\Delta x$. what the linear approximation does, therefore, is used the vertical change in the tangent line as an approximation to the vertical change in the graph of $f(x)$. when Δx is small, the two quantities are nearly equal.

Keep in mind the different roles played by Δf and $f'(a)\Delta x$. the quantity of interest is Δf and we estimate it by $f'(a)\Delta x$. most importantly, the linear approximation tells us that there is a nearly linear relationship between Δf and Δx when Δx is small.

linearization:
we can approximate $f(x)$ itself rather than the change Δf by rewriting the linear approximation in terms of the variable $x = a + \Delta x$. Then
$f(a + \Delta x) - f(a) \approx f'(a)\Delta x$
$f(x) - f(a) \approx f'(a)(x - a)$ (since $\Delta x = x - a$)
$f(x) \approx f(a) + f'(a)(x - a)$
the function on the right, denoted by $L(x)$, is called the linearization of $f(x)$ at $x = a$:
$L(x) = f'(a)(x - a) + f(a)$
we refer to $x = a$ as the center of the linearization. notice that $y = L(x)$ is the equation of the tangent line to the graph of $f(x)$ at $x = a$

approximating $f(x)$ **by its linearization** assume that f is differentiable at $x = a$ if x is close to a, then
$f(x) \approx L(x) = f'(a)(x - a) + f(a)$

extreme values on an interval let $f(x)$ be a function on an interval I and let $a \in I$. we say that $f(a)$ is the

- absolute minimum of $f(x)$ on I if $f(a) \le f(x)$ for all $x \in I$

- absolute maximum of $f(x)$ on I if $f(a) \ge f(x)$ for all $x \in I$

existence of extrema on a closed interval if $f(x)$ is a continuous function on a closed (bounded) interval $I = [a, b]$, then $f(x)$ takes on a minimum and a maximum value on I.

local extrema we say that $f(x)$ has a

- at $x = c$ if $f(c)$ is the minimum value of f on some open interval (in the domain of f) containing c.

- at $x = c$ if $f(c)$ is the maximum value of $f(x)$ on some open interval (in the domain of f) containing c.

critical points a number c in the domain of f is called a critical point if either $f'(c) = 0$ or $f'(c)$ DNE.

fermat's theorem on local extrema if $f(c)$ is a local min or max, then c is a critical point of f.

extreme values on a closed interval assume that $f(x)$ is continuous on $[a, b]$ and let $f(c)$ be the minimum or maximum value on $[a, b]$. then c is either a critical point or one of the endpoints a or b.

rolle's theorem assume that $f(x)$ is continous on $[a, b]$ and differentiable on (a, b). if $f(a) = f(b)$, then there exists a number c between a and b such that $f'(c) = 0$.

mean value theorem assume that f is continuous on the closed interval $[a, b]$ and differentiable on (a, b). then there exists at least one value c in (a, b) such that
$f'(c) = \frac{f(b) - f(a)}{b - a}$

the sign of the derivative let f be a differentiable function on the open interval (a, b).

- if $f'(x) > 0$ for $x \in (a, b)$, then f is increasing on (a, b).

- if $f'(x) < 0$ for $x \in (a, b)$, then f is decreasing on (a, b).

first derivative test for critical points assume that $f(x)$ is differentiable and let c be a critical point of $f(x)$. then:

- $f'(x)$ changes from + to - at $c \Rightarrow f(c)$ is a local maximum.

- $f'(x)$ changes from - to + at $c \Rightarrow f(c)$ is a local minimum.

test for concavity suppose that $f''(x)$ exists for all $x \in (a, b)$.

- if $f''(x) > 0$ for all $x \in (a, b)$, then f is concave up on (a, b).

- if $f''(x) < 0$ for all $x \in (a, b)$, then f is concave down on (a, b).

test for inflection points assume that $f''(x)$ exists for all $x \in (a, b)$ and let $c \in (a, b)$. if $f''(c) = 0$ and $f''(x)$ changes sign at $x = c$, then $f(x)$ has a point of inflection at $x = c$.

second derivative test assume that $f(x)$ is differentiable and let c be a critical point. if $f''(c)$ exists, then

- $f''(c) > 0 \Rightarrow f(c)$ is a local minimum

- $f''(c) < 0 \Rightarrow f(c)$ is a local maximum

- $f''(c) = 0 \Rightarrow f(c)$ inconclusive: $f(c)$ may be a local min, max, or neither

newton's method:
this is a procedure for finding numerical approximations to zeros of functions. numerical approximations are important because it is often impossible to find the zeros exactly. for example, the polynomial $f(x) = x^5 - x - 1$ has one real root c, but we can prove, using an advanced branch of mathematics called Galois Theory, that there is no algebraic formula for this root. newtons method shows that $c \approx 1.1673$, and with enough computation, we can compute c to any desired degree of accuracy.

in newton's method, we begin by choosing a number x_0, which we believe is close to a root of $f(x)$. this starting value x_0 is called the initial guess. newton's method then produces a sequence $x_0, x_1, x_2, \ldots$ of successive approximations that, in favorable situations, converge to a root.

steps:

- choose initial guess x_0 (close to the desired root if possible).

- generate successive approximations $x_1, x_2, \ldots$ where
  $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

derivatives:

- $\frac{d}{dx}(c) = 0$

- $\frac{d}{dx}x = 1$

- $\frac{d}{dx}(x^n) = nx^{-1}$ (power rule)

- $\frac{d}{dx}[cf(x)] = cf'(x)$

- $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$

- $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$

- $\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

- $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$

- $\frac{d}{dx}f(x)^n = nf(x)^{n-1}f'(x)$

- $\frac{d}{dx}\sin(x) = \cos(x)$

- $\frac{d}{dx}\cos(x) = -\sin(x)$

- $\frac{d}{dx}\tan(x) = \sec^2(x)$

- $\frac{d}{dx}\csc(x) = -\csc(x)\cot(x)$

- $\frac{d}{dx}\sec(x) = \sec(x)\tan(x)$

- $\frac{d}{dx} \cot(x) = -\csc^2(x)$

- $\frac{d}{dx} \sin^{-1}(x) = \frac{1}{\sqrt{(1-x^2)}}$

- $\frac{d}{dx} \cos^{-1}(x) = -\frac{1}{\sqrt{(1-x^2)}}$

- $\frac{d}{dx} \tan^{-1}(x) = \frac{1}{1+x^2}$

- $\frac{d}{dx}(e^x) = e^x$

- $\frac{d}{dx}(a^x) = (\ln a)a^x$

- $\frac{d}{dx} \ln \mid x \mid = \frac{1}{x}$

- $\frac{d}{dx} \log_a x = \frac{1}{(\ln a)x}$