Notes on Calculus

Paulo Eduardo Rauber

2015

1 Limits

Consider an open interval I and $a \in I$, and a function f defined on I, except possibly at a. The limit of f when x tends to a is equal to L, written as

$$\lim_{x \to a} f(x) = L,$$

if, for every $\epsilon > 0$, there is a $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - L| < \epsilon$, for every x.

Intuitively, this means that it is possible to get $f(x)$ within any $\epsilon > 0$ of L from any x within $\delta > 0$ of a.

Let c be a constant and suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist. The following properties hold:

$$\lim_{x \to a}[f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x).$$
$$\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x).$$
$$\lim_{x \to a}[f(x)g(x)] = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x).$$
$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}, \text{ if } \lim_{x \to a} g(x) \neq 0.$$

A function f is continuous at a if it is defined at a and $\lim_{x\to a} f(x) = f(a)$. This definition extends naturally to intervals. Intuitively, a function is continuous on an interval if its graph has no *jumps* on this interval.

The intermediate value theorem states that if f is continuous on the closed interval $[a, b]$, and $f(a) \neq f(b)$, then, for any N between $f(a)$ and $f(b)$, there is a $c \in (a, b)$ such that $f(c) = N$.

Let f be a function defined on an open interval that contains a, except possibly at a itself. Then

$$\lim_{x \to a} f(x) = \infty$$

denotes that, for every positive M, there is a positive δ such that $0 < |x - a| < \delta$ implies $f(x) > M$, for every x. Intuitively, it is possible to surpass any fixed M with $f(x)$ from any x within $\delta > 0$ of a. An analogous definition applies to $\lim_{x\to a} f(x) = -\infty$.

Let a be a constant and f a function defined for every $x > a$. Then

$$\lim_{x \to \infty} f(x) = L$$

if, for every $\epsilon > 0$, there is a $N > a$ such that $x > N$ implies $|f(x) - L| < \epsilon$. An analogous definition definition applies to $\lim_{x\to-\infty} f(x) = L$.

Consider the functions f, g and h defined on an interval I, except possibly at $a \in I$, and let $f(x) \leq g(x) \leq h(x)$ for every $x \in I$, except possibly when $x = a$. Then $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L$ implies $\lim_{x\to a} g(x) = L$.

2 Derivatives

The derivative f' of a function f is defined by

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

for the x for which the limit exists. If $f'(x)$ is defined, then f is differentiable at x. Intuitively, $f'(x)$ is the instantaneous rate of change of f at x, or the slope of a line tangent to the curve of f at x.

It can be shown that if f is differentiable at a, then f is continuous at a. The converse is not true.

In an alternative notation, and letting $y = f(x)$,

$$f'(x) = \frac{d}{dx} f(x) = \frac{dy}{dx}.$$

Let a, b, c and n be constants and e denote Euler's number. The following identities can be shown:

$$\begin{aligned}
\frac{d}{dx}(c) &= 0, \\
\frac{d}{dx}(x^n) &= nx^{n-1}, \\
\frac{d}{dx}[cf(x)] &= c\frac{d}{dx}f(x), \\
\frac{d}{dx}[f(x)+g(x)] &= \frac{d}{dx}f(x) + \frac{d}{dx}g(x), \\
\frac{d}{dx}[f(x)g(x)] &= f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x), \\
\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] &= \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}, \\
\frac{d}{dx}(e^x) &= e^x, \\
\frac{d}{dx}(a^x) &= a^x \ln(a), \\
\frac{d}{dx}\ln(x) &= \frac{1}{x} \\
\frac{d}{dx}\log_b(x) &= \frac{1}{x\ln(b)} \\
\frac{d}{dx}\sin(x) &= \cos(x), \\
\frac{d}{dx}\cos(x) &= -\sin(x).
\end{aligned}$$

Let g and f be functions, and let $F(x) = f(g(x))$. If g is differentiable at x and f is differentiable at $g(x)$, then $F'(x) = f'(g(x))g'(x)$. Alternatively, $(f \circ g)'(x) = f'(g(x))g'(x)$. This is the chain rule for differentiation.

Therefore, if $u = g(x)$ and $y = f(u)$, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

This provides a useful mnemonic when dx, du and dy are interpreted if they were real numbers, because the terms du would cancel each other. This should be used carefully, and should respect the differentiability conditions for the chain rule.

Suppose $f(x) = g(x)$ for every x on an open interval I. If f and g are differentiable on I, Then $f'(x) = g'(x)$ for every $x \in I$. This observation can be used for *implicit differentiation*. As an example, consider the equation of the circle $x^2 + y^2 = 1$, where $y = h(x)$. If the equation is true for every $x \in (-1, 1)$, $\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(1)$, which gives $2x + \frac{d}{dx}(y^2) = 0$. Consider only the upper semicircle $y = h(x) = \sqrt{1 - x^2}$. Then h is differentiable in $(-1, 1)$ and $\frac{d}{dy}(y^2)$ is defined for every y. By the chain rule, $2x + \frac{d}{dx}(y^2) = 2x + \frac{d}{dy}(y^2)\frac{dy}{dx} = 2x + 2y\frac{dy}{dx}$ and $\frac{dy}{dx} = -\frac{x}{y}$ on the interval $(-1, 1)$. The last step is valid because $y \neq 0$ for every $x \in (-1, 1)$. This is a very important differentiation technique in the cases where it is easier to establish that h is differentiable than to differentiate it explicitly.

Let f be a function defined on the domain D and let $c \in D$. Then $f(c)$ is the global maximum value of f on D if $f(c) \geq f(x)$ for all $x \in D$. When $f(c) \geq f(x)$ for every x on an open interval containing c, $f(c)$ is a local maximum value. An analogous definition applies to global and local minimum values.

The following is the extreme value theorem. If f is continuous on a closed interval $[a, b]$, there are $c, d \in [a, b]$ such that $f(c)$ is the global maximum value of f on $[a, b]$ and $f(d)$ is the global minimum value of f on $[a, b]$.

By Fermat's theorem, if f is differentiable and has a local maximum or minimum at c, then $f'(c) = 0$. This theorem is extremely important in applied calculus.

A critical point c of a function f is such that $f'(c) = 0$ or $f'(c)$ is not defined. Consider a function f differentiable on a closed interval $[a, b]$. It is possible to find its global maximum (or minimum) on $[a, b]$ by comparing $f(a)$ and $f(b)$ to $f(c)$, for every critical point $c \in [a, b]$.

If a function f is continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$, then $f'(c) = 0$ for some $c \in (a, b)$. This is Rolle's theorem.

The mean value theorem states the following. If f is continuous on $[a, b]$ and differentiable on (a, b), then there is a c in (a, b) such that

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

Intuitively, this theorem states that the slope of a line through $(a, f(a))$ and $(b, f(b))$ is equal to the slope of the curve f at some point c between a and b.

By the mean value theorem, if $f'(x) = 0$ for every $x \in (a, b)$, then $f(a) = f(b)$. Also, if $f'(x) = g'(x)$ for every $x \in (a, b)$, then $f(x) = g(x) + c$ for every $x \in (a, b)$.

Also by the mean value theorem, if $f'(x) > 0$ on (a, b) for every x, then f is increasing on (a, b). Analogously, if $f'(x) < 0$ on (a, b) for every x, then f is decreasing on (a, b).

The following is the first derivative test. If c is a critical point of a continuous function f, then f has a local maximum at c if f' changes from positive to negative at c. It has a local minimum if f' changes from negative to positive. If f' is defined and does not change sign at c, it has no minimum or maximum at c.

The second derivative test states the following. Let f be a continuous function, and let f'' be continuous on an open interval containing c. If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c. If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c.

Suppose f and g are differentiable on an open interval I and $g'(x) \neq 0$ for every $x \in I$, except possibly at $a \in I$. Suppose also that $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$. Then

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

This result also holds when $\lim_{x \to a} f(x) = \infty$ and $\lim_{x \to a} g(x) = \infty$, and is called L'Hospital's rule.

Consider a function f differentiable on the interval (a, b), and suppose we are interested on finding an x such that $f(x) = 0$. Newton's method is a procedure that starts at an arbitrary point $x_0 \in (a, b)$ and computes $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$, as long as $f'(x_i) \neq 0$ and $x_{i+1} \in (a, b)$, until convergence. If the procedure converges, $x_{i+1} = x_i$ and $f(x_{i+1}) = 0$. However, convergence is not guaranteed. Intuitively, each step of this procedure finds the intersection $(x_{i+1}, 0)$ of the line with slope $f'(x_i)$ that goes through $(x_i, f(x_i))$ with the line with slope 0 that goes through $(0, 0)$.

A function F is an antiderivative of a function f on an interval I if $F'(x) = f(x)$ for every $x \in I$. If F is an antiderivative of f on I, C is a constant and $G(x) = F(x) + C$, then G is also an antiderivative of f on I.

3 Integrals

Let f be a function defined on the interval $[a, b]$, and $\Delta x = \frac{b-a}{n}$. Consider the intervals $I_i = [a + (i-1)\Delta x, a + i\Delta x]$, for $i \in \{1, \dots, n\}$, and let x_i^* denote an element of I_i. The definite integral of f from a to b is

$$\int_a^b f(x)\ dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*) \Delta x,$$

if the limit exists and is the same for all choices of x_i^*, for all i. In this case, the function f is integrable on $[a, b]$. If f is continuous on $[a, b]$, then f is integrable on $[a, b]$. Intuitively, the definite integral of f from a to b corresponds to the signed area bellow the curve of f between a and b, or an accumulation of $f(x)$ for all x between a and b.

The following properties hold for definite integrals of an integrable function f:

$$\int_a^b f(x)\ dx = -\int_b^a f(x)\ dx,$$

$$\int_a^a f(x)\ dx = 0,$$

$$\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx,$$

$$\int_a^b [f(x) + g(x)]\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx,$$

where c is any constant.

If $f(x) \geq 0$ for every $x \in [a,b]$, then $\int_a^b f(x)\ dx \geq 0$. If $f(x) \geq g(x)$ for every $x \in [a,b]$, then $\int_a^b f(x)\ dx \geq \int_a^b g(x)\ dx$.

If $m \leq f(x) \leq M$ for every $x \in [a,b]$, then $m(b-a) \leq \int_a^b f(x)\ dx \leq M(b-a)$.

The mean value theorem for integrals states that given a function f continuous on $[a,b]$, there exists a number $c \in [a,b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\ dx.$$

Consider a function f and suppose $\int_a^t f(x)\ dx$ exists for every $t \geq a$. Then

$$\int_a^\infty f(x)\ dx = \lim_{t\to\infty}\int_a^t f(x)\ dx,$$

and an analogous definition can be made for $\lim_{t\to-\infty}\int_t^b f(x)\ dx$.

The following two statements correspond to the fundamental theorem of calculus.

Consider a function f continuous on $[a,b]$ and let $g(x) = \int_a^x f(t)\ dt$ for $x \in [a,b]$. Then g is continuous and differentiable on $[a,b]$ and $g'(x) = f(x)$. Intuitively, the instantaneous rate at which f is accumulated on g at x is equal to the value of f at x.

If f is continuous on $[a,b]$, then

$$\int_a^b f(x)\ dx = F(b) - F(a) = F(x)\Big]_a^b,$$

where F is any antiderivative of f.

We let the indefinite integral $\int f(x)\ dx$ of f denote any antiderivative of f.

Let a, c, k, n and C be constants. The following properties of indefinite integrals can be shown.

$$\int k\ dx = kx + C,$$
$$\int x^n\ dx = \frac{x^{n+1}}{n+1} + C,\ \text{if } n \neq -1,$$
$$\int \frac{1}{x}\ dx = \ln(|x|) + C,$$
$$\int e^x\ dx = e^x + C,$$
$$\int a^x\ dx = \frac{a^x}{\ln(|a|)} + C,$$
$$\int \sin(x)\ dx = -\cos(x) + C,$$
$$\int \cos(x)\ dx = \sin(x) + C.$$

There are two main techniques to evaluate more complicated indefinite integrals: the substitution rule and integration by parts.

Let $F(x) = f(g(x))$ for every x in an interval I. By the chain rule, if g is differentiable on I and f is differentiable on the image of g, then $F'(x) = f'(g(x))g'(x)$. Therefore:

$$\int f'(g(x))g'(x)\ dx = f(g(x)) + C.$$

If we let $u = g(x)$,

$$\int f'(u)\frac{du}{dx}\ dx = \int f'(u)\ du + C.$$

Once again, dx and du may serve as mnemonic devices, because the terms dx would cancel each other if they were real numbers. However, the differentiability conditions for f and g must hold.

Intuitively, the substitution rule can be seen as an inversion of the chain rule. It is often challenging to choose f and g so that the indefinite integral of interest results in $f(g(x)) + C$. As a general guideline, it is useful to look for f and g such that the neglected factors in the integrand would cancel $g'(x)$.

Integration by parts is based on the product rule for differentiation. Let f and g be differentiable functions. Then $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$. So:

$$\int f(x)g'(x) + f'(x)g(x)\ dx = f(x)g(x) + C,$$

and

$$\int f(x)g'(x)\ dx = f(x)g(x) - \int f'(x)g(x)\ dx + C.$$

Integration by parts is a good solution for evaluating an indefinite integral that is the product of two functions f and g', under the conditions that f is easy to differentiate, and g' and $f'g$ are easy to antidifferentiate. Once again, it may be challenging to choose f and g' with these properties.

4 Differential equations

A differential equation is an equation that involves an unknown function f and at least one of its derivatives. Differential equations are often useful to find models for phenomena that are naturally described by a rate of change.

The order of a differential equation is the order of the highest derivative in the equation. A first order differential equation can often be written as

$$f'(x) = F(x, f(x)),$$

where F is a known function and f is an unknown function. A solution for a differential equation is any function f that satisfies the equation. Finding the set of all solutions for a differential equation may be very challenging.

A first order differential equation is separable if it can be written as

$$f'(x) = \frac{g(x)}{h(f(x))},$$

where g and h are continuous functions. In this case,

$$\int f'(x)h(f(x))\ dx = \int g(x)\ dx + C,$$

for any choice of antiderativatives and some constant C. Letting $y = f(x)$ and using the chain rule,

$$\int f'(x)h(f(x))\ dx = \int \frac{dy}{dx}h(y)\ dx = H(f(x)) + C_1,$$

for some antiderivative H of h. Choosing $C_1 = 0$, it follows that

$$H(f(x)) = \int g(x)\ dx + C.$$

A function H^{-1} is the inverse of H in an interval if $H^{-1}(H(y)) = y$, for every y in that interval. If H is invertible on the image of f on I, every solution has the form

$$f(x) = H^{-1}\Big(\int g(x)\ dx + C\Big),$$

for any antiderivative of g and some C. Alternatively, choosing an antiderivative G of g,

$$f(x) = H^{-1}\Big(G(x) + C\Big),$$

for *any* C and some antiderivative H of h.

A first order differential equation is linear when it can be written as

$$f'(x) + f(x)p(x) = q(x),$$

where p and q are continuous functions. Suppose there is a differentiable function I such that $I'(x) = I(x)p(x)$ for every x in the domain of f. By multiplying I on both sides of the linear differential equation,

$$I(x)f'(x) + I(x)f(x)p(x) = I(x)q(x).$$

For any choice of antiderivatives and some C,

$$\int I(x)f'(x) + I(x)f(x)p(x)\ dx = \int I(x)q(x)\ dx + C.$$

.

By the product rule and the supposition about I',

$$I(x)f(x) + C_1 = \int I(x)q(x)\ dx + C.$$

. By choosing $C_1 = 0$ and assuming I is non-zero in its domain, every solution has the form:

$$f(x) = \frac{1}{I(x)}\Big[\int I(x)q(x)\ dx + C\Big],$$

for some C. Alternatively, let G denote some antiderivative of $I \cdot q$. Then

$$f(x) = \frac{1}{I(x)}\Big[G(x) + C\Big],$$

for *any* C.

Finding a suitable function I requires solving the first order differential equation $I'(x) = I(x)p(x)$. Because I must be non-zero in its domain, this is a separable differential equation with $g(x) = p(x)$ and $h(I(x)) = \frac{1}{I(x)}$. Therefore, $I(x) = e^{\int p(x)\ dx}$. By choosing any antiderivative P of p, a suitable I is $I(x) = e^{P(x)}$.

When a first order differential equation is not separable or linear, it is still possible to obtain useful information about the unknown function.

If the differential equation is in the form $f'(x) = F(x, f(x))$ for a known F, it is possible to evaluate (or approximate) F at every pair (x, y) in the Cartesian plane. Therefore, a graph can be created to illustrate the behavior of the derivative f' by supposing that the curve corresponding to f goes through a selected set of points in the plane. For each selected point, a small line segment with appropriate slope can be drawn. A given line segment would be tangent to the curve of f if the function went through the respective point. A graph resulting from this process is known as a direction field of f.

Euler's method is a method for approximating the curve of a function f that follows the differential equation $f'(x) = F(x, f(x))$ for a known F. For a chosen x_0 and y_0, let $x_{i+1} = x_i + h$ and $y_{i+1} = y_i + hF(x_i, y_i)$. For a sufficiently small $|h|$, the sequence of points $(x_0, y_0), \ldots, (x_n, y_n)$ can be connected to approximate the curve of f between x_0 and $x_0 + nh$. It is important to notice that approximation errors accumulate at each step.

5 Partial derivatives

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of multiple variables. In vector notation, let $\mathbf{x} = (x_1, \ldots, x_n)$ and $f(x_1, \ldots, x_n) = f(\mathbf{x})$. An open ball B around $\mathbf{a} \in \mathbb{R}^n$ is a set $B = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x} - \mathbf{a}|| < \delta\}$ for some $\delta > 0$.

Consider a function f defined on an open ball B around $\mathbf{a}$, except possibly at $\mathbf{a}$. The limit of f as $\mathbf{x}$ approaches $\mathbf{a}$ is L, denoted by

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = L,$$

if for every $\epsilon > 0$, there is an $\delta > 0$ such that $0 < ||\mathbf{x} - \mathbf{a}|| < \delta$ implies $|f(\mathbf{x}) - L| < \epsilon$, for every $\mathbf{x} \in B$. Intuitively, this happens if every $\mathbf{x}$ at a small (non-zero) distance from $\mathbf{a}$ corresponds to $f(\mathbf{x})$ at a small distance from L.

A function of multiple variables f is continuous at $\mathbf{a}$ if

$$f(\mathbf{a}) = \lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}).$$

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of multiple variables $x_1, \ldots, x_n$. The partial derivative of f with respect to x_i at $\mathbf{a} = (a_1, \ldots, a_n)$, denoted by $\frac{\partial}{\partial x_i} f(a_1, \ldots, a_n)$, is defined as

$$\frac{\partial}{\partial x_i} f(a_1, \dots, a_n) = \lim_{h \to 0} \frac{f(a_1, \dots, a_i + h, \dots, a_n) - f(a_1, \dots, a_i, \dots, a_n)}{h},$$

if the limit exists.

If $z = f(\mathbf{a})$, the following notation is also used for the partial derivative of f with respect to x_i at $\mathbf{a} = (a_1, \dots, a_n)$:

$$\frac{\partial}{\partial x_i} f(\mathbf{a}) = \frac{\partial z}{\partial x_i} = \partial_{x_i} f(\mathbf{a}) = \partial_{x_i} z.$$

If the partial derivatives $\frac{\partial}{\partial x_i} f(\mathbf{a})$ exist in an open ball containing $\mathbf{a}$ and are continuous on $\mathbf{a}$ for every x_i, then f is differentiable at $\mathbf{a}$.

Consider the function $g(x) = f(a_1, \dots, x, \dots, a_n)$, where x is placed as the i-th argument of f. Clearly, $\frac{\partial}{\partial x_i} f(\mathbf{a}) = g'(a_i)$. Therefore, to compute the partial derivative of f with respect to x_i at point $\mathbf{a}$, it is possible to treat a_j as a constant for every $i \neq j$ and differentiate f as usual.

Intuitively, $\frac{\partial}{\partial x_i} f(a_1, \dots, a_n)$ is the slope (or instantaneous rate of change) of g at the point a_i, where g is a function that corresponds to f with all of its arguments fixed to $\mathbf{a}$, except for x_i. The partial derivative of f with respect to x_i at point $\mathbf{a}$ can also be seen as the *derivative* of f in the *direction* of the standard basis vector $\hat{\mathbf{e}}_i$ at point $\mathbf{a}$.

Because $\frac{\partial}{\partial x_i} f$ is a function, its partial derivatives may also be defined. The partial derivative with respect to x_j of the partial derivative with respect to x_i of f at point $\mathbf{a}$ is denoted by

$$\frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(\mathbf{a}) = \frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{a}).$$

The notation is analogous for higher derivatives.

The following is Clairaut's theorem. If f is defined on an open ball B that contains $\mathbf{a} = (a_1, \dots, a_n)$, and $\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{a})$ and $\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{a})$ are both continuous on B, then $\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{a}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{a})$. Therefore, under simple conditions, partial differentiation with respect to possibly distinct variables is commutative.

Consider the function of two variables f defined on the domain D, and the surface $\mathcal{S} = \{(x, y, f(x, y)) \mid (x, y) \in D\}$. The directional vector between $\mathbf{p}_1 \in D$ and $\mathbf{p}_2 \in D$ is defined as the vector $\mathbf{p}_2 - \mathbf{p}_1$. Consider the directional vector $\mathbf{d}$ between $\mathbf{p} = (x_0, y_0, f(x_0, y_0))$ and $(x_0 + h, y_0, f(x_0 + h, y_0))$:

$$\mathbf{d} = (x_0 + h, y_0, f(x_0 + h, y_0)) - (x_0, y_0, f(x_0, y_0)) = (h, 0, f(x_0 + h, y_0) - f(x_0, y_0)).$$

Intuitively, this vector represents the direction in which an increase of h in x_0 takes the value of f, when y_0 is kept constant. For $h \neq 0$, the vector $\frac{\mathbf{d}}{h}$ has the same direction as $\mathbf{d}$:

$$\begin{aligned} \frac{\mathbf{d}}{h} &= \frac{(h, 0, f(x_0 + h, y_0) - f(x_0, y_0))}{h} \\ &= \Big(1, 0, \frac{f(x_0 + h, y_0) - f(x_0, y_0))}{h}\Big). \end{aligned}$$

Therefore,

$$\lim_{h \to 0} \frac{\mathbf{d}}{h} = \Big(1, 0, \frac{\partial}{\partial x} f(x_0, y_0)\Big) = \mathbf{u},$$

if the partial derivative with respect to x is defined at (x_0, y_0). Analogously, let $\mathbf{v} = \Big(0, 1, \frac{\partial}{\partial y} f(x_0, y_0)\Big)$. The vectors $\mathbf{u}$ and $\mathbf{v}$ are clearly linearly independent, and so they span a plane in $\mathbb{R}^3$.

Consider the vector $\mathbf{w} = \Big(\frac{\partial}{\partial x} f(x_0, y_0), \frac{\partial}{\partial y} f(x_0, y_0), -1\Big)$. Clearly, $\mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$. The plane tangent to the surface $\mathcal{S}$ at $(x_0, y_0, f(x_0, y_0))$ is defined as the set of $\mathbf{t} \in \mathbb{R}^3$ such that $\mathbf{w} \cdot (\mathbf{t} - \mathbf{p}) = 0$. Therefore, (x, y, z) belongs to the plane tangent to the surface $\mathcal{S}$ at $(x_0, y_0, f(x_0, y_0))$ if:

$$z = \frac{\partial}{\partial x} f(x_0, y_0)(x - x_0) + \frac{\partial}{\partial y} f(x_0, y_0)(y - y_0) + f(x_0, y_0),$$

when these partial derivatives are defined. The equation above can also be used to approximate f at a point near (x_0, y_0) by the corresponding tangent plane, and can be generalized to any number of variables.

Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ of multiple variables $x_1, \dots x_n$ and differentiable functions $g_i : \mathbb{R} \to \mathbb{R}$ of a single variable t, for every $1 \leq i \leq n$. The chain rule for partial derivatives states that

$$\frac{d}{dt} f(g_1(u), \dots, g_n(u)) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(g_1(u), \dots, g_n(u)) \frac{d}{dt} g_i(u).$$

Letting $x_i = g_i(t)$ and $z = f(x_1, \dots, x_n)$,

$$\frac{dz}{dt} = \sum_{i=1}^{n} \frac{\partial z}{\partial x_i} \frac{dx_i}{dt}.$$

The statement above is only valid if changes in t affect z only through changes in some of the x_i, and if changes in x_i do not directly affect x_j for $i \neq j$.

More generally, consider the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ of n variables $x_1, \dots, x_n$ and the differentiable functions $g_i : \mathbb{R}^m \to \mathbb{R}$ of m variables $t_1, \dots, t_m$, for $1 \leq i \leq n$. The chain rule also states that

$$\frac{\partial}{\partial t_j} f(g_1(u_1, \dots, u_m), \dots, g_n(u_1, \dots, u_m)) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(g_1(u_1, \dots, u_m), \dots, g_n(u_1, \dots, u_m)) \frac{\partial}{\partial t_j} g_i(u_1, \dots, u_m),$$

for every $1 \leq j \leq m$. Letting $x_i = g_i(t_1, \dots, t_m)$ and $z = f(x_1, \dots, x_n)$:

$$\frac{\partial z}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial z}{\partial x_i} \frac{\partial x_i}{\partial t_j}.$$

The statement above is only valid if changes in t_j affect z only through changes in some of the x_i, and if changes in x_i do not directly affect x_j for $i \neq j$.

The gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ of multiple variables $x_1, \dots, x_n$ is defined by

$$\nabla f(\mathbf{a}) = \Big(\frac{\partial}{\partial x_1} f(\mathbf{a}), \dots, \frac{\partial}{\partial x_n} f(\mathbf{a}) \Big).$$

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and a unit vector $\mathbf{u}$. The derivative $D_{\mathbf{u}} f(\mathbf{a})$ of f in the direction $\mathbf{u}$ at point $\mathbf{a}$ is defined as:

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

If f is differentiable, it can be shown that

$$D_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}.$$

Because $\mathbf{u}$ is a unit vector, $D_{\mathbf{u}} f(\mathbf{a}) = ||\nabla f(\mathbf{a})|| \cos \theta$, where θ is the angle between $\nabla f(\mathbf{a})$ and $\mathbf{u}$. Therefore, $\frac{\nabla f(\mathbf{a})}{||\nabla f(\mathbf{a})||}$ is the direction of maximum instantaneous increase of f at $\mathbf{a}$. Similarly, if $\nabla f(\mathbf{a}) \cdot \mathbf{u} = 0$, then $\mathbf{u}$ is orthogonal to $\nabla f(\mathbf{a})$ and f has derivative zero at $\mathbf{a}$ in the $\mathbf{u}$ direction.

A function f of multiple variables $x_1, \dots, x_n$ has a local minimum at $\mathbf{a}$ if $f(\mathbf{a}) \leq f(\mathbf{x})$ for every $\mathbf{x}$ in an open ball around $\mathbf{a}$. A local maximum is defined analogously. If $f(\mathbf{a}) \leq f(\mathbf{x})$ for every $\mathbf{x}$ in a given domain, then f has a global minimum in $\mathbf{a}$.

The following theorem is extremely important in practical applications of calculus. If f has a local minimum or maximum at $\mathbf{a}$, then $\nabla f(\mathbf{a}) = (\frac{\partial}{\partial x_1} f(\mathbf{a}), \dots, \frac{\partial}{\partial x_n} f(\mathbf{a})) = 0$. The converse statement is not generally true.

Every $\mathbf{a}$ for which $\nabla f(\mathbf{a}) = 0$ is called a critical point. A critical point that is not a local minimum or maximum is called a saddle point.

Consider the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ of variables $x_1, \dots, x_n$ for which the second partial derivatives also exist. The Hessian matrix $H(\mathbf{a})$ of f at $\mathbf{a}$ is defined as

$$H(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f(\mathbf{a}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(\mathbf{a}) & \cdots & \frac{\partial^2}{\partial x_n^2} f(\mathbf{a}) \end{pmatrix}.$$

If $\mathbf{a}$ is a critical point of f, the Hessian matrix of f at $\mathbf{a}$ is useful to determine whether $\mathbf{a}$ is a local minimum, maximum or saddle point. If all eigenvalues of H are positive, then $\mathbf{a}$ is a local minimum. If all eigenvalues of H are negative, then $\mathbf{a}$ is a local maximum. If H has negative and positive eigenvalues, then $\mathbf{a}$ is a saddle point. In other cases, the test is inconclusive. This is the generalization of the second derivative test presented in a previous section.

Finding the global minima (or maxima) of a function f often requires evaluating its local minima and the points on the boundary of its domain D (points for which no open ball is contained in D).

Many problems of practical interest can be modeled as a search for $\mathbf{a} = (a_1, \dots, a_n)$ such that $f(\mathbf{a}) = \min_{\mathbf{a}' \in \mathbb{R}^n} f(\mathbf{a}')$ under the restriction that $F(\mathbf{a}) = 0$. Such a minimization problem can often be solved by eliminating one degree of freedom. Concretely, suppose it is possible to write $x_n = h(x_1, \dots, x_{n-1})$ for any $\mathbf{x} \in \mathbb{R}^n$ such that $F(\mathbf{x}) = 0$. The minimization problem becomes finding $(a_1, \dots, a_{n-1})$ such that

$$f(a_1, \dots, a_{n-1}, h(a_1, \dots, a_{n-1})) = \min_{\mathbf{a}' \in \mathbb{R}^{n-1}} f(a'_1, \dots, a'_{n-1}, h(a'_1, \dots, a'_{n-1})),$$

with no additional restrictions, and letting $a_n = h(a_1, \dots, a_{n-1})$.

A constrained minimization problem is defined by a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and differentiable functions $g_1, \dots, g_m$. The set of feasible solutions is $X = \{\mathbf{x} \in \mathbb{R}^n \mid \forall i \in \{1, \dots, m\}, g_i(\mathbf{x}) = 0\}$. The vector $\mathbf{a} \in X$ is a local minimum subject to constraints if $\mathbf{a} \leq \mathbf{x}$, for every $\mathbf{x} \in X \cap B$, for every open ball B centered on $\mathbf{a}$. Notice that a local minimum subject to constraints is not necessarily a local minimum of f.

The Lagrange multiplier theorem states the following. If $\mathbf{a}$ is a local minimum subject to the constraints of the minimization problem defined above and the vectors $\nabla g_1(\mathbf{a}), \dots, \nabla g_m(\mathbf{a})$ are linearly independent, then there exist unique $\lambda_1, \dots, \lambda_m$ such that

$$\nabla f(\mathbf{a}) = \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{a}),$$

or, equivalently, $\nabla f(\mathbf{a}) \in \operatorname{span}(\nabla g_1(\mathbf{a}), \dots, \nabla g_m(\mathbf{a}))$.

The system of equations above can be used to find candidate solutions to the constrained minimization problem. It is not trivial to determine which candidates are in fact local minima subject to constraints, although it may suffice to know which are the global minima subject to constraints.

6 Multiple integrals

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ over variables x and y. Let f be defined on a rectangular region $R = \{(x, y) \mid x \in [a, b], y \in [c, d]\} = [a, b] \times [c, d]$, where a, b, c and d are constants. Let m and n be positive integers, and let $\Delta x = \frac{b-a}{m}$ and $\Delta y = \frac{d-c}{n}$. Consider also the rectangle $R_{i,j} \subseteq R$ defined as $R_{i,j} = [a+(i-1)\Delta x, a+i\Delta x] \times [c+(j-1)\Delta y, c+j\Delta y]$.

The double integral of f over R is defined as

$$\iint\limits_R f(x, y) \, dx \, dy = \lim_{m,n \to \infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x^*_{i,j}, y^*_{i,j}) \Delta x \Delta y,$$

when the limit in the right exists and is the same for all choices of $(x^*_{i,j}, y^*_{i,j}) \in R_{i,j}$, for all i and j. In that case, f is integrable over R. If f is continuous on R, then f is integrable over R.

Intuitively, if $f(x, y) \geq 0$ for all $(x, y) \in R$, the double integral of f over R can be interpreted as the volume below the surface of f over R.

Consider integrable functions f and g over R and let c be a constant. The following properties hold:

$$\iint\limits_R [f(x, y) + g(x, y)] \, dx \, dy = \iint\limits_R f(x, y) \, dx \, dy + \iint\limits_R g(x, y) \, dx \, dy$$

$$\iint\limits_R c f(x, y) \, dx \, dy = c \iint\limits_R f(x, y) \, dx \, dy$$

Furthermore, if $f(x, y) \geq g(x, y)$ for all $(x, y) \in R$, then $\iint\limits_R f(x, y) \, dx \, dy \geq \iint\limits_R g(x, y) \, dx \, dy$.

The following statement follows from Fubini's theorem. If f is continuous on $R = [a, b] \times [c, d]$, where a, b, c and d are constants, then

$$\iint\limits_R f(x, y)\, dx\, dy = \int_c^d \Big[\int_a^b f(x, y)\, dx \Big]\, dy = \int_a^b \Big[\int_c^d f(x, y)\, dy \Big]\, dx.$$

The two equations to the right are referred to as iterated integrals. Notice the commutativity under change of the limits of integration, which are constants. The square brackets are usually omitted when there is no ambiguity.

If f is continuous on $R = [a, b] \times [c, d]$ and $f(x, y) = g(x)h(y)$, then

$$\int\limits_R f(x, y)\, dx\, dy = \int\limits_R g(x)h(y)\, dx\, dy = \int_a^b g(x)\, dx \int_c^d h(y)\, dy.$$

Consider the rectangle $R = [a, b] \times [c, d]$, where a, b, c and d are constants, and let $D \subseteq R$. Consider also a continuous function f over variables x and y, defined on D. The double integral of f over D is defined as

$$\iint\limits_D f(x, y)\, dx\, dy = \iint\limits_R F(x, y)\, dx\, dy,$$

where

$$F(x, y) = \begin{cases} f(x, y) & \text{if } (x, y) \in D \\ 0 & \text{otherwise.} \end{cases}$$

This is a generalization of double integration to an arbitrary region D bounded by a rectangle R. Since F is not necessarily continuous, iterated integration does not apply directly.

Consider the region $D = D_1 \cup D_2$, with $D_1 \cap D_2$ containing any subset of the boundary points of the two sets. If f is integrable over D_1 and D_2, then

$$\iint\limits_D f(x, y)\, dx\, dy = \iint\limits_{D_1} f(x, y)\, dx\, dy + \iint\limits_{D_2} f(x, y)\, dx\, dy.$$

Consider a rectangle $R = [a, b] \times [c, d]$ for constants a, b, c and d. Consider also $D \subseteq R$ such that $D = \{(x, y) \mid x \in [a, b], g(x) \leq y \leq h(x)\}$, for continuous functions g and h. The double integral of f over D is given by

$$\iint\limits_D f(x, y)\, dx\, dy = \int_a^b \Big[\int_{g(x)}^{h(x)} f(x, y)\, dy \Big]\, dx.$$

Therefore, the integral over a region D that is between two continuous functions of x can also be expressed as an iterated integral. Notice that the commutativity of iterated integration does not appy in this case, because the limits of integration in the innermost integral are functions of x. An analogous property also holds for a region D that is between two continuous functions of y.

It is often possible to obtain the double integral of a function over a region by partitioning it into subregions and combining the resulting double integrals. It is also possible to perform double integration in polar coordinates, although the details are omitted in this text.

We now present a generalization of double integrals to functions of more than two variables.

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ over variables $\mathbf{x} = (x_1, \dots, x_n)$. Let f be defined on the hyperrectangular region $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$, where a_i and b_i are constants for every i. Consider also positive integers m_i, for $i \in \{1, \dots, n\}$, and let $\Delta x_i = \frac{b_i - a_i}{m_i}$.

Consider also the hyperrectangles $R_{i_1,\dots,i_n} \subseteq R$ defined as

$$R_{i_1,\dots,i_n} = [a_1 + (i_1 - 1)\Delta x_1, a_1 + i_1 \Delta x_1] \times \cdots \times [a_n + (i_n - 1)\Delta x_n, a_n + i_n \Delta x_n],$$

for $i_j \in \{1, \dots, m_j\}$.

The multiple integral of f over R is defined as

$$\int_R f(\mathbf{x})\ d\mathbf{x} = \int \cdots \int\limits_R f(x_1, \dots, x_n)\, dx_1 \dots\, dx_n = \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} f(x^*_{1,i_1,\dots,i_n}, \dots, x^*_{n,i_1,\dots i_n}) \Delta x_1 \cdots \Delta x_n,$$

when the limit in the right exists and is the same for all choices of $(x^*_{1,i_1,\ldots,i_n}, \ldots, x^*_{n,i_1,\ldots i_n}) \in R_{i_1,\ldots,i_n}$. In that case, f is integrable over R. If f is continuous on R, then f is integrable over R.

Consider integrable functions f and g over R and let c be a constant. The following properties hold:

$$\int_R f(\mathbf{x}) + g(\mathbf{x})\ d\mathbf{x} = \int_R f(\mathbf{x})\ d\mathbf{x} + \int_R g(\mathbf{x})\ d\mathbf{x}$$

$$\int_R cf(\mathbf{x})\ d\mathbf{x} = c\int_R f(\mathbf{x})\ d\mathbf{x}$$

Also, if $f(\mathbf{x}) \geq g(\mathbf{x})$ for any $\mathbf{x} \in R$, then $\int_R f(\mathbf{x})\ d\mathbf{x} \geq \int_R g(\mathbf{x})\ d\mathbf{x}$.

Iterated integrals can also be used to compute multivariate integrals.

If f is continuous on $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$, where a_i and b_i are constants for every i, then

$$\int_R f(\mathbf{x})\ d\mathbf{x} = \int_{a_1}^{b_1} \cdots \Big[\int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n\Big] \cdots dx_1.$$

Multiple integrals can also be defined over more general regions, although the details are omitted in this text.

License

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

References

[1] Stewart, James. *Calculus: Early Transcendentals, 7th edition*, 2012.

[2] Apostol, Tom M. *Calculus, Volume 1, 2nd edition*, Wiley, 1991.

[3] Apostol, Tom M. *Calculus, Volume 2, 2nd edition*, Wiley, 1969.

[4] Bertsekas, Dimitri P. *Nonlinear Programming, 2nd edition*, Athena Scientific, 1995.