I. BASIC CALCULUS

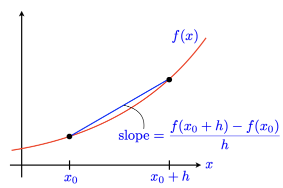
1 Differentiation

We begin with a recap of the main ideas of differentiation. Much of the material in this section will likely already be familiar to you.

Definition (Derivative of a function). We define the *derivative* of a function $f(x)$ with respect to its argument x as the *function* given by the *limit*

$$\frac{df}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \,. \tag{1}$$

As shown in the figure to the right, the value of df/dx at argument $x = x_0$ is the slope of the graph of $f(x)$ at the point x_0, and so determines the rate of change of $f(x)$ with respect to x there.

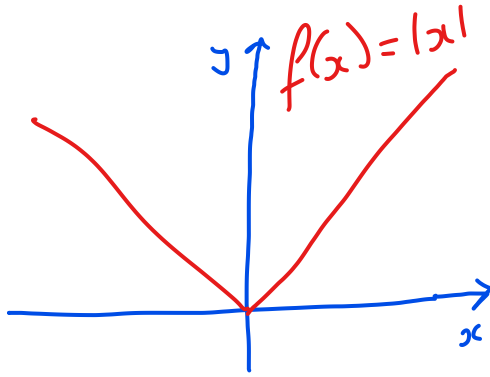


For the function $f(x)$ to be *differentiable* at the point x_0, and so for the function df/dx to be well-defined there, the left-hand limit (i.e., h is negative and approaches zero from below) and the right-hand limit (h is positive and approaches zero from above) must be defined and equal:

$$\lim_{h\to 0^-} \frac{f(x_0+h) - f(x_0)}{h} = \lim_{h\to 0^+} \frac{f(x_0+h) - f(x_0)}{h} \,.$$

This is actually quite a strong restriction on the "smoothness" of the function $f(x)$. As an example, $f(x) = |x|$ is not differentiable at $x = 0$ since

$$\lim_{h\to 0^-} \frac{|h|}{h} = -1 \quad \text{but} \quad \lim_{h\to 0^+} \frac{|h|}{h} = +1 \,.$$



Aside: limits

You will see much more about limits next term in *Analysis I*.

Informally, if $\lim_{x\to x_0} f(x) = A$, then $f(x)$ can be made arbitrarily close to A by making x sufficiently close to x_0. Note that we do *not* require $f(x_0)$ to equal A (or even to exist) – the limit is a statement about the behaviour of the function in the vicinity of x_0, but not actually at x_0.

A little more formally, for a function $f(x)$ defined on some open interval containing x_0 (but not necessarily at x_0), $\lim_{x\to x_0} f(x) = A$ means that

$$\boxed{\begin{array}{l}\text{for any } \epsilon > 0, \text{ there exists } \delta > 0 \text{ such that} \\ |f(x) - A| < \epsilon \text{ for all } 0 < |x - x_0| < \delta.\end{array}}$$

The right-hand limit, for example, is defined similarly but with $0 < |x - x_0| < \delta$ replaced by $0 < x - x_0 < \delta$.

We can also define limits at infinity. For example, $\lim_{x\to\infty} f(x) = A$ means that

$$\boxed{\begin{array}{l}\text{for any } \epsilon > 0, \text{ there exists } X > 0 \text{ such that} \\ |f(x) - A| < \epsilon \text{ for all } x > X.\end{array}}$$

Various properties of limits will be proven in *Analysis I*. Here, we simply state some of the most important properties.

- If $f(x)$ has a limit at a point, it is unique.
- If $\lim_{x\to x_0} f(x) = A$ and $\lim_{x\to x_0} g(x) = B$, then:
  - $\lim_{x\to x_0}[f(x) + g(x)] = A + B$;
  - $\lim_{x\to x_0}[f(x)g(x)] = AB$;
  - $\lim_{x\to x_0}[f(x)/g(x)] = A/B$ for $B \neq 0$.

If $B = 0$, the limit of the quotient does not exist if $A \neq 0$, but *may* exist in the *indeterminate* case $A = B = 0$.

The notation df/dx for the derivative of a function is due to Leibniz. Notice how the denominator shows what the argument of the function is. Other widely used notations for the derivative include $f'(x)$, due to Lagrange, and $\dot{f}$, due to Newton and usually reserved for differentiation with respect to time.

For sufficiently smooth functions, we can define higher derivatives recursively. For example, for the second derivative

$$\frac{d}{dt}\left(\frac{df}{dt}\right) = \frac{d^2 f}{dt^2} = f''(t) = \ddot{f}(t)\,.$$

For the nth derivative, the notation

$$\frac{d^n f}{dx^n} = f^{(n)}(x)$$

is sometimes used.

1.1 Big-O and little-o notation

Two very useful concepts in applied mathematics are big O (pronounced "Oh") and little o, which is sometimes written $\underline{o}$ to distinguish clearly between the two symbols.

These two concepts, sometimes called *order parameters*, are used to give comparative scalings between functions sufficiently close to some limiting point x_0 (which may be ∞).

Definition (O and o notations).

1. $f(x)$ is $o[g(x)]$ as $x \to x_0$ if

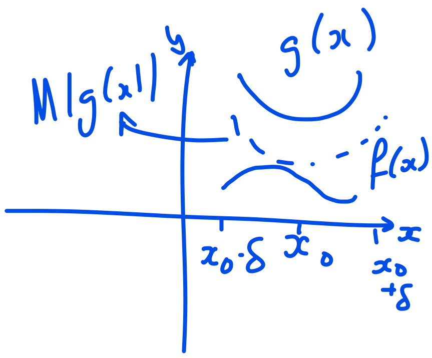
   $$\lim_{x\to x_0} \frac{f(x)}{g(x)} = 0\,,$$

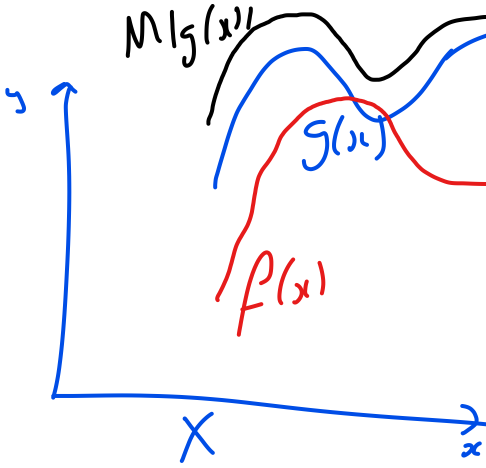
   commonly written as $f(x) = o[g(x)]$. Informally, this means that "$f(x)$ is much smaller than $g(x)$ as $x \to x_0$".

2. $f(x)$ is $O[g(x)]$ as $x \to x_0$ if $f(x)/g(x)$ is bounded as $x \to x_0$, i.e., there exists $\delta > 0$ and $M > 0$ such that for all x with $0 < |x - x_0| < \delta$,

   $$|f(x)| \le M|g(x)|\,.$$

   This is commonly written as $f(x) = O[g(x)]$.





These ideas can be extended to the behaviour at infinity. For example, $f(x)$ is $O[g(x)]$ as $x \to \infty$ if there exists $M > 0$ and $X > 0$ such that for all $x > X$ we have $|f(x)| \leq M|g(x)|$.

Note that there is an abuse of notation here when writing, for example, $f(x) = O[g(x)]$, since $O[g(x)]$ is not a function. Rather, we mean that $f(x)$ and $g(x)$ are in a class of functions with the required property of varying in a particular way as x_0 is approached. Writing $f(x) \in O[g(x)]$ is more appropriate, but $f(x) = O[g(x)]$ is commonplace.

Note from the definitions, $f = o(g) \Rightarrow f = O(g)$ but not vice versa. For example, if $f(x) = 2x$, we have $f(x) = O(x)$ since $f(x)/x = 2$, but then $f(x) \neq o(x)$.

Example. The function x^2 is $o(x)$ as $x \to 0$ since

$$\lim_{x\to 0} \frac{x^2}{x} = \lim_{x\to 0} x = 0\,.$$

Example. The function $x^2 + x$ is $O(x^2)$ as $x \to \infty$. This follows since for $x > 1$ we have $x^2 > x$, so that

$$|x^2 + x| < 2|x^2| \quad \text{for } x > 1\,.$$

Generally, a polynomial with largest power x^n will be $O(x^n)$ (or any larger power of x) as $x \to \infty$.

Some further examples:

- $x = o(\sqrt{x})$ as $x \to 0$;
- $\sin 2x = O(x)$ as $x \to 0$ as $\sin 2x \approx 2x$ for small x;
- $\sqrt{x} = o(x)$ as $x \to \infty$; and
- $\cos(x) = O(1)$ for all x as $|\cos x| \leq 1$.

Order parameters are frequently used in calculus to classify remainder terms before taking a limit. For example, we can write Eq. (1) as

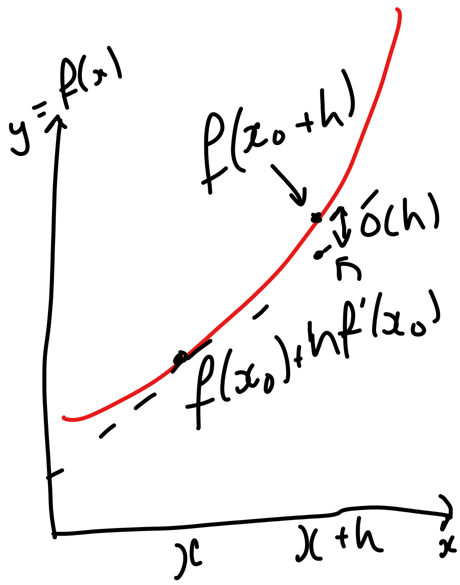
$$\left.\frac{df}{dx}\right|_{x_0} = \frac{f(x_0 + h) - f(x_0)}{h} + \frac{o(h)}{h} \quad \text{as } h \to 0\,. \qquad (2)$$

We can establish the truth of this by taking the limit $h \to 0$ of both sides. The left-hand side does not depend on h, while the limit of the first term on the right is, by definition, df/dx at x_0. It follows that the limit of the (final) remainder term, $\lim_{h\to 0} o(h)/h = 0$, which is consistent with the little-o notation.

Multiplying Eq. (2) across by h, we obtain

$$f(x_0+h) = f(x_0) + h \left.\frac{df}{dx}\right|_{x_0} + o(h) \quad \text{as } h \to 0\,. \qquad (3)$$

This identifies $f(x_0+h)$ with the value given by the tangent line at the point x_0 plus a remainder that is $o(h)$.



1.2 Rules of differentiation

Let us remind ourselves of the several useful rules of differentiation, and how they arise from the fundamental definitions presented above.

1.2.1 Chain rule

Consider the case where we want to differentiate a "function of a function" of the independent variable, i.e., $f(x) = F\left(g(x)\right)$. For example, we might have $f(x) = \sin(x^2 - x + 2)$, where $F(X) = \sin(X)$ and $g(x) = x^2 - x + 2$.

Theorem (Chain rule). Given $f(x) = F\left(g(x)\right)$, then

$$\frac{df}{dx} = F'\left(g(x)\right)\frac{dg}{dx} = \frac{dF}{dg}\frac{dg}{dx}\,. \qquad (4)$$

The first term on the right is the derivatve of the function F with respect to its argument, evaluated at $g(x)$.

For our specific example,

$$\frac{d}{dx}\sin(x^2 - x + 2) = \left[\cos(x^2 - x + 2)\right](2x - 1).$$

Proof (Chain rule). We have

$$\begin{aligned}\frac{df}{dx} &= \lim_{h\to 0}\frac{F(g(x+h)) - F(g(x))}{h}\\ &= \lim_{h\to 0}\frac{F(g(x)+hg'(x)+o(h)) - F(g(x))}{h},\end{aligned}$$

where we assume that g is differentiable. Now, if we write

$$X = g(x) \quad \text{and} \quad H = hg'(x) + o(h),$$

then

$$\begin{aligned}F\big(g(x)+hg'(x)+o(h)\big) &= F(X+H)\\ &= F(X) + HF'(X) + o(H),\end{aligned}$$

so that

$$\begin{aligned}\frac{df}{dx} &= \lim_{h\to 0}\frac{[hg'(x)+o(h)]\,F'(X) + o(H)}{h}\\ &= g'(x)F'(X) + \lim_{h\to 0}\frac{o(H)}{h}.\end{aligned}$$

The final term on the right is zero. To see this, consider the following two cases separately.

- $g'(x) = 0$. In this case, $H = o(h)$ as $h \to 0$ and so goes to zero "faster" than h as $h \to 0$. It follows that a term of $o(H)$ as $H \to 0$ goes to zero "faster still" as $h \to 0$ and so is certainly $o(h)$.
- $g'(x) \neq 0$. In this case, $H = O(h)$ but is certainly not $o(h)$ as $h \to 0$. This means that for small enough h, H is proportional to h and so a term of $o(H)$ goes to zero faster than linearly in h and so is $o(h)$ as $h \to 0$.

Finally, we recover the chain rule:

$$\frac{df}{dx} = g'(x)F'(g(x)).$$

1.2.2 Product rule

Consider the situation where $f(x) = u(x)v(x)$, i.e., f can be written as the product of two other functions u and v.

Theorem (Product rule). Given $f(x) = u(x)v(x)$,

$$\frac{df}{dx} = v\frac{du}{dx} + u\frac{dv}{dx}\,. \tag{5}$$

The proof of this is an exercise on *Examples Sheet 1.*

The *quotient rule* is a special case of the product rule, replacing $v \to 1/v$:

$$f = \frac{u}{v} \to f' = \frac{u'v - v'u}{v^2}\,. \tag{6}$$

1.2.3 Leibniz's rule

The product rule can be generalized to higher-order derivatives very straightforwardly by recursive application. Considering $f(x) = u(x)v(x)$, we have

$$\begin{aligned} f' &= u'v + uv'\,, \\ f'' &= u''v + u'v' + u'v' + uv'' \\ &= u''v + 2u'v' + uv''\,, \\ f''' &= u'''v + u''v' + 2u''v' + 2u'v'' + u'v'' + uv''' \\ &= u'''v + 3u''v' + 3u'v'' + uv'''\,. \end{aligned}$$

This should be reminscent to you of Pascal's triangle and the binomial theorem.

Theorem (Leibniz's rule). Given $f(x) = u(x)v(x)$,

$$\begin{aligned} f^{(n)}(x) &= \sum_{r=0}^{n} \binom{n}{r} u^{(n-r)}v^{(r)} \qquad (7) \\ &= u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{2!}u^{(n-2)}v'' \\ &+ \cdots + \frac{n!}{m!(n-m)!}u^{(n-m)}v^{(m)} + \cdots + uv^{(n)}\,. \end{aligned}$$

Here, recall, a superscript (n) denotes the nth derivative (with, for example, $u^{(0)} = u$), and the binomial coefficient,

$$\binom{n}{r} \equiv \frac{n!}{r!(n-r)!}\,, \tag{8}$$

denotes the number of combinations of r elements that can be taken from n elements without replacement.

Of course, the Leibniz rule relies on the functions u and v being n-times differentiable.

Proof (Leibniz's rule). We prove this by induction. Equation (7) reduces to the product rule when $n = 1$, and so is true in this case. We now show that it is true for $n+1$ if true for $n \geq 1$.

Differentiating the Leibniz rule for n with the product rule, we have

$$\begin{aligned} f^{(n+1)} &= \frac{d}{dx}\sum_{r=0}^{n}\binom{n}{r}u^{(n-r)}v^{(r)} \\ &= \sum_{r=0}^{n}\binom{n}{r}\left[u^{(n+1-r)}v^{(r)} + u^{(n-r)}v^{(r+1)}\right] . \end{aligned} \tag{9}$$

The last term on the right is

$$\begin{aligned} \sum_{r=0}^{n}\binom{n}{r}u^{(n-r)}v^{(r+1)} &= \sum_{r=0}^{n-1}\binom{n}{r}u^{(n-r)}v^{(r+1)} + uv^{(n+1)} \\ &= \sum_{r=1}^{n}\binom{n}{r-1}u^{(n+1-r)}v^{(r)} + uv^{(n+1)} , \end{aligned}$$

where we have relabelled $r \to r-1$ in passing to the second line. Combining with the first term on the right of Eq. (9), we have

$$\begin{aligned} f^{(n+1)} &= u^{(n+1)}v + \sum_{r=1}^{n}\left[\binom{n}{r} + \binom{n}{r-1}\right]u^{(n+1-r)}v^{(r)} + uv^{(n+1)} \\ &= u^{(n+1)}v + \sum_{r=1}^{n}\binom{n+1}{r}u^{(n+1-r)}v^{(r)} + uv^{(n+1)} \\ &= \sum_{r=0}^{n+1}\binom{n+1}{r}u^{(n+1-r)}v^{(r)} . \end{aligned} \tag{10}$$

Here, we have used *Pascal's rule*,

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1} ,$$

which follows from Pascal's triangle (or directly from the expression for the binomial coefficients, Eq. 8).

Equation (10) is Leibniz's rule for $n+1$, proving that if true for n, it is also true for $n+1$.

1.3 Taylor series

1.3.1 Taylor's theorem

Recall Eq. (3), which can be rewritten as

$$f(x_0+h) = f(x_0) + h \left.\frac{df}{dx}\right|_{x=x_0} + o(h) \quad \text{as } h \to 0\,.$$

Provided the first n derivatives of f exist, this can be generalised to *Taylor's theorem.*

Theorem (Taylor's theorem). For n-times differentiable $f(x)$, we have

$$f(x_0+h) = f(x_0) + h \left.\frac{df}{dx}\right|_{x=x_0} + \frac{h^2}{2!} \left.\frac{d^2f}{dx^2}\right|_{x=x_0} + \cdots + \frac{h^n}{n!} \left.\frac{d^nf}{dx^n}\right|_{x=x_0} + E_n\,, \quad (11)$$

where $E_n = o(h^n)$ as $h \to 0$.

In fact, if $f^{(n+1)}$ exists $\forall x \in (x_0, x_0+h)$ and $f^{(n)}$ is continuous in this range, it can be shown (see *Analysis I*) that $E_n = O(h^{n+1})$ as $h \to 0$. In particular,

$$E_n = \frac{f^{(n+1)}(x_n)}{(n+1)!} h^{n+1}\,, \quad (12)$$

for some x_n with $x_0 \le x_n \le x_0 + h$.

Note that $E_n = O(h^{n+1})$ is a stronger statement than $E_n = o(h^n)$. For example, for $0 < a < 1$,

$$h^{n+a} = o(h^n) \quad \text{as } h \to 0\,,$$

but

$$h^{n+a} \neq O(h^{n+1}) \quad \text{as } h \to 0\,.$$

Taylor's theorem is an exact statement that expresses the value of a function f at a point $x_0 + h$ in terms of the value of the function at x_0, its derivatives at x_0, and an error term E_n whose behaviour we know as h gets smaller.

1.3.2 Taylor Polynomials

If we write $x = x_0 + h$, then Eq. (11) can be rewritten as

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) + E_n\,. \quad (13)$$

The first $n+1$ terms on the right-hand side form the nth-order *Taylor polynomial* of $f(x)$ about the point $x = x_0$. Note that the coefficients of this polynomial are such that its first n derivatives match those of $f(x)$ at x_0.

The Taylor polynomial can be used to approximate functions in the vicinity of a point, with an error controlled by E_n. Note that this is a local approximation and does not necessarily tell us anything about the function far from the point (although it sometimes does).

Regarding the nth Taylor polynomial as a series, if the limit $n \to \infty$ exists (i.e., the series converges) we obtain the *Taylor series* of $f(x)$ about the point x_0.

1.4 L'Hôpital's rule

Taylor series representations are very useful to understand L'Hôpital's rule, which can be used to determine the value of limits of *indeterminate forms*.

Theorem (L'Hôpital's rule). Let $f(x)$ and $g(x)$ be differentiable at $x = x_0$, with continuous first derivatives there, and

$$\lim_{x\to x_0} f(x) = f(x_0) = 0 \quad \text{and} \quad \lim_{x\to x_0} g(x) = g(x_0) = 0\,.$$

Then if $g'(x_0) \neq 0$,

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}\,, \quad (14)$$

provided the limit on the right exists.

Proof. For this special case (the rule actually applies in much more general circumstances), we have

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)f'(x_0) + o(x - x_0) \\ &= 0 + (x - x_0)f'(x_0) + o(x - x_0)\,, \\ g(x) &= g(x_0) + (x - x_0)g'(x_0) + o(x - x_0) \\ &= 0 + (x - x_0)g'(x_0) + o(x - x_0)\,. \end{aligned}$$

$$\frac{\lim f' + o(x-x_0)/(x-x_0)}{\lim g' + o(x-x_0)/(x-x_0)}$$

It follows that

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f' + o(x - x_0)/(x - x_0)}{g' + o(x - x_0)/(x - x_0)} = \frac{f'(x_0)}{g'(x_0)}\,,$$

where in the last step we have used $g'(x_0) \neq 0$. Finally, since the first derivatives were assumed continuous at $x = x_0$, we have

$$\frac{f'(x_0)}{g'(x_0)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}\,.$$

The rule can be generalised to higher orders. For example, if $f(x_0) = f'(x_0) = 0$ and if $g(x_0) = g'(x_0) = 0$, then

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f''(x)}{g''(x)}\,,$$

by just applying l'hôpital again.

provided that the limit exists.

As a concrete example, consider

$$f(x) = 3\sin x - \sin 3x \quad \text{and} \quad g(x) = 2x - \sin 2x\,.$$

For these functions, $f(0) = g(0) = f'(0) = g'(0) = f''(0) = g''(0) = 0$. As an exercise, show that

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = 3 = \lim_{x\to 0} \frac{f'''(x)}{g'''(x)}\,.$$

2 Integration

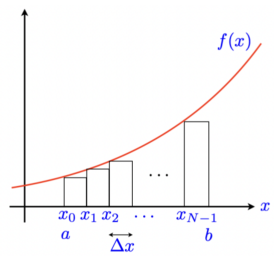
You will be familiar with integration as the "area under a curve" and also as the inverse of differentiation. We shall review both concepts in this section, as well as recapping some useful methods for integrating functions.

2.1 Integrals as Riemann sums

The "area under the curve" concept of an integral can be formalised as a *Riemann sum*. This is another topic that will be dealt with in detail in *Analysis I*, but let us briefly explore the idea here.

Definition (Integral). The *integral* of a (suitably well-defined) function $f(x)$ is the limit of a sum, e.g.,

$$\int_a^b f(x)\,dx \equiv \lim_{\Delta x\to 0}\sum_{n=0}^{N-1} f(x_n)\Delta x = \lim_{N\to\infty}\sum_{n=0}^{N-1} f(x_n)\Delta x\,, \tag{15}$$

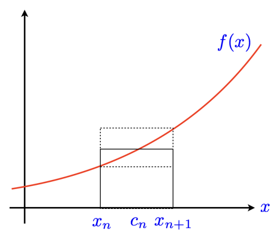


where $\Delta x = (b-a)/N$ and $x_n = a + n\Delta x$, as shown in the figure to the right.

The essence of the Riemann sum definition of the integral is that the limit should not depend exactly on how the rectangles are chosen (e.g., their widths do not have to be uniform, provided they all go to zero as $N \to 0$).

We now demonstrate that for sufficiently well-behaved functions, Eq. (15) coincides with the familar "area under the curve". We do this by first considering each rectangle in turn for finite N. Provided that $f(x)$ is continuous, the area under the curve between x_n and x_{n+1} is

$$A_n = (x_{n+1} - x_n) f(c_n)\,,$$



where $x_n \le c_n \le x_{n+1}$. This is an example of the *mean-value theorem*. We will not prove this here, but see the figure to the right for why this is plausible.

If $f(x)$ is differentiable, we have

$$\begin{aligned} f(c_n) &= f(x_n) + O(c_n - x_n) \quad \text{as } c_n - x_n \to 0 \\ &= f(x_n) + O(\Delta x) \quad (\text{since } c_n \le x_n + \Delta x)\,, \end{aligned}$$

from which it follows that

$$A_n = f(x_n)\Delta x + O(\Delta x^2) \quad \text{as } \Delta x \to 0\,.$$

Finally, the total area under the curve between $x = a$ and $x = b$ is

$$\begin{aligned}
A &= \lim_{N\to\infty} \sum_{n=0}^{N-1} A_n \\
&= \lim_{N\to\infty} \sum_{n=0}^{N-1} \left[f(x_n)\Delta x + O(\Delta x^2) \right] \\
&= \lim_{N\to\infty} \sum_{n=0}^{N-1} f(x_n)\Delta x + \lim_{N\to\infty} N O\left(\frac{(b-a)^2}{N^2} \right) \\
&= \lim_{N\to\infty} \sum_{n=0}^{N-1} f(x_n)\Delta x + \lim_{N\to\infty} O\left(\frac{(b-a)^2}{N} \right) \\
&= \lim_{N\to\infty} \sum_{n=0}^{N-1} f(x_n)\Delta x = \int_a^b f(x)\, dx \,.
\end{aligned}$$

2.2 Fundamental theorem of calculus

The concept of integration as the inverse of differentation is formulated in the *fundamental theorem of calculus*.

Theorem (Fundamental theorem of calculus). Let $F(x)$ be defined as

$$F(x) = \int_a^x f(t)\, dt \,.$$

Then

$$\frac{dF}{dx} = f(x) \,. \tag{16}$$

Proof. From the definition of the derivative, we have

$$\frac{dF}{dx} = \lim_{h\to 0} \frac{1}{h} \left[\int_a^{x+h} f(t)\, dt - \int_a^x f(t)\, dt \right] .$$

The two overlapping parts cancel, and so

$$\begin{aligned}\frac{dF}{dx} &= \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(t)\,dt \\ &= \lim_{h\to 0}\frac{1}{h}\left[f(x)h + O(h^2)\right] \\ &= \lim_{h\to 0}\left[f(x) + O(h)\right] \\ &= f(x)\,,\end{aligned}$$

where we used the mean-value theorem in passing to the second line.

Note that another way to interpret the Fundamental theorem of calculus is that the integral $F(x)$ is the solution of the differential equation

$$\frac{dF}{dx} = f(x) \quad \text{with } F(a) = 0\,. \tag{17}$$

As corollaries to the Fundamental theorem of calculus, we have

$$\frac{d}{dx}\int_x^b f(t)\,dt = -f(x)\,,$$

and, using the chain rule,

$$\frac{d}{dx}\int_a^{g(x)} f(t)\,dt = f(g(x))g'(x)\,.$$

Notation. We write *indefinite integrals* either as $\int f(x)\,dx$ or $\int^x f(t)\,dt$, where the unspecified lower limit gives rise to an integration constant.

2.3 Methods of integration

Integration is more difficult than differentiation. We cannot always evaluate integrals analytically in terms of simple (or not so simple!) functions. However, for those cases where we can, tricks such as integration by substitution or by parts are often helpful.

Identity	Term in integrand	Substitution
$\cos^2\theta + \sin^2\theta = 1$	$\sqrt{1-x^2}$	$x = \sin\theta$
$1 + \tan^2\theta = \sec^2\theta$	$1 + x^2$	$x = \tan\theta$
$\cosh^2 u - \sinh^2 u = 1$	$\sqrt{x^2-1}$	$x = \cosh u$
$\cosh^2 u - \sinh^2 u = 1$	$\sqrt{1+x^2}$	$x = \sinh u$
$1 - \tanh^2 u = \mathrm{sech}^2 u$	$1 - x^2$	$x = \tanh u$

Table 1: Useful trigonometric or hyperblic substitutions.

2.3.1 Integration by substitution

If the integrand contains a function of a function, integration by substitution is often useful.

Example. Consider

$$I = \int \frac{1-2x}{\sqrt{x-x^2}}\,dx\,.$$

Let $u = x - x^2$ so that $du = (1-2x)dx$; then

$$I = \int \frac{du}{\sqrt{u}} = 2\sqrt{u} + c = 2\sqrt{x-x^2} + c\,,$$

where c is a constant of integration.

Trigonometric (or hyperbolic) substitutions are often useful. If the function in the second column of Table 1 appears in the integrand, try proceeding by making the substitution in the third column and simplify using the idenity in the first column.

Example. Consider

$$I = \int \sqrt{2x - x^2}\,dx\,.$$

Since $2x - x^2 = 1 - (x-1)^2$, let us try $x - 1 = \sin\theta$ so that $dx = \cos\theta d\theta$. It follows that

$$\begin{aligned} I &= \int \cos^2\theta \, d\theta \\ &= \frac{1}{2}\int (1+\cos 2\theta)\, d\theta \\ &= \frac{1}{2}\theta + \frac{1}{4}\sin 2\theta + c \\ &= \frac{1}{2}\theta + \frac{1}{2}\sin\theta\cos\theta + c \\ &= \frac{1}{2}\text{arcsin}(x-1) + \frac{1}{2}(x-1)\sqrt{1-(x-1)^2} + c\,, \end{aligned}$$

where c is an integration constant and arcsin is the inverse sine function.

2.4 Integration by parts

"Integration by parts" exploits the product rule, which we write here in the form

$$uv' = (uv)' - u'v\,.$$

Integrating both sides gives rise to the following.

Theorem (Integration by parts). For functions u and v,

$$\int uv'\, dx = uv - \int u'v\, dx. \tag{18}$$

Example. Consider

$$I = \int_0^\infty xe^{-x}\, dx\,.$$

Let $u = x$ and $v' = e^{-x}$, so that $v = -e^{-x}$. We then have

$$\begin{aligned} I &= \left[-xe^{-x}\right]_0^\infty + \int_0^\infty e^{-x}\, dx \\ &= 0 + \left[-e^{-x}\right]_0^\infty = 1\,. \end{aligned}$$

Example. Consider

$$I = \int \ln x\, dx\,.$$

Let $u = \ln x$ and $v' = 1$, so that $v = x$. Integrating by parts gives

$$\begin{aligned} I &= x \ln x - \int x \left(\frac{1}{x}\right) dx \\ &= x \ln x - x + c\,, \end{aligned}$$

where c is an integration constant.

3 Partial differentiation

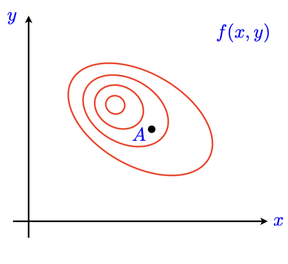
Here we generalize differentiation to functions of more than one variable.

3.1 Functions of several variables

Multivariate functions depend on more than one independent variable. Some physical examples include:

- the height of some terrain, which depends on both latitude and longitude;
- the density of air in this room, which depends on both position and time; and
- the energy of a thermodynamic system, which depends on its volume and temperature.

Considering a function of two variables, $f(x, y)$, we can represent it as a *contour plot* (see figure to the right) where f is constant on each contour line.



What is the slope of the function f at the point A? The answer obviously depends on the direction. The first thing to do is to figure out the slope along directions parallel to the coordinate axes, which leads us to the concept of *partial differentiation*.

3.2 Partial derivatives

Definition (Partial derivative). Given a function of several variables, for example, $f(x, y)$, the *partial derivative* of f with respect to x is the rate of change of f as x varies at fixed y. It is given by

$$\left.\frac{\partial f}{\partial x}\right|_y = \lim_{\delta x \to 0} \frac{f(x + \delta x, y) - f(x, y)}{\delta x} \, . \tag{19}$$

Note that the partial derivative of f with respect to x essentially corresponds to the slope of f experienced when moving purely left to right (in the positive x-direction).

Similarly, the partial derivative of $f(x, y)$ with respect to y is defined as the function

$$\left.\frac{\partial f}{\partial y}\right|_x = \lim_{\delta y \to 0} \frac{f(x, y + \delta y) - f(x, y)}{\delta y} \, , \tag{20}$$

i.e., the slope of f experienced when moving in the positive y-direction.

Note the "curly" partial derivative symbol, ∂, to distinguish from the ordinary derivative.

Example. Consider $f(x, y) = x^2 + y^3 + e^{xy^2}$. To compute the partial derivatives with respect to x, we simply hold y constant and differentiate regularly as if x were the only variable:

$$\left.\frac{\partial f}{\partial x}\right|_y = 2x + y^2 e^{xy^2} \, .$$

Similarly, for the partial derivative with respect to y,

$$\left.\frac{\partial f}{\partial y}\right|_x = 3y^2 + 2xye^{xy^2}.$$

We can also compute second derivatives

$$\begin{aligned} \left.\frac{\partial^2 f}{\partial x^2}\right|_y &= 2 + y^4 e^{xy^2} \, , \\ \left.\frac{\partial^2 f}{\partial y^2}\right|_x &= 6y + 2xe^{xy^2} + 4x^2y^2e^{xy^2} \, , \end{aligned}$$

as well as *mixed partial derivatives*

$$\frac{\partial}{\partial x}\left(\left.\frac{\partial f}{\partial y}\right|_x\right)\Bigg|_y = 2ye^{xy^2} + 2xy^3e^{xy^2}\,,$$
$$\frac{\partial}{\partial y}\left(\left.\frac{\partial f}{\partial x}\right|_y\right)\Bigg|_x = 2ye^{xy^2} + 2xy^3e^{xy^2}\,.$$

It is necessary to be careful to avoid ambiguity in which arguments specifically are being held fixed when the partial derivatives are being taken. However, it is often cumbersome to indicate this explicitly. In cases where *all* other variables are being held fixed, we shall omit the $|_y$, for example, and simply write $\partial f/\partial x$.

With this convention, for example,

$$\frac{\partial}{\partial x}\left(\left.\frac{\partial f}{\partial y}\right|_x\right)\Bigg|_y = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y}\,.$$

Notice that for the function in the example above,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x}\,, \qquad (21)$$

and so it does not matter in which order the partial derivatives are taken. This can be shown to hold true generally (*Schwarz's theorem* or *Clairaut's theorem*), provided that the function has continuous (mixed) second derivatives at the point of interest.

Finally, we note that an alternative subscript notation is sometimes used for partial derivatives. For example,

$$f_x \equiv \frac{\partial f}{\partial x}\,; \quad f_{xy} \equiv \frac{\partial^2 f}{\partial y \partial x}.$$

Note the ordering in the second case. The left-hand side should be interpreted as $(f_x)_y$, so that the partial derivative is first taken with respect to x before the resulting function is differentiated with respect to y. Of course, in most cases this is not important since $f_{xy} = f_{yx}$.

3.3 Multivariate chain rule

The chain rule (Eq. 4) tells us how to differentiate a "function of a function". How does this extend to multivariate functions? For example, for a given path $x(t)$ and $y(t)$, where t is a parameter along the path, the function $f(x, y)$ can be considered a function of t, i.e., $f((x(t), y(t))$. How do we calculate df/dt?

Consider the change in $f(x, y)$ under an arbitrary small displacement in any direction,

$$(x, y) \to (x + \delta x, y + \delta y)\,.$$

We have

$$\begin{aligned}\delta f &= f(x + \delta x, y + \delta y) - f(x, y)\\ &= [f(x + \delta x, y + \delta y) - f(x + \delta x, y)]\\ &\quad + [f(x + \delta x, y) - f(x, y)]\,.\end{aligned}$$

Taylor expanding the second term in square brackets, we have

$$f(x + \delta x, y) - f(x, y) = f_x(x, y)\delta x + o(\delta x)\,.$$

Similarly,

$$f(x+\delta x, y+\delta y)-f(x+\delta x, y) = f_y(x+\delta x, y)\delta y+o(\delta y)\,.$$

This involves the partial derivative f_y evaluated at $(x + \delta x, y)$. We can expand this about the point (x, y) using

$$f_y(x + \delta x, y) = f_y(x, y) + f_{yx}(x, y)\delta x + o(\delta x)\,.$$

Putting this together, we find

$$\begin{aligned}\delta f = [f_y(x, y) + f_{yx}(x, y)\delta x + o(\delta x)]\,\delta y + o(\delta y)\\ + f_x(x, y)\delta x + o(\delta x)\,.\end{aligned} \tag{22}$$

Taking the limit as $\delta x, \delta y \to 0$, and defining the *differential* of f as

$$df = \lim_{\delta x, \delta y \to 0} \delta f\,,$$

we obtain the multivariate chain rule in differential form.

Theorem (Chain rule for partial derivatives). The differential df of the function f is related to the differentials of its arguments dx and dy as

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy\,. \tag{23}$$

Note that the remaining terms in Eq. (22) all go to zero faster than δx or δy in the limit $\delta x, \delta y \to 0$.

We can now use the multivariate chain rule to calculate df/dt along the path $(x(t), y(t))$. Dividing by δt in Eq. (22) and then taking the limit, we have

$$\frac{d}{dt}f(x(t), y(t)) = \lim_{\delta t \to 0} \frac{\delta f}{\delta t} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\,. \tag{24}$$

Note that, since $x(t)$ and $y(t)$ are functions of t alone, their *ordinary* derivatives appear in this expression.

Another commonly occuring case is where the path is parameterised by one of the coordinates. For example, y may be specified as $y(x)$ so that along the path we have $f(x, y(x))$, which is a function of x. The rate of change of this function with respect to x is

$$\frac{df}{dx} = \frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}\,. \tag{25}$$

Note that f changes both because the first argument of the function (x) is changing *and* because the second argument (y) changes as x changes.

3.3.1 Integral form of the chain rule

The chain rule, Eq. (23), can be integrated along a path to get the change in the function f between the start- and end-points:

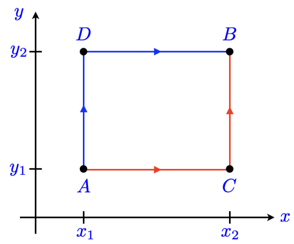
$$\Delta f = \int df = \int \left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy\right)\,.$$

If the path is parameterised as $(x(t), y(t))$, we have

$$\Delta f = \int \frac{df}{dt}\,dt = \int \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right) dt\,.$$

For given start- and end-points, the result will not depend on the particular path that is chosen.

As an example, consider integrating from point $A = (x_1, y_1)$ to $B = (x_2, y_2)$. If we consider the two paths shown in the figure to the right, for that going via $C = (x_2, y_1)$, we have



$$\begin{aligned}\Delta f &= \int_{x_1}^{x_2} \frac{\partial f}{\partial x}(x, y_1)\,dx + \int_{y_1}^{y_2} \frac{\partial f}{\partial y}(x_2, y)\,dy \\ &= [f(x_2, y_1) - f(x_1, y_1)] + [f(x_2, y_2) - f(x_2, y_1)] \\ &= f(x_2, y_2) - f(x_1, y_1)\,.\end{aligned}$$

For the path going via $D = (x_1, y_2)$, we have

$$\begin{aligned}\Delta f &= \int_{y_1}^{y_2} \frac{\partial f}{\partial y}(x_1, y)\,dy + \int_{x_1}^{x_2} \frac{\partial f}{\partial x}(x, y_2)\,dx \\ &= [f(x_1, y_2) - f(x_1, y_1)] + [f(x_2, y_2) - f(x_1, y_2)] \\ &= f(x_2, y_2) - f(x_1, y_1)\,,\end{aligned}$$

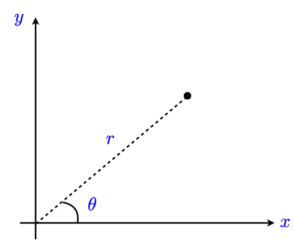
which is the same as going via C.

3.4 Applications of the multivariate chain rule

3.4.1 Change of variables

The chain rule naturally plays a central role when we change the independent variables, for example through a coordinate transformation.

Example. Consider transforming from Cartesian coordinates (x, y) to plane-polar coordinates (r, θ), with



$$x = r\cos\theta\,, \quad y = r\sin\theta\,. \tag{26}$$

The original function $f(x, y)$ can be thought of as a function of r and θ, i.e., $f(x(r, \theta), y(r, \theta))$. To compute the partial derivatives with respect to r and θ, we use the chain rule. For r, we have

$$\begin{aligned}\left.\frac{\partial f}{\partial r}\right|_\theta &= \left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial r}\right|_\theta + \left.\frac{\partial f}{\partial y}\right|_x \left.\frac{\partial y}{\partial r}\right|_\theta \\ &= \left.\frac{\partial f}{\partial x}\right|_y \cos\theta + \left.\frac{\partial f}{\partial y}\right|_x \sin\theta\,,\end{aligned}$$

and for θ,

$$\begin{aligned}\left.\frac{\partial f}{\partial \theta}\right|_r &= \left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial \theta}\right|_r + \left.\frac{\partial f}{\partial y}\right|_x \left.\frac{\partial y}{\partial \theta}\right|_r \\ &= -\left.\frac{\partial f}{\partial x}\right|_y r\sin\theta + \left.\frac{\partial f}{\partial y}\right|_x r\cos\theta\,.\end{aligned}$$

3.4.2 Implicit differentiation

Consider the expression $f(x, y, z) = c$, for some constant c. This defines a surface in 3D space, and so it *implicitly* defines a functional relationship between one of the coordinates x, y and z and the other two, i.e.,

$$z = z(x, y) \quad \text{or} \quad x = x(y, z) \quad \text{or} \quad y = y(x, z)\,.$$

Depending on the function $f(x, y, z)$, we may not be able to express these functional relationships in closed form. However, we can still evaluate their partial derivatives using implicit differentiation.

Example. Consider

$$xy + y^2 z + z^5 = 1\,. \tag{27}$$

Finding $x(y, z)$ is straightforward since x only appears linearly. To determine $y(x, z)$ we have to solve a quadratic equation. However, we cannot find $z(x, y)$ explicitly since this would require solving a quintic equation.

Despite this, we can still determine $\partial z/\partial x|_y$, for example, by taking the derivative of Eq. (27) with respect to x, holding y constant, using implicit differentiation:

$$y + y^2 \left.\frac{\partial z}{\partial x}\right|_y + 5z^4 \left.\frac{\partial z}{\partial x}\right|_y = 0\,,$$

so that

$$\left.\frac{\partial z}{\partial x}\right|_y = -\frac{y}{y^2+5z^4}\,.$$

Generally, given $f(x,y,z)=c$, the chain rule (extended to three variables) gives

$$0 = df = \left.\frac{\partial f}{\partial x}\right|_{y,z} dx + \left.\frac{\partial f}{\partial y}\right|_{x,z} dy + \left.\frac{\partial f}{\partial z}\right|_{x,y} dz\,.$$

Note that we cannot vary x, y and z independently as we must stay in the surface. We can find the rate of change of z with x at fixed y from

$$0 = \left.\frac{\partial f}{\partial x}\right|_{y,z} \underbrace{\left.\frac{\partial x}{\partial x}\right|_y}_{=1} + \left.\frac{\partial f}{\partial y}\right|_{x,z} \underbrace{\left.\frac{\partial y}{\partial x}\right|_y}_{=0} + \left.\frac{\partial f}{\partial z}\right|_{x,y} \left.\frac{\partial z}{\partial x}\right|_y\,.$$

Therefore

$$\left.\frac{\partial z}{\partial x}\right|_y = -\frac{\partial f/\partial x|_{y,z}}{\partial f/\partial z|_{x,y}}\,. \qquad (28)$$

We can similarly find that

$$\left.\frac{\partial x}{\partial y}\right|_z = -\frac{\partial f/\partial y|_{x,z}}{\partial f/\partial x|_{y,z}}\,, \qquad \left.\frac{\partial y}{\partial z}\right|_x = -\frac{\partial f/\partial z|_{x,y}}{\partial f/\partial y|_{x,z}}\,,$$

and so the relation

$$\left.\frac{\partial x}{\partial y}\right|_z \left.\frac{\partial y}{\partial z}\right|_x \left.\frac{\partial z}{\partial x}\right|_y = -1\,.$$

Note that normal *reciprocal rules* apply for partial derivatives, *provided* the same variables are being held constant. For example, for $f(x,y,z)=c$, similarly to Eq. (28) we have

$$\left.\frac{\partial x}{\partial z}\right|_y = -\frac{\partial f/\partial z|_{x,y}}{\partial f/\partial x|_{y,z}}\,,$$

so that

$$\left.\frac{\partial z}{\partial x}\right|_y = \frac{1}{\partial x/\partial z|_y}\,.$$

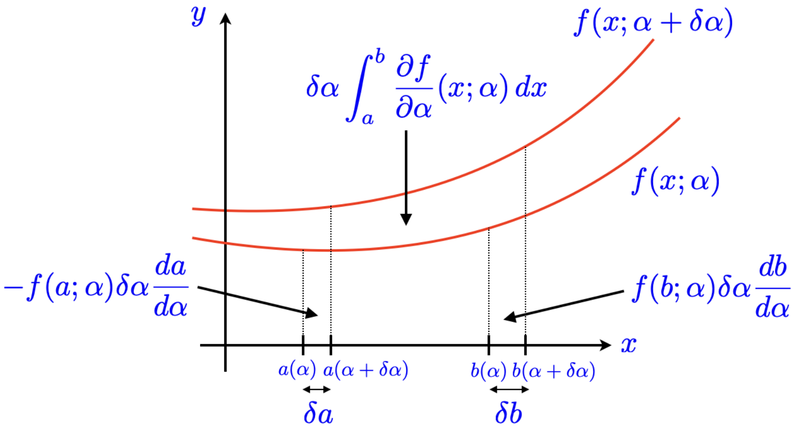


Figure 1: Contributions to the change in an integral when a parameter α appearing in the integrand, $f(x;\alpha)$, and the limits, $x = a(\alpha)$ and $x = b(\alpha)$, is varied.

3.4.3 Differentiation of an integral with respect to its parameters

Consider a family of functions $f(x;\alpha)$ where the parameter α labels the different members of the family. Define the integral

$$I(\alpha) = \int_{a(\alpha)}^{b(\alpha)} f(x;\alpha)\,dx\,,$$

where we have allowed the limits of the integral to depend on the parameter also. What is $dI/d\alpha$?

Theorem (Differentiation of an integral w.r.t. a parameter). The derivative of $I(\alpha)$ is

$$\frac{d}{d\alpha}\int_{a(\alpha)}^{b(\alpha)} f(x;\alpha)\,dx = \int_{a(\alpha)}^{b(\alpha)} \frac{\partial f}{\partial \alpha}(x;\alpha)\,dx + f(b;\alpha)\frac{db}{d\alpha} - f(a;\alpha)\frac{da}{d\alpha}\,. \quad (29)$$

Proof. We have

$$\begin{aligned}\frac{dI}{d\alpha} &= \lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\left[\int_{a(\alpha+\delta\alpha)}^{b(\alpha+\delta\alpha)} f(x;\alpha+\delta\alpha)\,dx \right.\\ &\qquad\qquad\qquad \left. - \int_{a(\alpha)}^{b(\alpha)} f(x;\alpha)\,dx\right]\\ &= \lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\left[\int_{a(\alpha)}^{b(\alpha)} f(x;\alpha+\delta\alpha) - f(x;\alpha)\,dx\right]\\ &\quad + \lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\int_{b(\alpha)}^{b(\alpha+\delta\alpha)} f(x;\alpha+\delta\alpha)\,dx\\ &\quad - \lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\int_{a(\alpha)}^{a(\alpha+\delta\alpha)} f(x;\alpha+\delta\alpha)\,dx\,. \end{aligned} \tag{30}$$

The first term on the right reduces to the integral of $\partial f/\partial\alpha$ between a and b. For the second term, we can use the mean-value theorem to write

$$\begin{aligned}\lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\int_{b(\alpha)}^{b(\alpha+\delta\alpha)} f(x;\alpha+\delta\alpha)\,dx &= \lim_{\delta\alpha\to 0}\left[f(\bar{x};\alpha+\delta\alpha)\right.\\ &\qquad \left.\times\left(\frac{b(\alpha+\delta\alpha)-b(\alpha)}{\delta\alpha}\right)\right],\end{aligned}$$

where, inside the limit, $b(\alpha)\le\bar{x}\le b(\alpha+\delta\alpha)$. Taking the limit gives

$$\lim_{\delta\alpha\to 0}\frac{1}{\delta\alpha}\int_{b(\alpha)}^{b(\alpha+\delta\alpha)} f(x;\alpha+\delta\alpha)\,dx = f(b;\alpha)\frac{db}{d\alpha}\,.$$

The third term in Eq. (30) is handled similarly, and putting the three terms together establishes Eq. (29). The origin of these three terms is illustrated in Fig. 1.

Example. Consider

$$I(\lambda) = \int_0^{\lambda} e^{-\lambda x^2}\,dx\,.$$

Then

$$\frac{dI}{d\lambda} = e^{-\lambda^3} - \int_0^{\lambda} x^2 e^{-\lambda x^2}\,dx\,.$$