

I like to do calculus using *differentials*. Differentials and the related *differential forms* are often used in applications, especially (but not only) to physics. The official textbook originally covers differentials incompletely and only in one minor application; it then uses them again for integration, primarily a notational convenience. But they are useful for much more. Now is the time to explain what they are.

Notation and terminology

If u is a variable quantity, then du is the **differential** of u . You can think of du as indicating an infinitely small (infinitesimal) change in the value of u , or the amount by which u changes when an infinitesimal change is made. A precise definition appears later in these notes.

Note that du is *not* d times u , and du is also *not* exactly a function of u . Rather, u (being a *variable* quantity) should itself be a function of some other quantities x, y, \dots , and du is also a function of a sort; so d is an *operator*: something that turns one function into another function. (However, an expression like $v du$ does involve multiplication: it is the quantity v multiplied by the differential of u .)

We often divide one differential by another; for example, $\frac{dy}{dx}$ is the result of dividing the differential of y by the differential of x . The textbook introduces this notation early to stand for the *derivative* of y with respect to x , and indeed it is that; but what the book doesn't tell you is that $\frac{dy}{dx}$ literally is dy divided by dx . Unfortunately, $\frac{d^2y}{dx^2}$, the second derivative, is *not* literally $d^2y = d(dy)$ divided by $dx^2 = (dx)^2$; for this reason, I prefer the notation $\left(\frac{d}{dx}\right)^2 y = \frac{d}{dx}\left(\frac{d}{dx}y\right) = \frac{d(dy/dx)}{dx}$.

Differentials and the rules of differentiation

One sometimes sees the Chain Rule expressed as

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx},$$

but the Chain Rule is a nontrivial fact that cannot be proved by simply cancelling factors. I prefer to state the Chain Rule as

$$df(u) = f'(u) du;$$

the point is that the *same* function f' appears regardless of which argument u we use.

Even this is more abstract than how the Chain Rule is applied. For example, suppose that you have discovered (say from the definition as a limit) that the derivative of $f(x) = \sin x$ is $f'(x) = \cos x$. Since $f'(x)$ may be defined as $\frac{df(x)}{dx}$, this derivative can be expressed in differential form without even bothering to name the functions involved:

$$(1) \quad d(\sin x) = \cos x dx.$$

Once you know this, you know something even more general:

$$(2) \quad d(\sin u) = \cos u du$$

for any other differentiable quantity u ; the Chain Rule is the power to derive equation (2) from equation (1). Thus, using $u = x^2$ (to continue the example),

$$d(\sin(x^2)) = \cos(x^2) d(x^2) = \cos(x^2)(2x dx) = 2x \cos(x^2) dx.$$

You may now divide both sides of this equation by dx if you wish, but the basic calculation involves only rules for differentials.

For the record, here are the rules for differentiation that you should already know, expressed using differentials:

- The Constant Rule: $dk = 0$ if k is constant.
- The Sum Rule: $d(u + v) = du + dv$.
- The Translate Rule: $d(u + k) = du$ if k is constant.
- The Difference Rule: $d(u - v) = du - dv$.
- The Product Rule: $d(uv) = v du + u dv$.
- The Multiple Rule: $d(ku) = k du$ if k is constant.
- The Quotient Rule: $d\left(\frac{u}{v}\right) = \frac{v du - u dv}{v^2}$.
- The Power Rule: $d(u^k) = ku^{k-1} du$ if k is constant.
- The Root Rule: $d(\sqrt[k]{u}) = \frac{\sqrt[k]{u} du}{ku^k}$ if k is constant.
- The Exponentiation Rule: $d(e^u) = e^u du$.
- The Logarithm Rule: $d(\ln u) = \frac{du}{u}$.
- The Sine Rule: $d(\sin u) = \cos u du$.
- The Cosine Rule: $d(\cos u) = -\sin u du$.
- The Tangent Rule: $d(\tan u) = \sec^2 u du$.
- The Cotangent Rule: $d(\cot u) = -\csc^2 u du$.
- The Secant Rule: $d(\sec u) = \tan u \sec u du$.
- The Cosecant Rule: $d(\csc u) = -\cot u \csc u du$.
- The Arcsine Rule: $d(\arcsin u) = \frac{du}{\sqrt{1-u^2}}$.
- The Arccosine Rule: $d(\arccos u) = -\frac{du}{\sqrt{1-u^2}}$.
- The Arctangent Rule: $d(\arctan u) = \frac{du}{u^2+1}$.
- The Arccotangent Rule: $d(\operatorname{arccot} u) = -\frac{du}{u^2+1}$.
- The Arcsecant Rule: $d(\operatorname{arcsec} u) = \frac{du}{|u|\sqrt{u^2-1}}$.
- The Arccosecant Rule: $d(\operatorname{arccsc} u) = -\frac{du}{|u|\sqrt{u^2-1}}$.
- The Chain Rule: $d(f(u)) = f'(u) du$ (where f is a differentiable function of one variable).
- The First Fundamental Theorem of Calculus: $d\left(\int_u^v f(t) dt\right) = f(v) dv - f(u) du$ (where f is a continuous function of one variable).

The last one might not be familiar to you in such a general form, but it can be handy.

Differential forms and smooth functions

Notice that every one of the rules above turns the differential on the left into a sum of terms (possibly only one term, or none in the case of the Constant Rule), each of which is an ordinary expression multiplied by a differential (or something algebraically equivalent to this). Such an expression is called a **differential form**, or more precisely an *exterior differential form of rank 1* (or a *1-form* for short) since there are other sorts of differential forms (some of which we will use later on).

It's time to actually give a *definition* of du . Since d is an operator, it must be applied to a function; so we should have $u = f(R)$ (by which I mean $u = f(x, y)$, $u = f(x, y, z)$, or whatever) for some function f . We say that the function f is **smooth** (infinitely differentiable) at some point P if, for every smooth parametrized curve C and real number $a \in \operatorname{dom} C$, if $C(a) = P$, then the composite function $f \circ C$ is smooth at a . (Notice that $f \circ C$ is an ordinary real-valued function of one real variable, so we know what it means for it to be infinitely differentiable.) If f is smooth at every point P in its domain, then we simply say

that f is *smooth*; then $f \circ C$ is smooth, for every smooth parametrized curve C , at every number a in the domain of C . In this case, we also say that the quantity $u = f(R)$ is *smooth* in R (that is smooth in x and y , smooth in x , y , and z , or whatever).

By the way, it doesn't matter whether f is defined on \mathbf{R}^2 or \mathbf{R}^3 or on some more abstract space; we only need to know what a smooth parametrized curve in the space is. However, I am only defining *infinite* differentiability here; to define mere *once* (or twice, etc) differentiability is actually more complicated, and it's not enough to know the smooth parametrized curves for that.

A very general notion of derivative of f is $(f \circ C)'(a)$, which we've just said exists when f and C are smooth and $a \in \text{dom } C$. More particularly, if f is smooth at P , then this defines an operation that, when given a smooth parametrized curve C and a real number a with $C(a) = P$, results in the real number $(f \circ C)'(a)$. Formally, we define the **differential** of f at P to be this operation:

$$C^*df(P)|_a = (f \circ C)'(a).$$

(You can think of the left-hand side of simply notation for this operation of df on C and a ; although C^* does have a meaning by itself, that would take us too far afield.)

If $u = f(R)$ is a differentiable quantity, then du simply means $df(R)$. More generally, given any differential form $M du + N dv + \dots$, this is an operation taking a smooth parametrized curve C and a real number $a \in \text{dom } C$ to $MC^*du|_a + NC^*dv|_a + \dots$. When we write an equation between differential forms, we mean that the result is the same for every C and a .

There are two good things about this definition. First of all, all of the usual rules of differentiation are actually true of it; because the definition ultimately refers to ordinary functions, we can prove each rule in the list on page 2 by using the corresponding result for ordinary functions. The other good thing about this definition is that we will sometimes want to apply a differential form to a curve, or rather to apply a differential form to a *vector*.

Specifically, given a point P and a vector \mathbf{v} , let $C(t) = P + t\mathbf{v}$; then C is a curve with $C(0) = P$, so $C^*df(P)|_0$ makes sense. We also write $\iota_{\mathbf{v}}df(P)$ for this, which hopefully does not seem too complicated. In a way, this is all that we really need, thanks to this theorem: if C is any smooth curve (and as usual f is a smooth function and $a \in \text{dom } C$), then $C^*df(P)|_a = \iota_{C'(a)}df(P)$.

If \mathbf{v} happens to be a unit vector, then $\iota_{\mathbf{v}}df(P)$ is called the **directional derivative** of f at P in the direction of \mathbf{v} . (Sometimes this term is used even when \mathbf{v} is *not* a unit vector, because that is also important and there doesn't seem to be any other term for it.) In particular, the directional derivatives parallel to the coordinate axes, that is $\iota_1df(P)$, $\iota_2df(P)$, and so on, are called the **partial derivatives** of f at P .

Partial derivatives

The partial derivatives can be viewed from another perspective. If $f(x, y, z)$ (for example) can be expressed using the usual operations (and possibly even if it cannot), then its differential will come out as

$$df(x, y, z) = f_1(x, y, z) dx + f_2(x, y, z) dy + f_3(x, y, z) dz$$

for some functions f_1 , f_2 , and f_3 . These functions are the **partial derivatives** of f . Since subscripts can be used for many things, a better notation for these is $D_i f$ for f_i ; compare the notation Df for f' in one-variable Calculus. For example, if $f(x, y, z) = x^2y + \sin(z^2)$, then

$$df(x, y, z) = 2xy dx + x^2 dy + 2z \cos(z^2) dz,$$

so

$$D_1 f(x, y, z) = 2xy,$$

$$D_2 f(x, y, z) = x^2, \text{ and}$$

$$D_3 f(x, y, z) = 2z \cos(z^2).$$

If instead we write u for $f(x, y, z)$, then we have a different notation for the coefficients on the differentials:

$$du = \left(\frac{\partial u}{\partial x} \right)_{y,z} dx + \left(\frac{\partial u}{\partial y} \right)_{x,z} dy + \left(\frac{\partial u}{\partial z} \right)_{x,y} dz.$$

So for example, if $u = x^2y + \sin(z^2)$, then

$$du = 2xy \, dx + x^2 \, dy + 2z \cos(z^2) \, dz,$$

so

$$\begin{aligned}\left(\frac{\partial u}{\partial x}\right)_{y,z} &= 2xy, \\ \left(\frac{\partial u}{\partial y}\right)_{x,z} &= x^2, \text{ and} \\ \left(\frac{\partial u}{\partial z}\right)_{x,y} &= 2z \cos(z^2).\end{aligned}$$

This $\left(\frac{\partial u}{\partial x}\right)_{y,z}$ is the **partial derivative of u with respect to x , fixing y and z** , which simply means whatever comes before dx in an expansion of du in the variables x , y , and z . All of this information is necessary to avoid ambiguity, although in practice people usually write simply $\frac{\partial u}{\partial x}$, call this the *partial derivative of u with respect to x* , and expect you to guess from context what the other variables are.

Of course, people also mix notation for f with notation for u , writing $D_x f$, f_x , $\frac{\partial f}{\partial x}$, and so on, as well as u_x , u_1 , $D_1 u$, and so on. Technically, notation with numbers makes sense only when applied to the name of a function, because the arguments of that function come in a specific order; while notation referring to the variables used does *not* make sense when applied to the name of a function, since one could use any variables as the arguments of the function (although it does make sense when applied to an expression such as $f(x, y, z)$, in which these variables have been specified). In practice, however, people usually use the variables x, y, z in that order; then there is no confusion.

If the partial derivatives of f are used as the components of a vector, then this defines a vector field, the **gradient** of f . In symbols,

$$\nabla f = \langle D_1 f, D_2 f, D_3 f \rangle.$$

(The symbol ‘ ∇ ’ is variously pronounced ‘Atled’, ‘Nabla’, and ‘Del’. People also write $\text{grad } f$ for ∇f .) In other words,

$$\nabla f(x, y, z) = \left\langle \frac{\partial f(x, y, z)}{\partial x}, \frac{\partial f(x, y, z)}{\partial y}, \frac{\partial f(x, y, z)}{\partial z} \right\rangle.$$

The gradient has the same information as the differential; $df(x, y, z) = \nabla f(x, y, z) \cdot d(x, y, z)$; for this reason, the gradient may also be written as $df(R)/ddR$, where (as before) $R = (x, y, z)$ (or you can use $\mathbf{r} = \langle x, y, z \rangle$ instead). Conversely, $\iota_{\mathbf{v}} df(x, y, z) = \mathbf{v} \cdot \nabla f(x, y, z)$. The differential is the more fundamental concept, since it doesn't require any geometric notions, but the gradient, being a vector, is easier to visualize.