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

We often divide one differential by another; for example, $\frac{\mathrm{d}y}{\mathrm{d}x}$ 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 $\frac{\mathrm{d}y}{\mathrm{d}x}$ literally is $\mathrm{d}y$ divided by $\mathrm{d}x$. Unfortunately, $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$, the second derivative, is *not* literally $\mathrm{d}^2y = \mathrm{d}(\mathrm{d}y)$ divided by $\mathrm{d}x^2 = (\mathrm{d}x)^2$; for this reason, I prefer the notation $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 y = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\,y\right) = \frac{\mathrm{d}(\mathrm{d}y/\mathrm{d}x)}{\mathrm{d}x}$.

Differentials and the rules of differentiation

One sometimes sees the Chain Rule expressed as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x},$$

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

$$\mathrm{d}f(u) = f'(u)\,\mathrm{d}u;$$

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 $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$, this derivative can be expressed in differential form without even bothering to name the functions involved:

$$\mathrm{d}(\sin x) = \cos x\,\mathrm{d}x. \tag{1}$$

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

$$\mathrm{d}(\sin u) = \cos u\,\mathrm{d}u \tag{2}$$

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),

$$\mathrm{d}\big(\sin(x^2)\big) = \cos(x^2)\,\mathrm{d}(x^2) = \cos(x^2)(2x\,\mathrm{d}x) = 2x\cos(x^2)\,\mathrm{d}x.$$

You may now divide both sides of this equation by $\mathrm{d}x$ 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: $\mathrm{d}k = 0$ if k is constant.
- The Sum Rule: $\mathrm{d}(u+v) = \mathrm{d}u + \mathrm{d}v$.
- The Translate Rule: $\mathrm{d}(u+k) = \mathrm{d}u$ if k is constant.
- The Difference Rule: $\mathrm{d}(u-v) = \mathrm{d}u - \mathrm{d}v$.
- The Product Rule: $\mathrm{d}(uv) = v\,\mathrm{d}u + u\,\mathrm{d}v$.
- The Multiple Rule: $\mathrm{d}(ku) = k\,\mathrm{d}u$ if k is constant.
- The Quotient Rule: $\mathrm{d}\left(\frac{u}{v}\right) = \frac{v\,\mathrm{d}u - u\,\mathrm{d}v}{v^2}$.
- The Power Rule: $\mathrm{d}\left(u^k\right) = ku^{k-1}\,\mathrm{d}u$ if k is constant.
- The Root Rule: $\mathrm{d}(\sqrt[k]{u}) = \frac{\sqrt[k]{u}\,\mathrm{d}u}{ku^k}$ if k is constant.
- The Exponentiation Rule: $\mathrm{d}(\mathrm{e}^u) = \mathrm{e}^u\,\mathrm{d}u$.
- The Logarithm Rule: $\mathrm{d}(\ln u) = \frac{\mathrm{d}u}{u}$.
- The Sine Rule: $\mathrm{d}(\sin u) = \cos u\,\mathrm{d}u$.
- The Cosine Rule: $\mathrm{d}(\cos u) = -\sin u\,\mathrm{d}u$.
- The Tangent Rule: $\mathrm{d}(\tan u) = \sec^2 u\,\mathrm{d}u$.
- The Cotangent Rule: $\mathrm{d}(\cot u) = -\csc^2 u\,\mathrm{d}u$.
- The Secant Rule: $\mathrm{d}(\sec u) = \tan u \sec u\,\mathrm{d}u$.
- The Cosecant Rule: $\mathrm{d}(\csc u) = -\cot u \csc u\,\mathrm{d}u$.
- The Arcsine Rule: $\mathrm{d}(\operatorname{asin} u) = \frac{\mathrm{d}u}{\sqrt{1-u^2}}$.
- The Arccosine Rule: $\mathrm{d}(\operatorname{acos} u) = -\frac{\mathrm{d}u}{\sqrt{1-u^2}}$.
- The Arctangent Rule: $\mathrm{d}(\operatorname{atan} u) = \frac{\mathrm{d}u}{u^2+1}$.
- The Arccotangent Rule: $\mathrm{d}(\operatorname{acot} u) = -\frac{\mathrm{d}u}{u^2+1}$.
- The Arcsecant Rule: $\mathrm{d}(\operatorname{asec} u) = \frac{\mathrm{d}u}{|u|\sqrt{u^2-1}}$.
- The Arccosecant Rule: $\mathrm{d}(\operatorname{acsc} u) = -\frac{\mathrm{d}u}{|u|\sqrt{u^2-1}}$.
- The Chain Rule: $\mathrm{d}(f(u)) = f'(u)\,\mathrm{d}u$ (where f is a differentiable function of one variable).
- The First Fundamental Theorem of Calculus: $\mathrm{d}\left(\int_u^v f(t)\,\mathrm{d}t\right) = f(v)\,\mathrm{d}v - f(u)\,\mathrm{d}u$ (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.

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). You should recognize this as a kind of *differential form*; more precisely, these are *exterior differential 1-forms*.

Differentiable functions

It's time to actually give a *definition* of $\mathrm{d}u$. 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 **differentiable** at some point P if there exists a row vector, which we will write as $\nabla f(P)$, such that, for every differentiable parametrized curve C and real number $a \in \operatorname{dom} C$, if $C(a) = P$, then the composite function $f \circ C$ is differentiable at a and furthermore $(f \circ C)'(a) = \nabla f(P) \cdot C'(a)$.

(Notice that $f \circ C$ is an ordinary real-valued function of one real variable, so we know what its derivative means.) If f is differentiable at every point P in its domain, then we simply say that f is *differentiable*; then $f \circ C$ is differentiable, for every differentiable parametrized curve C. Note that ∇f is a vector field defined where f is differentiable, called the **gradient** of f. (The symbol '∇' is variously pronounced 'Atled', 'Nabla', and 'Del'; people also write $\operatorname{grad} f$ for ∇f.)

If $u = f(R)$ and f is differentiable, then we write

$$\mathrm{d}u = \nabla f(R) \cdot \mathrm{d}R = \nabla f(R) \cdot \mathrm{d}\mathbf{r},$$

where $\mathbf{r}$ is R minus the origin, as usual. If you think of ∇f as a derivative of f, then this is simply taking the Chain Rule as a definition. There are two good things about this definition of $\mathrm{d}u$. 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 when we evaluate a differential at a given point and vector, then the result is one of the derivatives $(f \circ C)'(a)$ that appear in the definition above.

Specifically, given a point P and a vector $\mathbf{v}$, let $C(t) = P + t\mathbf{v}$; then C is a differentiable curve with $C(0) = P$ and $C'(0) = \mathbf{v}$, so

$$\mathrm{d}u|_{\substack{R=P \\ \mathrm{d}R=\mathbf{v}}} = \nabla f(P) \cdot \mathbf{v} = \nabla f\big(C(0)\big) \cdot C'(0) = (f \circ C)'(0)$$

when $u = f(R)$. If $\mathbf{v}$ happens to be a unit vector (a *direction*), then $\nabla f(P) \cdot \mathbf{v}$ is called the **directional derivative** of f at P in the direction of $\mathbf{v}$. In general, the directional derivative in the direction of $\mathbf{v}$ is $\nabla f(P) \cdot \mathbf{v}/|\mathbf{v}|$, although some people use the term 'directional derivative' for $\nabla f(P) \cdot \mathbf{v}$ in the general case (since it's important but there is no standard name for it), so be careful. In particular, the directional derivatives parallel to the coordinate axes —that is $\nabla f(P) \cdot \mathbf{i}$, $\nabla f(P) \cdot \mathbf{j}$, and (in 3 dimensions) $\nabla f(P) \cdot \mathbf{k}$— 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

$$\mathrm{d}f(x, y, z) = f_1(x, y, z)\,\mathrm{d}x + f_2(x, y, z)\,\mathrm{d}y + f_3(x, y, z)\,\mathrm{d}z$$

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 f_1, f_2, and f_3 is $D_1 f$, $D_2 f$, and $D_3 f$ (respectively); compare the notation Df for f' in single-variable Calculus. For example, if $f(x, y, z) = x^2 y + \sin(z^2)$, then

$$\mathrm{d}f(x, y, z) = 2xy\,\mathrm{d}x + x^2\,\mathrm{d}y + 2z\cos(z^2)\,\mathrm{d}z,$$

so

$$\begin{aligned} D_1 f(x, y, z) &= 2xy, \\ D_2 f(x, y, z) &= x^2, \text{ and} \\ D_3 f(x, y, z) &= 2z\cos(z^2). \end{aligned}$$

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

$$\mathrm{d}u = \left(\frac{\partial u}{\partial x}\right)_{y,z} \mathrm{d}x + \left(\frac{\partial u}{\partial y}\right)_{x,z} \mathrm{d}y + \left(\frac{\partial u}{\partial z}\right)_{x,y} \mathrm{d}z.$$

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

$$\mathrm{d}u = 2xy\,\mathrm{d}x + x^2\,\mathrm{d}y + 2z\cos(z^2)\,\mathrm{d}z,$$

so

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

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 $\mathrm{d}x$ in an expansion of $\mathrm{d}u$ 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.

We can reconstrut the gradient of f using its partial derivatives:

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

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; $\mathrm{d}f(x, y, z) = \nabla f(x, y, z) \cdot \mathrm{d}(x, y, z)$; for this reason, the gradient may also be written as $\mathrm{d}f(R)/\mathrm{d}R = \mathrm{d}f(R)/\mathrm{d}\mathbf{r}$ (although we cannot define division by a vector in general), where (as before) $R = (x, y, z)$ and $\mathbf{r} = R - (0, 0, 0) = \langle x, y, z\rangle$ (since $\mathrm{d}R = \mathrm{d}\mathbf{r}$). The differential is the more fundamental concept, since it doesn't require the geometric notions of length and angle that go into defining the dot product; however, the gradient, being a vector, is easier to visualize.