Differentials and differential forms

Differential forms are, broadly speaking, expressions that may have *differentials* in them. They have many uses in modern science and engineering, even though they are not traditionally covered explicitly in math class. They are covered somewhat, however, and they are there whenever you differentiate or integrate, even if you don't recognize them. They are especially prominent in multivariable Calculus, and I want to bring them to your attention; you'll find that symbols that otherwise seem meaningless or merely mnemonic can be understood literally (sometimes with slight changes) as differential forms.

Examples

The most basic examples of differential forms are differentials such as $\mathrm{d}x$ and $\mathrm{d}y$. In general, if u is any quantity that might change, then $\mathrm{d}u$ is intended to be a related quantity whose value is an infinitely small change in u, or rather the amount by which the value of u changes when an infinitely small (or arbitrarily small) change is made. (I will make this precise later on.)

Besides the differentials themselves, differential forms can be constructed by applying arithmetic operations, so $\mathrm{d}x + \mathrm{d}y$, $\mathrm{d}x\,\mathrm{d}y$, and $\sqrt{\mathrm{d}x}$ are all differential forms. In all of these expressions, we adopt an order of operations in which the differential operator d is applied before any arithmetic operator; for example, $\mathrm{d}x^2$ means $(\mathrm{d}x)^2$, not $\mathrm{d}(x^2)$ (which is $\mathrm{d}u$ when $u = x^2$ and turns out to equal $2x\,\mathrm{d}x$). Additionally, we can include ordinary quantities in these expressions, so $x + \mathrm{d}x$, $3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z$, and $x\ln(y/\mathrm{d}z)$ are also differential forms. We can also use differentials of differentials, such as d^2x (which means $\mathrm{d}(\mathrm{d}x)$, the differential of $\mathrm{d}x$), although we won't need such *higher-order* differentials in this course. Besides all of this, any ordinary expression counts as a differential form in a degenerate way; thus, x, y^2, and $2xy^3$ are also differential forms (of order zero).

Some differential forms are more useful than others. Of those listed above, besides the differentials and the non-differential quantities, the ones most likely to appear in a real problem are $\mathrm{d}x + \mathrm{d}y$ and $3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z$. These consist of any number of terms, each of which is the product of an ordinary quantity (possibly the constant 1) and the differential of an ordinary quantity. Differential forms with this property are most commonly found in practice. We will use other differential forms, such as $3x\,|\mathrm{d}y|$ and $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$; however, you might be able to see how even these forms are differential of *degree* 1 in a sense similar to the degree of a polynomial.

All of the examples so far are differential forms of *rank* 1; there are also differential forms of higher rank, such as $\mathrm{d}x \wedge \mathrm{d}y$, which are written using a new operation, the *wedge product*. We will not use these until later; these notes are only about differential forms of rank 1, or 1-forms for short. (Ordinary quantities have rank 0.)

Evaluating differential forms

In this class, we generally assume that any ordinary quantity (that is any 0-form) is a function of 2 or 3 ordinary variables, $P = (x, y)$ or $P = (x, y, z)$. Thus, we evaluate ordinary quantites (0-forms) by specifying specific values for the variables that comprise P. For example, to evaluate $u = x^2 + xy$ when $x = 2$ and $y = 3$, we may write

$$u|_{P=(2,3)} = (x^2 + xy)|_{(x,y)=(2,3)} = (2)^2 + (2)(3) = 10.$$

To evaluate a differential form (of order 1), we need not only a point (a value of P) but also a vector (a value of $\mathrm{d}P = \langle \mathrm{d}x, \mathrm{d}y\rangle$ or $\mathrm{d}P = \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z\rangle$). So for example, to evaluate $\alpha = 3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z$ when $x = 2$, $y = 3$, $z = 4$, $\mathrm{d}x = 0.05$, $\mathrm{d}y = -0.01$, and $\mathrm{d}z = 0$, we may write

$$\begin{aligned} \alpha|_{\substack{P=(2,3,4),\\ \mathrm{d}P=\langle 0.05,-0.01,0\rangle}} &= (3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z)|_{\substack{(x,y,z)=(2,3,4),\\ \langle \mathrm{d}x,\mathrm{d}y,\mathrm{d}z\rangle=\langle 0.05,-0.01,0\rangle}} \\ &= 3(0.05) + (2)^2(-0.01) + \mathrm{e}^{(3)}(0) = 0.11. \end{aligned}$$

(Differential forms are often denoted with Greek letters such as 'α', although they don't have to be.) We say that α has been evaluated *at* the point $P = (2, 3, 4)$ *along* the vector $\mathrm{d}P = \langle 0.05, -0.01, 0\rangle$. (The components of $\mathrm{d}P$ don't need to be small, since the definition makes sense in any case, but in applications that's usually what matters; after all, $\mathrm{d}P$ is supposed to be a *small* change in position.)

(To evaluate higher-order differential forms (those that involve higher-order differentials), we need to specify additional vectors such as $\mathrm{d}^2P = \langle \mathrm{d}^2x, \mathrm{d}^2y, \mathrm{d}^2z\rangle$, etc. However, we won't need that level of generality in this course.)

Differential forms as vectors

A differential form $\alpha = M\,\mathrm{d}x + N\,\mathrm{d}y + O\,\mathrm{d}z$ may be viewed as a dot product $\alpha = \langle M, N, O\rangle \cdot \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z\rangle = \mathbf{V} \cdot \mathrm{d}P$. For example, if $\alpha = 3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z$, then $\alpha = \langle 3, x^2, \mathrm{e}^y\rangle \cdot \mathrm{d}P$; conversely, if $\mathbf{V} = \langle 3, x^2, \mathrm{e}^y\rangle$, then

$$\mathbf{V} \cdot \mathrm{d}P = \langle 3, x^2, \mathrm{e}^y\rangle \cdot \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z\rangle = 3\,\mathrm{d}x + x^2\,\mathrm{d}y + \mathrm{e}^y\,\mathrm{d}z.$$

(We can recover $\mathbf{V}$ from α formally by evaluating α when $\mathrm{d}P$ is $\langle \mathbf{i}, \mathbf{j}\rangle$ or $\langle \mathbf{i}, \mathbf{j}, \mathbf{k}\rangle$, but there's probably no need to think about that explicitly.)

Even in circumstances where it makes no sense to interpret a change in the values of (x, y, z) as a vector in the geometric sense (with length and direction), in which case dot products involving them generally have no meaning, it is traditional to write differential forms in this way and to focus on $\mathbf{V}$ rather than on α as the object of study. In this case, we need to think of $\mathbf{V}$ as a *row* vector. Regardless of whether $\mathbf{V}$ has geometric significance as a vector, it can be helpful to visualize it as one.

When calculations with a row vector need to be performed, ultimately it is the differential form $\alpha = \mathbf{V} \cdot \mathrm{d}P$ that matters. It's more common to see $\mathbf{V} \cdot \mathrm{d}\mathbf{r}$, where as usual the vector $\mathbf{r} = P - \mathrm{O}$ (where O is $(0, 0)$ or $(0, 0, 0)$) satisfies $\mathrm{d}\mathbf{r} = \mathrm{d}P$. Sometimes $\mathbf{V} \cdot \mathrm{d}\mathbf{r}$ is even regarded as merely a mnemonic notation (especially in the context of defining integrals such as those in Section 15.2 of the textbook), but it can be taken literally, just as $\mathrm{d}y/\mathrm{d}x$ (which is also sometimes regarded as merely mnemonic) can be taken literally as the result of a division of differentials. In any case, people do write $\mathbf{V} \cdot \mathrm{d}\mathbf{r}$ (even in the textbook), so it can be nice to know what it means!

In the textbook, they also sometimes write $\mathrm{d}\mathbf{r} = \mathbf{T}\,\mathrm{d}s$, where $\mathrm{d}s$ (which is not really the differential of anything in space as a whole) is the magnitude $\mathrm{d}s = |\mathrm{d}\mathbf{r}|$ and $\mathbf{T} = \widehat{\mathrm{d}\mathbf{r}}$, the unit vector in the direction of $\mathrm{d}\mathbf{r}$. This is sometimes useful when thinking about things geometrically, but it's not necessary for purposes of calculation. In 2 dimensions, we can also take cross products (using the rule $\langle a, b\rangle \times \langle c, d\rangle = ad - bc$); for example, if $\mathbf{V} = \langle 3, x^2\rangle$, then

$$\mathbf{V} \times \mathrm{d}\mathbf{r} = \langle 3, x^2\rangle \times \langle \mathrm{d}x, \mathrm{d}y\rangle = 3\,\mathrm{d}y - x^2\,\mathrm{d}x.$$

(This requires that changes in x and y make sense as having a geometric length even when $\mathbf{V}$ is regarded as merely a row vector, so it doesn't come up as often.) If you use $\times\langle c, d\rangle = \langle d, -c\rangle$, so that $\mathbf{u} \times \mathbf{v} = \mathbf{u} \cdot \times\mathbf{v}$, then you can write $\mathbf{V} \times \mathrm{d}\mathbf{r}$ as $\mathbf{V} \cdot \times\mathrm{d}\mathbf{r}$; the book sometimes writes this as $\mathbf{V} \cdot \mathbf{n}\,\mathrm{d}s$, where $\mathrm{d}s = |\times\mathrm{d}\mathbf{r}| = |\mathrm{d}\mathbf{r}|$ again, and now $\mathbf{n} = \widehat{\times\mathrm{d}\mathbf{r}} = \times\mathbf{T}$ is the direction perpendicular and clockwise from $\mathrm{d}\mathbf{r}$. Again, sometimes this is useful when thinking about the geometry, but you don't need it for doing calculations.

This is all especially common when $\mathbf{V}$ is the output of a *vector field*, that is a vector-valued function of several variables. For example, if $\mathbf{F}(x, y) = \langle 3, x^2\rangle$, then

$$\mathbf{F}(x, y) \cdot \mathrm{d}\mathbf{r} = \langle 3, x^2\rangle \cdot \langle \mathrm{d}x, \mathrm{d}y\rangle = 3\,\mathrm{d}x + x^2\,\mathrm{d}y,$$

and

$$\mathbf{F}(x, y) \times \mathrm{d}\mathbf{r} = \langle 3, x^2\rangle \times \langle \mathrm{d}x, \mathrm{d}y\rangle = 3\,\mathrm{d}y - x^2\,\mathrm{d}x.$$

So in Section 15.2, which is really about integrating differential 1-forms along curves, the book spends most of its time talking about integrating vector fields along curves (and occasionally integrating them across curves in 2 dimensions). What's really going on is that you integrate the vector field $\mathbf{F}$ by integrating one of the two differential forms above (usually the first one).

Differentials and the rules of differentiation

In one-variable Calculus, 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

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

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

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

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

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

for any other differentiable quantity u; the Chain Rule is the power to derive this equation from the previous one. 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: $\mathrm{d}(u + C) = \mathrm{d}u$ if C 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: $\mathrm{d}\left(\dfrac{u}{v}\right) = \dfrac{v\,\mathrm{d}u - u\,\mathrm{d}v}{v^2}$.
- The Power Rule: $\mathrm{d}(u^n) = nu^{n-1}\,\mathrm{d}u$ if n is constant.
- The Exponentiation Rule: $\mathrm{d}(\exp u) = \exp u\,\mathrm{d}u$ (where $\exp u$ means e^u).
- The Logarithm Rule: $\mathrm{d}(\ln u) = \dfrac{\mathrm{d}u}{u}$.
- 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: $\mathrm{d}(\operatorname{asin} u) = \dfrac{\mathrm{d}u}{\sqrt{1 - u^2}}$ (where $\operatorname{asin} u$ means $\sin^{-1} u$).
- The Arccosine Rule: $\mathrm{d}(\operatorname{acos} u) = -\dfrac{\mathrm{d}u}{\sqrt{1 - u^2}}$.

- The Arctangent Rule: $\mathrm{d}(\operatorname{atan} u) = \dfrac{\mathrm{d}u}{u^2+1}$.
- The Arccotangent Rule: $\mathrm{d}(\operatorname{acot} u) = -\dfrac{\mathrm{d}u}{u^2+1}$.
- The Arcsecant Rule: $\mathrm{d}(\operatorname{asec} u) = \dfrac{\mathrm{d}u}{|u|\sqrt{u^2-1}}$.
- The Arccosecant Rule: $\mathrm{d}(\operatorname{acsc} u) = -\dfrac{\mathrm{d}u}{|u|\sqrt{u^2-1}}$.
- The Chain Rule: $\mathrm{d}(f(u)) = f'(u)\,\mathrm{d}u$ if f is a function of one variable that's differentiable at u.
- The First Fundamental Theorem of Calculus: $\mathrm{d}\Big(\int_{t=u}^{v} f(t)\,\mathrm{d}t\Big) = f(v)\,\mathrm{d}v - f(u)\,\mathrm{d}u$ if f is a function of one variable that's continuous between u and v.

(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). This is a kind of differential form; more precisely, these are *linear differential* 1*-forms* (which are also called *exterior differential* 1*-forms*).

Here is an example of how to use the rules, step by step, to find a differential. Specifically, I'll find the differential of $x^2y + \sin(z^2)$. (In one-variable Calculus, you might consider this if x, y, and z all happen to be functions of some other variable t; but in multivariable Calculus, the same calculation will apply even when the variables x, y, and z are all independent.)

$$\begin{aligned}
\mathrm{d}\Big(x^2y + \sin(z^2)\Big) &= \mathrm{d}(x^2y) + \mathrm{d}\Big(\sin(z^2)\Big) \\
&= y\,\mathrm{d}(x^2) + x^2\,\mathrm{d}y + \cos(z^2)\,\mathrm{d}(z^2) \\
&= y(2x^{2-1}\,\mathrm{d}x) + x^2\,\mathrm{d}y + \cos(z^2)(2z^{2-1}\,\mathrm{d}z) \\
&= 2xy\,\mathrm{d}x + x^2\,\mathrm{d}y + 2z\cos(z^2)\,\mathrm{d}z.
\end{aligned}$$

Here I've used, in turn, the sum rule, the product and sine rules (one in one term and the other in the other term), the power rule (in two places), and finally some algebra to simplify. Of course, you can usually do this much faster; with practice, you can jump immediately to the second-to-last line by applying the next rule whenever one rule results in a differential; then you only need one more step to simplify it algebraically. Often you can even do some of the algebra in your head immediately (like simplifying x^{2-1} to x, so that $\mathrm{d}(x^2)$ immediately becomes $2x\,\mathrm{d}x$).

Partial derivatives

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

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

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_1f, D_2f, and D_3f (respectively); compare the notation Df for f' that is sometimes used in single-variable Calculus. For example, if $f(x,y,z) = x^2y + \sin(z^2)$, then

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

(as I calcluated earlier), so

$$\begin{aligned}
D_1f(x,y,z) &= 2xy, \\
D_2f(x,y,z) &= x^2, \text{ and} \\
D_3f(x,y,z) &= 2z\cos(z^2).
\end{aligned}$$

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

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

(The symbol '∂' is a variation on the lowercase Greek Delta, 'δ'. It is usually not pronounced directly; instead, you read the entire expression as described below.) So for example, if $u = x^2y + \sin(z^2)$, then

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

again, 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 tells you how much u changes relative to the change in x as long as y and z remain the same. All of the information in this notation is necessary to avoid ambiguity, but in practice people usually write simply $\frac{\partial u}{\partial x}$, call this simply the partial derivative of u with respect to x, and expect you to guess from context what other variables are remaining fixed.

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.

Defining differentials

Recall from the handout on definitions for functions of several variables that the function f is **differentiable** at the point P_0 if there exists a row vector $\nabla f(P_0)$ such that, for every differentiable parametrized curve C and real number t_0, if $C(t_0)$ exists and equals P_0, then the composite function $f \circ C$ is differentiable at t_0 and furthermore $(f \circ C)'(t_0) = \nabla f(P_0) \cdot C'(t_0)$. This makes ∇f a vector field, called the **gradient** of f, that is defined wherever f is differentiable. (The symbol '∇' is variously pronounced 'Atled', 'Nabla', and 'Del'; people also write $\operatorname{grad} f$ for ∇f.)

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

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

where $\mathbf{r}$ is $P - \mathrm{O}$ (P 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 pages 3 and 4 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 $(f \circ C)'(t_0)$ that appear in the definition above.

Specifically, fixing a point P_0 and a vector $\mathbf{v}_0$, let $C(t) = P_0 + t\mathbf{v}_0$; then C is a differentiable curve with $C(0) = P_0$ and $C'(0) = \mathbf{v}_0$, so

$$\mathrm{d}u|_{\substack{P=P_0,\\ \mathrm{d}P=\mathbf{v}_0}} = \nabla f(P_0) \cdot \mathbf{v}_0 = \nabla f\big(C(0)\big) \cdot C'(0) = (f \circ C)'(0)$$

when $u = f(P)$. If $\mathbf{v}_0$ happens to be a unit vector (a *direction*), then $\nabla f(P_0) \cdot \mathbf{v}_0$ is called the **directional derivative** of f at P_0 in the direction of $\mathbf{v}_0$. In general, the directional derivative in the direction of $\mathbf{v}_0$ is $\nabla f(P_0) \cdot \hat{\mathbf{v}}_0$ (where $\hat{\mathbf{v}} = \mathbf{v}/|\mathbf{v}|$ is the unit vector in the direction of $\mathbf{v}$); however, be careful, because some people use the term 'directional derivative' for $\nabla f(P_0) \cdot \mathbf{v}_0$ in the general case (since it's important but there is no standard name for it). In particular, the directional derivatives parallel to the coordinate axes —that is $\nabla f(P_0) \cdot \mathbf{i}$, $\nabla f(P_0) \cdot \mathbf{j}$, and (in 3 dimensions) $\nabla f(P_0) \cdot \mathbf{k}$— are simply the partial derivatives of f at P_0.

Because $\mathrm{d}\big(f(P)\big) = \nabla f(P) \cdot \mathrm{d}P = \nabla f(P) \cdot \mathrm{d}\mathbf{r}$, the value of the gradient may also be written as $\nabla f(P) = \mathrm{d}\big(f(P)\big)\big/\mathrm{d}P = \mathrm{d}\big(f(P)\big)\big/\mathrm{d}\mathbf{r}$ (although we cannot define division by a vector in general). An even simpler notation for $\nabla f(P)$ would be $f'(P)$, but this is traditionally not used, because there are many notions of derivative of f (such as the directional derivatives and the partial derivatives); even though the gradient is the most general derivative, it is commonly thought that f' would be ambiguous in this context. (When we start differentiating vector fields near the end of this course, there will be another reason that it's convenient to have a symbol ∇ that we can manipulate more easily than the tiny tick mark on f'.)

Gradients

If f is a function of (say) 3 variables, then the definition of differential above states that

$$\mathrm{d}\big(f(x,y,z)\big) = \nabla f(x,y,z) \cdot \mathrm{d}(x,y,z) = \nabla f(x,y,z) \cdot \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z \rangle;$$

meanwhile, the definition of partial derivatives states that

$$\begin{aligned} \mathrm{d}\big(f(x,y,z)\big) &= D_1 f(x,y,z)\,\mathrm{d}x + D_2 f(x,y,z)\,\mathrm{d}y + D_3 f(x,y,z)\,\mathrm{d}z \\ &= \langle D_1 f(x,y,z), D_2 f(x,y,z), D_3 f(x,y,z) \rangle \cdot \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z \rangle. \end{aligned}$$

In other words,

$$\nabla f(x,y,z) = \langle D_1 f(x,y,z), D_2 f(x,y,z), D_3 f(x,y,z) \rangle = \left\langle \frac{\partial\big(f(x,y,z)\big)}{\partial x}, \frac{\partial\big(f(x,y,z)\big)}{\partial y}, \frac{\partial\big(f(x,y,z)\big)}{\partial z} \right\rangle.$$

Put more simply,

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

or even

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

The gradient has the same information as the differential, and the partial derivatives are the componenets of the gradient, so any one of these (the gradient, the partial derivatives, or the differential) may be used to solve any problem. The differential is usually the most useful for direct calculation, which is one reason why I use it heavily. However, if we have a geometric notion of length available to allow us to think of row vectors (such as the gradient) as the same as column vectors (the usual ones, going between points), then the gradient is easier to visualize.

For reference, here are a bunch of relationships between differentials, partial derivatives, and gradients, assuming that $u = f(x, y, z)$:

$$\mathrm{d}u = \left(\frac{\partial u}{\partial x}\right)_{y,z} \mathrm{d}x + \left(\frac{\partial u}{\partial y}\right)_{x,z} \mathrm{d}y + \left(\frac{\partial u}{\partial z}\right)_{x,y} \mathrm{d}z;$$
$$\mathrm{d}u = D_1 f(x,y,z)\,\mathrm{d}x + D_2 f(x,y,z)\,\mathrm{d}y + D_3 f(x,y,z)\,\mathrm{d}z;$$
$$D_1 f(x,y,z) = \left(\frac{\partial u}{\partial x}\right)_{y,z},\ D_2 f(x,y,z) = \left(\frac{\partial u}{\partial y}\right)_{x,z},\ D_3 f(x,y,z) = \left(\frac{\partial u}{\partial z}\right)_{x,y};$$
$$\nabla f(x,y,z) = \langle D_1 f(x,y,z), D_2 f(x,y,z), D_3 f(x,y,z)\rangle;$$
$$\mathrm{d}u = \nabla f(x,y,z) \cdot \langle \mathrm{d}x, \mathrm{d}y, \mathrm{d}z\rangle$$
$$\mathrm{d}u|_{\langle \mathrm{d}x,\mathrm{d}y,\mathrm{d}z\rangle = \mathbf{v}} = \nabla f(x,y,z) \cdot \mathbf{v}$$

Tangents and normal lines

If f is a function of 2 (or 3) variables and P_0 is a point in 2 (or 3) dimensions, then the level curve (or surface) of f through P_0 is given by the equation $f(P) = f(P_0)$, where $P = (x, y)$ (or (x, y, z), as usual). (The function f and the point P_0 have already been fixed, but the point P is allowed to vary, so this is an equation in our 2 (or 3) variables, as it should be.) If f is differentiable at P_0 and the gradient of f is nonzero at P_0, then this level curve (or surface) has a **tangent** line (or plane) through P_0, given by the equation $\nabla f(P_0) \cdot (P - P_0) = 0$. Finally, perpendicular to this tangent line (or plane), there is a **normal** line (always a line!) through P_0, with parametrization $P = P_0 + t\,\nabla f(P_0)$ in the parameter t.

Writing u for $f(P)$, the equation for the level curve (or surface) is $u = u|_{P=P_0}$. Writing Δu for $f(P + \Delta P) - f(P)$, a quantity that depends on both a point P and a vector ΔP, another equation for the level curve (or surface) is $\Delta u|_{P=P_0,\ \Delta P = P - P_0} = 0$. That is, you take the expression for Δu, which says how much u changes between two points, put P_0 in for the starting point P, and then put $P - P_0$ in for the difference ΔP between the two points. Since the value of u shouldn't change on the level curve (or surface), this difference Δu should be zero. (Notice that the meaning of P changes over the course of this substitution; originally it refers to the starting point, which we set to P_0, but afterwards it refers to another point on the level curve (or surface), so we set the displacement ΔP between the two points to $P - P_0$.)

The tangent line (or plane) is given by a very similar equation, except that now we look at how the curve (or surface) is changing infinitesimally at P_0 and extend this out to arbitrary distances. Thus, the equation $\Delta u = 0$ for the level curve (or surface) becomes $\mathrm{d}u = 0$ for the tangent line (or plane). However, we're still looking for the values of u in the same place, so the full equation is $\mathrm{d}u|_{P=P_0,\ \mathrm{d}P = P - P_0} = 0$. If you follow the definition of differential from page 5 and 6 above, then you'll see that this means precisely $\nabla f(P_0) \cdot (P - P_0) = 0$.

For example, if $u = xy$ and $P_0 = (2, 3)$, then the level curve is $xy = (2)(3)$, or simply $xy = 6$. (Replace x with 2 and y with 3 on the right-hand side.) Alternatively, $\Delta u = (x + \Delta x)(y + \Delta y) - xy = y\,\Delta x + x\,\Delta y + \Delta x\,\Delta y$, so the level curve is $(3)(x - 2) + (2)(y - 3) + (x - 2)(y - 3) = 0$. (Replace x with 2, y with 3, Δx with $x - 2$, and Δy with $y - 3$.) This also simplifies to $xy = 6$.

That was obviously more work than necessary for the level curve, but now apply the same technique to the differential to get the tangent line: $\mathrm{d}u = y\,\mathrm{d}x + x\,\mathrm{d}y$, so the tangent line is $(3)(x - 2) + (2)(y - 3) = 0$. (Replace x with 2, y with 3, $\mathrm{d}x$ with $x - 2$, and $\mathrm{d}y$ with $y - 3$.) This simplifies to $3x + 2y = 12$, and now we learnt something that we didn't know before.

Because the normal line depends on the geometric notion of angle (to tell you what's perpendicular to what), this can't be done as slickly using only differentials. Now we really do want to think of the gradient vector. All the same, since this can be read off of the differential so easily, you can still start with $\mathrm{d}u = y\,\mathrm{d}x + x\,\mathrm{d}y$. First, replace only x with 2 and y with 3 to get $3\,\mathrm{d}x + 2\,\mathrm{d}y$, then read off the gradient vector $\langle 3, 2\rangle$. Since we started at the point $(2, 3)$, the parametric equation is $P = (2, 3) + t\langle 3, 2\rangle$, or $(x, y) = (3t + 2, 2t + 3)$ in more detail.

None of this (beyond the level curve (or surface) itself) works right if the gradient $\nabla f(P_0)$ is zero or undefined. If the gradient is undefined, then of course we can't say anything using it; but if the gradient is zero, then these equations say that every point belongs to the tangent line (or plane) and only the point

P_0 belongs to the normal line. Of course, that would mean that they're not lines (or a plane and a line) at all! When the gradient is zero, the truth may be that there is no tangent or that there is a tangent but it really does consist of everything, or there may be an honest tangent line (or plane) after all; but in any case, these formulas won't help you know that!