Supplement on Differentials

Prepared by Jonathan Goodman for Calculus I, fall 2004, Courant Institute of Mathematical Sciences, NYU.

When calculus was young, the expression

$$f'(x) = \frac{df}{dx} \tag{1}$$

really meant that you find the derivative by dividing the differential of f by the differential of x. Even now, most users of calculus think of derivatives this way. A modern mathematician writes Δx for a change in x, and $\Delta f = f(x + \Delta x) - f(x)$. If Δx is very small, then f' and $\Delta f / \Delta x$ are very nearly equal. The inventors of calculus wrote dx for Δx and thought that if dx could be made infinitely small without being zero exactly, then (1) would be true exactly. A quantity that is infinitely small but not exactly zero was called an *infinitesimal*. The differentials dx and $df = f(x + dx) - f(x)$ are examples of infinitesimals.

The idea of an infinitesimals never was without controversy. Though Isaac Newton invented calculus, he distrusted it and spent years developing alternate and more complicated explanations of facts that he derived easily using calculus. His contemporary, philosopher George Berkeley[1], lampooned infinitesimals as "ghosts of departed quantities". Most modern mathematicians simply say there is a choice: Either $dx = 0$ and the quotient makes no sense, or $dx \neq 0$ and $df/dx \neq f'$. In either case, the formula (1) cannot be true literally and exactly.

Your calculus books have been teaching you the "modern" (invented early in the ninteenth century) solution to this dilemma. They define the derivative as the *limit* as Δx goes to zero of $\Delta f / \Delta x$. Though the value f' is never achieved exactly, it is approached with arbitrary precision for small enough Δx. In this way of understanding calculus, the infinitesimals dx and df need not really exist. The expression df/dx is not a quotient, but a way to denote the derivative. In this way, most of the basic objects in calculus got precise definitions and the theorems got rigorous proofs. So, why has the majority of the scientific world ignored the mathematicians' approach to calculus for two centuries?

A common joke illustrates the physicists' view of mathematical rigor. Two physicists soaring in a hot air balloon are blown into a cloud where they cannot see anything. When they emerge, they do not recognize the ground below. They call out to a man below: "Where are we?".

The man yells back: "You're in a balloon."

One of the physicists then says to the other: "That was a mathematician. What he said was absolutely true and absolutely useless."

The point is that the extra security of knowing that the results are absolutely correct is not worth the extra trouble of defining limits and giving rigorous proofs of things that seem almost obvious. The ninteenth century English applied

[1] Pronounced "BARK-lee", unlike the town in California named for him.

mathematician Heaviside was criticized for using his "operational calculus" that he did not understand well enough to define rigorously. He answered: "Should I stop eating because I do not understand digestion?"

The chain rule illustrates the additional complexity of a rigorous mathematical derivation. Suppose we have a quantity, f, that depends on another quantity, u, which in turn depends on x. We want to know what will happen to f is we make an infinitesimal change in x: we want the derivative $\frac{df}{dx}$. If we have a formula for $f(u)$ and for $u(x)$, we should be able to calculate $f'(u) = \frac{df}{du}$ and $u'(x) = \frac{du}{dx}$. Almost any scientist would tell us to multiply the numerator and denominator of the fraction $\frac{df}{dx}$ by du, to get

$$\frac{df}{dx} = \frac{df \cdot du}{dx \cdot du} = \frac{df}{du} \cdot \frac{du}{dx} = f'(u) \cdot u'(x) \ ,$$

which is the formula we call the *chain rule*. **Warning**: many math professors will take off points if a student writes something like this on a test.

In old school calculus, the d means "differential of" or "infinitesimal differential of" and can be applied to almost anything. For example, we might write

$$d(u^2) = 2udu \ . \tag{2}$$

This means that when du is infinitely small (but not necessarily zero), the change in u^2 is equal to $2u$ times the change in u. Modern calculus books, including the ones we are using, redefine the symbols d and du so that the formula (2) is true. If du is not a small or infinitesimal change in u, then what does (2) actually mean? The mathematical precision has made the formula more true and less useful.

Old school calculus makes other things things simpler to say and understand. For example, suppose we know $d(u^2)$ and want to know how much u changed. No problem. Just rearrange (2) to get

$$du = \frac{d(u^2)}{2u} \ .$$

The Salas, Hille, and Etgen book would say the same thing as follows: Let $t = f(u) = u^2$. Let $u = g(t)$ be the inverse function. Then $g'(t) = 1/f'(u) = 1/2u$. A routine algebraic manipulation vs. a piece of formalism and a theorem. Which is easier?

A typical scientist or engineer does not try to distinguish the infinitesimal from the very small. Instead, he or she adopts a more liberal definition of equality, the = sign[2]. Writing $P = Q$ does not have to mean that P is exactly equal to Q in some Platonic ideal sense. Rather, it can mean that it is useful for the purpose at hand to take P as a substitute for Q. For example, if $A = L^2$ is the area of a square with side L, and we measure the side of an actual square with a ruler, we might write $L = 9.34\ in$ and $A = 87.2\ in^2$. If pressed, we might

[2]The liberal former President Clinton got in trouble for attempting to redefine the word "is".

admit that a more accurate measurement could well give $L = 9.339$ *in*, and that our A is not even the exact mathematical square of our L. Still, 87.2 might be a more useful expression than, say, 87.236 because it is shorter, simpler, and perfectly adequate for the purpose.

Our typical scientist might view (2) not as saying that $d(u^2)$ is absolutely exactly equal to $2udu$ if, say, $u = 3$ and $du = .1$, but rather think that $2udu$ is an accurate enough representation of $d(u^2)$ for the purpose at hand (here, $2udu$ is 1.8 while $d(u^2)$ is 1.81, about half a percent more). This person might explain the chain rule (above) as follows: We want to express df, the change in f, in terms of dx. We know that $df = f'(u)du$, so we can get df if we know du. We also cam estimate du in terms of dx using $du = u'(x)dx$. Substituting the formula for du into the formula for df gives $df = f'(u) \cdot u'(x)dx$. Though none of these formulas is mathematically exact if $dx \neq 0$, they are accurate enough to give a useful formula for df in terms of dx.

The basic rule of differentials is that the square of a differential is so much smaller than the differential itself that it can be ignored *if the differential itself is present*. For example, we might write $dt + 3dt^2 = dt$. This is fine in the liberal sense of equality above, but be careful saying it to a mathematician. Also, be careful not to set $dt^2 = 0$ if there is no dt present. For example, the acceleration is $a = (d^2x)/(dt^2)$ – the numerator being the second differential of $x(t)$ (defined later in the our course) and the denominator being the square of dt. What scientists lose with the liberal sense of equality is the permanence of equality: If A ever is equal to B, then A always is equal to B. The liberal seems to violate this by taking $dt^2 = 0$ in working with $dt+3dt^2$ but not when working with d^2x/dt^2. Well, maybe we can say $dt + 3dt^2 = dt$ without saying $dt^2 = 0$. This shows once again that trying to redefine equality, the word *is*, leads to trouble. In doing so, each user of calculus takes the personal responsibility to know when $dt^2 = 0$ and when it does not. This is not such a heavy burden. Over the centuries, millions of people have got the right answer with the derivation $d(x^2) = (x + dx)^2 - x^2 = x^2 + 2xdx + dx^2 - x^2 = 2xdx + dx^2 = 2xdx$. The Leibnitz rule (product rule) is

$$d(uv) = (u + du)(v + dv) - uv = uv + du \cdot v + u \cdot dv + du \cdot dv - uv = udv + vdu \; .$$

Coming back to applications, here is a derivation of the fact that the tangent line to a circle at a point P is perpendicular to the line segment going from the center to P. The circle of radius r about and center (a, b) is the set of points satisfying the formula

$$(x - a)^2 + (y - b)^2 = r^2 \; .$$

Let $P = (x, y)$ be a point on the circle and $Q = (x+dx, y+dy)$ be a nearby point also on the circle. When we move from P to Q, the radius does not change. This means that $d(r^2) = 0$. Using this, we calculate:

$$\begin{aligned} 0 &= d(r^2) \\ &= d\left((x-a)^2\right) + d\left((y-b)^2\right) \\ &= d\left(x^2\right) - 2adx + d\left(a^2\right) + d\left(y^2\right) - 2bdy + d\left(b^2\right) \end{aligned}$$

$$
\begin{aligned}
&= 2xdx - 2adx + 2ydy - 2bdy \\
0 &= 2\left((x-a)dx + (y-b)dy\right) \ . \qquad (3)
\end{aligned}
$$

The slope of the segment from (a, b) to (x, y) is $m_1 = \Delta y/\Delta x = (y-b)/(x-a)$. The slope of the segment connecting P to Q is $m_2 = dy/dx$. The segments are perpendicular if $m_1 \cdot m_2 = -1$. We can derive this from (3): Cancel the 2 and get $(x-a)dx = -(y-b)dy$, then divide both sides by $(x-a)dx$.

A quantitative science book is more likely than not to use the language of differentials in the old school way. For example, the thermodynamics of a gas is given by the formula $de = pdV + TdS$, where: e is specific energy (energy per unit mass), p is the pressure, V is the volume, T is the temperature (above absolute zero), and S is the entropy. This means that you compute the change in energy of a piece of gas by multiplying the change in volume by the pressure, the change in entropy by the temperature, and adding. The book hopefully would remind the student that this is only accurate when dV and dS are sufficiently small. What happens when dV and dS are not small can be determined through a mathematical process called *integration*.

Another example is a simple model of growing bacteria: $dN = rNdt$, where N is the number of bacteria and t is time. This says that in a small time increment, the number of new bacteria, dN, is proportional to the current number and the time increment. Note here that dN should be small, but $dN < 1$ does not make physical sense since the number of bacteria is an integer.