How to Use Calculus Like a Physicist

Physics A300*

Fall 2004

The purpose of these notes is to make contact between the abstract descriptions you may have seen in your calculus classes and the applications of mathematical methods to problems in Physics.

1 Solving Ordinary First-Order Differential Equations By Integration

Often in Physics, one has an expression for the first time derivative of a quantity, along with its value at some initial time, and needs to find its value at all times. So for example, you may know

$$\dot{x}(t) = at \tag{1.1a}$$
$$x(0) = x_0 \tag{1.1b}$$

One standard way to solve this problem is to take the indefinite integral of $\dot{x}(t)$ to find the value of $x(t)$ up to an additive constant, then use the initial condition to set that constant. So for example in the problem stated in (1.1), we would find

$$x(t) = \int \dot{x}(t)\,dt = \frac{1}{2}at^2 + C \tag{1.2}$$

and then set the constant C via

$$x_0 = x(0) = C \tag{1.3}$$

which tells us

$$x(t) = x_0 + \frac{1}{2}at^2 \tag{1.4}$$

On the other hand, we could also solve the problem in one step with a definite integral, as follows. The Fundamental Theorems of Calculus tell us that for any t_1 and t_2,

$$\int_{t_1}^{t_2} \dot{x}(t)\,dt = x(t_2) - x(t_1) \tag{1.5}$$

or

$$x(t_2) = x(t_1) + \int_{t_1}^{t_2} \dot{x}(t)\, dt \tag{1.6}$$

If we set t_1 to 0 (the time at which the initial condition is known), everything on the right-hand side is given in the statement of the problem, e.g., (1.1), so we can integrate to find $x(t_2)$ for any t_2. We'd like to rename t_2 to t so that we can state the answer as $x(t)$, but we should be a little careful, because the integration variable is already called t, and it's a no-no to have the same variable appear in the same expression as the integration variable of a definite integral and also outside the integral (or in its limits). So we rename the integration variable from t to t', which lets us rename t_2 to t, and obtain the general expression

$$x(t) = x(0) + \int_0^t \dot{x}(t')\, dt' \tag{1.7}$$

in the example problem, this becomes

$$x(t) = x_0 + \int_0^t at'\, dt' = x_0 + \frac{1}{2}at'^2\Big|_0^t = x_0 + \frac{1}{2}at^2 \tag{1.8}$$

as before.

Both methods are valid, but the indefinite integral method saves you from having to define a lot of temporary integration constants.

2 Derivatives

2.1 Leibniz Notation

Many math books define the derivative of a function $f(x)$ with the notation

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \tag{2.1}$$

This is perfectly unambiguous, but conceals somewhat the physical nature of a derivative as a ratio of small changes. Thus it's sometimes useful to work with alternative notation, invented by Leibniz, which writes

$$\frac{df}{dx} = f'(x) = \lim_{\Delta x\to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x} \tag{2.2}$$

where $\Delta f = f(x+\Delta x) - f(x)$ is the change in $f(x)$ associated with a change in x.

2.2 Working With Differentials

The notation $\frac{df}{dx}$ represents the derivative, *not* the ratio of two numbers df and dx. However, it is often useful to manipulate it as though it were. In particular, we refer to dx as an *infinitesimal* change in x, and $df = f(x+dx) - f(x)$ is the corresponding infinitesimal change in $f(x)$. This is really just a notational convenience, but it's something that makes it a lot easier to do implicit differentiation, and, in multi-variable calculus, to obtain relationships involving partial derivatives.

2.2.1 Example: Differentiating $\ln x$ by Implicit Differentiation

To give an example of how using differentials simplifies things, consider the function

$$y = \ln x \tag{2.3}$$

which is implicitly defined by

$$x = e^y \tag{2.4}$$

If we know that

$$\frac{d}{dy} e^y = e^y \tag{2.5}$$

we can find $\frac{d}{dx} \ln x$ by implicit differentiation. First, we take the differential of both sides of (2.4), and obtain

$$dx = d(e^y) = e^y dy = x dy \tag{2.6}$$

If we allow ourselves manipulate the differentials algebraically as though they were numbers, we can solve for

$$\frac{dy}{dx} = \frac{1}{x} \tag{2.7}$$

which is indeed the derivative of $\ln x$. There is some non-trivial mathematics behind showing that you can treat differentials in this way, but the important thing for us as Physicists is that it not only seems like a reasonable thing to do, it works.

2.2.2 Using Differentials in Multi-Variable Calculus

If we have some physical quantity w which can be written as a function[1] $f(x, y, z)$ of three variables x, y, and z, the three partial derivatives of f are defined as

$$\frac{\partial w}{\partial x} = f_x(x, y, z) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y, z) - f(x, y, z)}{\Delta x} \tag{2.8a}$$

$$\frac{\partial w}{\partial y} = f_y(x, y, z) = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y, z) - f(x, y, z)}{\Delta y} \tag{2.8b}$$

$$\frac{\partial w}{\partial z} = f_z(x, y, z) = \lim_{\Delta z \to 0} \frac{f(x, y, z + \Delta z) - f(x, y, z)}{\Delta z} \tag{2.8c}$$

sometimes we'll write $\left(\frac{\partial w}{\partial x}\right)_{y,z}$, $\left(\frac{\partial w}{\partial y}\right)_{x,z}$, and $\left(\frac{\partial w}{\partial z}\right)_{x,y}$, to emphasize which variables are being held constant in the partial derivatives.

---

[1] In this section of the notes, I'm being extra-explicit about using different letters for the same "physical" quantities depending on which set of variables they're being considered functions of. This is the Mathematician's way of doing things. In fact, as Physicists, we wouldn't normally bother with this, and would write

$$w = w(x, y, z)$$

or

$$w(t) = w(x(t), y(t), z(t))$$

Now, we can summarize the three partial derivatives in a single expression involving "differentials", i.e., infinitesimal changes in the various variables. The basic idea is that if we make infinitesimal changes dx, dy, and dz in the variables x, y, and z, that will lead to an infinitesimal change

$$dw = f(x+dx, y+dy, z+dz) - f(x,y,z) \tag{2.9}$$

Off the bat, we can't relate this to any of the partial derivatives, since all three variables are changing at once. But if we add and subtract the right terms, we can focus on one variable changing at a time:

$$\begin{aligned} dw = {} & \overbrace{f(x+dx, y+dy, z+dz) - f(x+dx, y+dy, z)}^{f_z(x+dx,y+dy,z)\,dz} \\ & + \underbrace{f(x+dx, y+dy, z) - f(x+dx, y, z)}_{f_y(x+dx,y,z)\,dy} \\ & + \underbrace{f(x+dx, y, z) - f(x,y,z)}_{f_x(x,y,z)\,dx} \end{aligned} \tag{2.10}$$

Now the last term is just the partial derivative $\frac{\partial w}{\partial x}$ times the differential dx. The middle term is nearly the partial derivative $\frac{\partial w}{\partial y}$ times dy, except it's evaluated at $(x+dx, y, z)$ instead of (x,y,z). In fact,

$$f_y(x+dx, y, z) - f_y(x,y,z) = f_{yx}(x,y,z)\,dx \tag{2.11}$$

(where f_{yx} is a second derivative), but if we use this fact, we see that

$$f_y(x+dx, y, z)\,dy = f_y(x,y,z)\,dy + \cancel{f_{yx}(x,y,z)\,dy\,dx} \tag{2.12}$$

We can drop the correction term because it contains two differentials, and one of the principal rules of this differential formalism is that anything that's the product of two differentials can be ignored. (The idea is that we will eventually divide everything by a single differential and then let all infinitesimal quantities go to zero, so anything with two infinitesimal factors will vanish even after we divide out one of them.) A similar argument allows us not to care about the argument of $f_z = \frac{\partial w}{\partial z}$ and thus write

$$dw = f_x(x,y,z)\,dx + f_y(x,y,z)\,dy + f_z(x,y,z)\,dz = \frac{\partial w}{\partial x}\,dx + \frac{\partial w}{\partial y}\,dy + \frac{\partial w}{\partial z}\,dz \tag{2.13}$$

Note that this lets us calculate all the partial derivatives of w "at once": Suppose $w = x^2 + yz + z^2$. Then straightforward application of the chain and product rules fives us

$$dw = d(x^2) + d(yz) + d(z^2) = 2x\,dx + z\,dy + y\,dz + 2z\,dz = 2x\,dx + z\,dy + (y+2z)\,dz \tag{2.14}$$

from which we can read off the partial derivatives

$$\frac{\partial w}{\partial x} = 2x \tag{2.15a}$$

$$\frac{\partial w}{\partial y} = z \tag{2.15b}$$

$$\frac{\partial w}{\partial z} = y + 2z \tag{2.15c}$$

2.3 The Chain Rule

The chain rule as expressed by a mathematician is typically written as follows:

If $f = g \circ h$ is the function defined by the composition of g and h, i.e.,

$$f(x) = g(h(x)) \tag{2.16}$$

then the derivative of f is given in terms of the derivatives of g and h by

$$f'(x) = g'(h(x))\, h'(x) \tag{2.17}$$

Written that way, it's not exactly easy to remember. But let's rewrite it in Leibniz notation. First, for notational simplicity, let's define

$$y = h(x) \tag{2.18a}$$

and

$$z = f(x) = g(h(x)) = g(y) \tag{2.18b}$$

Now in Leibniz notation,

$$\frac{dz}{dx} = f'(x) \tag{2.19a}$$

$$\frac{dy}{dx} = h'(x) \tag{2.19b}$$

$$\frac{dz}{dy} = g'(y) \tag{2.19c}$$

and the chain rule (2.17) becomes

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \tag{2.20}$$

Again, the calculus-through-algebra outlook tells us that this is really obvious, since one can just cancel out the dys.

Note that our notation as Physicists is also simpler in that we don't need to define separate functions $f(x)$ and $g(y)$; we're usually interested more in the quantity z than in the functional form of its relationship to x or y.

2.4 The Chain Rule in Multi-Variable Calculus

Note that the chain rule is not always as simple as straightforward algebra when functions can depend on more than one variable. Consider the function $w = F(t) = f(x,y,z) = f(g(t),h(t),k(t))$. Written as the function $F(t)$, w depends only on one variable, so we can define an ordinary derivative

$$\frac{dw}{dt} = F'(t) \tag{2.21}$$

while, if we see it as a function $f(x, y, z)$ of three variables, we have to define partial derivatives with respect to all three individually:

$$\frac{\partial w}{\partial x} = f_x(x, y, z) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y, z)}{\Delta x} \tag{2.22a}$$

$$\frac{\partial w}{\partial y} = f_y(x, y, z) = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y, z)}{\Delta y} \tag{2.22b}$$

$$\frac{\partial w}{\partial z} = f_z(x, y, z) = \lim_{\Delta z \to 0} \frac{f(x, y, z + \Delta z)}{\Delta z} \tag{2.22c}$$

Of course $x = g(t)$, $y = h(t)$, and $z = k(t)$ are all functions of one variable, so they have ordinary derivatives

$$\frac{dx}{dt} = g'(t) \tag{2.23a}$$

$$\frac{dy}{dt} = h'(t) \tag{2.23b}$$

$$\frac{dz}{dt} = k'(t) \tag{2.23c}$$

So how does the chain rule apply to this multi-variable problem? We have to consider what happens when we make an infinitesimal change dt in t; we know that the corresponding infinitesimal change in w will be, by definition,

$$dw = F(t + dt) - F(t) = F'(t)\, dt = \frac{dw}{dt}\, dt \tag{2.24}$$

On the other hand, the change in t will lead to changes in x, y and z according to (2.23):

$$dx = g'(t)\, dt \tag{2.25a}$$

$$dy = h'(t)\, dt \tag{2.25b}$$

$$dz = k'(t)\, dt \tag{2.25c}$$

Substituting this into (2.13) gives

$$dw = f_x(g(t), h(t), k(t))\, g'(t)\, dt + f_y(g(t), h(t), k(t))\, h'(t)\, dt + f_z(g(t), h(t), k(t))\, k'(t)\, dt \tag{2.26}$$

and comparing this to (2.24) gives the multi-variable chain rule. A Mathematician would write this as the almost incomprehensible

$$F'(t) = f_x(g(t), h(t), k(t))\, g'(t) + f_y(g(t), h(t), k(t))\, h'(t) + f_z(g(t), h(t), k(t))\, k'(t) \tag{2.27}$$

In Leibniz notation, however, this becomes

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt} \tag{2.28}$$

It's almost as though we've "divided" (2.13) by dt. That's not really what we've done, but it's a good way to remember the form of the equation.