Where Have You Gone Infinitesimals?

Windham loves you more than you can know

Michael P. Windham

August 24, 2004

Also available at

http://www.southalabama.edu/mathstat/personal_pages/windham/

1. Introduction

In the seventeenth century people talked of infinitesimal quantities and instantaneous rates. It was all very intuitive, not at all rigorous, but you got the right answers, if you followed the rules. In the twenty-first century we have tensor bundles on smooth manifolds, which are very abstract, but completely rigorous and practically inscrutable. They describe the same thing, but how the two are related is not easy to see and no one seems to talk about it.

If you pick up a mathematics book on differential geometry, you get the tensor bundles. There may be references to the classical notation, but not the classical thinking and how it relates to the modern approach. There is one exception, Michael Spivak's *A Comprehensive Introduction to Differential Geometry*, but the forest often gets lost in the trees. You think you could get there in a physics book where a discussion might begin with "suppose you have an infinitesimal displacement dx." But, within half a page you are looking at contravariant and covariant vectors and equations for how they transform. You are in the world of tensor bundles and you don't even know how you got there.

So, how do you get from seventeenth century calculus to tensor bundles? I hope to answer that question. I am not going to give a historical chain of events, but rather a mathematical chain of events. These events may never have taken place the way I will describe them, but they could have.

2. Infinitesimal quantities and instantaneous rates

So back to the seventeenth century. You would find people having no qualms at all in talking about an infinitesimal amount, dx, of some quantity represented by the variable x. They might think of dx as the infinitesimal change in the variable from x to $x + dx$. In fact, they might even say dx is the displacement from the point x to the *next* point $x + dx$. I read just this statement in one of those physics books I mentioned. It was written in the twentieth century!

The seventeenth century folk were trying to deal with things that changed all the time, such as the speed of an object as it moves along. If the speed was constant then all was well. The relation between time, t, distance traveled, s, and speed, v, was just a linear calculation, $s = vt$. But, Galileo rolled enough balls down inclined planes to deduce that the distance the balls traveled was roughly proportional to the time squared, so how fast were the balls going?

Leibnitz might have put it like this. If $s = t^2$, to keep it simple, you need to know how far the ball travels in an instant of time, dt, and that would be an infinitesimal distance, ds, which you could obtain by

$$\begin{aligned} ds &= (t + dt)^2 - t^2 \\ &= t^2 + 2t\,dt + (dt)^2 - t^2 \\ &= 2t\,dt + (dt)^2 \end{aligned}$$

and then you "throw away the higher order term" $(dt)^2$ to get the result

$$ds = 2t\,dt \tag{1}$$

Throw away higher order terms! Why? Well, because it works! You can get the speed, v, by dividing ds by dt. For one thing, the relation between infinitesimal distance, instants of time and speed is linear. More to the point, at, say, $t = 3$, you have gone $s = 9$, and at that instant your speed is $v = 6$. Just draw a graph of $s = t^2$ and measure the slope of the graph at the point $(3, 9)$. In other words, infinitesimal calculus was not rigorous. There were rules that one followed because they gave the right answers, and if they did not, they were changed. People were a bit distressed by this lack of a firm foundation. (By the middle of the eighteenth century people had begun to formulate the idea of a limit, which allowed you to watch the higher order terms disappear gracefully, rather than be arbitrarily discarded.) Calculus was an experimental science. Nevertheless, it did work.

The result (1) is called a *differential equation* relating the *differential* of s, ds, to the differential of t, dt. The equation describes the effect of an infinitesimal change in time to produce an infinitesimal change in location. In general, $f(x)dx$ is an infinitesimal amount of something produced by an infinitesimal amount of x.

In one variable calculus, any infinitesimal quantity $f(x)dx$ is the differential of some variable.[1] That is, $f(x)dx = dy$ for some variable y, namely, $y = \int f(x)dx$. The Fundamental Theorem of Calculus says that these two equations are equivalent. When there is more than one independent variable an infinitesimal quantity may not be somebody's differential. If you had a force in plane with components y, $-x$ in the x and y directions, respectively, then $y\,dx - x\,dy$ would be the (infinitesimal) amount of work needed to move from (x, y) to $(x + dx, y + dy)$. If you wanted to calculate the total work to

[1] Assuming f is sufficiently nice, say continuous. I will always assume that functions are sufficiently nice for what I say to be true. So, send me email only if what I say is never true.

move along a path you would just add up the work to move through each point, namely $\int y\,dx - x\,dy$. The integral symbol is, after all, an elongated "s" for "sum." However, $y\,dx - x\,dy$ is not an infinitesimal change, that is, the differential of some variable. If $y\,dx - x\,dy = dz$ for a variable z, then $y = \partial z/\partial x$ and $-x = \partial z/\partial y$, so that $\partial^2 z/\partial y \partial x = 1$ and $\partial^2 z/\partial x \partial y = -1$ and that cannot happen since the two second order derivatives should be the same. We tend to think of calculus as the mathematics of change, but there is more to it than that.

So, what do you do with infinitesimal quantities? Well, if you divide one by the other you get an *instantaneous rate*, the amount of one quantity per unit of the other at a single point. Since these are infinitesimal quantities the rate can change from point to point. Continuing the work example, if dt is an instant of time then dividing by it you get

$$y\,\frac{dx}{dt} - x\,\frac{dy}{dt} \tag{2}$$

the amount of work done per unit time by the particle as it moves along from (x, y) to $(x + dx, y + dy)$, that is, the power required at that instant to move from that point to the next. It is an instantaneous rate in that the power at the next point will be different.

Great. Calculus is just arithmetic of infinitesimals. But wait! How big is dx, what is the length of a point, isn't it 0? You've just divided by zero! People wrote papers and articles back and forth, one side saying "this infinitesimal business was nonsense" and the other saying, "but it works." One of the former was Bishop Berkeley and one of the latter was Isaac Newton. Which name do you recognize? I rest my case.

Consider, if you will, the following two innocent looking formulas. The seventeenth century folk would say that if you have variables x and y, then you have

$$dy = \frac{dy}{dx}\,dx \qquad \text{and} \qquad \frac{1}{dy} = \frac{dx}{dy}\,\frac{1}{dx} \tag{3}$$

the formula on the left relates infinitesimals in terms of y to infinitesimals in terms of x, and the formula on the right relates rates with respect to y to rates with respect to x. Note that dividing through by an infinitesimal has become multiplying by its reciprocal. These formulas look like you are just doing arithmetic and to the seventeenth century people you were. There is a subtlety that should occur to us, but not necessarily the seventeenth century folk, namely that dx and dy are infinitesimals that have a heuristic existence and dy/dx and dx/dy are really derivatives that are real functions. The formulas don't look like much, but they are an important step toward the modern approach. For the sophisticated reader, the one on the left has to do with the cotangent bundle and the one on the right with the tangent bundle, but I get ahead of myself. For the rest of us, just let me say that in going from x to y the differential is *covariant* and the rate is *contravariant*, which you can say refers to which variable is on top and which is on the bottom of the derivative in the formula. Covariant means that you are going the same way, that is, going to the new y from the old x you need the derivative of the new with respect to the old. Contravariant is the other way around.

To keep going we need to look at a multivariable setting where you have more than one independent variable. We might as well go for broke and say we have variables $x_1, x_2, \ldots, x_n$. Infinitesimal quantities become[2]

$$\omega = \omega_1\, dx_1 + \ldots + \omega_n\, dx_n = \sum \omega_i\, dx_i$$

where ω is the total infinitesimal amount of a quantity and the term $\omega_i\, dx_i$ is the contribution to the total from the amount dx_i. Rates become

$$V = \sum v^i \frac{1}{dx_i}$$

where $v^i\ 1/dx_i$ is the amount per unit x_i contributed to the rate per unit V.

Finally, the analog of dividing one infinitesimal quantity by another is to multiply ω and V together to get the instantaneous amount of ω per unit V,

$$\begin{aligned} \omega \cdot V &= \sum \omega_i v^j \frac{dx_i}{dx_j} \\ &= \sum \omega_i v^i \end{aligned} \tag{4}$$

where we have used the seventeenth-century-like rule

$$\frac{dx_i}{dx_j} = \delta^i_j = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

which simply says that one unit of x_i produces one unit of x_i, that's deep, and and one unit of x_j does not produce any of x_i because they are independent of each other.

[2]I am not going to put indices on summation symbols. Just assume that any index inside the sum that does not appear outside the sum is summed over. For that matter, I am not going to put punctuation at the end of displayed formulas, so sue me.

A most important example may help pull things together. Suppose $\omega = dy$ an infinitesimal change in the variable y, then

$$\omega = dy = \sum \frac{\partial y}{\partial x_i} dx_i$$

so that

$$\omega \cdot V = \sum \frac{\partial y}{\partial x_i} v^i \tag{5}$$

is just the directional derivative of y in the direction of the vector $v = \langle v^1, \ldots, v^n \rangle$, the instantaneous rate of change of y along the direction of the vector v. I have momentarily lapsed into modern jargon by using the term vector, but I could not help myself.

It is fairly easy to imagine an infinitesimal quantity ω, but where does a rate operator V come from? Let me present an intuitively appealing example. Suppose you have a curve parameterized by a function $c(t) = (x_1(t), \ldots, x_n(t))$. The derivative of c, is

$$c'(t) = \left\langle \frac{dx_1}{dt}, \ldots, \frac{dx_n}{dt} \right\rangle$$

and $c'(t)$ is the tangent vector to the curve at the point $c(t)$ on the curve. To make the example more concrete you can think of $c(t)$ as the path of a particle moving through space, then $c'(t)$ is its velocity. If you have some quantity y you measure as you move along the curve, you have the instantaneous rate of change of the quantity with respect to t given by

$$\frac{dy}{dt} = \sum \frac{\partial y}{\partial x_i} \frac{dx_i}{dt} \tag{6}$$

so that the rates with respect to t are just

$$\frac{1}{dt} = \sum \frac{dx_i}{dt} \frac{1}{dx_i}$$

or $1/dt = V$, where $v^i = dx_i/dt$. Therefore, "per unit V" is per unit t along the curve and the vector v mentioned in the previous paragraph is the tangent vector to the curve. This example is more than appealing. Any V can be obtained this way. Suppose you have a V, pick a point $x_0 = (x_{01}, \ldots, x_{0n})$, then on an interval about $t = 0$, solve the initial value problem

$$\begin{aligned} c'(t) &= \langle v^1(c(t)), \ldots, v^n(c(t)) \rangle \\ c(0) &= x_0 \end{aligned}$$

to get a curve whose tangent vectors are given by V.

3. The modern approach

I think we are ready to see what became of infinitesimals. We all know that ratios of infinitesimals became derivatives, but that only gets you through Calculus I. A warning: we are going to take a big leap, but hang on, we will come down softly. Just remember the Mathematic Oath: Pick a result and find a starting point from which you can rigorously arrive at the result, even if the starting point is out in left field.

Before we get to the fate of infinitesimals, we need to look at variables. Essentially everything I have said so far has been about variables, independent variables and dependent variables. So the first step toward modern is to replace dependent variables with functions, $f : R^n \to R$. That was easy enough. But, wait we need to stick to nice functions, I will call them *smooth*, but I am not going to tell you what they are yet, even though it violates the Mathematic Oath. If you have heard the word smooth in this context, then go with what you have heard, but I may have a surprise for you later.

Next comes (or goes depending on your attitude) the rate V. At a point p in R^n, let $\mathcal{F}_p$ be the set of smooth functions defined on a neighborhood of p. A *derivation* at p is a function $V_p : \mathcal{F}_p \to R$ satisfying the following for f and g in $\mathcal{F}_p$ and a and b in R.

1. If $f = g$ on an open set containing p, then $V_p(f) = V_p(g)$.
2. $V_p(af + bg) = aV_p(f) + bV_p(g)$
3. $V_p(fg) = g(p)V_p(f) + f(p)V_p(g)$

The first condition just say a derivation acts locally on functions. The other two say that if it acts like a derivative, then it is a derivative. So, this approach to rates is not too bad, the rate operation is just an operator on functions, which is as it should be. The set of all derivations at p, for which I will write T_p, can be made into a real vector space in the usual way for sets of real valued functions. This is a pretty abstract approach to something that started out as dividing by an infinitesimal, but having been born in the twentieth century I could see the rationale behind it, particularly when I found out what comes next.

If $c : R \to R^n$ is a curve in R^n going through p with $c(0) = p$, then the function $V_p : \mathcal{F}_p \to R$ defined by

$$V_p(f) = \frac{df \circ c}{dt}(0) \tag{7}$$

is a derivation at p. When you compare this to (6) and the discussion around it, it comes as no surprise that a derivation is also called a *tangent vector* at p and T_p is called the *tangent space* at p. I feel better now. A *vector field* is just a bunch of tangent vectors glued together smoothly, that is, a vector field V is a function that assigns to each point p in an open set of R^n a tangent vector V_p in T_p so that for each smooth function f on the open set, the function $V(f)$ defined by $V(f)(p) = V_p(f)$ is also smooth.

In fact, to get the tangent bundle all you do is glue all the tangent spaces together, in a "smooth" way, but I think that is as close as I will get to bundles. You can chase them from here yourself.

Now at last we come to the demise of the infinitesimal. The infinitesimal quantity ω is replaced at a point p in R^n by a linear function $\omega_p : T_p \to R$. Now, that is definitely out in left field. This definition is determined not by what an infinitesimal quantity might mean to you, but is determined entirely by how it is used. How it is used is summarized completely by (4). In the work example, I said that to obtain the amount of work in going from one point to another you add up the infinitesimal amounts along the way by integrating. But, what you really integrate is the function in (2) along a path between to two points. The vector $\langle dx/dt, dy/dt\rangle$ is the tangent vector to the path along the way. You are using the linear relationship in (4) between the infinitesimal amount of work along the curve and the tangent vector to the curve to get the thing you really need to integrate. The directional derivative computed in (5) is another example. That result can be achieved by defining for a function f in $\mathcal{F}_p$ a linear function $df_p : T_p \to R$ by

$$df_p(V_p) = V_p(f) \tag{8}$$

Well, there you have it. Our warm, fuzzy little infinitesimal amount of something is replaced by a linear function. Why? I guess you could say because it works.

We might as well keep going. The set of all real-valued linear functions defined on T_p is denoted T_p^* and is called the *cotangent* space at p, the linear functions are called *cotangent vectors*. The cotangent space is also a real vector space in the obvious way and is well known in linear algebra as the dual space of T_p. When you build cotangent fields in a way analogous to the way you build vector fields (I leave the details as an exercise), you do not call them cotangent fields, you call them *differential forms* in deference to their now long lost ancestor, the infinitesimal differential. Glue the cotangent spaces together smoothly and you get the cotangent bundle.

What now, the linear functions on the cotangent space? We already have them, they are the tangent vectors, just define for each tangent vector V_p, the function $V_p : T_p^* \to R$ by

$$V_p(\omega_p) = \omega_p(V_p) \tag{9}$$

So, what does "smooth" really mean? Before I tell you I would like you to observe that in this modern approach not only did I do away with the dependent variables, I have never used any independent variables! In other words I have never referred to a coordinate system in R^n. You are supposed to be impressed, I have created multivariable calculus without any variables. If you liked what I have done so far, you will love how I am about to define "smooth functions." In fact, I will let you participate. You get to choose the starting point. Pick from one of the following and do it.

1. Choose a family of real-valued functions defined on open sets in R^n
2. Choose a family of curves from R to R^n

If you do number 1, then we will call your family the smooth functions and any curve c for which $f \circ c : R \to R$ is infinitely differentiable for any f among your smooth functions we will call a smooth curve. If you do number 2, we will call your curves smooth and the smooth functions will be the f's that satisfy $f \circ c$ is infinitely differentiable for any c among your smooth curves. I only hope that you choose enough, but not too many, smooth things in your family so that both families are useful or at least interesting. When you think about, all we really need are curves and functions, then (7) gives you tangent vectors and tangent vectors give you cotangent vectors. In fact, you might even be able to show that any derivation can be obtained from (7).

The problem with using a coordinate system is that there are so many of them. What this approach to the modern calculus shows is that the entities that constitute calculus do not depend on a coordinate system. If you do pick one, calculus is already there waiting for you. You might say that the symbol R^n itself implies that there already is a coordinate system, but not really. I could say that R^3 is euclidean space, or R^4 is spacetime. Well, OK, what would R^5 be, you say?

So, what is a coordinate system? It is just an assignment to each point p in R^n, a unique n-tuple of real numbers $(x_1, \ldots, x_n)$, so that every point is assigned to some n-tuple. You have, then, n functions, $x_i : R^n \to R$, call them the *coordinate functions*. If you want to have any fun, they should be among your smooth functions. What then, if you don't need a coordinate system, why have it? Without a coordinate system calculus is probably too abstract. You would be hard pressed to use it to pay the bills without the variables that we use to link the calculus to the real world. So, let's have one.

The coordinate function x_i gives us a differential form dx_i using (8) and then a vector field ∂x_i using (9). These guys are wonderful because their values at each point are bases for the cotangent and tangent spaces at the point. To prove this is messy, so I won't. In fact, they are what are called dual bases, which means $dx_i(\partial x_j) = \partial x_j(dx_i) = \delta^i_j$. So, we have that any vector field and differential form can be written

$$V = \sum v^i \partial x_i$$

and

$$\omega = \sum \omega_i dx_i$$

and

$$\omega(V) = V(\omega) = \sum \omega_i v^i$$

where ω_i and v^i are smooth functions. Home again. In fact, to really make you feel at home, it is common practice to write

$$\frac{\partial x_j}{\partial x_i} \quad \text{for} \quad \partial x_i(x_j)$$

and

$$\frac{\partial f}{\partial x_i} \quad \text{for} \quad \partial x_i(f)$$

and even

$$\frac{\partial}{\partial x_i} \quad \text{for} \quad \partial x_i$$

Perhaps you have noticed that I have used superscripts for the coefficients of tangent vectors and subscripts for the coefficients of cotangent vectors. There is a reason. It helps you remember how the basis vectors and coefficients in one system are related those in another. In particular, it happens that, for two coordinate systems $(x_1, \ldots, x_n)$ and $(\bar{x}_1, \ldots, \bar{x}_n)$

$$d\bar{x}_\alpha = \sum \frac{\partial \bar{x}_\alpha}{\partial x_i} dx_i \text{ so that } \bar{\omega}_\alpha = \sum \frac{\partial x_i}{\partial \bar{x}_\alpha} \omega_i$$

and

$$\frac{\partial}{\partial \bar{x}_\alpha} = \sum \frac{\partial x_i}{\partial \bar{x}_\alpha} \frac{\partial}{\partial x_i} \text{ so that } \bar{v}^\alpha = \sum \frac{\partial \bar{x}_\alpha}{\partial x_i} v^i$$

So, for a coefficient with a superscript inside a sum the corresponding variable is on the bottom of the partial derivative and for those with a subscript the corresponding variable is on the top. I could have even made that happen in the formulas for the differential forms by writing the coordinate functions x^i, but then I would have to declare that, since the superscript in $\frac{\partial}{\partial x^i}$ is in the bottom of the "fraction" consider it to be a subscript. Besides, I can live with v^2, but x^2 has been x squared too long for me. If an index inside a sum appears on the top and on the bottom, it is summed over. This cute fact is the basis for the infamous Einstein summation convention where you leave off the summation symbols entirely. In any case, note that the change of basis for cotangent spaces is covariant and the change of basis for the tangent space is contravariant (compare to (3)). Unfortunately the coefficients change the opposite way, so that the terminolgy can be confusing.

That's about it. You can take the rest of the trip yourself, but let me fulfill the promise I made in the beginning and get you started, although I will be fast and loose - so what else is new.

4. Manifolds and tensor bundles

A *n-dimensional manifold* M is a topological space together with homeomorphisms from open subsets of M to R^n, so that each point of M is in the domain of at least one of these homeomorphisms. What did he say?! All I said was you have a space with coordinate systems(s) drawn on it, at least locally. Take a sphere and draw lines of latitude and longitude on it and you are there. The manifold is *smooth* or *differentiable*, means that whenever two coordinate systems overlap, the function that takes one set of coordinates to the other is infinitely differentiable.

A *tensor* τ_p of type (N, M) at a point p in the manifold is just a function

$$\tau_p = \underbrace{T_p^* \times \ldots \times T_p^*}_{N \text{ copies}} \times \underbrace{T_p \times \ldots \times T_p}_{M \text{ copies}} \to R$$

that is linear in each of its variables. So, cotangent vectors are tensors of type $(0, 1)$ and tangent vectors of type $(1, 0)$. In fact, you can call the real numbers themselves tensors of type $(0, 0)$.

The tensors of type (N, M) are a real vector space in the usual way. You can even define a multiplication of tensors where the tensor product of τ_p of type (N, M) and σ_p of type (L, K) gives a tensor of type $(N + L, M + K)$ denoted $\tau_p \otimes \sigma_p$, but this is another story. Glue tensors of the same type at each point together smoothly and you have a *tensor field*. So, what are they good for? Well, if nothing else, they generalize and unify real valued functions (which are now tensor fields of type $(0, 0)$), vector fields and differential forms to an incredible, but esthetically pleasing level. If that were all, well ... Let me just say that they provide what you need to do differential geometry and talk about things like curvature. They are wonderful for physicists for describing physical laws. In fact, the most famous equation of all, the Einstein gravitational field equation an equality between two type $(0, 2)$ tensor fields, one describing the geometry of spacetime and the other describing the distribution of mass in it. Maxwell's equations for electromagnetism are tensor equations.

In a coordinate system you can build a basis for type (N, M) tensors at a point using the dx_i's and ∂x_j's. The coefficients are grotesque

$$\mathcal{T}_{i_1 \dots i_M}^{j_1 \dots j_N} \tag{10}$$

and the relation between coefficients in different coordinate systems is even worse

$$\bar{\mathcal{T}}_{\alpha_1 \dots \alpha_M}^{\beta_1 \dots \beta_N} = \sum \frac{\partial x_{i_1}}{\partial \bar{x}_{\alpha_1}} \cdots \frac{\partial x_{i_M}}{\partial \bar{x}_{\alpha_M}} \frac{\partial \bar{x}_{\beta_1}}{\partial x_{j_1}} \cdots \frac{\partial \bar{x}_{\beta_N}}{\partial x_{j_N}} \mathcal{T}_{i_1 \dots i_M}^{j_1 \dots j_N} \tag{11}$$

The only reason I mention this is that people will define tensors using these formulas, that is, a tensor field is a bunch of real valued functions (10) related by (11) in different coordinate systems. Physicists are the worst for doing this, but I have done it myself.

5. Fini

Enough already. You take it from here - really. I am tired of trying to type infinitesimal. I will say that I miss infinitesimals. They are so intuitively appealing. Actually, I lied, I don't miss infinitesimals at all. I use them all the time, I teach my students to use them. You can get the right answers, if you do. Even better, you can formulate the right questions about the world and that is the best of all.