THE IDEA OF DIFFERENTIABILITY FOR FUNCTIONS OF SEVERAL VARIABLES

Math 225

As we have seen, the definition of derivative for a Math 111 function $g\colon \mathbb{R} \to \mathbb{R}$ and for a curve $\gamma\colon \mathbb{R} \to E^n$ are the same, except for interpretation:

$$(1) \qquad g'(a) = \lim_{h\to 0} \frac{g(a+h)-g(a)}{h} \qquad \text{and} \qquad \gamma'(t_0) = \lim_{h\to 0} \frac{\gamma(t_0+h)-\gamma(t_0)}{h},$$

provided the limits exist. When they do exist, we say g and γ are differentiable at $a \in \mathbb{R}$ and $t_0 \in \mathbb{R}$, respectively. By letting $x = a+h$ in the first formula and $t = t_0 + h$ in the second, these can be written as

$$(2) \qquad g'(a) = \lim_{x\to a} \frac{g(x)-g(a)}{x-a} \qquad \text{and} \qquad \gamma'(t_0) = \lim_{t\to t_0} \frac{\gamma(t)-\gamma(t_0)}{t-t_0}.$$

The difference between these is that the numerator is a scalar in the definition of $g'(a)$, and is a vector in the definition of $\gamma'(a)$, leading to the interpretation of the slope of the tangent line in the first case and the velocity vector in the second.

Now suppose $f\colon E^n \to \mathbb{R}$ or $F\colon E^n \to E^k$, and let $P \in E^n$. We want to define what it means for $z = f(X)$ and $Z = F(X)$ to be differentiable at P. A straight forward mimic of (1) and (2) would be

$$f'(P) \stackrel{?}{=} \lim_{X\to P} \frac{f(X)-f(P)}{X-P} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{f(P+\mathbf{h})-f(P)}{\mathbf{h}},$$

where $\mathbf{h} = X - P$ is the displacement vector from P to X. This is nonsense because we can't divide by vectors.

Note that we *can* define partial derivatives and directional derivatives in this way because they are essentially functions of one variable. The partial derivative of $z = f(x,y)$ at $(x,y) = (a,b)$ with respect to x is

$$f_x(a,b) = \lim_{x\to a} \frac{f(x,b)-f(a,b)}{x-a} = \lim_{\Delta x\to 0} \frac{f(a+\Delta x,b)-f(a,b)}{\Delta x},$$

and the derivative of F at P along the vector $\mathbf{v}$ is

$$D_{\mathbf{v}}F(P) = \lim_{t\to 0} \frac{F(P+t\mathbf{v})-F(P)}{t}.$$

The solution is to realize that the differentiability of $g\colon \mathbb{R} \to \mathbb{R}$ and $\gamma\colon \mathbb{R} \to E^n$ can be viewed in another way that *can* be generalized to $f\colon E^n \to \mathbb{R}$ and $F\colon E^n \to E^k$. Informally (2) says that

$$g'(a) \approx \frac{g(x)-g(a)}{x-a} \qquad \text{and} \qquad \gamma'(t_0) \approx \frac{\gamma(t)-\gamma(t_0)}{t-t_0}$$

when x is close to a and t is close to t_0. Solving for $g(x)$ and $\gamma(t)$ we have

$$(3) \qquad g(x) \approx g(a) + g'(a)(x-a) \qquad \text{and} \qquad \gamma(t) \approx \gamma(t_0) + (t-t_0)\gamma'(t_0).$$

In the right hand sides we recognize $y = g(a) + g'(a)(x-a)$ as the equation of the tangent line to the graph of g from Math 111 and $X = \gamma(t_0) + \gamma'(t_0)(t-t_0)$ as a parameterization of the tangent line to the image of the curve γ. The idea is that a function is differentiable if it can be approximated by a linear function. The functions $f\colon E^n \to \mathbb{R}$ and $F\colon E^n \to E^k$ should be differentiable at P if there are good approximations

$$f(X) \approx f(P) + f'(P)(X-P) \qquad \text{and} \qquad F(X) \approx F(P) + F'(P)(X-P).$$

But what do these mean, and how can we interpret $f'(P)$ and $F'(P)$ if we don't know what differentiability should mean? It's not quite a logical circle. Once we know how good the approximation needs to be, we can actually replace the slope $g'(a)$ and the velocity vector $\gamma'(t_0)$ in (3) with an arbitrary slope m and vector $\mathbf{v}$, namely if the approximations

$$g(x) \approx g(a) + m(x-a) \qquad \text{and} \qquad \gamma(t) \approx \gamma(t_0) + (t-t_0)\mathbf{v}$$

are good enough in a suitable sense, then m and $\mathbf{v}$ are uniquely determined.

Why the Tangent Line is the Best Linear Approximation in the One-Variable Case. An arbitrary non-vertical line through $(a, g(a))$ has the form $y = g(a) + m(x-a)$, which you can get by the point-slope formula. Suppose we approximate $y = g(x)$ by this line near $x = a$:

$$g(x) \approx g(a) + m(x-a).$$

To analyze this approximation, let's look at the error it makes, namely

$$\epsilon = g(x) - g(a) - m(x-a).$$

We want to think of this as a function of how close x is to a, and so we switch to the $h = x - a$ notation:

$$(4) \qquad \epsilon(h) = g(a+h) - g(a) - mh.$$

If g is continuous at a, then

$$\lim_{h\to 0}\epsilon(h) = \lim_{h\to 0}\big(g(a+h) - g(a) - mh\big) = g(a) - g(a) - m0 = 0.$$

The fact that the error goes to zero as x goes to a is not a big deal ... it basically just says that the curve $y = g(x)$ and the line $y = g(a) + m(x - a)$ both go continuously through the same point when $x = a$. *Every* line that goes through the point $(a, g(a))$ does this—nothing is special about the tangent line.

Instead of looking at $\lim_{h\to 0}\epsilon(h)$, which is zero no matter what the slope of the line is, let's look at $\lim_{h\to 0}\frac{\epsilon(h)}{h}$. Since the denominator goes to zero, this will be less likely to go to 0. Assuming the limit in (1) exists, we have

$$\lim_{h\to 0}\frac{\epsilon(h)}{h} = \lim_{h\to 0}\frac{g(a+h) - g(a) - mh}{h} = \lim_{h\to 0}\frac{g(a+h) - g(a)}{h} - m = g'(a) - m.$$

Notice that this is zero if and only if $m = g'(a)$. In other words, $\lim_{h\to 0}\frac{\epsilon(h)}{h} = 0$ if and only if the line is what was defined in Math 111 to be the tangent line. Here is the interpretation of $\lim_{h\to 0}\frac{\epsilon(h)}{h} = 0$ that I like. Not only is the error going to zero, it is going to zero much faster that h is going to zero, that is, much faster than x is approaching a.

We have derived this under the assumption that g is differentiable at $x = a$, but this is really unnecessary. Let's not assume that g is differentiable, but only that there is some line for which $\lim_{h\to 0}\frac{\epsilon(h)}{h} = 0$. Then

$$\lim_{h\to 0}\frac{g(a+h) - g(a)}{h} = \lim_{h\to 0}\left(\frac{g(a+h) - g(a)}{h} - m\right) + m = \lim_{h\to 0}\frac{\epsilon(h)}{h} + m = m,$$

and so g turns out to be differentiable after all, and the value of $g'(a)$ is the slope of the line.

What we have found is summarized in the following theorem.

Theorem. *Let $g\colon \mathbb{R} \to \mathbb{R}$. Consider the approximation $g(x) \approx g(a) + m(x - a)$ near $x = a$ and its error $\epsilon(h) = g(a+h) - g(a) - mh$. Then g is differentiable at a if and only if there is some slope m such that the error goes to zero so quickly that $\lim_{h\to 0}\frac{\epsilon(h)}{h} = 0$. Furthermore, if g is differentiable at a, the slope m is unique, and is equal to $g'(a)$ as defined in (1).*

A similar theorem holds for curves.

Theorem. *Let $\gamma\colon \mathbb{R} \to E^n$. Consider the approximation $\gamma(t) \approx \gamma(t_0) + (t - t_0)\mathbf{v}$ near $t = t_0$ and its error $\epsilon(h) = \gamma(t + t_0) - \gamma(t_0) - h\mathbf{v}$. Then γ is differentiable at t_0 if and only if there is some vector $\mathbf{v}$ such that the error goes to zero so quickly that $\lim_{h\to 0} \frac{\epsilon(h)}{h} = \mathbf{0}$. Furthermore, if γ is differentiable at t_0, the vector $\mathbf{v}$ is unique, and is equal to $\gamma'(t_0)$ as defined in* (1).

Note that the error $\epsilon(h)$ is a vector in this case, and it measures how well the parameterized line $X = \gamma(t_0) + (t - t_0)\mathbf{v}$ approximates $\gamma(t)$. This parameterized line goes through the right point $\gamma(t_0)$ at the right time t_0, but could be going in the wrong direction or at the wrong speed. Only one line could possibly have the same speed and direction as γ, the one with velocity $\gamma'(t_0)$.

A key point in these theorems is that they characterize the differentiability of g and γ *without* having to mention $g'(a)$ and $\gamma'(t_0)$ (the last clause could be left off the statements without changing the theorems).

One additional observation, which will help make this work in more variables: $\lim_{h\to 0} \frac{\epsilon(h)}{h} = 0$ if and only if $\lim_{h\to 0} \frac{\epsilon(h)}{|h|} = 0$ if and only if $\lim_{h\to 0} \frac{|\epsilon(h)|}{|h|} = 0$.

The Best Linear Approximation in More Variables. What can replace the approximations

$$g(x) \approx g(a) + m(x - a) \qquad \text{and} \qquad \gamma(t) \approx \gamma(t_0) + (t - t_0)\mathbf{v}$$

in the cases of $f\colon E^n \to \mathbb{R}$ and $F\colon E^n \to E^k$? For $f\colon E^n \to \mathbb{R}$, the appropriate approximation is

$$f(X) \approx f(P) + \mathbf{w} \cdot (X - P),$$

in which $\mathbf{w}$ is a vector on E^n and $\mathbf{w} \cdot (X - P)$ is dot product.

As much as I like doing things without coordinates, the coordinate version of $F\colon E^n \to E^k$ is easier to state. For $F\colon \mathbb{R}^n \to \mathbb{R}^k$ (note the change from E^n to $\mathbb{R}^n$), the appropriate approximation is

$$F(X) \approx F(P) + M(X - P),$$

in which M is a $k \times n$ matrix and $M(X - P)$ is matrix multiplication.

Part of the statements in the *theorems* for the one-variable case become *definitions* in the several-variable case.

Definition. Let $f\colon E^n \to \mathbb{R}$. Consider the approximation $f(X) \approx f(P) + \mathbf{w} \cdot (X - P)$ near $X = P$ and its error $\epsilon(\mathbf{h}) = f(P + \mathbf{h}) - f(P) - \mathbf{w} \cdot \mathbf{h}$. We say that f *is differentiable at* P if there is some vector $\mathbf{w}$ such that the error goes to zero so quickly that $\lim_{\mathbf{h}\to \mathbf{0}} \frac{\epsilon(\mathbf{h})}{\|\mathbf{h}\|} = 0$.

Definition. Let $F\colon \mathbb{R}^n \to \mathbb{R}^k$. Consider the approximation $F(X) \approx F(P) + M(X - P)$ near $X = P$ and its error $\epsilon(\mathbf{h}) = F(P + \mathbf{h}) - F(P) - M\mathbf{h}$, where M is a $k \times n$ matrix. We say that F *is differentiable at* P if there is some matrix M such that the error goes to zero so quickly that $\lim_{\mathbf{h}\to\mathbf{0}} \frac{\epsilon(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$. (Note that the error $\epsilon(h)$ is a vector in this case.)

The "furthermore" parts of the theorems are still theorems.

Theorem. *Suppose* $f\colon E^n \to \mathbb{R}$ *is differentiable at* P*. Then the vector* $\mathbf{w}$ *in the definition of differentiability is unique. In a coordinate system, the coefficients of* $\mathbf{w}$ *are the partial derivatives of* f *at* P*.*

Theorem. *Suppose* $F\colon \mathbb{R}^n \to \mathbb{R}^k$ *is differentiable at* P*. Then the matrix* M *in the definition of differentiability is unique. The columns of* M *are the partial derivatives of* F *at* P*. (Remember that the partial derivatives of* F *are vectors.)*

Anything that is unique deserves a name and some notation.

Definition. If $f\colon E^n \to \mathbb{R}$ is differentiable at P, then the vector $\mathbf{w}$ in the definition of differentiability is called the *gradient* of f at P, and is denoted by $\nabla f(P)$.

Definition. If $F\colon \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at P, then the matrix M in the definition of differentiability is called the *derivative matrix* of F at P, and is denoted by $F'(P)$.

You should be getting the idea that differential calculus is the mathematics of finding linear approximations to functions. If a function is differentiable at a point, it has a unique best linear approximation. Here is a summary:

(5)

$g\colon \mathbb{R} \to \mathbb{R}$,	$g(x) \approx g(a) + g'(a)(x - a)$
$\gamma\colon \mathbb{R} \to E^n$,	$\gamma(t) \approx \gamma(t_0) + (t - t_0)\gamma'(t_0)$
$f\colon E^n \to \mathbb{R}$,	$f(X) \approx f(P) + \nabla f(P) \cdot (X - P)$
$F\colon \mathbb{R}^n \to \mathbb{R}^k$,	$F(X) \approx F(P) + F'(P)(X - P)$

I hope it is clear that you want to view these as being variations on a common theme, rather than different.

Differentiability and Directional Derivatives. One important fact about differentiability is that it unifies directional derivatives, or rather, derivatives along vectors.

Theorem. *Suppose $f\colon E^n \to \mathbb{R}$ is differentiable at P. Then*

$$D_{\mathbf{v}} f(P) = \nabla f(P) \cdot \mathbf{v}.$$

Theorem. *Suppose $F\colon \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at P. Then*

$$D_{\mathbf{v}} F(P) = F'(P)\mathbf{v}.$$

Differentiability and Tangent Lines and Planes. The linear approximations in (5) lead directly to equations of tangent lines and planes of the objects defined by the functions and mappings. The two familiar examples are graphs of $g\colon \mathbb{R} \to \mathbb{R}$ and parameterized curves $\gamma\colon \mathbb{R} \to E^n$. Recall that for a parameterized curve, the associated object is its image.

Function	Object	Tangent Object
$g\colon \mathbb{R} \to \mathbb{R}$	Curve Graph of $y = g(x)$	Line Graph of $y = g(a) + g'(a)(x - a)$
$\gamma\colon \mathbb{R} \to E^n$	Parametric curve Image of $X = \gamma(t)$	Parametric line Image of $X = \gamma(t_0) + (t - t_0)\gamma'(t_0)$

The main idea is that the tangent object bears the same relationship to its equation as the object bears to its equation, that is, the tangent object to a graph is the graph of the linear approximation, the tangent object to a parameterized image is the parameterized image of the linear approximation, and the tangent object to a level set is an appropriate level set of the linear approximation.

Function	Object	Tangent Object
$f\colon \mathbb{R}^2 \to \mathbb{R}$	Surface Graph of $z = f(X)$	Plane Graph of $z = f(P) + \nabla f(P) \cdot (X - P)$
$f\colon \mathbb{R}^2 \to \mathbb{R}$	Curve Level set of $z = f(X)$ for const. $z = f(P)$	Line Level set of $z = f(P) + \nabla f(P) \cdot (X - P)$ for const. $z = f(P)$ Simplifies to $0 = \nabla f(P) \cdot (X - P)$
$f\colon \mathbb{R}^3 \to \mathbb{R}$	Surface Level set of $w = f(X)$ for const. $w = f(P)$	Plane Level set of $w = f(P) + \nabla f(P) \cdot (X - P)$ for const. $w = f(P)$ Simplifies to $0 = \nabla f(P) \cdot (X - P)$
$F\colon \mathbb{R}^2 \to E^3$	Parametric surface Image of $X = F(T)$ where $T = (s,t)$	Parametric plane Image of $X = F(T_0) + F'(T_0)(T - T_0)$ where $T_0 = (s_0, t_0)$

Note that the tangent object equations for level sets simplify. For $f\colon \mathbb{R}^2 \to \mathbb{R}$, you get a level curve by setting z equal to a constant in $z = f(X)$. If you want the level curve through P, the constant is $f(P)$. To get the tangent line, you set z equal to the same constant in $z = f(P)+\nabla f(P)\cdot(X-P)$. Then $f(P)$ cancels from both sides of the equation, leaving $0 = \nabla f(P) \cdot (X - P)$.

The last item in the chart needs further discussion to make the tangent plane look parametric. The parametric plane should be of the form $X = F(s_0, t_0)+(s-s_0)\mathbf{v}+(t-t_0)\mathbf{w}$, where $\mathbf{v}$ and $\mathbf{w}$ have something to do with the partial derivatives of F. Indeed, the columns of the matrix $F'(T_0) = F'(s_0, t_0)$ are the partial derivatives, and so the matrix product becomes

$$\begin{aligned} F'(T_0)(T - T_0) &= F'(s_0, t_0) \begin{pmatrix} s - s_0 \\ t - t_0 \end{pmatrix} = \begin{pmatrix} F_s(s_0, t_0) & F_t(s_0, t_0) \end{pmatrix} \begin{pmatrix} s - s_0 \\ t - t_0 \end{pmatrix} \\ &= (s - s_0)F_s(s_0, t_0) + (t - t_0)F_t(s_0, t_0). \end{aligned}$$

The parametric plane is then

$$X = F(s_0, t_0) + (s - s_0)F_s(s_0, t_0) + (t - t_0)F_t(s_0, t_0).$$

Examples will be forthcoming in N:/Math/Math225/CurvesAndSurfaces.nb.

Robert L. Foote, Fall 2007