Differential Calculus of Real-Valued Functions: A Brief Guide

G. B. Folland

Since Apostol talks about vector-valued functions right away, whereas I prefer to work with scalar-valued functions for a while and proceed to the vector-valued case later, there are some places where the way he says things is at variance with the way I do. This is a little guide to help you with translating one into the other. In what follows, all functions are real-valued, defined on $\mathbb{R}^n$ or an open subset of $\mathbb{R}^n$, unless explicitly described otherwise.

Directional Derivatives (Definition 12.1): The derivative of f at $\mathbf{c}$ in the direction $\mathbf{u}$ is

$$f'(\mathbf{c};\mathbf{u}) = \lim_{h\to 0}\frac{f(\mathbf{c}+h\mathbf{u})-f(\mathbf{c})}{h} = \frac{d}{dt}f(\mathbf{c}+t\mathbf{u})\bigg|_{t=0}.$$

Partial derivatives: The partial derivatives of f are the directional derivatives in the directions of the unit coordinate vectors:

$$D_j f(\mathbf{c}) = \partial_j f(\mathbf{c}) = \frac{\partial f}{\partial x_j}(\mathbf{c}) = f'(\mathbf{c},\mathbf{e}_j),$$

where $\mathbf{e}_j$ is the vector whose jth component is 1 and whose other components are 0. (When $n=3$, $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$ are often called $\mathbf{i},\mathbf{j},\mathbf{k}$.)

Differentiability, Total Derivatives, Gradients (Definition 12.2): The function f is *differentiable* at $\mathbf{c}$ if there is a vector $\mathbf{a}$ such that

$$f(\mathbf{c}+\mathbf{v}) = f(\mathbf{c}) + \mathbf{a}\cdot\mathbf{v} + o(\|\mathbf{v}\|),$$

in which case the vector $\mathbf{a}$ is called the *gradient* of f at $\mathbf{c}$ and is denoted by $\nabla f(\mathbf{c})$. The *total derivative* of f at $\mathbf{c}$ is the linear function that maps the vector $\mathbf{v}$ to the number $\nabla f(\mathbf{c})\cdot\mathbf{v}$. Apostol denotes it by $f'(\mathbf{c})$; thus, $f'(\mathbf{c})(\mathbf{v}) = \nabla f(\mathbf{c})\cdot\mathbf{v}$. (Other people have different notations for the total derivative, and it's also called the *differential* or *Fréchet derivative* of f.)

Theorems 12.3 and 12.5: If f is differentiable at $\mathbf{c}$, its directional derivatives at $\mathbf{c}$ all exist and are given by $f'(\mathbf{c},\mathbf{u}) = \nabla f(\mathbf{c})\cdot\mathbf{u}$. In particular, taking $\mathbf{u}=\mathbf{e}_j$, we see that the partial derivative $D_j f(\mathbf{c})$ is the jth component of $\nabla f(\mathbf{c})$; that is, $\nabla f(\mathbf{c})$ is the vector whose components are the partial derivatives of f at $\mathbf{c}$.

These results are not reversible: the existence of the partial derivatives $D_j f(\mathbf{c})$ does not guarantee the differentiability (or even continuity) of f at $\mathbf{c}$. But:

Theorem 12.11: If the partial derivatives $D_j f$ $(1\le j\le n)$ exist *and are continuous* on some open set containing $\mathbf{c}$, then f is differentiable at $\mathbf{c}$. (Apostol notes that it's enough for all but one of the partial derivatives to be continuous, but that is a minor technicality. The statement just given is easier to remember and much more important in practice.)

If the partial derivatives $D_j f$ exist and are continuous on some open set U, we say that f is *of class* C^1 on U, or that f is *a* C^1 *function* on U, or that $f\in C^1(U)$. Thus, if f is of class C^1 on U then f is differentiable at every point of U.

Next we come to the chain rule, for which we do have to consider vector-valued functions. But for the basic version we need only vector-valued functions of a scalar variable, which were introduced back in Chapter 5.

Chain Rule, first version (special case of Theorem 12.7): Suppose $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ is differentiable at a and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{b} = \mathbf{g}(a)$. Then $h = f \circ \mathbf{g}$ is differentiable at a, and

$$h'(a) = \nabla f(\mathbf{b}) \cdot \mathbf{g}'(a).$$

If we now make $\mathbf{g}$ a function of several variables, we can apply this result to the function of one of the variables obtained by fixing all the others, and thereby get a result about partial derivatives:

Chain Rule, second version (another corollary of Theorem 12.7): Suppose $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^n$ is differentiable at $\mathbf{a}$ and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{b} = \mathbf{g}(\mathbf{a})$. Then the partial derivatives of $h = f \circ \mathbf{g}$ at $\mathbf{a}$ exist and are given by

$$D_j h(\mathbf{a}) = \nabla f(\mathbf{b}) \cdot D_j \mathbf{g}(\mathbf{a}).$$

In particular, if $\mathbf{g}$ is of class C^1 on an open set U and f is of class C^1 on an open set V that includes $\mathbf{g}(U)$, then h is of class C^1 (and hence is differentiable) on U.

We'll get the chain rule in its full glory when we consider vector-valued functions more systematically. For now, here's one more important result:

Mean Value Theorem (Theorem 12.9): Suppose f is differentiable on an open set U, and $\mathbf{x}$ and $\mathbf{y}$ are two points in U such that the line segment joining $\mathbf{x}$ and $\mathbf{y}$ lies in U. Then there is some point $\mathbf{z}$ on this line segment such that

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla f(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{x}).$$