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 \( c \) in the direction \( u \) is

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

**Partial derivatives**: The partial derivatives of \( f \) are the directional derivatives in the directions of the unit coordinate vectors:

\[
D_j f(c) = \frac{\partial f}{\partial x_j}(c) = f'(c, e_j),
\]

where \( e_j \) is the vector whose \( j \)th component is 1 and whose other components are 0. (When \( n = 3 \), \( e_1, e_2, e_3 \) are often called \( i, j, k \).)

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

\[
f(c + v) = f(c) + a \cdot v + o(\|v\|),
\]

in which case the vector \( a \) is called the gradient of \( f \) at \( c \) and is denoted by \( \nabla f(c) \). The total derivative of \( f \) at \( c \) is the linear function that maps the vector \( v \) to the number \( \nabla f(c) \cdot v \). Apostol denotes it by \( f'(c) \); thus, \( f'(c)(v) = \nabla f(c) \cdot v \). (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 \( c \), its directional derivatives at \( c \) all exist and are given by \( f'(c, u) = \nabla f(c) \cdot u \). In particular, taking \( u = e_j \), we see that the partial derivative \( D_j f(c) \) is the \( j \)th component of \( \nabla f(c) \); that is, \( \nabla f(c) \) is the vector whose components are the partial derivatives of \( f \) at \( c \).

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

**Theorem 12.11**: If the partial derivatives \( D_j f (1 \leq j \leq n) \) exist and are continuous on some open set containing \( c \), then \( f \) is differentiable at \( c \). (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 \( g: \mathbb{R} \to \mathbb{R}^n \) is differentiable at \( a \) and \( f: \mathbb{R}^n \to \mathbb{R} \) is differentiable at \( b = g(a) \). Then \( h = f \circ g \) is differentiable at \( a \), and

\[
h'(a) = \nabla f(b) \cdot g'(a).
\]

If we now make \( g \) 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 \( g: \mathbb{R}^k \to \mathbb{R}^n \) is differentiable at \( a \) and \( f: \mathbb{R}^n \to \mathbb{R} \) is differentiable at \( b = g(a) \). Then the partial derivatives of \( h = f \circ g \) at \( a \) exist and are given by

\[
D_j h(a) = \nabla f(b) \cdot D_j g(a).
\]

In particular, if \( g \) is of class \( C^1 \) on an open set \( U \) and \( f \) is of class \( C^1 \) on an open set \( V \) that includes \( g(U) \), 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 \( x \) and \( y \) are two points in \( U \) such that the line segment joining \( x \) and \( y \) lies in \( U \). Then there is some point \( z \) on this line segment such that

\[
f(y) - f(x) = \nabla f(z) \cdot (y - x).
\]