DERIVATIVES AS MATRICES

1. Derivatives of Real-valued Functions

In Math 8, you studied partial derivatives, differentiable functions, gradients, tangent planes, and directional derivatives. But do you know what we mean by the *derivative* (not just partial or directional but the actual derivative) of a function $f : \mathbf{R}^2 \to \mathbf{R}$? Chances are, this notion wasn't actually mentioned, although it was lurking behind the scenes with the gradient.

1.1. Math 8 review. Let's first recall what you learned about differentiable functions $z = f(x, y)$. To keep a geometric picture in mind, let's view the graph of f as the surface of a mountain. Think of the positive x and y axes as pointing east and north, respectively, and think of z as elevation. If you stand at the point (x_0, y_0, z_0) and look due east, the mountain slope is given by $\frac{\partial f}{\partial x}((x_0, y_0))$. Similarly, $\frac{\partial}{\partial y}((x_0, y_0))$ gives the slope heading north.

(i) The gradient of f, which can be denoted either as $grad(f)$ or $\nabla(f)$ is the vector $grad(f) = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle$. The gradient points in the direction of steepest ascent.

(ii) The equation for the tangent plane to the graph of f at (x_0, y_0, z_0), where $z_0 = f(x_0, y_0)$, is given by

$$z - z_0 = \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

(iii) The *directional derivative* of f at (x_0, y_0) in the direction of the unit vector $\mathbf{u} = \langle u_1, u_2 \rangle$ was *defined* to be

$$D_{\mathbf{u}} f(x_0, y_0) = \frac{d}{dt}_{|t=0} f((x_0, y_0) + t\mathbf{u}).$$

Note that $(x_0, y_0) + t\mathbf{u}$ describes a straight line in the x, y-plane through the point (x_0, y_0) in the direction of the unit vector $\mathbf{u}$. The directional derivative $D_{\mathbf{u}} f(x_0, y_0)$ gives the rate of change in elevation as one walks at unit speed along the mountaintop in the direction $\mathbf{u}$; i.e., it gives the slope of the mountain at the point (x_0, y_0) in the direction $\mathbf{u}$. E.g, if $\mathbf{u} = \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \rangle$, then $D_{\mathbf{u}} f(x_0, y_0)$ is the slope in the northeast direction at (x_0, y_0).

(iv) Next you used the idea that the tangent line to any path on the graph of f must lie on the tangent plane to get a much simpler formula for the directional derivative:

$$D_{\mathbf{u}} f(x_0, y_0) = grad(f)((x_0, y_0)) \cdot \mathbf{u}.$$

1.2. Example. Let $f(x, y) = x^2 y^3$ and let $(x_0, y_0) = (2, 1)$. We have

$$\frac{\partial f}{\partial x}(x_0, y_0) = 4, \quad \frac{\partial f}{\partial y}((x_0, y_0)) = 12.$$

I.e., at the point $(2, 1, 4)$ on the mountain, the slope heading east is 4 and the slope heading north is 12. Thus:

- $grad(f)((x_0, y_0)) = \langle 4, 12 \rangle$.
- The tangent plane at $(2, 1, 4)$ is
$$z - 4 = 4(x - 2) + 12(y - 1).$$
- In, say, the direction of the unit vector $\mathbf{u} = \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \rangle$, we have
$$D_{\mathbf{u}} f(x_0, y_0) = \langle 4, 12 \rangle \cdot \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \rangle = \frac{16}{\sqrt{2}} = 8\sqrt{2}.$$
I.e., the slope heading northeast is $8\sqrt{2}$.

Strange!!! To compute the slope in the northeast direction (or any other specified direction for that matter) the only info we needed was the two partial derivatives 4, 12 (i.e., the slopes in the two specific directions east and north). Why should the slope in those two directions tell you the slopes in every other direction? That probably doesn't jive with your experience when hiking on a mountain!

What's going on here is that differentiability is a very strong condition. To say that a function f is differentiable (and not just that it has partial derivatives) means precisely that the plane in 1.1(ii) really is tangent to the graph of f. (There are some functions that have partial derivatives at a point (x_0, y_0) but for which the plane in 1.1(ii) is not actually tangent to the graph of f. Such functions aren't differentiable. This won't come up for us though. Nice functions like $f(x, y) = x^2y^3$ that have *continuous* partial derivatives are always differentiable.)

We've talked about differentiable functions and their partial and directional derivatives, but we still haven't said what we mean by the derivative of f. We address that now. We first introduce matrix notation.

1.3. Notation. A *row matrix* is a matrix with only one row, e.g., $\begin{bmatrix} 3 & 1 & 4 \end{bmatrix}$ is a row matrix. Similarly a *column matrix* has only one column, e.g.,
$$\begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$$
Often it is convenient to view row or column matrices as vectors or as points. For example, the two matrices above may be identified with the vector $< 3, 1, 4 >$ or the point $(3, 1, 4)$ in $\mathbf{R}^3$. In what follows, points will be identified with column matrices. For this reason, we will use lower case bold face letters such as $\mathbf{x}$ to denote column matrices.

1.4. Example. If we permit ourselves to view elements of $\mathbf{R}^2$ as points, as vectors, and as column matrices, then the following three expressions all describe the same function:
$$L(x, y) = 3x + 4y$$
$$L(x, y) = \langle 3, 4 \rangle \cdot \langle x, y \rangle$$
$$L(x, y) = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

(The third expression is given by multiplying the matrices.) We will also write the third expression as

$$L(\mathbf{x}) = A\mathbf{x}$$

where

$$A = \begin{bmatrix} 3 & 4 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Definition 1.1. If $f : \mathbf{R}^2 \to \mathbf{R}$ is differentiable at (x_0, y_0), then we define the derivative to be

$$f'(x_0, y_0) = \begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \end{bmatrix}.$$

If we view this row matrix as a vector, then it is the familiar gradient $\nabla f(x_0, y_0)$. Similarly, for a differentiable function $f : \mathbf{R}^3 \to \mathbf{R}$, we have

$$f'(x_0, y_0, z_0) = \begin{bmatrix} f_x(x_0, y_0, z_0) & f_y(x_0, y_0, z_0) & f_z(x_0, y_0, z_0) \end{bmatrix}$$

Notice that the derivative contains all the information you need to compute all directional derivatives.

1.5. Example. Let $f(x, y) = x^2y^3$, and let $(x_0, y_0) = (2, 1)$. Then

$$f'(2, 1) = \begin{bmatrix} 4 & 12 \end{bmatrix}.$$

Observe that the tangent plane

$$z - 4 = 4(x - 2) + 12(y - 1)$$

may be written as

$$z - 4 = \begin{bmatrix} 4 & 12 \end{bmatrix} \begin{bmatrix} x - 2 \\ y - 1 \end{bmatrix}.$$

1.6. Tangent plane. As in the example above, for an arbitrary differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$, the tangent plane to the graph at (x_0, y_0, z_0) may be written

$$z - z_0 = f'(x_0, y_0) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \tag{1.7}$$

If we write

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } \mathbf{x}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

then Equation 1.7 may be rewritten as

$$z - z_0 = f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0). \tag{1.8}$$

(By $f'(\mathbf{x}_0)$ we mean $f'((x_0, y_0))$.)

Note the resemblance between the expression 1.8 for the tangent plane to the graph of a function $f : \mathbf{R}^2 \to \mathbf{R}$ and the familiar expression for the tangent line to the graph of a real-valued function of one variable.

2. Functions for which the domain and range are higher-dimensional.

Consider a function $f : \mathbf{R}^2 \to \mathbf{R}^2$, for example $f(x, y) = (x^2y, 2x + 5y^2)$. Such a function is made up of two real-valued functions, called the *component functions* of f. In this example, the component functions are $f_1(x, y) = x^2y$ and $f_2(x, y) = 2x + 5y^2$. If we write elements of $\mathbf{R}^2$ as column matrices, then we can also write the function f as

$$f(x, y) = \begin{bmatrix} x^2y \\ 2x + 5y^2 \end{bmatrix}$$

To define the derivative of this function, we first note that the two component functions $f_1(x, y) = x^2y$ and $f_2(x, y) = 2x + 5y^2$ are differentiable and we know how to write down their derivatives as row matrices:

$$f_1'(x, y) = \begin{bmatrix} 2xy & x^2 \end{bmatrix}$$
$$f_2'(x, y) = \begin{bmatrix} 2 & 10y \end{bmatrix}$$

E.g., at the point $(1, 1)$,

$$f_1'(1, 1) = \begin{bmatrix} 2 & 1 \end{bmatrix}$$
$$f_2'(1, 1) = \begin{bmatrix} 2 & 10 \end{bmatrix}$$

To get the derivative of f, we put together the two row matrices given by the derivatives of the component functions:

$$f'(1, 1) = \begin{bmatrix} 2 & 1 \\ 2 & 10 \end{bmatrix}$$

More generally:

Definition 2.1. We say a function $f : \mathbf{R}^n \to \mathbf{R}^m$ is *differentiable* if each of the component functions is differentiable. In that case, the derivative at a point $(a_1, \dots, a_n)$ in $\mathbf{R}^n$ is given by the $m \times n$ matrix whose ith row is the derivative of the ith component function.

2.1. Example. Define a function $f : \mathbf{R}^2 \to \mathbf{R}^3$ by $f(x, y) = (x^2y, 3xy, 5x + 4y))$, or in matrix form,

$$f(x, y) = \begin{bmatrix} x^2y \\ 3xy \\ 5x + 4y \end{bmatrix}.$$

Let $(x_0, y_0) = (1, 1)$. The component functions are $f_1(x, y) = x^2y$, $f_2(x, y) = 3xy$, and $f_3(x, y) = 5x + 4y$. We have $f_1'(x, y) = \begin{bmatrix} 2xy & x^2 \end{bmatrix}$ so

$$f_1'(1, 1) = \begin{bmatrix} 2 & 1 \end{bmatrix}.$$

Similarly

$$f_2'(1, 1) = \begin{bmatrix} 3 & 3 \end{bmatrix}$$

and

$$f_3'(1, 1) = \begin{bmatrix} 5 & 4 \end{bmatrix}$$

Thus

$$f'(1, 1) = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 5 & 4 \end{bmatrix}.$$

2.2. Exercise. Find the derivatives (as matrices) of each of the following functions at the indicated point:

(1) $f(x, y, z) = (e^{3x-y-z}, x^2y^2)$ at the point $(1, 1, 2)$
(2) $f(x, y) = ((x + y)^3, xsin(y))$ at $(2, 0)$.
(3) $f(t) = (t^2, t^3, t^4)$ at $t_0 = 1$. (If you view f as a parametrized curve, you have the familiar notion of the tangent vector at the point with parameter $t_0 = 1$. How does this compare to the derivative that you just computed?)

In the examples above, we wrote down the derivative (as a matrix) by first computing the partials. Conversely, if you are given the derivative matrix, then you can read off the partial derivatives of all the component functions.

2.3. Example. Suppose $f : \mathbf{R}^3 \to \mathbf{R}^2$ is differentiable at $\mathbf{x}_0 = (x_0, y_0, z_0)$ and that its derivative at $\mathbf{x}_0$ is

$$f'(x_0, y_0, z_0) = \begin{bmatrix} 2 & 5 & 1 \\ 3 & 1 & 4 \end{bmatrix}$$

Writing $(u, v) = f(x, y, z) = (f_1(x, y, z), f_2(x, y, z)$ or in matrix form

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_1(x, y, z) \\ f_2(x, y, z) \end{bmatrix}$$

then at the point $\mathbf{x}_0$, we can read off from the derivative matrix that

$$\frac{\partial}{\partial x} f_1(\mathbf{x}_0) = 2.$$

(This is also written as $\frac{\partial u}{\partial x} = 2$.) Similarly at $\mathbf{x}_0$, we have

$$\frac{\partial u}{\partial y} = \frac{\partial}{\partial y} f_1 = 5$$

$$\frac{\partial u}{\partial z} = \frac{\partial}{\partial z} f_1 = 1$$

and

$$\frac{\partial v}{\partial x} = \frac{\partial}{\partial x} f_2 = 3$$

etc.

3. The Chain Rule

Recall the chain rule for real-valued functions of one variable:

1. Familiar chain rule. *If $g : \mathbf{R} \to \mathbf{R}$ is differentiable at x_0 and $f : \mathbf{R} \to \mathbf{R}$ is differentiable at $y_0 = g(x_0)$, then $f \circ g$ is differentiable at x_0 and*

$$(f \circ g)'(x_0) = f'(y_0)g'(x_0).$$

Using matrices, the chain rule in higher dimensions looks identical to the familiar chain rule. As before, we use bold face letters like $\mathbf{x}$ to denote points (or vectors) in $\mathbf{R}^n$.

1. Chain rule in higher dimensions. *If $g : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $\mathbf{x}_0$ and $f : \mathbf{R}^m \to \mathbf{R}^p$ is differentiable at $\mathbf{y}_0 = g(\mathbf{x}_0)$, then $f \circ g$ is differentiable at $\mathbf{x}_0$ and*

$$(f \circ g)'(\mathbf{x}_0) = f'(\mathbf{y}_0)g'(\mathbf{x}_0).$$

3.1. Example. Define $g : \mathbf{R}^2 \to \mathbf{R}^3$ by

$$g(s,t) = (s^2t, s + 2t^2, st)$$

or in column form

$$g(s,t) = \begin{bmatrix} s^2t \\ s + 2t^2 \\ st \end{bmatrix}$$

and define $f : \mathbf{R}^3 \to \mathbf{R}$ by

$$f(x,y,z) = e^{2x-y+z}.$$

Let's compute the derivative of $f \circ g$ at $(s_0, t_0) = (1,1)$.

Note that $g(1,1) = (1,3,1)$. The chain rule says that

$$(f \circ g)'(1,1) = f'(1,3,1)g'(1,1).$$

We compute:

$$g'(s,t) = \begin{bmatrix} 2st & s^2 \\ 1 & 4t \\ t & s \end{bmatrix}$$

so

$$g'(1,1) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix}$$

Similarly

$$f'(1,3,1) = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}$$

Thus

$$(f \circ g)'(1,1) = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \end{bmatrix}$$

Writing $w = f \circ g(s,t)$, we can read off from this derivative matrix that at $(1,1)$, $\frac{\partial w}{\partial s} = 4$ and $\frac{\partial w}{\partial t} = -1$.

Let's compare this computation with the version of the chain rule given in Section 14.5 of Stewart. Writing $w = f(x,y,z)$ and $(x,y,z) = g(s,t)$, the chain rule in Stewart says that

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial w}{\partial z}\frac{\partial x}{\partial s}.$$

At $(s,t) = (1,1)$ and $(x,y,z) = (1,3,1)$, we get

$$\frac{\partial w}{\partial s} = 2(2) + (-1)(1) + (1)(1) = 4. \tag{3.2}$$

This agrees with what we obtained using the matrices. Notice that the computation in Equation 3.2 is equivalent to multiplying the first (and only) row of the matrix $f'(1,3,1)$ by the first column of the matrix $g'(1,1)$. Similarly, the formula in Stewart for $\frac{\partial w}{\partial t}$ corresponds to multiplying the row of $f'(1,3,1)$ by the second column of $g'(1,1)$. The matrix multiplication gives us both partials at once. (If you are only interested in one of the partials, then the formula in Stewart is a bit faster. If you want all the partials, the matrix method is more convenient.)

3.3. EXAMPLE. Let g be as in the previous example, and let $f : \mathbf{R}^3 \to \mathbf{R}^2$ be given by
$$f(x, y, z) = (e^{2x-y+z}, xyz).$$
Then
$$f'(1, 3, 1) = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 1 & 3 \end{bmatrix}$$
so
$$(f \circ g)'(1, 1) = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 10 & 10 \end{bmatrix}$$
Writing $(u, v) = f(x, y, z) = (f \circ g)(s, t)$, we can read off all the partials at $(s, t) = (1, 1)$:
$$\frac{\partial u}{\partial s} = 4$$
$$\frac{\partial u}{\partial t} = -1$$
$$\frac{\partial v}{\partial s} = 10$$
$$\frac{\partial v}{\partial t} = 10$$
3.4. EXERCISE. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be given by $f(x, y) = (e^x y^2, (x+1)y^3)$ and let $g : \mathbf{R} \to \mathbf{R}^2$ be given by $g(t) = (sin(t), e^t)$. First use the matrix version of the Chain Rule to find $(f \circ g)'(0)$. Then compute $(f \circ g)'(0)$ by the method in Section 14.5 and compare your answers.

3.5. EXERCISE. This exercise refers to problem 25 in Section 14.5 of Stewart.

(1) Express the given functions as $N = f(p, q, r)$ and $(p, q, r) = g(u, v, w)$.
(2) Write down the derivative matrices for f and g at the indicated point (i.e., $(u, v, w) = (2, 3, 4)$ and $(p, q, r) = g(2, 3, 4)$, which you can compute).
(3) Use the matrix version of the Chain Rule to compute the derivative $(f \circ g)'(2, 3, 4)$.
(4) Read off from your matrix in (3) the partial derivatives that are requested in problem 25.