DERIVATIVES AS MATRICES; CHAIN RULE

1. Derivatives of Real-valued Functions

Let's first consider functions $f : \mathbf{R}^2 \to \mathbf{R}$. Recall that if the partial derivatives of f exist at the point (x_0, y_0), then we can write down the candidate for the tangent plane:

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \tag{1.1}$$

We defined f to be differentiable if this plane really is a tangent plane to the graph of f in the sense that the tangent line to every smooth curve in the graph through the point (x_0, y_0, z_0) lies in this plane. As mentioned in class, one can show that if the partial derivatives are continuous, then the function is differentiable.

However, while we talked about differentiable functions, we never defined the actual *derivative* of a function from $\mathbf{R}^2$ to $\mathbf{R}$.

Definition 1.1. If f 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 at a vector, then it is the gradient $\nabla f(x_0, y_0)$.) Similarly, for a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, the derivative is the row matrix with n entries consisting of the n partial derivatives. E.g., for a function of three variables, we have

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

1.2. Example. Let $f(x, y) = x^2 y^3$, and let $(x_0, y_0) = (2, 1)$. Then $\frac{\partial f}{\partial x} = 2xy^3$ and $\frac{\partial f}{\partial y} = 3x^2 y^2$, so $\frac{\partial f}{\partial x}(2, 1) = 4$ and $\frac{\partial f}{\partial y}(2, 1) = 12$. Thus

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

The matrix $f'(2, 1)$ defines a linear transformation:

$$L(x, y) = \begin{bmatrix} 4 & 12 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4x + 12y.$$

Observe that the tangent line

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

may be written as

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

As in this example, for an arbitrary differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$, the tangent plane (Equation 1.1) to the graph at (x_0, y_0, z_0) may be written

$$z - z_0 = f'(x_0, y_0) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \tag{1.3}$$

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 1.3 may be rewritten as

$$z - z_0 = f'(x_0, y_0)(\mathbf{x} - \mathbf{x}_0). \tag{1.4}$$

Note the resemblance between this expression for the tangent plane to the graph of a function $f : \mathbf{R}^2 \to \mathbf{R}^2$ 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 both higher-dimensional.

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ consists of m-real-valued functions on $\mathbf{R}^n$:

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

The real-valued functions $f_1, \ldots, f_m$ are called the *component functions*.

2.1. Example. Define a function $f : \mathbf{R}^2 \to \mathbf{R}^3$ by $f(x, y) = (x^2y, 3xy, 5x + 4y))$. Then the component functions are $f_1(x, y) = x^2y$, $f_2(x, y) = 3xy$, and $f_3(x, y) = 5x + 4y$.

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 $(a_1, \ldots, a_n)$ in $\mathbf{R}^n$ is given by the $m \times n$ matrix whose ith row is the derivative of the ith component function.

2.2. Example. Define f as in Example 2.1. Let $(a_1, a_2) = (1, 1)$. Then $f_1'(x, y) = \begin{bmatrix} 2xy & x^2 \end{bmatrix}$ 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

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

2.3. 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?)

Again consider an arbitrary differentiable function $f : \mathbf{R}^n \to \mathbf{R}^m$. Denote elements of $\mathbf{R}^n$ by column vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

and elements of $\mathbf{R}^m$ by columns

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$$

Let

$$\mathbf{x}_0 = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

be a chosen point in $\mathbf{R}^n$ and set

$$\mathbf{u}_0 = f(a_1, \dots, a_n)$$

(viewed as a column matrix). Then the analog of the tangent plane, or best *linear approximation* to the graph at $(a_1, \dots, a_n)$, is given by:

$$\mathbf{u} - \mathbf{u}_0 = f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \tag{2.4}$$

2.5. Example. In Example 2.1, $f(1,1) = (1,3,9)$, and the best linear approximation at $(1,1)$ is given as follows:

$$\mathbf{u} - \mathbf{u}_0 = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 5 & 4 \end{bmatrix} (\mathbf{x} - \mathbf{x}_0)$$

I.e.,

$$\begin{bmatrix} u_1 - 1 \\ u_2 - 3 \\ u_3 - 9 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 1 \end{bmatrix} = \begin{bmatrix} 2(x-1) + 1(y-1) \\ 3(x-1) + 3(y-1) \\ 5(x-1) + 4(y-1) \end{bmatrix} \tag{2.6}$$

where the first equality is the expression for the tangent plane, and the second equality is obtained by multiplying matrices. If we compare the matrices on the left and right sides, the first row tells us that

$$u - 1 = 2(x-1) + 1(y-1).$$

Note that this is the equation of the tangent plane to the function $f_1(x,y) = x^2 y$ at $(1,2)$. Similarly, the second and third rows give the tangent planes to the second and third component functions of f.

2.7. Exercise. Write down the linear approximation to the graph of each of the functions in Exercise 2.3 at the indicated point. (Your answers should be expressed similarly to Equation 2.6.)

2.8. Remark. The most common mistake students make in writing down a derivative matrix is to switch the rows and columns. Note that the rows of the derivative matrix are the derivatives (gradients) of the component functions. The columns correspond to the independent variable; e.g., in the first column we see the partials of the component functions with respect to the first variable. A good way to keep this straight is to view the derivative matrix as the matrix of a linear transformation L (the one that gives us the linear approximation). Since the original function f has domain $\mathbf{R}^n$, we want L to have domain $\mathbf{R}^n$ and range $\mathbf{R}^m$ as well. That means we need to be able to multiply the derivative matrix by a column with n entries. So the derivative matrix must have n columns. Similarly, for the range to be $\mathbf{R}^m$, it must have m rows.

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.

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

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

Writing $(u, v) = f(x, y, z) = (f_1(x, y, z), f_2(x, y, z)$, then at the point $\mathbf{x}_0$, we can read off from the matrix that

$$\frac{\partial}{\partial x} f_1(x_0, y_0 z_0) = 2.$$

(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

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

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 15.5 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 15.5 and compare your answers.

3.5. Exercise. This exercise refers to problem 26 in Section 15.5. (same in old edition).

(1) Express the given functions as $Y = f(u,v,w)$ and $(u,v,w) = g(r,s,t)$.
(2) Write down the derivative matrices for f and g at the indicated point (i.e., $(r,s,t) = (1,0,1)$ and $(u,v,w) = g(1,0,1)$, which you can compute).
(3) Use the matrix version of the Chain Rule to compute the derivative $(f \circ g)'(1,0,1)$.
(4) Read off from your matrix in (3) the partial derivatives that are requested in problem 26.