

# Notes on Multivariable Differentiation

## MENU, Winter 2013

These notes summarize the main properties and uses of multivariable derivatives. Most of this is in the book, but some it gets lost in the various notations the book uses, especially when dealing with arbitrary functions  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ . The geometric interpretation of second partial derivatives, however, is not really mentioned in a book. (This is a crying shame, since the geometric origins of calculus are key to understanding what it is you're actually doing.)

Note that these notes don't say anything about Chapter 1 material which will be on the midterm, such as general stuff about planes, distances, and polar/cylindrical/spherical coordinates. Be sure to review this stuff as well. Here are the problems from the Midterm 2 practice problems which I think are the most important to look at: 1abcdeij, 2abcdehijkl, 3, 4, 5, 6, 7, 8, 9, 12.

## Functions

The basic object of study is a function of several variables  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ : such a function would take  $n$  inputs and give  $m$  outputs. Mainly, we'll be interested in functions  $\mathbb{R}^2 \rightarrow \mathbb{R}$  and  $\mathbb{R}^3 \rightarrow \mathbb{R}$ . For functions  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , the *graph* of  $f$  is the set of points  $(x, y, z)$  where the  $z$  coordinate equals  $f(x, y)$ . Geometrically, the graph is the surface given by the equation  $z = f(x, y)$ .

Note that in order for  $f$  to actually be a function (i.e. any input only gives one output), the graph must pass the so-called "vertical line test": the vertical line passing through the point  $(x, y, 0)$  in the  $xy$ -plane can only intersect the graph of  $f$  in one location, since otherwise there would be two different outputs for the single input  $(x, y)$ .

**Fact.** *The level curve of the function  $f(x, y)$  at  $z = k$  is the set of points  $(x, y)$  in the  $xy$ -plane satisfying the equation  $f(x, y) = k$ . Geometrically, this is the intersection of the graph of  $z = f(x, y)$  with the horizontal plane at  $z = k$ .*

*The section of  $f(x, y)$  by  $x = c$  is the set of points in  $\mathbb{R}^3$  where  $x = c$  and  $y$  and  $z$  satisfy  $f(c, y) = z$ . Geometrically, this is the intersection of the graph of  $z = f(x, y)$  with the vertical plane at  $x = c$ . Similarly, the section by  $y = c$  is the set of points satisfying  $y = c$  and  $f(x, c) = z$ , which geometrically is the intersection of the graph with the vertical plane at  $y = c$ .*

The same sorts of definitions make sense for other surfaces (for example, quadric surfaces) as well. The idea of a level curve also makes sense for functions of three variables  $g(x, y, z)$ , only we get level *surfaces*  $g(x, y, z) = k$  instead of level curves. The level curves and sections described above help to visualize the graphs of the corresponding functions and help to analyze their behavior. Given a drawing of a bunch of level curves with their corresponding  $z = c$  values, we imagine the graph as being traced out by these curves at varying heights  $z = c$ .

Just as with single-variable functions, we can talk about limits of multivariable functions. The new idea is that there are many possible directions from which to approach a given point as opposed to the case of a single-variable function:

**Fact.** *If  $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$  exists, then approaching  $(a, b)$  along any possible curve passing through  $(a, b)$  should give the same value as this limit. So, if there is a curve from which we can approach  $(a, b)$  where the limit does not exist, or if there are two curves from which we can approach  $(a, b)$  which give different values for the limit, then  $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$  does not exist.*

This fact, together with changing coordinates, are the main tools we have for showing that limits do not exist. As for showing that limits do exist, again either we can change coordinates or use the following:

**Fact.** If  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is continuous at  $(a, b)$ , then  $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$ .

Again, this only applies when our function is continuous at the point we're approaching, which is usually the case when our function is made up by multiplying, adding, and composing continuous things. The only thing to watch out for is when you have an expression with a denominator which is 0 at the point you're approaching.

## Derivatives

The nice thing about dealing with multivariable derivatives is that partial derivatives are usually pretty simple to compute:

**Fact.** Given a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , to compute  $\frac{\partial f}{\partial x} = f_x$  we think of  $y$  as a constant and differentiate with respect to  $x$  as we normally would. Similarly, to compute  $\frac{\partial f}{\partial y} = f_y$  we think of  $x$  as a constant and differentiate with respect to  $y$  as you normally would. The same is true for functions  $\mathbb{R}^n \rightarrow \mathbb{R}$  in general.

Geometrically,  $f_x(x, y)$  is the “slope in the  $x$ -direction” at the point  $(x, y)$ : if you stand on the point of the graph of  $f$  corresponding to the point  $(x, y)$  and face in the positive  $x$ -direction  $\mathbf{i}$ ,  $f_x(x, y)$  is the slope of the piece of the graph you are facing. Similarly,  $f_y(x, y)$  is the “slope in the  $y$ -direction”, or better yet the slope in the  $\mathbf{j}$  direction at  $(x, y)$ .

Given these partial derivatives, we can construct the candidate for the tangent plane to the graph of  $f$  at a point  $(a, b)$ :

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Another way to remember this equation is by simply remembering that the normal vector to this plane is  $(f_x(a, b), f_y(a, b), -1)$  and that the plane passes through the point  $(a, b, f(a, b))$ :

**Fact.** At a point  $(a, b)$ , the normal vector of the tangent plane at  $(a, b)$  is  $(f_x(a, b), f_y(a, b), -1)$ , and thus the tangent plane is given by the equation

$$(f_x(a, b), f_y(a, b), -1) \cdot (x - a, y - b, z - f(a, b)) = 0.$$

If you work out this dot product you get precisely the equation of the tangent plane above. The only catch now is whether or not this “candidate” for the tangent plane is actually the tangent plane:

**Fact.**  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is differentiable at  $(a, b)$  if

$$\lim_{(x,y) \rightarrow (a,b)} \frac{f(x, y) - (\text{value for } z \text{ we get from candidate tangent plane})}{\sqrt{(x - a)^2 + (y - b)^2}} = 0$$

Geometrically this is saying that the tangent plane is actually correct, in that it provides a good linear approximation to the function in the sense that the numerator above goes to 0 faster than the distance between  $(x, y)$  and  $(a, b)$  in the denominator.

The gradient  $\nabla f$  of  $f$  encodes the information needed to form this tangent plane, and in general the matrix of partial derivatives  $Df$  of  $f$  encodes the same for higher-dimensional analogues of the tangent plane.

Higher-order derivatives are just as simple to compute; for instance

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right)$$

means we differentiate  $f_x$  with respect to  $x$ , while

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right)$$

means we differentiate  $f_x$  with respect to  $y$ .

**Fact.** Geometrically,  $f_{xx}$  measure the “concavity” of the graph of  $f$  in the  $x$ -direction while  $f_{yy}$  measures the “concavity” in the  $y$ -direction. The mixed partial  $f_{yx}$  measures the rate at which the  $x$ -directional slopes are changing as you move in the  $y$  direction and  $f_{xy}$  measures the rate at which the  $y$ -directional slopes are changing as you move in the  $x$ -direction. For “nice” functions (i.e. ones whose second partial derivatives are continuous), these mixed partials are the same.

Finally, we come to the chain rule, which you should view as analogous to the single-variable chain rule: if  $y = g(x)$  and  $z = f(y)$ , then the derivative of  $(f \circ g)(x) = f(g(x))$  is

$$\frac{dz}{dx} = f'(y)g'(x) = f'(g(x))g'(x).$$

The only difference now is that we add together terms similar to these, one for each “intermediate” variable. All versions of the chain rule can be summarized using the version expressed in terms of matrix multiplication:

**Fact.** If  $g : \mathbb{R}^k \rightarrow \mathbb{R}^n$  and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are differentiable functions, the matrix of partial derivatives of the composition  $f \circ g$  at a point  $\mathbf{x}$  in  $\mathbb{R}^k$  is the product

$$D(f \circ g)(\mathbf{x}) = Df(\mathbf{y})Dg(\mathbf{x})$$

of the matrices of partial derivatives of  $f$  and  $g$  respectively, where  $\mathbf{y}$  is the point  $\mathbf{y} = g(\mathbf{x})$ . Note that the order in which you multiply these matrices is important.

As special cases of this, if  $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a function  $g(s, t) = (x(s, t), y(s, t))$  and  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a function  $z = f(x, y)$ , then the matrix of partial derivatives of  $f \circ g$  (i.e.  $z$  expressed in terms of  $s$  and  $t$ ) is

$$(Df)(Dg) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s} & \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} \end{pmatrix},$$

meaning that

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t},$$

as the chain rule says should happen.

As another special case, consider functions  $y = g(x)$  and  $z = f(y)$ , each of a single variable. The matrix of partial derivatives of each of these are  $1 \times 1$  matrices:

$$Dg = (g'(x)) \quad \text{and} \quad Df = (f'(y)),$$

so the derivative of their composition  $(f \circ g)(x) = f(g(x))$  is the  $1 \times 1$  matrix

$$(Df)(Dg) = (f'(y)) (g'(x)) = (f'(y)g'(x)) = (f'(g(x))g'(x)),$$

which is precisely the single-variable chain rule.

## Gradients

Recall that the gradient of function  $z = f(x, y)$  gives an easy way to compute directional derivatives:

**Fact.** *The directional derivative of  $f$  at the point  $(x, y)$  in the direction of the unit vector  $\mathbf{u}$  is given by*

$$D_{\mathbf{u}}f(x, y) = \nabla f(x, y) \cdot \mathbf{u} = \|\nabla f(x, y)\| \cos \theta$$

*where  $\theta$  is the angle between  $\nabla f(x, y)$  and  $\mathbf{u}$ . Geometrically, this gives the rate of change (or slope) of  $f$  when standing at the point on the graph of  $f$  corresponding to  $(x, y)$  and facing in the direction of  $\mathbf{u}$ . In particular,*

$$D_{\mathbf{i}}f(x, y) = f_x(x, y) \text{ and } D_{\mathbf{j}}f(x, y) = f_y(x, y).$$

The formula for directional derivatives given above expressed in terms of  $\cos \theta$  leads to the geometric interpretations of the gradient itself:

**Fact.** *At any point  $(x, y)$ ,  $\nabla f(x, y)$  itself points in the direction in which  $f$  is increasing most rapidly (i.e. the direction of maximum rate of change). That maximum rate of increase itself is equal to  $\|\nabla f(x, y)\|$ .*

*The direction in which  $f$  decreases most rapidly is given by  $-\nabla f(x, y)$ , and the directions in which  $f$  does not change are given by those perpendicular to  $\nabla f(x, y)$ .*

Apart from the geometric interpretations of the gradient related to directional derivatives, also keep in mind the following:

**Fact.** *At any point  $(x, y)$ ,  $\nabla f(x, y)$  is perpendicular to the level curve of  $f$  passing through the point  $(x, y)$ . Similarly, for a function  $g(x, y, z)$  of three variables  $\nabla g(x, y, z)$  is perpendicular to the level surface of  $g$  passing through  $(x, y, z)$ .*

This gives us a way to find tangent planes to surfaces which are not given as the graph of a function of two variables. Note that this works for surfaces which *are* given as the graph of a function as well: a graph  $z = f(x, y)$  can be viewed as the level surface at 0 of the function

$$g(x, y, z) = f(x, y) - z,$$

in which case a normal vector to the tangent plane at  $(x, y, z)$  is given by

$$\nabla g(x, y, z) = (f_x(x, y), f_y(x, y), -1),$$

which is the same normal vector we get finding equations of tangent planes the old way.