Math 208H

Topics for the first exam

Our first exam will be **Thursday, February 16, in Avery 118**. The room has been reserved for **4:00pm to 10:00pm**; you should choose the 60 to 100 minute time interval that best suits your schedule, to take the exam. The exam will be designed so that (to the best of the instructor's knowledge) it can reasonably be finished in 50 to 60 minutes; you should perhaps plan for some extra time so that you can be confident that you have answered every question as best you can.

Vectors

In one-variable calculus, we make a distinction between speed and velocity; velocity has a direction (left or right), while speed doesn't. Speed is the *size* of the velocity. This distinction is even more important in higher dimensions, and leads to the notion of a *vector*.

Basically, a vector $\vec{v}$ is an arrow pointing *from* one point in the plane (or 3-space or ...) *to* another. A vector is thought of as pointing frm its *tail* to its *head*. If it points from P to Q, we call the vector $\vec{v} = \overrightarrow{PQ}$.

A vector has both a *size* (= length = distance from P to Q) and a *direction*. Vectors that have the same size and point in the same direction are often thought of as the same, even if they have different tails (and heads). Put differently, by picking up the vector and translating it so that its tail is at the origin (0,0), we can identify $\vec{v}$ with a point in the plane, namely its head (x, y), and write $\vec{v} = \langle x, y\rangle$. If $\vec{v}$ goes from (a, b) to (c, d), then we would have $\vec{v}= \langle c - a, d - b\rangle$. The length of $\vec{v} = \langle a, b\rangle$ is then $||\vec{v}|| = \sqrt{a^2 + b^2}$.

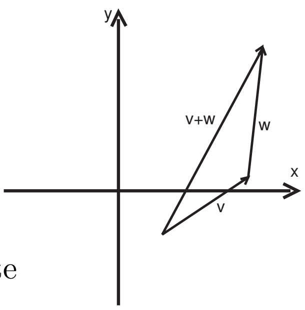
In 3-space we have three special vectors, pointing in the direction of each coordinate axis (in the plane there are, analogously, two); these are called

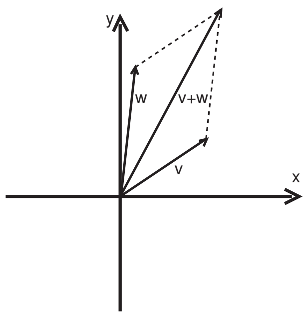
$$\vec{i} = \langle 1, 0, 0\rangle,\ \vec{j} = \langle 0, 1, 0\rangle, \text{ and } \vec{k} = \langle 0, 0, 1\rangle$$

These come in especially handy when we start to add vectors. There are several different points of view to vector addition:

(1) move the vector $\vec{w}$ so that its head is on the tail of $\vec{v}$; then the vector $\vec{v} + \vec{w}$ has tail equal to the tail of $\vec{v}$ and head equal to the head of $\vec{w}$;

(2) move $\vec{v}$ and $\vec{w}$ so that their tails are both at the origin, and build the parallelogram which has sides equal to $\vec{v}$ and $\vec{w}$; then $\vec{v} + \vec{w}$ is the vector that goes from the origin to the opposite corner of the parallelogram;





(3) if $\vec{v} = \langle a, b\rangle$ and $\vec{w} = \langle c, d\rangle$, then $\vec{v} + \vec{w} = \langle a + c, b + d\rangle$

We can also subtract vectors; if they share the same tail, $\vec{v} - \vec{w}$ is the vector that points from the head of $\vec{w}$ to the head of $\vec{v}$ (so that $\vec{w} + (\vec{v} - \vec{w}) = \vec{v}$). In coordinates, we simply subtract the coordinates.

We can also *rescale* vectors = multiply them by a constant factor; $a\vec{v}$ = vector pointing in the same direction, but a times as long. (We use the convention that if $a < 0$, then $a\vec{v}$ points in the *opposite* direction from $\vec{v}$.)

Using coordinates, this means that $a\langle x, y\rangle = \langle ax, ay\rangle$. To distinguish a from the coordinates or the vector, we call a a *scalar*. One consequence of this formula is that $||c\vec{v}|| = |c| \cdot ||\vec{v}||$.

All of these operations satisfy all of the usual properties you would expect:

$\vec{v} + \vec{w} = \vec{w} + \vec{v}$ $\qquad$ $(\vec{v} + \vec{w}) + \vec{u} = \vec{v} + (\vec{w} + \vec{u})$

$a(b\vec{v}) = (ab)\vec{v}$ $a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w}$

If all that we are interested in about a vector is its *direction*, then we can choose a vector of length one pointing in the same direction:

$\vec{u} = \dfrac{\vec{v}}{||\vec{v}||}$ = unit vector pointing in the same diection as $\vec{v}$.

Of course there is nothing special in all of this about vectors in the plane; all of these ideas work for vectors in 3-space. The only thing we really need to determine is the right formula for *length*: a few applications of the Pythagorean theorem leads us to

$||\langle a, b, c\rangle|| = (a^2 + b^2 + c^2)^{1/2}$

Dot products

One thing we haven't done yet is multiply vectors together. It turns out that there are two ways to reasonably do this, serving two very different sorts of purposes.

The first, the dot product, is intended to measure the extent to which two vectors $\vec{v}$ and $\vec{w}$ are pointing in the same direction. It takes a pair of vectors $\vec{v} = \langle v_1, \ldots, v_n\rangle$ and $\vec{w} = \langle w_1, \ldots, w_n\rangle$, and gives us a *scalar* $\vec{v} \bullet \vec{w} = v_1w_1 + \cdots + v_nw_n$.

Note that $\vec{v} \bullet \vec{v} = v_1^2 + \cdots + v_n^2 = ||\vec{v}||^2$. In general, $\vec{v} \bullet \vec{w} = ||\vec{v}|| \cdot ||\vec{w}|| \cdot \cos(\theta)$, where θ is the angle between the vectors $\vec{v}$ and $\vec{w}$ (when they have the same tail); this can be seen by comparing the Law of Cosines to the formula

$||\vec{v} - \vec{w}||^2 = ||\vec{v}||^2 + ||\vec{w}||^2 - 2\vec{v} \bullet \vec{w}$

This in turn allows us to compute this angle:

The *angle* Θ between v and w = the angle (between 0 and π with $\cos(\Theta) = \langle v, w\rangle/(||v|| \cdot ||w||)$

The dot product satisfies some properties which justify calling it a product:

$\vec{v} \bullet \vec{w} = \vec{w} \bullet \vec{v}$

$(k\vec{v}) \bullet \vec{w} = k(\vec{v} \bullet \vec{w})$

$\vec{v} \bullet (\vec{w} + \vec{u}) = \vec{v} \bullet \vec{w} + \vec{v} \bullet \vec{u}$

Two vectors are orthogonal (= perpendicular) if the angle θ between them is $\pi/2$, so $\cos(\theta)$=0; this means that $\vec{v} \bullet \vec{w} = 0$. We write $\vec{v} \perp \vec{w}$.

Since $|\cos\theta| \leq 1$, we always have $|\vec{v} \bullet \vec{w}| \leq ||\vec{v}||||\vec{w}||$. This is the *Cauchy-Schwartz inequality.* From this we can also deduce the *Triangle inequality* : $||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||$.

Projecting one vector onto another:

The idea is to figure out how much of one vector $\vec{v}$ points *in* the direction of another vector $\vec{w}$. The dot product measures to what extent they are pointing in the same direction, so it is only natural that it plays a role.

What we wish to do is to write $\vec{v} = c\vec{w} + \vec{u}$, where $\vec{u} \perp \vec{w}$ (i.e., write $\vec{v}$ as the part pointing in the direction of $\vec{w}$ and the part $\perp \vec{w}$). By solving the equation $(\vec{v} - c\vec{w}) \bullet \vec{w} = 0$, we find that $c = (\vec{v} \bullet \vec{w})/(\vec{w} \bullet \vec{w})$.

We write $c\vec{w} = \text{proj}_{\vec{w}}\vec{v} = \dfrac{\vec{v} \bullet \vec{w}}{\vec{w} \bullet \vec{w}}\vec{w} = \dfrac{\vec{v} \bullet \vec{w}}{||\vec{w}||}\dfrac{\vec{w}}{||\vec{w}||}$ = (orthogonal) projection of $\vec{v}$ onto $\vec{w}$.

$\vec{u} = \vec{v} - c\vec{w}$ = the part of $\vec{v}$ perpendicular to $\vec{w}$.

The cross product

The dot product takes two vectors and spits out a scalar. For vectors in 3-space, there is another product, which spits out another vector. The basica idea is that given two vectors in 3-space, there is a third vector which is perpendicular to both of them. Given the two vectors

$\vec{v} = \langle a_1, a_2, a_3\rangle$, $\vec{w} = \langle b_1, b_2, b_3\rangle$

we can solve the pair of equations $a_1x + a_2y + a_3z = 0$, $b_1x + b_2y + b_3z = 0$ to find that

$\langle a_2b_3 - a_3b_2, -(a_1b_3 - a_3b_1), a_1b_2 - a_2b_1\rangle$ is a solution.

We call this vector the *cross product* $\vec{v} \times \vec{w}$ for $\vec{v}$ and $\vec{w}$.

How do you remember this formula? Most people remember it using the notation

$$\vec{v} \times \vec{w} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \vec{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \vec{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \vec{k}$$

where $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ is the *determinant* of the 2×2 matrix

The cross product satisfies several useful equalities:

$\vec{v} \bullet (\vec{v} \times \vec{w}) = 0, \vec{w} \bullet (\vec{v} \times \vec{w}) = 0$ $\quad$ $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \bullet \vec{w})\vec{v} - (\vec{u} \bullet \vec{v})\vec{w}$

$(k\vec{v}) \times \vec{w} = k(\vec{v} \times \vec{w})$ $\quad$ $\vec{v} \times (\vec{w} + \vec{u}) = \vec{v} \times \vec{w} + \vec{v} \times \vec{u}$

$\vec{u} \bullet (\vec{v} \times \vec{w}) = \vec{v} \bullet (\vec{w} \times \vec{u}) = \vec{w} \bullet (\vec{u} \times \vec{v})$ (the *triple product*) $\quad$ $\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}$

For our standard vectors in 3-space we have $\quad \vec{i} \times \vec{j} = \vec{k}$, $\vec{j} \times \vec{k} = \vec{i}$, $\vec{k} \times \vec{i} = \vec{j}$

Our formula for the cross product was worked out just by solving a pair of equations; any other multiple of our vector would have been perpendicular to $\vec{v}$ and $\vec{w}$, too. But in a precise sense, the formula we came up with is the right one, because the length of our vector has geometric significance:

$||\vec{v} \times \vec{w}|| = ||\vec{v}||||\vec{w}|| \sin\theta$, where $\theta =$ angle between $\vec{v}$ and $\vec{w}$.

But! The area of a parallelogram with sides equal to the vectors $\vec{v}$ and $\vec{w}$ is

Area = (base)$\times$(height) = $||\vec{w}|| \cdot h = ||\vec{w}|| \cdot ||\vec{v}|| \cdot \sin(\theta)$ (from trigonometry).

So: $||\vec{v} \times \vec{w}|| = ||\vec{v}|| \cdot ||\vec{w}|| \cdot \sin(\theta) =$ the area of that parallelogram!

The cross product can be used to carry out many calculations which we will find useful. For example, to compute the distance d from a point P to the line through the points Q and R, we find that (setting $\vec{v} = \overrightarrow{QP}$, $\vec{w} = \overrightarrow{QR}$) using right triangles we have

$d = ||\vec{v}|| \sin\theta = (||\vec{v}||||\vec{w}|| \sin\theta)/||\vec{w}|| = ||\vec{v} \times \vec{w}||/||\vec{w}||$

Also, to compute the volume of a parallelopiped with sides $\vec{u}, \vec{v}, \vec{w}$, we can compute

volume = (area of base)$\cdot$(height) = $||\vec{u} \times \vec{v}|| \cdot (||\vec{w}||| \cos\psi|)$

where $\psi =$ angle between $\vec{u} \times \vec{v}$ and $\vec{w}$, so

volume = $|(\vec{u} \times \vec{v}) \bullet \vec{w}|$ = absolute value of the triple product!

Lines and planes in 3-space

Just as with lines in the plane, we can parametrize lines in space, given a point on the line, P, and the direction $\vec{v}$ that the line is travelling:

$L(t) = (x(t), y(t), z(t)) = P + \vec{v}t = (x_0 + at, y_0 + bt, z_0 + ct)$

This involves a (somewhat arbitrary) parameter t to describe; we can find a more *symmetric* description of the line by determining, for each coordinate, what t is and setting them all equal to one another:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

To determine if and where two lines in space intersect, if we use the parametrized forms, we need to remember that the two lines might pass through that same point at different times, and so we really need to use different names for the parameters:

$P + \vec{v}t = Q + \vec{w}s$

This gives us three equations (each of the three coordinates) with two variables; it therefore usually does not have a solutions. Two lines in 3-space that do not meet are called *skew*. If two lines do meet, then we can treat them much like in the plane; we can, for example, determine the angle at which they meet by computing the angle between their direction vectors $\vec{v}, \vec{w}$.

For planes, three points P, Q and R that do not lie on a single line will have exactly one plane through them. To describe that plane, we can think of it as all points X so that $\overrightarrow{PX}$ can be expressed as a combination of $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. This in turn means that $\overrightarrow{PX}$ is perpendicular to anything that is simultaneously perpendicular to both $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. But the cross product is such a vector; and so we can describe the plane by insisting that

$$\overrightarrow{PX} \bullet (\overrightarrow{PQ} \times \overrightarrow{PR}) = 0$$

If we write $\overrightarrow{PQ} \times \overrightarrow{PR} = \langle a, b, c\rangle$ and $\overrightarrow{PX} = \langle x - x_0, y - y_0, z - z_0\rangle$, then this equation becomes

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

What is really needed to describe this plane, in some sense, is the point $P = (x_0, y_0, z_0)$ and the vector $\vec(N) = \langle a, b, c\rangle =$ the *normal* vector to the plane. In other words, to completely describe a plane we can also use knowledge of a single point that the plane passes through, P, and what direction "up" is, namely the vector $\vec(N)$ perpendicular to the plane (i.e, the vector perpendicular to every vector lying in the plane). We can then write the equation for the plane as

$$\langle x, y, z\rangle \bullet \vec{N} = P \bullet \vec{N}$$

Note that if we are given the equation for the plane, we can quickly read off its normal vector; it is the coefficients of x, y, and z.

To find the plane through the three points P, Q, and R in 3-space, look at the vectors $\vec{v} = \overrightarrow{PQ}$ and $\vec{w} = \overrightarrow{PR}$. These are vectors between points in our plane, and so they give a pair of directions in the plane. They then must both be perpendicular to the normal vector $\vec{n}$ for the plane. But we know that they are both perpendicular to $\vec{v} \times \vec{w}$, and so $\vec{v} \times \vec{w}$ must be perpendicular to the plane. In other words, we can choose our normal vector to *be* $\vec{v} \times \vec{w}$. Using one of our original points (P, say) as a point in the plane, we can write down the equation for the plane using our dot product equation, above.

Intersecting planes: typically, two planes will intersect in a line (unless they are parallel, i.e., their normals are multiples of one another). We can find the parametric equation for the line by solving each equation of the plane for x, say, as an expression in y and z. Setting these two expressions equal, we can express y, say, as a function of z. Plugging back into our original expression for x, we get x as a function of z. So x, y, and z have all been expressed in terms of a single variable, z, which is exactly what a parametric equation does! The direction vector for this line, it is worth pointing out, is the cross product of the normals to the two planes; this direction vector points in a direction lying in both planes, and so much be perpendicular to both normals.

Functions of two variables

Function of one variable: one number in, one number out. Picture a black box; one input and one output.

Function of several variables: several inputs, one output. Picture a quantity which depends on several different quantities. E.g., distance from the origin in the plane:

$$\text{distance} = \text{d} = \sqrt{x^2 + y^2}$$

depends on both the x- and y-coordinates of our point.

Our goal is to understand functions of several variables, in much the same way that the tools of calculus allow us to understand functions of one variable. And our secret weapon is going to be to *think of a function of several variables as a function of* **one variable** *(at a time!)*, so that we can use those tools to good effect.

Graphs of functions of two variables

We know what such a graph *is*; but how do we see what it looks like? One answer is to think of it as a function of one variable (at a time!).

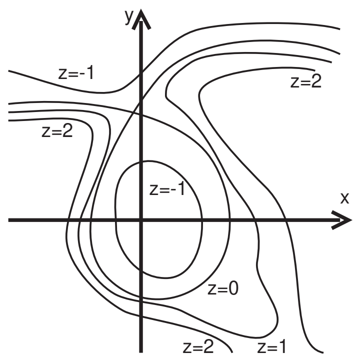
If we set $y = c$=constant, and look at $z = f(x, c)$, we are looking at a function of one variable, x, which we can (in theory) graph. This graph is what we would see when the *plane* $y = c$ meets the graph $z = f(x, y)$; this is a (vertical) *cross section* of our graph (parallel to the plane $y = 0$, the xz-plane). Similarly, if we set $x = d$=constant, and look at the graph of $z = f(d, y)$ (as a function of y), we are seeing vertical cross sections of our original graph, parallel to the yz-plane. Several of these x- and y-cross sections together can give a very good picture of the general shape of the graph of our function z=$f(x, y)$.
Some of the simplest functions to describe are linear functions; functions havng equations of the form z=$ax + by + c$. Their cross sections are all lines; the cross sections x =const all have the same slope b, and the y-cross sections all have slope a.
Another simple type of function is *cylinders*; these are functions like $f(x, y) = y^2$ which, although we think of them as functions of x and y, the output does not depend on one of the inputs. Cross sections of such functions, setting equal to a constant whichever variable does not change the value of the function, will all be identical, so the graph looks likecopies of the exact same function, stacked side-by-side.

Contour diagrams

The cross sections of the graph of a function are obtained by slicing our graph with vertical planes (parallel to one of our coordinate planes). But we could also use *horizontal* planes as well, that is, the planes z=constant. In other words, we graph $f(x, y) = c$ for different values of the constant c. These are called *contour lines* or *level curves* for the function f, since they represent all of the points on the graph of f which lie on the same horizontal level (the term contour line is borrowed from topographic maps; the lines represent the level curves of the height of land). These have the advantage that they can be graphed *together* in the xy-plane, for different values of c, because the level curves corresponding to different values of c cannot meet (a point (x, y) on both level curves would satisfy $c = f(x, y) = d$, so f would not be a function....)
A collection of level curves also gives a good picture of what the graph of our function f looks like; we can imagine wandering through the domain of our function, reading off the value of the function f by looking at what level curve we are standing on.

We usually draw level curves for equally spaced values of c; that way, if the level curves are close together, we know that the function is changing values rapidly in that region, while if they are far apart, the values of the function are not varying by a large amount in that area.

We usually, for convenience, draw the level curves of f on a single xy-plane (since we can keep them somewhat separate), labelling each curve with its z-value. We could reconstruct a picture of the graph of f by drawing the level curve $f(x, y) = c$ on the horizontal plane $z = c$ in 3-space.



Functions of more than two variables

There is of course no reason to stop with two variables for a function. An expression like F =$F(M, m, r) = GMm/r^2$ can be thought of as a function describing F as a function of M, m, and r (and G!). When we think of the graph of this function (as a function of the first three variables), its graph will live in 4-space! However, we can still get an impression of what the function looks like, by graphing $F(M, m, r) = c =$ constant, for various values of c. These are *level surfaces* for the function f. We can get a picture of what the level surfaces look like by taking cross sections! Or we could look at each level surface's level *curves*.

Partial derivatives

In one-variable calculus, the derivative of a function $y = f(x)$ is defined as the limit of difference quotients:

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

and interpreted as an instantaneous rate of chage, or slope of tangent line.

But a function of two variables has *two variables*; which one do you increment by h to get your difference quotient? The answer is **both of them**, one at a time. In other words, a function of two variables $z = f(x, y)$ has *two* (partial) derivatives:

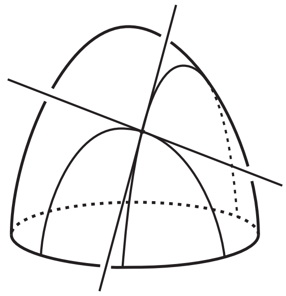
$$\frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(x+h, y) - f(x, y)}{h} \quad \text{and} \quad \frac{\partial f}{\partial y} = \lim_{h\to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

Essentially, $\frac{\partial f}{\partial x}$ is the derivative of f, thought of solely as a function of x (i.e., pretending that y is a constant), while $\frac{\partial f}{\partial y}$ is the derivative of f, thought of solely as a function of y.

Different viewpoint, same result:

For one variable calculus $f'(x)$ is the slope of the tangent line to the graph of f. As we shall see, The graph of a function of two variables has something we would naturally call a tangent *plane*, and one way to describe a plane is by computing its x- and y-slopes, i.e, the rate of change of f solely in the x- and y- directions. But this is precisely what the limits above calculate; so $\frac{\partial f}{\partial x}$ will be the x-slope of the tangent plane, and $\frac{\partial f}{\partial y}$ will be the y-slope.

The basic picture here is:



Just as with one variable, there are lots of different notations for describing the partial derivatives:
for $z = f(x, y)$,

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(f) = \frac{\partial z}{\partial x} = \frac{\partial}{\partial x}(z) = D_x(f) = D_x(z) = f_x = z_x$$

The algebra of partial derivatives

The basic idea is that since a partial derivative is 'really' the derivative of a function of one *variable* (the other 'variable' is *really* a constant), all of our usual differentiation rules can be applied. so, e.g,

$$\frac{\partial}{\partial x}(f+g) = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x} \qquad \frac{\partial}{\partial y}(f+g) = \frac{\partial f}{\partial y} + \frac{\partial g}{\partial y}$$

$$\frac{\partial}{\partial x}(c \cdot f) = c\frac{\partial f}{\partial x} \qquad \frac{\partial}{\partial y}(c \cdot f) = c\frac{\partial f}{\partial y}$$

$$\frac{\partial}{\partial x}(f \cdot g) = \frac{\partial f}{\partial x}g + f\frac{\partial g}{\partial x} \qquad \text{(etc.)}$$

$$\frac{\partial}{\partial x}(f/g) = (\frac{\partial f}{\partial x}g - f\frac{\partial g}{\partial x})/g^2 \qquad \text{(etc.)}$$

$$\frac{\partial}{\partial x}(h(f(x,y))) = h'(f(x,y)) \cdot \frac{\partial f}{\partial x} \qquad \text{(etc.)}$$

In the end the way we should get used to taking a partial derivative is exactly the same as for functions of one variable; just read from the outside in, applying each rule as it is appropriate. The *only* difference now is that when taking a derivative of a function $z = f(x, y)$, we need to remember that

$$\frac{\partial}{\partial x}(y) = 0 \qquad \text{and} \qquad \frac{\partial}{\partial y}(x) = 0$$

Linear approximations

In some sense, the culmination of one-variable calculus is the observation that any function can be approximated by a polynomial; and the polynomial of degree n that 'best' approximates f near the point a is the one which has the same (higher) derivatives as f at a, up to the nth derivative. This leads to the definition of the *Taylor polynomial* :

$p_n(x) = f(a) + f'(a)(x-a) + \cdots + \dfrac{f^{(n)}(a)}{n!}(x-a)^n$

Functions of two variables are not much different; we just replace the word 'derivative' with '*partial* derivative'! So for example, the best linear approximation is

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

This, we shall see, is the formula for the tangent plane to the graph of f at the point $(a, b, f(a, b))$.

Differentiability

In one-variable calculus, 'f is differentiable' is just another way of saying 'the derivative of f exists'. But with several variables, differentiablility means **more** than that all of the partial derivatives exist.

A function of several variables is *differentiable* at a point if the tangent plane to the graph of f at that point makes a good approximation to the function, near the point of tangency. In the words of the previous paragraph, $L-f$ shrinks to 0 *faster* than a linear function would. In other words, the 'best' linear approximation, above, is also a *good* linear approximation.

The basic fact, that we will keep using, is that if the partial derivatives of f don't just *exist* at a point, but are also **continuous** near the point, then f is differentiable in this more precise sense. (The proof of this fact is a little delicate...)

The Chain Rule

If f is a function of the variables x and y, and both x and y depend on a single variable t, then in a certain sense, f is a function of t; $f(x, y) = f(x(t), y(t))$; it is a *composition.* To find its derivative *with respect to* t, we can turn to our notion of a linear approximation: for $(a, b) = (x(t_0), y(t_0))$,

$f(x, y) \approx f(a, b) + f_x(a, b)(x-a) + f_y(a, b)(y-b)$, while

$x(t) - a \approx x'(t_0)(t-t_0)$ and $y(t) - b \approx y'(t_0)(t-t_0)$, so

$f(x, y) \approx f(a, b) + f_x(a, b)x'(t_0)(t-t_0) + f_y(a, b)y'(t_0)(t-t_0)$. But since $f(x(t), y(t))$ should also be $\approx (df/dt)(t_0)(t-t_0)$, we can conclude that

$\dfrac{df}{dt} = \dfrac{\partial f}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial f}{\partial y}\dfrac{dy}{dt}$

This is the (or rather, one of the) Chain Rule(s) for functions of several variables. A similar line of reasoning, for example, would lead us to:

If $z = f(u, v, w)$ and $u = u(x, y)$, $v = v(x, y)$, and $w = w(x, y)$, then

$\dfrac{\partial f}{\partial x} = \dfrac{\partial f}{\partial u}\dfrac{\partial u}{\partial x} + \dfrac{\partial f}{\partial v}\dfrac{\partial v}{\partial x} + \dfrac{\partial f}{\partial w}\dfrac{\partial w}{\partial x}$ (by thinking of f as a (composite) function of x alone). A similar formula would hold for $\dfrac{\partial f}{\partial y}$.

In general, we can imagine a composition of functions of several variables as a picture with each variable linked by a line going up to functions it is a variable of, and linked by a line going down to variables it is a *function* of, with the original function f at the top. To find the derivative of f with respect to a variable, one finds all paths leading down from f to the variable, multiplying together all of the partial derivatives of one varaible w.r.t.

the variable below it, and adding these products together, one for each path. This can, as before, be verified using differentials. (Thinking two 'levels' at a time, this is essentially a repeated use of the ideas used above.)

Second Order Partial Derivatives

Just as in one variable calculus, a (partial) derivative is a function; so it has its own partial derivatives. These are called *second partial derivatives.*

We write $\frac{\partial}{\partial x}(\frac{\partial f}{\partial x}) = \frac{\partial^2}{\partial x^2}(\mathrm{f}) = \frac{\partial^2 f}{\partial x^2} = f_{xx} = (f_x)_x$, and similarly for y, and $\frac{\partial}{\partial y}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2}{\partial y \partial x}(\mathrm{f}) = (f_x)_y = f_{xy}$, and similarly for $\frac{\partial^2}{\partial x \partial y}$(these are called the *mixed partial derivatives.*

This leads to the slightly confusing convention that $\frac{\partial^2 f}{\partial x \partial y} = f_{yx}$ while $\frac{\partial^2 f}{\partial y \partial x} = f_{xy}$, but as luck would have it:

Fact: If f_{xy} and f_{yx} are both continuous, then they are equal [[Mixed partials are equal.]] So while at first glance a function of two variables would seem to have four second partials, it 'really' has only three. (Similarly, a function of three variables 'really' has six second partials, and not nine.)

In one-variable calculus, the second derivative measures concavity, or the rate at which the graph of f bends. The second partials f_{xx} and f_{yy} measure the bending of the graph of f in the x- and y-directions, while f_{xy} measures the rate at which the x-slope of f changes as you move in the y-direction, i.e., the amount that the graph is *twisting* as you walk in the y direction. The statement that $f_{xy} = f_{yx}$ then says that the amount of twisting in the y-direction is *always* the *same* as the amount of twisting in the x-direction, at any point, which is by no means obvious!

Tangent planes

In one-variable calculus, we can convince ourselves that a function has a tangent line at a point by zooming in on that point of the graph; the closer we look, the 'straighter' the graph appears to be. At extreme magnification, the graph looks just like a line - its tangent line.

Functions of two variables are really no different; as we zoom in, the graph of our function f starts to look like a plane - the graph's *tangent plane*. Finding the equation of this tangent plane is really a matter of determining its x- and y-slopes, which is precisely what the partial derivatives of f do. The x-slope is the rate of change of the function in the x-direction, i.e., the partial derivative with respect to x; and similarly for the y-slope.

So the equation for the tangent plane to the graph of $z = f(x, y)$ at the point

$$(a, b, f(a, b)) \text{ is} \qquad z = \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) + f(a, b)$$

And just as with one-variable calculus, one use we put this to is to find good approximations to $f(x, y)$ at points near (a, b);

$$f(x, y) \approx \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) + f(a, b), \qquad \text{for } (x, y) \text{ near } (a, b)$$

As with one variable, this also goes hand-in-hand with the idea of *differentials*:

$$df = f_x(a, b)dx + f_y(a, b)dy = \text{differential of } f \text{ at } (a, b)$$

And as before, $f(x, y) - f(a, b) \approx df$, when $dx = x - a$ and $dy = y - b$ are small.

Directional derivatives and the gradient

$\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ measure the instantaneous rate of change of f in the x- and y-directions, respectively. But what if we want to know the rate of change of f in the direction of the vector $3\vec{i} - 4\vec{j}$? By thinking of the partial derivatives in a slightly different way, we can get a clue to how to answer this question.

By writing $f_x(a,b) = \lim_{h\to 0} \dfrac{f((a,b)+h(1,0)) - f(a,b)}{h}$

and $f_y(a,b) = \lim_{h\to 0} \dfrac{f((a,b)+h(0,1)) - f(a,b)}{h}$,

we can make the two derivatives *look* the same; which motivates us to define the *directional derivative* of f at (a,b), in the direction of the vector $\vec{u}$, as

$f_{\vec{u}}(a,b) = D_{\vec{u}}(a,b) = \lim_{h\to 0} \dfrac{f((a,b)+h\vec{u}) - f(a,b)}{h}$

But running to the limit definition all of the time would take up way too much of our time; we need a better way to calculate directional derivatives! We can figure out how to do this using differentials:

For $\vec{u} = (u_1, u_2)$, $f((a,b) + h\vec{u} \approx df = f_x(a,b)hu_1 + f_y(a,b)hu_2$, so

$$\frac{f((a,b)+h\vec{u}) - f(a,b)}{h} \approx f_x(a,b)u_1 + f_y(a,b)u_2$$

and so taking the limit, we find that $f_{\vec{u}}(a,b) = f_x(a,b)u_1 + f_y(a,b)u_2 = (f_x(a,b), f_y(a,b)) \bullet \vec{u}$.

The vector $(f_x(a,b), f_y(a,b))$ is going to come up often enough that we will give it its own name;

$$(f_x(a,b), f_y(a,b)) = \nabla(f)(a,b) = \text{grad}(f)(a,b) = \text{the } \textit{gradient} \text{ of } f$$

So the derivative f in the direction of $\vec{u}$ is the dot product of $\vec{u}$ with the gradient of f. This means that (when θ is the angle between ∇f and $\vec{u}$), $D_{\vec{u}}(f) = ||Delf|| \cdot ||\vec{u}|| \cdot \cos(\theta)$ $= ||\nabla f|| \cos(\theta)$. This is the largest when $\cos(\theta) = 1$, i.e., $\theta = 0$ i.e., $\vec{u}$ points in the same direction as ∇f . So ∇f points in the direction of largest increase for the function f, at every point (a,b) . Its length is this maximum rate of increase.

On the other hand, when $\vec{u}$ points in the same direction as the level curve for the point (a,b) (i.e., it is tangent to the level curve), then the rate of change of f in that direction is 0; so $\nabla f \bullet \vec{u} = 0$, i.e, $\nabla f \perp \vec{u}$. This means that ∇f is perpendicular to the level curves of f, at every point (a,b).

Gradients for functions of 3 variables

For functions of 3 variables, everything works pretty much the same. We can make a similar construction of the directional derivative of $w = f(x,y,z)$; using the differential of f,

$$df = f_x(a,b)dx + f_y(a,b)dy + f_z(a,b)dz$$

we can compute that $D_{\vec{u}}(f) = \nabla f \bullet \vec{u}$, where $\nabla f = (f_x, f_y, f_z)$ is the gradient of f. For the exact same reasons, this means that ∇f points in the direction of maximal increase for f, and ∇f is perpendicular to the *level surfaces* for f. This can be used to find the equation for the tangent plane to a level surface, since it is the plane orthognal to ∇f:

$$(f_x(a,b,c), f_y(a,b,c), f_z(a,b,c)) \bullet (x-a, y-b, z-c) = 0$$

We can use the gradient of functions of 3 variables to help us understand the graphs of functions of two variables, since we can think of the graph of a function of two variables, $z = f(x,y)$, as a *level curve* of a function of 3 variables

$$g(x,y,z) = f(x,y) - z = 0$$

The gradient of g is perpendicular to its level curves, so it is perpendicular to the graph of f, so gives us the normal vector for the tangent plane to the graph of f. Computing, we find that

$$\nabla g = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1) = \vec{n}$$

which means that the equation for the tangent plane to the graph of $z = f(x,y)$ at the point $(a,b,f(a,b))$ is

$$(\frac{\partial f}{\partial x}(a,b),\frac{\partial f}{\partial y}(a,b),-1)\bullet(x-a,y-b,z-f(a,b))=0$$

Local Extrema

The partial derivatives of f measuire the rate of change of f in each of the coordinate directions. So they are giving us partial information (no pun intended) about how thew function f is rising and falling. And just as in one-variable calculus, we ought to be able to turn this into a procedure for findong out when a function is at its maximum or minimum. The basic idea is that at a max or min for f, then, thinking of f just as a function of x, we would *still* think we were at a max or min, so the derivative, as a function of x, will be 0 (if it is defined). In other words, $f_x = 0$. similarly, we would find that $f_y = 0$, as well. following one-variable theory, therefore, we say that

A point (a, b) is a **critical point** for the function f if $f_x(a, b)$ and $f_y(a, b)$ are *each* either 0 or undefined. (A similar notion would hold for functions of more than two variables.)

Just as with the one-variable theory, then, if we wish to find the max or min of a function, what we first do is find the critical points; *if* the function has a max or min, it will occur at a critical point.

And just as before, we have a 'Second Derivative Test' for figuring out the difference between a (local) max and a (local) min (or *neither*, which we will call a *saddle point*). The point is that at a critical point, f looks like its quadratic approximation, which (simplifying things somewhat) is described as $Q(x, y) = Dx^2 + Exy + Fy^2$ (since the first derivatives are 0). By completing the square, we can see that the actual shape of the graph of Q is basically described by *one number*, called the descriminant, which (in terms of partial derivatives) is given by

$$D = f_{xx}(\text{a,b})f_{yy}(\text{a,b})-(f_{xy}(\text{a,b}))^2$$

(Basically, Q looks like one of x^2+y^2 (local min), $-x^2-y^2$ (local max), or x^2-y^2 (saddle), and D tells you if the signs are the same ($D > 0$) or opposite ($D < 0$) . More specifically, if, at a critical point (a, b),

$D > 0$ and $f_{xx} > 0$ then (a, b) is a local min; if
$D > 0$ and $f_{xx} < 0$ then (a, b) is a local max; and if
$D < 0$, then (a, b) is a saddle point
(We get no information if $D = 0$.)

Global Extrema: Unconstrained Optimization

Critical points help us find local extrema. To find *global* extrema, we take our cue from one-variable land, where the procedure was (1) Identify the domain, (2) find critical points *inside* the domain, (3) plug critical points and *endpoints* into f, (4) biggest is the max, smallest is the min.

For two variables, we do (essentially) *exactly the same thing*:

(1) Identify the domain
(2) Find critical points in the *interior* of the domain
(3) Identify the (potential) max and min values on the *boundary* of the domain (more about this later!)
(4) Plug the critical points, and your potential points on the boundary
(5) biggest is max, smnallest is min

This works if the domain is *closed* and *bounded* (think, e.g., of a closed interval in the x direction and a closed interval in the y direction, or the inside of a circle in the plane (including the circle)). Usually, in practice, we don't have such nice domains; but we usually know from physical considerations that our function *has* a max or min (e.g., find the maximum volume you can enclose in a box made from 300 square inches of cardboard...), and so we *still* know that it has to occur at a critical point of our function.