Math 208

Topics for the second exam

(Technically, everything covered on the first exam, *plus*)

Chapter 13: Differentiation

§6: 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 fuinction 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 differentials:

$df = f_x dx + f_y dy$, while $dx = \dfrac{dx}{dt}dt$ and $dy = \dfrac{dy}{dt}dt$. Putting these together we get

$df = (\dfrac{\partial f}{\partial x}\dfrac{dx}{dt}+\dfrac{\partial f}{\partial y}\dfrac{dy}{dt})dt = \dfrac{df}{dt}dt$, which implies 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 would lead us to:

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

$\dfrac{\partial f}{\partial x}= \dfrac{\partial f}{\partial u}\dfrac{\partial u}{\partial x}+\dfrac{\partial f}{\partial v}\dfrac{\partial v}{\partial x}$. 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.

§7: 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 $\dfrac{\partial}{\partial x}(\dfrac{\partial f}{\partial x}) = \dfrac{\partial^2}{\partial x^2}(\mathrm{f}) = \dfrac{\partial^2 f}{\partial x^2} = f_{xx} = (f_x)_x$, and similarly for y, and

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

This leads to the slightly confusing convention that $\dfrac{\partial^2 f}{\partial x\partial y} = f_{yx}$ while $\dfrac{\partial^2 f}{\partial y\partial x} = f_{xy}$, 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!

§8: **Taylor 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* :

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

Functions of two variables are not much different; we just replace the word 'derivative' with '*partial* derivative'! So for example, the degree one Taylor polynomial is

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

which is nothing more than our old formula for the tangent plane to the graph of f at the point $(a,b,f(a,b))$.

We will soon need the second degree version (which for simplicity we will write for the point $(a,b) = (0,0)$) :

$$Q(x,y) = L(x,y) + \frac{f_{xx}(0,0)}{2}x^2 + f_{xy}(0,0)xy + \frac{f_{yy}(0,0)}{2}y^2 = L(x,y) + Ax^2 + Bxy + Cy^2$$

As before, L and Q are the 'best' linear and quadratic approximations to f, near the point (a,b), in a sense that can be made precise; basically, $L-f$ shrinks to 0 like a quadratic, near (a,b), while $Q-f$ shrinks like a cubic (which shrinks to 0 *faster*, when your input is small).

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.

The basic fact, that we 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.

Chapter 14: Optimization: Local and Global Extrema

§1: **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* thew 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 second degree Taylor polynomial, which (simplifying things somewhat) is described as $Q(x,y) = Ax^2 + Bxy + Cy^2$ (since the first derivatives are 0). 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

$D = f_{xx}f_{yy} - (f_{xy})^2$

(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 $D = 0$.)

§2: 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, smnallewst is min

This works if the domain is *closed* and *bounded* (think, e.g., of a closed interval in the x direction and a closd intervasl in the y direction, or the inside of a circle in the plane). 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.

§3: Constrained Optimization: Lagrange Multipliers

Most optimization problems that arise naturally at not unconstrained; we are usually trying to maximize one function while *satisfying* another. Even the problem above is best phrased this way; maximize *volume* subject to the *constraint* that surface area equals 300. We can use the one-variable calculus trick of solving the constraint for one variable, and plugging this into the function we wish to maximize, **or** we can take a completely different (and often better) approach:

The basic idea is that if we think of our constraint as describing a level curve (or surface) of a function g, then we are trying to maximize or minimize f among all the points of the level curve. If the level curves of f are cutting *across* our level curve of g, it's easy to see that we can icrease or decrease f while still staying on the level curve of g. So at a max or min, the level curve of f has to be *tangent* to our constraining level curve of g. This in turn means:

At a max or min of f subject to the constraint g, $\nabla f = \lambda \nabla g$ (for some real number λ)

We must also satisfy the constraint : $g(x,y) = c$.

So to solve a constrained optimization problem (m,ax.min of f subject to the constraint $g(x,y) = c$) we solve

$\nabla f = \lambda \nabla g$ **and** $g(x,y) = c$

This in turn allows us to finish our procedure for finding global extrema, since step (3) can be interpreted as a constrained optimization problem (max or min on the *boundary*). In these terms,

To optimize f subject to the condition $g(x, y) \leq c$, we (1) solve $\nabla f = 0$ and $g(x, y) < c$, (2) solve $\nabla f = \lambda \nabla g$ and $g(x, y) = c$, (3) plug all of these points into f, (4) the biggest is the max, the smallest is the min.

[This works fine, unless the region $g(x, y) \leq c$ runs off to infinity; but often, physical considerations will still tell us that one of our critical points is an optimum.]

Chapter 15: Integrating Functions of Several Variables

§1: The Definite Integral of a Function of Two Variables

In an entirely formal sense, the intergal of a function of one variable is a great big huge sum of little tiny numbers; we add up things of the form $f(c_i)\Delta x_i$, where we cut the interval $[a, b]$ we are integrating over into little intervals of length Δx_i, and pick points c_i in each interval. In esssence, the integral is the sum of areas of very thin rectangles, which leads us to iterpret the integral as the area under the graph of f.

For functions of two variables, we do the exact same thing. To integrate a function f over a rectangle in the plane, we cut the rectangle into lots of tiny rectangles, with side lengths Δx_i and Δy_j, pick a point in each rectangle, and then add up $f(x_i, y_j)\Delta x_i \Delta y_j$. This gives an *approximation* to the actual integral; letting the little side length go to zero, we arrive at what we would call the integral of f over the rectangle R, which we denote by

$\int_R f\ dA$ (where dA denotes the 'differential of area' $dxdy$ (or $dydx$)

The idea is that if we think of f as measuring height above the rectangle, then $f(x_i, y_j)\Delta x_i \Delta y_j$ is the volume of a thin rectangular box; letting the Δ's go to zero, the integral would then measure the *volume* under the graph of f, lying over the rectangle R.

If the region R *isn't* a rectangle, we can still use this method of defining an integral; we simply *cover* R with tiny rectangles, take the same sum, and let the Δ's go to 0.

Of course, we have no reason to believe that as the Δ's go to 0, this collection of sums will *converge* to a single number. But it is a basic fact that if the function f is *continuous*, and the region R isn't too ugly, then these sums always will converge.

§2: Iterated Integrals

Of course, the preceding approach is no way to *compute* a double integral! Instead, we (as usual) steal an idea from one-variable calculus.

The idea is that we already *know* how to compute volumes, and so we *implicitly* know how to compute double integrals! We can compute the volume of a region by integrating the *area of a slice*. You can do this two ways; (thinking in terms of the region R in the plane) you can slice R into horizontal lines, and integrate the area of the slices dy, or you can slice R into vertical lines, and integrate the slices dx.

but each *slice* can be interpreted as an integral; the area of a horizontal slice is the integral of f, thought of as *just a function of* x, and the area of a vertical slice is the integral of f, thought of as *just a function of* y. This leads to *two* ways to compute our integral:

$\int_R f\ dA = \int_c^d (\int_a^b f(x, y)\ dx)\ dy$ (for horiz slices) $= \int_a^b (\int_c^d f(x, y)\ dy)\ dx$ (for vert slices)

In each case, the *inner* integral is thought of as the integral of a function of *one variable*. It just happens to be a *different* variable in each case. In the case of a rectangle, the limits of integration are just numbers, as we have written it. In the case of a more complicated region R, the inner limits of integration might depend on where we cut. The idea is that a slice along a horizontal line is a slice along $y =$ constant, and the endpoints of the integral might *depend on* y; for a slice along a vertical line (x = constant), the endpoints might depend on x .

So, e.g., to integrate a function f over the region lying between the graphs of $y = 4x$ and $y = x^3$, we would compute either

$\int_0^2 (\int_{x^3}^{4x} f(x, y)\ dy)\ dx$ or $\int_0^8 (\int_{y/4}^{y^{1/3}} f(x, y)\ dx)\ dy$

Which should we compute? Whichever one is easier! They give the *same* number!

§3: **Triple Integrals**

Triple integrals are just like double integrals, only more so. We can define them as a limit of a huge sum; here the terms in the sum would be the value of the function f time the volume of a tiny rectangular box. The usual interpretation of a triple integral arises by thinking of the function f as giving the density of the matter at each point of a solid region W in 3-space. Since density times volume is mass, the integral of f over the region W would compute the mass of the solid object occupying the region W. In the special case that f is the function 1, the integral will compute the volume of the region W.

Again, as with double integrals, the way we really comupute a triple integral is as a (triply) iterated integral. You pick a direction to slice (x=constant, y=constant, or z=constant) W up, and compute the integral of f over each slice. Each of these is a double integral (computed as an iterated integral), whose value *depends* on the variable you sliced along. To compute the integral over W, you integrate these double integrals over the last variable, getting three iterated integrals.

Put slightly differently, you can evaluate a triple integral by integrating out each variable, one at a time. Typically, we start with z, since our region W is usually described as the region lying between the graphs of two functions, given as z=blah and z=bleh . The idea is to first, for each fixed value of x and y, integrate the function f, dz, from blah to bleh. (Ther resulting values depend on x and y, i.e., are a *function* of x and y.) Then we integrate over the region, R, in the plane consisting of the points (x, y) such that the vertical line *hits* the region W. We usually call this region R the *shadow* of W in the x-y plane. In symbols

$\int_W f\ dV = \int_R (\int_{a(x,y)}^{e(x,y)} f(x, y, z)\ dz)\ dA$

For example, the integral of a function over the region lying above the x-y plane and inside the sphere of radius 2, centered at the origin, would be computed as

$\int_R (\int_0^{\sqrt{4-x^2-y^2}} f(x, y, z)\ dz)\ dA = \int_{-2}^2 (\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} (\int_0^{\sqrt{4-x^2-y^2}} f(x, y, z)\ dz)\ dy)\ dx$

where R is the shadow of W (in this case, the disk of radius 2,m centered at the origin, in the x-y plane).