Math 208H

Topics for the second exam

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

Vector-valued functions

Basic idea: think of a parametric curve in 3-space.

$\vec{r}(t) = (x(t), y(t), z(t))$
If we think of t as time, then what $\vec{r}$ does is give us a point in 3-space at each moment of time. Thinking of $\vec{r}$ as the position of a particle, the particle sweeps out a path or curve, C, in 3-space as time passes.

Example: *lines*; they can be described as having a starting place and a direction they travel, and so can be parametrized by $\vec{r}(t) = P + t\vec{v}$, where P is the starting point and $\vec{v}$ is the direction (for example, the difference of two points lying *along* the line).

Vector function calculus

We can extend the concept of a limit to vector-valued functions by thinking in terms of distance; $\vec{r}(t)$ approaches L as t goes to a if the distance between $\vec{r}(t)$ and L tends to 0. This in turn is the same as insisting that each coordinate function $x(t), y(t), z(t)$ tends to the corresponding corrdinate of L as t goes to a. So in particular, a vector function $\vec{r}(t)$ is *continuous* at a if each of its coordinate functions x, y, z are continuous at a.
When we think of t as time, we can imagine ourselves as travelling along the parametrized curve $\vec{r}(t)$, and so at each point we can make sense of both *velocity* and *acceleration*. Velocity, which is the instantaneous rate of change of position, can be calculated as the limit of the usual difference quotient, using the ideas above; but since limits can really be computed one coordinate at a time, the derivative of $\vec{r}(t) = x(t), y(t), z(t)$ is $\vec{v}(t) = \vec{r}'(t)$ $= x'(t), y'(t), z'(t)$.

Some basic properties:

$(\vec{r} + \vec{s})'(t) = \vec{r}'(t) + \vec{s}'(t)$
$(f(t)\vec{r}(t))' = f'(t)\vec{r}(t) + f(t)\vec{r}'(t)$
$(\vec{r} \bullet \vec{s})'(t) = \vec{r}'(t) \bullet \vec{s}(t) + \vec{r}(t) \bullet \vec{s}'(t)$
$(\vec{r} \times \vec{s})'(t) = \vec{r}'(t) \times \vec{s}(t) + \vec{r}(t) \times \vec{s}'(t)$
Similarly, acceleration can be computed as $\vec{a}(t) = \vec{r}''(t) = x''(t), y''(t), z''(t)$; it is the rate of change of the velocity of $\vec{r}(t)$.
One useful fact: if the length of the velocity (i.e., its *speed*), $||\vec{v}(t)||$ is constant, then $\vec{a}(t)$ is always perpendicular to $\vec{v}(t)$
And speaking of length, we can compute the *length* of a parametrized curve by integrating its speed: the length of the parametrized curve $\vec{r}(t)$, $a \leq t \leq b$, is

$$\text{Length} = \int_a^b ||\vec{v}(t)||\, dt$$

Since vector functions have derivatives, which are also vector functions, they therefore have *antiderivatives*; $\vec{R}(t)$ is the antiderivative of $\vec{r}(t)$ if $\vec{R}'(t) = \vec{r}(t)$. Since derivatives can be computed by taking the derivative of each coordinate function, its antiderivative can be computed by taking the antiderivative of each coordinate.

Motion in space

Newton's second law states that the (mass times the) acceleration of a particle is equal to the (vector) sum of all of the forces acting on the particle. This means that if we know all of the forces acting on an object, we know its acceleration. But if we know an object's acceleration (and it's velocity for one t), we can recover its velocity by integrating:

$$\vec{r}'(t) = \vec{r}'(t_0) + \int_{t_0}^{t} \vec{r}''(t)\ dt$$

Then, in turn, if we know the position of the object at one t, we can recover the position of the object, by integrating (again):

$$\vec{r}(t) = \vec{r}(t_0) + \int_{t_0}^{t} \vec{r}'(t)\ dt$$

So knowing the forces acting on an object, together with its initial position and initial velocity, allows us to determine its position at any time t.

Typical forces we may encounter:
F_g = force due to gravity = $(0, 0, -g)$, where g = 32 ft/sec^2 = 9.8 m/sec^2
F_w = force due to the wind = any particular constant vector $\vec{v}$

Functions of Several Variables

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 basic tool is going to be to *think of a function of several variables as a function of* **one variable** *(at a time!)*, s ithat we can use those tools to good effect.

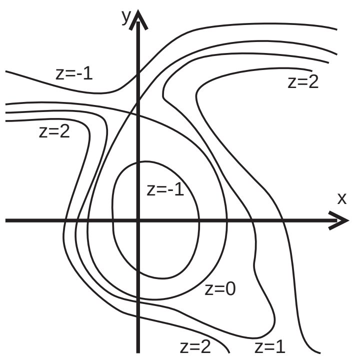
Graphs of functions of two variables

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 doman 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 appart, 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 simply 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*.

Differentiation

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:

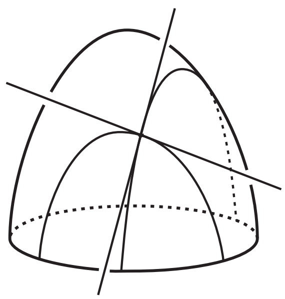
$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} \text{ and } \frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

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,

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

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

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

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

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

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

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

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

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

$\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ 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 $f_x(a,b) = \displaystyle\lim_{h\to 0} \frac{f((a,b)+h(1,0)) - f(a,b)}{h}$

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

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

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

[Technically, we need $\vec{u}$ to be a unit vector, $||\vec{u}|| = 1$; for other vectors $\vec{v}$, we would define $D_{\vec{v}}(f) = D_{\vec{v}/||\vec{v}||}(f)$.]

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.

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

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

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 differentials:

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

$df = (\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt})dt = \frac{df}{dt}dt$, which implies that $\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$

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

$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}$. A similar formula would hold for $\frac{\partial f}{\partial y}$.

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.

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!

Linear and quadratic 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 best linear approximation 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: the most general degree 2 polynomial is

$$A + Bx + Cy + Dx^2 + Exy + Fy^2$$

When we (by computing derivatives) determine the one with the same first and second partial derivative as $z = f(x,y)$ at (a,b), we find that it is

$$Q(x,y) = L(x,y) + \frac{f_{xx}(a,b)}{2}(x-a)^2 + f_{xy}(a,b)(x-a)(y-b) + \frac{f_{yy}(a,b)}{2}(y-b)^2$$

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. 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...)

Optimization: Local and Global Extrema

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.

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 increase 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 (max/min of f subject to the constraint $g(x, y) = c$) we solve

$\nabla f = \lambda \nabla g$ **and** $g(x, y) = c$ for x, y, and λ. All of the pairs (x, y) that arise are candidates for the max/min; and the max and min must occur at some of these points. [Technically, as before, we must also include points along $g(x, y) = c$ where ∇f is undefined; we won't run into this possibility in practice, however.]

This also works for functions of more than two variables; the procedure is exactly the same. In all of these cases, the real work is in solving the resulting equations! A basic technique that often works is to solve each of the coordinate equations in $\nabla f = \lambda \nabla g$ for λ ; the other halves of the equations are then all equal to one another (since they all equal λ).

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.]