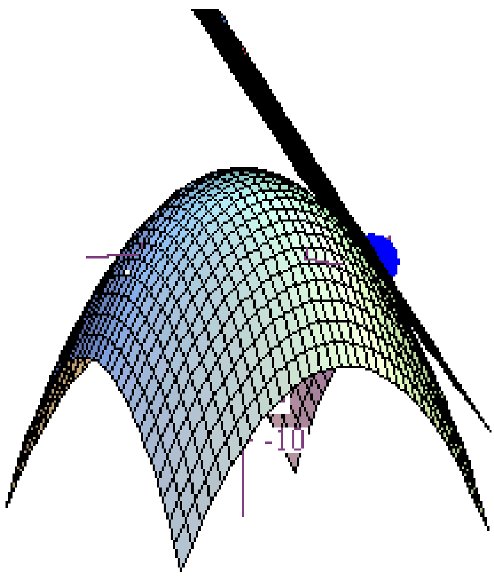
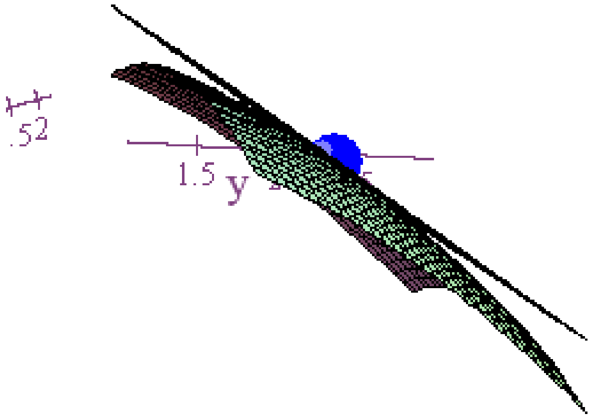
CHAPTER 7

DERIVATIVES

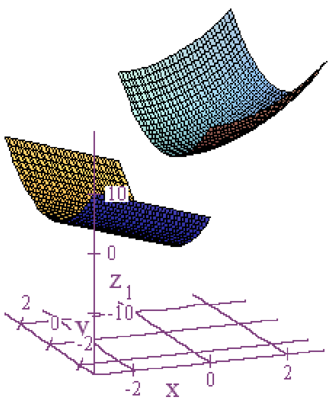
At last we have arrived! We're finally ready to look at the wide, wide world of derivatives of multivariable functions! And as you might suspect, just as graphing is richer and more complex in three dimensions, so will the concept of a derivative have more dimensions to it (pun intended!). In a nutshell, though, we will say that a function of several variables is differentiable at a point if we can define a nonvertical tangent plane at that point. For example, below is a graph of $z = f(x, y) = -x^2 - y^2 + 4$ with a tangent plane plotted at the point $(1, 2, -1)$.

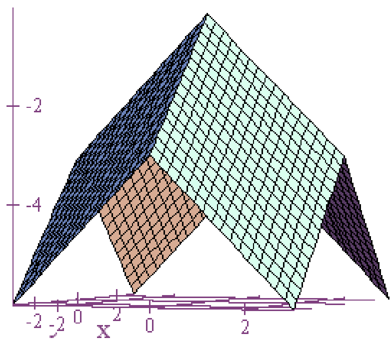


When we look at this tangent plane, we see that every single tangent line at the point $(1,2,-1)$ is contained in that plane, and we also realize that the tangent plane is a good approximation for the function when our input values are close to $x=1$ and $y=2$. Also, because we can define a tangent plane at this point, that means that as we zoom in on this point the graph of $z=f(x,y)=-x^2-y^2+4$ resembles the tangent plane more and more. That property is what we call local linearity, and you can easily see this property for many functions in two dimensions using a standard graphing calculator. Just take function such as $y=f(x)=x^2$, zoom in on any point on the graph, and your curve will resemble a straight line more and more as you zoom in. In other words, $y=f(x)=x^2$ is locally linear. Here is the graph of $z=f(x,y)=-x^2-y^2+4$ zoomed in around the point $(1,2,-1)$ so that you can see the local linearity starting to take effect.



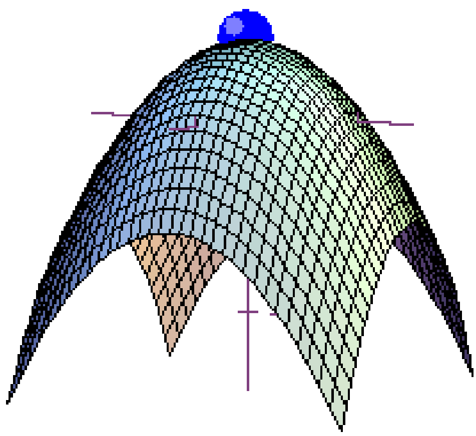
The appropriate question to ask now is what sorts of things would prevent a graph from being differentiable at a point? Normally, the same sorts of things that cause nondifferentiability back in two dimensions. In other words, breaks in continuity and sharp points. Here are a couple of examples.



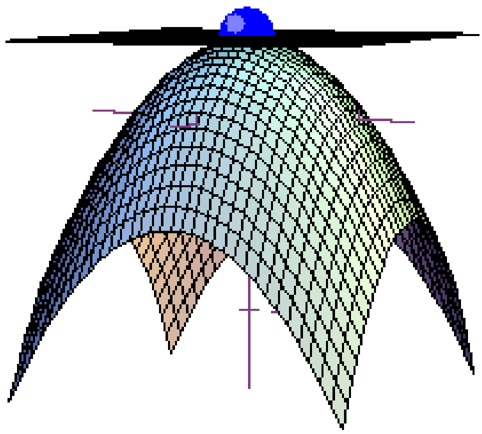


In the first graph, we're not going to be able to define tangent planes where the breaks in the graph occur, and in the second surface, we can't define tangent planes at the sharp edges or at the sharp point at the top of the graph. As you've seen before, derivatives require smoothness in order to be defined, and wherever a graph is not smooth, it's not going to be differentiable.

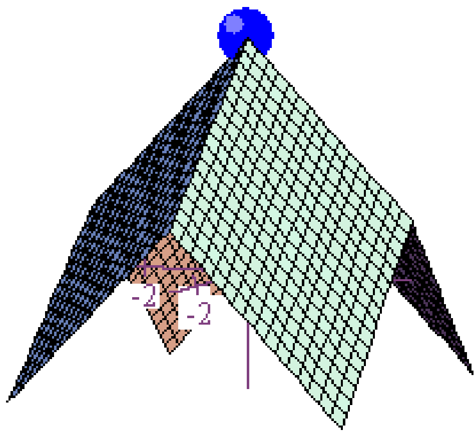
Now let's go back to the graph of $z = f(x,y) = -x^2 - y^2 + 4$. What's obvious here is that the point with coordinates $(0,0,4)$ represents the top of a hill. When that happens we call it a local maximum. Furthermore, if there are no other points or hilltops on the graph that are higher, then we also call it an absolute maximum.



If we now add the tangent plane to this point, then we can visually see something very important. Namely, that it's horizontal. This also means that that all the tangent lines in all directions have slope zero.



On the other hand, if we look at the graph of $z = -|x| - |y| + 4$, then we can see that it also has a local and absolute maximum at $(0,0,4)$, but no tangent plane exists at this point since it's a sharp point. This also tells us that tangent lines will generally fail to exist in some if not all directions when we can't define a tangent plane.



Back in first semester calculus you probably saw that if a function $y = f(x)$ had a local maximum or minimum at a point, then either the derivative was zero at that point or it was undefined. A similar criteria exists for functions of the form $z = f(x, y)$. If a local maximum or minimum exists for a function of this sort, then either the slopes of the tangent lines in both the direction of the x-axis and the y-axis are zero, or else one of these slopes fails to exist. If we find points that meet this criteria, then those are our candidates for local maximums and minimums.

Let's now go back to the function $z = f(x, y) = -x^2 - y^2 + 4$ and rethink how we might find tangent lines at a point. Previously, back in chapter 3, we went through a somewhat complicated procedure. That is, we picked a point, we fixed either the x-

or y-value, we graphed the resulting equation in two dimensions, we took a derivative to find the slope of the tangent line at a particular point, and then we moved the whole thing back into three dimensions. Trust me, there's an easer way to do it, and we're all about life being easy! Instead of fixing a y-value and then differentiating with respect to x, let's just pretend that y is fixed and go ahead and find our derivative. Thus, if $z = f(x, y) = -x^2 - y^2 + 4$, but we pretend that y is a fixed constant, then when we differentiate, we get back $-2x$. In other words, if y is a constant, then the derivative of $-y^2$ is zero just as the derivative of the constant term 4 is zero. Now let's go back and pretend this time that it's x which is fixed to a constant value and y that is variable. The derivative of z this time will be $-2y$.

Already this may be getting a little confusing if we try to do too much of it in our head, so let's try and introduce some notation in order to simplify things. First, since z is a function of two variables, that means there are two derivatives we could compute here, one with respect to x and one with respect to y. Since each derivative by itself tells only part of the story, we call each derivative a partial derivative, and we have two basic notations that we'll use for partial derivatives. For the partial derivative of $z = f(x, y)$ with respect to x, we'll use z_x or f_x or $\frac{\partial z}{\partial x}$ or $\frac{\partial f}{\partial x}$. Similarly, for the partial derivative of $z = f(x, y)$ with respect to y, we'll use z_y or f_y or $\frac{\partial z}{\partial y}$ or

$\frac{\partial f}{\partial y}$. The latter notation is similar to the usual notation for a derivative except that we use a somewhat stylized version of the letter "d" to denote that it's a partial derivative. If you use this notation, write it correctly so that people will know that it is a partial derivative that you are talking about! Also, both of these partial derivatives are technically defined in terms of limits such as the ones below.

$$z_x = \frac{\partial z}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}$$

$$z_y = \frac{\partial z}{\partial y} = \lim_{\Delta y \to 0} \frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}$$

In practice, though, you find the partial derivative with respect to x by treating y as fixed, and you find the partial derivative with respect to y by treating x as fixed. Also, the partial derivative of z with respect to x can be interpreted as either the instantaneous rate of change of z with respect to a change in x, or as the slope of the tangent line in the direction of the x-axis if you evaluate this partial derivate at a particular point. Likewise, the partial derivative of z with respect to y can be interpreted as either the instantaneous rate of change of z with respect to a change in y, or as the slope of the tangent line in the direction of the 0-axis. Since it's important to be very good at computing partial derivatives, below are several examples. Study them well, and make sure you understand what is going on.

1. $z = f(x, y) = x^2 + y^2$

$$\frac{\partial z}{\partial x} = 2x$$
$$\frac{\partial z}{\partial y} = 2y$$

2. $z = f(x, y) = x^2 - y^2$

$$\frac{\partial z}{\partial x} = 2x$$
$$\frac{\partial z}{\partial y} = -2y$$

3. $z = f(x, y) = xy$

$$\frac{\partial z}{\partial x} = y$$
$$\frac{\partial z}{\partial y} = x$$

4. $z = f(x, y) = x + y$

$$\frac{\partial z}{\partial x} = 1$$
$$\frac{\partial z}{\partial y} = 1$$

5. $z = f(x, y) = \dfrac{y}{x}$

$$\frac{\partial z}{\partial x} = \frac{-y}{x^2}$$
$$\frac{\partial z}{\partial y} = \frac{1}{x}$$

6. $z = f(x, y) = \ln(xy)$

$$\frac{\partial z}{\partial x} = \frac{y}{xy} = \frac{1}{x}$$
$$\frac{\partial z}{\partial y} = \frac{x}{xy} = \frac{1}{y}$$

7. $z = f(x,y) = e^{5xy^2}$

$$\frac{\partial z}{\partial x} = e^{5xy^2} \cdot 5y^2 = 5y^2 e^{5xy^2}$$

$$\frac{\partial z}{\partial y} = e^{5xy^2} \cdot 10xy = 10xye^{5xy^2}$$

8. $z = f(x,y) = 4x^2 + y^2$

$$\frac{\partial z}{\partial x} = 8x$$

$$\frac{\partial z}{\partial y} = 2y$$

9. $z = f(x,y) = -\left(x^2 + y^2\right)$

$$\frac{\partial z}{\partial x} = -2x$$

$$\frac{\partial z}{\partial y} = -2y$$

10. $z = f(x,y) = x^2$

$$\frac{\partial z}{\partial x} = 2x$$

$$\frac{\partial z}{\partial y} = 0$$

11. $z = f(x,y) = \sqrt{xy}$

$$\frac{\partial z}{\partial x} = \frac{1}{2}\left(xy\right)^{-\frac{1}{2}} \cdot y = \frac{y}{2\sqrt{xy}}$$

$$\frac{\partial z}{\partial y} = \frac{1}{2}\left(xy\right)^{-\frac{1}{2}} \cdot x = \frac{x}{2\sqrt{xy}}$$

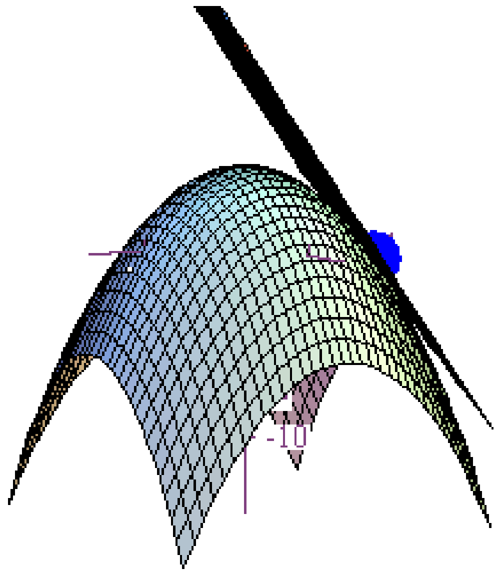
Now let's return to an exploration of $z = f(x,y) = -x^2 - y^2 + 4$ and see how we might construct the tangent plane at the point $\left(1,2,-1\right)$. Recall now that back in chapter 1 we said that if an equation for a plane was written in the form $z = Ax + By + C$, then A

would be the slope of the plane in the direction of the *x*-axis, and *B* would be the slope of the plane in the direction of the *y*-axis. We can now find these slopes for $z = f(x, y) = -x^2 - y^2 + 4$ by taking partial derivatives and evaluating the results at the point $(1, 2, -1)$. Clearly, $z_x = -2x$ and $z_y = -2y$. If we evaluate these partial derivatives at $x = 1$ and $y = 2$, we get $z_x(1) = -2(1) = -2$, and $z_y(2) = -2(2) = -4$. Thus, the equation for our tangent plane is starting to look like $z = -2x - 4y + C$. To now find the value of *C*, just plug in the coordinates of the point $(1, 2, -1)$ for *x*, *y*, and *z*.

$$-1 = -2(1) - 4(2) + C \Rightarrow C = 9$$

$$z = -2x - 4y + 9$$

If we now graph our parabolid, our point, and our surface together, then we can see it worked.

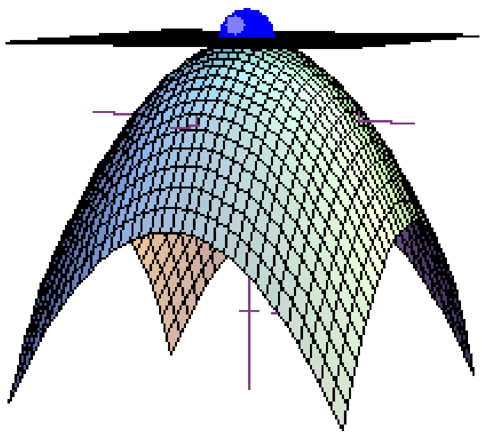


Now let's backtrack a second and think in terms of how we would use our partial derivatives to find the location of our local maximum. As we alluded to previously, at such an extreme point either both partial derivatives equal zero of one of them fails to exist. Well, if we set each of the partial derivatives for $z = f(x, y) = -x^2 - y^2 + 4$ equal to zero, then we're in business.

$$\frac{\partial z}{\partial x} = -2x = 0 \Rightarrow x = 0$$

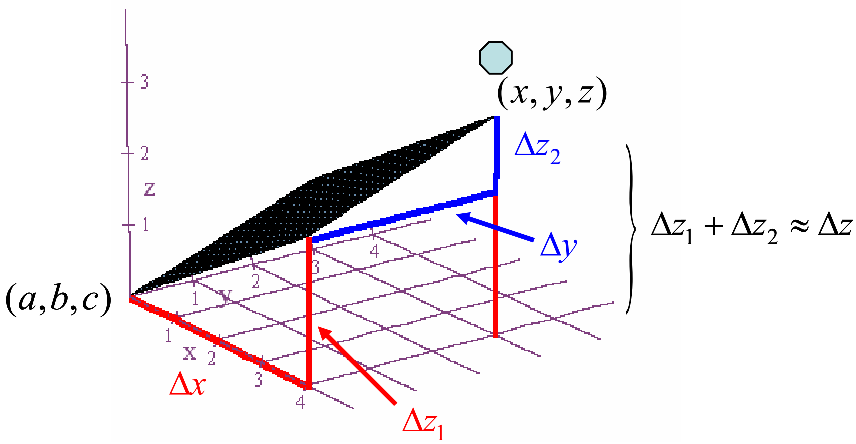
$$\frac{\partial z}{\partial y} = -2y = 0 \Rightarrow y = 0$$

Both partials are equal to zero when $x = 0$ and $y = 0$, so we'll call the point $(0,0)$ a <u>critical point</u>, and from the graph above we can see that we will have a local (and absolute!) maximum at this point. Also, technically speaking, the local maximum is just the z-value that we get at our critical point, in this case $z = f(0,0) = 4$. However, I always like to know not only the z-coordinate, but the x- and y-coordinates, too. Thus, I'll always give my answers as coordinates for a point, and I'll refer, in this example, to the point $(0,0,4)$ as a local maximum point. And finally, when we add the tangent plane $z = 4$ to our graph, we see that it is perfectly horizontal at the point $(0,0,4)$.



Later on, we'll learn an algebraic test for determining whether a critical point results in a local maximum or a local minimum or something else, but for now we'll combine our algebraic methods with graphical insights.

At this point, we want to move on to another topic and define what's known as the total differential. It is basically a formula that shows the relationship between small changes in x and y and the corresponding change in z, and among other things, it can be used for doing approximations related to the output of our function. We'll begin our development by looking at a diagram that's similar to some we've seen before.



What we are looking at here is a portion of a tangent plane that we might have at a point (a,b,c) on the surface of some function $z = f(x,y)$. The graph of $z = f(x,y)$, however, is not depicted here. All we are looking at is the tangent plane. Also, assume that (x,y,z) is another point on the graph of $z = f(x,y)$ that is close to (a,b,c). Then the change in z is $\Delta z = z - c$, and using the tangent plane we can approximate this change in z by $\Delta z \approx \Delta z_1 + \Delta z_2$. Also, from our diagram above, we can see that the slope of the plane in the direction of the positive x-axis is $\frac{\Delta z_1}{\Delta x}$ and the slope of the plane in the direction of the positive y-axis is $\frac{\Delta z_2}{\Delta y}$. However, these slopes are also equal, respectively, to $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ for our function $z = f(x, y)$. Thus,

$$\frac{\Delta z_1}{\Delta x} = \frac{\partial z}{\partial x} \Rightarrow \Delta z_1 = \frac{\partial z}{\partial x}\Delta x$$

$$\frac{\Delta z_2}{\Delta y} = \frac{\partial z_2}{\partial y} \Rightarrow \Delta z_2 = \frac{\partial z}{\partial y}\Delta y$$

And the consequence of this is that,

$$z - c = \Delta z \approx \Delta z_1 + \Delta z_2 = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y .$$

If we write these changes in x, y, and z as differentials, then we get an expression that we call the total differential.

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

However, since we are mainly going to use this expression for approximations, the previous form, $z - c = \Delta z \approx \Delta z_1 + \Delta z_2 = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$, is a little better. For example, if we take $z = f(x, y) = -x^2 - y^2 + 4$ and $x = 1$ and $y = 2$, and if we want to change our input to $x = 1.1$ and $y = 2.3$ then we can use the total differential to approximate both the new function value and the change that occurs in z. For the change in z use the formula $\Delta z \approx \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$. To finish the computation, we have to specify the change in x, the change in y, and the value of both partial derivatives at our first point, $(1,2)$.

The results are below.

$$\Delta x = 0.1$$
$$\Delta y = 0.3$$
$$\frac{\partial z}{\partial x} = -2x,\ \frac{\partial z(1,2)}{\partial x} = -2$$
$$\frac{\partial z}{\partial y} = -2y,\ \frac{\partial z(1,2)}{\partial y} = -4$$
$$\Delta z \approx \frac{\partial z(1,2)}{\partial x}\Delta x + \frac{\partial z(1,2)}{\partial y}\Delta y = (-2)(0.1) + (-4)(0.3) = -1.4$$

We can also approximate the new z value by rewriting our formula above as

$z \approx \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y + c$. In this case, we get $z(1.1,\ 2.3) \approx (-2)(0.1) + (-4)(0.3) - 1 = -2.4$. If we computer the actually value of $z(1.1,\ 2.3)$, we get

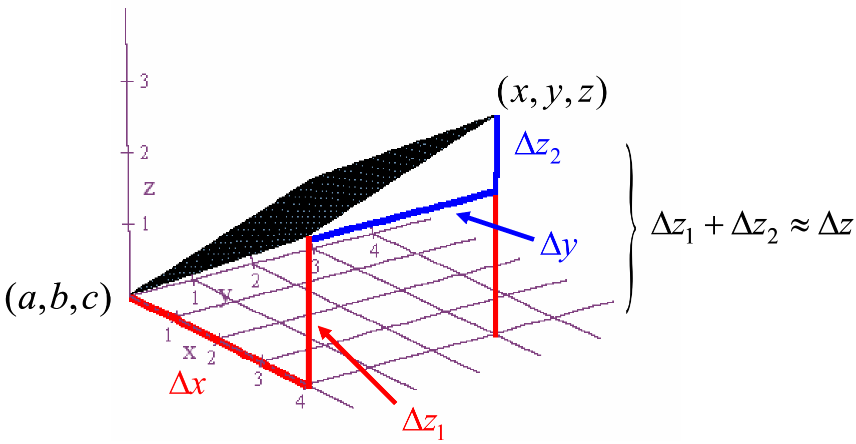
$z(1.1,\ 2.3) = -(1.1^2) - (2.3^2) + 4 = -2.5$ which means the actual change in z is -1.5.

Well, we can see that our approximations are pretty close, and that's the whole point. Nonetheless, the total differential has some other applications, too. It helps us see at a glance what the chain rule should look like for a function of several variables.

Recall that in first semester calculus you occasionally had a function that could be considered a composition of two or more functions, and then you had to use the chain rule in order to get the derivative. Likewise, for functions of two or more variables there also exists a version of the chain rule, several in fact. Thus, it'll be lots of fun! From our derivation of the total differential, you can pretty much guess the correct form for the chain rule. For example, suppose that z is a function of two variables and that x and y are both functions of one variable, t. In other words, $x = x(t)$, $y = y(t)$, and $z = f(x, y) = f\left(x(t), y(t)\right)$, and we want to find $\frac{dz}{dt}$. Well, since the total differential is $dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$, a good (and correct!) guess for $\frac{dz}{dt}$ would be

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

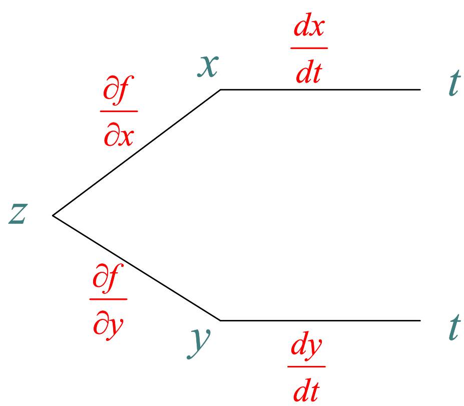
Let's go back to my favorite diagram and look at this in a little more detail.



Recall that from this diagram we derived the formula $\Delta z \approx \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$. Now just divide everything by Δt to get $\frac{\Delta z}{\Delta t} \approx \frac{\partial z}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial z}{\partial y}\frac{\Delta y}{\Delta t}$. And finally, take limits as $\Delta t \to 0$ and you get,

$$\frac{dz}{dt} = \lim_{\Delta t \to 0}\frac{\Delta z}{\Delta t} = \lim_{\Delta t \to 0}\left(\frac{\partial z}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial z}{\partial y}\frac{\Delta y}{\Delta t}\right) = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

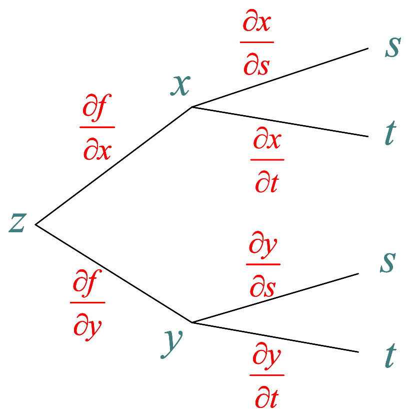
And that's our chain rule! There are a couple of additional points you need to be aware of, though. First, when do we use partial derivative notation and when do we use regular derivative notation? Well, the rule is that if a variable is a function of more than one input, then you use partial derivative notation, and if it's a function of a single input, then you use regular derivative notation. Up above, at first glance we have z as a function of two variables, x and y. However, each of these variables can be written as a function of just one variable, t, and so ultimately z is a function of a single input t. That's why we write $\frac{dz}{dt}$ as an ordinary derivative. We can also draw diagrams such as the one below to help us.



$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

To get the appropriate chain rule, we multiply along the branches and then add together all the results.

If we have different functions, however, then we might need to use different versions of the chain rule. For example, suppose we have $z = f(x, y)$, $x = x(s,t)$, and $y = y(s,t)$ and that we want to find the derivative of z with respect to t. Then this derivative will be a partial derivative since we won't be able to express z solely as a function of t. Our tree diagram for the chain rule in this case is as follows.



$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

Now let's suppose that $z = x^2 + y^2$, $x = t^2$, and $y = t^3$, and that we want to find $\frac{dz}{dt}$.

Then by the chain rule,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x)(2t) + (2y)(3t^2) = (2t^2)(2t) + (2t^3)(3t^2) = 4t^3 + 6t^5$$

A question you should be asking yourself now is, "Couldn't we have also done this just by first writing x and y in terms of t and then differentiating without having to use the chain rule?" The answer to this question is "yes" as you can see below.

$$z = x^2 + y^2 = (t^2)^2 + (t^3)^2 = t^4 + t^6$$

$$\frac{dz}{dt} = 4t^3 + 6t^5$$

However, the chain rule is also going to be important to us for the role that it plays in helping us prove some key theorems and to derive other important results. For example, suppose we have the equation $xy^3 + x^2 + x^3 y = 0$, and suppose that y is a function of x. Then we can use the chain rule to help us implicitly find the derivative of y with respect to x. First, if we think of the left-hand side of this equation as defining a function $f(x, y) = xy^3 + x^2 + x^3 y$, then the chain rule tells us that

$$\frac{df}{dx} = \frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}.$$

However, since the right-hand side of our original equation is equal to zero, we can set $\frac{\partial f}{\partial x} = 0$ to get,

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0.$$

From here it's a simple task to solve for $\frac{dy}{dx}$ to get,

$$\frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}$$

See how easy the chain rule made that derivation?

Another easy fact to now derive from the chain rule is the product rule for derivatives that everyone learns back in calculus of a single variable. The only thing we'll do different is that instead of thinking of $y = y(x) = f(x) \cdot g(x)$ as a function of one variable, we'll now think of it as a function of two, *f* and *g*. In other words, $y = y(f, g) = f \cdot g = f(x) \cdot g(x)$. A quick application of the chain rule for functions of several variables will immediately yield from this the familiar product rule.

$$\frac{dy}{dx} = \frac{\partial y}{\partial f}\frac{df}{dx} + \frac{\partial y}{\partial g}\frac{dg}{dx} = g\frac{df}{dx} + f\frac{dg}{dx} = f\frac{dg}{dx} + g\frac{df}{dx}.$$

Now we want to look at another useful tool that is known as the gradient. For a function $z = f(x, y)$ we define the gradient as $grad\ f = \nabla f = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j}$. Notice that the gradient is a vector that is formed using our first partial derivatives. Also, the notation "∇f" is generally read as "del f." Furthermore, if we have a function of three variables such as $w = f(x, y, z)$, then we would have $\nabla f = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial z}\hat{j}$. One of the important applications of the gradient vector is the following theorem.

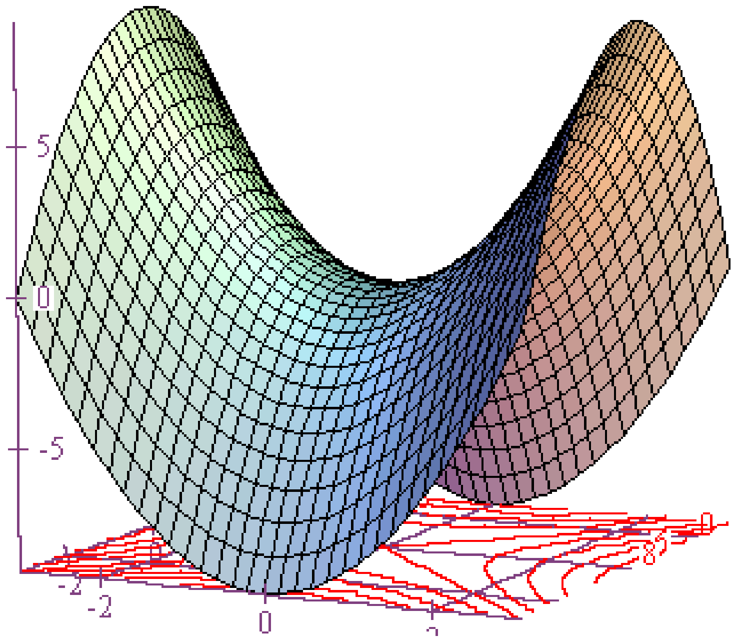
<u>Theorem:</u> Let $z = f(x, y)$ be differentiable at (a,b), and suppose $f(a,b) = c$. Also, let C be the level curve $f(x, y) = c$ that passes through (a,b) in the *xy*-plane. If C is smooth with a smooth parametrizatoin $\vec{r}(t)$ and if $\nabla f(a,b) \neq 0$, then $\nabla f(a,b)$ is normal to C at (a,b). In other words, ∇f is perpendicular to $\vec{r}(t)$ at (a,b).

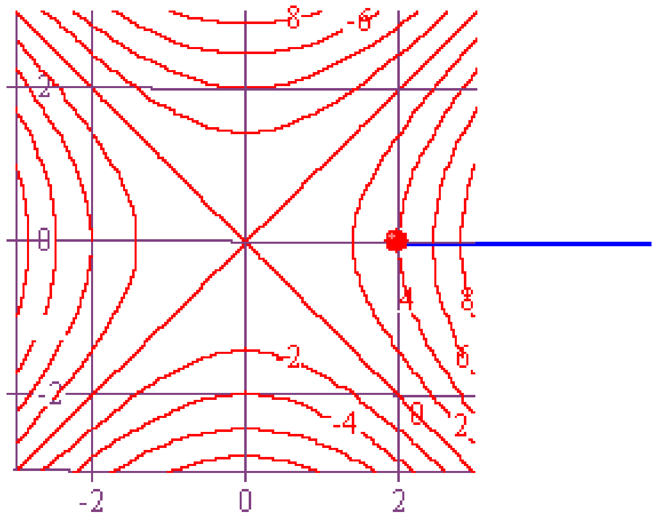
<u>Proof:</u> Let $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j}$ with $t \in I$, an interval, be a smooth parametrization for C. Then $f(x, y) = f\left(x(t), y(t)\right) = c$ when $t \in I$. Hence,

$$0 = \frac{dc}{dt} = \frac{df\left(x(t), y(t)\right)}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}$$

Therefore, ∇f is normal to C at (a,b). ■

Notice how we used the chain rule in the above proof. That's one of the reasons it's so important. Let's now do a construction to make this all a little more concrete. For our function, let's use $z = x^2 - y^2$, the level curve $z = 4$, and for input we'll use the point $P = \left(2,0\right)$. Then $\nabla z = \frac{\partial z}{\partial x}\hat{i} + \frac{\partial z}{\partial y}\hat{j} = 2x\hat{i} - 2y\hat{j}$, and $\nabla z(2,0) = 4\hat{i}$. To plot this vector as a displacement vector in the *xy*-plane starting at $\left(2,0\right)$, we can use the following parametric equations: $x = 2 + 4t,\ y = 0,\ 0 \leq t \leq 1$. And below we have the graph of our surface as well as the gradient vector at the point $\left(2,0\right)$ on the level curve corresponding to $z = 4$.


5
0
-5
-2
0


-8
-6
2
0
-2
-4
4
8
6
-2
0
2

Another success! It sure looks to me like the gradient is perpendicular to our level curve at the specified point.

Now here's something which is both a little different and very important. Let's do some constructions this time with the function $z = f(x,y) = -x^2 - y^2 + 4$ at the point $(1,2,-1)$. We looked at this function earlier in this chapter and found that the equation for the tangent plane at the point $(1,2,-1)$ was $z = -2x - 4y + 9$ where $-2 = \frac{\partial f(1,2)}{\partial x}$ and $-4 = \frac{\partial f(1,2)}{\partial y}$, the slopes of tangent lines of $z = f(x,y) = -x^2 - y^2 + 4$, respectively, in the directions of the positive *x*- and *y*-axis when evaluated at the point $(1,2,-1)$. If we now rewrite our equation as $0 = -2x - 4y - z + 9$, then from past discussions we know that the coefficients of *x*, *y*, and *z* give us a vector that is normal to this plane. In other words, the vector $\vec{v} = -2\hat{i} - 4\hat{j} - \hat{k}$ is perpendicular to the tangent plane to our surface $z = f(x,y) = -x^2 - y^2 + 4$ at the point $(1,2,-1)$. But notice this, if we set $w = -x^2 - y^2 - z + 4$, then the surface $0 = -x^2 - y^2 - z + 4$ is just a level surface for this function of three variables. Furthermore, the gradient of *w* is $\nabla w = -2x\hat{i} - 2y\hat{j} - \hat{k}$, and evaluated at the point $(1,2,-1)$ we get $\nabla w(1,2,-1) = -2\hat{i} - 4\hat{j} - \hat{k}$. But this is the same vector $\vec{v} = -2\hat{i} - 4\hat{j} - \hat{k}$ that is perpendicular to our tangent plane! In other words, we have the following general result.

<u>Theorem:</u> If $w = c$ is a level surface for the function $w = f(x, y, z)$, then the gradient of *w*, ∇w, evaluated at a point *P* on this level surface is perpendicular to the tangent plane at *P* on that surface.

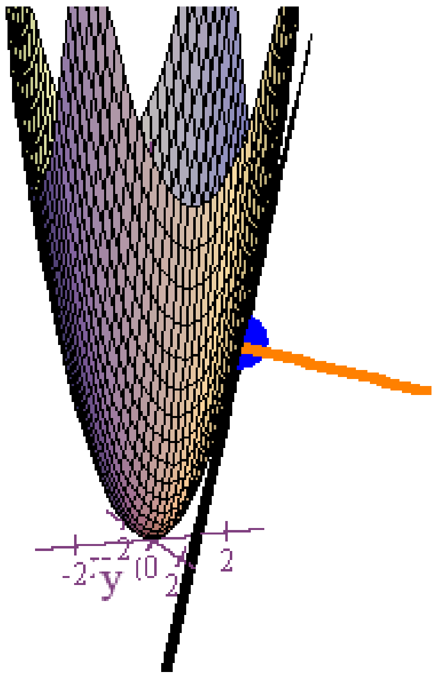
Let's do another construction as an example because this gives us yet another way to find a tangent plane. We can first find a gradient vector that is normal to the tangent plane, and then use the dot product to find the equation for the tangent plane. Here's how. Suppose $z = x^2 + y^2$ and $P = (1,2,5)$. Then *P* is a point on the surface of *z*. Notice, also, that we can write the equation for our surface as $0 = x^2 + y^2 - z$. Now let $w = x^2 + y^2 - z$. Then we can think of $0 = x^2 + y^2 - z$ as just a level surface for the function $w = x^2 + y^2 - z$. Furthermore, $\nabla w = 2x\hat{i} + 2y\hat{j} - \hat{k}$, and hence, $\nabla w(1,2,5) = 2\hat{i} + 4\hat{j} - \hat{k}$ is normal to our surface at the point *P*. Now consider the tangent plane to our surface at the point $P = (1,2,5)$, and suppose $Q = (x, y, z)$ is another point in that plane. Then since $\nabla w(1,2,5) = 2\hat{i} + 4\hat{j} - \hat{k}$ is perpendicular to the plane, it is also perpendicular to the displacement vector $\overrightarrow{PQ} = (x-1)\hat{i} + (y-2)\hat{j} + (z-5)\hat{k}$ Therefore, $\nabla w(1,2,5) \bullet \overrightarrow{PQ} = 0$. But this gives us

$$\begin{aligned} 0 &= \left(2\hat{i} + 4\hat{j} - \hat{k}\right) \bullet \left((x-1)\hat{i} + (y-2)\hat{j} + (z-5)\hat{k}\right) = 2(x-1) + 4(y-2) - (z-5) \\ &= 2x + 4y - z - 5 \end{aligned}$$

In other words, $z = 2x + 4y - 5$ is the tangent plane to our surface at $P = (1,2,5)$.

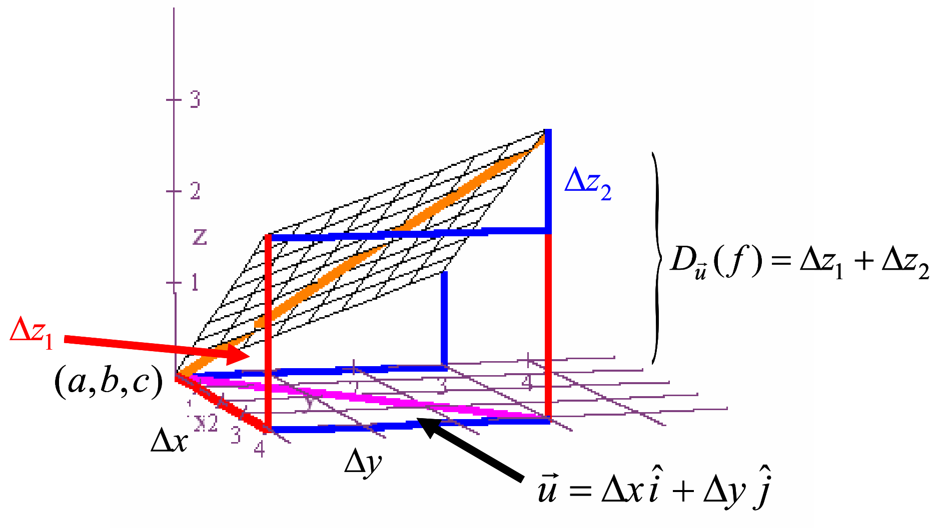
Now let's graph everything. Our surface is $z = x^2 + y^2$, our tangent plane is $z = 2x + 4y - 5$, and the parametric equations for graphing our gradient vector at the point $P = (1,2,5)$ are,

$$\begin{aligned} x &= 1 + 2t \\ y &= 2 + 4t \\ z &= 5 - t \\ 0 &\le t \le 1 \end{aligned}$$



And there we have it! Level surface, tangent plane, point, and gradient vector that is perpendicular to the level surface.

So far we've looked only at the slopes of tangent lines in the directions of the positive x-axis and the positive y-axis. These values correspond to $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. However, suppose we want to find the slope of a tangent line in some other direction. If we do, then we'll refer to this slope as a directional derivative. Furthermore, if a tangent plane contains all possible tangent lines to a surface at a point, and if the tangent plane can be determined from the values of $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ at that point, then we should be able to find the value of a directional derivative directly from the values of $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. And that's exactly what's going to happen. Let's start by looking at yet another variation of my favorite diagram.



In this diagram, we are looking at a tangent plane to some surface at the point (a,b,c), and the orange line represents a tangent line in the direction of a unit vector $\vec{u} = \Delta x\hat{i} + \Delta y\hat{j}$ that is shown in magenta above. Since the length of $\vec{u} = \Delta x\hat{i} + \Delta y\hat{j}$ is 1, it follows that the slope of our tangent line is given just by the height of the right triangle depicted with $\vec{u} = \Delta x\hat{i} + \Delta y\hat{j}$ as the base. In other words, by $\Delta z_1 + \Delta z_2$. But as we've seen previously with this diagram, $\Delta z_1 = \frac{\partial z(a,b)}{\partial x}\Delta x$ and $\Delta z_2 = \frac{\partial z(a,b)}{\partial y}\Delta y$.

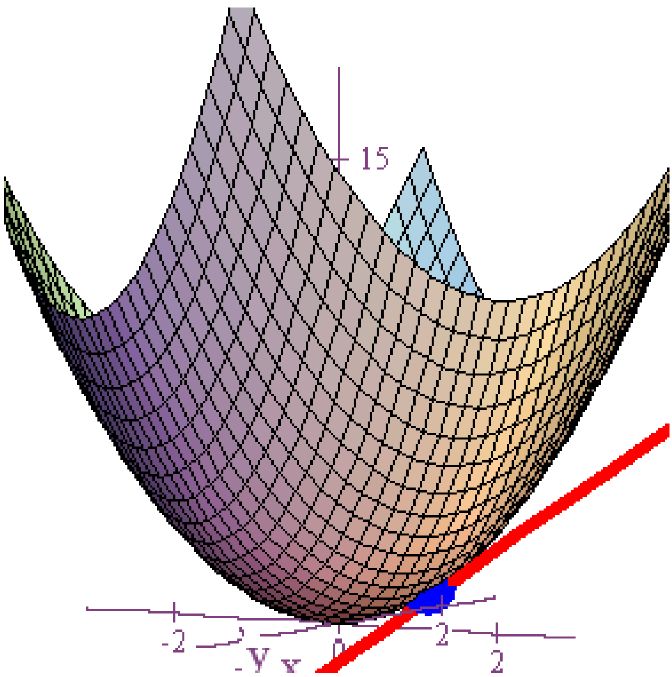
Therefore, the value of the directional derivative of $z = f(x,y)$ in the direction of $\vec{u}$ is

$$D_{\vec{u}}(f) = \Delta z_1 + \Delta z_2 = \frac{\partial f(a,b)}{\partial x}\Delta x + \frac{\partial f(a,b)}{\partial y}\Delta y = \nabla f(a,b)\bullet\vec{u}$$

We usually abbreviate this formula as $D_{\vec{u}}(f) = \nabla f \bullet \vec{u}$. For example, if $z = f(x,y) = x^2 + y^2$ and $\vec{u} = \frac{\sqrt{2}}{2}\vec{i} + \frac{\sqrt{2}}{2}\hat{j}$, then $D_{\vec{u}}(f) = \nabla f \bullet \vec{u} = x\sqrt{2} + y\sqrt{2}$. If we now evaluate this derivative at the point $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, the we get $D_{\vec{u}} f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \nabla f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\bullet\vec{u} = 1 + 1 = 2$. Thus, at the point $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1\right)$ on the surface of $z = f(x,y) = x^2 + y^2$, the tangent line in the direction of $\vec{u} = \frac{\sqrt{2}}{2}\vec{i} + \frac{\sqrt{2}}{2}\hat{j}$ has slope 2. If we now want to add the graph of this tangent line to our surface graph, it's not too difficult if you think in terms of vectors. Our point $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1\right)$ on the surface corresponds to the position vector $\vec{w} = \frac{1}{\sqrt{2}}\hat{i} + \frac{1}{\sqrt{2}}\hat{j} + \hat{k}$. Since our tangent line has slope 2, if we add our unit vector $\vec{u} = \frac{\sqrt{2}}{2}\vec{i} + \frac{\sqrt{2}}{2}\hat{j}$ to this position vector and then go up 2 units by adding on the vector $\vec{v}=2\hat{k}$, then everything should terminate at another point on the tangent line. In other words, the vector $\vec{u}+\vec{v}$ is parallel to our tangent line. Hence, we can describe the tangent line parametrically as

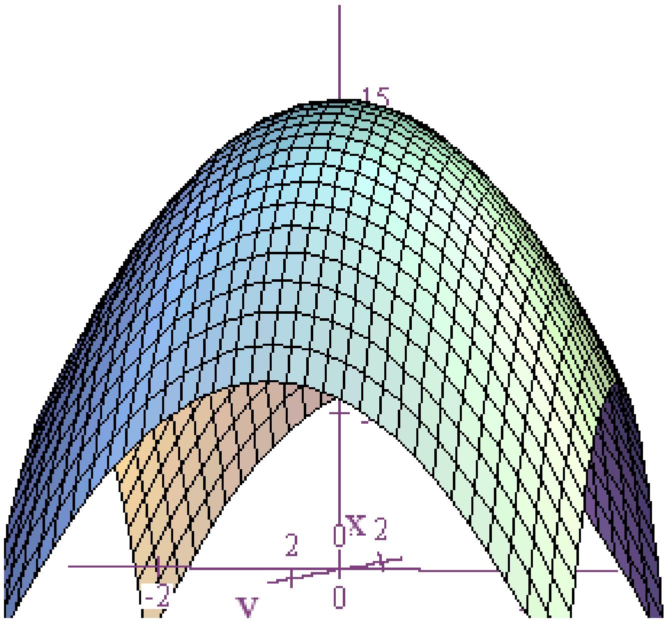
$\vec{w}+t(\vec{u}+\vec{v})=\left(\frac{1}{\sqrt{2}}i+\frac{1}{\sqrt{2}}\hat{j}+\hat{k}\right)+t\left(\frac{\sqrt{2}}{2}\hat{i}+\frac{\sqrt{2}}{2}\hat{j}+2\hat{k}\right)$. This, in turn, gives us the following values for x, y, and z.

$$
\begin{aligned}
x &= \frac{1}{\sqrt{2}}+\frac{t}{\sqrt{2}} \\
y &= \frac{1}{\sqrt{2}}+\frac{t}{\sqrt{2}} \\
z &= 1+2t \\
-\infty &< t < \infty
\end{aligned}
$$

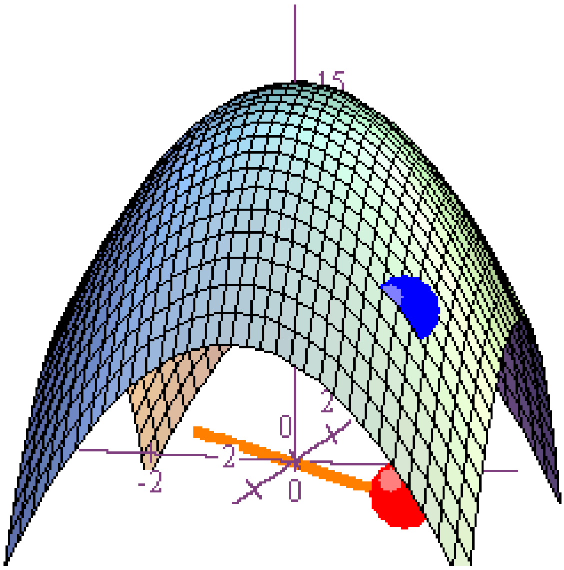


WOW! It actually worked!

If we go back to the formula we derived for computing the directional derivative as a dot product, $D_{\vec{u}}(f) = \nabla f \bullet \vec{u}$, then there are even more wonderful things we can deduce. For instance, using our alternate formula for computing the dot product, we have $D_{\vec{u}}(f) = \nabla f \bullet \vec{u} = \|\nabla f\|\|\vec{u}\|\cos\theta = \|\nabla f\| \cdot 1 \cdot \cos\theta = \|\nabla f\|\cos\theta$. Among other things, this formula tells us that the directional derivative will take on its maximum positive value at a point on the surface when the angle between the gradient vector and the unit vector is zero, and this derivative will take on it's negative value of largest magnitude when the angle between the gradient vector and the unit vector is $180°$ (or π). In other words, if you are trying to climb a hill, then the gradient vector points in the direction of steepest ascent. I must confess that this really confused me when I first heard it mentioned decades ago in multivariable calculus because I thought to myself, "If I want to go up the hill as fast as possible, shouldn't I go up in the direction of 'up'?" Well, yes, that's true. If you want to go up, then go up. However, the type of direction we are speaking of with regard to the gradient vector is more like a compass direction. Consequently, if you are facing north and there is a mountain in front of you, then the direction you go in to quickly ascend the mountain is north, not up. And if you want to descend the mountain as quickly as possible, then you go in the opposite direction, south. As an example, let's look at the mountain $z = -x^2 - y^2 + 15$ below.



The top of this mountain is at $(0,0,15)$, and if we want to ascend the mountain as quickly as possible, then our direction in the plane should always be pointing towards the *z*-axis. Now let's find the gradient vector and evaluate it at the point $(2,2,7)$. We have that $\nabla z = -2x\hat{i} - 2y\hat{j}$ and $\nabla z(2,2,) = -4\hat{i} - 4\hat{j}$. In the graph below, we've plotted a blue dot at $(2,2,7)$ and a red dot right below it at $(2,2,0)$. Furthermore, we've attached our gradient vector to the red dot, and sure enough, it's pointing right toward the *z*-axis. So as we predicted, if we want to ascend this hill as quickly as possible, then our compass direction should always be towards the *z*-axis.



At this point we've talked a lot about derivatives of $z = f(x, y)$ with respect to x and with respect to y, but if you think back to your first calculus course then you know that we didn't stop there. After all, once we've differentiated a function once, what could be more fun than differentiating it again! For a function of the form $z = f(x, y)$, there are two first partial derivatives we can compute. How many second partial derivatives, however, are possible? A moment's reflection should tell us that there are four second partial derivatives. This, of course, is because if you differentiate with respect to x the first time, then you can differentiate with respect to y or x the second time. Likewise, if you differentiate with respect to y the first time, then you can differentiate with respect to either variable the second time. Thus, for $z = f(x, y)$, there are four second partial derivatives that are available.

What notation do we use for the second partial derivatives? Well, just as we had two different notations for first partial derivatives, such as $\frac{\partial z}{\partial x}$ and z_x, we also have two notations for the second partial derivatives. We write the four second partial derivatives either as $\frac{\partial^2 z}{\partial x^2}$, $\frac{\partial^2 z}{\partial y^2}$, $\frac{\partial^2 z}{\partial y \partial x}$, or $\frac{\partial^2 z}{\partial x \partial y}$, or we write them z_{xx}, z_{yy}, z_{xy}, or z_{yx}. The latter two derivatives in each list are called "mixed partials" because we differentiate first with respect to one variable and then the next. However, there is an order difference you have to be aware of. In the notation $\frac{\partial^2 z}{\partial y \partial x}$, we take our derivatives from right to left. In other words, $\frac{\partial^2 z}{\partial y \partial x}$ means that you first differentiate with respect to *x* and then with respect to *y*. This is because the notation $\frac{\partial^2 z}{\partial y \partial x}$ is just shorthand for $\frac{\partial\left(\frac{\partial z}{\partial x}\right)}{\partial y}$. On the other hand, when we write z_{yx}, we perform the operations in order from left to right. Thus, this one means that we should first differentiate with respect to *y* and then with respect to *x*. In summary, $\frac{\partial^2 z}{\partial y \partial x} = z_{xy}$,

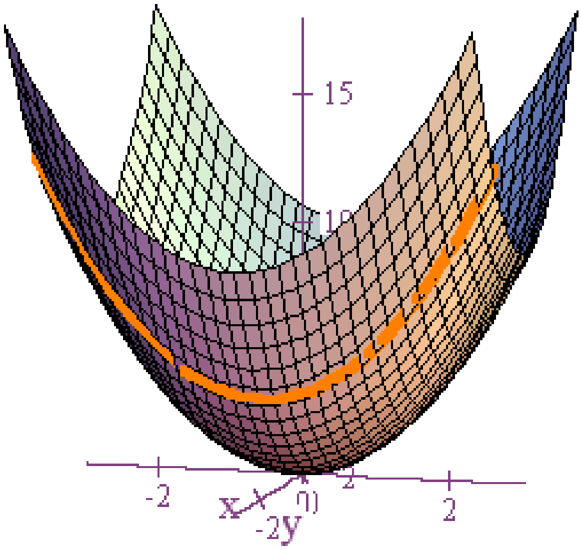
$\frac{\partial^2 z}{\partial x \partial y} = z_{yx}$, $\frac{\partial^2 z}{\partial x^2} = z_{xx}$, and $\frac{\partial^2 z}{\partial y^2} = z_{yy}$.

Let's now take a simple example. If $z = x^2 + y^2$, then $z_x = 2x$ and $z_y = 2y$. If we now compute our second partials, we get $z_{xx} = 2$, $z_{xy} = 0$, $z_{yx} = 0$, and $z_{yy} = 2$. For clarity, we often like to arrange these second partial derivatives in a matrix where they always appear in the following order.

$$\begin{pmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

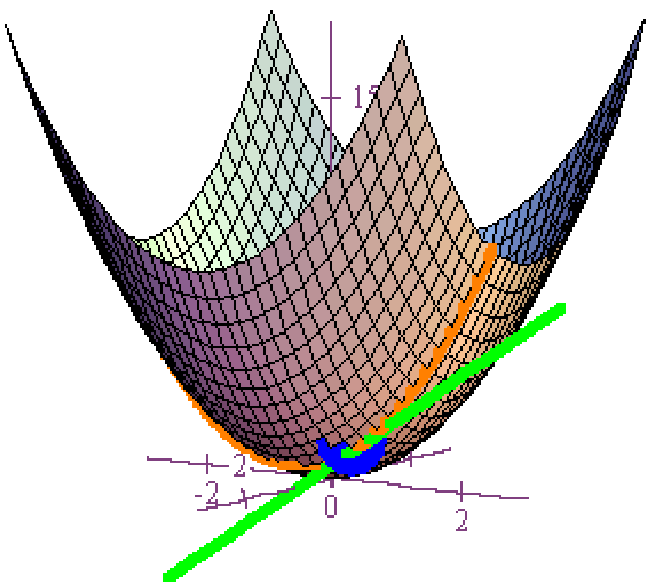
Notice in this example that $z_{xx} = z_{yy}$ and $z_{xy} = z_{yx}$. The first equality is simply by accident, but the second one, $z_{xy} = z_{yx}$, happens almost all the time. In fact, there is a theorem that says that if the mixed partials are continuous at a point in the interior of the domain of our function, then the mixed partials will be equal at that point. And since in calculus, we tend to deal with functions that are continuous almost everywhere, we tend to almost always have equality of the mixed partials. Thus, we can use this property as a check to make sure that we've computed our mixed partials correctly.

How do we interpret second partial derivatives? Again, pretty much the same as an ordinary derivative. Just as in first semester calculus, z_x tells us the rate at which z is changing with respect to x, and if we take a cross-section by fixing a value for y, then for that fixed y value, z_x will tell us over what intervals, with respect to x, z is increasing and where it is decreasing. If we continue on to the second partial z_{xx}, then this will tell us something about the concavity of our cross-section. For example, with $z = x^2 + y^2$, we had $z_x = 2x$. This tells us that regardless of what we set y equal to, the z values in the corresponding cross-section will be decreasing when x is negative, increasing when x is positive, and we'll have a horizontal tangent line when $x = 0$. Differentiating with respect to x a second time gives us $z_{xx} = 2$. This is now telling us that now matter what value we fix y at, the cross-section will be concave-up. Below we see a cross-section corresponding to $y = -2$, and as predicted, the graph of the cross-section is decreasing for $x < 0$ and increasing for $x > 0$, and for all values of x, the graph is concave-up.

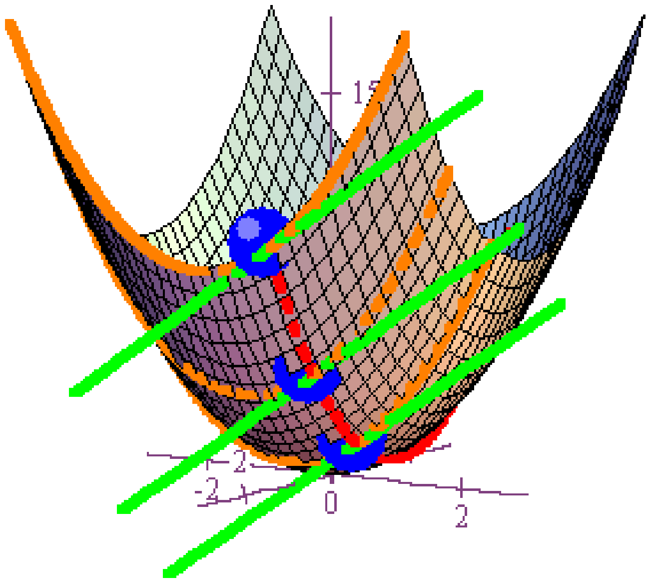


In a similar manner, z_{yy} will tell us something about the concavity of a cross-section obtained by fixing a value of x.

Now how do we interpret the mixed partials z_{xy} and z_{yx}. On the one hand, z_{xy} tells us the rate at which $\frac{\partial z}{\partial x}$ changes with respect to a change in y, and z_{yx} tells us the rate at which $\frac{\partial z}{\partial y}$ changes with respect to a change in x. That part is clear simply from our understanding that derivatives are always instantaneous rates of change. But on the other hand, these mixed partials are much harder to visualize geometrically. In the case of $z = x^2 + y^2$, it's pretty easy. For this function, we have $z_{xy} = 0 = z_{yx}$. This means that if we first find z_x and then look to see how this rate changes as we let y change, we discover that it doesn't change at all. The rate of change is zero. For example, let's take $z = x^2 + y^2$, fix things at a point $P = (1,1,2)$ on the graph, and look at the cross-section corresponding to $x = 1$. Below is the graph of our surface along with the cross-section, and the tangent line at the point $P = (1,1,2)$.

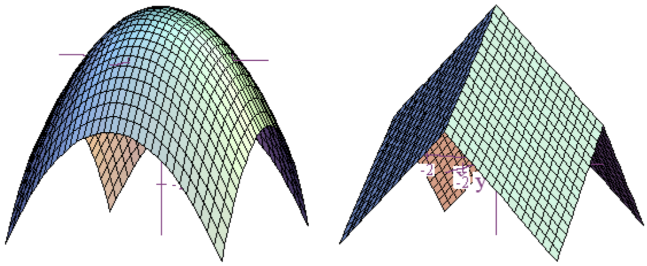


If we now repeat this construction at the fixed values $x = 2$ and $x = 3$, we'll see that the slopes of the tangent lines corresponding to $y = 1$ don't change. And this happens because $z_{yx} = 0$. In other words, the slopes of the tangent lines in the direction of the y-axis don't change as x changes, the rate of change with respect to x is zero. Here's a graph of what we're talking about.

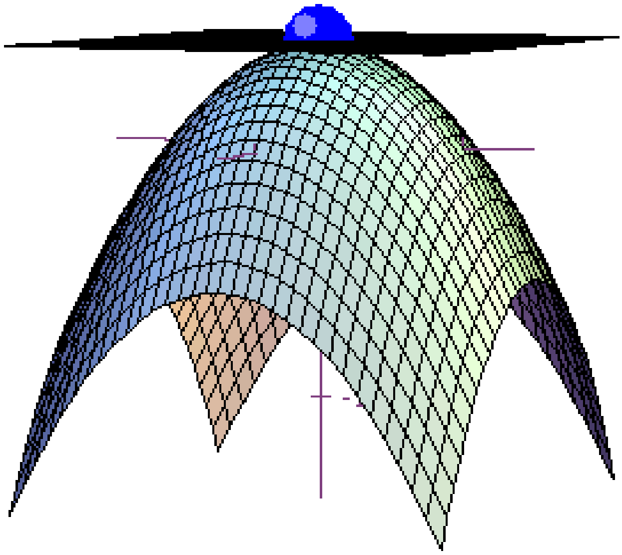


The red curve above corresponds to the cross-section $y = 1$. As we let x vary along this curve, the slopes of the tangent lines in the direction of the y-axis remain unchanged. In other words, their rate of change is zero. This is what $z_{yx} = 0$ is trying to tell us. If this is hard to visualize, don't worry to much about it. It takes practice. Furthermore, the interpretations of z_{xx} and z_{yy} as indicators of concavity will be far more important to us in the long run.

It's finally time for us to discuss the main application of derivatives of functions of several variables. Namely, how do we use these derivatives to find extreme values. Let's revisit two examples we looked at earlier,



The surface on the left is the graph of $z=-\left(x^2+y^2\right)+4$, and the surface on the right is the graph of $z=-\left(|x|+|y|\right)+4$. It's clear that both of these surfaces have a maximum point at $(0,0,4)$. However, it's also clear that we can't define tangent lines at this point on the second surface since we have a sharp corner at that location. Consequently, neither $\frac{\partial z}{\partial x}$ nor $\frac{\partial z}{\partial y}$ exist at that point. On the other hand, both partial derivatives exist at the point $(0,0,4)$ on the graph of $z=-\left(x^2+y^2\right)+4$, and both of our first partial derivatives will be zero at this point resulting in horizontal tangent lines that subsequently give rise to a horizontal tangent plane.



What we are seeing here is similar to what you saw in your first calculus course, and we now summarize the results below using the concept of the partial derivative.

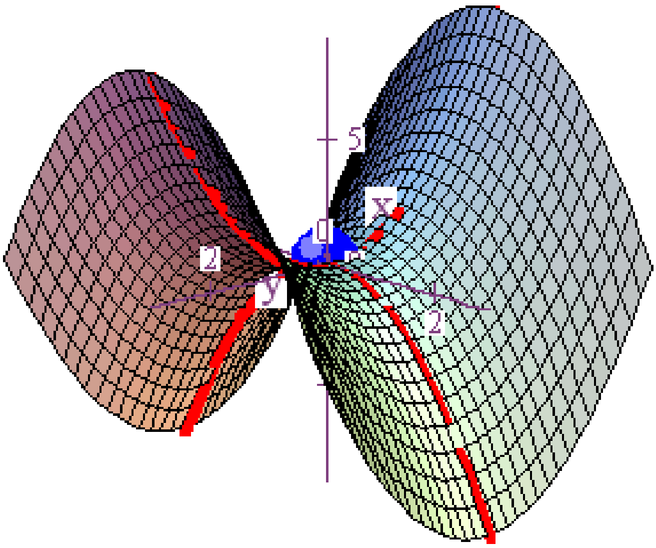
Definition: Let (a,b) be a point contained in an open region R on which a function $z = f(x,y)$ is defined. Then (a,b) is a critical point if either of the following conditions is true:

1. $z_x(a,b) = 0 = z_y(a,b)$
2. $z_x(a,b)$ does not exist
3. $z_y(a,b)$ does not exist.

<u>Theorem:</u> If $z = f(x, y)$ has a relative maximum or a relative minimum at a point (a,b) contained within an open region *R* on which $z = f(x, y)$ is defined, then (a,b) is a critical point.

In other words, critical points are going to be points where either both first derivatives are zero or else one of the first derivatives fails to exist, and if we have a local extreme value, then it has to occur at a critical point. Before we go over a test for helping us classify what type of critical point we have, let's look at a very special kind of point called a <u>saddle point</u>. The classic example of this is found in the graph of $z = x^2 - y^2$. For this function, we have $z_x = 2x$ and $z_y = -2y$. Furthermore,

$\left.\begin{matrix} z_x = 0 \\ z_y = 0 \end{matrix}\right\} \Rightarrow \begin{matrix} 2x = 0 \\ -2y = 0 \end{matrix} \Rightarrow \begin{matrix} x = 0 \\ y = 0 \end{matrix}$. Thus, $(0,0)$ is a critical point. However, if we look at the corresponding point, $(0,0,0)$, on our surface graph, then we easily see that it's neither a maximum nor a minimum point.



Instead, what we see is that this point is at the bottom of one parabolic cross-section and at the top of another parabolic cross-section. When we have a point like this which is a critical point, but movement in one direction causes z to increase while movement in another direction causes z to decrease, then we call the point on our surface a saddle point. The terminology obviously derives from the saddle shape of the graph above. Thus, when we try to classify the points on our surface corresponding to critical points (a,b), sometimes they will be relative maximum points, sometimes they will be relative minimum points, and sometimes the will be saddle points.

Below is a simple test that will allow us to classify most critical points we come across. The proof of this theorem is rather messy, and it is usually left out of calculus books. However, we'll give a proof at the end of this chapter. Nonetheless, feel free to skip it and move on if you wish.

<u>Second Partials Test:</u> Suppose $z = f(x,y)$ has continuous second partial derivatives on an open region containing a point (a,b) such that $z_x(a,b) = 0 = z_y(a,b)$, and let

$$D = \begin{vmatrix} z_{xx}(a,b) & z_{xy}(a,b) \\ z_{yx}(a,b) & z_{yy}(a,b) \end{vmatrix} = z_{xx}(a,b)z_{yy}(a,b) - z_{xy}(a,b)z_{yx}(a,b).$$

Then:

1. If $D > 0$ and $z_{xx}(a,b) > 0$, $f(a,b)$ is a relative minimum.
2. If $D > 0$ and $z_{xx}(a,b) < 0$, $f(a,b)$ is a relative maximum.
3. If $D < 0$, $(a,b,f(a,b))$ is a saddle point.
4. If $D = 0$, the test is inconclusive.

Let's start with a very simple example, $z = f(x,y) = x^2 + y^2$. If we take the first partial derivatives, we get $z_x = 2x$ and $z_y = 2y$. Setting both of these equal to zero and solving for *x* and *y* yields,

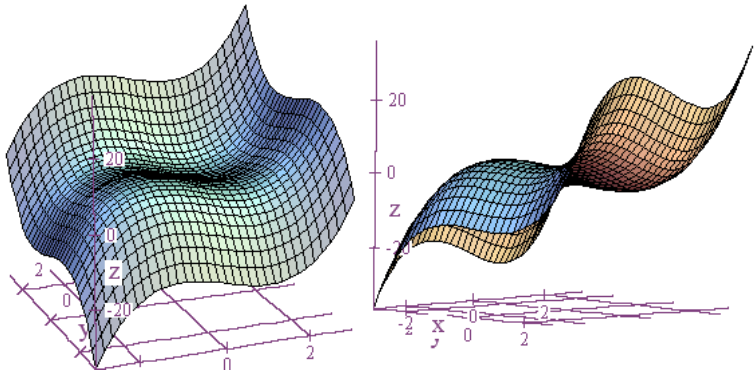
$$\begin{matrix} 2x = 0 \\ 2y = 0 \end{matrix} \Rightarrow \begin{matrix} x = 0 \\ y = 0 \end{matrix}$$

Therefore, $(0,0)$ is our critical point. We now need to find our second partials matrix and evaluate its determinant at the critical point.

$$\begin{pmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow D(0,0) = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = (2)(2) - (0)(0) = 4 > 0$$

There are two things to notice here. First, in our second partials matrix we found $z_{xy} = 0 = z_{yx}$. Since we expect to find $z_{xy} = z_{yx}$, we're probably right on track. Additionally, if these two second mixed partials were not equal to each other, then we should suspect that we've made an error. The second important thing to notice is that $D(0,0) = 4 > 0$. This automatically means that we have either a local maximum or a local minimum. To determine which, our second partials test tells us to look at the sign of z_{xx} at our critical point. In this case, we have $z_{xx}(0,0) = 2 > 0$. We can interpret this second derivative as meaning that a particular cross-section of our surface is concave-up at the critical point. Therefore, the critical point is at the bottom of this cross-section, and the point $(0,0,f(0,0)) = (0,0,0)$ is a local minimum point.

Now let's look at $z = f(x,y) = x^3 - 3x + y^3 - 3y$. The graph below suggests that there is one local maximum, one local minimum, and two saddle points.



We'll find all the critical points, but apply our second partials test to just one of them as an example. If we take our first partial derivatives, we get,

$$z_x = 3x^2 - 3 = 3\left(x^2 - 1\right) = 3(x+1)(x-1)$$
$$z_y = 3y^2 - 3 = 3\left(y^2 - 1\right) = 3(y+1)(y-1)$$

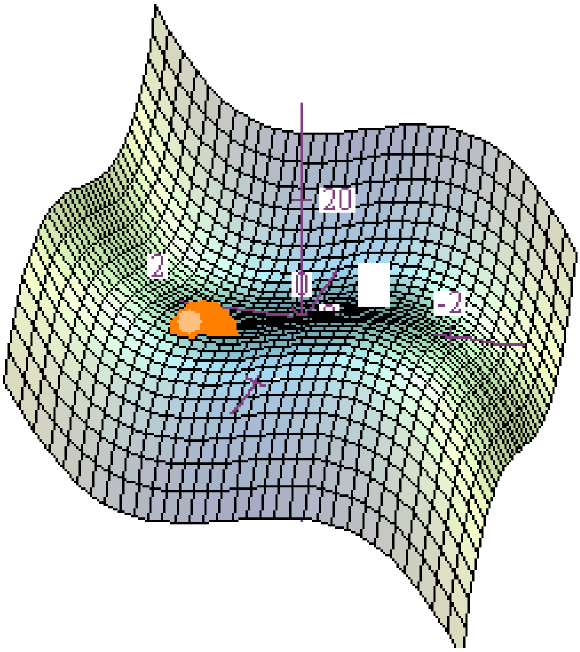
Setting each partial derivative equal to zero results in,

$$\left.\begin{array}{l} z_x = 0 \\ z_y = 0 \end{array}\right\} \Rightarrow \begin{array}{l} 3(x+1)(x-1) = 0 \\ 3(y+1)(y-1) = 0 \end{array} \Rightarrow \begin{array}{l} x = -1,1 \\ y = -1,1 \end{array}$$

Therefore, our critical points are $(-1,-1)$, $(-1,1)$, $(1,-1)$, and $(1,1)$. Let's do our second partials test just at the point $(-1,1)$ to see what happens. Setting up our second partials matrix and computing the value of $D(-1,1)$, we get,

$$\begin{pmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{pmatrix} = \begin{pmatrix} 6x & 0 \\ 0 & 6y \end{pmatrix} \Rightarrow D(-1,1) = \begin{vmatrix} -6 & 0 \\ 0 & 6 \end{vmatrix} = (-6)(6) - (0)(0) = -36 < 0$$

And this is as far as we need to go. Since $D(-1,1) < 0$, we automatically know from our second partials test that $(-1,1,f(-1,1)) = (-1,1,0)$ is a saddle point. Also, a quick look at this point on our graph confirms this conclusion.



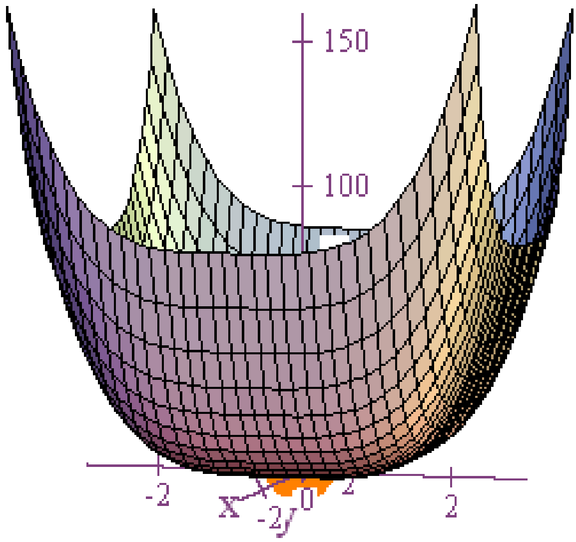
Now let's look at just one more function, $z = f(x,y) = x^4 + y^4$. Taking first partial derivatives, we have $z_x = 4x^3$ and $z_y = 4y^3$. Hence, our critical point is $(0,0)$ since,

$$\left.\begin{matrix} z_x = 0 \\ z_y = 0 \end{matrix}\right\} \Rightarrow \begin{matrix} 4x^3 = 0 \\ 4y^3 = 0 \end{matrix} \Rightarrow \begin{matrix} x = 0 \\ y = 0 \end{matrix}$$

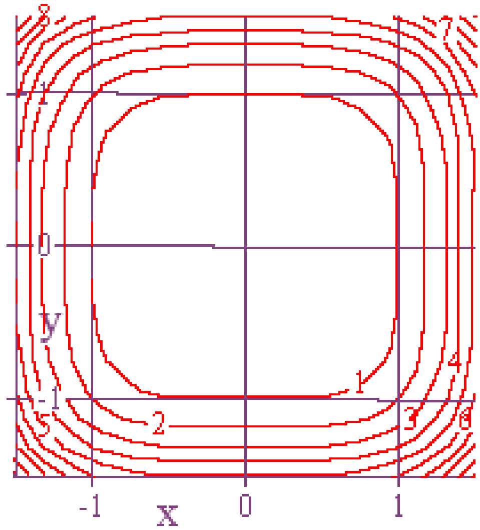
If we now perform the second partials test, we get,

$$\begin{pmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{pmatrix} = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix} \Rightarrow D(0,0) = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = (0)(0) - (0)(0) = 0$$

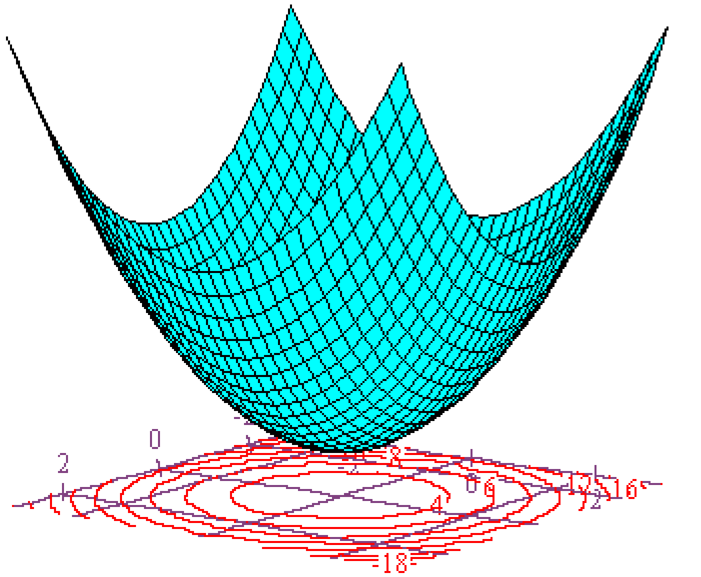
Unfortunately, our second partials test tells us that the test is inconclusive if $D = 0$. However, a glance at the graph will confirm that the point $(0,0,0)$ is a local minimum point.



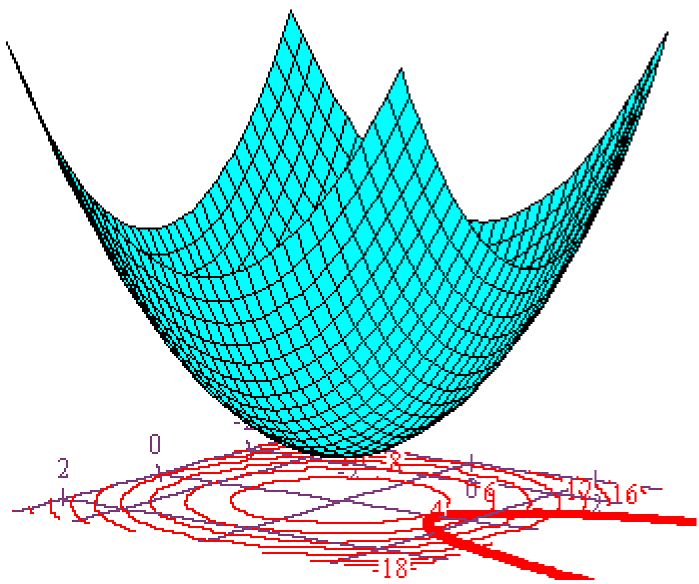
We can also get confirmation by looking at the level curves for this function. Below you can easily see the wagons circling tighter and tighter around the point $(0,0)$, and the *z*-values decrease as we approach the origin.



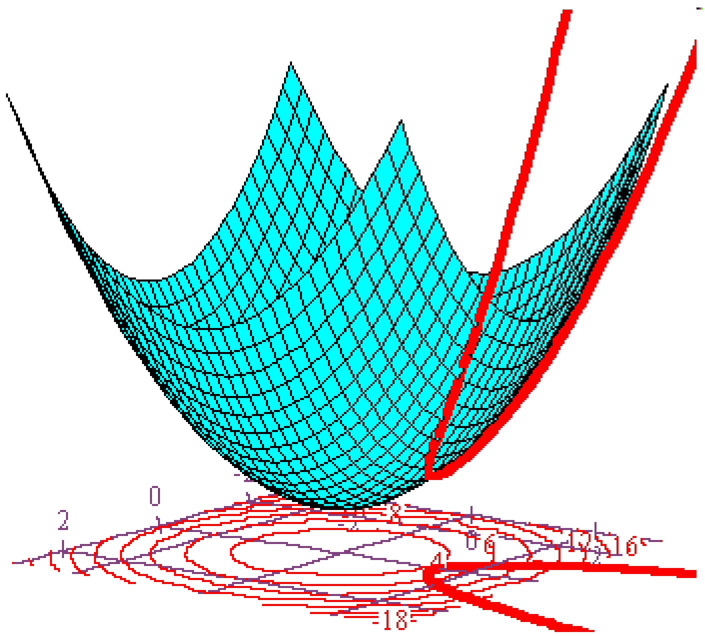
We now want to move on to another type of optimization problem, problems where are input values are restricted to some type of constraint curve. These problems will involve a technique known as *Lagrange multipliers*. To begin, consider the graph of a function $z = f(x, y)$ below .



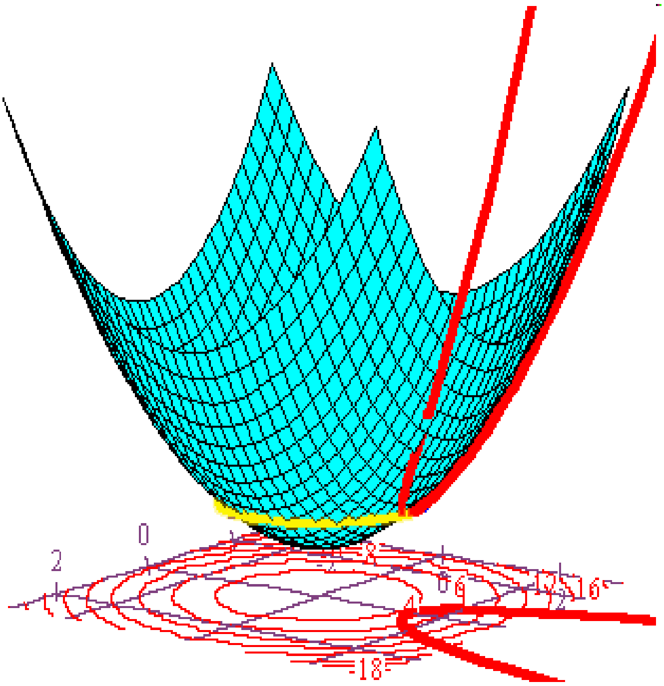
In looking at this graph, it appears that it is defined for all values of *x* and *y*, and that a local (and absolute) minimum will occur at the origin. Now let's add a curve to the *xy*-plane.



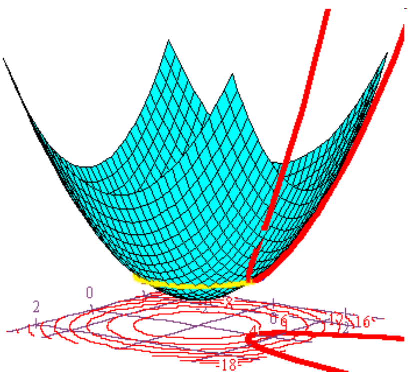
If we restrict our input values just to the points on this curve, then, graphically, the result will be a corresponding curve that lies on our surface.

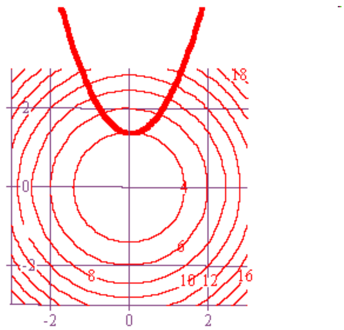


We can see that, with this restricted input, we now have our minimum point residing at a different location. However, what is also (hopefully) obvious is that there is going to be a contour curve on the surface that touches our surface curve right at this minimum point.

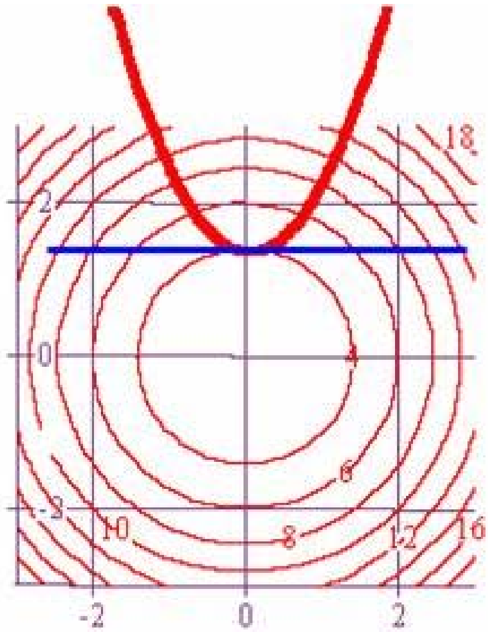


If we now move this contour curve down to the *xy*-plane, then it will also touch our constraint curve at a single point.





In other words, the constraint curve in the *xy*-plane and the level curve in the *xy*-plane are tangent to one another, and thus, they have a common tangent line.



Now, what does this all mean to us? It means this. Suppose our we consider our constraint curve as just a level curve for some other function $g(x, y)$. Then as we can see above, there is a level curve for $f(x, y)$ such that $f(x, y)$ and $g(x, y)$ have a common tangent line at the point in the *xy*-plan that corresponds to the minimum point of the curve on our surface graph. However, recall that a gradient vector evaluated at a point is normal to the level curve at that point. Hence, it follows that, at this point of tangency, ∇f is parallel to ∇g. Recall, too, that two vectors are parallel if and only if one is a scalar multiple of the other. If we denote this scalar by the Greek letter lambda, λ, then we get $\nabla f = \lambda \nabla g$. At the partial derivative level we write this as $f_x = \lambda g_x$ and $f_y = \lambda g_y$. This number λ is what is known as a *Lagrange* *multiplier*, named after the great mathematician Joseph Lagrange who discovered this technique. At this point, though, if we add the constraint equation, written as $g(x, y) = c$, to the other two, then we arrive at the following system that we need to solve for *x*, *y*, and λ in order to find our extreme point. In other words, solve the system below, and you're done!

$$\begin{aligned} f_x &= \lambda g_x \\ f_y &= \lambda g_y \\ g(x, y) &= c \end{aligned}$$

Of course, solving this system is sometimes easier said than done, but let's take a simple example. Suppose $z = f(x, y) = x^2 + y^2$ and our constraint curve is the graph of $y - x^2 = 4$. Then we can think of our constraint curve as a level curve for the function $g(x, y) = y - x^2$. We now have the following equations to play around with.

$$\left.\begin{aligned} f_x &= \lambda g_x \\ f_y &= \lambda g_y \\ g(x, y) &= c \end{aligned}\right\} \Rightarrow \begin{aligned} 2x &= -\lambda 2x \\ 2y &= \lambda \\ y - x^2 &= 4 \end{aligned}$$

From the first equation we get that either $x = 0$ or $\lambda = -1$. If $x = 0$, then substitution into the last equation tells us that $y = 4$. On the other hand, if $\lambda = -1$, then substitution into the second equation tells us that $y = -\frac{1}{2}$. If we plug this value into the third equation, we get $-\frac{1}{2} - x^2 = 4 \Rightarrow -x^2 = \frac{9}{2} \Rightarrow x^2 = -\frac{9}{2}$. This, however, can't happen since for any real number $x^2 \geq 0$. Hence, our minimum value occurs when

$x = 0$ and $y = 4$, and the coordinates of our minimum point are $(0,4,16)$. And it's just that simple! Of course, once you understand the method of Lagrange multipliers, the method is simple, but it's solving the actual equations that often times gets very difficult. The problem is that our equations are often nonlinear, and methods that work well on one problem may not help at all on another.

Here's an example that involves a function of three variables. The method of Lagrange multipliers is, nonetheless, the same. We just have to set up an additional equation for the extra variable.

Problem: Use Lagrange multipliers to find the minimum distance between the point $(1,2,3)$ and the plane $4x + 5y + 6z = 20$.

Solution: We'll let the equation for our plane, $4x + 5y + 6z = 20$, be our constraint, and we'll set $z = f(x, y) = (x-1)^2 + (y-2)^2 + (z-3)^2$. Basically, we're setting $f(x, y)$ equal to the square of the distance between the point $(1,2,3)$ and another point (x, y, z). Certainly, if we can minimize the square of the distance, then we've also found the solution to the minimal distance. However, by working with the square of the distance, the derivative process will be simpler. And finally, we'll set

$g(x, y) = 4x + 5y + 6z$ so that $4x + 5y + 6z = 20$ can be thought of as a level surface for this function. Now let's find some derivatives!

$$f_x = 2(x-1) = 2x-2$$
$$g_x = 4$$

$$f_y = 2(y-2) = 2y-4$$
$$g_y = 5$$

$$f_z = 2(z-3) = 2z-6$$
$$g_z = 6$$

From these derivatives, we apply lagrange multipliers to get the following equations,

$$2x-2 = 4\lambda$$
$$2y-4 = 5\lambda$$
$$2z-6 = 6\lambda$$

If we solve these equations for *x*, *y*, and *z*, then we get,

$$x = \frac{4\lambda+2}{2}$$
$$y = \frac{5\lambda+4}{2}$$
$$z = \frac{6\lambda+6}{2}$$

We can now substitute these expressions for our variables in our constraint equation, and this will allow us to solve for λ.

$$4x+5y+6z=4\left(\frac{4\lambda+2}{2}\right)+5\left(\frac{5\lambda+4}{2}\right)+6\left(\frac{6\lambda+6}{2}\right)=20$$
$$\Rightarrow 16\lambda+8+25\lambda+20+36\lambda+36=40$$
$$\Rightarrow 77\lambda=-24$$
$$\Rightarrow \lambda=-\frac{24}{77}$$

Hence,

$$x=\frac{4\lambda+2}{2}=\frac{59}{77}$$
$$y=\frac{5\lambda+4}{2}=\frac{94}{77}$$
$$z=\frac{6\lambda+6}{2}=\frac{159}{77}$$

Thus, the point $\left(\frac{29}{77},\frac{94}{77},\frac{159}{77}\right)$ should be the point on the plane $4x+5y+6z=20$ that is closest to the point $(1,2,3)$. Let's do a little check, though, before we compute the minimum distance. First to verify that $\left(\frac{29}{77},\frac{94}{77},\frac{159}{77}\right)$ is in the plane, note that $4\left(\frac{29}{77}\right)+5\left(\frac{94}{77}\right)+6\left(\frac{159}{77}\right)=\frac{116}{77}+\frac{470}{77}+\frac{954}{77}=\frac{1540}{77}=20$. Second, it's obvious that if the line segment from $(1,2,3)$ to $\left(\frac{29}{77},\frac{94}{77},\frac{159}{77}\right)$ represents the shortest distance between the point $(1,2,3)$ and the plane $4x+5y+6z=20$, then this line segment should be perpendicular to our plane. Let's do a check of this both algebraically and visually. We already know one vector perpendicular to our plane, the one we obtain from the coefficients of x, y, and z. In other words, $\vec{v} = 4\hat{i} + 5\hat{j} + 6\hat{k}$. Parametric equations for the line that passes through $\left(\frac{29}{77}, \frac{94}{77}, \frac{159}{77}\right)$ in the direction of this vector are,

$$\begin{aligned}
x &= \frac{29}{77} + 4t \\
y &= \frac{94}{77} + 5t \\
z &= \frac{159}{77} + 6t \\
-\infty &< t < \infty
\end{aligned}$$

Now let's see what value of t will make x equal to 1.

$$\frac{29}{77} + 4t = 1 \Rightarrow 4t = 1 - \frac{29}{77} = \frac{48}{77} \Rightarrow t = \frac{12}{77}$$

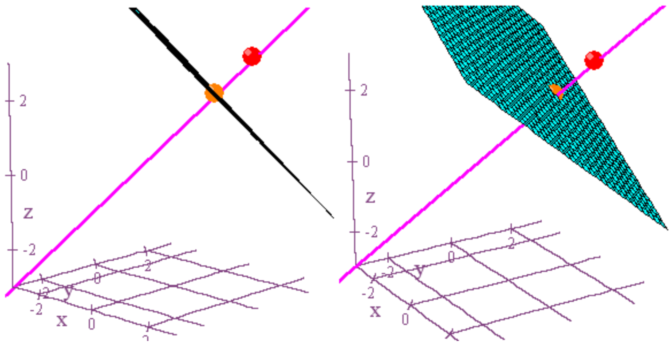
If we set $t = \frac{12}{77}$ and find the corresponding values for y and z, then this will show that the point $(1,2,3)$ is on the line that passes through $\left(\frac{29}{77}, \frac{94}{77}, \frac{159}{77}\right)$ and is perpendicular to the plane $4x + 5y + 6z = 20$.

$$x = \frac{29}{77} + 4\left(\frac{12}{77}\right) = \frac{77}{77} = 1$$

$$y = \frac{94}{77} + 5\left(\frac{12}{77}\right) = \frac{154}{77} = 2$$

$$z = \frac{159}{77} + 6\left(\frac{12}{77}\right) = \frac{231}{77} = 3$$

Now, for good measure, let's get some visual confirmation.



Looks confirmed to me! The minimum distance is now

$$d = \sqrt{\left(1 - \frac{29}{77}\right)^2 + \left(2 - \frac{94}{77}\right)^2 + \left(3 - \frac{159}{77}\right)^2} = \sqrt{\frac{2304}{5929} + \frac{3600}{5929} + \frac{5184}{5929}} = \sqrt{\frac{144}{77}}$$

$$= \frac{12}{\sqrt{77}} = \frac{12\sqrt{77}}{77} \approx 1.3675$$

At the start of this section on Lagrange multipliers, we gave a rather visual proof of why it would work for functions of the form $z = f(x, y)$. We would like to give an alternate proof now that is another example of how the chain rule is frequently important in proving theorems in multivariable calculus.

Theorem: Let $f(x,y)$ and $g(x,y)$ have continuous first partial derivatives, and suppose that $f(x,y)$ has an extreme value at the interior point (x_0, y_0) on a smooth constraint curve represented by $g(x,y)=c$. If $\nabla g(x_0,y_0) \neq \vec{0}$, then there is a real number λ such that $\nabla f(x_0,y_0) = \lambda \nabla g(x_0,y_0)$.

Proof: Let $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j}$ be a smooth parametrization for the constraint curve, and suppose $f(x_0,y_0) = f(x(t_0), y(t_0))$ is an extreme value. Then since $f(x,y)$ is differentiable along this curve, $0 = \dfrac{df(x,y)}{dt} = \dfrac{\partial f}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial f}{\partial y}\dfrac{dy}{dt} = \nabla f \bullet \dfrac{d\vec{r}}{dt}$ when the derivatives are evaluated at $t = t_0$. Therefore, $\nabla \mathrm{f}(x_0,y_0) \perp \vec{r}\,'(t_0)$. But since $\vec{r}(t)$ is a level curve for $w = g(x,y)$, $\nabla \mathrm{g}\ (x_0,y_0)$ is also perpendicular to $\vec{r}\,'(t_0)$. Therefore, $\nabla \mathrm{f}(x_0,y_0) \,\|\, \nabla g(x_0,y_0) \Rightarrow \nabla \mathrm{f}(x_0,y_0) = \lambda \nabla g(x_0,y_0)$. ■

So, that's it for now for Lagrange multipliers, and we're just about through with derivatives. There's just one more thing left to look at, the dreaded proof of the second partials test, so here goes. Gird your loins, and get ready for the ride!

<u>Theorem (Second Partials Test):</u> Suppose (a,b) is a point such that

$f_x(a,b) = 0 = f_y(a,b)$, and let

$$D = D(a,b) = f_{xx}(a,b) f_{yy}(a,b) - \left[f_{xy}(a,b) \right]^2 = \begin{vmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{xy}(a,b) & f_{yy}(a,b) \end{vmatrix}.$$

1. If $D > 0$ and $f_{xx}(a,b) > 0$, then $f(a,b)$ is a local minimum.
2. If $D > 0$ and $f_{xx}(a,b) < 0$, then $f(a,b)$ is a local maximum.
3. If $D < 0$, then $(a,b,f(a,b))$ is a saddle point.
4. If $D = 0$, then we know nothing.

<u>Proof:</u> Let $D(a,b) = f_{xx}(a,b) \cdot f_{yy}(a,b) - \left[f_{xy}(a,b) \right]^2$. Notice that, for the sake of simplicity, in the statement and execution of this proof we are assuming the equality of the two mixed partials, $f_{xy} = f_{yx}$. Also, we will abbreviate the above equality as $D = f_{xx} f_{yy} - f_{xy}{}^2$. Additionally, we will assume that every expression we write down is being evaluated at the point (a,b). Since all directional derivatives at this point will have the value zero, our strategy will be to look at concavity as determined by the sign of the second derivative in order to decide whether $(a,b,f(a,b))$ is a maximum point, a minimum point, or a saddle point.

Now let $\vec{u} = h\hat{i} + k\hat{j}$ be a unit vector. Then $D_{\vec{u}} f = \nabla f \bullet \vec{u} = f_x h + f_y k$. Also,

$D^2{}_{\vec{u}} f = D_{\vec{u}}\left(D_{\vec{u}} f\right) = \nabla\left(D_{\vec{u}} f\right) \bullet \vec{u} = \left[\left(f_{xx}h + f_{yx}k\right)\hat{i} + \left(f_{xy}h + f_{yy}k\right)\hat{j}\right] \bullet \left[h\hat{i} + k\hat{j}\right]$

$= f_{xx}h^2 + f_{yx}hk + f_{xy}hk + f_{yy}k^2 = f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2$. We can rewrite this last expression by completing the square.

$$
\begin{aligned}
D^2{}_{\vec{u}} f &= f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2 = f_{xx}\left(h^2 + \frac{2f_{xy}}{f_{xx}}hk\right) + f_{yy}k^2 \\
&= f_{xx}\left(h^2 + \frac{2f_{xy}}{f_{xx}}hk + \left[\frac{f_{xy}k}{f_{xx}}\right]^2\right) + f_{yy}k^2 - \frac{f_{xy}{}^2k^2}{f_{xx}} \\
&= f_{xx}\left(h + \frac{f_{xy}k}{f_{xx}}\right)^2 + \frac{k^2}{f_{xx}}\left(f_{xx}f_{yy} - f_{xy}{}^2\right)
\end{aligned}
$$

Thus, $D^2{}_{\vec{u}} f = f_{xx}\left(h + \frac{f_{xy}k}{f_{xx}}\right)^2 + \frac{k^2}{f_{xx}}\left(f_{xx}f_{yy} - f_{xy}{}^2\right)$. Consequently, if $D = f_{xx}f_{yy} - f_{xy}{}^2 > 0$ and $f_{xx} > 0$, then $D^2{}_{\vec{u}} f > 0$ for all unit vectors $\vec{u}$. Thus, any plane that passes through $z = f(x,y)$ and contains the point $\left(a,b,f(a,b)\right)$ and is perpendicular to the *xy*-plane will result in a cross-section with $z = f(x,y)$ that is concave up. Therefore, $\left(a,b,f(a,b)\right)$ is a minimum point. If $D = f_{xx}f_{yy} - f_{xy}{}^2 > 0$ and $f_{xx} < 0$, then $D^2{}_{\vec{u}} f < 0$ for all unit vectors $\vec{u}$, and the argument is similar that $\left(a,b,f(a,b)\right)$ is a maximum point.

Now suppose that $D = f_{xx}f_{yy} - f_{xy}^{\;2} < 0$, and recall that $D^2_{\vec{u}} f = f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2$.

Suppose $f_{xx} \neq 0$, and note that

$$f_{xx}D_{\vec{u}}^{\;2} f = f_{xx}\left(f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2\right) = f_{xx}^{\;2}h^2 + 2f_{xx}f_{yy}hk + f_{xx}f_{yy}k^2$$

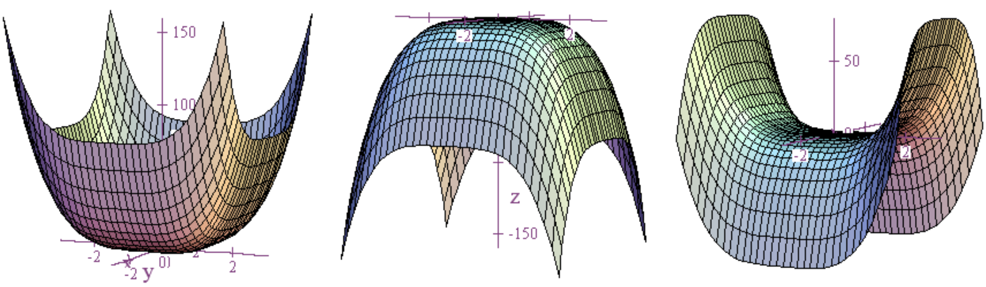
$$= f_{xx}^{\;2}h^2 + 2f_{xx}f_{xy}hk + f_{xy}^{\;2}k^2 + f_{xx}f_{yy}k^2 - f_{xy}^{\;2}k^2 = \left(f_{xx}h + f_{xy}k\right)^2 + \left(f_{xx}f_{yy} - f_{xy}^{\;2}\right)k^2.$$

Hence, $f_{xx}D_{\vec{u}}^{\;2} f > 0$ when $h \neq 0$ and $k = 0$, and $f_{xx}D_{\vec{u}}^{\;2} f < 0$ when $f_{xx}h + f_{xy}k = 0$ and $k \neq 0$. This implies that $D_{\vec{u}}^{\;2} f$ is positive in one direction and negative in another, thus implying that the graph of $z = f(x,y)$ is concave-up in one direction at $(a,b,f(a,b))$ and concave-down in another direction. Therefore, $(a,b,f(a,b))$ is a saddle point. Also, if $f_{yy} \neq 0$, then a similar argument may be used to arrive at the same conclusion that $(a,b,f(a,b))$ is a saddle point. One may ask, though, in the argument above how it is that we know that we can have both $f_{xx}h + f_{xy}k = 0$ and $k \neq 0$. Well, if $k \neq 0$, then $f_{xx}h + f_{xy}k = 0 \Rightarrow \dfrac{h}{k} = -\dfrac{f_{xy}}{f_{xx}}$. However, since (h,k) is a point on the unit circle, we have that $\dfrac{h}{k} = \cot\theta$ where θ is the angle made with the positive x-axis. Furthermore, since $\cot\theta$ takes on every real number value as θ goes from 0 to 2π, it certainly, at some point, takes on the value $-\dfrac{f_{xy}}{f_{xx}}$. And from this we get that $f_{xx}h + f_{xy}k = 0$.

Now, suppose again that $D = f_{xx}f_{yy} - f_{xy}^{\ 2} < 0$, and recall that

$D^2_{\vec{u}}f = f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2$. If $f_{xx} = 0 = f_{yy}$, then $D^2_{\vec{u}}f = f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2$ reduces to $D^2_{\vec{u}}f = 2f_{xy}hk$ and $D = -f_{xy}^{\ 2} < 0$. Hence, $f_{xy} \neq 0$, and $D_{\vec{u}}^{\ 2}f$ will have different signs for the unit vectors $\vec{u}_1 = \frac{1}{\sqrt{2}}\hat{i} + \frac{1}{\sqrt{2}}\hat{j}$ and $\vec{u}_2 = \frac{1}{\sqrt{2}}\hat{i} - \frac{1}{\sqrt{2}}\hat{j}$. Therefore, $(a, b, f(a,b))$ is a saddle point.

The final thing we need to show is that if $D = 0$, then the test is inconclusive. This can be done simply by examining the graphs of $z = x^4 + y^4$, $z = -x^4 - y^4$, and $z = x^4 - y^4$. You can easily show that each of these functions has $(0,0)$ as a critical point, and each function results in $D = 0$. However, the graph of the first function displays a local minimum at the critical point, the graph of the second shows a local maximum, and the third graph has a saddle point at $(0,0)$. Thus, these three examples show that anything can happen when $D = 0$.



By golly, I think we're done! ■