3 Differentiation

Now that we have a basic understanding of multivariable functions, we want to apply calculus to them. Our goal in this section is to define and understand the derivative, which measures the rate at which a function is changing.

3.1 The Partial Derivative

Already during this class, we have often talked about how quickly a function is changing when you change one of the input variables. This is exactly the single-variable calculus derivative and can be defined accordingly.

Definition 3.1. Let f be a function of two variables. Then we define the *partial derivatives at the point* (a, b) by

$$\begin{aligned} \frac{\partial f}{\partial x}(a,b) &= \lim_{h \to 0} \frac{f(a+h,b) - f(a,b)}{h} = f_x(a,b) \\ \frac{\partial f}{\partial y} &= \lim_{h \to 0} \frac{f(a,b+h) - f(a,b)}{h} = f_y(a,b). \end{aligned}$$

If we allow (a, b) to vary, we get functions $f_x(x, y)$ and $f_y(x, y)$.

We will sometimes write $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. If we want to represent these derivatives evaluated at a point, we will write $\frac{\partial z}{\partial x}\Big|_{(a,b)}$ and $\frac{\partial z}{\partial y}\Big|_{(a,b)}$.

Remark 3.2. This isn't just analogous to the single-variable calculus derivative; it is exactly identical. If we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ and we hold the second variable fixed at $y = b$, then we get a single-variable function defined by $f_b(x) = f(x, b)$. Then $f_x(a, b) = f_b'(a)$ is just the single-variable derivative of this single-variable function.

The interesting part here is not that we can define the partial derivatives, which are basically old news. The interesting thing is that we can get the answers to genuinely multi-varaible questions out of these essentially single-variable tools.

Example 3.3. Suppose a differentiable function $f(x, y)$ has the following values:

$y \setminus x$	0	1	2	3	4	5
0	120	135	155	160	160	150
1	125	128	135	160	175	160
2	100	110	120	145	190	170
3	85	90	110	135	155	180

Then we can estimate the partial derivatives off the chart. For instance, we can estimate that $f_x(3,1)$ is about 20: since $f(4,1) - f(3,1) = 15$ and $f(3,1) - f(2,1) = 25$. Similarly, we can estimate $f_y(3,1) \approx -7.5$ since $f(3,1) - f(3,0) = 0$ and $f(3,2) - f(3,1) = -15$.

One way to understand partial derivatives is to think about the units of the function. For instance, in your homework (problem 12.3.26) you looked at a function $H(x,t)$ that took position and time as inputs, and had temperature as an output. Then $H_x(x,t)$ has units of degrees per meter—how quickly temperature changes when you move a foot away. And $H_t(x,t)$ has units of degrees per minute—how quickly temperature changes over time.

Partial derivatives are easy and quite boring to calculate. Since we're looking at $f(x,y)$ as a function of a single variable, while holding the other constant, we can pretend it's simply a single-variable function, and treat the other variable like a constant.

Example 3.4. Let $f(x,y) = x^2 + y^2$. Then $f_x(x,y) = 2x$ and $f_y(x,y) = 2y$.

Let $g(x,y) = \sin(xy)$. Then $g_x(x,y) = \cos(xy) \cdot y$ and $g_y(x,y) = \cos(xy) \cdot x$.

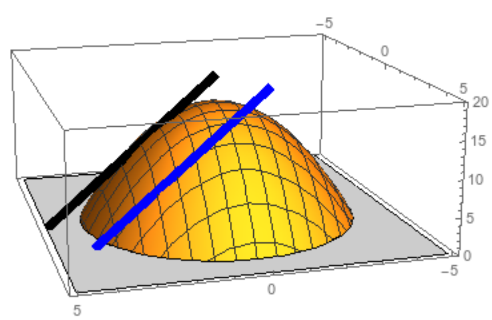
Let $h(x,y) = \frac{x^2}{y^3-3y}$. Then $h_x(x,y) = \frac{2x}{y^3-3y}$ and $h_y(x,y) = -\frac{x^2(3y^2-3)}{(y^3-3y)^2}$.

We can graphically understand partial derivatives by thinking about the cross-section.

Example 3.5. Let $f(x,y) = 16 - x^2 - y^2$. Then $f_x(x,y) = -2x$. Thus $f_x(2,0) = -4$, and the cross-section at 0 is $f(x,0) = 16 - x^2$ and has tangent line $z - 12 = -4(x-2)$.

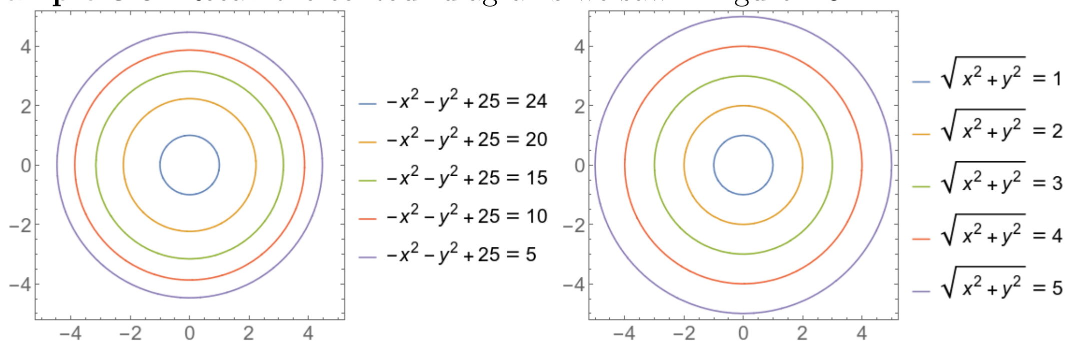
Similarly, if we look at the point $(2,2)$, we see that the cross-section is $f(x,2) = 12 - x^2$ and the derivative is $f_x(2,2) = -4$, so the tangent line is $z - 8 = -4(x-2)$.

Notice that the slopes of both lines are the same, since here $f_x(x,y)$ doesn't depend on y.



In section 1.2.4 we talked about reading contour diagrams and thinking about in which directions the function was changing. We can interpret this in terms of partial derivatives.

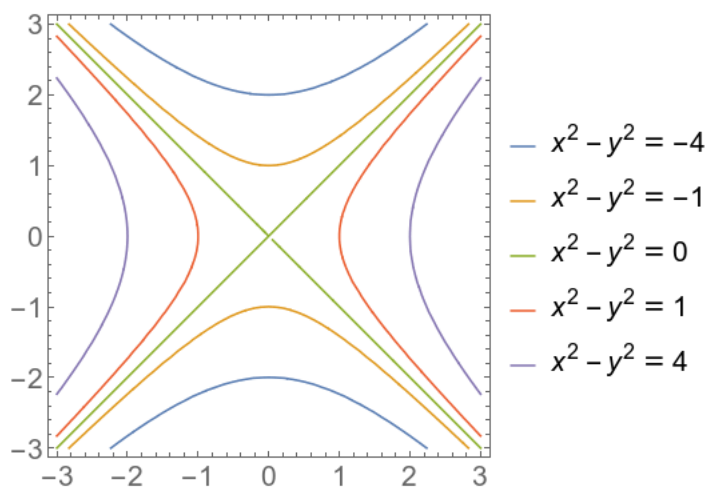
Example 3.6. Recall the contour diagrams we saw in figure 1.9:



We can ask questions like $f_x(1,0)$ and $g_x(1,0)$. Looking at the graph, we see that $f_x(1,0) \approx -4$ since it changes from 24 to 20 between $(1,0)$ and $(2,0)$. We can see that $f_y(1,0)$ is slightly smaller, since going from $(1,0)$ to $(1,1)$ doesn't quite get us from 24 to 20.

Similarly, $g_x(-2,0)$ is about -1, since $g(-3,0) = 3, g(-2,0) = 2$, and $g(-1,0) = 1$. $g_y(-2,0)$ is positive but less than 1.

Example 3.7. In the picture below, is $f_x(0,2)$ positive, negative, or zero? Is $f_y(0,2)$ positive, negative, or zero?



$f_x(0,2)$ is zero, since the curve is flat there and moving to the left or right shouldn't increase or decrease the output.

$f_y(0,2)$ is negative since the output gets lower as we go up away from the origin.

We can also define the partial derivatives in three (or more) dimensions; the only thing that changes is that the picture becomes more difficult to draw.

Example 3.8. Let $f(x, y, z) = x^2 + xyz + y/z$. Then we have

$$f_x(x, y, z) = 2x + yz$$
$$f_y(x, y, z) = zy + 1/z$$
$$f_z(x, y, z) = xy - y/z^2.$$

3.2 Local Linear Approximation

In many ways, the most important application of the derivative is the ability to approximate a function with a linear function. The basic idea is the same as the idea from single-variable calculus. If you zoom in enough on a 1-variable function, it will loook mostly like a line; if you zoom in on a 2-variable function, it will look like a plane.

Definition 3.9. Roughly speaking, the *tangent plane* to a surface at the point (x, y, z) is a plane that passes through the point (x, y, z), and touches the surface only at that point.

Proposition 3.10. *If $f(x, y)$ is differentiable at the point (a, b), then the equation of the tangent plane to $z = f(x, y)$ at the point (a, b) is*

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

From the equation form, we see that this plane must pass through the point $(a, b, f(a, b))$. Further, the slope in the x direction is $f_x(a, b)$, which is the rate at which f is changing when you change x. Similarly, $f_y(a, b)$ is the slope in the y direction.

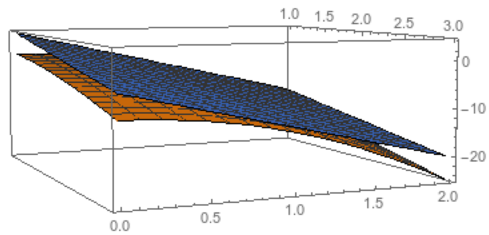
Example 3.11. Let's find the tangent plane to the function $f(x, y) = -x^2 - 4y^2$ at the point $(2, 1, -8)$.

We compute

$$f_x(x, y) = -2x \qquad f_x(2, 1) = -4$$
$$f_y(x, y) = -8y \qquad f_y(2, 1) = -8.$$

Since $f(2, 1) = -8$, the equation for the tangent plane is

$$z = -8 - 4(x - 2) - 8(y - 1)$$



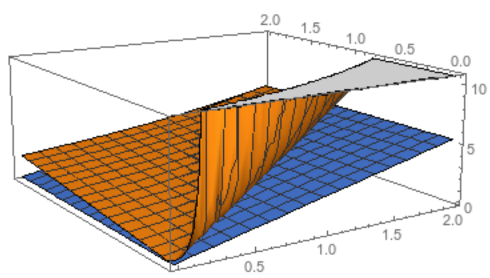
Example 3.12. Let's find the tangent plane to the function $g(x, y) = ye^{x/y}$ at the point $(1, 1)$.

We compute

$$\begin{aligned}
g_x(x, y) &= ye^{x/y}\frac{1}{y} = e^{x/y} \\
g_x(1, 1) &= e \\
g_y(x, y) &= e^{x/y} + ye^{x/y}\frac{-x}{y^2} = e^{x/y} - \frac{x}{y}e^{x/y} \\
g_y(1, 1) &= e - e = 0.
\end{aligned}$$

Since $g(1, 1) = e$, the equation for the tangent plane is

$$z = e + e(x - 1).$$



As with linear functions in single-variable calculus, we can use the tangent plane to approximate the values of a funtion.

Example 3.13. Let's estimate $g(1.1, 1)$.

We know that

$$g(x,y) \approx e + e(x-1)$$
$$g(1.1,1) \approx e + e(1.1-1) = e + .1e = 1.1e.$$

Using Mathematica, we compute that $g(1.1,1) \approx 3.00417$, and $1.1e \approx 2.99011$, so this is pretty good.

Definition 3.14. The *tangent plane approximation* to a function $f(x,y)$ near the point (a,b) is given by

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

The *linear approximation* to a function $f(x,y,z)$ near the point (a,b,c) is given by

$$f(x,y,z) \approx f(a,b,c) + f_x(a,b,c)(x-a) + f_y(a,b,c)(y-b) + f_z(a,b,c)(z-c).$$

Sometimes this is phrased in terms of the differential.

Definition 3.15. The *differential* df of a function f at a point (a,b) is a linear function in the variables dx and dy, given by

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

We will sometimes write $df = f_x dx + f_y dy$.

We can interpret the differential as being, for each point (a,b), a linear function that takes in a change in the x and y coordinates and outputs a change in the z coordinate. Thus

$$f(a+dx, b+dy) \approx f(a,b) + df(dx,dy) = f_x(a,b)dx + f_y(a,b)dy.$$

3.3 Gradients and directional derivatives

In the previous sections we used the partial derivatives to tell us how $f(x,y)$ will change as we change the input variables x and y. We'd like to generalize this further, and see what happens when we change the input in an arbitrary direction.

Definition 3.16. Let $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$ be a unit vector. Then we define the *directional derivative* of f in the direction $\vec{u}$ at the point (a,b,c) to be

$$f_{\vec{u}}(a,b,c) = \lim_{h\to 0} \frac{f(a+hu_1, b+hu_2, c+hu_3) - f(a,b)}{h}$$

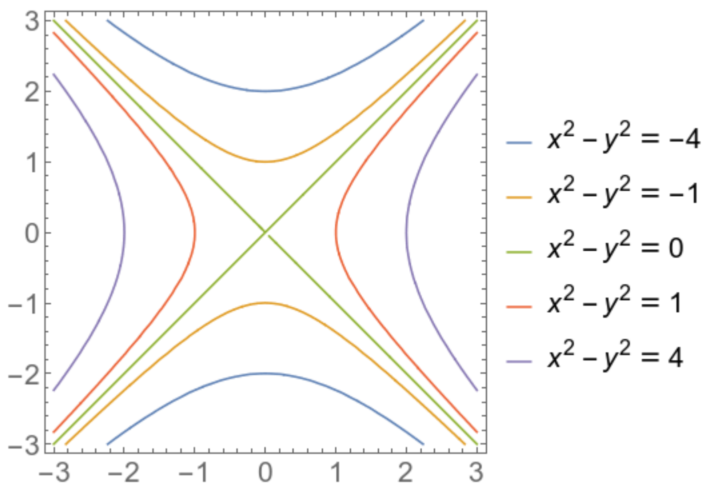
to be the rate of change of f in the direction $\vec{u}$.

If $\vec{v}$ is a non-zero non-unit vector, then we say the directional derivative in the direction of $\vec{v}$ is the directional derivative in the direction of $\frac{\vec{v}}{\|\vec{v}\|}$.

Conceptually, here we're seeing what happens if we change the input in the direction $\vec{u}$ with a change of size h, and then letting the size of the change go to zero.

Remark 3.17. If $\vec{u} = \vec{i}$, then $f_{\vec{u}} = f_x$. Similarly $f_{\vec{j}} = f_y$ and $f_{\vec{k}} = f_z$.

Example 3.18. Let's look at some of our contour plot from section 3.1 again.



We can ask for directional derivatives at a point. If we look at the point $(1, 1)$, we can see the derivative in the $\vec{i}$ direction is positive, and the derivative in the $\vec{j}$ direction is negative; these are just the partial derivatives we've already discussed.

But we can also see that the derivative in the $\vec{i} + \vec{j}$ direction is zero, since it follows directly along the contour.

Now think about the point $(1, -3)$. Is the directional derivative in the $\vec{i} + \vec{j}$ direction positive or negative? It should be positive, again, since we're climing up past the -4 contour towards the -1 contour.

What direction should we go to have a zero directional derivative? It's hard to be exact, but it looks like it should be down-right, and more right than down (following roughly parallel to the blue contour). In fact, we can compute that the exact direction is $3\vec{i} - \vec{j}$; we will see how to compute this later in this section.

We can compute these directional derivatives directly from the definition.

Example 3.19. Let $f(x) = x^2 - y^2$ (the function whose contour plot is in example 3.18). Let's compute the directional derivative in the $\vec{i} + \vec{j}$ direction at the point $(1, -3)$. Our unit vector in that direction is $\vec{u} = \frac{1}{\sqrt{2}}\vec{i} + \frac{1}{\sqrt{2}}\vec{j}$, and we compute

$$\begin{aligned}
f_{\vec{u}}(1,-3) &= \lim_{h\to 0} \frac{f(1+h/\sqrt{2}, -3+h/\sqrt{2}) - f(1,-3)}{h} \\
&= \lim_{h\to 0} \frac{(1+h/\sqrt{2})^2 - (-3+h/\sqrt{2})^2 - (1^2 - (-3)^2)}{h} \\
&= \lim_{h\to 0} \frac{1+\sqrt{2}h + h^2/2 - (9 - 3\sqrt{2}h + h^2/2) - (-8)}{h} \\
&= \lim_{h\to 0} \frac{4\sqrt{2}h}{h} = \lim_{h\to 0} 4\sqrt{2} = 4\sqrt{2}.
\end{aligned}$$

Computing the directional derivative directly from the limit definition is completely possible, but it's tedious. Just as we found easy ways to compute the single-variable derivative, we would like easy ways to compute the directional derivative of a multivariable function.

Fortunately, the partial derivatives give us enough information to do this. By local linearity, we see that

$$\begin{aligned}
f(a+hu_1, b+hu_2) &\approx f(a,b) + hu_1 f_x(a,b) + hu_2 f_y(a,b) \\
\frac{f(a+hu_1, b+hu_2) - f(a,b)}{h} &\approx \frac{hu_1 f_x(a,b) + hu_2 f_y(a,b)}{h} = u_1 f_x(a,b) + u_2 f_y(a,b).
\end{aligned}$$

Since this approximation should get increasingly good as h gets small, we conclude that

$$f_{\vec{u}}(a,b) = \lim_{h\to 0} \frac{f(a+hu_1, b+hu_2) - f(a,b)}{h} = u_1 f_x(a,b) + u_2 f_y(a,b).$$

Example 3.20. Let's work out our previous example this way. If $f(x,y) = x^2 - y^2$, we see that $f_x(x,y) = 2x$ and $f_y(x,y) = -2y$. Thus $f_x(1,-3) = 2$ and $f_y(1,-3) = 6$. Then we have

$$f_{\vec{u}}(1,-3) = \frac{1}{\sqrt{2}} \cdot 2 + \frac{1}{\sqrt{2}} \cdot 6 = \frac{8}{\sqrt{2}} = 4\sqrt{2}$$

as we got before.

In this computation, you may notice that we have something that looks like a dot product of $\vec{u}$ with a vector containing the partial derivatives. This leads us to define an object that we will use in almost all of our derivative calculations in the future.

Definition 3.21. If $f(x,y)$ is differentiable at (a,b), then the *gradient vector* of f at (a,b) is

$$\operatorname{grad} f(a,b) = \nabla f(a,b) = f_x(a,b)\vec{i} + f_y(a,b)\vec{j}.$$

Similarly, if $f(x,y,z)$ is differentiable at (a,b,c), then the gradient vector is

$$\operatorname{grad} f(a,b,c) = \nabla f(a,b,c) = f_x(a,b,c)\vec{i} + f_y(a,b,c)\vec{j} + f_z(a,b,c)\vec{k}.$$

Remark 3.22. We sometimes say that

$$\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}.$$

This is just another way of writing the same definition, but is really notationally convenient.

Proposition 3.23. *If f is differentiable at (a, b, c) and $\vec{u}$ is a unit vector, then*

$$f_{\vec{u}}(a, b, c) = \nabla f(a, b, c) \cdot \vec{u}.$$

Example 3.24. Let $f(x, y) = xy - \sin(x)$. Then the gradient is

$$\nabla f(x, y) = (y - \cos(x))\vec{i} + x\vec{j}$$

and the gradient at the point $(\pi, 1)$ is

$$\nabla f(\pi, 1) = 2\vec{i} + \pi\vec{j}.$$

The directional derivative in the direction $3/5\vec{i} + 4/5\vec{j}$ is

$$(2\vec{i} + \pi\vec{j}) \cdot (3/5\vec{i} + 4/5\vec{j}) = \frac{6 + 4\pi}{5}.$$

The gradient tells us basically everything we want to know about the derivative of the function f; in many ways it "is" the derivative. (From a linear algebra perspective, ∇f is the matrix corresponding to the local linearization of f).

Proposition 3.25. *If f is differentiable at (a, b, c) and $\nabla f(a, b, c) \neq \vec{0}$, then:*

- *$\nabla f(a, b, c)$ is in the direction of maximum increase for f.*
- *$\|\nabla f(a, b, c)\|$ is the maximum rate of increase of f in any direction.*
- *$\nabla f(a, b, c)$ is perpendicular to the level sets of f.*

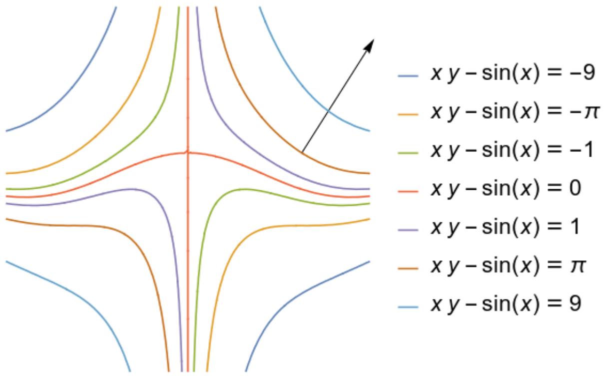
Proof. The rate of increase in the direction of a unit vector $\vec{u}$ is

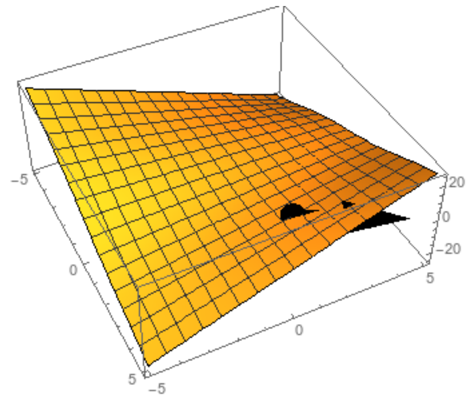
$$\nabla f(a, b, c)\dot{\vec{u}} = \|\nabla f(a, b, c)\| \cdot \|\vec{u}\| \cos\theta = \|\nabla f(a, b, c)\| \cos\theta.$$

This is maximized when $\theta = 0$, which is when $\nabla f(a, b, c)$ and $\vec{u}$ point in the same direction; the maximum value is $\|\nabla f(a, b, c)\|$.

$\nabla f(a, b, c)$ is the normal vector to the tangent plane (or line) at (a, b, c), and thus is perpendicular to the tangent plane. Thus it is perpendicular to the level set. $\square$

Example 3.26. We can look at the contour diagram and the graph for the function $f(x, y) = xy - \sin(x)$ from example 3.24.



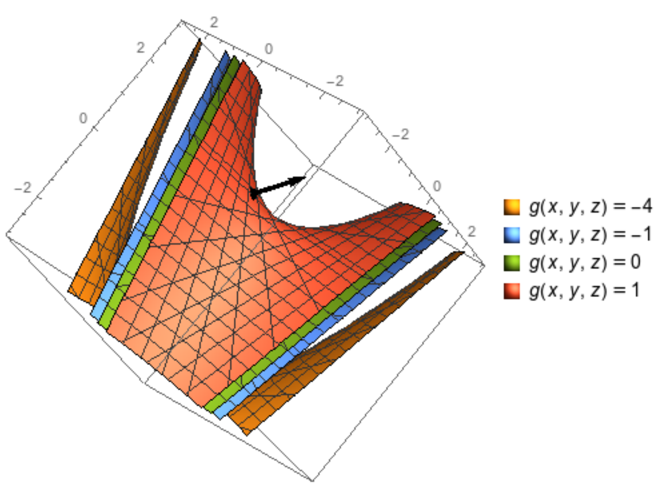


We see in the contour diagram that the gradient vector is perpendicular to the contour, and is in the direction of greatest increase. We can see the latter again in the three-dimensional graph—but this is much harder to read and see what's happening.

Example 3.27. Let's do a three-variable example next. Let $g(x, y, z) = xy + z$. Then

$$\nabla g(x, y, z) = y\vec{i} + x\vec{j} + 1\vec{k}.$$

At the point $(-1, 0, 1)$, we have $\nabla g(x, y, z) = -\vec{j} + \vec{k}$. Thus the direction of greatest increase is $-\vec{j} + \vec{k}$ and the magnitude of the increase in that direction is $\sqrt{2}$.



What if we want the directional derivative in the direction of, say $\vec{v} = 2\vec{i} + \vec{k}$? Then we have

$$\vec{u} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{2}{\sqrt{5}}\vec{i} + \frac{1}{\sqrt{5}}\vec{k}$$
$$f_{\vec{u}}(-1, 0, 1) = (-\vec{j} + \vec{k}) \cdot \vec{u} = 0 \cdot \frac{2}{\sqrt{5}} - 1 \cdot 0 + \frac{1}{\sqrt{5}} = \frac{1}{\sqrt{5}}.$$

3.4 The Chain Rule

We'd like an analogue of the single-variable chain rule for multivariable functions. In the single-variable case, we ask how much f changes when you change x, and then how much g changes when you change $f(x)$, and multiply those together: $\frac{d}{dx}g(f(x)) = \frac{dg}{dx}(f(x)) \cdot \frac{df}{dx}(x)$.

The intuition in the multivariable case is basically the same; we track what effect changing each input has, and multiply them through. The expressions are more complicated pretty purely because there are more levers we can push on to change things.

To build some intuition, we'll start with the case where our composite isn't *really* a multivariable function: f depends on two variables, but each of those variables depends only on some variable t. This corresponds to, say, the force acting on a particle over time, when the force depends on position in space and the position in space depends on time.

Proposition 3.28 (Parametrized Chain Rule)**.** *If f, g, h are differentiable, and $x = g(t)$ and $y = h(t)$, then*

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

Conceptually, what's happening here is that we look at every way that f can change, and then see how t affects each of those factors; then we add all the separate changes together. (This is making some implicit assumption that things are almost linear—but every time we use the derivative, we're making that assumption).

Sketch. We know that $\Delta f \approx \frac{\partial f}{\partial x} \cdot \Delta x + \frac{\partial f}{\partial y} \cdot \Delta y$. But further we know that $\Delta x \approx \frac{dx}{dt} \cdot \Delta t$ and $\Delta y \approx \frac{dy}{dt} \cdot \Delta t$. Putting this together gives us

$$\Delta f \approx \frac{\partial f}{\partial x}\frac{dx}{dt}\Delta t + \frac{\partial f}{\partial y}\frac{dy}{dt}\Delta t$$
$$\frac{\Delta f}{\Delta t} \approx \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

and taking the limit as t goes to zero gives us what we want. □

Example 3.29. Suppose $z = y\cos(x)$, where $x = t^2$ and $y = t^3$. Then

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\ &= (-y\sin(x)) \cdot 2t + \cos(y) \cdot 3t^2 \\ &= -t^3\sin(t^2) \cdot 2t + \cos(t^3) \cdot 3t^2.\end{aligned}$$

We can generalize this sort of chain rule behavior to chaining together functions of many variables. In general, we have

$$\frac{\partial z}{\partial t} = \sum_{x_i} \frac{\partial z}{\partial x_i} \cdot \frac{\partial x_i}{\partial t}.$$

That is, for each variable that z depends on, we multiply together the way z depends on the variable and the way the variable depends on t, and then add these all together to get the total change.

Example 3.30. Let $f(x, y) = x^2 y$ where $x = 4u + v$ and $y = u^2 - v^2$. Compute $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$.

$$\begin{aligned}
\frac{\partial f}{\partial u} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = 2xy \cdot 4 + x^2 \cdot 2u \\
&= 2(4u + v)(u^2 - v^2)4 + (4u + v)^2 2u \\
\frac{\partial f}{\partial v} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = 2xy \cdot 1 + x^2(-2v) \\
&= 2(4u + v)(u^2 - v^2) + (4u + v)^2(-2v).
\end{aligned}$$

Example 3.31. Suppose we have a function f that can be expressed as a function of x and y, or of u and v, where $x = u + v$ and $y = u - v$. (This is called a change of basis). We can express the partial derivatives in terms of each other.

We have

$$\begin{aligned}
\frac{\partial f}{\partial u} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = \frac{\partial f}{\partial x} \cdot 1 + \frac{\partial f}{\partial y} \cdot 1 \\
&= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \\
\frac{\partial f}{\partial v} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = \frac{\partial f}{\partial x} \cdot 1 + \frac{\partial f}{\partial y} \cdot (-1) \\
&= \frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}.
\end{aligned}$$

If we want to go the opposite way, and express $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ in terms of $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$, then we have two options. One is to observe that $u = \frac{x+y}{2}$ and $v = \frac{x-y}{2}$, and then use the chain rule again:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} = \frac{1}{2}\frac{\partial f}{\partial u} + \frac{1}{2}\frac{\partial f}{\partial v} \\
\frac{\partial f}{\partial y} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} = \frac{1}{2}\frac{\partial f}{\partial u} - \frac{1}{2}\frac{\partial f}{\partial v}.
\end{aligned}$$

Alternatively, we could have taken the expressions we already had and rearranged them. We knew that

$$\frac{\partial f}{\partial u} + \frac{\partial f}{\partial v} = 2\frac{\partial f}{\partial x}$$
$$\frac{\partial f}{\partial u} - \frac{\partial f}{\partial v} = 2\frac{\partial f}{\partial y}$$

and dividing by 2 gives us the same answer we got before.

3.5 Second Partials

So far we've spoken explicitly only about the first-order derivatives of f. But each derivative gives us a new function, which we can also take the derivatives of. In single variable calculus this gives us "the" second derivative. In multivariable calculus, just as there is more than one first derivative, there is more than one second derivative.

Definition 3.32. We define the *second-order partial derivatives* of $f(x, y)$ to be

$$\frac{\partial^2 z}{\partial^2 x} = f_{xx} = (f_x)_x \qquad \frac{\partial^2 z}{\partial x \partial y} = f_{yx} = (f_y)_x$$
$$\frac{\partial^2 z}{\partial y \partial x} = f_{xy} = (f_x)_y \qquad \frac{\partial^2 z}{\partial^2 y} = f_{yy} = (f_y)_y$$

Example 3.33. Let $f(x, y) = xy^2 + 3x^2e^y$. Then

$$f_x(x, y) = y^2 + 6xe^y \qquad f_y(x, y) = 2xy + 3x^2e^y$$

so we compute

$$f_{xx}(x, y) = 6e^y \qquad f_{yx}(x, y) = 2y + 6xe^y$$
$$f_{xy}(x, y) = 2y + 6xe^y \qquad f_{yy}(x, y) = 2x + 3x^2e^y.$$

You may have noticed a repetition here. Though there are four distinct mixed partials to compute, we only got three separate answers. This isn't an accident.

Theorem 3.34. *If f_{xy} and f_{yx} are continuous at the point (a, b), and (a, b) is an interior point of their domain, then*

$$f_{xy}(a, b) = f_{yx}(a, b).$$

These second-order partials measure how quickly the derivatives—the first partials—change when we change our input. This is similar to your homework problem 14.1.24, which asked how the partial derivatives changed as you moved from point A to point B.

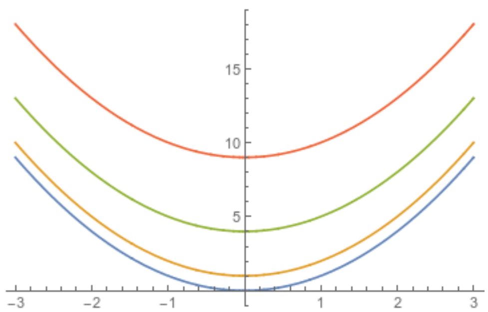
For example, if f_{xx} is positive, that means that the function gets steeper in the x direction as you increase x. If f_{xy} is positive, that means the function gets steeper in the x diretion as you increase y.

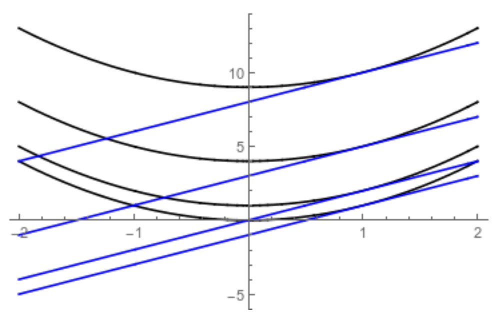
Example 3.35. Consider the function $f(x, y) = x^2 + y^2$. We see that

$$f_{xx}(x, y) = 2 \qquad f_{xy}(x, y) = 0 \qquad f_{yy}(x, y) = 2.$$

What does this tell us? Well, at any point, moving one unit in the x direction increases the x slope by about two; and similarly, moving one unit in the y direction increases the y slope by about two.

But moving in the y direction doesn't affect the x slope at all, and vice versa. Geometrically, this is because all the cross sections are identical parabolas at different heights: their levels will be different, but their slopes will all be the same at the same x value. So changing y doesn't change the x slope at all.





We can use these second partial derivatives to improve our approximations. In section 3.2 we talked about linear approximation, which the linear function that best approximates our function near a given point. With second partials, we can construct the *second-degree Taylor polynomial* or *quadratic approximation.*

Definition 3.36. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined near (a, b). The *Taylor polynomial of degree 1* for f near (a, b) is

$$T_1(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

The *Taylor polynomial of degree 2* is

$$\begin{aligned} T_2(x, y) = f(a, b) &+ f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &+ \frac{f_{xx}(a, b)}{2}(x - a)^2 + f_{xy}(a, b)(x - a)(y - b) + \frac{f_{yy}(a, b)}{2}(y - b)^2. \end{aligned}$$

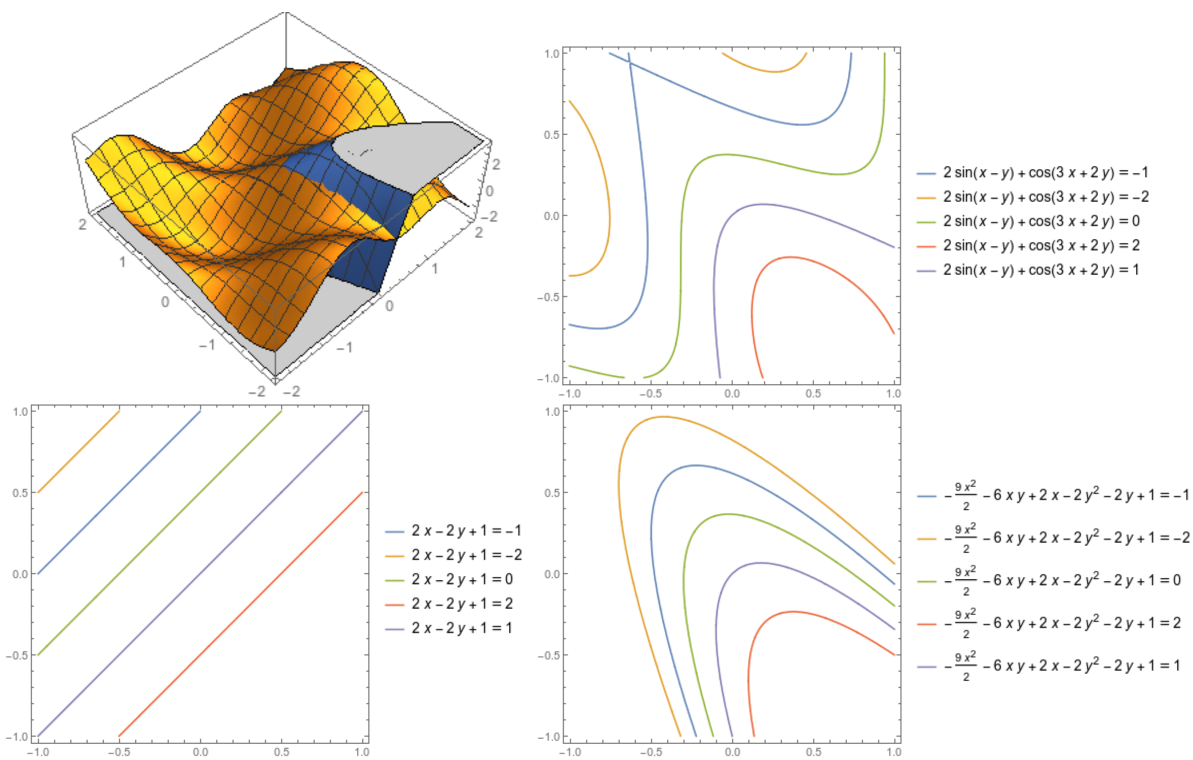
These approximations are used quite often in physics and in any other sort of numeric approximation work. It's possible to go to third-order and higher, and this works exactly like you'd expect; but third-order approximations are rarely actually useful. If the quadratic approximation isn't good enough, you generally want to just use a better tool instead.

Example 3.37. Let's find a quadratic approximation to $\cos(3x + 2y) + 2\sin(x - y)$ near $(0, 0)$.

$$\begin{aligned}
f(x, y) &= \cos(3x + 2y) + 2\sin(x - y) & f(0, 0) &= 1 \\
f_x(x, y) &= -3\sin(3x + 2y) + 2\cos(x - y) & f_x(0, 0) &= 2 \\
f_y(x, y) &= -2\sin(3x + 2y) - 2\cos(x - y) & f_y(0, 0) &= -2 \\
f_{xx}(x, y) &= -9\cos(3x + 2y) + 2\sin(x - y) & f_{xx}(0, 0) &= -9 \\
f_{xy}(x, y) &= -6\cos(3x + 2y) + 2\sin(x - y) & f_{xy}(0, 0) &= -6 \\
f_{yy}(x, y) &= -4\cos(3x + 2y) - 2\sin(x - y) & f_{yy}(0, 0) &= -4
\end{aligned}$$

so the quadratic approximation is

$$T_2(x, y) = 1 + 2x - 2y - 9x^2/2 - 6xy - 2y^2.$$



Suppose we want to find $\cos(.3 - .2) + 2\sin(.1 + .1)$. Then we have

$$f(.1, -.1) \approx T_2(.1, -.1) = 1 + .2 + .2 - .09/2 + .06 - .02 = 1.395.$$

Plugging in, the true answer is ≈ 1.39234, so this is pretty good.

Example 3.38. Let's find a quadratic approximation to e^{xy} near $(0, 2)$.

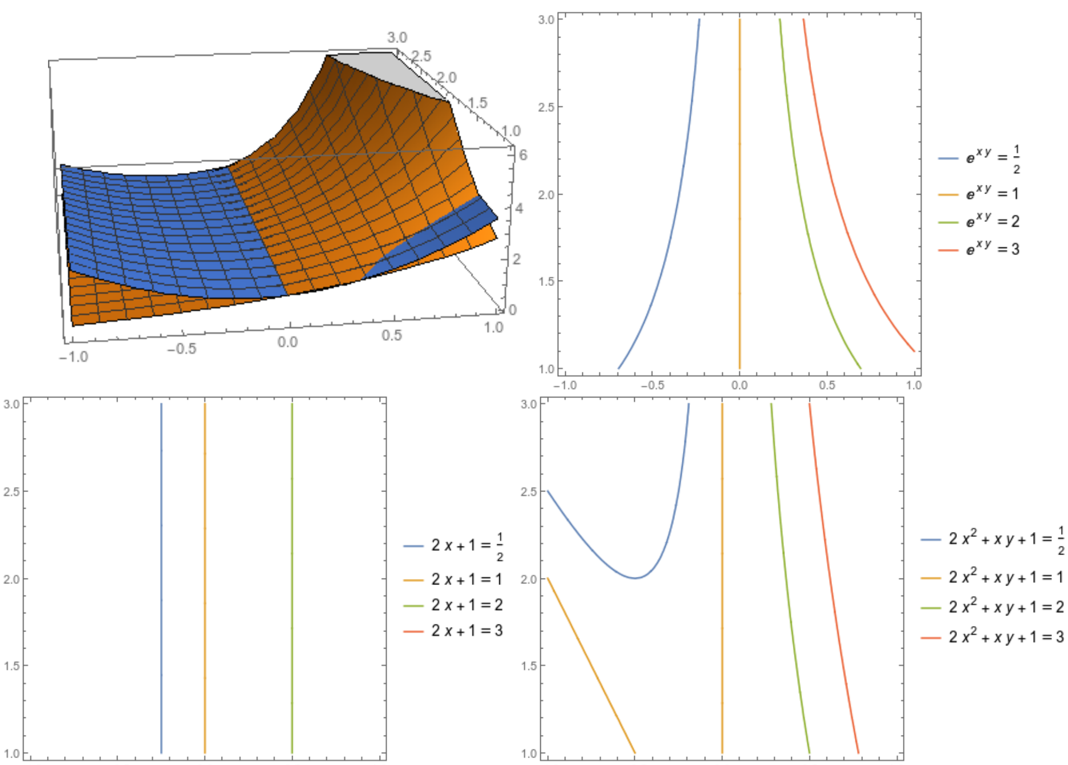
We compute

$$\begin{aligned} f(x,y) &= e^{xy} & f(0,2) &= 1 \\ f_x(x,y) &= ye^{xy} & f_x(0,2) &= 2 \\ f_y(x,y) &= xe^{xy} & f_y(0,2) &= 0 \\ f_{xx}(x,y) &= y^2e^{xy} & f_{xx}(0,2) &= 4 \\ f_{xy}(x,y) &= e^{xy} + xye^{xy} & f_{xy}(0,2) &= 1 \\ f_{yy}(x,y) &= x^2e^{xy} & f_{yy}(0,2) &= 0 \end{aligned}$$

Thus we can compute the Taylor polynomial:

$$\begin{aligned} T_2(x,y) &= 1 + 2x + 0(y-2) + 4x^2/2 + 1 \cdot x(y-2) + 0(y-2)^2/2 \\ &= 1 + 2x + 2x^2 + x(y-2) = 1 + 4x^2 + xy \end{aligned}$$

(We can multiply it out like in that last step; we generally shouldn't).



We can see that the linear approximation is still trying but not quite there.

We can also estimate, say, $e^{(-.1)\cdot 2.2} = e^{-.22}$. We have

$$e^{-.22} = f(-.1, 2.2) \approx T_2(-.1, 2.2) = 1 + .02 - .22 = .80.$$

Alternatively, we could write

$$e^{-.22} = f(-.1, 2.2) \approx T_2(-.1, 2.2) = 1 - .2 + .02 - .1(.2) = .8.$$

The true answer is about .8025.