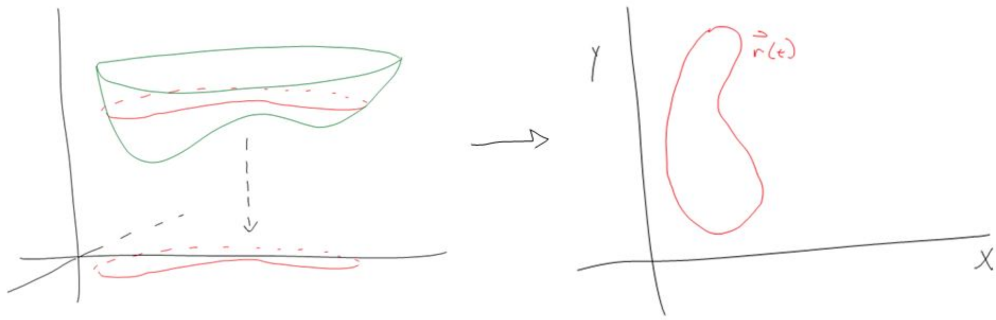
Gradients and the Directional Derivative

In 14.3, we discussed the partial derivatives $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$, which tell us the rate of change of the height of the surface defined by f in the "x direction" and the "y direction", respectively. These are rather severe restrictions on our knowledge of the behavior of the surface; we would like to be able to determine the surface's rate of change in *any* chosen direction, not just along the x or y axis. We will develop the tools to do so in this section.

The Gradient Vector

If the graph of $f(x, y)$ is a sufficiently nice surface, we can think about its level curves at height c, $f(x, y) = c$. A slightly different way to think about a chosen level curve $f(x, y) = c$ is to think about a vector function $\vec{r}(t) = g(t)\vec{i} + h(t)\vec{j}$ that traces out the same curve in the xy plane.



In particular, this allows us to think of $f(x, y) = c$ as $f(g(t), h(t)) = c$; in other words, we turn f into a function of one variable, t. Let's differentiate the relationship with respect to t:

$$\frac{d}{dt} f(g(t), h(t)) = \frac{\partial f}{\partial x}\frac{dg}{dt} + \frac{\partial f}{\partial y}\frac{dh}{dt}.$$

Since $f(g(t), h(t)) = c$ and $\dfrac{d}{dt}c = 0$ (c is a constant), we have

$$\frac{\partial f}{\partial x}\frac{dg}{dt} + \frac{\partial f}{\partial y}\frac{dh}{dt} = 0.$$

Notice that we can rewrite the relationship on the left hand side of the equality using the dot product:

$$\frac{\partial f}{\partial x}\frac{dg}{dt} + \frac{\partial f}{\partial y}\frac{dh}{dt} = \left(\frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j}\right) \cdot \left(\frac{dg}{dt}\vec{i} + \frac{dh}{dt}\vec{j}\right).$$

In particular, putting the two relationships together, we have

$$\left(\frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j}\right) \cdot \left(\frac{dg}{dt}\vec{i} + \frac{dh}{dt}\vec{j}\right) = 0 \tag{1}$$

for any differentiable multivariate function f.

Since $\vec{r}(t) = g(t)\vec{i} + h(t)\vec{j}$, the function

$$\frac{dg}{dt}\vec{i} + \frac{dh}{dt}\vec{j}$$

is precisely $\frac{dr}{dt}$, which is a vector that is tangent to the level curve and pointing in the direction of increasing t.

The function

$$\frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j}$$

on the left-hand side of the dot product is important enough that we will give it a name:

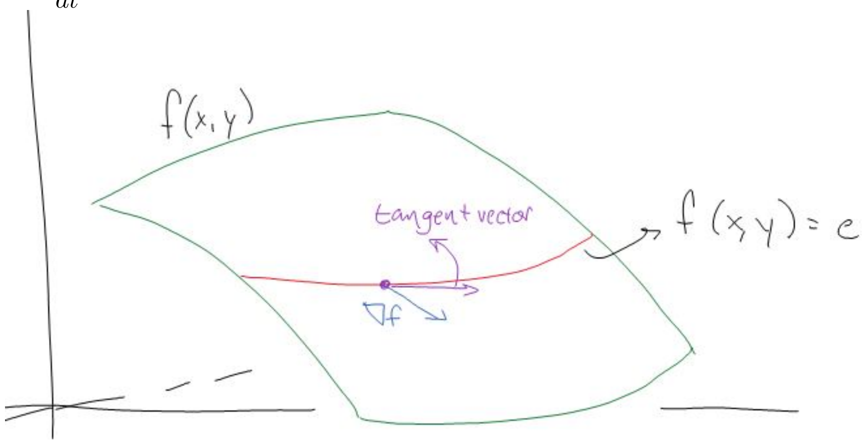
Definition 0.0.1. The **gradient** ∇f of the function $f(x, y)$ is defined to be the *vector* function

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

So the equation in line (1) becomes

$$\nabla f \cdot \frac{dr}{dt} = 0;$$

since the two vector functions have dot product 0, ∇f and $\frac{dr}{dt}$ are always orthogonal. As noted above, $\frac{dr}{dt}$ is tangent to the level curve, so that ∇f is *normal* to the level curve.



Theorem 0.0.2. Given a differentiable function $f(x, y)$, the value the gradient ∇f at (x_0, y_0) is a vector that is normal to the level curve through $f(x_0, y_0)$.

This fact will become extremely useful once we discuss the directional derivative.

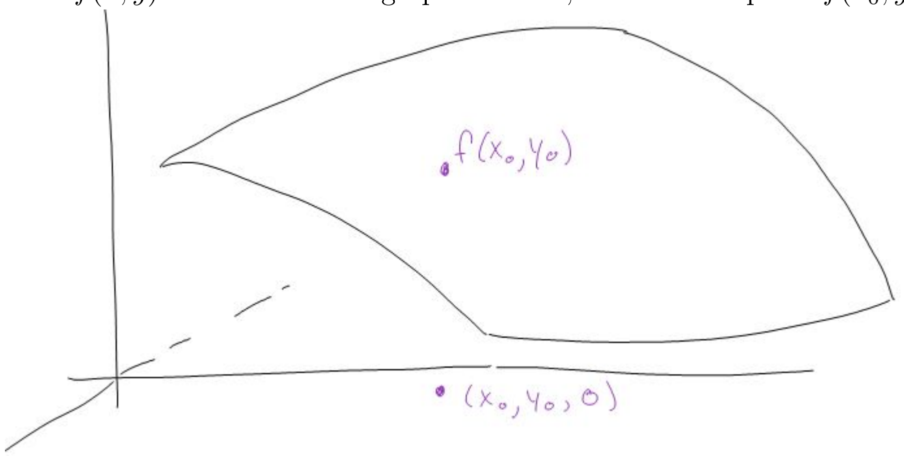
The following rules will help us to evaluate gradient vectors efficiently:

Theorem 0.0.3. If $f(x, y)$ and $g(x, y)$ are differentiable functions and k is any constant, then:

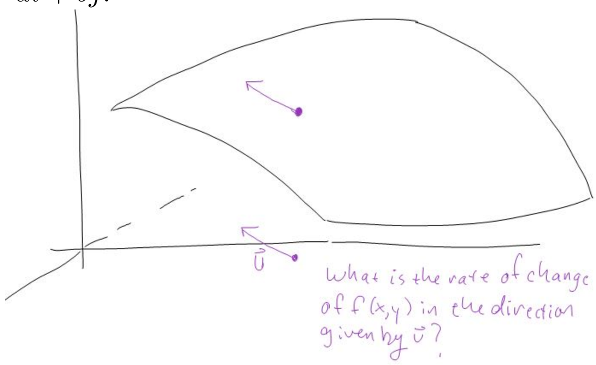
1. $\nabla(kf) = k(\nabla f)$
2. $\nabla(f \pm g) = \nabla f \pm \nabla g$
3. $\nabla(fg) = f(\nabla g) + g(\nabla f)$
4. $\nabla\left(\dfrac{f}{g}\right) = \dfrac{g(\nabla f) - f(\nabla g)}{g^2}$.

The Directional Derivative

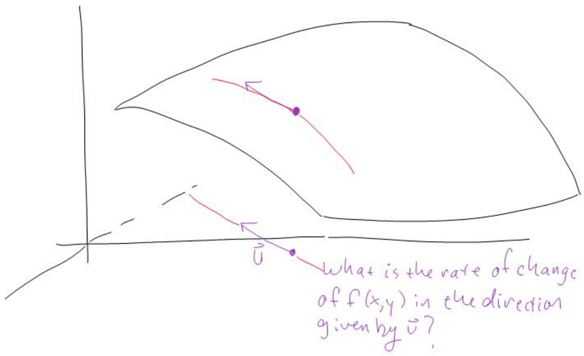
We now turn to the directional derivative. As indicated before, we wish to have a way to discuss the rate of change of the function $f(x, y)$ in *any* direction, not just along the x or y axis. Given a function $f(x, y)$ whose surface is graphed below, consider the point $f(x_0, y_0)$:



Directly below $f(x_0, y_0)$ is the point $(x_0, y_0, 0)$ in the xy plane. We may wish to determine the rate of change of the surface at $f(x_0, y_0)$ in a particular direction, say given by the unit vector $\vec{u} = a\vec{i} + b\vec{j}$:



We can think about the line in the xy plane in the direction given by $\vec{u}$ through (x_0, y_0):



This line has parametric equations given by

$$x = x_0 + ta,\ \ y = y_0 + tb,$$

where t is a parameter.

Varying t causes the x and y variables to move along this line; plugging the values for x and y along the line into $f(x, y)$ corresponds to moving along the surface in the direction given by the line. In all, we think of f as being controlled by t,

$$f(x, y) = f(x_0 + ta, y_0 + tb);$$

i.e. as a function in terms of the single variable t. Thus we can differentiate f using ideas from single variable calculus.

Definition 0.0.4. The *directional derivative* of the function f at $P_0 = (x_0, y_0)$ in the direction of the unit vector $\vec{u} = a\vec{i} + b\vec{j}$ is

$$D_{\vec{u}}f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h},$$

if the limit exists.

The value for the directional derivative of f in the direction $\vec{u}$ at $f(x_0, y_0)$ is precisely:

1. The slope of the line in the direction $\vec{u}$ that is tangent to the surface at $f(x_0, y_0)$ (the direction of slope is determined using the right-hand rule with $\vec{u}$ and $\vec{k}$)

2. The rate of change of the height of f at $f(x_0, y_0)$ in the direction given by $\vec{u}$.

If we choose $\vec{u}$ to be the unit vector $\vec{i}$, the formula for the directional derivative becomes

$$D_{\vec{i}}f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h},$$

which is precisely the partial derivative of f with respect to x. This makes sense, since "directional derivative in the direction $\vec{i}$" means that we should differentiate f in the x direction.

As we might expect, calculating the directional derivative using the limit definition above is quite tedious. While we will not go through the details here, it turns out that the directional derivative can be calculated much more readily using the gradient.

Theorem 3. If $f(x,y)$ is differentiable and $\vec{u} = \langle a, b\rangle$ is a unit vector, then the derivative of f in the direction $\vec{u}$ is precisely

$$D_{\vec{u}} f(x,y) = \nabla f(x,y) \cdot \vec{u}.$$

In other words, to calculate the directional derivative of f in the direction $\vec{u}$ at the point (x,y), we can evaluate ∇f at (x,y), then take the dot product of this vector with $\vec{u}$.

This formula gives us quite a bit of information about the role of ∇f. Recall that an alternate form of the dot product formula is $\vec{u}\cdot\vec{v} = |\vec{u}||\vec{v}|\cos\theta$, where θ is the angle between the two vectors. In our case, we are considering

$$D_{\vec{u}} f(x,y) = \nabla f(x,y) \cdot \vec{u} = |\nabla f(x,y)||\vec{u}|\cos\theta.$$

Since $\vec{u}$ is a unit vector, the formula reduces to

$$\nabla f(x,y) \cdot \vec{u} = |\nabla f(x,y)|\cos\theta.$$

Since $-1 \le \cos\theta \le 1$, the maximum value for $D_{\vec{u}} f$ (i.e. maximum rate of change of f) occurs when $\cos\theta = 1$, i.e. $\theta = 0$, and the minimum value for $D_{\vec{u}} f$ occurs when $\cos\theta = -1$, i.e. $\theta = \pi$, so that

$$-|\nabla f| \le D_{\vec{u}} f \le |\nabla f|.$$

Thus we have the following properties:

1. At any point (x_0, y_0), f increases fastest when $\vec{u}$ is pointing in the direction $\nabla f(x,y)$, and the largest value for the directional derivative at (x_0, y_0) is
$$D_{\vec{u}} = |\nabla f|.$$

2. At any point (x_0, y_0), f decreases fastest when $\vec{v}$ is pointing in the direction $-\nabla f(x,y)$, and the smallest value for the directional derivative at (x_0, y_0) is
$$D_{\vec{v}} = -|\nabla f|.$$

3. At any point (x_0, y_0), if the direction vector $\vec{w}$ is orthogonal to $\nabla f(x,y)$, then the directional derivative in the direction $\vec{w}$ is precisely 0, which means that $\vec{v}$ points along the level curve of f at height $f(x_0, y_0)$.

In particular, the maximum rate of change of f at any point $f(x_0, y_0)$ will always be in the direction orthogonal to the level curve through $f(x_0, y_0)$. An interesting application of this fact can be seen on topographical maps; rivers are always perpendicular to the level curves on the map, since they flow in the direction of most rapid elevation decrease.

Examples:
Given $f(x,y) = e^{x^2+y^2}$, find the rate of change of f at $(\sqrt{2}, \sqrt{2})$ in the direction $2\vec{\imath} + \vec{\jmath}$.

Since the formula for directional derivative is

$$D_{\vec{u}} f = \nabla f(x,y) \cdot \vec{u},$$

we will need to know the gradient vector at $(\sqrt{2}, \sqrt{2})$, and will need to find a unit vector pointing in the direction $2\vec{i} + \vec{j}$. The gradient is

$$\nabla f = 2xe^{x^2+y^2}\vec{i} + 2ye^{x^2+y^2}\vec{j};$$

its value at $(\sqrt{2}, \sqrt{2})$ is

$$\nabla f(\sqrt{2}, \sqrt{2}) = 2e^4\sqrt{2}\vec{i} + 2e^4\sqrt{2}\vec{j}.$$

Since the length of $2\vec{i} + \vec{j}$ is $\sqrt{5}$, which is clearly not a unit vector, we will set

$$\vec{u} = \frac{2}{\sqrt{5}}\vec{i} + \frac{1}{\sqrt{5}}\vec{j}.$$

Then $D_{\vec{u}}f$ is

$$\begin{aligned}
(2e^4\sqrt{2}\vec{i} + 2e^4\sqrt{2}\vec{j})\cdot\left(\frac{2}{\sqrt{5}}\vec{i} + \frac{1}{\sqrt{5}}\vec{j}\right) &= (2e^4\sqrt{2})\left(\frac{2}{\sqrt{5}}\right) + (2e^4\sqrt{2})\left(\frac{1}{\sqrt{5}}\right) \\
&= \frac{4e^4\sqrt{2}}{\sqrt{5}} + \frac{2e^4\sqrt{2}}{\sqrt{5}} \\
&= \frac{6e^4\sqrt{2}}{\sqrt{5}} \\
&= 6e^4\sqrt{\frac{2}{5}}.
\end{aligned}$$

So the rate of change of f in the direction $2\vec{i} + \vec{j}$ is $6e^4\sqrt{\frac{2}{5}}$.

Given $f(x, y) = x^2 + y^2$, determine ∇f and find the maximum rate of change of f at $(3, 4)$. Then sketch the gradient vector at $(3, 4)$ on the level curve that passes through this point.

The gradient is

$$\nabla f = 2x\vec{i} + 2y\vec{j}.$$

Since the length of the gradient at (x_0, y_0) is the maximum value for the derivative of $f(x, y)$ at (x_0, y_0), we can find the maximum rate of change of f at $(3, 4)$ by evaluating ∇f at $(3, 4)$, then finding the length of this vector:

$$\nabla f(3, 4) = 6\vec{i} + 8\vec{j},$$

and

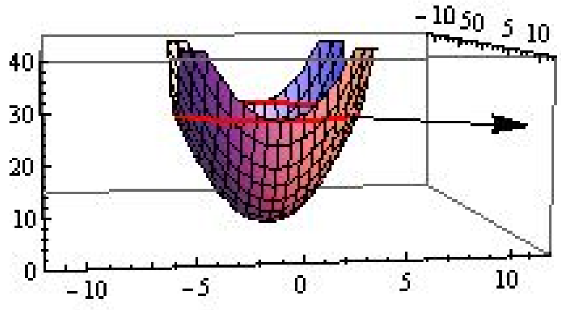
$$|\nabla f(3, 4)| = \sqrt{36 + 64} = \sqrt{100} = 10.$$

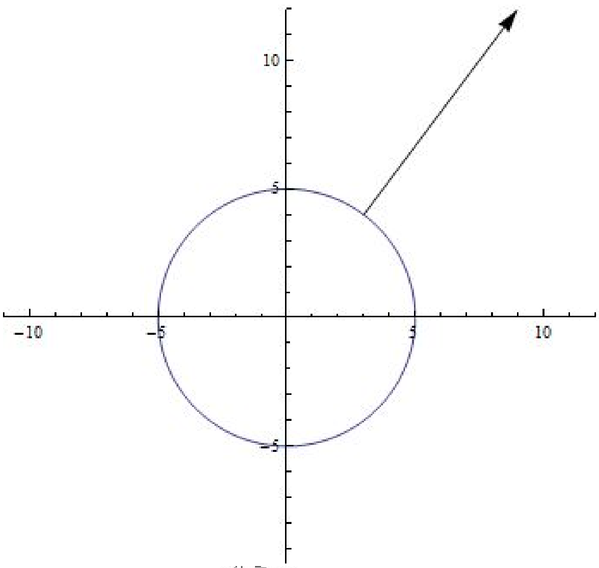
In other words, if we could stand at $f(3, 4)$ on the surface f, we would see the steepest ascent of f in the direction $6\vec{i} + 8\vec{j}$, and the rate of change in this direction would be 10.

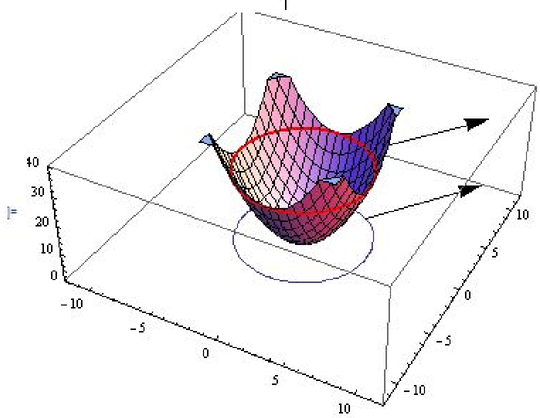
The level curve of f that passes through $(3, 4)$ is at height $f(3, 4) = 9 + 16 = 25$. So the level curve is the set of all (x, y) so that $x^2 + y^2 = 25$, i.e. the circle of radius 5 centered at the origin. We need to sketch the gradient vector

$$\nabla f(3, 4) = 6\vec{i} + 8\vec{j}$$

at this point:







Notice that the gradient vector is indeed orthogonal to the level curve.

The temperature in degrees Celsius on the surface of a metal plate is $T(x, y) = 20 - 4x^2 - y^2$, where x and y are measured in centimeters. In what direction from $(2, -3)$ does the temperature increase most rapidly, and what is the rate of increase?

The most rapid rate of increase of T occurs in the direction of $\nabla f = -8x\vec{i} - 2y\vec{j}$; at the point $(2, -3)$, the gradient points in the direction $\nabla f(2, -3) = -16\vec{i} + 6\vec{j}$. The rate of increase at the point in this direction is

$$|\nabla f(2, -3)| = \sqrt{256 + 36} = \sqrt{292} \approx 17.09^\circ \text{ per centimeter.}$$

If $f(x, y, z)$ is a function of three independent variables, the same definitions for gradient and directional derivative apply almost exactly; we simply need to extend them to the third variable.

Definition 0.0.5. If $f(x, y, z)$ is a differentiable function and $\vec{u} = u_1\vec{i} + u_2\vec{j} + u_3\vec{k}$ is a unit vector, then the gradient of f is

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

and the directional derivative of f in the direction of f in the direction $\vec{u}$ is

$$D_u f = \nabla f \cdot \vec{u} = \frac{\partial f}{\partial x}u_1 + \frac{\partial f}{\partial y}u_2 + \frac{\partial f}{\partial z}u_3.$$

Recall that, in the two variable case, $\nabla f\big|_{(x_0,y_0)}$ is normal to the level curve through $f(x_0, y_0)$. In the three variable case, $\nabla f\big|_{(x_0,y_0,z_0)}$ is normal to the *level surface* through $f(x_0, y_0, z_0)$.

Tangent Planes to Level Surfaces

If $f(x, y, z)$ is a function of three independent variables, then $f(x, y, z) = k$ is a level surface of f. Thus it makes sense to discuss "tangent planes" of the level surface–in other words, planes tangent to the level surface at a given point.

In order to find an equation for the plane, we must know (1) a point through which the plane passes, and (2) a vector normal to the plane. Fortunately, if $f(x, y, z)$ is a function of three variables, then ∇f is always normal to the level *surface* $f(x, y, z) = k$, so we may use ∇f as the normal vector for the plane.

More formally,

Definition 0.0.6. The *tangent plane* to the level surface $f(x, y, z) = k$ at $P_0 = (x_0, y_0, z_0)$ is the plane through (x_0, y_0, z_0) that is normal to $\nabla f\big|_{P_0}$.

If we want the plane through $P_0 = (x_0, y_0, z_0)$, and the gradient of f is given by

$$\nabla f = f_x\vec{i} + f_y\vec{j} + f_z\vec{k},$$

then

$$\nabla f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0)\vec{i} + f_y(x_0, y_0, z_0)\vec{j} + f_z(x_0, y_0, z_0)\vec{k};$$

thus $\langle f_x(x_0, y_0, z_0), f_y(x_0, y_0, z_0), f_z(x_0, y_0, z_0)\rangle$ is normal to the surface, and we can use this vector to write the equation for the plane tangent to $f(x, y, z) = c$ at P_0. We have all the information we need to write the equation for the plane (as well as for the normal line to the surface through P_0):

Theorem 19. The equation of the plane tangent to the level surface $f(x,y,z) = k$ at $P_0 = (x_0, y_0, z_0)$ is

$$f_x(x_0,y_0,z_0)(x-x_0) + f_y(x_0,y_0,z_0)(y-y_0) + f_z(x_0,y_0,z_0)(z-z_0) = 0.$$

The parametric equations for the line through $P_0 = (x_0, y_0, z_0)$ normal to the level surface $f(x,y,z) = k$ are

$$x = x_0 + f_x(x_0,y_0,z_0)t,\ \ y = y_0 + f_y(x_0,y_0,z_0)t,\ \ z = z_0 + f_z(x_0,y_0,z_0)t.$$

Examples:

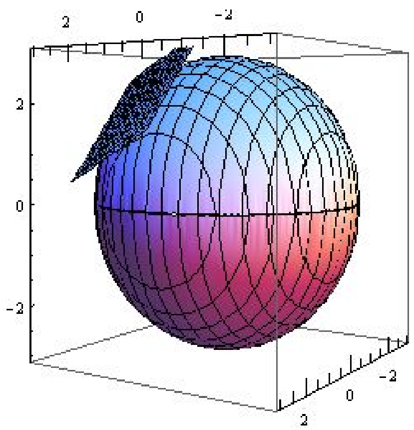
Let $f(x,y,z) = x^2+y^2+z^2$. Determine the formula for the plane that passes through the point $(1,2,2)$ and is tangent to the level surface of f at $w = 9$.

To use the formula, we will need to know the values for the partials $f_x = 2x$, $f_y = 2y$, and $f_z = 2z$ at $(1,2,2)$. We have

$$f_x(1,2,2) = 2,\ \ f_y(1,2,2) = 4,\ \text{ and } f_z(1,2,2) = 4.$$

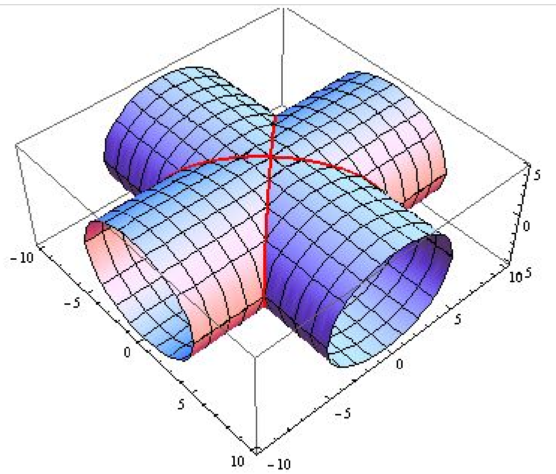
So the formula for the tangent plane is

$$2(x-1) + 4(y-2) + 4(z-2) = 0.$$

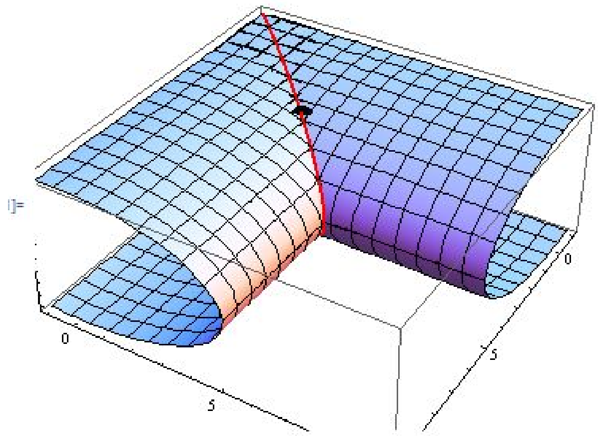


The surfaces $x^2+z^2 = 25$ and $y^2+z^2 = 25$ intersect along two curves in space. Find the equation for a line tangent to the curve that passes through the point $(3,3,4)$, then find the angle between the gradients of the two surfaces at this point.

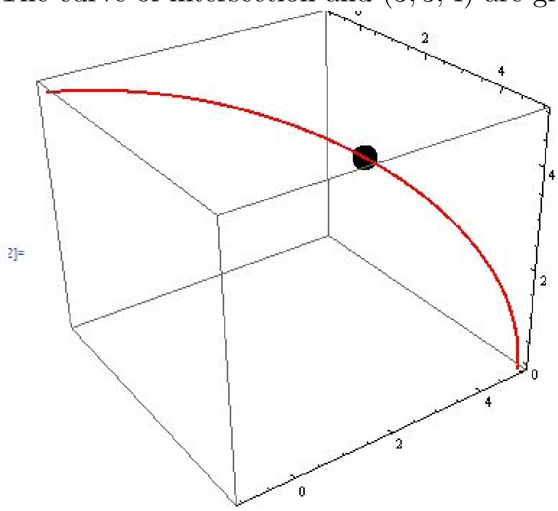
The surfaces and their lines of intersection are graphed below:



Let's zoom in on the point $(3, 3, 4)$, marked in black:



The curve of intersection and $(3, 3, 4)$ are graphed without the surfaces:



To find the equation for the line, we need to know a point on the line ($(3, 3, 4)$ will do) as well as a vector parallel to the line.

We can think of the surfaces as level surfaces of the functions $f(x, y, z) = x^2 + z^2$ and $g(x, y, z) = y^2 + z^2$. Then the gradients ∇f and ∇g are orthogonal to the level surfaces of f and g, respectively. In particular, if we take the cross product $\nabla f \times \nabla g$, the resulting vector will be orthogonal to each of ∇f and ∇g, thus parallel to the line tangent to the curve of intersection.

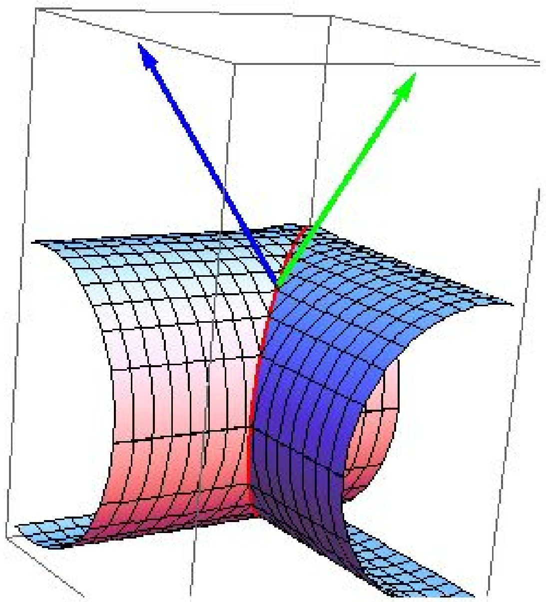
The gradients are

$$\nabla f = 2x\vec{i} + 2z\vec{k} \text{ and } \nabla g = 2y\vec{j} + 2z\vec{k}.$$

Then the gradient vectors at $P_0 = (3, 3, 4)$ are

$$\nabla f\big|_{P_0} = 6\vec{i} + 8\vec{k} \text{ and } \nabla g\big|_{P_0} = 6\vec{j} + 8\vec{k}.$$

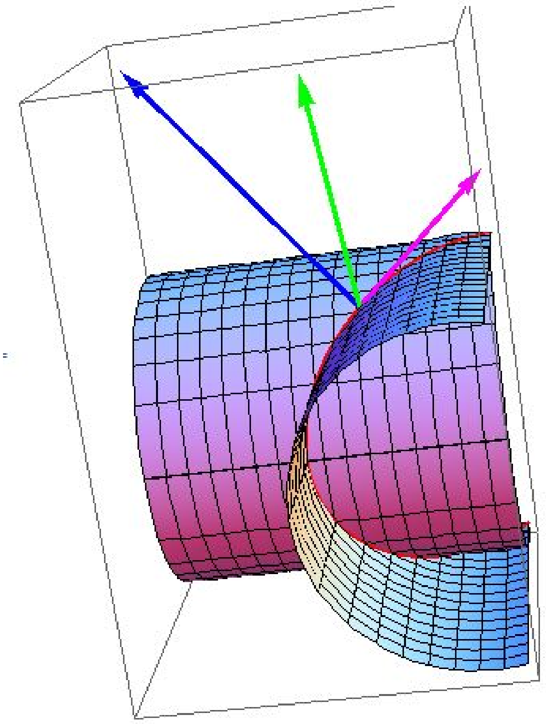
Here the gradients are graphed on the surfaces:



The cross product $\nabla f\big|_{P_0} \times \nabla g\big|_{P_0}$ is

$$\nabla f\big|_{P_0} \times \nabla g\big|_{P_0} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 6 & 0 & 8 \\ 0 & 6 & 8 \end{vmatrix} = -48\vec{i} - 48\vec{j} + 36\vec{k}.$$

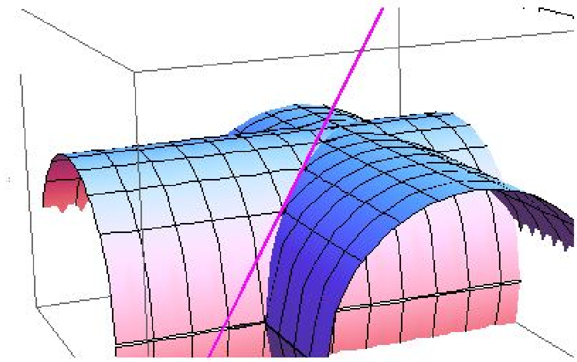
The vector $\nabla f\big|_{P_0} \times \nabla g\big|_{P_0}$ is graphed in purple below:



This vector is parallel to the line tangent to the curve of intersection, so the parametric equations for this line are

$$x = 3 - 48t, \; y = 3 - 48t, \text{ and } z = 4 + 36t.$$

The line is graphed below in purple:



To determine the angle between the gradients ∇f and ∇g at $(3, 3, 4)$, we can use the dot product formula

$$\nabla f \cdot \nabla g = |\nabla f||\nabla g| \cos\theta.$$

To use the formula, we need to find $\nabla f \cdot \nabla g$ at $(3, 3, 4)$, which is given by

$$\nabla f|_{P_0} \cdot \nabla g|_{P_0} = 6 \cdot 0 + 0 \cdot 6 + 8 \cdot 8 = 64.$$

We also need to calculate the lengths of each of the gradients at $(3, 3, 4)$:

$$|\nabla f|_{P_0} = \sqrt{36 + 48} = 2\sqrt{21},$$

and

$$\left|\nabla g\right|_{P_0} = \sqrt{36+48} = 2\sqrt{21}.$$

So the cosine of the angle between the gradients is

$$\cos\theta = \frac{64}{84}.$$

So the angle between the gradients is

$$\theta = \cos^{-1}\left(\frac{64}{84}\right) \approx .7 \text{ radians} \approx 40^\circ.$$