Chapter 4

differentiation

In single variable calculus we learn from the outset that the derivative of a function describes the slope of the function at a point. On the other hand, we also learned that the derivative at a point can be used to construct the best linear approximation to the function. In particular, the derivative at a point shows how the change in the independent variable Δx gives an approximate change $\Delta y = f'(a)\Delta x$. This characterization of the derivative is the one which most readily generalizes to many dimensions. In particular, we generalize Δy and Δx to become vectors and $f'(a)$ is a matrix when $f : \mathbb{R}^n \to \mathbb{R}^m$. I'll explain how the derivative matrix[1] $f'(a)$ is the natural extension of our single-variable calculus to the general case.

The nuts and bolts of this derivative matrix are made from what we call *partial derivatives* of the *component functions*. The partial derivative in turn is naturally defined in the context of directional derivatives. The directional derivative takes the multivariate function and restricts it to a particular line in the domain. By making this restriction we find a way to *do* single variable-type calculations on a multivariate function. Much of the calculation presented in this chapter is little more than single-variable calculus with a few simple rules adjoined. However, connecting the partial derivatives to the general derivative involves multivariate limits and some analysis that is beyond the required content of this course. That said, I include some of those arguments in these notes in the interest of logical completeness.

Most modern treatments ignore the need to discuss the general concept of differentiation and instead just show students an assortment of various partial derivative calculations. I've found students who are thinking are usually unsatisfied with the popular approach because there is no big picture behind the partial differentiation. It's just a seemingly random collection of adhoc rules. This need not be. If we submit ourselves to a little linear algebaic terminology there is a beautiful and quite general context in which all the partial derivatives find a natural purpose.

[1]often called the Jacobian matrix

4.1 directional derivatives

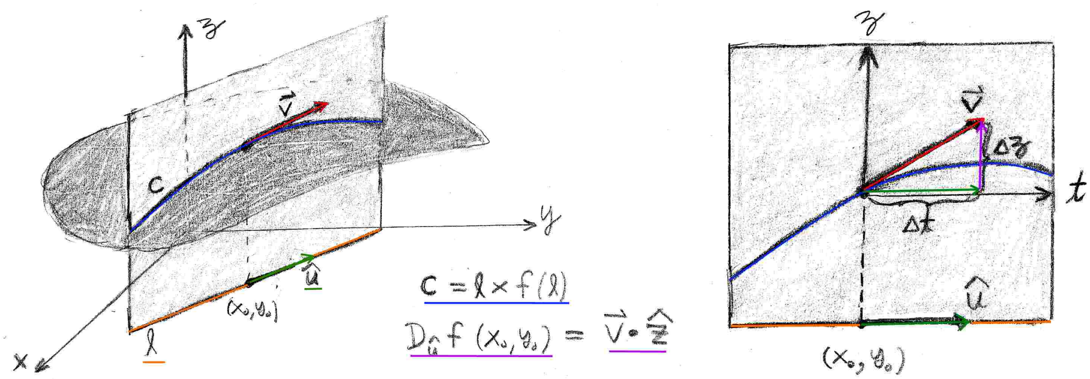
We begin our discussion with a function $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$. Consider a fixed point $(x_o, y_o) \in dom(f)$. Furthermore, picture $\widehat{u} = \langle a, b \rangle$ as a **unit**-vector (it's green) in the domain of f attached to (x_o, y_o) and construct the path in $dom(f)$ with direction $\langle a, b \rangle$ and base-point (x_o, y_o):

$$\vec{r}(t) = \langle x_o + ta,\ y_o + tb \rangle$$

If we feed this path (the orange line) to the function f then we can construct a curve in $\mathbb{R}^3$ which lies on the graph $z = f(x, y)$ and passes through the point $(x_o, y_o, f(x_o, y_o))$. In particular,

$$\vec{\gamma}(t) = \Big(x_o + ta,\ y_o + tb,\ f(x_o + ta,\ y_o + tb) \Big)$$

parametrizes the curve (in blue) formed by the intersection of the graph $z = f(x, y)$ and the vertical plane which contains $\widehat{z}$ and $a\,\widehat{x} + b\,\widehat{y}$.



In the picture above you can see that we identify the xy-plane embedded in $\mathbb{R}^3$ with the plane $\mathbb{R}^2$ which contains $dom(f)$. A natural choice of coordinates on vertical slice containing $\langle a, b, 0 \rangle$ and $\langle 0, 0, 1 \rangle$ is given by t, z. For the sake of discussion let $g(t) = f(x_o + ta,\ y_o + tb)$ and consider the graph $z = g(t)$. This is a context to which ordinary single-variate calculus applies. The derivative $g'(0)$ describes the slope of the tangent line in the tz-plane at $(0, g(0))$. Of course, from the three-dimensional perspective, $g'(0)$ gives the z-component of the velocity-vector (the red arrow) to the path $t \mapsto \vec{\gamma}(t)$. So what? Well, what is that quantity's meaning for $z = f(x, y)$? It's simply the following:

> **The value of $\frac{d}{dt}\big[\ f(x_o + ta,\ y_o + tb)\big]\Big|_{t=0}$ describes the rate of change in $f(x, y)$ in the direction $\langle a, b \rangle$ at the point (x_o, y_o).**

This is why we are interested in this calculation. The directional derivative of f in the $\langle a, b \rangle$ direction at (x_o, y_o) is precisely the slope described above.

Definition 4.1.1.

Let $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function with $\vec{p}_o = (x_o, y_o) \in dom(f)$ and suppose $\widehat{u} = \langle a, b\rangle$ is a unit-vector. If the limit below exists, then we define the **directional derivative** of f at $\vec{p}_o$ in the $\widehat{u}$-direction by

$$D_{\widehat{u}} f(\vec{p}_o) = \lim_{t\to 0}\left[\frac{f(\,\vec{p}_o + t\widehat{u}\,) - f(\,\vec{p}_o\,)}{t}\right] = \lim_{t\to 0}\left[\frac{f(x_o + ta,\ y_o + tb) - f(x_o, y_o)}{t}\right].$$

The definition above can also be written in terms of a derivative followed by an evaluation:

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ f(x_o + ta,\ y_o + tb)\ \bigg]\bigg|_{t=0}.$$

We pause to look at a few examples.

Example 4.1.2. Problem: *Suppose* $f(x, y) = 25xy$ *and calculate the rate of change in* f *at* $(1, 2)$ *in the direction of the* $\langle 3, 4\rangle$*-vector.*

Solution: *we identify this is a directional derivative problem. We need a point and a unit vector. The point is* $p_o = (1, 2)$*. However,* $||\langle 3, 4\rangle|| = \sqrt{9 + 16} = 5$ *hence we need to rescale the given vector before we calculate. Just divide by* 5 *to obtain* $\widehat{u} = \langle 3/5, 4/5\rangle$*. Calculate,*

$$f(\,\vec{p}_o + t\widehat{u}\,) = f(1 + 3t/5, 2 + 4t/5) = 25(1 + 3t/5)(2 + 4t/5) = (5 + 3t)(10 + 4t)$$

Differentiate, and then evaluate,

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ (5 + 3t)(10 + 4t)\ \bigg]\bigg|_{t=0} = \bigg[\ 3(10 + 4t) + 4(5 + 3t)\ \bigg]\bigg|_{t=0} = 30 + 20 = 50.$$

Naturally if you would rather calculate the difference quotient and take the limit you are free to do that. I choose to use the tools we've already developed, no sense in reinventing the wheel here. Incidentally, we will find a better way to package this calculation so you should look at this example as a means to better understand the definition. It is not computationally ideal. Neither is what follows, but these help bring understanding to later calculations so here we go.

Example 4.1.3. Problem: *Suppose* $f(x, y) = 25xy$ *and calculate the rate of change in* f *at* $(1, 2)$ *in the direction of the* **(a.)** $\langle 1, 0\rangle$*-vector,* **(b.)** $\langle 0, 1\rangle$*-vector.*

Solution of (a.): *I'll get straight to it here, identify* $\widehat{u} = \langle 1, 0\rangle$ *and* $\vec{p}_o = (1, 2)$ *and calculate*

$$f(\vec{p}_o + t\,\widehat{x}) = f(1 + t, 2) = 25(1 + t)(2) = 50 + 50t$$

Therefore,

$$D_{\widehat{x}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ 50 + 50t\ \bigg]\bigg|_{t=0} = 50.$$

Solution of (b.): *Identify* $\widehat{u} = \langle 0, 1\rangle$ *and* $\vec{p}_o = (1, 2)$ *and calculate*

$$f(\vec{p}_o + t\,\widehat{y}) = f(1, 2+t) = 25(1)(2+t) = 50 + 25t$$

Therefore,

$$D_{\widehat{y}} f(\vec{p}_o) = \frac{d}{dt}\left[\ 50 + 25t\ \right]\bigg|_{t=0} = 25.$$

Notice $50 = \frac{3}{5}(50) + \frac{4}{5}(25)$ hence the previous examples are related in a curious manner:

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{3}{5} D_{\widehat{x}} f(\vec{p}_o) + \frac{4}{5} D_{\widehat{y}} f(\vec{p}_o).$$

In other words, the pattern we see is:

$$D_{\langle a,b\rangle} f(\vec{p}_o) = \langle a, b\rangle \bullet \langle D_{\widehat{x}} f(\vec{p}_o), D_{\widehat{y}} f(\vec{p}_o)\rangle.$$

The directional derivatives in the coordinate directions are apparently important. We may be able to build the directional derivative in other directions[2]. This leads us to the topic of the next section. However, for the sake of logical completeness I define directional derivatives for functions of more than two variables. The visualization of the slopes implicit in the definition below are beyond most of our visual acumen.

Definition 4.1.4.

Let $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function with $\vec{p}_o \in dom(f)$ and suppose $\widehat{u} \in \mathbb{R}^n$ is a unit-vector. If the limit below exists, then we define the **directional derivative** of f at $\vec{p}_o$ in the $\widehat{u}$-direction by

$$D_{\widehat{u}} f(\vec{p}_o) = \lim_{t\to 0}\left[\frac{f(\,\vec{p}_o + t\widehat{u}\,) - f(\,\vec{p}_o\,)}{t}\right] = \frac{d}{dt}\left[\ f(\,\vec{p}_o + t\widehat{u}\,)\ \right]\bigg|_{t=0}.$$

We will calculate a few such directional derivatives in the section after the next once we understand the two-dimensional case in some depth.

[2] it turns out this is not generally true, but the exceptions are rare in applications

4.2 partial differentiation in $\mathbb{R}^2$

We continue the discussion of the last section concerning the change in functions of two variables. The formulas and concepts readily generalize to $n \geq 3$ however we postpone such discussion until we have settled the $n = 2$ theory.

Definition 4.2.1.

Let $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function with $(x_o, y_o) \in dom(f)$. If the directional derivative below exists, then we define the **partial derivative** of f at (x_o, y_o) with respect to x by

$$\frac{\partial f}{\partial x}(x_o, y_o) = (D_{\widehat{x}} f)(x_o, y_o).$$

Likewise, we define the **partial derivative** of f at (x_o, y_o) with respect to y by

$$\frac{\partial f}{\partial y}(x_o, y_o) = (D_{\widehat{y}} f)(x_o, y_o)$$

provided the directional derivative $(D_{\widehat{y}} f)(x_o, y_o)$ exists.

Notice that $(x_o, y_o) \mapsto \frac{\partial f}{\partial x}(x_o, y_o)$ and $(x_o, y_o) \mapsto \frac{\partial f}{\partial y}(x_o, y_o)$ define new multivariate functions provided the given function f possesses the necessary directional derivatives. We define higher derivatives by successive partial differentiation in the natural way: $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right]$. Derivatives such as $\frac{\partial^2 f}{\partial y^2}$ and $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are similarly defined. A brief notation for partial derivatives is as follows:

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{xy} = (f_x)_y = \frac{\partial^2 f}{\partial y \partial x}, \ etc...$$

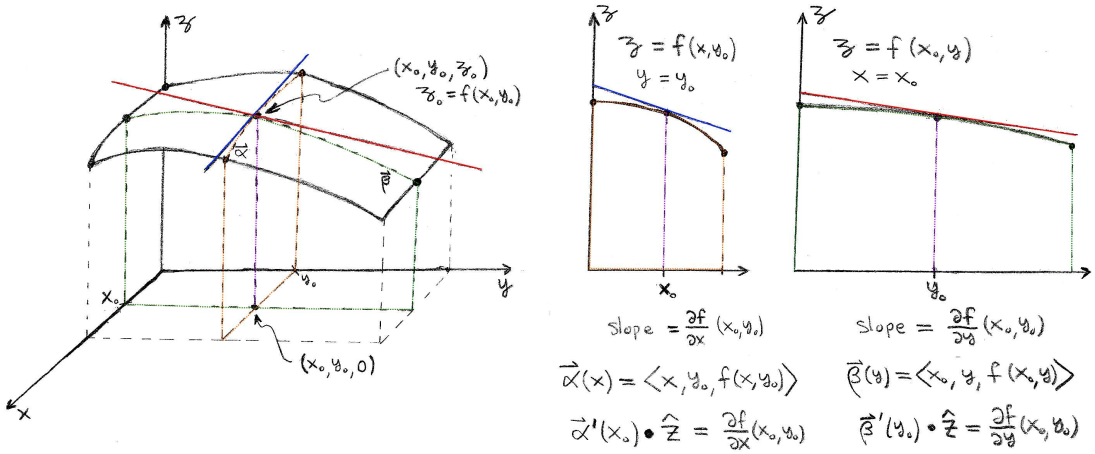
It is usually the case that $f_{xy} = f_{yx}$ but the proof of that statement is nontrivial and can be found in most advanced calculus texts. Given the connection of the partial derivative and the directional derivative we have the following conceptual guidelines:

> f_x **gives the rate of change in** f **in the** x**-direction.**
> f_y **gives the rate of change in** f **in the** y**-direction.**

It is also useful to rewrite the definition of the partial derivatives explicitly in terms of derivatives.

$$\boxed{\frac{\partial f}{\partial x}(x_o, y_o) = \frac{d}{dt}\bigg[\ f(x_o + t, y_o)\ \bigg]\bigg|_{t=0} \qquad \frac{\partial f}{\partial y}(x_o, y_o) = \frac{d}{dt}\bigg[\ f(x_o, y_o + t)\ \bigg]\bigg|_{t=0}.}$$

The geometry is revealed in the diagram below:



Well, how do these really work? The proposition below explains the working calculus of partial derivatives. It is really very simple.

Proposition 4.2.2.

Assume f, g are functions from $\mathbb{R}^2$ to $\mathbb{R}$ whose partial derivatives exist. Then for $c \in \mathbb{R}$,

1. $(f+g)_x = f_x + g_x$ and $(f+g)_y = f_y + g_y$.
2. $(cf)_x = cf_x$ and $(cf)_y = cf_y$.
3. $(fg)_x = f_x g + f g_x$ and $(fg)_y = f_y g + f g_y$.

Moreover, if $h : dom(h) \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable function then (4.)

$$\frac{\partial}{\partial x}\Big[h(f(x,y))\Big] = \frac{dh}{dt}\bigg|_{f(x,y)} \frac{\partial f}{\partial x} = \frac{dh}{df}\frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial}{\partial y}\Big[h(f(x,y))\Big] = \frac{dh}{dt}\bigg|_{f(x,y)} \frac{\partial f}{\partial y} = \frac{dh}{df}\frac{\partial f}{\partial y}.$$

5. $\dfrac{\partial x}{\partial x} = 1, \quad \dfrac{\partial x}{\partial y} = 0, \quad \dfrac{\partial y}{\partial x} = 0, \quad \dfrac{\partial y}{\partial y} = 1.$

Proof: the proofs of 1,2,3 follow immediately from the corresponding properties of single-variable differentiation. Let's work on the x-part of (4.)

$$\begin{aligned} \frac{\partial}{\partial x}\Big[h(f(x,y))\Big] &= \frac{d}{dt}\Big[\ h(f(x_o+t, y_o))\ \Big]\bigg|_{t=0} \\ &= \left(\frac{dh}{dt}\bigg|_{f(x_o+t,y_o)} \frac{d}{dt}\Big[\ f(x_o+t, y_o)\ \Big]\right)\bigg|_{t=0} \\ &= \frac{dh}{dt}\bigg|_{f(x,y)} \frac{\partial f}{\partial x}. \end{aligned}$$

We find that (4.) follows from the chain-rule of single-variable calculus. The proof in the y-variable is nearly the same. The proof of (5.) requires understanding of the definition. Let $F(x, y) = x$ and calculate

$$\frac{\partial F}{\partial x} = \frac{d}{dt}\left[F(x+t, y) \right]\Big)\Big|_{t=0} = \frac{d}{dt}\left[x+t \right]\Big)\Big|_{t=0} = 1.$$

$$\frac{\partial F}{\partial y} = \frac{d}{dt}\left[F(x, y+t) \right]\Big)\Big|_{t=0} = \frac{d}{dt}\left[x \right]\Big)\Big|_{t=0} = 0.$$

Likewise, let $G(x, y) = y$ and calculate,

$$\frac{\partial G}{\partial x} = \frac{d}{dt}\left[G(x+t, y) \right]\Big)\Big|_{t=0} = \frac{d}{dt}\left[y \right]\Big)\Big|_{t=0} = 0.$$

$$\frac{\partial G}{\partial y} = \frac{d}{dt}\left[G(x, y+t) \right]\Big)\Big|_{t=0} = \frac{d}{dt}\left[y+t \right]\Big)\Big|_{t=0} = 1.$$

Which concludes the proof of (5.) $\square$

Example 4.2.3. *Can you identify which property of the proposition I use in each line below?*

$$\begin{aligned}
\frac{\partial}{\partial x}\left[2^{x^2+y^2} \right] &= \ln(2)2^{x^2+y^2}\frac{\partial}{\partial x}(x^2+y^2) \\
&= \ln(2)2^{x^2+y^2}\left[\frac{\partial}{\partial x}(x^2) + \frac{\partial}{\partial x}(y^2)\right] \\
&= \ln(2)2^{x^2+y^2}\left[2x\frac{\partial x}{\partial x} + 2y\frac{\partial y}{\partial x}\right] \\
&= \boxed{2\ln(2)x2^{x^2+y^2}.}
\end{aligned}$$

Example 4.2.4. *Can you identify which property of the proposition I use in each line below?*

$$\begin{aligned}
\frac{\partial}{\partial x}[\ \sin(x^2y)\ ] &= \cos(x^2y)\frac{\partial}{\partial x}[\ x^2y\ ] \\
&= \cos(x^2y)\left[\frac{\partial}{\partial x}[x^2]y + x^2\frac{\partial y}{\partial x}\right] \\
&= \cos(x^2y)\left(2x\frac{\partial x}{\partial x}y\right) \\
&= \boxed{2xy\cos(x^2y).}
\end{aligned}$$

Similarly, you can calculate:

$$\frac{\partial}{\partial y}[\ \sin(x^2y)\ ] = \boxed{x^2\cos(x^2y).}$$

In practice I rarely write as many steps as I just offered in the examples above.

Example 4.2.5. *Power functions and exponential functions are different.*

$$\frac{\partial}{\partial x}\left[\ x^y\ \right] = yx^{y-1} \quad \textit{whereas} \quad \frac{\partial}{\partial y}\left[\ x^y\ \right] = \ln(x)x^y$$

$$\frac{\partial}{\partial y}\left[\ y^x\ \right] = xy^{x-1} \quad \textit{whereas} \quad \frac{\partial}{\partial x}\left[\ y^x\ \right] = \ln(y)y^x.$$

Example 4.2.6.

$$\frac{\partial}{\partial x}\left[\ x^{y^2}\ \right] = y^2x^{y^2-1} \quad \textit{whereas} \quad \frac{\partial}{\partial y}\left[\ x^{y^2}\ \right] = \ln(x)x^{y^2}\frac{\partial y^2}{\partial y} = \boxed{2y\ln(x)x^{y^2}.}$$

Example 4.2.7.

$$\begin{aligned}\frac{\partial}{\partial x}\left[\ \sin\big(x^2y\cosh(x)\big)\ \right] &= \cos\big(x^2y\cosh(x)\big)\frac{\partial}{\partial x}\big(x^2y\cosh(x)\big). \\ &= \boxed{\big(2xy\cosh(x) + x^2y\sinh(x)\big)\cos\big(x^2y\cosh(x)\big).}\end{aligned}$$

Can I skip the middle step in the example above? Some days yes. Should you? Probably not.

Example 4.2.8.

$$\begin{aligned}\frac{\partial}{\partial y}\left[\ \sin\big(\cos(\sqrt{xy})\big)\ \right] &= \cos\big(\cos(\sqrt{xy})\big)\frac{\partial}{\partial y}\big(\cos(\sqrt{xy})\big). \\ &= \cos\big(\cos(\sqrt{xy})\big)\big(-\sin(\sqrt{xy})\big)\frac{\partial}{\partial y}\sqrt{xy}. \\ &= \cos\big(\cos(\sqrt{xy})\big)\big(-\sin(\sqrt{xy})\big)\frac{1}{2\sqrt{xy}}\frac{\partial}{\partial y}[xy]. \\ &= \boxed{-\frac{1}{2}\sqrt{\frac{x}{y}}\sin\big(\sqrt{xy}\,\big)\cos\big(\cos(\sqrt{xy})\,\big).}\end{aligned}$$

Example 4.2.9. .

$\boxed{E49}$ $F(x,y) = x^2 + y^2.$

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}\left[x^2 + y^2\right] = \frac{\partial}{\partial x}\left[x^2\right] + \frac{\partial}{\partial x}\cancelto{0}{\left[y^2\right]} = \boxed{2x}$$

: y is constant with respect to x

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\left[x^2 + y^2\right] = \boxed{2y}$$

: we regard x as constant as we perform the $\partial/\partial y$ operation.

Example 4.2.10. .

E50 $F(x,y) = xe^{xy}$

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}(xe^{xy}) = \frac{\partial x}{\partial x}e^{xy} + x\frac{\partial}{\partial x}(e^{xy}) \quad \text{: product rule}$$

$$\boxed{F_x = e^{xy} + xye^{xy}} \quad \text{: chain rule, remember } y \text{ is regarded constant}$$

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}(xe^{xy}) = \underbrace{x\frac{\partial}{\partial y}(e^{xy}) = xe^{xy}\frac{\partial}{\partial y}(xy)}_{\text{here I wrote out the chain rule, not always need, but may help in messy cases.}} = \boxed{x^2e^{xy} = F_y}$$

Example 4.2.11. .

E51 $z^2 = \sin(xy) + x + \ln(y)$. Suppose that x & y are independent and z is dependent; $z = z(x,y)$. We use implicit differention to find implicit formulas for z_x and z_y.

$$\frac{\partial}{\partial x}[z^2] = 2z\frac{\partial z}{\partial x}$$

$$\frac{\partial}{\partial x}[\sin(xy) + x + \ln(y)] = y\cos(xy) + 1$$

But these are equal so $\boxed{\dfrac{\partial z}{\partial x} = \dfrac{1}{2z}[y\cos(xy) + 1]}$ question: why is this implicit?

Likewise we calculate,

$$2z\frac{\partial z}{\partial y} = x\cos(xy) + \frac{1}{y} \quad \therefore \quad \boxed{\frac{\partial z}{\partial y} = \frac{1}{2z}\left[x\cos(xy) + \frac{1}{y}\right]}$$

Example 4.2.12. .

E54 Let $f(x,y) = xy^2$.

$$f_{xx} \equiv \frac{\partial^2 f}{\partial x^2} \equiv \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right] = \frac{\partial}{\partial x}[y^2] = 0.$$

$$f_{yx} \equiv \frac{\partial^2 f}{\partial x \partial y} \equiv \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right] = \frac{\partial}{\partial x}[2xy] = 2y.$$

$$f_{xy} \equiv \frac{\partial^2 f}{\partial y \partial x} \equiv \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial x}\right] = \frac{\partial}{\partial y}[y^2] = 2y.$$

$$f_{yy} \equiv \frac{\partial^2 f}{\partial y^2} \equiv \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}\right] = \frac{\partial}{\partial y}[2yx] = 2x.$$

interesting, is it always the case that $f_{xy} = f_{yx}$?

Example 4.2.13. .

$$f(x,t) = x^2 e^{-ct} \quad \text{where } c \text{ is a constant.}$$

$$\begin{aligned}
f_{ttt} &= \frac{\partial}{\partial t}\frac{\partial}{\partial t}\frac{\partial}{\partial t}\left(x^2 e^{-ct}\right) \\
&= x^2 \frac{\partial}{\partial t}\frac{\partial}{\partial t}\left(-ce^{-ct}\right) \\
&= x^2 \frac{\partial}{\partial t}\left(c^2 e^{-ct}\right) \\
&= \boxed{-x^2 c^3 e^{-ct} = f_{ttt}}
\end{aligned}$$

$$\begin{aligned}
f_{txx} &= \frac{\partial}{\partial t}\frac{\partial}{\partial x}\frac{\partial}{\partial x}\left(x^2 e^{-ct}\right) \\
&= \frac{\partial}{\partial t}\frac{\partial}{\partial x}\left(2x e^{-ct}\right) \\
&= \frac{\partial}{\partial t}\left(2e^{-ct}\right) \\
&= \boxed{-2c\, e^{-ct} = f_{txx}}
\end{aligned}$$

Example 4.2.14. .

Let $f(x,y) = \ln\left(x + \sqrt{x^2+y^2}\right)$ find $f_x(3,4)$

$$\begin{aligned}
f_x &= \frac{\partial}{\partial x}\left[\ln\left(x + \sqrt{x^2+y^2}\right)\right] \\
&= \frac{1}{x+\sqrt{x^2+y^2}}\frac{\partial}{\partial x}\left[x + \sqrt{x^2+y^2}\right] \\
&= \frac{1}{x+\sqrt{x^2+y^2}}\left(1 + \frac{1}{2\sqrt{x^2+y^2}}\frac{\partial}{\partial x}\left(x^2+y^2\right)\right) \\
&= \frac{1}{x+\sqrt{x^2+y^2}}\left(1 + \frac{x}{\sqrt{x^2+y^2}}\right)
\end{aligned}$$

Thus, $f_x(3,4) = \dfrac{1}{3+\sqrt{9+16}}\left(1 + \dfrac{3}{\sqrt{9+16}}\right) = \dfrac{1}{8}\left[\dfrac{5}{5} + \dfrac{3}{5}\right]$

$$\therefore \boxed{f_x(3,4) = \frac{1}{5}}$$

Example 4.2.15. .

$$f(x,y) = x^3y^5 + 2x^4y$$

$$f_x = 3x^2y^5 + 8x^3y \qquad f_y = 5x^3y^4 + 2x^4$$

$$\boxed{\begin{array}{ll} f_{xx} = 6xy^5 + 24x^2y & f_{yy} = 20x^3y^3 \\ f_{xy} = 15x^2y^4 + 8x^3 & f_{yx} = 15x^2y^4 + 8x^3 \end{array}}$$

Example 4.2.16. .

$$f(s,t) = st^2/(s^2+t^2)$$

$$\frac{\partial f}{\partial s} = t^2 \frac{\partial}{\partial s}\left[\frac{s}{s^2+t^2}\right] = t^2\left[\frac{s^2+t^2-2s\cdot s}{(s^2+t^2)^2}\right] = \boxed{\frac{t^2(t^2-s^2)}{(s^2+t^2)^2}}$$

$$\frac{\partial f}{\partial t} = s\frac{\partial}{\partial t}\left[\frac{t^2}{s^2+t^2}\right] = s\left[\frac{2t(s^2+\cancel{t^2}) - \cancel{2t\cdot t^2}}{(s^2+t^2)^2}\right] = \boxed{\frac{2ts^3}{(s^2+t^2)^2}}$$

Example 4.2.17. .

$$f(x,y) = \int_y^x \cos(t^2)\,dt = F(x) - F(y) \quad \text{where} \quad F'(u) = \cos(u^2).$$

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left[\int_y^x \cos(t^2)\,dt\right] = \frac{\partial}{\partial x}\left(F(x) - \cancel{F(y)}^{\,0}\right) = \frac{\partial F}{\partial x}(x) = \cos(x^2).$$

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\left[\int_y^x \cos(t^2)\,dt\right] = \frac{\partial}{\partial y}\left(\cancel{F(x)}^{\,0} - F(y)\right) = \frac{\partial F}{\partial y}(y) = \cos(y^2).$$

Example 4.2.18. .

$$\text{Let } w = \sin\alpha\cos\beta$$

$$\frac{\partial w}{\partial \alpha} = \frac{\partial}{\partial\alpha}\left[\sin\alpha\cos\beta\right] = \cos\beta\frac{\partial}{\partial\alpha}\left[\sin\alpha\right] = \boxed{\cos\beta\cos\alpha}$$

$$\frac{\partial w}{\partial \beta} = \frac{\partial}{\partial\beta}\left[\sin\alpha\cos\beta\right] = \sin\alpha\frac{\partial}{\partial\beta}\left[\cos\beta\right] = \boxed{-\sin\alpha\sin\beta}$$

Example 4.2.19. .

find $\partial z/\partial x$ and $\partial z/\partial y$,

$$x - z = \tan^{-1}(yz)$$

Differentiate with respect to x (hold y fixed),

$$\frac{\partial}{\partial x}(x-z) = \frac{\partial}{\partial x}\left[\tan^{-1}(yz)\right]$$

$$1 - \frac{\partial z}{\partial x} = \frac{1}{1+(yz)^2}\frac{\partial}{\partial x}[yz] = \left[\frac{y}{1+y^2z^2}\right]\frac{\partial z}{\partial x}$$

$$1 = \left[1 + \frac{y}{1+y^2z^2}\right]\frac{\partial z}{\partial x} \quad \therefore \quad \boxed{\frac{\partial z}{\partial x} = \frac{1}{1+\frac{y}{1+y^2z^2}}}$$

could simplify a bit more,

$$\boxed{z_x = \frac{1+y^2z^2}{1+y^2z^2+y}}$$

Likewise,

$$\frac{\partial}{\partial y}(x-z) = \frac{\partial}{\partial y}\left[\tan^{-1}(yz)\right]$$

$$-\frac{\partial z}{\partial y} = \frac{1}{1+y^2z^2}\frac{\partial}{\partial y}[yz]$$

$$-\frac{\partial z}{\partial y} = \frac{1}{1+y^2z^2}\left[z + y\frac{\partial z}{\partial y}\right]$$

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

$$\therefore \quad \frac{\partial z}{\partial y} = \left(\frac{1}{1+\frac{y}{1+y^2z^2}}\right)\left(\frac{-z}{1+y^2z^2}\right)$$

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

Example 4.2.20. .

find 2^{nd} partials of $z = y\tan(2x)$

$$\frac{\partial z}{\partial y} = \tan(2x) \quad \text{then} \quad \frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left[\tan(2x)\right] = 0.$$

$$\frac{\partial z}{\partial x} = 2y\sec^2(2x) \quad \text{then} \quad \frac{\partial^2 z}{\partial x^2} = \frac{\partial}{\partial x}\left[2y\sec^2(2x)\right]$$

$$= 2y\cdot 2\sec(2x)\cdot\frac{\partial}{\partial x}\left[\sec(2x)\right]$$

$$= 4y\sec(2x)\sec(2x)\tan(2x)\cdot 2$$

$$= \boxed{8y\sec^2(2x)\tan(2x) = \frac{\partial^2 z}{\partial x^2}}$$

You can calculate the remaining 2^{nd} order partial derivative which is z_{xy}.

4.2.1 directional derivatives and the gradient in $\mathbb{R}^2$

Now that we have a little experience in partial differentiation let's return to the problem of the directional derivative. We saw that

$$D_{\langle a,b\rangle} f(x_o, y_o) = \langle\ f_x(x_o, y_o),\ f_y(x_o, y_o)\ \rangle \bullet \langle a, b\rangle$$

for the particular example we considered. Is this always true? Is it generally the case that we can build the directional derivative in the $\langle a, b\rangle$-direction from the partial derivatives? If you just try most functions that come to the nonpathological mind then you'd be tempted to agree with this claim. However, many counter-examples exist. We only need one to debunk the claim.

Example 4.2.21. *Suppose that*

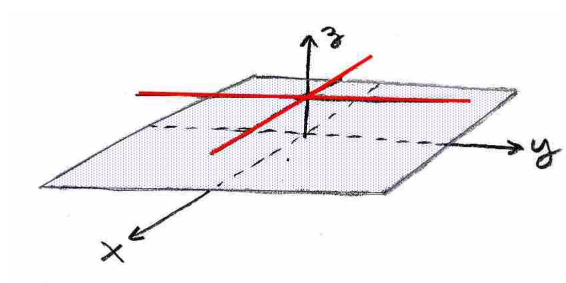
$$f(x, y) = \begin{cases} x+1 & y = 0 \\ y+1 & x = 0 \\ 0 & xy \neq 0 \end{cases}$$

Clearly $f_x(0, 0) = 1$ and $f_y(0, 0) = 1$ however the directional derivative is given by

$$D_{\langle a,b\rangle} f(0, 0) = \lim_{t\to 0} \frac{f(ta, tb) - f(0, 0)}{t} = \lim_{t\to 0} \frac{-1}{t}$$

which diverges. The directional derivative in any non-coordinate direction does not exist since the function jumps from 0 to 1 at the origin along any line except the axes.

Example 4.2.22. *This example is even easier: let $f(x, y) = \begin{cases} 1 & y = 0 \\ 1 & x = 0 \\ 0 & xy \neq 0 \end{cases}$. In this case I can graph the function and it is obvious that $f_x(0, 0) = 0$ and $f_y(0, 0) = 0$ yet all the directional derivatives in non-coordinate directions fail to exist.*



We can easily see the discontinuity of the function above is the source of the trouble. It is sometimes true that a function is discontinuous and the formula holds. However, the case which we really want to consider, the type of functions for which the derivatives considered are most meaningful, are called **continuously differentiable**. You might recall from single-variable calculus that when a function is differentiable at a point but the derivative function is discontinuous it led to bizzare features for the linearization. That continues to be true in the multivariate case.

Definition 4.2.23.

A function $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ is said to be **continuously differentiable** at (x_o, y_o) iff the partial derivative functions $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ are continuous at (x_o, y_o). We say $f \in C^1(x_o, y_o)$. If all the second-order partial derivatives of f are continuous at (x_o, y_o) then we say $f \in C^2(x_o, y_o)$. If continuous partial derivatives of arbitrary order exist at (x_o, y_o) then we say f is **smooth** and write $f \in C^\infty(x_o, y_o)$.

The continuity of the partial derivative functions implicitly involves multivariate limits and this is what ultimately makes this criteria quite strong.

Proposition 4.2.24.

Suppose f is continuously differentiable at (x_o, y_o) then the directional derivative at (x_o, y_o) in the direction of the unit vector $\langle a, b\rangle$ is given by:

$$D_{\langle a,b\rangle} f(x_o, y_o) = \langle\ f_x(x_o, y_o),\ \ f_y(x_o, y_o)\ \rangle \bullet \langle a, b\rangle$$

Proof: delayed until the next section. $\square$

At this point it is useful to introduce a convenient notation which groups all the partial derivatives together in a particular vector of functions.

Definition 4.2.25.

If the partial derivatives of f exist then we define

$$\nabla f = \langle f_x,\ f_y\rangle = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y}.$$

we also use the notation $grad(f)$ and call this the **gradient** of f.

The upside-down triangle ∇ is also known as *nabla*. Identify that $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y$ is a vector of operators, it takes a function f and produces a vector field ∇f. This is called the **gradient vector field** of f. We'll think more about that after the examples. For a continuously differentiable function we have the following beautiful formula for the directional derivative:

$$\boxed{D_{\langle a,b\rangle} f(x_o, y_o) = (\nabla f)(x_o, y_o) \bullet \langle a, b\rangle.}$$

This is the formula I advocate for calculation of directional derivatives. This formula most elegantly summarizes how the directional derivative works. I'd make it the definition, but the discontinuous[3] counter-Example 4.2.21 already spoiled our fun.

[3] I don't mean to say there are no continuous counter examples,I'd wager there are examples of continuous functions whose partial derivatives exist but are discontinuous. Then the formula fails because some non-coordinate directions fail to possess a directional derivative.

Example 4.2.26. *Suppose* $f(x,y) = x^2 + y^2$. *Then*

$$\nabla f = \langle 2x, 2y \rangle.$$

Calculate the directional derivative of f *at* (x_o, y_o) *in the* $\langle a,b\rangle$*-direction:*

$$D_{\langle a,b\rangle} f(x_o,y_o) = \langle 2x_o, 2y_o\rangle \bullet \langle a,b\rangle = 2x_o a + 2y_o b.$$

It is often useful to write $D_{\langle a,b\rangle} f(x_o,y_o) = (\nabla f)(x_o,y_o) \bullet \langle a,b\rangle$ in terms of the angle θ between the $\nabla f(x_o,y_o)$ and $\langle a,b\rangle$:

$$\boxed{D_{\langle a,b\rangle} f(x_o,y_o) = ||(\nabla f)(x_o,y_o)|| \cos\theta.}$$

With this formula the following are obvious:

1. $(\theta = 0)$ when $\langle a,b\rangle$ is parallel to $(\nabla f)(x_o,y_o)$ the direction $\langle a,b\rangle$ points towards **maximum increase** in f
2. $(\theta = \pi)$ when $\langle a,b\rangle$ is antiparallel to $(\nabla f)(x_o,y_o)$ the direction $\langle a,b\rangle$ points towards **maximum decrease** in f
3. $(\theta = \pi/2)$ when $\langle a,b\rangle$ is perpendicular to $(\nabla f)(x_o,y_o)$ the direction $\langle a,b\rangle$ points towards where f remains **constant**.

Example 4.2.27. Problem: *if* $f(x,y) = x^2 + y^2$. *Then in what direction(s) is(are)* f **(a.)** *increasing the most at* $(2,3)$, **(b.)** *decreasing the most at* $(2,3)$, **(c.)** *not increasing at* $(2,3)$*?*

Solution of (a.): f *increases most in the* $(\nabla f)(2,3)$*-direction. In particular,* $(\nabla f)(2,3) = \langle 4,6\rangle$. *If you prefer a unit-vector then rescale* $\langle 4,6\rangle\rangle$ *to* $\widehat{u} = \frac{1}{\sqrt{13}}\langle 2,3\rangle$. *The magnitude* $||(\nabla f)(2,3)|| = \sqrt{13}$ *is the rate of increase in the* $\widehat{u} = \frac{1}{\sqrt{13}}\langle 2,3\rangle$*-direction.*

Solution of (b.): f *decreases most in the* $-(\nabla f)(2,3)$*-direction. In particular,* $-(\nabla f)(2,3) = \langle -4,-6\rangle$. *If you prefer a unit-vector then rescale* $\langle -4,-6\rangle\rangle$ *to* $\widehat{u} = \frac{1}{\sqrt{13}}\langle -2,-3\rangle$. *The rate of decrease is also* $\sqrt{13}$ *in magnitude.*

Solution of (c.): f *is constant in directions which are perpendicular to* $(\nabla f)(2,3)$. *A unit-vector which is perpendicular to* $(\nabla f)(2,3) = \langle 4,6\rangle$ *satisfied two conditions:*

$$(\nabla f)(2,3) \bullet \langle a,b\rangle = 4a + 6b = 0 \qquad \text{and} \qquad a^2 + b^2 = 1$$

These are easily solved by solving the orthogonality condition for $b = -\frac{2}{3}a$ *and substituting it into the unit-length condition:*

$$1 = a^2 + b^2 = a^2 + \frac{4}{9}a^2 = \frac{13}{9}a^2 \quad \Rightarrow \quad a^2 = \frac{9}{13} \quad \Rightarrow \quad a = \pm\frac{3}{\sqrt{13}} \quad \Rightarrow b = \mp\frac{2}{\sqrt{13}}.$$

Therefore, we find f *is constant in either the* $\langle 3/\sqrt{13}, -2/\sqrt{13}\rangle$ *or the* $\langle -3/\sqrt{13}, 2/\sqrt{13}\rangle$ *direction.*

Example 4.2.28. **Problem:** *find a point (x_o, y_o) at which the function $f(x,y) = x^2 + y^2$ is constant in all directions.*

Solution: *We need to find a point (x_o, y_o) at which $(\nabla f)(x_o, y_o)$ is perpendicular to all unit-vectors. The only vector which is perpendicular to all other vectors is the zero vector. We seek solutions to $(\nabla f)(x_o, y_o) = \langle 2x_o, 2y_o \rangle = \langle 0, 0 \rangle$. The only solution is $x_o = 0$ and $y_o = 0$. Apparently the graph $z = f(x,y)$ levels out at the origin since $f(x,y)$ stays constant in all directions near $(0,0)$.*

Definition 4.2.29.

We say (x_o, y_o) is a **critical point** of f if $(\nabla f)(x_o, y_o)$ does not exist or $(\nabla f)(x_o, y_o) = \langle 0, 0 \rangle$.

The term critical point is appropriate here since these are points where the function f *may* have a local maximum or minimum. Other possibilities exist and we'll spend a few lectures this semester developing tools to carefully discern what the geometry is near a given critical point.

Example 4.2.30. .

Let $f(x,y) = y^2/x$ and consider $P = (1,2)$ and the unit vector $\hat{u} = \langle 2/3, \sqrt{5}/3 \rangle$ find

a.) $\nabla f = \langle f_x, f_y \rangle$ thus, $\boxed{\nabla f = \langle -y^2/x^2, 2y/x \rangle}$

b.) $\nabla f(1,2) = \langle -4/1, 4/1 \rangle = \boxed{\langle -4, 4 \rangle = \nabla f(1,2)}$

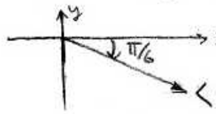
c.)
$$\begin{aligned}(D_{\hat{u}} f)(P) &= (\nabla f)(P) \cdot \hat{u} \\ &= \langle -4, 4 \rangle \cdot \langle 2/3, \sqrt{5}/3 \rangle \\ &= -8/3 + 4\sqrt{5}/3 \\ &= \boxed{\tfrac{1}{3}(4\sqrt{5} - 8)}.\end{aligned}$$

Example 4.2.31. .

Let $f(x,y) = -\sqrt{5x-4y}$ find Df at $(4,1)$ in $\theta = -\pi/6$ direction

$$(\nabla f)(x,y) = \left\langle \frac{5}{2\sqrt{5x-4y}}, \frac{-2}{\sqrt{5x-4y}} \right\rangle \Rightarrow (\nabla f)(4,1) = \left\langle \frac{5}{2\sqrt{16}}, \frac{-2}{\sqrt{16}} \right\rangle$$

Thus $(\nabla f)(4,1) = \langle 5/8, -1/2 \rangle$. Now lets find the unit vector in the $\theta = -\pi/6$ direction, here θ is the usual polar coordinate.



$$\langle \cos(-\pi/6), \sin(-\pi/6) \rangle = \left\langle \frac{\sqrt{3}}{2}, \frac{-1}{2} \right\rangle = \vec{u}$$

$$\begin{aligned} D_{\vec{u}} f(4,1) &= (\nabla f)(4,1) \bullet \left\langle \frac{\sqrt{3}}{2}, \frac{-1}{2} \right\rangle \\ &= \langle 5/8, -1/2 \rangle \bullet \langle \sqrt{3}/2, -1/2 \rangle \\ &= \boxed{\frac{5\sqrt{3}}{16} + \frac{1}{4}} \end{aligned}$$

Example 4.2.32. .

Let $f(x,y) = 5xy^2 - 4x^3y$.

$$\nabla f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle = \langle 5y^2 - 12x^2y, 10xy - 4x^3 \rangle$$

$$(\nabla f)(1,2) = \langle 20 - 12(1)(2), 10(1)(2) - 4(1) \rangle = \langle -4, 16 \rangle$$

The rate of change of f at the point $P = (1,2)$ in the $\vec{u} = \langle \frac{5}{13}, \frac{12}{13} \rangle$ direction is the directional derivative, we should check that $\vec{u} = \hat{u}$ notice $|\vec{u}| = \sqrt{\frac{1}{13^2}(5^2 + 12^2)} = \sqrt{\frac{1}{169}(169)} = 1$ $\therefore$ $\vec{u}$ is unit vector. Thus,

$$\begin{aligned} (D_{\vec{u}} f)(P) &= (\nabla f)(P) \bullet \vec{u} \\ &= \langle -4, 16 \rangle \bullet \langle 5/13, 12/13 \rangle \\ &= \frac{1}{13}(-20 + 192) = \boxed{172/13} \end{aligned}$$

Example 4.2.33. .

Let $f(x,y) = \ln(x^2+y^2)$ find $(D_{\hat{v}}f)(2,1)$ for $\vec{v} = \langle -1, 2\rangle$.
Notice that $|\vec{v}| = \sqrt{5}$ thus $\hat{v} = \frac{1}{\sqrt{5}}\langle -1, 2\rangle$. You can check $|\hat{v}| = 1$.

$$\nabla f = \left\langle \frac{2x}{x^2+y^2}, \frac{2y}{x^2+y^2} \right\rangle$$

$$(\nabla f)(2,1) = \left\langle \frac{4}{5}, \frac{2}{5} \right\rangle$$

$$(\nabla f)(2,1)\cdot\left(\frac{1}{\sqrt{5}}\langle -1, 2\rangle\right) = \frac{1}{\sqrt{5}}\frac{1}{5}\langle 4, 2\rangle \cdot \langle -1, 2\rangle = \frac{1}{5\sqrt{5}}(-4+4) = 0$$

Thus we find $\boxed{(D_{\hat{v}}f)(2,1) = 0}$

Example 4.2.34. .

Find directions in which $f(x,y) = x^2 + \sin(xy)$ has directional derivative at $(1,0)$ with value 1. That is find $\hat{u}$ such that $(D_{\hat{u}}f)(1,0) = 1$. For our convenience let us define a, b unknowns such that $\hat{u} = \langle a, b\rangle$. $(a^2+b^2=1)$

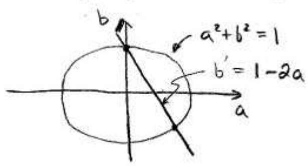
$$\begin{aligned} \nabla f &= \left\langle \frac{\partial}{\partial x}[x^2+\sin(xy)], \frac{\partial}{\partial y}[x^2+\sin(xy)] \right\rangle \\ &= \left\langle 2x + \cos(xy)\frac{\partial}{\partial x}[xy], \cos(xy)\frac{\partial}{\partial y}[xy] \right\rangle \\ &= \langle 2x + y\cos(xy), x\cos(xy) \rangle \end{aligned}$$

chain-rule.

Now $(\nabla f)(1,0) = \langle 2, 1\rangle$. We wish to study $(D_{\hat{u}}f)(1,0) = 1$, that is,

$$(\nabla f)(1,0)\cdot\langle a,b\rangle = \langle 2,1\rangle \cdot \langle a, b\rangle = 2a+b = 1$$

The eq$^{\underline{n}}$ $2a+b$ has ∞ly many sol$^{\underline{n}}$'s <u>But</u> we also demand that $a^2+b^2=1$ since we wish to find the <u>directions</u> in which $(D_{\hat{u}}f)(1,0) = 1$.



- you can see we get two sol$^{\underline{n}}$'s from the graph.
- algebraically we find them as follows,

$$\begin{aligned} 1 = a^2+b^2 &= a^2 + (1-2a)^2 \quad \text{: substituting} \\ &= a^2+1-4a+4a^2 \\ &= 5a^2-4a+1 \end{aligned}$$

$$\therefore\ 5a^2-4a = a(5a-4) = 0$$

$$\underline{a = 0 \quad \text{or} \quad a = 4/5}$$

Thus $\hat{u} = \langle a, b\rangle = \langle a, 1-2a\rangle$ should be $\boxed{\langle 0, 1\rangle \text{ or } \langle 4/5, 3/5\rangle}$.

4.2.2 gradient vector fields

We've seen that the value of ∇f at a particular point reveals both the magnitude and the direction of the change in the function f. The gradient vector field is simply the vector field which a differentiable function f generates through the gradient operation.

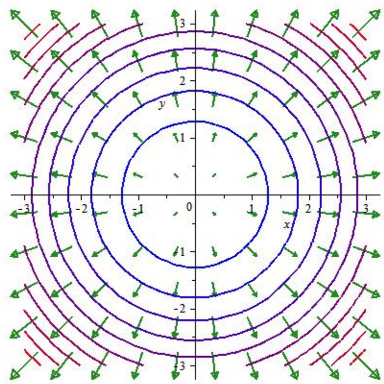
Definition 4.2.35.

If f is differentiable on $U \subseteq \mathbb{R}^2$ then ∇f defines the gradient vector field on U. We assign to each point $\vec{p} \in U$ the vector $\nabla f(\vec{p})$.

Example 4.2.36. *Let $f(x, y) = x^2 + y^2$. We calculate,*

$$\nabla f(x, y) = \langle \partial_x(x^2 + y^2), \partial_y(x^2 + y^2)\rangle = \langle 2x, 2y\rangle$$

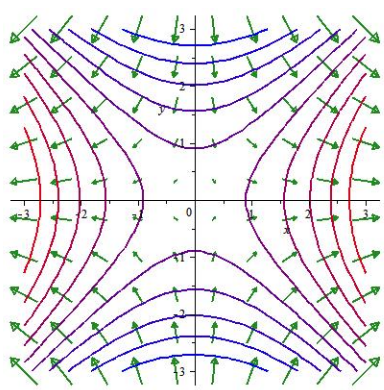
This gradient vector field is easily described; at each point $\vec{p}$ we attach the vector $2\vec{p}$.



Example 4.2.37. *Let $f(x, y) = x^2 - y^2$. We calculate,*

$$\nabla f(x, y) = \langle \partial_x(x^2 - y^2), \partial_y(x^2 - y^2)\rangle = \langle 2x, -2y\rangle$$

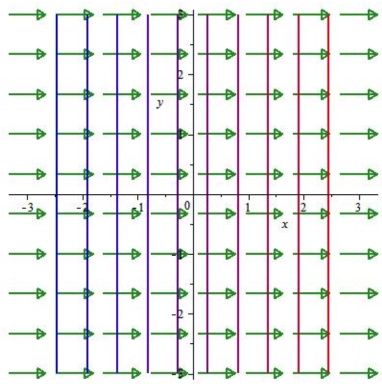
This gradient vector field is not so easily described, however, most CAS will provide nice plots if you are willing to invest a little time.



Example 4.2.38. *Let $f(x,y) = x$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(x), \partial_y(x) \rangle = \langle 1, 0 \rangle = \widehat{x}$$

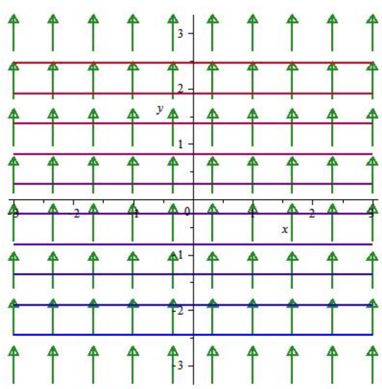
Therefore, $\nabla x = \widehat{x}$. Interesting. The gradient operation reproduces the unit-vector in the direction of increasing x.



Example 4.2.39. *Let $f(x,y) = y$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(y), \partial_y(y) \rangle = \langle 0, 1 \rangle = \widehat{y}$$

Therefore, $\nabla y = \widehat{y}$. Interesting. The gradient operation reproduces the unit-vector in the direction of increasing y.



Naturally, we are tempted to derive other unit-vector-fields by this method. In the examples above we were a bit lucky, generally when you take the gradient of a coordinate function you'll need to normalize it. But, this is a very nice **algebraic** method to derive the frame of a non-cartesian coordinate system. In particular, if y_1, y_2 are coordinates then there exist differentiable functions f_1, f_2 such that $y_1 = f_1(x,y)$ and $y_2 = f_2(x,y)$ we can calculate the unit-vectors

$$\widehat{y_1} = \frac{\nabla f_1}{||\nabla f_1||} \qquad \text{and} \qquad \widehat{y_2} = \frac{\nabla f_2}{||\nabla f_2||}.$$

Let's see how this method produces the frame for polar coordinates. I initially claimed it could be derived from geometry alone. That is true, but this is also nice:

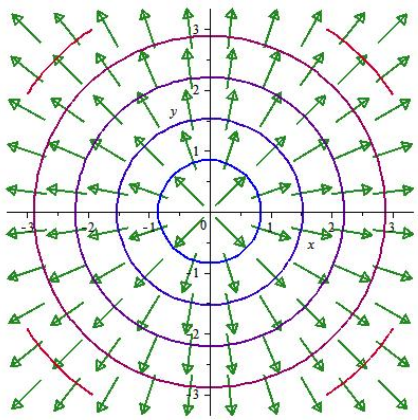
Example 4.2.40. *Consider polar coordinates r, θ, these were defined by $r = \sqrt{x^2+y^2}$ and $\theta = \tan^{-1}\big[y/x\big]$ for $x > 0$. Calculate,*

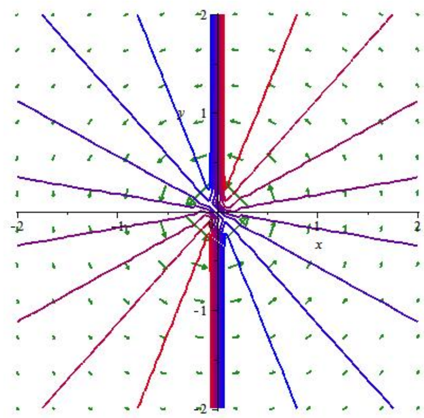
$$\nabla r = \left\langle \frac{\partial}{\partial x}\sqrt{x^2+y^2},\ \frac{\partial}{\partial y}\sqrt{x^2+y^2} \right\rangle = \left\langle \frac{x}{\sqrt{x^2+y^2}},\ \frac{y}{\sqrt{x^2+y^2}} \right\rangle = \left\langle \frac{x}{r},\ \frac{y}{r} \right\rangle$$

But, $x = r\cos\theta$ and $y = r\sin\theta$ thus we derive $\nabla r = \langle \cos\theta, \sin\theta\rangle$. Since $||\nabla r|| = 1$ we find $\boxed{\widehat{r} = \langle \cos\theta, \sin\theta\rangle.}$ *The unit-vector in the direction of increasing θ is likewise calculated,*

$$\nabla \theta = \left\langle \frac{\partial}{\partial x}\tan^{-1}\big[y/x\big],\ \frac{\partial}{\partial y}\tan^{-1}\big[y/x\big] \right\rangle = \left\langle \frac{-y}{x^2+y^2},\ \frac{x}{x^2+y^2} \right\rangle = \left\langle \frac{-y}{r^2},\ \frac{x}{r^2} \right\rangle.$$

In this case we find $\nabla\theta = \frac{1}{r}\langle -\sin\theta, \cos\theta\rangle$. Gradients and level curves of r and θ are plotted below[4]*:*





The gradient of θ is not a unit-vector so we have to normalize. Since $||\nabla\theta|| = \frac{1}{r}$ we derive $\boxed{\widehat{\theta} = \langle -\sin\theta, \cos\theta\rangle.}$

This is a very nice calculation for coordinates which are not easy to visualize.

Another nice application of the gradient involves level curves. Consider this: a level curve is the set of points which solves $f(x, y) = k$ for some value k. If we consider a point (x_o, y_o) on the level curve $f(x, y) = k$ then the gradient vector $(\nabla f)(x_o, y_o)$ will be perpendicular to the tangent line of the level curve. Remember that when $\theta = \pi/2$ we find a direction in which $f(x, y)$ stays constant near (x_o, y_o). What does this mean? Let's summarize it:

The gradient vector field ∇f is perpendicular to the level curve $f(x, y) = k$.

If you are less than satisfied with my geometric justification for this claim then you'll be happy to hear we can prove it with a simple calculation. However, we need a chain-rule which we have yet to justify. Therefore, further justification is postponed until a later section. That said, let's look at a few examples to appreciate the power of this statement:

[4] notice how the software chokes on $x = 0$

Example 4.2.41. *Suppose $V(x,y) = \frac{1}{\sqrt{x^2+y^2}}$ represents the voltage due to a point-charge at the origin. Electrostatics states that the electric field $\vec{E} = -\nabla V$. Geometrically this has a simple meaning; the electric field points along the normal direction to the level-curves of the voltage function. In other words, the electric field is normal to the equipotential lines. What is an "equipotential line", it's a line on which the voltage assumes a constant value. This is nothing more than a level-curve of the voltage function. For the given potential function, using $r = \sqrt{x^2+y^2}$,*

$$\nabla V = \langle \partial_x(1/r), \partial_y(1/r)\rangle = \langle (-1/r^2)\partial_x r,\ (-1/r^2)\partial_y r\rangle = \frac{-1}{r^2}\langle \partial_x r,\ \partial_y r\rangle = -\frac{1}{r^2}\widehat{r}.$$

Equipotentials $V = V_o = 1/r$ are simply circles $r = 1/V_o$ and the electric field is a purely radial field $\vec{E} = \frac{1}{r^2}\widehat{r}$.

Example 4.2.42. *Consider the ellipse $f(x,y) = x^2/a^2 + y^2/b^2 = k$. At any point on the ellipse the vector field*

$$\nabla f = \frac{2x}{a^2}\,\widehat{x} + \frac{2y}{b^2}\,\widehat{y}$$

points in the normal direction to the ellipse.

Example 4.2.43. *Consider the hyperbolas $g(x,y) = x^2y^2 = k$. At any point on the hyperbolas the vector field*

$$\nabla g = 2xy^2\,\widehat{x} + 2x^2y\,\widehat{y}$$

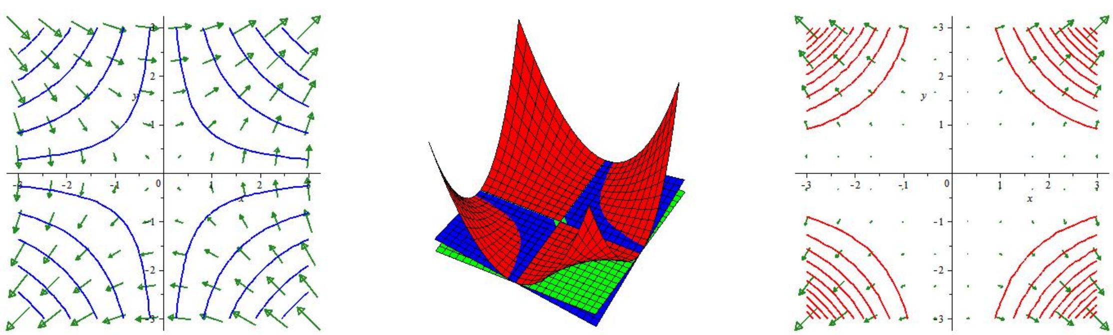
points in the normal direction to the hyperbola. Notice that for $k > 0$ we have $y^2 = k/x^2$ hence $y = \pm\sqrt{k}/x$. When $k = 0$ we find solutions $x = 0$ and $y = 0$. The gradient vector field is identically zero on the coordinate axes in this case. I plot it after the next example for the sake of side-by-side comparison

Example 4.2.44. *Suppose we have a level curve $f(x,y) = xy = k$. This either gives a hyperbola ($k \neq 0$) or the coordinate axes ($k = 0$). The gradient vector field is a bit more descriptive in this case:*

$$\nabla f = y\,\widehat{x} + x\,\widehat{y}.$$

In this case the exceptional solution $x = 0$ has $\nabla f\big|_{x=0} = y\,\widehat{x}$ and $y = 0$ has $\nabla f\big|_{y=0} = x\,\widehat{y}$. The origin $(0,0)$ is the only critical point for f in this example.

I plot ∇f on the left and ∇g on the right together with a few level curves. The picture in the middle has $z = x^2y^2$ in red and $z = xy$ in blue with $z = 0$ in green for reference.

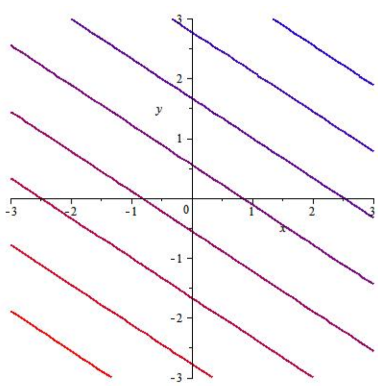


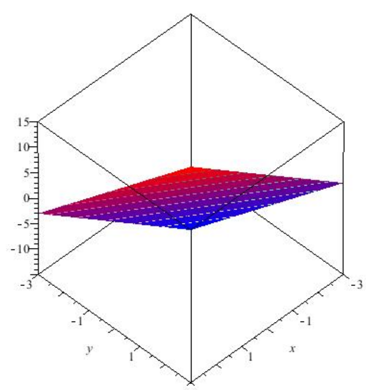
The last pair of examples goes to show that a given set of points can be described by many different level-functions. In particular notice that $xy = 1$ is covered by $x^2y^2 = 1$ but the level functions $f(x, y) = xy$ and $g(x, y) = x^2y^2$ change to other levels in rather distinct fashions. Just compare the gradient vector fields. Or, use a CAS[5] to graph $z = f(x, y)$ and $z = g(x, y)$. Those graphs will intersect along the curve $(x, 1/x, 1)$ for $x > 0$. Do they intersect anywhere else?

4.2.3 contour plots

Perhaps you've studied a *topographical map* before. The topographical map uses a two-dimensional chart to plot a three-dimensional landscape. We can make a similar diagram for graphs of the form $z = f(x, y)$. To form such a plot we simply imagine projecting the graph at a few representative z-values down or up to the xy-plane. This is an invaluable tool since we have much better two-dimensional visualization than we do three. Few people can draw excellent three dimesnional perspective, but the contour plot requires no understanding of perspective. We just slice and project. Moreover, we can use the gradient vector field as a sort of *compass*[6]. The gradient vector field in the domain of $f(x, y)$ points toward higher contours. I use the term *higher* with the idea of traveling from $f(x, y) = k_1$ to $f(x, y) = k_2$ where $k_1 < k_2$. If $f(x, y)$ was actually the altitude function then the term upward would be literally accurate. Usually the term has nothing to do with an actual height, that's just a mental picture for us to help think through the math.

Example 4.2.45. *Suppose $f(x, y) = 2x + 3y$. The graph $z = f(x, y)$ is the plane $z = 2x + 3y$. Contours are level curves of the form $2x + 3y = k$. These contours are simply lines with x-intercept $k/2$ and y-intercept $k/3$. See the plot and graph below to appreciate how the contour plot and graph complement one another. Also, note there is no critical point in this example and the gradient vector field $\nabla f = \langle 2, 3\rangle$ is constant in the domain of f.*



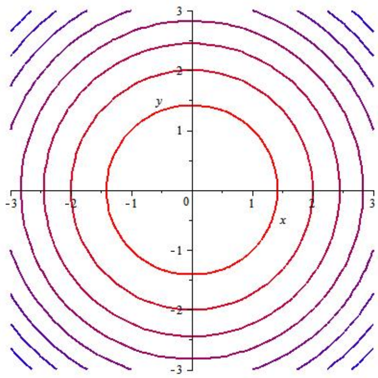
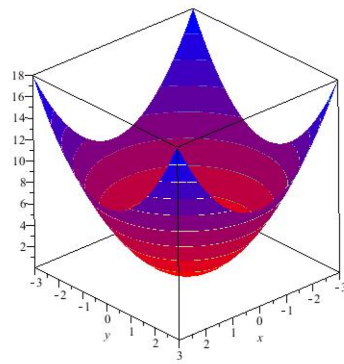


Example 4.2.46. *Suppose $f(x, y) = x^2 + y^2$. The graph $z = f(x, y)$ is the quadratic surface known as a paraboloid. Contours are level curves of the form $x^2 + y^2 = k$. These solutions of $x^2 + y^2 = k$ form circles of radius $\sqrt{k}$ for $k > 0$ and a solitary point $(0, 0)$ for $k = 0$. There are no contours*

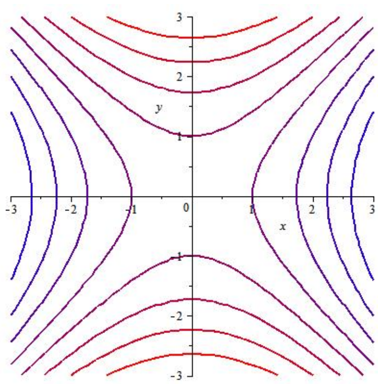
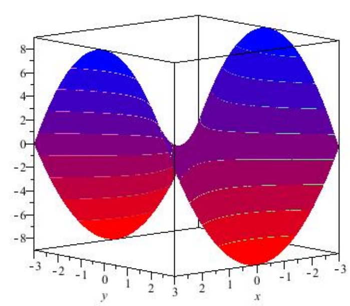
[5] I used Maple to create these graphs, of course you could use Mathematica or any other plotting tool, I have links to free ones on my website... I do expect you use something to aid your visualization.

[6] thanks to Dr. Monty Kester for this particular slogan

with $k < 0$. Once more see how the graph a contour plot complement one another. Furthermore, observe that $\nabla f = \langle 2x, 2y \rangle$ is zero at the origin which is the only critical point. It's clear from the contours or the graph that $f(0,0)$ is a local minimum for f. In fact, it's clear it is the global minimum for the function.

Example 4.2.47. *Suppose $f(x,y) = x^2 - y^2$. The graph $z = f(x,y)$ is the quadratic surface known as a hyperboloid. Contours are level curves of the form $x^2 - y^2 = k$. These solutions of $x^2 - y^2 = k$ form hyperbolas which open up/down for $k < 0$ and open left/right for $k > 0$. If $k = 0$ the $x^2 - y^2 = 0$ yields the special case $y = \pm x$, these are asymptotes for all the hyperbolas from $k \neq 0$. Once more see how the graph and contour plot complement one another. Furthermore, observe that $\nabla f = \langle 2x, -2y \rangle$ is zero at the origin which is the only critical point. It's clear from the contours or the graph that $f(0,0)$ is a not a local minimum or maximum for f. This sort of critical point is called a* **saddle point**.

Example 4.2.48. *Suppose $f(x,y) = \cos(x)$. The graph $z = f(x,y)$ is sort-of a wavy plane. Contours are solutions of the level curve equation $\cos(x) = k$. In this case y is free, however we only find non-empty solution sets for $-1 \leq k \leq 1$. For a particular $k \in [-1,1]$ we have the level-curve $\{(x,y) \mid \cos(x) = k\}$. Note that the cosine curve will reach k twice for each 2π interval in x. Let me pick on a few special values,*

$$k = 0, \quad \textit{solve} \quad \cos(x) = 0, \quad \textit{to obtain} \ \ x = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \dots$$

The $k = 0$ contours are of the form $x = \frac{\pi}{2}(2n-1)$ for $n \in \mathbb{Z}$, there are infinitely many such contours and they are disconnected from one another. Another case which is easy to think through without a calculator,

$$k = 1/2, \quad \textit{solve} \quad \cos(x) = 1/2, \quad \textit{to obtain} \ \ x = -\frac{\pi}{3} + 2\pi n, \quad \textit{or} \ \ x = \frac{\pi}{3} + 2\pi n$$

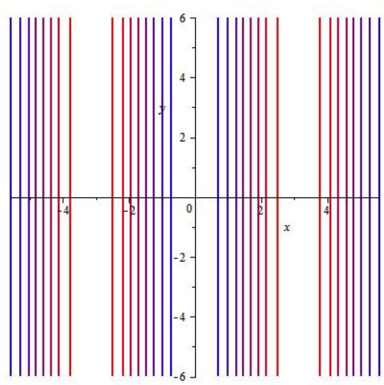
for $n \in \mathbb{Z}$. Once more the level-curves are vertical lines. Continuing, study $k = 1$,

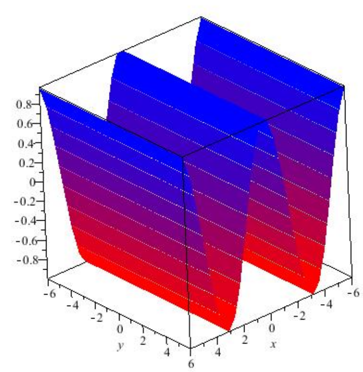
$$k = 1, \quad \textit{solve} \quad \cos(x) = 1, \quad \textit{to obtain} \ \ x = 2\pi n, \textit{for} \ \ n \in \mathbb{Z}.$$

Likewise:

$$k = -1, \quad \textit{solve} \quad \cos(x) = -1, \quad \textit{to obtain} \ \ x = (2n-1)\pi, \textit{for} \ \ n \in \mathbb{Z}.$$

Observe the gradient $\nabla f = \langle -\sin(x), 0\rangle$ is zero along the $k = \pm 1$ contours. The points on $k = 1$ give a local maximum whereas the points on $k = -1$ give local minima for f. This is a special sort of critical point since they are not isolated, no matter how close we zoom in there are always infinitely many critical points in a neighborhood of a given critical point.





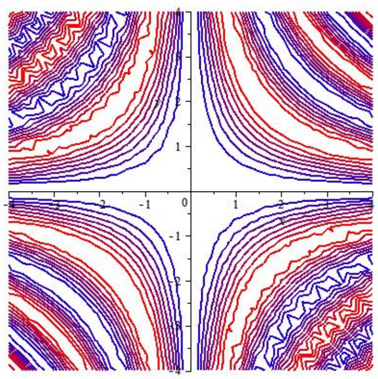
Example 4.2.49. *Suppose* $f(x,y) = \cos(xy)$. *Calculate* $\nabla f = \langle -y\sin(xy), -x\sin(xy)\rangle$ *it follows that solutions of* $xy = n\pi$ *for* $n \in \mathbb{Z}$ *give critical points of* f. *Contours are given by the level-curves* $\cos(xy) = k$ *which have nonempty solutions for* $k \in [-1,1]$. *For example, note that* $\cos(xy) = 1$ *has solution* $xy = 2j\pi$ *for some* $j \in \mathbb{Z}$. *In particular,*

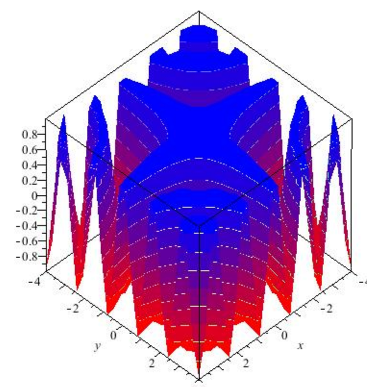
$$xy = 0,\ xy = \pm 2\pi,\ xy = \pm 4\pi,\ \dots \quad \Rightarrow \quad y = 0, \quad x = 0,\ y = \pm\frac{2\pi}{x}, \quad y = \pm\frac{4\pi}{x},\ \dots$$

On the other hand, $\cos(xy) = -1$ *has solution* $xy = (2m-1)\pi$ *for some* $m \in \mathbb{Z}$. *In particular,*

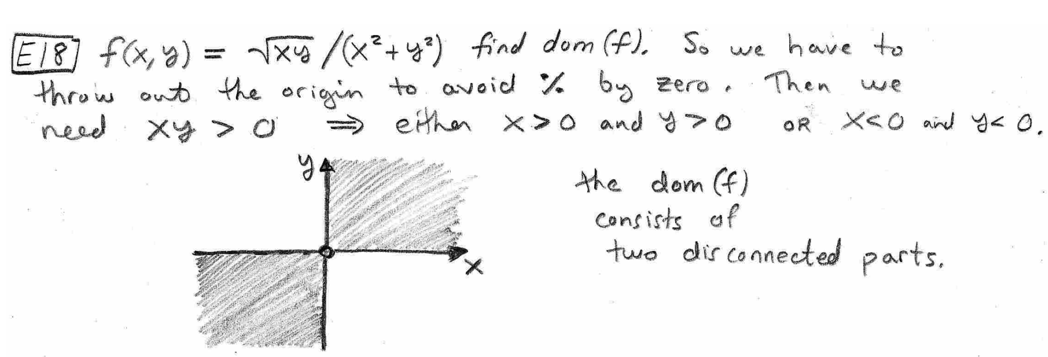
$$xy = \pm\pi, \quad xy = \pm 3\pi,, \quad xy = \pm 5\pi \quad \dots \quad \Rightarrow \quad y = \pm\frac{\pi}{x}, \quad y = \pm\frac{3\pi}{x}, \quad y = \pm\frac{5\pi}{x},\ \dots$$

The contours are simply a family of hyperbolas which take the coordinate axes as asymptotes. This is a great example to see both why contour plots help us visualize the graph which we'd rather not illustrate three-dimensionally. Of course we can use a CAS to directly picture $z = f(x,y)$, *but such pictures rarely yield the same sort of detailed information a well-drawn contour plot.*





Example 4.2.50. *Nice CAS (in this section I used Maple, but all mature CAS's have built-in contour tools) plots are a luxury we don't always have. Notice we can do much just with hand-drawn sketches. I trade color-coding for explicit level labels.*



4.3 partial differentiation in $\mathbb{R}^3$ and $\mathbb{R}^n$

Definition 4.3.1.

Let $f : dom(f) \subseteq \mathbb{R}^3 \to \mathbb{R}$ be a function with $(x_o, y_o, z_o) \in dom(f)$. If the directional derivative below exists, then we define the **partial derivative** of f at $\vec{p}_o = (x_o, y_o, z_o)$ with respect to x, y, z by

$$\frac{\partial f}{\partial x}(\vec{p}_o) = (D_{\widehat{x}} f)(\vec{p}_o), \qquad \frac{\partial f}{\partial y}(\vec{p}_o) = (D_{\widehat{y}} f)(\vec{p}_o), \qquad \frac{\partial f}{\partial y}(\vec{p}_o) = (D_{\widehat{z}} f)(\vec{p}_o)$$

respective. We also use the notations $\frac{\partial f}{\partial x} = \partial_x f = f_x$, $\frac{\partial f}{\partial y} = \partial_y f = f_y$ and $\frac{\partial f}{\partial z} = \partial_z f = f_z$. Generally, if $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is a function with $\vec{p}_o \in dom(f)$ and the limit below exists, then we define the **partial derivative** of f at $\vec{p}_o$ with respect to x_j by

$$\frac{\partial f}{\partial x_j}(\vec{p}_o) = (D_{\widehat{x}_j} f)(\vec{p}_o).$$

The notation $\frac{\partial f}{\partial x_j} = \partial_j f$ is at times useful.

Once more we have natural interpretations for these partial derivatives:

f_x gives the rate of change in f in the x-direction.
f_y gives the rate of change in f in the y-direction.
f_z gives the rate of change in f in the z-direction.

It is useful to rewrite the definition of the partial derivatives explicitly in terms of derivatives.

$$\frac{\partial f}{\partial x}(x_o, y_o, z_o) = \frac{d}{dt}\bigg[\ f(x_o + t, y_o, z_o)\ \bigg]\bigg|_{t=0}$$

$$\frac{\partial f}{\partial y}(x_o, y_o, z_o) = \frac{d}{dt}\bigg[\ f(x_o, y_o + t, z_o)\ \bigg]\bigg|_{t=0}$$

$$\frac{\partial f}{\partial z}(x_o, y_o, z_o) = \frac{d}{dt}\bigg[\ f(x_o, y_o, z_o + t)\ \bigg]\bigg|_{t=0}.$$

Partial differentiation is just differentiation where we hold all but one of the **independent variables** constant. Notice that z in the context above is an independent variable. In constrast, when we studied $z = f(x, y)$ the variable z was a **dependent variable**. The symbols x, y, z are not reserved. They have multiple meanings in multiple contexts and you must have the correct conceptual framework if you are to make the correct computations. When z, x are independent we have $\frac{\partial z}{\partial x} = 0$. If z, x are dependent then it is generally some function. In any event, the following proposition should be entirely unsurprising at this point:

Proposition 4.3.2.

Assume f, g are functions from $\mathbb{R}^3$ to $\mathbb{R}$ whose partial derivatives exist. Then for $c \in \mathbb{R}$,

1. $(f+g)_x = f_x + g_x$ and $(f+g)_y = f_y + g_y$ and $(f+g)_z = f_z + g_z$.
2. $(cf)_x = cf_x$ and $(cf)_y = cf_y$ and $(cf)_z = cf_z$.
3. $(fg)_x = f_x g + f g_x$ and $(fg)_y = f_y g + f g_y$ and $(fg)_z = f_z g + f g_z$.

Moreover, if $h : dom(h) \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable and $x_1 = x,\ x_2 = y,\ x_3 = z$,

4. $$\frac{\partial}{\partial x_j}\big[h(f(x_1, x_2, x_3))\big] = \frac{dh}{dt}\Big|_{f(x_1,x_2,x_3)}\frac{\partial f}{\partial x_j} = \frac{dh}{df}\frac{\partial f}{\partial x_j}$$
5. $$\frac{\partial x_i}{\partial x_j} = \delta_{ij} \qquad \text{where } x_1 = x, x_2 = y, x_3 = z.$$

Proof: The proofs are nearly identical to those given in the $n = 2$ case. However, I will offer a proof of (5.) for arbitrary n. Suppose $f(x_1, x_2, \dots, x_n) = x_i = \vec{x} \bullet \widehat{x}_i$ and calculate

$$\frac{\partial f}{\partial x_j} = \lim_{t\to 0}\left[\frac{f(\vec{x}) - f(\vec{x} + t\,\widehat{x}_j)}{t}\right] = \lim_{t\to 0}\left[\frac{x_i - (\vec{x} + t\,\widehat{x}_j)\bullet \widehat{x}_i}{t}\right] = \lim_{t\to 0}\left[\frac{x_i - x_i + t\delta_{ij}}{t}\right] = \delta_{ij}.$$

Therefore, $\partial_j x_i = \delta_{ij}$ for all $i, j \in \mathbb{N}_n$. In particular, this result applies to the case $n = 3$ hence the proof of (5.) is complete. Naturally this proposition equally well applies to $f : \mathbb{R}^n \to \mathbb{R}$. The proofs are nearly identical to the $n = 2$ case, we just have a few sums to sort through. I leave those to the reader. $\square$

Example 4.3.3. . *1*

$\boxed{\text{E52}}$ Let $g(x,y,z) = xy^2z^3 + \sin(xyz)$ then

$g_x = y^2z^3 + yz\cos(xyz)$: y, z treated as constants.

$g_y = 2xyz^3 + xz\cos(xyz)$: x, z treated as constants.

$g_z = 3xy^2z^2 + xy\cos(xyz)$: x, y treated as constants.

Example 4.3.4. . *2*

Let $f(x,y,z) = x/(y+z)$ find $f_z(3,2,1) = \frac{\partial f}{\partial z}\Big|_{(3,2,1)}$

$$\frac{\partial f}{\partial z}\Big|_{(3,2,1)} = \frac{-x}{(y+z)^2}\Big|_{(3,2,1)} = \frac{-3}{(2+1)^2} = \frac{-3}{9} = \boxed{\frac{-1}{3}}$$

Example 4.3.5. . *3*

$\boxed{E53}$ Suppose $r = \sqrt{x^2+y^2+z^2}$.

$$\frac{\partial r}{\partial x} = \frac{1}{2\sqrt{x^2+y^2+z^2}}\frac{\partial}{\partial x}\left[x^2+y^2+z^2\right] = \frac{x}{\sqrt{x^2+y^2+z^2}} = \frac{x}{r}$$

Likewise $\partial r/\partial y = y/r$ and $\partial r/\partial z = z/r$.

Example 4.3.6. . *4*

$u = \sqrt{x_1^2+x_2^2+\cdots+x_n^2}$. Let $1 \le k \le n$ then,

$$\frac{\partial u}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\sqrt{x_1^2+x_2^2+\cdots+x_n^2}\right]$$

$$= \frac{1}{2\sqrt{x_1^2+x_2^2+\cdots+x_n^2}}\frac{\partial}{\partial x_k}\left[\cancel{x_1^2}^{\,0}+\cancel{x_2^2}^{\,0}+\cdots+\cancel{x_k^2}^{\,2x_k}+\cdots+\cancel{x_n^2}^{\,0}\right]$$

$$= \boxed{\frac{x_k}{\sqrt{x_1^2+\cdots+x_n^2}}}$$

Example 4.3.7. . *5*

Verify $u = \dfrac{1}{\sqrt{x^2+y^2+z^2}}$ solves $u_{xx} + u_{yy} + u_{zz} = 0$.

if $W \equiv \sqrt{x^2+y^2+z^2}$ then $W_x = \dfrac{x}{W}$

$$\frac{\partial}{\partial x}(u) = \frac{\partial}{\partial x}\left[\frac{1}{W}\right] = \frac{-1}{W^2}\frac{\partial W}{\partial x} = \frac{-1}{W^2}\frac{x}{W} = \frac{-x}{W^3}$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left[\frac{-x}{W^3}\right] = \frac{-1W^3 + x \cdot 3W^2 W_x}{W^6} = \frac{-W^3 + 3x^2 W}{W^6} = u_{xx} = \frac{-W^2 + 3x^2}{W^5}$$

Likewise, u_{yy} and u_{zz} have same form with $x \to y$ or z,

$$u_{xx} + u_{yy} + u_{zz} = \frac{-W^2+3x^2}{W^5} + \frac{-W^2+3y^2}{W^5} + \frac{-W^2+3z^2}{W^5}$$

$$= \frac{-3W^2 + 3(x^2+y^2+z^2)}{W^5}$$

$$= \frac{\cancel{-3(x^2+y^2+z^2)} + \cancel{3(x^2+y^2+z^2)}}{W^5}$$

we'll explain later.

$$= 0. \quad \therefore \boxed{u \text{ solves } u_{xx} + u_{yy} + u_{zz} = \nabla^2 u = 0.}$$

Remark: I'm pretty-sure that introducing W makes life easier here.

4.3.1 directional derivatives and the gradient in $\mathbb{R}^3$ and $\mathbb{R}^n$

The idea of Example 4.2.21 equally well transfer to functions of three or more variables. We usually require the functions we analyze to be continuously differentiable since that avoids certain pathological examples:

Definition 4.3.8.

A function $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be **continuously differentiable** at $\vec{p}_o \in dom(f)$ iff the partial derivative functions $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n}$ are continuous at $\vec{p}_o$. We say $f \in C^1(\vec{p}_o)$. If all the second-order partial derivatives of f are continuous at $\vec{p}_o$ then we say $f \in C^2(\vec{p}_o)$. If continuous partial derivatives of arbitrary order exist at $\vec{p}_o$ then we say f is **smooth** and write $f \in C^\infty \vec{p}_o$.

We'll see an example in the next section where the formula below holds for a multivariate functions which is not even continuously differentiable, however the geometric analysis which flows from this formula is most meaningful for continuously differentiable functions.

Proposition 4.3.9.

Suppose $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $\vec{p}_o \in \mathbb{R}^n$ then the directional derivative at $\vec{p}_o$ in the direction of the unit vector $\widehat{u}$ is given by:

$$D_{\widehat{u}} f(\vec{p}_o) = \Big\langle \ \partial_1 f(\vec{p}_o), \ \partial_2 f(\vec{p}_o), \dots, \partial_n f(\vec{p}_o) \ \Big\rangle \bullet \widehat{u}.$$

Proof: delayed until the next section. $\square$

At this point it is useful to introduce a convenient notation which groups all the partial derivatives together in a particular vector of functions. Notice that the length of the gradient vector depends on the context in which it is used.

Definition 4.3.10.

If the partial derivatives of f exist then we define

$$\nabla f = \langle \partial_1 f, \ \partial_1 f, \ \dots, \ \partial_n f \rangle = \widehat{x}_1 \frac{\partial f}{\partial x_1} + \widehat{x}_2 \frac{\partial f}{\partial x_2} + \cdots + \widehat{x}_n \frac{\partial f}{\partial x_n}.$$

we also use the notation $grad(f)$ and call this the **gradient** of f.

The upside-down triangle ∇ is also known as *nabla*. Identify that for $\mathbb{R}^3$ $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y + \widehat{z}\partial_z$. The operator ∇ takes a function f and produces a vector field ∇f. This is called the **gradient vector field** of f. For a continuously differentiable function we have the following beautiful formula for the directional derivative:

$$\boxed{D_{\widehat{u}} f(\vec{p}_o) = (\nabla f)(\vec{p}_o) \bullet \widehat{u}.}$$

Technically this isn't the definition, but pragmatically this is almost always what we use to work out problems. We can also write the dot-product in terms of lengths and the angle between the gradient vector $(\nabla f)(\vec{p}_o)$ and the unit-direction vector $\widehat{u}$:

$$\boxed{D_{\widehat{u}} f(\vec{p}_o) = ||(\nabla f)(\vec{p}_o)|| \cos\theta.}$$

Just like the $n = 2$ case we can use the gradient vector field to point us in the directions in which f either increases, decreases or simply stays constant.

Example 4.3.11. Problem: *Suppose $f(x, y, z) = x^2 + y^2 + z^2$. Does f increase at a rate of 10 in any direction at the point $(1, 2, 3)$?*

Solution: *Note $\nabla f(x, y, z) = \langle 2x, 2y, 2z\rangle$ thus $\nabla f(1, 2, 3) = \langle 2, 4, 6\rangle$. The magnitude of $\nabla f(1, 2, 3)$ is $||\nabla f(1, 2, 3)|| = \sqrt{4 + 16 + 36} = \sqrt{56}$ and that is the maximum rate possible. Therefore, the answer is no. This function increases at a rate of $\sqrt{56}$ in the direction $\frac{1}{\sqrt{14}}\langle 1, 2, 3\rangle$.*

Example 4.3.12. Problem: *Suppose $f(x, y, z) = 2x + y + 2z$. Does f increase at a rate of 2 in any direction at the point $(1, 1, 1)$?*

Solution: *Note $\nabla f(x, y, z) = \langle 2, 1, 2\rangle$ thus $\nabla f(1, 1, 1) = \langle 2, 1, 2\rangle$. The magnitude $||\nabla f(1, 1, 1)|| = \sqrt{9} = 3$ and that is the maximum rate possible. Therefore, the answer is yes. Now let's find the direction(s) in which this occurs. Solve:*

$$D_{\langle a,b,c\rangle} f(1, 1, 1) = \langle 2, 1, 2\rangle \bullet \langle a, b, c\rangle = 2a + b + 2c = 2$$

subject the unit-vector condition $a^2 + b^2 + c^2 = 1$. I'll eliminate c by solving the linear equation for $c = \frac{1}{2}(2 - 2a - b)$ and substituting:

$$a^2 + b^2 + \frac{1}{4}(2 - 2a - b)^2 = 1.$$

This give an ellipse in a, b-space. Apparently there is not just one direction where f increases at a rate of 2. There are infinitely many. For example, we can easily solve the ellipse equation for its b-intercepts by putting $a = 0$,

$$b^2 + \frac{1}{4}(2 - b)^2 = 1 \quad \Rightarrow \quad 4b^2 + 4 - 4b + b^2 = 4 \quad \Rightarrow \quad 5b^2 - 4b = 0 \quad \Rightarrow \quad b(5b - 4) = 0.$$

We find the points $(0, 0)$ and $(0, 4/5)$ are on the ellipse. Returning to the plane equation we find the c-value for these points by substituting them into the equation $c = \frac{1}{2}(2 - 2a - b)$:

$$(0, 0): \quad c = \frac{1}{2}(2 - 2a - b) = 1 \qquad \& \qquad (0, 4/5): \quad c = \frac{1}{2}(2 - 4/5) = \frac{1}{2} \cdot \frac{6}{5} = \frac{3}{5}.$$

Thus, we find the direction vectors $\langle 0, 0, 1\rangle$ and $\langle 0, \frac{4}{5}, \frac{3}{5}\rangle$ point where f increases at a rate of 2. You can probably see a few more possibilities by just thinking about $(\nabla f)(1, 1, 1) = \langle 2, 1, 2\rangle$ directly. For example, I see $\langle 1, 0, 0\rangle$ also works.

The two-dimensional analogue of this problem is much easier since we have to solve the intersection of a line and the unit-circle. In that case there are either $0, 1$ or 2 solutions. The three dimensional case is much more interesting. If f models the temperature at the point (x, y, z) then this calculation shows there are many directions in which the temperature increases at a rate of 2.

Example 4.3.13. .

find $(D_{\hat{u}} f)(P)$ at $P = (0,0,0)$ in $\vec{V} = \langle 5, 1, -2 \rangle$ direction.

$$f(x,y,z) = xe^y + ye^z + ze^x$$

$$\nabla f = \left\langle \frac{\partial}{\partial x}(xe^y + ye^z + ze^x), \frac{\partial}{\partial y}(xe^y + ye^z + ze^x), \frac{\partial}{\partial z}(xe^y + ye^z + ze^x) \right\rangle$$

$$= \langle e^y + ze^x, xe^y + e^z, ye^z + e^x \rangle$$

Thus we can calculate,

$$(\nabla f)(0,0,0) = \langle 1+0, 0+1, 0+1 \rangle = \langle 1, 1, 1 \rangle.$$

Finally, we need to normalize $\vec{V}$,

$$\hat{V} = \frac{1}{|\vec{V}|}\vec{V} = \frac{1}{\sqrt{25+1+4}}\langle 5, 1, -2 \rangle = \frac{1}{\sqrt{30}}\langle 5, 1, -2 \rangle.$$

Thus,

$$(D_{\hat{V}} f)(0,0,0) = (\nabla f)(0,0,0) \cdot \left[\frac{1}{\sqrt{30}}\langle 5, 1, -2 \rangle\right]$$

$$= \frac{1}{\sqrt{30}}\langle 1,1,1 \rangle \cdot \langle 5, 1, -2 \rangle = \boxed{\frac{4}{\sqrt{30}}}$$

Example 4.3.14. .

Let $\vec{V} = \langle 1,2,3 \rangle$ then $\hat{V} = \frac{1}{\sqrt{14}}\langle 1,2,3 \rangle$ has $|\hat{V}| = 1$.

Suppose $f(x,y,z) = x/(y+z)$. Find $(D_{\hat{V}} f)(4,1,1)$,

$$\nabla f = \left\langle \frac{1}{y+z}, \frac{-x}{(y+z)^2}, \frac{-x}{(y+z)^2} \right\rangle$$

$$(\nabla f)(4,1,1) = \langle 1/2, -4/4, -4/4 \rangle = 1/2\langle 1, -2, -2 \rangle$$

$$(\nabla f)(4,1,1) \cdot \hat{V} = \frac{1}{2}\frac{1}{\sqrt{14}}\langle 1,-2,-2 \rangle \cdot \langle 1,2,3 \rangle = \frac{1}{2\sqrt{14}}(1-4-6) = \boxed{\frac{-9}{2\sqrt{14}}}$$

Therefore we find $\boxed{(D_{\hat{V}} f)(4,1,1) = -9/2\sqrt{14}}$

Example 4.3.15. .

Let $f(x,y,z)=\sqrt{x^2+y^2+z^2}$ find the max. rate of change at $(3,6,-2)$ and the direction in which it occurs.

$$\nabla f = \langle f_x, f_y, f_z \rangle$$
$$= \left\langle \frac{x}{\sqrt{x^2+y^2+z^2}}, \frac{y}{\sqrt{x^2+y^2+z^2}}, \frac{z}{\sqrt{x^2+y^2+z^2}} \right\rangle$$
$$= \frac{1}{\sqrt{x^2+y^2+z^2}} \langle x, y, z \rangle.$$

Notice for $(3,6,-2)$ we have $\sqrt{x^2+y^2+z^2} = \sqrt{9+36+4} = 7$,

$$\nabla f(3,6,-2) = \frac{1}{7}\langle 3, 6, -2\rangle.$$

$$|\nabla f(3,6,-2)| = \frac{1}{7}\sqrt{3^2+6^2+2^2} = \boxed{1 = |\nabla f(3,6,-2)|}$$

Max Rate of Change

This occurs in the $\nabla f(3,6,-2)$ direction which is the $\boxed{\frac{1}{7}\langle 3, 6, -2\rangle - \text{direction}}$

Example 4.3.16. .

Consider $f(x,y,z)=x^2+y^2+z^2$. Find the directional derivative of f at $(2,1,3)$ in the direction of the origin. That is the $(-2,-1,-3)$ direction, we need a unit vector so $\div$ by length $\sqrt{4+1+9}$ to construct,

$$\hat{u} = \frac{1}{\sqrt{14}}(-2,-1,-3)$$

We find the gradient of f,

$$\nabla f = \langle 2x, 2y, 2z\rangle \Rightarrow (\nabla f)(2,1,3) = \langle 4, 2, 6\rangle.$$

Hence,

$$(D_{\hat{u}} f)(2,1,3) = \frac{-1}{\sqrt{14}}\langle 4,2,6\rangle \cdot \langle 2,1,3\rangle = \frac{-1}{\sqrt{14}}(8+2+18) = \frac{-28}{\sqrt{14}}$$

Since $\frac{28}{\sqrt{14}} = \frac{2(14)\sqrt{14}}{\sqrt{14}\sqrt{14}} = 2\sqrt{14}$ we find $\boxed{(D_{\hat{u}}f)(2,1,3) = -2\sqrt{14}}$

Moreover, we extend the definition of critical point to the general case in the obvious way:

Definition 4.3.17.

We say $\vec{p}_o$ is a **critical point** of f if $(\nabla f)(\vec{p}_o)$ does not exist or $(\nabla f)(\vec{p}_o) = \vec{0}$.

The function in Example 4.3.11 the origin $(0,0,0)$ is the only critical point. On the other hand, the function in Example 4.3.12 has no critical point.

4.3.2 gradient vector fields in $\mathbb{R}^3$ and $\mathbb{R}^n$

We can calculate the gradient vector field for functions on $\mathbb{R}^n$ with $n \geq 1$ but, visualization is beyond most of us if $n > 3$. I mainly focus on the $n = 3$ case here and we see how the gradient aids our understanding of non-cartesian coordinate systems. Then we examine how the gradient vector field naturally provides a normal vector field to a level surface.

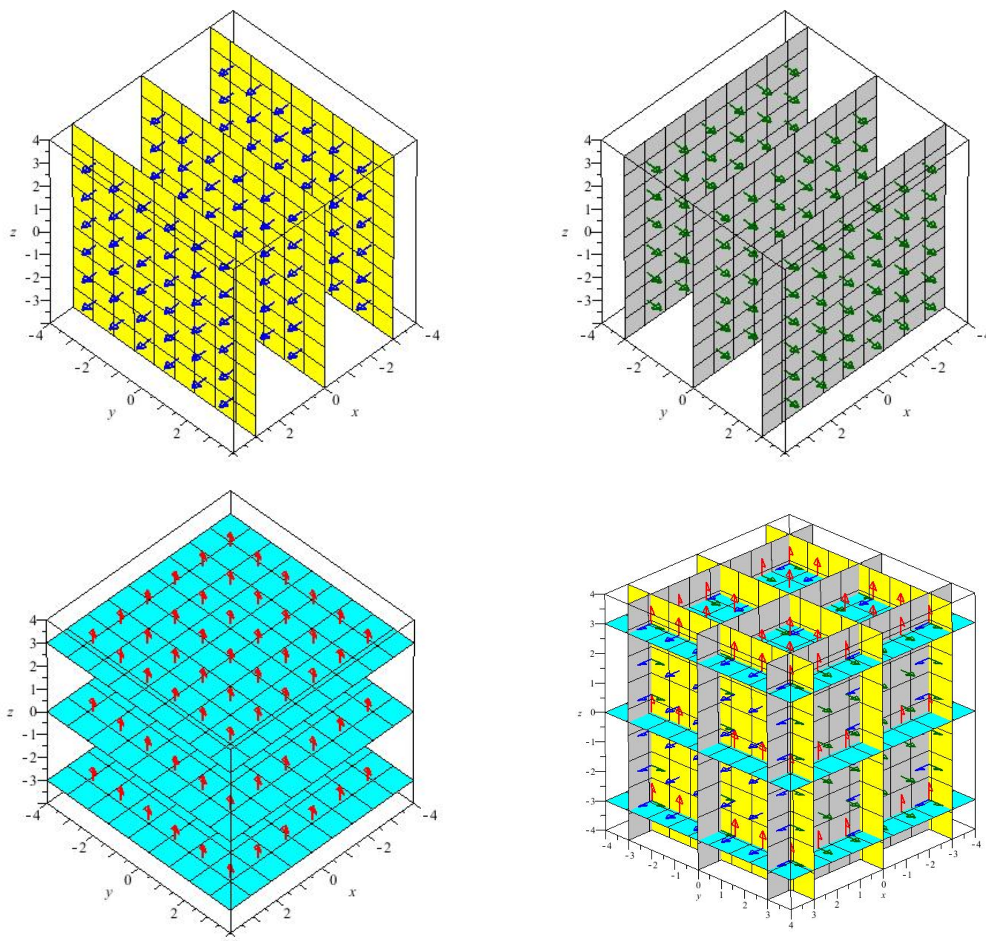
Definition 4.3.18.

If f is differentiable on $U \subseteq \mathbb{R}^n$ then ∇f defines the gradient vector field on U. We assign to each point $\vec{p} \in U$ the vector $\nabla f(\vec{p})$.

Example 4.3.19. *If x, y, z denote the coordinate functions on $\mathbb{R}^3$ then we find*

$$\nabla x = \langle 1, 0, 0, \rangle = \widehat{x},$$
$$\nabla y = \langle 0, 1, 0, \rangle = \widehat{y},$$
$$\nabla z = \langle 0, 0, 1, \rangle = \widehat{z}.$$

These define constant vector fields on $\mathbb{R}^3$.



Generally, the gradient vector fields of the coordinate functions of a non-cartesian coordinate system provide a vector fields which point in the direction of increasing coordinates. To obtain unit-vectors we simply normalize the gradient vector fields. In particular, if $y_1, y_2, \dots y_n$ are coordinates on $\mathbb{R}^n$ then there exist differentiable functions $f_1, f_2, \dots, f_n$ such that $y_j = f_j(x_1, x_2, \dots, x_n)$ for $j = 1, 2, \dots, n$. We can define:

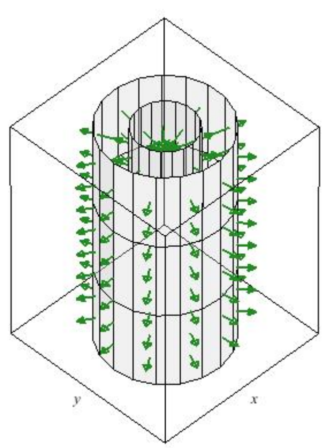
$$\widehat{y_1} = \frac{\nabla f_1}{||\nabla f_1||} \quad \text{and} \quad \widehat{y_2} = \frac{\nabla f_2}{||\nabla f_2||}, \ \dots, \quad \widehat{y_n} = \frac{\nabla f_n}{||\nabla f_n||}.$$

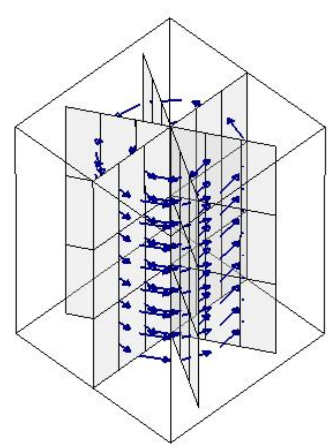
I mention this general idea for the interested reader. We are primarily interested in the cylindrical and spherical three dimensional coordinate systems. That's just a custom, we could easily extend these techniques to orthonormal coordinates based on ellipses or hyperbolas. If we are willing to give up on nice distance formulas we could even use coordinates based on tilted lines which meet at angles other than 90 degrees.

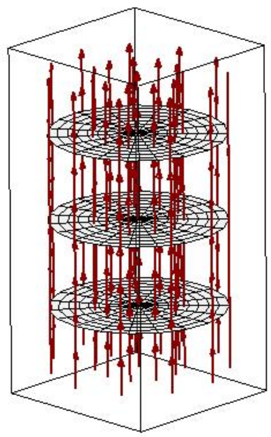
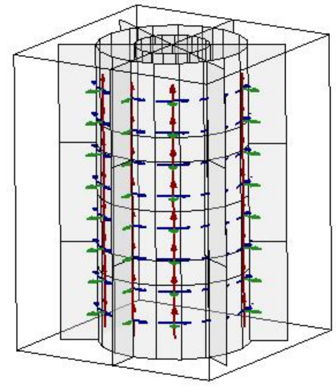
Example 4.3.20. *For cylindrical coordinates r, θ, z we can easily derive (following the same calculational steps as the polar two-dimensional case)*

$$\begin{aligned} \widehat{r} &= \tfrac{1}{||\nabla r||}\nabla r = \widehat{r} = \langle \cos(\theta), \sin(\theta), 0\rangle \\ \widehat{\theta} &= \tfrac{1}{||\nabla \theta||}\nabla \theta = \langle -\sin(\theta), \cos(\theta), 0\rangle \\ \widehat{z} &= \tfrac{1}{||\nabla z||}\nabla z = \langle 0, 0, 1\rangle \end{aligned}$$

The difference between the calculations above and the polar coordinate case is that cylindrical coordinates are three dimensional and that means the gradient vector fields of the coordinate functions are three dimensional vector fields. I advocated a geometric derivation of these cylindrical unit vectors earlier in this course, but we now have computational method which requires almost no geometric intuition.





Example 4.3.21. *Suppose ρ, ϕ, θ denote spherical coordinates. Recall*[7]

$$\rho = \sqrt{x^2+y^2+z^2}, \quad \phi = \cos^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}} \right), \quad \theta = \tan^{-1}\left(\frac{y}{x} \right)$$

You can calculate that

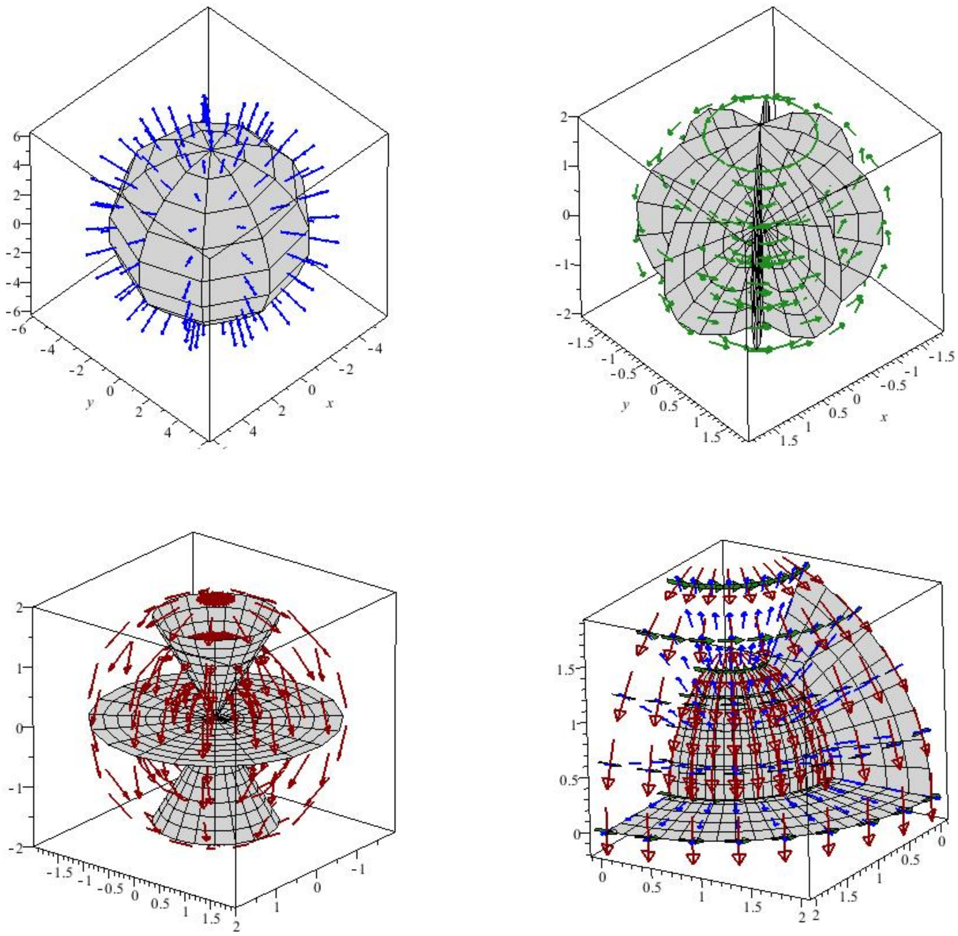
$$\begin{aligned}
\widehat{\rho} &= \tfrac{1}{||\nabla\rho||}\nabla\rho = \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\
\widehat{\phi} &= \tfrac{1}{||\nabla\phi||}\nabla\phi = -\cos(\phi)\cos(\theta)\,\widehat{x} - \cos(\phi)\sin(\theta)\,\widehat{y} + \sin(\phi)\,\widehat{z} \\
\widehat{\theta} &= \tfrac{1}{||\nabla\theta||}\nabla\theta = -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}.
\end{aligned}$$

I'll walk you through the ρ calculation. To begin you can show that $\nabla\rho = \langle x/\rho,\ y/\rho,\ z/\rho\rangle$. But, we also know $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$ and $z = \rho\cos\phi$. Therefore,

$$\nabla\rho = \langle \cos\theta\sin\phi,\ \sin\theta\sin\phi,\ \cos\phi\rangle.$$

But, $||\nabla\rho|| = 1$. We derive that $\widehat{\rho} = \langle \cos\theta\sin\phi,\ \sin\theta\sin\phi,\ \cos\phi\rangle$. Perhaps I asked you to verify the formulas for $\widehat{\phi}, \widehat{\theta}$ in your homework. Making nice pictures of the spherical frame is an art I have yet to master... here's my best for now:

[7] these formulas only apply for certain octants, however, the ambiguity for the remaining octants only involves shifting the angular formulas by a constant. As you continue to read you'll notice that differentiation ultimately will kill any such constant so these formulas suffice.



Another nice application of the gradient involves level surfaces. Consider this: a level surface is the set of points which solves $f(x, y, z) = k$ for some value k. If we consider a point (x_o, y_o, z_o) on the level surface $f(x, y, z) = k$ then the gradient vector $(\nabla f)(x_o, y_o, z_o)$ will be perpendicular to the tangent plane of the level surface. Remember that when $\theta = \pi/2$ we find a direction in which $f(x, y, z)$ stays constant near (x_o, y_o, z_o). What does this mean? Let's summarize it:

The gradient vector field ∇f is normal to the level surface $f(x, y, z) = k$.

I use geometric intuition to make this claim here. We will offer a better proof later in this chapter. For now, let's try to appreciate the geometry.

Example 4.3.22. *Suppose* $V(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$ *represents the voltage due to a point-charge at the origin. Electrostatics states that the electric field* $\vec{E} = -\nabla V$*. Geometrically this has a simple meaning; the electric field points along the normal direction to the level-surfaces of the voltage function*[8]*. In other words, the electric field vectors are normal to the equipotential surfaces where they are attached. What is an "equipotential surface", it's a surface on which the voltage assumes a constant value. This is nothing more than a level-surface of the voltage function. For the given potential function, using* $\rho = \sqrt{x^2+y^2+z^2}$,

$$\begin{aligned}
\nabla V &= \langle \partial_x(1/\rho), \partial_y(1/\rho), \partial_z(1/\rho)\rangle \\
&= \langle (-1/\rho^2)\partial_x\rho,\ (-1/\rho^2)\partial_y\rho,\ (-1/\rho^2)\partial_z\rho\rangle \\
&= \frac{-1}{\rho^2}\langle \partial_x\rho,\ \partial_y\rho,\ \partial_z\rho\rangle \\
&= -\frac{1}{\rho^2}\widehat{\rho}.
\end{aligned}$$

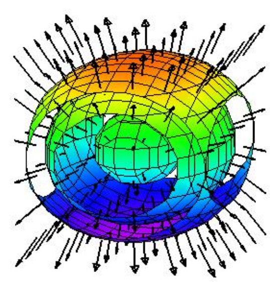
Equipotentials $V = V_o = 1/\rho$ *are simply spheres* $\rho = 1/V_o$ *and the electric field is a purely radial field* $\vec{E} = \frac{1}{r^2}\widehat{\rho}$.

Example 4.3.23. *Consider the ellipsoid* $f(x,y,z) = x^2/a^2 + y^2/b^2 + z^2/c^2 = k$*. At any point on the ellipse the vector field*

$$\nabla f = \frac{2x}{a^2}\,\widehat{x} + \frac{2y}{b^2}\,\widehat{y} + \frac{2z}{c^2}\,\widehat{z}$$

points in the normal direction to the ellipsoid.

It amazes me how easy it is to find a formula to assign a normal-vector to an arbitrary point on an ellipse. Imagine solving that problem without calculus.



[8]The voltage function is the electric potential or simply the potential function in this context

4.4 the general derivative

Thus far we have primarily discussed partial derivatives in their connection to the rate of change of a given function in a particular direction. However, we would like to characterize the change in the function as a whole. Moreover, even in the one-dimensional case the derivative was closely tied to the best linear approximation to the function. In the single variable case it is as simple as this: the best linear approximation to a differentiable function at a point is the linearization of the function at that point whose graph is the tangent line. The slope of the tangent line is the value of the derivative function at the point. How do these ideas generalize? I take an n-dimensional approach in the beginning of this section because little is gained by talking in lower dimensions for the basic definitions.

Definition 4.4.1.

Suppose that U is open and $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping the we say that $\vec{F}$ is **differentiable** at $\vec{a} \in U$ iff there exists a linear mapping $\vec{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to 0} \frac{\vec{F}(\vec{a}+\vec{h}) - \vec{F}(\vec{a}) - \vec{L}(\vec{h})}{||\vec{h}||} = 0.$$

In such a case we call the linear mapping $\vec{L}$ the **differential at** $\vec{a}$ and we denote $\vec{L} = d\vec{F}_{\vec{a}}$. The matrix of the differential is called the **derivative of** $\vec{F}$ **at** $\vec{a}$ and we denote $[d\vec{F}_{\vec{a}}] = \vec{F}'(\vec{a}) \in \mathbb{R}^{\ m\times n}$ which means that $d\vec{F}_{\vec{a}}(\vec{v}) = \vec{F}'(\vec{a})\vec{v}$ for all $\vec{v} \in \mathbb{R}^n$.

4.4.1 matrix of the derivative

If we know a function is differentiable at a point then we can calculate the formula for $\vec{L}$ in terms of partial derivatives. In particular, if $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then the differential $d\vec{F}_{\vec{a}}$ has the derivative matrix $\vec{F}'(\vec{a})$ which has components expressed in terms of partial derivatives of the component functions:

$$[\ d\vec{F}_{\vec{a}}\ ]_{ij} = \partial_j F_i = \frac{\partial F_i}{\partial x_j}(\,\vec{a}\,)$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. This result is proved in advanced calculus. Let me expand this claim in detail for a few common cases: in each case we note $\vec{L}(\vec{a}+\vec{h}) \approx \vec{F}(a) + \vec{F}'(\vec{a})\vec{h}$

1. **function on** $\mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$, $\boxed{L(a+h) \approx f(a) + f'(a)h}$ the derivative matrix is just the derivative $f'(a)$ at the point.
2. **path into** $\mathbb{R}^n$, $\vec{r} : \mathbb{R} \to \mathbb{R}^n$, $\boxed{\vec{r}(a+h) \approx \vec{r}(a) + \vec{r}'(a)h}$. The derivative matrix is just the velocity vector $\vec{r}'(a)$ viewed as an $n \times 1$ matrix (it's a column vector).

3. **multivariate real-valued function**, $f : \mathbb{R}^n \to \mathbb{R}$, $\boxed{f(\vec{a}+\vec{h}) \approx f(\vec{a}) + (\nabla f)(\vec{a})\vec{h}}$. The derivative matrix is just the gradient vector $(\nabla f)(\vec{a})$ viewed as an $1 \times n$ matrix (it's a row vector).

4. **coordinate change mapping**, $\vec{T} : \mathbb{R}^3 \to \mathbb{R}^3$, $\boxed{\vec{T}(\vec{a}+\vec{h}) \approx \vec{T}(\vec{a}) + \vec{T}'(\vec{a})\vec{h}}$. The derivative matrix is a 3×3 matrix. In particular, if we denote $\vec{T} = \langle x, y, z\rangle$ and use u, v, w for cartesian coordinates in the domain of $\vec{T}$

$$\vec{T}'(\vec{a}) = [\partial_u \vec{T} | \partial_v \vec{T} | \partial_w \vec{T}] = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{bmatrix}$$

For two-dimensional coordinate change, $\vec{T} : \mathbb{R}^2 \to \mathbb{R}^2$ we again write $\boxed{\vec{T}(\vec{a}+\vec{h}) \approx \vec{T}(\vec{a}) + \vec{T}'(\vec{a})\vec{h}}$ but the matrix $\vec{T}'(\vec{a})$ is just a 2×2 matrix

$$\vec{T}'(\vec{a}) = [\partial_u \vec{T} | \partial_v \vec{T}] = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

Example 4.4.2. *Let $f(x) = \sqrt{x}$. The linearization at $x = 4$ is given by $L(x) = 2 + \frac{1}{4}(x-4)$ since $f'(4) = \frac{1}{2\sqrt{4}} = \frac{1}{4}$. We could also express L by $L(4+h) = 2 + h/4$. As an application, note the approximation $\sqrt{5} \approx 2 + 1/4 = 2.25$.*

Example 4.4.3. *Let $\vec{r}(t) = \langle t, t^2, sin(10t)\rangle$ for $t \in [0, 2]$. The linearization of $\vec{r}$ at $t = 1$ is given by $\vec{L}(1+h) = \vec{r}(1) + h\vec{r}\,'(1)$. In particular,*

$$\vec{L}(1+h) = \langle 1+h, 1+2h, sin(10) + 10hcos(10)\rangle.$$

Example 4.4.4. .

Find linearization of $f(x,y) = x/y$ at $(6,3)$.

Notice $\frac{\partial f}{\partial x} = 1/y$ and $\frac{\partial f}{\partial y} = -x/y^2$. These are continuous at $(6,3)$ so $f(x,y)$ is differentiable at $(6,3)$.

$$L(x,y) = f(6,3) + \frac{\partial f}{\partial x}\Big|_{(6,3)}(x-6) + \frac{\partial f}{\partial y}\Big|_{(6,3)}(y-3)$$

$$= 6/3 + \frac{1}{3}(x-6) - \frac{6}{9}(y-3)$$

$$= \boxed{\frac{1}{3}x - \frac{2}{3}y + 2 = L(x,y)}$$

Example 4.4.5. .

$\boxed{E72}$ Find the linearization of $f(x,y) = x^2+y^2$ at $(1,2)$. Then approximate $f(2,2)$ and compare to the real-value. We found the tangent plane's eq^{2} in $\boxed{E71}$ so we already know

$$\boxed{L(x,y) = 5 + 2(x-1) + 4(y-2)}$$

We approximate f via L,

$$f(2,2) \cong L(2,2) = 5 + 2(2-1) + 4(0) = 7$$

Of course we can just evaluate $f(2,2) = 2^2+2^2 = 8$ to see we have an absolute error of $8-7=1$.

4.4.2 tangent space as graph of linearization

In the section after this I wrestle with why these are good definitions. For now I'll state them without justification.

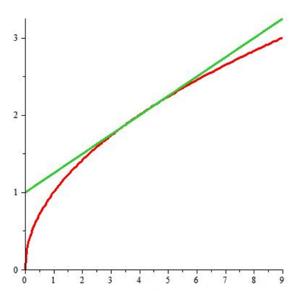
1. $f : \mathbb{R} \to \mathbb{R}$ has tangent line at $(a, f(a))$ with equation $y = f(a) + f'(a)(x-a)$.

2. $\vec{r} : \mathbb{R} \to \mathbb{R}^n$ has tangent line at $\vec{r}(a)$ with natural parametrization $\vec{l}(h) = \vec{r}(a) + \vec{r}\,'(a)h$.

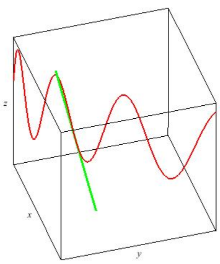
3. $f : \mathbb{R}^2 \to \mathbb{R}$ has tangent plane at $(a, b, f(a,b)$ with equation $z = f(a,b) + (\nabla f)(a,b) \bullet \langle x - a, y - b\rangle$.

These are the cases of interest, in case 2 we usually deal with $n=2$ or $n=3$ in this course. The following triple of examples mirror those given in the last section. The overall theme is simple: the tangent space to a graph of a function is the graph of the linearization of that function. There are several other viewpoints on the tangent space of a surface and we devote an entire section to that a little later in this chapter. Here I just want you to get what we mean when we say a derivative *gives the best linear approximation to a function.*

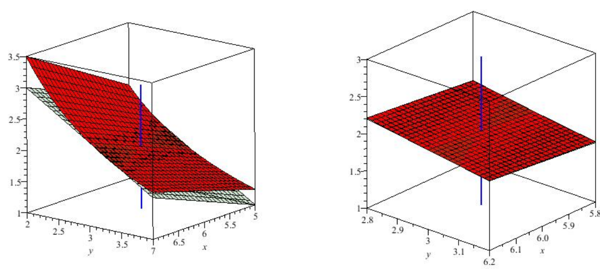
Example 4.4.6. *We continue Example 4.4.2, $f(x) = \sqrt{x}$ and the linearization at $x = 4$ is given by $L(x) = 2 + \frac{1}{4}(x-4)$. The tangent line is the graph $y = L(x)$ which is in green, whereas the $y = f(x)$ is in red.*



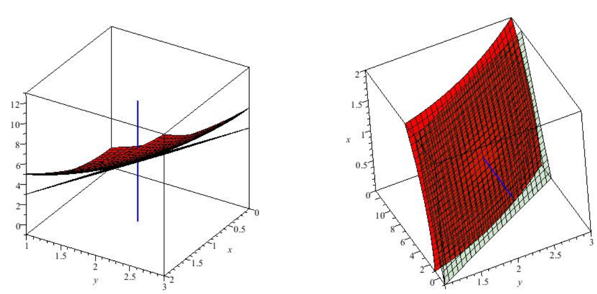
Example 4.4.7. *We continue Example 4.4.3, $\vec{r}(t) = \langle t, t^2, sin(10t)\rangle$ for $t \in [0,2]$ and the linearization of $\vec{r}$ at $t = 1$ is given by $\vec{L}(1+h) = \langle 1+h, 1+2h, sin(10) + 10hcos(10)\rangle$. Once more we plot the curve in red and the tangent line parametrized by $\vec{L}$ in green:*



Example 4.4.8. *Continue Example 4.4.4, $f(x,y) = x/y$ and the tangent plane to $z = x/y$ at $(6,3)$ is the solution set of $z = x/3 - 2y/3 + 2$. Below I illustrate the tangent plane, the blue line goes through the point of tangency. See how the surface is locally flat, note the right picture is zoomed further in towards the point of tangency.*



Example 4.4.9. *Continue Example 4.4.5, $f(x,y) = x^2 + y^2$ and the tangent plane to $z = x^2 + y^2$ at $(1,2)$ is the solution set of $z = 5 + (x-1) + 4(y-2)$. Below I illustrate the tangent plane, the blue line goes through the point of tangency. See how the surface is locally flat, these are just two views of the same scale, I put a rotating animation of this on the webpage, take a look.*



4.4.3 existence and connections to directional differentiation

Existence is usually more troublesome than calculation. But, that is no reason to ignore it. In this subsection I attempt to give you a better sense of what it means for a function to be differentiable at a point. Geometrically we eventually come to the simple realization that a function is differentiable iff it is well-approximated by its linearization. This in turn is tied to the proper definition of the tangent plane. We already gave formulas for important cases in the last subsection, my goal here is to explain why we use those definitions and not something else. Before we get to those more subtle topics, I begin by demonstrating the general derivative recovers single-variable differentiation:

Example 4.4.10. *Suppose $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at x. It follows that there exists a linear function $df_x : \mathbb{R} \to \mathbb{R}$ such that*[9]

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0.$$

Since $df_x : \mathbb{R} \to \mathbb{R}$ is linear there exists a constant matrix m such that $df_x(h) = mh$. In this silly case the matrix m is a 1×1 matrix which otherwise known as a real number. Note that

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0 \quad \Leftrightarrow \quad \lim_{h\to 0^\pm} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0.$$

In the left limit $h \to 0^-$ we have $h < 0$ hence $|h| = -h$. On the other hand, in the right limit $h \to 0^+$ we have $h > 0$ hence $|h| = h$. Thus, differentiability suggests that $\lim_{h\to 0^\pm} \frac{f(x+h)-f(x)-df_x(h)}{\pm h} = 0$. But we can pull the minus out of the left limit to obtain $\lim_{h\to 0^-} \frac{f(x+h)-f(x)-df_x(h)}{h} = 0$. Therefore,

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{h} = 0.$$

We seek to show that $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = m$.

$$m = \lim_{h\to 0} \frac{mh}{h} = \lim_{h\to 0} \frac{df_x(h)}{h}$$

A theorem from calculus I states that if $\lim(f - g) = 0$ and $\lim(g)$ exists then so must $\lim(f)$ and $\lim(f) = \lim(g)$. Apply that theorem to the fact we know $\lim_{h\to 0} \frac{df_x(h)}{h}$ exists and

$$\lim_{h\to 0} \left[\frac{f(x+h) - f(x)}{h} - \frac{df_x(h)}{h} \right] = 0.$$

It follows that

$$\lim_{h\to 0} \frac{df_x(h)}{h} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

Consequently,

$$df_x(h) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \quad \text{defined } f'(x) \text{ in calc. I.}$$

[9] unless we state otherwise, $\mathbb{R}^n$ is assumed to have the euclidean norm, in this case $||x||_{\mathbb{R}} = \sqrt{x^2} = |x|$.

Therefore, $\boxed{df_x(h) = f'(x)h}$*. In other words, if a function is differentiable in the sense we defined at the beginning of this section then it is differentiable in the terminology we used in calculus I. Moreover, the derivative at x is precisely the matrix of the differential. If we use the notation $y = f(x)$ and $h = dx$ then we recover formula for the differential often taught in first semester calculus:*

$$dy_x(dx) = \frac{dy}{dx}(x)dx$$

Or, more compactly, $dy = \frac{dy}{dx}dx$ where dy is the change in y corresponding to the change dx in x. These seemingly heuristic statements take a rigorous meaning in the boxed equation above.

Of course, what really makes the general derivative interesting is its ability to tackle problems such as given below:

Example 4.4.11. *Suppose $F : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by $F(x, y) = (xy,\ x^2,\ x+3y)$ for all $(x, y) \in \mathbb{R}^2$. Consider the difference function $\triangle F$ at (x, y):*

$$\triangle F = F((x, y) + (h, k)) - F(x, y) = F(x + h, y + k) - F(x, y)$$

Calculate,

$$\triangle F = \big((x + h)(y + k),\ (x + h)^2,\ x + h + 3(y + k)\big) - \big(xy,\ x^2,\ x + 3y\big)$$

Simplify by cancelling terms which cancel with $F(x, y)$:

$$\triangle F = \big(xk + hy,\ 2xh + h^2,\ h + 3k)\big)$$

Identify the linear part of $\triangle F$ as a good candidate for the differential. I claim that:

$$L(h, k) = \big(xk + hy,\ 2xh,\ h + 3k\big).$$

is the differential for f at (x,y). Observe first that we can write

$$L(h, k) = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}.$$

therefore $L : \mathbb{R}^2 \to \mathbb{R}^3$ is manifestly linear. Use the algebra above to simplify the difference quotient below:

$$\lim_{(h,k)\to(0,0)} \left[\frac{\triangle F - L(h, k)}{||(h, k)||} \right] = \lim_{(h,k)\to(0,0)} \left[\frac{(0, h^2, 0)}{||(h, k)||} \right]$$

Note $||(h, k)|| = \sqrt{h^2 + k^2}$ therefore we fact the task of showing that $(0,\ h^2/\sqrt{h^2 + k^2},\ 0) \to (0, 0, 0)$ as $(h, k) \to (0, 0)$. Recall from our study of limits that we can prove the vector tends to $(0, 0, 0)$ by showing the each component tends to zero. The first and third components are obviously zero however the second component requires study. Observe that

$$0 \leq \frac{h^2}{\sqrt{h^2 + k^2}} \leq \frac{h^2}{\sqrt{h^2}} = |h|$$

Clearly $\lim_{(h,k)\to(0,0)}(0) = 0$ *and* $\lim_{(h,k)\to(0,0)} |h| = 0$ *hence the squeeze theorem for multivariate limits shows that* $\lim_{(h,k)\to(0,0)} \frac{h^2}{\sqrt{h^2+k^2}} = 0$. *Therefore,*

$$dF_{(x,y)}(h,k) = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}.$$

Fortunately we can usually avoid explicit limit calculations due to the nice proposition below.

Example 4.4.12. *Again consider* $F(x,y) = (xy,\ x^2,\ x+3y)$. *Identify* $F_1(x,y) = xy$, $F_2(x,y) = x^2$ *and* $F_3(x,y) = x+3y$. *Calculate,*

$$[F'(x,y)] = \begin{bmatrix} \partial_x F_1 & \partial_y F_1 \\ \partial_x F_2 & \partial_y F_2 \\ \partial_x F_3 & \partial_y F_3 \end{bmatrix} = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix}$$

In single-variable calculus we learn that differentiability implies continuity. However, continuity does not imply differentiability at a given point. The same is true for multivariate functions.

Proposition 4.4.13.

If $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then $\vec{F}$ is continuous at $\vec{a}$.

The proof is given in advanced calculus. It's not too difficult. $\square$

The general derivative also reproduces all the directional derivatives we previously discussed.

Proposition 4.4.14.

If $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then the directional derivative $D_{\vec{v}}\vec{F}(\vec{a})$ exists for each $\vec{v} \in \mathbb{R}^n$ and $D_{\vec{v}}\vec{F}(\vec{a}) = d\vec{F}_{\vec{a}}(\vec{v})$.

The proof is given in advanced calculus. It's not terribly difficult. $\square$

We should consider the example below. It may challenge some of your misconceptions. It shows that directional differentiation at a point does not give us enough to build the derivative. In fact, the example below has **all** directional derivatives and yet the function is not even continuous.

Example 4.4.15. *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *be defined by* $f(x,y) = \frac{x^2y}{x^4+y^2}$ *for* $(x,y) \neq (0,0)$ *and* $f(0,0) = 0$. *We proved in Example 3.3.4 that this function is not continuous at* $(0,0)$. *Given the proposition above we also may infer the function is not differentiable at* $(0,0)$. *You might expect this indicates at least some directional derivative fails to exist. Let's investigate. We turn to the problem of* *calculating the directional derivative of this function in the unit-vector $\langle a, b\rangle$ direction, suppose $b \neq 0$ to begin,*

$$D_{\langle a,b\rangle} f(0,0) = \frac{d}{dt}\left[f(at,bt)\right]\bigg|_{t=0} = \frac{d}{dt}\left[\frac{a^2bt^3}{a^4t^4+b^2t^2}\right]\bigg|_{t=0} = \left[\frac{a^2b(a^4t^2+b^2)-a^2bt(2ta^4)}{(a^4t^2+b^2)^2}\right]\bigg|_{t=0} = \frac{a^2}{b}.$$

On the other hand, if $b = 0$ then we know $a \neq 0$ since $\langle a, b\rangle$ is a unit-vector[10] *hence $f(at, bt) = \frac{a^2bt^3}{a^4t^4+b^2t^2} = 0$ and it follows $D_{\langle a,0\rangle} f(0, 0) = 0$. We find the directional derivatives of f exist in all directions.*

Notice that the directional derivatives do jump from one value to another as we travel around the unit-circle. In particular, as we traverse the arc of the circle through the point $\langle 1, 0\rangle$ we have $\langle a, b\rangle$ go from vectors with $b > 0$ which have $\frac{a^2}{b} \to \infty$ to vectors with $b < 0$ which have $\frac{a^2}{b} \to -\infty$. In the middle, we hit $\langle 1, 0\rangle$ where $D_{\langle a,0\rangle} f(0, 0) = 0$. These directional derivatives may exist but they certainly do not continuously paste together. It turns out that continuity of the directional derivatives in the coordinate directions is a sufficient condition to eliminate the trouble of the previous example.

Definition 4.4.16.

A mapping $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **continuously differentiable** at $a \in U$ iff all the partial derivative mappings $\partial F_i/\partial x_j$ exist on an open set containing a and are continuous at a.

Continuous differentiability is typically easier than differentiability to check. The reason is that partial derivatives are straightforward to calculate. On the other hand, it is sometimes challenging to find the linearization and actually check the appropriate limit vanishes. It follows that the proposition below is welcome news:

Proposition 4.4.17.

If F is continuously differentiable at a then F is differentiable at a

The proof is somewhat involved. The main construction involves breaking a vector into a sum of vector components. Then continuity of the partial derivatives paired with a mean value theorem argument goes to prove the differentiability of the mapping. Again, details are given in my advanced calculus notes (or any good text on the subject). □

There do exist functions which are differentiable at a point and yet fail to be continuously differentiable at that point. In single variable calculus I usually present the example: Let $f(0) = 0$ and

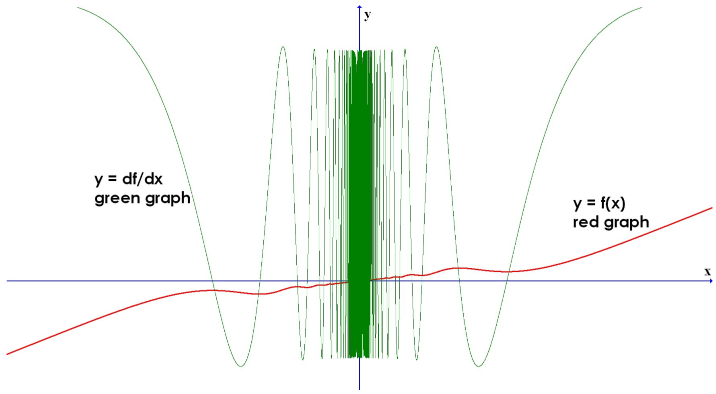
$$f(x) = \frac{x}{2} + x^2 \sin\frac{1}{x}$$

for all $x \neq 0$. I can be shown that the derivative $f'(0) = 1/2$. Moreover, we can show that $f'(x)$ exists for all $x \neq 0$, we can calculate:

$$f'(x) = \frac{1}{2} + 2x\sin\frac{1}{x} - \cos\frac{1}{x}$$

[10] if $a = 0$ and $b = 0$ then $||\langle a, b\rangle|| = 0 \neq 1$

Notice that $dom(f') = \mathbb{R}$. Note then that the tangent line at $(0,0)$ is $y = x/2$.



The lack of continuity for the derivative means that the tangent line at the origin does not well-approximate the graph near the point of tangency. In other words, the linearization is not a good approximation near the point of tangency. This is not just a single-variable phenonmenon. Pathological multivariate examples exist. For example,

Example 4.4.18. *Let $f(0,y) = 0$ and*

$$f(x,y) = x^2 \sin\frac{1}{x}$$

for all $(x,y) \in \mathbb{R}^2$ such that $x \neq 0$. You can show that $D_{\widehat{u}}f(0,0) = 0$ for all unit vectors u. This means that the tangent vectors to any path $t \to (at, bt, f(at, bt))$ reside in the xy-plane. It appears the set of all tangent vectors fill out the xy-plane. However, I'm not sure what happens with non-linear paths in the domain. I suspect the curves on the graph $z = f(x,y)$ built from composing a smooth, but non-linear, path $\gamma : \mathbb{R} \to \mathbb{R}^2$ with f might result in a path $f \circ \gamma$ which is not even differentiable at the origin.

Let's investigate the differentiability of f at $(0,0)$. Given the triviality of all the directional derivatives we suspect $L(h,k) = 0$. Consider,

$$\frac{|f(h,k) - f(0,0) - L(h,k)|}{||(h,k)||} = \frac{|h^2\sin(1/h)|}{\sqrt{h^2+k^2}} = \frac{|h\sin(1/h)|}{\sqrt{1+k^2/h^2}} \leq |h\sin(1/h)| \leq |h|.$$

It follows that f is differentiable at $(0,0)$ since we have $|h| \to 0$ as $(h,k) \to (0,0)$ along any path. Therefore, my suspicion was incorrect. Even nonlinear paths composed with f yield a differentiable path. However, this does give us another example of a function which is differentiable at $(0,0)$ but is not continuously differentiable. If you're wondering it is clear that f_x is not continuous along the entire y-axis. Given our experience in the single variable case we suspect the linearization does not approximate the function in a natural way as we leave the point of tangency. We need the continuity of the partial derivatives to insure the function does not wildly misbehave in the locality of the tangent point.

I haven't proved it yet but I suspect the function below is not differentiable. It gives and example of a function which is continuous but is not differentiable at zero. However, both partial derivatives exist at $(0,0)$, they're just not continuous.

Example 4.4.19. *Let us define $f(0,0) = 0$ and*

$$f(x,y) = \frac{x^2 y}{x^2 + y^2}$$

for all $(x,y) \neq (0,0)$ in $\mathbb{R}^2$. It can be shown[11] *that f is continuous at $(0,0)$. Moreover, since $f(x,0) = f(0,y) = 0$ for all x and all y it follows that f vanishes identically along the coordinate axis. Thus the rate of change in the $\widehat{x}$ or $\widehat{y}$ directions is zero. We can calculate that*

$$\frac{\partial f}{\partial x} = \frac{2xy^3}{(x^2+y^2)^2} \qquad \textit{and} \qquad \frac{\partial f}{\partial y} = \frac{x^4 - x^2y^2}{(x^2+y^2)^2}$$

Consider the path to the origin $t \mapsto (t,t)$ gives $f_x(t,t) = 2t^4/(t^2+t^2)^2 = 1/2$ hence $f_x(x,y) \to 1/2$ along the path $t \mapsto (t,t)$, but $f_x(0,0) = 0$ hence the partial derivative f_x is not continuous at $(0,0)$. Therefore, this function has discontinuous partial derivatives. It is not continuously differentiable.

Let's return to the question of directional derivatives and differentiability. It is tempting to think that the reason the function in Example 4.4.15 failed to be differentiable is that the tangent vectors to the curves $t \mapsto (at, bt, f(at, bt))$ failed to fill out a plane. This suspicion is further encouraged by Example 4.4.18 where we see the function is differentiable and the tangent vectors to the curves $t \mapsto (at, bt, f(at, bt))$ do fill out the xy-plane. However, this suspicion is false. Think back to our experience with multivariate limits in Example 3.3.2. Differentiability also concerns a multivariate limit so intuitively we may expect something could be hidden if we only think about straight-line approaches to the limit point. I suspect that if we had that the tangents to $t \mapsto (\vec{r}(t), f(\vec{r}(t)))$ fill out a plane for all differentiable paths $\vec{r}$ with $\vec{r}(0) = \langle 0, 0\rangle$ then it would follow f is differentiable. I don't have a proof of this claim in the notes at the present time.

Why all this fuss? Let me try to clarify the confusion which pushed me to this discussion:

1. some authors define the tangent plane to be the union of all tangent vectors at a point.

2. other authors say the tangent plane is a plane which well-approximates the graph of the function near the point of tangency.

Item (2.) begs some questions, what exactly do we mean by "well-approximates". Is the nearness to the graph the concept captured by mere differentiability or is it the stronger version captured by continuous differentiability? Item (1.) is dangerous since it would *seem* that looking at all possible directional derivatives should give a complete picture of the tangent vectors at a point. We just

[11] you did this one in homework... or at least you were supposed to...

argued this is not the case[12]. It is possible for all tangents to curves built from linear paths to exist whereas the tangent vectors to a path built from a nonlinear path may not even exist. If we are to use item (1.) as a definition we must clarify it a bit:

> **The tangent plane to the graph $z = f(x, y)$ is formed by the union of all possible tangent vectors of curves $f \circ \vec{\gamma}$ where $\vec{\gamma}$ is a smooth curve in $dom(f)$ which pass at $t = 0$ through the xy-coordinates of the point of tangency. If there exists a smooth curve $\vec{\gamma}$ such that $f \circ \vec{\gamma}$ is not differentiable at $t = 0$ then the tangent plane fails to exist.**

This is just my comment here, I haven't seen this elsewhere. Most authors don't bother with these details or deliberations. In fact, many authors assume continuous differentiability in their definitions. In any event, it seems clear to me that we should prefer a slightly more careful version of (2.) since it has far less technical trouble. With all of this in mind we define (I expand on the most important case to this course after this general definition),

Definition 4.4.20. *general tangent space to a graph.*

Suppose that U is open and $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping which is differentiable at $\vec{a} \in U$ then the linear mapping $\vec{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to 0} \frac{\vec{F}(\vec{a} + \vec{h}) - \vec{F}(\vec{a}) - \vec{L}(\vec{h})}{||\vec{h}||} = 0.$$

defines the **tangent space** at $(\vec{a}, \vec{F}(\vec{a}))$ to $graph(\vec{F}) = \{(\vec{x}, \vec{F}(\vec{x})) \mid dom(\vec{F})\}$ with equations $\vec{z} = \vec{F}(\vec{a}) + \vec{L}(\vec{x} - \vec{a})$ in $\mathbb{R}^n \times \mathbb{R}^m$. We use the notation $\vec{z} \in \mathbb{R}^m$ whereas $\vec{x}, \vec{a} \in \mathbb{R}^n$ in the equation above.

In particular, for $f : \mathbb{R}^2 \to \mathbb{R}$ we have $L(x - x_o, y - y_o) = (\nabla f)(x_o, y_o) \bullet \langle x - x_o, y - y_o \rangle$ and the tangent plane has equation:

$$\boxed{z = f(x_o, y_o) + (x - x_o) f_x(x_o, y_o) + (y - y_o) f_y(x_o, y_o).}$$

The assumption of differentiability of f at (x_o, y_o) insures that the tangent plane $z = f(x_o, y_o) + L(x, y) \approx f(x, y)$ for points near (x_o, y_o). In other words, the graph $z = f(x, y)$ looks like a plane if we zoom in close to the point $(x_o, y_o, f(x_o, y_o))$. In fact, many authors simply define differentiability in view of this concept:

$$\boxed{\textbf{A function is differentiable at } \vec{p} \textbf{ iff it has a tangent plane at } \vec{p}.}$$

This is less than satisfactory if the text you're reading nowhere defines the tangent plane. I won't name names. The boxed statement is true, but it is not a definition. Not here at least.

[12] I have an example if you ask

In the case a function is differentiable but not continuously differentiable we have the situation that there is a tangent plane, but it fails to well-approximate the graph near the point of tangency.

Continuous differentiability is needed for many of the calculations we perform in the remainder of this course. I conclude this section with an example of how it may happen that $f_{xy} \neq f_{yx}$ at a point which is merely differentiable. On the other hand, Clairaut's Theorem states that $f_{xy} = f_{yx}$ for continuously differentiable functions.

Example 4.4.21. .

A CURIOUS EXAMPLE: WHY $f_{xy} \neq f_{yx}$ ALWAYS.

$$f(x,y) = \begin{cases} (x^3y - xy^3)/(x^2+y^2) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

When $(x,y) \neq (0,0)$ its a simple matter to differentiate,

$$f_x = \frac{(3x^2y - y^3)(x^2+y^2) - 2x(x^3y - xy^3)}{(x^2+y^2)^2} = \frac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^2}$$

$$f_y = \frac{x^5 - 4y^2x^3 - y^4x}{(x^2+y^2)^2}$$

$$f_{xy} = \frac{x^6 + 9x^4y^2 - 9x^2y^4 - y^6}{(x^2+y^2)^3} = f_{yx}(x,y) \quad \text{for } (x,y) \neq 0.$$

At the origin we need to use the def$^{\underline{n}}$ of partial differentiation,

$$f_x(0,0) = \lim_{h \to 0}\left[\frac{f(h,0) - f(0,0)}{h}\right] = \lim_{h\to 0}\left[\frac{0-0}{h}\right] = 0.$$

$$f_y(0,0) = \lim_{h \to 0}\left[\frac{f(0,h) - f(0,0)}{h}\right] = \lim_{h\to 0}\left[\frac{0-0}{h}\right] = 0.$$

$$f_{xy}(0,0) \equiv \frac{\partial f_x}{\partial y}(0,0) = \lim_{h\to 0}\left[\frac{f_x(0,h) - f_x(0,0)}{h}\right] = \lim_{h\to 0}\left[\frac{-h^5/(h^2)^2 - 0}{h}\right] = -1.$$

$$f_{yx}(0,0) \equiv \frac{\partial f_y}{\partial x}(0,0) = \lim_{h\to 0}\left[\frac{f_y(h,0) - f_y(0,0)}{h}\right] = \lim_{h\to 0}\left[\frac{h^5/h^4 - 0}{h}\right] = 1.$$

Therefore $f_{xy} \neq f_{yx}$ since at $(0,0)$ they disagree. You might object that this to picky on our part, well sorry its math. The trouble here is that f_{xy} is not continuous at $(0,0)$, everywhere else it is and in all those places $f_{xy}(x,y) = f_{yx}(x,y)$ $\forall (x,y) \neq (0,0)$.

Theorem 4.4.22. *Clairaut's Theorem:*

If $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ is a function where $dom(f)$ contains an open disk D centered at (a,b) and the function f_{xy} and f_{yx} are both continuous on D then

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

The proof is found in most advanced calculus texts. Finally, I should mention that the concerns and examples of this section readily generalize to functions from $\mathbb{R}^m$ to $\mathbb{R}^n$.

4.4.4 properties of the derivative

Suppose $\vec{F}_1 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $\vec{F}_2 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $\vec{a} \in U$ then $\vec{F}_1 + \vec{F}_2$ is differentiable at $\vec{a}$ and $d(\vec{F}_1 + \vec{F}_2)_a = (d\vec{F}_1)_a + (d\vec{F}_2)_a$ which means for the Jacobian matrices we also have $(\vec{F}_1 + \vec{F}_2)'(\vec{a}) = \vec{F}_1'(\vec{a}) + \vec{F}_2'(\vec{a})$. Likewise, if $c \in \mathbb{R}$ then $d(c\vec{F}_1)_a = c(d\vec{F}_1)_a$ hence for the Jacobian matrices we have $(c\vec{F}_1)'(\vec{a}) = c(\vec{F}_1'(\vec{a})$. Nothing terribly surprising here. What is much more fascinating is the following general version of the chain rule:

Proposition 4.4.23.

If $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at $\vec{a}$ and $\vec{G} : V \subseteq \mathbb{R}^p \to \mathbb{R}^m$ is differentiable at $\vec{F}(\vec{a}) \in V$ then $\vec{G} \circ \vec{F}$ is differentiable at $\vec{a}$ and $d(\vec{G} \circ \vec{F})_{\vec{a}} = (d\vec{G})_{\vec{F}(\vec{a})} \circ d\vec{F}_{\vec{a}}$. Moreover, in Jacobian matrix notation,

$$(\vec{G} \circ \vec{F})'(\vec{a}) = \vec{G}'(\vec{F}(\vec{a}))\vec{F}'(\vec{a}).$$

In words, the Jacobian matrix of the composite of $\vec{G}$ with $\vec{F}$ is simply the matrix product of the Jacobian matrices of $\vec{G}$ with the Jacobian matrix of $\vec{F}$. Unfortunately, not all students really learned matrix algebra in highschool so this statement lacks the power it should have in your mind. This proposition builds the foundation for the multivariate version of u-substution. All the chain rules in the next section are derivable from this general proposition. For this reason I offer no proofs in the next section. The calculations in the next section all follow from the calculation below[13]:

Proof: $\approx$ Suppose $\vec{F} : dom(\vec{F}) \subseteq \mathbb{R}^n \to \mathbb{R}^p$ and $\vec{G} : dom(\vec{G}) \subseteq \mathbb{R}^p \to \mathbb{R}^m$. Let $\vec{x}_o \in \mathbb{R}^n$ for which $\vec{F}(\vec{x}_o) = \vec{y}_o \in dom(\vec{G})$ and suppose that $\vec{F}$ is differentiable at $\vec{x}_o$ and $\vec{G}$ is differentiable at $\vec{y}_o$. We seek to show that $\vec{G} \circ \vec{F}$ is differentiable at $\vec{x}_o$ with Jacobian matrix $\vec{G}'(\vec{y}_o)\vec{F}'(\vec{x}_o)$. Observe that the existence of $\vec{G}'(\vec{y}_o) \in \mathbb{R}^{\,m \times p}$ and $\vec{F}'(\vec{x}_o) \in \mathbb{R}^{\,p \times n}$ follow from the differentiability of $\vec{G}$ at $\vec{y}_o$ and $\vec{F}$ at $\vec{x}_o$. In particular, if $||\vec{k}|| \approx 0$ then

$$\vec{G}(\vec{y}_o + \vec{k}) \approx \vec{G}(\vec{y}_o) + \vec{G}'(\vec{y}_o)\vec{k}.$$

Likewise, if $||\vec{h}|| \approx 0$ then

$$\vec{F}(\vec{x}_o + \vec{h}) \approx \vec{F}(\vec{x}_o) + \vec{F}'(\vec{x}_o)\vec{h}.$$

Suppose $\vec{h}$ is given such that $||\vec{h}|| \approx 0$. It follows that $\vec{F}'(\vec{x}_o)\vec{h} \approx 0$. Let $\vec{k} = \vec{F}'(\vec{x}_o)\vec{h}$ and note that

$$\underbrace{\vec{G}(\vec{F}(\vec{x}_o + \vec{h}) \approx \vec{G}(\vec{F}(\vec{x}_o) + \vec{F}'(\vec{x}_o)\vec{h})}_{continuity\ of\ G\ at\ y_o} = \vec{G}(\vec{y}_o + \vec{k}) \approx \vec{G}(\vec{y}_o) + \vec{G}'(\vec{y}_o)\vec{k}$$

[13] this is a plausibility argument, not a formal proof, all the $\approx$ symbols are shorthands for a more detailed estimation which is not given in these notes, however, you guessed it, can be found in a good advanced calculus text.

Therefore, for $||\vec{h}|| \approx 0$,

$$\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) \approx \vec{G}(\vec{F}(\vec{x}_o)) + \vec{G}\,'(\vec{F}(\vec{x}_o))\vec{F}\,'(\vec{x}_o)\vec{h}.$$

Thus $\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) - \vec{G}(\vec{F}(\vec{x}_o)) - \vec{G}\,'(\vec{F}(\vec{x}_o))\vec{F}\,'(\vec{x}_o)\vec{h} \approx 0$. In fact, if we worked out the careful details we could show that

$$\lim_{\vec{h}\to 0} \frac{\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) - \vec{G}(\vec{F}(\vec{x}_o)) - \vec{G}\,'(\vec{F}(\vec{x}_o))\vec{F}\,'(\vec{x}_o)\vec{h}}{||\vec{h}||} = 0$$

and it follows that $(\vec{G}\circ\vec{F})'(\vec{x}_o) = \vec{G}\,'(\vec{F}(\vec{x}_o))\vec{F}\,'(\vec{x}_o)$. Technically, this is not a proof, but perhaps it makes the rule a bit more plausible. The chain rule is primarily a consequence of matrix multiplication when we look at it the right way. □.

Example 4.4.24. . *14*

$\boxed{\text{E69}}$ Suppose $f: \mathbb{R}^2 \to \mathbb{R}$ is differentiable. Define the polar coordinate change map; $\mathbb{X}(r,\theta) \equiv (r\cos\theta, r\sin\theta)$ this means $\mathbb{X}: \mathbb{R}^2 \to \mathbb{R}^2$ it takes $(r,\theta) \longmapsto (x(r,\theta), y(r,\theta))$. Consider $g = f\circ\mathbb{X}$. Then $Dg = Df \circ D\mathbb{X}$ where $\mathbb{X} = (x,y)$

$$Df = \left[\ \frac{\partial f}{\partial x} \ \ \frac{\partial f}{\partial y}\ \right] \qquad D\mathbb{X} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

Hence if $w = f\circ\mathbb{X} = g$

$$Dg(r,\theta) = \left[\ \frac{\partial f}{\partial x} \ \ \frac{\partial f}{\partial y}\ \right]\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

$$\| $$

$$\left[\frac{\partial w}{\partial r}, \frac{\partial w}{\partial \theta}\right] = \left[\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y},\ -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}\right]$$

Thus,

$$\frac{\partial w}{\partial r} = \cos\theta\frac{\partial w}{\partial x} + \sin\theta\frac{\partial w}{\partial y} \longrightarrow \boxed{\frac{\partial}{\partial r} = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}}$$

$$\frac{\partial w}{\partial \theta} = -r\sin\theta\frac{\partial w}{\partial x} + r\cos\theta\frac{\partial w}{\partial y} \longrightarrow \boxed{\frac{\partial}{\partial \theta} = -r\sin\theta\frac{\partial}{\partial x} + r\cos\theta\frac{\partial}{\partial y}}$$

Example 4.4.25. . *13*

$\boxed{E70}$ Let $z = f(x,y) = x^2 - 3y^2$ and let $x = uv$ & $y = u+v^2$

Calculate $\partial z/\partial u$ and $\partial z/\partial v$.

$$\frac{\partial z}{\partial u} = \frac{\partial}{\partial u}\Big[f(x(u,v), y(u,v))\Big] = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = (2x)(v) - 6y(1).$$

$$\frac{\partial z}{\partial v} = \frac{\partial}{\partial v}\Big[f(x(u,v), y(u,v))\Big] = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = (2x)(u) - 6y(2v).$$

this is simple enough, you can use a tree-diagram if you like, but I've never needed them, you just identify the intermediate variables and sort-of "conserve partials". Lets see how this is done in the matrix/Jacobian formalism. We define

$$X(u,v) \equiv (x(u,v), y(u,v)) = (uv, u+v^2).$$

Thus, notice $x_1 = u$ and $x_2 = v$ while $x = f_1$, $y = f_2$ and $X = f$

$$DX = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \quad \text{while} \quad Df = \left[\tfrac{\partial f}{\partial x} \ \ \tfrac{\partial f}{\partial y}\right]$$

Then $z = f \circ X$ so $z = z(u,v)$

$$\begin{aligned} Dz = \left[\tfrac{\partial z}{\partial u}, \tfrac{\partial z}{\partial v}\right] &= (Df)(DX) = \left[\tfrac{\partial f}{\partial x} \ \ \tfrac{\partial f}{\partial y}\right] \\ &= \left[\tfrac{\partial f}{\partial x}, \tfrac{\partial f}{\partial y}\right]\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \\ &= \Big[\underbrace{\tfrac{\partial f}{\partial x}\tfrac{\partial x}{\partial u} + \tfrac{\partial f}{\partial y}\tfrac{\partial y}{\partial u}}_{\frac{\partial z}{\partial u}}, \underbrace{\tfrac{\partial f}{\partial x}\tfrac{\partial x}{\partial v} + \tfrac{\partial f}{\partial y}\tfrac{\partial y}{\partial v}}_{\frac{\partial z}{\partial v}}\Big] \end{aligned}$$

4.5 chain rules

In this section we explain how the chain rule generalizes to functions of several variables. Before I get to that, recall we already learned one new chain-rule for space curves:

$$\boxed{\frac{d}{dt}\big[\ \vec{r}(u(t))\ \big] = \frac{d\vec{r}}{du}\frac{du}{dt}.}$$

For example, if $\vec{r}(t) = \langle t, t^2, t^3\rangle$ and $u(t) = \sin t$ then

$$\frac{d}{dt}\big[\ \vec{r}(u(t))\ \big] = \langle 1, 2u, 3u^2\rangle \cos t = \langle \cos t, 2\sin t\cos t, 3\sin^2 t\cos t\rangle.$$

This chain rule was important to understand how the Frenet Serret formulas are reformulated for non-unit-speed curves. It was the source of the speed factors ds/dt in those equations.

Next consider the composite of $f : \mathbb{R}^2 \to \mathbb{R}$ and $\vec{r} : \mathbb{R} \to \mathbb{R}^2$ where $\vec{r} = \langle x, y\rangle$. Here's the rule:

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}}$$

In this case the **independent variable** is t and the **intermediate variables** are x, y. All of the expressions above are understood to be functions of t. A more pedantic statement of the same rule is as follows:

$$\boxed{\frac{d}{dt}f(\vec{r}(t)) = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt}.}$$

Example 4.5.1. *Suppose $f(x, y) = x^2 - xy$ and $x = e^t, y = t^2$ then*

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (2x - y)e^t - x(2t) = (2e^t - t^2)e^t - 2te^t.$$

Now, some of you will doubtless note that we could just as well subsitute $x = e^t$ and $y = t^2$ at the outset and just do ordinary differentiation on $g(t) = (e^t)^2 - t^2e^t$. Will you obtain the same answer? Yes. Is that the right method to count on generally? No. Otherwise, why would I teach you the new rule?

Example 4.5.2. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and $t \mapsto \vec{r}(t) = \langle x(t), y(t)\rangle$ defines a smooth path. What is the geometric relation between the tangent vector to the path and the gradient vector field of f? Use the chain rule,*

$$\frac{d}{dt}\big[f(\vec{r}(t))\big] = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt} = \nabla f(\vec{r}(t)) \bullet \left\langle \frac{dx}{dt}, \frac{dy}{dt}\right\rangle = \nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}.$$

This doesn't really tell us much of anything for an arbitrary function and path. However, if we suppose the path parametrizes a level curve $f(x, y) = k$ then we find something nice. To say $\vec{r}$

parametrizes $f(x,y)=k$ is to insist $f(\vec{r}(t))=k$ for all t. Differentiate this equation and we again use the chain rule on the l.h.s. whereas $\frac{d}{dt}(k)=0$. Thus,

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

We find the gradient vector field is normal to the tangent vector field of the level curve. The chain rule has given us the calculational tool to verify what we argued geometrically earlier in this chapter.

Notice we can't just substitute in the formulas for $x(t)$ and $y(t)$ in the example above. Why? Because we are not given them. The chain rule allows us to discover general relationships which may not be obvious if we always just work at the level of the independent variable.

Example 4.5.3. . *7*

Let $z = x\ln(x+2y)$, $x = \sin t$, $y = \cos t$

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

$$= \left[\ln(x+2y) + \frac{x}{x+2y}\right]\frac{dx}{dt} + \left[\frac{2x}{x+2y}\right]\frac{dy}{dt}$$

$$= \boxed{\left[\ln(\sin t + 2\cos t) + \frac{\sin t}{\sin t + 2\cos t}\right]\cos t - \frac{2\sin^2 t}{\sin t + 2\cos t}}$$

Example 4.5.4. . *26*

$\boxed{E56}$ Let $z = xy$ and again suppose $x = e^t$ & $y = \sin t$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = \boxed{e^t \sin t + e^t \cos t = \frac{dz}{dt}}$$

- Sometimes z is taken as the <u>dependent</u> variable. Other times z is playing the role of an <u>intermediate</u> variable.

Example 4.5.5. . *27*

$\boxed{E55}$ Let $W = xy$ and suppose $x = e^t$ and $y = \sin t$

$$\frac{dW}{dt} = \frac{\partial W}{\partial x}\frac{dx}{dt} + \frac{\partial W}{\partial y}\frac{dy}{dt} = ye^t + x\cos t = \boxed{e^t \sin t + e^t \cos t = \frac{dW}{dt}}$$

The chain rule below is a natural generalization of what we just discussed: if $\vec{r} = \langle x, y, z\rangle$ and[14] $f = f(x, y, z)$ then

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}.}$$

In this case the **independent variable** is t and the **intermediate variables** are x, y, z. All of the expressions above are understood to be functions of t. A more pedantic statement of the same rule is as follows:

$$\frac{d}{dt}f(\vec{r}(t)) = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt} + \frac{\partial f}{\partial z}(\vec{r}(t))\frac{dz}{dt}.$$

Example 4.5.6. *Suppose $\vec{r}(t) = \langle \cos t \sin t, \sin t \sin t, \cos t\rangle$ and $f(x, y, z) = x^2 + y^2 + z^2$. This means $x = \cos t \sin t$, $y = \sin^2 t$ and $z = \cos t$. Calculate,*

$$\begin{aligned}
\frac{d}{dt}f(\vec{r}(t)) &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} \\
&= 2x(\cos^2 t - \sin^2 t) + 2y(2\sin t\cos t) + 2z(-\sin t) \\
&= 2\cos t\sin t(\cos^2 t - \sin^2 t) + 2\sin^2 t(2\sin t\cos t) + 2\cos t(-\sin t) \\
&= 2\cos t\sin t(1 - 2\sin^2 t) + 2\sin^2 t(2\sin t\cos t) + 2\cos t(-\sin t) \\
&= 0.
\end{aligned}$$

Why is this? Simple. The path given by $x = \cos t \sin t$, $y = \sin^2 t$ and $z = \cos t$ parametrizes a curve which lies on the sphere $x^2 + y^2 + z^2 = 1$. It follows that $f(\cos t \sin t, \sin^2 t, \cos t) = 1$ hence differentiation by t yields zero. Geometrically we find the gradient vector field $\nabla f = \langle 2x, 2y, 2z\rangle$ is normal to the tangent vector field of the curve wherever they intersect.

Of course there are many other curves which reside in the level surface $f(x, y, z) = 1$. I just picked one to illustrate that the gradient vectors are normal to the curves on the surface. We can argue this in general.

Example 4.5.7. *Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is differentiable and $t \mapsto \vec{r}(t) = \langle x(t), y(t), z(t)\rangle$ defines a smooth path. Use the chain rule, omitting explicit point dependence on the partials,*

$$\frac{d}{dt}\big[f(\vec{r}(t))\big] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \nabla f(\vec{r}(t)) \bullet \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right\rangle = \nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}.$$

If we suppose the path $t \mapsto \vec{r}(t)$ parametrizes a curve which is on the level surface $f(x, y, z) = k$ then $f(\vec{r}(t)) = k$ for all t. Differentiate this equation and we again use the chain rule on the l.h.s. whereas $\frac{d}{dt}(k) = 0$. Thus,

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

We find the gradient vector field is normal to the tangent vector field of an arbitrary curve on the surface. If the function f is continuously differentiable then it follows that the union of all such

[14] this notation means that $f : \mathbb{R}^3 \to \mathbb{R}$, it is a bit sloppy, but it is also popular and I suppose I should expose you to it.

tangent vectors forms the tangent space to the level surface. The gradient vector at the point of tangency gives the normal to the tangent plane.

For example, a sphere of radius R centered at (a, b, c) has equation $(x-a)^2+(y-b)^2+(z-c)^2 = R^2$. This sphere is naturally viewed as a level surface of $F(x, y, z) = (x-a)^2 + (y-b)^2 + (z-c)^2$. We calculate,

$$\nabla F(x, y, z) = \langle 2(x-a), 2(y-b), 2(z-c)\rangle$$

The equation of the tangent plane at (x_o, y_o, z_o) on this sphere is

$$2(x_o - a)(x - x_o) + 2(y_o - b)(y - y_o) + 2(z_o - c)(z - z_o) = 0.$$

In particular, if $a = b = c = 0$ then we have a tangent plane

$$2x_o(x - x_o) + 2y_o(y - y_o) + 2z_o(z - z_o) = 0.$$

For this case the vector pointing to (x_o, y_o, z_o) and the normal vector $\langle 2x_o, 2y_o, 2z_o\rangle$ point along the same line.

Example 4.5.8. *. 25*

$\boxed{\text{E57}}$ Let $w = xyz$ and suppose $x = t$, $y = t^2$, $z = t^3$

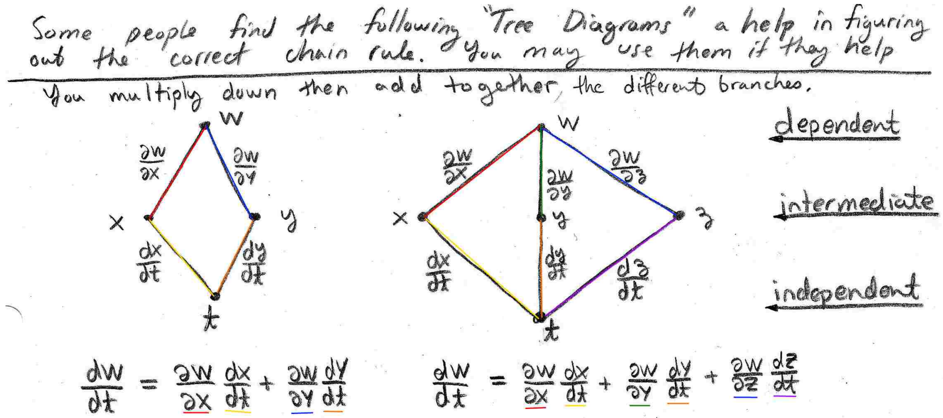
$$\begin{aligned}\frac{dw}{dt} &= \frac{d}{dt}\Big[w(x(t), y(t), z(t))\Big] \\ &= \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt} \\ &= yz + xz(2t) + xy(3t^2) \\ &= t^5 + 2t^5 + 3t^5 \\ &= \boxed{6t^5 = \frac{dw}{dt}} \quad \left(\text{which is good since } w = t^6 \text{ so } \frac{dw}{dt} = 6t^5\right)\end{aligned}$$

Moving on to our next case, if $\vec{r} = \langle x_1, x_2, \dots, x_n\rangle$ and[15] $f = f(x_1, x_2, \dots, x_n)$ then

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}.}$$

In this case the **independent variable** is t and the **intermediate variables** are $x_1, x_2, \dots, x_n$. All of the expressions above are understood to be functions of t. I'm not a big fan, but, another trick to remember the chain-rules above is given by the mneumonic device below:

[15] this notation means that $f : \mathbb{R}^n \to \mathbb{R}$.



Example 4.5.9. *Suppose $f(\vec{x}) = \vec{x} \bullet \vec{x}$ where $\vec{x} \in \mathbb{R}^n$. Moreover, suppose $\vec{r} : \mathbb{R} \to \mathbb{R}^n$ is a path which parametrizes the level set $f(\vec{x}) = R^2$ (this is a higher-dimensional sphere). We have $f(\vec{r}(t)) = R^2$ for all t. Differentiate to find*

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

Once more we find the tangents to curves on the level set are orthogonal to the gradient vector field. Don't ask me to draw the picture here. The tangent space is an $(n-1)$-dimensional hyperplane embedded in $\mathbb{R}^n$, the normal vector field ∇f always points in the one remaining dimension if there are no critical points for f.

Another case[16] is $\vec{F} = \vec{F}(x_1, x_2, \dots, x_n)$ composed with a path. In particular, if $\vec{F} = \langle F_1, F_2, \dots, F_m \rangle : \mathbb{R}^n \to \mathbb{R}^m$ is composed with $\vec{r} = \langle x_1, x_2, \dots, x_n \rangle : \mathbb{R} \to \mathbb{R}^n$ then we have the chain rule

$$\boxed{\frac{d\vec{F}}{dt} = \frac{\partial \vec{F}}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial \vec{F}}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial \vec{F}}{\partial x_n}\frac{dx_n}{dt} = \left\langle \nabla F_1 \bullet \frac{d\vec{r}}{dt},\ \nabla F_2 \bullet \frac{d\vec{r}}{dt},\ \dots, \nabla F_m \bullet \frac{d\vec{r}}{dt} \right\rangle.}$$

Another nice way to think of this rule is as follows:

$$\begin{aligned} \frac{d}{dt}\langle F_1, F_2, \dots, F_m \rangle &= \left\langle \frac{dF_1}{dt}, \frac{dF_2}{dt}, \dots, \frac{dF_m}{dt} \right\rangle \\ &= \left\langle \frac{\partial F_1}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial F_1}{\partial x_n}\frac{dx_n}{dt},\ \dots,\ \frac{\partial F_m}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial F_m}{\partial x_n}\frac{dx_n}{dt} \right\rangle \end{aligned}$$

[16] somewhat rare

Example 4.5.10. *Suppose $\vec{F}(x,y,z) = \langle xy,\ y+z^2\rangle$ and suppose $x = t, y = t^2, z = t^3$. Calculate,*

$$\begin{aligned} \frac{d\vec{F}}{dt} &= \left\langle \frac{\partial F_1}{\partial x}\frac{dx}{dt} + \frac{\partial F_1}{\partial y}\frac{dy}{dt} + \frac{\partial F_1}{\partial z}\frac{dz}{dt},\ \frac{\partial F_2}{\partial x}\frac{dx}{dt} + \frac{\partial F_2}{\partial y}\frac{dy}{dt} + \frac{\partial F_2}{\partial z}\frac{dz}{dt} \right\rangle \\ &= \left\langle y\frac{dx}{dt} + x\partial y\frac{dy}{dt} + 0\frac{dz}{dt},\ 0\frac{dx}{dt} + \frac{dy}{dt} + 2z\frac{dz}{dt} \right\rangle \\ &= \big\langle y + x(2t),\ 2t + 2z(3t^2)\big\rangle \\ &= \big\langle 3t^2,\ 2t + 6t^5\big\rangle \end{aligned}$$

All of the examples up to this point have considered chain rules for functions of just one indpendent variable which we have denoted by t for the sake of conceptual uniformity. We now consider differentiation of composite functions of two or more independent variables.

Suppose $f = f(x,y)$ and $x = x(u,v)$ and $y = y(u,v)$. This means $f : \mathbb{R}^2 \to \mathbb{R}$ and $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$. We have two interesting partial derivatives to compute:

$$\boxed{\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} \qquad \& \qquad \frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.}$$

In this case the **independent variables** are u, v and the **intermediate variables** are x, y. All of the expressions above are understood to be functions of u, v. To be a bit more pedantic we can use $\vec{r}(u,v) = \langle x(u,v), y(u,v)\rangle$ and write

$$\boxed{\frac{\partial}{\partial u}\big[f(\vec{r}(u,v))\big] = \nabla f(\vec{r}(u,v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \& \qquad \frac{\partial}{\partial v}\big[f(\vec{r}(u,v))\big] = \nabla f(\vec{r}(u,v)) \bullet \frac{\partial \vec{r}}{\partial v}.}$$

Notation aside, these rules are very natural extensions of what we have already seen.

Example 4.5.11. *Suppose $z = e^{xy}$ and $x = u^2 + v^2$ and $y = uv$. Calculate,*

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} = ye^{xy}(2u) + xe^{xy}(v) = \big[3u^2v + v^3\big]e^{u^3v + uv^3}.$$

$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v} = ye^{xy}(2v) + xe^{xy}(u) = \big[3uv^2 + u^3\big]e^{u^3v + uv^3}.$$

Once more, if we ignored the chain rule and instead directly substituted the expressions for u, v at the outset then we will still obtain the same result. However, if we are faced with extremely ugly formulas for $x(u,v)$ or $y(u,v)$ then this is a useful organizaing principle. Or we could encounter the situtation that formulas for $x(u,v)$ and $y(u,v)$ are not given and the chain rule still helps uncover general patterns.

Example 4.5.12. *Another application of the chain rule is coordinate change for the differentiation operators. For example, suppose $x = r\cos\theta, y = r\sin\theta$. How do we convert a partial derivative with respect to x for an equivalent differentiation in terms of the polar coordinates? Suppose f is an arbitrary function on $\mathbb{R}^2$, notice by the chain rule,*

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}.$$

But, these relations hold for any function f hence we find the following operator equations:

$$\boxed{\frac{\partial}{\partial r} = \frac{x}{\sqrt{x^2+y^2}}\frac{\partial}{\partial x} + \frac{y}{\sqrt{x^2+y^2}}\frac{\partial}{\partial y}} \qquad \boxed{\frac{\partial}{\partial \theta} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}}$$

Algebra challenge: solve the operator equations above for $\partial/\partial x$ and $\partial/\partial y$. Then compare your answers to what we obtain from the chain rules below:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial x} \qquad \& \qquad \frac{\partial f}{\partial y} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial y}$$

We need the formulas $r = \sqrt{x^2+y^2}$ and $\theta = \tan^{-1}(y/x) + c$ where c is a constant that is either zero for $x > 0$ or π for $x < 0$. (ok, maybe constant is the wrong word, but it certainly differentiates to zero at most points). Calculate that $\frac{\partial r}{\partial x} = \frac{x}{r} = \cos\theta$ and $\frac{\partial r}{\partial y} = \frac{y}{r} = \sin\theta$. Also,

$$\frac{\partial \theta}{\partial x} = \frac{\partial}{\partial x}\big[\tan^{-1}(y/x) + c\big] = \frac{1}{1+y^2/x^2}\cdot\frac{-y}{x^2} = \frac{-y}{x^2+y^2} = -\frac{\sin\theta}{r}$$

$$\frac{\partial \theta}{\partial y} = \frac{\partial}{\partial y}\big[\tan^{-1}(y/x) + c\big] = \frac{1}{1+y^2/x^2}\cdot\frac{1}{x} = \frac{x}{x^2+y^2} = \frac{\cos\theta}{r}.$$

Now substitute these back into the chain rules,

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\cos\theta - \frac{\partial f}{\partial \theta}\frac{\sin\theta}{r} \qquad \& \qquad \frac{\partial f}{\partial y} = \frac{\partial f}{\partial r}\sin\theta + \frac{\partial f}{\partial \theta}\frac{\cos\theta}{r}$$

We obtain,

$$\boxed{\frac{\partial}{\partial x} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}} \qquad \& \qquad \boxed{\frac{\partial}{\partial y} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}.}$$

We may also be faced with the problem of changing coordinates for higher derivatives. The differentiatial equation $\nabla^2\Phi = 0$ is called Laplace's equation. It is important to the theory of electrostatics as well as fluid flow. In cartesian coordinates $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y$ and it follows that the Laplacian operator $\nabla \bullet \nabla = \partial_x^2 + \partial_y^2$. (we'll explore this sort of differentiation more at the end of this course). The example below builds off the results of the previous example, keep that in mind.

Example 4.5.13. Problem: Write Laplace's equation in polar coordinates.

$$\begin{aligned}
\nabla^2\Phi &= \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} \\
&= \frac{\partial}{\partial x}\left[\frac{\partial\Phi}{\partial x}\right] + \frac{\partial}{\partial y}\left[\frac{\partial\Phi}{\partial y}\right] \\
&= \left[\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}\right]\left[\cos\theta\frac{\partial\Phi}{\partial r} - \frac{\sin\theta}{r}\frac{\partial\Phi}{\partial\theta}\right] \\
&\quad + \left[\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta}\right]\left[\sin\theta\frac{\partial\Phi}{\partial r} + \frac{\cos\theta}{r}\frac{\partial\Phi}{\partial\theta}\right] \\
&= \boxed{\frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2}}
\end{aligned}$$

I invite the reader to fill in the details missing in the last step.

Example 4.5.14. *. 9*

Let $W(s,t) = F(u(s,t), v(s,t))$
where F, u, v are differentiable and

$u(1,0) = 2$ $\quad$ $v(1,0) = 3$ $\quad$ $F_u(2,3) = -1$

$u_s(1,0) = -2$ $\quad$ $v_s(1,0) = 5$ $\quad$ $F_v(2,3) = 10$

$u_t(1,0) = 6$ $\quad$ $v_t(1,0) = 4$

Find $W_s(1,0)$ and $W_t(1,0)$ using given data,

$$\frac{\partial W}{\partial s} = \frac{\partial}{\partial s}\left[F(u(s,t), v(s,t))\right] = \frac{\partial F}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial s} = F_u u_s + F_v v_s$$

$$\begin{aligned}
W_s(1,0) = \left.\frac{\partial W}{\partial s}\right|_{(1,0)} &= F_u(u(1,0), v(1,0))\, u_s(1,0) + F_v(u(1,0), v(1,0))\, v_s(1,0) \\
&= F_u(2,3)\cdot(-2) + F_v(2,3)\cdot(5) \\
&= (-1)(-2) + (10)(5) \\
&= \boxed{52 = W_s(1,0)}
\end{aligned}$$

Likewise,

$$\begin{aligned}
W_t(1,0) &= F_u(2,3)\, u_t(1,0) + F_v(2,3)\, v_t(1,0) \\
&= (-1)(6) + (10)(4) \\
&= \boxed{34 = W_t(1,0)}
\end{aligned}$$

Example 4.5.15. . *5*

find $\partial z/\partial x$ and $\partial z/\partial y$ via implicit differentiation.

Suppose that $x^2+y^2+z^2 = 3xyz$, assume $z = z(x,y)$,

$$\frac{\partial}{\partial x}\left[x^2+y^2+z^2\right] = \frac{\partial}{\partial x}\left[3xyz\right]$$

$$2x + 2z\frac{\partial z}{\partial x} = 3xy\frac{\partial z}{\partial x} + 3yz$$

$$\frac{\partial z}{\partial x}(2z - 3xy) = 3yz - 2x \quad \therefore \quad \boxed{\frac{\partial z}{\partial x} = \frac{3yz-2x}{2z-3xy}}$$

Likewise

$$2y + 2z\frac{\partial z}{\partial y} = 3xy\frac{\partial z}{\partial y} + 3xz$$

$$\frac{\partial z}{\partial y}(2z-3xy) = 3xz - 2y \quad \therefore \boxed{\frac{\partial z}{\partial y} = \frac{3xz-2y}{2z-3xy}}$$

Example 4.5.16. . *23*

$\boxed{\text{E59}}$ Suppose that $z = f(x,y)$ has continuous $f_x, f_y, f_{xy}, f_{yx}, f_{xx}$ etc...
and $x = r^2+s^2$ and $y = 2rs$ then note,

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = 2r\frac{\partial z}{\partial x} + 2s\frac{\partial z}{\partial y}$$

We wish to compute $\partial^2 z/\partial r^2$.

$$\begin{aligned}
\frac{\partial^2 z}{\partial r^2} &= \frac{\partial}{\partial r}\left[2r\frac{\partial z}{\partial x} + 2s\frac{\partial z}{\partial y}\right] \\
&= 2\frac{\partial z}{\partial x} + 2\frac{\partial}{\partial r}\left[\frac{\partial z}{\partial x}\right] + 2s\frac{\partial}{\partial r}\left[\frac{\partial z}{\partial y}\right] \\
&= 2\frac{\partial z}{\partial x} + 2\left[\frac{\partial^2 z}{\partial x^2}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y\partial x}\frac{\partial y}{\partial r}\right] + 2s\left[\frac{\partial^2 z}{\partial x\partial y}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y^2}\frac{\partial y}{\partial r}\right] \\
&= 2\frac{\partial z}{\partial x} + 2\left[\frac{\partial^2 z}{\partial x^2}\cdot 2r + \frac{\partial^2 z}{\partial y\partial x}\cdot 2s\right] + 2s\left[\frac{\partial^2 z}{\partial x\partial y}\cdot 2r + \frac{\partial^2 z}{\partial y^2}\cdot 2s\right] \\
&= 2f_x + 4rf_{xx} + 4sf_{xy} + 4srf_{yx} + 4s^2 f_{yy} \\
&= \boxed{2f_x + 4rf_{xx} + 4s^2 f_{yy} + 4s(1+r)f_{xy}} \quad \text{using Clairaut's Th}^{\underline{m}}.
\end{aligned}$$

(This is the best we can do w/o an explicit formula for $f(x,y)$.)

Example 4.5.17. . *24*

$\boxed{E58}$ Suppose $w = e^x \sin(y)$ and $y = st^2$, $x = \ln(s-t)$

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s}$$: identified that w is function of x & y.

$$= e^x \sin(y)\frac{1}{s-t} + e^x \cos(y)\, t^2$$: note $e^x = e^{\ln(s-t)} = s-t$.

$$= \boxed{\sin(st^2) + (s-t)t^2\cos(st^2) = \frac{\partial w}{\partial s}}$$

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t}$$

$$= e^x \sin(y)\left(\frac{-1}{s-t}\right) + e^x\cos(y)(2st)$$: again $e^x = s-t$,

$$= \boxed{-\sin(st^2) + 2st(s-t)\cos(st^2) = \frac{\partial w}{\partial t}}$$

Next, consider $F = F(x, y, z)$ and $\vec{r} = \vec{r}(u, v)$. In particular, we wish to differentiate the composite of $F : \mathbb{R}^3 \to \mathbb{R}$ with $\vec{r} : \mathbb{R}^2 \to \mathbb{R}^3$. The chain rules in this case are as follows:

$$\boxed{\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial u} \qquad \& \qquad \frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial v}.}$$

You can write these rules in terms of gradients and partial derivatives of vectors,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v}.$$

I explained in the preceding section that we can derive chain rules from the general derivative. For example,

$$F = F(x, y, z)$$

$$\vec{X} = \vec{X}(s,t) = \langle x(s,t),\, y(s,t),\, z(s,t) \rangle$$

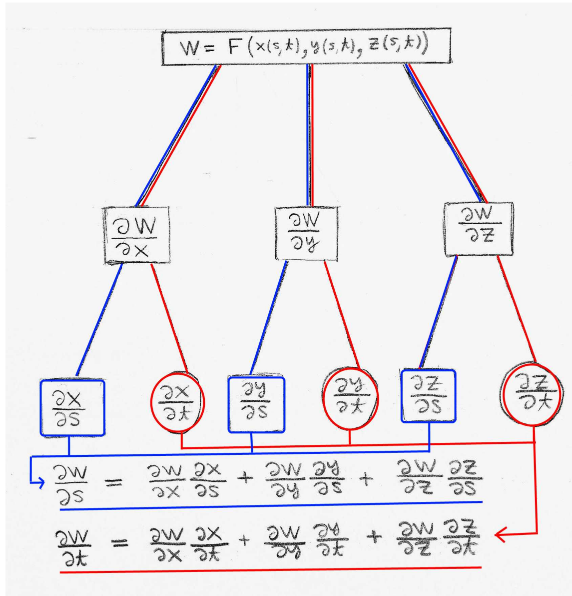
$$F' = \left[\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right]$$

$$\vec{X}' = \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix}$$

$$(F \circ \vec{X})' = F'\vec{X}' = \left[\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right]\begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix}$$

$$(F \circ \vec{X})' = \Big[\underbrace{\frac{\partial F}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial s}}_{\frac{\partial w}{\partial s}},\ \underbrace{\frac{\partial F}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial t}}_{\frac{\partial w}{\partial t}}\Big]$$

Some people would rather use these silly tree diagrams to remember the chain rule:



Personally, I much prefer to calculate with understanding as opposed to inventing new and unecessary mnuemonics. To each his own, you just need to find the way that works for you. Context is everything with chain rules.

Example 4.5.18. *. 3*

$$z = e^r\cos\theta \quad \text{with } r = st, \quad \theta = \sqrt{s^2+t^2}$$

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial r}\frac{\partial r}{\partial t} + \frac{\partial z}{\partial \theta}\frac{\partial \theta}{\partial t}$$

$$= (e^r\cos\theta)s - e^r\sin\theta\left(\frac{t}{\sqrt{s^2+t^2}}\right)$$

$$= \boxed{e^{st}\left[s\cos\sqrt{s^2+t^2} - \frac{t}{\sqrt{s^2+t^2}}\sin\sqrt{s^2+t^2}\right] = \frac{\partial z}{\partial t}}$$

$$\frac{\partial z}{\partial s} = (e^r\cos\theta)t - (e^r\sin\theta)\frac{s}{\sqrt{s^2+t^2}}$$

$$\boxed{\frac{\partial z}{\partial s} = e^{st}\left(t\cos\sqrt{s^2+t^2} - \frac{s}{\sqrt{s^2+t^2}}\sin\sqrt{s^2+t^2}\right)}$$

Example 4.5.19. . 4

$$z = x^2 + xy^3$$
$$x = uv^2 + w^3$$
$$y = u + ve^w$$

Find z_u, z_v, z_w when $u = 2$, $v = 1$, $w = 0$

Let's see we'll need to know

$$\frac{\partial x}{\partial u} = v^2 + w^3 \qquad x_u(2,1,0) = 1$$

$$\frac{\partial x}{\partial v} = 2uv \qquad x_v(2,1,0) = 4$$

$$\frac{\partial x}{\partial w} = 3w^2 \qquad x_w(2,1,0) = 0$$

$$\frac{\partial y}{\partial u} = 1 \qquad y_u(2,1,0) = 1$$

$$\frac{\partial y}{\partial v} = e^w \qquad y_v(2,1,0) = e^0 = 1$$

$$\frac{\partial y}{\partial w} = ve^w \qquad y_w(2,1,0) = 1e^0 = 1.$$

Also notice $x(2,1,0) = 2$ and $y(2,1,0) = 3$.

and $z_x(2,1,0) = 2x + y^3 = 4 + 27 = 31$.

$z_y(2,1,0) = 3xy^2 = 3(2)(9) = 54$.

Finally at the point $(2,1,0)$ in uvw-space,

$$\frac{\partial z}{\partial u} = z_x \frac{\partial x}{\partial u} + z_y \frac{\partial y}{\partial u} = (31) + (54) = \boxed{85 = \left.\frac{\partial z}{\partial u}\right|_{(2,1,0)}}$$

$$\frac{\partial z}{\partial v} = z_x \frac{\partial x}{\partial v} + z_y \frac{\partial y}{\partial v} = (31)4 + (54) = \boxed{178 = \left.\frac{\partial z}{\partial v}\right|_{(2,1,0)}}$$

$$\frac{\partial z}{\partial w} = z_x \frac{\partial x}{\partial w} + z_y \frac{\partial y}{\partial w} = (31)(0) + (54)1 = \boxed{54 = \left.\frac{\partial z}{\partial w}\right|_{(2,1,0)}}$$

Example 4.5.20. . *10*

I'll find $\frac{\partial z}{\partial u}$ for $z = x^2 + xy^3$ with $\begin{array}{l} x = uv^2 + w^3 \\ y = u + ve^w \end{array}$ when $u = 2$, $v = 1$ and $w = 0$,

$$\begin{aligned}
\frac{\partial z}{\partial u} &= \frac{\partial}{\partial u}\Big(z\,(x(u,v,w),\ y(u,v,w))\Big) \\
&= \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} \\
&= (2x + y^3)v^2 + (3xy^2)\cdot 1
\end{aligned}$$

Notice $x(2,1,0) = 2(1) + 0^3 = 2$ and $y(2,1,0) = 2 + 1 \cdot e^0 = 3$.

$$\left.\frac{\partial z}{\partial u}\right|_{(2,1,0)} = \underbrace{(2(2) + 3^3)1^2}_{31} + \underbrace{3(2)(3)^2}_{54} = \boxed{85 = \left.\frac{\partial z}{\partial u}\right|_{(2,1,0)}}$$

Example 4.5.21. . *2*

$$\begin{array}{lll}
z = x^2y^3 & z_x = 2xy^3 & z_y = 3x^2y^2 \\
x = s\cos t & x_s = \cos t & x_t = -s\sin t \\
y = s\sin t & y_s = \sin t & y_t = s\cos t
\end{array}$$

$$\begin{aligned}
\frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \\
&= (2xy^3)(-s\sin t) + (3x^2y^2)(s\cos t) \\
&= -2(s\cos t)(s\sin t)^4 + 3(s\sin t)^2(s\cos t)^3 \\
&= \boxed{s^5\left[-2\cos t\sin^4 t + 3\sin^2 t\cos^3 t\right] = \frac{\partial z}{\partial t}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \\
&= (2xy^3)\cos t + (3x^2y^2)\sin t \\
&= 2(s\cos t)(s\sin t)^3\cos t + 3(s\cos t)^2(s\sin t)^2\sin t \\
&= \boxed{s^2\left[2\cos^2 t\sin^3 t + 3\cos^2 t\sin^3 t\right] = \frac{\partial z}{\partial s}}
\end{aligned}$$

Example 4.5.22. . *6*

$$x - z = \tan^{-1}(yz) \quad \text{find } \partial z/\partial x \text{ and } \partial z/\partial y,$$

$$\frac{\partial}{\partial x}\left[x - z\right] = \frac{\partial}{\partial x}\left[\tan^{-1}(yz)\right]$$

$$1 - \frac{\partial z}{\partial x} = \frac{1}{1+(yz)^2}\frac{\partial}{\partial x}\left[yz\right] = \frac{y}{1+(yz)^2}\frac{\partial z}{\partial x}$$

$$\frac{\partial z}{\partial x}\left[\frac{y}{1+(yz)^2} + 1\right] = 1$$

$$\frac{\partial z}{\partial x} = \frac{1}{\frac{y}{1+(yz)^2}+1} = \boxed{\frac{1+y^2z^2}{y+1+y^2z^2} = \frac{\partial z}{\partial x}}$$

Next take $\partial/\partial y$,

$$-\frac{\partial z}{\partial y} = \frac{1}{1+(yz)^2}\frac{\partial}{\partial y}(yz)$$

$$= \frac{1}{1+y^2z^2}\left[z + y\frac{\partial z}{\partial y}\right]$$

$$\Rightarrow \frac{\partial z}{\partial y}\left[\frac{y}{1+y^2z^2}+1\right] = \frac{-z}{1+y^2z^2}$$

$$\Rightarrow \frac{\partial z}{\partial y}\left(y + 1 + y^2z^2\right) = -z$$

$$\Rightarrow \boxed{\frac{\partial z}{\partial y} = \frac{-z}{y+1+y^2z^2}}$$

Example 4.5.23. . *12*

I'll find $\frac{\partial z}{\partial u}$ for $z = x^2 + xy^3$ with $x = uv^2 + w^3$, $y = u + ve^w$ when $u = 2$, $v = 1$ and $w = 0$,

$$\frac{\partial z}{\partial u} = \frac{\partial}{\partial u}\Big(z\,(x(u,v,w),\, y(u,v,w))\Big)$$

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

$$= (2x+y^3)v^2 + (3xy^2)\cdot 1$$

Notice $x(2,1,0) = 2(1) + 0^3 = 2$ and $y(2,1,0) = 2 + 1\cdot e^0 = 3$.

$$\left.\frac{\partial z}{\partial u}\right|_{(2,1,0)} = \underbrace{(2(2)+3^3)1^2}_{31} + \underbrace{3(2)(3)^2}_{54} = \boxed{85 = \left.\frac{\partial z}{\partial u}\right|_{(2,1,0)}}$$

Example 4.5.24. . *1*

$$\left(\frac{\partial f}{\partial x} = f'(xy)\right)$$

$$z = f(xy)$$

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x}\frac{\partial}{\partial x}(xy) = \boxed{y\frac{\partial f}{\partial x} = z_x}$$

$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial y}\frac{\partial}{\partial y}(xy) = \boxed{x\frac{\partial f}{\partial y} = z_y}$$

$$z = f(x/y)$$

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x}\frac{\partial}{\partial x}\left(\frac{x}{y}\right) = \boxed{\frac{1}{y}\frac{\partial f}{\partial x} = z_x}$$

$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial y}\frac{\partial}{\partial y}\left(\frac{x}{y}\right) = \boxed{\frac{-x}{y^2}\frac{\partial f}{\partial y} = z_y}$$

Example 4.5.25. . *11*

I'll find $\frac{\partial z}{\partial u}$ for $z = x^2 + xy^3$ with $x = uv^2 + w^3$, $y = u + ve^w$ when $u = 2$, $v = 1$ and $w = 0$,

$$\frac{\partial z}{\partial u} = \frac{\partial}{\partial u}\Big(z\big(x(u,v,w),\, y(u,v,w)\big)\Big)$$

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

$$= (2x + y^3)v^2 + (3xy^2)\cdot 1$$

Notice $x(2,1,0) = 2(1) + 0^3 = 2$ and $y(2,1,0) = 2 + 1\cdot e^0 = 3$.

$$\frac{\partial z}{\partial u}\bigg|_{(2,1,0)} = \underbrace{\left(2(2) + 3^3\right)}_{31} 1^2 + \underbrace{3(2)(3)^2}_{54} = \boxed{85 = \frac{\partial z}{\partial u}\bigg|_{(2,1,0)}}$$

4.6 tangent spaces and the normal vector field

In this section we wish to analyze the tangent space for a smooth surface. We assume the surface in consideration is smooth so that the calculations are not complicated by exceptional cases. In particular, we wish to analyze a surface S in three particular views:

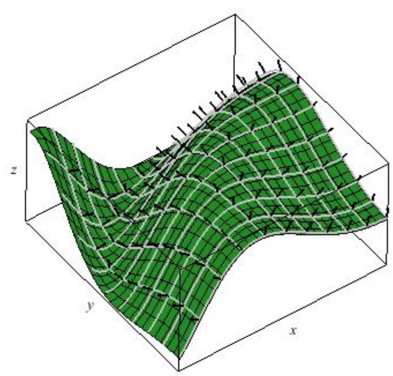
1. as a **level surface** the set S is the solution set of $F(x, y, z) = 0$

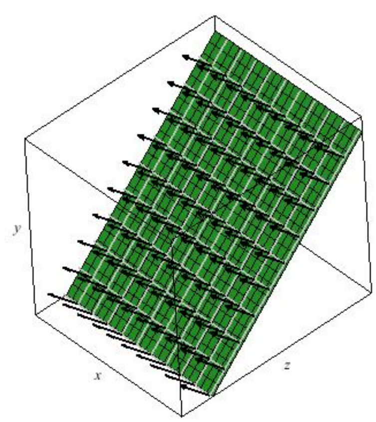
2. as a **parametrized surface** we see S as the image of $\vec{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$

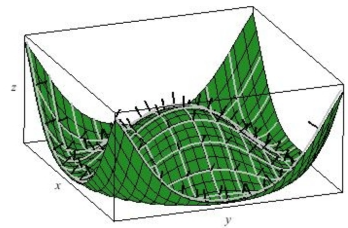
3. as a **graph** we see S as the solution set of $z = f(x, y)$

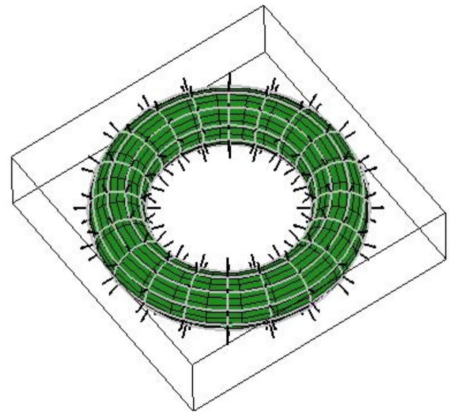
As we have discussed previously it is only sometimes possible to cover all of S as a graph. Moreover, each view has it's advantages. My goal in this section is to explain how to find the tangent space and normal vector field for S in each of these views. We've already done a lot of calcation towards these questions in the last section.

For your viewing enjoyment I have included a few figures of surfaces which have coordinate curves in gray and little normal vectors in black. I have animations of these on the webpage, perhaps it helps bring to life the fact the normals pick out a side of orientable surfaces.









Our goal in this section is to find formulas for the little black arrows.

4.6.1 level surfaces and tangent space

In Example 4.5.7 we proved that curves in the solution set of $F(x, y, z) = k$ have tangent vectors which are perpendicular to ∇F. It follows that the normal vector for the tangent plane at $(x_o, y_o, z_o) \in S$ is simply $\nabla F(x_o, y_o, z_o)$. The tangent plane has equation:

$$\boxed{\nabla F(x_o, y_o, z_o) \bullet \langle x - x_o, y - y_o, z - z_o\rangle = 0.}$$

The **normal vector field on** S is given by the assignment

$$(x, y, z) \to \nabla F(x, y, z)$$

for each $(x, y, z) \in S$.

Remark 4.6.1.

The choice of level function matters. If we multiply the equation by a negative quantity the direction of the gradient flips over and hence the normal vector field flips to the other side of the surface. As an example, $F(x, y, z) = x^2 + y^2 + z^2 = 1$ has $\nabla F = \langle 2x, 2y, 2z\rangle$ whereas $G(x, y, z) = -x^2 - y^2 - z^2 = -1$ has $\nabla G = \langle -2x, -2y, -2z\rangle$. We say $F = 1$ is the sphere **oriented outwards** whereas $G = -1$ is the sphere **oriented inwards**.

Example 4.6.2. .

Find eq$^{\underline{n}}$ of tangent plane and normal line to the surface $x^2 - 2y^2 + z^2 + yz = 2$ at $(2, 1, -1)$.

Notice $2^2 - 2(1)^2 + (-1)^2 + (1)(-1) = 4 - 2 + 1 - 1 = 2$ so the point is indeed on the surface as claimed. This is a level surface let $F(x, y, z) = x^2 - 2y^2 + z^2 + yz - 2$ then $F = 0$ is the surface.

$$\nabla F = \langle 2x, -4y + z, 2z + y \rangle$$

$$(\nabla F)(2,1,-1) = \langle 4, -4-1, -2+1 \rangle = \langle 4, -5, -1 \rangle.$$

This is the normal to the level surface at $(2,1,-1)$. Hence

$\boxed{4(x-2) - 5(y-1) - (z+1) = 0}$ tangent plane to $F = 0$ at $(2, 1, -1)$.

We have a direction $\langle 4, -5, -1 \rangle$ and a point $(2,1,-1)$ the line through this point with that direction is simply

$$\boxed{r(t) = \langle 2,1,-1 \rangle + t\langle 4, -5, -1 \rangle}$$

4.6.2 parametrized surfaces and tangent space

Suppose a surface S can either be viewed as a level surface $F(x, y, z) = k$ or as a parametrized surface by the mapping $\vec{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$. In particular, if we denote the parameters by u, v and write $\vec{r} = \langle x, y, z \rangle$ then these viewpoints are connected by the equation $F(\vec{r}(u, v)) = k$ for all $u, v \in D$. The chain rules in this case are as follows:

$$\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial u} \qquad \& \qquad \frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial v}.$$

You can write these rules in terms of gradients and partial derivatives of vectors,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v}.$$

Differentiate $F(\vec{r}(u, v)) = k$ with respect to u or v to obtain,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} = 0 \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v} = 0.$$

The vectors $\frac{\partial \vec{r}}{\partial u}$ and $\frac{\partial \vec{r}}{\partial v}$ are perpendicular to $\nabla F(\vec{r}(u, v))$. We envision all three of these vectors attached at the point $\vec{r}(u, v)$ of S. The curves

$$\vec{\alpha}(u) = \vec{r}(u, v_o) \qquad \vec{\beta}(v) = \vec{r}(u_o, v)$$

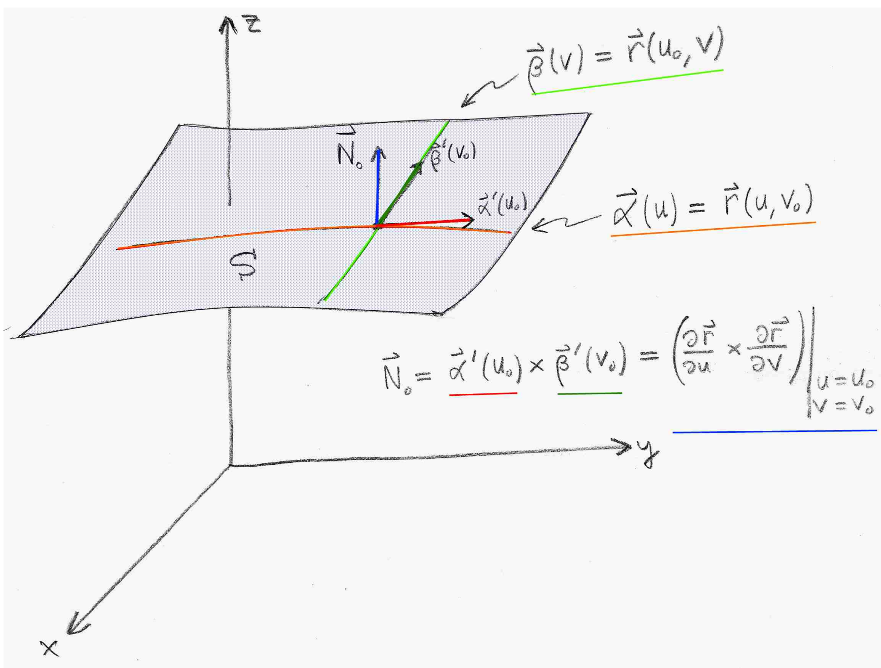
are the coordinate curves through $\vec{r}(u_o, v_o)$. The tangent vectors at $\vec{r}(u_o, v_o)$ to these curves are given by

$$\vec{\alpha}\,'(u_o) = \frac{d}{du}\big[\vec{r}(u, v_o)\big]_{u_o} = \frac{\partial \vec{r}}{\partial u}(u_o) \;\;\&\;\; \vec{\beta}\,'(v_o) = \frac{d}{dv}\big[\vec{r}(u_o, v)\big]_{v_o} = \frac{\partial \vec{r}}{\partial v}(v_o)$$

Therefore, the tangent vectors to the coordinate curves are perpendicular to the gradient vector of the corresponding level curve. In three dimensional space it follows that the cross product of $\vec{\alpha}\,'(u_o)$ with $\vec{\beta}\,'(v_o)$ must be colinear to $\nabla F(\vec{r}(u_o, v_o))$. Therefore, we define

$$\boxed{\vec{N}(u, v) = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}.}$$

The vector field $\vec{r}(u, v) \to \vec{N}(u, v)$ defines the **normal vector field** of $\vec{r}$. If a surface S has a non-vanishing normal vector field then it is said to be **oriented**. Clearly it is easier to calculate the normal in the level surface formulation since gradients are way easier than cross products. However, we will find that the parametric viewpoint is an essential part of the definition of the surface integral for a vector field. The diagram below indicates how a particular vector in the normal vector field is calculated in the parametric setting:



Remark 4.6.3.

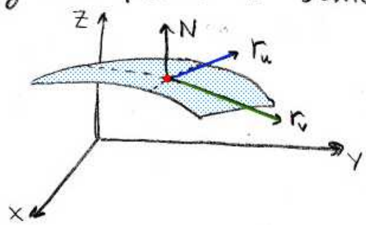
The ordering of the parameters matters. If we swap the order of the parameters it flips the normal vector field. Suppose S_1 is oriented by $\vec{r}_1(u, v) = \langle f(u, v), g(u, v), h(u, v)\rangle$ and $\vec{r}_2(v, u) = \langle f(u, v), g(u, v), h(u, v)\rangle$. The normal vector field induced from $\vec{r}_1$ by our conventions is $\vec{N}_1 = \partial_u \vec{r}_1 \times \partial_v \vec{r}_1$ whereas the normal vector field induced from $\vec{r}_2$ is $\vec{N}_2 = \partial_v \vec{r}_2 \times \partial_u \vec{r}_2$. Since $\vec{r}_1(u, v) = \vec{r}_2(v, u)$ it follows that $\vec{N}_1 = -\vec{N}_2$. My point? Beware the order.

Example 4.6.4. .

Suppose we have a parametrically defined surface,

$$\vec{r}(u,v) = \langle u^2,\ 2u\sin v,\ u\cos v\rangle$$

find the tangent plane at $\vec{r}(1,0) = \langle 1, 0, 1\rangle$. This case is different than the early problems in this section. basically the picture is something like the following



$$\vec{r}_u = \frac{\partial\vec{r}}{\partial u} \qquad \vec{r}_v = \frac{\partial\vec{r}}{\partial v} \qquad \vec{N} = \vec{r}_u \times \vec{r}_v$$

Now all of this should be done at the point of intersest where $u=1$ and $v=0$. Calculate then,

$$\frac{\partial\vec{r}}{\partial u} = \langle 2u,\ 2\sin v,\ \cos v\rangle \qquad \frac{\partial\vec{r}}{\partial u}(1,0) = \langle 2, 0, 1\rangle$$

$$\frac{\partial\vec{r}}{\partial v} = \langle 0,\ 2u\cos v,\ -u\sin v\rangle \qquad \frac{\partial\vec{r}}{\partial v}(1,0) = \langle 0, 2, 0\rangle$$

$$(\vec{r}_u \times \vec{r}_v)\Big|_{(1,0)} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 2 & 0 & 1 \\ 0 & 2 & 0 \end{vmatrix} = -2\hat{i} + 4\hat{k} = \langle -2, 0, 4\rangle$$

We have the normal $\vec{N} = \langle -2, 0, 2\rangle$ and a point $(1,0,1)$ the tangent plane has the eq$^{\underline{n}}$,

$$\boxed{-2(x-1) + 4(z-1) = 0}$$

Example 4.6.5. *Suppose the plane $F(x,y,z) = a(x-x_o) + b(y-y_o) + c(z-z_o) = 0$ contains non-colinear vectors $\vec{A}$ and $\vec{B}$. Note $\nabla F = \langle a,b,c\rangle$, the normal derived from the level function matches the natural normal suggested by the equation for the plane. Next, consider the parametrization naturally induced from $\vec{A}, \vec{B}$ and the base-point (x_o, y_o, z_o),*

$$\vec{r}(u,v) = \langle x_o, y_o, z_o\rangle + u\vec{A} + v\vec{B}.$$

In this case calculation of the tangent vectors to the coordinate curves is easy:

$$\frac{\partial\vec{r}}{\partial u} = \vec{A} \qquad \frac{\partial\vec{r}}{\partial v} = \vec{B}$$

Thus $\vec{N}(u,v) = \vec{A}\times\vec{B}$. The normal vector field to a plane is a constant vector field. Geometry indicates that $\vec{A}\times\vec{B} = \lambda\langle a,b,c\rangle$ for some nonzero constant λ.

Example 4.6.6. .

Consider the parametrized surface

$$\vec{r}(u,v) = \langle u, \ln(uv), v \rangle$$

$$\vec{r}_u = \langle 1, \tfrac{1}{uv}\cdot v, 0 \rangle = \langle 1, \tfrac{1}{u}, 0 \rangle$$

$$\vec{r}_v = \langle 0, \tfrac{1}{v}, 1 \rangle$$

$$\vec{r}_u \times \vec{r}_v = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 1/u & 0 \\ 0 & 1/v & 1 \end{vmatrix} = \langle \tfrac{1}{u}, -1, \tfrac{1}{v} \rangle = \vec{N}(u,v)$$

we could find the tangent plane most anywhere now, but lets consider $u=1, v=1$ where

$$\vec{r}(1,1) = \langle 1, \ln(1), 1 \rangle = \langle 1, 0, 1 \rangle.$$

$$\vec{N}(1,1) = \langle 1, -1, 1 \rangle$$

Hence the tangent plane is,

$$\boxed{x - 1 - (y-0) + z - 1 = 0 = x - y + z - 2}$$

4.6.3 tangent plane to a graph

The graphical viewpoint is connected to the level-surface view and the parametric view by the following: given that S is the solution set of $z = f(x,y)$ we can

1. write S as the level surface $F(x,y,z) = 0$ for $F(x,y,z) = z - f(x,y)$.

2. write S as a parametrized surface with parameters x, y and $\vec{r}(x,y) = \langle x, y, f(x,y) \rangle$.

Notice there is some ambiguity in the normal vectors which are induced. If we chose $-F = 0$ then that flips over the normal and if we swapped the order of the parameters x, y then that would also flip the normal vector $\vec{N}(x,y)$. These ambiguities must be dealt with as we do calculations on surfaces. Picking an orientation specifies a side to the surface. Equivalently, an **oriented surface is a set of points paired with a normal vector field on the surface.**

Remark 4.6.7.

Question: if S is oriented and we describe S_1 by $F(x,y,z) = 0$ for $F(x,y,z) = z - f(x,y)$ then is the same oriented surface as the parametrized surface S_2 with parameters x, y and $\vec{r}(x,y) = \langle x, y, f(x,y)\rangle$?

Solution: to begin note that as point-sets it is clear that $S_1 = S_2$ so the question reduces to the problem of ascertaining if the normal vector fields match-up. Calculate, from the level-surface viewpoint the normal vector field at (x,y,z) on S_1 is

$$\vec{N}(x,y,z) = \nabla F = \langle -f_x, -f_y, 1\rangle$$

On the other hand, from the parametric viewpoint we calculate for $(x,y) \in dom(f)$,

$$\frac{\partial \vec{r}}{\partial x} = \langle 1, 0, f_x\rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial y} = \langle 0, 1, f_y\rangle$$

and the cross-product

$$\begin{aligned}
\vec{N}(x,y) &= \frac{\partial \vec{r}}{\partial x} \times \frac{\partial \vec{r}}{\partial y} \\
&= (\,\widehat{x} + f_x\,\widehat{z}) \times (\,\widehat{y} + f_y\,\widehat{z}) \\
&= \,\widehat{x} \times \,\widehat{y} + f_y\,\widehat{x} \times \,\widehat{z} + f_x\,\widehat{z} \times \,\widehat{y} \\
&= \,\widehat{z} - f_y\,\widehat{y} - f_x\,\widehat{x} \\
&= \langle -f_x, -f_y, 1\rangle.
\end{aligned}$$

Therefore, if we change viewpoints as advocated at the beginning of the subsection we will maintain the natural orientation. This is the reason I wrote $F(x,y,z) = z - f(x,y)$ as opposed to $G(x,y,z) = f(x,y) - z$.

Example 4.6.8. .

E71 Find the eq$^{\underline{n}}$ of the tangent plane at $(1,2,5)$ for $f(x,y) = x^2 + y^2$. We calculate $f_x(1,2)$ and $f_y(1,2)$,

$$f_x(1,2) = 2x\big|_{(1,2)} = 2 \qquad f(1,2) = 1^2 + 2^2 = 5$$

$$f_y(1,2) = 2y\big|_{(1,2)} = 4$$

Thus the tangent plane is $\boxed{z = 5 + 2(x-1) + 4(y-2)}$

Example 4.6.9. .

Consider $z = \exp(x^2 - y^2)$ find tangent plane to surface at $(1,-1,1)$. We calculate,

$$\frac{\partial z}{\partial x} = 2x\exp(x^2-y^2) \quad \therefore \quad \frac{\partial z}{\partial x}\bigg|_{(1,-1)} = 2e^0 = 2$$

$$\frac{\partial z}{\partial y} = -2y\exp(x^2-y^2) \quad \therefore \quad \frac{\partial z}{\partial y}\bigg|_{(1,-1)} = -2(-1)e^0 = 2$$

Then the eq$^{\underline{n}}$ of the tangent plane to $z = f(x,y)$ at (x_o, y_o),

$$z - z_o = \frac{\partial f}{\partial x}\bigg|_{(x_o,y_o)}(x - x_o) + \frac{\partial f}{\partial y}\bigg|_{(x_o,y_o)}(y - y_o)$$

In particular, for $(x_o, y_o, z_o) = (1,-1,1)$,

$$\boxed{z - 1 = 2(x-1) + 2(y+1)}$$

<u>Remark</u>: I prefer to look at $F(x,y,z) = \exp(x^2-y^2) - z$ then $\nabla F \equiv \left\langle \frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right\rangle$ gives the normal to the tangent planes of the surface $F = 0$. In this problem, we've already calculated F_x and F_y at $(1,-1,1)$ and $F_z = -1$ thus

$$\nabla F\big|_{(1,-1,1)} = \langle 2, 2, -1 \rangle$$

then the plane at $(1,-1,1)$ is simply

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

which is the same as we found via the text's formula.

Example 4.6.10. .

Find tangent plane to $z = 4x^2 - y^2 + 2y$ at $(-1,2,4)$.

Let $z = f(x,y) = 4x^2 - y^2 + 2y$

$$\frac{\partial f}{\partial x} = 8x \qquad f_x(-1,2) = -8$$

$$\frac{\partial f}{\partial y} = -2y + 2 \qquad f_y(-1,2) = -2$$

Thus $z = f(-1,2) + f_x(-1,2)(x+1) + f_y(-1,2)(y-2)$

Gives $\boxed{z = 4 - 8(x+1) - 2(y-2)}$

4.7 partial differentiation with side conditions

Every chain rule in the preceding section follows as a subcase of the chain rule for the general derivative. In this section the rigorous justification is given by the implicit or inverse function theorems. I will not even state those here[17]. I discuss them in advanced calculus and those notes are available if you'd like to read about the theoretical underpinning for the calculations in this section. I will show how to formally calculate in this section. In other words, I will teach you symbol pushing techniques. To begin, we define the total differential.

Definition 4.7.1.

If $f = f(x_1, x_2, \dots, x_n)$ then $df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n$.

Example 4.7.2. *Suppose $E = pv + t^2$ then $dE = vdp + pdv + 2tdt$. In this example the dependent variable is E whereas the independent variables are p, v and t.*

Example 4.7.3. Problem: *what are $\partial F/\partial x$ and $\partial F/\partial y$ if we know that $F = F(x,y)$ and $dF = (x^2+y)dx - \cos(xy)dy$.*
Solution: *if $F = F(x,y)$ then the total differential has the form $dF = F_x dx + F_y dy$. We simply compare the general form to the given $dF = (x^2+y)dx - \cos(xy)dy$ to obtain:*

$$\frac{\partial F}{\partial x} = x^2 + y, \qquad \frac{\partial F}{\partial y} = -\cos(xy).$$

Example 4.7.4. .

$u = e^{-t}\sin(s+2t)$ so u is a fnct. of t and s,

$$\begin{aligned} du &= \frac{\partial u}{\partial s}ds + \frac{\partial u}{\partial t}dt \\ &= e^{-t}\cos(s+2t)\,ds + \left[-e^{-t}\sin(s+2t) + e^{-t}\cdot 2\cos(s+2t)\right]dt \\ &= \boxed{e^{-t}\cos(s+2t)\,ds + e^{-t}\left[2\cos(s+2t) - \sin(s+2t)\right]dt = du} \end{aligned}$$

$w = xy\exp(xz)$ find the total-differential

$$\begin{aligned} dw &= \frac{\partial w}{\partial x}dx + \frac{\partial w}{\partial y}dy + \frac{\partial w}{\partial z}dz \\ &= \left[y\exp(xz) + xyz\exp(xz)\right]dx + \left[x\exp(xz)\right]dy + \left[x^2y\exp(xz)\right]dz \\ &= \boxed{e^{xz}\left\{(y + xyz)dx + x\,dy + x^2y\,dz\right\} = dw} \end{aligned}$$

[17] I do give the easiest version of the implicit function theorem later in this section, but it does not really play a computational role beyond the question of existence

Example 4.7.5. *Differentials are useful for error estimation.*

E74 Let $R = R_1 + R_2 = g(R_1, R_2)$. That is assume the resistors are in series this time. Find dR in this case,

$$dR = \frac{\partial R}{\partial R_1}\Big|_{(10,2)} dR_1 + \frac{\partial R}{\partial R_2}\Big|_{(10,2)} dR_2$$

$$= dR_1 + dR_2$$

$$= 1 + 0.5$$

$$= \boxed{1.5 = dR} \quad (\text{Mostly from } \Delta R_1)$$

Example 4.7.6. *Here's another error estimation calculation.*

E73 It is known that if we place resistors R_1 and R_2 in parallell then the effective resistance R of the system is

$$R = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1}$$

we can view $R = f(R_1, R_2)$, it is a function of two variables. Now suppose $R_1 = 10\Omega \pm 1\Omega$ and $R_2 = 2\Omega \pm 0.5\Omega$, here $\Omega \equiv$ "ohm" and we'll drop them for convenience sake. Interpretation:

$$dR_1 = 1 \quad \text{and} \quad dR_2 = 0.5$$

what is the order of our uncertainty in R then? Essentially upto a convention or two its the total differential in R,

$$dR = \frac{\partial R}{\partial R_1}\Big|_{(10,2)} dR_1 + \frac{\partial R}{\partial R_2}\Big|_{(10,2)} dR_2$$

$$= \left(\frac{-1}{\left(\frac{1}{R_1}+\frac{1}{R_2}\right)^2}\right)\left(\frac{-1}{R_1^2}\right)\Big|_{(10,2)} dR_1 + \left(\frac{-1}{\left(\frac{1}{R_1}+\frac{1}{R_2}\right)^2}\right)\left(\frac{-1}{R_2^2}\right)\Big|_{(10,2)} dR_2$$

$$= \frac{dR_1}{(1+R_1/R_2)^2}\Big|_{(10,2)} + \frac{dR_2}{(R_2/R_1+1)^2}\Big|_{(10,2)}$$

$$= \frac{1}{36} + \left(\frac{25}{36}\right)(0.5) = \frac{13.5}{36} = \boxed{0.375} \quad (\text{Mostly from } \Delta R_2.)$$

Example 4.7.7. *What? You want more error estimation? Here, take this.*

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

where $R_1 = 25\Omega$, $R_2 = 40\Omega$, $R_3 = 50\Omega$ with errors of 0.5% in each case. Estimate max error in calculated value of R.

Strategy: If $f(x,y,z)$ then $df = f_x dx + f_y dy + f_z dz$ is the total differential of f. This gives us the max error in f to be df based on assumed errors of dx, dy, dz in x, y, z respective. In this problem $x \sim R_1$, $y \sim R_2$, $z \sim R_3$ $f \sim R$.

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

Calculate: for $k = 1, 2, 3$,

$$\frac{\partial R}{\partial R_k} = \frac{-1}{\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right)^2} \frac{\partial}{\partial R_k}\left[\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right]$$

$$= -R^2\left[\frac{-1}{R_1^2}\delta_{1k} - \frac{1}{R_2^2}\delta_{2k} - \frac{1}{R_3^2}\delta_{3k}\right] \quad : \quad \frac{\partial R_j}{\partial R_k} = \delta_{jk}$$

$$\delta_{jk} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

Thus, look at each case and obtain

$$\frac{\partial R}{\partial R_1} = \frac{R^2}{R_1^2}, \quad \frac{\partial R}{\partial R_2} = \frac{R^2}{R_2^2}, \quad \frac{\partial R}{\partial R_3} = \frac{R^2}{R_3^2}$$

Thus,

$$dR = \frac{\partial R}{\partial R_1} dR_1 + \frac{\partial R}{\partial R_2} dR_2 + \frac{\partial R}{\partial R_3} dR_3$$

$$= R^2\left[(1/R_1^2)dR_1 + (1/R_2^2)dR_2 + (1/R_3^2)dR_3\right]$$

We had R^2 in each $\partial R/\partial R_k$ so I factored it out.

Notice $\Delta R_1 = 0.005 R_1$, $\Delta R_2 = 0.005 R_2$ and $\Delta R_3 = 0.005 R_3$ so [using Δ instead of d now to emphasize finite increments]

$$\Delta R = R^2\left(\frac{1}{R_1^2}0.005 R_1 + \frac{1}{R_2^2}0.005 R_2 + \frac{1}{R_3^2}0.005 R_3\right)$$

$$= 0.005 R^2 \underbrace{\left(1/R_1 + 1/R_2 + 1/R_3\right)}_{1/R} = \underline{0.005 R = \Delta R}$$

And we know $R_1 = 25\Omega$, $R_2 = 40\Omega$, $R_3 = 50\Omega$ so

$$R = \frac{1}{\frac{1}{25\Omega} + \frac{1}{40\Omega} + \frac{1}{50\Omega}} = 11.76\Omega$$

Thus,

$$\Delta R \cong (0.005)(11.76\Omega) \cong \boxed{0.05882\Omega = \Delta R}$$

Of course we ought to pay significant figures a little respect here so $\boxed{\Delta R = 0.059\Omega}$ which amounts to a max error of $\approx 0.05\%$ in the total resistance.

It is very likely I will do something different in lecture.

Example 4.7.8. *Suppose $w = xyz$ then $dw = yzdx + xzdy + xydz$. On the other hand, we can solve for $z = z(x, y, w)$*

$$z = \frac{w}{xy} \qquad \Rightarrow \qquad dz = -\frac{w}{x^2y}dx - \frac{w}{xy^2}dy + \frac{1}{xy}dw. \ \star$$

If we solve $dw = yzdx + xzdy + xydz$ directly for dz we obtain:

$$dz = -\frac{z}{x}dx - \frac{z}{y}dy + \frac{1}{xy}dw \quad \star\star.$$

Are $\star$ and $\star\star$ consistent? Well, yes. Note $\frac{w}{x^2y} = \frac{xyz}{x^2y} = \frac{z}{x}$ and $\frac{w}{xy^2} = \frac{xyz}{xy^2} = \frac{z}{y}$.

Which variables are independent/dependent in the example above? It depends. In this initial portion of the example we treated x, y, z as independent whereas w was dependent. But, in the last half we treated x, y, w as independent and z was the dependent variable. Consider this, if I ask you what the value of $\frac{\partial z}{\partial x}$ is in the example above then this question is ambiguous!

$$\underbrace{\frac{\partial z}{\partial x} = 0}_{z \text{ indpendent of } x} \qquad \text{verses} \qquad \underbrace{\frac{\partial z}{\partial x} = \frac{-z}{x}}_{z \text{ depends on } x}$$

Obviously this sort of ambiguity is rather unpleasant. A natural solution to this trouble is simply to write a bit more when variables are used in multiple contexts. In particular,

$$\underbrace{\frac{\partial z}{\partial x}\bigg|_{y,z} = 0}_{\textit{means } x,y,z \textit{ independent}} \qquad \text{is different than} \qquad \underbrace{\frac{\partial z}{\partial x}\bigg|_{y,w} = \frac{-z}{x}}_{\textit{means } x,y,w \textit{ independent}}.$$

The key concept is that all the other independent variables are held fixed as an indpendent variable is partial differentiated. Holding y, z fixed as x varies means z does not change hence $\frac{\partial z}{\partial x}\big|_{y,z} = 0$. On the other hand, if we hold y, w fixed as x varies then the change in z need not be trivial; $\frac{\partial z}{\partial x}\big|_{y,w} = \frac{-z}{x}$. Let me expand on how this notation interfaces with the total differential.

Definition 4.7.9.

If w, x, y, z are variables then

$$dw = \frac{\partial w}{\partial x}\bigg|_{y,z} dx + \frac{\partial w}{\partial y}\bigg|_{x,z} dy + \frac{\partial w}{\partial z}\bigg|_{x,y} dz.$$

Alternatively,

$$dx = \frac{\partial x}{\partial w}\bigg|_{y,z} dw + \frac{\partial x}{\partial y}\bigg|_{w,z} dy + \frac{\partial x}{\partial z}\bigg|_{w,y} dz.$$

The larger idea here is that we can identify partial derivatives from the coefficients in equations of differentials. I'd say a differential equation but you might get the wrong idea... Incidentally, there is a whole theory of solving differential equations by clever use of differentials, it's called the *method of characteristics*. I have books if you are interested.

Example 4.7.10. *Suppose $w = x + y + z$ and $x + y = wz$ then calculate $\frac{\partial w}{\partial x}\big|_y$ and $\frac{\partial w}{\partial x}\big|_z$. Notice we must choose dependent and independent variables to make sense of partial derivatives in question.*

1. *suppose w, z both depend on x, y. Calculate,*

$$\frac{\partial w}{\partial x}\bigg|_y = \frac{\partial}{\partial x}\bigg|_y (x+y+z) = \frac{\partial x}{\partial x}\bigg|_y + \frac{\partial y}{\partial x}\bigg|_y + \frac{\partial z}{\partial x}\bigg|_y = 1 + 0 + \frac{\partial z}{\partial x}\bigg|_y \quad \star$$

To calculate further we need to eliminate w by substituting $w = x + y + z$ into $x + y = wz$; thus $x + y = (x + y + z)z$ hence $dx + dy = (dx + dy + dz)z + (x + y + z)dz$

$$(2z + x + y)dz = (1 - z)dx + (1 - z)dy \quad \star\star$$

Therefore,

$$dz = \frac{1-z}{2z+x+y}dx + \frac{1-z}{2z+x+y}dy = \frac{\partial z}{\partial x}\bigg|_y dx + \frac{\partial z}{\partial y}\bigg|_x dy \quad \Rightarrow \quad \frac{\partial z}{\partial x}\bigg|_y = \frac{1-z}{2z+x+y}.$$

Returning to $\star$ we derive

$$\boxed{\frac{\partial w}{\partial x}\bigg|_y = 1 + \frac{1-z}{2z+x+y}.}$$

2. *suppose w, y both depend on x, z. Calculate,*

$$\frac{\partial w}{\partial x}\bigg|_z = \frac{\partial}{\partial x}\bigg|_z (x+y+z) = \frac{\partial x}{\partial x}\bigg|_z + \frac{\partial y}{\partial x}\bigg|_z + \frac{\partial z}{\partial x}\bigg|_z = 1 + \frac{\partial y}{\partial x}\bigg|_z + 0$$

To complete this calculation we need to eliminate w as before, using $\star\star$,

$$(1-z)dy = (1-z)dx - (2z+x+y)dz \quad \Rightarrow \quad \frac{\partial y}{\partial x}\bigg|_z = 1.$$

Therefore,

$$\boxed{\frac{\partial w}{\partial x}\bigg|_z = 2.}$$

I hope you can begin to see how the game is played. Basically the example above generalizes the idea of implicit differentiation to several equations of many variables. This is actually a pretty important type of calculation for engineering. The study of thermodynamics is full of variables which are intermittently used as either dependent or independent variables. The so-called equation of state can be given in terms of about a dozen distinct sets of state variables.

Example 4.7.11. *The ideal gas law states that for a fixed number of particles n the pressure P, volume V and temperature T are related by $PV = nRT$ where R is a constant. Calculate,*

$$\frac{\partial P}{\partial V}\bigg|_T = \frac{\partial}{\partial V}\left[\frac{nRT}{V}\right]\bigg|_T = -\frac{nRT}{V^2},$$

$$\frac{\partial V}{\partial T}\bigg|_P = \frac{\partial}{\partial T}\left[\frac{nRT}{P}\right]\bigg|_T = \frac{nR}{P},$$

$$\frac{\partial T}{\partial P}\bigg|_V = \frac{\partial}{\partial P}\left[\frac{PV}{nR}\right]\bigg|_T = \frac{V}{nR}.$$

You might expect that $\frac{\partial P}{\partial V}\big|_T \frac{\partial V}{\partial T}\big|_P \frac{\partial T}{\partial P}\big|_V = 1$. Is it true?

$$\frac{\partial P}{\partial V}\bigg|_T \frac{\partial V}{\partial T}\bigg|_P \frac{\partial T}{\partial P}\bigg|_V = -\frac{nRT}{V^2}\cdot\frac{nR}{P}\cdot\frac{V}{nR} = \frac{-nRT}{PV} = -1.$$

This is an example where naive cancellation of partials fails.

The example above is merely a special case of a general result shown below.

Example 4.7.12. *You can repeat the example above for x, y, z constrained by $F(x, y, z) = 0$. The differential of F is*

$$dF = F_x dx + F_y dy + F_z dz$$

Solve for dx, dy or dz to derivatives of $x = x(y, z)$, $y = y(x, z)$ or $z = z(x, y)$,

$$dx = -\frac{F_y}{F_x}dy - \frac{F_z}{F_x}dz \qquad \Rightarrow \qquad \frac{\partial x}{\partial y}\bigg|_z = -\frac{F_y}{F_x} \quad \& \quad \frac{\partial x}{\partial z}\bigg|_y = -\frac{F_z}{F_x}$$

$$dy = -\frac{F_x}{F_y}dx - \frac{F_x}{F_y}dz \qquad \Rightarrow \qquad \frac{\partial y}{\partial x}\bigg|_z = -\frac{F_x}{F_y} \quad \& \quad \frac{\partial y}{\partial z}\bigg|_x = -\frac{F_z}{F_y}$$

$$dz = -\frac{F_x}{F_z}dx - \frac{F_y}{F_z}dy \qquad \Rightarrow \qquad \frac{\partial z}{\partial x}\bigg|_y = -\frac{F_x}{F_z} \quad \& \quad \frac{\partial z}{\partial y}\bigg|_x = -\frac{F_y}{F_z}$$

Notice that the factors will cancel if we choose the right triple from the list above:

$$\frac{\partial x}{\partial y}\bigg|_z \frac{\partial y}{\partial z}\bigg|_x \frac{\partial z}{\partial x}\bigg|_y = -\frac{F_y}{F_x} \cdot \frac{F_z}{F_y} \cdot \frac{F_x}{F_z} = -1.$$

The identity above reliably holds if all the partial derivatives of F are nonzero. We need $F_x \neq 0$, $F_y \neq 0$ and $F_z \neq 0$. Incidentally, and not coincidentally, the implicit function theorem[18] *needs precisely these three conditions to solve for $x = x(y, z)$, $y = y(x, z)$ and $z = z(x, y)$ respective.*

Example 4.7.13. *Here's a different take on the example as above.*

$\boxed{E68}$ Show if $f(x,y,z) = 0$ then $\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1$. We begin by exploiting $f(x,y,z)=0$ to give a few differential relations,

$$\frac{\partial f}{\partial x} = 0 = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \quad \therefore \left(\frac{\partial z}{\partial x}\right)_y = \frac{-\partial f/\partial x}{\partial f/\partial z} \quad \left(\frac{\partial y}{\partial x} = 0\right)$$

$$\frac{\partial f}{\partial y} = 0 = \frac{\partial f}{\partial x}\frac{\partial x}{\partial y} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} \quad \therefore \left(\frac{\partial x}{\partial y}\right)_z = \frac{-\partial f/\partial y}{\partial f/\partial x} \quad \left(\frac{\partial z}{\partial y} = 0\right)$$

$$\frac{\partial f}{\partial z} = 0 = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} + \frac{\partial f}{\partial z} \quad \therefore \left(\frac{\partial y}{\partial z}\right)_x = \frac{-\partial f/\partial z}{\partial f/\partial y} \quad \left(\frac{\partial x}{\partial z} = 0\right)$$

Thus,

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = \left(\frac{-f_y}{f_x}\right)\left(\frac{-f_z}{f_y}\right)\left(\frac{-f_x}{f_z}\right) = -1.$$

---

[18] covered in advanced calculus

Example 4.7.14. . *21*

② SET-UP: Suppose $F(x,y,z) = 0$ implicitly defines $z = f(x,y)$, so that $F(x,y,f(x,y)) = 0$. Then

$$\frac{dF}{dx} = 0 = \frac{\partial F}{\partial x}\underbrace{\frac{\partial x}{\partial x}}_{1} + \frac{\partial F}{\partial y}\underbrace{\frac{\partial y}{\partial x}}_{0} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} \quad \therefore \boxed{\frac{\partial z}{\partial x} = \frac{-F_x}{F_z}}$$

Likewise we can derive that

$$\frac{dF}{dy} = 0 = \frac{\partial F}{\partial x}\underbrace{\frac{\partial x}{\partial y}}_{0} + \frac{\partial F}{\partial y}\underbrace{\frac{\partial y}{\partial y}}_{1} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} \quad \therefore \boxed{\frac{\partial z}{\partial y} = \frac{-F_y}{F_z}}$$

E62 Consider $x^2+y^2+z^2 = 1$ find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. We suppose that $z = z(x,y)$, define $F(x,y,z) = x^2+y^2+z^2-1$. Then,

$$\frac{\partial z}{\partial x} = \frac{-F_x}{F_z} = \frac{-2x}{2z} = \boxed{\frac{-x}{z} = \frac{\partial z}{\partial x}}$$

$$\frac{\partial z}{\partial y} = \frac{-F_y}{F_z} = \frac{-2y}{2z} = \boxed{\frac{-y}{z} = \frac{\partial z}{\partial y}}$$

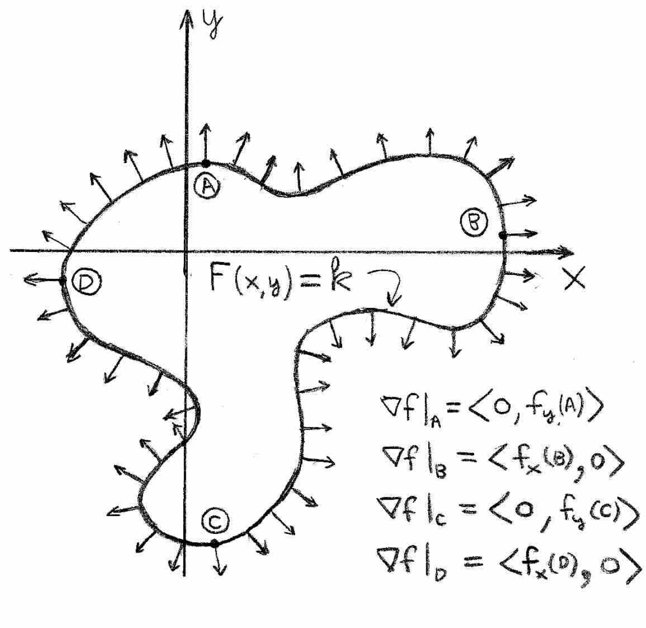
You might be curious above level curves or volumes given the interesting results above for the level surface $F(x,y,z) = 0$. Consider the curve case first. Here we recover the implicit differentiation of single-variable calculus.

Example 4.7.15. *Suppose $F(x,y) = 0$ then $dF = F_x dx + F_y dy = 0$ and it follows that $dx = -\frac{F_y}{F_x}dy$ or $dy = -\frac{F_x}{F_y}dx$. Hence, $\frac{\partial x}{\partial y} = -\frac{F_y}{F_x}$ and $\frac{\partial y}{\partial x} = -\frac{F_x}{F_y}$. Therefore,*

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial x} = \frac{F_y}{F_x} \cdot \frac{F_x}{F_y} = 1$$

for (x,y) such that $F_x \neq 0$ and $F_y \neq 0$. The condition $F_x \neq 0$ suggests we can solve for $y = y(x)$ whereas the condition $F_y \neq 0$ suggests we can solve for $x = x(y)$.

If you pause to think about the geometry of $F(x, y) = 0$ as it relates to $\nabla F = \langle F_x, F_y \rangle$ you can see why the conditions $F_x \neq 0$ and $F_y \neq 0$ are necessary.



There is no way to find y as single-valued function of x on a open set about the point where $F_y = 0$. Likewise, when $F_x = 0$ this means that there may be no way to find x as a single-valued function of y for a neighborhood centered at the point in question. If the point where $F_x = 0$ or $F_y = 0$ is on the edge of an interval then there is still hope, but the implicit function theorem does not apply. For example, $y - x^2 = 0$ for $x \geq 0$ can be solved for x as a function of y by $x = \sqrt{y}$. On the other hand, we cannot solve $y - x^2 = 0$ for x as a function of y in an open set centered about $x = 0$, each y value must return two x-values and that is not a function. Ok, enough about this.

Example 4.7.16. .

Given $F(x,y) = 0$ we can show that

$\frac{dy}{dx} = \frac{-\partial F/\partial x}{\partial F/\partial y}$. Let $F(x,y) = \sin(x)\cos(y) - \sin(x) - \cos(y)$.

Calculate then,

$$\frac{\partial F}{\partial x} = \cos(x)\cos(y) - \cos(x) \qquad \frac{\partial F}{\partial y} = -\sin(x)\sin(y) + \sin(y)$$

Therefore,

$$\frac{dy}{dx} = \frac{-\cos(x)\cos(y) + \cos(x)}{-\sin(x)\sin(y) + \sin(y)} = \boxed{\frac{\cos(x)}{\sin(y)}\left[\frac{1-\cos(y)}{1-\sin(y)}\right] = \frac{dy}{dx}}$$

Example 4.7.17. . *22*

E60 $\sin(x)\cos(y) + y^2 = x^3$: suppose $y = y(x)$ find $\frac{dy}{dx}$

$$\cos(x)\cos(y) - \sin(x)\sin(y)\frac{dy}{dx} + 2y\frac{dy}{dx} = 3x^2$$

$$\therefore \frac{dy}{dx} = \frac{3x^2 - \cos(x)\cos(y)}{2y - \sin(x)\sin(y)}.$$

We may arive at this result through another approach, perhaps easier.

1) Set-up : Suppose $F(x,y) = 0$ implicitly defines $y = f(x)$ such that $F(x, f(x)) = 0$. Now differentiate w.r.t. x, (here "t" $= x$)

$$\frac{dF}{dx} = \frac{d}{dx}(0) = 0 = \frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0$$

$$\therefore \boxed{\frac{dy}{dx} = \frac{-\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}}$$

E61 Lets revisit E60 to begin define $F(x,y) = x^3 - \sin(x)\cos(y) - y^2 = 0$,

$$\frac{dy}{dx} = \frac{-F_x}{F_y} = \frac{-(3x^2 - \cos(x)\cos(y))}{\sin(x)\sin(y) - 2y} = \frac{3x^2 - \cos(x)\cos(y)}{2y - \sin(x)\sin(y)}.$$

The solution set of $F(x,y,z,w) = 0$ gives a volume embedded in four-dimensional space. In invite the reader to demonstrate

$$\left.\frac{\partial x}{\partial y}\right|_{z,w}\left.\frac{\partial y}{\partial z}\right|_{x,w}\left.\frac{\partial z}{\partial w}\right|_{x,y}\left.\frac{\partial w}{\partial x}\right|_{y,z} = 1.$$

Again, this formula is only valid if all the partial derivatives of F are nontrivial at the point in question. In the next example we see why this identity holds in thermodynamics:

Example 4.7.18. . *19*

E65 Again suppose $PV = NRT$, but this time suppose that only R is constant so P, V, N, T are variables.

$$\left(\frac{\partial P}{\partial T}\right)_{V,N} = \frac{\partial}{\partial T}\left[\frac{NRT}{V}\right]\Bigg|_{V,N \text{ fixed}} = \frac{NR}{V}$$

$$\left(\frac{\partial T}{\partial V}\right)_{P,N} = \frac{\partial}{\partial V}\left[\frac{PV}{NR}\right]\Bigg|_{P,N \text{ fixed}} = \frac{P}{NR}$$

$$\left(\frac{\partial V}{\partial N}\right)_{P,T} = \frac{\partial}{\partial N}\left[\frac{NRT}{P}\right]\Bigg|_{P,T \text{ fixed}} = \frac{RT}{P}$$

$$\left(\frac{\partial N}{\partial P}\right)_{T,V} = \frac{\partial}{\partial P}\left[\frac{PV}{RT}\right]\Bigg|_{T,V \text{ fixed}} = \frac{V}{RT}$$

$$\left(\frac{\partial P}{\partial T}\right)_{V,N}\left(\frac{\partial T}{\partial V}\right)_{P,N}\left(\frac{\partial V}{\partial N}\right)_{P,T}\left(\frac{\partial N}{\partial P}\right)_{T,V} = \frac{\cancel{NR}}{\cancel{V}}\frac{\cancel{P}}{\cancel{NR}}\frac{\cancel{RT}}{\cancel{P}}\frac{\cancel{V}}{\cancel{RT}} = 1.$$

Example 4.7.19. . *18*

$\boxed{E66}$ Suppose $U = f(P,V,T)$ = internal energy of a gas that obeys the Ideal Gas Law $PV = nRT$ (n, R constants).

$$\left(\frac{\partial U}{\partial P}\right)_V = \frac{\partial}{\partial P}\Big[f(P,V,T)\Big]\Big|_{V\text{-fixed}} = \frac{\partial f}{\partial P}\frac{\partial P}{\partial P} + \frac{\partial f}{\partial T}\frac{\partial T}{\partial P} = \boxed{\frac{\partial f}{\partial P} + \frac{\partial f}{\partial T}\frac{V}{nR}}$$

$$\left(\frac{\partial U}{\partial T}\right)_V = \frac{\partial}{\partial T}\Big[f(P,V,T)\Big]\Big|_{V\text{-fixed}} = \frac{\partial f}{\partial P}\frac{\partial P}{\partial T} + \frac{\partial f}{\partial T}\frac{\partial T}{\partial T} = \boxed{\frac{\partial f}{\partial P}\left(\frac{nR}{V}\right) + \frac{\partial f}{\partial T}}$$

I've used $T = PV/nR$ and $P = nRT/V$ to calculate $\partial T/\partial P$ & $\partial P/\partial T$. And its better to write $\frac{\partial U}{\partial P} + \frac{\partial U}{\partial T}\frac{V}{nR} = \left(\frac{\partial U}{\partial P}\right)_V$ and $\frac{\partial U}{\partial P}\frac{nR}{V} + \frac{\partial U}{\partial T} = \left(\frac{\partial U}{\partial T}\right)_V$.

Example 4.7.20. . *20*

$\boxed{E64}$

Given $K = \frac{1}{2}mv^2$ show that $\frac{\partial K}{\partial m}\frac{\partial^2 K}{\partial v^2} = K$. To begin lets make this problem statement more precise, show

$$\boxed{\left(\frac{\partial K}{\partial m}\right)_v \left(\frac{\partial^2 K}{\partial v^2}\right)_m = K}$$

very well, lets begin

$$\left(\frac{\partial K}{\partial m}\right)_v = \frac{\partial}{\partial m}\left[\frac{1}{2}mv^2\right]\Big|_{v\text{-fixed}} = \frac{1}{2}v^2$$

$$\left(\frac{\partial K}{\partial v}\right)_m = \frac{\partial}{\partial v}\left[\frac{1}{2}mv^2\right]\Big|_{m\text{-fixed}} = mv \qquad (\text{momentum !})$$

$$\left(\frac{\partial^2 K}{\partial v^2}\right)_m = \frac{\partial}{\partial v}\Big[mv\Big]\Big|_{m\text{-fixed}} = m.$$

Therefore we find that $K = \frac{1}{2}mv^2 = \boxed{\left(\frac{\partial K}{\partial m}\right)_v \left(\frac{\partial^2 K}{\partial v^2}\right)_m = K}$.

Example 4.7.21. . *17*

$\boxed{E67}$ Suppose that $x^2 + y^2 = r^2$ and $x = r\cos\theta$, $y = r\sin\theta$,

$$\left(\frac{\partial x}{\partial r}\right)_\theta = \frac{\partial}{\partial r}\Big[r\cos\theta\Big]\Big|_{\theta\text{ fixed}} = \cos\theta$$

$$\left(\frac{\partial r}{\partial x}\right)_y = \frac{\partial}{\partial x}\Big[r\Big]\Big|_{y\text{-fixed}} = \frac{\partial}{\partial x}\Big[\sqrt{x^2+y^2}\Big]\Big|_{y\text{-fixed}} = \frac{x}{\sqrt{x^2+y^2}}$$

Just trying to eulicidate the notation.

Finally, for the unsatisfied reader I remind you once more that these calculations are justified by the implicit function theorem of advanced calculus. Here is a brief discussion of the simplest version of the theorem:

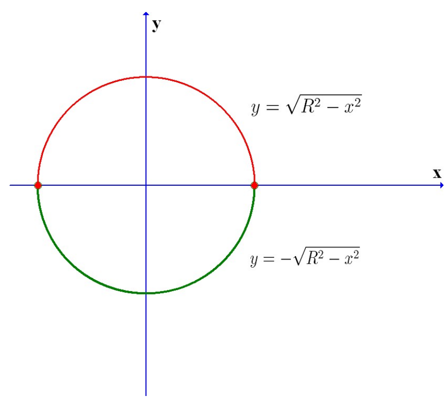
Theorem 4.7.22. *sometimes a level curve can be locally represented as the graph of a function.*

Suppose (x_o, y_o) is a point on the level curve $F(x, y) = k$ hence $F(x_o, y_o) = k$. We say the level curve $F(x, y) = k$ is **locally represented by a function** $y = f(x)$ at (x_o, y_o) iff $F(x, f(x)) = k$ for all $x \in B_\delta(x_o)$ for some $\delta > 0$. Claim: if

$$\frac{\partial F}{\partial y}(x_o, y_o) = \left(\frac{d}{dy}F(x_o, y)\right)\bigg|_{y=y_o} \neq 0$$

and the $\frac{\partial F}{\partial y}$ is continuous near (x_o, y_o) then $F(x, y) = k$ is locally represented by some function near (x_o, y_o).

The theorem above is called the **implicit function theorem** and its proof is nontrivial. Its proper statement is given in Advanced Calculus (Math 332). I'll just illustrate with the circle: $F(x, y) = x^2 + y^2 = R^2$ has $\frac{\partial F}{\partial y} = 2y$ which is continuous everywhere, however at $y = 0$ we have $\frac{\partial F}{\partial y} = 0$ which means the implicit function theorem might fail. On the circle, $y = 0$ when $x = \pm R$ which are precisely the points where we cannot write $y = f(x)$ for just one function. For any other point we may write either $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$ as a local solution of the level curve.



4.8 gradients in curvelinear coordinates

In this section we derive formulas for the gradient in polar, cylindrical and spherical coordinates. These formulas are important since many problems are more naturally phrased in polar, cylindrical or spherical coordinates.

4.8.1 polar coordinates

Our goal is to convert $\nabla f = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y}$ to polar coordinates. The unit-vectors for polar coordinates are given by

$$\begin{aligned} \widehat{r} &= \cos\theta\,\widehat{x} + \sin\theta\,\widehat{y} \\ \widehat{\theta} &= -\sin\theta\,\widehat{x} + \cos\theta\,\widehat{y}. \end{aligned} \tag{4.1}$$

We need to solve the equations above for $\widehat{x}$, $\widehat{y}$. I'll use multiplication by inverse:

$$\left[\begin{array}{c} \widehat{r} \\ \widehat{\theta} \end{array}\right] = \left[\begin{array}{cc} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{array}\right]\left[\begin{array}{c} \widehat{x} \\ \widehat{y} \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{c} \widehat{x} \\ \widehat{y} \end{array}\right] = \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right]\left[\begin{array}{c} \widehat{r} \\ \widehat{\theta} \end{array}\right] = \left[\begin{array}{c} \cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta} \\ \sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta} \end{array}\right]$$

Therefore, $\widehat{x} = \cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta}$ and $\widehat{y} = \sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta}$. Recall that we worked out in the chain rule section that $\frac{\partial}{\partial x} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}$ and $\frac{\partial}{\partial y} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}$. Let's put these together,

$$\begin{aligned} \nabla f &= \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} \\ &= \big(\cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta}\big)\left[\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right]f + \big(\sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta}\big)\left[\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}\right]f \\ &= \big(\cos^2\theta + \sin^2\theta\big)\widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\big(\cos^2\theta + \sin^2\theta\big)\frac{\partial f}{\partial \theta} \\ &= \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial \theta} \end{aligned}$$

Therefore, $\boxed{\nabla = \widehat{r}\frac{\partial}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial}{\partial \theta}}$ and $\boxed{\nabla f = \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial \theta}.}$

Example 4.8.1. *Suppose* $f(r,\theta) = r^3$ *then* $\nabla r^3 = \widehat{r}\frac{\partial r^3}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial r^3}{\partial \theta} = 3\widehat{r}\,r^2$.

The geometry of the function above is fairly clear in polar coordinates. If we did the same calculation in cartesians then you'd face the trouble of sorting through the derivatives of $f(x,y) = (x^2+y^2)^{3/2}$ paired with sorting out the radial pattern hidden in the $\widehat{x}$, $\widehat{y}$ notation.

Example 4.8.2. *Suppose* $f(r,\theta) = \theta$ *then* $\nabla\theta = \widehat{r}\frac{\partial\theta}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial\theta}{\partial\theta} = \frac{1}{r}\widehat{\theta}$.

4.8.2 cylindrical coordinates

There is not much to do here. We follow the same calculations as in the polar case with the slight modication of adjoining a z coordinate. It's not hard to see that we'll find $\nabla f = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}$ converts to

$$\boxed{\nabla f = \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial \theta} + \widehat{z}\frac{\partial f}{\partial z}.}$$

Example 4.8.3. *Suppose $f(r,\theta,z) = rz\theta$ then we calculate,*

$$\nabla f = \widehat{r}\frac{\partial (rz\theta)}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial (rz\theta)}{\partial \theta} + \widehat{z}\frac{\partial (rz\theta)}{\partial z} = z\theta\,\widehat{r} + z\widehat{\theta} + r\theta\,\widehat{z}.$$

4.8.3 spherical coordinates

We could derive the formula for ∇f in spherical coordinates in the same way as we did for polar and cylindrical coordinates. However, I take a different approach to illustrate a few calculation techniques. The basic observation is this: ∇f is a vector field and we can write it as a sum of the spherical unit-vector fields at each point in space;

$$\nabla f = (\nabla f \bullet \widehat{\rho})\,\widehat{\rho} + (\nabla f \bullet \widehat{\phi})\,\widehat{\phi} + (\nabla f \bullet \widehat{\theta})\,\widehat{\theta}$$

Hence the problem reduces to converting $\nabla f \bullet \widehat{\rho}$, $\nabla f \bullet \widehat{\phi}$ and $\nabla f \bullet \widehat{\theta}$ to spherical coordinates. Recall that unit vectors in the direction of increasing $\rho,\ \phi,\ \theta$ by $\widehat{\rho},\ \widehat{\phi},\ \widehat{\theta}$ are given by:

$$\begin{aligned} \widehat{\rho} &= \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\ \widehat{\phi} &= -\cos(\phi)\cos(\theta)\,\widehat{x} - \cos(\phi)\sin(\theta)\,\widehat{y} + \sin(\phi)\,\widehat{z} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}. \end{aligned} \tag{4.2}$$

We calculate: (remember $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$ and $z = \rho\cos\phi$ in order to understand the chain rule calculation below)

$$\begin{aligned} \nabla f \bullet \widehat{\rho} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(\sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z}\right) \\ &= \sin(\phi)\cos(\theta)\frac{\partial f}{\partial x} + \sin(\phi)\sin(\theta)\frac{\partial f}{\partial y} + \cos(\phi)\frac{\partial f}{\partial z} \\ &= \frac{\partial x}{\partial \rho}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial \rho}\frac{\partial f}{\partial y} + \frac{\partial z}{\partial \rho}\frac{\partial f}{\partial z} \\ &= \frac{\partial f}{\partial \rho} \end{aligned}$$

Continuing, calculate the ϕ-component of ∇f

$$\begin{aligned}
\nabla f \bullet \widehat{\phi} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(-\cos(\phi)\cos(\theta)\,\widehat{x} - \cos(\phi)\sin(\theta)\,\widehat{y} + \sin(\phi)\,\widehat{z}\right) \\
&= -\cos(\phi)\cos(\theta)\frac{\partial f}{\partial x} - \cos(\phi)\sin(\theta)\frac{\partial f}{\partial y} + \sin(\phi)\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho}\frac{\partial x}{\partial \phi}\frac{\partial f}{\partial x} + \frac{1}{\rho}\frac{\partial y}{\partial \phi}\frac{\partial f}{\partial y} + \frac{1}{\rho}\frac{\partial z}{\partial \phi}\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho}\frac{\partial f}{\partial \phi}
\end{aligned}$$

One more component to go:

$$\begin{aligned}
\nabla f \bullet \widehat{\theta} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(-\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}\right) \\
&= -\sin(\theta)\frac{\partial f}{\partial x} + \cos(\theta)\frac{\partial f}{\partial y} \\
&= \frac{-\rho\sin(\phi)\sin(\theta)}{\rho\sin(\phi)}\frac{\partial f}{\partial x}\frac{-\rho\sin(\phi)\cos(\theta)}{\rho\sin(\phi)}\frac{\partial f}{\partial y} \\
&= \frac{1}{\rho\sin(\phi)}\frac{\partial x}{\partial \theta}\frac{\partial f}{\partial x} + \frac{1}{\rho\sin(\phi)}\frac{\partial x}{\partial \theta}\frac{\partial f}{\partial x} + \frac{1}{\rho\sin(\phi)}\frac{\partial y}{\partial \theta}\frac{\partial f}{\partial y} + \frac{1}{\rho\sin(\phi)}\frac{\partial z}{\partial \theta}\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial \theta}.
\end{aligned}$$

Therefore, we find

$$\boxed{\nabla f \;=\; \widehat{\rho}\,\frac{\partial f}{\partial \rho} + \widehat{\phi}\,\frac{1}{\rho}\frac{\partial f}{\partial \phi} + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial \theta}.}$$

Spherical coordinate formulas are important for studying applications with spherical symmetry.

Example 4.8.4. *In spherical coordinates the potential due to a point charge is simply $V(\rho,\phi,\theta) = \frac{1}{\rho}$. The theory of electrostatics says this generates an electric field $\vec{E} = -\nabla V$. We find the field easily using our formula for the gradient in sphericals,*

$$\vec{E} = \nabla V = \widehat{\rho}\,\frac{\partial V}{\partial \rho} + \widehat{\phi}\,\frac{1}{\rho}\underbrace{\frac{\partial V}{\partial \phi}}_{zero} + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\underbrace{\frac{\partial V}{\partial \theta}}_{zero} = -\frac{1}{\rho^2}\widehat{\rho}.$$

More generally, the spherical gradient formula allows us to evaluate how a given function changes in spherical coordinates.

Example 4.8.5. *Suppose $f(x,y,z) = y/x$. To find how f changes in spherical coordinates we convert to sphericals[19]; $f(\rho,\phi,\theta) = \tan\theta$. It is clear that f is constant in ρ and ϕ. In particular,*

$$\nabla f = \widehat{\rho}\,\frac{\partial}{\partial \rho}\big[\tan\theta\big] + \widehat{\phi}\,\frac{1}{\rho}\frac{\partial}{\partial \phi}\big[\tan\theta\big] + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\frac{\partial}{\partial \theta}\big[\tan\theta\big] = \frac{\sec^2(\theta)}{\rho\sin(\phi)}\,\widehat{\theta}.$$

[19] this is a slight abuse of notation, the function is not really f with this modication. Instead, we should perhaps

This means that $\tan\theta$ *increases at the* $\frac{\sec^2(\theta)}{\rho\sin(\phi)}$ *rate in the* $\widehat{\theta}$*-direction.*

Remark 4.8.6.

There are other slicker methods to derive the formulas in this section. My goal here is not to be particularly clever. I merely wish to obtain these formulas for our future use and hopefully to illustrate once more the structure of vector algebra and the chain rules of multivariate calculus. If you'd like to know about alternate ways to derive these formulas I have a source or two for further reading.

denote it $\tilde{f}$ where to be technical $\tilde{f}(\rho,\phi,\theta) = f(\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$. The underlying motivation for this abuse is the idea that f is really an object which exists w/o regard to the particulars of the coordinate system we use, so it's appropriate to use the same letter for both the cartesian and spherical. Well, perhaps, but they are not the same actual function. This is similar to the problem of the sine function. $sin(90)$ and $sin(\pi/2)$ are usually both taken to be 1 but this is an overloading of the symbol sin. The degree-based sine function and the radian-based sine function are in fact different functions on $\mathbb{R}$.