

# NOTES ON VECTOR CALCULUS

(compiled on April 29, 2021 at 02:28)

## Contents

<b>1</b>	<b>Functions, Paths, and Derivatives in <math>\mathbb{R}^n</math></b>	<b>3</b>
1.1	Functions on $\mathbb{R}^n$	4
1.1.1	The Graph of a Function	5
1.1.2	Level Sets of a Function	7
1.2	Paths in $\mathbb{R}^n$	9
1.3	Compositions of Functions and Paths	16
1.4	Vectors	18
1.5	Velocities	20
1.5.1	Examples	22
1.5.2	Derivative of $f(\vec{p}(t))$	23
1.5.3	Vectors at a Point	24
1.6	Vector Fields	25
1.7	Partial Derivatives	28
1.7.1	Higher Partial	30
1.7.2	Smoothness	31
1.7.3	Multivariable Chain Rule	32
1.8	Domains	34
<b>2</b>	<b>Coordinate Systems and Spaces</b>	<b>35</b>
2.1	Coordinate Systems on $\mathbb{R}^n$	35
2.1.1	Some Useful Coordinate Systems on $\mathbb{R}^2$ and $\mathbb{R}^3$	37
2.1.2	Understanding Coordinate Systems	40

2.1.3	Re-expressing Functions and Paths in New Coordinates . . . . .	45
2.1.4	Coordinate Bases . . . . .	47
2.2	Spaces . . . . .	56
2.2.1	Examples of Coordinates on Abstract Spaces . . . . .	57
2.2.2	Smooth Structure . . . . .	63
<b>3</b>	<b>Derivatives</b>	<b>64</b>
3.1	Speed . . . . .	64
3.2	Norm and Dot Product for Vectors at a Point . . . . .	66
3.2.1	A Polar Example . . . . .	67
3.2.2	Dot Products in Non-Cartesian Coordinates . . . . .	69
3.2.3	Polar Example Picture . . . . .	69
3.3	The Gradient $\vec{\nabla}f$ . . . . .	70
3.3.1	Direction of $\vec{\nabla}f$ . . . . .	72
3.3.2	Directional Derivatives . . . . .	76
3.3.3	Norm of $\vec{\nabla}f$ . . . . .	78
3.3.4	$\vec{\nabla}f$ in Non-Cartesian Coordinates . . . . .	79
3.3.5	The Notation for Coordinate Basis Vectors . . . . .	86
3.4	The Derivative of a Function $\vec{F}(\vec{x})$ . . . . .	87
3.4.1	Defining the Derivative of $\vec{F}(\vec{x})$ . . . . .	88
3.4.2	Derivatives are Linear Approximations . . . . .	95
3.4.3	The Chain Rule . . . . .	98
3.4.4	The Jacobian Matrix of a Coordinate Transformation . . . . .	100
3.5	Optimization . . . . .	103
<b>4</b>	<b>Integrals</b>	<b>109</b>
4.1	Integrals of Functions over Regions . . . . .	110

4.2	Line Integrals and Conservative Vector Fields	123
4.3	Green's Theorem	132
<b>A</b>	<b>Appendix</b>	<b>136</b>
A.1	Parameterized Lines	136
A.2	Norm and Dot Product (Lengths and Angles)	136
A.2.1	The Dot Product	137
A.2.2	Algebra of Dot Products	139
A.2.3	Angles	139
A.2.4	Unit Vectors and Normalization	143
A.2.5	Projection: The Component of One Vector Along Another	143
A.3	Equations of Planes	144
A.4	Cross Products	146

# 1 Functions, Paths, and Derivatives in $\mathbb{R}^n$

In this section we will look at two generalizations of a function of one variable  $f(x)$ . One generalization allows for multiple inputs— a function of *multiple variables*  $f(x_1, x_2, \dots, x_n)$ . Another generalization allows for multiple outputs— a function with *multiple components*  $(x_1(t), x_2(t), \dots, x_n(t))$ .

$$\begin{array}{ccc}
 f(x, y, z) & \vec{p}(t) = & \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}
 \end{array}
 \begin{array}{l}
 \left. \begin{array}{c} \curvearrowright \\ \curvearrowleft \\ \curvearrowright \end{array} \right\} \text{multiple} \\
 \left. \begin{array}{c} \curvearrowleft \\ \curvearrowright \\ \curvearrowleft \end{array} \right\} \text{components} \\
 \text{multiple variables}
 \end{array}$$

The symbol  $\mathbb{R}^n$  refers to the set of points  $(x_1, x_2, \dots, x_n)$  in  $n$ -dimensional space, also called *euclidean space*. In the  $n = 2$  or  $n = 3$  situations we will often just use  $(x, y)$  or  $(x, y, z)$  to label points. When denoting an object which is actually a list of multiple components, we will use an arrow above the variable name to remind us of this fact. So for example if we were to give a single name to the point  $(x_1, x_2, \dots, x_n)$ , a good choice would be to just call it  $\vec{x}$ .

A function  $f$  of multiple variables could be thought of as actually being a function of one variable, but where that variable is a point in  $\mathbb{R}^n$  rather than simply being one number. Thinking this way, one would call  $f$  a *function on*  $\mathbb{R}^n$  and one would write  $f(\vec{x})$ . This is shorthand for  $f((x_1, x_2, \dots, x_n))$ , but one always drops the extra parentheses and simply writes  $f(x_1, x_2, \dots, x_n)$ . Similarly, a function  $\vec{p}(t)$  with multiple components could be thought of as actually having only one output, but that output is a point in  $\mathbb{R}^n$  rather than simply being one number. Thinking this way, one would call  $\vec{p}(t)$  a *path in*  $\mathbb{R}^n$ . Note the arrow in the notation  $\vec{p}(t)$  which reminds us that it is a function whose output has multiple components.

Let's look at how to visualize these gadgets, starting with functions. Later we will also look into how derivatives work for them.

## 1.1 Functions on $\mathbb{R}^n$

Whenever a numerical quantity depends on a bunch of other numerical quantities  $x_1, x_2, \dots, x_n$ , the relationship can be expressed by specifying a function  $f(x_1, x_2, \dots, x_n)$  of multiple variables. Some examples:

- $e(x, y)$  could be the elevation of a terrain above latitude-longitude coordinates  $x, y$ .
- $T(x, t)$  could be the temperature of a metal rod at a distance  $x$  along the rod, and at time  $t$ .
- $V(x, y, z)$  could be electric potential at a point  $(x, y, z)$  in three dimensional space.
- $\rho(x, y, z)$  could be the mass density of some object in kilograms per cubic meter, also as a function of three dimensional space.
- $T(x, y, z, t)$  could be the air temperature in a room where points in the room are described by  $(x, y, z)$  and where temperature can also vary with time  $t$ .
- $T(V, P)$  could be the temperature of a gas, as a function of its volume and pressure.
- $E(x, v)$  could be the energy of a particle traveling in one dimension, based on its position  $x$  and its velocity  $v$ .

For some of the examples the inputs of the function do not represent points in *physical* space. For example  $T(V, P)$  is a function of volume-pressure pairs  $(V, P)$ , which are not really points in the plane, right? Well, we can still think of them like they are— they are points of the “ $V$ - $P$ -plane” and that's just as good mathematically as the abstract “ $x$ - $y$ -plane.” From now on we will think of  $f(x_1, x_2, \dots, x_n)$  as a function on  $\mathbb{R}^n$  ( $n$ -dimensional euclidean space), even though the list of inputs  $(x_1, x_2, \dots, x_n)$  might not always represent a point in physical space. And don't forget that I might sometimes write  $f(\vec{x})$  to save having to list all the variables.

Here are some examples with specified definitions:

- $f(x_1, x_2, \dots, x_n) = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$  is the function that averages  $n$  numbers.

- $f(x, y) = x^2$  is the function that squares its first input.
- $f(x, y) = x + y$  is the function that adds its two inputs.
- $f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$  is the function that tells us the distance to the origin of a point in three dimensional space.
- $f(x, y, z) = x^3 e^{x+y} + \sin^5(yz)$  is some complicated function.

When visualizing a function of one variable like  $f(x) = x^2$ , we could just graph it and understand what  $f$  is doing from the parabola that we see. How do we visualize a function of *multiple* variables?

### 1.1.1 The Graph of a Function

Let's think about what we're really doing when we graph a function  $f(x)$  of one variable. We draw two axes, and we call the horizontal one the  $x$ -axis. Then, for each point along the  $x$ -axis, we plot a point at a height  $f(x)$  above it (or below if  $f(x)$  is negative). Really, we are just plotting the point  $(x, f(x))$  in the two dimensional plane. When we plot those for all possible inputs  $x$ , we see the points trace out some curve, like the parabola in the case of  $f(x) = x^2$ . The *graph* of  $f$  is essentially just the set of points of the form  $(x, f(x))$  in the plane. This extends easily to a function of multiple variables. In fact, let's make it an official definition:

#### Definition of Graph

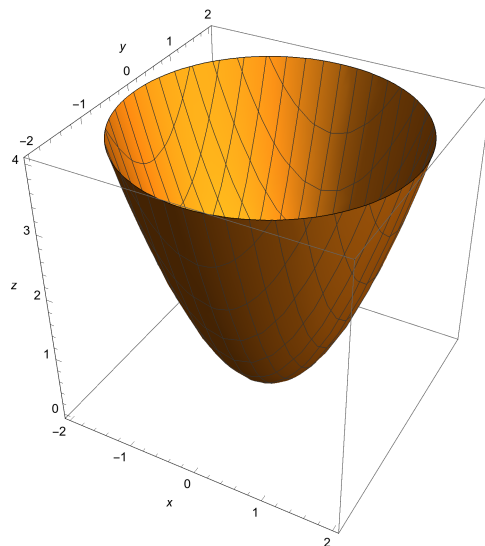
The *graph* of a function  $f(x_1, x_2, \dots, x_n)$  on  $\mathbb{R}^n$  is the set of points  $(x_1, x_2, \dots, x_n, f(x_1, x_2, \dots, x_n))$  in  $\mathbb{R}^{n+1}$ .

The first thing to notice is that the graph of a function of  $n$  variables is something that lives in  $(n + 1)$ -dimensional space.

As an example, let's look at the graph of the function

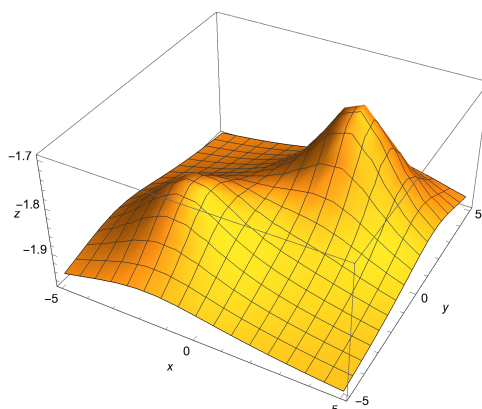
$$f(x, y) = x^2 + y^2.$$

It's a function of 2 variables, so its graph lives in 3-dimensional space. To plot it, imagine that the  $x$ - $y$ -plane is the space of inputs to the function, and then plot a point at a height  $z = f(x, y)$  above each point in the  $x$ - $y$ -plane. The points trace out a surface:



It looks kind of like a parabola, but with some kind of circular symmetry. This makes sense, considering that the expression  $x^2 + y^2$  is the square of the distance of  $(x, y)$  from the origin.

Here's the graph of  $f(x, y) = -e^{-\frac{1}{(x-2)^2+(y-2)^2+3}} - e^{-\frac{1}{(x+2)^2+(y+2)^2+5}}$ :



That picture would be hard to produce without using a computer to make plots. (I'm making these plots in Mathematica, which is freely available to you as a UCSB student.) Whenever I'm thinking of a function of two variables, I imagine some sort of hilly terrain like this, and the function is just reporting elevation above various points in the input plane.

What about the graph of a function  $f(x, y, z)$  of *three* variables? Well that would live in  $\mathbb{R}^4$ , so we have no hope of drawing that. If you tried to graph something like  $f(x, y, z)$  then you would need three axes for the input space,  $x$ ,  $y$ , and  $z$ , and then you would need a fourth direction in which to plot the functional value "above" each  $(x, y, z)$ . But no such spatial direction is available to our imagination. Regardless of how difficult it is to visualize, the graph of  $f$  still exists and has a simple mathematical description— it's just the set of points of the form  $(x, y, z, f(x, y, z))$ .

Still,  $f(x, y, z)$  is not such a crazy object to imagine. All  $f$  is doing is assigning a number to every point in a three dimensional input space. You could just imagine a three-dimensional space with a

little number hanging at every point. I like to imagine that where the numbers are big the space is “hot” and where they are small or negative the space is “cold”. Let’s try this on the function  $f(x, y, z) = x^2 + y^2 + z^2$ . It’s coldest at the origin  $(0, 0, 0)$ , where  $f$  takes the value 0. It gets hotter and hotter the further you go from the origin. Actually  $f$  is just giving the squared distance of a point from the origin. So if you were to start at the origin and travel in any direct at a constant speed, then you would feel temperature increase quadratically over time.

### 1.1.2 Level Sets of a Function

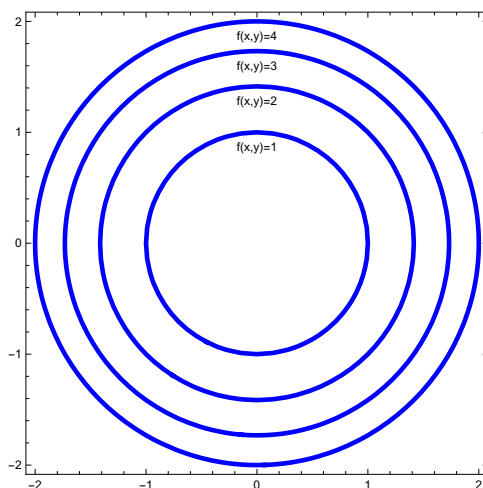
The level sets of a function are sets of inputs that all yield the same output. More precisely,

#### Definition of Level Set

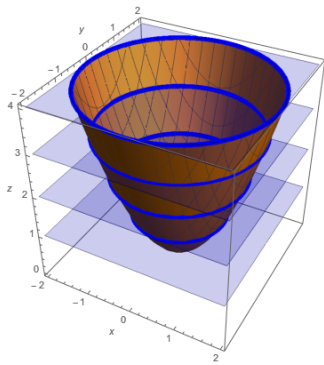
Assume you have a function  $f(\vec{x})$  on  $\mathbb{R}^n$ . For any fixed real number  $k$ , the set of points  $\vec{x}$  satisfying  $f(\vec{x}) = k$  is called a *level set* of  $f$ .

There are many level sets for a given function, one for each choice of  $k$ .

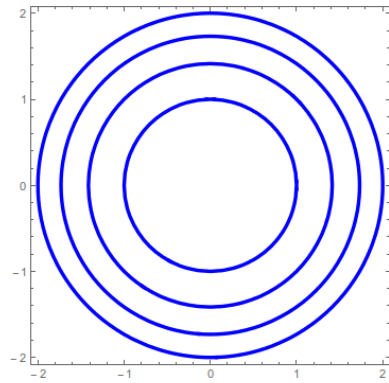
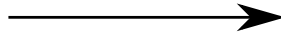
Let’s see what the level sets are for the function  $f(x, y) = x^2 + y^2$ , whose graph we looked at in the previous section. Each level set is the set of solutions to  $x^2 + y^2 = k$  for a choice of  $k$ . When  $k$  is chosen to be negative, there are no solutions and the level set we get is just the empty set. When  $k$  is 0, the only solution is  $x = y = 0$ , and so our level set contains just one point,  $(0, 0)$ . When  $k$  is positive, the solutions to  $x^2 + y^2 = k$  trace out a circle of radius  $\sqrt{k}$  centered at the origin. Here’s a picture of the level sets for  $k = 1, 2, 3, \dots$ :



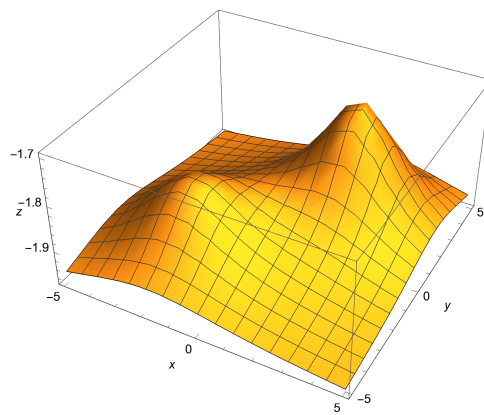
Since we can visualize the graph of  $f(x, y) = x^2 + y^2$ , we can actually use that to pick out the level sets. They are just the cross-sections of the graph at various heights:



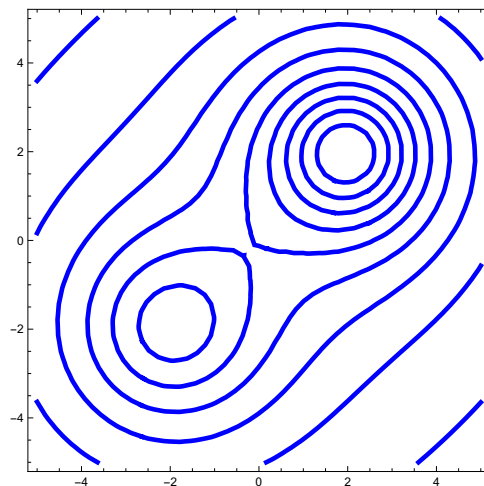
Slice the graph of  $f$  at various constant heights  $k$  to get level sets



Remember the plot of  $f(x, y) = -e^{-\frac{1}{(x-2)^2+(y-2)^2+3}} - e^{-\frac{1}{(x+2)^2+(y+2)^2+5}}$  from the last section?



Here's what its level sets look like:

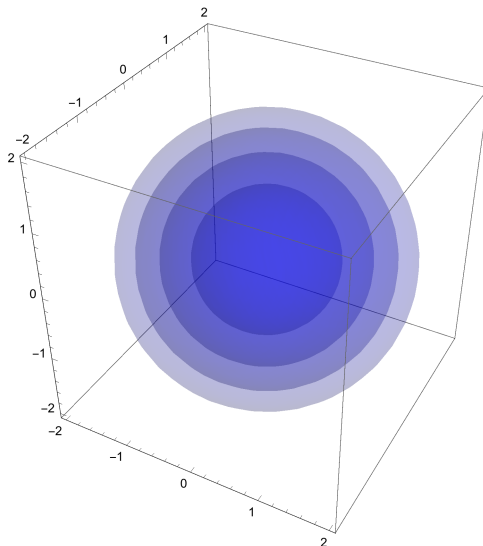


See how the two pictures relate? You can still pick out hills and valleys from the level sets picture, even though it's shown in only two dimensions. If you've ever used a topographic map to understand a terrain— that's basically what this is.



One advantage of using level sets to visualize functions is that you only need to picture things in  $n$ -dimensional space in order to visualize the level sets of a function of  $n$  variables (as opposed to the graph, for which you need to picture things in  $(n + 1)$ -dimensional space). In particular, we can use level sets to visualize a function of *three* variables.

Let's try getting the level sets for  $f(x, y, z) = x^2 + y^2 + z^2$ . When  $k$  is negative, the equation  $x^2 + y^2 + z^2 = k$  has no solution and so we just get the empty set. When  $k$  is 0 we get only the origin  $(0, 0, 0)$ . When  $k$  is positive, the equation  $x^2 + y^2 + z^2 = k$  describes a sphere of radius  $\sqrt{k}$  centered at the origin. So the level sets of  $f(x, y, z) = x^2 + y^2 + z^2$  are concentric spheres of various radii centered at the origin.



## 1.2 Paths in $\mathbb{R}^n$

Suppose that  $\vec{p}(t)$  is a function of one variable with multiple outputs. We can label the outputs by writing

$$\vec{p}(t) = (p_1(t), p_2(t), \dots, p_n(t)),$$

where  $p_1, p_2, \dots, p_n$  are ordinary functions (each having one numerical output). Actually, since we have been referring to points in  $\mathbb{R}^n$  as “ $(x_1, x_2, \dots, x_n)$ ”, let's refer to the component functions using the names of the coordinates:

$$\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t)).$$

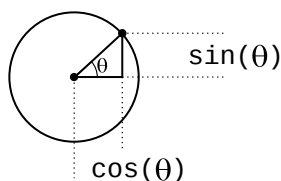
For each choice of input  $t$ , we get a point  $\vec{p}(t)$  in  $\mathbb{R}^n$ . If we think of  $t$  as “time” and we let  $t$  run over all possible values, then we can watch the points  $\vec{p}(t)$  trace out some *curve* in  $\mathbb{R}^n$ . Thus a function of one variable with  $n$  outputs can be thought of as a *path* in  $\mathbb{R}^n$ .

The ordinary functions  $x_1(t), x_2(t), \dots, x_n(t)$  are called the *component functions* of  $\vec{p}$ . Sometimes you will hear people call  $\vec{p}(t)$  a “vector-valued” function—this is okay but I will explain later why I am avoiding such terminology.

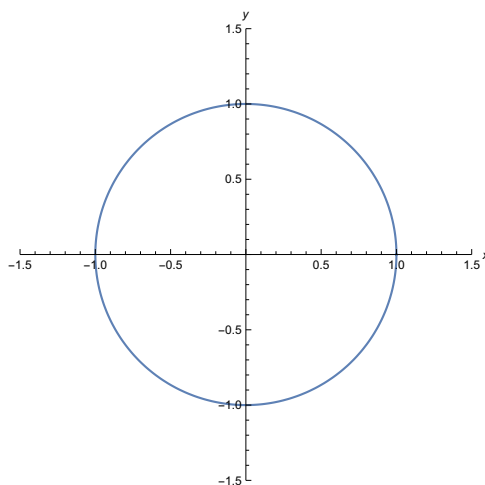
Let's look at an example. Consider the path

$$\vec{p}(t) = (\cos(t), \sin(t)).$$

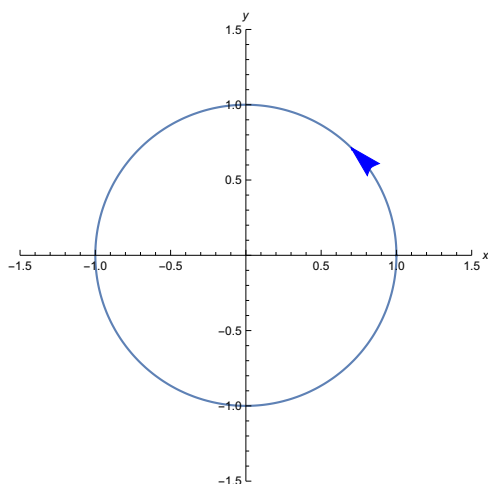
This is a path in  $\mathbb{R}^2$ . To visualize this path, we should remind ourselves of the meanings of  $\sin$  and  $\cos$  in terms of the unit circle: Given a point on the unit circle at an angle of  $\theta$  from the  $x$ -axis in the counter-clockwise direction, the  $x$ -coordinate of the point is  $\cos(\theta)$  and the  $y$ -coordinate is  $\sin(\theta)$ .



Ah, so  $(\cos(t), \sin(t))$  simply gives us the point on the unit circle that makes an angle  $\theta = t$  from the  $x$ -axis in the counter-clockwise direction. Thinking of  $t$  as time and letting it run over all possible values, we watch as  $\vec{p}(t)$  traces out all the points on the unit circle.

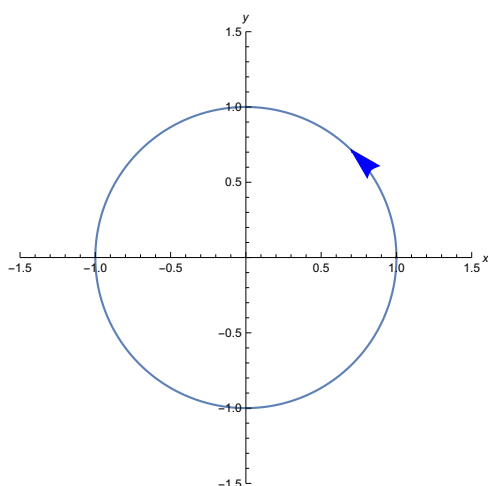


Some things are missing from this picture. If we really let  $t$  run over all real numbers, then as  $t$  steadily increases the path  $\vec{p}(t)$  will continue to repeat its trek around the unit circle forever. That information is not contained in the picture. Another thing that's missing from the picture: When you *increase*  $t$ , the point  $\vec{p}(t)$  runs around the circle in *counter-clockwise* direction. We could include this information in the picture by adding a little arrow:

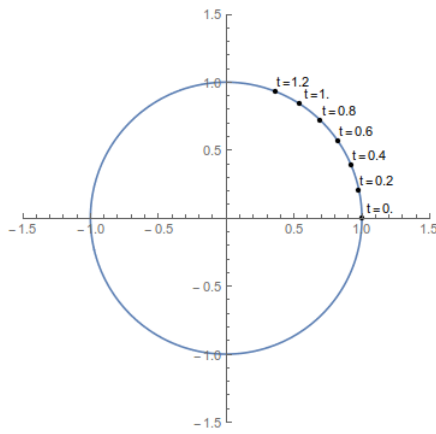


We call this the *orientation* of the path  $\vec{p}$ .

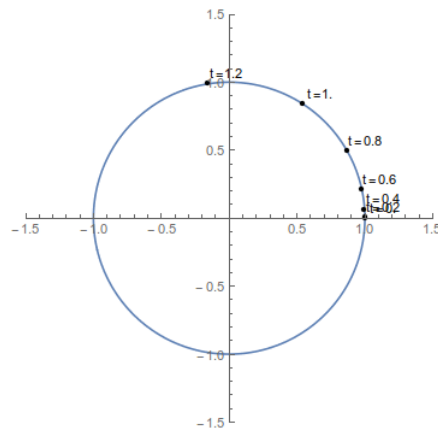
There's yet another bit information missing from the picture. Lets try looking at the path  $\vec{q}(t) = (\cos(t^3), \sin(t^3))$ . All we did was replace  $t$  by  $t^3$ . As  $t$  runs over all possible real numbers, so does  $t^3$ . And so  $\vec{q}(t) = (\cos(t^3), \sin(t^3))$  still runs around the unit circle:



The picture is exactly the same as before! The issue is that we are only plotting the *outputs* of  $\vec{p}$  and  $\vec{q}$ , while ignoring *which* inputs actually map to those outputs. In other words, the picture is displaying the *locations* that each path encounters, but it is not showing us *when* (i.e. for what  $t$ ) they encounter each location. We could somewhat remedy this by including some labels showing us where  $\vec{p}(t)$  is for a few values of  $t$ :



$$\begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}$$



$$\begin{bmatrix} \cos(t^3) \\ \sin(t^3) \end{bmatrix}$$

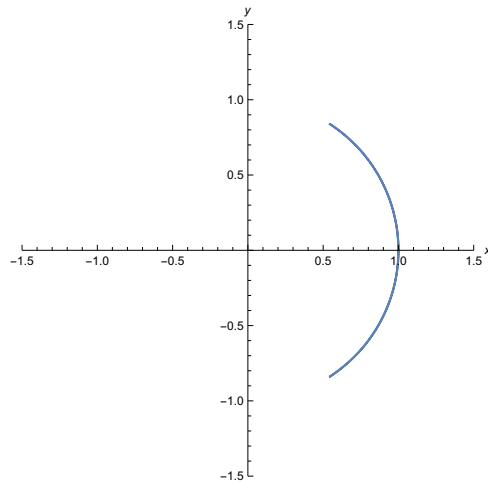
Some terminology to make it easier to talk about this stuff in the future:

### Some Terminology

Given a path  $\vec{p}(t)$ , the variable  $t$  is called the *parameter* of the path. The set of points that are outputs of the function  $\vec{p}$  is called the *image* of the path. The path  $\vec{p}$  itself is in turn called a *parameterization* of that image. If  $t$  is replaced by some function of  $t$  (like in the example we just considered where  $t$  was replaced by  $t^3$ ) such that the image of the path doesn't change, then we call the new path a *reparameterization* of the old one.

Often when reparameterizing a path, we try to avoid confusion by using a new name for the new parameter. So in the above example, we would define a new parameter  $s$  by the relation  $s^3 = t$ , and then we would describe the reparameterization  $\vec{q}$  of  $\vec{p}$  by saying that  $\vec{q}(s) = \vec{p}(s^3) = (\cos(s^3), \sin(s^3))$ .

One small warning about reparameterization: It doesn't always work to just replace  $t$  by any old function of  $t$ . For example, if we start with our circular path  $\vec{p}(t) = (\cos(t), \sin(t))$  and we try to replace  $t$  by  $\sin(t)$ , then the new path  $(\cos(\sin(t)), \sin(\sin(t)))$  looks like this:

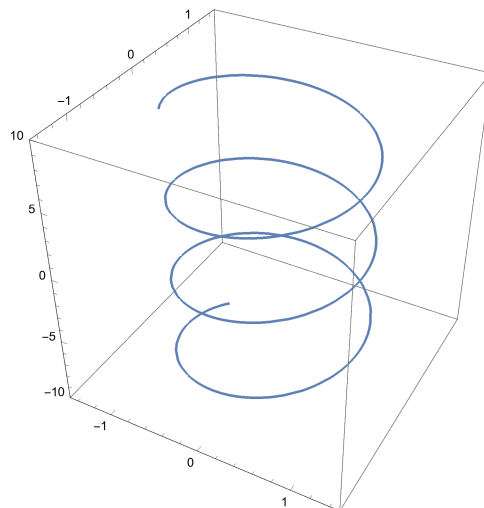


That's truly a different image. In this case we wouldn't call the new path a reparameterization of the old, for a reparameterization should change only the parameter and not the image of the path. The issue is that as  $t$  runs over all real numbers, the function  $\sin(t)$  is only allowed to run over values between  $-1$  and  $1$ . In order for a reparameterization to be legitimate, each value of the old parameter must correspond to some unique value of the new parameter.

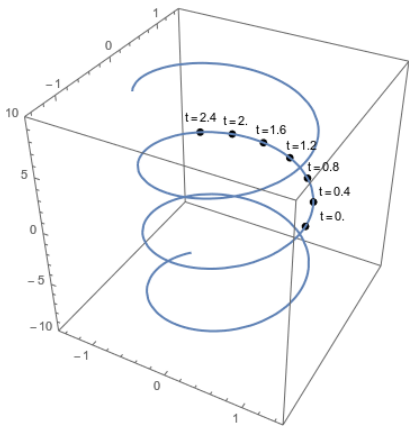
Let's look at a couple of other examples. Consider the path

$$\vec{p}(t) = (\cos(t), \sin(t), t).$$

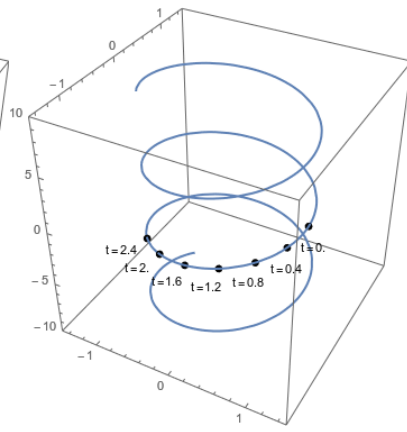
This is a path in  $\mathbb{R}^3$ , which we will think of as  $x$ - $y$ - $z$ -space. If one only considers what the  $x$  and  $y$  coordinates of the path are doing, then it's the same as what we just considered: the path runs around a unit circle in the counter-clockwise direction as  $t$  increases. However, the path now also has a steadily increasing  $z$ -coordinate as  $t$ -increases. This traces out a helix:



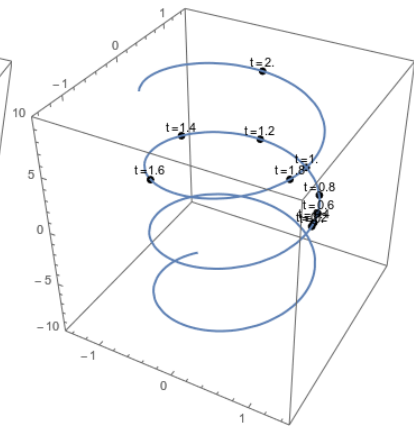
Just for fun, here are two additional parameterizations of the same thing:



$p(t)$



$p(-t)$

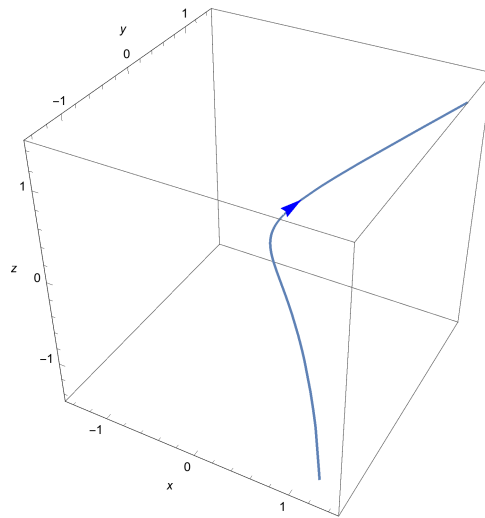


$p(t^3)$

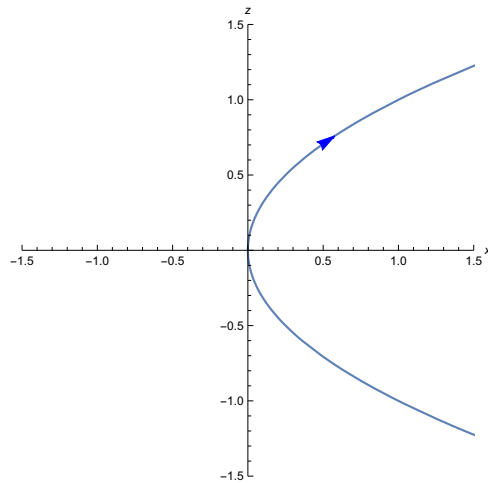
Now consider the path

$$\vec{p}(t) = (t^2, t^3, t).$$

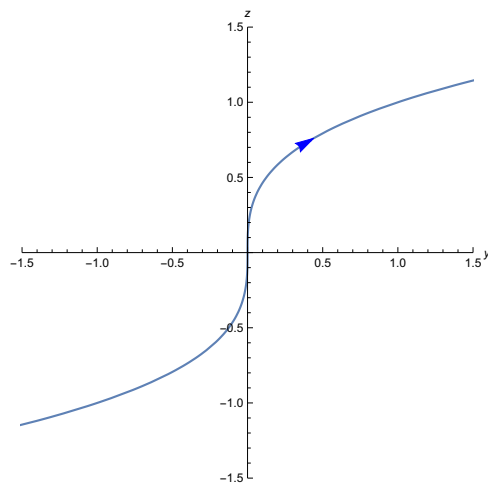
Here's a picture:



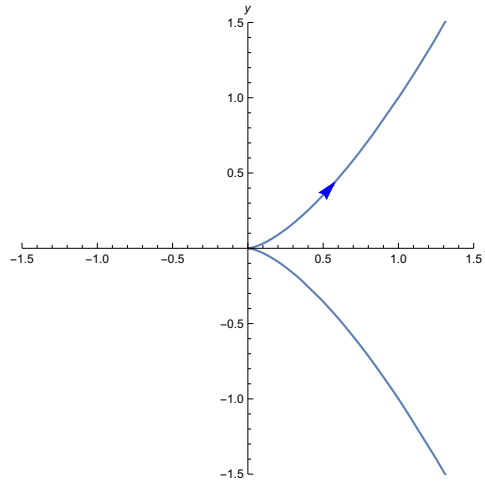
It's a little hard to see what's going on. If you look directly at the  $x$ - $z$  plane, then you see the path given by  $x = t^2$  and  $z = t$ , ignoring  $y$ :



This picture makes sense— it’s showing us that the points along the path all satisfy  $x = z^2$  and therefore trace out a parabola. What we’re looking at, the path  $(t^2, t)$ , is called the *projection* of  $\vec{p}(t)$  to the  $x$ - $z$  plane. The projection of  $\vec{p}(t)$  to the  $y$ - $z$  plane shows a path given by  $y = t^3$  and  $z = t$ :



This also makes sense— it’s showing us that the points along the path all satisfy  $y = z^3$  and therefore trace out a cubic. The projection of  $\vec{p}(t)$  to the  $x$ - $y$  plane shows a path given by  $x = t^2$  and  $y = t^3$ :



Now what shape are we seeing? Well  $y = t^3$  can be solved uniquely to tell us that a point with a  $y$ -coordinate of  $y$  is encountered at the unique parameter value  $t = y^{1/3}$ . The  $x$ -coordinate of such a point is then  $x = t^2 = y^{2/3}$ . So the picture is showing us points in the  $x$ - $y$  plane that satisfy  $x = y^{2/3}$ . Maybe a nicer way to put it is that it's showing us points that satisfy  $x^3 = y^2$ .

### 1.3 Compositions of Functions and Paths

A function on  $\mathbb{R}^n$  is a function that takes  $n$  inputs and gives one output:

$$(x_1, x_2, \dots, x_n) \mapsto f(x_1, x_2, \dots, x_n).$$

A path in  $\mathbb{R}^n$  is a function that takes one input and yields  $n$  outputs:

$$t \mapsto \vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t)).$$

These two gadgets plug into each other quite neatly:

$$t \mapsto \vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t)) \mapsto f(\vec{p}(t)) = f(x_1(t), x_2(t), \dots, x_n(t)).$$

The resulting function  $f(\vec{p}(t))$  is a function of only *one* variable which gives only *one* output. It is called a *composite* of the functions  $f$  and  $\vec{p}$ . I'll say it again: it's a function of only *one* variable which gives only *one* output! That means we can graph it and visualize it easily, and we can even do ordinary *single-variable* calculus on it (which we soon will do).

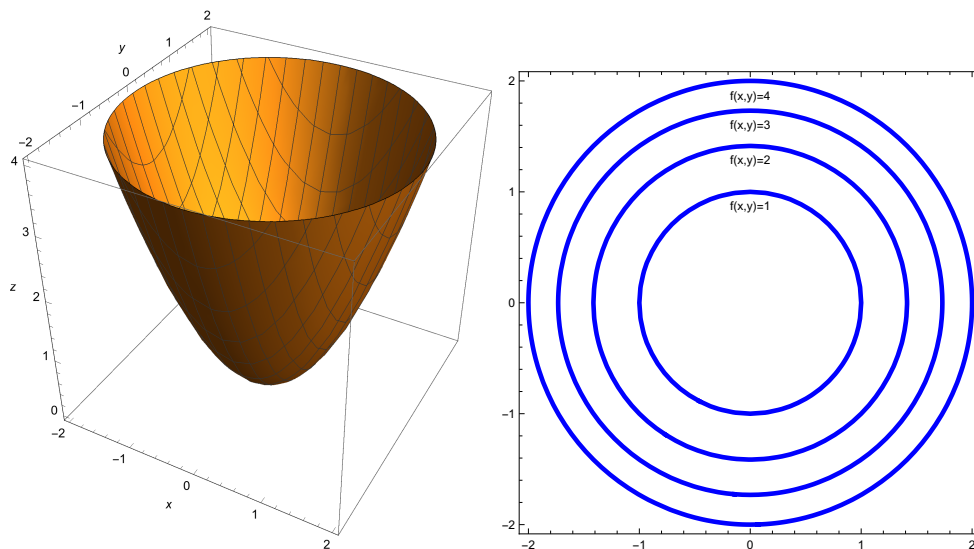
A few examples attaching some possible physical meanings to this idea:

- Suppose  $e(x, y)$  is the elevation of a terrain above latitude-longitude coordinates  $x, y$ . You are hiking in this terrain and you follow a path given by  $(x(t), y(t))$ , where  $t$  is time. What is your elevation as a function of time? The answer is the composite function  $e(x(t), y(t))$ .
- Suppose  $T(x, y, z)$  is the air temperature in a room where points in the room are described by  $(x, y, z)$ . A fly is flying around the room and following a path given by  $(x(t), y(t), z(t))$ , where  $t$  is time. What temperature does the fly feel as a function of time? The answer is the composite function  $T(x(t), y(t), z(t))$ .

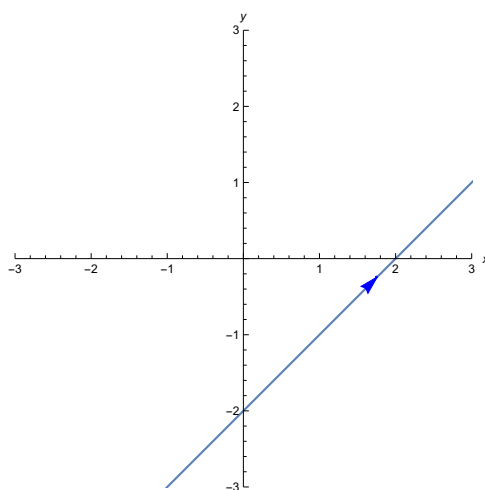


- Suppose  $T(V, P)$  is the temperature of a gas, as a function of its volume and pressure. You know that the volume of the gas at a time  $t$  is  $V(t)$  and the pressure of the gas at time  $t$  is  $P(t)$ . What is the temperature of the gas at time  $T$ ? The answer is the composite function  $T(V(t), P(t))$ . Here we composed the function  $T(V, P)$  with the path in the  $V$ - $P$ -plane given by  $(V(t), P(t))$ .

For a more specific example, let's revisit the function  $f(x, y) = x^2 + y^2$ , whose graph and level sets looked like this:



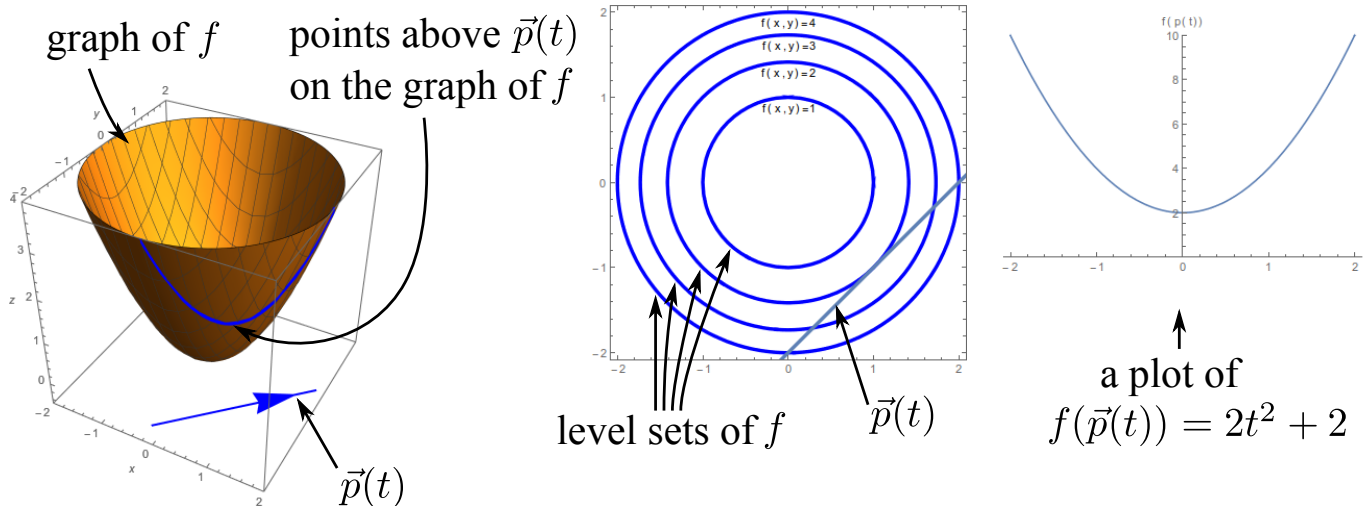
Consider the path  $\vec{p}(t) = (t + 1, t - 1)$ , whose image looks like this:



The composite function is

$$\begin{aligned}
 f(\vec{p}(t)) &= f(t + 1, t - 1) \\
 &= (t + 1)^2 + (t - 1)^2 \\
 &= 2t^2 + 2.
 \end{aligned}$$

What we are seeing is a function that gives *the values of  $f$  along the path  $\vec{p}$* :



## 1.4 Vectors

What is a *vector*?

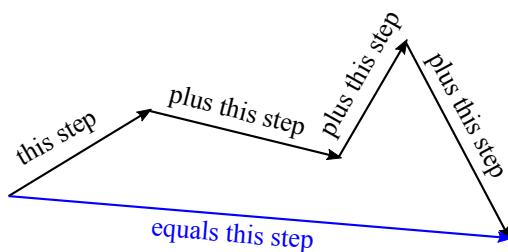
“A list of numbers,” says the computer scientist.

“A geometric object with a magnitude and a direction,” says the physicist.

“An element of a vector space,” says the mathematician.

There is some truth in each answer. Different viewpoints are appropriate in different situations. Of course, as usual, it is the mathematician’s answer that unifies all answers as one concept. However, even in this math course we will find it useful to specialize our use of the word *vector* to a particular viewpoint. Let’s start with a review of the general mathematician’s view. Later, in section 1.5.3, we will describe how vectors are thought of in the specific setting of vector calculus.

A *vector space* is a set of things that can be thought of as steps in space, things that can be added to each other



and that can be scaled by numbers (which we then call *scalars*).

$$2 \left( \begin{array}{c} \nearrow \\ \end{array} \right) = \begin{array}{c} \nearrow \\ \end{array}$$

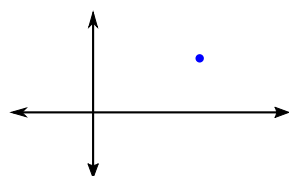
$$-1 \left( \begin{array}{c} \nearrow \\ \end{array} \right) = \begin{array}{c} \nwarrow \\ \end{array}$$

That is the abstract viewpoint, but as you probably learned in linear algebra (e.g. math 4A at UCSB), every finite dimensional vector space looks pretty much like  $\mathbb{R}^n$  for some  $n$ — assuming we take addition and scaling to operate as follows:

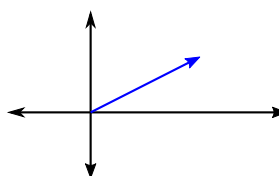
$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} = \begin{bmatrix} x_1 + x'_1 \\ x_2 + x'_2 \\ \vdots \\ x_n + x'_n \end{bmatrix},$$

$$\alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}.$$

When treating  $\mathbb{R}^n$  as a vector space, we no longer think of its elements as merely being points; we instead think of them as representing a *step* from the origin out to some point. So how we think about the object  $(2, 1)$ , for example, depends on context:



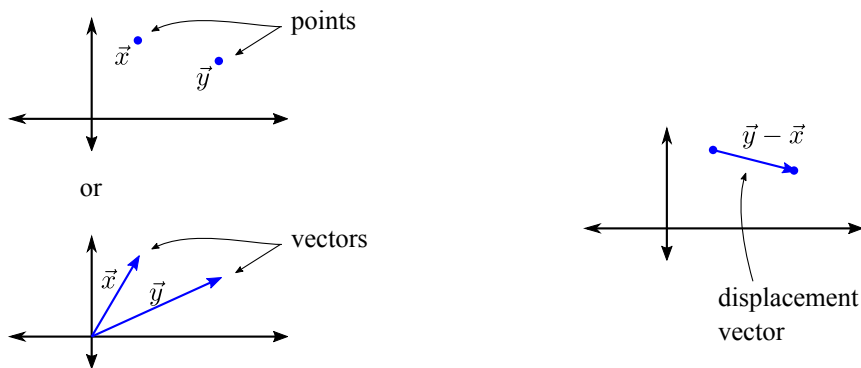
$(2, 1)$  as a *point*  
 $\mathbb{R}^2$  as a 2D (geometric) space



$(2, 1)$  as a *vector*  
 $\mathbb{R}^2$  as a 2D vector space

Now it should be more clear why I wanted to avoid the terminology “vector-valued function” to describe  $\vec{p}(t)$  in section 1.2. While a path  $\vec{p}(t)$  does have outputs that have multiple components, they’re usually thought of as *points* rather than vectors. So I’d rather call a path a “point-valued function.”

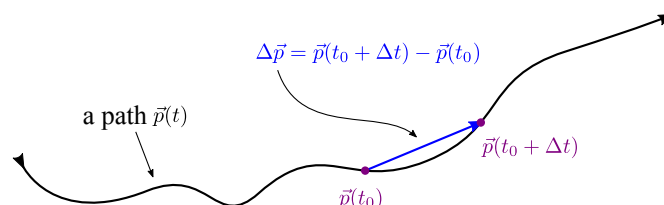
One particularly useful vector to form is the *difference*  $\vec{y} - \vec{x}$ . Even if we happened to be thinking of  $\vec{x}$  and  $\vec{y}$  as just *points* in  $\mathbb{R}^n$ , the vector difference  $\vec{y} - \vec{x}$  is still a meaningful *vector*: it is the step that displaces you from  $\vec{x}$  to  $\vec{y}$ .



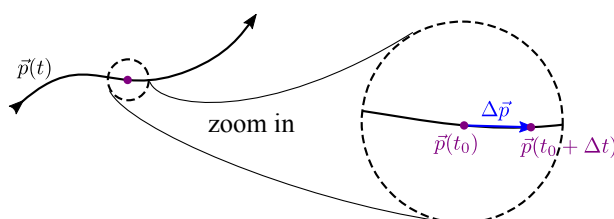
If you're not very familiar with the definition of the *derivative* of a function in single-variable calculus, then now would be a good time to pause and review that. The definition  $f'(t) = \lim_{\Delta t \rightarrow 0} \frac{f(t+\Delta t) - f(t)}{\Delta t}$  should be okay with you before moving to the next section.

## 1.5 Velocities

Consider a path  $\vec{p}(t)$  in  $\mathbb{R}^n$ . It will really help in this discussion to think of  $t$  as *time*. Let's focus on a particular  $t = t_0$  and some later  $t = t_0 + \Delta t$ . The vector difference  $\Delta \vec{p} = \vec{p}(t_0 + \Delta t) - \vec{p}(t_0)$  is a vector that represents the step from  $\vec{p}(t_0)$  to  $\vec{p}(t_0 + \Delta t)$ .



Now imagine that  $\Delta t$  is very small, so small that the path looks straight.



The vector  $\Delta \vec{p}$  points pretty much in the direction that the path  $\vec{p}(t)$  is traveling when  $t = t_0$ , and its length is the distance traversed during the time interval  $[t_0, t_0 + \Delta t]$ . Scaling by  $\frac{1}{\Delta t}$  keeps the direction the same, but makes the length be the average speed during the time interval  $[t_0, t_0 + \Delta t]$ . In the limit as  $\Delta t \rightarrow 0$ , the vector  $\frac{1}{\Delta t} \Delta \vec{p}$  takes on the following properties: (1) It points *exactly* in the direction that  $\vec{p}(t)$  is traveling at  $t = t_0$ , and (2) its length becomes the *instantaneous* speed of  $\vec{p}(t)$  at  $t = t_0$ . Except, certain things could go wrong:

Issue 1: The limit might not exist at all.

Issue 2: The limit might be  $\vec{0}$ , in which case it has no direction at all.

Issue 1 can occur when a path jumps around or turns a sharp corner. Issue 2 can occur when a path stops in one place. If the limit

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0))$$

exists for all  $t_0$ , then we say the path  $\vec{p}(t)$  is *differentiable*. This limit is how we define the *velocity* of a path.

### Definition of Velocity Vector

Let  $\vec{p}(t)$  be a path in  $\mathbb{R}^n$  and let  $t_0$  be a parameter at which it is defined. If the limit

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0))$$

exists, then it is called the *derivative* or *velocity* of  $\vec{p}(t)$  at  $t = t_0$ , and it is denoted by  $\vec{p}'(t_0)$  or  $\left. \frac{d\vec{p}}{dt} \right|_{t=t_0}$ .

We will avoid issue 1 above by assuming all our paths are differentiable. Issue 2 may not seem like a big deal, but it's actually pretty annoying. We will want velocity vectors to always provide us with the *direction* in which a path  $\vec{p}(t)$  is traveling. We will avoid issue 2 by assuming it away as well. Here are the assumptions:

### Assumptions

Whenever I say  $\vec{p}(t)$  is a *path*, assume that I'm talking about a *differentiable* path for which the derivative is never  $\vec{0}$ .

Further similar assumptions may come later, depending on what we need to develop the theory.

The limit involved in the definition of velocity is a limit of the vector-valued function  $\frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0))$  of  $\Delta t$ . It turns out that because of the way limits work in  $\mathbb{R}^n$ , the limit can be carried out on each component function individually. Naming the component functions by  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ , this enables the following calculation:

$$\begin{aligned} \vec{p}'(t_0) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0)) \\ &= \lim_{\Delta t \rightarrow 0} \begin{bmatrix} \frac{1}{\Delta t} (x_1(t_0 + \Delta t) - x_1(t_0)) \\ \frac{1}{\Delta t} (x_2(t_0 + \Delta t) - x_2(t_0)) \\ \vdots \\ \frac{1}{\Delta t} (x_n(t_0 + \Delta t) - x_n(t_0)) \end{bmatrix} = \begin{bmatrix} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (x_1(t_0 + \Delta t) - x_1(t_0)) \\ \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (x_2(t_0 + \Delta t) - x_2(t_0)) \\ \vdots \\ \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (x_n(t_0 + \Delta t) - x_n(t_0)) \end{bmatrix} = \begin{bmatrix} x'_1(t_0) \\ x'_2(t_0) \\ \vdots \\ x'_n(t_0) \end{bmatrix}. \end{aligned}$$

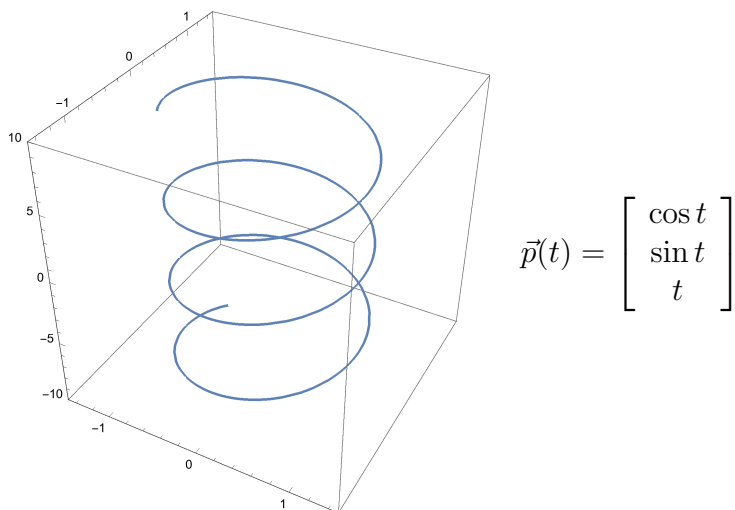
So the velocity of a path can be obtained by simply differentiating all the component functions! Worth boxing in:

## Computing the Derivative of a Path

Let  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  be a path in  $\mathbb{R}^n$ . Then its derivative is given by  $\vec{p}'(t) = (x_1'(t), x_2'(t), \dots, x_n'(t))$ .

### 1.5.1 Examples

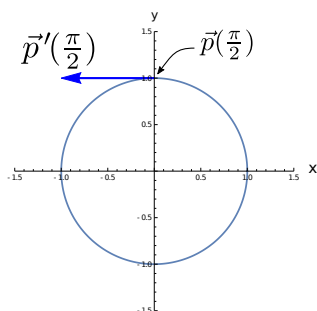
Let's take the helix from earlier as an example.



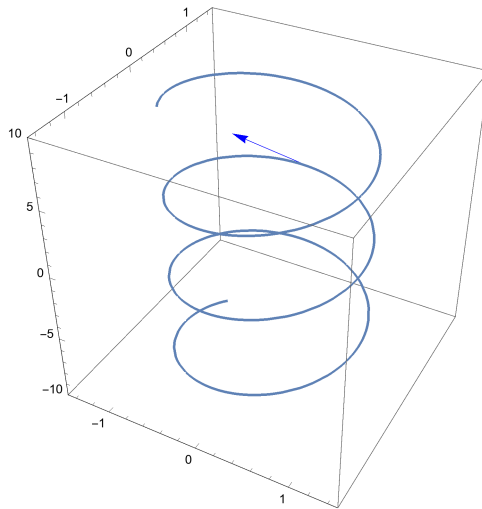
The velocity is given by

$$\vec{p}'(t) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix}.$$

So for example the velocity at  $t = \frac{\pi}{2}$  is  $(-1, 0, 1)$ . The  $-1$  and  $0$  for the first two components make sense—looking down at the helix from above (projecting to the  $x$ - $y$ -plane) shows



and indeed we see that  $\vec{p}'(t)$  is heading directly in the negative  $x$ -direction. Now consider what's happening in the  $z$  direction—the path is steadily rising and this fact is reflected by the constant  $z$ -component of  $1$  in the velocity vector. Here's the full 3D picture with the velocity vector shown at  $t = \frac{\pi}{2}$ :



What happens if we consider the same helix with a different parameterization? Let  $\vec{q}(t) = (\cos(2t), \sin(2t), 2t)$ . Now the velocity is

$$\vec{q}'(t) = \begin{bmatrix} -2 \sin(2t) \\ 2 \cos(2t) \\ 2 \end{bmatrix} = 2 \begin{bmatrix} -\sin(2t) \\ \cos(2t) \\ 1 \end{bmatrix} = 2\vec{p}'(2t).$$

So replacing the  $t$  by  $2t$  caused the velocity vectors to all get doubled. Replacing  $t$  by  $2t$  in  $\vec{p}(t)$  means we are traversing the same helix except *twice as fast* (thinking of  $t$  as time). So it makes sense that the velocity vectors got doubled. Actually, we could have seen this directly from the chain rule, which works as you would expect for paths:

$$\frac{d}{ds} \vec{p}(t(s)) = \frac{dt}{ds} \frac{d\vec{p}}{dt}.$$

Our computation would then have been

$$\frac{d}{dt} \vec{p}(2t) = \frac{d}{dt}(2t) \frac{d\vec{p}}{dt} = 2 \frac{d\vec{p}}{dt}.$$

We can see from the chain rule that reparameterization always has the effect of simply rescaling velocity vectors.

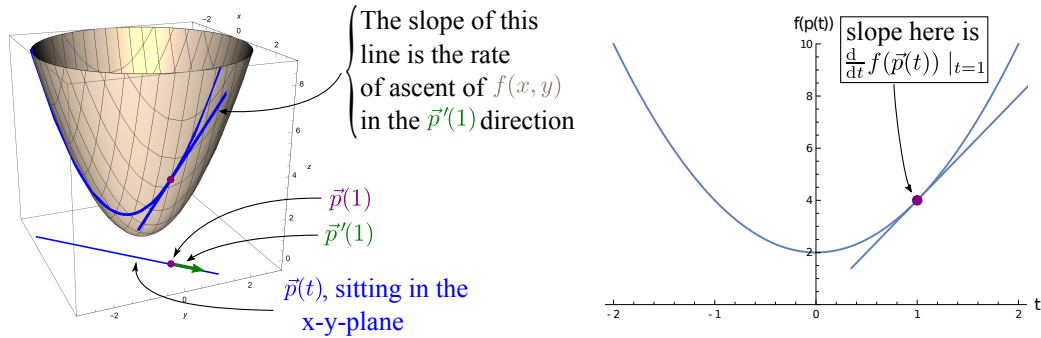
### 1.5.2 Derivative of $f(\vec{p}(t))$

Recall the composite function we talked about in section 1.3: given a function  $f(\vec{x})$  on  $\mathbb{R}^n$  and a path  $\vec{p}(t)$  in  $\mathbb{R}^n$ , the function  $f(\vec{p}(t))$  is a function of just one variable with just one output. The function  $f(\vec{p}(t))$  tracks the values of  $f$  along the path  $\vec{p}$ . The derivative  $\frac{d}{dt} f(\vec{p}(t))$  then tells us about the *rate of change of  $f$  along the path  $\vec{p}(t)$* .

Let's focus on a specific  $t = t_0$ . Let  $\vec{x}_0 = \vec{p}(t_0)$ . Then  $\frac{d}{dt} f(\vec{p}(t)) \big|_{t=t_0}$  tells us about the rate of change of  $f$  that would be observed if one focused on values of  $f$  that are near  $\vec{x}_0$  and along the direction in which  $\vec{p}(t)$  travels through  $\vec{x}_0$ . That direction is dictated by the vector  $\vec{p}'(t_0)$ . So really  $\frac{d}{dt} f(\vec{p}(t))$

is telling us something about the rate of change of  $f(\vec{x})$  at  $\vec{x}_0$  in the direction of  $\vec{p}'(t_0)$ . There will be more on this idea later in the course, but I just want to prime you to start thinking about it.

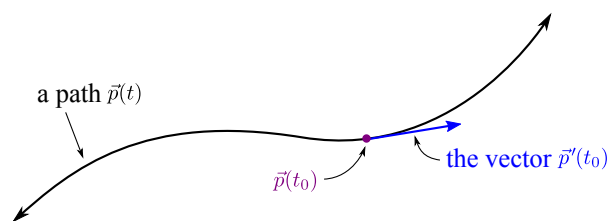
Here's an example using the same setup from page 17: Let  $f(x, y) = x^2 + y^2$  and let  $\vec{p}(t) = (t+1, t-1)$ . Let's focus on  $t = 1$ , just to have a specific point to look at. Now  $\vec{p}'(t) = (1, 1)$ , so  $\vec{p}'(1) = (1, 1)$ . Recall from page 17 that  $f(\vec{p}(t)) = 2t^2 + 2$ . So  $\frac{d}{dt}f(\vec{p}(t)) = 4t$  and  $\frac{d}{dt}f(\vec{p}(t))|_{t=1} = 4$ . This number is telling us about the rate of change of  $f(x, y)$  if one considers values of  $f(x, y)$  near  $\vec{p}(1) = (2, 0)$  and along the path  $\vec{p}(t)$ , so in the direction  $\vec{p}'(1) = (1, 1)$ .



Let me say all that again with a physical example in mind: If  $f(x, y)$  represents the elevation of a terrain over longitude and latitude coordinates  $(x, y)$ , and if  $\vec{p}(t)$  represents the path you take through the terrain on your hike, then  $\frac{d}{dt}f(\vec{p}(t))|_{t=1} = 4$  is your rate of ascent per unit time at time  $t = 1$ . This rate of ascent is telling you something about how steep the terrain (the graph of  $f(x, y)$ ) is in the direction you happen to be going at  $t = 1$  (which is the direction  $\vec{p}'(1) = (1, 1)$ ).

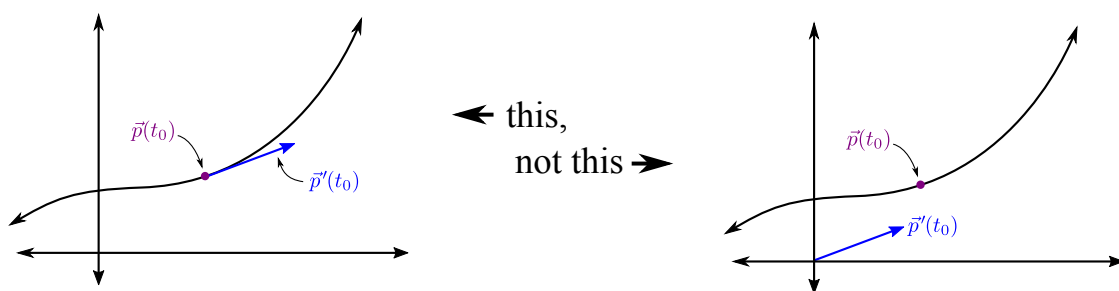
### 1.5.3 Vectors at a Point

Look at how we depict the velocity vector of a path:

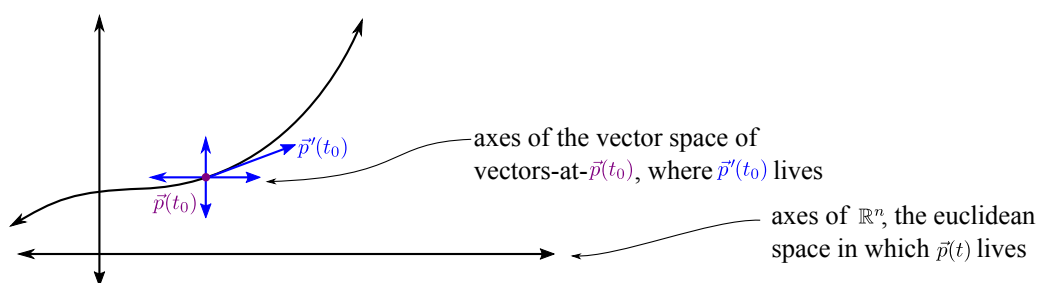


You may have noticed that the arrow for the velocity vector  $\vec{p}'(t_0)$  is depicted a little bit differently from how we were depicting vectors in section 1.4, where all vectors had their tail-end at the origin. This is because section 1.4 was a pure linear algebra setting where a single vector space was being considered. In our current setting, however, our vector  $\vec{p}'(t_0)$  is obtained by looking at the behavior of  $\vec{p}(t)$  near  $\vec{p}(t_0)$ . The vector  $\vec{p}'(t_0)$  therefore rightly sits at the point  $\vec{p}(t_0)$ , and not at the origin of the euclidean space in which the path  $\vec{p}(t)$  lives.





What's really going on is that there are many, many vector spaces in the same picture! Let me use the term *vector-at- $\vec{x}$*  to refer to any vector that is a possible velocity for a path passing through a specific point  $\vec{x}$  in  $\mathbb{R}^n$ . At each point  $\vec{x}$  in  $\mathbb{R}^n$ , there's an entire  $n$ -dimensional vector space (an entire copy of  $\mathbb{R}^n$ ) consisting of vectors-at- $\vec{x}$ . In particular,  $\vec{p}'(t_0)$  is a vector-at- $\vec{p}(t_0)$ . You can think of its tail end as being at the origin of the vector space of vectors-at- $\vec{p}(t_0)$ .



I will admit that this is a bit confusing. Understand the following for now, and you will eventually see how this way of thinking works:

### Vectors at a Point

Every point  $\vec{x}$  in euclidean space has a vector space attached to it, namely the space of vectors-at- $\vec{x}$ . If  $\vec{p}(t)$  is a path then the velocity  $\vec{p}'(t_0)$  is a vector-at- $\vec{p}(t_0)$ .

## 1.6 Vector Fields

Now that we've talked about vectors at a point, we can define *vector fields*:

### Definition of Vector Field

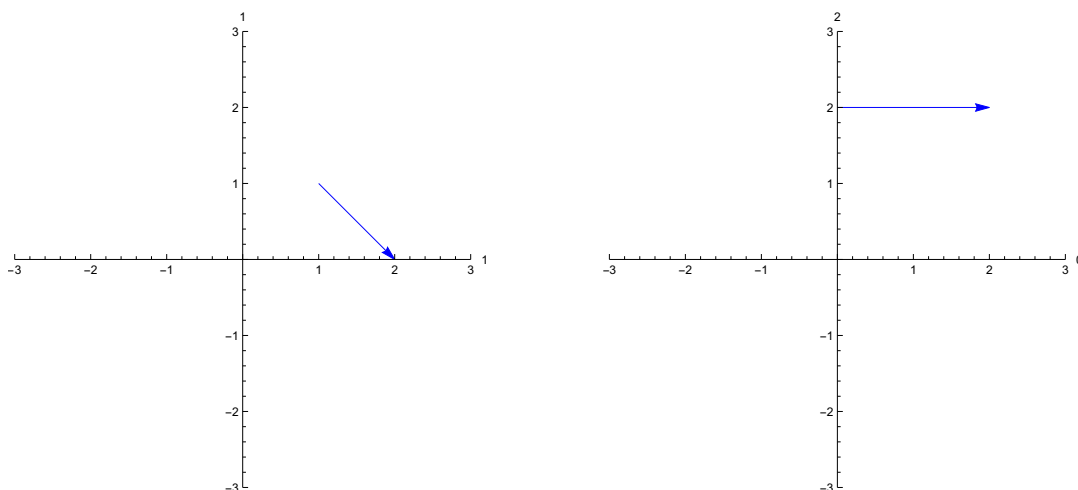
A *vector field* on  $\mathbb{R}^n$  is a function  $\vec{F}(\vec{x})$  which takes as an input a point in  $\mathbb{R}^n$  and yields as an output a vector at that point.

We will also introduce some kind of differentiability assumptions into the definition later, but we're not ready to state those yet. As suggested by the notation  $\vec{F}(\vec{x})$ , a vector field is a function of *multiple* variables with *multiple* components. And the number of components *equals* the number

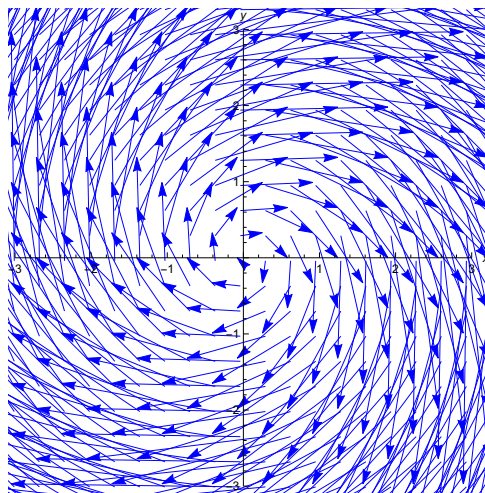
of variables, because  $\vec{F}(\vec{x})$  is always supposed to be a vector-at- $\vec{x}$ , which is an object that has  $n$  components if  $\vec{x}$  is a point in  $\mathbb{R}^n$ .

Pay attention to the difference in how we think about what goes into  $\vec{F}(\vec{x})$  and what comes out of it: *points* of  $\mathbb{R}^n$  go in, *vectors* (at those points) come out. (This is not the only way to think of a function of multiple variables with multiple components. One could also think of it as taking in points and outputting *points*. But if we view the function that way then we would not call it a vector field.)

Here's an example:  $\vec{F}(x, y) = (y, -x)$ . It assigns to every point in  $\mathbb{R}^2$  a vector at that point. So for example  $(1, 1)$  gets assigned  $(1, -1)$ ,  $(0, 2)$  gets assigned  $(2, 0)$ , etc.

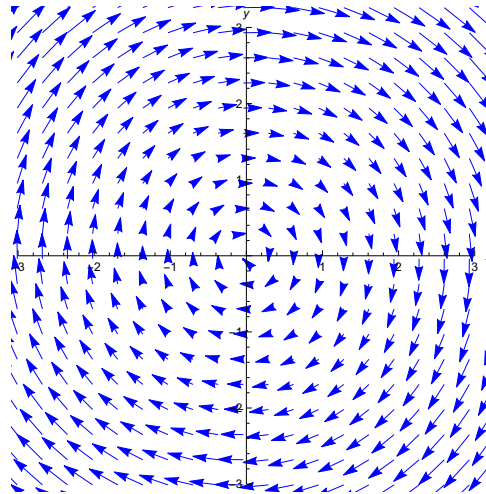


Let's plot the vectors that  $\vec{F}(\vec{x})$  attaches to a whole bunch of points all in the same picture:

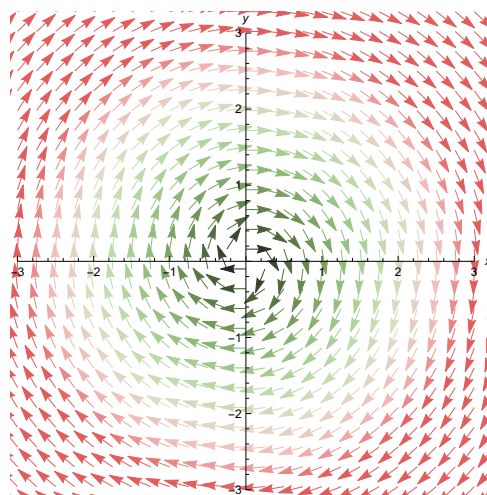


Obviously we cannot display the value of  $\vec{F}(\vec{x})$  at *all* the points shown; that would just fill the diagram with blue ink. But you can get the general idea of what  $\vec{F}(\vec{x})$  is doing from the discrete grid of vectors shown.

The collisions of arrows at the outer edges of the plot is pretty annoying. The collisions are not “real,” they are just a graphical artifact having to do with how we chose to display magnitudes of vectors. We could remedy the situation by scaling down how we display all the vectors



or by making them all have the same length and using some other feature such as color to represent the relative sizes of vectors:



(redder vectors are bigger and greener vectors are smaller)

Here are some things a vector field might be used to represent:

- A velocity field for a fluid. One could model a fluid as a vector field  $\vec{v}(\vec{x})$  where  $\vec{v}(\vec{x})$  is the velocity of bits of fluid at the location  $\vec{x}$ .
- A force field (like gravitational or electric force). One could have a function  $\vec{F}(\vec{x})$  that gives the force imparted on a particle located at  $\vec{x}$ . But something is off with this example. Remember how I defined the term vector-at- $\vec{x}$ ? A vector-at- $\vec{x}$  is a vector that arises as the velocity of some path that passes through  $\vec{x}$ . That doesn't sound at all like what a force vector is! In truth, the force field function  $\vec{F}(\vec{x})$  is not a vector field, but something else. We usually cheat and treat a force field like a vector field anyway, and that is usually okay. If you ever decide to

do physics seriously, then at some point it would be good to understand what kind of object a force field really is.

- A system of first order ordinary differential equations (ODEs). This example will make sense to you if you've had a course on the topic (such as math 4B at UCSB). Say you are trying to determine some functions  $x_1(t), x_2(t), \dots, x_n(t)$  according to some law that they are supposed to obey. These functions could be anything that evolves in time—populations of various species in an ecosystem, prices of various goods in a marketplace, position and momentum coordinates of some particles, etc. If you state the law that governs their time-evolution as a system of first order ODEs, then it would look like

$$\begin{aligned}x_1'(t) &= f_1(x_1(t), x_2(t), \dots, x_n(t)) \\x_2'(t) &= f_2(x_1(t), x_2(t), \dots, x_n(t)) \\&\vdots \\x_n'(t) &= f_n(x_1(t), x_2(t), \dots, x_n(t))\end{aligned}$$

for some functions  $f_1, f_2, \dots, f_n$  on  $\mathbb{R}^n$ . Or, using our notation for multiple components to abbreviate this, it's just

$$\vec{x}'(t) = \vec{f}(\vec{x}(t)).$$

The function  $\vec{f}(\vec{x})$  that appears here is a vector field on  $\mathbb{R}^n$ . Really, a system of first order ODEs in  $n$  unknown functions is exactly the same thing as a vector field on  $\mathbb{R}^n$ !

## 1.7 Partial Derivatives

How do we take a derivative of a function of multiple variables? We saw one option in section 1.5.2: compose it with a path and then apply an ordinary derivative to the resulting function of one variable. Now we will see another option: *partial derivatives* (which will soon turn out to be a special case of the first option).

### Definition of Partial Derivative

Let  $f(x_1, x_2, \dots, x_n)$  be a function on  $\mathbb{R}^n$ . The *partial derivative of  $f$  with respect to  $x_1$* , denoted by  $\frac{\partial f}{\partial x_1}$ , is the function on  $\mathbb{R}^n$  whose value at  $(x_1, x_2, \dots, x_n)$  is

$$\lim_{\Delta x_1 \rightarrow 0} \frac{f(x_1 + \Delta x_1, x_2, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{\Delta x_1}.$$

In other words, it is the function obtained by taking the derivative of  $f$  with respect to  $x_1$ , while holding constant all the other variables  $x_2, \dots, x_n$ . The other partial derivatives  $\frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}$  are defined similarly.

Let's use the function  $f(x, y, z) = 4xyz + 3x^2z + y^2$  on  $\mathbb{R}^3$  to demonstrate:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 4yz + 6xz \\ \frac{\partial f}{\partial y} &= 4xz + 2y \\ \frac{\partial f}{\partial z} &= 4xy + 3x^2.\end{aligned}$$

Pretty simple. Just follow the usual derivative rules, pretending that everything is a constant other than the variable with respect to which you are taking a derivative.

What is the meaning of the partial derivative? The derivative of a function of one variable  $f(x)$  is supposed to tell you how the output of the function responds when you vary the input. If you vary the input  $x$  by a small amount  $\Delta x$ , then the derivative tells you how the output changes:

$$f(x_0 + \Delta x) \approx f(x_0) + \frac{df}{dx}(x_0) \Delta x.$$

It's the same deal with partial derivatives. The *partial* derivative with respect to  $x$  of the function  $f(x, y, z)$  on  $\mathbb{R}^3$  is telling us how the output of the function responds to variations in  $x$  *while the other variables are held fixed*:

$$f(x_0 + \Delta x, y_0, z_0) \approx f(x_0, y_0, z_0) + \frac{\partial f}{\partial x}(x_0, y_0, z_0) \Delta x.$$

Similarly,

$$f(x_0, y_0 + \Delta y, z_0) \approx f(x_0, y_0, z_0) + \frac{\partial f}{\partial y}(x_0, y_0, z_0) \Delta y,$$

and so on for  $\frac{\partial f}{\partial z}$  and so on for functions on  $\mathbb{R}^n$  in general. If  $f(x_1, x_2, \dots, x_n)$  is a function on  $\mathbb{R}^n$  then the partial derivative  $\frac{\partial f}{\partial x_i}$  tells us about the rate of change of  $f$  *in the direction of the  $x_i$  coordinate* in  $\mathbb{R}^n$ .

Partial derivatives are actually a special case of what we saw in section 1.5.2. Let me demonstrate with a function  $f(x, y, z)$  on  $\mathbb{R}^3$ . Suppose we are interested in the value of the partial derivative  $\frac{\partial f}{\partial x}$  at a specific point  $(x_0, y_0, z_0)$ . Take the path  $\vec{p}(t) = (x_0 + t, y_0, z_0)$ , which can alternatively be written as  $\vec{p}(t) = (x_0, y_0, z_0) + t(1, 0, 0)$ . This is a path that starts at  $(x_0, y_0, z_0)$  and travels with velocity vector  $(1, 0, 0)$  (straight in the  $x$  direction with unit speed). Now consider the derivative of the composite  $f(\vec{p}(t))$ :

$$\begin{aligned}\frac{d}{dt}f(\vec{p}(t)) \Big|_{t=0} &= \frac{d}{dt}f(x_0 + t, y_0, z_0) \Big|_{t=0} \\ &= \lim_{\Delta t \rightarrow 0} \frac{f(x_0 + \Delta t, y_0, z_0) - f(x_0, y_0, z_0)}{\Delta t} \\ &= \frac{\partial f}{\partial x}(x_0, y_0, z_0).\end{aligned}$$

What we are seeing is that the partial derivative of  $f$  with respect to  $x$  is just the rate of change of  $f$  *along a certain path*, namely the path that heads off straight in the  $x$  direction.

### 1.7.1 Higher Partial

For a function of one variable  $f(x)$ , the so-called “higher derivatives” are the result of iterating the process of differentiation:  $f(x)$ ,  $f'(x)$ ,  $f''(x)$ , and so on. Iterating the process of partial differentiation leads to the “higher partials.”

If  $f(x, y)$  is a function on  $\mathbb{R}^2$ , then the first-order partial derivatives are  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$ . There are *four* second-order partial derivatives:  $\frac{\partial}{\partial x} \frac{\partial f}{\partial x}$ ,  $\frac{\partial}{\partial x} \frac{\partial f}{\partial y}$ ,  $\frac{\partial}{\partial y} \frac{\partial f}{\partial x}$ , and  $\frac{\partial}{\partial y} \frac{\partial f}{\partial y}$ . Usually we would write these using the respective notations

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y \partial x}, \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}.$$

Similarly, there would be  $3 \times 3 = 9$  second-order partial derivatives for a function of 3 variables, 16 for a function of four variables, and so on. And for a function of two variables there would be  $2 \times 2 \times 2 = 8$  *third-order* partials, and so on.

Note that in notation like  $\frac{\partial^2 f}{\partial x \partial y}$ , you would read the bottom from right to left to determine the order in which partial derivatives are being applied; this is because  $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \frac{\partial f}{\partial y}$  has the  $\frac{\partial}{\partial y}$  being applied to  $f$  first, followed by the  $\frac{\partial}{\partial x}$  being applied to the result.

Speaking of the order of application, it turns out not to matter. Usually.

**Theorem:** Order doesn't matter for partial derivative operators

Let  $f(x, y)$  be a function on  $\mathbb{R}^2$ . If  $\frac{\partial^2 f}{\partial x \partial y}$  and  $\frac{\partial^2 f}{\partial y \partial x}$  both exist and are continuous, then they are equal. It follows that for a function  $f(\vec{x})$  on  $\mathbb{R}^n$ , the order of application doesn't matter for any pair of partial derivative operators  $\frac{\partial}{\partial x_i}$  and  $\frac{\partial}{\partial x_j}$  as long as both double partials exist and are continuous.

Soon in the next section we will put down some global assumptions that will hold throughout the course and that will make all double partials exist and be continuous. So you can use  $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$  throughout our course.

Let's look at the theorem in action using the example  $f(x, y, z) = 4xyz + 3x^2z + y^2$  from earlier. Recall that we had

$$\begin{aligned} \frac{\partial f}{\partial x} &= 4yz + 6xz \\ \frac{\partial f}{\partial y} &= 4xz + 2y. \end{aligned}$$

Now apply  $\frac{\partial}{\partial y}$  to  $\frac{\partial f}{\partial x}$  and apply  $\frac{\partial}{\partial x}$  to  $\frac{\partial f}{\partial y}$ :

$$\begin{aligned} \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y} \frac{\partial f}{\partial x} = \frac{\partial}{\partial y} (4yz + 6xz) = 4z \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial y} = \frac{\partial}{\partial x} (4xz + 2y) = 4z. \end{aligned}$$

Indeed, they are equal. Neat!

We will not prove this theorem here, but if you are curious then just wait until we start doing some integration. At that point there will be an easy way to prove it.

### 1.7.2 Smoothness

A function of any number of variables is called *smooth* if all its derivatives of all orders exist at all points where the function is defined. So for example  $f(x)$  is smooth if  $f(x)$ ,  $f'(x)$ ,  $f''(x)$ , etc. *all exist for all  $x$* . For a function  $f(x, y, z, w)$  to be smooth, things like

$$\frac{\partial^{38} f}{\partial x^{15} \partial y^4 \partial z^8 \partial w^{11}}$$

must exist—any partial derivative of any order that you can concoct must exist.

If a function has multiple *components* (meaning that its outputs have multiple components), then we call it smooth if each of its component functions is smooth. So for example a path  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  is smooth if each of the functions  $x_1(t), x_2(t), \dots, x_n(t)$  is smooth in the above sense. And a function of multiple variables with multiple components  $\vec{F}(\vec{x}) = (F_1(\vec{x}), F_2(\vec{x}), \dots, F_n(\vec{x}))$  is smooth if each of its component functions is smooth in the above sense. This terminology applies whether we are thinking of  $\vec{F}(\vec{x})$  as a vector field or just as an  $\mathbb{R}^n$ -valued function on  $\mathbb{R}^m$ .

I introduced this terminology just so we can lay down the following assumptions for the rest of the course:

We assume that everything is smooth.

In this course, unless we explicitly state otherwise, all paths, functions, and vector fields are assumed to be smooth. (We are also keeping the assumption from before that paths have nonzero velocity.) In other words, you can take the word “function” to mean “smooth function,” take the word “path” to mean “smooth path with nonvanishing derivative,” and take the word “vector field” to mean “smooth vector field.”

Making these assumptions now will save us the headache of constantly fussing over differentiability hypotheses in theorems. A lot of functions are smooth anyway. And you can go back and worry about non-smooth versions of our theorems later, if you have the need.

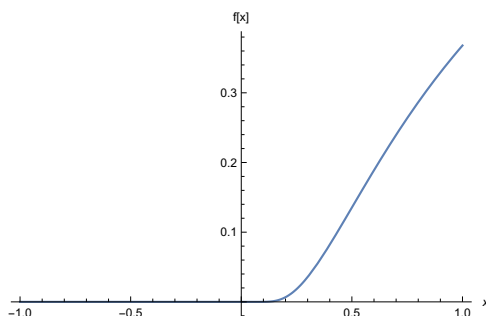
To give you some intuition about the smoothness condition, and to convince you that a lot of functions are smooth, let’s look at smoothness for functions of one variable. First of all, if  $f(x)$  is a polynomial, then  $f(x)$  is smooth (its derivative is given by using the power rule, which always just yields another polynomial). That’s already a lot of functions, but the same goes for the “infinite sum” version of polynomials: functions that equal their own Taylor series everywhere. So our usual transcendental friends  $\sin(x)$ ,  $\cos(x)$ ,  $e^x$ , and  $\log(x)$  are all smooth. Sums, products, quotients, and composites of smooth functions are smooth. But there’s quite a bit more still.

Here’s a fun example. It is possible to have a smooth function  $f(x)$  take the value 0 for  $x < 0$ , and

then suddenly (but smoothly!) start ramping up and doing other things. Here's how it's defined:

$$f(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \leq 0. \end{cases}$$

Normally with piecewise-defined functions, one would worry about a lack of differentiability at points where the pieces meet— in this case at  $x = 0$ . But in this case  $e^{-1/x}$  and every single one of its derivatives all limit to 0 as  $x$  approaches 0. So this function ends up being smooth. A plot of  $f(x)$ :



### 1.7.3 Multivariable Chain Rule

We now know how to take derivatives of functions  $f(\vec{x})$  (take partial derivatives) and how to take derivatives of paths  $\vec{p}(t)$  (take velocity). In section 1.5.2, we explored the meaning of the derivative of the single-component single-variable composite function  $f(\vec{p}(t))$ . This ought to be related to derivatives of  $f$  and of  $\vec{p}$  via some kind of chain rule. It is:

#### The Multivariable Chain Rule

Let  $f(x_1, x_2, \dots, x_n)$  be a function on  $\mathbb{R}^n$  and let  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  be a path in  $\mathbb{R}^n$ . Then

$$\frac{d}{dt} f(x_1(t), x_2(t), \dots, x_n(t)) = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} + \dots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dt}.$$

Let's start with a demonstration, and then I will try to motivate this form of the chain rule. We will reuse the same example from sections 1.3 and 1.5.2: Let  $f(x, y) = x^2 + y^2$  and let  $\vec{p}(t) = (t+1, t-1)$ . Remember that we often like to refer the component functions of a path by using the names of coordinates:  $\vec{p}(t) = (x(t), y(t))$ . Now by the multivariable chain rule, we have

$$\frac{d}{dt} f(\vec{p}(t)) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} = (2x(t))(1) + (2y(t))(1) = (2(t+1)) + (2(t-1)) = 4t,$$

which is the same as what we got in section 1.5.2 without using the chain rule! You may have noticed that when we filled in  $\frac{\partial f}{\partial x}$  we didn't just write  $2x$ ; rather we wrote  $2x(t)$ , putting in the appropriate component of the path. Doing this is just part of the chain rule that I left out of the



notation just to make it easier on the eyes<sup>1</sup>.

On to the motivation. Consider a function  $f(x, y)$  on  $\mathbb{R}^n$ . If you change  $x$  by a little bit  $\Delta x$ , the response of  $f(x, y)$  is governed roughly<sup>2</sup> by  $\frac{\partial f}{\partial x}$ :

$$f(x + \Delta x, y) \approx f(x, y) + \frac{\partial f}{\partial x} \Delta x.$$

Similarly, changing  $y$  by a little bit  $\Delta y$  yields a response of  $f(x, y)$  that is governed roughly by  $\frac{\partial f}{\partial y}$ :

$$f(x, y + \Delta y) \approx f(x, y) + \frac{\partial f}{\partial y} \Delta y.$$

What if *both* changes  $\Delta x$  and  $\Delta y$  are made? How does  $f(x, y)$  respond? Well, you just *add up* the contribution to the change in  $f(x, y)$  that is due to changing  $x$  and the contribution that is due to changing  $y$ :<sup>3</sup>

$$f(x + \Delta x, y + \Delta y) \approx f(x, y) + \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y.$$

Now suppose you have a path  $\vec{p}(t) = (x(t), y(t))$ . What we're really interested in is how  $f(\vec{p}(t)) = f(x(t), y(t))$  responds to variations in  $t$ . Let's deal with the  $\vec{p}(t)$  part first: we have

$$\begin{aligned}x(t + \Delta t) &\approx x(t) + \frac{dx}{dt} \Delta t \\y(t + \Delta t) &\approx y(t) + \frac{dy}{dt} \Delta t.\end{aligned}$$

So putting everything together we have the following for how  $f(x(t), y(t))$  responds to a change  $\Delta t$ :

$$\begin{aligned}f(x(t + \Delta t), y(t + \Delta t)) &\approx f\left(x(t) + \frac{dx}{dt} \Delta t, y(t) + \frac{dy}{dt} \Delta t\right) \\&\approx f(x(t), y(t)) + \frac{\partial f}{\partial x} \left(\frac{dx}{dt} \Delta t\right) + \frac{\partial f}{\partial y} \left(\frac{dy}{dt} \Delta t\right) \\&= f(x(t), y(t)) + \left(\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}\right) \Delta t.\end{aligned}$$

And there's the multivariable chain rule. The thing in big parentheses on the last line must be the derivative  $\frac{d}{dt} f(x(t), y(t))$ .

There is something very practical to the chain rule: it provides a relationship between the rate of change of a function and the rates of change of its inputs. Suppose your happiness  $H(h, r, g)$  is a function of only three things: your hunger level  $h$ , your restedness level  $r$ , and your grade  $g$  in math 6A. Since this is a silly example, the units are all made up— we could just vaguely call them “units of happiness,” “units of hunger,” etc. Now suppose you knew that your hunger is currently

---

<sup>1</sup>People do this all the time with the single variable chain rule:  $(g(f(x)))' = g'(f(x))f'(x)$  is often written as  $\frac{dg}{dx} = \frac{dg}{df} \frac{df}{dx}$ . It's really more like  $\frac{dg}{dx} = \frac{dg}{df} \Big|_{f=f(x)} \frac{df}{dx}$ , but that's just not as pretty. Context usually suggests the intended interpretation of the symbols.

<sup>2</sup>By “roughly” I mean “to first order.” That is, if one considers only the aspect of the response of  $f(x, y)$  to a change  $\Delta x$  that is *linear* in  $\Delta x$ , then that response is governed by  $\frac{\partial f}{\partial x}$ .

<sup>3</sup>What you're seeing here is basically the first-order terms in the *multivariable Taylor series* of  $f(x, y)$ .

changing at a rate of  $-3$  (units of hunger per unit time), your restedness is changing at a rate of  $-2$  (units of restedness per unit time) and your grade is changing at a rate of  $0.1$  (grade points per unit time). Then what is the current time-rate-of-change of your happiness? The chain rule says it is:

$$\frac{dH}{dt} = -3\frac{\partial H}{\partial h} - 2\frac{\partial H}{\partial r} + 0.1\frac{\partial H}{\partial g}.$$

So to figure it out all you need fill in is the partial derivatives of  $H$ . But remember that each partial derivative of  $H$  is a *function* of  $h$ ,  $r$ , and  $g$ . So you would also have to know the *current* values of  $h$ ,  $r$ , and  $g$  so that you could fill in the *current* values of the partial derivatives of  $H$ .

You might come up with a more accurate function  $H(h, r, g, t)$  to describe your happiness, where  $t$  is time. For example maybe you are grumpier during certain parts of the day and happier during others. Here we have *explicit* time dependence, and the chain rule would say

$$\frac{dH}{dt} = \frac{dh}{dt}\frac{\partial H}{\partial h} + \frac{dr}{dt}\frac{\partial H}{\partial r} + \frac{dg}{dt}\frac{\partial H}{\partial g} + \frac{\partial H}{\partial t}.$$

One way to think of this is that it is just applying the chain rule while using the fact that  $\frac{dt}{dt} = 1$ . If you were thinking of a particular path through  $h$ - $r$ - $g$ -space, then you can upgrade it to a path through  $h$ - $r$ - $g$ - $t$ -space by introducing a  $t$ -component which is just  $t$  itself. Note the difference between the symbols  $\frac{dH}{dt}$  and  $\frac{\partial H}{\partial t}$ . The first one is the *total* rate of change of  $H(h(t), r(t), g(t), t)$  in time, taking into account the ways in which  $h$ ,  $r$ , and  $g$  are changing in time. The second is the *partial* derivative of  $H(h, r, g, t)$  with respect to  $t$ , and it tells you how  $H$  changes with  $t$  if  $h$ ,  $r$ , and  $g$  are all held constant.

## 1.8 Domains

Functions and vector fields do not necessarily have to be defined on all of  $\mathbb{R}^n$  and paths do not necessarily have to be defined on all of  $\mathbb{R}$ . The domain<sup>4</sup> of definition may be restricted for certain reasons:

1. The definition of a function or path may involve other functions/operations that have restricted domain. Also, one might only care about a certain segment of a path, or one might want a path to stay within a certain region.
2. The coordinates  $x_1, x_2, \dots, x_n$  may be used to represent something in such a way that it only makes sense to consider points within a certain subregion of  $\mathbb{R}^n$

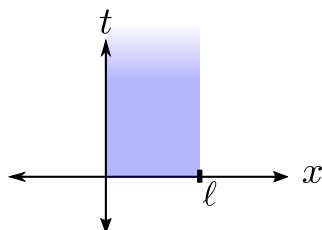
An example of the first situation: One might want to work with  $f(x, y, z) = \log\left(\frac{z}{(x-2)^2+(y+1)^2}\right)$ . Due to the division, this function would not be defined at points of the form  $(2, -1, z)$ . Due to the log, it would not be defined at points of the form  $(x, y, z)$  for which  $z \leq 0$ . The set of remaining points, those points  $(x, y, z)$  in  $\mathbb{R}^3$  for which  $(x, y) \neq (2, -1)$  and  $z > 0$ , can be taken to be the domain of  $f$ . If we call that set  $A$ , then we would say  $f$  is a function *on*  $A$ , rather than *on*  $\mathbb{R}^n$ . Now consider a path  $\vec{p}(t) = (\sqrt{t}, t, 1 - t)$ . Due to the  $\sqrt{t}$ , this path is only defined for  $t \geq 0$ . Also, we might

---

<sup>4</sup>The *domain* of a function is the set of things on which it is defined— the set of valid inputs to it.

insist that the path stay within  $A$ , possibly because we want to be able to consider the composite  $f(\vec{p}(t))$ . Then we would also insist that  $1 - t > 0$ , so  $t < 1$ . So this is a situation where we would want to take the domain of the path to be the interval  $[0, 1)$ .

An example of the second situation: Suppose that  $T(x, t)$  is the temperature of a metal rod at a distance  $x$  along the rod, and at time  $t$ . Then the domain might only contain  $(x, t)$  for which  $x$  is in the interval  $[0, \ell]$ , where  $\ell$  is the length of the rod. Maybe also you only care about  $t \geq 0$ , because you've got an initial temperature distribution on the rod and you are only concerned with how that evolves forward in time according to your model. Then the domain of  $T(x, t)$  is a strip in  $\mathbb{R}^2$  that extends infinitely in one direction only:



## 2 Coordinate Systems and Spaces

A coordinate system assigns labels to points in a space so that we can more easily talk about it. It is sort of an artificial construct, since the points being labeled are often not even distinguishable from each other. Sometimes the stroke of insight needed to make an impossible problem possible is to express everything in just the right coordinates. Sometimes the initial setup of a problem inherently suggests certain coordinates, and one eventually has to transform everything so that it is expressed in better coordinates for solving the problem. This is why it will be important for us to understand how to transform expressions of objects from coordinate system to another.

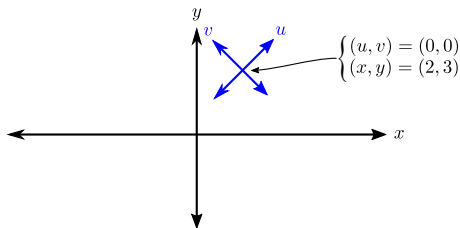
In section 2.1, we will begin our study of coordinate transformations in  $\mathbb{R}^n$ . Since  $\mathbb{R}^n$  has cartesian coordinates built-in as a sort of “default setting,” putting new coordinates on it is always a matter of transforming from cartesian coordinates to the new coordinates. Then, in the second half of section 2, we will extend the idea of coordinate systems to describe a given family of objects as “points in a space,” where the space will no longer have to be  $\mathbb{R}^n$  or even a part of  $\mathbb{R}^n$ .

### 2.1 Coordinate Systems on $\mathbb{R}^n$

A coordinate system is a way of labeling points by lists of real numbers. With the way we defined  $\mathbb{R}^n$ , points *already have* such a labeling because they *are* the labels—elements of  $\mathbb{R}^n$  are already lists of real numbers. This kind of default labeling that  $\mathbb{R}^n$  has is one example of *cartesian coordinates*. Translating or applying invertible linear transformations can give other cartesian coordinates.

For example,  $\mathbb{R}^2$  has its default cartesian coordinates, where we label any element  $(x, y)$  of  $\mathbb{R}^2$  by the list of coordinates  $(x, y)$ . So  $(3, 4)$  gets the label  $(3, 4)$ ,  $(1, 2)$  gets the label  $(1, 2)$ , etc. Let's give

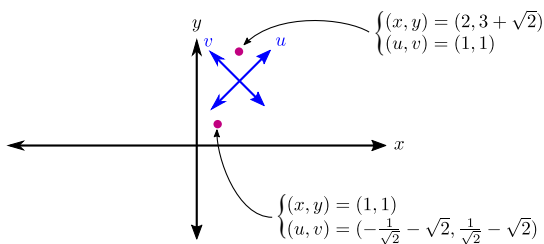
$\mathbb{R}^2$  new coordinates  $(u, v)$ , where  $(u, v)$  has its origin shifted to the point  $(x, y) = (2, 3)$  and where the coordinate axes are rotated by counterclockwise by  $45^\circ$ .



Now each point has two names, its  $(x, y)$ -name and its  $(u, v)$ -name. For example the point that is called  $(2, 3)$  in the  $(x, y)$  coordinate system has the name  $(0, 0)$  in the  $(u, v)$  coordinate system. We say that  $(0, 0)$  are the *u-v-coordinates* of the point  $(x, y) = (2, 3)$ . What about the point  $(x, y) = (1, 1)$ ? What are its *u-v-coordinates*? Or how about the reverse sort of problem: What are the *x-y-coordinates* of the point  $(u, v) = (1, 1)$ ? Whenever we want to work with new coordinates, we need a transformation rule that tells us how points get renamed. In this case, the transformation rules in both directions are:

$$\begin{aligned} x &= \frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{2}}v + 2 & u &= \frac{x-2}{\sqrt{2}} + \frac{y-3}{\sqrt{2}} \\ y &= \frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v + 3 & v &= -\frac{x-2}{\sqrt{2}} + \frac{y-3}{\sqrt{2}} \end{aligned}$$

Now we can answer those questions:



Note that it is useful to have transformation rules going in both directions— getting  $(x, y)$  in terms of  $(u, v)$  and getting  $(u, v)$  in terms of  $(x, y)$ . A crucial feature of coordinate systems is that they should assign to each point a *unique* label, so it should always be possible to have transformation rules in both directions.

The *u-v*-coordinate system assigns to each point in  $\mathbb{R}^2$  a pair of numbers. Looking at just the *u* coordinate, it assigns a number to each point in  $\mathbb{R}^2$ . We already know about a kind of object that does just that: a function on  $\mathbb{R}^2$ ! Really, coordinates on a space are just functions on that space. The coordinate transformation rules are the definitions of those functions:

$$\begin{aligned} u(x, y) &= \frac{x-2}{\sqrt{2}} + \frac{y-3}{\sqrt{2}} \\ v(x, y) &= -\frac{x-2}{\sqrt{2}} + \frac{y-3}{\sqrt{2}}. \end{aligned}$$

Even the  $x$ - $y$ -coordinate system is just a pair of functions. The functions  $x$  and  $y$  look really simple:

$$\begin{aligned}x(x, y) &= x \\y(x, y) &= y,\end{aligned}$$

but that's because we are expressing them in  $x$ - $y$ -coordinates! If we express the functions  $x$  and  $y$  in  $u$ - $v$  coordinates instead, then we see the other coordinate transformation rule:

$$\begin{aligned}x(u, v) &= \frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{2}}v + 2 \\y(u, v) &= \frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v + 3.\end{aligned}$$

We can now set up our general framework for coordinate systems:

### Coordinate Systems on $\mathbb{R}^n$ (or part of $\mathbb{R}^n$ )

A coordinate system on a subset  $U$  of  $\mathbb{R}^n$  is a list of functions

$$\begin{aligned}u_1 &= u_1(x_1, x_2, \dots, x_n) \\u_2 &= u_2(x_1, x_2, \dots, x_n) \\&\vdots \\u_n &= u_n(x_1, x_2, \dots, x_n)\end{aligned}$$

such that each point  $(x_1, x_2, \dots, x_n)$  in  $U$  gets assigned a unique list of coordinates  $(u_1, \dots, u_n)$ . The functions  $u_i(\vec{x})$  give you the transformation rule for converting points  $(x_1, x_2, \dots, x_n)$  from  $x_i$ -coordinates to  $u_i$ -coordinates. There should also be a transformation rule going in the other direction, giving you  $x_i$ -coordinates in terms of  $u_i$ -coordinates:

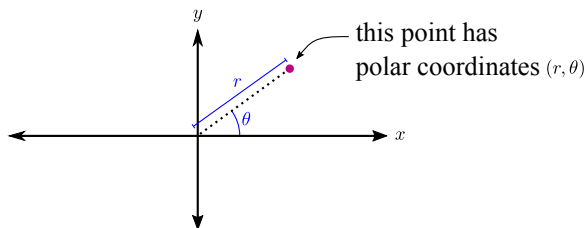
$$\begin{aligned}x_1 &= x_1(u_1, u_2, \dots, u_n) \\x_2 &= x_2(u_1, u_2, \dots, u_n) \\&\vdots \\x_n &= x_n(u_1, u_2, \dots, u_n)\end{aligned}$$

In the example above, the new coordinates  $(u, v)$  were defined on all of  $\mathbb{R}^2$ . But in general, new coordinates can be introduced that are only defined on a portion  $U$  of  $\mathbb{R}^n$ . For example when we define polar coordinates on  $\mathbb{R}^2$ , we will see that they are not defined at certain places (the set  $U$  on which they are defined turns out to be  $\mathbb{R}^2$  with a certain ray deleted).

#### 2.1.1 Some Useful Coordinate Systems on $\mathbb{R}^2$ and $\mathbb{R}^3$

In this section we will define three famous non-cartesian coordinate systems. The first is the *polar* coordinate system on  $\mathbb{R}^2$ .

**Polar Coordinates** Instead of specifying a point by giving its  $x$  and  $y$  coordinates, we can specify it by giving its distance  $r$  from  $(x, y) = (0, 0)$  and the angle  $\theta$  that it makes with the  $x$ -axis:



The transformation rules for going between polar and cartesian coordinates are as follows:

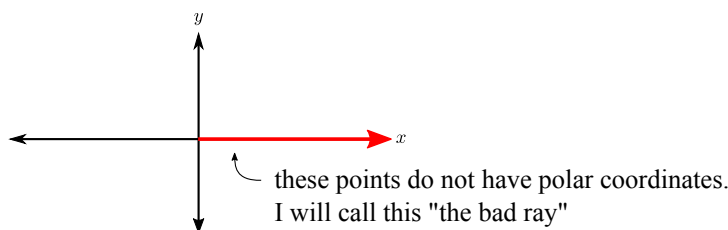
$$\begin{aligned} x &= r \cos(\theta) & r &= \sqrt{x^2 + y^2} \\ y &= r \sin(\theta) & \theta &= \text{atan2}(y, x). \end{aligned}$$

You could derive these yourself by just using the meaning attributed to  $r$  and  $\theta$ . The function  $\text{atan2}(y, x)$  is a version of  $\arctan(y/x)$  that corrects for the fact that  $\arctan(y/x)$  doesn't notice the relative signs of  $x$  and  $y$ <sup>5</sup>.

There are some restrictions on the values that  $r$  and  $\theta$  are allowed to take:

$$\begin{aligned} r &> 0 \\ 0 &< \theta < 2\pi \end{aligned}$$

Note in particular that the ray of points that *would* have gotten the polar coordinates  $r = 0$  or  $\theta = 0$  are being excluded:



<sup>5</sup>The function  $\text{atan2}(y, x)$  is defined like this:

$$\text{atan2}(y, x) \begin{cases} \arctan(y/x) & \text{if } y > 0 \text{ and } x \neq 0 \\ \pi & \text{if } y = 0 \text{ and } x < 0 \\ \arctan(y/x) + \pi & \text{if } y < 0 \text{ and } x \neq 0 \\ \pi/2 & \text{if } y > 0 \text{ and } x = 0 \\ 3\pi/2 & \text{if } y < 0 \text{ and } x = 0, \end{cases}$$

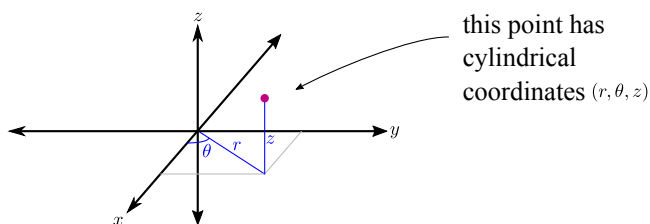
assuming a version of  $\arctan$  that takes values in the range  $(0, \pi)$ . Basically,  $\text{atan2}(y, x)$  corrects for the fact that  $\arctan(y/x)$  doesn't notice the relative signs of  $x$  and  $y$ . If we had just used  $\theta = \arctan(y/x)$ , then the points  $(x, y) = (1, 1)$  and  $(x, y) = (-1, -1)$  would have gotten the same coordinates  $(r, \theta) = (\sqrt{2}, \pi/4)$ , which is a big no-no for coordinate systems. The correct  $\theta$  coordinate for  $(-1, -1)$  should be  $\text{atan2}(-1, -1) = \pi + \pi/4$ .

These restrictions arise for a few reasons:

- *The meanings of the coordinates.* In this case, the coordinate  $r$  represents distance from the origin, so it doesn't make sense to allow it to be negative. This leads to the restriction  $r \geq 0$ .
- *The requirement that each point get a unique set of coordinates.* The  $r$ -coordinate for the point  $(x, y) = (0, 0)$  would clearly have to be 0. But what should be its  $\theta$  coordinate? The same question arises for any point of the form  $(x, y) = (x, 0)$  with  $x \geq 0$ : It would clearly have an  $r$ -coordinate equal to  $x$ , but what should  $\theta$  be? Should it be 0 or  $2\pi$ ? Since we don't know what to choose, we just kick out from the polar coordinate system all points of the form  $(x, y) = (x, 0)$  with  $x \geq 0$ ; let's call the set of those points "the bad ray." The set  $U$  on which polar coordinates are defined is the complement: the set of points in  $\mathbb{R}^2$  that are not in the bad ray. You might wonder: Why kick out the points for which we are unsure which coordinates to choose? Why not just choose something? Why not choose  $\theta = 0$  on the bad ray? Read on.
- *The requirement that coordinate transformations be smooth.* Remember that we want all our functions to be smooth for the rest of the course—this includes the coordinate functions  $r(x, y)$  and  $\theta(x, y)$ . We want to use coordinates to do calculus after all, so we don't want coordinates to interfere with the differentiability of the objects we study! If we had chosen  $\theta = 0$  or  $\theta = 2\pi$  for the points in the bad ray, then  $\theta(x, y)$  would fail to even be a continuous function! Passing across the bad ray would lead to a discontinuous "jump" in  $\theta$  values.

Okay, on to the other two famous coordinate systems, both on  $\mathbb{R}^3$ . For each one I will give you the meanings of the coordinates, the coordinate transformation rules going in both directions, and the restrictions on the values the coordinates can take.

**Cylindrical Coordinates** These are a lot like polar coordinates, but extended to  $\mathbb{R}^3$  by including an additional  $z$ -coordinate. A point  $(x, y, z)$  in  $\mathbb{R}^3$  is specified in cylindrical coordinates by giving  $(r, \theta, z)$ , where  $(r, \theta)$  is actually just the polar coordinate of the point  $(x, y)$  in  $\mathbb{R}^2$  and, where  $z$  in cylindrical coordinates is identical to the cartesian coordinate  $z$ . In this case, the cylindrical coordinate  $r$  could be interpreted as "distance from the  $z$ -axis."



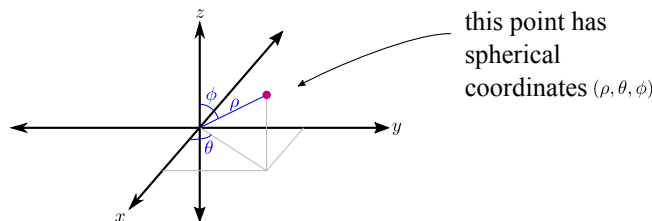
The transformation rules for going between cylindrical and cartesian coordinates are as follows:

$$\begin{aligned} x &= r \cos(\theta) & r &= \sqrt{x^2 + y^2} \\ y &= r \sin(\theta) & \theta &= \text{atan2}(y, x) \\ z &= z & z &= z. \end{aligned}$$

The range of values that  $(r, \theta, z)$  can take is as follows:

$$\begin{aligned} r &> 0 \\ 0 &< \theta < 2\pi \\ z &\text{ can be anything.} \end{aligned}$$

**Spherical Coordinates** Let's start our specification of a point in  $\mathbb{R}^3$  by pinning down its distance  $\rho$  from the point  $(x, y, z) = (0, 0, 0)$ . I don't want to use  $r$  for that coordinate in order to avoid confusing it with the polar and cylindrical  $r$ , so I am using the Greek equivalent  $\rho$ . Now if  $\rho$  is specified, what else needs to be specified to determine which point we are talking about? Well, once  $\rho$  is specified, there is a sphere of possible points we could be talking about. To pin down a point on the sphere, we can use two angles  $\theta$  and  $\phi$ . The first angle  $\phi$  is the angle our point makes with the  $z$ -axis. Specifying  $\phi$  is like choosing a circle of "latitude" on the sphere. After that, one more angle  $\theta$  is used to choose which point on that circle we are talking about.



The  $\theta$  coordinate of  $(x, y, z)$  is really the same one from the cylindrical coordinate system: it is the polar  $\theta$  coordinate of  $(x, y)$ . The transformation rules for going between spherical and cartesian coordinates are as follows:

$$\begin{aligned} x &= \rho \sin(\phi) \cos(\theta) & \rho &= \sqrt{x^2 + y^2 + z^2} \\ y &= \rho \sin(\phi) \sin(\theta) & \theta &= \text{atan2}(y, x) \\ z &= \rho \cos(\phi) & \phi &= \text{atan2}(\sqrt{x^2 + y^2}, z). \end{aligned}$$

Try to derive these yourself! We will probably do it in lecture also. The range of values that  $(\rho, \theta, \phi)$  can take is as follows:

$$\begin{aligned} \rho &> 0 \\ 0 &< \theta < 2\pi \\ 0 &< \phi < \pi \end{aligned}$$

Why does  $\phi$  only get to do a half-turn? Again it is due to the requirement that each point get a unique set of coordinates. Think about how allowing  $\phi$  to go beyond  $\pi$  would give some points multiple  $(\rho, \theta, \phi)$  labels.

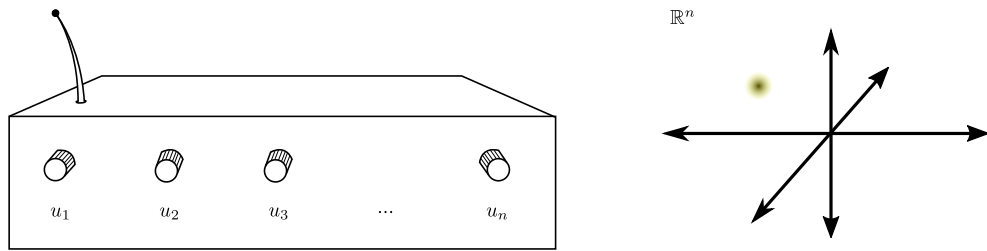
## 2.1.2 Understanding Coordinate Systems

We are going to discuss two basic ways to understand and visualize coordinate systems:



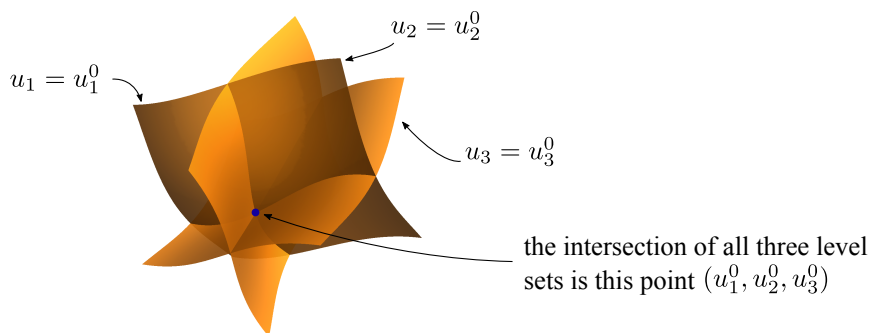
- Look at what happens when you fix one coordinate and let all the other coordinates vary. This amounts to considering the *level sets of coordinate functions*.
- Look at what happens when you vary one coordinate and keep all the others fixed. This will generate paths which I will call *coordinate paths*.

Suppose we have a coordinate system  $\vec{u}(\vec{x})$  on some subset  $U$  of  $\mathbb{R}^n$ . I like to imagine that I've got a box with  $n$  control knobs on it labeled  $u_1, \dots, u_n$ . As I turn the knobs, a little glowing point flies around in  $U$  taking the various positions  $(u_1, u_2, \dots, u_n)$  that the knobs are set to. Each point in  $U$  corresponds to one unique knob arrangement.



There are two things I could look at to try to understand how a particular knob such as  $u_1$  is working.

I could set  $u_1$  to a certain value<sup>6</sup>  $u_1^0$ , and look at which points can be reached by turning all the other knobs in every possible way. The set of points that can be reached are all the points that have a  $u_1$ -coordinate of  $u_1^0$ . Since  $u_1$  itself is really a function  $u_1(\vec{x})$  on  $\mathbb{R}^n$ , I would actually be looking at a level set  $u_1(\vec{x}) = u_1^0$ . If I do that for lots of values of  $u_1^0$ , then I can have a better understanding of the  $u_1$  knob. If I also go through that process for  $u_2$ , then for  $u_3$ , etc., then I would have a pretty full picture! Any point  $(u_1^0, u_2^0, \dots, u_n^0)$  is the unique point in the intersection of the level sets  $u_1(\vec{x}) = u_1^0, u_2(\vec{x}) = u_2^0$ , etc. If this paragraph isn't clear yet, wait until we do examples with polar and spherical coordinates in just a moment, then come back and reread this paragraph.

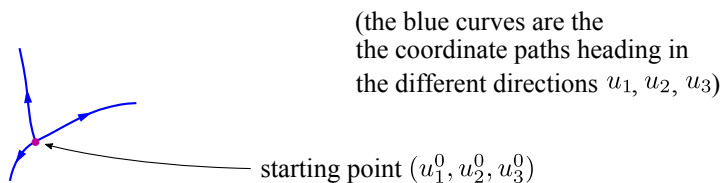


<sup>6</sup>The superscript in  $u_1^0$  is not a zeroth power. It's just a superscript.

The second thing I could do is fix the  $u_2, \dots, u_n$  knobs to a certain setting  $u_2^0, \dots, u_n^0$ , and then turn just the  $u_1$  knob and watch the point fly around. It should traverse some path, and I will call such a path a *coordinate path*. So for example if the initial setting was  $(u_1^0, u_2^0, \dots, u_n^0)$ , then the  $u_1$  coordinate path starting there would be the path

$$(u_1(t), u_2(t), \dots, u_n(t)) = (u_1^0 + t, u_2^0, \dots, u_n^0).$$

This is the expression of the path in the  $u_i$  coordinate system. You would have to use the coordinate transformation rules to turn it into a path in cartesian coordinates.



### Coordinate Paths

Suppose you have a coordinate system  $u_1, u_2, \dots, u_n$  and a point  $(u_1^0, \dots, u_n^0)$  expressed in that coordinate system. Then the *coordinate path* in the  $u_1$  direction starting at  $(u_1^0, \dots, u_n^0)$  is the path given by

$$(u_1(t), u_2(t), \dots, u_n(t)) = (u_1^0 + t, u_2^0, \dots, u_n^0).$$

The coordinate path in the  $u_2$  direction starting at  $(u_1^0, \dots, u_n^0)$  is the path given by

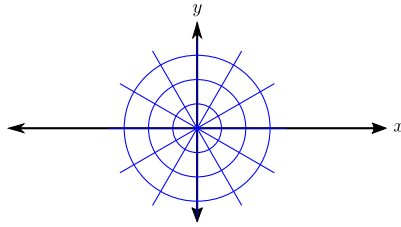
$$(u_1(t), u_2(t), \dots, u_n(t)) = (u_1^0, u_2^0 + t, u_3^0, \dots, u_n^0).$$

And so on.

**Polar Coordinates Example** Let's try these ideas out on polar coordinates. (Now I am imagining my control box with two knobs, one labeled  $r$  and the other labeled  $\theta$ , controlling the position of a point in  $\mathbb{R}^2$ .) What are the level sets for  $r$ ? For a constant  $r_0$ , setting  $r(x, y) = r_0$  is setting  $\sqrt{x^2 + y^2} = r_0$ . The points satisfying this are the points on a circle of radius  $r_0$ . (We didn't actually have to bring  $x$  and  $y$  into the picture to deduce this. Just using the definition of  $r$  as the distance of a point from  $(x, y) = (0, 0)$ , we can conclude that  $r = r_0$  defines a circle of radius  $r_0$ .) So the level sets for  $r$  are circles of various radii.

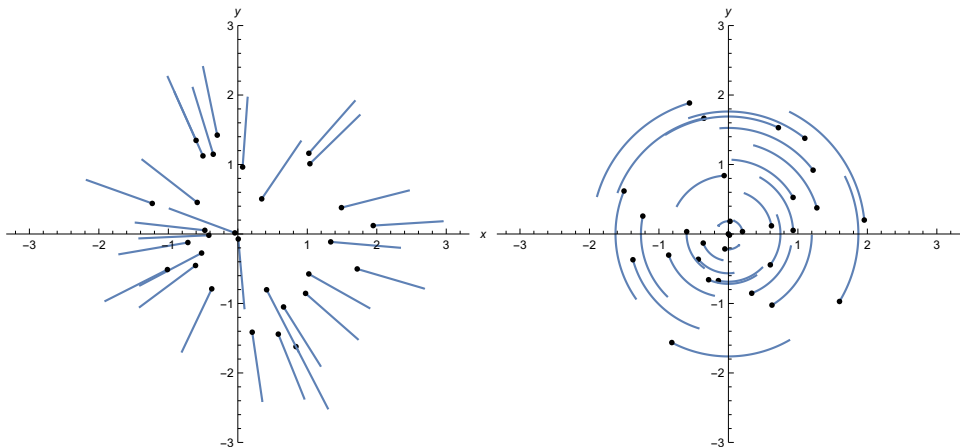
For a constant  $\theta_0$ , what is the set of points satisfying  $\theta = \theta_0$ ? Since it's kind of a pain to think about the function  $\text{atan2}$ , let's just reason through it in terms of the original meaning of the  $\theta$  coordinate. The  $\theta$  coordinate measures the angle that a point makes with the  $x$ -axis. What is the set of points that make a constant angle  $\theta_0$  with the  $x$ -axis? It's a ray emanating from the origin  $(x, y) = (0, 0)$  and making an angle  $\theta_0$  with the  $x$ -axis. So the level sets for  $\theta$  are rays emanating from the origin at various angles.

Now overlay the pictures together: circles of various radii  $r_0$  and rays at various angles  $\theta_0$ . You get the familiar "polar coordinate grid!"



To specify a point in polar coordinates, you choose a circle (by choosing a radius  $r_0$ ) and you choose a ray (by choosing an angle  $\theta_0$ ). The unique point (“grid point”) where the circle and ray intersect is the point  $(r, \theta) = (r_0, \theta_0)$  that you specified.

Now let’s examine coordinate paths for polar coordinates. A coordinate path along the  $r$  coordinate would be the path  $(r(t), \theta(t)) = (r_0 + t, \theta_0)$ . This would sit in the  $\theta = \theta_0$  ray and travel outward along the ray away from the origin, starting at a distance  $r_0$  from the origin. A coordinate path along the  $\theta$  coordinate would be the path  $(r(t), \theta(t)) = (r_0, \theta_0 + t)$ . This would sit on the  $r = r_0$  circle and travel around in the direction of increasing  $\theta$  (so counterclockwise), starting at angle  $\theta_0$  off of the  $x$ -axis.



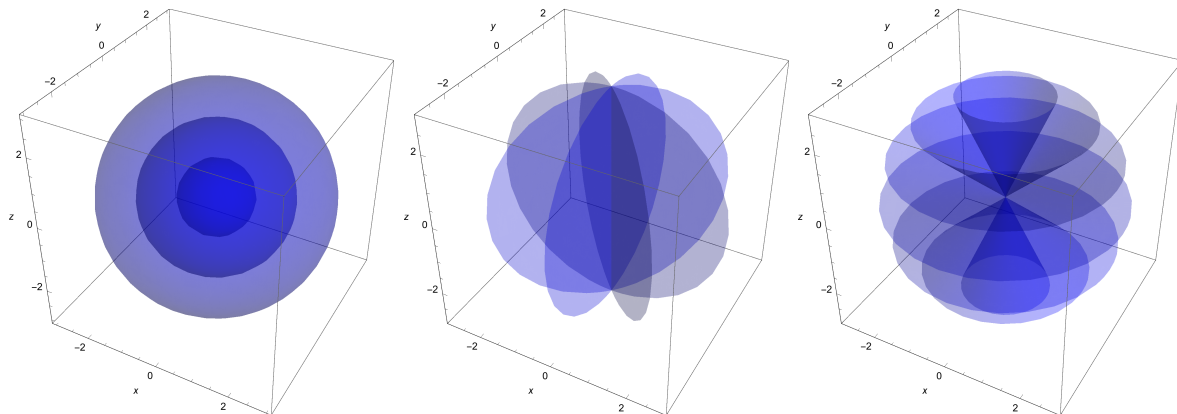
(Here I’ve plotted a random selection of coordinate paths, in the  $r$  direction on the left and in the  $\theta$  direction on the right. There’s a dot at the starting point of each coordinate path, and I let each one run from  $t = 0$  to  $t = 1$ .)

**Spherical Coordinates Example** Let’s do the same kind of analysis but for spherical coordinates. (Now I’m imagining my control box with three knobs, labeled  $\rho$ ,  $\theta$ , and  $\phi$ , controlling the position of a point in  $\mathbb{R}^3$ ). For a constant  $\rho_0$ , the set of points satisfying  $\rho = \rho_0$  is the set of points that have distance  $\rho_0$  from the origin  $(x, y, z) = (0, 0, 0)$ . This is a sphere of radius  $\rho_0$ . So the level sets for  $\rho$  are spheres of various radii.

For a constant  $\theta_0$ , the set of points satisfying  $\theta = \theta_0$  is the set of points  $(x, y, z)$  for which the polar angle of  $(x, y)$  is  $\theta = \theta_0$ . We already saw that the set of points  $(x, y)$  with  $\theta = \theta_0$  are the points along a ray that makes angle  $\theta_0$  with the  $x$ -axis. The set of points  $(x, y, z)$  such that  $(x, y)$  are on that ray is a sort of flap emanating from the  $z$ -axis such that the angle it makes with the  $x$ - $z$  plane is  $\theta_0$ . So the level sets for  $\theta$  are flaps emanating from the  $z$ -axis at various angles.

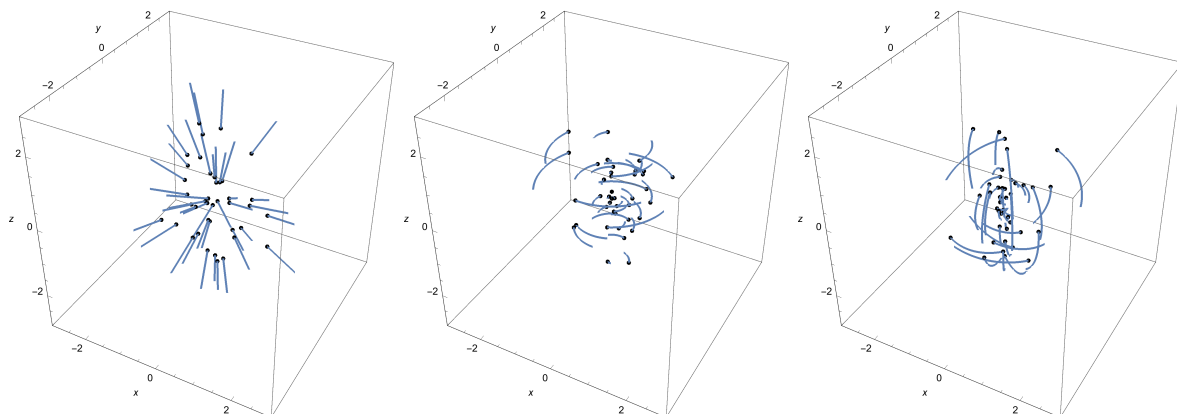
For a constant  $\phi_0$ , the set of points satisfying  $\phi = \phi_0$  is the set of points that make an angle  $\phi_0$  with the  $z$ -axis. More precisely, they are points such that when you extend a line from them to the origin, that line makes an angle of  $\phi_0$  with the  $z$ -axis. Can you see that these points make a *cone*? The level sets for  $\phi$  are cones that open at various angles.

Now put it all together. Here are the separate pictures of level sets:



I will not overlay them because the picture gets pretty messy, but try to overlay all these in your mind. What you are imagining could be called the “spherical coordinate grid.” To specify a point in spherical coordinates, you choose a sphere (by choosing a radius  $\rho_0$ ), you choose a flap (by choosing an angle  $\theta_0$ ), and you choose a cone (by choosing an angle  $\phi_0$ ). The unique point (“grid point”) where the sphere, flap, and cone intersect is the point  $(\rho, \theta, \phi) = (\rho_0, \theta_0, \phi_0)$  that you specified.

Now for coordinate paths for spherical coordinates. A coordinate path along the  $\rho$  coordinate would be the path  $(\rho(t), \theta(t), \phi(t)) = (\rho_0 + t, \theta_0, \phi_0)$ . This would sit in the ray of intersection between the  $\theta = \theta_0$  flap and the  $\phi = \phi_0$  cone, and it would travel outward along the ray away from the origin, starting at a distance  $\rho_0$  from the origin. A coordinate path along the  $\theta$  coordinate would be the path  $(\rho(t), \theta(t), \phi(t)) = (\rho_0, \theta_0 + t, \phi_0)$ . This would sit in the sphere of radius  $\rho_0$  and in the “latitude” circle at angle  $\phi_0$  from the  $z$ -axis. It would travel along the circle in the direction of increasing  $\theta$  (so counterclockwise to someone looking down from above) starting from  $\theta = \theta_0$ . A coordinate path along the  $\phi$  coordinate would be the path  $(\rho(t), \theta(t), \phi(t)) = (\rho_0, \theta_0, \phi_0 + t)$ . This would sit in the sphere of radius  $\rho_0$  and in the “meridian” at angle  $\theta_0$ . It would travel along that longitude line in the direction of increasing  $\phi$  (so towards the “south pole”) starting from  $\phi = \phi_0$ .



(The plots show a random selection of coordinate paths, in the  $\rho$  direction on the left, in the  $\theta$  direction in the middle, and in the  $\phi$  direction on the right. There's a dot at the starting point of each coordinate path, and I let each one run from  $t = 0$  to  $t = 1$ .)

### 2.1.3 Re-expressing Functions and Paths in New Coordinates

Up till now, we've only been using the "default" cartesian coordinates on  $\mathbb{R}^n$ , and we've expressed everything in terms of those coordinates. This includes functions, paths, partial derivatives of functions, velocity vectors of paths, and vector fields. In this section, we will look at how functions and paths can be expressed in a new coordinate system. The other stuff is a little more subtle and we will address it in a later section.

Functions and paths are easy to convert from one coordinate system to another. A coordinate system is the component functions that make up the transformation  $\vec{u}(\vec{x})$ , and it always has an inverse transformation  $\vec{x}(\vec{u})$ . These coordinate transformation rules are how we switch back and forth between expressions of points in  $u_i$ -coordinates and in  $x_i$ -coordinates:

$$\begin{aligned} (x_1, x_2, \dots, x_n) & \text{ becomes } (u_1(x_1, x_2, \dots, x_n), u_2(x_1, x_2, \dots, x_n), \dots, u_n(x_1, x_2, \dots, x_n)) \\ (u_1, u_2, \dots, u_n) & \text{ becomes } (x_1(u_1, u_2, \dots, u_n), x_2(u_1, u_2, \dots, u_n), \dots, x_n(u_1, u_2, \dots, u_n)). \end{aligned}$$

If a function  $f(x_1, x_2, \dots, x_n)$  on  $\mathbb{R}^n$  is given to you in terms of  $x_i$ -coordinates, you can convert it to  $u_i$ -coordinates by just plugging in the coordinate transformation in the only way that makes sense:

$$f(x_1(u_1, u_2, \dots, u_n), x_2(u_1, u_2, \dots, u_n), \dots, x_n(u_1, u_2, \dots, u_n)).$$

If a path  $(x_1(t), x_2(t), \dots, x_n(t))$  is given to you in terms of  $x_i$ -coordinates, you can also convert it to  $u_i$ -coordinates by just plugging in the coordinate transformation in the only way that makes sense:

$$(u_1(x_1(t), x_2(t), \dots, x_n(t)), u_2(x_1(t), x_2(t), \dots, x_n(t)), \dots, u_n(x_1(t), x_2(t), \dots, x_n(t))).$$

Okay that is an obnoxious amount of parentheses; let me write it vertically:

$$\begin{bmatrix} u_1(x_1(t), x_2(t), \dots, x_n(t)) \\ u_2(x_1(t), x_2(t), \dots, x_n(t)) \\ \vdots \\ u_n(x_1(t), x_2(t), \dots, x_n(t)) \end{bmatrix}.$$

Observe that re-expressing a path in new coordinates requires the coordinate transformation from old to new coordinates, while re-expressing a function in new coordinates requires the inverse coordinate transformation, from new to old.

**Examples** The function  $f(x, y) = x + y^2$  in polar coordinates is:

$$f(x(r, \theta), y(r, \theta)) = f(r \cos(\theta), r \sin(\theta)) = r \cos(\theta) + r^2 \sin^2(\theta).$$

Now that we know this, we might want to write the following:

$$f(r, \theta) = r \cos(\theta) + r^2 \sin^2(\theta).$$

I find this a little jarring to look at, but some people are okay with it. We just said that  $f(x, y) = x + y^2$ , so wouldn't it follow from the rules of function evaluation that  $f(r, \theta) = r + \theta^2$ ?! Well, sometimes people want you to look at the actual letters chosen for the variable names, and use that information to decide on the coordinates in which to express the function. So you would be expected to know from the appearance of the letters “ $r$ ” and “ $\theta$ ” in “ $f(r, \theta)$ ” that the function  $f(x, y) = x + y^2$  should be converted into polar coordinates and then evaluated at  $(r, \theta)$  to give  $f(r, \theta) = r \cos(\theta) + r^2 \sin^2(\theta)$ . One solution to resolve the ambiguity might be to decorate the function with some notation to indicate the coordinates being used to express it; for example something like

$$\begin{aligned} f_{\text{cartesian}}(x, y) &= x + y^2 \\ f_{\text{polar}}(r, \theta) &= f_{\text{cartesian}}(x(r, \theta), y(r, \theta)) = r \cos(\theta) + r^2 \sin^2(\theta). \end{aligned}$$

But I've never seen anyone do that. I will only do it if I think the ambiguity is confusing things.

Another example. The function  $f(x, y) = e^{-(x^2+y^2)}$  in polar coordinates is:

$$f(r \cos(\theta), r \sin(\theta)) = e^{-(r^2 \cos^2(\theta) + r^2 \sin^2(\theta))} = e^{-r^2}.$$

We could have seen this without doing the full substitution. If we had in mind the relation  $x^2 + y^2 = r^2$  among the coordinates then we could immediately replace the  $x^2 + y^2$  with  $r^2$  to get that  $f(r, \theta) = r^2$ .

Note that expressing  $f(x, y) = e^{-(x^2+y^2)}$  in polar coordinates has greatly simplified it to  $e^{-r^2}$ , and now we can much more readily imagine what the function is doing.

Another example, going the other way. What is the function  $f(r, \theta) = 2r + \cos(\theta)$  in cartesian  $(x, y)$ -coordinates? Well, it's just

$$f(r(x, y), \theta(x, y)),$$

but the function  $\theta(x, y) = \text{atan2}(y, x)$  is pretty annoying to work with, and I'm not really going to have fun simplifying the expression  $\cos(\text{atan2}(y, x))$ . One way around this is to notice that  $\cos(\theta)$  is closely related to  $x$ . We have  $x = r \cos(\theta)$ , so  $\cos(\theta) = x/r = \frac{x}{\sqrt{x^2+y^2}}$ . Now we can write the answer:

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

Let's try a path now. Consider the path  $(x(t), y(t)) = (t^2, t^3)$ . The same path in polar coordinates is given by

$$\begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} r(x(t), y(t)) \\ \theta(x(t), y(t)) \end{bmatrix} = \begin{bmatrix} \sqrt{x(t)^2 + y(t)^2} \\ \text{atan2}(y(t), x(t)) \end{bmatrix} = \begin{bmatrix} \sqrt{t^4 + t^6} \\ \text{atan2}(t^3, t^2) \end{bmatrix}.$$

Now for an example going the other way. Consider the path  $(r(t), \theta(t)) = (3, t)$ , which is given to us in polar coordinates. It's easy to see what this path is doing. It's staying in the circle  $r = 3$  and its  $\theta$ -coordinate is increasing steadily with time. So it's just running counterclockwise around the circle of radius 3. Now convert it to cartesian coordinates:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x(r(t), \theta(t)) \\ y(r(t), \theta(t)) \end{bmatrix} = \begin{bmatrix} r(t) \cos(\theta(t)) \\ r(t) \sin(\theta(t)) \end{bmatrix} = \begin{bmatrix} 3 \cos(t) \\ 3 \sin(t) \end{bmatrix}.$$

And there is our familiar parameterized circle of radius 3, in cartesian coordinates.

**The Composite  $f(\vec{p}(t))$**  What happens to a composite  $f(\vec{p}(t))$  of a function and a path when we change coordinates? For example, suppose that  $f(x, y)$  and  $\vec{p}(t) = (x(t), y(t))$  are both given to us in cartesian coordinates. What happens if we convert them both to polar coordinates, and *then* consider the composite  $f(\vec{p}(t))$ ? The answer is: nothing! It's the same! To see why, think about the coordinate conversion process. For a path, converting it to new coordinates is a matter of applying a coordinate transformation *after* applying the path function:

$$t \text{ (parameter)} \xrightarrow{\text{path function}} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ (point in old coordinates)} \\ \xrightarrow{\text{coord. transf.}} \begin{bmatrix} u(x(t), y(t)) \\ v(x(t), y(t)) \end{bmatrix} \text{ (point in new coordinates).}$$

For a function, converting it to new coordinates is a matter of applying an *inverse* coordinate transformation *before* applying the function:

$$\begin{bmatrix} u \\ v \end{bmatrix} \text{ (point in new coordinates)} \xrightarrow{\text{inverse coord. transf.}} \begin{bmatrix} x(u, v) \\ y(u, v) \end{bmatrix} \text{ (point in old coordinates)} \\ \xrightarrow{\text{function } f} f \left( \begin{bmatrix} x(u, v) \\ y(u, v) \end{bmatrix} \right) \text{ (number output).}$$

Now chain these together, as we do when we form the composite of a function and a path:

$$t \text{ (parameter)} \xrightarrow{\text{path function}} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ (point in old coordinates)} \\ \xrightarrow{\text{coord. transf.}} \begin{bmatrix} u(x(t), y(t)) \\ v(x(t), y(t)) \end{bmatrix} \text{ (point in new coordinates)} \\ \xrightarrow{\text{inverse coord. transf.}} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ (back to same point in old coordinates)} \\ \xrightarrow{\text{function } f} f \left( \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \right) \text{ (number output).}$$

The coordinate transformation and inverse coordinate transformation appear in the middle and cancel each other's effects. The result is the same as if we had just composed the function and path without changing coordinates:

$$t \text{ (parameter)} \xrightarrow{\text{path function}} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ (point)} \\ \xrightarrow{\text{function } f} f \left( \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \right) \text{ (number output).}$$

While our expressions of functions and paths may depend on the coordinates in which we choose to write them, the composite of a function with a path is a truly *coordinate independent* construction.

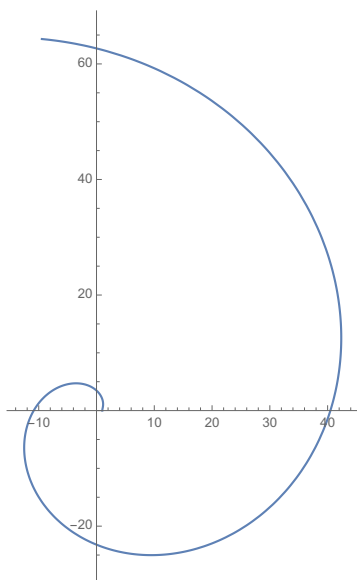
## 2.1.4 Coordinate Bases

Our goal in this section is to figure out how to express the velocity of a path in terms of new coordinates. This will lead us to the description of a *coordinate basis*, which is, at each point, a special vector space basis that somehow lines up with a given coordinate system.

Let's start with an example path and look at a description of its velocity in both cartesian coordinates and polar coordinates. I will describe the example path in polar coordinates:

$$\vec{p}(t) = \begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} t^2 + 1 \\ t \end{bmatrix}.$$

It's pretty easy to think about what this path does. Here is a plot running from  $t = 0$  to  $t = 8$ :



Let's convert it to cartesian coordinates so that we have both coordinate expressions of the path in hand:

$$\vec{p}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} (t^2 + 1) \cos(t) \\ (t^2 + 1) \sin(t) \end{bmatrix}.$$

Now, we know how to obtain and think about the velocity vector in cartesian coordinates:

$$\vec{p}'(t) = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} 2t \cos(t) - (t^2 + 1) \sin(t) \\ 2t \sin(t) + (t^2 + 1) \cos(t) \end{bmatrix}.$$

What if we took the derivative without first transforming to cartesian coordinates? That is, what if we took the derivatives of the components in *polar* coordinates? That should give some kind of "polar version" of the velocity information right? We get

$$\begin{bmatrix} r'(t) \\ \theta'(t) \end{bmatrix} = \begin{bmatrix} 2t \\ 1 \end{bmatrix}.$$

This is somehow also the velocity vector of  $\vec{p}(t)$ . One might even write

$$\vec{p}'(t) \text{ " = " } \begin{bmatrix} 2t \cos(t) - (t^2 + 1) \sin(t) \\ 2t \sin(t) + (t^2 + 1) \cos(t) \end{bmatrix} \text{ " = " } \begin{bmatrix} 2t \\ 1 \end{bmatrix}.$$

But that should make you cringe a little. Certainly those column vectors are not equal in the sense that their components match up. They both represent the same underlying object  $\vec{p}'(t)$ , but in



different coordinates. We could decorate the column vectors with something to fix this horrible notation:

$$\vec{p}'(t) = \begin{bmatrix} 2t \cos(t) - (t^2 + 1) \sin(t) \\ 2t \sin(t) + (t^2 + 1) \cos(t) \end{bmatrix}_{\text{cartesian}} = \begin{bmatrix} 2t \\ 1 \end{bmatrix}_{\text{polar}}.$$

That works, but there's a much nicer solution. I propose the following notation: Use  $\partial_x$  and  $\partial_y$  to respectively denote the vectors  $(1, 0)$  and  $(0, 1)$  when those vectors are to be interpreted as velocity vectors in *cartesian* coordinates. And use  $\partial_r$  and  $\partial_\theta$  to respectively denote the vectors  $(1, 0)$  and  $(0, 1)$  when those vectors are to be interpreted as velocity vectors in *polar* coordinates. In other words,

$$\begin{aligned} \partial_r &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\text{polar}}, & \partial_\theta &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\text{polar}} \\ \partial_x &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\text{cartesian}}, & \partial_y &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\text{cartesian}} \end{aligned}$$

Now we can write

$$\begin{aligned} \vec{p}'(t) &= (2t \cos(t) - (t^2 + 1) \sin(t)) \partial_x + (2t \sin(t) + (t^2 + 1) \cos(t)) \partial_y \\ &= 2t \partial_r + \partial_\theta. \end{aligned}$$

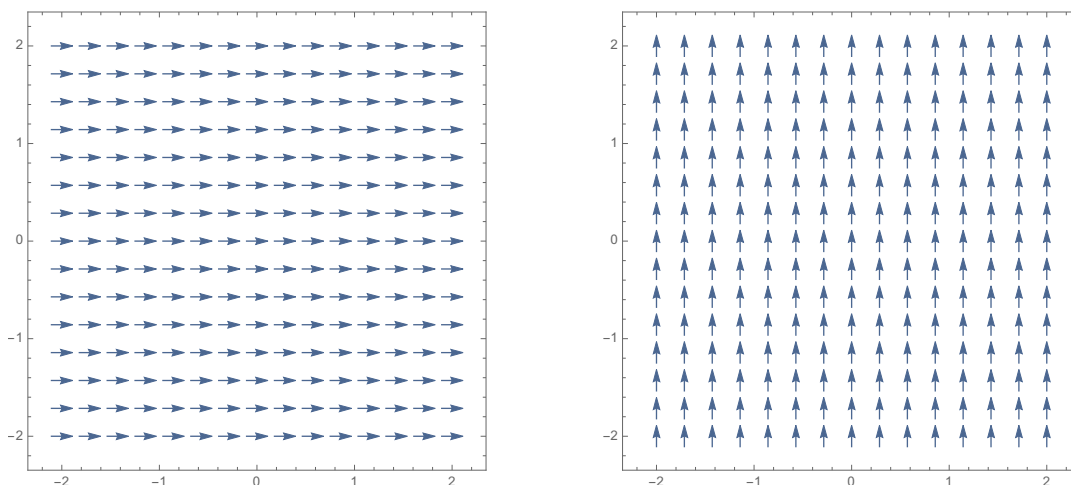
Much better.

We can make one more improvement to the precision of the notation. Remember from section 1.5.3 that  $\vec{p}'(t)$  is not just any old column vector, but specifically a vector-at- $\vec{p}(t)$ . So what did it mean when I said  $\partial_x$  denotes the vector  $(1, 0)$  interpreted as a velocity *vector* in cartesian coordinates? A velocity vector *where*? The answer is: *Everywhere*, because actually  $\partial_x$  is a vector *field*. At any point  $\vec{q}$  in  $\mathbb{R}^n$ , the value of  $\partial_x$  at  $\vec{q}$  is the vector  $(1, 0)$  interpreted as a velocity vector-at- $\vec{q}$  in cartesian coordinates. The same goes for  $\partial_y$ ,  $\partial_r$ , and  $\partial_\theta$ . So really we should write

$$\begin{aligned} \vec{p}'(t) &= (2t \cos(t) - (t^2 + 1) \sin(t)) \partial_x(\vec{p}(t)) + (2t \sin(t) + (t^2 + 1) \cos(t)) \partial_y(\vec{p}(t)) \\ &= 2t \partial_r(\vec{p}(t)) + \partial_\theta(\vec{p}(t)), \end{aligned}$$

but that is usually too much effort and we would just stick to the bare  $\partial_x$ ,  $\partial_y$ ,  $\partial_r$ , and  $\partial_\theta$ , letting context explain what is meant.

We already know how to plot the vector fields  $\partial_x$  and  $\partial_y$ :



You might have a guess as to how the plots of  $\partial_r$  and  $\partial_\theta$  would look, but hold that thought for now. It will become clear soon enough.

So far we haven't done much except write things in a certain way. But it's quite a bit of progress in this subject where notation is often half the battle. Let's make the notation official for arbitrary coordinate systems.

### Notation for Coordinate Vector Fields

Given a coordinate system  $u_1, u_2, \dots, u_n$ , we denote by  $\partial_{u_i}$  the vector field whose value at a point  $\vec{q}$  is the vector-at- $\vec{q}$  which has the following description if it is interpreted as a velocity vector of a path that was expressed in the  $u_1, u_2, \dots, u_n$  coordinate system: it has 0 for all its components except the  $i^{\text{th}}$  component, which is 1. We refer to  $\partial_{u_1}, \partial_{u_2}, \dots, \partial_{u_n}$  as the *coordinate basis vectors* for the coordinate system  $u_1, u_2, \dots, u_n$ . (Really we should call them coordinate basis vector *fields*.)

For example,  $\partial_{u_1}$  is defined to be

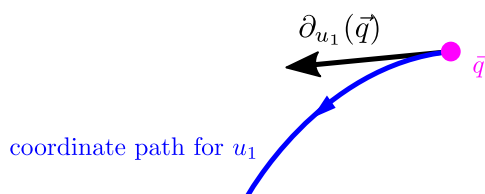
$$\partial_{u_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{\text{the } u_1, \dots, u_n \text{ coordinate system}}$$

in the notation we were temporarily using above.

**In Terms of Coordinate Paths** We can “see” exactly which vector  $\partial_{u_1}$  is at each point by using coordinate paths. If we can find a path  $\vec{p}(t)$  that starts at the point  $\vec{q}$  and that starts with velocity  $(1, 0, \dots, 0)$  (when the velocity is described in  $u_i$  coordinates), then that should show us how  $\partial_{u_1}$  looks (because it would be the vector tangent to that path at  $\vec{q}$  and pointing in its direction of motion). It's very easy to form such a path in  $u_i$  coordinates. Assuming that  $\vec{q}$  has  $u_i$  coordinates of  $(u_1^0, u_2^0, \dots, u_n^0)$ , the following path works:

$$\vec{p}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_n(t) \end{bmatrix} = \begin{bmatrix} u_1^0 + t \\ u_2^0 \\ \vdots \\ u_n^0 \end{bmatrix}.$$

Take a  $t$  derivative and observe that you indeed get  $(1, 0, \dots, 0)$ , which is now also known as  $\partial_{u_1}(\vec{q})$ , or just  $\partial_{u_1}$ . This path is the coordinate path for  $u_1$ , starting at  $\vec{q}$ . So  $\partial_{u_1}(\vec{q})$  is the initial velocity of the coordinate path for  $u_1$  that starts at  $\vec{q}$ .

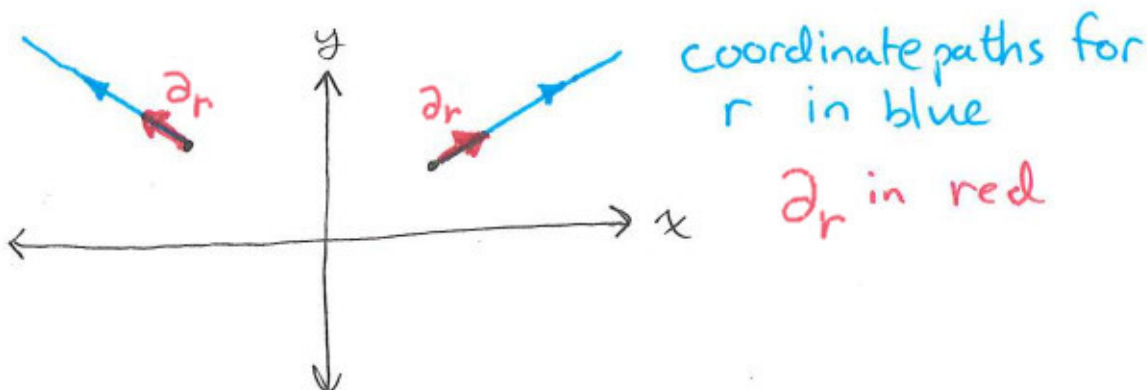


Similarly for  $\partial_{u_i}$ , and worth boxing in:

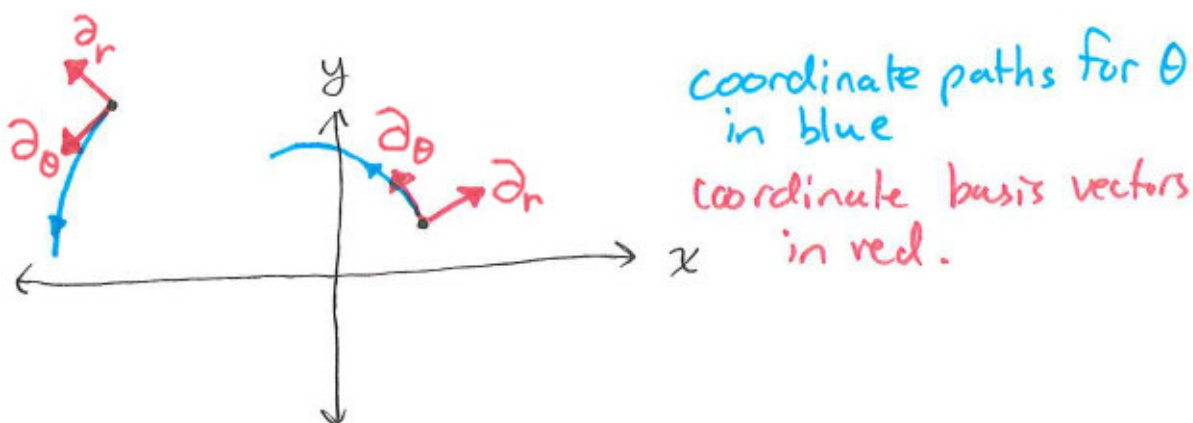
Description of  $\partial_{u_i}$  in terms of coordinate paths

Given a coordinate system  $u_1, \dots, u_n$ , the vector  $\partial_{u_i}(\vec{q})$  is the initial velocity of the coordinate path for  $u_i$  that starts at  $\vec{q}$ .

So in the example of polar coordinates on  $\mathbb{R}^2$ , this gives us a way to look at  $\partial_r$  and  $\partial_\theta$ . The vector  $\partial_r$  is the initial velocity of the coordinate path  $(r(t), \theta(t)) = (r_0 + t, \theta_0)$ , assuming we are looking at  $\partial_r$  at the point that has polar coordinates  $(r_0, \theta_0)$ . Here is a picture at a couple of different choices of  $(r_0, \theta_0)$ :



Note that it looks different in different places—again remember that  $\partial_r$  is really a vector field. As for  $\partial_\theta$ , also at the point that has polar coordinates  $(r_0, \theta_0)$ , it is the initial velocity of the coordinate path  $(r(t), \theta(t)) = (r_0, \theta_0 + t)$ .



Now we can make sense of the velocity in polar coordinates of the example from the beginning of this section. For the path

$$\vec{p}(t) = \begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} t^2 + 1 \\ t \end{bmatrix}$$

we had a velocity of

$$\vec{p}'(t) = 2t \partial_r + \partial_\theta.$$

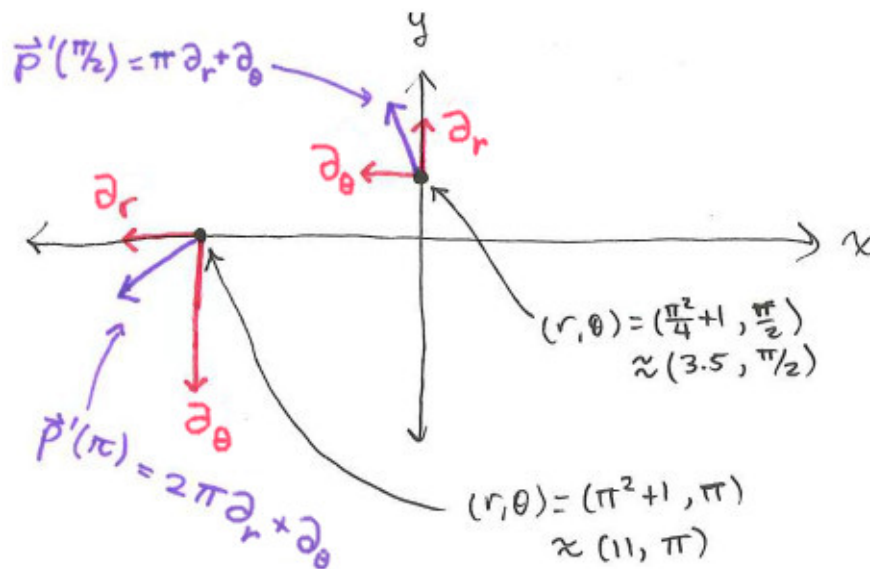
To get a specific vector, let's evaluate at  $t = \pi/2$ . There we have

$$\vec{p}(\pi/2) = \begin{bmatrix} \pi^2/4 + 1 \\ \pi/2 \end{bmatrix} \quad \vec{p}'(\pi/2) = \pi \partial_r + \partial_\theta.$$

And at  $t = \pi$  we have

$$\vec{p}(\pi) = \begin{bmatrix} \pi^2 + 1 \\ \pi \end{bmatrix} \quad \vec{p}'(\pi) = 2\pi \partial_r + \partial_\theta.$$

Given that  $\partial_r$  and  $\partial_\theta$  point along coordinate paths, we can depict these vectors based on their components:



Observe that in the velocity expression  $\vec{p}'(t) = 2t\partial_r + \partial_\theta$ , the time dependence is not just in the “2t” but also in  $\partial_r$  and  $\partial_\theta$ ! It just wasn't as clear because the expression is really  $\vec{p}'(t) = 2t \partial_r(\vec{p}(t)) + \partial_\theta(\vec{p}(t))$  and we are being lazy in how we write it out.

**Transforming Coordinate Basis Vectors** Suppose we want to convert the velocity vector  $\vec{p}'(t) = 2t \partial_r + \partial_\theta$  into cartesian coordinates. Well we already had the answer at the beginning of the section: convert the original path  $\vec{p}(t)$  to cartesian coordinates and then take the velocity of that result to obtain

$$\vec{p}'(t) = (2t \cos(t) - (t^2 + 1) \sin(t)) \partial_x + (2t \sin(t) + (t^2 + 1) \cos(t)) \partial_y.$$

But there ought to be a way to go back and forth between the expressions  $2t \partial_r + \partial_\theta$  and  $(2t \cos(t) - (t^2 + 1) \sin(t)) \partial_x + (2t \sin(t) + (t^2 + 1) \cos(t)) \partial_y$  without having to use the original path. To achieve this, we need to relate  $\partial_r$  and  $\partial_\theta$  to  $\partial_x$  and  $\partial_y$ .

Let's resolve this problem in general, actually, because the answer is really useful. Suppose you have a coordinate system  $u_1, \dots, u_n$  and another coordinate system  $x_1, \dots, x_n$  (it could be cartesian coordinates but doesn't have to be). Then what is  $\partial_{u_1}$  in terms of  $\partial_{x_1}, \partial_{x_2}, \dots, \partial_{x_n}$ ? These are all vector *fields* I'm trying to relate, so let me specify a point  $\vec{u}^0 = (u_1^0, u_2^0, \dots, u_n^0)$  in  $u_i$ -coordinates at which I want to relate them. We can use the coordinate path description of  $\partial_{u_1}$  to calculate:

$$\begin{aligned} \partial_{u_1}(\vec{u}^0) &= \frac{d}{dt} \Big|_{t=0} \begin{bmatrix} u_1 = u_1^0 + t \\ u_2 = u_2^0 \\ \vdots \\ u_n = u_n^0 \end{bmatrix} \\ &= \frac{d}{dt} \Big|_{t=0} \begin{bmatrix} x_1 = x_1(u_1^0 + t, u_2^0, \dots, u_n^0) \\ x_2 = x_2(u_1^0 + t, u_2^0, \dots, u_n^0) \\ \vdots \\ x_n = x_n(u_1^0 + t, u_2^0, \dots, u_n^0) \end{bmatrix} \\ &= \sum_{i=1}^n \frac{d}{dt} \Big|_{t=0} x_i(u_1^0 + t, u_2^0, \dots, u_n^0) \partial_{x_i}(\vec{u}^0) \\ &= \sum_{i=1}^n \frac{\partial x_i}{\partial u_1} \Big|_{\vec{u}^0} \partial_{x_i}(\vec{u}^0). \end{aligned}$$

Removing the evaluation at  $\vec{u}^0$  to get an overall equality of vector fields, we have

$$\partial_{u_1} = \sum_{i=1}^n \frac{\partial x_i}{\partial u_1} \partial_{x_i}.$$

I chose to do the calculation for  $\partial_{u_1}$  to make it easier to write out, but the same can be done for  $\partial_{u_2}, \dots, \partial_{u_n}$ . What we have derived here is the coordinate transformation law for coordinate basis vectors.

#### Transformation Law for Coordinate Basis Vectors

Given coordinate systems  $u_1, \dots, u_n$  and  $x_1, \dots, x_n$ , the following relationship holds among the coordinate basis vector fields:

$$\partial_{u_j} = \sum_{i=1}^n \frac{\partial x_i}{\partial u_j} \partial_{x_i}$$

for all  $j$ . At each point  $\vec{q}$ , the sets  $\partial_{u_1}(\vec{q}), \dots, \partial_{u_n}(\vec{q})$  and  $\partial_{x_1}(\vec{q}), \dots, \partial_{x_n}(\vec{q})$  are bases for the vector space of vectors-at- $\vec{q}$ , and the relationship above is the change-of-basis rule.

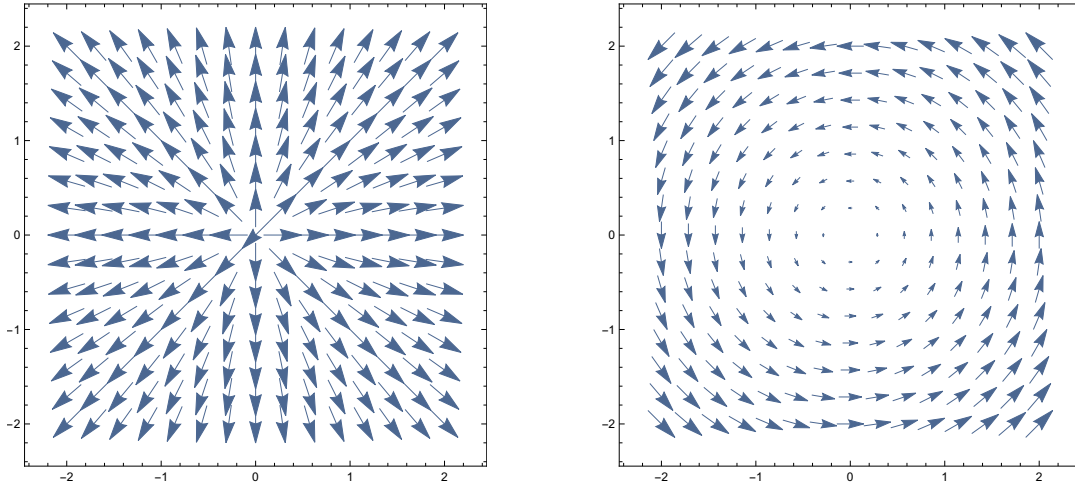
Now we can use this result to relate  $\partial_r$  and  $\partial_\theta$  to  $\partial_x$  and  $\partial_y$ :

$$\begin{aligned} \partial_r &= \frac{\partial x}{\partial r} \partial_x + \frac{\partial y}{\partial r} \partial_y = \frac{\partial}{\partial r} (r \cos(\theta)) \partial_x + \frac{\partial}{\partial r} (r \sin(\theta)) \partial_y = \cos(\theta) \partial_x + \sin(\theta) \partial_y \\ \partial_\theta &= \frac{\partial x}{\partial \theta} \partial_x + \frac{\partial y}{\partial \theta} \partial_y = \frac{\partial}{\partial \theta} (r \cos(\theta)) \partial_x + \frac{\partial}{\partial \theta} (r \sin(\theta)) \partial_y = -r \sin(\theta) \partial_x + r \cos(\theta) \partial_y \end{aligned}$$

One could also describe the coefficient functions of  $\partial_x$  and  $\partial_y$  in cartesian coordinates. We do the conversion (as we learned how to do in section 2.1.3) and get

$$\begin{aligned}\partial_r &= \frac{x}{\sqrt{x^2 + y^2}}\partial_x + \frac{y}{\sqrt{x^2 + y^2}}\partial_y \\ \partial_\theta &= -y\partial_x + x\partial_y\end{aligned}$$

Now we can easily tell a computer to plot the polar coordinate basis vector fields  $\partial_r$  and  $\partial_\theta$ :



As a demonstration, let's use the expressions we got for  $\partial_r$  and  $\partial_\theta$  to relate the two velocity vectors from the example at the beginning of this section:

$$\begin{aligned}\vec{p}'(t) &= 2t\partial_r + \partial_\theta \\ &= 2t(\cos(\theta(t))\partial_x + \sin(\theta(t))\partial_y) + (-r(t)\sin(\theta(t))\partial_x + r(t)\cos(\theta(t))\partial_y) \\ &= 2t(\cos(t)\partial_x + \sin(t)\partial_y) + (-(t^2 + 1)\sin(t)\partial_x + (t^2 + 1)\cos(t)\partial_y) \\ &= (2t\cos(t) - (t^2 + 1)\sin(t))\partial_x + (2t\sin(t) + (t^2 + 1)\cos(t))\partial_y.\end{aligned}$$

It works! This is the same velocity vector we got earlier by re-expressing the entire path in cartesian coordinates and taking a derivative. Note that since this particular path has a simpler polar expression, I used the version of the transformation law that has coefficient functions written in terms of polar coordinates. That is, I made substitutions like  $\partial_\theta = -r\sin(\theta)\partial_x + r\cos(\theta)\partial_y$  rather than  $\partial_\theta = -x\partial_x + y\partial_y$ .

We have expressed  $\partial_r$  and  $\partial_\theta$  in terms of  $\partial_x$  and  $\partial_y$ ; now what about going the other way? Well we could use the coordinate basis vector transformation law again,

$$\begin{aligned}\partial_x &= \frac{\partial r}{\partial x}\partial_r + \frac{\partial \theta}{\partial x}\partial_\theta \\ \partial_y &= \frac{\partial r}{\partial y}\partial_r + \frac{\partial \theta}{\partial y}\partial_\theta,\end{aligned}$$

but that requires thinking about the  $\text{atan2}$  in  $\theta(x, y)$ . Actually it's not that bad, because when you take a derivative it behaves just like  $\arctan(y/x)$ , whose derivative you can look up. An alternative approach is to take our expressions for  $\partial_r$  and  $\partial_\theta$  in terms of  $\partial_x$  and  $\partial_y$  and *solve* for  $\partial_x$  and  $\partial_y$ . I will leave the answer as a homework problem.

**Transforming Vector Components** Before ending this section, I want to write down the transformation law in another way which can be more useful. When we wanted to rewrite  $v_r \partial_r + v_\theta \partial_\theta$  (where we had specific coefficients in the place of  $v_r$  and  $v_\theta$ ), we just replaced  $\partial_r$  and  $\partial_\theta$  by their expressions in terms of  $\partial_x$  and  $\partial_y$ . Then we had four terms that needed to be shuffled around to get something of the form  $v_x \partial_x + v_y \partial_y$ . In the end, we had converted the polar velocity *components*  $(v_r, v_\theta)$  to cartesian velocity *components*  $(v_x, v_y)$ . We can do the shuffling around of those terms *in general*, and then write down a transformation law that tells you how the *components* of velocity change with a change of coordinates.

Suppose you have two coordinate systems  $x_1, \dots, x_n$  and  $u_1, \dots, u_n$ . Suppose you have a vector-at-a-point

$$\vec{v} = \sum_{j=1}^n v_j^{(x)} \partial_{x_j} = \sum_{j=1}^n v_j^{(u)} \partial_{u_j}$$

and you want to know how the coefficients  $v_1^{(x)}, \dots, v_n^{(x)}$  are related to the coefficients  $v_1^{(u)}, \dots, v_n^{(u)}$ . Use the coordinate basis vector transformation law and shuffle things around:

$$\begin{aligned} \vec{v} &= \sum_{j=1}^n v_j^{(u)} \partial_{u_j} \\ &= \sum_{j=1}^n v_j^{(u)} \left( \sum_{i=1}^n \frac{\partial x_i}{\partial u_j} \partial_{x_i} \right) \\ &= \sum_{j=1}^n \sum_{i=1}^n v_j^{(u)} \frac{\partial x_i}{\partial u_j} \partial_{x_i} \\ &= \sum_{i=1}^n \left( \sum_{j=1}^n \frac{\partial x_i}{\partial u_j} v_j^{(u)} \right) \partial_{x_i} \end{aligned}$$

Comparing with the expression

$$\vec{v} = \sum_{i=1}^n v_i^{(x)} \partial_{x_i}$$

and using the fact that the  $\partial_{x_i}$  form a *basis*, we can equate coefficients:

$$v_i^{(x)} = \sum_{j=1}^n \frac{\partial x_i}{\partial u_j} v_j^{(u)}.$$

If you consider the vector components  $v_i^{(x)}$  and  $v_j^{(u)}$  to be the components of the velocity of a path, then what we are writing down is actually just a manifestation of the chain rule! To see how, just write  $v_i^{(x)}$  as  $\frac{dx_i}{dt}$  and write  $v_j^{(u)}$  as  $\frac{du_j}{dt}$ . Then the transformation rule we just wrote down is nothing but the chain rule, where the functions involved are coordinates:

$$\frac{dx_i}{dt} = \sum_{j=1}^n \frac{\partial x_i}{\partial u_j} \frac{du_j}{dt}.$$

Let's box in our result:

## Transformation Law for Vector Components

Assume you have two coordinate systems  $u_1, \dots, u_n$  and  $x_1, \dots, x_n$ , and you have a vector  $\vec{v}$  with components  $v_1^{(x)}, \dots, v_n^{(x)}$  and  $v_1^{(u)}, \dots, v_n^{(u)}$  in the sense that

$$\vec{v} = \sum_{j=1}^n v_j^{(x)} \partial_{x_j} = \sum_{j=1}^n v_j^{(u)} \partial_{u_j}.$$

Then the components are related as follows:

$$v_i^{(x)} = \sum_{j=1}^n \frac{\partial x_i}{\partial u_j} v_j^{(u)}.$$

Here is a demonstration of the principle at work on the example from the beginning of this section:

We had the vector  $2t \partial_r + \partial_\theta$  and we wanted to express it in cartesian coordinates. The components are  $v_r = 2t$  and  $v_\theta = 1$ . We want the components  $v_x$  and  $v_y$  that fit into

$$2t \partial_r + \partial_\theta = v_x \partial_x + v_y \partial_y$$

We also should know where this vector is located: since it arose as the velocity of the path  $(r(t), \theta(t)) = (t^2 + 1, t)$ , it is located at the point  $(r, \theta) = (t^2 + 1, t)$ . The transformation law tells us  $v_x$  and  $v_y$  in terms of  $v_r$  and  $v_\theta$ :

$$\begin{aligned} v_x &= \frac{\partial x}{\partial r} v_r + \frac{\partial x}{\partial \theta} v_\theta \\ &= \cos(\theta) v_r - r \sin(\theta) v_\theta \\ v_y &= \frac{\partial y}{\partial r} v_r + \frac{\partial y}{\partial \theta} v_\theta \\ &= \sin(\theta) v_r + r \cos(\theta) v_\theta. \end{aligned}$$

To apply this general rule to our specific vector  $2t \partial_r + \partial_\theta$ , we put in the components  $v_r = 2t$  and  $v_\theta = 1$  and we also put in the fact that it is a vector-at- $(r = t^2 + 1, \theta = t)$ :

$$\begin{aligned} v_x &= \cos(t)(2t) - (t^2 + 1) \sin(t)(1) = 2t \cos(t) - (t^2 + 1) \sin(t) \\ v_y &= \sin(t)(2t) + (t^2 + 1) \cos(t)(1) = 2t \sin(t) + (t^2 + 1) \cos(t). \end{aligned}$$

This agrees with our previous calculations.

## 2.2 Spaces

When faced with a system that you want to analyze, one of the first steps you take almost every time is to *choose coordinates*. Think about the first sentence that begins a typical analysis:

“Let  $\theta$  denote the angle the pendulum string makes with the vertical line here. Now...”

“Let  $S_1, S_2, \dots, S_n$  be the stock prices...”



“Let  $T$  be the temperature of the gas and let  $P$  be its pressure...”

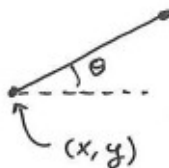
We usually just think of this as “choosing variable names” or just “setting up notation.” But it is often deeper than that— it is a choice of coordinates.

Remember how we defined coordinates? They are functions  $u_1, u_2, \dots, u_n$  on a part of  $\mathbb{R}^n$  such that each point  $\vec{x}$  gets a unique name  $(u_1(\vec{x}), u_2(\vec{x}), \dots, u_n(\vec{x}))$ . Choosing variable names in the setup of a problem often achieves the same thing, except instead of giving unique names to points in  $\mathbb{R}^n$  it ends up giving names to the various possible *states of a system*. This is a more general notion of coordinates than just having coordinates on  $\mathbb{R}^n$ .

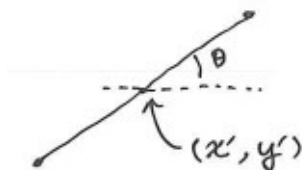
### 2.2.1 Examples of Coordinates on Abstract Spaces

In this section, we explore the more general notion of coordinates by looking at some examples.

**A stick in the plane.** Suppose you are interested in modeling the motion of a stick of fixed length  $\ell$  flying around in a two dimensional plane. Maybe you want to introduce some influences such as gravity on the motion of the stick, but before you can even state laws that govern the stick’s motion you need some mathematical way to represent the possible states of the stick. A natural thing to do is to use  $(x, y)$  to represent the position of one end of the stick in the plane, and then use an angle  $\theta$  to somehow represent the orientation of the stick.



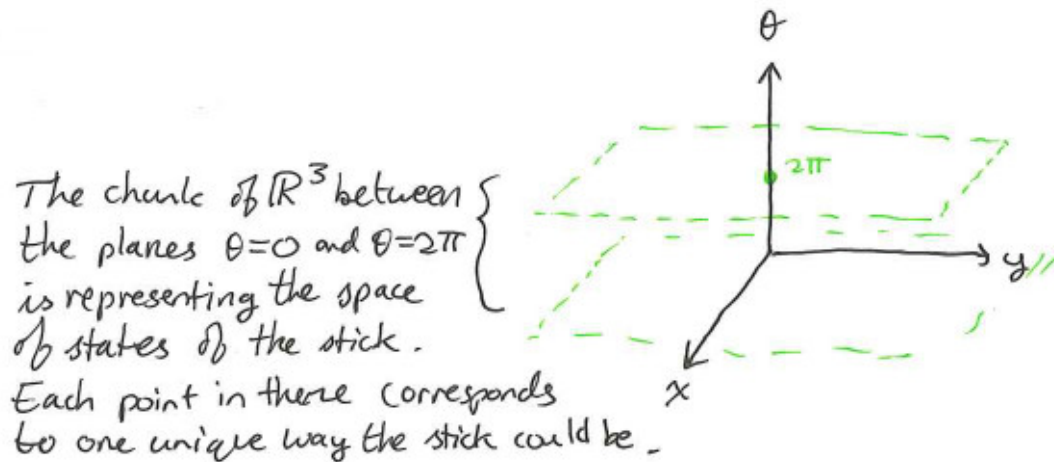
What this amounts to is a choice of coordinates  $x, y, \theta$  on the space of stick configurations. Since exactly three independent coordinates are needed to specify a unique configuration of the stick, we say that the space of stick configurations is *three dimensional*. The coordinates  $x, y, \theta$  really are functions, but on something abstract rather than on euclidean space— they are functions on the space of stick configurations. The function  $x$ , for example, takes as an input a stick configuration— not a number, not a list of numbers, not a matrix— it takes a stick configuration and it outputs the  $x$  coordinate of that stick configuration, which is the cartesian  $x$  coordinate of one of the endpoints of the stick. There were other ways to choose coordinates, for example you could track the midpoint of the stick rather than one end of it:



And there is some coordinate transformation that you could work out relating the two systems:

$$\begin{aligned}x' &= x + \frac{\ell}{2} \cos(\theta) \\y' &= y + \frac{\ell}{2} \sin(\theta) \\ \theta &= \theta.\end{aligned}$$

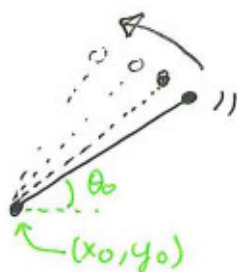
The coordinate systems  $x, y, \theta$  and  $x', y', \theta$  are giving us ways of treating the space of stick configurations like a chunk of  $\mathbb{R}^3$ .



With coordinates, we can describe functions and paths in the space of stick configurations. Let's look at some coordinate paths for fun. The  $\theta$  coordinate path in the  $x, y, \theta$  system starting from some configuration  $(x_0, y_0, \theta_0)$  is the path

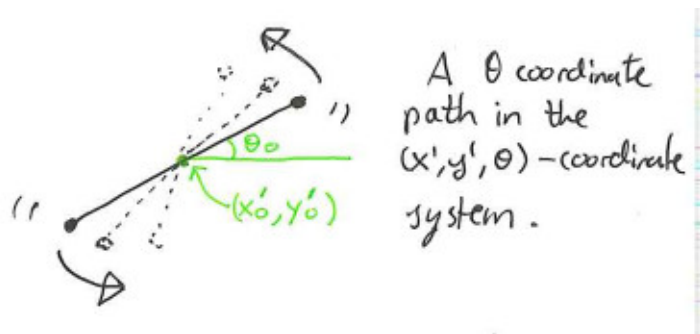
$$\begin{bmatrix} x(t) \\ y(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ \theta_0 + t \end{bmatrix}.$$

This is a path through the space of stick configurations, so if we think of  $t$  as time then it represents a way for the stick to behave over time. We see the stick spinning around with one endpoint in place.



A  $\theta$  coordinate path in the  $(x, y, \theta)$ -coordinate system.

Note that this looks different from a  $\theta$  coordinate path in the  $x', y', \theta$  system.

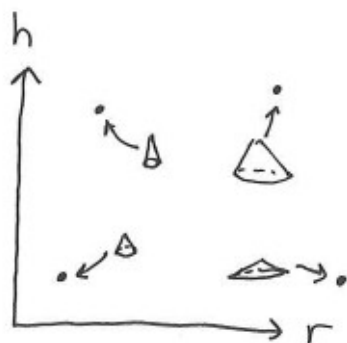


This difference is due to the fact that coordinate paths are all about varying one coordinate while holding the other coordinates fixed. What that looks like depends not only on the coordinate being varied, but also on what the other coordinates are that get fixed.

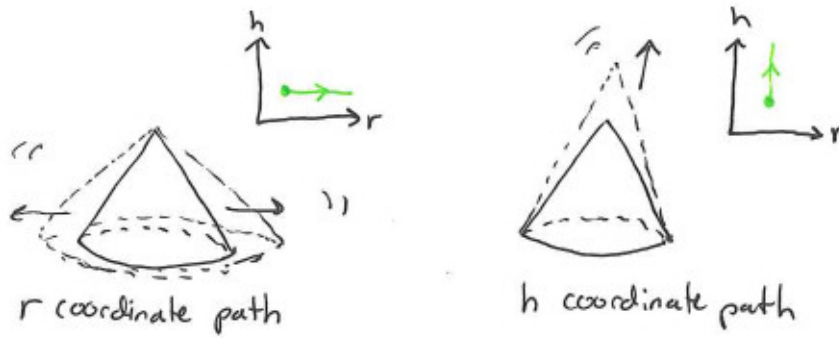
**The space of right circular cones.** A right circular cone can be given by specifying its base radius  $r$  and its height  $h$ .



This immediately imposes a coordinate system  $r, h$  on the space of all possible right circular cones. It makes the space look like a chunk of  $\mathbb{R}^2$ .

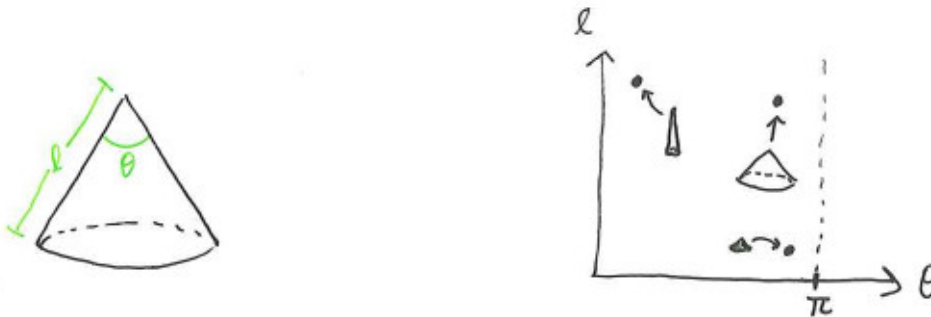


The coordinates  $r$  and  $h$  are functions on the space of cones. Input a cone into  $r$ , and it will output for you the base radius of the cone. Since exactly two independent coordinates  $r$  and  $h$  are needed to uniquely specify a cone, we say that the space of cones is *two dimensional*. It's fun to think about coordinate paths:



Here's an example of a function on the space of cones: volume. Volume  $V$  is a function for which the input would be a cone (a point in the space of cones) and for which the output would be its numerical volume. In the  $r, h$  coordinate system, the volume function is given by  $V(r, h) = \frac{1}{3}\pi r^2 h$ . It would look different in different coordinates.

You could describe a cone by its opening angle  $\theta$  and the length of its side  $l$ :

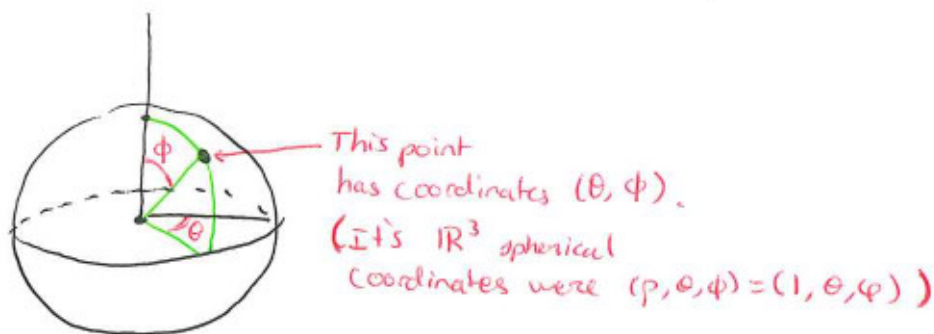


This gives some different coordinates on the space of cones, which you will play with in the homework.

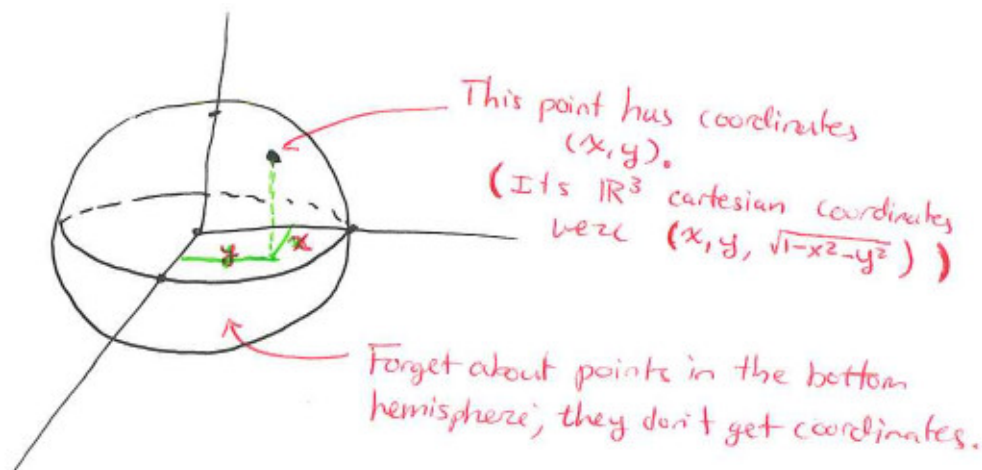
**The space of 20 pixel by 20 pixel grayscale images.** A particular such image can be represented by giving the grayscale value of each of the 400 pixels. Let's say that grayscale value ranges continuously from 0 (white) to 1 (black). This is a choice of coordinates  $x_1, x_2, \dots, x_{400}$ , where the coordinate  $x_i$  stands for the grayscale value of the  $i^{\text{th}}$  pixel (imagine labeling the pixels by the numbers 1 to 400). Since a point in the space of all possible  $20 \times 20$  images can be represented uniquely by exactly 400 independent coordinates  $(x_1, \dots, x_{400})$ , we can say that the space of  $20 \times 20$  images is a *four hundred dimensional space*. That sounds crazy the first time you say it, but it's not really that crazy to think about. A point in four hundred dimensional space is a simple thing to imagine—just imagine one possible  $20 \times 20$  grayscale image. A path in this space is literally an animation! (If you interpret the parameter as time.) But it wouldn't exactly be our usual notion of animation, since each individual pixel value would have to vary smoothly in time. Some fun things to think about: What do the coordinate paths look like? What are the level sets of coordinate functions?

**A sphere.** The points on a sphere form a space on which we can put coordinates. Note that I am not talking about the space of all possible spheres like in the example with cones above. I am talking about a space whose points are the points on one particular sphere. The unit sphere in  $\mathbb{R}^3$  is the set of points  $(x, y, z)$  satisfying  $x^2 + y^2 + z^2 = 1$ . This specifies what the points are, but it does not give coordinates! While it is true that each point on the sphere gets a unique name  $(x, y, z)$ , the  $x$ ,  $y$ , and  $z$  values are not independent. Knowing the  $x$  and  $y$  values of a point on the sphere almost pins down the  $z$  value completely (it leaves only up to two possible  $z$  values), so  $x, y, z$  does not count as a coordinate system for the sphere. Here are two of the many possible ways to put coordinates on a sphere:

The first is to draw inspiration from spherical coordinates, which are a coordinate system on  $\mathbb{R}^3$ . Just forget about  $\rho$  (because it's fixed at 1 throughout the unit sphere anyway) and use  $\theta$  and  $\phi$  to specify points on the sphere.



The second is to just use  $x$  and  $y$  coordinates, and pick one of the two possible  $z$  coordinates. The limitation is that we would only coordinatize half the sphere this way. But it does work for that half at least.



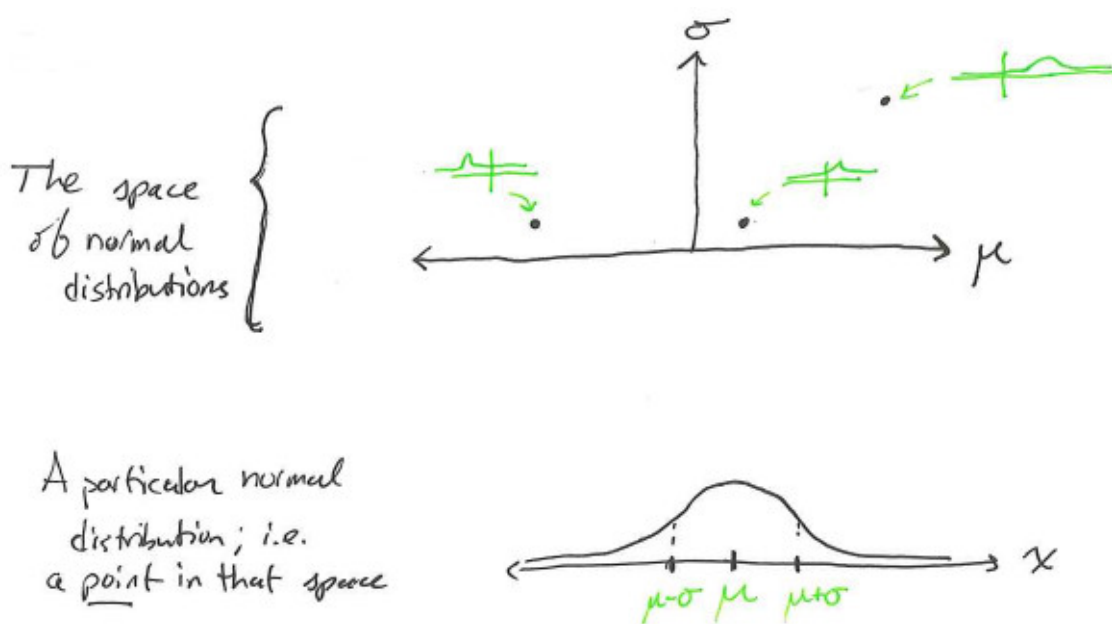
No matter how we set up our coordinate system, one thing will always be true for the sphere: you will need exactly two independent coordinates to uniquely specify a point on the sphere. Thus we

say the sphere is *two dimensional*. It is living inside a three dimensional space, but it is itself two dimensional.

The sphere is one example of a general way to form interesting spaces. The sphere is a certain level set of the function  $f(x, y, z) = x^2 + y^2 + z^2$  on  $\mathbb{R}^3$ . Often it happens that you can form an interesting space by taking a level set of a function and trying to put some coordinates on it.

You may have noticed that the previous examples of interesting spaces were all spaces in which the points were “configurations” of something. Well, trivially you could think of points on a sphere as the possible configurations of... a point on a sphere. More interestingly you could think of a sphere as the space of “possible directions that one could point to in three dimensions.” Or, for a mechanical example, the configurations of a Foucault pendulum can be represented as points of a sphere.

**The space of normal distributions.** In probability theory, a normal distribution is a statistical distribution that has probability density  $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation. In our description of space of all possible normal distributions, we have already imposed some kind of coordinates  $\mu$  and  $\sigma$ .



**Any finite dimensional vector space.** Any vector space of dimension  $n$  has many possible bases, each consisting of  $n$  linearly independent vectors. A particular choice of basis  $e_1, e_2, \dots, e_n$  provides a way to coordinatize the vector space as follows: Any vector  $v$  can be written uniquely as a linear combination of the basis vectors:

$$v = x_1e_1 + \dots + x_n e_n,$$

where  $x_1, \dots, x_n$  are scalars. Let the coordinates of  $v$  then be  $(x_1, \dots, x_n)$ . The coordinates  $x_1, \dots, x_n$  are functions on the vector space, where  $x_i$  assigns to any vector the coefficient that

appears in front of  $e_i$  in the unique expansion of that vector as a linear combination of  $e_1, \dots, e_n$ .

Actually, if you consider  $\mathbb{R}^n$  as a vector space and use the standard basis (the one consisting of vectors that have a zero for each entry except one entry which is 1), then this process yields the “default” cartesian coordinates.

### 2.2.2 Smooth Structure

This section is optional reading, containing material that you are not required to know for the course.

Unlike  $\mathbb{R}^n$  or chunks of  $\mathbb{R}^n$ , the spaces we explored above were *abstract* spaces. There was not really a way to do calculus on them until we put down specific coordinates. This is unlike  $\mathbb{R}^n$ , on which we could immediately start doing calculus because its points are expressed in cartesian coordinates by default.

When we defined coordinate systems on  $\mathbb{R}^n$ , there was something hiding in the definition due to the fact that it involved functions  $u_1, \dots, u_n$ . We agreed earlier that whenever we say “function” in this course, we mean “smooth function.” That does include coordinate functions. But it’s not clear what “smooth” means in many of the abstract examples described above.

Let’s use the space of cones as an example. We gave it coordinates  $r$  and  $h$ . These are functions in that

$$\begin{aligned}r(\text{a cone}) &= \text{the radius of the cone's base} \\h(\text{a cone}) &= \text{the height of the cone.}\end{aligned}$$

But what does it mean for a function to be smooth, or even just differentiable, when its inputs are cones? How does one take a derivative with respect to... hmm... a *cone-valued* variable?

The answer is: we don’t, at least not until coordinates are in place in order to give that idea meaning. For an abstract space that isn’t just some chunk of  $\mathbb{R}^n$ , part of the construction of the space has to include some coordinate systems on it. Instead of requiring that those coordinate systems have smooth coordinate functions, which is initially a meaningless concept, we require that the coordinate *transformations* from one coordinate system to another be smooth functions. Then we *define* smoothness in terms of coordinates: we say that a function on the abstract space is smooth if expressing it in any coordinate system yields an infinitely differentiable function.

So, for example, the volume function on the space of cones is smooth because if you express it in  $r, h$ , coordinates and get  $V(r, h) = \frac{1}{3}\pi r^2 h$ , then *that* is a smooth function in the traditional sense (all its partial derivatives of all orders exist). The  $(r, h)$  coordinate system provides the space of cones with a notion of “smooth function.” The same goes for paths. A path in the space of cones is smooth if, when expressed as  $(r(t), h(t))$ , the functions  $r(t)$  and  $h(t)$  are smooth in the traditional sense.

This framework seems like it’s favoring one coordinate system above all others, but it’s not quite like that. The volume function on the space of cones is also smooth if you look at it in terms of the  $\theta, \ell$

coordinates (the opening angle and side length). That’s because it could be obtained from  $V(r, h)$  by using a coordinate transformation,  $V(r(\theta, \ell), h(\theta, \ell))$ , and because coordinate transformation functions are always required to be smooth.

Requiring coordinate transformation functions to be smooth is a way of ensuring that there is “internal consistency” among coordinate systems with respect to the class of functions each coordinate system considers to be smooth. All coordinate systems that are defined at a given point will either agree or disagree *together* about the smoothness of any given function at the given point.

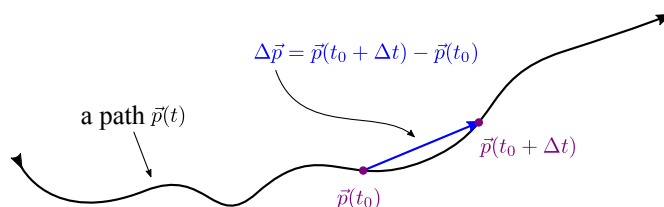
A batch of coordinate systems for which the coordinate transformation functions are smooth is called a *smooth structure*, or an *atlas*. When an abstract space is equipped with a smooth structure, we gain the ability to do calculus on it. That means we can do with the space everything we’ve done so far with  $\mathbb{R}^n$ : we can talk about smooth functions on it, smooth paths in it, and smooth vector fields. We can take velocities of paths, and we can take partial derivatives of functions with respect to various coordinates. Coordinate systems are the bridge that joins concepts to calculations.

### 3 Derivatives

This part of the course will revisit derivatives of functions and paths using some of the fancier tools that we now have. Our exploration in section 1 was very much focused on *cartesian components*. We used cartesian coordinates to break things down, and then mostly just used ordinary calculus to analyze the pieces. In section 1, a path was nothing more than a *list* of ordinary functions of one variable, and a function on  $\mathbb{R}^n$  was nothing more than a function of *multiple variables*. But we realized in section 2 that functions and paths are objects in their own right, which have different expressions in terms of different coordinates. This section will provide a similar treatment for *derivatives* of functions and paths— derivatives will be objects in their own right, with different expressions in different coordinates.

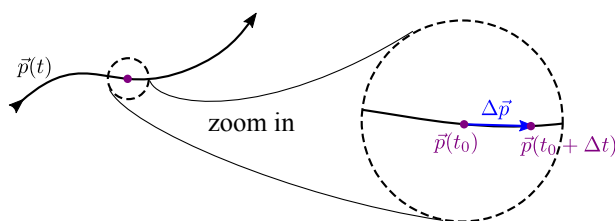
#### 3.1 Speed

This section begins the same way that section 1.5 begins (and in fact I am lazily copying and pasting from there). Consider a path  $\vec{p}(t)$  in  $\mathbb{R}^n$ . It will really help in this discussion to think of  $t$  as *time*. Let’s focus on a particular  $t = t_0$  and some later  $t = t_0 + \Delta t$ . The vector difference  $\Delta\vec{p} = \vec{p}(t_0 + \Delta t) - \vec{p}(t_0)$  is a vector that represents the step from  $\vec{p}(t_0)$  to  $\vec{p}(t_0 + \Delta t)$ .



Now imagine that  $\Delta t$  is very small, so small that the path looks straight.





The vector  $\Delta \vec{p}$  points pretty much in the direction that the path  $\vec{p}(t)$  is traveling when  $t = t_0$ , and its length is the distance traversed during the time interval  $[t_0, t_0 + \Delta t]$ . (End lazy-copy-paste.)

So the distance traversed during the time interval  $[t_0, t_0 + \Delta t]$  is  $\|\Delta \vec{p}\|$ . If we want to know *how fast* the path  $\vec{p}(t)$  happens to be traveling at time  $t = t_0$ , then we are interested in the limit of the ratio  $\frac{\text{distance traversed}}{\text{size of time interval}}$  in the limit as the time interval becomes arbitrarily short. In other words, the *speed* of the path at  $t = t_0$  is given by

$$\lim_{\Delta t \rightarrow 0} \frac{\|\Delta \vec{p}\|}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\|\vec{p}(t_0 + \Delta t) - \vec{p}(t_0)\|}{\Delta t}.$$

A scalar can always be pushed into a norm, so this is the same as

$$\lim_{\Delta t \rightarrow 0} \frac{\|\Delta \vec{p}\|}{\Delta t} = \lim_{\Delta t \rightarrow 0} \left\| \frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0)) \right\|.$$

Since taking norm is a continuous operation (it's just a square root with a bunch of adding and squaring), the limit can be pushed inside the norm:

$$\lim_{\Delta t \rightarrow 0} \frac{\|\Delta \vec{p}\|}{\Delta t} = \left\| \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{p}(t_0 + \Delta t) - \vec{p}(t_0)) \right\|.$$

What we are seeing inside the norm is exactly the thing we defined to be the velocity vector  $\vec{p}'(t_0)$ ! So

$$\lim_{\Delta t \rightarrow 0} \frac{\|\Delta \vec{p}\|}{\Delta t} = \|\vec{p}'(t_0)\|.$$

We just discovered that the speed of a path is the norm of its velocity.

Velocity is a vector-at-a-point, so what does norm actually mean in this situation? It's simple if  $\vec{p}(t)$  is expressed in *cartesian* coordinates. If the path  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  is in cartesian coordinates, then its velocity is simply  $(x'_1(t), x'_2(t), \dots, x'_n(t))$ , which we now know to write as  $x'_1(t)\partial_{x_1} + x'_2(t)\partial_{x_2} + \dots + x'_n(t)\partial_{x_n}$ . The norm of this is  $\sqrt{(x'_1(t))^2 + (x'_2(t))^2 + \dots + (x'_n(t))^2}$ .

This is probably all as you expected, but I want it to be clear that things only look like this because we are using cartesian coordinates. The argument we made above required us to consider the length of the displacement vector  $\Delta \vec{p}$ . It is due to the pythagorean theorem that taking *the square root of the sum of the squares of the components* of  $\Delta \vec{p}$  gives the *length* of  $\Delta \vec{p}$ . But the pythagorean theorem only applies in cartesian coordinates! (Think about it. If you have two points in the plane with polar coordinate differences  $\Delta r$  and  $\Delta \theta$ , then is the squared distance between them  $(\Delta r)^2 + (\Delta \theta)^2$ ? Certainly not!)

Summarizing:

- It is always true that speed is the norm of velocity.
- It is not always true that the norm of velocity is simply the square root of the sum of the squares of its components. This only works on *cartesian* components; that is, it only works if the velocity vector was expressed in cartesian coordinates.

### Speed of a Path

The speed of a path is the norm of its velocity. So if  $\vec{p}(t)$  is a path in  $\mathbb{R}^n$ , then its speed is given by  $\|\vec{p}'(t_0)\|$ .

If the path is expressed in cartesian coordinates as  $\vec{p}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ , then its speed is given by

$$\|x'_1(t)\partial_{x_1} + x'_2(t)\partial_{x_2} + \dots + x'_n(t)\partial_{x_n}\| = \sqrt{(x'_1(t))^2 + (x'_2(t))^2 + \dots + (x'_n(t))^2}$$

## 3.2 Norm and Dot Product for Vectors at a Point

In appendix A.2, you saw how dot products and norms work for vectors in  $\mathbb{R}^n$ . I never actually told you how dot products and norms work for vectors-at-a-point. But the previous section made it clear, at least for norms, that vectors-at-a-point should be treated in the same way as vectors in  $\mathbb{R}^n$ ... *as long as they are written in cartesian coordinates*. For example, the norm of a vector  $v_x\partial_x + v_y\partial_y$  is  $\sqrt{v_x^2 + v_y^2}$ . The same should go for dot products. The dot product of  $v_x\partial_x + v_y\partial_y$  and  $w_x\partial_x + w_y\partial_y$  is  $v_xw_x + v_yw_y$ . One way to encode this information is to just write down the dot products of the coordinate basis vectors:

$$\partial_x \cdot \partial_x = 1 \quad \partial_y \cdot \partial_y = 1 \quad \partial_x \cdot \partial_y = 0$$

This information alone is sufficient to compute all dot products (and hence also all norms, since  $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$ ) for vectors-at-points in  $\mathbb{R}^2$ :

$$\begin{aligned} (v_x\partial_x + v_y\partial_y) \cdot (w_x\partial_x + w_y\partial_y) &= v_xw_x\partial_x \cdot \partial_x + v_xw_y\partial_x \cdot \partial_y + v_yw_x\partial_y \cdot \partial_x + v_yw_y\partial_y \cdot \partial_y \\ &= v_xw_x(1) + v_xw_y(0) + v_yw_x(0) + v_yw_y(1) \\ &= v_xw_x + v_yw_y \end{aligned}$$

To write a nice, succinct definition of dot product for vectors-at-points, we will just state the dot products of coordinate basis vectors.

### Dot Product of Vectors in Cartesian Coordinates

If  $x_1, x_2, \dots, x_n$  denote the cartesian coordinates on  $\mathbb{R}^n$ , then:

$$\partial_{x_i} \cdot \partial_{x_j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

This information is enough for you to know how to take dot products (and hence also norms) of all vectors-at-points in  $\mathbb{R}^n$ . It's just a fancy way of saying "compute dot products the same way you did in appendix A.2, but make sure to use cartesian coordinates when you do it." It may seem like overkill right now to phrase things in terms of coordinate basis vectors, but you will come to appreciate it once we look at what happens in non-cartesian coordinates.

### 3.2.1 A Polar Example

In this little section we will take a path that has been written nicely in polar coordinates, and we will ask: *what is its speed?* We already know one way to answer such a question: convert to cartesian coordinates and then use the approach we figured out in section 3.1. But our goal here will be to find a way to arrive at the answer in a "purely polar" way.

Consider the path  $\vec{p}(t)$  in  $\mathbb{R}^2$  that is given in polar coordinates as

$$\begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} r_0 + t \\ \theta_0 + 2t \end{bmatrix}.$$

Here  $r_0$  and  $\theta_0$  are constants, and it's easy to convince yourself that this path is some kind of outward-going spiral starting at  $(r, \theta) = (r_0, \theta_0)$ . If you want to plot it, you might make use of an expression of the path in cartesian coordinates:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} (r_0 + t) \cos(\theta_0 + 2t) \\ (r_0 + t) \sin(\theta_0 + 2t) \end{bmatrix}.$$

The velocity vector is easy to write down in the original polar coordinates:

$$\frac{d\vec{p}}{dt} = \partial_r + 2\partial_\theta.$$

So is the speed of the path  $\sqrt{1^2 + 2^2} = \sqrt{5}$ ? No! The velocity in cartesian coordinates is

$$\frac{d\vec{p}}{dt} = [\cos(\theta_0 + 2t) - 2(r_0 + t) \sin(\theta_0 + 2t)] \partial_x + [\sin(\theta_0 + 2t) + 2(r_0 + t) \cos(\theta_0 + 2t)] \partial_y,$$

which tells us that the true speed of the path is

$$\begin{aligned} \left\| \frac{d\vec{p}}{dt} \right\| &= \sqrt{[\cos(\theta_0 + 2t) - 2(r_0 + t) \sin(\theta_0 + 2t)]^2 + [\sin(\theta_0 + 2t) + 2(r_0 + t) \cos(\theta_0 + 2t)]^2} \\ &= \sqrt{\cos^2(\theta_0 + 2t) + 4(r_0 + t)^2 \sin^2(\theta_0 + 2t) - 4(r_0 + t) \sin(\theta_0 + 2t) \cos(\theta_0 + 2t) +} \\ &\quad \sin^2(\theta_0 + 2t) + 4(r_0 + t)^2 \cos^2(\theta_0 + 2t) + 4(r_0 + t) \sin(\theta_0 + 2t) \cos(\theta_0 + 2t)} \\ &= \sqrt{1 + 4(r_0 + t)^2}. \end{aligned}$$

The square root on the second line continues onto the third line. What a horrendous calculation! All we were computing is the norm  $\left\| \frac{d\vec{p}}{dt} \right\|$ , which is the square root of  $\frac{d\vec{p}}{dt} \cdot \frac{d\vec{p}}{dt}$ . There is a smarter way to break up this calculation before we bring in the cartesian coordinates:

$$\begin{aligned} \frac{d\vec{p}}{dt} \cdot \frac{d\vec{p}}{dt} &= (\partial_r + 2\partial_\theta) \cdot (\partial_r + 2\partial_\theta) \\ &= \partial_r \cdot \partial_r + 4\partial_r \cdot \partial_\theta + 4\partial_\theta \cdot \partial_\theta \end{aligned}$$

All we need to know is those three dot products:  $\partial_r \cdot \partial_r$ ,  $\partial_r \cdot \partial_\theta$ , and  $\partial_\theta \cdot \partial_\theta$ . We can use cartesian coordinates to compute these three dot products once and for all, and then we will gain an awesome ability forever: we will be able to get speeds of paths expressed in polar coordinates without having to do anything with cartesian coordinates!

To compute the three dot products, write everything in terms of cartesian coordinate basis vectors, as we learned to do in section 2.1.4:

$$\begin{aligned}\partial_r &= \frac{\partial x}{\partial r} \partial_x + \frac{\partial y}{\partial r} \partial_y = \cos(\theta) \partial_x + \sin(\theta) \partial_y \\ \partial_\theta &= \frac{\partial x}{\partial \theta} \partial_x + \frac{\partial y}{\partial \theta} \partial_y = -r \sin(\theta) \partial_x + r \cos(\theta) \partial_y.\end{aligned}$$

Then the dot products are readily computable:

$$\begin{aligned}\partial_r \cdot \partial_r &= \cos^2(\theta) + \sin^2(\theta) = 1 \\ \partial_r \cdot \partial_\theta &= -r \sin(\theta) \cos(\theta) + r \sin(\theta) \cos(\theta) = 0 \\ \partial_\theta \cdot \partial_\theta &= r^2 \sin^2(\theta) + r^2 \cos^2(\theta) = r^2.\end{aligned}$$

Now we can compute dot products of vectors written in polar coordinates!

$$(v_r \partial_r + v_\theta \partial_\theta) \cdot (w_r \partial_r + w_\theta \partial_\theta) = v_r w_r + r^2 v_\theta w_\theta.$$

That  $r^2$  is pretty interesting. If all you knew about your two vectors is that their components in polar coordinates are  $v_r, v_\theta$  and  $w_r, w_\theta$ , then it is *impossible* for you to know their dot product. You must also know *where* these vectors are located, because the value of the  $r$  coordinate at that location figures into the dot product calculation! The way dot product works in polar coordinates depends on where you are looking.

Let's test drive the result on our example spiral path. The velocity of  $\vec{p}(t)$  was  $\partial_r + 2 \partial_\theta$ , and the norm of this is

$$\begin{aligned}\|\partial_r + 2 \partial_\theta\| &= \sqrt{(\partial_r + 2 \partial_\theta) \cdot (\partial_r + 2 \partial_\theta)} \\ &= \sqrt{\partial_r \cdot \partial_r + 4 \partial_r \cdot \partial_\theta + 4 \partial_\theta \cdot \partial_\theta} \\ &= \sqrt{1 + 4r^2}\end{aligned}$$

Remember that when we take the velocity of a path  $\vec{p}(t)$  at a particular time  $t$ , we get a vector-at- $\vec{p}(t)$ . That tells us what to do with the  $r$  that appears at the end of the calculation:

$$\begin{aligned}\left\| \frac{d\vec{p}}{dt} \right\| &= \|\partial_r + 2 \partial_\theta\| \\ &= \sqrt{1 + 4r(t)^2} \\ &= \sqrt{1 + 4(r_0 + t)^2}.\end{aligned}$$

Yay, it's the same speed we got from the horrendous calculation that used cartesian coordinates!

### 3.2.2 Dot Products in Non-Cartesian Coordinates

At this point you are probably seeing that in order to compute dot products and norms in any coordinates  $u_1, u_2, \dots, u_n$ , you just need to know the dot products  $\partial_{u_i} \cdot \partial_{u_j}$  of coordinate basis vectors. To get those, we express coordinate basis vectors in terms of cartesian coordinates,

$$\partial_{u_i} = \sum_{k=1}^n \frac{\partial x_k}{\partial u_i} \partial_{x_k},$$

and then we take a cartesian-style dot product:

$$\begin{aligned} \partial_{u_i} \cdot \partial_{u_j} &= \left( \sum_{k=1}^n \frac{\partial x_k}{\partial u_i} \partial_{x_k} \right) \cdot \left( \sum_{\ell=1}^n \frac{\partial x_\ell}{\partial u_j} \partial_{x_\ell} \right) \\ &= \sum_{k=1}^n \sum_{\ell=1}^n \frac{\partial x_k}{\partial u_i} \frac{\partial x_\ell}{\partial u_j} \partial_{x_k} \cdot \partial_{x_\ell} \\ &= \sum_{k=1}^n \frac{\partial x_k}{\partial u_i} \frac{\partial x_k}{\partial u_j}. \end{aligned}$$

#### Dot Products in Any Coordinate System

Let  $x_1, \dots, x_n$  be a cartesian coordinate system on  $\mathbb{R}^n$ . Given any coordinate system  $u_1, \dots, u_n$ , the dot products of the coordinate basis vectors are as follows:

$$\partial_{u_i} \cdot \partial_{u_j} = \sum_{k=1}^n \frac{\partial x_k}{\partial u_i} \frac{\partial x_k}{\partial u_j}.$$

So, for example, if you wanted to know the dot product of the spherical coordinate basis vectors  $\partial_\rho$  and  $\partial_\theta$ , then it would be

$$\partial_\rho \cdot \partial_\theta = \frac{\partial x}{\partial \rho} \frac{\partial x}{\partial \theta} + \frac{\partial y}{\partial \rho} \frac{\partial y}{\partial \theta} + \frac{\partial z}{\partial \rho} \frac{\partial z}{\partial \theta}.$$

If you wanted to know the norm of  $\partial_\rho$  then this would be

$$\|\partial_\rho\| = \sqrt{\partial_\rho \cdot \partial_\rho} = \sqrt{\frac{\partial x^2}{\partial \rho} + \frac{\partial y^2}{\partial \rho} + \frac{\partial z^2}{\partial \rho}}.$$

### 3.2.3 Polar Example Picture

We saw in section 3.2.1 that the way dot product works for vectors written in polar coordinates depends on *where* you are looking. If you had two vectors whose components in polar coordinates are  $v_r, v_\theta$  and  $w_r, w_\theta$ , then their dot product is

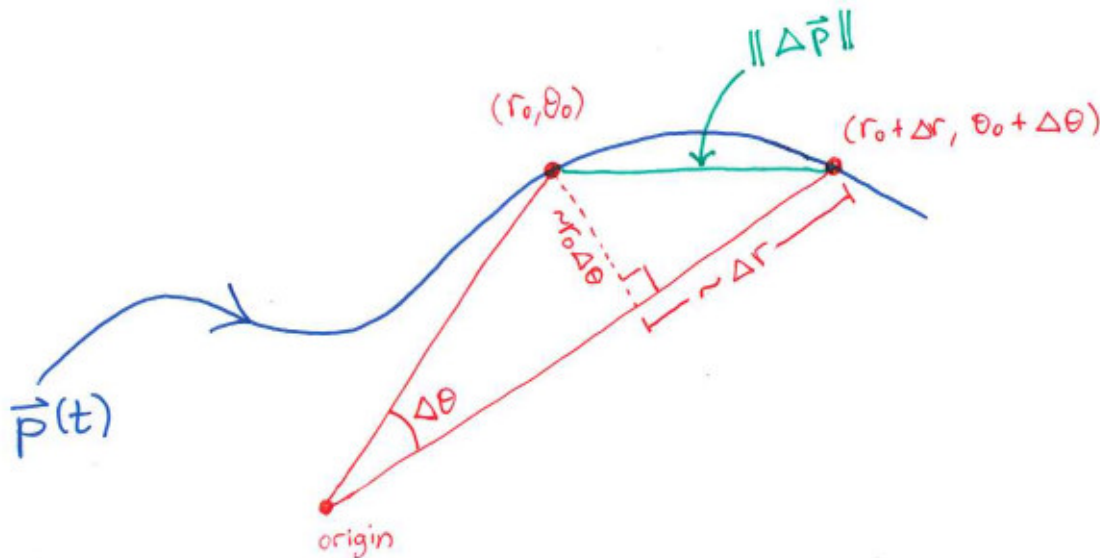
$$(v_r \partial_r + v_\theta \partial_\theta) \cdot (w_r \partial_r + w_\theta \partial_\theta) = v_r w_r + r^2 v_\theta w_\theta.$$

You need to fill in the value of the  $r$  coordinate for *where those vectors are located*.

It follows that the expression for *norm* in terms of components also depends on where you are looking, as you can see by the appearance of  $r$  in the expression:

$$\|v_r \partial_r + v_\theta \partial_\theta\| = \sqrt{(v_r \partial_r + v_\theta \partial_\theta) \cdot (v_r \partial_r + v_\theta \partial_\theta)} = \sqrt{v_r^2 + r^2 v_\theta^2}.$$

There is a nice pictorial explanation for the  $r$  there. Go back to the definition of speed as  $\frac{\text{distance traversed}}{\text{size of time interval}}$  in the limit as the time interval becomes arbitrarily short, and then try to use polar coordinates to express the distance traversed. If the path  $\vec{p}(t)$  is located at  $(r_0, \theta_0)$  at a certain time, and it travels to  $(r_0 + \Delta r, \theta_0 + \Delta\theta)$  after some time  $\Delta t$ , then what is the distance  $\|\Delta\vec{p}\|$  that it has traversed? Draw the appropriate picture to see where to apply the pythagorean theorem to answer this:



Convince yourself that the legs of the right triangle in which  $\|\Delta\vec{p}\|$  is a hypotenuse have approximate lengths of  $\Delta r$  and  $r_0\Delta\theta$  as displayed (with the approximation becoming exact in the limit  $\Delta t \rightarrow 0$ ). Then  $\|\Delta\vec{p}\|$  is approximately  $\sqrt{(\Delta r)^2 + (r_0\Delta\theta)^2}$ , and the speed at which the path passes by  $(r_0, \theta_0)$  is

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{\|\Delta\vec{p}\|}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \frac{\sqrt{(\Delta r)^2 + (r_0\Delta\theta)^2}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \sqrt{\left(\frac{\Delta r}{\Delta t}\right)^2 + r_0^2 \left(\frac{\Delta\theta}{\Delta t}\right)^2} \\ &= \sqrt{\left(\frac{dr}{dt}\right)^2 + r_0^2 \left(\frac{d\theta}{dt}\right)^2}. \end{aligned}$$

See the  $r_0$  appearing in there? You need to know the actual *position* of the path when it has velocity  $\frac{dr}{dt} \partial_r + \frac{d\theta}{dt} \partial_\theta$  in order to know its speed.

### 3.3 The Gradient $\vec{\nabla} f$

Suggestion: review sections 1.3, 1.5.2, and 1.7.3 before reading this one.

Let  $f(\vec{x})$  denote a function on  $\mathbb{R}^n$ , expressed in cartesian coordinates. Given a path  $\vec{p}(t) = \vec{x}(t)$  in cartesian coordinates, the chain rule gives us a way to express the rate of change of  $f(\vec{x})$  along  $\vec{x}(t)$ :

$$\frac{d}{dt}f(\vec{x}(t)) = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dt}.$$

It's a sum of products, built out of two lists of things: the partials of  $f$  and the components of the velocity of  $\vec{x}(t)$ . It's a dot product!

$$\frac{d}{dt}f(\vec{x}(t)) = \left( \frac{\partial f}{\partial x_1} \partial_{x_1} + \cdots + \frac{\partial f}{\partial x_n} \partial_{x_n} \right) \cdot \left( \frac{dx_1}{dt} \partial_{x_1} + \cdots + \frac{dx_n}{dt} \partial_{x_n} \right)$$

I used coordinate basis vectors in the notation because we are already familiar with the vector on right side: it is just the velocity vector  $\frac{d\vec{x}}{dt}$  of the path. But what is the vector on the left side? It is a new character in our story, and it deserves a name. We denote it by  $\vec{\nabla} f$  and call it the *gradient* of  $f$ :

$$\vec{\nabla} f = \frac{\partial f}{\partial x_1} \partial_{x_1} + \cdots + \frac{\partial f}{\partial x_n} \partial_{x_n}.$$

The coefficients of the  $\partial_{x_i}$  are functions on  $\mathbb{R}^n$  and not just scalars, so  $\vec{\nabla} f$  is not a single vector but a vector *field*. You can think of  $\vec{\nabla}$  as an operator that turns a function  $f$  into a vector field  $\vec{\nabla} f$ .

#### Definition of Gradient

Assume that  $f(x_1, \dots, x_n)$  is a function on  $\mathbb{R}^n$  written in cartesian coordinates. Then the gradient  $\vec{\nabla} f$  of  $f$  is defined to be the following vector field:

$$\vec{\nabla} f = \frac{\partial f}{\partial x_1} \partial_{x_1} + \cdots + \frac{\partial f}{\partial x_n} \partial_{x_n}.$$

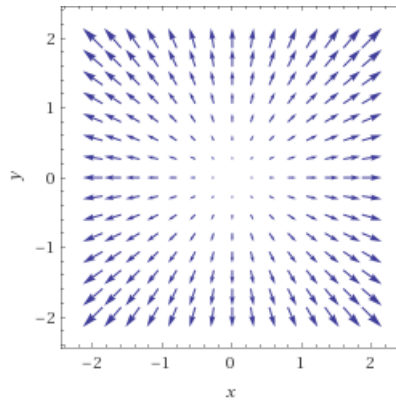
(Cartesian coordinates are important for this to work. Later we will see why.)

At each point  $\vec{x}_0$  in  $\mathbb{R}^n$ , the gradient of  $f$  gives us a specific vector-at- $\vec{x}_0$  known as  $\vec{\nabla} f(\vec{x}_0)$ . What does it mean? In what direction is this vector pointing? And what is the meaning of its norm? The rest of this section will explore these questions.

**Example** The function  $f(x, y) = x^2 + y^2$  has the following gradient vector field:

$$\vec{\nabla} f = \frac{\partial f}{\partial x} \partial_x + \frac{\partial f}{\partial y} \partial_y = 2x \partial_x + 2y \partial_y.$$

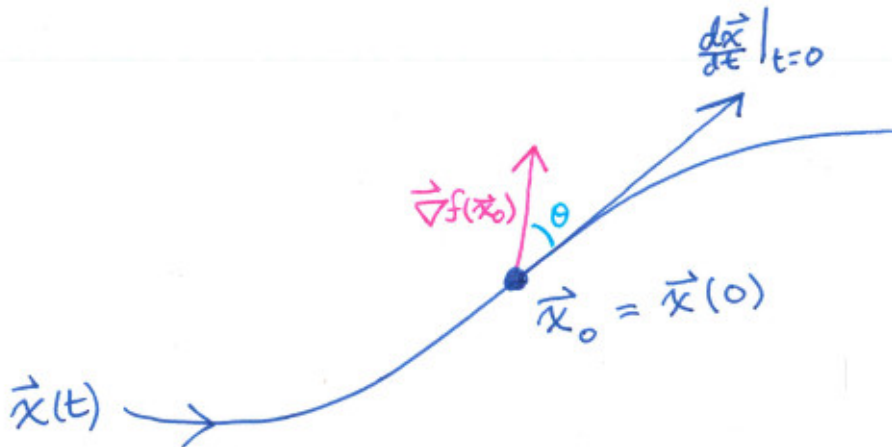
Note that this is a vector field on  $\mathbb{R}^2$ . Here is a plot of  $\vec{\nabla} f$ :



### 3.3.1 Direction of $\vec{\nabla} f$

Let's try to understand the direction of " $\vec{\nabla} f(\vec{x}_0)$ ."

Assume we've got a function  $f(\vec{x})$  on  $\mathbb{R}^n$  and a path  $\vec{p}(t) = \vec{x}(t)$  in  $\mathbb{R}^n$ , both written in cartesian coordinates, and assume that the path passes through the point  $\vec{x}_0$  at  $t = 0$ . Zoom in on the point  $\vec{x}_0$  and focus on the vector space of vectors-at- $\vec{x}_0$ . We currently have two vectors of interest living there, the velocity of  $\vec{x}(t)$  at  $t = 0$  and the gradient of  $f$  at  $\vec{x}_0$ . Let  $\theta$  denote the angle between them.



The gradient vector  $\vec{\nabla} f(\vec{x}_0)$  was defined in the first place because it lets us express the chain rule as a dot product:

$$\left. \frac{d}{dt} f(\vec{x}(t)) \right|_{t=0} = \left( \vec{\nabla} f(\vec{x}_0) \right) \cdot \left( \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right).$$

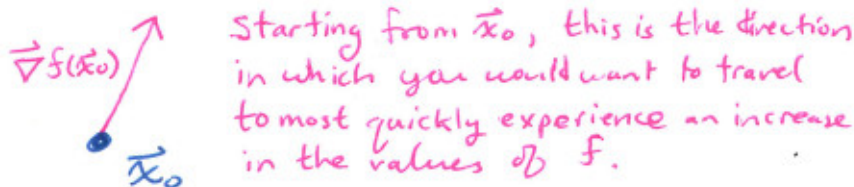
In appendix A.2.3, we reasoned from

$$\left( \vec{\nabla} f(\vec{x}_0) \right) \cdot \left( \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right) = \left\| \vec{\nabla} f(\vec{x}_0) \right\| \left\| \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right\| \cos(\theta)$$

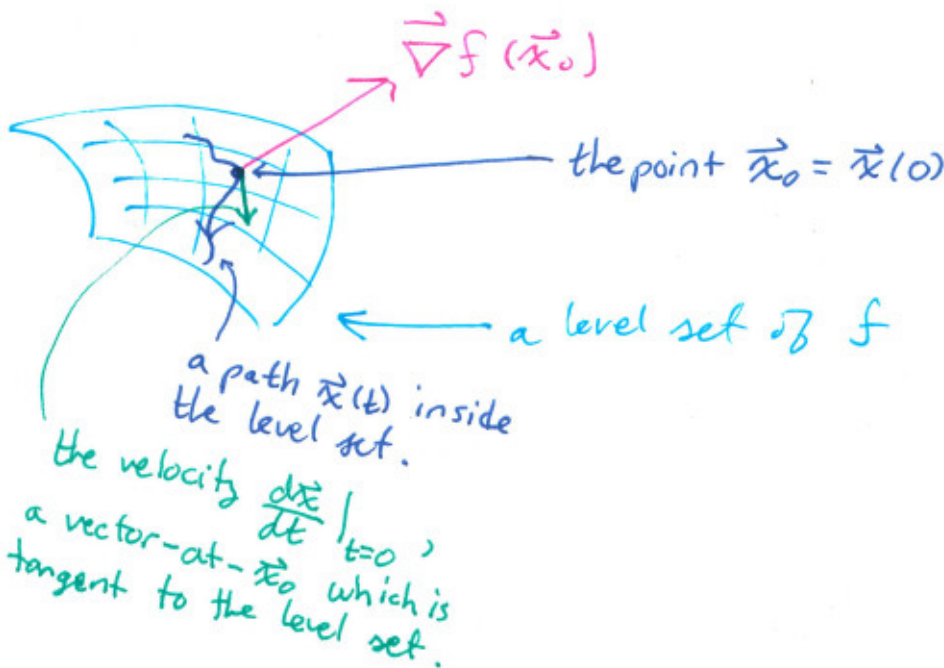
that the dot product is the largest it can be when  $\theta = 0$ , i.e. when the vectors  $\vec{\nabla} f(\vec{x}_0)$  and  $\left. \frac{d\vec{x}}{dt} \right|_{t=0}$  are perfectly aligned in the same direction. So if we fix the speed  $\left\| \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right\|$  at which the path passes



through  $\vec{x}_0$ , the direction in which the path should travel in order to maximize the dot product is the direction of  $\vec{\nabla}f(\vec{x}_0)$ . In other words, of all paths starting at  $\vec{x}_0$  that have a given speed, the ones that head in the  $\vec{\nabla}f(\vec{x}_0)$  direction are the ones for which  $\left. \frac{d}{dt}f(\vec{x}(t)) \right|_{t=0}$  is maximized. This is telling us exactly what the direction of  $\vec{\nabla}f(\vec{x}_0)$  means: *it is the direction along which  $f(\vec{x})$  increases the fastest, starting from  $x_0$ .*



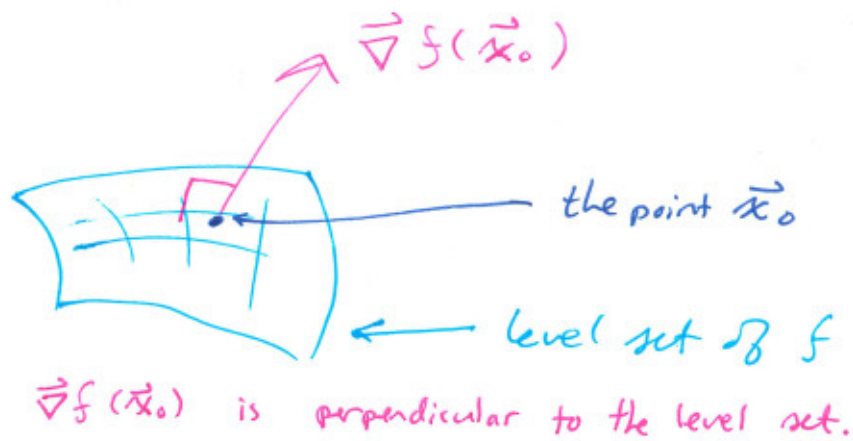
There is a second way in which we can use our understanding of the dot product to describe the directionality of  $\vec{\nabla}f(\vec{x}_0)$ . Suppose that a path  $\vec{x}(t)$  is passing through  $\vec{x}_0$  at  $t = 0$ , but assume also that  $\vec{x}(t)$  stays within the level set of  $f$  that contains  $\vec{x}_0$ . This means that  $f(\vec{x}(t))$  is a constant, and that  $\left. \frac{d\vec{x}}{dt} \right|_{t=0}$  always points in a direction tangent to the level set of  $f$ .



Since  $f(\vec{x}(t))$  is a constant, its derivative  $\left. \frac{d}{dt}f(\vec{x}(t)) \right|_{t=0}$  must be 0. The chain rule then says:

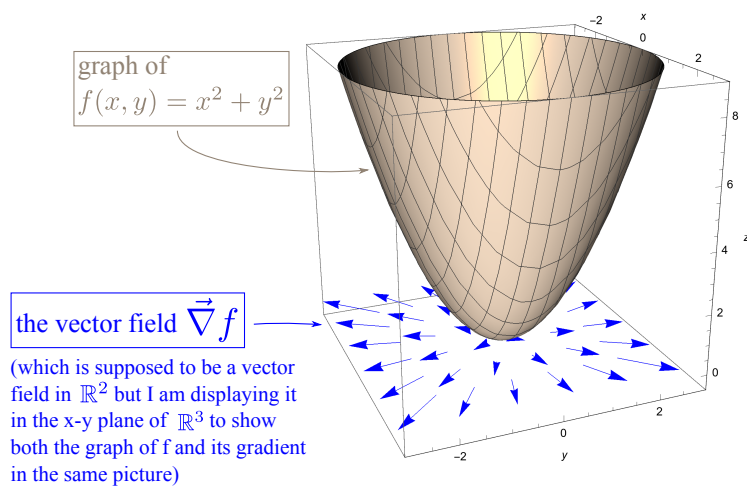
$$0 = \left. \frac{d}{dt}f(\vec{x}(t)) \right|_{t=0} = \left( \vec{\nabla}f(\vec{x}_0) \right) \cdot \left( \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right).$$

A vanishing dot product can only mean one thing: orthogonality! This is telling us that  $\vec{\nabla}f(\vec{x}_0)$  is orthogonal to  $\left. \frac{d\vec{x}}{dt} \right|_{t=0}$ . This holds whenever  $\vec{x}(t)$  is a path passing through  $\vec{x}_0$  that stays within the level set of  $f$ . Every single vector tangent to the level set of  $f$  at  $\vec{x}_0$  is the velocity of some such path. So this is telling us that  $\vec{\nabla}f(\vec{x}_0)$  is orthogonal to every vector that is tangent to the level set of  $f$  at  $\vec{x}_0$ . In short, the vector  $\vec{\nabla}f(\vec{x}_0)$  is orthogonal to the level set of  $f$  at  $\vec{x}_0$ .



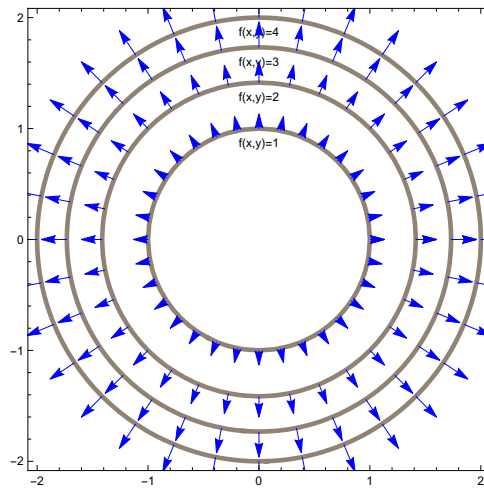
The pictures are showing you what things look like in the case of  $\mathbb{R}^3$ , but this stuff works in  $\mathbb{R}^n$ . The concepts of orthogonality and tangency work in the same way, even though the pictures are harder to draw.

**Example** Again consider the function  $f(x, y) = x^2 + y^2$ , which has  $\vec{\nabla}f = 2x\partial_x + 2y\partial_y$ . At the specific point  $(2, 0)$ , we have  $\vec{\nabla}f(2, 0) = 4\partial_x$ . This is telling you that if you start from the point  $(2, 0)$  in  $\mathbb{R}^2$ , then the  $\partial_x$  direction is the correct direction to start traveling if you want to experience the fastest possible increase in the values of  $f$ . We saw from the plot of  $\vec{\nabla}f$  that it just points radially outward at each point. This is telling you that no matter where you start, you should travel directly away from the origin if you want to experience the fastest possible increase in the values of  $f$ . The graph of  $f(x, y)$  should make this fact pretty obvious.

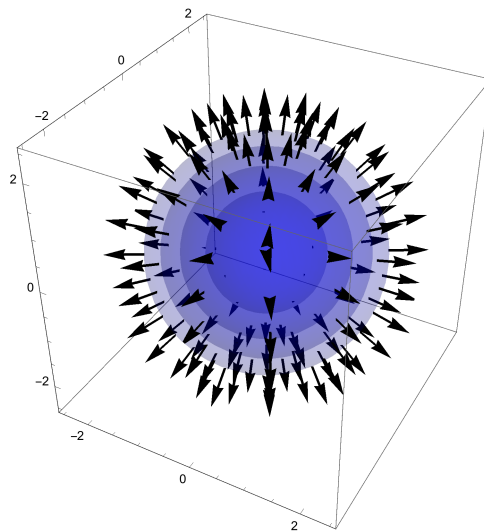


The vector field  $\vec{\nabla}f$  is pointing radially outward, reflecting the obvious fact that one should vary the input to  $f$  radially outward from any point in order to observe the fastest possible increase in its output.

You can also see that  $\vec{\nabla}f$  is orthogonal to the level sets.



**Another Example** Now consider the function  $f(x, y, z) = x^2 + y^2 + z^2$ . Just like in the previous example,  $f(\vec{x})$  is the squared distance of  $\vec{x}$  from the origin— it’s just defined on  $\mathbb{R}^3$  this time, rather than  $\mathbb{R}^2$ . Its level sets are spheres of various radii. The gradient vector field is  $\vec{\nabla} f = 2x \partial_x + 2y \partial_y + 2z \partial_z$ , and this points radially outward. “Radially outward” is indeed the direction in which one should travel to most quickly increase the output of  $f$ . And the gradient vectors pointing radially outward are indeed orthogonal to the level sets, which are spheres centered at the origin.



Summarizing our understanding of the direction of  $\vec{\nabla} f$ :

## The Direction of $\vec{\nabla} f$

Let  $f(\vec{x})$  be a function on  $\mathbb{R}^n$  and let  $\vec{x}_0$  be a point in  $\mathbb{R}^n$  where it is defined. Then  $\vec{\nabla} f(\vec{x}_0)$  is a vector-at- $\vec{x}_0$  which points...

- in the direction along which  $f(\vec{x})$  increases the fastest, if one started from  $\vec{x}_0$ .
- in the direction orthogonal to all directions in which  $f(\vec{x})$  stays constant, i.e. in the direction orthogonal to the level set of  $f$  at  $\vec{x}_0$ .

These ideas in the summary box are not dependent on coordinates— they are valid geometric facts no matter what coordinates one uses to express them. The only coordinate-dependent thing we've used so far is the method for actually *computing*  $\vec{\nabla} f$  by simply working out  $\frac{\partial f}{\partial x_1} \partial_{x_1} + \cdots + \frac{\partial f}{\partial x_n} \partial_{x_n}$ , which requires cartesian coordinates.

### 3.3.2 Directional Derivatives

We saw in section 1.5.2 that the derivative  $\frac{d}{dt} f(\vec{x}(t))$  tells us the rate of change of the function  $f(\vec{x})$  along the path  $\vec{p}(t) = \vec{x}(t)$ . This is a quantity that scales with the speed  $\left\| \frac{d\vec{x}}{dt} \right\|$  of the path— if the path  $\vec{x}(t)$  goes twice as fast, then the values of  $f(\vec{x}(t))$  will be observed to change twice as fast. Let's try to cancel this effect by dividing out the speed of the path:

$$\frac{\frac{d}{dt} f(\vec{x}(t))}{\left\| \frac{d\vec{x}}{dt} \right\|}.$$

To help us think about what this is, let's assume that  $\vec{x}(0) = \vec{x}_0$  and consider the passage of a small amount of time  $\Delta t$ . The function  $f$  changes by some amount  $\Delta f = f(\vec{x}(\Delta t)) - f(\vec{x}_0)$  and the path is displaced by some vector  $\Delta \vec{x} = \vec{x}(\Delta t) - \vec{x}_0$ . Now the expression we built above can be described as follows:

$$\frac{\frac{d}{dt} f(\vec{x}(t)) \big|_{t=0}}{\left\| \frac{d\vec{x}}{dt} \big|_{t=0} \right\|} = \frac{\frac{d}{dt} f(\vec{x}(t)) \big|_{t=0}}{\left\| \frac{d\vec{x}}{dt} \big|_{t=0} \right\|} \frac{\Delta t}{\Delta t} \approx \frac{\Delta f}{\|\Delta \vec{x}\|},$$

with the approximation becoming exact in the limit  $\Delta t \rightarrow 0$ . This is giving us the amount of increase in  $f$  per unit *distance*, rather than per unit  $t$ . Specifically, it's giving us the rate of change of  $f$  per unit distance travelled from  $\vec{x}_0$  along the direction  $\frac{d\vec{x}}{dt} \big|_{t=0}$ . It's called a *directional derivative*, and the chain rule tells us how to compute it without having to explicitly combine  $f(\vec{x})$  and  $\vec{x}(t)$ :

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta f}{\|\Delta \vec{x}\|} = \frac{\frac{d}{dt} f(\vec{x}(t)) \big|_{t=0}}{\left\| \frac{d\vec{x}}{dt} \big|_{t=0} \right\|} = \frac{1}{\left\| \frac{d\vec{x}}{dt} \big|_{t=0} \right\|} \left( \vec{\nabla} f(\vec{x}_0) \right) \cdot \left( \frac{d\vec{x}}{dt} \bigg|_{t=0} \right).$$

The quantity we just wrote down is the directional derivative of  $f$  at the point  $\vec{x}_0$  in the direction of  $\frac{d\vec{x}}{dt} \big|_{t=0}$ . The choice of direction  $\frac{d\vec{x}}{dt} \big|_{t=0}$  can just as well be specified by a vector-at- $\vec{x}_0$  rather than a path-passing-through- $\vec{x}_0$ . Here is the full definition in these terms:

## Directional Derivatives

The *directional derivative* of  $f(\vec{x})$  at  $\vec{x}_0$  in the direction  $\vec{v}$  is denoted by  $\text{Dir}(f, \vec{x}_0, \vec{v})$  and it can be computed as follows:

$$\text{Dir}(f, \vec{x}_0, \vec{v}) = \frac{1}{\|\vec{v}\|} \left( \vec{\nabla} f(\vec{x}_0) \right) \cdot \vec{v}.$$

It gives the rate of increase in  $f$  per unit *distance* travelled from  $\vec{x}_0$  along the direction  $\vec{v}$ .

Usually you would just use a unit vector for  $\vec{v}$  so that the directional derivative simply becomes  $\left( \vec{\nabla} f(\vec{x}_0) \right) \cdot \vec{v}$ .

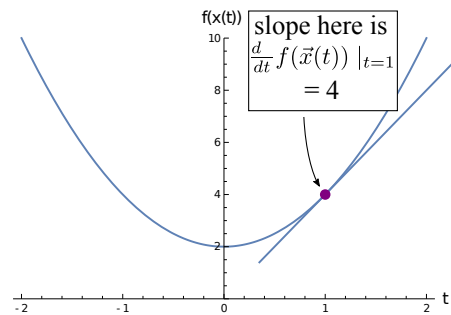
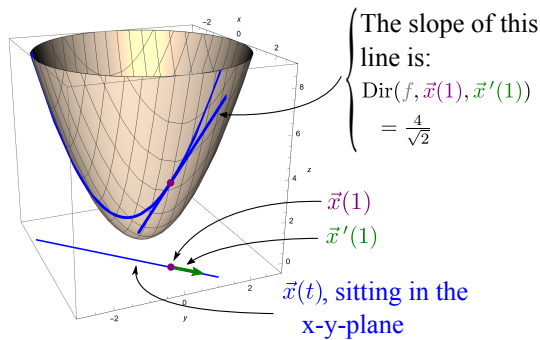
**Example** This example will be a repeat of the example at the end of section 1.5.2. Let  $f(x, y) = x^2 + y^2$  and let  $\vec{x}(t) = (t + 1, t - 1)$ . In section 1.5.2, we computed explicitly that  $\frac{d}{dt} f(\vec{x}(t)) \big|_{t=1} = 4$ . Later in section 1.7.3, we got the same answer using the chain rule, which we would now express as follows:

$$\begin{aligned} \frac{d}{dt} f(\vec{x}(t)) \big|_{t=1} &= \vec{\nabla} f(\vec{x}(1)) \cdot (\vec{x}'(1)) \\ &= \vec{\nabla} f(2, 0) \cdot (\partial_x + \partial_y) \\ &= (2x \partial_x + 2y \partial_y) \big|_{(2,0)} \cdot (\partial_x + \partial_y) \\ &= (4 \partial_x) \cdot (\partial_x + \partial_y) \\ &= 4. \end{aligned}$$

We interpreted this number 4 as *the rate of change of  $f(x, y)$  if one considers values of  $f(x, y)$  near  $\vec{x}(1) = (2, 0)$  and in the direction of the path, which is  $\vec{x}'(1) = \partial_x + \partial_y$* . But 4 is not quite the directional derivative! The directional derivative of  $f(x, y)$  at  $(2, 0)$  in the direction  $\partial_x + \partial_y$  is:

$$\begin{aligned} \text{Dir}(f, (2, 0), \partial_x + \partial_y) &= \frac{1}{\|\partial_x + \partial_y\|} \vec{\nabla} f(2, 0) \cdot (\partial_x + \partial_y) \\ &= \frac{1}{\sqrt{(\partial_x + \partial_y) \cdot (\partial_x + \partial_y)}} (4 \partial_x) \cdot (\partial_x + \partial_y) \\ &= \frac{4}{\sqrt{2}}. \end{aligned}$$

The difference is that the true directional derivative of  $f$ , which is  $\frac{4}{\sqrt{2}}$ , tells us about how  $f$  changes per unit *distance* of change in its input, whereas the quantity  $\frac{d}{dt} f(\vec{x}(t)) \big|_{t=1}$ , which is 4, tells us about how  $f$  changes per unit change in  $t$ .



Let me say all that again with a physical example in mind: If  $f(x, y)$  represents the elevation of a terrain over longitude and latitude coordinates  $(x, y)$ , and if  $\vec{x}(t)$  represents the path you take through the terrain on your hike, then  $\frac{d}{dt} f(\vec{x}(t)) \big|_{t=1} = 4$  is your rate of ascent per unit time at time  $t = 1$ . This rate of ascent is related to both how fast you happened to be hiking and how steep the terrain (the graph of  $f(x, y)$ ) is in the direction you happen to be going at  $t = 1$  (which is the direction  $\vec{p}'(1) = (1, 1)$ ). The actual steepness of the terrain in that direction is given by the directional derivative  $\text{Dir}(f, \vec{x}(1), \vec{x}'(1)) = \frac{4}{\sqrt{2}}$ .

### 3.3.3 Norm of $\vec{\nabla} f$

We understand the direction of  $\vec{\nabla} f$ , but what about its norm? Actually,  $\|\vec{\nabla} f\|$  arises as a particular case of a directional derivative:

$$\begin{aligned} \text{Dir}(f, \vec{x}_0, \vec{\nabla} f(\vec{x}_0)) &= \frac{1}{\|\vec{\nabla} f(\vec{x}_0)\|} (\vec{\nabla} f(\vec{x}_0)) \cdot (\vec{\nabla} f(\vec{x}_0)) \\ &= \frac{1}{\|\vec{\nabla} f(\vec{x}_0)\|} \|\vec{\nabla} f(\vec{x}_0)\|^2 \\ &= \|\vec{\nabla} f(\vec{x}_0)\|. \end{aligned}$$

The rate of change of  $f$  starting from  $\vec{x}_0$  along the direction of  $\vec{\nabla} f(\vec{x}_0)$  is  $\|\vec{\nabla} f(\vec{x}_0)\|$ . We already identified the direction of  $\vec{\nabla} f(\vec{x}_0)$  as the direction in which the rate of change of  $f$  is maximized. So  $\|\vec{\nabla} f(\vec{x}_0)\|$  is the rate of change of  $f$  in the direction along which the rate of change of  $f$  is maximized. We could have also spotted this directly from the chain rule. If  $\vec{x}(t)$  is chosen to be a path that starts at  $\vec{x}_0$  and travels with unit speed in the direction of  $\vec{\nabla} f(\vec{x}_0)$ , then

$$\left. \frac{d}{dt} f(\vec{x}(t)) \right|_{t=0} = (\vec{\nabla} f(\vec{x}_0)) \cdot \left( \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right) = \|\vec{\nabla} f(\vec{x}_0)\| \left\| \left. \frac{d\vec{x}}{dt} \right|_{t=0} \right\| \cos(0) = \|\vec{\nabla} f(\vec{x}_0)\|.$$

### Norm of $\vec{\nabla} f$

The number  $\|\vec{\nabla} f(\vec{x}_0)\|$  is the rate of change of  $f$  in the direction along which the rate of change of  $f$  is maximized. It is the fastest rate of increase of  $f$  that can be experienced by starting from  $\vec{x}_0$  and traveling in some direction. (And the direction that lets you experience it is  $\vec{\nabla} f(\vec{x}_0)$ .)

In terms of the graph of  $f$ , the gradient vector  $\vec{\nabla} f$  points in the direction of steepest ascent, and its magnitude tells you the slope in that direction.

**Example** Again let  $f(x, y) = x^2 + y^2$ . The gradient  $2x \partial_x + 2y \partial_y$  points radially outward and it has norm

$$\|\vec{\nabla} f\| = \sqrt{(2x \partial_x + 2y \partial_y) \cdot (2x \partial_x + 2y \partial_y)} = \sqrt{4x^2 + 4y^2} = 2\sqrt{x^2 + y^2}.$$

At a particular point like  $(3, 4)$ , this is telling you that, considering all the directions one could choose to head away from  $(3, 4)$ , the maximum rate of increase of  $f$  that could be observed is  $\|\vec{\nabla} f(3, 4)\| = 2\sqrt{3^2 + 4^2}$ . If you look at  $2\sqrt{x^2 + y^2}$  in polar coordinates then it is very easy to see what it is saying:

$$\|\vec{\nabla} f\| = 2r.$$

This is telling us that the maximum slope of  $f(x, y)$  (which we know is the slope in the radial direction because  $\vec{\nabla} f$  is always pointing radially outward) increases as you get further from the origin. This is obvious from the graph of  $f(x, y) = x^2 + y^2$ .

### 3.3.4 $\vec{\nabla} f$ in Non-Cartesian Coordinates

Given a function  $f(x, y)$  written in cartesian coordinates, we defined the gradient of  $f$  to be the vector that gets dotted with velocity in the chain rule:

$$\frac{d}{dt} f(x(t), y(t)) = [ \quad ??? \quad ] \cdot [x'(t) \partial_x + y'(t) \partial_y] ,$$

$$\frac{d}{dt} f(x(t), y(t)) = \left[ \begin{array}{l} \text{whatever fits here} \\ \text{is called “}\vec{\nabla} f\text{”} \end{array} \right] \cdot [x'(t) \partial_x + y'(t) \partial_y] ,$$

$$\frac{d}{dt} f(x(t), y(t)) = [ \vec{\nabla} f(x(t), y(t)) ] \cdot [x'(t) \partial_x + y'(t) \partial_y] .$$

That's why  $\vec{\nabla} f = \frac{\partial f}{\partial x} \partial_x + \frac{\partial f}{\partial y} \partial_y$ . If  $f$  is expressed in polar coordinates as  $f(r, \theta)$  and you have a path  $r(t), \theta(t)$ , then the chain rule works in a similar way:

$$\frac{d}{dt} f(r(t), \theta(t)) = \frac{\partial f}{\partial r} r'(t) + \frac{\partial f}{\partial \theta} \theta'(t).$$

The chain rule, which doesn't know or care about coordinates, still gives you a sum of products. It looks something like a dot product, so you might guess that  $\vec{\nabla}f = \frac{\partial f}{\partial r}\partial_r + \frac{\partial f}{\partial \theta}\partial_\theta$ , but watch how that fails:

$$\begin{aligned} \left[ \frac{\partial f}{\partial r}\partial_r + \frac{\partial f}{\partial \theta}\partial_\theta \right] \cdot [r'(t)\partial_r + \theta'(t)\partial_\theta] &= \frac{\partial f}{\partial r}r'(t)\partial_r \cdot \partial_r + \left( \frac{\partial f}{\partial r}\theta'(t) + \frac{\partial f}{\partial \theta}r'(t) \right) \partial_r \cdot \partial_\theta + \frac{\partial f}{\partial \theta}\theta'(t)\partial_\theta \cdot \partial_\theta \\ &= \frac{\partial f}{\partial r}r'(t) + r(t)^2\frac{\partial f}{\partial \theta}\theta'(t), \end{aligned}$$

which is *not* the same as  $\frac{d}{dt}f(r(t), \theta(t)) = \frac{\partial f}{\partial r}r'(t) + \frac{\partial f}{\partial \theta}\theta'(t)$ ! There's an extra  $r^2$  showing up because  $\partial_r \cdot \partial_\theta = r^2$ . We can make things work by introducing a  $\frac{1}{r^2}$  to cancel it:

$$\begin{aligned} \left[ \frac{\partial f}{\partial r}\partial_r + \frac{1}{r(t)^2}\frac{\partial f}{\partial \theta}\partial_\theta \right] \cdot [r'(t)\partial_r + \theta'(t)\partial_\theta] \\ &= \frac{\partial f}{\partial r}r'(t)\partial_r \cdot \partial_r + \left( \frac{\partial f}{\partial r}\theta'(t) + \frac{1}{r(t)^2}\frac{\partial f}{\partial \theta}r'(t) \right) \partial_r \cdot \partial_\theta + \frac{1}{r(t)^2}\frac{\partial f}{\partial \theta}\theta'(t)\partial_\theta \cdot \partial_\theta \\ &= \frac{\partial f}{\partial r}r'(t) + \frac{\partial f}{\partial \theta}\theta'(t). \end{aligned}$$

And now it works! Now that we have

$$\frac{d}{dt}f(r(t), \theta(t)) = \left[ \frac{\partial f}{\partial r}\partial_r + \frac{1}{r(t)^2}\frac{\partial f}{\partial \theta}\partial_\theta \right] \cdot [r'(t)\partial_r + \theta'(t)\partial_\theta],$$

we have found a vector field that is written in polar coordinates and that is worthy of the name "gradient of  $f$ ":

$$\vec{\nabla}f = \frac{\partial f}{\partial r}\partial_r + \frac{1}{r^2}\frac{\partial f}{\partial \theta}\partial_\theta.$$

Instead of guessing the wrong thing and then awkwardly trying to fix it, we could have also discovered this formula by using the vector transformation laws. In an earlier homework problem, you used the transformation law to compute expressions for  $\partial_x$  and  $\partial_y$  in terms of  $\partial_r$  and  $\partial_\theta$ . Starting with the expression we know for  $\vec{\nabla}f$  in cartesian coordinates and then substituting in those expressions from the homework problem, we get:

$$\begin{aligned} \vec{\nabla}f &= \frac{\partial f}{\partial x}\partial_x + \frac{\partial f}{\partial y}\partial_y \\ &= \frac{\partial f}{\partial x}\left(\frac{x}{\sqrt{x^2+y^2}}\partial_r - \frac{y}{x^2+y^2}\partial_\theta\right) + \frac{\partial f}{\partial y}\left(\frac{y}{\sqrt{x^2+y^2}}\partial_r + \frac{x}{x^2+y^2}\partial_\theta\right) \\ &= \frac{\partial f}{\partial x}\left(\frac{x}{r}\partial_r - \frac{y}{r^2}\partial_\theta\right) + \frac{\partial f}{\partial y}\left(\frac{y}{r}\partial_r + \frac{x}{r^2}\partial_\theta\right) \\ &= \frac{1}{r}\left(x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}\right)\partial_r + \frac{1}{r^2}\left(x\frac{\partial f}{\partial y} - y\frac{\partial f}{\partial x}\right)\partial_\theta. \end{aligned}$$

This is certainly working to give us an expression for  $\vec{\nabla}f$  in polar coordinates, but the partials  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$  are showing up, so this formula assumes that  $f$  was written in cartesian coordinates to start with. To get something involving  $\frac{\partial f}{\partial r}$  and  $\frac{\partial f}{\partial \theta}$  instead, we can assume that  $f$  was written in



polar coordinates as  $f(r, \theta)$  and we can use the chain rule on  $f(r(x, y), \theta(x, y))$ . Continuing the calculation:

$$\begin{aligned}\vec{\nabla} f &= \frac{1}{r} \left( x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} \right) \partial_r + \frac{1}{r^2} \left( x \frac{\partial f}{\partial y} - y \frac{\partial f}{\partial x} \right) \partial_\theta \\ &= \frac{1}{r} \left( x \left( \frac{\partial f}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial x} \right) + y \left( \frac{\partial f}{\partial r} \frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial y} \right) \right) \partial_r + \\ &\quad \frac{1}{r^2} \left( x \left( \frac{\partial f}{\partial r} \frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial y} \right) - y \left( \frac{\partial f}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial x} \right) \right) \partial_\theta \\ &= \frac{1}{r} \left( \left( x \frac{\partial r}{\partial x} + y \frac{\partial r}{\partial y} \right) \frac{\partial f}{\partial r} + \left( x \frac{\partial \theta}{\partial x} + y \frac{\partial \theta}{\partial y} \right) \frac{\partial f}{\partial \theta} \right) \partial_r + \\ &\quad \frac{1}{r^2} \left( \left( x \frac{\partial r}{\partial y} - y \frac{\partial r}{\partial x} \right) \frac{\partial f}{\partial r} + \left( x \frac{\partial \theta}{\partial y} - y \frac{\partial \theta}{\partial x} \right) \frac{\partial f}{\partial \theta} \right) \partial_\theta.\end{aligned}$$

Now compute certain pieces of this expression:

$$\begin{aligned}x \frac{\partial r}{\partial x} + y \frac{\partial r}{\partial y} &= x \frac{x}{\sqrt{x^2 + y^2}} + y \frac{y}{\sqrt{x^2 + y^2}} = \sqrt{x^2 + y^2} = r \\ x \frac{\partial \theta}{\partial x} + y \frac{\partial \theta}{\partial y} &= -x \frac{y}{x^2 + y^2} + y \frac{x}{x^2 + y^2} = 0 \\ x \frac{\partial r}{\partial y} - y \frac{\partial r}{\partial x} &= x \frac{y}{\sqrt{x^2 + y^2}} - y \frac{x}{\sqrt{x^2 + y^2}} = 0 \\ x \frac{\partial \theta}{\partial y} - y \frac{\partial \theta}{\partial x} &= x \frac{x}{x^2 + y^2} + y \frac{y}{x^2 + y^2} = 1.\end{aligned}$$

And finally we can finish the calculation:

$$\begin{aligned}\vec{\nabla} f &= \frac{1}{r} \left( \left( x \frac{\partial r}{\partial x} + y \frac{\partial r}{\partial y} \right) \frac{\partial f}{\partial r} + \left( x \frac{\partial \theta}{\partial x} + y \frac{\partial \theta}{\partial y} \right) \frac{\partial f}{\partial \theta} \right) \partial_r + \\ &\quad \frac{1}{r^2} \left( \left( x \frac{\partial r}{\partial y} - y \frac{\partial r}{\partial x} \right) \frac{\partial f}{\partial r} + \left( x \frac{\partial \theta}{\partial y} - y \frac{\partial \theta}{\partial x} \right) \frac{\partial f}{\partial \theta} \right) \partial_\theta \\ &= \frac{1}{r} \left( (r) \frac{\partial f}{\partial r} + (0) \frac{\partial f}{\partial \theta} \right) \partial_r + \frac{1}{r^2} \left( (0) \frac{\partial f}{\partial r} + (1) \frac{\partial f}{\partial \theta} \right) \partial_\theta \\ &= \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta.\end{aligned}$$

Great, we got the same answer. There's a general formula lurking behind this calculation, which could give us an expression for the gradient of a function in any coordinates. But as you might guess from the example calculation, the formula is not very pretty:

$$\vec{\nabla} f = \sum_{j=1}^n \left[ \sum_{k=1}^n \frac{\partial f}{\partial u_k} \left[ \sum_{i=1}^n \frac{\partial u_k}{\partial x_i} \frac{\partial u_j}{\partial x_i} \right] \right] \partial_{u_j}.$$

**Example** Let's revisit the function  $f(x, y) = x^2 + y^2$  from all our earlier examples. In polar coordinates, the same function can be written as  $f(r, \theta) = r^2$ . Now we have a way to get the

gradient without ever having to switch back to cartesian coordinates:

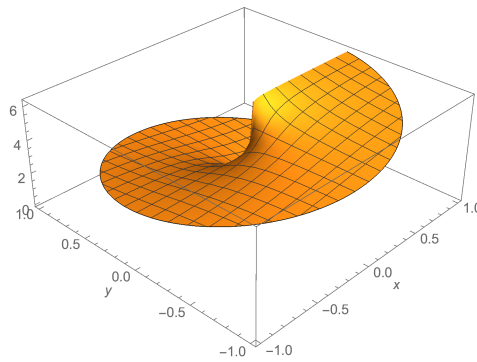
$$\vec{\nabla} f = \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta = 2r \partial_r.$$

And we see again that it is pointing radially outward everywhere. Its norm is given by

$$\|\vec{\nabla} f\| = \|2r \partial_r\| = 2r \|\partial_r\| = 2r,$$

which is the same answer we got earlier using cartesian coordinates.

**Another Example** Here is a fun one: What is  $\vec{\nabla}\theta$ ? Remember, the symbol for the coordinate  $\theta$  actually stands for the *function* that assigns to points in  $\mathbb{R}^2$  their  $\theta$  coordinates. In polar coordinates this is a very simple function to write down:  $\theta(r, \theta) = \theta$ . It has a pretty cool graph:



We have  $\frac{\partial \theta}{\partial r} = 0$  and  $\frac{\partial \theta}{\partial \theta} = 1$ , so

$$\begin{aligned} \vec{\nabla} \theta &= \frac{\partial \theta}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial \theta}{\partial \theta} \partial_\theta \\ &= \frac{1}{r^2} \partial_\theta. \end{aligned}$$

The fact that  $\vec{\nabla}\theta$  points in the  $\partial_\theta$  direction is telling us the following fact: Starting from any point in the plane, if you want to increase your theta coordinate as fast as possible then you should head in the  $\partial_\theta$  direction. It sounds really obvious, but just wait until the next example! The norm of  $\vec{\nabla}\theta$  is also interesting:

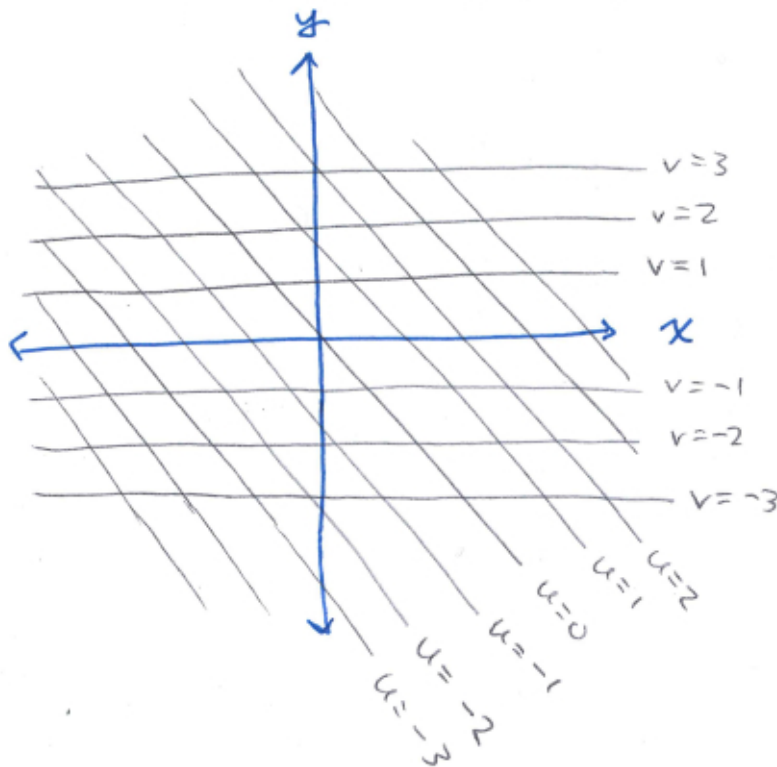
$$\|\vec{\nabla}\theta\| = \left\| \frac{1}{r^2} \partial_\theta \right\| = \frac{1}{r^2} \|\partial_\theta\| = \frac{1}{r}$$

This is telling us that the maximum rate of change of  $\theta$  coordinate (which can be achieved by traveling in the  $\partial_\theta$  direction) is inversely proportional to the distance from the origin. In other words: If you are very close to the origin, then you can achieve very rapid changes in  $\theta$  with only tiny movements. Makes sense, right? This should also be evident from the graph of  $\theta$  above.

**Another Example** Let's try working with a different set of coordinates:

$$\begin{aligned} u &= x + y & x &= u - v \\ v &= y & y &= v. \end{aligned}$$

This is a coordinate system on  $\mathbb{R}^2$ , and it's very similar to the default cartesian coordinates. But it distorts angles and lengths so it's a little different, with the differences becoming apparent once we consider dot products. Since the gradient is defined in terms of its role in a dot product, it will also look very different in these coordinates. First, have a look at the coordinate grid. Convince yourself that the picture makes sense (this is just an application of the techniques from section 2.1.2).



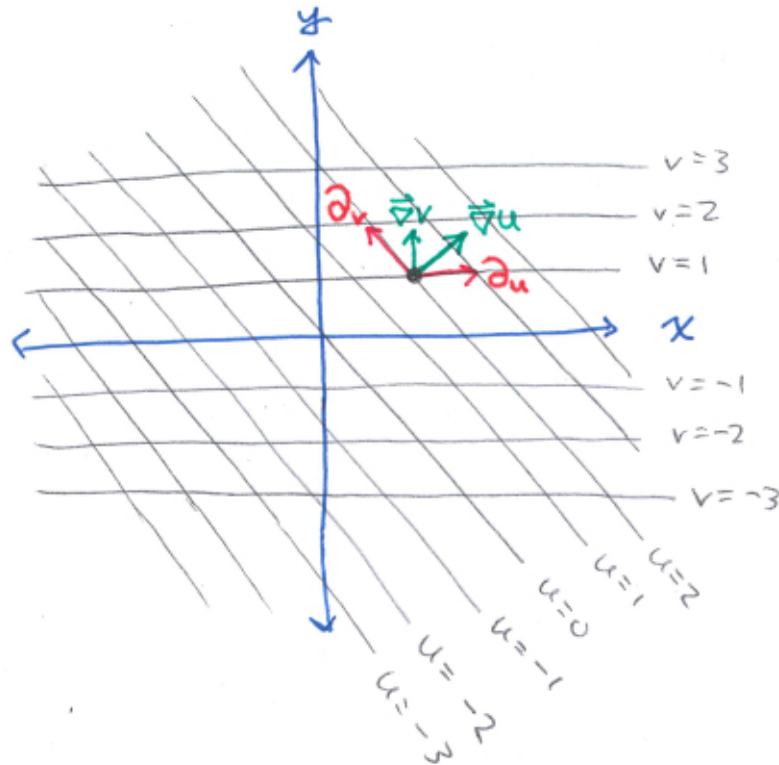
If we had a function  $f(u, v)$  on  $\mathbb{R}^2$  that was expressed in  $u, v$  coordinates, it might be nice to know how to take its gradient without having to go through  $x, y$  coordinates. Here is the formula:

$$\vec{\nabla} f = \left[ 2 \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v} \right] \partial_u + \left[ \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v} \right] \partial_v.$$

Just accept this formula— we are only interested in interpreting the results here. This says in particular that  $\vec{\nabla} u = 2\partial_u + \partial_v$  and  $\vec{\nabla} v = \partial_u + \partial_v$ . You could have gotten this information by going through  $x, y$  coordinates: First use  $u = x + y$  to get  $\vec{\nabla} u = \partial_x + \partial_y$ , and then use  $\partial_x = \frac{\partial u}{\partial x} \partial_u + \frac{\partial v}{\partial x} \partial_v$  and  $\partial_y = \frac{\partial u}{\partial y} \partial_u + \frac{\partial v}{\partial y} \partial_v$  to rewrite  $\vec{\nabla} u = \partial_x + \partial_y$  in terms of  $\partial_u$  and  $\partial_v$ . (Try this calculation and make sure you correctly get  $\vec{\nabla} u = 2\partial_u + \partial_v$  and  $\vec{\nabla} v = \partial_u + \partial_v$ .)

What is  $\vec{\nabla} u = 2\partial_u + \partial_v$  telling us? It's saying that if you start at some point in  $\mathbb{R}^2$  and you want to increase your  $u$  coordinate as fast as possible, then you should head in the  $2\partial_u + \partial_v$  direction.

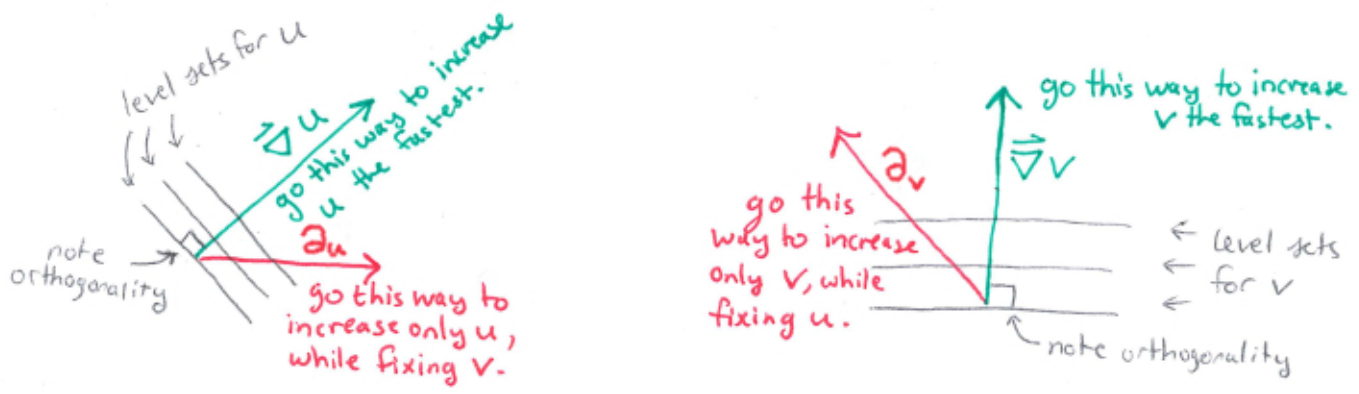
Surprised? You might have expected that the best way to increase your  $u$  coordinate as fast as possible is to head in the  $\partial_u$  direction. But no! That's not really what  $\partial_u$  means. Remember,  $\partial_u$  is supposed to point in the direction that you should travel to increase your  $u$  coordinate while keeping all other coordinates (i.e. your  $v$  coordinate) fixed. In this situation, "increasing  $u$  while holding  $v$  fixed" is not achieved by the same direction as "increasing  $u$  as quickly as possible." Look:



Here I have picked some point at random and shown four vectors at that point:

$$\begin{aligned} \partial_u &= \partial_x, \\ \partial_v &= \partial_y - \partial_x, \\ \vec{\nabla}u &= 2\partial_u + \partial_v = \partial_x + \partial_y, \quad \text{and} \\ \vec{\nabla}v &= \partial_u + \partial_v = \partial_y. \end{aligned}$$

See how the red vectors point along the coordinate paths? For example  $\partial_u$  is pointing in the direction that one would travel to increase their  $u$  coordinate while keeping their  $v$  coordinate fixed. To understand the direction of the green vectors, ask yourself: which way should I travel to increase my  $u$  coordinate as fast as possible? You would probably want to go directly towards the next higher level set for  $u$ , right? So you would want to travel orthogonally to the level set you are on, and this is why  $\vec{\nabla}u$  looks the way it does. Similarly, notice that  $\vec{\nabla}v$  is orthogonal to level sets for  $v$ .



**Gradient in Cylindrical and Spherical Coordinates** Since the calculation for what a gradient looks like in different coordinates is not so fun, it is useful to have the result handy for commonly used coordinate systems. For cylindrical coordinates on  $\mathbb{R}^3$  we have

$$\vec{\nabla} f = \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta + \frac{\partial f}{\partial z} \partial_z.$$

For spherical coordinates on  $\mathbb{R}^3$  we have

$$\vec{\nabla} f = \frac{\partial f}{\partial \rho} \partial_\rho + \frac{1}{\rho^2 \sin^2(\phi)} \frac{\partial f}{\partial \theta} \partial_\theta + \frac{1}{\rho^2} \frac{\partial f}{\partial \phi} \partial_\phi.$$

### Summary: Gradient in Any Coordinate System

The gradient of a function  $f(u_1, u_2, \dots, u_n)$ , which could be expressed in any coordinates, is defined to be whatever vector field fills the blank in the statement of the chain rule:

$$\frac{d}{dt} f(\vec{u}(t)) = \left[ \begin{array}{l} \text{whatever fits here} \\ \text{is called } \vec{\nabla} f \end{array} \right] \cdot \frac{d\vec{u}}{dt}.$$

One way to compute  $\vec{\nabla} f$  is to first convert  $f$  to cartesian coordinates, where the formula for gradient is the simplest:

$$\vec{\nabla} f = \frac{\partial f}{\partial x_1} \partial_{x_1} + \dots + \frac{\partial f}{\partial x_n} \partial_{x_n}.$$

Then if you want it in different coordinates, just use the vector transformation law to rewrite this expression. If you don't want to have to go through cartesian coordinates every time you compute a gradient, then you can use the vector transformation law and some chain rule to derive a general formula. Here are general formulas in terms of polar, cylindrical, and spherical coordinates:

$$\begin{aligned} \vec{\nabla} f(r, \theta) &= \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta \\ \vec{\nabla} f(r, \theta, z) &= \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta + \frac{\partial f}{\partial z} \partial_z \\ \vec{\nabla} f(\rho, \theta, \phi) &= \frac{\partial f}{\partial \rho} \partial_\rho + \frac{1}{\rho^2 \sin^2(\phi)} \frac{\partial f}{\partial \theta} \partial_\theta + \frac{1}{\rho^2} \frac{\partial f}{\partial \phi} \partial_\phi. \end{aligned}$$

There are much more efficient ways to get these gradient formulas, and we do not have time to cover them in this short summer course. Mainly I just expect you to know what these formulas mean and how to use them— I will not ask you to derive them yourself.

### 3.3.5 The Notation for Coordinate Basis Vectors

This section is optional reading, in case you are curious as to why the notation  $\partial_u$  for a coordinate basis vector looks like a partial derivative symbol.

The reason is directional derivatives. The main thing that we do with vectors-at-points is to use them alongside functions in order to get the rate of change of a function along a certain direction. If you think of coordinate basis vectors like  $\partial_u$  as though they were partial derivative operators like “ $\frac{\partial}{\partial u}$ ,” then they end up fitting together with functions in just the right way to give you directional derivatives. This is why the vector  $\partial_u$  has a partial symbol in its notation. Let me demonstrate with a couple of examples.

**Example 1** Suppose you wanted to know the rate of change of a function  $f(x, y)$  along the direction  $\partial_x$ . This would be the partial derivative  $\frac{\partial f}{\partial x}$ . You can get this by viewing  $\partial_x$  as a partial derivative operator  $\frac{\partial}{\partial x}$ , and then combining it with  $f(x, y)$  to get  $\frac{\partial f}{\partial x}$ .

**Example 2** Suppose you wanted to know the rate of change of  $f(x, y)$  along the direction  $2\partial_x - 3\partial_y$ . Then you would dot the vector  $2\partial_x - 3\partial_y$  with the gradient  $\vec{\nabla}f$ :

$$\left(\frac{\partial f}{\partial x}\partial_x + \frac{\partial f}{\partial y}\partial_y\right) \cdot (2\partial_x - 3\partial_y)$$

(Here we are assuming that  $x, y$  are cartesian coordinates.) You would also want to divide by the norm of  $2\partial_x - 3\partial_y$  in order to get a true directional derivative, but let’s ignore that for the purpose of this discussion. When you work out the dot product you get

$$2\frac{\partial f}{\partial x} - 3\frac{\partial f}{\partial y}.$$

It is just as though we took the partial derivative “operator”  $2\frac{\partial}{\partial x} - 3\frac{\partial}{\partial y}$  and then you “applied” it to the function  $f$ . The notation  $\vec{v} = 2\partial_x - 3\partial_y$  is meant to remind you that taking directional derivatives in the  $\vec{v}$  direction involves an application of the  $2\frac{\partial}{\partial x} - 3\frac{\partial}{\partial y}$  operator (and also dividing by the norm of  $\vec{v}$  if you want a true directional derivative).

**Example 3** Suppose that  $f(r, \theta)$  is given to you in polar coordinates, and you want to know the rate of change of  $f$  along the direction  $\partial_r + 2\partial_\theta$ . Then you would dot the vector  $\partial_r + 2\partial_\theta$  with the gradient  $\vec{\nabla}f$ :

$$\left(\frac{\partial f}{\partial r}\partial_r + \frac{1}{r^2}\frac{\partial f}{\partial \theta}\partial_\theta\right) \cdot (\partial_r + 2\partial_\theta)$$

When you work out the dot product you get

$$\frac{\partial f}{\partial r} + 2\frac{\partial f}{\partial \theta}.$$

It is the same as applying the partial derivative operator  $\frac{\partial}{\partial r} + 2\frac{\partial}{\partial \theta}$  to the function  $f$ , and this is why the vector  $\partial_r + 2\partial_\theta$  is written the way it is.

**Other notations for coordinate basis vectors** If you are a physics or engineering student, you may very well see coordinate basis vectors written with some other notation:

- $\hat{i}$ ,  $\hat{j}$ , and  $\hat{k}$  are sometimes used in place of  $\partial_x$ ,  $\partial_y$ , and  $\partial_z$ . This notation is only used to talk about cartesian coordinates. It's inspired by quaternions, which provide an alternative way to do vector algebra.
- Sometimes  $\hat{u}$  is used as a normalized version of  $\partial_u$ . So for example  $\hat{x}$  and  $\hat{y}$  can be used in place of  $\partial_x$  and  $\partial_y$ . And  $\hat{r}$  is used in place of  $\partial_r$  in polar coordinates... but be careful with  $\partial_\theta$ ! Since  $\hat{\theta}$  is supposed to be *normalized*, the relationship is  $\hat{\theta} = \frac{1}{r}\partial_\theta$ . This is because  $\|\partial_\theta\| = r$ .
- There are plenty of other notations that people use for the *normalized* version of  $\partial_u$ . Ones that I've seen include a boldface  $\mathbf{u}$ , or  $e_u$ , or  $\hat{e}_u$ , or  $\vec{e}_u$ , or  $\mathbf{e}_u$ .

### 3.4 The Derivative of a Function $\vec{F}(\vec{x})$

We've considered two kinds of derivatives so far: derivatives of paths (velocities) and derivatives of functions (gradients). This section considers the case of a function  $\vec{F}(\vec{x})$  that has both multiple variables *and* multiple components. We've already worked with one such kind of function: a coordinate transformation. For example, you can think of the polar-to-cartesian coordinate transformation as a single function of two variables ( $r$  and  $\theta$ ) and with two outputs ( $x$  and  $y$ ):

$$\vec{F}(r, \theta) = \begin{bmatrix} x(r, \theta) \\ y(r, \theta) \end{bmatrix} = \begin{bmatrix} r \cos(\theta) \\ r \sin(\theta) \end{bmatrix}.$$

But a function of multiple variables with multiple components does not have to be a coordinate transformation. In fact, in this section we will not think too much about coordinates—just consider functions as machines into which a list of numbers goes in and a list of numbers comes out. We will form pictures by just imagining that everything is in cartesian coordinates.

The functions  $\vec{F}(\vec{x})$  of this section take *points* to *points*. If you are considering a function of  $n$  variables with  $m$  components

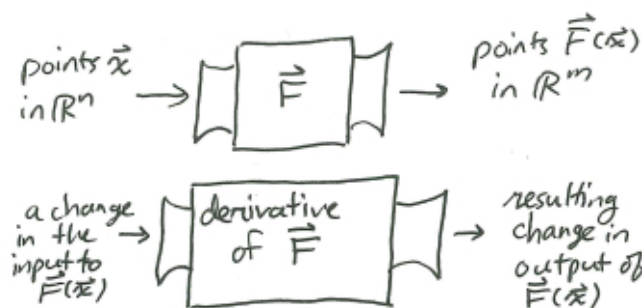
$$\vec{F}(\vec{x}) = \begin{bmatrix} F_1(x_1, \dots, x_n) \\ \vdots \\ F_m(x_1, \dots, x_n) \end{bmatrix},$$

then the way to think about it in this section is that whenever you input a point of  $\mathbb{R}^n$  into  $\vec{F}$ , it outputs a point of  $\mathbb{R}^m$ . Do not confuse the functions  $\vec{F}(\vec{x})$  of this section with *vector fields*. When

you input a point into a vector field, you get a vector at that point. When you input a point into the kind of function we are talking about here, you get another point<sup>7</sup>.

### 3.4.1 Defining the Derivative of $\vec{F}(\vec{x})$

The idea of a derivative is that it's supposed to capture the way in which a function's output responds to changes in its input. In the case of a function of  $n$  variables, the inputs are coming from the space  $\mathbb{R}^n$  and there are therefore many directions in which one could change an input. In the case of a function with  $m$  components, the outputs land in the space  $\mathbb{R}^m$  and there are therefore many directions in which an output can change in response to a change in input. This gives us the following clue about the derivative of a function of  $n$  variables with  $m$  components: whatever the derivative is, it should be an object that can be given a direction and amount of input change, and that outputs a direction and amount of output change.



To begin our exploration of this idea, we need a way of describing a changing input. We can use paths for this. Think of a path  $\vec{x}(t)$  starting at  $\vec{x}(0) = \vec{x}_0$  as a particular way of varying the input to a function  $\vec{F}(\vec{x})$ . To look at how the function responds, we consider the outputs  $\vec{F}(\vec{x}(t))$  that it gives along the path. This is another path, this time starting at  $\vec{F}(\vec{x}_0)$ . The path  $\vec{F}(\vec{x}(t))$  is the *path of outputs* of  $\vec{F}$  that corresponds to the path of inputs  $\vec{x}(t)$ . If  $\vec{F}(\vec{x})$  is a function of  $n$  variables with  $m$  components, then it turns paths  $\vec{x}(t)$  in  $\mathbb{R}^n$  into paths  $\vec{F}(\vec{x}(t))$  in  $\mathbb{R}^m$ . Somehow, the derivative of  $\vec{F}$  will have to relate paths of inputs to paths of outputs. Let us look at an example of how functions can affect paths.

Consider the following function of 2 variables with 2 components:

$$\vec{F}(x, y) = \begin{bmatrix} 2x^2 - xy + 1 \\ -(x + 2)2^{-y/8} \end{bmatrix}.$$

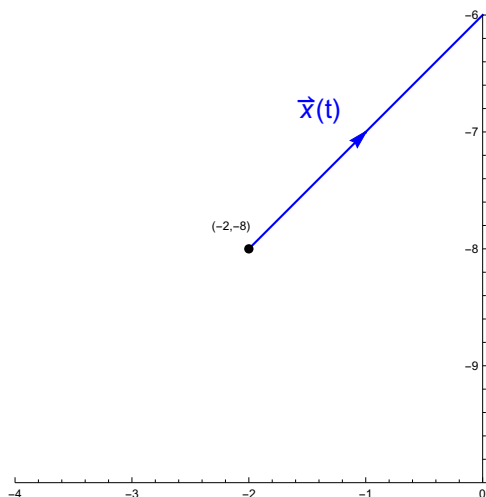
First, think about the *type* of the function: it takes points in  $\mathbb{R}^2$  to points in  $\mathbb{R}^2$ . So it also takes *paths* in  $\mathbb{R}^2$  to *paths* in  $\mathbb{R}^2$ . For example consider the path

$$\vec{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} -2 + t \\ -8 + t \end{bmatrix}.$$

<sup>7</sup>If I did want to express that  $\vec{F}(\vec{x})$  stood for a vector field, then I would write  $\vec{F}(\vec{x}) = F_1(\vec{x})\partial_{x_1} + \dots + F_n(\vec{x})\partial_{x_n}$ . Oh, and note that for vector fields one is forced to have  $n = m$  (the number of components of the output must equal the number of components of the input).



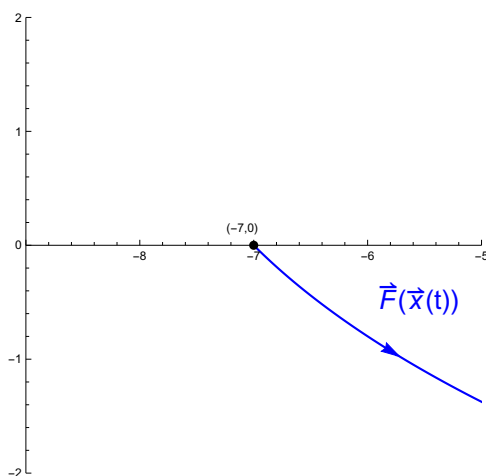
This path starts at  $\vec{x}_0 = (-2, -8)$  and heads off in the  $\partial_x + \partial_y$  direction in a straight line.



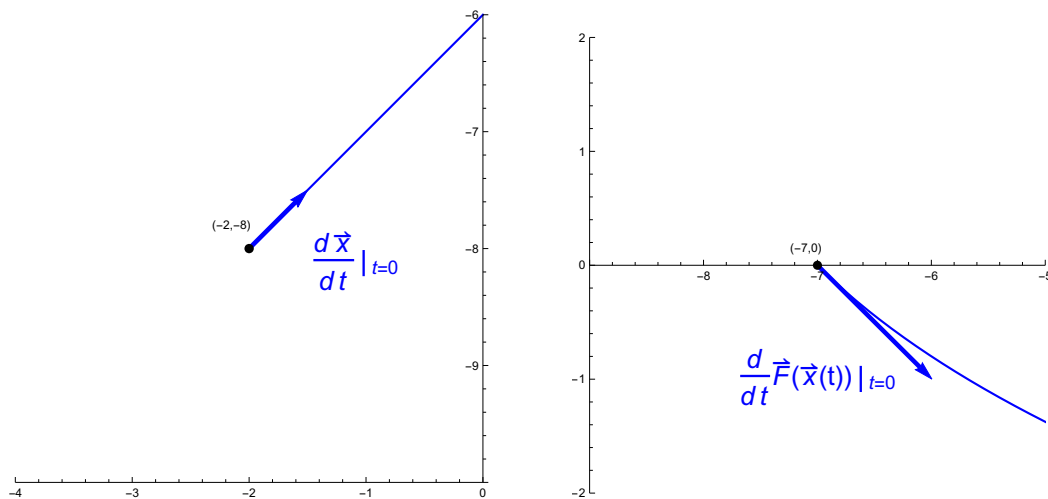
Imagine that you are starting by inputting  $(-2, -8)$  into  $\vec{F}$  to get  $\vec{F}(-2, -8) = (-7, 0)$ , and then you change the input along the path  $\vec{x}(t)$ . The output changes according to the path:

$$\vec{F}(x(t), y(t)) = \begin{bmatrix} 2(-2+t)^2 - (-2+t)(-8+t) + 1 \\ -t2^{-(8+t)/8} \end{bmatrix}$$

which looks like



To get information about what is happening near  $\vec{x}_0$  only, we examine what happens to the initial velocity of the path:

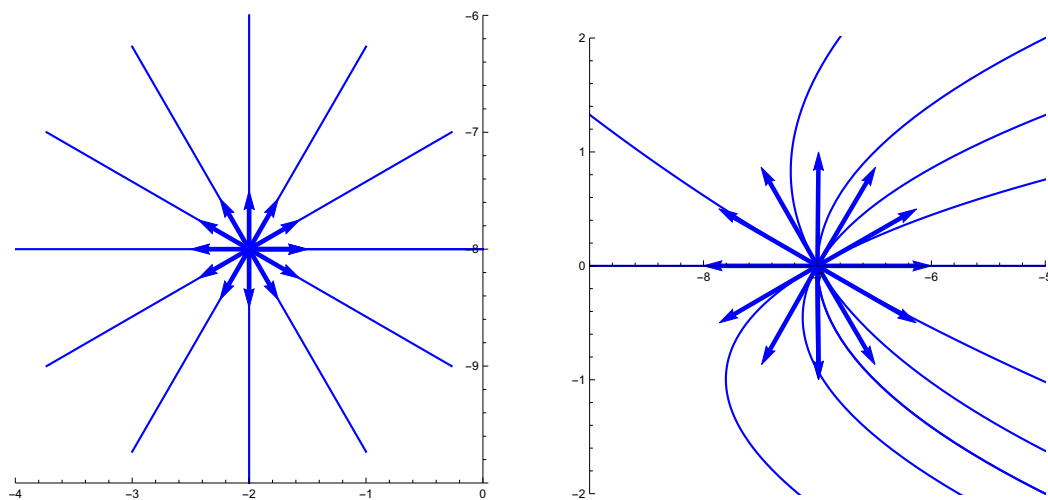


We calculate the initial velocity of  $\vec{F}(\vec{x}(t))$  to be

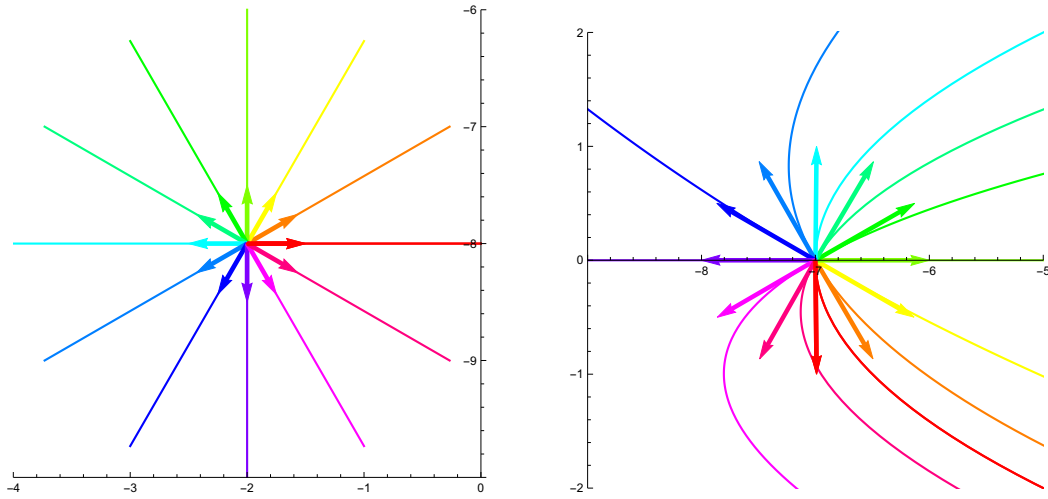
$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \vec{F}(x(t), y(t)) &= \frac{d}{dt} \Big|_{t=0} \begin{bmatrix} 2(-2+t)^2 - (-2+t)(-8+t) + 1 \\ -t2^{-(8+t)/8} \end{bmatrix} \\ &= [4(-2+t) - (-2+t) - (-8+t)] \partial_x + [-2^{-(8+t)/8} + \frac{1}{8} \log(2)t2^{-(8+t)/8}] \partial_y \Big|_{t=0} \\ &= [2] \partial_x + [-2] \partial_y. \end{aligned}$$

This tells us that if you start from  $\vec{x}_0 = (-2, -8)$  and change the input in the  $\partial_x + \partial_y$  direction, then you can expect the output to change in the  $[2] \partial_x + [-2] \partial_y$  direction. Okay, but that was just one path. To get an overall view of how  $\vec{F}$  responds to input variations near  $\vec{x}_0$ , we need to do the above calculation for *all possible paths* that start at  $\vec{x}_0$ .

Let me start with a picture instead of a calculation. Let's take a bunch of paths heading off from  $\vec{x}_0 = (-2, -8)$  and heading off in various directions, and then look at what  $\vec{F}$  does to them:



Okay it's kind of hard to see which paths go to which paths, so let me color them:



It looks like  $\vec{F}(\vec{x})$  has the effect of rotating paths by  $90^\circ$  clockwise, at least near  $\vec{x}_0$ . The picture gives us even more information if we look not only at the *direction* of the response of  $\vec{F}(\vec{x})$ , but also at the *magnitude* of the response. The arrows I included in these plots are the initial velocities of the paths. By looking at the initial velocities, you can see that in addition to rotating paths,  $\vec{F}(\vec{x})$  also has the effect of making them run roughly twice as fast. Summarizing our observations: *If you change the input to  $\vec{F}(\vec{x})$  starting from  $\vec{x}_0$ , then the output responds by changing twice as fast and in a direction  $90^\circ$  clockwise from the direction in which you changed the input.* This statement is what the derivative of  $\vec{F}(\vec{x})$  should capture.

Let's now establish the statement by a formal calculation. If  $\vec{x}(t)$  is any path starting at  $\vec{x}_0 = (-2, -8)$ , say with initial velocity  $\vec{x}'(t) = v_x \partial_x + v_y \partial_y$ , then the path  $\vec{F}(\vec{x}(t))$  has initial velocity

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \vec{F}(\vec{x}(t)) &= \frac{d}{dt} \Big|_{t=0} \begin{bmatrix} F_1(x(t), y(t)) \\ F_2(x(t), y(t)) \end{bmatrix} \\ &= \left[ \frac{d}{dt} \Big|_{t=0} F_1(x(t), y(t)) \right] \partial_x + \left[ \frac{d}{dt} \Big|_{t=0} F_2(x(t), y(t)) \right] \partial_y \\ &= \left[ \frac{\partial F_1}{\partial x} \Big|_{\vec{x}_0} v_x + \frac{\partial F_1}{\partial y} \Big|_{\vec{x}_0} v_y \right] \partial_x + \left[ \frac{\partial F_2}{\partial x} \Big|_{\vec{x}_0} v_x + \frac{\partial F_2}{\partial y} \Big|_{\vec{x}_0} v_y \right] \partial_y. \end{aligned}$$

We used the chain rule in the last step. In order to finish the calculation we need certain partial derivatives:

$$\begin{aligned} \frac{\partial F_1}{\partial x} &= 4x - y & \frac{\partial F_1}{\partial y} &= -x \\ \frac{\partial F_2}{\partial x} &= -2^{-y/8} & \frac{\partial F_2}{\partial y} &= \frac{1}{8} \log(2)(x+2)2^{-y/8}. \end{aligned}$$

Evaluating at  $\vec{x}_0 = (-2, -8)$ , we get

$$\begin{aligned} \frac{\partial F_1}{\partial x} \Big|_{\vec{x}_0} &= 0 & \frac{\partial F_1}{\partial y} \Big|_{\vec{x}_0} &= 2 \\ \frac{\partial F_2}{\partial x} \Big|_{\vec{x}_0} &= -2 & \frac{\partial F_2}{\partial y} \Big|_{\vec{x}_0} &= 0. \end{aligned}$$

Now we finish the calculation:

$$\frac{d}{dt} \Big|_{t=0} \vec{F}(\vec{x}(t)) = 2v_y \partial_x - 2v_x \partial_y.$$

What we have found is that if you change the input to  $\vec{F}(\vec{x})$  starting from  $\vec{x}_0$  with an initial velocity  $v_x\partial_x + v_y\partial_y$ , then the output responds by changing along a path starting from  $\vec{F}(\vec{x}_0)$  with initial velocity  $2v_y\partial_x - 2v_x\partial_y$ . If you focus on just what is happening to velocity components, the transformation  $(v_x, v_y) \mapsto (2v_y, -2v_x)$  can be described in terms of a matrix acting on column vectors:

$$\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 2v_y \\ -2v_x \end{bmatrix}.$$

Let me repeat our summary of observations from above: *If you change the input to  $\vec{F}(\vec{x})$  starting from  $\vec{x}_0$ , then the output responds by changing twice as fast and in a direction  $90^\circ$  clockwise from the direction in which you changed the input.* The matrix  $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$  encodes this information exactly! Its action on vectors is to scale them by 2 and rotate them  $90^\circ$  clockwise. In other words the matrix  $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$  represents *what  $\vec{F}$  does to initial velocities at  $\vec{x}_0$* . This is what we call *the derivative of  $\vec{F}$  at  $\vec{x}_0$* . Finally, a definition:

#### Definition of Derivative

Given a function  $\vec{F}(\vec{x})$  of  $n$  variables with  $m$  components, its derivative at a point  $\vec{x}_0$  is an  $m \times n$  matrix  $DF(\vec{x}_0)$  that does the following to column vectors:

$$DF(\vec{x}_0) \begin{bmatrix} \text{column vector of compo-} \\ \text{nents of initial velocity of a} \\ \text{path } \vec{x}(t) \text{ of inputs} \end{bmatrix} = \begin{bmatrix} \text{column vector of compo-} \\ \text{nents of initial velocity of} \\ \text{the path } \vec{F}(\vec{x}(t)) \text{ of outputs} \end{bmatrix}.$$

Writing the same thing out symbolically,  $DF$  can be described as follows: Given a path  $\vec{x}(t)$  of inputs to  $F$  that starts at some point  $\vec{x}_0$  with some initial velocity  $\vec{v}$ , the initial velocity of the path of outputs  $\vec{F}(\vec{x}(t))$  is the result of applying the matrix  $DF(\vec{x}_0)$  to  $\vec{v}$ :

$$\left. \frac{d}{dt} \vec{F}(\vec{x}(t)) \right|_{t=0} = DF(\vec{x}_0)(\vec{v}).$$

Without evaluating at a specific point  $\vec{x}_0$ , the derivative  $DF$  of  $\vec{F}(\vec{x})$  is a matrix-valued function on  $\mathbb{R}^n$ .

What we just found in our example is that

$$DF(-2, -8) = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}.$$

If we want to know  $DF$  at any  $\vec{x}$ , then we look at the initial velocity of  $\vec{F}(\vec{x}(t))$  when  $\vec{x}(t)$  starts at

some unspecified  $(x, y)$  and has initial velocity  $v_x \partial_x + v_y \partial_y$ :

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \vec{F}(\vec{x}(t)) &= \frac{d}{dt} \Big|_{t=0} \begin{bmatrix} F_1(x(t), y(t)) \\ F_2(x(t), y(t)) \end{bmatrix} \\ &= \left[ \frac{d}{dt} \Big|_{t=0} F_1(x(t), y(t)) \right] \partial_x + \left[ \frac{d}{dt} \Big|_{t=0} F_2(x(t), y(t)) \right] \partial_y \\ &= \left[ \frac{\partial F_1}{\partial x} v_x + \frac{\partial F_1}{\partial y} v_y \right] \partial_x + \left[ \frac{\partial F_2}{\partial x} v_x + \frac{\partial F_2}{\partial y} v_y \right] \partial_y. \end{aligned}$$

The transformation  $(v_x, v_y) \mapsto \left( \frac{\partial F_1}{\partial x} v_x + \frac{\partial F_1}{\partial y} v_y, \frac{\partial F_2}{\partial x} v_x + \frac{\partial F_2}{\partial y} v_y \right)$  can be described in terms of a matrix acting on column vectors as follows:

$$\begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial x} v_x + \frac{\partial F_1}{\partial y} v_y \\ \frac{\partial F_2}{\partial x} v_x + \frac{\partial F_2}{\partial y} v_y \end{bmatrix}.$$

Thus the derivative of  $\vec{F}$  at  $\vec{x}$  is the following matrix-valued function:

$$DF(\vec{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 4x - y & -x \\ -2^{-y/8} & \frac{1}{8} \log(2)(x + 2)2^{-y/8} \end{bmatrix}.$$

We've actually just found a general formula for computing derivatives.

### Computing Derivatives

Given a function  $\vec{F}(\vec{x})$  of  $n$  variables with  $m$  components, its derivative is the following  $[m \times n$  matrix]-valued function:

$$DF(\vec{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \frac{\partial F_m}{\partial x_2} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

**Another Example** Consider the following function of 3 variables with 2 components:

$$\vec{F}(x, y, z) = \begin{bmatrix} xy \\ yz \end{bmatrix}.$$

This function takes points from  $\mathbb{R}^3$  and gives points in  $\mathbb{R}^2$ , so you can think of it as being some way of crushing  $\mathbb{R}^3$  into the plane. Its derivative is given by

$$DF(x, y, z) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{bmatrix} = \begin{bmatrix} y & x & 0 \\ 0 & z & y \end{bmatrix}.$$

At the point  $(1, 1, 1)$  this takes the value

$$DF(1, 1, 1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

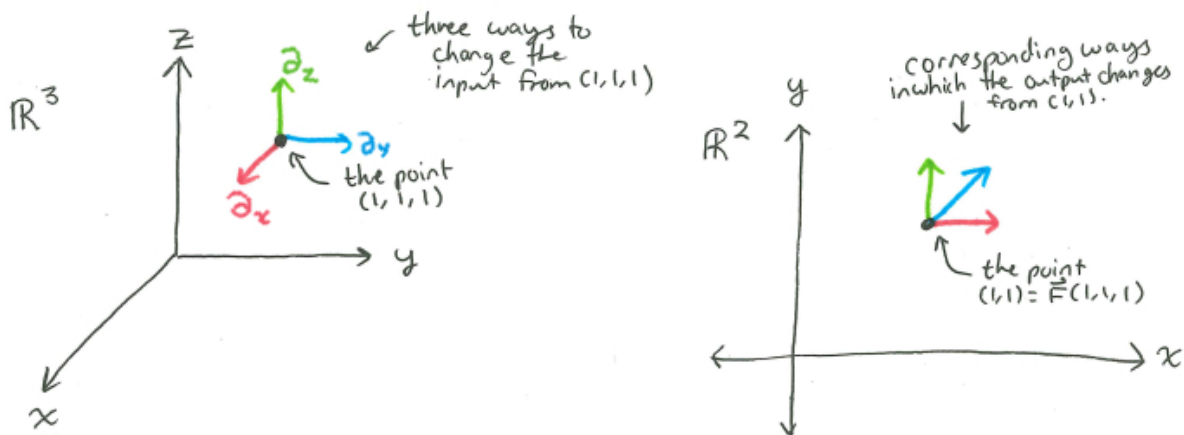
Here is perhaps an easier way to see what this matrix is doing to vectors at  $(1, 1, 1)$ :

$$DF(1, 1, 1) \partial_x = \partial_x$$

$$DF(1, 1, 1) \partial_y = \partial_x + \partial_y$$

$$DF(1, 1, 1) \partial_z = \partial_y.$$

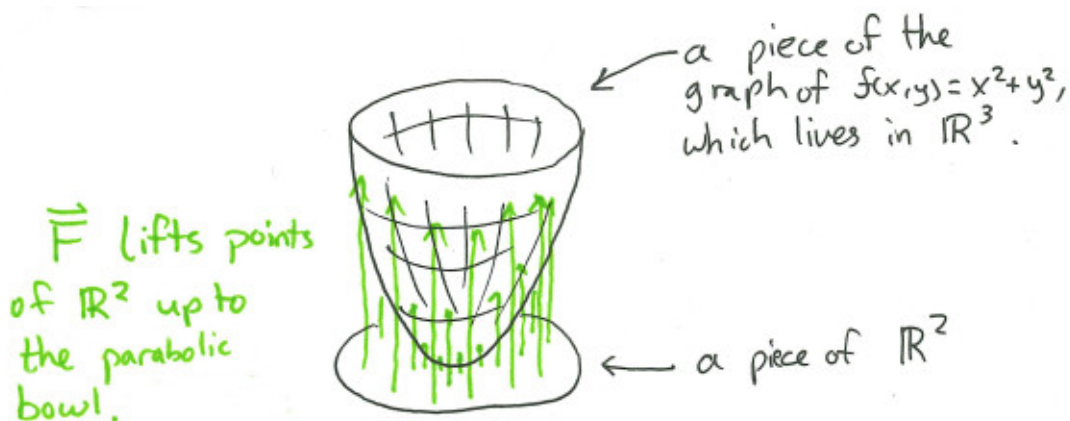
This is saying that if you start from  $(1, 1, 1)$  and start to increase the first coordinate, then the first coordinate of the output  $(F_1(\vec{x}), F_2(\vec{x}))$  would start to increase. And if you start to increase the third coordinate, then second coordinate of the output would start to increase. And if you start to increase the second coordinate, the the output responds by moving in the  $\partial_x + \partial_y$  direction (so both  $F_1$  and  $F_2$  would increase at the same rate).



**Another Example** Consider the following function of 2 variables with 3 components:

$$\vec{F}(x, y) = \begin{bmatrix} x \\ y \\ x^2 + y^2 \end{bmatrix}.$$

This function takes points from  $\mathbb{R}^2$  and gives points in  $\mathbb{R}^3$ , so you can think of it as being some way of inserting  $\mathbb{R}^2$  into  $\mathbb{R}^3$ . Actually, the output  $(x, y, x^2 + y^2)$  is the point above  $(x, y)$  on the graph of  $f(x, y) = x^2 + y^2$ . So this function is “lifting” points of  $\mathbb{R}^2$  onto a certain surface in  $\mathbb{R}^3$ , namely the parabolic bowl that we have seen is the graph of  $f(x, y) = x^2 + y^2$ .



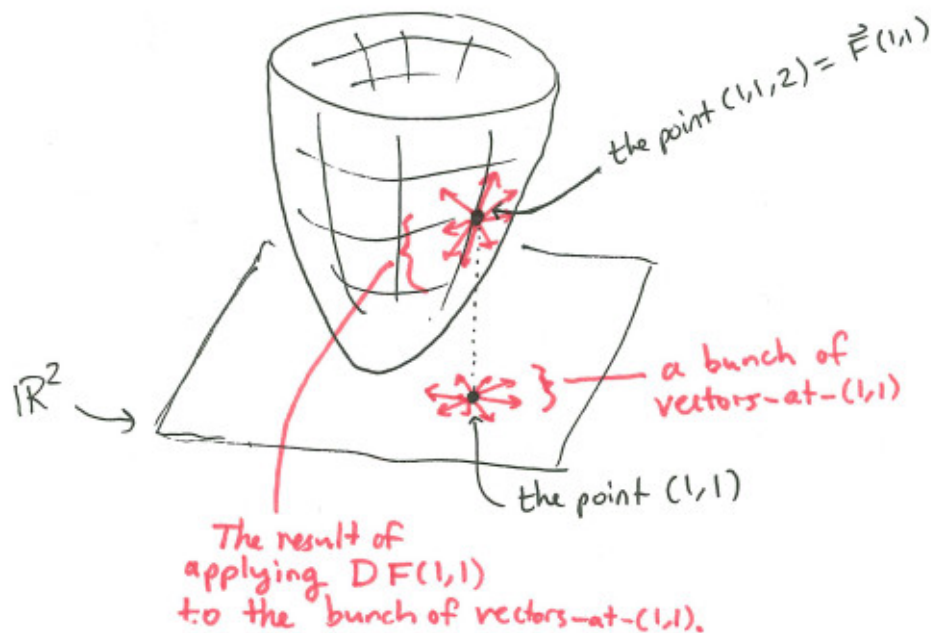
The derivative of  $\vec{F}(\vec{x})$  is given by

$$DF(x, y, z) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \\ \frac{\partial F_3}{\partial x} & \frac{\partial F_3}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2x & 2y \end{bmatrix}.$$

At the point  $(1, 1)$  this takes the value

$$DF(1, 1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}.$$

Since the output of  $\vec{F}$  is always on the parabolic bowl (the graph of  $f(x, y) = x^2 + y^2$ ), any path of the form  $\vec{F}(\vec{x}(t))$  must be a path that stays on parabolic bowl. Given any path  $\vec{x}(t)$  in  $\mathbb{R}^2$  that starts at  $(1, 1)$ , the path  $\vec{F}(\vec{x}(t))$  starts at  $\vec{F}(1, 1) = (1, 1, 2)$  and lives in the parabolic bowl. The velocity of a path living inside a surface is always tangent to that surface. So the initial velocity of  $\vec{F}(\vec{x}(t))$  is always some vector-at- $(1, 1, 2)$  that is tangent to the parabolic bowl. This explains what the matrix  $DF(1, 1)$  must do: it must “lift” (two dimensional) vectors-at- $(1, 1)$  to (three dimensional) vectors-at- $(1, 1, 2)$  that are tangent to the parabolic bowl.



### 3.4.2 Derivatives are Linear Approximations

Let  $\vec{F}(\vec{x})$  be a function of  $n$  variables with  $m$  components. Consider some particular input  $\vec{x}_0$  and a nearby input  $\vec{x}_0 + \Delta\vec{x}$ . If something is going to call itself “the derivative of  $\vec{F}$ ” then it sure as hell better have something to say about how  $\vec{F}(\vec{x}_0 + \Delta\vec{x})$  differs from  $\vec{F}(\vec{x}_0)$ . To see what it says, make up a path that starts at  $\vec{x}_0$  and heads in the  $\Delta\vec{x}$  direction in a straight line. Call the path  $\vec{x}(t)$  and

assume that  $\vec{x}(0) = \vec{x}_0$  and  $\vec{x}(\Delta t) = \vec{x}_0 + \Delta\vec{x}$ . The initial velocity  $\frac{d}{dt}\vec{F}(\vec{x}(t))|_{t=0}$  is (by definition of velocity) the limit of

$$\frac{1}{\Delta t}(\vec{F}(\vec{x}_0 + \Delta\vec{x}) - \vec{F}(\vec{x}_0))$$

as  $\Delta t \rightarrow 0$ . It is also (by definition of  $DF$ ) equal to  $DF(\vec{x}_0)(\frac{1}{\Delta t}\Delta\vec{x})$ , since the initial velocity of  $\vec{x}(t)$  is  $\frac{1}{\Delta t}\Delta\vec{x}$ . Hence we have

$$\frac{1}{\Delta t}(\vec{F}(\vec{x}_0 + \Delta\vec{x}) - \vec{F}(\vec{x}_0)) = DF(\vec{x}_0)(\frac{1}{\Delta t}\Delta\vec{x})$$

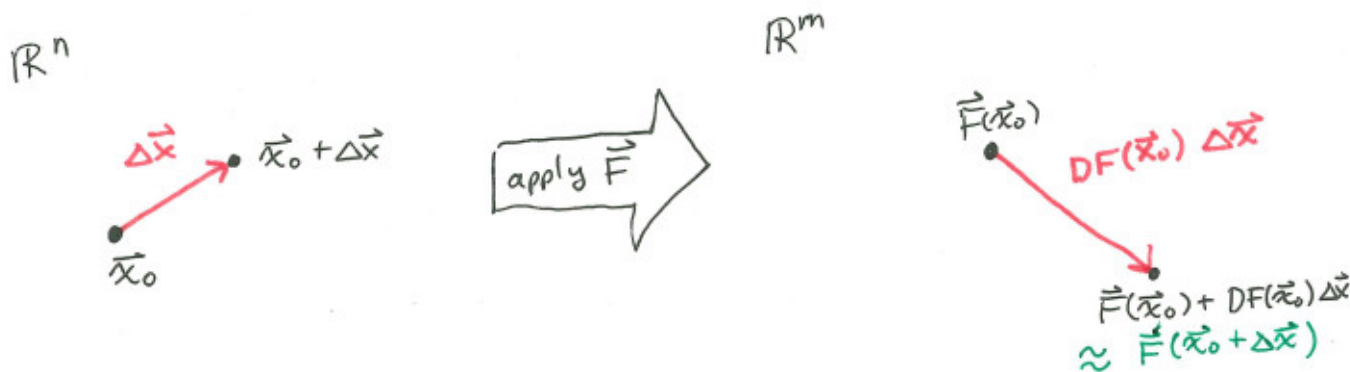
holding in the limit  $\Delta t \rightarrow 0$ , which is also the limit  $\Delta\vec{x} \rightarrow 0$ . The scalar  $\frac{1}{\Delta t}$  pulls out of matrix application:

$$DF(\vec{x}_0)(\frac{1}{\Delta t}\Delta\vec{x}) = \frac{1}{\Delta t}DF(\vec{x}_0)(\Delta\vec{x}),$$

so we can cancel the  $\frac{1}{\Delta t}$  above and we get

$$\vec{F}(\vec{x}_0 + \Delta\vec{x}) - \vec{F}(\vec{x}_0) = DF(\vec{x}_0)(\Delta\vec{x})$$

holding in the limit  $\Delta\vec{x} \rightarrow 0$ . This is telling us that near  $\vec{x}_0$ , the function  $\vec{F}$  acts approximately just like the *matrix*  $DF(\vec{x}_0)$  acts.



To be more precise, if you consider small vector displacements  $\Delta\vec{x}$  from  $\vec{x}_0$ , then the values that  $\vec{F}$  takes at those points are given by small vector displacements from  $\vec{F}(\vec{x}_0)$  that are approximately what you would get by applying the matrix  $DF(\vec{x}_0)$  to  $\Delta\vec{x}$ . The matrix  $DF(\vec{x}_0)$  represents the best *linear* function that captures the behavior of  $\vec{F}$  near  $\vec{x}_0$ .

### Derivatives are Linear Approximations

Given a function  $\vec{F}(\vec{x})$  of  $n$  variables with  $m$  components, the derivative matrix  $DF(\vec{x})$  at a point  $\vec{x}$  is the best linear approximation for the behavior of  $\vec{F}$  near  $\vec{x}$ . To first order in  $\Delta\vec{x}$ ,  $\vec{F}(\vec{x} + \Delta\vec{x}) \approx \vec{F}(\vec{x}) + DF(\vec{x})(\Delta\vec{x})$ .

**Example** Let's revisit the function of 2 variables with 2 components from the previous section:

$$\vec{F}(x, y) = \begin{bmatrix} 2x^2 - xy + 1 \\ -(x + 2)2^{-y/8} \end{bmatrix}.$$



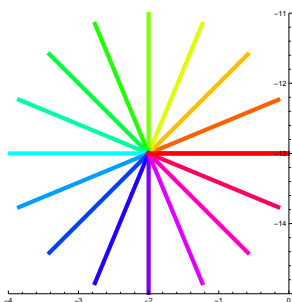
We found that

$$DF(\vec{x}) = \begin{bmatrix} 4x - y & -x \\ -2^{-y/8} & \frac{1}{8} \log(2)(x + 2)2^{-y/8} \end{bmatrix}.$$

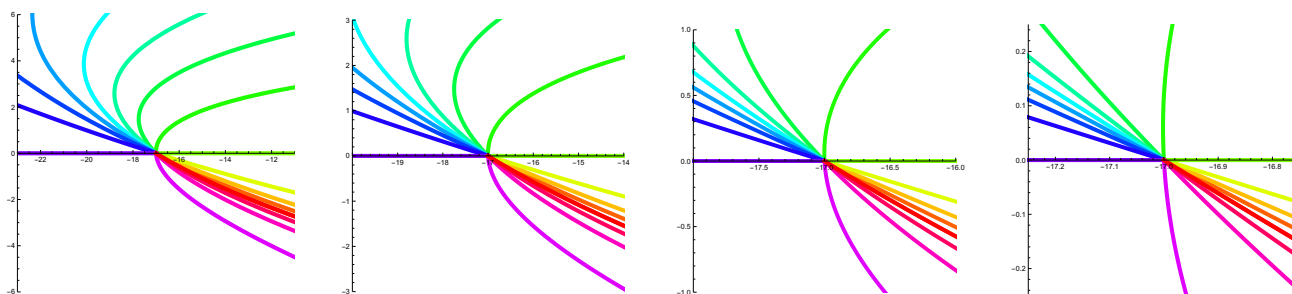
This time instead of  $(-2, -8)$ , let's plug in  $(x, y) = (-2, -13)$  (just to mix things up). We get

$$DF(-2, -13) = \begin{bmatrix} 5 & 2 \\ -2^{13/8} & 0 \end{bmatrix}.$$

This matrix represents the way in which  $\vec{F}(\vec{x})$  changes in response to changes in  $\vec{x}$  starting from  $(-2, -13)$ . It would be nice to see graphically how the action of this matrix on small displacement vectors  $\Delta\vec{x}$  is an approximation of  $\vec{F}(\vec{x})$  at points near  $\vec{x}_0 = (-2, -13)$ . To see it graphically, start with a bunch of paths (paths of inputs) radiating outward from  $(-2, -13)$ :



Then look at what  $\vec{F}$  does to those paths, and gradually zoom the picture in closer and closer to the point  $\vec{F}(-2, -13)$  from which the output paths emanate:



Eventually, once you zoom in close enough, the action of  $\vec{F}(\vec{x})$  on paths looks indistinguishable from the action of the matrix  $DF(-2, -13)$  on the initial velocity vectors-at- $\vec{F}(-2, -13)$ .

Exercise: If derivatives are really “best linear approximations”, then functions that are already linear should look like their own derivatives. Prove to yourself that if

$$\vec{F}(x, y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

then

$$DF(x, y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

for all  $(x, y)$ . This is analogous to the fact that  $\frac{d}{dx} ax = a$  in single variable calculus.

### 3.4.3 The Chain Rule

The chain rule is concerned with breaking down the derivative of a composite  $\vec{F}(\vec{G}(\vec{x}))$ . Such a composite only makes sense if the number of components of  $G$  matches the number of variables that  $F$  expects. Assume that  $\vec{G}(\vec{x})$  is a function of  $n$  variables with  $m$  components and  $\vec{F}(\vec{x})$  is a function of  $m$  variables with  $k$  components. Then how is the derivative of  $\vec{F}(\vec{G}(\vec{x}))$  related to  $DF$  and  $DG$ ?

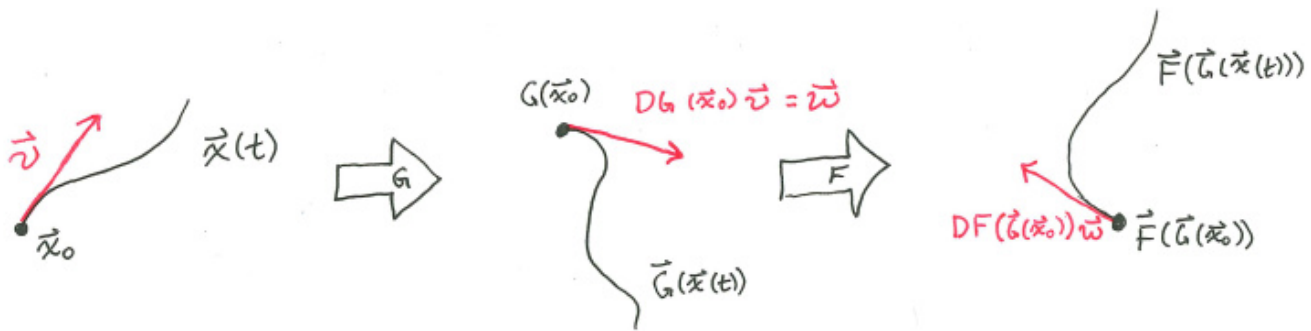
A quick notational aside. We usually refer to a function by a notation like “ $\vec{F}(\vec{x})$ ,” but this notation also makes it look like we are trying to evaluate  $F$  at a specific point  $\vec{x}$  to get a specific output  $\vec{F}(\vec{x})$ . If we want to refer purely to the function, without any specific input plugged into it, then we just write “ $F$ .” What if a function is built by composition like  $\vec{F}(\vec{G}(\vec{x}))$ ? Then there is still a way to refer purely to the composite function without any specific input plugged into it:  $F \circ G$ . The definition of the “ $\circ$ ” symbol is that  $F \circ G$  is the function whose value at  $\vec{x}$  is the value of  $F$  at  $\vec{G}(\vec{x})$ :

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

Okay, back to the main question: How is  $D(F \circ G)$  related to  $DF$  and  $DG$ ? By definition, the matrix functions  $DF$  and  $DG$  describe the effects of  $F$  and  $G$  on the initial velocities of paths. So let’s consider some path of inputs  $\vec{x}(t)$  starting at some  $\vec{x}_0$  with initial velocity  $\frac{d\vec{x}}{dt} |_{t=0} = \vec{v}$ . By the definition of  $D(F \circ G)$  we have

$$\begin{aligned} D(F \circ G)(\vec{x}_0)(\vec{v}) &= \frac{d}{dt} |_{t=0} (F \circ G)(\vec{x}(t)) \\ &= \frac{d}{dt} |_{t=0} \vec{F}(\vec{G}(\vec{x}(t))). \end{aligned}$$

The path  $\vec{G}(\vec{x}(t))$  of outputs of  $G$  is being used here as a path of inputs for  $F$ .



Since the path  $\vec{G}(\vec{x}(t))$  starts at  $\vec{G}(\vec{x}_0)$  with initial velocity

$$\frac{d}{dt} |_{t=0} \vec{G}(\vec{x}(t)) = DG(\vec{x}_0)(\vec{v}),$$

inputting the path  $\vec{G}(\vec{x}(t))$  into  $F$  gives the initial velocity

$$\frac{d}{dt} |_{t=0} \vec{F}(\vec{G}(\vec{x}(t))) = DF(\vec{G}(\vec{x}_0))(DG(\vec{x}_0)(\vec{v})).$$

And so this becomes our chain rule:

$$D(F \circ G)(\vec{x}_0)(\vec{v}) = DF(\vec{G}(\vec{x}_0))(DG(\vec{x}_0)(\vec{v})).$$

The right hand side is a product of matrices– it starts with the vector  $\vec{v}$ , then applies the matrix  $DG(\vec{x}_0)$  to that, and then applies the matrix  $DF(\vec{G}(\vec{x}_0))$  to *that*. This is the same as just taking the product matrix  $DF(\vec{G}(\vec{x}_0)) \cdot DG(\vec{x}_0)$  and applying it to  $\vec{v}$ . So the equality above says that the matrices  $D(F \circ G)(\vec{x}_0)$  and  $DF(\vec{G}(\vec{x}_0)) \cdot DG(\vec{x}_0)$  give the same result when you apply them to any vector  $\vec{v}$ . This implies that they are equal as matrices. Finally, we have found the chain rule:

### The Chain Rule

Assume that  $\vec{G}(\vec{x})$  is a function of  $n$  variables with  $m$  components and  $\vec{F}(\vec{x})$  is a function of  $m$  variables with  $k$  components. Then the derivative of  $\vec{F}(\vec{G}(\vec{x}))$  is related to the  $DF$  and  $DG$  in the following way:

$$D(F \circ G)(\vec{x}_0) = DF(\vec{G}(\vec{x}_0)) \cdot DG(\vec{x}_0).$$

The left hand side is a  $k \times n$  matrix, and the right hand side is the product of a  $k \times m$  matrix with an  $m \times n$  matrix.

**Example** Consider the functions

$$\vec{G}(x, y) = \begin{bmatrix} x \\ y \\ x^2 + y^2 \end{bmatrix}$$

and

$$\vec{F}(x, y, z) = \begin{bmatrix} xy \\ yz \end{bmatrix}$$

from some earlier examples. The composite  $\vec{F}(\vec{G}(x, y))$  is a function of 2 variables with 2 components given as follows:

$$\vec{F}(\vec{G}(x, y)) = \vec{F}(x, y, x^2 + y^2) = \begin{bmatrix} xy \\ y(x^2 + y^2) \end{bmatrix} = \begin{bmatrix} xy \\ x^2y + y^3 \end{bmatrix}.$$

We can take the derivative directly:

$$D(F \circ G)(x, y) = \begin{bmatrix} y & x \\ 2xy & x^2 + 3y^2 \end{bmatrix}.$$

Or we can use the chain rule. First, we compute all the ingredients that go into the chain rule:

$$\begin{aligned} DG(x, y) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2x & 2y \end{bmatrix} \\ DF(x, y, z) &= \begin{bmatrix} y & x & 0 \\ 0 & z & y \end{bmatrix} \\ DF(\vec{G}(x, y)) &= DF(x, y, x^2 + y^2) = \begin{bmatrix} y & x & 0 \\ 0 & x^2 + y^2 & y \end{bmatrix} \end{aligned}$$

And then we can apply the chain rule:

$$D(F \circ G)(x, y) = DF(\vec{G}(x, y)) \cdot DG(x, y) = \begin{bmatrix} y & x & 0 \\ 0 & x^2 + y^2 & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2x & 2y \end{bmatrix} = \begin{bmatrix} y & x \\ 2xy & x^2 + 3y^2 \end{bmatrix}.$$

**Special Cases  $f(\vec{x})$  and  $\vec{p}(t)$**  I will show you that the notion of derivative defined in this section has all previous notions (velocities and gradients) as special cases. Also, the version of the chain rule that we saw in section 1.7.3 is a special case of the general chain rule we are seeing here.

A function  $f(\vec{x})$  on  $\mathbb{R}^n$  is a function of  $n$  variables with just *one* component. So its derivative is a  $1 \times n$  matrix-valued function:

$$Df = \left[ \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right].$$

You could call this single-row matrix the “gradient” of  $f(\vec{x})$ , as long as you know that this row vector stands for

$$\frac{\partial f}{\partial x_1} \partial_{x_1} + \cdots + \frac{\partial f}{\partial x_n} \partial_{x_n},$$

and as long as  $x_1, \dots, x_n$  are *cartesian* coordinates. (This discussion can still work if they are not cartesian coordinates, but it’s more complicated.)

A path  $\vec{p}(t)$  in  $\mathbb{R}^n$  is a function of 1 variable with  $n$  components. So its derivative is an  $n \times 1$  matrix-valued function:

$$Dp = \begin{bmatrix} \frac{dp_1}{dt} \\ \vdots \\ \frac{dp_n}{dt} \end{bmatrix}.$$

We are already used to writing a list of components for the velocity of a path, knowing that it really stands for  $\frac{dp_1}{dt} \partial_{x_1} + \cdots + \frac{dp_n}{dt} \partial_{x_n}$ .

Now consider what the chain rule has to say about the derivative of a function  $f(\vec{x})$  along a path  $\vec{p}(t)$ :

$$\begin{aligned} D(f \circ p)(t) &= Df(\vec{p}(t)) \cdot Dp(t) \\ &= \left[ \frac{\partial f}{\partial x_1} \Big|_{\vec{p}(t)} \quad \cdots \quad \frac{\partial f}{\partial x_n} \Big|_{\vec{p}(t)} \right] \cdot \begin{bmatrix} \frac{dp_1}{dt} \\ \vdots \\ \frac{dp_n}{dt} \end{bmatrix} \\ &= \left[ \frac{\partial f}{\partial x_1} \Big|_{\vec{p}(t)} \frac{dp_1}{dt} + \cdots + \frac{\partial f}{\partial x_n} \Big|_{\vec{p}(t)} \frac{dp_n}{dt} \right]. \end{aligned}$$

The result is a  $1 \times 1$  matrix with the one entry being exactly what we would expect from the chain rule that we studied in section 1.7.3.

### 3.4.4 The Jacobian Matrix of a Coordinate Transformation

I mentioned at the beginning of this section that *coordinate transformations* are the main examples in this course of functions of multiple variables with multiple components. The transformation from

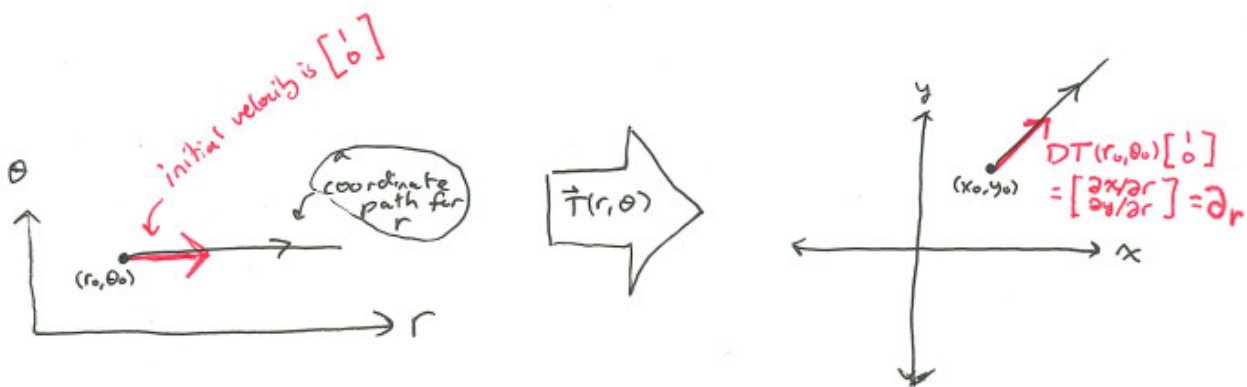
polar to cartesian coordinates is the following function of 2 variables and 2 components:

$$\vec{T}(r, \theta) = \begin{bmatrix} x(r, \theta) \\ y(r, \theta) \end{bmatrix} = \begin{bmatrix} r \cos(\theta) \\ r \sin(\theta) \end{bmatrix}.$$

The derivative is the following  $2 \times 2$  matrix-valued function:

$$DT = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{bmatrix}.$$

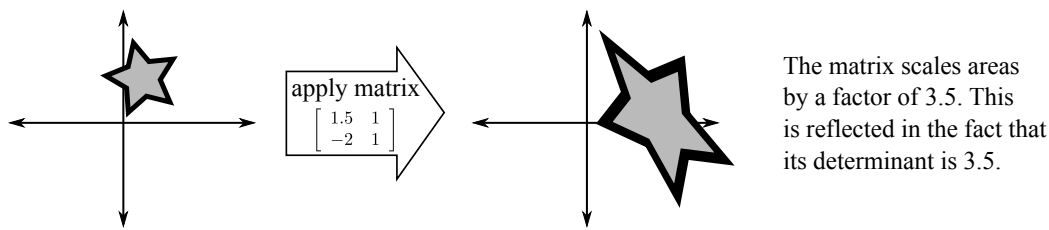
The derivative of a coordinate transformation has a special name: it's called the *Jacobian matrix* of the transformation. The partial derivatives appearing as the matrix entries are probably looking familiar to you— they are the coefficients in the expansion of  $\partial_r$  and  $\partial_\theta$  in terms of  $\partial_x$  and  $\partial_y$ . This is no coincidence! Remember that  $\partial_r$  is, at each point, the initial velocity of a coordinate path in the  $r$  direction. Such a coordinate path could be considered to be a path of inputs for the transformation function  $\vec{T}(r, \theta)$ , and its initial velocity  $\partial_r$  is represented by the column vector  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ . Applying the transformation  $\vec{T}(r, \theta)$  to the coordinate path of  $r$  would give us an expression for coordinate path for  $r$  in terms of  $x, y$  coordinates. The components of velocity for that path are what you get when you apply  $DT$  to  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ; that is, they are the entries of the first column of the matrix  $DT$ .



So the Jacobian matrix of a coordinate transformation contains within it the information needed to transform vectors from one coordinate system to the other. Actually, we could have derived the vector transformation law in this way to start with.

Another useful aspect of a Jacobian matrix is its *determinant*, which we can refer to as the *Jacobian determinant* of the coordinate transformation.

Recall from linear algebra that the determinant of a linear transformation is the scale factor by which it scales volumes (or areas, or whatever). For example:

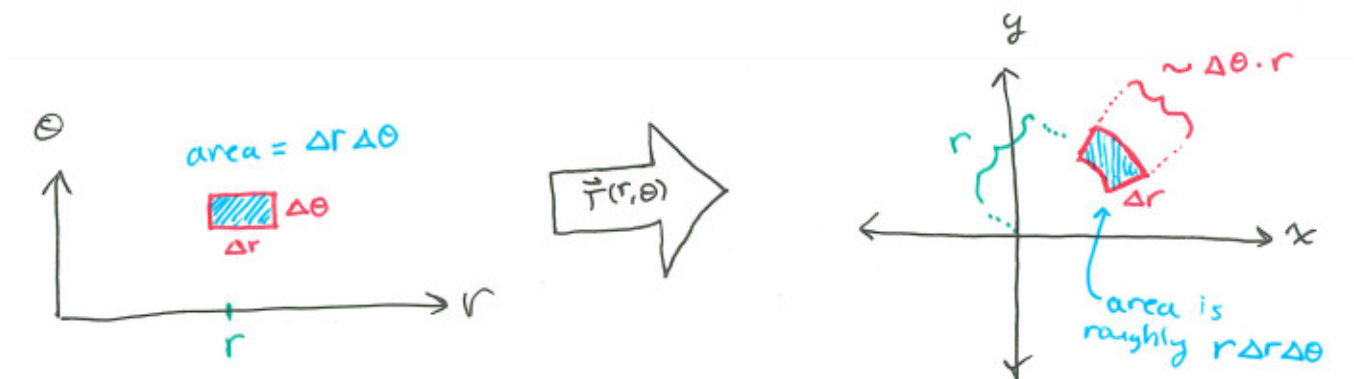


Determinants can be negative, in which case you should take absolute value to get the scale factor by which volumes/areas are scaled.

Since the derivative of a function is a linear approximation of the function at each point, the determinant of the Jacobian matrix  $DT$  ends up giving us the scale factor, at each point, by which the coordinate transformation distorts volumes. Let's use polar coordinates as an example. The Jacobian determinant is

$$\det(DT) = \det \left( \begin{bmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{bmatrix} \right) = r \cos^2(\theta) - (-r \sin^2(\theta)) = r.$$

This is telling us about the area distortion caused by polar coordinates. The bigger  $r$  is (the farther from the origin you are looking), the more the transformation  $\vec{T}$  scales up areas. Here's a picture showing you how a little box of side lengths  $\Delta r$  and  $\Delta \theta$  gets its area scaled by the amount  $\det(DT) = r$ :



The box does have to be small for the approximate scale factor to be accurate, because the derivative matrix captures the “infinitesimal” aspect of the behavior of the transformation near each point.

Summary:

## Jacobian Matrix and Jacobian Determinant

The derivative of a coordinate transformation is called its *Jacobian matrix*. If the coordinate transformation functions are  $u_1(x_1, \dots, x_n), \dots, u_n(x_1, \dots, x_n)$ , then the Jacobian matrix is given by

$$\begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial u_n}{\partial x_1} & \cdots & \frac{\partial u_n}{\partial x_n} \end{bmatrix}.$$

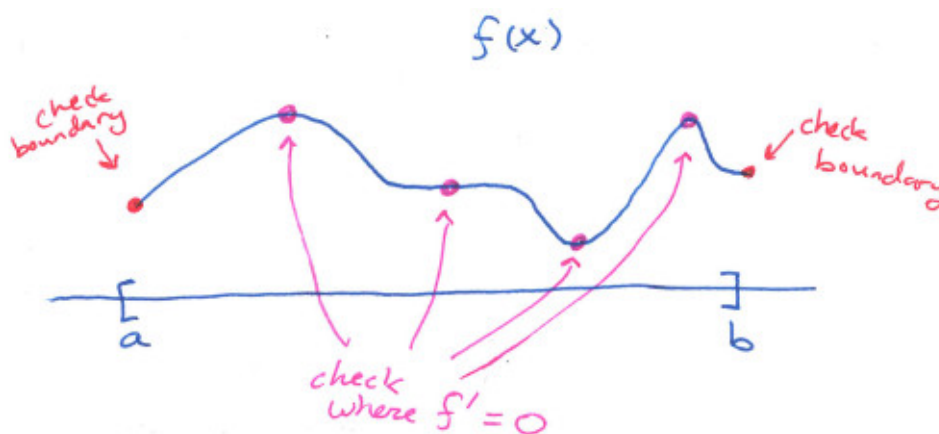
The determinant of this matrix is called the *Jacobian determinant* of the coordinate transformation. Since the Jacobian matrix is a matrix-valued function, the Jacobian determinant is a scalar function. The absolute value of the Jacobian determinant is the scale factor by which the coordinate transformation distorts volumes at each point.

## 3.5 Optimization

This section is about extremizing functions of multiple variables. The question we are trying to answer is: *Given a function  $f(x, y)$  of two variables, how can we choose inputs  $x$  and  $y$  such that the output  $f(x, y)$  is maximized or minimized?* Most of our discussion will apply to functions of any number of variables, but we will focus on functions of only two variables in the end.

**Single Variable Optimization** First, we recall how optimization goes in single variable calculus. Say you've got a function  $f(x)$  defined for  $x$  in the interval  $[a, b]$ , and you want to find where  $f(x)$  takes on a maximum value. If you were some kind of oracle-like superbeing that could gather and analyze uncountably infinite amounts of information, then you could take the following approach: check the value of  $f(x)$  for each  $x$  in  $[a, b]$  and then pick out the  $x$  that gets you the biggest output. But that's silly—you don't need to check every possible  $x$  because you know calculus. If  $f'(x)$  were positive, then  $f(x + \Delta x)$  is bigger than  $f(x)$  for a sufficiently small  $\Delta x$ . And if  $f'(x)$  were negative then  $f(x - \Delta x)$  is bigger than  $f(x)$  for a sufficiently small  $\Delta x$ . So unless  $f'(x)$  is zero, the input  $x$  could not possibly be an input that maximizes the output. Thus it suffices in our search for maxima to only consider values of  $x$  for which  $f'(x) = 0$ . These inputs are special, so we give them a name: they are *critical points*. They're not always maxima or minima, but at least we know the maxima and minima are among them and so they are all we have to check. This is what leads to the calculus-based strategy for optimization of a function: solve  $f'(x) = 0$  to get critical points, then analyze critical points to see which ones are maximizing or minimizing values.

There's a catch in the argument above with the  $\Delta x$ 's. The argument doesn't work when  $x = a$  or  $x = b$  (the "boundary case"). So just those two points have to be analyzed specially.



**Multivariable Optimization** Optimization for a function of multiple variables will end up working pretty much the same way in that we set a derivative equal to zero to locate critical points. The main difference is that there's a bit more finesse needed to actually analyze the critical points.

Let me start by clarifying what I mean by maxima and minima:

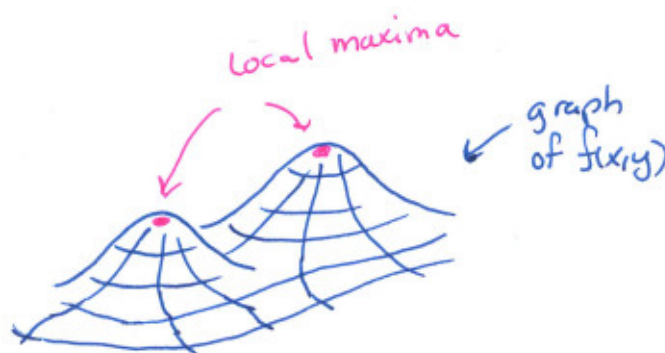
#### Definition of Local Extrema

A function  $f(\vec{x})$  is said to have a *local maximum* at  $\vec{x}_0$  if  $f(\vec{x}) \leq f(\vec{x}_0)$  for all  $\vec{x}$  in some neighborhood of  $\vec{x}_0$  and in the domain of  $f$ .

A function  $f(\vec{x})$  is said to have a *local minimum* at  $\vec{x}_0$  if  $f(\vec{x}) \geq f(\vec{x}_0)$  for all  $\vec{x}$  in some neighborhood of  $\vec{x}_0$  and in the domain of  $f$ .

Local minima and local maxima are both called *local extrema*.

The phrase “a neighborhood of  $\vec{x}_0$ ” precisely means “for some positive  $r$ , some bunch of points that are within a distance  $r$  of  $\vec{x}_0$ .” The key thing to notice is the word *local*. We are interested in all the little hilltops and ditches in the graph of  $f$ , not just the tallest peaks and lowest valleys.



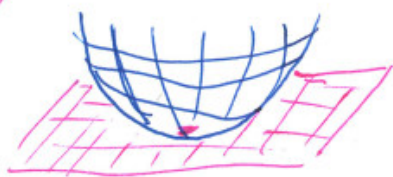
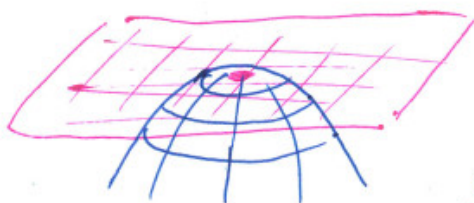
The first observation we have to make in order to locate extrema is that they can only occur for inputs  $\vec{x}_0$  where  $\vec{\nabla} f(\vec{x}_0) = \vec{0}$ . If  $\vec{\nabla} f(\vec{x}_0)$  were not zero then traveling from  $\vec{x}_0$  in the direction  $\vec{\nabla} f(\vec{x}_0)$  would immediately start to increase the values of  $f$ , while traveling in the direction  $-\vec{\nabla} f(\vec{x}_0)$  would immediately start to decrease the values of  $f$ , so  $\vec{x}_0$  could not possibly be a local extremum. This is the key observation for optimization in both single-variable and multi-variable calculus. Okay,



but the argument we made only applies in a situation where you can just start from  $\vec{x}_0$  and head in any direction freely and still be able to look at values of  $f$ — so the argument doesn't apply when  $\vec{x}_0$  is on the boundary of the domain of  $f$ .

Local extrema can only be at critical points or boundary points

If  $\vec{x}_0$  is a local extremum for  $f(\vec{x})$ , then it must be either that  $\vec{\nabla}f(\vec{x}_0) = \vec{0}$  or that  $\vec{x}_0$  is on the boundary of the domain of  $f$ .



The tangent plane to the graph of  $f(x,y)$  is flat at local minima and maxima. This is because  $\vec{\nabla}f$  vanishes; i.e. all directional derivatives vanish.

This makes the following terminology useful:

Definition of critical point

If  $\vec{\nabla}f(\vec{x}_0) = \vec{0}$  then  $\vec{x}_0$  is called a *critical point* for  $f$ .

So we now have a strategy for finding the extrema of a function  $f(x,y)$  of two variables:

1. Compute  $\vec{\nabla}f$ .
2. Solve the system of equations  $\vec{\nabla}f = \vec{0}$  to get critical points.
3. Analyze each critical point to find out if it is a local minimum, a local maximum, or neither.
4. Don't forget to check the boundary of the domain of  $f$  for other local extrema.

It's all pretty straightforward, except for step 3. Next, we look into how to carry out step 3.

**Analyzing Critical Points** In single-variable calculus you do this by checking the second derivative: if  $x_0$  is a critical point of  $f(x)$  then

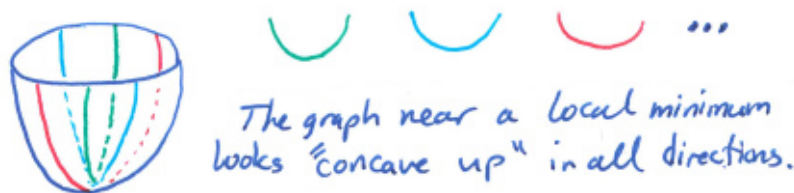
- if  $f''(x_0) > 0$ , then  $x_0$  is a local minimum,
- if  $f''(x_0) < 0$ , then  $x_0$  is a local maximum,
- if  $f''(x_0) = 0$ , then further analysis is needed to find out what's going on, but  $x_0$  might not be a local extremum at all.

In our multi-variable situation there are a lot of things that could be called “the second derivative” For a function  $f(x, y)$  there are three partial second derivatives  $\frac{\partial^2 f}{\partial x^2}$ ,  $\frac{\partial^2 f}{\partial x \partial y}$  and  $\frac{\partial^2 f}{\partial y^2}$ . And those are actually just special cases of “double directional derivatives” like  $\text{Dir}(\text{Dir}(f, (x, y), \vec{v}), (x, y), \vec{w})$ . Just like the gradient of a function is a vector object that holds within it all derivative information for all directions, there is an object that holds within it all *double*-derivative information for all *pairs* of directions. It is called the *Hessian matrix* of  $f(x, y)$ , and it looks like this in cartesian coordinates:

$$Hf(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

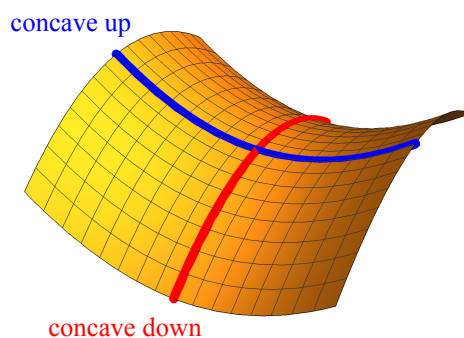
The double directional derivative  $\text{Dir}(\text{Dir}(f, (x, y), \vec{v}), (x, y), \vec{w})$  ends up being the matrix product  $\vec{w}^T(Hf)\vec{v}$ .

If  $\vec{x}_0 = (x_0, y_0)$  is a critical point of  $f(x, y)$ , it would definitely be a local minimum if  $f$  looked “concave up” in all directions  $\vec{v}$  at  $\vec{x}_0$ . This corresponds to all double directional derivatives  $\text{Dir}(\text{Dir}(f, (x, y), \vec{v}), (x, y), \vec{v})$  being positive, which corresponds to the matrix product  $\vec{v}^T(Hf)\vec{v}$  being positive for all nonzero column vectors  $\vec{v}$ . This would be an annoying condition to check were it not for a wonderfully easy criterion: If either of the diagonal entries of the matrix  $Hf(x_0, y_0)$  are positive and the *determinant* of the matrix is positive, then  $\vec{v}^T(Hf)\vec{v}$  is always positive for nonzero  $\vec{v}$ .



There’s a similar analysis of the local maximum situation, in which one looks for “concave-downness” in all directions at  $\vec{x}_0$ .

But here comes the main subtlety of multiple variables: you could have “concave up” behavior in some directions from  $\vec{x}_0$  while also having “concave down” behavior in others. The graph of a function near such a point looks like a saddle shape (or a pringle potato chip), so in this case  $\vec{x}_0$  would be called a *saddle point* for  $f$ . It’s definitely not a local extremum, because  $f$  increases in some directions and decreases in others. It turns out that detecting a saddle point is easy in two dimensions:  $\vec{x}_0 = (x_0, y_0)$  is a saddle point if the determinant of the Hessian matrix is negative.



You might think that if  $\frac{\partial^2 f}{\partial x^2}$  and  $\frac{\partial^2 f}{\partial y^2}$  were both positive at the critical point  $\vec{x}_0$  then the function looks concave up in every direction. But no! Even with both of those double partials being positive, there could be *other* directions other than the  $\partial_x$  and  $\partial_y$  directions in which the function looks concave down. You really do need to check the determinant of the Hessian matrix to know for sure whether or not you are looking at a saddle point.

Here is the multi-variable second derivative test, in full, for functions of two variables:

#### Analyzing Critical Points Using Second Derivatives

Assume that  $\vec{x}_0 = (x_0, y_0)$  is a critical point for  $f(x, y)$ . Let

$$h = \det(\text{H}f(\vec{x}_0)) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2$$

be the determinant of the Hessian matrix at  $\vec{x}_0$ . Then

- if  $\frac{\partial^2 f}{\partial x^2} |_{\vec{x}_0} > 0$  and  $h > 0$ , then  $\vec{x}_0$  is a local minimum,
- if  $\frac{\partial^2 f}{\partial x^2} |_{\vec{x}_0} < 0$  and  $h > 0$ , then  $\vec{x}_0$  is a local maximum,
- if  $h < 0$  then  $\vec{x}_0$  is a saddle point,
- if  $h = 0$  then further analysis is needed to find out what is going on.

**An Example** Let  $f(x, y) = x^2 - y^2$ , defined on all of  $\mathbb{R}^2$ . What are the local extrema? There is no boundary of the domain to worry about, so we only need to consider critical points of  $f(x, y)$ . The gradient is  $\vec{\nabla} f = 2x \partial_x - 2y \partial_y$ , so critical points can be found by solving the system of equations

$$\begin{aligned} 2x &= 0 \\ -2y &= 0. \end{aligned}$$

The only solution is  $(x, y) = (0, 0)$ , so the only critical point is  $(0, 0)$ . This implies that the only place where a local extremum can occur for  $f(x, y)$  is at  $(0, 0)$ , if it even occurs there. Now we need to analyze what kind of critical point  $(0, 0)$  is: is it a local minimum, a local maximum, a saddle

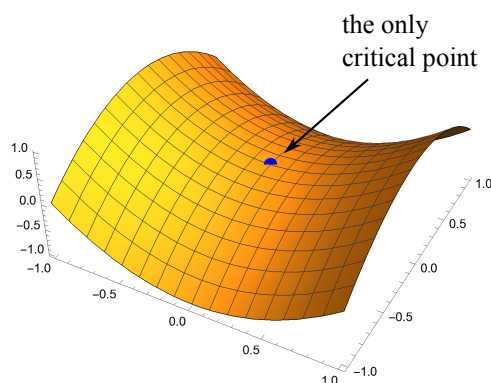
point, or something else? For this we compute the Hessian matrix for  $f(x, y)$ :

$$Hf(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

We need the determinant at the critical point:

$$\det(Hf(0, 0)) = -4.$$

Since it is negative, we conclude that the critical point  $(0, 0)$  is a saddle point. Thus  $f(x, y)$  has no local extrema. Here's the graph of  $f(x, y)$ :



**Another Example** Let  $f(x, y) = x^2 + y^2 - 2x - 4y + 6$ , but assume that it is only defined for  $x, y \geq 0$ . Since the function is only defined in the first quadrant of the plane, there will be boundary points to worry about. Let's just focus on finding critical points to start with. The gradient is  $\vec{\nabla} f = (2x - 2) \partial_x + (2y - 4) \partial_y$ , so critical points can be found by solving the system of equations

$$2x - 2 = 0$$

$$2y - 4 = 0.$$

The only solution is  $(x, y) = (1, 2)$ , so the only critical point is  $(1, 2)$ . And this is indeed a valid input to  $f$  because  $(1, 2)$  is in the first quadrant, the domain of the function. Now we need to analyze what kind of critical point  $(1, 2)$  is: is it a local minimum, a local maximum, a saddle point, or something else? For this we compute the Hessian matrix for  $f(x, y)$ :

$$Hf(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

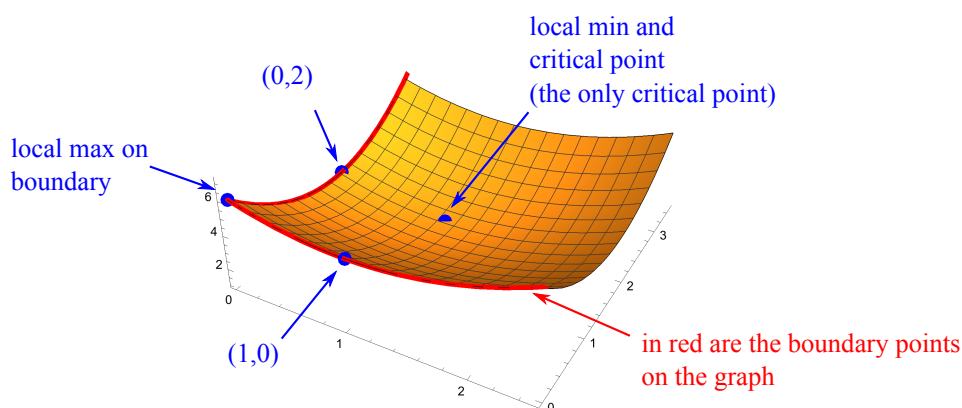
We need the determinant at the critical point:

$$\det(Hf(1, 2)) = 4.$$

Since it is positive, we conclude that the critical point  $(1, 2)$  is some kind of local extremum— either a minimum or a maximum. Since  $\frac{\partial^2 f}{\partial x^2} |_{(1,2)} = 2$  is positive, we conclude that  $(1, 2)$  is a local *minimum*.

So far we've found one local extremum, but there could be further extrema on the boundary of the domain of  $f(x, y)$ . The boundary of the domain is the set of points where either  $x = 0$  and  $y > 0$  or  $y = 0$  and  $x > 0$ . Let  $g(y) = f(0, y) = y^2 - 4y + 6$ , defined for  $y \geq 0$ . And let  $h(x) = f(x, 0) = x^2 - 2x + 6$ , defined for  $x \geq 0$ . These are functions that take on the values of  $f$  along the boundary. We can find potential extrema of  $f$  along the boundary by extremizing  $g$  and  $h$ . Since  $g'(y) = 2y - 4$ , the function  $g(y)$  has one critical point  $y = 2$ . Since  $g''(2) = 2 > 0$ , the critical point  $y = 2$  is a local minimum for  $g(y)$ . This suggests that  $(0, 2)$  might be a local minimum (though not a critical point!) for  $f(x, y)$ . But it's not, because you can *decrease* the output of  $f(x, y)$  by traveling in the  $\partial_x$  direction from  $(0, 2)$ :  $\text{Dir}(f, (0, 2), \partial_x) = \frac{\partial f}{\partial x} |_{(0,2)} = -2 < 0$ . Since  $h'(x) = 2x - 2$ , the function  $h(x)$  has one critical point  $x = 1$ . Since  $h''(1) = 2 > 0$ , the critical point  $x = 1$  is a local minimum for  $h(x)$ . This suggests that  $(1, 0)$  might be a local minimum for  $f(x, y)$ . But it's not, because you can *decrease* the output of  $f(x, y)$  by traveling in the  $\partial_y$  direction from  $(1, 0)$ :  $\text{Dir}(f, (1, 0), \partial_y) = \frac{\partial f}{\partial y} |_{(1,0)} = -4 < 0$ . Finally, both  $g(y)$  and  $h(x)$  have a boundary to their own domains! It is  $y = 0$  and  $x = 0$ , respectively. So the origin  $(0, 0)$  just needs to be treated specially. If  $v_x$  and  $v_y$  are positive, then  $\text{Dir}(f, (0, 0), v_x \partial_x + v_y \partial_y)$  is  $\frac{-2v_x - 4v_y}{\sqrt{v_x^2 + v_y^2}}$ , which is negative. So the function decreases in every direction heading away from the origin into the first quadrant of the plane. Thus  $(0, 0)$  is a local maximum.

Here's a picture of the graph of  $f(x, y)$  pointing out the two local extrema that we found:



As you can see, the critical points are usually easy to analyze, while boundary points can be quite tricky.

## 4 Integrals

In this section we will explore several types of integrals, starting with the one most directly analogous to integration in single variable calculus.

## 4.1 Integrals of Functions over Regions

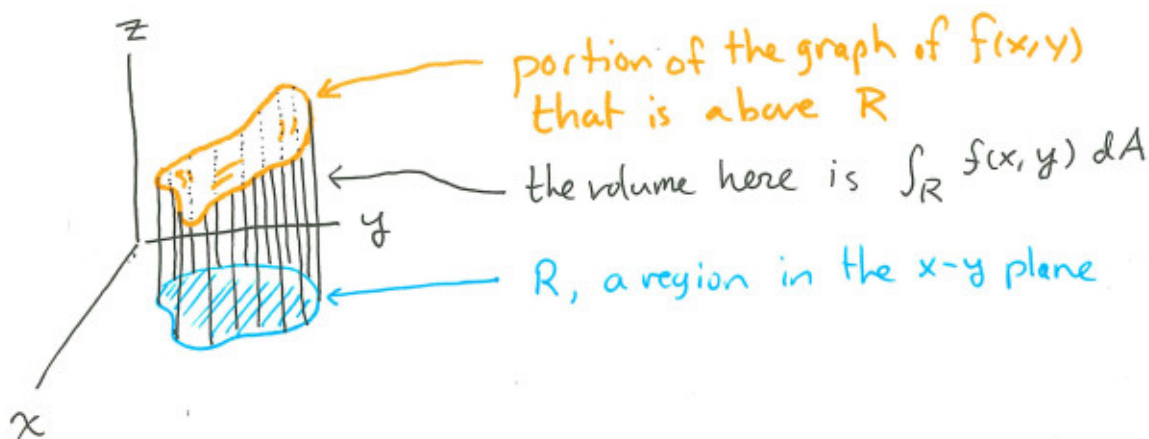
Recall from single variable calculus that  $\int_a^b f(x) dx$  is the area “under the graph” of the function  $f(x)$ , which we can approximate by adding up areas of rectangles. The “adding up rectangles” approximation gets closer and closer to the true value of  $\int_a^b f(x) dx$  as one uses more and narrower rectangles. The definition of  $\int_a^b f(x) dx$  is that it is the limit of the approximations by adding up rectangles.



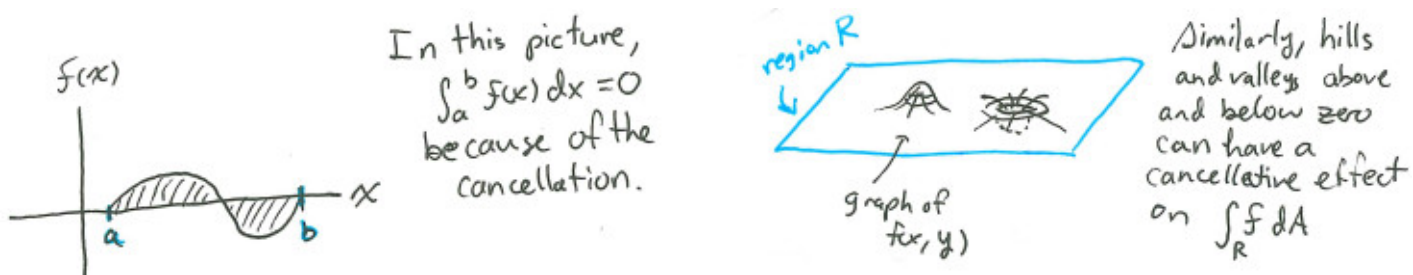
The integral of a function of two variables is defined in a very similar way. For a function  $f(x, y)$  on  $\mathbb{R}^2$ , the graph is a surface in  $\mathbb{R}^3$ . The integral is the volume “under the surface.” Since the domain is  $\mathbb{R}^2$  instead of  $\mathbb{R}^1$ , a *region of integration* can be much more complicated than a simple interval like  $[a, b]$ . If the region of integration is some piece  $R$  of  $\mathbb{R}^2$ , then the integral of  $f(x, y)$  on  $R$  is denote by

$$\int_R f(x, y) dA$$

and it is the volume “under the graph” of  $f(x, y)$  within the region  $R$ . The “ $dA$ ” indicates that this is an *area integral*, i.e. an integral of a function over some region in  $\mathbb{R}^2$ .

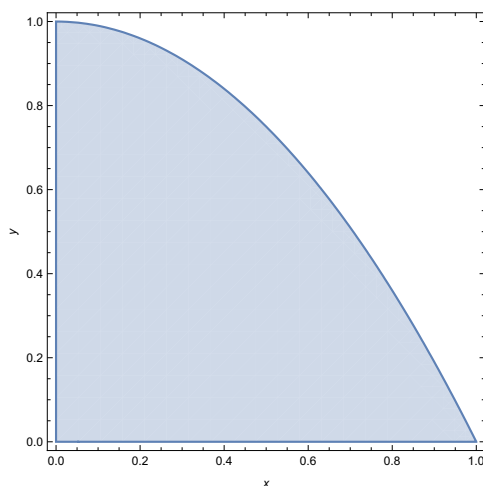


I’ve been putting scare-quotes around the phrase “under the graph” because it’s not strictly correct: any negative values taken by the function in the region of integration contribute “negative area/volume” to the integral.

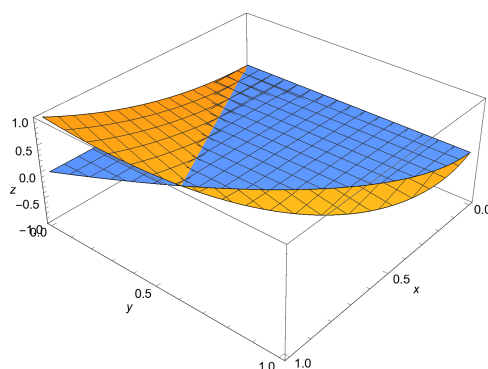


The way to actually compute an area integral is as a *double integral*— that is, as two single integrals. You can slice up the region along one coordinate, do the single-variable integral along all those slices, and then integrate the result along the other variable. It's easy to say but somewhat tricky to do correctly— the main challenge here is to describe the region of integration in a way that makes the slicing work. An example will demonstrate.

**Example** Let  $f(x, y) = x^2 - 2xy$ . Let's try to compute the integral of  $f$  over the region bounded by  $y = 1 - x^2$  in the first quadrant. Let's call this region  $R$ . Here is a picture of  $R$ :

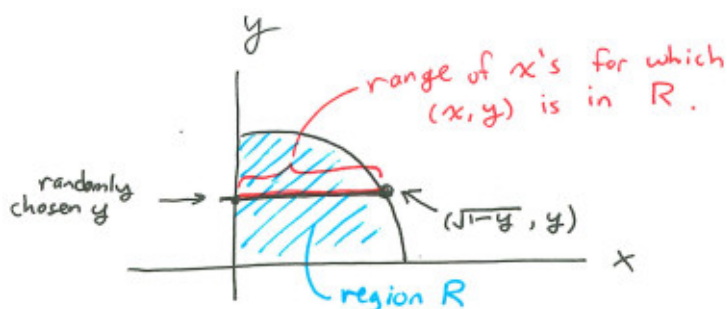


Here the graph of the function:



The graph is shown in orange, and in blue you see the surface  $z = 0$ . The picture only includes information above the region  $R$ , since that's all we care about here. We want to compute  $\int_R f dA$ ,

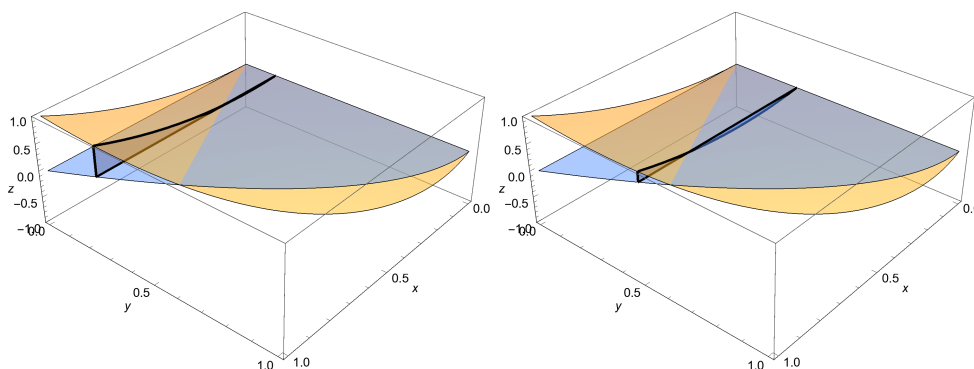
which is the “volume” between the orange and blue surfaces in the picture (keep in mind that for the portion where the orange surface dips below the blue surface, there is a negative contribution to the overall “volume”). For each fixed value of  $y$ , we can view  $f(x, y) = x^2 - 2xy$  as a function of  $x$  alone. (For example if you fix  $y$  to be 0.5, then you get the function  $x^2 - x$  of  $x$  alone). We can integrate this function of one variable with respect to  $x$  to get “ $\int f(x) dx$ ” as a function of  $y$  alone. But we need to know the range over which to integrate  $x$ . For a fixed  $y$ , what range of  $x$ 's is allowed for  $(x, y)$  to be within the region  $R$ ? Since  $R$  is described as the region in the first quadrant bounded by  $y = 1 - x^2$ , the smallest  $x$  can be is 0 and the biggest  $x$  can be is some value that places  $(x, y)$  on the curve  $y = 1 - x^2$ . For a fixed  $y$ , the value of  $x$  that places  $(x, y)$  on the curve  $y = 1 - x^2$  is  $x = \sqrt{1 - y}$ . So, for a fixed  $y$ , the range of  $x$ 's that makes  $(x, y)$  sit inside the region  $R$  is  $0 \leq x \leq \sqrt{1 - y}$ .



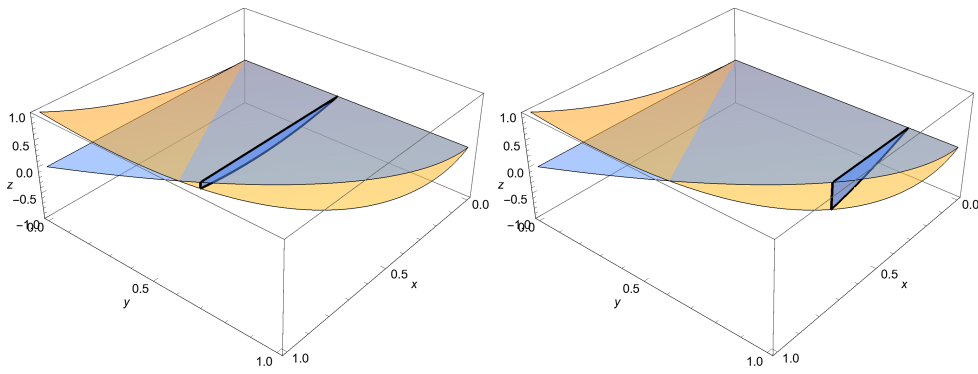
Now for each  $y$  we can do the integral over  $x$ :

$$\begin{aligned}
 \int_{x=0}^{x=\sqrt{1-y}} f(x, y) dx &= \int_{x=0}^{x=\sqrt{1-y}} (x^2 - 2xy) dx \\
 &= \left( \frac{1}{3}x^3 - x^2y \right) \Big|_{x=0}^{x=\sqrt{1-y}} \\
 &= \left( \frac{1}{3}(\sqrt{1-y})^3 - (\sqrt{1-y})^2y \right) - \left( \frac{1}{3}(0)^3 - (0)^2y \right) \\
 &= \left( \frac{1}{3}(1-y)^{3/2} - (1-y)y \right) \\
 &= \frac{1}{3}(1-y)^{3/2} - y + y^2.
 \end{aligned}$$

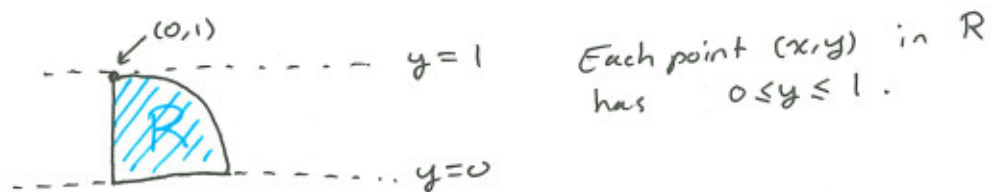
Note that the result is only a function of  $y$ . This is not the final answer, but it's a step along the way. For each fixed  $y$ , we have computed the area under the graph of the function  $f(x, y)$ , viewed as a function of only  $x$  with  $y$  being a constant. Graphically, we have computed the areas of regions like the ones depicted below







(In the pictures I am showing what we computed for four sample values of  $y$ .) If we now “add up” all those areas for all  $y$ , i.e. integrate with respect to  $y$ , then we should get the total volume under the graph of  $f$ . The range of  $y$ 's over which there is a nonempty range of  $x$ 's is  $0 \leq y \leq 1$ .



We integrate over this range of  $y$ 's to get the volume we want:

$$\begin{aligned} \int_{y=0}^{y=1} \left( \frac{1}{3}(1-y)^{3/2} - y + y^2 \right) dy &= \left( -\frac{2}{15}(1-y)^{5/2} - \frac{1}{2}y^2 + \frac{1}{3}y^3 \right) \Big|_{y=0}^{y=1} \\ &= \left( -\frac{2}{15}(1-1)^{5/2} - \frac{1}{2}(1)^2 + \frac{1}{3}(1)^3 \right) - \left( -\frac{2}{15}(1-0)^{5/2} - \frac{1}{2}(0)^2 + \frac{1}{3}(0)^3 \right) \\ &= -1/30. \end{aligned}$$

And thus we have computed the area integral

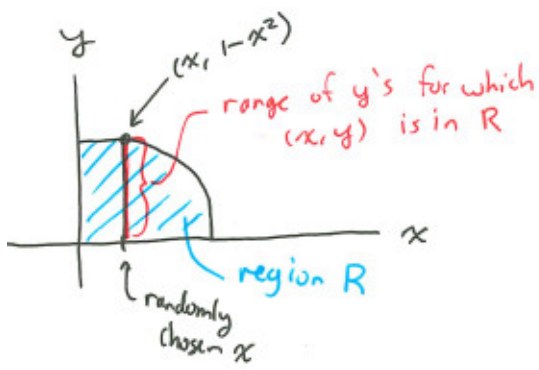
$$\int_R f dA = -1/30.$$

Summarizing our entire calculation, it went like this:

$$\begin{aligned} \int_R f dA &= \int_{y=0}^{y=1} \int_{x=0}^{x=\sqrt{1-y}} (x^2 - 2xy) dx dy \\ &= \int_{y=0}^{y=1} \left( \frac{1}{3}(1-y)^{3/2} - y + y^2 \right) dy = -1/30. \end{aligned}$$

The trickiest bit was describing the region  $R$ . We determined the ranges over which to integrate  $x$  and  $y$  by describing  $R$  as *the set of points  $(x, y)$  for which  $0 \leq y \leq 1$  and  $0 \leq x \leq \sqrt{1-y}$ .*

Another description of  $R$  is that it is *the set of points  $(x, y)$  for which  $0 \leq x \leq 1$  and  $0 \leq y \leq 1-x^2$ .*



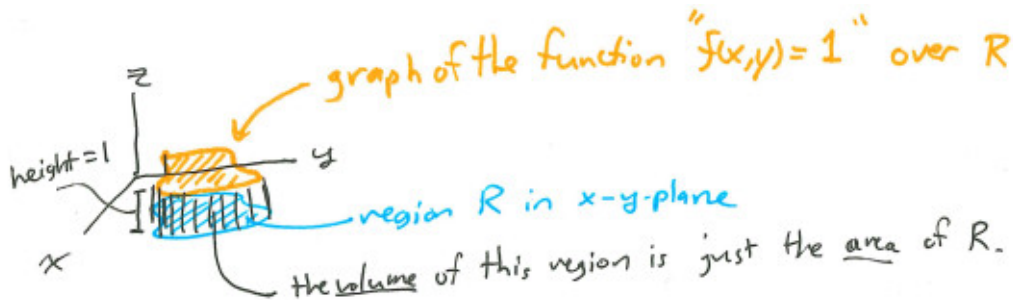
Each point  $(x, y)$  in  $R$  has  $0 \leq x \leq 1$

This leads an alternative approach to computing  $\int_R f dA$ , but doing the double integral in the opposite order:

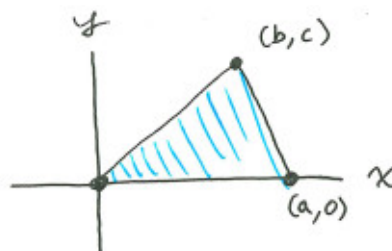
$$\begin{aligned} \int_R f dA &= \int_{x=0}^{x=1} \int_{y=0}^{y=1-x^2} (x^2 - 2xy) dy dx \\ &= \int_{x=0}^{x=1} -x^5 - x^4 + 2x^3 + x^2 - x dx = -1/30. \end{aligned}$$

Both orders of integration should give the same answer, because they both represent the same thing: the *two dimensional* integral  $\int_R f dA$ . Sometimes one order of integration will be easy to compute by hand while the other is harder or even impossible!

**Integrating the function "1".** In single variable calculus, the integral  $\int_a^b 1 dx$  is  $b - a$ , which is the length of the region  $[a, b]$  of integration. Analogously for area integrals, the integral  $\int_R 1 dA$  is the *area* of the region  $R$ .

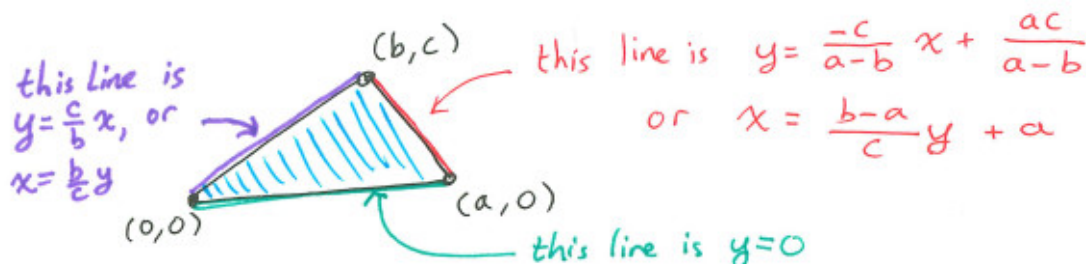


As a demonstration, let's use this to find a formula for the area of the triangle with vertices at  $(0, 0)$ ,  $(a, 0)$  and  $(b, c)$ . Assume that  $a$ ,  $b$ , and  $c$  are all positive and that  $b \leq a$ .



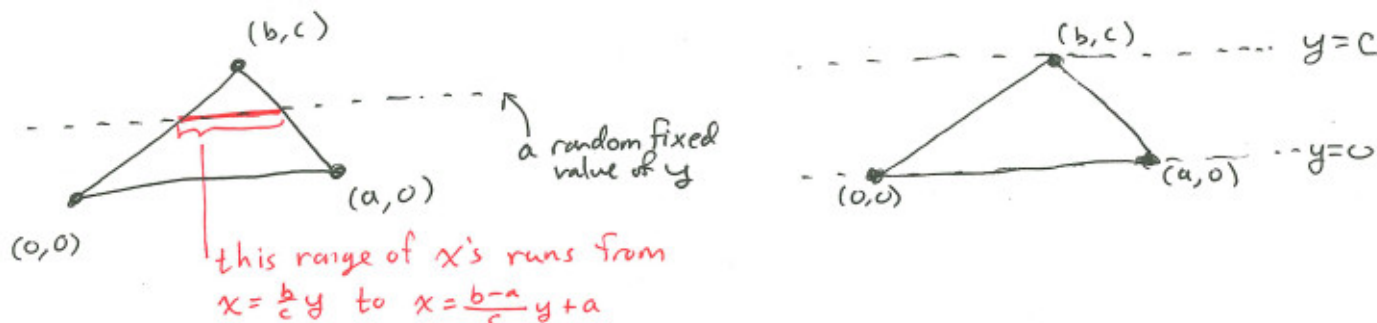
Okay, you probably already know that the area of the triangle is “one half of the base times the height,” but for fun let’s pretend you didn’t know that and figure it out by integrating the function 1 over the triangle.

First, the region  $R$  is the triangle itself. We need to decide whether we want to do an  $x$  integral first, or a  $y$  integral first. For the sake of demonstration, let’s do both orders of integration! Before we start, let’s write down equations for the boundary of this region.



You will see in a moment why these are helpful to have. Let’s begin the calculation.

*Doing an  $x$  integral, then a  $y$  integral.* For any fixed  $y$ , the range of  $x$ ’s within the triangle runs from  $x = \frac{b}{c}y$  to  $x = \frac{b-a}{c}y + a$ . The overall range of  $y$ ’s for the whole triangle has  $y$  running from  $y = 0$  to  $y = c$ . In other words, the triangle (let’s call it  $T$ ) has the following description:  $T$  is the set of points  $(x, y)$  for which  $0 \leq y \leq c$  and  $\frac{b}{c}y \leq x \leq \frac{b-a}{c}y + a$ .



We can now calculate:

$$\begin{aligned}
 \text{area of } T &= \int_T 1 \, dA \\
 &= \int_{y=0}^{y=c} \int_{x=\frac{b}{c}y}^{x=\frac{b-a}{c}y+a} 1 \, dx \, dy \\
 &= \int_{y=0}^{y=c} \left( \frac{b-a}{c}y + a - \frac{b}{c}y \right) dy \\
 &= \int_{y=0}^{y=c} \left( a - \frac{a}{c}y \right) dy \\
 &= \left( ay - \frac{a}{2c}y^2 \right) \Big|_{y=0}^{y=c} \\
 &= ac - \frac{a}{2c}c^2 \\
 &= \frac{1}{2}ac.
 \end{aligned}$$

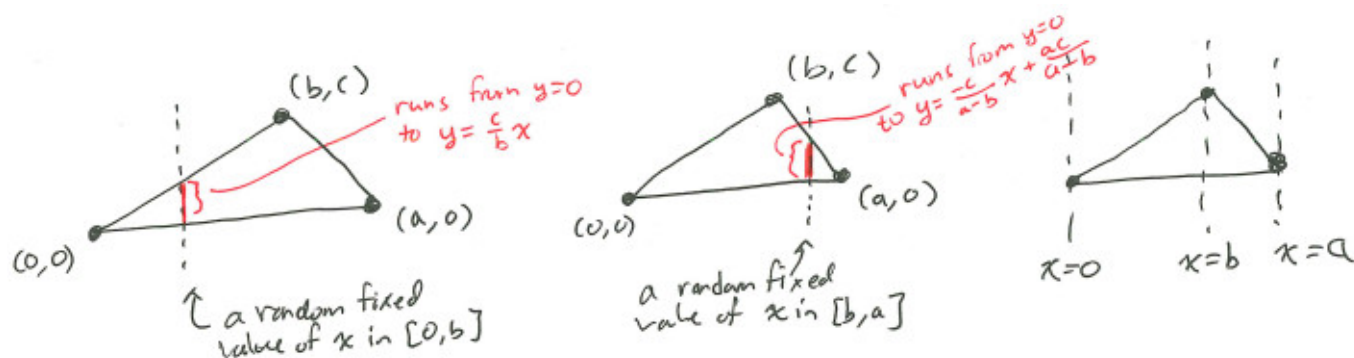
Yay! We have recovered, in an overly complicated way, the area formula “one half base times height” that we already knew. Now let’s try the other order, for even more fun.

*Doing a y integral, then an x integral.* For any fixed  $x$ , the range of  $y$ ’s within the triangle runs from  $y = 0$  to  $y = \frac{c}{b}x$ ... or it runs from  $y = 0$  to  $y = \frac{-c}{a-b}x + \frac{ac}{a-b}$ . Wait, so which is it? It *depends on x*. The function defining the boundary of  $y$ -values in terms of  $x$  is a piecewise function here. For any fixed  $x$  between 0 and  $b$ , the range of  $y$ ’s within the triangle runs from  $y = 0$  to  $y = \frac{c}{b}x$ . For any fixed  $x$  between  $b$  and  $a$ , the range of  $y$ ’s within the triangle runs from  $y = 0$  to  $y = \frac{-c}{a-b}x + \frac{ac}{a-b}$ . The overall range of  $x$ ’s for the whole triangle has  $x$  running from  $x = 0$  to  $x = a$ , but we are effectively splitting it into two ranges: one with  $x$  running from 0 to  $b$  and the other with  $x$  running from  $b$  to  $a$ . In the end our triangle  $T$  has been given the following description:  $T$  is the set of points  $(x, y)$  for which either

$$0 \leq x \leq b \text{ and } 0 \leq y \leq \frac{c}{b}x$$

or

$$b \leq x \leq a \text{ and } 0 \leq y \leq \frac{-c}{a-b}x + \frac{ac}{a-b}.$$



We can now calculate: It has to be done as two separate integrals, one for each piece in the piecewise

description of the  $y$  bounds as a function of  $x$ :

$$\begin{aligned}
 \text{area of } T &= \int_T 1 \, dA \\
 &= \int_{x=0}^{x=b} \int_{y=0}^{y=\frac{c}{b}x} 1 \, dy \, dx + \int_{x=b}^{x=a} \int_{y=0}^{y=\frac{c}{b-a}x + \frac{ac}{a-b}} 1 \, dy \, dx \\
 &= \int_{x=0}^{x=b} \frac{c}{b}x \, dx + \int_{x=b}^{x=a} \left( \frac{c}{b-a}x + \frac{ac}{a-b} \right) dx \\
 &= \frac{c}{2b}b^2 + \left( \frac{c}{2(b-a)}x^2 + \frac{ac}{a-b}x \right) \Big|_{x=b}^{x=a} \\
 &= \frac{bc}{2} + \left( \frac{c}{2(b-a)}a^2 + \frac{ac}{a-b}a \right) - \left( \frac{c}{2(b-a)}b^2 + \frac{ac}{a-b}b \right) \\
 &= \frac{1}{2}ac.
 \end{aligned}$$

Yuck! We got the same answer, but a lot more simplification was needed to get there (I skipped a lot of the algebra at the end). It turned out that the other order of integration was much nicer.

**Volume Integrals and Beyond.** Above we have talked about *area integrals*, which are integrals of functions of *two variables* over subregions of their domain. The idea extends quite readily to functions of three or more variables. In the case of a function of three variables, we get *volume integrals*, which can be computed as *triple integrals*.

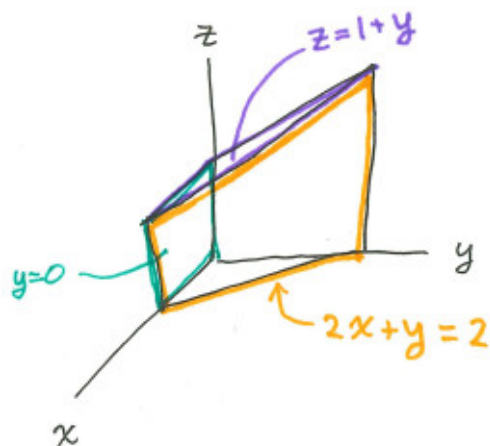
The integral of a function  $f(x, y, z)$  on  $\mathbb{R}^3$  over a region  $R$  inside  $\mathbb{R}^3$  is denoted by  $\int_R f \, dV$ , with the “ $dV$ ” telling you that it’s a volume integral. But what does it mean? The integral of a function  $f(x)$  is the “area under the curve,” the integral of a function  $f(x, y)$  is the “volume under the surface,” and the integral of  $f(x, y, z)$  is...?

Well, you could try to describe it as some kind of a four-dimensional volume-thingy under the three dimensional graph of  $f(x, y, z)$  in  $\mathbb{R}^4$ . But don’t. Ultimately, integrals are not really about getting “area under the curve” or “volume under the surface” at all. They are about turning *densities* into *amounts*.

By “density” I don’t necessarily mean “mass density.” Imagine that  $f(x, y, z)$  is the density of *something*, it doesn’t matter what. Let’s say that  $f(x, y, z)$  gives the density of *things* per cubic meter at  $(x, y, z)$ . Then the volume integral  $\int_R f(x, y, z) \, dV$  is the *total amount of things in the region  $R$* . Similarly, if  $f(x, y)$  is the density of things per square meter and  $R$  is a region in  $\mathbb{R}^2$ , then the area integral  $\int_R f(x, y) \, dA$  total amount of things in the region  $R$ . You can replace “thing” by whatever you want here: mass, electric charge, probability, chemical concentration, etc.

Let’s now compute a volume integral as an example. There’s nothing new here— just a more complicated version of what we did before.

**Example** Suppose that a piece of material is shaped like the region  $R$  shown below, and that its mass density is given by the function  $f(x, y, z) = x + z$ . Our task to compute the mass.



The mass is  $\int_R (x + z) dV$ . We have to write this out as a triple integral, and the hard part is to describe the region of integration in a way that makes it possible to write down the integration limits. I will integrate with respect to  $z$  first, getting a function of  $x$  and  $y$ , then I will integrate with respect to  $y$ , getting a function of  $x$ , and finally I will integrate with respect to  $x$ , getting a numerical value for the mass. For any suitable choice of  $x$  and  $y$ , the  $z$  coordinate runs from 0 to  $1 + y$  in order for  $(x, y, z)$  to be inside  $R$ . Now we need to determine what are the suitable choices of  $x$  and  $y$  (those that have  $(x, y, z)$  in  $R$  for some  $z$ ). For any suitable choice of  $x$ , the  $y$  coordinate runs from 0 to  $2 - 2x$  (get this by solving  $2x + y = 2$  for  $y$ ) in order for  $(x, y, z)$  to be inside  $R$  for some  $z$ . Now we need to determine what are the suitable choices of  $x$  (those that have  $(x, y, z)$  in  $R$  for some  $y$  and  $z$ ). The  $x$  coordinate runs from 0 to 1 (get this by looking at the picture; what is the maximum value of  $x$  in the region?) in order for  $(x, y, z)$  to be inside  $R$  for some  $y$  and  $z$ . After all this rambling, we have the following excellent and useful description of the region  $R$ : *it is the set of points  $(x, y, z)$  for which  $0 \leq x \leq 1$ ,  $0 \leq y \leq 2 - 2x$ , and  $0 \leq z \leq 1 + y$ .* Now we can turn the crank and get an answer:

$$\begin{aligned}
 \int_R (x + z) dV &= \int_0^1 \int_0^{2-2x} \int_0^{1+y} (x + z) dz dy dx \\
 &= \int_0^1 \int_0^{2-2x} \left. \left( xz + \frac{1}{2}z^2 \right) \right|_{z=0}^{z=1+y} dy dx \\
 &= \int_0^1 \int_0^{2-2x} \left( x(1+y) + \frac{1}{2}(1+y)^2 \right) dy dx \\
 &= \int_0^1 \int_0^{2-2x} \left( xy + x + \frac{1}{2}y^2 + y + \frac{1}{2} \right) dy dx \\
 &= \int_0^1 \left( \frac{1}{2}xy^2 + xy + \frac{1}{6}y^3 + \frac{1}{2}y^2 + \frac{1}{2}y \right) \Big|_{y=0}^{y=2-2x} dx \\
 &= \int_0^1 \left( \frac{1}{2}x(2-2x)^2 + x(2-2x) + \frac{1}{6}(2-2x)^3 + \frac{1}{2}(2-2x)^2 + \frac{1}{2}(2-2x) \right) dx \\
 &= \int_0^1 \left( \frac{2}{3}x^3 - 5x + \frac{13}{3} \right) dx \\
 &= \left( \frac{1}{6}x^4 - \frac{5}{2}x^2 + \frac{13}{3}x \right) \Big|_{x=0}^{x=1} \\
 &= \frac{1}{6} - \frac{5}{2} + \frac{13}{3} \\
 &= 2.
 \end{aligned}$$

The mass is 2.

**Non-Cartesian Coordinates.** You saw earlier that choosing a good order in which to do multiple integrals can make the difference between an easy calculation and a hard one. Another thing you can choose cleverly to simplify an integral calculation (or to make it at all possible!) is *coordinates*. This is analogous to “*u*-substitution” from single variable calculus.

There two reasons you might choose a certain coordinate system for a particular integration problem<sup>8</sup>  $\int_R f dV$ :

1. The function being integrated might have a simpler description in a certain coordinate system.
2. The region of integration might have a cleaner description in a certain a coordinate system.

However there’s one extra thing you have to worry about if you use non-cartesian coordinates: the distortion of areas/volumes due to the coordinate system. We discussed this in section 3.4.4, which gives us a way to quantify the distortion. (Reread that section if it’s faded from memory.) The Jacobian determinant turns out to be the *correction factor* that needs to be included in your integrals to account for the artificial overestimates and underestimates of areas/volumes caused by your choice of coordinates.

Specifically, the correction factor to include in your integrand if you use coordinates  $u_1, \dots, u_n$  is the absolute value of the Jacobian determinant of the transformation from  $u_1, \dots, u_n$  to cartesian coordinates. Let me use  $J$  for the Jacobian determinant:

$$J = \det \left( \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_n} \end{bmatrix} \right).$$

The correction factor you want is  $|J|$ . If you want to write out an integral

$$\int_R f dV$$

in terms of the coordinates  $u_1, \dots, u_n$ , then you have to expand three things in terms of the coordinates: the region  $R$ , the function  $f$ , and the symbol “ $dV$ .” The last item is where  $|J|$  comes in– the symbol  $dV$  expands as  $dV = |J|du_1du_2 \cdots du_n$ . The symbol  $dV$  is sometimes thought of as an infinitesimal element of volume. For example when you write a three dimensional integral in cartesian coordinates, it ends up expanding as  $dV = dx dy dz$ , in some order depending on the order in which you do your integration. This is thought of as the volume of a little “coordinate box” with side lengths  $dx$ ,  $dy$ , and  $dz$ . For any other coordinate system  $u, v, w$ , just include the factor  $|J|$  to get something that correctly represents the volume of the little coordinate box that is swept out by  $du$ ,  $dv$ , and  $dw$ .

---

<sup>8</sup>I will use the generic notation  $\int_R f dV$  to stand for area integrals (which I would normally write as  $\int_R f dA$  to be specific) or volume integrals. Actually it can really stand for integrals in any dimension. The word “volume” is usually used as a generic term.

It will be good to have the Jacobian determinants handy for the famous coordinate systems. For polar coordinates, we computed in section 3.4.4 that  $J = r$ . Since  $r$  is always positive,  $|J| = r$ . So if you want to write out an integral in polar coordinates, expand  $\int_R f dA$  as

$$\int \int f(r, \theta) r dr d\theta.$$

(You would have to figure out the limits of integration in terms of polar coordinates as well.) Think of it like this:  $dA = r dr d\theta$ .

For spherical coordinates (you computed this in HW5 problem 39) we have  $|J| = \rho^2 \sin(\phi)$ . So we just write out the symbol  $dV$  as  $\rho^2 \sin(\phi) d\rho d\theta d\phi$  (possibly with a different order depending on the order if integration you are going for).

For cylindrical coordinates we have  $|J| = r$ . So we just write out the symbol  $dV$  as  $r dr d\theta dz$  (possibly with a different order depending on the order if integration you are going for).

**Example.** Consider the function  $f(x, y) = x^2 + y^2$ . Let's integrate this function over the unit disk centered at the origin. In cartesian coordinates, things do not look very pretty:

$$\int_{\text{unit disk}} (x^2 + y^2) dA = \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} (x^2 + y^2) dx dy.$$

Let's do it in polar coordinates instead. We have three things to write out in polar coordinates: the region of integration, the function, and the symbol  $dA$ . The region of integration, the unit disk, is simply the region  $r \leq 1$  in terms of polar coordinates. The coordinate  $\theta$  can take its full range of 0 to  $2\pi$ . Precisely, we have the following description of the unit disk: *it is the set of points  $(r, \theta)$  for which  $0 \leq r \leq 1$  and  $0 \leq \theta \leq 2\pi$* . Now for the function. That's easy:  $x^2 + y^2 = r^2$ . Finally,  $dA$ . For polar coordinates  $dA = r dr d\theta$ . Now we compute

$$\begin{aligned} \int_{\text{unit disk}} (x^2 + y^2) dA &= \int_0^1 \int_0^{2\pi} r^2 (r d\theta dr) \\ &= \int_0^1 \int_0^{2\pi} r^3 d\theta dr \\ &= \int_0^1 r^3 \theta \Big|_{\theta=0}^{\theta=2\pi} dr \\ &= 2\pi \int_0^1 r^3 dr \\ &= \frac{\pi}{2} r^4 \Big|_{r=0}^{r=1} = \frac{\pi}{2}. \end{aligned}$$

**Example: Computing the area of a circle.** Remember that when you integrate the function "1" you are computing the area/volume of the region of integration. Let's do this to compute the

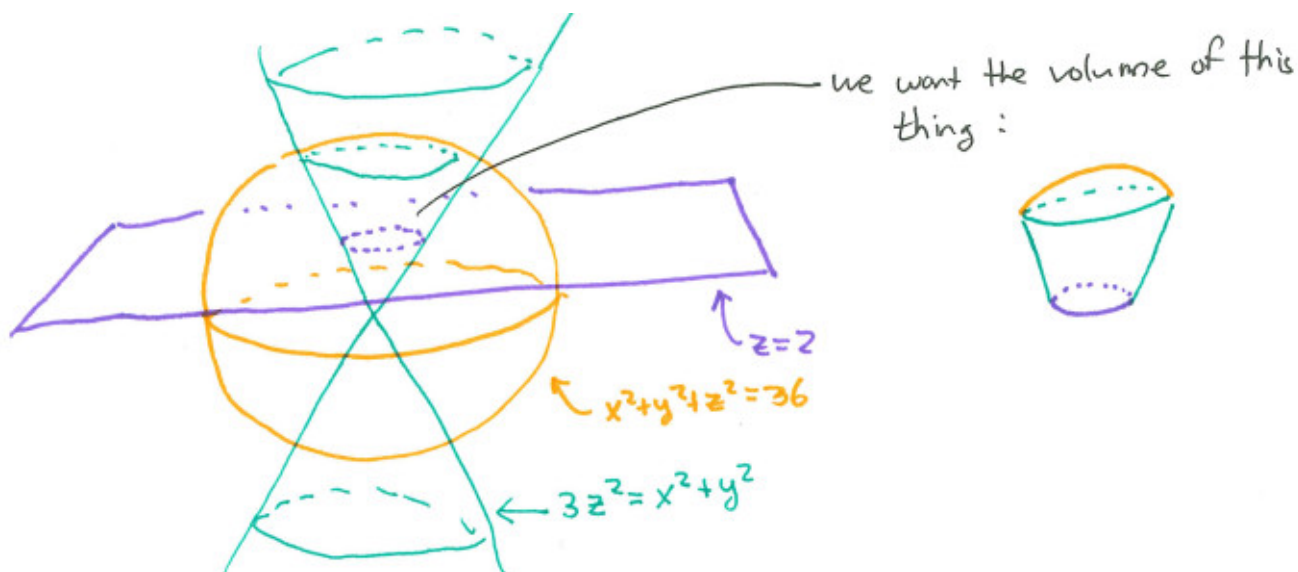


area of a disk of radius  $r_0$

$$\begin{aligned} \int_{\text{disk of radius } r_0} 1 \, dA &= \int_0^{r_0} \int_0^{2\pi} 1 \, (r \, d\theta \, dr) \\ &= 2\pi \int_0^{r_0} r \, dr \\ &= \pi r^2 \Big|_{r=0}^{r=r_0} = \pi r_0^2. \end{aligned}$$

We have derived the famous formula for the area of a circle!

**Example.** Let us find the volume inside the cone  $3z^2 = x^2 + y^2$ , above the plane  $z = 2$ , and inside the sphere  $x^2 + y^2 + z^2 = 36$ .



The hard part is describing the region in such a way that we get the limits on the integrals. I drew a picture here but let me try to describe the region without the picture, just to show you how an analysis might look in a situation where you don't already know what everything looks like<sup>9</sup> There's just one portion where you need the picture: to interpret the phrase "inside the cone  $3z^2 = x^2 + y^2$ " you should have a good picture. The picture shows that the "inside" of the cone is given by  $3z^2 \geq x^2 + y^2$ .

Let's call the region  $R$ . Our current description of  $R$  is that it is the set of points  $(x, y, z)$  such that

- $3z^2 \geq x^2 + y^2$  and
- $z \geq 2$  and
- $x^2 + y^2 + z^2 \leq 36$ .

<sup>9</sup>For example, in higher dimensions than 3 you would not get to rely so much on a picture. Still, I do recommend making the most of whatever pictures you can produce when describing regions!

This is not even close to good enough for us to write out a triple integral in cartesian coordinates. And cartesian coordinates are not a good choice for this problem anyway! Let's use spherical coordinates. Converting the three inequalities, we obtain the following description of  $R$ : it is the set of points  $(\rho, \theta, \phi)$  such that

- $3\rho^2 \cos^2(\phi) \geq \rho^2 \sin^2(\phi)$  and
- $\rho \cos(\phi) \geq 2$  and
- $\rho^2 \leq 36$ .

The three conditions simplify a bit to

- $3 \cos^2(\phi) \geq \sin^2(\phi)$  and
- $\rho \cos(\phi) \geq 2$  and
- $\rho \leq 6$ .

Simplifying the first condition further requires taking a square root... which branch of the square root should we take? Well  $z \geq 2$  for all points in  $R$ , so since  $z$  is always positive  $\phi$  is in the range 0 to  $\pi/2$ . This ensures that  $\cos(\phi)$  and  $\sin(\phi)$  are positive. Now we can simplify the first condition to  $\tan(\phi) \leq \sqrt{3}$ , or just  $\phi \leq \arctan(\sqrt{3})$ . Now  $R$  is described as the set of points  $(\rho, \theta, \phi)$  such that

- $\phi \leq \arctan(\sqrt{3})$  and
- $\rho \cos(\phi) \geq 2$  and
- $\rho \leq 6$ .

That nice and simple first condition is suggesting that we use  $\phi$  as the "outermost" integral. For any  $\phi$  between 0 and  $\arctan(\sqrt{3})$ , what must be true of  $\rho$  and  $\theta$  in order for  $(\rho, \theta, \phi)$  to be in  $R$ ? Well  $\theta$  is not even mentioned in the condition so it can be allowed its full range from 0 to  $2\pi$ . As for  $\rho$ , the third condition tells us it must be in the range 0 to 6. And it remains to satisfy the second condition, which imposes a further restriction on  $\rho$ :  $\rho \geq \frac{2}{\cos(\phi)} = 2 \sec(\phi)$ . Now we have the following excellent description of  $R$ : it is the set of points  $(\rho, \theta, \phi)$  such that

- $0 \leq \phi \leq \arctan(\sqrt{3})$  and
- $2 \sec(\phi) \leq \rho \leq 6$  and
- $0 \leq \theta \leq 2\pi$ .

Since I am planning to do the  $\rho$  integral *inside* the  $\phi$  integral, the range for  $\rho$  is allowed to depend on  $\phi$ . If I were planning to do the  $\theta$  integral inside both of the other integrals, the range for  $\theta$  would be allowed to depend on both  $\phi$  and  $\rho$ . Luckily here, things are not so complicated and the  $\theta$  bounds turned to be constant. In fact, only  $\rho$  has a nonconstant bound. This is why spherical

coordinates are a superior approach in this problem! Now we can go ahead and compute the volume of  $R$ , remembering that  $dV = \rho^2 \sin(\phi) d\rho d\theta d\phi$ :

$$\begin{aligned}
 \int_0^{\arctan(\sqrt{3})} \int_{2\sec(\phi)}^6 \int_0^{2\pi} \rho^2 \sin(\phi) d\theta d\rho d\phi &= 2\pi \int_0^{\arctan(\sqrt{3})} \int_{2\sec(\phi)}^6 \rho^2 \sin(\phi) d\rho d\phi \\
 &= \frac{2\pi}{3} \int_0^{\arctan(\sqrt{3})} \rho^3 \sin(\phi) \Big|_{\rho=2\sec(\phi)}^{\rho=6} d\phi \\
 &= \frac{2\pi}{3} \int_0^{\arctan(\sqrt{3})} (6^3 - 8 \sec^3(\phi)) \sin(\phi) d\phi \\
 &= \frac{2\pi}{3} 6^3 \int_0^{\arctan(\sqrt{3})} \sin(\phi) d\phi - \frac{16\pi}{3} \int_0^{\arctan(\sqrt{3})} \frac{\sin(\phi)}{\cos^3(\phi)} d\phi \\
 &= \frac{2\pi}{3} 6^3 (-\cos(\arctan(\sqrt{3})) + 1) - \frac{8\pi}{3} \left( \frac{1}{\cos^2(\arctan(\sqrt{3}))} - \frac{1}{\cos^2(0)} \right) \\
 &= \frac{2\pi}{3} 6^3 (1/2) - \frac{8\pi}{3} (3) \\
 &= 64\pi.
 \end{aligned}$$

## Summary.

### Summary on integration of functions over regions

An integral of a function  $f$  over a region  $R$  is denoted by  $\int_R f dV$  (or sometimes  $\int_R f dA$  if  $f$  is a function of only two variables). When you want to compute such an integral, you write it as multiple single-variable integrals. You get to choose

- the coordinate system in which to express everything
- the order in which to do the single integrals

and you should choose them strategically. Once you make those choices, expand each of the three ingredients in the notation  $\int_R f dV$ :

- Express the region  $R$  in terms of your chosen coordinates in such a way that you can determine limits for the single integrals.
- Express the function  $f$  in terms of your chosen coordinates.
- Write  $dV$  as  $|J| du_1 \cdots du_n$ , where  $u_1, \dots, u_n$  are your chosen coordinates and  $J$  is the Jacobian determinant of the transformation from your coordinate system to cartesian coordinates.

## 4.2 Line Integrals and Conservative Vector Fields

We now introduce a new kind of integral, the integral of a *vector field* along a *path*. It is called a *line integral*. Given a vector field  $\vec{F}(\vec{x})$  and a path  $\vec{x}(t)$ , we denote<sup>10</sup> the *line integral of  $\vec{F}(\vec{x})$  along*

<sup>10</sup>There are two pieces of input data to a line integral: the vector field and the path. The vector field sits in the integrand position, and the path is indicated by the  $d\vec{x}$ , or additionally it can be indicated as a subscript to the integral symbol. Here I'm showing both places where the path could be indicated in the notation.

$\vec{x}(t)$  by

$$\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x}$$

and we define it to be the ordinary integral

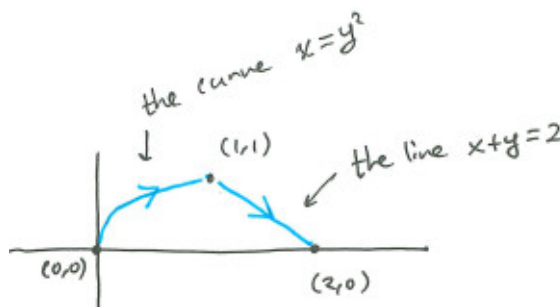
$$\int_a^b \vec{F}(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt} dt,$$

where  $(a, b)$  is the interval on which the path is defined. Recall from section A.2.3 that a dot product measures the extent to which two vectors point in the same direction (though it also scales with the lengths of both vectors involved). A line integral adds up, along a path, all the dot products of the path velocities with the vector field. Thus a line integral measures the extent to which a path travels *along* a vector field (though it also scales with the norm of the vector field— but as we will soon see it does not scale up with the speed of the path).

The main application of line integrals is to compute *work* in physics. If  $\vec{F}(\vec{x})$  is a force field and  $\vec{x}(t)$  is the trajectory of a particle that is subjected to the force, then the work done by the force field on the particle is given by the line integral  $\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x}$ . The notation suggests what is going on: Integrating the scalar function  $\vec{F}(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt}$  over  $t$  is like adding up  $\vec{F}(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt} \Delta t$  over lots of finely subdivided time intervals  $\Delta t$ , which is like adding up  $\vec{F}(\vec{x}(t)) \cdot \Delta\vec{x}$  over lots of finely subdivided segments  $\Delta\vec{x}$  of the path. Over each little segment  $\Delta\vec{x}$  of the path, the work done by the force field equals the force vector  $\vec{F}(\vec{x})$  at the location  $\vec{x}$  of the path dotted with the displacement vector  $\Delta\vec{x}$  of the path. A line integral is a way of taking a parameterized path and adding up all those little bits of work to get the total work.

Even if the physics application is not meaningful for you, it turns out that line integrals are interesting from a purely mathematical perspective as well. They will be crucial to answering the following (mathematical) question: Given a vector field  $\vec{F}(\vec{x})$ , is it possible to find a scalar function  $f(\vec{x})$  such that  $\vec{F} = \vec{\nabla} f$ ?

**Example Computation of a Line Integral.** Let  $\vec{F}$  be the vector field  $(6x + 2y)\partial_x + (2x)\partial_y$ . Let  $\vec{x}(t)$  be the trajectory<sup>11</sup> depicted below from the origin  $(0, 0)$  to the point  $(2, 0)$ .



<sup>11</sup>Technically this is not a smooth path and I did say everything we consider in this course will be smooth. Well, it is a path made out of two smooth pieces connected together, and that will be good enough for us when we consider line integrals. “Piecewise-smooth” paths are okay in this section. Also, in this section we will allow paths to have velocity zero sometimes

We would like to compute the line integral of  $\vec{F}$  along the trajectory. If  $\vec{x}(t)$  is a parameterization of that trajectory for  $a < t < b$ , then what we want is simply  $\int_a^b \vec{F}(\vec{x}(t)) \cdot \vec{x}'(t) dt$ , so we just need to parameterize the trajectory. It has to be done in two parts, and we will have to compute the integral of each part separately, adding the results. The first part, running from  $(0, 0)$  to  $(1, 1)$ , can be parameterized as follows:

$$\vec{x}_1(t) = \begin{bmatrix} x_1(t) \\ y_1(t) \end{bmatrix} = \begin{bmatrix} t^2 \\ t \end{bmatrix}, \text{ defined for } 0 \leq t \leq 1.$$

The second part, running from  $(1, 1)$  to  $(2, 0)$ , can be parameterized as follows:

$$\vec{x}_2(t) = \begin{bmatrix} x_2(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} t \\ 2 - t \end{bmatrix}, \text{ defined for } 1 \leq t \leq 2.$$

These are not the only parameterizations— we made lots of choices here by choosing these parameterizations. This may worry you in the upcoming calculation— you might think that the answer depends on our choices— but you will see later in this section that the answer is mostly independent of our choices of parameterization! Now the first integral is of the function

$$\vec{F}(\vec{x}_1(t)) \cdot \vec{x}_1'(t) = \begin{bmatrix} 6x_1(t) + 2y_1(t) \\ 2x_1(t) \end{bmatrix} \cdot \begin{bmatrix} 2t \\ 1 \end{bmatrix} = \begin{bmatrix} 6t^2 + 2t \\ 2t^2 \end{bmatrix} \cdot \begin{bmatrix} 2t \\ 1 \end{bmatrix} = 12t^3 + 6t^2$$

from  $t = 0$  to  $t = 1$ . This evaluates to  $(3t^4 + 2t^3)|_{t=0}^{t=1} = 5$ . The second integral is of the function

$$\vec{F}(\vec{x}_2(t)) \cdot \vec{x}_2'(t) = \begin{bmatrix} 6x_2(t) + 2y_2(t) \\ 2x_2(t) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 6t + 2(2 - t) \\ 2t \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2t + 4$$

from  $t = 1$  to  $t = 2$ . This evaluates to  $(t^2 + 4t)|_{t=1}^{t=2} = 7$  So we've got the line integral as the sum of the line integrals over the two pieces:

$$\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x} = \int_{\vec{x}_1(t)} \vec{F} \cdot d\vec{x} + \int_{\vec{x}_2(t)} \vec{F} \cdot d\vec{x} = 5 + 7 = 12.$$

**An Alternative Notation** Sometimes a line integral will be presented in the format

$$\int_{\vec{x}(t)} (f(x, y) dx + g(x, y) dy),$$

where  $\vec{x}(t)$  is some given path and  $f$  and  $g$  are some given scalar functions. How is this a line integral? What is the vector field? The idea of this notation is that the  $d\vec{x}$  from our original notation  $\int \vec{F} \cdot d\vec{x}$  is some kind of infinitesimal vector made out of the components  $dx$  and  $dy$ , and that the “dot product”  $\vec{F} \cdot d\vec{x}$  has been written out as  $f(x, y) dx + g(x, y) dy$ . Thus the scalar functions  $f$  and  $g$  appearing in this notation are giving you the components of the vector field<sup>12</sup>.

<sup>12</sup>Actually it is due to the simplicity of cartesian coordinates that the dot product works out this way. In this course I will only use this notation with cartesian coordinates.

**Example Where the Other Notation is Used.** Suppose you are asked to compute the line integral  $\int (y dx + (y - x) dy)$  along the straight line running from  $(0, 2)$  to  $(0, -2)$ . What's the vector field here? The component functions of the vector field can be read off as the coefficients of  $dx$  and  $dy$  in the integrand; the vector field is  $y \partial_x + (y - x) \partial_y$ . Now we can just parameterize the path and proceed normally. One good parameterization would be  $\vec{x}(t) = (0, -2t)$  defined for  $-1 < t < 1$ . The line integral is now

$$\int_{\vec{x}(t)} \begin{bmatrix} y \\ y - x \end{bmatrix} \cdot d\vec{x} = \int_{-1}^1 \begin{bmatrix} -2t \\ -2t - 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -2 \end{bmatrix} dt = \int_{-1}^1 4t dt = 2t^2 \Big|_{t=-1}^{t=1} = 0.$$

Another way to view this calculation is that, after choosing the path  $\vec{x}(t) = (x(t), y(t)) = (0, -2t)$  for  $-1 < t < 1$ , we can replace  $dx$  by  $\frac{dx}{dt} dt$  and  $dy$  by  $\frac{dy}{dt} dt$  to get a single integral:

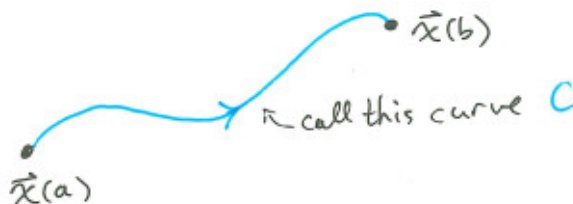
$$\int (y dx + (y - x) dy) = \int_{-1}^1 \left( y(t) \frac{dx}{dt} + (y(t) - x(t)) \frac{dy}{dt} \right) dt = \int_{-1}^1 -2(y(t) - x(t)) dt = \int_{-1}^1 4t dt = 0.$$

Yet another way would be to use  $y$  itself as the parameter of the path, to note that  $x = 0$  and  $dx = 0$  along the path (because the path always has  $dx = 0$  and it is vertical, so any displacement along the path has zero  $x$ -component), and to let the parameter  $y$  run its course from 2 to  $-2$ :

$$\int (y dx + (y - x) dy) = \int_2^{-2} y dy = \frac{1}{2} y^2 \Big|_{y=2}^{y=-2} = 0.$$

The different viewpoints and notations take some getting used to, but just remember that every time you see a line integral, there are somehow a vector field and a path involved and the integral is of the scalar function  $\vec{F}(\vec{x}(t)) \cdot \vec{x}'(t)$  for some parameter  $t$ .

**Parameterization Doesn't Matter, Mostly.** Suppose you have two parameterized paths  $\vec{x}(t)$  and  $\vec{x}(t(s))$  that represent the same curve  $C$  in  $\mathbb{R}^n$ . Here  $\vec{x}(t)$  is one parameterization and  $\vec{x}(t(s))$  is a reparameterization, with  $t(s)$  being the function that expresses the parameter  $s$  in terms of the parameter  $t$ . (Revisit section 1.2 or section 1.5.1 if you're rusty on this.) Suppose the curve has endpoints such that the path  $\vec{x}(t)$  should only be defined from  $t = a$  to  $t = b$ .



Then in terms of the  $s$  parameter, we should allow  $\vec{x}(t(s))$  to run from  $s = c$  to  $s = d$ , where  $c$  and  $d$  are such that  $t(s = a) = c$  and  $t(s = b) = d$  (this can be arranged because a reparameterization should be an invertible function). Given a vector field  $\vec{F}(\vec{x})$ , let's describe the line integral  $\int_C \vec{F} \cdot d\vec{x}$

in terms of the parameterization  $\vec{x}(t(s))$  of  $C$ :

$$\begin{aligned} \int_{\vec{x}(t(s))} \vec{F} \cdot d\vec{x} &= \int_c^d \vec{F}(\vec{x}(t(s))) \cdot \left( \frac{d}{ds} \vec{x}(t(s)) \right) ds \\ &= \int_c^d \vec{F}(\vec{x}(t(s))) \cdot \frac{d\vec{x}}{dt} \frac{dt}{ds} ds \\ &= \int_a^b \vec{F}(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt} dt. \end{aligned}$$

The last line was obtained by “ $u$ -substitution;” the last two lines are equal by the change-of-variables rule of single variable calculus. The last line is just  $\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x}$ , so we have shown that line integrals only care about the curve  $C$  and not the particular parameterization you use.

Well, almost. There’s one aspect of a parameterization that line integrals care about– the *orientation* of a curve. This is the information depicted by the little arrow we draw on the curve. Basically, it amounts to a specification of which is the starting endpoint and which is the ending endpoint of the curve. Getting it backwards gets you a minus sign, since it would flip around all the velocities, and since negating one vector in a dot product negates the dot product.

(You might be wondering: Where does orientation show up in the argument we made above? We assumed in the argument that  $a < b$ . If the reparameterization were such that  $d < c$ , then it would be flipping orientation. In the case  $d < c$  we would actually write  $\int_{\vec{x}(t(s))} \vec{F} \cdot d\vec{x}$  as  $\int_d^c \vec{F}(\vec{x}(t(s))) \cdot \left( \frac{d}{ds} \vec{x}(t(s)) \right) ds$ , which is off by a minus sign from what we wanted.)

The key point here is this: Line integrals actually involve only a vector field and an *oriented* curve. A particular parameterization for the curve can be chosen to *compute* a line integral, but it doesn’t matter which one you choose as long as it travels along the curve from the starting point to the ending point.

### Definition of Line Integral

The line integral of a vector field  $\vec{F}(\vec{x})$  along an oriented curve  $C$  is denote by

$$\int_C \vec{F} \cdot d\vec{x}$$

and it can be computed as

$$\int_a^b \vec{F}(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt} dt$$

for any parameterization  $\vec{x}(t)$  of  $C$  that that traverses  $C$  from start to end when  $t$  goes from  $a$  to  $b$ .

**The Fundamental Theorem of Calculus** Recall the fundamental theorem from single variable calculus: it says that, for a function  $f(x)$ , the integral of the derivative  $\frac{df}{dx}$  from  $a$  to  $b$  is  $f(b) - f(a)$ . For a function  $f(x_1, \dots, x_n)$ , the derivative of  $f$  is its gradient  $\vec{\nabla}f$ . In vector calculus, the

fundamental theorem of calculus also tells you about the integral of a derivative, but the correct kind of integral to use is a *line integral*. It goes like this:

### Fundamental Theorem of Calculus

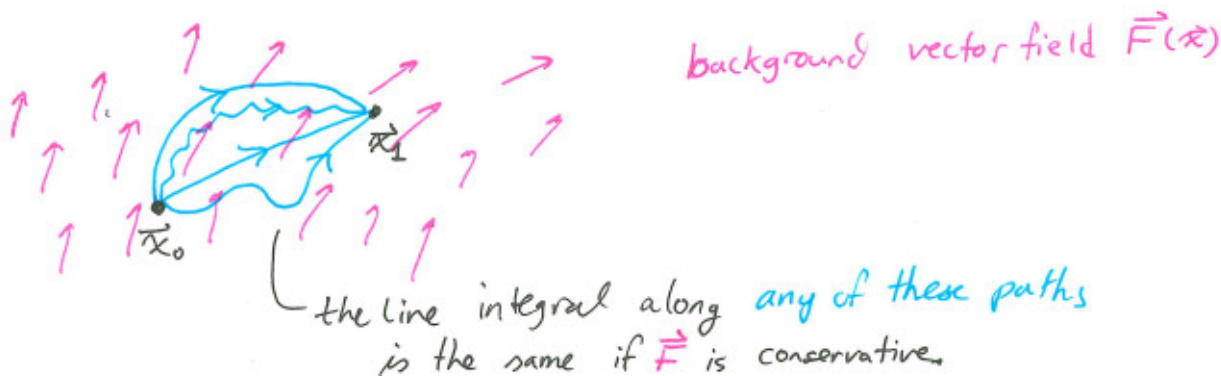
Given a scalar function  $f(x_1, \dots, x_n)$  on  $\mathbb{R}^n$  and a curve  $C$  starting at  $\vec{x}_0$  and ending at  $\vec{x}_1$ , the line integral of  $\vec{\nabla}f$  along  $C$  is the net change in  $f$  from  $\vec{x}_0$  to  $\vec{x}_1$ :

$$\int_C \vec{\nabla}f \cdot d\vec{x} = f(\vec{x}_1) - f(\vec{x}_0).$$

This follows from the fact that for any little segment  $\Delta\vec{x}$  of the curve,  $\vec{\nabla}f \cdot \Delta\vec{x}$  is the change  $\Delta f$  in  $f$  across the segment, and so the integral adds up all the little changes  $\Delta f$  to result in a net change  $f(\vec{x}_1) - f(\vec{x}_0)$ . You can also get this result right out of the chain rule and the single-variable fundamental theorem of calculus. If  $\vec{x}(t)$  is a parameterization of  $C$  with  $a \leq t \leq b$ , then

$$\int_{\vec{x}(t)} \vec{\nabla}f \cdot d\vec{x} = \int_a^b \vec{\nabla}f(\vec{x}(t)) \cdot \frac{d\vec{x}}{dt} dt = \int_a^b \frac{d}{dt} f(\vec{x}(t)) dt = f(\vec{x}(b)) - f(\vec{x}(a)).$$

**Conservative Vector Fields** A vector field is called *conservative* when *all line integrals of the vector field depend on only on the endpoints of curves*. That is, the vector field  $\vec{F}(\vec{x})$  is conservative if, given any two points  $\vec{x}_0$  and  $\vec{x}_1$ , any line integral  $\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x}$  which has  $\vec{x}(t)$  running from  $\vec{x}_0$  to  $\vec{x}_1$  has the same value. (The terminology “conservative” comes from physics, where it refers to force fields for which energy is conserved.)



There is an alternative equivalent way to describe the condition of conservativeness: Any line integral along a *loop* (a path that starts and ends in the same place) is zero.

The fundamental theorem of calculus offers us the best examples of conservative vector fields: *any gradient vector field is conservative*. Since  $\int_C \vec{\nabla}f \cdot d\vec{x}$  is just the net change of  $f$  along the endpoints of  $C$ , it is independent of which curve  $C$  is chosen to connect the endpoints. Hence  $\vec{\nabla}f$  is always conservative.

When a vector field is conservative people will sometimes just write the endpoints of a curve for a line integral, and leave it to you to choose a path that joins those endpoints. So for example you



might see the notation

$$\int_{\vec{x}_0}^{\vec{x}_1} \vec{F} \cdot d\vec{x}$$

for a line integral of  $\vec{F}$  along any path joining  $\vec{x}_0$  to  $\vec{x}_1$ . This notation only makes sense if  $\vec{F}$  is conservative; otherwise it should not be used.

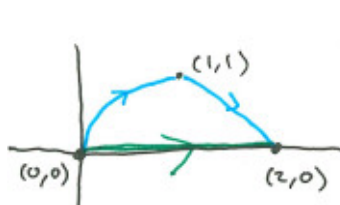
### Conservative Vector Fields

A vector field  $\vec{F}(\vec{x})$  is called *conservative* when all line integrals of the vector field depend on only on the endpoints of curves. An equivalent condition is for all line integrals along loops to vanish. By the fundamental theorem of calculus, the gradient of a scalar function is always a conservative vector field.

**Example of Conservativeness.** Let's revisit the vector field  $(6x + 2y)\partial_x + (2x)\partial_y$  from the first example in this section. This time let  $\vec{x}(t)$  be a straight-line trajectory from  $(0, 0)$  to  $(2, 0)$ . The path can be parameterized as  $(t, 0)$  for  $0 \leq t \leq 2$ . The line integral is then

$$\int_{\vec{x}(t)} \vec{F} \cdot d\vec{x} = \int_0^2 \begin{bmatrix} 6t + 2(0) \\ 2t \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt = \int_0^2 6t dt = 3t^2 \Big|_{t=0}^{t=2} = 12.$$

We got 12 for the line integral, which is the same answer we got in the first example of this section when we were considering the same vector field but a *different path* from  $(0, 0)$  to  $(2, 0)$ .

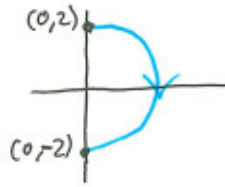


The line integral of  $\begin{bmatrix} 6x+2y \\ 2x \end{bmatrix}$  along the blue path and the green path is the same (both are 12).

Does that mean  $(6x + 2y)\partial_x + (2x)\partial_y$  is conservative? Not necessarily. To check if  $(6x + 2y)\partial_x + (2x)\partial_y$  is conservative, we would have to check *all possible paths* joining  $(0, 0)$  to  $(2, 0)$ , and then that still wouldn't be enough because we would have to do this for *all possible pairs of endpoints*. In this case, however, there is an easier way for me to convince you that  $(6x + 2y)\partial_x + (2x)\partial_y$  is in fact conservative. Let  $f(x, y) = 3x^2 + 2xy$ , and notice that  $(6x + 2y)\partial_x + (2x)\partial_y$  is  $\vec{\nabla} f$ . The fundamental theorem of calculus now *guarantees* that  $(6x + 2y)\partial_x + (2x)\partial_y$  is conservative. In particular, the fundamental theorem of calculus tells us that the line integral of  $(6x + 2y)\partial_x + (2x)\partial_y$  along *any path* that runs from  $(0, 0)$  to  $(2, 0)$  is simply

$$f(2, 0) - f(0, 0) = 12 - 0 = 12.$$

**Example of NonConservativeness.** Let's revisit the vector field  $y\partial_x + (y - x)\partial_y$  from the second example we considered in this section. Recall that we had computed the line integral of  $y\partial_x + (y - x)\partial_y$  along the straight-line path from  $(0, 2)$  to  $(0, -2)$ , and we got 0 for that. Let's compute the line integral from  $(0, 2)$  to  $(0, -2)$  along a different path, this time a circular arc as shown below.



If we just started with the setup

$$\int (y dx + (y - x) dy)$$

then it is natural to start by trying to use  $x$  or  $y$  as a parameter. Let's say you wanted to use  $y$  as a parameter. Then you would note that, along the curve,  $x = \sqrt{4 - y^2}$ , and so also  $dx = \frac{-y}{\sqrt{4 - y^2}} dy$ .

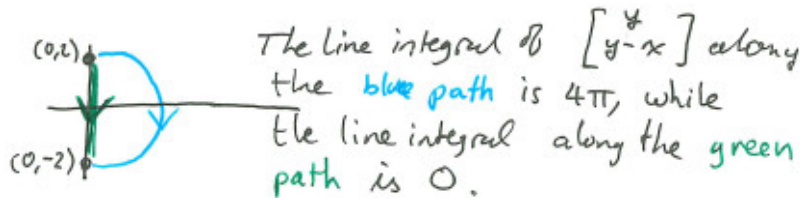
Then you get the single variable integral

$$\int (y dx + (y - x) dy) = \int_2^{-2} \left( y \frac{-y}{\sqrt{4 - y^2}} dy + (y - \sqrt{4 - y^2}) dy \right)$$

with respect to  $y$ . But this integral is too hard, mainly because choosing  $y$  as a parameter just isn't a good way to parameterize a circular arc! A better approach would be to parameterize it as  $(2 \cos(t), -2 \sin(t))$  for  $-\pi/2 \leq t \leq \pi/2$ . Then the line integral is

$$\begin{aligned} \int (y dx + (y - x) dy) &= \int_{-\pi/2}^{\pi/2} \begin{bmatrix} -2 \sin(t) \\ -2 \sin(t) - 2 \cos(t) \end{bmatrix} \cdot \begin{bmatrix} -2 \sin(t) \\ -2 \cos(t) \end{bmatrix} dt \\ &= \int_{-\pi/2}^{\pi/2} (4 \sin^2(t) + 4 \sin(t) \cos(t) + 4 \cos^2(t)) dt \\ &= 4 \int_{-\pi/2}^{\pi/2} (1 + \sin(t) \cos(t)) dt \\ &= 4 \int_{-\pi/2}^{\pi/2} dt + 4 \int_{-\pi/2}^{\pi/2} \sin(t) \cos(t) dt \\ &= 4\pi + 2 \sin^2(t) \Big|_{t=-\pi/2}^{t=\pi/2} = 4\pi. \end{aligned}$$

This shows that  $y \partial_x + (y - x) \partial_y$  is not a conservative vector field!



In particular, this shows that  $y \partial_x + (y - x) \partial_y$  could not possibly be the gradient of a scalar function  $f(x, y)$ . (If it were the gradient of  $f(x, y)$ , then both of the line integrals we just looked at would be the same, because they would both be  $f(0, 2) - f(0, -2)$ .)

**Recovering  $f$  from  $\vec{\nabla}f$ .** Two examples ago, I told you to notice that  $(6x + 2y)\partial_x + (2x)\partial_y$  is the gradient of the function  $3x^2 + 2xy$ , as a way to convince you that  $(6x + 2y)\partial_x + (2x)\partial_y$  is a conservative vector field. How did I find the function  $3x^2 + 2xy$ ? In other words, if I have a vector field  $\vec{F}$  and I know it is  $\vec{\nabla}f$  for some  $f$ , how can I find the  $f$ ? The key is to use the fundamental theorem of calculus—integrate to undo the derivative (or more accurately “line-integrate to undo the gradient”).

If  $\vec{F} = \vec{\nabla}f$  then we know that  $\vec{F}$  is conservative and  $\int_{\vec{x}_0}^{\vec{x}_1} \vec{F} \cdot d\vec{x} = f(\vec{x}_1) - f(\vec{x}_0)$ . So just fix a point  $\vec{x}_0$  and compute the following line integral for arbitrary  $\vec{x}$ :

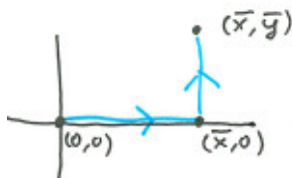
$$\int_{\vec{x}_0}^{\vec{x}} \vec{F} \cdot d\vec{x} = f(\vec{x}) - f(\vec{x}_0).$$

You end up recovering  $f(\vec{x})$  up to a possible shift by a constant  $f(\vec{x}_0)$ . It makes sense that this is the best we can do, since  $\vec{\nabla}$  will not notice shifting  $f$  by a constant. You have to make a choice of starting point  $\vec{x}_0$  to compute all these line integrals, and different choices just amount to different constant shifts of the function you get.

As an example, let’s use this technique on the vector field  $(6x + 2y)\partial_x + (2x)\partial_y$ . I’ll pick the origin  $(0, 0)$  as our starting point. Now we’d like to compute

$$\int_{(0,0)}^{(\bar{x},\bar{y})} ((6x + 2y)dx + (2x)dy)$$

for arbitrary  $(\bar{x}, \bar{y})$ . To compute this we need to choose a path from  $(0, 0)$  to  $(\bar{x}, \bar{y})$ . You can choose any path (because  $\vec{F}$  is conservative), but one of the most straight-forward ones to use is a path that first travels from  $(0, 0)$  to  $(\bar{x}, 0)$ , then travels from  $(\bar{x}, 0)$  to  $(\bar{x}, \bar{y})$ .

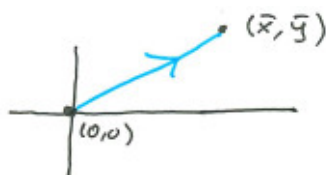


I could call the first part “ $(t, 0)$  for  $0 \leq t \leq \bar{x}$ ” and the second part “ $(\bar{x}, t)$  for  $0 \leq t \leq \bar{y}$ ,” but the notationally simpler approach is to just name the parameter  $x$  for the first part and  $y$  for the second part. We get:

$$\int_{(0,0)}^{(\bar{x},\bar{y})} ((6x + 2y)dx + (2x)dy) = \int_0^{\bar{x}} (6x + 2(0))dx + \int_0^{\bar{y}} 2\bar{x}dy = 3x^2|_0^{\bar{x}} + 2\bar{x}y|_0^{\bar{y}} = 3\bar{x}^2 + 2\bar{x}\bar{y}.$$

And thus we have arrived at a function  $3x^2 + 2xy$  whose gradient is  $((6x + 2y)\partial_x + (2x)\partial_y)$ .

Just to emphasize that you that you could really have used any path to get this answer, let me use a straight line path directly from  $(0, 0)$  to  $(\bar{x}, \bar{y})$ .



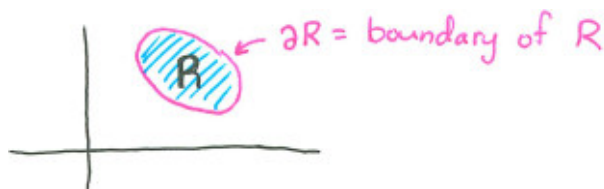
Such a path can be parameterized as  $(t\bar{x}, t\bar{y})$  with  $0 \leq t \leq 1$ . We get

$$\begin{aligned} \int_{(0,0)}^{(\bar{x},\bar{y})} ((6x+2y)dx + (2x)dy) &= \int_0^1 \begin{bmatrix} 6t\bar{x} + 2t\bar{y} \\ 2t\bar{x} \end{bmatrix} \cdot \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} dt \\ &= \int_0^1 6t\bar{x}^2 + 4t\bar{x}\bar{y} dt \\ &= (3t^2\bar{x}^2 + 2t^2\bar{x}\bar{y})|_0^1 \\ &= 3\bar{x}^2 + 2\bar{x}\bar{y}. \end{aligned}$$

### 4.3 Green's Theorem

We will look at a new analogue of the fundamental of theorem of calculus, a theorem that ends up linking two types of integrals we've learned about: area integrals (section 4.1) and line integrals (section 4.2).

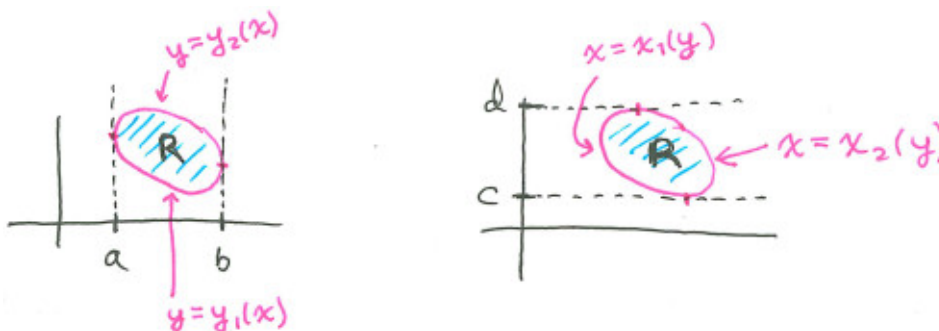
Consider a region  $R$  in  $\mathbb{R}^2$  like the one depicted below.



The boundary of the region is a curve in the plane which we will refer to as " $\partial R$ ." We assume certain things about the region:

- We assume that  $\partial R$  is a simple curve. (A simple curve is one that does not cross itself. So for example  $\partial R$  could not be a figure 8 curve.)
- We assume that  $\partial R$  can be cut in two places so that each of the two pieces has the description  $(x, y(x))$  for some function  $y(x)$ .
- We assume that  $\partial R$  can be cut in two places so that each of the two pieces has the description  $(x(y), y)$  for some function  $x(y)$ .

Here is a depiction of the last two assumptions, with some names for the functions that describe  $\partial R$ :



Now suppose we have a vector field  $\vec{F}(x, y) = P(x, y)\partial_x + Q(x, y)\partial_y$ . We will show that the area integral of  $\frac{\partial P}{\partial y}$  over the region  $R$  is equal to the line integral of the vector field  $P(x, y)\partial_x$  along  $\partial R$ . I will use the notation  $P_y$  for the partial derivative  $\frac{\partial P}{\partial y}$  to make the calculation more clear. To make use of the fundamental theorem of calculus, we should do the area integral by doing a  $y$  integral followed by an  $x$  integral:

$$\begin{aligned}\int_R P_y(x, y) dA &= \int_a^b \int_{y_1(x)}^{y_2(x)} P_y(x, y) dy dx \\ &= \int_a^b (P(x, y_2(x)) - P(x, y_1(x))) dx.\end{aligned}$$

I claimed that this is the line integral of  $P\partial_x$ , but how is this so? Think about how you would compute the line integral of  $P(x, y)\partial_x$  along  $\partial R$ . First of all,  $\partial R$  needs an orientation—let's go with counterclockwise. Next, we need to parameterize it. We have the parameterization  $(x, y_1(x))$  for the bottom part, where  $x$  is the parameter running from  $a$  to  $b$ . We have the parameterization  $(x, y_2(x))$  for the top part, where  $x$  is again the parameter but this time running from  $b$  to  $a$ . Thus

$$\begin{aligned}\int_{\partial R} (P\partial_x) \cdot d\vec{x} &= \int_{\text{bottom part}} (P\partial_x) \cdot d\vec{x} + \int_{\text{top part}} (P\partial_x) \cdot d\vec{x} \\ &= \int_a^b P(x, y_1(x))\partial_x \cdot (\partial_x + y_1'(x)\partial_y) dx + \int_b^a P(x, y_2(x))\partial_x \cdot (\partial_x + y_2'(x)\partial_y) dx \\ &= \int_a^b P(x, y_1(x)) dx + \int_b^a P(x, y_2(x)) dx \\ &= \int_a^b P(x, y_1(x)) dx - \int_a^b P(x, y_2(x)) dx \\ &= \int_a^b (P(x, y_1(x)) - P(x, y_2(x))) dx.\end{aligned}$$

Thus we see that

$$\int_R P_y(x, y) dA = - \int_{\partial R} (P\partial_x) \cdot d\vec{x},$$

assuming  $\partial R$  is oriented counterclockwise. We will now carry out a similar computation to show that the area integral of  $\frac{\partial Q}{\partial x}$  over the region  $R$  is equal to the line integral of the vector field  $Q(x, y)\partial_y$  along  $\partial R$ :

$$\begin{aligned}\int_R Q_x(x, y) dA &= \int_c^d \int_{x_1(y)}^{x_2(y)} Q_x(x, y) dx dy \\ &= \int_c^d (Q(x_2(y), y) - Q(x_1(y), y)) dy.\end{aligned}$$

Now we examine the counterclockwise line integral of  $Q(x, y)\partial_y$  along  $\partial R$ . We have the parameterization  $(x_1(y), y)$  for the left part, where  $y$  is the parameter running from  $d$  to  $c$ . We have the parameterization  $(x_2(y), y)$  for the right part, where  $y$  is again the parameter but this time running

from  $c$  to  $d$ . Thus

$$\begin{aligned}
 \int_{\partial R} (Q\partial_y) \cdot d\vec{x} &= \int_{\text{left part}} (Q\partial_y) \cdot d\vec{x} + \int_{\text{right part}} (Q\partial_y) \cdot d\vec{x} \\
 &= \int_d^c Q(x_1(y), y) \partial_y \cdot (x_1'(y) \partial_x + \partial_y) dy + \int_c^d Q(x_2(y), y) \partial_y \cdot (x_2'(y) \partial_x + \partial_y) dy \\
 &= \int_d^c Q(x_1(y), y) dy + \int_c^d Q(x_2(y), y) dy \\
 &= - \int_c^d Q(x_1(y), y) dy + \int_c^d Q(x_2(y), y) dy \\
 &= \int_c^d (Q(x_2(y), y) - Q(x_1(y), y)) dy.
 \end{aligned}$$

Thus we see that

$$\int_R Q_x(x, y) dA = \int_{\partial R} (Q\partial_y) \cdot d\vec{x},$$

assuming  $\partial R$  is oriented counterclockwise. The line integral of the vector field  $\vec{F} = P\partial_x + Q\partial_y$  is the sum of the two line integrals we just worked with:

$$\int_{\partial R} \vec{F} \cdot d\vec{x} = \int_{\partial R} (P\partial_x) \cdot d\vec{x} + \int_{\partial R} (Q\partial_y) \cdot d\vec{x}.$$

Replacing the line integrals of  $P\partial_x$  and  $Q\partial_y$  by the area integrals to which we just equated them, we have

$$\int_{\partial R} \vec{F} \cdot d\vec{x} = - \int_R P_y(x, y) dA + \int_R Q_x(x, y) dA = \int_R (Q_x - P_y) dA.$$

This is known as *Green's Theorem*.

### Green's Theorem

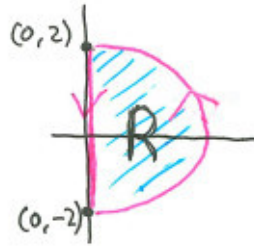
Given a vector field  $\vec{F} = P(x, y)\partial_x + Q(x, y)\partial_y$  on  $\mathbb{R}^2$  and given a region  $R$  with a simple counterclockwise-oriented boundary  $\partial R$ ,

$$\int_{\partial R} \vec{F} \cdot d\vec{x} = \int_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

For the sake of our calculation, we assumed that  $\partial R$  can be cut into pieces that have descriptions like  $(x, y(x))$  or  $(x(y), y)$ . Even without this assumption, Green's theorem still holds (but it is harder to prove).

Any time you want to calculate a line integral of a vector field in  $\mathbb{R}^2$ , Green's theorem gives you the option to instead calculate the area integral of a certain scalar function. In practice you should choose whichever method leads to an easier calculation.

**Example.** Consider the vector field  $\vec{F} = y\partial_x + (y - x)\partial_y$ , and consider the curve which is the boundary of the right half of a disk of radius 2 at the origin, with counterclockwise orientation.



Let  $R$  denote the right half of the disk, so that the curve can be called  $\partial R$ . To see Green's theorem in action, we should consider the line integral of  $\vec{F} = y \partial_x + (y - x) \partial_y$  along  $\partial R$ . We pretty much computed the pieces in the examples of section 4.1— here is a summary of the overall calculation:

$$\begin{aligned} \int_{\partial R} \vec{F} \cdot d\vec{x} &= \int_{\text{arc}} \vec{F} \cdot d\vec{x} + \int_{\text{straight line}} \vec{F} \cdot d\vec{x} \\ &= \int_{-\pi/2}^{\pi/2} \vec{F}(2 \cos(t), 2 \sin(t)) \cdot \begin{bmatrix} -2 \sin(t) \\ 2 \cos(t) \end{bmatrix} dt + \int_2^{-2} \vec{F}(0, y) dy \\ &= (-4\pi) + 0 \\ &= -4\pi. \end{aligned}$$

Now Green's theorem tells us this is supposed to be the same thing as the area integral of the function  $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ , where  $P(x, y) = y$  and  $Q(x, y) = y - x$ . We have

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = -1 - 1 = -2.$$

So Green's theorem says that we could have obtained that answer  $-4\pi$  by simply integrating the function  $-2$  over  $R$ . Indeed we have

$$\begin{aligned} \int_R (-2) dA &= -2 \int_R 1 dA \\ &= -2(\text{area of } R) \\ &= -2 \frac{1}{2} (\text{area of disk of radius } 2) \\ &= -4\pi. \end{aligned}$$

# A Appendix

I will put things here that are useful but do not fit into the main flow of the lectures.

## A.1 Parameterized Lines

If you want a path in  $\mathbb{R}^n$  that starts<sup>13</sup> at the point  $\vec{x}_0$  and travels at a constant velocity  $\vec{v}$ , then it's this:

$$\vec{p}(t) = \vec{x}_0 + t\vec{v}.$$

To verify that this is correct, just look at the starting place and compute the derivative. We have  $\vec{p}(0) = \vec{x}_0$ , so the starting point checks out. As for the derivative, we have

$$\begin{aligned}\vec{p}'(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{p}(t + \Delta t) - \vec{p}(t)) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\vec{x}_0 + (t + \Delta t)\vec{v} - \vec{x}_0 - t\vec{v}) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\Delta t \vec{v}) \\ &= \lim_{\Delta t \rightarrow 0} \vec{v} \\ &= \vec{v},\end{aligned}$$

so the velocity checks out as well:  $\vec{p}'(t) = \vec{v}$  for all  $t$ .

If you want a linear path that goes through two specified points, then you can just use one of the points as the starting point and use their difference vector as the velocity. So for example if you want a linear path going through  $\vec{x}_1$  and  $\vec{x}_2$ , then you can just write down the parameterized line that starts at  $\vec{x}_1$  and has velocity vector  $\vec{x}_2 - \vec{x}_1$ :

$$\vec{p}(t) = \vec{x}_1 + t(\vec{x}_2 - \vec{x}_1).$$

Of course, the answer here is not unique since we are just asking for a path that goes through  $\vec{x}_1$  and  $\vec{x}_2$  and we are not specifying *when* (for what  $t$ ) it reaches those points or how fast it needs to travel.

Finally, here is a nifty way to quickly write down a linear path that passes through the point  $\vec{x}_1$  at  $t = 0$  and passes through  $\vec{x}_2$  at  $t = 1$ :

$$\vec{p}(t) = (1 - t)\vec{x}_1 + t\vec{x}_2.$$

See that it works by evaluating at  $t = 0$  and  $t = 1$ .

## A.2 Norm and Dot Product (Lengths and Angles)

Recall from section 1.4 that we have two uses of the word “vector” in this course: (1) There's the abstract notion of vector, where we have one unspecified vector space in mind, usually  $\mathbb{R}^n$ . When

---

<sup>13</sup>The *starting point* of a path  $\vec{p}(t)$  is the point  $\vec{p}(0)$ .



treating elements of  $\mathbb{R}^n$  (i.e. lists of  $n$  real numbers) as vectors, we think of them not as points but as steps emanating from the origin. (2) Then there's the notion of vector that is specific to vector calculus, for which we have been using the term "vector-at-a-point." These can also be represented by lists of real numbers (elements of  $\mathbb{R}^n$ ), but we think of them as being located at a particular point in space (where "space" is often also  $\mathbb{R}^n$  but could be one of the more exotic things appearing in section 2.2). We found in section 2.1.4 that representing vectors-at-a-point by using lists of real numbers is pretty confusing when different coordinate systems get involved, so we adopted the superior " $\partial_u$ " notation for coordinate basis vectors. Now we express vectors-at-a-point as linear combinations of coordinate basis vectors.

This appendix section is about *lengths* of vectors and *angles* between vectors. It will be presented in terms of the abstract version of vectors, viewpoint (1) in the paragraph above. Then when we return to the main text we will look at what these ideas mean for vector calculus, viewpoint (2) in the paragraph above.

### A.2.1 The Dot Product

The dot product is an operation that takes two vectors and produces a scalar. In other words it takes two lists of numbers and combines them to somehow give you just one number. The definition is very simple:

#### Definition of Dot Product

Given two elements  $\vec{x} = (x_1, x_2, \dots, x_n)$  and  $\vec{y} = (y_1, y_2, \dots, y_n)$  of  $\mathbb{R}^n$ , their *dot product* is denoted by  $\vec{x} \cdot \vec{y}$  and it is defined to be the scalar

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

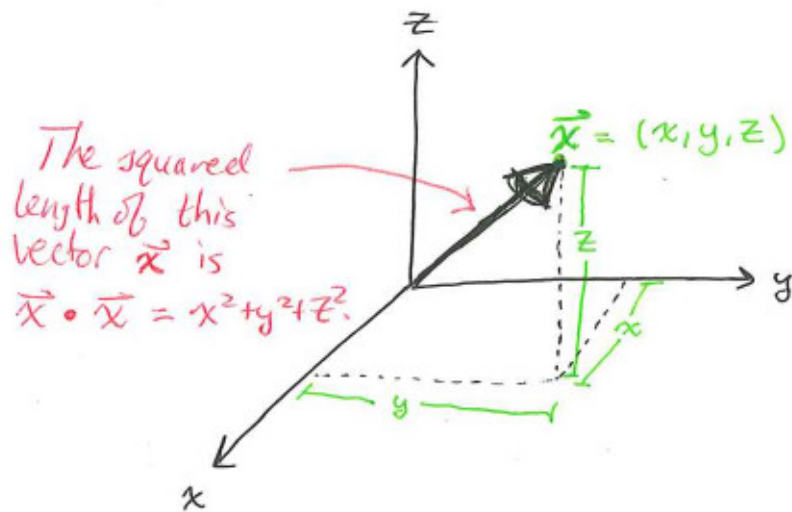
The first thing to notice is what happens when you dot a vector with itself. If  $\vec{x} = (x_1, x_2, \dots, x_n)$ , then

$$\vec{x} \cdot \vec{x} = x_1^2 + x_2^2 + \dots + x_n^2.$$

It gives the sum of the squares of the components. That should be a familiar looking construction; maybe it will look more familiar in two or three dimensions:

$$\begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2 \qquad \begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x^2 + y^2 + z^2$$

By the pythagorean theorem (or the "distance formula"), dotting a vector with itself gives us the square of its distance from the origin! Since we are thinking of a vector as an arrow emanating from the origin, I should really say this is giving us the squared *length* of the arrow.



We will use this observation to formally define the length of a vector in  $\mathbb{R}^n$ , which often goes by other names such as *norm* or *magnitude*. I tend to like using *norm*.

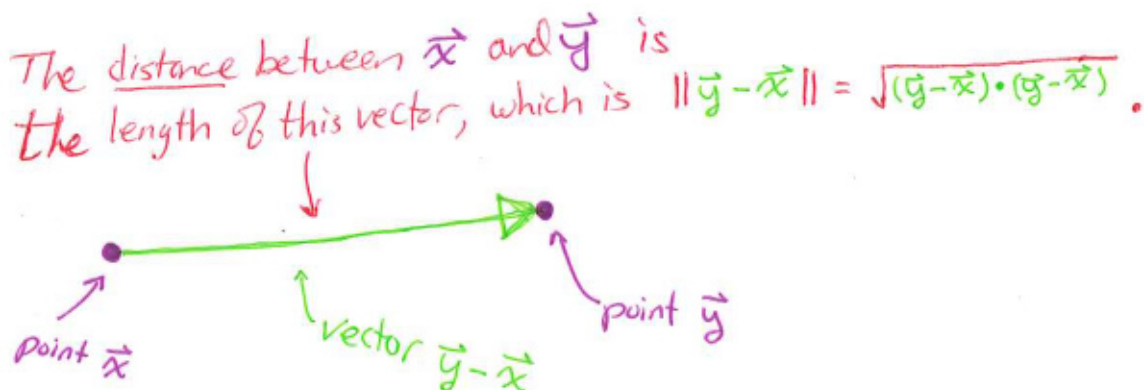
#### Definition of Norm

The *norm* (or *length* or *magnitude*) of a vector  $\vec{x}$  in  $\mathbb{R}^n$  is denoted by  $\|\vec{x}\|$  and it is defined to be

$$\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}}.$$

Note that the square root never gives any trouble because  $\vec{x} \cdot \vec{x}$ , being a sum of squares, can never be negative.

This notation gives us a convenient way to talk about the distance between two points in  $\mathbb{R}^n$ . Recall from section 1.4 that the vector difference  $\vec{y} - \vec{x}$  between two *points*  $\vec{x}$  and  $\vec{y}$  gives a meaningful *vector* which is the displacement vector from  $\vec{x}$  to  $\vec{y}$ . The norm  $\|\vec{y} - \vec{x}\|$  of that displacement vector is then the distance from  $\vec{x}$  to  $\vec{y}$ .



## A.2.2 Algebra of Dot Products

There's a good reason for the dot product symbol to look like a fat multiplication symbol. Dot products interact with vector addition in such a way that they follow the same rules that multiplication of scalars follows in its interaction with addition of scalars.

$$\vec{x} \cdot (\vec{v} + \vec{w}) = \vec{x} \cdot \vec{v} + \vec{x} \cdot \vec{w}.$$

By the way, dot product takes the same precedence as multiplication in the order of operations, so the expression on the right hand side above is equivalent to  $(\vec{x} \cdot \vec{v}) + (\vec{x} \cdot \vec{w})$ . There is no ambiguity when you string together fat dots and little dots: if  $\alpha$  is a scalar then both interpretations of  $\alpha \cdot \vec{v} \cdot \vec{w}$  are equal:

$$\alpha \cdot (\vec{v} \cdot \vec{w}) = (\alpha \cdot \vec{v}) \cdot \vec{w}.$$

Little dots (scaling or multiplication) can be omitted from notation, so we could write  $\alpha \vec{v} \cdot \vec{w}$  for the above. But you should never omit a fat dot.

Dot products also share with multiplication the property of *commutativity*:

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}.$$

Be careful not to just assume this for every new operation you learn about. Later we will introduce another operation on vectors that fails to be commutative.

The algebraic rules given here are all ones you can prove to yourself in a few steps by just using the definition of the dot product. You might want to go back and think about why each rule works.

Here is some advice. When you work with expressions involving vectors and scalars, it will be very helpful to keep track of the *types* of things. When you spot a new expression like

$$(\vec{x} \cdot \vec{y}) \frac{\vec{v}}{\|\vec{w} + \vec{v}\|} - \|\vec{y}\| \frac{\vec{x} + \vec{y}}{\vec{w} \cdot \vec{x} + 2\vec{y} \cdot (\|\vec{x}\| \vec{y} + \vec{w})},$$

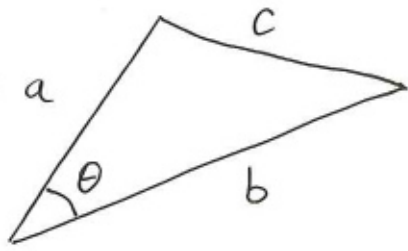
go through a *type check* before you even try to understand it geometrically. Which subexpressions are vectors and which are scalars? Which of the operations involved are vector addition and which are just scalar addition? Where are the dot products and where are the scalings? Try doing this for the expression above.

Since we defined norms in terms of dot products, norms get a lot of nice properties that you can check on the fly when you need them. For example, scaling a vector scales its norm:

$$\|\alpha \vec{x}\| = \sqrt{(\alpha \vec{x}) \cdot (\alpha \vec{x})} = \sqrt{\alpha^2 (\vec{x} \cdot \vec{x})} = |\alpha| \sqrt{\vec{x} \cdot \vec{x}} = |\alpha| \|\vec{x}\|.$$

## A.2.3 Angles

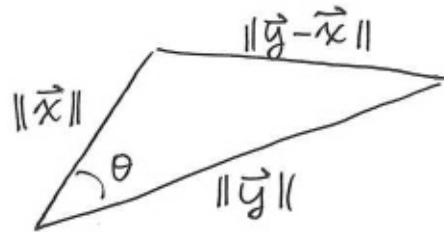
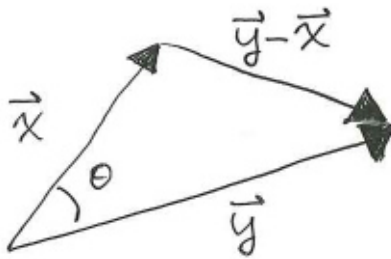
By using the pythagorean theorem, we managed to make geometric sense of  $\vec{x} \cdot \vec{x}$ . What about a dot product of two *different* vectors like  $\vec{x} \cdot \vec{y}$ ? To get at  $\vec{x} \cdot \vec{y}$  geometrically we will need a souped up version of the pythagorean theorem, known as the *law of cosines*:



Law of Cosines:

$$c^2 = a^2 + b^2 - 2ab \cos \theta.$$

You may have seen this before in high school geometry. See how it reduces to the pythagorean theorem when  $\theta$  is a right angle? Pretty soon you can forget about this law, because we will replace it with something much prettier involving dot products. We will apply the law of cosines to relate the angle  $\theta$  between two vectors  $\vec{x}$  and  $\vec{y}$  to their dot product. The angle  $\theta$  sits in a triangle whose sides are the vectors  $\vec{x}$ ,  $\vec{y}$ , and  $\vec{y} - \vec{x}$ . This would be a triangle whose sides have *lengths* given by  $\|\vec{x}\|$ ,  $\|\vec{y}\|$  and  $\|\vec{y} - \vec{x}\|$ .



The law of cosines says that

$$\|\vec{y} - \vec{x}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\|\vec{x}\|\|\vec{y}\|\cos(\theta).$$

Writing out some of the norms in terms of dot products, this is saying that

$$(\vec{y} - \vec{x}) \cdot (\vec{y} - \vec{x}) = \vec{x} \cdot \vec{x} + \vec{y} \cdot \vec{y} - 2\|\vec{x}\|\|\vec{y}\|\cos(\theta).$$

Do some algebra on the left hand side, using distributivity to expand things out and commutativity to combine  $\vec{x} \cdot \vec{y}$  and  $\vec{y} \cdot \vec{x}$ :

$$\vec{y} \cdot \vec{y} - 2\vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{x} = \vec{x} \cdot \vec{x} + \vec{y} \cdot \vec{y} - 2\|\vec{x}\|\|\vec{y}\|\cos(\theta).$$

Cancelling the repeated terms on both sides and then dividing by  $-2$  gives

$$\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\|\cos(\theta).$$

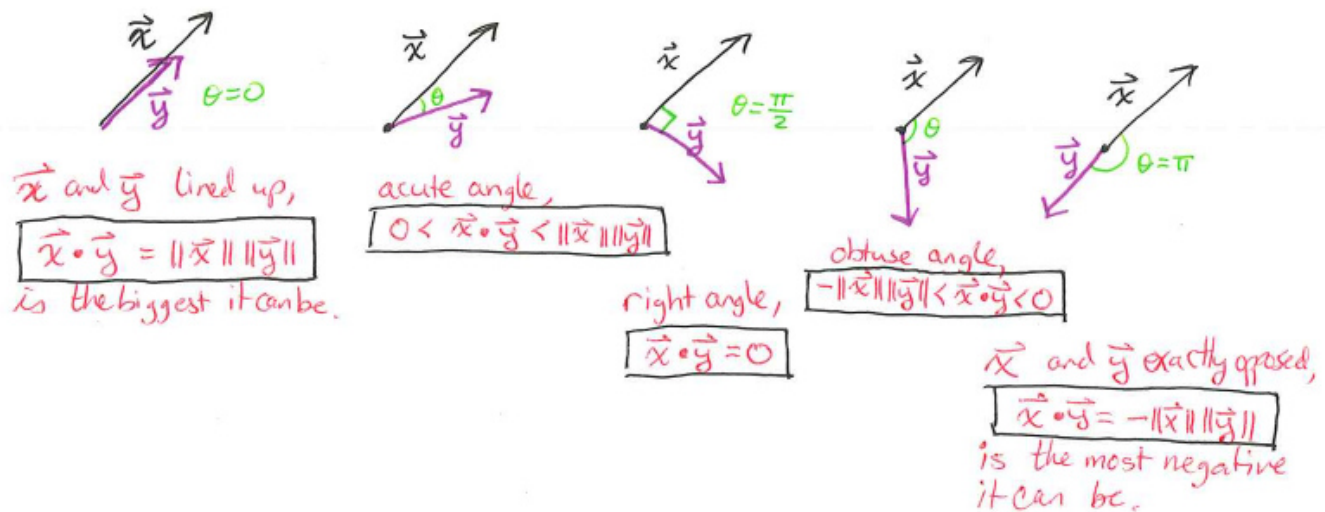
This is how we can geometrically understand how dot products work! The dot product of two vectors gives the cosine of the angle between them, scaled up by the lengths of both of the vectors.

### Dot Products in Terms of Angles

The dot product of two vectors  $\vec{x}$  and  $\vec{y}$  is related to the angle between them as follows:

$$\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\|\cos(\text{the angle between } \vec{x} \text{ and } \vec{y}).$$

Here is a way to interpret this result. Assume that you've got two vectors  $\vec{x}$  and  $\vec{y}$  and that their length is fixed, but you get to swing them around and change their direction to whatever you want. Then in the product  $\|\vec{x}\| \|\vec{y}\| \cos(\theta)$ , only the  $\cos(\theta)$  part changes while the  $\|\vec{x}\| \|\vec{y}\|$  part stays the same. Also, the  $\|\vec{x}\| \|\vec{y}\|$  part is non-negative, so the sign of  $\vec{x} \cdot \vec{y}$  is the same as the sign of  $\cos(\theta)$ . Now as you swing around  $\vec{x}$  and  $\vec{y}$ , how does  $\vec{x} \cdot \vec{y}$  respond? Since  $\cos(\theta)$  is biggest when  $\theta = 0$ , the dot product  $\vec{x} \cdot \vec{y}$  is the biggest (and is positive) when you line up the vectors  $\vec{x}$  and  $\vec{y}$  so that the angle between them is 0. If you deviate from perfectly lining up  $\vec{x}$  and  $\vec{y}$ , the dot product responds by decreasing. So you can think of  $\vec{x} \cdot \vec{y}$  as a measure of the extent to which  $\vec{x}$  and  $\vec{y}$  point in the same direction. Think about how the  $\cos(\theta)$  part of  $\vec{x} \cdot \vec{y} = \|\vec{x}\| \|\vec{y}\| \cos(\theta)$  leads to the following situations:



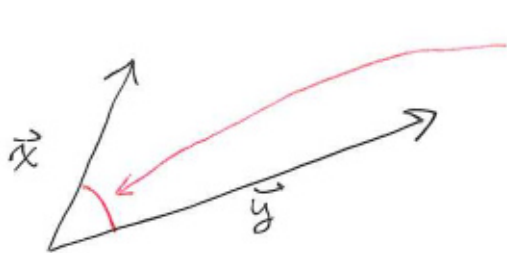
The middle situation, with  $\vec{x} \cdot \vec{y} = 0$ , has some special terminology<sup>14</sup>.

#### Definition of Orthogonality

Two vectors  $\vec{x}$  and  $\vec{y}$  are said to be *orthogonal* if  $\vec{x} \cdot \vec{y} = 0$ .

The dot product is such an easy thing to compute. You just multiply components and then add. And dimensionality doesn't add any computational difficulty; even dot products of vectors in  $\mathbb{R}^{2358}$  can be obtained by just multiplying and adding. Yet the dot product gives us a simple way to talk about things that may have previously seemed quite complicated. Angles are now a piece of cake.

<sup>14</sup>We already have the word "perpendicular" to describe when two vectors make a right angle with one another. The point of having this new terminology "orthogonal" is to focus on dot products rather than angles. There is one situation where the terms differ: the zero vector. One would never use the term "perpendicular" on the zero vector because the zero vector has no direction. But we do still say that that zero vector is *orthogonal* to every other vector, simply because the dot product says so:  $\vec{0} \cdot \vec{x} = 0$  for all  $\vec{x}$ .



This angle  $\theta$  satisfies

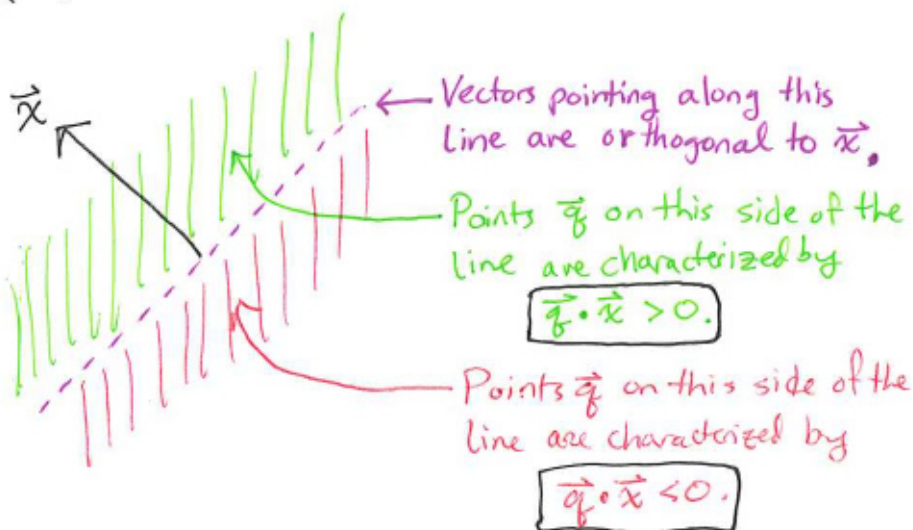
$$\vec{x} \cdot \vec{y} = \|\vec{x}\| \|\vec{y}\| \cos \theta, \text{ so}$$

if you want to know what it is,

it's just  $\theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|}\right)$

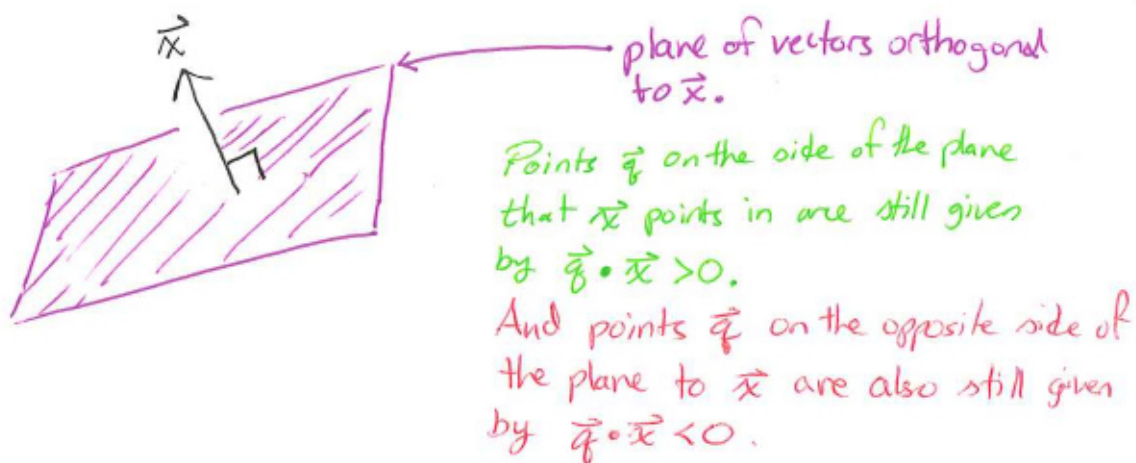
And now we also have a way of talking about the set of points on the “same side” or “opposite side” to a given vector, with the orthogonal vectors lining the barrier in between.

In  $\mathbb{R}^2$ :



The pictures are two dimensional but the ideas work in  $\mathbb{R}^3$  and in  $\mathbb{R}^n$  for any  $n$ !

In  $\mathbb{R}^3$ :



## A.2.4 Unit Vectors and Normalization

A *unit vector* is a vector  $\vec{x}$  for which  $\|\vec{x}\| = 1$ . Sometimes it's useful to use a unit vector to describe a direction, if you want your description of the direction to not also contain a choice of length. For example  $(1, 0, 0)$  and  $(2, 0, 0)$  are both good descriptions of the “positive  $x$  direction in  $x$ - $y$ - $z$ -space,” but somehow  $(1, 0, 0)$  is a purer way to capture that concept because it doesn't have the extra baggage of having length 2. Similarly  $(1, 1, 0)$  and  $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$  both describe the same direction in  $\mathbb{R}^3$ , but  $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$  is nicer in that it doesn't have the extra baggage of having length  $\sqrt{2}$ .

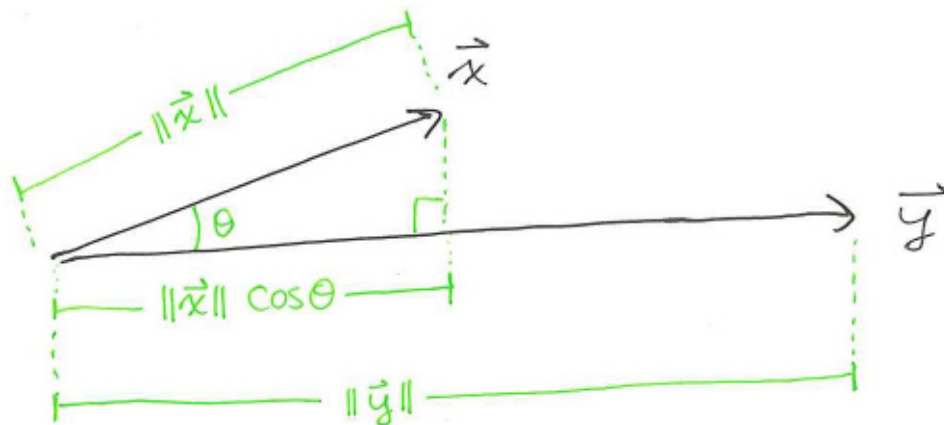
Suppose you have a vector  $\vec{x}$  in  $\mathbb{R}^n$  that describes a direction in which you are interested, but  $\|\vec{x}\|$  is not 1. There is a very easy way to get a vector that points in the same direction but has norm 1: just scale by  $\frac{1}{\|\vec{x}\|}$ ! The process is called *normalization*, and the result  $\frac{1}{\|\vec{x}\|}\vec{x}$  is called the *normalization* of  $\vec{x}$ . To check that it works, just use the fact that scalars pop out of the norm operation:

$$\left\| \frac{1}{\|\vec{x}\|} \vec{x} \right\| = \frac{1}{\|\vec{x}\|} \|\vec{x}\| = 1.$$

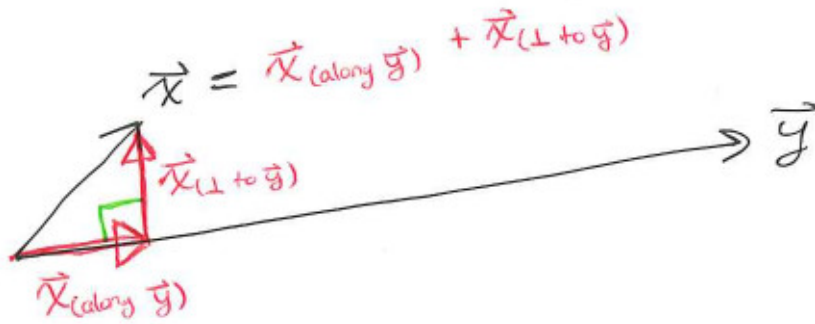
What if you want a vector that points in the same direction as  $\vec{x}$  but that has length 57.23? Just normalize, and then scale by the length you want it to have:  $\frac{57.23}{\|\vec{x}\|}\vec{x}$ .

## A.2.5 Projection: The Component of One Vector Along Another

Let's revisit  $\vec{x} \cdot \vec{y} = \|\vec{x}\| \|\vec{y}\| \cos(\theta)$ , but this time focus on a different portion of it than just  $\cos(\theta)$ . This time I want to focus on the  $\|\vec{x}\| \cos(\theta)$  portion. Here is a picture in which this portion appears as the length of a certain segment:



The scalar  $\|\vec{x}\| \cos(\theta)$  is the length of what you could call “the component of  $\vec{x}$  along the direction of  $\vec{y}$ .” Think of it this way: you can split  $\vec{x}$  into a sum of two parts, the part  $\vec{x}_{(\text{along } \vec{y})}$  along  $\vec{y}$  and the part  $\vec{x}_{(\perp \text{ to } \vec{y})}$  orthogonal to  $\vec{y}$ :



We saw in the picture above that  $\|\vec{x}_{(\text{along } \vec{y})}\| = \|\vec{x}\| \cos \theta$ . To someone who is not familiar with the magic of dot products, this is an annoying thing to compute since it involves trigonometry. But we have a better way, because we can replace  $\|\vec{x}\| \cos \theta$  by an expression involving the dot product. Just solve for  $\|\vec{x}\| \cos \theta$  in  $\vec{x} \cdot \vec{y} = \|\vec{x}\| \|\vec{y}\| \cos(\theta)$ :

$$\|\vec{x}_{(\text{along } \vec{y})}\| = \|\vec{x}\| \cos \theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{y}\|}.$$

Now that we have the *norm* of  $\vec{x}_{(\text{along } \vec{y})}$ , we can almost write down a formula for it. We just need to know its direction, but that's obvious— it should be along the direction of  $\vec{y}$ ! We saw in section A.2.4 how to describe a vector pointing in the direction of  $\vec{y}$  with any desired norm: we just normalize  $\vec{y}$  and then scale by the norm we want. Thus

$$\vec{x}_{(\text{along } \vec{y})} = \frac{\vec{x} \cdot \vec{y}}{\|\vec{y}\|} \frac{1}{\|\vec{y}\|} \vec{y} = \frac{\vec{x} \cdot \vec{y}}{\|\vec{y}\|^2} \vec{y} = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \vec{y}.$$

If you want a formula for  $\vec{x}_{(\perp \text{ to } \vec{y})}$  as well, then you can just solve for it in the decomposition  $\vec{x} = \vec{x}_{(\text{along } \vec{y})} + \vec{x}_{(\perp \text{ to } \vec{y})}$ :

$$\vec{x}_{(\perp \text{ to } \vec{y})} = \vec{x} - \vec{x}_{(\text{along } \vec{y})} = \vec{x} - \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \vec{y}.$$

**Exercise:** Prove that  $\vec{x} - \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \vec{y}$  really is orthogonal to  $\vec{y}$ .

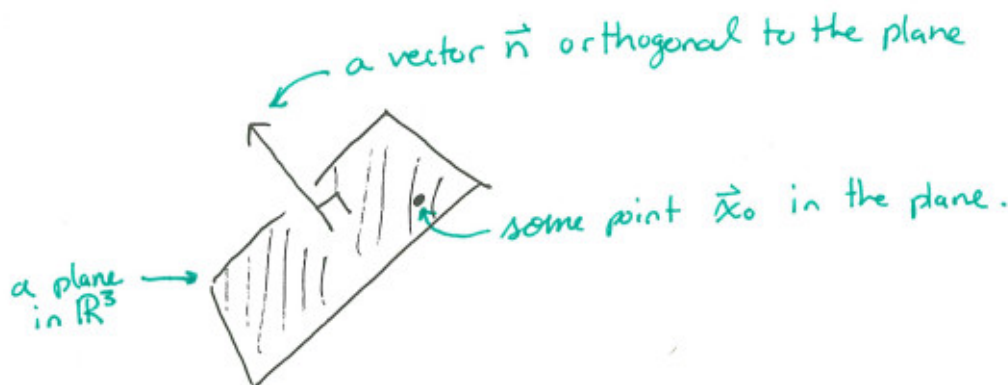
These formulas give us some more insight into what dot products are doing. If  $\vec{y}$  is a *unit* vector, then  $\vec{x} \cdot \vec{y}$  is equal to  $\frac{\vec{x} \cdot \vec{y}}{\|\vec{y}\|} = \|\vec{x}\| \cos(\theta) = \|\vec{x}_{(\text{along } \vec{y})}\|$ . So dotting  $\vec{x}$  with a unit vector  $\vec{y}$  gives the length of the component of  $\vec{x}$  along the direction of  $\vec{y}$ , or as I sometimes like to read it “the  $\vec{y}$ -ness of  $\vec{x}$ .” If  $\vec{y}$  is not a unit vector, then dotting  $\vec{x}$  with it gives you the same thing except it is also scaled by  $\|\vec{y}\|$ .

### A.3 Equations of Planes

How would you specify a plane in  $\mathbb{R}^3$ ? That is, if you have a certain plane in  $\mathbb{R}^3$  in mind then how can you communicate to me which plane you are thinking of? One way is for you to give me three points on the plane. That works, but its not as nice as the following way: Give just one point on the plane, and give a vector orthogonal to the plane. Specifying a vector orthogonal to the plane is a way of giving the plane's orientation in space. If you *only* give a plane's orientation, then that still does not pin down where it is (you can imagine translating a plane around in  $\mathbb{R}^3$  while fixing



its orientation). If you additionally gave just *one point* on the plane, then that completely pins down where it is.

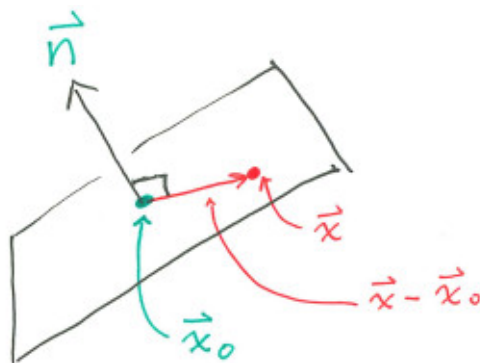


### Specifying a Plane in $\mathbb{R}^3$

A plane in  $\mathbb{R}^3$  can be specified by giving a point  $\vec{x}_0$  on it and a vector  $\vec{n}$  orthogonal to it. The vector  $\vec{n}$  is called a *normal vector* for the plane.

*Normal*, like *orthogonal*, is just another word for *perpendicular*.

This way of specifying planes, by a point and a normal vector, makes it very easy to write down an equation for the plane. Suppose you have a plane with a normal vector  $\vec{n}$  and you know that the point  $\vec{x}_0$  is on the plane. Given a point  $\vec{x}$  in  $\mathbb{R}^3$ , what equation does  $\vec{x}$  have to satisfy to be *on the plane*? Draw the normal vector with its tail end at  $\vec{x}_0$  and consider what must be true of random points  $\vec{x}$  on the plane:



The displacement vector  $\vec{x} - \vec{x}_0$  must always be perpendicular to  $\vec{n}$ . For  $\vec{x}$  to be on the plane, the exact condition that needs to be satisfied is  $\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0$ . And that's it for the equation of a plane! Note that even though vectors are involved, it is still just one scalar equation.

Other ways of writing the equation can be derived. If  $\vec{n} = (a, b, c)$  and  $\vec{x}_0 = (x_0, y_0, z_0)$ , then the points  $(x, y, z)$  on the plane are ones satisfying

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

This could be rearranged to

$$ax + by + cz - ax_0 - by_0 - cz_0 = 0,$$

or just

$$ax + by + cz + d = 0,$$

if we define  $d$  to be the constant  $-ax_0 - by_0 - cz_0$ . Anytime you see an equation like  $ax + by + cz + d = 0$ , that is the equation of a plane (well, as long as  $a, b, c$  are not all zero). (This should remind you of how  $ax + by + c = 0$  is a way of specifying a line in the plane.) If a plane is given to you by an equation of the form  $ax + by + cz + d = 0$ , then you can easily see the entries of a normal vector for the plane: they are just the coefficients  $a, b, c$  in front of  $x, y$ , and  $z$ .

**Example** The plane in  $\mathbb{R}^3$  going through the point  $(1, 2, 4)$  with a normal vector of  $(3, -5, 2)$  is the set of points  $(x, y, z)$  satisfying the equation

$$\begin{bmatrix} 3 \\ -5 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} x - 1 \\ y - 2 \\ z - 4 \end{bmatrix} = 0,$$

which could also be written as

$$3(x - 1) - 5(y - 2) + 2(z - 4) = 0.$$

This could be rearranged to  $3x - 5y + 2z - 1 = 0$  or  $3x - 5y + 2z = 1$ . If the equation  $3x - 5y + 2z = 1$  were given to us from the start, then we can know right away that it's a plane and we can know its orientation: it has a normal vector of  $(3, -5, 2)$  because the coefficients in front of  $x, y, z$  are  $3, -5, 2$ .

Note that normal vectors for planes are only determined up to a scaling. The purpose of a normal vector is only to specify a direction. If we replace the normal vector  $(3, -5, 2)$  by its double  $(6, -10, 4)$  then we end up with the equation

$$6(x - 1) - 10(y - 2) + 4(z - 4) = 0.$$

(which is also  $6x - 10y + 4z = 2$ ), which describes exactly the same set of points.

### Equations of Planes in $\mathbb{R}^3$

The plane in  $\mathbb{R}^3$  that passes through  $\vec{x}_0$  and has  $\vec{n}$  as a normal vector is the set of points  $\vec{x}$  satisfying  $\vec{n} \cdot (\vec{x} - \vec{x}_0) = 0$ . This can be rewritten in the form  $ax + by + cz + d = 0$  for some constants  $a, b, c, d$ . Any such equation is the equation of a plane, and  $(a, b, c)$  is a normal vector for the plane.

## A.4 Cross Products

Here we introduce a new binary operation on vectors, the *cross product*, which, unlike the dot product, yields a vector rather than a scalar. Also unlike the dot product, the cross product is very specific to  $\mathbb{R}^3$ .

The cross product of two vectors  $\vec{x}_1 = (x_1, y_1, z_1)$  and  $\vec{x}_2 = (x_2, y_2, z_2)$  is denoted by  $\vec{x} \times \vec{y}$  and it is the *vector*

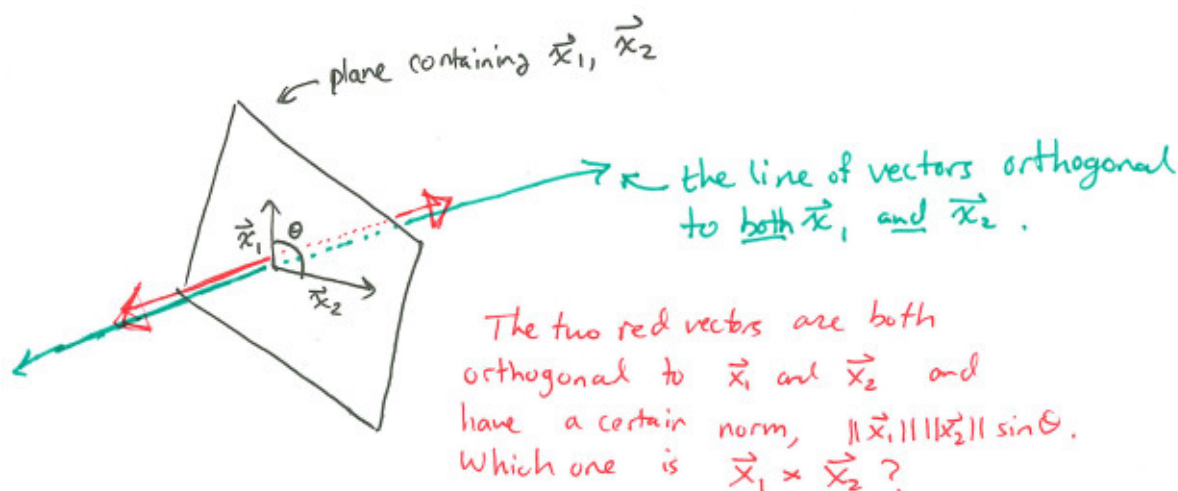
$$\vec{x}_1 \times \vec{x}_2 = \begin{bmatrix} y_1 z_2 - y_2 z_1 \\ -(x_1 z_2 - x_2 z_1) \\ x_1 y_2 - x_2 y_1 \end{bmatrix}.$$

The main reason for introducing this thing is that its norm and its direction are both pretty useful. The cross product  $\vec{x} \times \vec{y}$  is a vector *orthogonal* to both  $\vec{x}$  and  $\vec{y}$ , so taking a cross product is a convenient way to get a vector orthogonal to two other vectors. The norm  $\|\vec{x} \times \vec{y}\|$  measures the area of the parallelogram with sides  $\vec{x}$  and  $\vec{y}$ , a fact that will later be handy for writing certain kinds of integrals. More specifically, the norm of a cross product is

$$\|\vec{x} \times \vec{y}\| = \|\vec{x}\| \|\vec{y}\| \sin(\text{the angle between } \vec{x} \text{ and } \vec{y}).$$

The appearance of the *sine* of the angle between the vectors (as opposed to cosine, which appeared in dot products) has the following implication: A cross product between two vectors vanishes when the vectors are parallel, and its norm is maximized when the vectors are perpendicular.

Now you've *almost* got a complete geometric characterization of the cross product. You know that  $\vec{x}_1 \times \vec{x}_2$  points in a direction orthogonal to both  $\vec{x}_1$  and  $\vec{x}_2$ . This leaves a line of possible vectors that it could be, the line of vectors in  $\mathbb{R}^3$  that are orthogonal to both  $\vec{x}_1$  and  $\vec{x}_2$ . You also know that the length of  $\vec{x}_1 \times \vec{x}_2$  is  $\|\vec{x}_1\| \|\vec{x}_2\| \sin(\theta)$ , where  $\theta$  is the angle between  $\vec{x}_1$  and  $\vec{x}_2$ . But this still leaves two possibilities for what  $\vec{x}_1 \times \vec{x}_2$  could be.



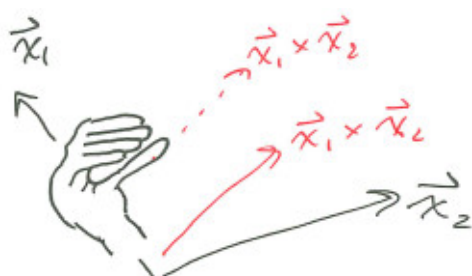
Here we just have to go ahead and pick a convention, and stick to that convention throughout the course. The convention we choose here is the *right hand rule* (the other convention is the *left hand rule*). Here's how it goes: To know which of the two possible vectors is  $\vec{x}_1 \times \vec{x}_2$ , start by pointing the fingers of your right hand in the direction of  $\vec{x}_1$ , the first vector in the product  $\vec{x}_1 \times \vec{x}_2$ .



Then curl your fingers so that they point in the direction of  $\vec{x}_2$ , the second vector in the product  $\vec{x}_1 \times \vec{x}_2$ .



Your thumb now points in the direction of  $\vec{x}_1 \times \vec{x}_2$ .



(the thumb is pointing in a direction perpendicular to both  $\vec{x}_1$  and  $\vec{x}_2$ , and into the page)

The right hand rule (and the left hand rule, if we had gone with that) has the following important implication:

$$\vec{x}_1 \times \vec{x}_2 = -\vec{x}_2 \times \vec{x}_1.$$

The vector cross product is a noncommutative operation. The order in which you cross vectors together matters!

Other than this weirdness, cross products behave like a multiplication when it interacts with vector addition,

$$\vec{x}_1 \times (\vec{x}_2 + \vec{x}_3) = \vec{x}_1 \times \vec{x}_2 + \vec{x}_1 \times \vec{x}_3,$$

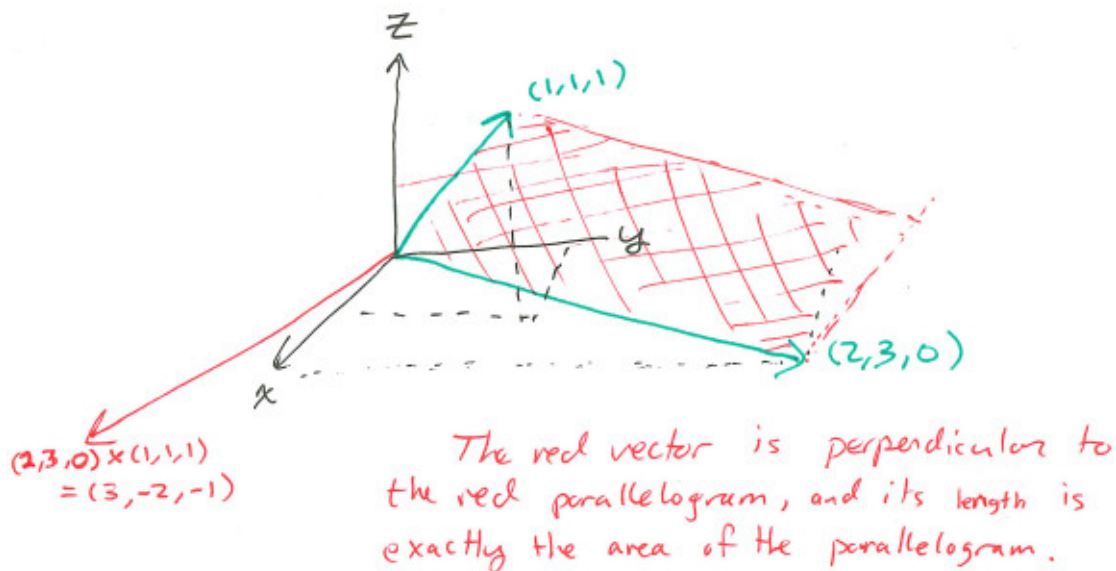
or vector scaling,

$$\vec{x}_1 \times (\alpha\vec{x}_2) = (\alpha\vec{x}_1) \times \vec{x}_2 = \alpha(\vec{x}_1 \times \vec{x}_2).$$

**Example** If vectors are given explicitly in terms of their components, then computing the cross product is just a matter of applying the formula. So for example the cross product between  $(2, 3, 0)$  and  $(1, 1, 1)$  is

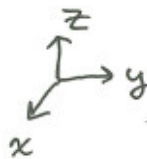
$$\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3(1) - 0(1) \\ -(2(1) - 0(1)) \\ 2(1) - 3(1) \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}.$$

This is giving us a vector orthogonal to both  $(2, 3, 0)$  and  $(1, 1, 1)$ , as you could check for yourself by taking dot products.

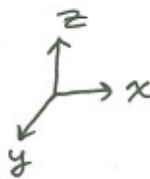


And the norm  $\|(3, -2, -1)\| = \sqrt{14}$  is telling us the area of the parallelogram that has  $(2, 3, 0)$  and  $(1, 1, 1)$  as sides.

**Right handed coordinate system** Let's use  $\partial_x, \partial_y, \partial_z$  for the standard basis vectors in  $\mathbb{R}^3$ . Use the right hand rule to convince yourself that  $\partial_x \times \partial_y = \partial_z$  and  $\partial_x \times \partial_z = -\partial_y$ . You should have no trouble doing this if you are imagining the coordinate axes like this:



But try computing the same thing with the coordinate axes drawn like this:



You end up getting  $\partial_x \times \partial_y = -\partial_z$  and  $\partial_x \times \partial_z = \partial_y$ ! The right hand rule is actually giving you the wrong answer here. The formula for the cross product makes it very clear that

$$\partial_x \times \partial_y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \partial_z \quad \partial_x \times \partial_z = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = -\partial_y.$$

The problem is that the second way of drawing coordinate axes is “left-handed.” A cartesian coordinate system is called “right-handed” if  $\partial_x \times \partial_y = \partial_z$ . As long as your coordinate axes are “right-handed,” the right hand rule will always correctly tell you the direction of a cross product.

### Cross Products Summary

The cross product  $\vec{x}_1 \times \vec{x}_2$  is

$$\vec{x}_1 \times \vec{x}_2 = \begin{bmatrix} y_1 z_2 - y_2 z_1 \\ -(x_1 z_2 - x_2 z_1) \\ x_1 y_2 - x_2 y_1 \end{bmatrix}.$$

The order matters—switching the order and writing  $\vec{x}_2 \times \vec{x}_1$  instead would produce a minus sign. The norm of a cross product is the product of the norms of the vectors times the sine of the angle between them. The direction of a cross product is normal to the plane containing the two vectors, and in the direction suggested by the right hand rule.