MATH 111: TECHNIQUES IN CALCULUS II

S. JAMSHIDI

Date: April 13, 2015.

Objectives

- ☐ I can distinguish between measurements which are stocks (quantities) versus measurements which are flows (rates).
- ☐ I understand that a function is a mathematical description of a relationship between two or more measurements.
- ☐ I understand the difference between an independent and dependent variable.

What does it mean to do mathematics? What exactly are we doing when we do math? These aren't easy questions to answer. Some of the responses I heard in class were:

- Math is the act of solving quantitative problems.
- Math involved analyzing data.
- Math is the logical treatment of variables.
- Math requires critical thinking.

All of these capture an important component of math, especially applied math (mathematics used to characterize our observations in reality).

Definition 1.1

For this class, mathematics is the attempt to describe and understand relationships between two or more measurements.

It is important to note that the subject of mathematics is much broader than this, but the math we cover in this class can be described as above.

There are two key words in our definition that demand our attention: *relationship* and *measurement*.

We can classify measurements as being of two types:

(1) **Stocks**
(2) **Flows**

A **stock** is an an amount of something existing at a particular point in time. Here are some examples of stocks.

Example 1.2: The amount of junk food in pounds in your house right now.

Example 1.3: The amount of money in US dollars your bank account tomorrow morning.

Example 1.4: The number of chipotle burritos in the US right now.

Each of these is an example of a stock. They are accumulated amounts of something (junk food, money, chipotle burritos) at some point in time (right now, tomorrow morning, right now). Because these are measured items, they each have a corresponding unit we use to *quantify* them: pounds, US dollars, chipotle burritos.

A **flow** is a measurement of how a stock changes over time. Let us list some flows associated to our stocks in the above examples.

Example 1.5: The amount of junk food (in pounds) you eat in a week.

Example 1.6: The amount of money (in US dollars) you make in a month.

Example 1.7: The number of chipotle burritos eaten every 5 minutes in the US.

Each of these amounts will tell us how the stocks previously listed will change. Not all stocks describe the same movementment. In examples 1.5 and 1.7, we are looking at **outflows**, flows that shrink the stock. In example 1.6, we are looking at an **inflow**, a flow that increases the stock.

For most stocks, we can find both outflows an inflows associated to them.

Check your Understanding

For examples 1.2 and 1.4, list an inflow. For example 1.3, list an outflow.

Now that we've discussed how we can classify measurements, let's look at the term "relationship." In previous math classes, a relationship between two or more quantities has usually been a **function**. For this class, it is no different. When we discuss a relation between a set of measurements, we will be referring to a function.

When we write a function, we are implying what quantity depends on what other quantity. Let us look at an example to illustrate this point.

Example 1.8: The number of dogs a person has[1] is related to the amount of money they spend on dog food per month.[2] If dog food for one dog costs \$50 per month, how can we describe this relationship?

There are two ways we can write the equation of a function relating these two measurements. The first is

$$y = 50x$$

where x is the number of dogs owned, and y is the amount of money spend on dog food per month.

The second possible equation is

$$x = \frac{y}{50}.$$

Of the two equations, which do you think is better? They are both mathematically equivalent, but the first one is more meaningful. By writing "$y =$," we are implying that y depends on x. That is, x causes changes in y. For this situation, that means the number of dogs we have influences the amount of money we spend on dog food per month.

For the second equation, we wrote "$x =$." So we were implying that the number of dogs you have depends on the amount of money you spend on dog food. It is like saying, "When a person buys dog food, a dog magically appears at home." But doesn't work that way. It's the amount of dogs a person has that influences how much the spend. Not the other way around.

To specify that y depends on the value of x, we call y the **dependent variable**. We therefore call x the **independent variable**.

[1]Notice that this is a stock.

[2] Notice that this is a flow, not related to the stock. It's an outflow of the owners bank account!

This distinction isn't mathematical, per se. It is psychological. When we write a function, we are implying a *direction of influence*. In the equation

$$y = 50x,$$

I am implying that the stock of dogs a person has influences the outflow of money in their bank account.

Remark 1.9: A more rigorous understanding of a function is *a map where every individual input produces only one output*. Requiring that relationships between measurements have this property encourages us to make better models.[3]

For example, if 2 dogs meant we could either spend \$40, \$100, or \$120 on dog food a month (that's three outputs from our one input), then the relationship is **not** a function. Physically, it means we probably need to consider more than just the number of dogs. We could consider dog sizes or dog food brands. If we did, we would need more *independent variables*.

Summary of Ideas: Lecture 1

- For this class, we will think of mathematics as the attempt to characterize a relationship between two or more measurements.
- Measurements are either stocks or flows.
- A **stock** is an an amount of something existing at a particular point in time.
- A **flow** is a measurement of how a stock changes over time.
- A flow can describe an **inflow**, one that grows a stock, or an **outflow**, one that shrinks a stock.
- For this class, functions will describe how two measurements are related to each other.
- One variable will always be described as depending on the other variable(s). We call this variable **dependent**. The remaining variable(s) are called **independent**.

[3] A model is a description of some kind of physical system. The functions that we use are "models," because they help us predict reality.

Objectives

- ☐ I understand the distinction between independent variable(s) and the corresponding dependent variable as well as why that distinction was chosen for the situation.
- ☐ I can define a linear equation, recognize when two measurements may have a linear relationship, and (with a calculator) I can find the best fit line to model that relationship.

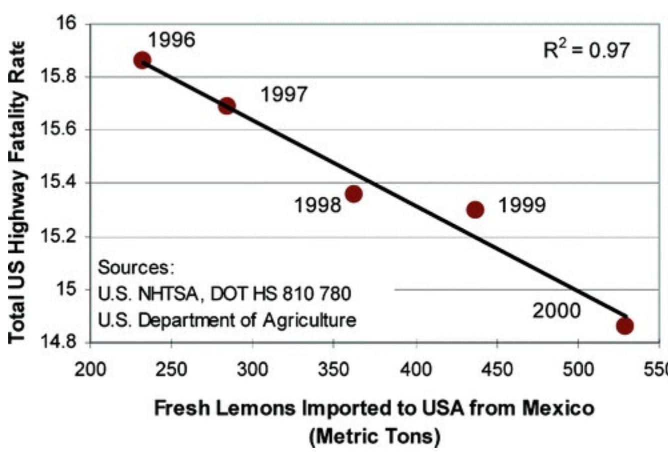
Last time, we defined how we will approach mathematics—namely, as a subject that attempts to describe and understand relationships between two or more measurements. We then discussed the two kinds of measurements: stocks and flows. In particular, flows can come in two forms. They can be inflows, rates that grow a stock, or outflows, rates that shrink a stock.

We describe and understand the relationships between measurements using functions. When we write a function, however, we are implying that some measurement (or set of measurements) is influencing one in particular. The measurement being influenced is called dependent. All other variables are called independent.

We now pick up where we left off by pointing out that although a function will indicate a direction of influence, that influence is not always *causal*.

Sometimes a causal relationship cannot be determined either because we do not know enough or it simply is not there. For example, there is a correlation between the average temperature of a day and the number of violent incidents a city experiences. Since we know violence between people will not change the temperature of a city, we treat violence as the dependent variable and temperature as the independent variable. While the temperature *influences* the number of violent incidents, we cannot say it *causes* them because we do not know a mechanism for it. It is possible that temperature causes some third factor (like encouraging people to go outside) which has a causal relationship with violence.

The image below is a more humorous correlation where causation is unlikely.



When no causal relationship can be justified, we have an obligation to write something along the lines of this: *While our function describes a relationship between two or more measurements,* *this relationship may not be causal or even meaningful.* If such a problem were to arise in the homework or exam, I expect you to write a statement to this effect.

2.1. **Determining the Dependent Variable.** If we don't know a causal mechanism, how can we determine which variable is the dependent variable? The answer is: look at the context. In many cases, there is one variable which does not have the capacity to influence the others. Below are some examples.

Example 2.1: Suppose you want to relate the amount of time a student spends drinking during the semester with their GPA at the end of the semester. Their GPA must be the dependent variable because it comes *after* the drinking was done. Since no one can go back in time, a student's GPA can't influence their amount of drinking.

Example 2.2: There exists a correlation between a person's age and their likelihood of dying by suicide. Suicide risk can, in no way, influence age. Therefore, your age must influence suicide risk.

In some cases, however, no variable is clearly the dependent variable. For example, homeless rates and crime are often correlated, but neither one is clearly dependent on the other. In those cases, your choice of dependent variable will depend on the question you are asking. For example, if you ask, "How will reducing homelessness impact crime rates?" then you will treat crime rates as dependent on homelessness.

2.2. **Linear Relationship.** If two measurements have a **linear relationship**, that means data plot (the graph of the data) is described by approximated by line. A somewhat imprecise way of determining a linear relationship is to graph the data. If the graph appears to cluster in the pattern of a line, then the measurements may have a linear relationship. We now go over a historical example to explain how to graph these points with a calculator.

During summer evenings, it is common to hear the crickets chirping. In the late 1800s, physicist and inventor Amos Dolbear conjectured that there was a relationship between the frequency of cricket chirps (the number of chirps per second) and the ambient temperature. He collected data like the chart below.

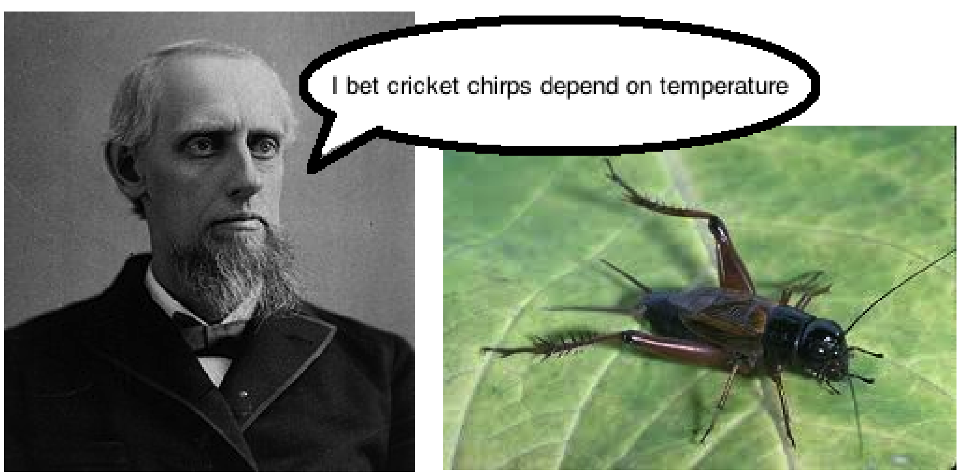


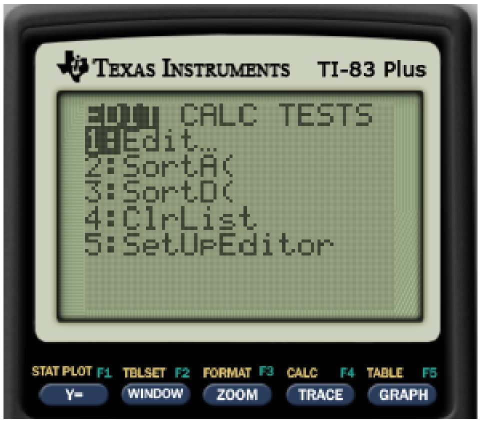
Table 2.3: Temperature versus Chirping Frequency

Temperature	Chirps per Second
20.0	88.6
16.0	71.6
19.8	93.3
18.4	84.3
17.1	80.6
15.5	75.2
14.7	69.7
15.7	71.6
15.4	69.4
16.3	83.3
15.0	79.6
17.2	82.6
16.0	80.6
17.0	83.5
14.4	76.3

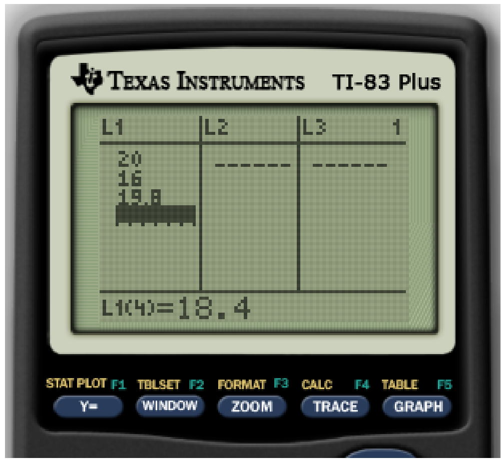
First, we need to determine which is the dependent variable. It is more sensible that the chirps should depend on temperature rather than temperature depending on the chirps (chirps don't generate a significant amount of heat). Therefore,

independent variable = time
dependent variable = chirps per second

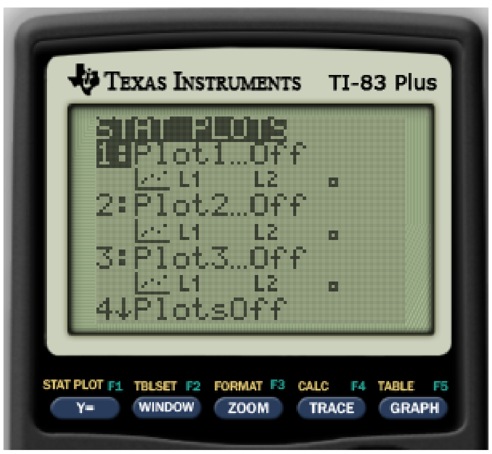
To input the above data in a TI-83 or TI-84, click `STAT` and select the `EDIT` menu at the top. Then select the first entry `1:Edit...` by pressing `ENTER`.



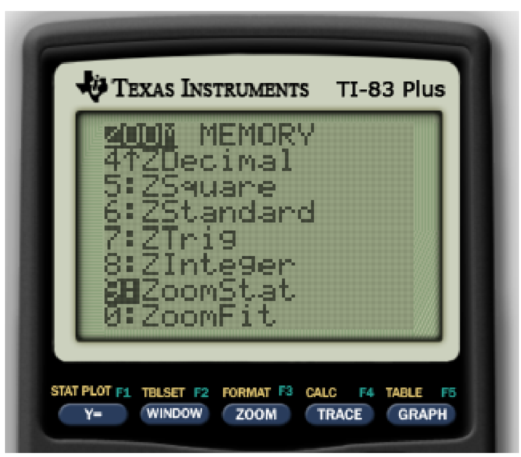
Now use the arrows to navigate the through the columns. In the `L1` column, type in the entries of the independent variable. In the `L2` column, type in the entries of the dependent variable. Once you're done, you can hit `2nd` then `MODE` to `QUIT`.



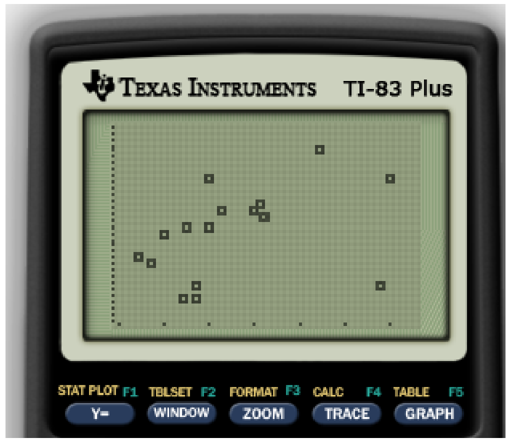
Now we would like to graph the data. To this, hit `2nd` then `Y=` to access the `STAT PLOT` menu. Select the first menu and highlight the following settings according to the picture below. Once you're done, you can hit `2nd` then `MODE` to `QUIT`.



Before you hit `GRAPH`, I suggest you hit `ZOOM` and select `9:ZoomStat` to adjust.

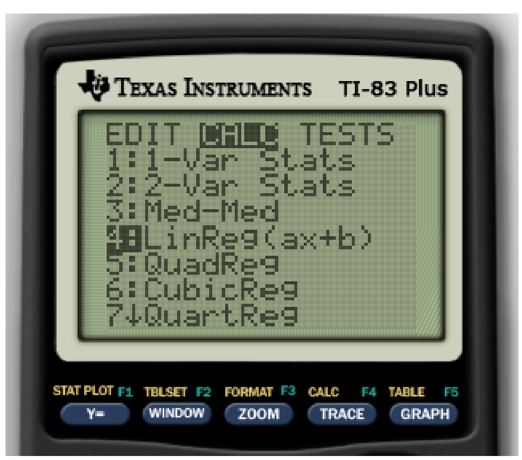


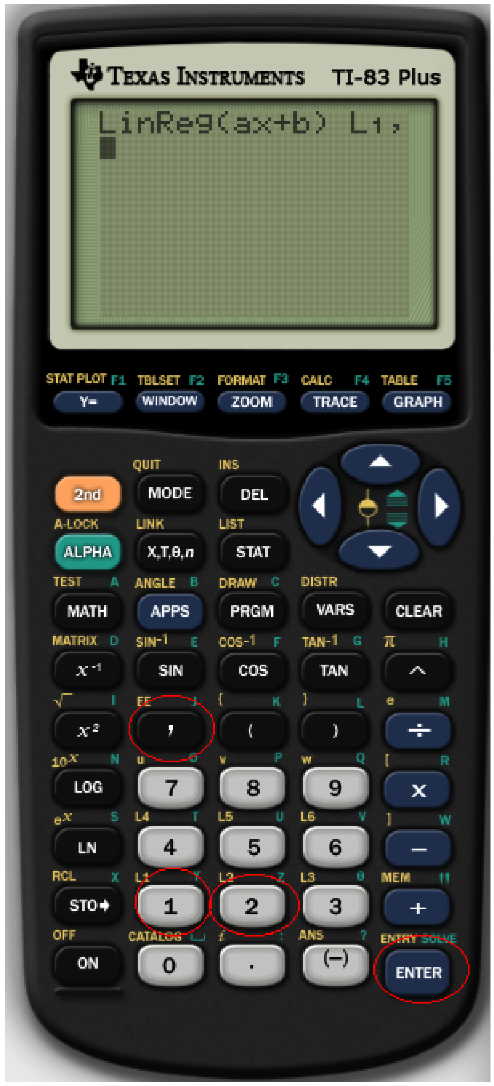
Below is the resulting scatter plot once you hit `ENTER`.



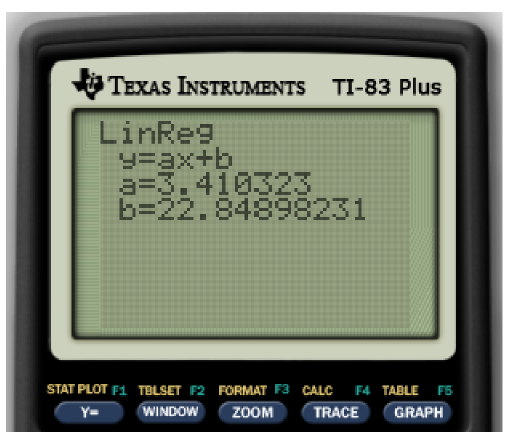
The picture appears to have a linear relationship. If we suspect a linear relationship, then the next step is to define the **best fit line** to describe the relationship. For this class, we simply covered how to use a calculator to find this line. A little later in the course, we will actually go over the mathematical process of finding the line.

To find a line, press `STAT` and select the `CALC` menu at the top. Then select the first entry `4:LinReg(ax+b)`. Once you do, it will display `LinReg(ax+b)` with the cursor blinking. Press `2nd` then `1` to print L_1. Then press `,` and press `2nd` then `2` to print L_2.



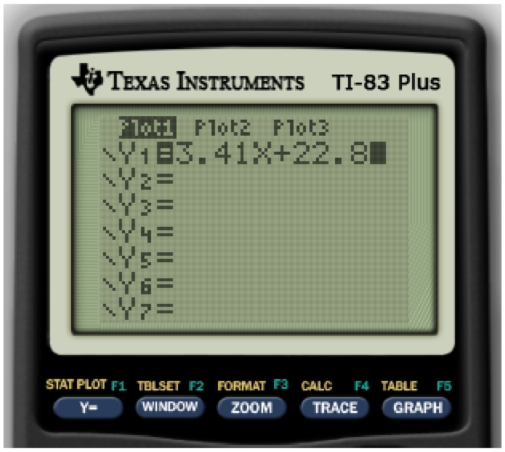


Hit $\boxed{\texttt{ENTER}}$. You should then get a print out of the line that best fits the data.

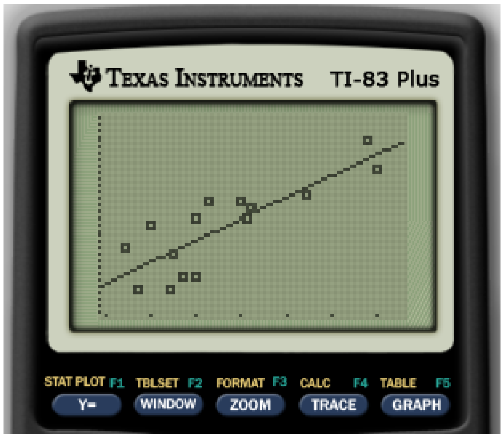


Our data went only up to the 10th decimal place, so we can treat this function as $y = 3.41x + 22.8$, where x is the temperature and y is chirps per second.

To compare the data against the line, we can press $\boxed{\texttt{Y=}}$ and type the function of the line. To type "x", press the $\boxed{\texttt{X,T,}\Theta\texttt{,n}}$ key.



Press $\boxed{\texttt{GRAPH}}$ and see how it compares.



2.2.1. *Interpretation of our Findings.* If we look at the picture, we see a line that appears to explain a significant relationship between temperature and chirping. If we take the derivative of the equation, we get

$$\frac{dy}{dx} = 3.41.$$

Remember from Math 110 (or your previous calculus course) that a derivative explains the amount y changes in relation to x at a particular point. Getting the derivative 3.41 tells us that for every 1 degree change in temperature (x), we should see an increase of 3.41 chirps per second (y).

The next question to ask is this: Is this relationship consistent with the data? That is, should we believe a real-life relationship exists between these two measurements. In general, this is a difficult question to answer. For now, we will determine this visually by comparing if the data appear to follow the line well. When we graphed the line with the data in the calculator, we could see a linear trend (if we ignored the point in the bottom right corner). The more the data agrees with the line we see, the more we should believe this relationship exists in real life. The less it agrees, the less we should believe these two measurements share a relationship.

We will make this ideal more concrete later on.

Summary of Ideas: Lecture 2

- Functions imply a causal relationship. Sometimes that is true and sometimes it is not. When we are not sure or when we believe it is not true, we have an obligation to point that out in our work.
- When modeling a relationship between two measurements, we treat one variable as depending on the other. We determine this based on the context or the question we ask.
- We can plot data with the help of a calculator. If that data appears to grow in the shape of a line, we suspect two the measurements have a linear relationship.
- We can find that line with the help of a calculator. We can then compare it with the actual data.
- We can always find a line regardless of the data. We should only suspect a relationship exists between two measurements if the data appears to follow the line well.

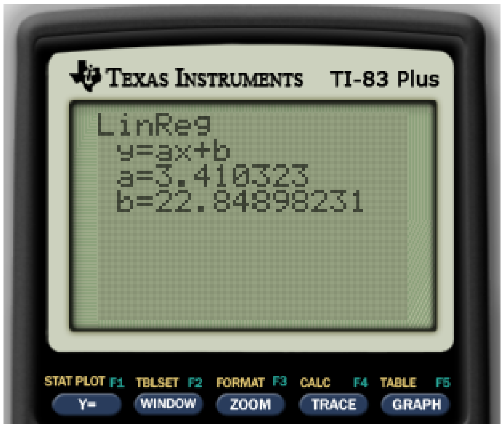
3. LECTURE 3

No new material was covered.

Objectives

- ☐ I can determine the precision of a measurement and reflect that precision in my work.
- ☐ I understand that not all data is ideal for determining a relationship. When possible, I can find better data.
- ☐ By visual inspection, I can isolate an outlier and justify its removal.
- ☐ I understand that every time I use data to determine a relationship, I am making a set of assumptions.

4.1. **Precision.** In the last section, we looked at the frequency of chirps per second versus temperature. Both the temperature and the chirp frequency had no more than three digits; however, when we calculated our best fit line, the calculator displayed numbers containing many digits.



Do we need that many digits? No, we don't need them. More importantly, they communicate a level of precision that we do not have based on out data.

Precision reflects the extent to which our data is accurate. In the case of temperature, we were given data that was three digits, like 17.2° C. This means that the equipment used could only measure the accuracy of the temperature up until the tenths place. We say that this measurement has 3 significant figures. If our precision allows for three significant figures but the final digit is zero, we keep that digit there to reflect the precision. For example, we could have the measurement 20.0°. As it is written, it still has 3 significant figures.

When we find a function describing how two measurements are related, we will choose coefficients to reflect that precision. Therefore, we write

$$y = 3.41x + 22.8$$

to keep our three significant figures.

If we have two measurements with a different number of significant figures, we will always choose the *smaller* of the two.

Check your Understanding

The figures $ $21,000.10$ and $45°$F have how many significant figures? If there was a function related these two measurements, how many digits should our coefficients have?

4.2. **Improving Your Data.** Not all data is created equal. In much of your future work, you will likely have access to *lots* of data, but much of it will be useless to you. We therefore need to spend some time discussing what constitutes good data. All the data presented here can be found at `https://migbirdapps.fws.gov`, a website managed by the US Fish & Wildlife Service.

Let's suppose you want to know if the frequency of mourning dove calls is affected by temperature. Why would you want to know this? Much of population estimation depends on indirect measurements like samples of animals seen or heard. If temperature impacts how much an animal is heard, we may overestimate or underestimate their population based on our observations. So understand how mourning dove calls are influenced by temperature can help us better predict their populations from sound.

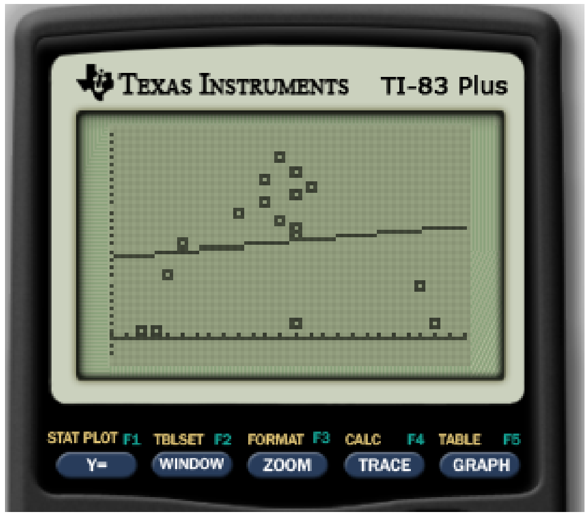
For a small project like this, you'll need to work off the data available. First, you decide to look at the temperature versus bird calls heard throughout Pennsylvania on selected days in May 2010. Below is the data.

Table 4.1:

Temperature (F)	Mourning Dove Call-Count
60	14
60	18
50	1
52	12
58	17
59	23
70	2
60	2
51	8
61	19
60	13
59	15
69	7
58	20
49	1
56	16
60	21

FIGURE 1. This data is taken from the US Fish & Wildlife Service, Division of Migratory Bird Management. The samples listed were taken throughout PA in May 2010.

When you graph it, you see no obvious relationship, especially when compared with the best fit line.



Does that mean no relationship exists in real life? Not necessarily. Although the data suggests there is no relationship, we need to consider whether this data is appropriate for the question we are asking.

We want to know whether temperature impacts mourning dove vocalizations. The data is a collection of observed mourning dove calls in various regions throughout Pennsylvania. These observations were taken in each area on one day in May 2010. What potential problems exist with using data like this?

(1) Different regions will likely have different population sizes of mourning doves.
(2) These differences in population may have a stronger impact on dove vocalizations than temperature, which would hide the relationship in the data.
(3) If we only look at measurements in the same month, we may not observe dramatic changes in temperature. That makes it harder to detect the relationship.

Ultimately, the data we used was not ideal for the problem. We'd like to use data more appropriate to our question. Below is data also taken from the US Fish and Wildlife Service. All the observations were taken in one park region in Chester County, PA. The measurements were taken on one day in May from 2001 to 2010.

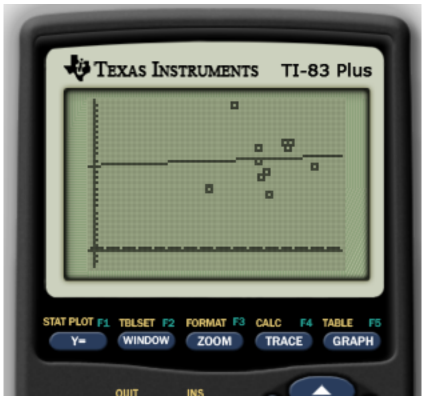
This data lacks some of the problems of the previous one. We are restricting our observation to the same park land and the measurements are taken across many years so we can get a variety of temperatures. The data still is not ideal. We do not know if the population of mourning doves remained constant during those years in that park. We also do not have many samples. The smaller your sample size, the more likely you are to see a pattern that is not there.

Still, we can do some analysis on the data we have to determine if there is a relationship. Below is a print out of the data along with a best fit line.

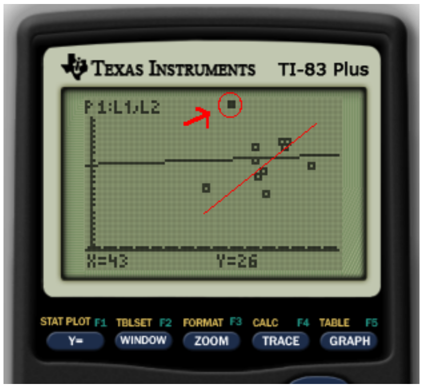
Table 4.2:

Temperature (F)	Mourning Dove Call-Count
54	10
59	19
52	13
68	15
51	18
36	11
51	16
43	26
53	14
61	19
60	18

FIGURE 2. This data is taken from the US Fish & Wildlife Service, Division of Migratory Bird Management. The samples listed were taken from Chester County, PA in May from 2001 to 2010.



Upon visual inspection, the data appears to have a nice linear relationship. Unfortunately, there appears to be one point, $(43, 26)$, which is very far from the rest of the data. Furthermore, the best fit line does not match nicely with the data we observe. In the image below, the point $(43, 26)$ is circled and a red line is drawn to highlight the trend of the rest of the data.



Points that are considered too far from the data set are called **outliers**. There are a variety of possible reasons for an outlier to occur:

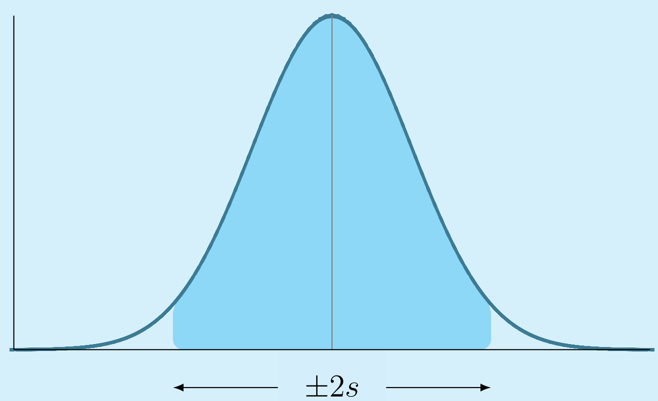
(1) **Human Error**: Those doing the measuring—in this case, counting the bird calls heard—can make mistakes. It is possible that the count of 26 bird calls was inaccurate.
(2) **Equipment Error**: Sometimes equipment is faulty. In data where each sample could be taken using different equipment (like measurements that consist of one sample per year), it is possible that a mismeasurement resulted from a technical issue.
(3) **A Fluke**: The measurement could be accurate but something unusual could be happening in the physical situation that could skew the data, like having someone nearby feeding the doves.
(4) **Nature**: Finally, it could be an accurate measurement with no weird circumstances. It is possible that the outlier really reflects what can happen in certain instances.

Because of that last reason, we need to be very careful when we consider removing a data point. If we can't definitively point to circumstances that call the measurement into question, then we run the risk of coming to inaccurate conclusions. That is, if we don't know the data point is bad, we should be careful about removing it and possibly seeing a pattern that is not there.

In this case, we did not collect the data nor do we know the circumstances under which it was collect. So if we remove the outlier, we run some risk. To reduce our risk, we will define a mathematical procedure to determine if an outlier should be removed.

Connecting Back to Past Content

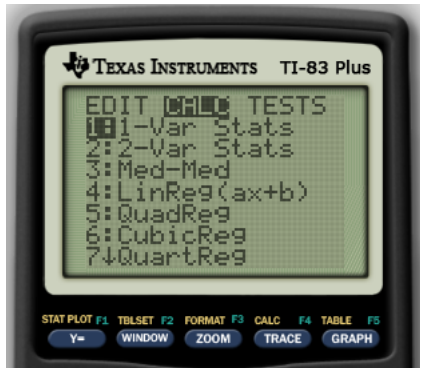
If you have taken statistics, then you understand that all measurements are just samples of a larger distribution. We will use this underlying fact to try to reduce our risk of throwing away good data to 5%.
The following procedure will require we find the mean, μ, and the sample standard deviation, s, of the dependent variable. Then we consider any values of the dependent variable smaller than $\mu - 2s$ or larger than $\mu + 2s$ to be outliers. If most of the data is reliable, the points removed are less than 5% likely to happen, meaning these points are more likely to come from an error.



Note that if most of the data is bad, this technique will not work.

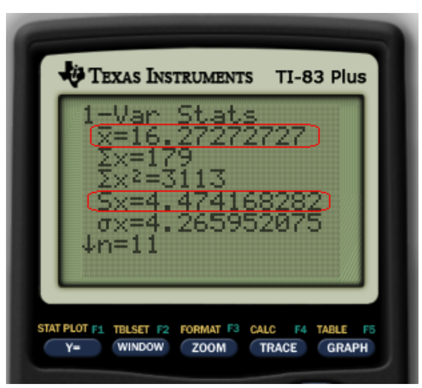
4.2.1. *Determining Outliers.* To determine outliers, we will calculate the average (or **mean**) of the dependent variable's measurements and the **sample standard deviation** (something that measures the width of the data's spread).

To calculate the mean (μ) and the sample standard deviation (s) using a TI-83 or TI-84, click `STAT` and select the `CALC` menu at the top. Then select the first entry `1:` `1-Var Stats` by pressing `ENTER`.



You will now see the phrase `1-Var Stats` on the home screen of the calculator. Before hitting enter, type the list of the *dependent* variable after. Here, our dependent variable is list L_2. Then press `ENTER`.

You'll see a display with a set of statistics. The ones we need are the mean, denoted as $\bar{x}$, and the sample standard deviation, denoted Sx. Although the calculator uses different symbols, we will refer to the mean as μ and the sample standard deviation as s.

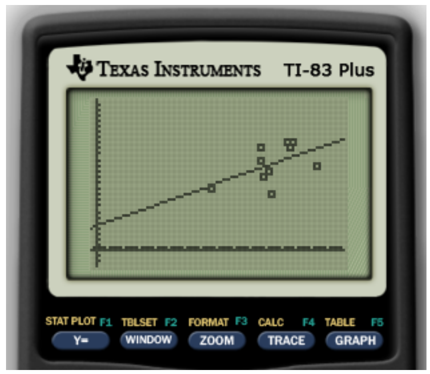


The print out above is for the data regarding dove calls in Chester County, PA. Based on the data, we will consider a point an outlier if the dependent variable is

- *smaller* than $\mu - 2s = 16 - (2 \times 4.5) = 7$, or
- *larger* than $\mu + 2s = 16 + (2 \times 4.5) = 23$.

The point we suspected to be an outlier is (43, 26). Notice that the dependent variable's value for this point is 26, which is bigger than 23! Therefore, we can remove this point with small risk.

Once we remove this point, we get a line that appears to better explain the relationship between the two measurements.



4.3. **Assumptions.** When making conclusions from raw data, it is important that we can identify our underlying assumptions. This is a list of what we need to be true in order for our findings to be meaningful. The precise statements will change from situation to situation, so here we will list the assumption for the problem above (dove calls versus temperature).

Assumptions

(1) The data is reliable; it was measured by attentive researchers with working equipment.
(2) The data was collected in roughly the same location every year.
(3) The population of mourning doves did not change significantly from year to year.
(4) The data set is large enough to define a relationship well.
(5) The data follows a "normal distribution" (a statistical assumption that means our method of selecting outliers is a good one).

Assumptions 1, 4 and 5 are needed in every problem involving raw data.

Summary of Ideas: Lecture 4

- Calculations should represent the precision of the measurements taken by keeping the same number of significant figures.
- *How* data is collected matters.
- Data can contain extreme values called **outliers**. Outliers can come about from an error or from nature.
- Any value that is **larger** than the $\mu + 2 \times s$ or **smaller** than $\mu - 2s$ is considered to be an **outlier**. They are often removed to perform calculations better.
- Every time we draw information from raw data, we are employing assumptions including that the data can be relied upon and that the data set is sufficiently large.

5. LECTURE 5

Objectives

- ☐ For two or more linear equations, I can find that solution (if one exists).
- ☐ I can determine the meaning of a solution to a system of linear equations.
- ☐ I can rewrite a system of linear equations as a linear expression with vectors and matrices.
- ☐ I can solve matrix equations with the help of my calculator.

Up until now, we have looked at finding linear relationships between two measurements from a set of collected data. Now, we will build the theory we need to understand *how* our calculator finds this line. This technique will be very useful when we look at nonlinear relationships (relationships that are described by something other than a line) as well as related systems where no measurement is clearly dependent like predator and prey populations.

5.1. **Systems of Linear Equations.** A **system of linear equations** is a set of lines in terms of the same variables. Physically, it means we have a set of relationships between the same measurements in different situations. The goal in these problems is always the same: find values of x and y that satisfy *all* the equations in the set.

Example 5.1: A new teacher in State College decides to form an archery club for middle schoolers and high schoolers. In a news article, it explains that middle schoolers begin with a 24 pound recursive bow (this is the draw weight of the shot) while high schoolers begin with a 45 pound recursive bow. The article states that 112 students are involved in the archery club and that Dick's Sporting Goods donated all the bows, which are equal in value to $22,700. You find out online that 24 pound recursive bows are $140 each and 45 pound recursive bows are $270 each. How many middle schoolers and high schoolers are involved in the clubs?

Solution 5.2: The first step to any word problem is to identify the measurements we do not know but wish to find.

There are two unknowns: the number of middle schoolers in the club and the number of high schoolers in the club. We'll use variables x and y for each one, respectively.

What information to we have about these two groups? We know they total 112 because that is the size of the club. Therefore,

$$x + y = 112.$$

The other piece of information we know is that the total number of bows costs $22,700. We know the high schoolers' bows were $270 each while the middle schoolers' bows were $140 each. Therefore,

$$140x + 270y = 22700$$

since $140x$ will equal the *total* amount of money spent on middle schooler bows and $270y$ will equal the *total* amount of high schooler bows.

Notice that we don't have a use for the figures "24 pound" or "45 pound." In this context, these numbers tell us the amount of force needed to use the bows, so they can't help us determine *how many* middle schoolers or high schoolers in the club.

Now that we've exhausted all the numbers listed, we will take our equations and solve for x and y. There are 2 methods to do this:

- **Substitution:** This method requires using one equation to solve for a variable and then plugging that expression into the other equation. Suppose we solve for y in the first equation. Then we get $y = 112 - x$. This expression is substituted into the second equation:

$$\begin{aligned} 140x + 270y &= 22700 \\ 140x + 270(112 - x) &= 22700 \\ 140x + 30240 - 270x &= 22700 \\ -130x + 30240 &= 22700 \end{aligned}$$

  After simplifying, we get the expression $-130x + 30240 = 22700$. When we solve for x in this equation, we get $x = 58$. There are a number of ways I can find y, but the easiest is to use the first equation we wrote: $y = 112 - x$. When we plug in 58 for x, we get $y = 112 - 58 = 54$.

  Hence, there were 58 middle schoolers and 54 high schoolers.

- **Elimination:** This method requires that you combine the equations so as to eliminate one of the variables. Here, we multiply the entire equation $x+y = 112$ by 140 and then subtract the two equations to eliminate x.

$$\begin{array}{rl} & 140(x + y = 112) \\ - & 140x + 270y = 22700 \\ \hline \end{array} \implies \begin{array}{rl} & 140x + 140y = 15680 \\ - & 140x + 270y = 22700 \\ \hline & -130y = -7020 \end{array}$$

  Now we can solve for y using the equation $-130y = -7020$. When we do, we get $y = 54$. We then plug 54 for y in one of the previous equations. The easiest is $x + 54 + 112$, which gives us $x = 58$.

  Hence, we get the same answer as we do with substitution.

Remark 5.3: There is no significant difference between using elimination and substitution. For problems with many equations, elimination can be faster, but substitution generally proves to be easier for humans. In the second half of these notes, we will learn how to solve these equations with matrices.

Remark 5.4: Notice that when we write equations like $x + y = 112$, we are treating x and y as both independent variables! We technically have a third *dependent* variable z, the population of the entire club, which depends on x and y. We don't explicitly use this variable because we are only concerned with one value: $z = 112$. Later on in the course, we will look at equations like $x + y = z$ and consider many possible solutions.

Example 5.5: Suppose you find an injured Allegheny woodrat, a threatened species in Pennsylvania. You decide to nurse it back to health and then releasing it into the wild. Suppose an ideal

Allegheny woodrat diet contains 12% protein, 4.5% fat, 15% fiber. At the corner store, you have the following options:

- canned dog food: 10% protein, 7.0% fat, 3.0% fiber
- hamster food: 12% protein, 5.0% fat, 18% fiber
- rabbit food: 13 % protein, 3.5% fat, 19% fiber
- (domesticated) rat food: 15% protein, 4.0% fat, 7.0% fiber

What mixture of dog food, hamster food, and rabbit food will be appropriate for the woodrat? Find the amounts for one cup (a week's worth of food).

Solution 5.6:(Partial) First, we need to determine the variables. These are the measurements we can adjust. Since we cannot adjust the percentage of protein in each food, our variables must be the amounts of each food type. Let x be the amount of dog food we will use in the mixture, y the amount of hamster food used, z the amount of rabbit food and w the amount of domesticated rat food used.

The mixture will include dog food (x), hamster food (y), rabbit food (z), and domesticated rat food (w) and needs to add to one cup. Hence

$$x + y + z + w = 1.$$

We now write equations for protein, fat, and fiber.

$$\begin{aligned}
&\text{Protein: } .10x + .12y + .13z + .15w = .12\\
&\text{Fat: } \quad\ .070x + .050y + .035z + .040w = .045\\
&\text{Fiber: } .030x + .18y + .19z + .070w = .15
\end{aligned}$$

From here, we can determine the values of x, y, z and w using substitution or elimination. As you can see, however, it will get complicated because we have four variables and four equations.

5.2. **Matrix Equations.** Matrices are a way to summarize a system of equations. In the previous example, we have four equations that used the same variables:

$$\begin{aligned}
&x + y + z + w = 1\\
&.10x + .12y + .13z + .15w = .12\\
&.070x + .050y + .035z + .040w = .045\\
&.030x + .18y + .19z + .070w = .15
\end{aligned}$$

A matrix is a kind of table that keeps track of the coefficients in a system of linear equations. The first row corresponds to the coefficients of the first equation and each column corresponds to only one variable. So the above set of equations can be written as

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ .1 & .12 & .13 & .15 \\ .070 & .050 & .035 & .040 \\ .030 & .18 & .19 & .070 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ .12 \\ .045 \\ .15 \end{pmatrix}$$

This arrangement may not seem revolutionary. All we did was:

- take the coefficients and put them in a table, called a **matrix**
- take the variables and put them in a column, called a **vector**, and
- take the solutions from the problem and put them in a column, also a **vector**.

On the surface, it doesn't seem like a big deal. These matrices and vectors, however, can be treated as numbers *and* as functions; this opens up a lot of useful math that makes up this course. For now, we will manipulate these objects using our calculator.

Before we move on to how to techniques with a calculator, we should mention how matrices add and multiply.

5.3. **Facts About Matrices & Vectors.** Before we move on to manipulating matrices with the help of a calculator, let's first talk about some basic facts about these objects. After all, we will be using them later on.

Matrix Size 5.7

Matrices can come in any size.

The following are all different examples of matrices:

$$\begin{pmatrix} 1 & 2 \\ \pi & 0 \end{pmatrix} \qquad \begin{pmatrix} 4.325 & 97000 & e \\ 0 & 24 & \frac{3\pi}{4} \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 1 & 3 \\ 2 & 0 & \frac{3\pi}{4} & 7 \\ 1 & 1 & 0 & -1 \\ -4 & -4 & 3 & 10 \\ 9 & 0 & -\pi & 5,280 \\ 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad (4)$$

We see from the above example that matrices don't have to be squares. They can be longer or taller rectangular tables. The last example shows us that "normal numbers" like 4 are, in fact, just a special kind of matrix.

Vector Size 5.8

Vectors are columns of numbers with no limit in length. Vectors are a special kind of matrix.

Below are examples of vectors:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 4.325 \\ \pi \\ 5280 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 2 \\ 1 \\ -4 \\ 9 \\ 1 \\ 0 \end{pmatrix} \tag{4}$$

We can make two interesting observations about vectors. The first is that any matrix is just a bunch of vectors put together. The second is that "normal numbers," like 4, are a special kind of vector.

Adding & Subtracting Matrices and Vectors 5.9

We can only add or subtract matrices/vectors of the same size. When we add/subtract them, we do so entry-wise.

For example, when we add the two matrices below, we are just adding the entries of each corresponding part in the table.

$$\begin{pmatrix} 2 & 3 \\ 1 & 0 \\ 4 & -1 \\ -2 & -4 \end{pmatrix} + \begin{pmatrix} 1 & 4 \\ -2 & 12 \\ 10 & 8 \\ 3 & 9 \end{pmatrix} = \begin{pmatrix} 2+1 & 3+4 \\ 1+(-2) & 0+12 \\ 4+10 & -1+8 \\ -2+3 & -4+9 \end{pmatrix} = \begin{pmatrix} 3 & 7 \\ -1 & 12 \\ 14 & 7 \\ 1 & 5 \end{pmatrix}$$

Similarly, when we subtract two matrices, we subtract entry-wise.

$$\begin{pmatrix} 2 \\ 4 \\ 3 \\ 0 \\ -2 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ -2 \\ -5 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 2-1 \\ 4-0 \\ 3-(-5) \\ 0-2 \\ -2-(-3) \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 8 \\ -2 \\ 1 \end{pmatrix}$$

Multiplying Matrices 5.10

* To multiply two matrices, we take entries in the row of the first matrix and multiply them by the entries in the column of the second matrix and add up those products.
* For this to work, we need the number of columns of the first matrix to equal the number of rows of the second matrix.

As you can see, multiplying matrices can be a little confusing when you first see them. When in doubt, always go back to our original set-up: a system of linear equations.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ .1 & .12 & .13 & .15 \\ .070 & .050 & .035 & .040 \\ .030 & .18 & .19 & .070 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$$

When we multiply the entries of the row by the entries of the column and sum them, we end up with our linear equations. When we multiply, we get the sum of the coefficients times the corresponding variable.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.10 & 0.12 & 0.13 & 0.15 \\ 0.070 & 0.050 & 0.035 & 0.040 \\ 0.030 & 0.18 & 0.19 & 0.070 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} x+y+z+w \\ 0.10x+0.12y+0.13z+0.15w \\ 0.070x+0.050y+0.035z+0.040w \\ 0.030x+0.18y+0.19z+0.070w \end{pmatrix}$$

To see this process written out step-by-step, you can visit `https://www.youtube.com/watch?v=aKhhYguY0DQ`[4]

Check your Understanding

Explain why the only vectors you can multiply are just those that are 1 entry (i.e. "normal" number).

A natural next question you might ask is: How do I divide two matrices? In a sense, we don't divide them. We multiply by the inverse matrix (which is just like division). In this course, we will not worry about how to find a matrix's inverse by hand. We will instead do this with a calculator.

5.4. **Matrix Equations on the Calculator.** The goal of this section is find a solution to a system of equations like this.

[4] The speaker says these are "human-created definitions" for matrix multiplication. If you decide to study math further, you'll learn that in some sense, there is no other way matrix multiplication could work. Matrices are a very special way to represent many mathematical objects. So his statement is debatable.

$$
\begin{aligned}
&x+y+z+w=1\\
&.10x+.12y+.13z+.15w=.12\\
&.070x+.050y+.035z+.040w=.045\\
&.030x+.18y+.19z+.070w=.15
\end{aligned}
$$

That is, we find numerical values for each variable x, y, z and w so that when we plug those numbers into the equations above, we get the solutions 1, .12, .045, and .15 (respectively).

Our technique uses matrix equations. We turn this system of linear equations into the following equation:

$$
\begin{pmatrix} 1 & 1 & 1 & 1 \\ .1 & .12 & .13 & .15 \\ .070 & .050 & .035 & .040 \\ .030 & .18 & .19 & .070 \end{pmatrix}
\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}
=
\begin{pmatrix} 1 \\ .12 \\ .045 \\ .15 \end{pmatrix}
$$

When mathematicians read the above statement, they read it as

$$A\vec{\gamma}=\vec{b}$$

where

$$\vec{\gamma}=\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \qquad \text{and} \qquad \vec{b}=\begin{pmatrix} 1 \\ .12 \\ .045 \\ .15 \end{pmatrix}.$$

Solving the equation $A\vec{\gamma}=\vec{b}$ amounts to “dividing” both sides by A. For matrices, that’s multiplying by the inverse:

$$A^{-1}A\vec{\gamma}=A^{-1}\vec{b}$$

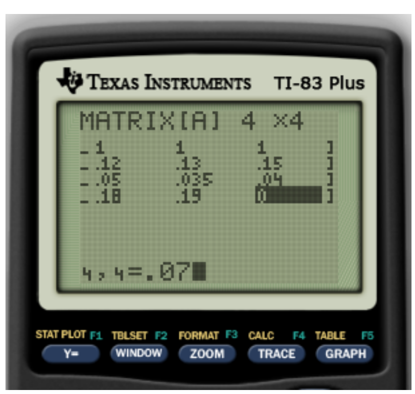
When we multiply a number (like 2) with its inverse ($\frac{1}{2}$), we always get 1. For matrices, it is the same, except ”1” is a special kind of square matrix with 1s along the diagonal entries and zeros everywhere else. For now, though, we will treat this like we do with any other algebra problem. If it becomes 1, it does not need to be written. So we get the equation:

$$\vec{\gamma}=A^{-1}\vec{b}$$

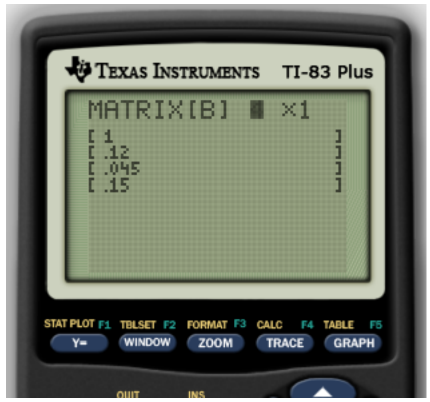
This is what we calculate. So we will find $A^{-1}\vec{b}$ using our calculator. The first step is to enter A and $\vec{b}$ into our calculator. We begin by pressing `2nd` and `x`$^{-1}$ to access the `MATRIX` menu. Navigate at the top using the right arrow to the `EDIT` menu. Make sure `[A]` is highlighted and hit enter.

When you enter the matrix menu for `[A]`, you'll need to first enter the dimensions of the matrix. The first number is the number of rows. The second is the number of columns. For A, we need 4 rows and 4 columns. Enter the values that correspond to the same location:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ .1 & .12 & .13 & .15 \\ .070 & .050 & .035 & .040 \\ .030 & .18 & .19 & .070 \end{pmatrix}$$



Once you're done, press 2nd and MODE to quit. Repeat the process to enter the values in $\vec{b}$ by returning to the MATRIX menu, navigating to EDIT and selecting [B]. This will have 4 rows and 1 column.

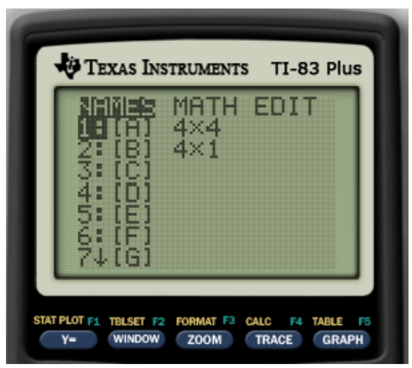


Once you're done, press 2nd and MODE to quit. The calculator now knows the matrix A and the vector $\vec{b}$. What is left is to compute

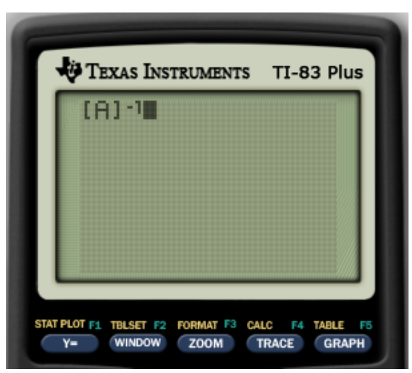
$$A^{-1}\vec{b}$$

to find the solution to the system of linear equations.

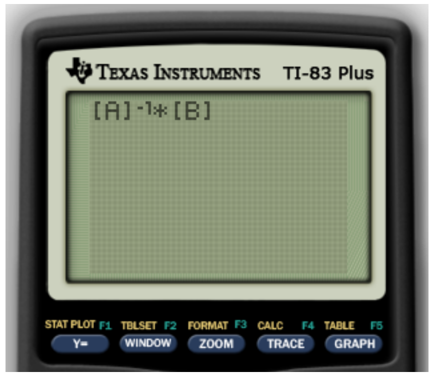
To do this, we will return to the MATRIX menu by pressing 2nd and x^{-1}. This time, we stay in the NAMES menu and select 1: [A] 4 x 4.



Then [A] will appear on the home screen. Press x^{-1} to display $[A]^{-1}$.



Press the button for multiplication and return to the MATRIX menu, selecting [B] under NAMES. Once you do, your home screen should look like the image below.



When you press enter, you'll see a display of a vector for the values of $\vec{\gamma}$. Here is what we get:

That means

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0.29 \\ 0 \\ 0.74 \\ -0.05 \end{pmatrix}$$

Going back to the original problem, our solution suggest that the ideal mixture for the rat is .29 cups of dog food, 0 cups of hamster food, .74 cups of rabbit food, and -.05 cups domesticated rat food. This last figure, as we know, is not possible. There is no way we can subtract rat food from a mixture of dog food and rabbit food.

So what went wrong? We found a solution but it isn't realistic. That's because no where in our technique did we include the restriction that x, y, z, and w be positive. Once we do, we're looking at an *optimization* problem. That's where we want to find the "best" solution with some set of physical limitations. The main point of this class will be tackling these kinds of problems.

For now, however, we will look at simple systems of linear equations. Sometimes our solutions will make physical sense, but sometimes they won't.

Summary of Ideas: Lecture 5

- A **system of linear equations** is a set of lines such that the variables used in all the equations represent the same physical measurements.
- There are two methods of solving systems of linear equations by hand: **substitution** and **elimination**.
- Numerically, we solve these problems using matrices.
- A **matrix** is a table of numbers that can be manipulated like a "normal number." A **vector** is a matrix with only one column.
- Two matrices can be added or subtracted only if they have the same dimensions (same number of rows and columns). We add matrices entry-wise.
- Two matrices can be multiplied if the number of columns of the first matrix equals the number of rows of the second one.
- To solve a system of linear equations using matrices, we define a matrix based on the coefficients, A, a vector with all the variables $\vec{\gamma}$, and a vector with the solutions to the linear equations $\vec{b}$. Then we calculate

$$\vec{\gamma} = A^{-1}\vec{b}$$

6. Lecture 6

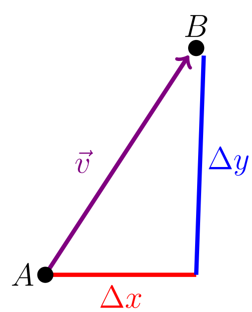
Objectives

- ☐ I understand how to use vectors to understand displacement.
- ☐ I can find the magnitude of a vector.
- ☐ I can sketch a vector.
- ☐ I can add and subtract vector.
- ☐ I can multiply a vector by a scalar.

During the last lecture, we discussed vectors and matrices in the context of systems of linear equations. Now, we will learn a little more about these objects and a different ways of understanding them.

6.1. **Vectors in Physics.** So far, we've learned that vectors are matrices with one column and that each entry is a value for a distinct variable. Because a variable can be any kind of measurement, vectors can represent almost anything.

In physics, vectors are used to indicate a **displacement** of an object from one point A (the initial point) to another point B (the terminal point). Physically, we can represent this as an arrow from A to B. As we saw in the previous section, vectors are typically symbolized by a lower-case letter with an arrow over it like $\vec{v}$. If it represents the displacement from A to B, then we sometimes write it as $\vec{AB}$.



We write this vector as

$$\vec{AB} = \vec{v} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

How is this related to the previous lecture? Here, our first variable represents change in position in the x direction and our second variable represents change in position in the y direction. The *net* movement is represented by the vector itself, represented by the violet arrow.

Notice that this is a right triangle, where the arrow is the hypotenuse (the longest leg). The length of $\vec{v}$, denoted $|\vec{v}|$, will tell us the distance from A to B. It can be found using the *pythagorean theorem*[5]:

$$|\vec{v}| = \sqrt{x^2 + y^2}$$

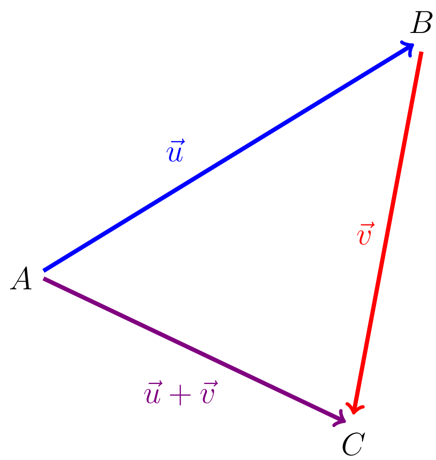
For a generic vector, the "length" of the vector is referred to as **magnitude**.

[5] Recall the identity $a^2 + b^2 = c^2$ from geometry

Check your Understanding

Explain why the zero vector, a vector with only zeros in its entries, should be a point.

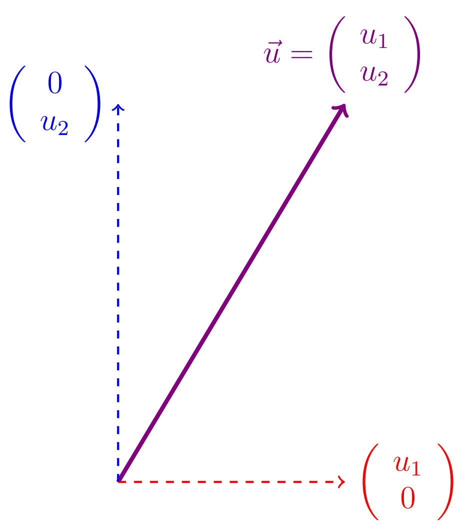
The sum of two vectors can be thought of as a combination of two displacements. For instance, if the vector $\vec{u}$ represents the displacement from point A to point B, and the vector $\vec{v}$ represents the displacement from the point B to point C, then their sum $\vec{u}+\vec{v}$ represents the total displacement from point A to point C. (This is called The Triangle Law.)



We add two vectors component-wise. For $\vec{u}=\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $\vec{v}=\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$,

$$\vec{u}+\vec{v}=\begin{pmatrix} x_1+x_2 \\ y_1+y_2 \end{pmatrix}$$

By this same logic, any vector can be broken up into its horizontal and vertical vectors. That is



Everything works the same way in three dimensions. Below is an example. You can think of the x direction as being forward/backward, the y direction as left/right, and the z direction as up/down.

—

6.1.1. *Examples.* **Example 6.1:** A group of scientists are modeling how bats fly in confined spaces. A large 20ft by 20ft by 20ft room is fitted with synchronized cameras hooked to computers which can extrapolate the position of the bat from the images. The middle of the room is the point $(0,0,0)$. Two photos taken in succession show the bat at point $A(1,3,5)$ and then 10 seconds later at $B(-1,0,7)$.

(1) Define the vector from A to B.
(2) Approximate the speed (in feet per second) at which the bat traveled from A to B.
(3) Sketch the room and the vector.

Solution 6.2:

(1) To measure a total distance traveled, we want to subtract our final destination from our starting point.

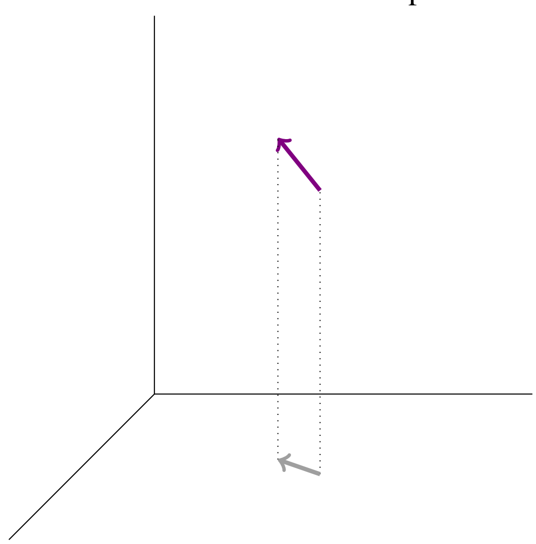
$$\vec{AB} = \begin{pmatrix} -1-1 \\ 0-3 \\ 7-5 \end{pmatrix} = \boxed{\begin{pmatrix} -2 \\ -3 \\ 2 \end{pmatrix}}$$

(2) To determine the speed, we need to know the distance traveled. First, we should remember that all measurements are taken in feet. Next, we need to determine the length of the distance between A and B. This is equal to the *magnitude* of the vector found above. Therefore, we calculate

$$|\vec{AB}| = \sqrt{(-2)^2 + (-3)^2 + (2)^2} = \sqrt{4+9+4} = \sqrt{17} \approx 4 \text{ ft}.$$

The bat traveled approximately 4 feet in 10 seconds or 24 feet per minute.

(3) We want to sketch this movement in 3 dimensions, which isn't easy. It looks roughly like the image below. The shadow and lines are added to help see the depth.



Example 6.3: What is $\vec{v}+\vec{u}$ and $\vec{v}-\vec{u}$ for $\vec{v} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$ and $\vec{u} = \begin{pmatrix} -1 \\ 4 \\ -5 \end{pmatrix}$

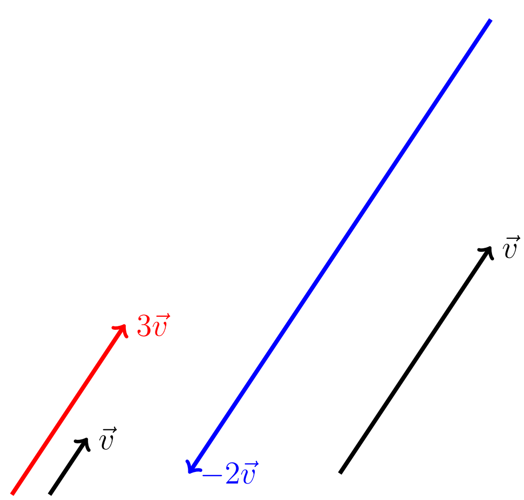
Adding and subtracting vectors works component-wise. Be careful with your signs!

$$\vec{v}+\vec{u} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ 4 \\ -5 \end{pmatrix}$$
$$= \begin{pmatrix} 2+(-1) \\ 3+4 \\ 1+(-5) \end{pmatrix}$$
$$= \boxed{\begin{pmatrix} 1 \\ 7 \\ -4 \end{pmatrix}}$$

$$\vec{v}-\vec{u} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} -1 \\ 4 \\ -5 \end{pmatrix}$$
$$= \begin{pmatrix} 2-(-1) \\ 3-4 \\ 1-(-5) \end{pmatrix}$$
$$= \boxed{\begin{pmatrix} 3 \\ -1 \\ 6 \end{pmatrix}}$$

Constants, or what we've called "normal numbers," are called **scalars**. The magnitude of a vector is a scalar. There is also the notion of multiplying a vector by a scalar. For any vector $\vec{v}$ and any scalar c, the scalar multiple $c\vec{v}$ is obtained by multiplying the length of $\vec{v}$ by c and keeping the

- same direction if $c > 0$, or the
- opposite direction if $c < 0$.



Notice that multiplying by a scalar adjusts the magnitude, but doesn't rotate the arrow (it can only flip it).

In the same way we can focus on a **scalar component** of a vector (its magnitude), we can focus on just the direction of a vector. To do this, we consider a vector that points in the same direction but has length 1. This is called a **unit vector**.

6.1.2. *Examples.* **Example 6.4:** Two bottle rockets are launched. The second is four times as fast as the first. They travel in the same direction. The first travels from $(0, 0, 0)$ to $(1, 1, 2)$ in four seconds. Where does the second end up in four seconds if it begins at $(1, 1, 0)$?

To answer this, we begin by defining the vector of displacement for the first rocket. That is

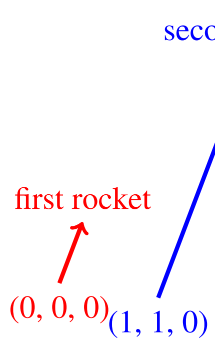
$$\vec{v} = \begin{pmatrix} 1-0 \\ 1-0 \\ 2-0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

If the second moves 4 times as fast as the first, then its vector of displacement just multiply each component by the scalar value.

$$4\vec{v} = \begin{pmatrix} 4 \\ 4 \\ 8 \end{pmatrix}$$

This, however, does not tell us the location of the second rocket. It only tells us how it moves in four seconds. If the rocket begins at $(1, 1, 0)$ then it increase by 4 steps in the x direction, 4 steps in the y direction, and 8 steps in the z direction. Therefore, in four seconds the rocket is at

$$(1+4, 1+4, 0+8) = \boxed{(5, 5, 8)}$$



Example 6.5: Find the unit vector in the direction $\vec{v} = \begin{pmatrix} 1, 2, 1 \end{pmatrix}$.

How can we find a unit vector that points in the same direction?
First, let's find the length of $\vec{v}$.

$$|\vec{v}| = \sqrt{1^2 + 2^2 + 1^2} = \sqrt{6}$$

If we divide $\vec{v}$ by its magnitude, we get the vector

$$\vec{u} = \left(\frac{1}{\sqrt{6}}\right)\vec{v} = \begin{pmatrix} \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \end{pmatrix}$$

Let's check our work. To do this, we ask ourselves, "What is the magnitude of this vector?"

$$|\vec{u}| = \sqrt{\left(\frac{1}{\sqrt{6}}\right)^2 + \left(\frac{2}{\sqrt{6}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2} = \sqrt{\frac{1}{6} + \frac{4}{6} + \frac{1}{6}} = 1$$

Notice that $\vec{u}$ is just the vector $\vec{v}$ multiplied by a positive scalar, so the direction doesn't change. So our unit vector is

$$\boxed{\vec{u} = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)}$$

Summary of Ideas: Lecture 6

- Vectors can be understood as a description of movement from one point to another.
- The **magnitude** of a vector, denoted $|\vec{v}|$, is its length. We can find it using the pythagorean theorem:

$$|\vec{v}| = \sqrt{x^2 + y^2 + z^2}$$

- Multiplying by a scalar changes the length of a vector. If that scalar is negative, the vector flips its direction.

7. Lecture 7

Objectives

- ☐ I know that a vector can be filled with any kind of information.
- ☐ I can define a unit vector.
- ☐ I can calculate the dot product of two vectors.

In the last lecture, we thought of vectors as being descriptions in space. In reality, a vector can contain any kinds of data. They simply summarize what is happening in a system.

In the context of physics, most systems can be summarized by position and possibly time. So a common vector might be

$$\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}$$

where t represents time and x, y, and z represent spacial positions. These are **inputs** of a system and should be a set of **independent variables**.

Unless you're modeling the physical movement of people, animals, or goods, you generally won't use vectors to represent space and time. Instead, you'll populate it with data with the independent variables that describe the system you are studying. For example, if you are studying gas milage, you might look at a system like this

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

where x_1 represents the weight of the car, x_2 represents the incline of the road, x_3 represents, x_3 represents the outside temperature, x_4 represents the number of cylinders in the engine, x_5 represents the airpressure of your tires, and x_6 represents the aerodynamics of the car.

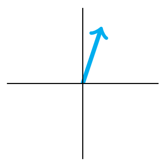
With mechanical systems, like gas mileage of a car, it's not very difficult to determine your variables. You need to know how the system works, which can be figured out by taking it apart.

In the biological, geological and social sciences, it is incredibly difficult to determine your variables! It's not always obvious which variables to use when modeling a system because we don't always know whether a variable is independent of the system or not. For example, suppose we want to model what causes a person to use heroine. We might look at variables like income or the percentage of people in their social network who use heroine. Neither of these, however, is clearly an independent variable. Because heroine use can get in the way of a job, it can influence a person's income. Similarly, heroine users are more likely to hang out with other heroine uses because they may share the risk of purchasing an illegal drug and/or prefer to use the drug in groups.

When a vector doesn't represent movement in space, what does direction mean? It depends. If the variables are different kinds of measurements like a percentage of people in a social network versus an annual income, the direction is meaningless. Direction compares the amounts with one another. If the measurements are not comparable, than direction means nothing. On the other hand, if the two variables are comparing measurements that exist on the same scale, i.e.

(1) measurements all have the same units (like feet) or
(2) values are all percentages,

then the direction allows us to see which variable dominates. For example, in the picture below, the y variable dominates (is larger) over the x variable because the arrow is closer to being parallel to the y-axis.



The magnitude will (in general) not be very helpful for two measurements on the same scale. For measurements which are not on the same scale, there are instances when the magnitude provides important information. For the purposes of this class, however, we will only use magnitude when

(1) discussing spacial vectors, or
(2) when wanting to find a **unit vector**.

7.1. **Unit Vectors.** A unit vector is a vector whose magnitude is 1. For example

$$\vec{v} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

is a unit vector because

$$|\vec{v}| = \sqrt{\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2} = 1.$$

Any vector can be made into a unit vector if we divide by it's magnitude. For example

$$\vec{w} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

is not a unit vector. It's magnitude is

$$|\vec{w}| = \sqrt{(1)^2 + (2)^2 + (3)^2} = \sqrt{14}.$$

But if we divide $\vec{w}$ by $\sqrt{14}$, we get a unit vector.

$$\frac{\vec{w}}{\sqrt{14}} = \begin{pmatrix} \frac{1}{\sqrt{14}} \\ \\ \frac{2}{\sqrt{14}} \\ \\ \frac{3}{\sqrt{14}} \end{pmatrix}$$

We can verify that this is a unit vector:

$$\left|\frac{\vec{w}}{\sqrt{14}}\right| = \sqrt{\left(\frac{1}{\sqrt{14}}\right)^2 + \left(\frac{2}{\sqrt{14}}\right)^2 \left(\frac{3}{\sqrt{14}}\right)^2} = 1$$

We typically try to find a unit vector when we want to only understand direction. This idea will become clearly in later lectures on optimization.

7.2. **Dot Products.** Today we will cover *how* to calculate a dot product, but we won't discuss its meaning until a little later in the course.

A **dot product** is a scalar value that results from taking the sum of the products of respective entries. Given two vectors

$$\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix} \qquad \vec{w} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix}$$

then their dot product is

$$\vec{v} \cdot \vec{w} = x_1y_1 + x_2y_2 + x_3y_3 + \dots$$

Example 7.1:1. What is the dot product of $\vec{u} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}$

Solution 7.2:

$$\vec{u} \cdot \vec{v} = (1)(-1) + (2)(0) + (-3)(-1) = -1 + 3 = \boxed{2}$$

Summary of Ideas: Lecture 7

- Vectors can contain any kind of information. They are typically used to summarize a system.
- In general, magnitude and direction are meaningless.
- A unit vector is any vector of length one.
- Given a any vector $\vec{v}$, we can find a unit vector with the same direction as $\vec{v}$. It is
$$\frac{\vec{v}}{|\vec{v}|}$$
- The **dot product** is a scalar value produced from two vectors. The formula is:
$$\vec{v} \cdot \vec{w} = x_1y_1 + x_2y_2 + x_3y_3 + \dots$$

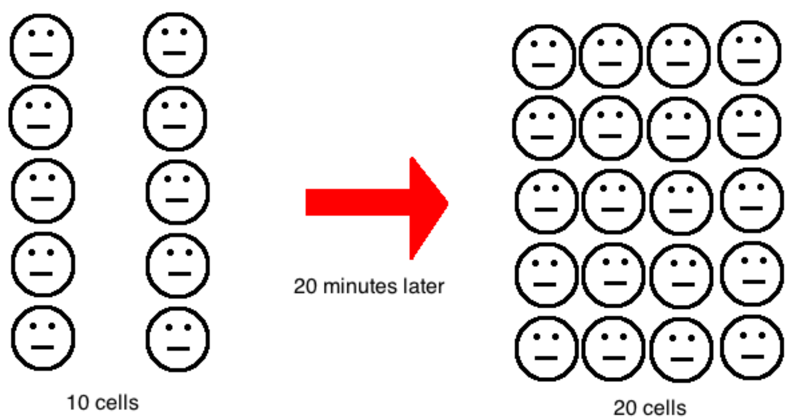
8. LECTURE 8

Objectives

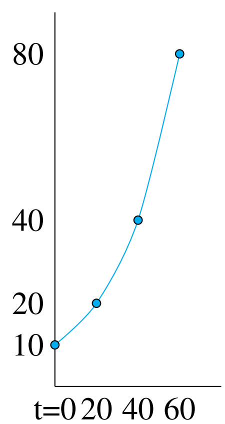
- ☐ I understand that two measurements can have a nonlinear relationship.
- ☐ I know the shapes of linear, exponential, logarithmic, logistic, and polynomial functions (up to degree 4).
- ☐ I can fit data with any of the above function-types using a graphic calculator.
- ☐ I can find the residuals to help me determine if a particular function type is appropriate.

Up until now, we have focused on linear relationships, many interesting relationships in nature are not linear. Most populations do not grow linearly.

Why? Well, let's consider the very simple model of bacteria like E. coli. E. coli cells split into two every 20 minutes. So if we begin with 10 cells, we assume that after 20 minutes, we get double that amount: 40 cells.



While this may sound linear on the surface, let's see what happens after a few 20-minute intervals:



As you can see, this is not a line at all! This is an exponential function. In fact, exponential growth describes all unconstrained population growth.

Exponential Growth 8.1

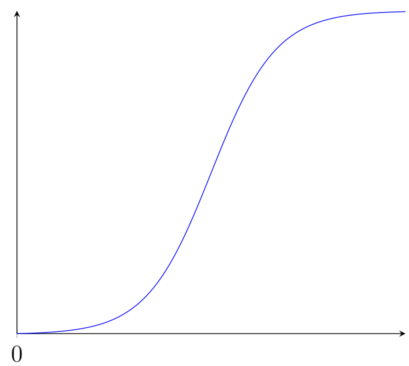
Biological populations, without constraints like limits on resources and space, grow according to some exponential function.

Exponential growth is of the form

$$y = ae^{rt}$$

where a is the initial population and r is the growth rate per time step.

In our model of E. coli growth, we don't consider death rates nor how those death rates change with respect to external limitations, like insufficient food or space. Growth begins exponential until those limitations are felt by the population. It then levels off and flattens out to the maximum population. Below is an image of a logistic graph.



A logistic function is of the form

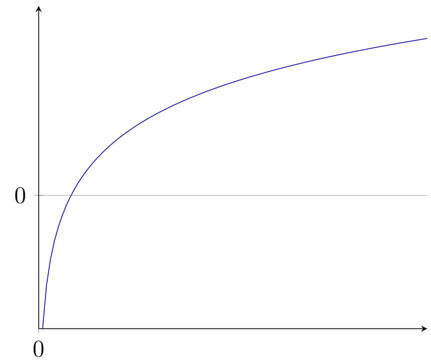
$$y = \frac{L}{1 + e^{-kx}}$$

where L is the initial population. We will revisit this function and why it is written this way during the differential equations portion of this class.

Logistical Growth 8.2

Biological populations with constraints (like limited food or space) grow according to some logistical function.

There are two other kinds of nonlinear functions the graphing calculator is equipped to handle. The first is logarithmic functions. Logarithmic functions are inverses of exponential functions. They are a class of functions that go to infinite very slowly. They appear as if they eventually become flat, but don't be fooled!



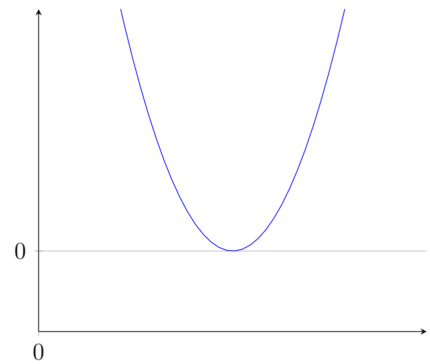
Logarithmic functions are of the form

$$y = a\ln(x) + b$$

Lastly, your calculator can fit polynomial functions. Polynomials are functions of the form

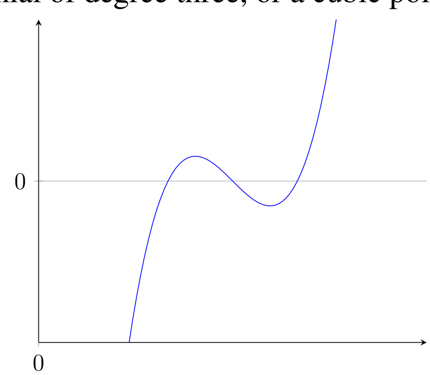
$$y = ax^n + bx^{n-1} + cx^{n-2} + \dots$$

where n is the degree of the polynomial and $a, b, c, \dots$ are constants. These are functions that change direction (e.g. positive slope to negative slope) $n - 1$ times. For example, the function graphed below is a degree-2 polynomial, or a quadratic.



Notice that it changes direction 1 at the vertex of. Recall from calculus that these points where the direction changes are the critical points of the function, where the derivative is zero.

Here is example of a polynomial of degree three, or a cubic polynomial.



Again, notice that it changes direction twice, precisely at the critical point

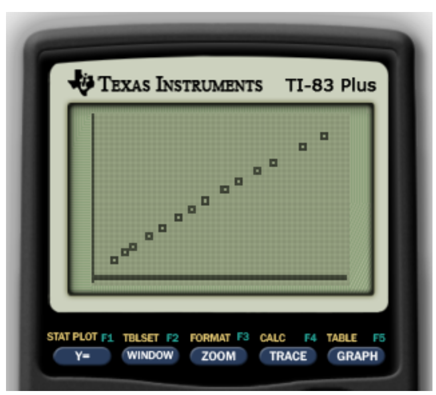
Because we can control how many times a polynomial can change direction by taking a higher degree, we can always fit data to a polynomial. While we can get a good fit, however, it does not mean that this fit is physically realistic. While we will not focus on it too much in this lecture, we will generally try to determine which function to use not just by the data collected but also by the physical situation.

8.0.1. *Best Fit Curves.* We now discuss how to use a graphing calculator to find a best fit curve. As you will see, it is the same process as finding the best fit line. The steps outlined here will skip some steps as they were covered in previous lectures. If you do not remember them, please check back on earlier notes.

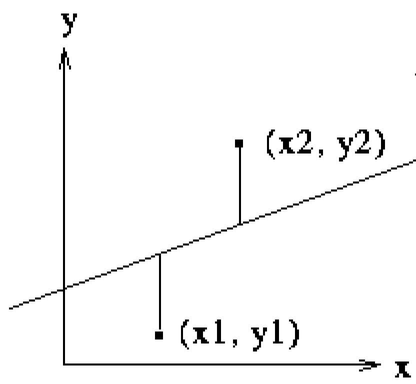
Let's begin with a sample data set. When plugging these values into your calculator, be sure to put the x values into column L_1 and the y values into column L_2.

y	x
10.07E0	77.6E0
14.73E0	114.9E0
17.94E0	141.1E0
23.93E0	190.8E0
29.61E0	239.9E0
35.18E0	289.0E0
40.02E0	332.8E0
44.82E0	378.4E0
50.76E0	434.8E0
55.05E0	477.3E0
61.01E0	536.8E0
66.40E0	593.1E0
75.47E0	689.1E0
81.78E0	760.0E0

Notice that when we plot the data, it appears linear. If you look at the functions described above, many of them can look linear if we only look at a small segment.



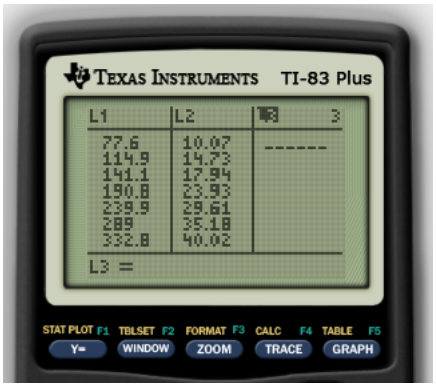
How do we distinguish a linear relationship against a nonlinear one? The answer is to look at the **residuals**. A residual is the difference (in the y-value) between a data point and the best fit curve.



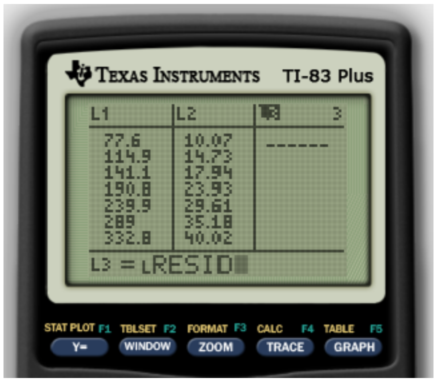
We use these distances to see if we can detect some curvature about the line. Here is how we graph residuals using your calculator.

- Find the best fit line: $\boxed{\texttt{STAT}}$ → CALC → `LinReg(ax + b)`
- Make a list of residuals: $\boxed{\texttt{STAT}}$ → EDIT → `Edit...`

Navigate to the L_3 column and navigate to the top.

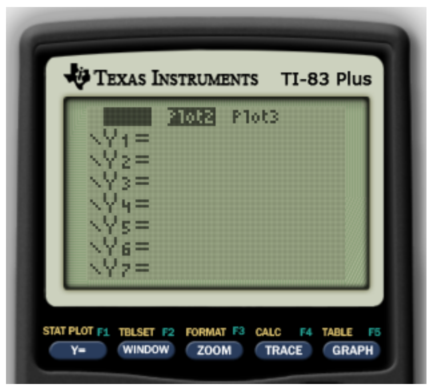


To access the residuals, press $\boxed{\texttt{2nd}}$ and $\boxed{\texttt{STAT}}$. Under the NAMES menu, select `RESID`.

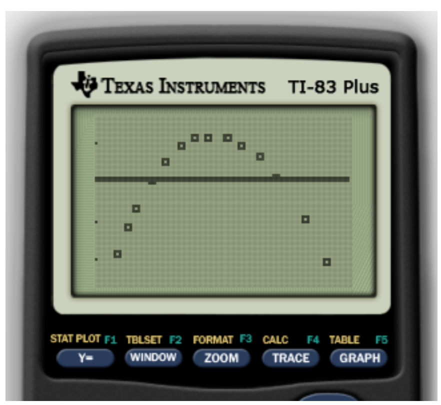


Press $\boxed{\texttt{ENTER}}$ and the third column will fill with the residuals. To graph them, first quit by pressing $\boxed{\texttt{2nd}}$ and $\boxed{\texttt{MODE}}$. Navigate to the `STAT PLOT` menu and edit `Plot2`. Select ON and change the `Ylist` to L_3

Quit this menu. Then press Y=. At the top, you will see `Plot1` and `Plot2` highlighted. Navigate upward to `Plot1` and hit enter to unhighlight it.



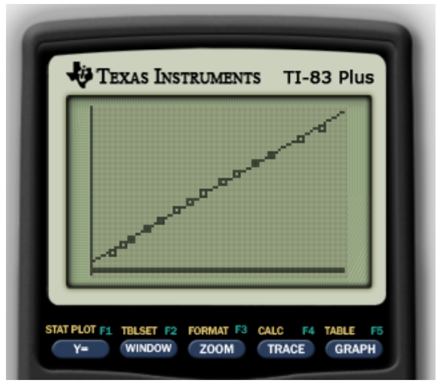
Then press ZOOM and select `ZoomStat`. Then you will see the residuals.



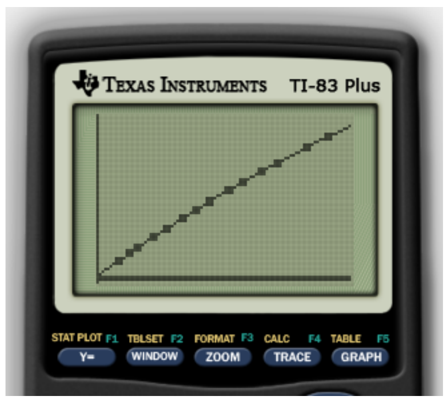
The graph of the residuals tells us that the data curves around the line, suggesting that a line is *not* appropriate for this data. Instead, we should try a nonlinear function.

Nonlinear best-fit curves can be graphed just like the lines. The functions available on your calculator are listed below. Each has the accompanying graph for the data set in this lecture.

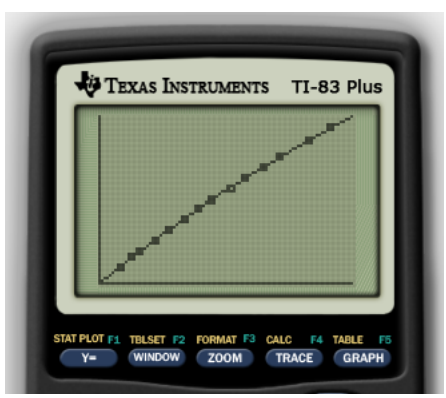
(1) `LinReg(ax+b)` for lines



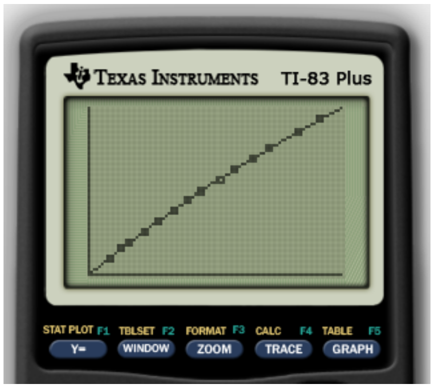
(2) `QuadReg` for polynomials of degree 2, i.e. quadratics



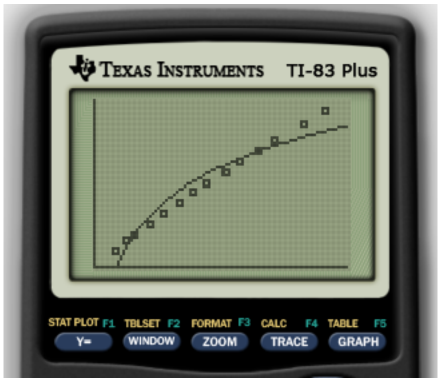
(3) `CubicReg` for polynomials of degree 3, i.e. cubics



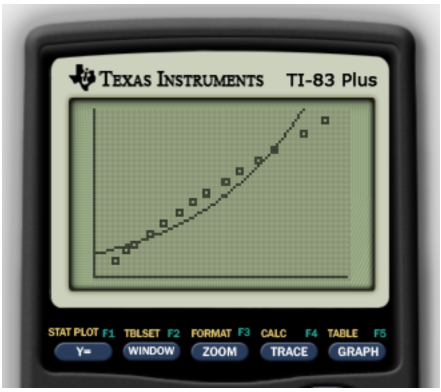
(4) `QuartReg` for polynomials of degree 4, i.e. quartics



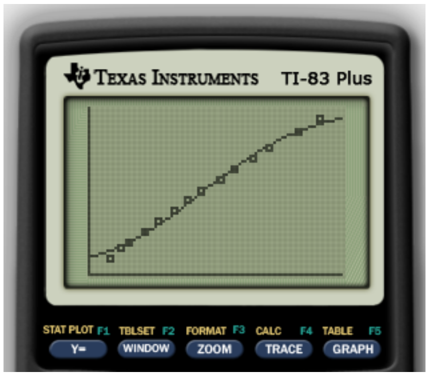
(5) `LnReg` for natural logs



(6) `ExpReg` for exponentials



`Logistic` for logistiscs



When comparing all the possible functions, the polynomials will generally fit the best (as you can see above). We generally want to pick the function based on *where* the data comes from rather than by just the graph.

One last calculator trick that may make testing out these various functions faster is the following. When you want to plot a best-fit curve, begin by pressing `Y=`. From there, press `VARS` and select `Statistics`. Navigate over to the `EQ` and select `RegEQ`. Then the last best-fit curve you calculated will appear in the `Y=` menu.

Summary of Ideas: Lecture 8

- Some pairs of measurements, like population and time, do not have a linear relationship.
- The nonlinear relationships we cover here are exponential, logarithmic, logistic, and polynomial.
- Populations grow exponentially unless there are constraints on resources. Then the growth is logistic.
- Polynomials can generally fit any kind of data very well, although they do not generally describe a lot of natural phenomenon.
- We use residuals to detect curvature of data around a best fit line.

Objectives

- ☐ I understand that a matrix is a map that sends vectors to vectors.
- ☐ Given data for two measurements, I can fit that data to *any* kind of function.
- ☐ Given data for any number of measurements, I can fit that data to *any* kind of function.

We first introduced matrices as a tool to find a solution to a system of linear equations. That is, we turned a set of linear equations like these

$$\begin{aligned} 2x - 3y + z &= 10 \\ 2y + 2z &= 4 \\ 2x \qquad - 3z &= 1 \end{aligned}$$

into an equation like this

$$\begin{pmatrix} 2 & -3 & 1 \\ 0 & 2 & 2 \\ 2 & 0 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \\ 1 \end{pmatrix}$$

When written with matrices, we can think of the expression as

$$A\vec{x} = \vec{b}$$

In the above equation, notice that the matrix A sends $\vec{x}$ to $\vec{b}$. When we solve the equation (when we calculate $\vec{x} = A^{-1}\vec{b}$) we get the solution

$$\begin{pmatrix} \frac{26}{7} \\ -\frac{1}{7} \\ \frac{15}{7} \end{pmatrix}.$$

We know this is the solution to our set of equations above. That is, if we plug in these values into our original set of equations, we find that they evaluate to exactly the values we want.

$$2\left(\frac{26}{7}\right) - 3\left(-\frac{1}{7}\right) + \left(\frac{15}{7}\right) = 10$$

$$2\left(-\frac{1}{7}\right) + 2\left(\frac{15}{7}\right) = 4$$

$$2\left(\frac{26}{7}\right) \qquad\qquad - 3\left(\frac{15}{7}\right) = 1$$

Alternatively, we can also say that the following equation is true as well.

$$\begin{pmatrix} 2 & -3 & 1 \\ 0 & 2 & 2 \\ 2 & 0 & -3 \end{pmatrix} \begin{pmatrix} 26/7 \\ -1/7 \\ 15/7 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \\ 1 \end{pmatrix}$$

In other words, the matrix

$$\begin{pmatrix} 2 & -3 & 1 \\ 0 & 2 & 2 \\ 2 & 0 & -3 \end{pmatrix}$$

sends the vector

$$\begin{pmatrix} 26/7 \\ -1/7 \\ 15/7 \end{pmatrix}$$

to the vector

$$\begin{pmatrix} 10 \\ 4 \\ 1 \end{pmatrix}.$$

Although we introduced matrices as a table of numbers, it's better to think of a matrix as a map between vectors. When we multiply a vector with a matrix[6], we get another vector.

9.1. **How Your Calculator Fits Curves to Data.** How does your calculator come up with these equations? How does it know which line or other function *best* fits the data?

It does this using matrices! Essentially, it tries to solve a problem very similar to the problem we solve with the system of linear equations. Unfortunately, it's not *exactly* the same because a few things go wrong. Let's see how this works and what goes wrong with an example.

To begin this process, you need two things:

(1) a data set, and
(2) a function to fit to the data set.

For our example, suppose you have the data below

[6]Remember that in order to multiply a matrix with a vector, we need the matrix's number of columns to be equal to the vector's number of rows.

x	y
1	3
2	20.9
3	43.3
4	68.6
5	96.3
6	125.9

and you suspect that the data should be fit to a function of the form below.

$$y = ax\ln(x) + b\ln(x) + cx + d$$

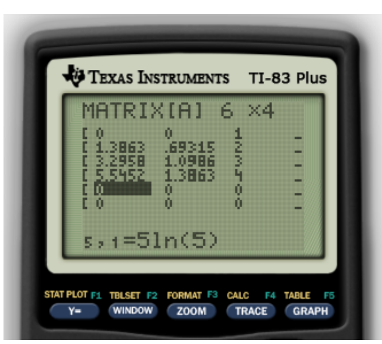
Notice that this function *does not* exist in your calculator. How do you do it? The trick is to change your perception. Rather than thinking of x and y as variables, we can think of the coefficients a, b, c and d as variables. If we use our data set, we can define a set of *linear* equations. We can plug in the values of x and y from our data and try to solve for the coefficients.

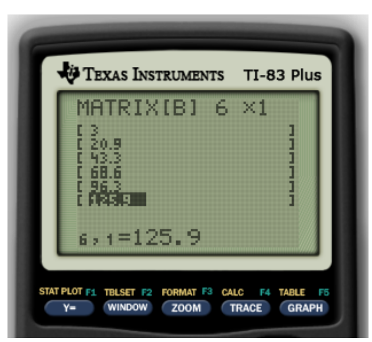
$$\begin{aligned}
a(1)\ln(1) + b\ln(1) + c(1) + d &= 3\\
a(2)\ln(2) + b\ln(2) + c(2) + d &= 20.9\\
a(3)\ln(3) + b\ln(3) + c(3) + d &= 43.3\\
a(4)\ln(4) + b\ln(4) + c(4) + d &= 68.6\\
a(5)\ln(5) + b\ln(5) + c(5) + d &= 96.3\\
a(6)\ln(6) + b\ln(6) + c(6) + d &= 125.9
\end{aligned}$$

This system of linear equations can be written as we have in the past.

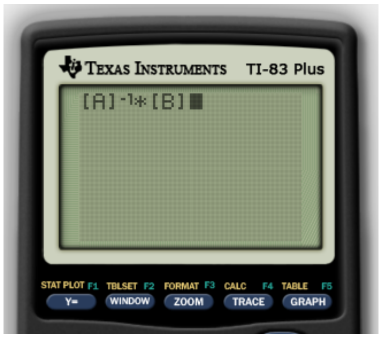
$$\begin{pmatrix} 0 & 0 & 1 & 1\\ 2\ln(2) & \ln(2) & 2 & 1\\ 3\ln(3) & \ln(3) & 3 & 1\\ 4\ln(4) & \ln(4) & 4 & 1\\ 5\ln(5) & \ln(5) & 5 & 1\\ 6\ln(6) & \ln(6) & 6 & 1 \end{pmatrix} \begin{pmatrix} a\\ b\\ c\\ d \end{pmatrix} = \begin{pmatrix} 3\\ 20.9\\ 43.3\\ 68.6\\ 96.3\\ 125.9 \end{pmatrix}$$

Suppose we attempt to solve this equation as we have other systems of linear equations. Remember that we access the `MATRIX` menu by pressing $\boxed{\texttt{2nd}}$ and $\boxed{x^{-1}}$. Under the `EDIT` menu, change the entries of matrices `[A]` and `[B]`.





Next, we quit by pressing `2nd` and `MODE`. Then we type in the equation by returning to the `MATRIX` menu and selecting matrices `[A]` and `[B]` under the `NAMES` menu. If we enter our normal equation, we get an error message.



Why is there an error message? Essentially, it's because we have too many equations for the number of unknowns we have.

For example, suppose we have two equations with one variable:

$$2x = 4$$
$$3x = 6.1$$

What is the solution? It has no solution! The first equation implies that $x = 2$ but the second equation implies that $x = 61/30 \approx 2.03333$. Because x cannot have two different values, the system has no true solution.

The same problem is occurring with our data. we only have 4 variables to determine, but we have too many equations. When we graph our best-fit curves, the curve doesn't go through *all* points. It simply needs to follow the pattern of the data and minimize the data.

Because we cannot find a true solution, we can approximate our solution with a different equation:

$$\vec{x} = (A^T A)^{-1}(A^T \vec{b}) \tag{9.1}$$

This equation is a *projection*. When we learn about projects for vectors, we will return to this equation and explain why it is a projection. For now, we will simply use the equation and take for granted what it achieves.

What does the notation mean? A^T is called the *transpose* of A. A transpose reorganizes a matrix so that its columns are rows and vice versa.

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 2\ln(2) & \ln(2) & 2 & 1 \\ 3\ln(3) & \ln(3) & 3 & 1 \\ 4\ln(4) & \ln(4) & 4 & 1 \\ 5\ln(5) & \ln(5) & 5 & 1 \\ 6\ln(6) & \ln(6) & 6 & 1 \end{pmatrix} \qquad \text{means}$$

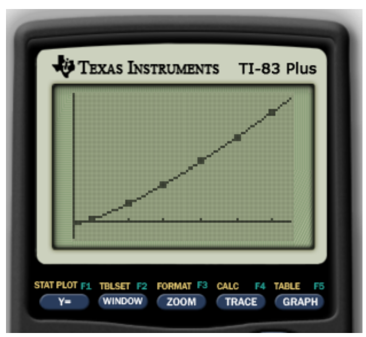
$$A^T = \begin{pmatrix} 0 & 2\ln(2) & 3\ln(3) & 4\ln(4) & 5\ln(5) & 6\ln(6) \\ 0 & \ln(2) & \ln(3) & \ln(4) & \ln(5) & \ln(6) \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Now, let us use our calculator to find the solution to this equation. You'll need to access the transpose by accessing the `MATRIX` menu. Under `MATH` select transpose, which is indicated by a little "T."

According to our calculation, our best-fit for the data with the given equation is

$$y = 9.898x\ln(x) + 2.726\ln(x) + 2.320x + 0.676$$

To check our work, let's see how this function fits with our data.



9.2. **Multi-Variable Best-Fit Curves.** Now that we know how to fit data to a function, we can consider multiple independent variables. Suppose we have the data set

x	y	z
1	1	3
2	1	9
1	2	4
2	2	16
3	1	27
3	2	27
1	3	5
2	3	23
3	3	39

and we wish to fit the data to the function

$$z = ay\ln(x) + bxy + cy + d.$$

Then we treat the coefficients a, b, c, and d as variables by plugging in data points. We get the following set of equations:

$$\begin{aligned}
a(1)\ln(1) + b(1)(1) + c(1) + d =& 3\\
a(1)\ln(2) + b(2)(1) + c(1) + d =& 9\\
a(2)\ln(1) + b(1)(2) + c(2) + d =& 4\\
a(2)\ln(2) + b(2)(2) + c(2) + d =& 16\\
a(1)\ln(3) + b(1)(1) + c(1) + d =& 27\\
a(2)\ln(3) + b(2)(3) + c(2) + d =& 27\\
a(3)\ln(1) + b(1)(3) + c(3) + d =& 5\\
a(3)\ln(2) + b(2)(3) + c(3) + d =& 23\\
a(3)\ln(3) + b(3)(3) + c(3) + d =& 39
\end{aligned}$$

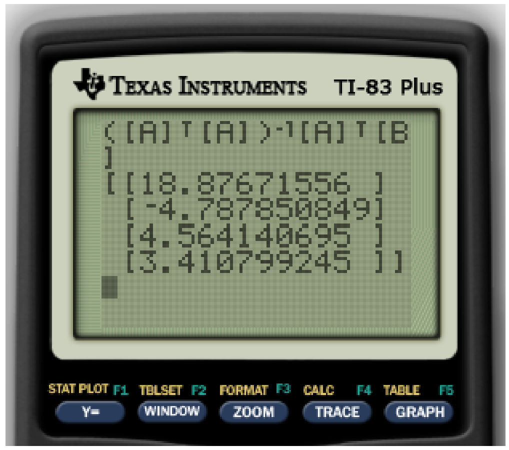
Notice that the only difference is plugging in more than one variable! We simply plug in the values according to the format $z = ay\ln(x) + bxy + cy + d$.

Then our matrix expression is

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ \ln(2) & 2 & 1 & 1 \\ 0 & 2 & 2 & 1 \\ 2\ln(2) & 4 & 2 & 1 \\ \ln(3) & 1 & 1 & 1 \\ 2\ln(3) & 6 & 2 & 1 \\ 0 & 3 & 3 & 1 \\ 3\ln(2) & 6 & 3 & 1 \\ 3\ln(3) & 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 3 \\ 9 \\ 4 \\ 16 \\ 27 \\ 27 \\ 5 \\ 23 \\ 39 \end{pmatrix}$$

Finally, we enter this information into our calculator and solve $\vec{x} = (A^TA)^{-1}(A^T\vec{b})$.

When we do, we get the solution



That means the equation that best fits the data is

$$z = 18.9y\ln(x) - 4.6xy + 4.6y + 3.4.$$

Unfortunately, our graphing calculator cannot graph functions in higher dimensions. So we cannot compare our graph with our data.

Summary of Ideas: Lecture 9

- Matrices are functions that take in a vector and produce a vector.
- We can find a best-fit curve by setting it up as a system of linear equations.
- Generally, we cannot solve this system. It often has more equations than unknown variables.
- To solve, we use an equation based on projections. It is

$$\vec{x} = (A^T A)^{-1}(A^T\vec{b})$$

- To fit data with more than one independent variable, we can use the *same* procedure.

Objectives

- ☐ I understand the difficulty in finding an appropriate function for a data set in general.
- ☐ In some cases, I can define a function type that may fit a data set well.

Last time, we learned the procedures for how to fit a particular function to a data set. Where does such a function come from? If we have a data set, how can we determine which function(s) to try?

Unfortunately, there is no good answer to this question. Throughout the sciences, researchers have defined important relationships between physical measurements and come up with mathematical expressions like those below.

- **Law of Gravitation:** $F = G\frac{m_1 m_2}{r^2}$
- **Boyle's Law (pressure & volume):** $P_1 V_1 = P_2 V_2$
- **Hardy-Weinberg Law (population genetics):** $(p+q)^2 = 1$

Generally speaking, these laws are discovered is by lots of observations and by trial and error. The functions tried are often based on researchers' expectations. Once a consistently good equation is found, researchers then attempt to justify why such an equation makes sense for the situation.

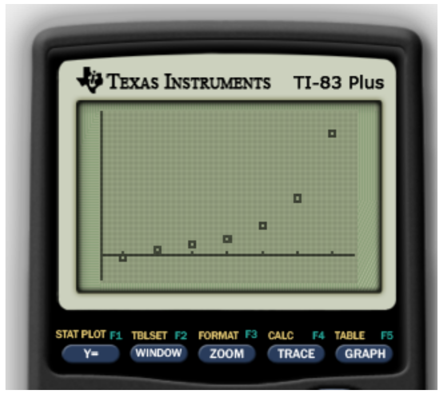
Here, we will spend some time discussing how we might determine a function for a data set with no physical intuition.

10.1. **Combining Different Fits for One Independent Variable.** Suppose you have the following data set:

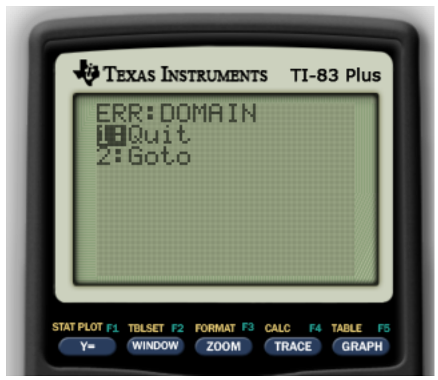
x	**y**
1	-11.32
2	12.85
3	27.69
4	46.65
5	81.10
6	157. 19
7	344.44

I have generated this data from a particular function. Let's see if we can guess it.

The first step is to graph the data and study the general shape. Below is a picture of the data.



Given its shape, we might begin by trying to fit an exponential equation. If we try to calculate a best-fit with `ExpReg`, we get an error message.



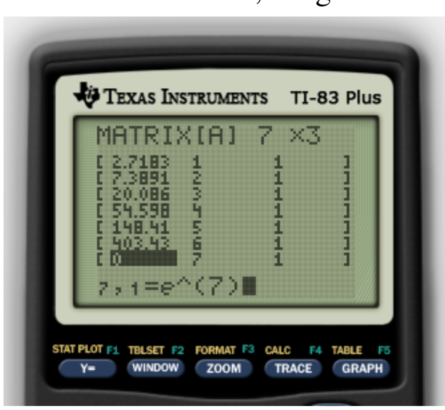
The error is an issue with the domain. That means one of the y-values is not appropriate for the type of function. This would be -11.32.

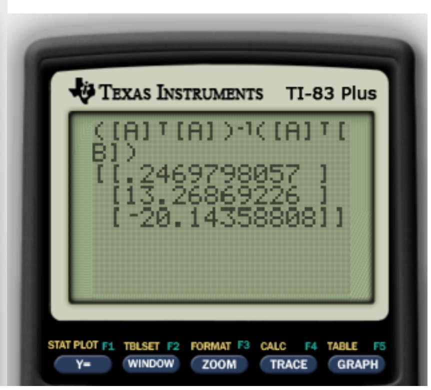
We may suppose that it is an exponential shifted down by a constant (about -8.6). A quick calculation shows us that if this were the case, then the y-value for $x = 2$ would also be negative. Therefore, we must have an additional term. We are not sure what term this might be, so we may try a linear one.

That gives us the form

$$y = ae^x + bx + c.$$

If we fit this function to the data, we get the following:

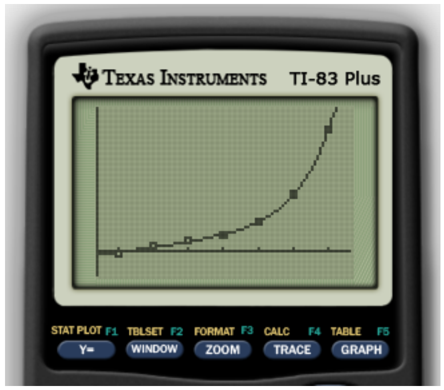




That gives us the function

$$y = .25e^x + 13x - 20.$$

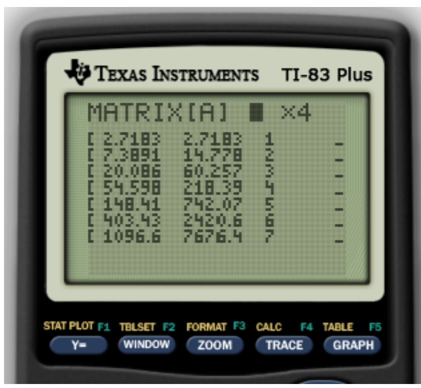
Let's see how well this matches the function.



This is actually quite good, but we may be able to do better. Let's consider a cross-terms between x and e^x. That is,

$$y = axe^x + be^x + cx + d.$$

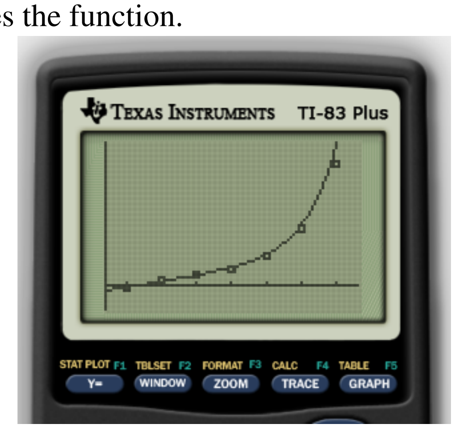
If we fit this function to the data, we get the following:



That gives us the function

$$y = .037xe^x + .028e^x + 16x - 24.$$

Let's see how well this matches the function.



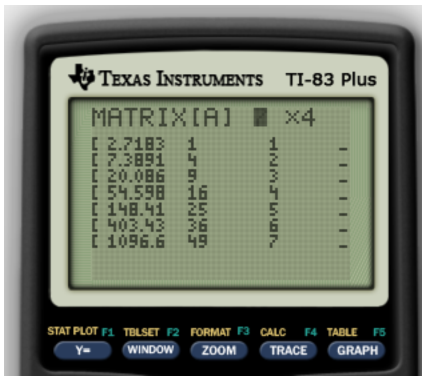
This fit is worse than the first. Notice that the curve is increasing faster than the points. That suggests that xe^x might be a bad term to add in since it is the fastest growing term.[7]

[7] You can take the derivative to verify this.

Another possible function to consider is one with a higher-order polynomial term. That is,

$$y = ae^x + bx^2 + cx + d.$$

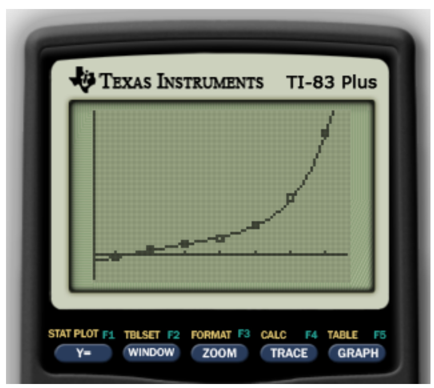
If we fit this function to the data, we get the following:



That gives us the function

$$y = .27e^x - 1.7x^2 + 23x - 32.$$

Let's see how well this matches the data.



This is pretty good. We can, of course, keep going and test a variety of other functions next. For example,

- $y = ae^x + bx^3 + cx^2 + dx + e$
- $y = ae^x + bx^2 + cx + d\sqrt{x} + e$
- $y = ae^x + bx^2 + cx + \frac{d}{x} + e$
- $y = ae^x + b\ln(x) + cx^2 + dx + e$

Any of these would be worth trying, but it is not very clear when one has the best possible answer. What is a great fit?

For this data set, it turns out the numbers come from the equation

$$y = .25e^x + x^2 + 27 - \frac{40}{x}.$$

If you study what we did, we were able to figure out the largest term e^x, but beyond that, we had a lot of trouble. This is pretty typical.

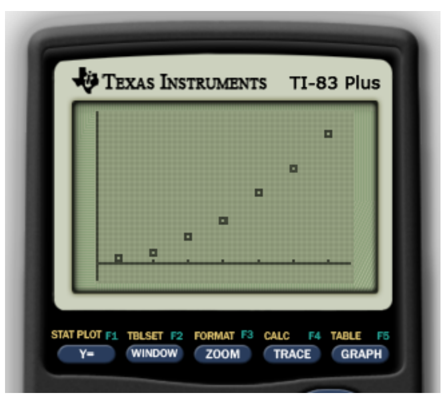
> **"Procedures" to Fit Data (One Variable) 10.1**
>
> 1.) Graph the data and try to determine a basic function type that may describe the general trend.
> 2.) Add in extra terms based on what you see and what you think may improve the fit.
> 3.) There is no clear stopping point. There is no clear way to verify that you have a good fit.

10.2. **Combining Different Fits for Many Independent Variables.** For multiple variables, it's a little more complicated. The idea is to consider each independent variable by itself. Then, once we have a set of functions, we will try to combine all of them into a multivariable function.

For example, suppose you have the data

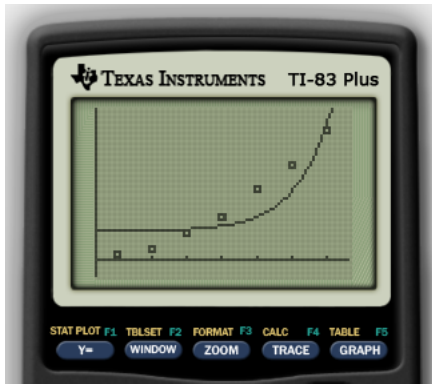
x_1	x_2	y
1	2	2.7
2	1	5.0
3	4	14.2
4	3	21.4
5	6	35.0
6	5	46.7
7	8	64.6

Then our first step is to find a function that models how x_1 influences y. As we saw above, that begins with a graph.



How might we describe this function? This looks exponential, so let's try fitting an exponential of the form $ae^{x_1} + b$.

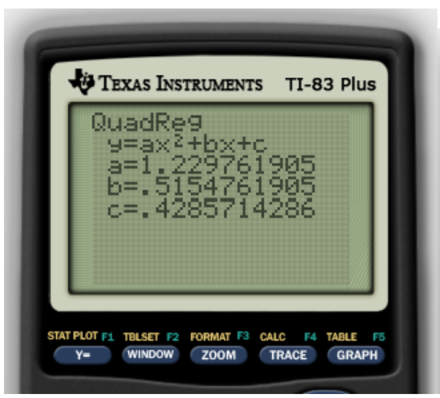
When we do, we get $y = .051e^{x_1} + 14.4$. If we compare this with the graph, we get the image below.

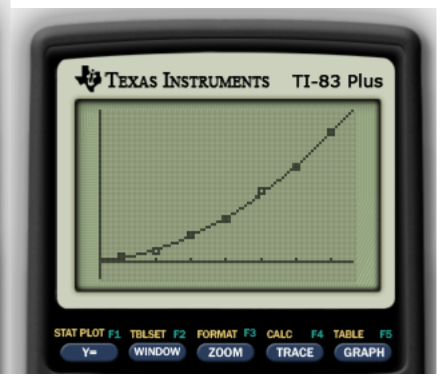


As you can see, it's a terrible fit! Our next attempt might be some other curved, increasing function like a quadratic. For that, we can use the calculator's program. It gives us the following function.

$$y = 1.3x_1^2 + .52x_1 + .42$$

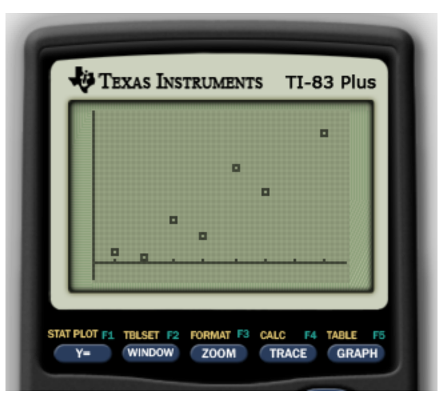
Let's see how that holds up:



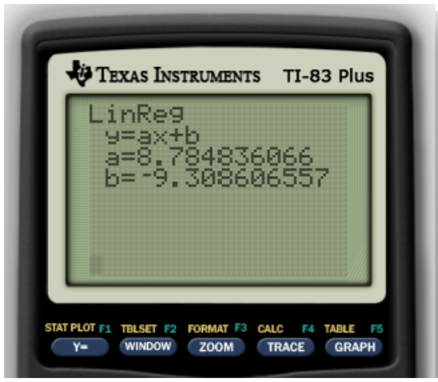


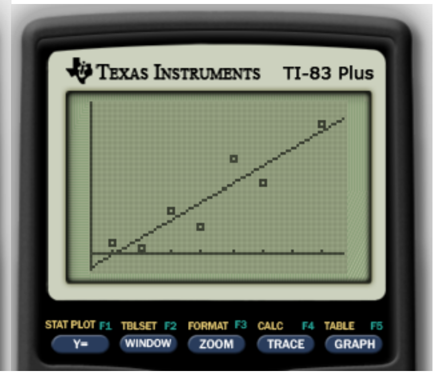
That looks quite good. So we will say that the relationship between y and x_1 is defined by a *quadratic* relationship.

Now let's see how y is related to x_2. First, we graph.



It's difficult to determine from the picture what kind of relationship exists. The only reasonably guess is linear. So let's find a best-fit line and compare.





Overall, this looks like a good fit. So we will assume that x_2 and y share a linear relationship. Therefore, our function will be constructed from these two function types.

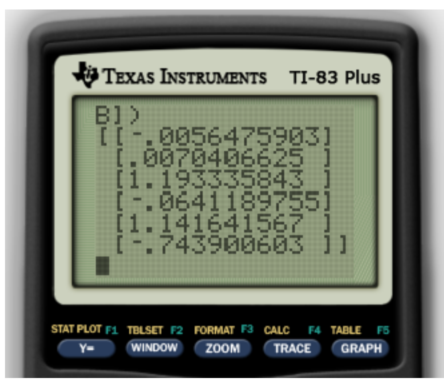
That is, we will combine

$$y = ax_1^2 + bx_1 + c \quad \text{and} \quad y = ax_2 + b$$

by looking at all possible combinations of terms. We consider each term in one equation, like x_1^2, multiplied by each term in the second equation, $x_1^2 x_2$, and by itself x_1^2. That gives us the function.

$$y = ax_1^2 x_2 + bx_1 x_2 + cx_1^2 + dx_1 + ex_2 + f$$

It is very possible that this function is way more than we need. So, we will try to fit it to the data. If any coefficient is zero or very close to zero, we will throw out that term.



In other words, we get the best-fit curve

$$y = -.006x_1^2 x_2 + .007x_1 x_2 + 1.2x_1^2 + -.06x_1 + 1.1x_2 + -.74$$

Now, we throw out the terms close to zero and try to re-fit the data. That is, we should instead consider

$$y = ax_1^2 + bx_1 + cx_2 + d.$$

When we do, we find the following coefficients.

In order words, our best-fit function for this data is

$$y = 1.1x_1^2 + .26x_1 + 1.1x_2 - .94$$

The actual function from which I got these numbers was the one below.

$$y = x_1^2 + x_1 \ln(x_2) + 1$$

It's not a perfect prediction. In reality, it is very hard to find the actual relationships. This is why, for example, it took humans a long time to realize that the sun was the center of the solar system. The geocentric models explained a lot of observations well enough. That is, until we developed a better telescopes and a better understanding of related topics, like gravity. Then, discrepancies led to finding a better model.

In the case of these two function, they don't differ too much if x_1 and x_2 are within the ranges of 1 through 7. But they will differ a lot for values far away from this range. Hence we are limited in how we use these functions for prediction. We cannot construct perfect predictions using just math. Having an understanding of a physical system is crucial to choosing better models.

In other words, expertise and mathematical know-how is *much* better than just mathematical know-how.

“Procedures” to Fit Data (Mulfi-Variable) 10.2

1.) Pick one independent variable.
2.) Graph the dependent variable against that independent variable and determine a basic function type that may describe the general trend.
3.) Add in extra terms based on what you see and what you think may improve the fit.
4.) Once you are satisfied with the function type you have selected, repeat this process for another independent variable.
5.) After you have defined one function-type for each independent variable, consider all possible combinations of the terms in all the functions. For example, if you have
$$y = a\ln(x_1) + bx_1 + c \quad \text{and} \quad y = ax_2 + \frac{b}{x_2} + c$$
then combine the two by considering each possible combination
$$y = ax_2\ln(x_1) + \frac{b\ln(x_1)}{x_2} + cx_1x_2 + \frac{dx_1}{x_2} + e$$
6.) Fit this multivariable function to the data.

10.3. **Next Time.** Up to this point, we have discussed how observations can give rise to single and multiple variable functions. From this point onward, we discuss how we use calculus to analyze these higher-dimensional functions.

Summary of Ideas: Lecture 10

- There is no standard way to find a good function for a data set. Ideally, you should use your knowledge of the physical system to guide you.
- You should also study the graph and try a number of possible functions.
- For multiple variables, you’ll need to evaluate each independent variable separately and then combine your findings at the end. If any coefficient is close to zero, you should probably throw it out.

Objectives

- ☐ I know how to graph level curves when there are two input variables.
- ☐ I can find the line of steepest descent, steepest ascent, or zero height change.

In this lecture, we will study functions that take in multiple scalar inputs, like x and y, but produce just one scalar output

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

These are called *functions of several variables*. They are the main object of study in multivariate calculus.

The first step in understanding any function is being able to graph it. Unfortunately, graphing functions in more than three dimensions quite tricky. One popular method for graphs of three and four dimensions is to graph their *level curves*.

Definition 11.1: A **level curve** of a multivariate function is any function constructed by choosing a specific value for the dependent variable.

Each level curve gives us a height of the multivariate function. For functions of three variables, we can use this information to construct the original function.

Example 11.2: Graph the level curves of

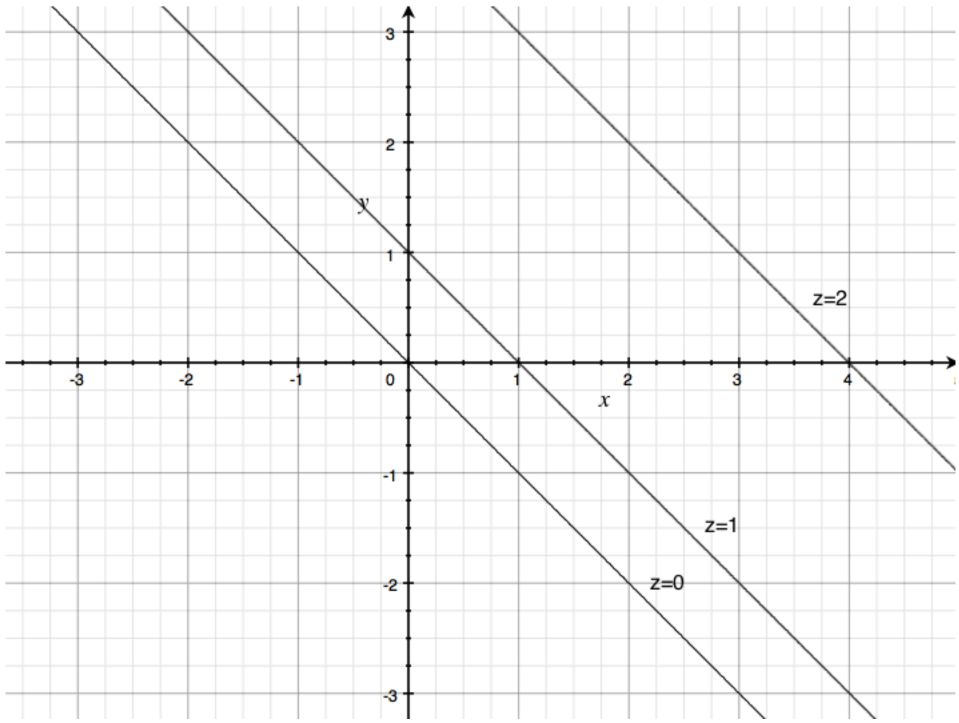
$$f(x, y) = \sqrt{x + y}$$

Use that information to sketch the 3 dimensional graph.

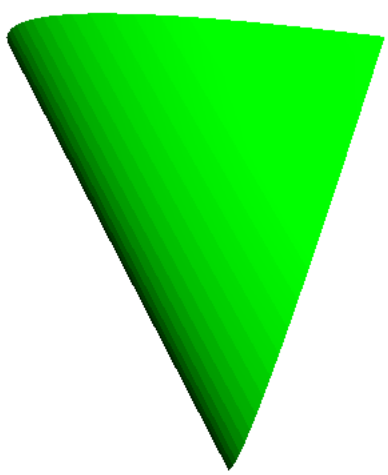
Solution 11.3: Let's pick values for f and write the respective functions.

f	Function
0	$0 = \sqrt{x+y} \implies 0 = x + y$
1	$1 = \sqrt{x+y} \implies 1 = x + y$
2	$2 = \sqrt{x+y} \implies 4 = x + y$
3	$3 = \sqrt{x+y} \implies 9 = x + y$

Now, we graph all these lines on the same graph.



From this, we see that the curve is increasing in height but the increasing is slowing down. Here is the graph of the three dimensional surface seen from two different angles.



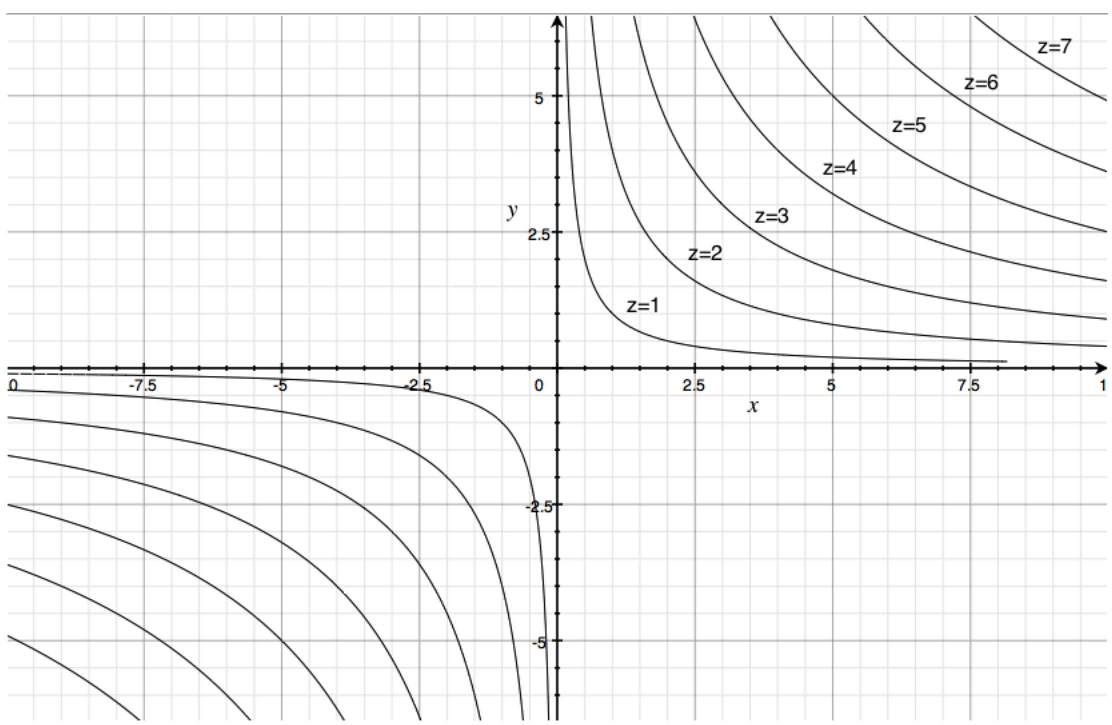
Example 11.4: Graph the level curves of

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

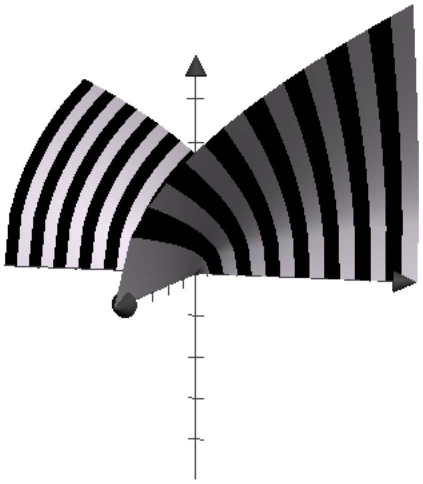
Use that information to sketch the 3 dimensional graph.

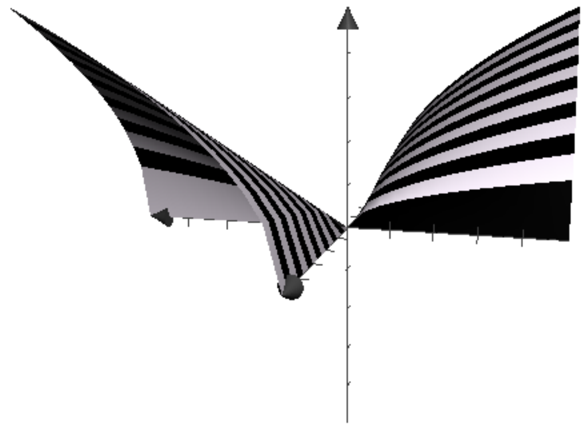
Solution 11.5: First, we pick z values. In the picture below, I have graphed $z = 1$, $z = 2$, ..., and $z = 7$.

f	Function
0	$0 = \sqrt{xy}$
1	$1 = \sqrt{xy}$
2	$2 = \sqrt{xy}$
3	$3 = \sqrt{xy}$
4	$4 = \sqrt{xy}$
5	$5 = \sqrt{xy}$
6	$6 = \sqrt{xy}$



Given the shape of the level curve, our graph is then





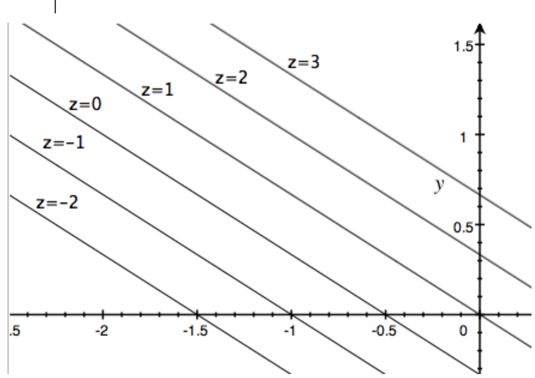
Example 11.6: Graph the level curves of

$$f(x,y) = 2x + 3y + 1$$

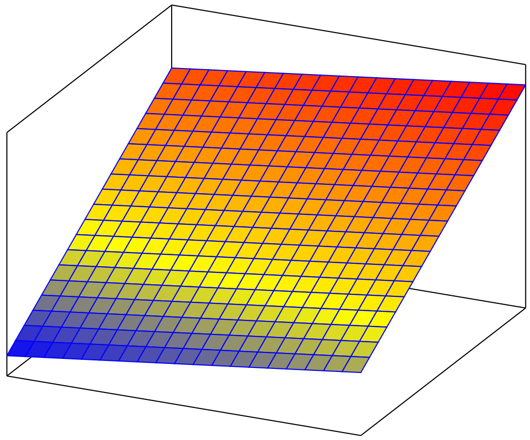
Use that information to sketch the 3 dimensional graph.

Solution 11.7: First, we pick z values.

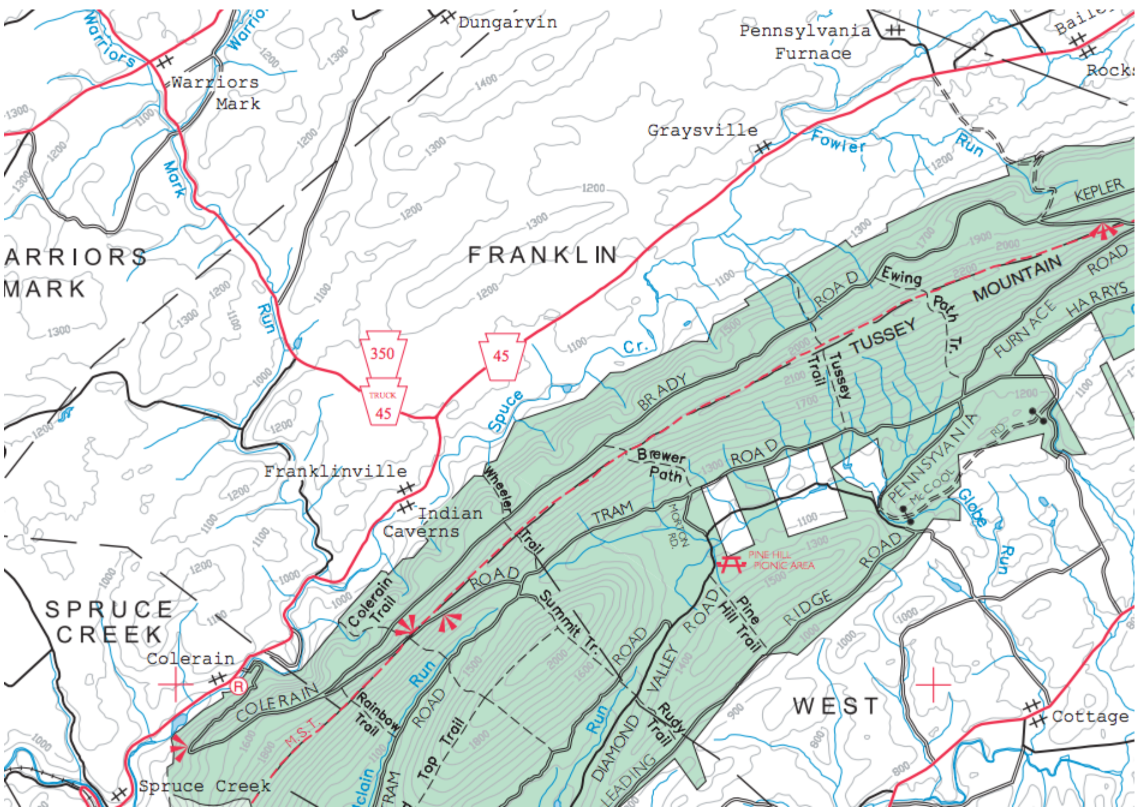
f	Function
-2	$-2 = 2x + 3y + 1 \implies y = -\frac{2x+3}{3}$
-1	$-1 = 2x + 3y + 1 \implies y = -\frac{2x+2}{3}$
0	$0 = 2x + 3y + 1 \implies y = -\frac{2x+1}{3}$
1	$1 = 2x + 3y + 1 \implies y = -\frac{2x}{3}$
2	$2 = 2x + 3y + 1 \implies y = -\frac{2x-1}{3}$
3	$3 = 2x + 3y + 1 \implies y = -\frac{2x-2}{3}$



Given the shape of the level curve, our graph is a plane.



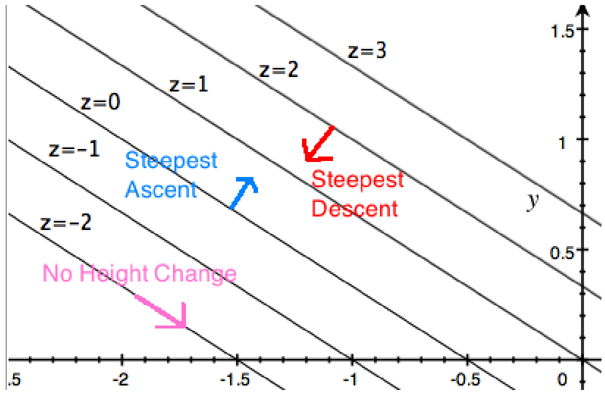
These are the lines one sees on a topographic map. Below is an example of a topographic map of Tussey Mountain.



Notice the lines looping around with jagged edges. Around the word "FRANKLIN," we see a loop labeled 1300. This tells us that on that line, the elevation is 1300ft.

This technique is a way of understanding the dependent variable as a height. To understand how the dependent variable changes with respect to the independent variables, we will consider the level curves.

Consider the level curves of the plane in the previous example. What vector points toward the steepest ascent? What vector points toward the steepest descent? As you can see from the image below, both directions are perpendicular to the level curves. Therefore, any vector perpendicular to the level curve is pointing in the direction of the largest change. For the same reason, any vector that is parallel to a level curve points toward the direction of no change.



In more general terms, when we consider the level curves of an arbitrary function, the largest changes in the dependent variable are experienced in directions perpendicular to the level curves. No changes in height are experienced along the level curve.

Summary of Ideas: Lecture 11

- The **level curves** of a multivariate function are the lines for various values of the dependent variable f.
- Drawing level curves is a technique for graphing three-dimensional surfaces.
- The directions of steepest ascent and descent are perpendicular to the level curves.
- Directions that are parallel to level curves are where the heights do not change.

Objectives

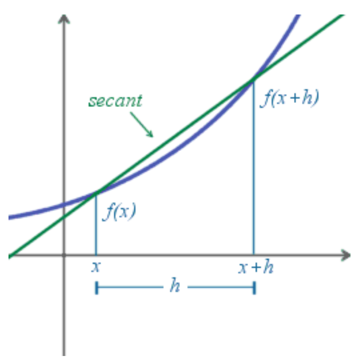
- ☐ I can approximate a partial derivative from raw data.
- ☐ I understand the notation for partial derivatives.
- ☐ I can identify if a partial derivative is increase of decreasing by its image.
- ☐ I know how to take a partial derivative with respect to a variable.

At the end of the previous section, we discussed directions relative to a level curve and the extent to which the dependent variable changes. Before we can discuss how to properly define these arrows (which are vectors), we need to understand *how much* the height change relative to each independent variable. This is information we get from a derivative.

When you studied derivatives in Math 110, you learned that they tell us the extent to which the dependent variable changes with respect to the independent variable. The limit definition of a derivative was

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

This limit is constructed by finding the slope of a secant line from x to $x+h$. That is, the numerator comes from the change in y and the denominator comes from the change in x.



In real life, we often don't have a function that is modeling reality. Instead, we are trying to determine a rate of change (a derivative) from raw data. When this is the case, we cannot find the true derivative. Instead, we *approximate* that derivative using a secant line. Let's consider the data set below. It describes the *humidex*, a function that tells you how hot it *feels* given the actual temperature (T) and humidity (H).

Relative humidity (%)

Actual temperature (°F)

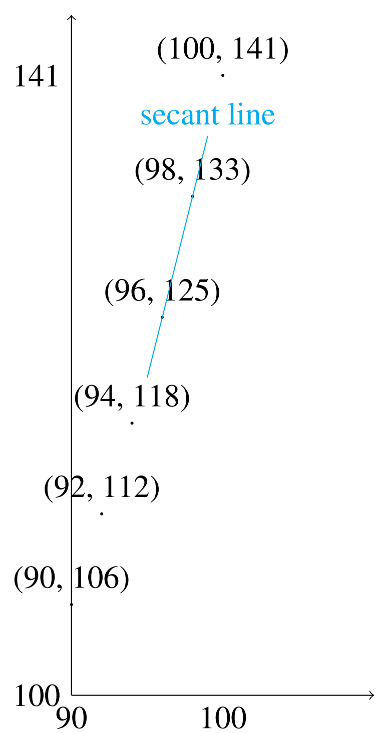
T \ H	50	55	60	65	70	75	80	85	90
90	96	98	100	103	106	109	112	115	119
92	100	103	105	108	112	115	119	123	128
94	104	107	111	114	118	122	127	132	137
96	109	113	116	121	125	130	135	141	146
98	114	118	123	127	133	138	144	150	157
100	119	124	129	135	141	147	154	161	168

What is the derivative in higher dimensions?

$$g(T) = f(T, 70)$$

$$g'(96) = \lim_{h \to 0} \frac{g(96 + h) - g(96)}{h} = \lim_{h \to 0} \frac{f(96 + h, 70) - f(96, 70)}{h}$$

Suppose we wanted to understand the rate at which temperature increases relative to humidity at $96°F$ and 70% humidity. That means, we want to understand $g'(96)$ where $g(T) = f(T, 70)$. Recall from the previous section that when we plug in a constant for one variable, we are actually considering a two-dimensional function. Hence, we can graph it and understand it. In the case of $g(T)$, we do not know its actual function, but we can plot its data.



We could try to fit the data to a function and find that function's derivative. A way to estimate the derivative directly is to approximating using the secant line that connects $g(96)$ with the next available data point, $g(98)$. That is,

$$g'(96) \approx \frac{g(x+h) - g(x)}{h} = \frac{133 - 125}{2} = 4$$

So the derivative is approximately 4. That means when the temperature is $96°F$ and the humidity is 70%, then the temperature feels like it increases by 4 degrees for every one-degree increase in T, the actual temp.

What did we do? We found derivative with respect to only one independent variable while keeping the other constant. We call this a **partial derivative**.

Definition 12.1: A **partial derivative** is a derivative taken with respect to one independent variable, treating all other independent variables as constants.

To denote the specific derivative, we use subscripts. For example, the derivative of f with respect to x is denoted f_x.

What does it mean when we take a partial derivative? A partial derivative is the slope of a a tangent line that only changes with respect to one variable. In the image below, the slope of T_1 is the partial derivative with respect to x at (a, b, c). This is denoted by $f_x(a, b)$. The slope of T_1 is $f_y(a, b)$.

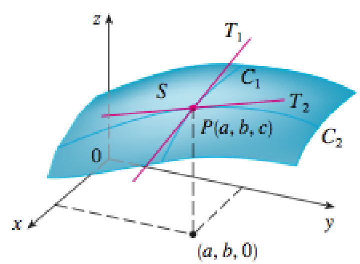


FIGURE 1
The partial derivatives of f at (a, b) are the slopes of the tangents to C_1 and C_2.

How can we tell if the slope is positive or negative? The arrows of the axes point in the direction that the variable is increase. As y increases, T_2 rises. Therefore, $f_y(a, b) > 0$. As x increases, T_1 falls. Therefore $f_x(a, b) < 0$.

Algebraically, finding a partial derivative means taking a derivative with respect to *only one* variable, and treating all others as constants.

Let's look at some examples of calculating partial derivatives from functions.

For the following examples, the color blue will indicate a portion of the function that is treated as a constant. Think of these portions as being numbers. The portions that have changed (because of a derivative) are in red.

Example 12.2:1. Find the first partial derivatives of the function

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

Since there is only two variables, there are two first partial derivatives. First, let's consider f_x. In this case, y is fixed and we treat it as a constant. So, e^{-y} is just a constant.

$$\boxed{f_x(x, y) = -\frac{2}{x^3}e^{-y}}$$

Now, find f_y. Here, x is fixed so $\frac{1}{x^2}$ is just a constant.

$$\boxed{f_y(x,y) = -e^{-y}\frac{1}{x^2}}$$

Example 12.3:2. Find the first partial derivatives of the function

$$f(x,y) = x^4y^3 + 8x^2y$$

Again, there are only two variables, so there are only two partial derivatives. They are

$$\boxed{f_x(x,y) = 4x^3y^3 + 16xy}$$

and

$$\boxed{f_y(x,y) = 3x^4y^2 + 8x^2}$$

Summary of Ideas: Lecture 12

- Derivatives can be approximated using the limit definition. That is
$$f'(x) \approx= \frac{f(x+h) - f(x)}{h}$$
- We can determine if a partial derivative is positive or negative by considering the change in the dependent variable as the relevant independent variable increases.
- A partial derivative with respect to a variable, takes the derivative of the function with respect to <u>that variable</u> and treats <u>all other variables as constants.</u>

Objectives

- ☐ I can use Clairaut's Theorem to make my calculations easier.
- ☐ I can take higher derivatives.
- ☐ I can check if a function is a solution to a partial differential equation.

Higher order derivatives are calculated as you would expect. We still use subscripts to describe the second derivative, like f_{xx} and f_{yy}. These describe the concavity of f in the x and y direction, just like in Math 110.

Interestingly, we can get mixed second derivatives like f_{xy} and f_{yx}. These tell us how f_x changes with y and how f_y changes with x. Even more interesting is the fact that these are both equal!

Theorem 13.1. (Clairaut's Theorem)

$$f_{xy} = f_{yx}$$

What's the point of knowing this theorem? It means that you can switch the order of derivatives based on whatever would be easiest.

Clairaut's Theorem extends to higher derivatives. If we were looking at taking two derivatives with respect to x and one with respect to y, we would have three possible ways to do this

$$f_{yxx} = f_{xyx} = f_{xxy}$$

Example 13.2:1. Find f_{xxx}, f_{xyx} for

$$f(x, y) = (2x + 5y)^7$$

Solution 13.3: Let's begin by finding f_x and use that to find f_{xx} and f_{xxx}

$$f_x = 2[7(2x + 5y)^6]$$

Remember that $5y$ is just treated as a constant. Notice that we could work towards finding f_{xyx} by finding f_{xy} from the above equation. If we use Clairaut's Theorem, however, we can skip a step by calculating f_{xxy} instead. Now, let's calculate f_{xx}.

$$f_{xx} = 2 \cdot 2[7 \cdot 6(2x + 5y)^5] = 168(2x + 5y)^5$$

Using f_{xx}, we can find f_{xxx} and f_{xxy}. They are

$$\boxed{f_{xxx} = 2 \cdot 168[5(2x + 5y)^4] = 1680(2x + 5y)^4}$$

and

$$\boxed{f_{xyx} = f_{xxy} = 5 \cdot 168[5(2x + 5y)^4] = 4200(2x + 5y)^4}$$

Example 13.4:2. Find f_{xyz} for

$$f(x, y, z) = e^{xyz^2}$$

Solution 13.5: This is a good example to pay close attention to because it illustrates how complicated these partial derivatives can get.

Let's first find f_x. It is

$$f_x = yz^2 e^{xyz^2}$$

Notice the coefficients. Because y and z are treated as constants, they need to be brought out front by the chain rule. For the next derivative, we will have to use the product rule. What does this tell us? It tells us that it's probably better to take f_z first since we won't get that pesky z^2.

$$f_z = 2zxye^{xyz^2}$$

Notice that taking the derivative with respect to x or y next will result in the same amount of work. Let's just pick x next.

$$f_{zx} = (2zxy)(yz^2 e^{xyz^2}) + (2zy)(e^{xyz^2}) = 2xy^2z^3e^{xyz^2} + 2zye^{xyz^2}$$

The parentheses are in place to indicate how I broke up the variables to take the derivatives. Now let's calculate the last derivative, the partial derivative with respect to y.

$$f_{zxy} = (2z)(e^{xyz^2}) + (2zy)(xz^2e^{xyz^2}) + (2xy^2z^3)(xz^2e^{xyz^2}) + (4xyz^3)(e^{xyz^2})$$

After we simplify, we get the final answer

$$\boxed{f_{zxy} = 2ze^{xyz^2}\left[1 + 3xyz^2 + x^2y^2z^4\right]}$$

You may have heard of **partial differential equations**. These are equations that use derivatives of an unknown function as variables. The goal is to try to figure out the original function. For example, our understanding of waves is based on partial differential equations. Specifically, we look at something called **the wave equation**

$$u_{tt} = a^2 u_{xx}.$$

Learn More About This

Where does the wave equation come from, you ask? You can find out more about the wave equation and how it is derived by watching this video: `https://www.youtube.com/watch?v=ck-r_qmNNG0`.
Towards the end of this course, we'll discuss how to solve systems of differential equations. If time permits, we will discuss some partial differential equations like the wave equation. If you'd like to learn more about differential equations, I recommend the videos made by the Khan Academy. You can find those here: `https://www.khanacademy.org/math/differential-equations`

Example 13.6:3. Determine if $u = x^2 + (at)^2$ is a solution to the wave equation

$$u_{tt} = a^2 u_{xx}$$

for $t > 0$ and all values of x.

Solution 13.7: To do this, we need to find u_{tt} and u_{xx} and show that the equation holds.

$$\begin{aligned} u_t &= a[2(at)] = 2a^2t \\ \implies u_{tt} &= 2a^2 \end{aligned}$$

$$\begin{aligned} u_x &= 2x \\ \implies u_{xx} &= 2 \end{aligned}$$

Plugging into the wave equation, we get

$$\begin{aligned} \left[2a^2\right] &= a^2\left[2\right] \\ \implies 2a^2 &= 2a^2 \end{aligned}$$

Since our result is trivially true, then we know $u = x^2 + (at)^2$ is a solution to the wave equation.

Example 13.8:3. Show that $u = a^2x^2\ln(t)$ is **not** a solution to the heat equation

$$u_t = a^2u_{xx}$$

for $t > 0$ and all values of x.

Solution 13.9: To do this, we need to find u_t and u_{xx} and show that the equation holds.

$$u_t = \frac{a^2x^2}{t}$$

$$\begin{aligned} u_x &= 2xa^2\ln(t) \\ \implies u_{xx} &= 2a^2\ln(t) \end{aligned}$$

Plugging into the heat equation, we get

$$\begin{aligned} \frac{a^2x}{t} &= 2a^2\ln(t) \\ \frac{x}{t} &= 2\ln(t) \end{aligned}$$

This statement is FALSE. This will not hold for all values of x and t.

Summary of Ideas: Lecture 13

- We can take a partial derivative more than once. The number of times we take a partial derivative determines the order of the derivative.
- The order in which we take partial derivatives does not matter. That is, $f_{xyz} = f_{yzx} = f_{zyx} = f_{yxz} = f_{zxy} = f_{xzy}$.
- We can determine if a function is a solution to a partial differential equation by plugging it into the equation.

Objectives

- ☐ I understand what a gradient vector is and what it tells you.
- ☐ I can use the gradient to identify important features of a surface like steepest ascent.

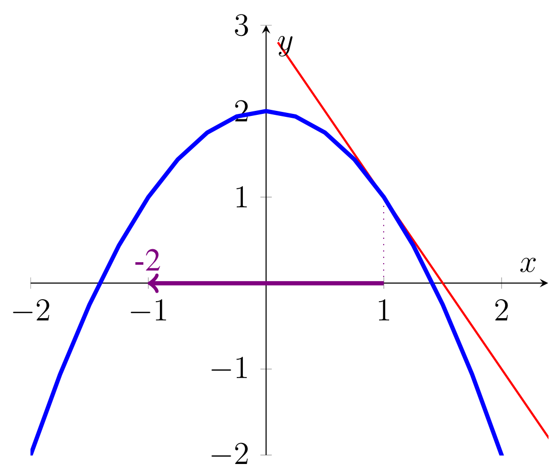
In lecture 11, we learned the direction of *steepest* change occurred perpendicularly to the level curves. But how much change can we see? What is this direction precisely? In this lecture, we will determine that direction.

The derivative of a function $f(x)$ at a point x is the slope of the tangent line of f at x. Loosely, we might say that it is the slope of f at that point x. As we will see, the derivative will also tell us a direction of steepest ascent.

When we are thinking of a two-dimensional function, there are only two choices–left or right–and the sign of the derivative tells us this direction.

For example, if you're given the function $f(x) = -x^2+2$, you know the derivative of f at $x = 1$ will be -2. If we represented this derivative as a vector (an arrow) on the x-axis, then

(1) it points in the direction of **steepest ascent** and
(2) it's length represents the steepness of the incline.



For higher dimensions, the approach is similar. We want to construct vectors that point in the direction of steepest ascent whose length represents the steepness of the incline. These vectors are constructed from partial derivatives! This vector is called the gradient vector.

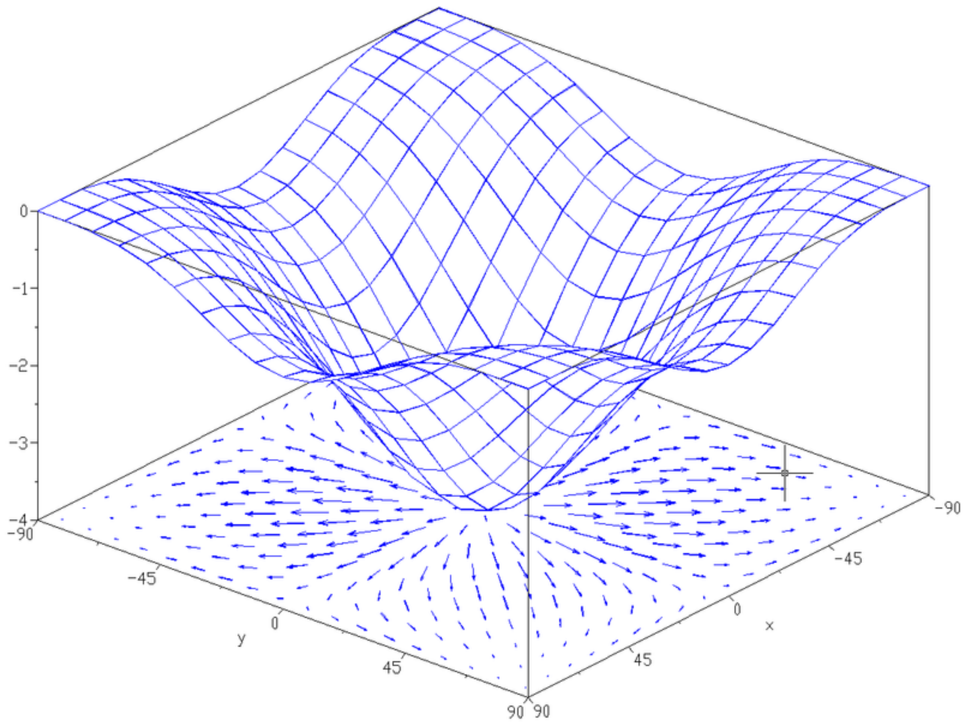
Definition 14.1: The **gradient vector** of a function f, denoted ∇f or $grad(f)$, is a vectors whose entries are the partial derivatives of f. That is,

$$\nabla f(x,y) = \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

For higher dimensions, we have

$$\nabla f(x_1, \ldots, x_n) = \begin{pmatrix} f_{x_n} \\ \vdots \\ f_{y_n} \end{pmatrix}$$

Here is a picture of a three dimensional surface. The arrows at the bottom represent the gradient vectors. They each point in the direction where it is steepest. Longer the vector (the greater its magnitude), the steeper the surface is at that point.



As we will see in the examples below, the gradient vector is always perpendicular to some level curve.

Example 14.2: Find the gradient vector of

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

What are the gradient vectors at $(1, 2)$, $(2, 1)$ and $(0, 0)$?

Solution 14.3: We begin with the formula.

$$\nabla f = \begin{pmatrix} f_x \\ f_y \end{pmatrix} = \boxed{\begin{pmatrix} 2x \\ 2y \end{pmatrix}}$$

Now, let us find the gradient at the following points.

- $\nabla f(1, 2) = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$

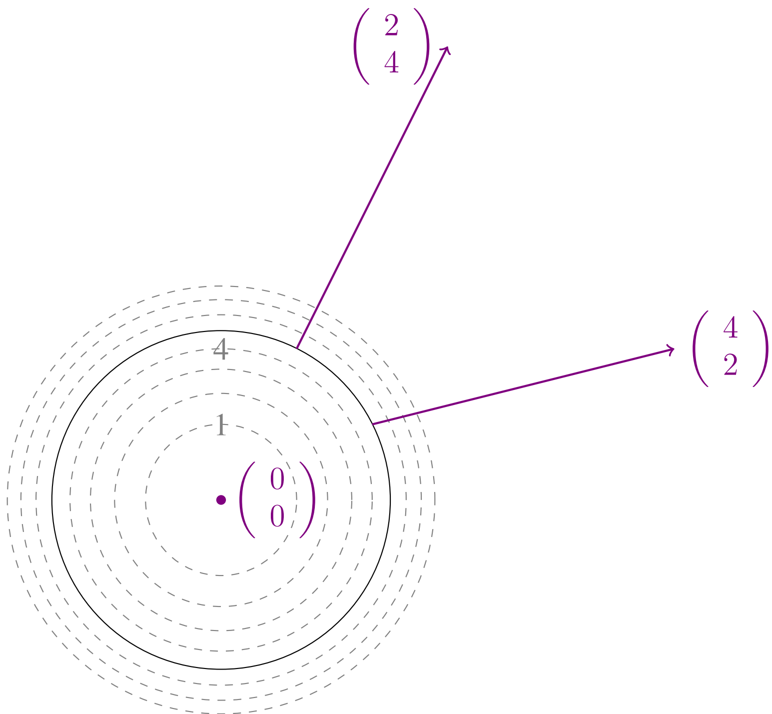
- $\nabla f(2, 1) = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

- $\nabla f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

How steep are these gradients? Let's calculate those lengths.

- $|\nabla f(1,2)| = \sqrt{2^2 + 4^2} = \sqrt{20} = 2\sqrt{5}$
- $|\nabla f(2,1)| = \sqrt{4^2 + 2^2} = \sqrt{20} = 2\sqrt{5}$
- $|\nabla f(0,0)| = \sqrt{0^2 + 0^2} = \sqrt{0} = 0$

Now, let's verify that these vectors are perpendicular to their level curve. Notice that $(1,2)$ corresponds to the level curve of height $f(1,2) = 1^2 + 2^2 = 5$. Similarly, $f(2,1) = 5$.



Example 14.4: Find the gradient vector of

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

What are the gradient vectors at $(1,1)$, $(0,-1)$ and $(0,0)$?

Solution 14.5:

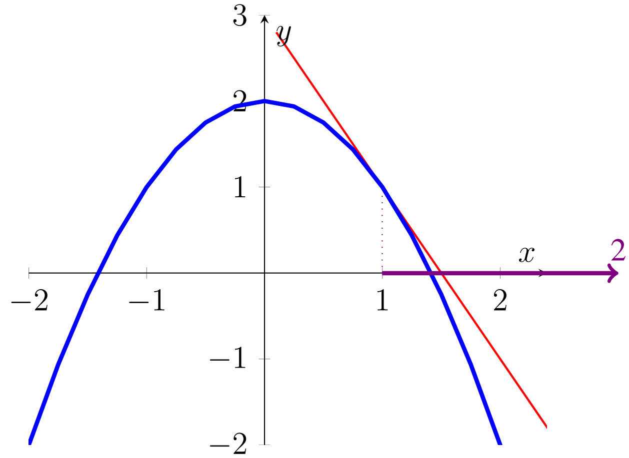
$$\nabla f = \begin{pmatrix} f_x \\ f_y \end{pmatrix} = \boxed{\begin{pmatrix} 2y + 2x \\ 2x + 1 \end{pmatrix}}$$

Now, let us find the gradient at the following points.

- $\nabla f(1,1) = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$
- $\nabla f(0,-1) = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

- $\nabla f(0,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Suppose we want to identify the direction of steepest descent? In the two-dimensional case, we simply switch the sign of the derivative. For example, if we return to the function $f(x) = -x^2 + 2$, the arrow of steepest descent points in the positive direction and is length 2.



For higher dimensions, it is the same. That is, $-\nabla f$ always points in the direction of steepest descent.

Example 14.6: Find the direction of steepest ascent

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

at $(1,1)$, $(0,-1)$ and $(0,0)$? How steep are they?

Solution 14.7: The directions of steepest ascent are

- $-\nabla f(1,1) = \begin{pmatrix} -4 \\ -3 \end{pmatrix}$
- $-\nabla f(0,-1) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$
- $-\nabla f(0,0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

The derivatives are

- $|-\nabla f(1,1)| = \sqrt{(-4)^2(-3)^2} = \sqrt{25} = 5$
- $|-\nabla f(0,-1)| = \sqrt{2^2(-1)^2} = \sqrt{5}$
- $|-\nabla f(0,0)| = \sqrt{0^2 + (-1)^2} = \sqrt{1} = 1$

The next obvious question is this:

How do we find how steep a surface is in other directions? To do that, we will need to understand *projections*. We will use the dot product to project the gradient vector in a direction to understand how steeply the surface is increasing/decreasing in that direction.

Summary of Ideas: Lecture 14

- The **gradient vector** of a function f, denoted ∇f or $grad(f)$, is a vectors whose entries are the partial derivatives of f.
$$\nabla f(x,y) = \begin{pmatrix} f_x(x,y) \\ f_y(x,y) \end{pmatrix}$$
It is the generalization of a derivative in higher dimensions.
- The gradient points in the direction of **steepest ascent**.
- $-\nabla f$ points in the direction of **steepest descent**.

Objectives

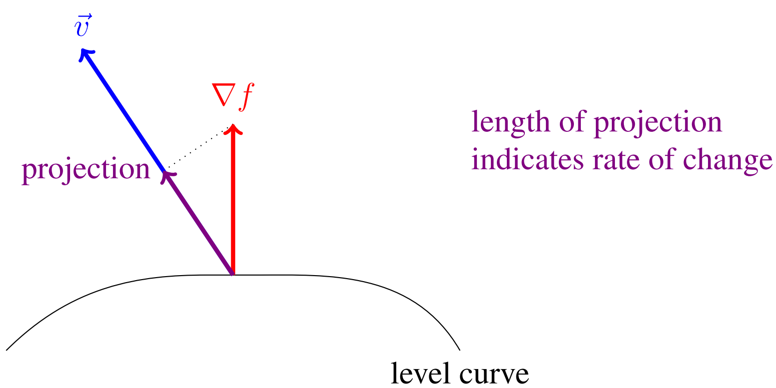
- ☐ I can calculate the dot product of two vectors and interpret its meaning.
- ☐ I can find the projection of one vector onto another one.

In the last few lectures, we've learned that

- vectors perpendicular to a level curve point in the direction of steepest ascent or descent;
- the gradient vector, ∇f, points in the direction of steepest ascent, and its negative, $-\nabla f$ points in the direction of steepest descent; and
- the magnitude of the gradient, $|\nabla f|$ tells us the rate at which the dependent variable increases in the direction of ∇f.

This last point can be rephrased as the derivative of f in the direction of ∇f is $|\nabla f|$.

What about in other directions? What's the derivative of f a point in a direction $\vec{v}$, which different from ∇f and $-\nabla f$? The answer to this question is: *the projection of ∇f onto $\vec{v}$ and find the magnitude of that projection.* This is depicted below.



To understand this, we'll need to first learn about projections. Before we begin, let's take a moment to recall the following facts:

- The **dot product** of two vectors, $\vec{v}$ and $\vec{w}$, is obtained by taking the sum of the product of the entries

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_1 w_2 + v_3 w_3 + \ldots$$

- A **unit vector** is a vector with magnitude equal to one.

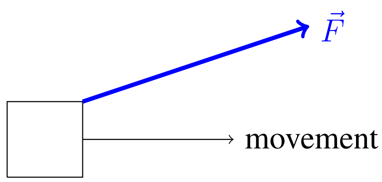
- To find a unit vector $\vec{u}$ that points in the *same* direction as an arbitrary vector $\vec{v}$, we must divide by the vector's magnitude. That is,

$$\vec{u} = \frac{\vec{v}}{|\vec{v}|} = \begin{pmatrix} \frac{v_1}{|\vec{v}|} \\ \frac{v_2}{|\vec{v}|} \\ \frac{v_3}{|\vec{v}|} \\ \vdots \end{pmatrix}$$

15.1. **Dot Products.** Because vectors describe movement, we might ask the question:

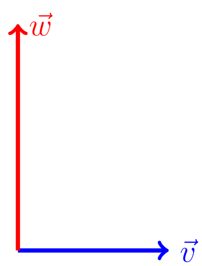
"To what extent do two vectors move together?"

Consider a box being dragged horizontally by a force applied at an angle. Both its movement and its force are described using vectors. We are interested in the amount these two vectors share the same direction.

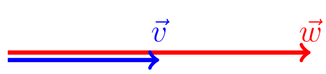


We will use this notion to understand the **dot product**. This will be a scalar value.[8] Before we go into the formulas, let's talk a little bit about this idea of moving together.

If I have two perpendicular vectors, they don't move in the same direction at all. These vectors are referred to as **orthogonal**. We will represent the amount they move together with 0. That is, $\vec{w} \cdot \vec{v} = 0$.



Parallel vectors move entirely in the same direction. We will represent the amount they move together with the product of their magnitudes. That is, $\vec{w} \cdot \vec{v} = |\vec{w}||\vec{v}|$.



From these two examples, we can see that the angle between the two vectors plays a part. Since $0°$ corresponds to 1 and $90°$ corresponds to 0, we can deduce that our equation for the dot product is

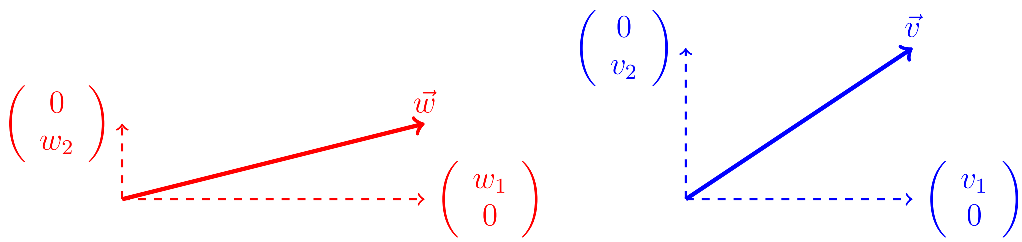
$$\vec{w} \cdot \vec{v} = |\vec{w}||\vec{v}| \cos\theta$$

[8] In fact, every notion of a measurement will be a scalar value.

where θ is the angle between them. From a few lectures previous, we know there exists another formula. For vectors $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$,

$$\vec{w} \cdot \vec{v} = w_1 v_1 + w_2 v_2$$

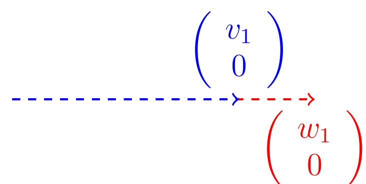
How might these two formulas be related? Let's think about the components of each vector.



Then we can FOIL out the separate components.

$$\vec{w} \cdot \vec{v} = \begin{pmatrix} w_1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \begin{pmatrix} w_1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ v_2 \end{pmatrix} + \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ v_2 \end{pmatrix} = w_1 v_1 + 0 + 0 + w_2 v_2$$

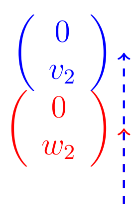
For the components that are parallel, we will get the product of the magnitudes. For the components that are perpendicular, we will get zero. Let's look at each pairing of vectors to see which is which.



$$\begin{pmatrix} u_1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ 0 \end{pmatrix} = w_1 v_1$$

$$\begin{pmatrix} w_1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ v_2 \end{pmatrix} = 0$$

$$\begin{pmatrix} 0 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ 0 \end{pmatrix} = 0$$



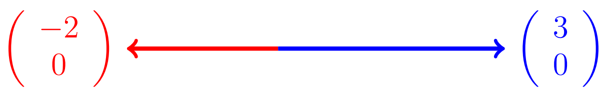
$$\begin{pmatrix} 0 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ v_2 \end{pmatrix} = w_2 v_2$$

Basically, when we break up a vector by its respective components and look at the dot product, we see why these two equations agree. The perpendicular components will have a zero dot product while the parallel components will only look at the product.

What happens when two parallel components point in opposite directions? For example, when

$$\begin{pmatrix} -2 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

Then the dot product can be expressed as either $|-2||3|\cos 180° = -6$ and $(0) + (-2)(3) = -6$. Where did this $180°$ come from? If the vectors are parallel, then isn't the angle between them zero? No, it turns out the angle is measure when we place vectors from end to end. Remember, these objects have no location, so we can shift them anywhere.



15.2. **Projections.** With the use of dot products, we can talk about projections. There are two notions of a projection:

- a scalar projection
- a vector projection

The **scalar projection** of $\vec{a}$ onto $\vec{b}$ indicates the amount $\vec{a}$ moves in the particular direction of $\vec{b}$. It is denoted $\text{comp}_{\vec{b}}\vec{a}$. The formula is

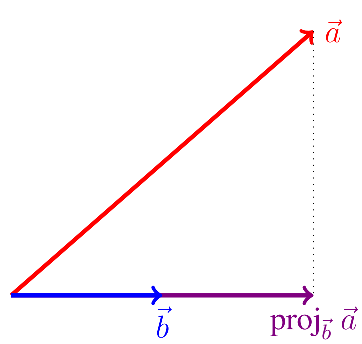
$$\text{comp}_{\vec{b}}\vec{a} = \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|}$$

Why does this formula make sense? The dot product tells us the amount in which the two vectors move in the same direction. Because we only want a sense of how much $\vec{a}$ moves, we divide by the magnitude of $|\vec{b}|$.

The **vector projection** of $\vec{a}$ onto $\vec{b}$ is a vector representative of the amount $\vec{a}$ moves in the direction of $\vec{b}$. It is denoted $\text{proj}_{\vec{b}}\,\vec{a}$. The formula is $\text{comp}_{\vec{b}}\,\vec{a}$ times the unit vector in the direction of $\vec{b}$.

$$\text{proj}_{\vec{b}}\,\vec{a} = (\text{comp}_{\vec{b}}\,\vec{a})\frac{\vec{b}}{|\vec{b}|} = \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|}\frac{\vec{b}}{|\vec{b}|} = \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|^2}\,\vec{b}$$

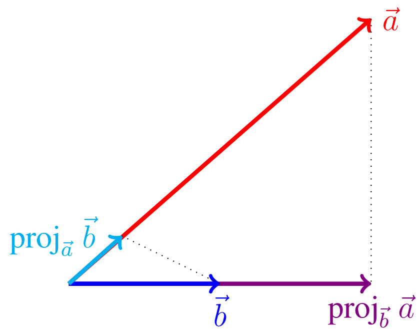
Notice that this is just the scalar projection multiplied by the unit vector in the direction of $\vec{b}$. This makes sense because the vector projection of $\vec{a}$ onto $\vec{b}$ should be a vector in the direction of $\vec{b}$ whose magnitude reflects the amount at which $\vec{a}$ points in the direction of $\vec{b}$.



One more thing to point out is that

$$\text{proj}_{\vec{b}}\,\vec{a} \neq \text{proj}_{\vec{a}}\,\vec{b}.$$

Let's look at a picture of the projection of $\text{proj}_{\vec{a}}\,\vec{b}$ to illustrate this fact.



The magnitude of our projection is $\text{comp}_{\vec{a}}\,\vec{b}$, so we can also deduce from this illustration that

$$\text{comp}_{\vec{b}}\,\vec{a} \neq \text{comp}_{\vec{a}}\,\vec{b}.$$

In other words, order matters!

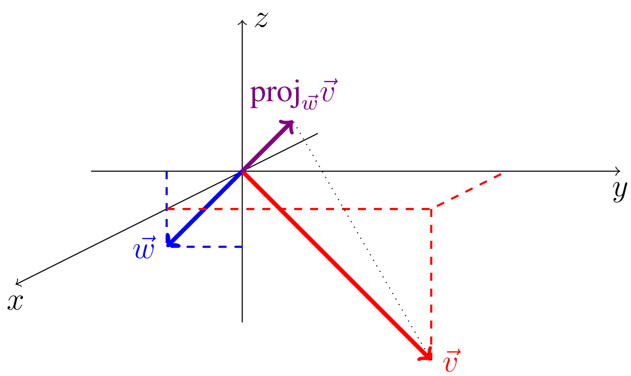
Example 15.1: Find the scalar and vector projections of $\vec{v}$ onto $\vec{w}$ where

$$\vec{v} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \text{ and } \vec{w} = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}$$

The question asks for $\vec{v}$ onto $\vec{w}$, so we're looking to find $\text{comp}_{\vec{w}}\vec{v}$ and $\text{proj}_{\vec{w}}\vec{v}$. When you sit down to memorize the formulas, remember that the vector you project onto will be the one that appears the most.

$$\text{comp}_{\vec{w}}\vec{v} = \frac{\vec{v}\cdot\vec{w}}{|\vec{w}|} = \frac{(1)(0)+(3)(-1)+(-2)(-1)}{\sqrt{0^2+(-1)^2+(-1)^2}} = \boxed{-\frac{1}{\sqrt{2}}}$$

The scalar projection reflects the amount in which $\vec{v}$ travels in the direction of $\vec{w}$. If the value is negative, it tells me that the vector $\text{proj}_{\vec{w}}\vec{v}$ is traveling in the opposite direction to $\vec{w}$.



Now, let's describe the vector pictured.

$$\text{proj}_{\vec{w}}\vec{v} = -\frac{1}{\sqrt{2}} \frac{\begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}}{\sqrt{2}} = -\frac{1}{2} \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} = \boxed{\left(\ 0, \tfrac{1}{2}, \tfrac{1}{2}\ \right)}$$

Summary of Ideas: Lecture 15

- The **dot product** is a measurement of how much two vectors move together. We have two formulas for it.

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos\theta \quad \text{and} \quad \vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + u_3 v_3$$

- We can also use the dot product to measure the amount one vector ($\vec{a}$) moves in the direction of another vector ($\vec{b}$):

$$\text{comp}_{\vec{b}}\vec{a} = \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|}$$

  This is called the **scalar projection**.
- To find the **vector projection**, we take the scalar projection and multiply by the unit vector we are projecting onto.

$$\text{proj}_{\vec{b}}\vec{a} = \text{comp}_{\vec{b}}\vec{a} \frac{\vec{b}}{|\vec{b}|}$$

Objectives

- ☐ I understand how to find the rate of change in any direction.
- ☐ I understand in what direction the maximum rate of change happens.

So far, we've learned the definition of the gradient vector and we know that it tells us the direction of steepest ascent. Its magnitude indicates the rate of change of the dependent variable in the direction of the gradient.

A natural question to ask is, "What is the rate of change of the dependent variable in the direction of an arbitrary vector $\vec{v}$?" In other words, how fast does the surface ascend in the direction of $\vec{v}$?

The rate of change in the direction of $\vec{v}$ is called the *directional derivative*. That is the *scalar projection of the gradient onto* $\vec{v}$.

Definition 16.1: The **directional derivative**, denoted $D_v f(x,y)$, is a derivative of a multivariable function in the direction of a vector $\vec{v}$. It is the scalar projection of the gradient onto $\vec{v}$.

$$D_v f(x,y) = \text{comp}_v \nabla f(x,y) = \frac{\nabla f(x,y) \cdot \vec{v}}{|\vec{v}|}$$

It's best to understand concepts with a picture. So let's draw one. Consider the function

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

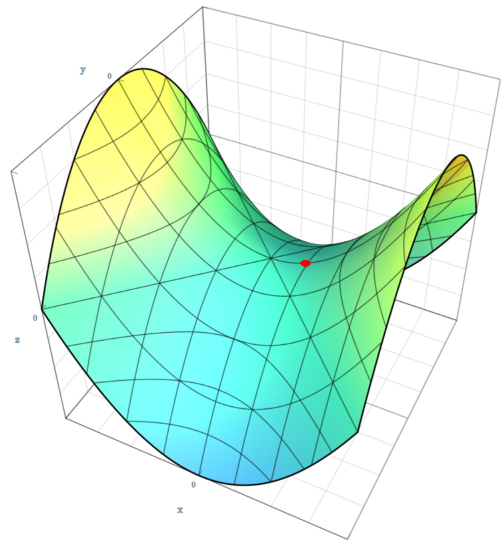
The gradient of f is

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix}.$$

At the point $(1,0)$, the direction of steepest ascent is

$$\nabla f(1,0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

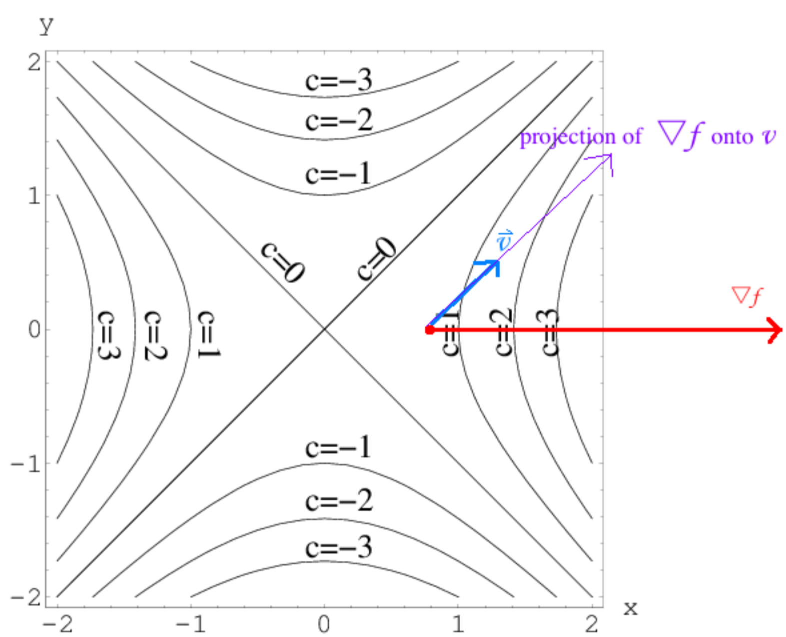
In that direction, f has a slope of $|\nabla f(1,0)| = \sqrt{(2)^2} = 2$.



What is the slope at $(1,0)$ in the direction of

$$\vec{v} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}?$$

$$D_v f(x,y) = \text{comp}_v \nabla f(1,0) = \frac{\begin{pmatrix} 2 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}}{\sqrt{0^2+1^2}} = \frac{1}{1} = 1$$



Let's look at some examples.

Example 16.2: Find the directional derivative of

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

in the direction of $\vec{v} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ at the point $(1,2)$.

First, we find the gradient.

$$f_x(x,y) = \frac{d}{dx}\left(\frac{x}{x^2+y^2}\right)$$

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

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

$$f_y(x,y) = \frac{d}{dy}\left(\frac{x}{x^2+y^2}\right)$$

$$= \frac{-2xy}{(x^2+y^2)^2}$$

The gradient is then

$$\nabla f(1,2) = \begin{pmatrix} \frac{4-1}{(4+1)^2} \\ \\ -\frac{4}{(4+1)^2} \end{pmatrix} = \begin{pmatrix} \frac{3}{25} \\ \\ -\frac{4}{25} \end{pmatrix} = \frac{1}{25}\begin{pmatrix} 3 \\ -4 \end{pmatrix}$$

We now find the magnitude of $\vec{v}$. We get

$$|\vec{v}| = \sqrt{9+25} = \sqrt{34}$$

The directional derivative is then

$$D_v f(1,2) = \frac{\nabla f(1,2)\cdot\vec{v}}{|\vec{v}|} = \frac{1}{25\sqrt{34}}\begin{pmatrix} 3 \\ -4 \end{pmatrix}\cdot\begin{pmatrix} 3 \\ 5 \end{pmatrix} = \frac{1}{25\sqrt{34}}(9-20) = \boxed{-\frac{11}{25\sqrt{34}}}$$

Example 16.3: Find the directional derivative of

$$f(x,y,z) = \sqrt{xyz}$$

in the direction of $\vec{v} = \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix}$ at the point $(3,2,6)$.

First, we find the partial derivatives to define the gradient.

$$f_x(x,y,z) = \frac{yz}{2\sqrt{xyz}}$$
$$f_y(x,y,z) = \frac{xz}{2\sqrt{xyz}}$$
$$f_z(x,y,z) = \frac{xy}{2\sqrt{xyz}}$$

The gradient is

$$\nabla f(3,2,6) = \begin{pmatrix} \frac{12}{2(6)} \\ \\ \frac{18}{2(6)} \\ \\ \frac{6}{2(6)} \end{pmatrix} = \begin{pmatrix} 1 \\ \\ \frac{3}{2} \\ \\ \frac{1}{2} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

The magnitude of $\vec{v} = \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix}$ is

$$|\vec{v}| = \sqrt{1+4+4} = 3$$

Therefore, the directional derivative is

$$D_v f(3,2,6) = \frac{\nabla f(3,2,6) \cdot \vec{v}}{|\vec{v}|} = \frac{1}{3(2)}\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix} = \frac{1}{6}(-2-6+2) = \boxed{-1}$$

The next natural question is:

In what direction is the derivative maximum?

As we just saw, the directional derivative is calculated by taking the scalar projection of ∇f onto a vector $\vec{v}$. Define θ be the angle between $\vec{v}$ and ∇f. Then,

$$\frac{\nabla f \cdot \vec{v}}{|\vec{v}|} = \frac{|\nabla f||\vec{u}|\cos(\theta)}{|\vec{v}|} = |\nabla f|\cos(\theta)$$

This is maximized if $\theta = 0$. From this, we know the following:

- The maximum rate of change (the largest directional derivative) is $|\nabla f|$.
- This occurs when $\vec{v}$ is parallel to ∇f, i.e. when they point in the same direction.

That makes sense since ∇f is the vector pointing toward steepest ascent, so it should be the direction with the largest derivative.

Observe, also that...

- No change occurs when $\theta = 90°$ or when $\theta = -90°$. In other words, directions perpendicular to the gradient are constant height.
- The rate of steepest *descent* happens when $\theta = 180°$. It's rate of change is $-|\nabla f|$,

Let's look at two examples.

Example 16.4: Find the maximum rate of change of f at the given point and the direction in which it occurs.

$$f(s,t) = te^{st}, \quad \text{at } (0,2)$$

The maximum rate of change is $|\nabla f(0,2)|$. Let's first find the gradient.

$$\nabla f = \begin{pmatrix} te^{st} \\ ste^{st} + e^{st} \end{pmatrix}$$

Then

$$|\nabla f(0,2)| = \sqrt{(2)^2 + 1^2} = \boxed{\sqrt{5}}$$

The direction is

$$\nabla f(0,2) = \boxed{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}$$

Remark 16.5: For this problem, it may not have been clear which component was the first and which was the second since s and t are atypical variables. For clues about the order, look at how the ordered pairs are defined in the function. It was written as "$f(s,t)$," which tells us our gradient vector should be $\begin{pmatrix} f_s \\ f_t \end{pmatrix}$.

Example 16.6: Find the maximum rate of change of f at the given point and the direction in which it occurs.

$$f(x,y,z) = \sqrt{x^2 + y^2 + z^2}, \quad (3,6,-2)$$

As above, the maximum rate of change is $|\nabla f(3,6,-2)|$.

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2+z^2}} \\ \\ \frac{y}{\sqrt{x^2+y^2+z^2}} \\ \\ \frac{z}{\sqrt{x^2+y^2+z^2}} \end{pmatrix}$$

Then

$$\nabla f(3,6,-2) = \begin{pmatrix} \frac{3}{7} \\ \\ \frac{6}{7} \\ \\ -\frac{2}{7} \end{pmatrix}$$

The $|\nabla f(3,6,-2)| = 1/7\sqrt{9+36+4} = \boxed{1}$

The direction is

$$\boxed{\begin{pmatrix} \frac{3}{7} \\ \\ \frac{6}{7} \\ \\ -\frac{2}{7} \end{pmatrix}}$$

Summary of Ideas: Lecture 16

- The **gradient vector** of a function f, denoted ∇f or $grad(f)$, is a vectors whose entries are the partial derivatives of f.

  $$\nabla f(x, y) = \left(\ f_x(x, y), f_y(x, y) \ \right)$$

  It is the generalization of a derivative in higher dimensions.
- The gradient points in the direction of **steepest ascent**.
- The **directional derivative**, denoted $D_v f(x, y)$, is a derivative of a $f(x, y)$ in the direction of a vector $\vec{v}$. It is the scalar projection of the gradient onto $\vec{v}$.

  $$D_v f(x, y) = \text{comp}_v \nabla f(x, y) = \frac{\nabla f(x, y) \cdot \vec{v}}{|\vec{v}|}$$

  This produces a vector whose magnitude represents the rate a function ascends (how steep it is) at point (x, y) in the direction of $\vec{v}$.
- The **maximum directional derivative** is always $|\nabla f|$.
- This happens in the direction of ∇f

Objectives

☐ Review Partial Derivatives

For Lecture 17, the following exercises were done in class.

17.1. Partial Derivatives.

(1) Find f_x and f_y for
$$f(x,y) = xy^2 - x^3y.$$

To find f_x, we treat x as a variable and y as a constant. Therefore,
$$\boxed{f_x(x,y) = y^2 - 3x^2y}$$

By the same logic,
$$\boxed{f_y(x,y) = 2xy - x^3}$$

(2) Find f_x and f_y for
$$f(x,y) = \ln(x + \sqrt{x^2+y^2}).$$

To find f_x, we treat x as a variable and y as a constant. Here, we will need to use the chain rule twice. The "outside" function is the natural log function. The middle function is x plus the square root function. Therefore,
$$f_x(x,y) = \frac{1}{x+\sqrt{x^2+y^2}} \cdot (1 + \frac{1}{2}(x^2+y^2)^{-\frac{1}{2}} \cdot 2x) = \boxed{\frac{1 + x(x^2+y^2)^{-\frac{1}{2}}}{x+\sqrt{x^2+y^2}}}$$

The approach for f_y is the same, except x is treated as a constant.
$$f_y(x,y) = \frac{1}{x+\sqrt{x^2+y^2}} \cdot (0 + \frac{1}{2}(x^2+y^2)^{-\frac{1}{2}} \cdot 2y) = \boxed{\frac{y(x^2+y^2)^{-\frac{1}{2}}}{x+\sqrt{x^2+y^2}}}$$

(3) Find f_x, f_y, and f_z for

$$f(x, y, z) = xz - 5x^2y^3z^4.$$

$$\boxed{f_x(x, y) = z - 10xy^3z^4}$$

$$\boxed{f_y(x, y) = 0 - 15x^2y^2z^4}$$

$$\boxed{f_z(x, y) = x - 20x^2y^3z^3}$$

(4) Find f_{xxx} and f_{xyx} for

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

$$f_x(x, y) = 4x^3y^2 - 3x^2y$$

$$f_{xx}(x, y) = 12x^2y^2 - 6xy$$

$$f_{xxx}(x, y) = \boxed{24xy^2 - 6y}$$

$$f_{xyx}(x, y) = f_{xxy}(x, y) = \boxed{24x^2y - 6x}$$

(5) Find z_{uvw} for

$$z = u\sqrt{v-w}.$$

$$\begin{aligned} z_u &= \sqrt{v-w} \\ z_{uv} &= \frac{1}{2}(v-w)^{-\frac{1}{2}} \cdot 1 \\ z_{uvw} &= -\frac{1}{4}(v-w)^{-\frac{3}{2}} \cdot (-1) = \boxed{\frac{1}{4}(v-w)^{-\frac{3}{2}}} \end{aligned}$$

(6) Find w_{zyx} and w_{xxy} for

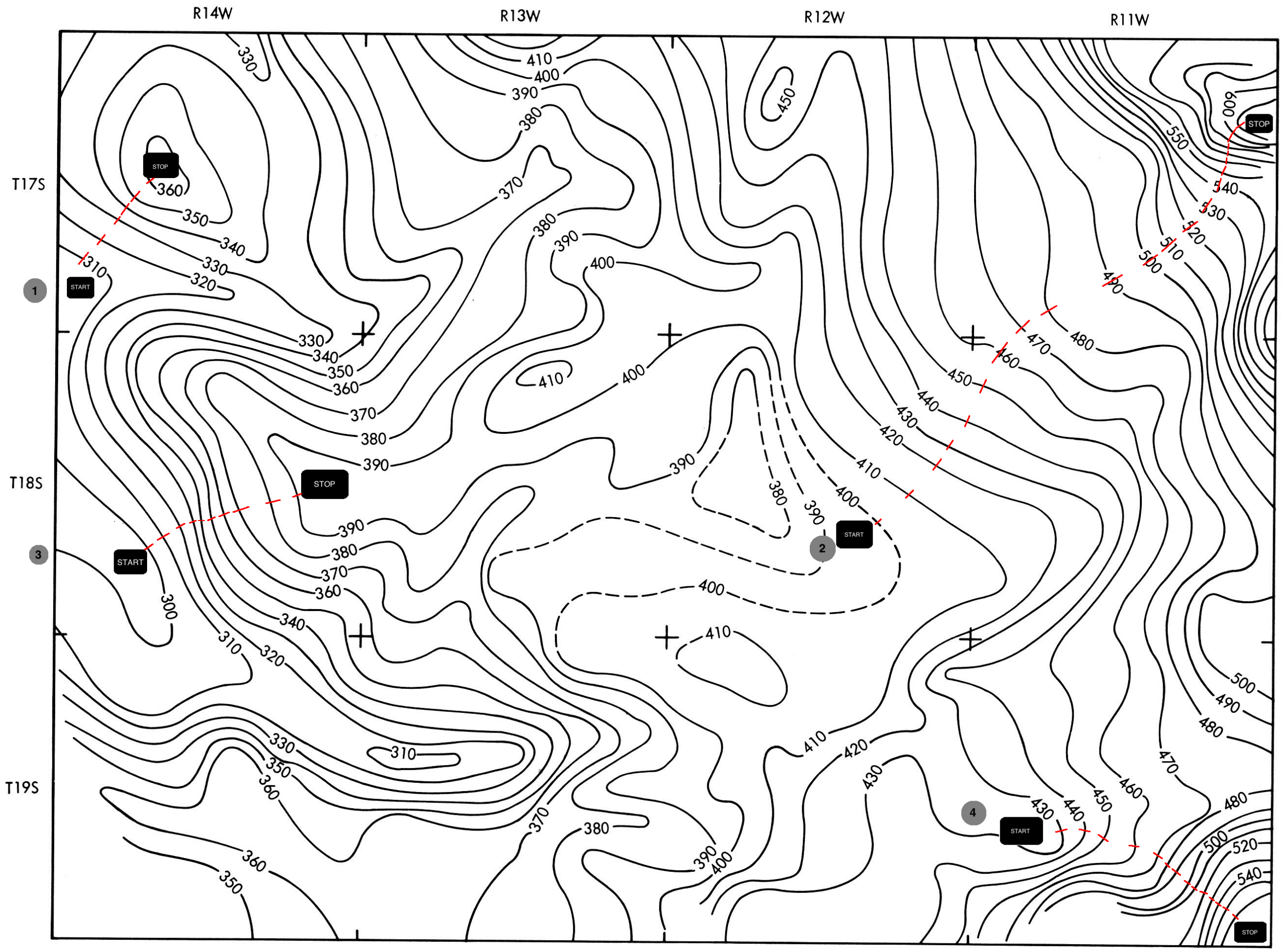
$$w = \frac{x}{y+2z}.$$

$$\begin{aligned} w_x &= \frac{1}{y+2z} \\ w_{xy} &= (-1)(y+2z)^{-2} \cdot 1 \\ w_{zyx} = w_{xyz} &= (-1)(-2)(y+2z)^{-3} \cdot 2 = \boxed{4(y+2z)^{-3}} \\ w_{xxy} = w_{xyx} &= \boxed{0} \end{aligned}$$

Objectives

☐ Review visually finding paths of steepest ascent.

For Lecture 18, we found paths of steepest ascent. The gradient is always perpendicular to a level curve. Since the gradient points toward the path of steepest ascent, such paths must be perpendicular to the level curve.



19. Lecture 19

Objectives

- ☐ More Questions

In this class, we answered questions

20. Lecture 20

Objectives

- ☐ Test your knowledge with 5 questions

The exam was administered during this class. Solutions posted on the following pages.

Objectives

- ☐ I can use the chain rule to find the derivative of a function with respect to a parameter like time.
- ☐ I know to write the result of the derivative in terms of the parameter(s) only.

Up to this point, we have focused on partial derivatives of a function $f(x, y)$ based on the variables x and y. In practice, however, the independent variables actually dependent on some "hidden variable" like time. We call these "hidden variables," **parameters**.

In this section, we will discuss how to take the derivative of the dependent variable with respect to some parameter. We will also introduce some new notation for partial derivatives.

Let's begin our discussion by motivating it with an example. On the last quiz, you were given the Cobb-Douglas production function

$$P(L, K) = AL^{\alpha}K^{\beta}$$

which models how a country's yearly production P is related to the number of labor hours L and the amount of invested capital K. Put in layman's terms, *the amount of money a country makes annually depends on how much people work (work hours) and what tools they have available to them (invested capital)*. The constants A, α and β relate to how efficiently a country can use its labor and capital investment. For example, if the population is filled with unproductive workers (maybe due to lack of education), then α will be low. If there aren't enough people who know how to use the capital investment (for example, if there are more busses than bus drivers), then β will be low. If there is a lot of corruption or crime, A will be low.

When we look at a partial derivative like P_L, we are seeing how much P changes if we change L by one unit. This is important for policy-makers to discern.

Policy makers may also want to understand how much production changes with time, $\frac{dP}{dt}$. For factors beyond anyone's control, L and K will change with time. For example, an aging population will cause L to go down with time. If a country can't continue to re-invest in its capital, like Cuba due to embargoes, the capital they have will slowly go down with time (tools rust, machines break, etc). When countries want to understand trends with time, P_L and P_K do not provide enough information.

To understand the derivative of P with respect to t, we will need to use the chain rule. Let us remind ourselves of how the chain rule works with two dimensional functions. If we are given the function $y = f(x)$, where x is a function of time: $x = g(t)$. Then the derivative of y with respect to t is the derivative of y with respect to x multiplied by the derivative of x with respect to t

$$\frac{dy}{dt} = \frac{df}{dx}\frac{dx}{dt} = f'(g(t))g'(t)$$

The technique for higher dimensions works similarly. The only difficulty is that we need to consider <u>all</u> the variables dependent on the relevant parameter (time t).

Suppose $z = f(x, y)$ and $x = g(t), y = h(t)$. Based on the one variable case, we can see that dz/dt is calculated as

$$\frac{dz}{dt} = f_x \frac{dx}{dt} + f_y \frac{dy}{dt}$$

In this context, it is more common to see the following notation.

$$f_x = \frac{\partial f}{\partial x}$$

The symbol ∂ is referred to as a "partial," short for partial derivative.

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

The procedures for calculating the derivative with respect to a parameter will be the following:

(1) Find the partial derivatives, f_x and f_y.
(2) Plug in the "inside functions" (i.e. plug in $g(t)$ for x when $x = g(t)$) into the partial derivatives
(3) Find the derivatives of the "inside functions."
(4) Calculate

$$\frac{dz}{dt} = f_x \frac{dx}{dt} + f_y \frac{dy}{dt}$$

Notice that we plug in the "inside functions." Why? Because it's less confusing to have our final answer completely in terms of the parameter variable. If we neglected this step, we would have an expression with x, y, and t.

Example 21.1: Use the chain rule to find dz/dt for

$$z = \ln(4x + y), \quad x = 5t^4, y = \frac{1}{t}$$

Solution 21.2: We begin by finding all the necessary partial derivatives. We put everything in terms of t by plugging in the functions for x and y. This gives us the following equations:

$$\frac{\partial z}{\partial x} = \frac{-4}{4x+y} = \frac{-4}{4[5t^4] + [1/t]} = \frac{-4}{\frac{20t^4+1}{t}} = \frac{-4t}{20t^4+1}$$

$$\frac{\partial z}{\partial y} = \frac{-1}{4x+y} = \frac{-1}{(4[5t^4] + [1/t])} = \frac{-1}{\frac{20t^4+1}{t}} = \frac{-t}{20t^4+1}$$

$$\frac{dx}{dt} = 20t^3$$

$$\frac{dy}{dt} = \frac{-1}{t^2}$$

Now, we just plug into the formula

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

to get

$$\frac{dz}{dt} = \frac{-4t}{20t^4+1} \cdot 20t^3 + \frac{-t}{20t^4+1} \cdot \frac{-1}{t^2}$$

$$\boxed{\frac{dz}{dt} = \frac{-80t^4}{20t^4+1} + \frac{t}{20t^6+t^2} = \frac{t-80t^6}{20t^6+t^2}}$$

Example 21.3: The temperature at a point (x, y) is $T(x, y)$, measured in degrees Celsius. A bug crawls so that its position after t second is given by $x = \sqrt{1+t}$, $y = 2 + \frac{t}{3}$, where x and y are measured in centimeters. The temperature function satisfies $T_x(2, 3) = 4$ and $T_y(2, 3) = 3$. How fast is the temperature rising on the bugs path after 3 seconds?

Solution 21.4: At first glance, it may appear that we lack enough information to answer this question. Surprisingly, we have just enough.

Because the symbols are different than above, it's helpful to begin by writing out the derivative of T with respect to time:

$$\frac{dT}{dt} = T_x \cdot \frac{dx}{dt} + T_y \cdot \frac{dy}{dt}$$

Now, we wish to determine this when $t = 3$. That means $x = \sqrt{1+3} = \sqrt{4} = 2$ and $y = 2 + \frac{3}{3} = 3$. Hence, we want to calculate all these derivatives with the assumption that $x = 2, y = 3$, and $t = 3$.

$$\begin{aligned}
&T_x(2,3) = 4 \text{ (this was given)}\\
&T_y(2,3) = 3 \text{ (this was also given)}\\
&\frac{dx}{dt}(3) = \frac{1}{2}(1+t)^{-1/2}\Big|_{t=3} = \frac{1}{2}(1+3)^{-1/2} = \frac{1}{2\sqrt{4}} = \frac{1}{4}\\
&\frac{dy}{dt}(3) = \frac{1}{3}
\end{aligned}$$

Now we plug into our expression $\frac{dT}{dt} = T_x \cdot \frac{dx}{dt} + T_y \cdot \frac{dy}{dt}$ and get

$$\frac{dT}{dt} = (4) \cdot \left(\frac{1}{4}\right) + (3) \cdot \left(\frac{1}{3}\right) = \boxed{2}$$

Thus, the temperature rises two degrees on the bug's path after 3 seconds.

Summary of Ideas

- Sometimes x and y are functions of one or more parameters. We may find the derivative of a function with respect to that parameter using the **chain rule**.
- The formula for calculating such a derivative is

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \qquad \text{and} \qquad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

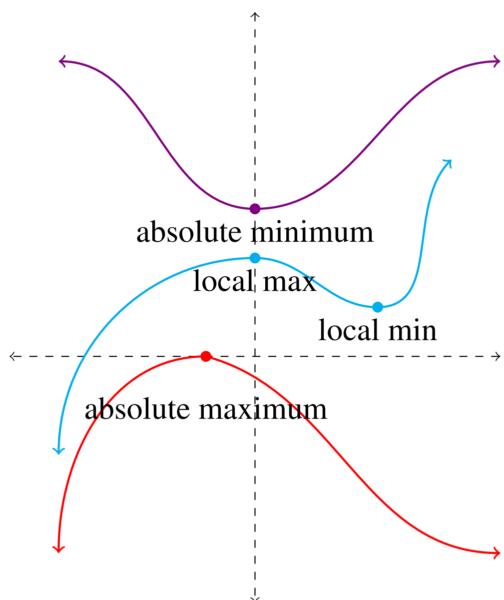
Objectives

- ☐ I can define critical points.
- ☐ I know the difference between local and absolute minimums/maximums.

In many physical problems, we're interested in finding the values (x, y) that maximize or minimize $f(x, y)$.

Recall from your first course in calculus that **critical points** are values, x, at which the function's derivative is zero, $f'(x) = 0$. These x-values either maximized $f(x)$ (a **maximum**) or minimized $f(x)$ (a **minimum**).

We did not simply call a critical point a maximum or a minimum, however. Sometimes your critical point is local, meaning it may not the highest/lowest value achieved by the *entire* function, but it's the highest/lowest point "near by." See the image below for clarification.



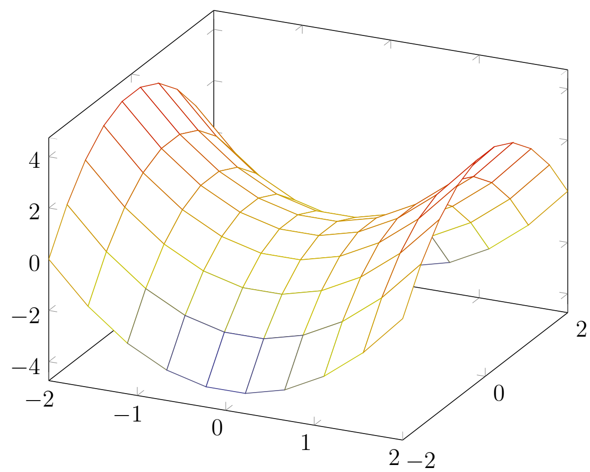
It's important to note that all maximums and minimums are *local*. To be an absolute maximum or minimum, you have to know the heights and trends of the entire function.

To classify critical points as maximums or minimums, we look at the second derivative. The point was called a minimum if $f''(x_0) > 0$ and it was called a maximum if $f''(x_0) < 0$. I like the mnemonic, "concave up (+) is like a cup; concave down (-) is like a frown."

For functions of two variables, $z = f(x, y)$, we do something similar.

Definition 22.1: A point (a, b) is a **critical point** of $z = f(x, y)$ if the gradient, ∇f, is the zero vector or if it is undefined.

Critical points in three dimensions can be maximums, minimums, or saddle points. A **saddle point** mixes a minimum in one direction with a maximum in another direction, so it's neither (see the image below).



Once a point is identified as a critical point, we want to be able to classify it as one of the three possibilities. Like you did in calculus, you will look at the second derivative. In higher dimensions, this is the determinant of a matrix containing <u>all possible</u> second derivatives, denoted as d. This matrix is called the **Hessian matrix**.

$$d = \det \begin{vmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{vmatrix}$$

A determinant is a process that takes in a matrix and produces a number. For a 2×2 matrix, it means multiplying the diagonal entries and subtracting the product of the off diagonal entries:

$$d = f_{xx}(a,b) f_{yy}(a,b) - [f_{xy}(a,b)]^2$$

The critical point is classified by the value of d.

- If $d > 0$ and $f_{xx}(a,b) > 0$, then the point is a **(local) minimum**. For f_{xx} we can think, "concave up (+) is like a cup."
- If $d > 0$ and $f_{xx}(a,b) < 0$, then the point is a **(local) maximum**. For f_{xx} we can think, "concave down (-) is like a frown."
- If $d < 0$, then the point is a **saddle point**.
- If $d = 0$, then we call the critical point **degenerate** because the test was inconclusive.

For degenerate points, often there is a clever argument one can make that let's you classify it. In Example 22.4, this is precisely what happens.

Example 22.2: Find and classify all critical values for the following function.

$$f(x,y) = xy - 2x - 2y - x^2 - y^2$$

First, we need to find the zeros of the partial derivatives. Those partials are

- $f_x(x,y) = y - 2 - 2x$
- $f_y(x,y) = x - 2 - 2y$

Set both of these partial derivatives to zero.

- $0 = y - 2 - 2x$
- $0 = x - 2 - 2y$

Then we solve the system of equations.

$$x = 2 + 2y \implies y = 2 + 2(2 + 2y) \implies y = 2 + 4 + 4y$$

Then $-3y = 6$ gives us that $y = -2$. We can plug in to find x

$$x = 2 + 2(-2) = -2$$

The solution is $\boxed{(-2, -2).}$ That is our critical point.

Now, we need to classify it. Let's find the second partial derivatives:

- $f_{xx}(x, y) = -2$
- $f_{yy}(x, y) = -2$
- $f_{xy}(x, y) = 1$

Then

$$d = (-2)(-2) - 1 = 3$$

Since $d = 3 > 0$ and $f_{xx} = -2 < 0$, then we have a $\boxed{\text{local maximum.}}$

Example 22.3: Find and classify all critical values for the following function.

$$f(x, y) = x^3 - 12xy - 8y^3$$

First, we need to find the zeros of the partial derivatives. Those partials are

- $f_x(x, y) = 3x^2 - 12y$
- $f_y(x, y) = -12x - 24y^2$

Set both of these partial derivatives to zero.

- $y = (1/4)x^2$
- $x = -2y^2$

Next, we solve the system of equations.

$$y = (1/4)(-2y^2)^2 \implies y = y^4$$

If $y = 1$, then $x = -2$. If $y = 0$, then $x = 0$. The critical points $\boxed{(-2, 1)}$ and $\boxed{(0, 0).}$

We now calculate the second derivatives to classify the critical point.

- $f_{xx}(x, y) = 6x$
- $f_{yy}(x, y) = -48y$
- $f_{xy}(x, y) = -12$

Then

$$d = (6x)(-48y) - (12)^2 = -288xy - 144$$

Now, let's determine d for the point $(-2, 1)$.

$$d(-2, 1) = 576 - 144 = 432 > 0$$

Since $d(-2, 1) = 432 > 0$ and $f_{xx}(-2, 1) = 6(-2) = -12 < 0$, then $(-2, 1)$ a $\boxed{\text{local maximum.}}$

$$d(0, 0) = 0 - 144 < 0$$

Thus, $(0, 0)$ is a $\boxed{\text{saddle point.}}$

Example 22.4: Find and classify all critical values for the following function.

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

As before, we begin by finding the partial derivatives:

- $f_x(x,y) = \frac{1}{2}(x^2+y^2)^{-1/2} \cdot 2x = x(x^2+y^2)^{-1/2} = \frac{x}{\sqrt{x^2+y^2}}$
- $f_y(x,y) = \frac{1}{2}(x^2+y^2)^{-1/2} \cdot 2y = y(x^2+y^2)^{-1/2} = \frac{y}{\sqrt{x^2+y^2}}$

Set both of these equations equal to zero.

- $0 = \frac{x}{\sqrt{x^2+y^2}} \implies x = 0$
- $0 = \frac{y}{\sqrt{x^2+y^2}} \implies y = 0$

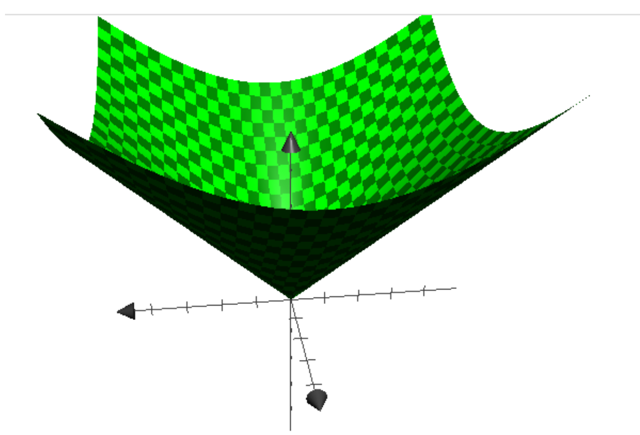
Hence, our only critical point is $(0,0)$.
Let's now calculate the second derivatives:

- $f_{xx}(x,y) = x \cdot (-\frac{1}{2})(x^2+y^2)^{-3/2} \cdot 2x + 1 \cdot x(x^2+y^2)^{-1/2} = \frac{-x^2+x(x^2+y^2)}{(x^2+y^2)^{3/2}}$
- $f_{xy}(x,y) = x \cdot (-\frac{1}{2})(x^2+y^2)^{-3/2} \cdot 2y + 0 \cdot x(x^2+y^2)^{-1/2} = \frac{-xy}{(x^2+y^2)^{3/2}}$
- $f_{yy}(x,y) = y \cdot (-\frac{1}{2})(x^2+y^2)^{-3/2} \cdot 2y + 1 \cdot y(x^2+y^2)^{-1/2} = \frac{-y^2+y(x^2+y^2)}{(x^2+y^2)^{3/2}}$

All of the equations above are undefined at $(0,0)$. So the critical point is **degenerate**.

How can we determine if it is a maximum, a minimum, or a saddle point? Let's return to the original function $f(x,y) = \sqrt{x^2+y^2}$. Notice that it is always positive. So at $(0,0)$, it's at it's lowest value. Hence, the point is a minimum.

If we graph the function, we see that this is correct.



Summary of Ideas

- Critical points are those points (x, y) such that $\nabla f(x, y) = 0$ or if ∇f is undefined.
- They are classified as local maximums, minimums and saddles using the determinant of the **Hessian matrix**.

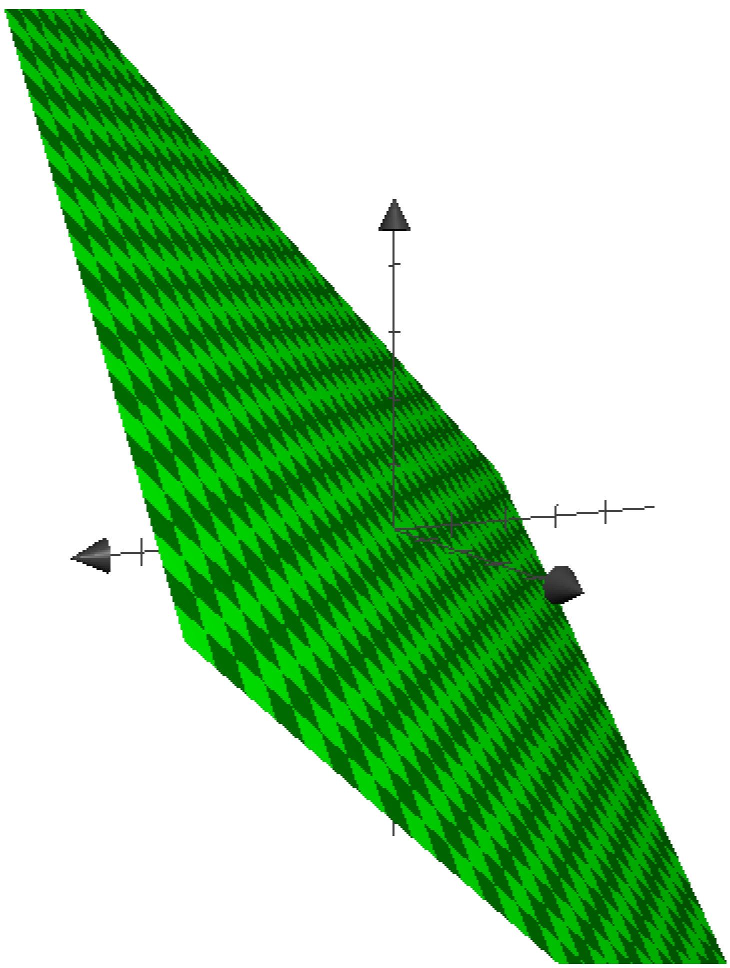
$$d = \det \begin{vmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{vmatrix}$$

- We classify according the following rules:
  - If $d > 0$ and $f_{xx}(a, b) > 0$, then the point is a **(local) minimum**.
  - If $d > 0$ and $f_{xx}(a, b) < 0$, then the point is a **(local) maximum**.
  - If $d < 0$, then the point is a **saddle point**.
  - If $d = 0$, then the point is **degenerate** and it could be anything.

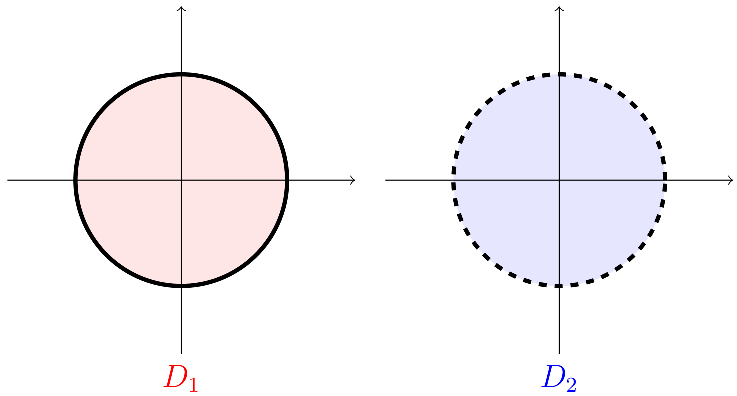
Objectives

☐ I know what a closed region is.
☐ I can classify critical points over a closed domain.
☐ I can determine if a critical point is an absolute maximum or absolute minimum.

In the last lecture, we defined absolute maximum and an absolute minimum. These were critical points that were the highest and lowest values (resp.) a function takes on. These are often positive and negative infinity. For example, for an infinite plane has an absolute maximum of ∞ and an absolute minimum of $-\infty$.



When will the absolute maximum or absolute minimum be finite? If the **domain** (the set of independent variables) is **closed**, then f has a finite **absolute minimum** and **absolute maximum**. A **closed domain** is a set of independent variables that includes its boundary points. For example $D_1 = \{(x, y)|x^2+y^2 \leq 2\}$ is closed because it includes its boundary while $D_2 = \{(x, y)|x^2+y^2 < 2\}$ is not closed because it does not.



To find the absolute maximum and absolute minimum, follow these steps:

(1) Find the the critical points of f on D.
(2) Find the extreme values of f on the boundary of D.
(3) The largest of the values from steps 1 and 2 is the absolute maximum value; the smallest of these values is the absolute minimum value.

The most challenging part of these problems will be considering the values of f on the boundary. You will reduce the problem in one of two ways. Let's consider the domain D_1 above and try to maximize $f(x, y) = xy$ on its boundary.

- One method is to solve one variable in terms of another. The boundary is $2 = x^2 + y^2$, so we could solve and say $y = \pm\sqrt{2 - x^2}$. Then we can plug in for y to get $f(x, y) = f(x) = \pm x\sqrt{2 - x^2}$. The boundary's critical points are precisely those values of x for which

$$0 = f'(x) = \mp\frac{2(x^2 - 1)}{\sqrt{2 - x^2}}$$

  This is only true when $x = \pm 1$. We then find the corresponding values of y and find the extreme points on the boundary are

$$(1, 1), (1, -1), (-1, -1), \text{ and } (-1, 1)$$

- Alternatively, we could parameterize the boundary. That means we pick $x = \sqrt{2}\sin t$ and $y = \sqrt{2}\cos t$. Then we get $f(t) = x(t)y(t) = 2\sin t\cos t = \sin 2t$. We find the critical points of $f(t)$ by solving $0 = f'(t) = 2\cos 2t$. Here, we need to consider all values of t between 0 and 2π because that is a full rotation around the boundary. Therefore, this is true if $t = \pi/4, 3\pi/4, 5\pi/4$, and $7\pi/4$. That means our extreme points are $(\sqrt{2}\sin t, \sqrt{2}\cos t)$ for those values of t. That is,

$$(1, 1), (1, -1), (-1, -1), \text{ and } (-1, 1)$$

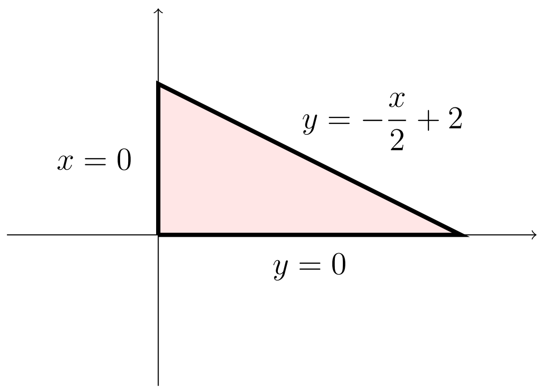
Either approach will work. I recommend you use the one that makes the most sense to you in the given problem.

Before we look at examples, let's briefly discuss terminology. When we say "critical points," we mean points where the derivative or gradient equals zero ($f'(x) = 0$ or $\nabla f = \vec{0}$). We use the term extreme value to just mean the biggest or smallest. The distinction is that an extreme value may not make the derivative zero, but it still may give the largest value.

Example 23.1: Find the absolute maximum and minimum values of the function on D, where D is the enclosed triangular region with vertices $(0,0)$, $(0,2)$, and $(4,0)$.

$$f(x,y) = x + y - xy$$

Let's first draw a picture of D to help us visualize everything.



Solution 23.2: First, we find the critical points on D. We begin by finding the partials and setting them equal to zero

- $f_x(x,y) = 1 - y = 0$
- $f_y(x,y) = 1 - x = 0$

The only critical point on D is $\boxed{(1,1)}$. Notice that $f(1,1) = 1$

Next, we find the extreme points on the boundary. We will use the information in our picture to help us.

From $(0,0)$ to $(0,2)$, the line is $x = 0$. We can then plug in $(0,y)$, where $0 \le y \le 2$. When we plug in these values, we see that

$$f(0,y) = 0 + y - (0)(y) = y.$$

The maximum value this can be is 2, which is achieved at $\boxed{(0,2)}$. The minimum value is 0, which is achieved at $\boxed{(0,0).}$

From $(0,0)$ to $(4,0)$, the line is $y = 0$. We can then plug in $(x,0)$, where $0 \le x \le 4$. Along this line, the values are

$$f(x,0) = x$$

The maximum value is 4, which is achieved at $\boxed{(4,0).}$ The minimum value is 0, which is achieved at $\boxed{(0,0).}$

From $(4,0)$ to $(0,3)$, the line that defines it is $y = -x/2 + 2$ for $0 \le x \le 4$. Let's plug in.

$$f(x) = x + \left(-\frac{x}{2} + 2\right) - x\left(-\frac{x}{2} + 2\right) = \frac{x^2}{2} - \frac{3x}{2} + 2$$

The critical points are the values of x such that

$$0 = f'(x) = x - \frac{3}{2}$$

The critical point is then $\boxed{(3/2, 5/4)}$ and $f(3/2, 5/4) = 7/8$. We do not need to check the end points since we already know those values.

Now, let's take a moment to study all the critical points we've found:

- $f(1, 1) = 1$
- $f(0, 2) = 2$
- $f(0, 0) = 0$
- $f(4, 0) = 4$
- $f(3/2, 5/4) = 7/8$

Therefore, the absolute maximum happens at $\boxed{\boxed{(4, 0)}}$ and the absolute minimum happens at $\boxed{\boxed{(0, 0).}}$

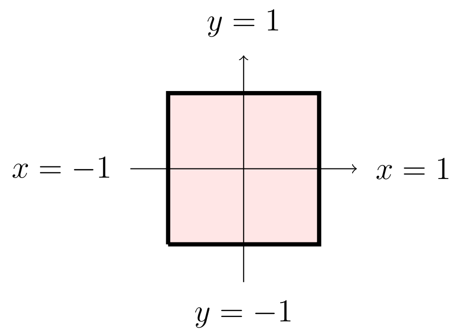
Example 23.3: Find the absolute maximum and minimum values of f on the set D, where

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

and

$$D = \{(x, y) \mid |x| \leq 1, |y| \leq 1\}$$

Solution 23.4: We start with the same process as before. While it's not required, it's always good to start with a picture of your domain.



Let us first find the critical points in D. We solve the following equations

- $f_x(x, y) = 2x + 2xy = 2x(1 + y) = 0$
- $f_x(x, y) = 2y + x^2 = 0$

The critical points are therefore $\boxed{(0, 0)}$ and $\boxed{(\sqrt{2}, -1)}$. We should take a moment to observe that $f(0, 0) = 4$ and $f(\sqrt{2}, -1) = 5$.

Now, let's consider the extreme points of the boundary.

Let's begin with the section $x = -1$. Plugging in gives us the function

$$f(-1, y) = 1 + y^2 + y + 4 = y^2 + y + 5$$

where $-1 \leq y \leq 1$. What is the maximum and minimum value on this side of the box? We can find this by looping for its critical points!

This function achieves its critical point when

$$f'(y) = 2y + 1 = 0 \implies y = -1/2$$

so at $\boxed{(-1, -1/2)}$ is a critical point on the boundary. Let's take note that $f(-1, -1/2) = 19/4$ or 4.75.

When we use derivatives to find critical points, we must also check the end points. Why? Because this is the same process. We're finding the critical points over a close set (a line segment). In this case, we get $\boxed{(-1,-1)}$ and $\boxed{(-1,1)}$ as potential critical points. Note, their values are $f(-1,-1) = 5$ and $f(-1,1) = 7$.

Now let's check $x = 1$. Then,

$$f(y) = 1 + y^2 + y + 4 = y^2 + y + 5$$

By the exact same work above, we know $\boxed{(1,-1/2)}$ will be a critical point for the boundary. We, again, take note that $f(1,-1/2) = 19/4$. We also need to check the corners because these are extreme points: $f(1,-1) = 5$ and $f(1,1) = 7$.

At this point, we've looked at *all* the corners. For the two remaining boundary lines, we can skip this step.

Let us now check $y = 1$. This gives us

$$f(x) = x^2 + 1 + x^2 + 4 = 2x^2 + 5$$

If we consider its derivative, $f'(x) = 4x$, we see that we have a critical point at $\boxed{(4,1).}$ We should check that $f(0,1) = 4$.

Finally, we check $y = -1$. This gives us

$$f(x) = x^2 + 1 - x^2 + 4 = 5$$

This is a flat line, so it has no critical points!

Now, let's tally all the points we found.

Critical points of the Surface in D

- $f(0,0) = 4$
- $f(\sqrt{2},-1) = 5$

Critical points on the boundary

- (On $x = -1$): $f(-1,-1/2) = 19/4 = 4.75$
- (On $x = 1$): $f(1,-1/2) = 19/4$
- (On $y = 1$): $f(0,1) = 4$

Corner Values

- $f(-1,-1) = 5$
- $f(-1,1) = 7$
- $f(1,-1) = 5$
- $f(1,1) = 7$

Therefore, the absolute maximum is achieved at two locations: $\boxed{(-1,1) \text{ and } (1,1)}$ and the absolute minimum is also at two locations: $\boxed{(-1,-1/2) \text{ and } (1,-1/2).}$

Finding absolute maximums and absolute minimums can be quite challenging. As the domains become more complex, so do the calculations. Let's look at one such example.

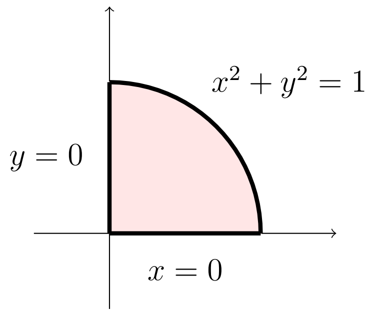
Example 23.5: Find the absolute maximum and minimum values of f on the set D, where

$$f(x,y) = xy^3$$

and

$$D = \{(x,y) \mid x \geq 0, y \geq 0, x^2 + y^2 \leq 1\}$$

Solution 23.6: This domain will be just a quarter of a circle since we restrict both x and y to be greater than zero.



Let us first find the critical points of the function.

- $f_x(x, y) = y^3 = 0$
- $f_y(x, y) = 3xy^2 = 0$

Notice that $y = 0$ is a sufficient condition to get a critical point. That is, all points $\boxed{(x, 0)}$ are critical! To stay in our domain, we'll consider $0 \leq x \leq \sqrt{1 - y^2} = 1$.

$$f(x, 0) = 0$$

Now, let's consider the boundary. You may have noticed that we already considered $y = 0$ above. Let's look at $x = 0$.

$$f(0, y) = 0$$

This doesn't have critical points, and all values are equal to zero (like on $y = 0$). While this may be a frustrating result, we know this problem must have a maximum and minimum. Let's check the last portion of the boundary.

The easiest way to find the critical point on the arc is to parametrize. In particular, $x = \cos t$, $y = \sin t$, and let $0 \leq t \leq \pi/2$.

Since this is easier, let's leave that for you to try on your own. Here, we will use the alternative approach $x = \sqrt{1 - y^2}$ where $0 \leq y \leq 1$.

Let's find the critical points by plugging this equation into $f(x, y)$.

$$\begin{aligned}
& f(\sqrt{1 - y^2}, y) = y^3 \sqrt{1 - y^2} \\
\implies \quad & f'(\sqrt{1 - y^2}, y) = 3y^2 \cdot \sqrt{1 - y^2} + y^3 \cdot \frac{1}{2}(1 - y^2)^{-1/2} \cdot -2y \\
& = 3y^2 \sqrt{1 - y^2} + \frac{y^3 \cdot -y}{\sqrt{1 - y^2}} \\
& \\
& = 3y^2 \sqrt{1 - y^2} \cdot \frac{\sqrt{1 - y^2}}{\sqrt{1 - y^2}} + \frac{-y^4}{\sqrt{1 - y^2}}
\end{aligned}$$

$$f'(\sqrt{1-y^2}, y) = \frac{3y^2(1-y^2) - y^4}{\sqrt{1-y^2}} = 0$$

The above equation is zero when the numerator is zero. That means

$$0 = 3y^2(1-y^2) - y^4 = 3y^2 - 3y^4 - y^4 = 3y^2 - 4y^4 = y^2(3-4y^2)$$

There are two possibilities: $y = 0$ or $y = \frac{\sqrt{3}}{2}$. By plugging into $x = \sqrt{1-y^2}$, we can get two critical points, $\boxed{(1,0)}$ and $\boxed{\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)}$.

Now, let's list our critical points and their corresponding values.

- $f(x, 0) = 0$
- $f(1, 0) = 0$
- $f\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \left(\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right)^3 = \frac{3\sqrt{3}}{4} \approx 1.3$

Hence, the absolute minimum 0 occurs at $\boxed{(x,0)}$ and $\boxed{(1,0)}$. The absolute maximum ≈ 1.3 occurs at $\boxed{\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)}$.

Summary of Ideas

- Over a closed region D, you can find the absolute minimum and absolute maximum. These are the smallest and largest values achieved by $f(x, y)$, respectively.
- To find these values, we first find the critical points on D. We then restrict f to the boundary of D and find the extreme values. We can solve one variable in terms of another and plug in the expression or we can parametrize the path and plug in $(x(t), y(t))$.

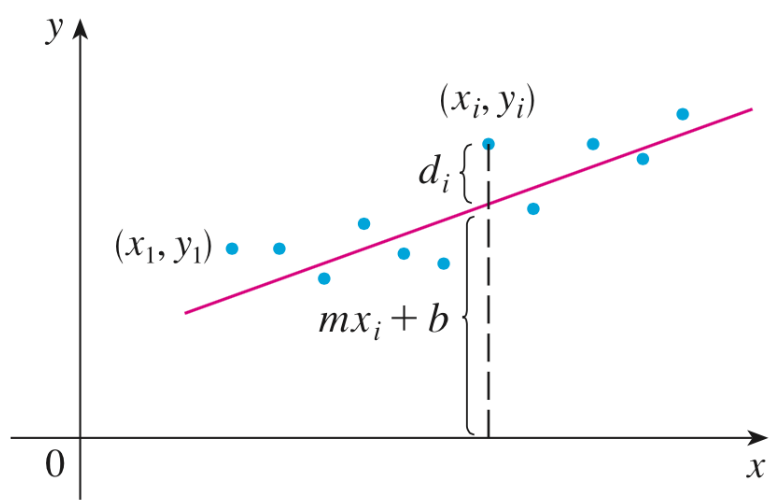
Objectives

☐ I know that finding a best-fit line/curve is an optimization problem.

In some sense, we've been using optimization all along in this class. The process of fitting data to a line (or curve) is exactly an optimization process known as the **Method of Least Squares**. In the example below, we explain the set up and its connection with the equation

$$A^T A\vec{x} = A^T\vec{b}$$

Example 24.1: (How We Fit Data to a Line.) Suppose you have a set of points $(x_1, y_1), \ldots (x_n, y_n)$ and you believe these points follow a linear relationship $y = ax + b$. How can we find the best values for the coefficients a and b so that it fits the data? We try to minimize the distance between each observed value, y_i, and the predicted value based on the line, $ax_i + b$. We can define these distances as $d_i = y_i - (ax_i + b)$.



The **method of least squares** is what we use to fit a line to data. It turns out to be an optimization problem. That is, we try to pick a and b so that

$$S = d_1^2 + d_2^2 + \ldots + d_n^2$$

is minimized!

Show that S is minimized when

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

and

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

Thus, the best fit line is found by solving these two equations with two unknowns, a and b.

Solution 24.2: Based on the description, the problem is to minimize

$$S = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

We, therefore, want to find values of a and b so that

$$\nabla S = \begin{pmatrix} S_a \\ S_b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So let's calculate each partial derivative.

$$\begin{aligned} 0 = S_a &= \left[\sum_{i=1}^{n} (y_i - ax_i - b)^2 \right]_a \\ &= \sum_{i=1}^{n} 2(y_i - ax_i - b) \cdot (-x_i) \\ &= -2 \sum_{i=1}^{n} x_i (y_i - ax_i - b) \end{aligned}$$

So from $0 = S_a$, we get

$$0 = -2 \sum_{i=1}^{n} x_i (y_i - ax_i - b) \implies a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

Now let's calculate the remaining partial derivative.

$$\begin{aligned} 0 = S_b &= \left[\sum_{i=1}^{n} (y_i - ax_i - b)^2 \right]_b \\ &= \sum_{i=1}^{n} 2(y_i - ax_i - b) \cdot (-1) \\ &= -2 \sum_{i=1}^{n} (y_i - ax_i - b) \end{aligned}$$

So from $0 = S_b$, we get

$$0 = -2 \sum_{i=1}^{n} (y_i - ax_i - b) \implies a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i$$

Notice that $\sum_{i=1}^{n} 1 = 1 + 1 + \ldots + 1 = n$. Hence, we get

$$a \sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

So we have proved the claim.

Summary of Ideas

- The method of least squares, which is what we use when we find best-fit curves, is an optimization problem.

Objectives

☐ I understand how the best-fit line/curve is related to the equation
$$\vec{x} = (A^T A)^{-1}(A^T \vec{b})$$

In the last class, we stated that the method of least squares in an optimization problem. In particular, we showed that the sum of squares

$$S = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)$$

is minimized when

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

and

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

For weeks now, we have been fitting data to functions, but we never used the equations stated above. Instead, we used

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}.$$

Today, we show that these two are the *same* equation.

First, observe that $\vec{x} = (A^T A)^{-1} A^T \vec{b}$ can be rewritten as

$$A^T A\vec{x} = A^T \vec{b}$$

if we multiply both sides of the equation by $(A^T A)$.

Example 25.1: Show how $A^T A\vec{x} = A^T \vec{b}$ is equivalent to solving

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

and

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

For ease of calculation, assume $n = 4$.

Solution 25.2: Let's write out each equation with the assumption that $n = 4$.

$$a\sum_{i=1}^{4} x_i + 4b = \sum_{i=1}^{4} y_i$$

means

$$a(x_1 + x_2 + x_3 + x_4) + 4b = y_1 + y_2 + y_3 + y_4$$

and

$$a\sum_{i=1}^{4} x_i^2 + b\sum_{i=1}^{4} x_i = \sum_{i=1}^{4} x_i y_i$$

means

$$a(x_1^2 + x_2^2 + x_3^2 + x_4^2) + b(x_1 + x_2 + x_3 + x_4) = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4.$$

Now, let's write out our usual matrices. Remember that the function is $y = ax + b$.

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} a \\ b \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

Now, let's calculate

$$A^T A = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 + x_4^2 & x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 & 4 \end{pmatrix}$$

Therefore,

$$\begin{aligned} A^T A\vec{x} &= \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 + x_4^2 & x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 & 4 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} a(x_1^2 + x_2^2 + x_3^2 + x_4^2) + b(x_1 + x_2 + x_3 + x_4) \\ a(x_1 + x_2 + x_3 + x_4) + 4b \end{pmatrix} \end{aligned}$$

Finally, we calculate $A^T\vec{b}$.

$$\begin{aligned} A^T\vec{b} &= \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \\ &= \begin{pmatrix} x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \\ y_1 + y_2 + y_3 + y_4 \end{pmatrix} \end{aligned}$$

Putting this all together, $A^T A\vec{x} = A^T\vec{b}$ is

$$\begin{pmatrix} a(x_1^2 + x_2^2 + x_3^2 + x_4^2) + b(x_1 + x_2 + x_3 + x_4) \\ a(x_1 + x_2 + x_3 + x_4) + 4b \end{pmatrix} = \begin{pmatrix} x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \\ y_1 + y_2 + y_3 + y_4 \end{pmatrix}$$

which is the same as

$$a(x_1 + x_2 + x_3 + x_4) + 4b = y_1 + y_2 + y_3 + y_4$$

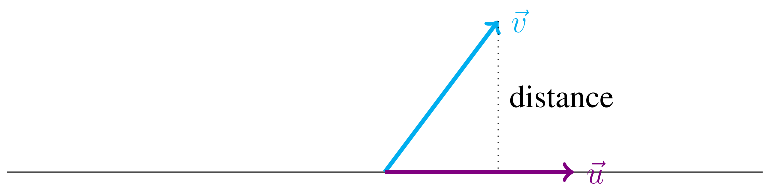
and

$$a(x_1^2 + x_2^2 + x_3^2 + x_4^2) + b(x_1 + x_2 + x_3 + x_4) = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4.$$

If we let n be an arbitrary number (not just 4), we would see that $A^T A\vec{x} = A^T\vec{b}$ is

$$\begin{pmatrix} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i \\ a\sum_{i=1}^{n} x_i + nb \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

Finally, I have mentioned in class numerous times that $A^T A\vec{x} = A^T\vec{b}$ is a "projection." This seems contradictory because we have just shown that fitting data to a line (or curve) is actually a minimization problem. It turns out, all projections are minimization problems. For example, when we project one vector $\vec{v}$ onto another vector $\vec{u}$, we are *minimizing* the distance from the tip of $\vec{v}$ to the line that contains $\vec{u}$:



Summary of Ideas

- The equation
$$\vec{x} = (A^T A)^{-1}(A^T\vec{b})$$
is equivalent to the equations that minimize
$$S(a, b):$$
$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$
and
$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

PENN STATE UNIVERSITY, UNIVERSITY PARK, PA 16802