

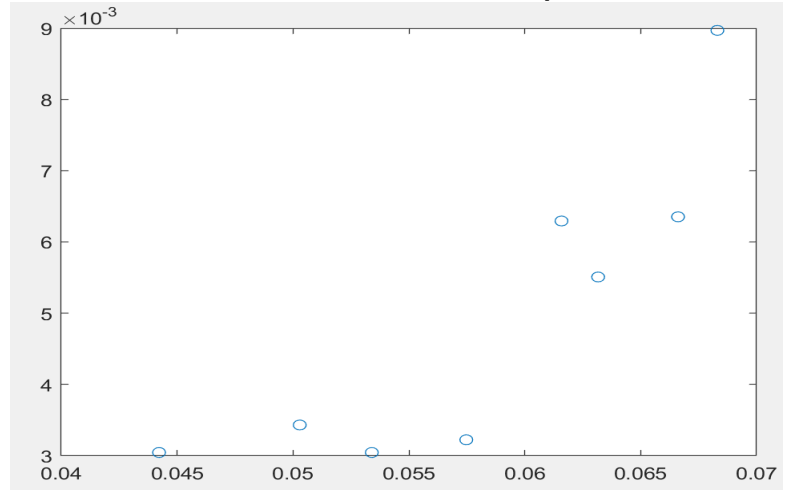
Section 8.5 Least Squares Solutions

The concept of "Least Squares" will come up in many different forms for many different problems, but at it is always a minimization problem. Let's frame it for this class through an example we studied at the beginning of the semester.

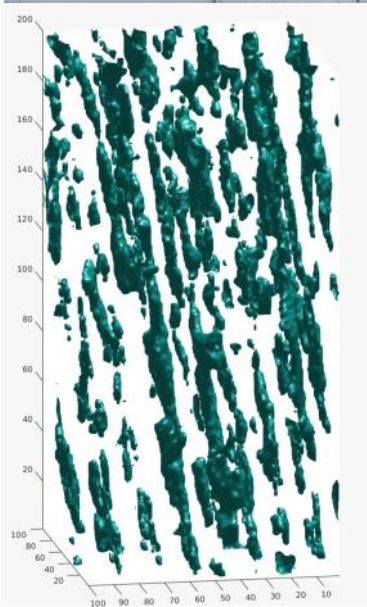
Physical Modeling

Brine Volume Fraction ϕ	Conductivity σ
0.068309	0.00897
0.066612	0.006353
0.05029	0.003431
0.04422	0.003045
0.057476	0.003223
0.063163	0.005508
0.0534	0.003045
0.061583	0.006294

$$\text{Archies Law } \sigma = a\phi^m$$

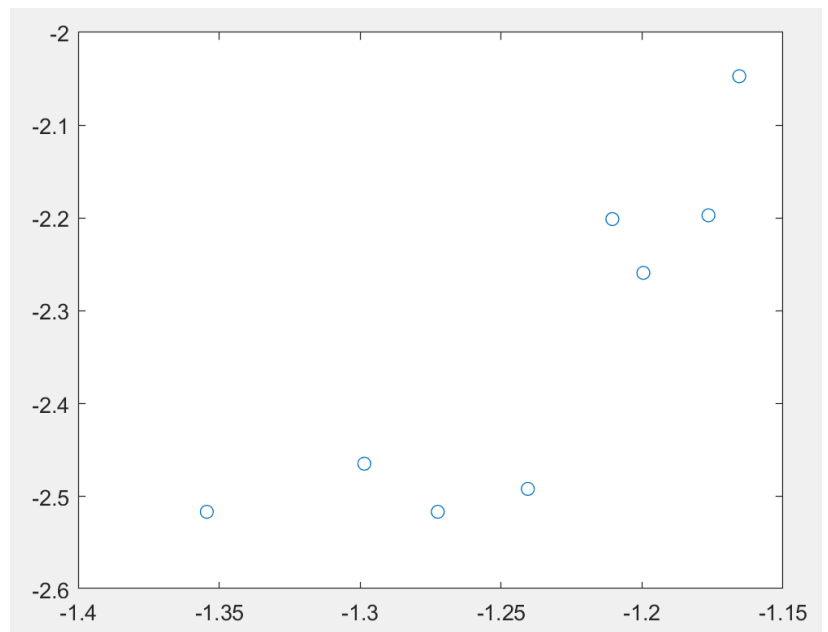


We can change this to a linear equation!



$\log_{10}(\phi)$	$\log_{10}(\sigma)$
-1.16552	-2.04721
-1.17645	-2.19704
-1.29852	-2.46455
-1.35438	-2.51647
-1.24052	-2.49179
-1.19954	-2.25902
-1.27246	-2.51637
-1.21054	-2.20106

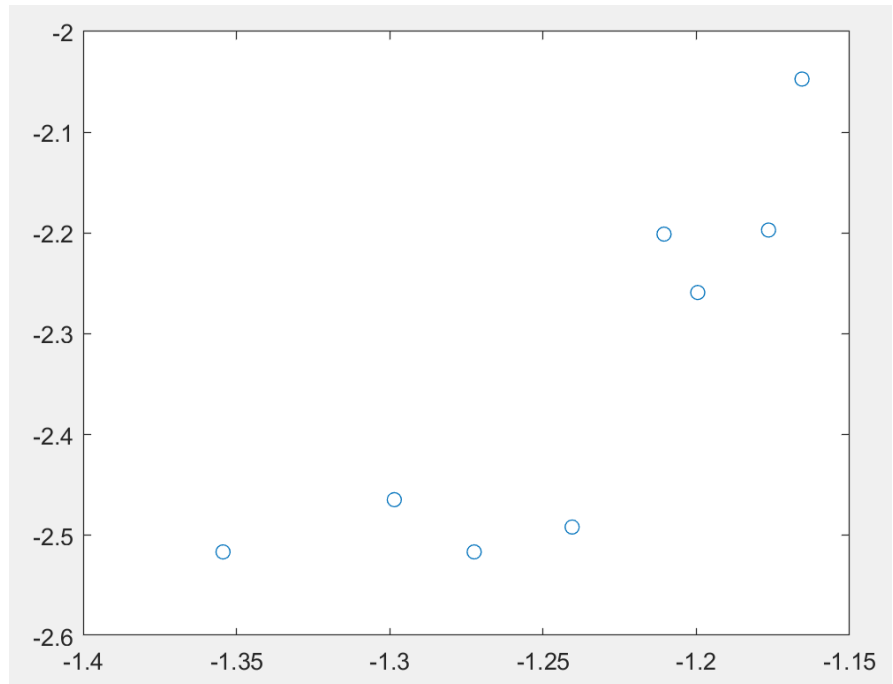
Sea ice brine microstructure.



Section 8.5 Least Squares Solutions

Last time we just chose two of the points and found the line that went through them. This time let's ask the question: What is the slope and y intercept of the line that is as close to all of the points as possible.

$\log_{10}(\phi)$	$\log_{10}(\sigma)$
-1.16552	-2.04721
-1.17645	-2.19704
-1.29852	-2.46455
-1.35438	-2.51647
-1.24052	-2.49179
-1.19954	-2.25902
-1.27246	-2.51637
-1.21054	-2.20106



$$\begin{aligned}
 -1.6552m + b &= -2.04721 \\
 -1.17645m + b &= -2.19704 \\
 -1.29852m + b &= -2.46455 \\
 -1.35438m + b &= -2.51647 \\
 -1.24052m + b &= -2.49179 \\
 -1.19954m + b &= -2.25902 \\
 -1.27246m + b &= -2.51637 \\
 -1.21054m + b &= -2.20106
 \end{aligned}$$

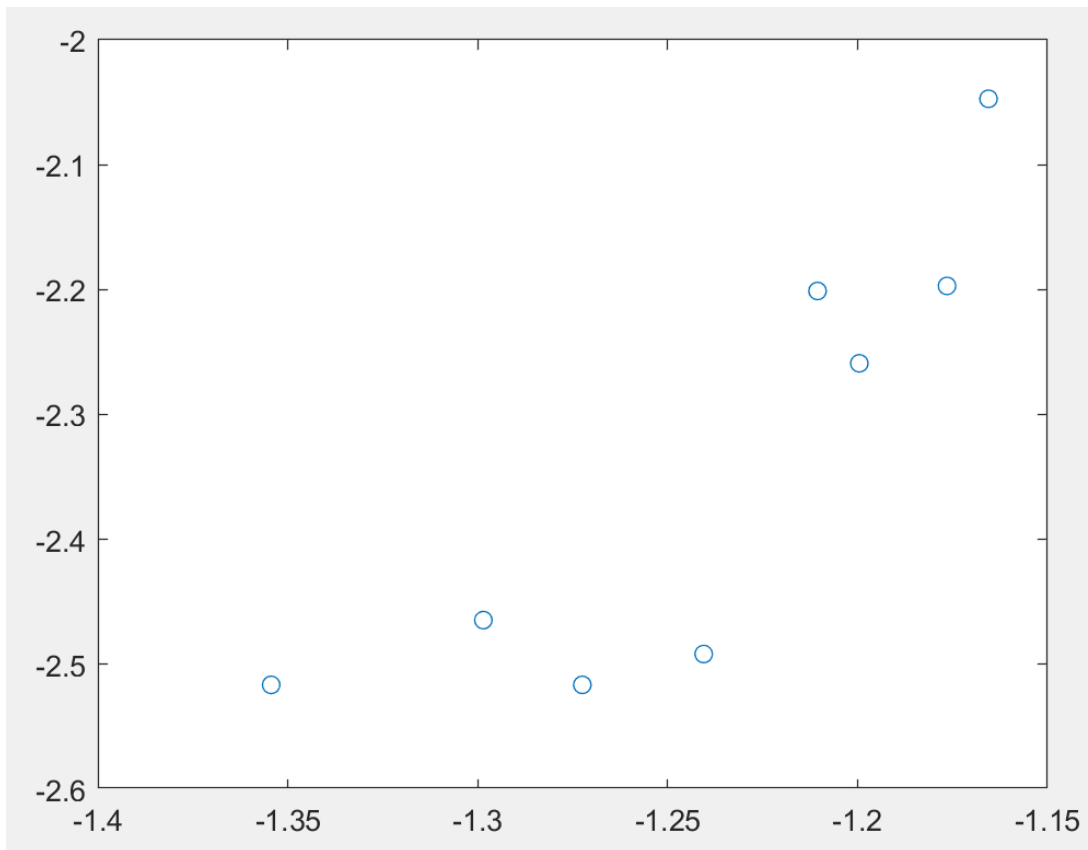
A =

-1.1655	1.0000	-2.0472
-1.1764	1.0000	-2.1970
-1.2985	1.0000	-2.4646
-1.3544	1.0000	-2.5164
-1.2405	1.0000	-2.4917
-1.1995	1.0000	-2.2590
-1.2725	1.0000	-2.5164
-1.2105	1.0000	-2.2011

$$\begin{array}{rcl}
 & 1 & 0 & 0 \\
 & 0 & 1 & 0 \\
 & 0 & 0 & 1 \\
 rref(A) = & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0
 \end{array}$$

Here there is no solution since the points we have don't actually fall on a line, that is, there is no m, b such that $\log(\sigma) = m\log(\phi) + b$ based solely on this data. We could just choose two points, but the question would be which two? Further we want to consider all of the data so what can we do?

Section 8.5 Least Squares Solutions



We can't solve $A\vec{x} = \vec{y}_{data}$ so instead we will consider the problem $A\vec{x} = \vec{y}_{ls}$ where $\vec{y}_{ls}$ is made of y values on a line where that line minimizes the sum of the squares of the distances between those y values and the y data points. That is, we are going to select from $col(A)$ the vector that best captures the data.

Let $\vec{y}_{ls} \in col(A)$ such that $A\vec{x} = \vec{y}_{ls}$

$$\vec{y}_{ls} = \min_{\vec{y} \in col(A)} \|\vec{y}_{data} - \vec{y}\| = \min_{\vec{y} \in col(A)} \left(\|y_{data_1} - y_1\|^2 + \|y_{data_2} - y_2\|^2 + \dots + \|y_{data_k} - y_k\|^2 \right)$$

From this system we can get the least squares solution to the problem. That is $\vec{x}_{ls}$ is the solution to $A\vec{x} = \vec{y}_{ls}$ which we know will exist since $\vec{y}_{ls} \in col(A)$ so we just need to find this vector!

In order to find this magic $\vec{y}_{ls}$ we use the following theorem

Let $\vec{y}$ be a vector, and let $\vec{y}_s \in S$ be the vector in S that is closest to $\vec{y}$. Then,

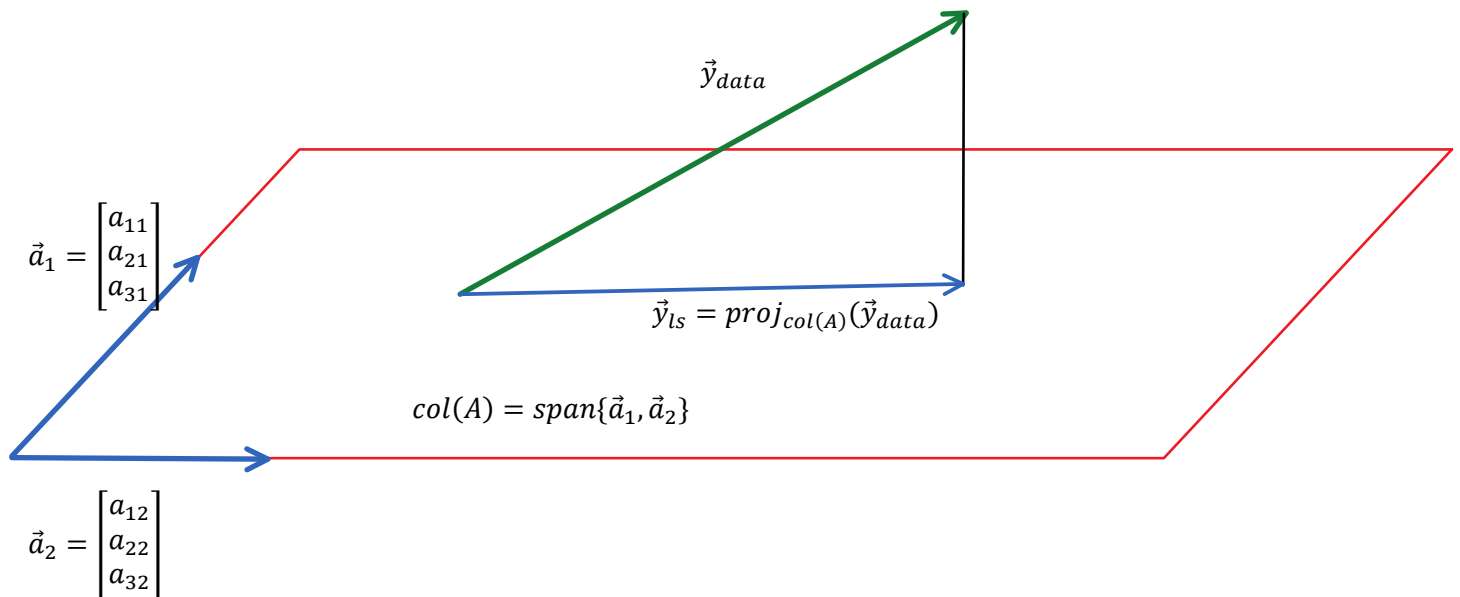
$$\vec{y}_s = proj_S(\vec{y})$$

So with this theorem if $S = col(A)$ the vector we want above is just the $proj_{col(A)}(\vec{y}_{data})$ which we can calculate easily! We would just Gram-Schmidt the column vectors of A to get an orthonormal basis and then project $\vec{y}_{data}$ on to the orthonormal basis.

Section 8.5 Least Squares Solutions

A diagram of a situation we can visualize. Suppose we measure 3 data points to which we want to fit

a line. $\vec{y}_{data} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ our system, $A\vec{x} = \vec{y}_{data}$ would be: $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$.



We can solve $A\vec{x} = \vec{y}_{ls}$ because $\vec{y}_{ls} \in col(A) = span\{\vec{a}_1, \vec{a}_2\}$, but note there would be no linear combination of $\{\vec{a}_1, \vec{a}_2\}$ that could get us to $\vec{y}_{data}$. So, instead we project it and find the vector in the column space of A which is closest to $\vec{y}_{data}$.

An algorithm for doing this would be the following.

1. Apply the Gram-Schmidt method to the column vectors of A, $\{\vec{a}_1, \dots, \vec{a}_m\}$, to obtain an orthonormal basis, $\{\vec{v}_1, \dots, \vec{v}_m\}$ for $col(A)$.

2. Now that we have an orthonormal basis for $col(A)$, call it we can project $\vec{y}_{data}$ using the formula

$$\vec{y}_{ls} = proj_{col(A)}(\vec{y}_{data}) = (\vec{v}_1 \cdot \vec{y}_{data})\vec{v}_1 + (\vec{v}_2 \cdot \vec{y}_{data})\vec{v}_2 + \dots + (\vec{v}_m \cdot \vec{y}_{data})\vec{v}_m$$

3. Solve the new system $A\vec{x} = \vec{y}_{ls}$ any way you like to obtain the least squares solution $\vec{x}_{ls}$

Section 8.5 Least Squares Solutions

```
%initialize the data
phi=[0.068309;0.066612;0.05029;0.04422;0.057476;0.063163;0.0534;0.061583];
sigma=[0.00897;0.006353;0.003431;0.003045;0.003223;0.005508;0.003045;0.006294];
% take log of each data point
lphi=log10(phi);
lsigma=log10(sigma);
%Build coefficient matrix representing each log(sigma)=mlog(phi)+b
A=[lphi ones(8,1)];
% Augment A with the right handside data, that way we could rref for a
% solution
B=[lphi ones(8,1) lsigma];
% Not nesscary since we know there is not solution, but we could try
rref(B)
%Since there is no solution we will seek the least squares solution
%First we need an orthornormal basis cor col(A), with out the augmentation.
%Here we perform Gram-Schmidt to find it
v1=A(:,1);
v2=A(:,2)-(dot(A(:,1),A(:,2))/norm(A(:,1))^2)*A(:,1);
v1=v1/norm(v1);
v2=v2/norm(v2);
%Once we have the basis for col(A) we project the right hand side vector
%onto col(A) to find the vector in col(A) closest to the y data
yls=(dot(lsigma,v1)/norm(v1)^2)*v1+(dot(lsigma,v2)/norm(v2)^2)*v2;
%Once we have the vector in col(A) closest to our data we just replace it
%in our augmented matrix and solve!
B(:,3)=yls
LS=rref(B)
%Plot the line we find with the data
x=[min(lphi):0.001:max(lphi)];
y=LS(1,3)*x+LS(2,3);
plot(lphi,lsigma,'o')
hold on
plot(x,y)
hold off
```

B =

-1.1655	1.0000	-2.0472
-1.1764	1.0000	-2.1970
-1.2985	1.0000	-2.4646
-1.3544	1.0000	-2.5164
-1.2405	1.0000	-2.4917
-1.1995	1.0000	-2.2590
-1.2725	1.0000	-2.5164
-1.2105	1.0000	-2.2011

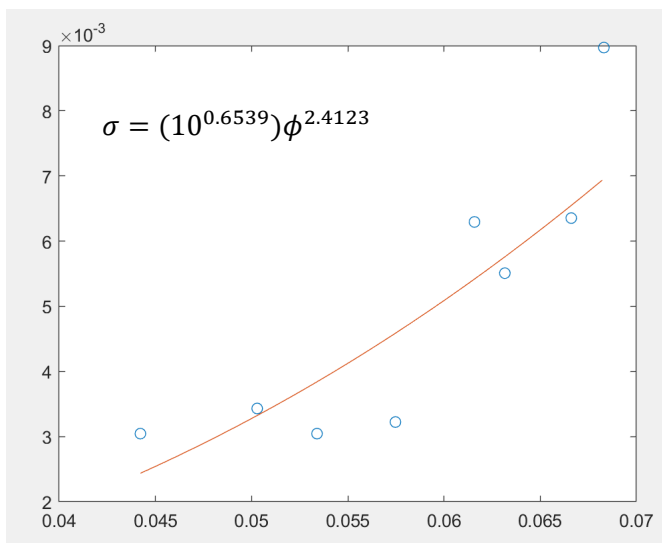
No solution to the system above
so we project the data to the
column space of A and solve that

B =

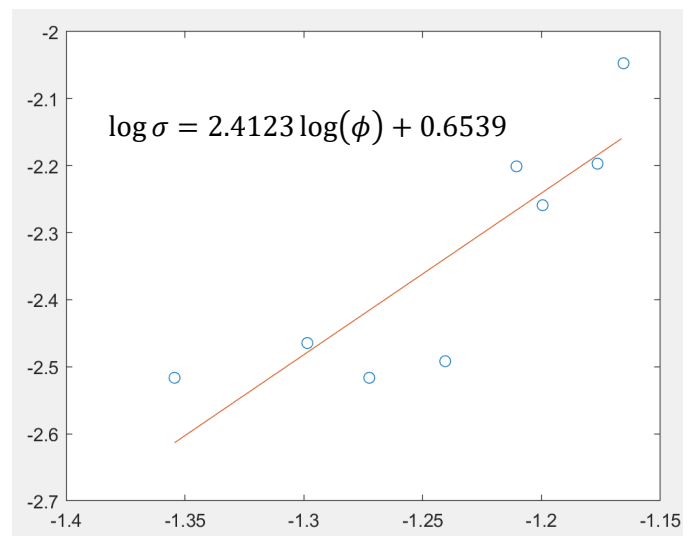
-1.1655	1.0000	-2.1576
-1.1764	1.0000	-2.1840
-1.2985	1.0000	-2.4785
-1.3544	1.0000	-2.6132
-1.2405	1.0000	-2.3385
-1.1995	1.0000	-2.2397
-1.2725	1.0000	-2.4156
-1.2105	1.0000	-2.2662

LS =

1.0000	0	2.4123
0	1.0000	0.6539
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0



Going back to the original variables



The best fit to the log- log data

Not bad! This is in fact the best possible solution we could find given the data.

Section 8.5 Least Squares Solutions

In the work we just did we took a system $A\vec{x} = \vec{y}$ which had no solution and projected the vector $\vec{y}$ onto $col(A)$ to find a right hand side, $\vec{y}_{ls}$ that permits a solution to exist and is as close to the original $\vec{y}$ as possible. Then we just solve the new system. Hopefully this is somewhat intuitive that this should give us the best possible solution we could have. However, there is another way that we can think of this and it produces the exact same "best solution" $\vec{x}_{ls}$.

If A is an $n \times m$ matrix and $\vec{y} \in \mathbb{R}^n$ then the least squares solution to $A\vec{x} = \vec{y}$ is a vector $\vec{x}_{ls} \in \mathbb{R}^m$ such that :

$$\|A\vec{x}_{ls} - \vec{y}\| \leq \|A\vec{x} - \vec{y}\| \text{ for all } \vec{x} \in \mathbb{R}^m$$

This definition is just saying that the least squares solution should be the vector $\vec{x}$ that minimizes the distance between $A\vec{x}$ and $\vec{y}$. One thing to note, if the system has an exact solution, then it is that solution that we would find since it makes the distance as small as possible, namely 0.

Another way to find $\vec{x}_{ls}$.

The least squares solution to $A\vec{x} = \vec{y}$ is equal to the set of solutions to the system

$$A^T A \vec{x} = A^T \vec{y}$$

Which we call the normal equations. If A has linearly independent columns, then there is a unique least squares solution given by

$$\vec{x}_{ls} = (A^T A)^{-1} A^T \vec{y}$$

If the columns of A are not linearly independent, then we have infinitely many solutions given by the same formula above.

$(A^T A)^{-1} A^T$ is an example of the Moore-Penrose Pseudo Inverse. (Something fun to google.)

We could solve our problem from before much more easily!

A =

```
-1.1655  1.0000
-1.1764  1.0000
-1.2985  1.0000
-1.3544  1.0000
-1.2405  1.0000
-1.1995  1.0000
-1.2725  1.0000
-1.2105  1.0000
```

xls=inv(A'*A)*A'*lsigma

xls =

```
2.4123
0.6539
```

Section 8.5 Least Squares Solutions

The least squares solution to $A\vec{x} = \vec{y}$ is equal to the set of solutions to the system

$$A^T A \vec{x} = A^T \vec{y}$$

Which we call the normal equations. If A has linearly independent columns, then there is a unique least squares solution given by

$$\vec{x}_{ls} = (A^T A)^{-1} A^T \vec{y}$$

If the columns of A are not linearly independent, then we have infinitely many solutions given by the same formula above.

$(A^T A)^{-1} A^T$ is an example of the Moore-Penrose Pseudo Inverse. (Something fun to google.)

Proof: Consider the problem we solved before,

$$A\vec{x} = \text{proj}_{\text{col}(A)}(\vec{y})$$

Let $\vec{x}_{ls}$ be the solution to that problem. So that we have the equality

$$A\vec{x}_{ls} = \text{proj}_{\text{col}(A)}(\vec{y})$$

We want to show that if $\vec{x}_{ls}$ solves the above equation then it solves $A^T A \vec{x} = A^T \vec{y}$.

In indeed $\vec{x}_{ls}$ is a solution to $A^T A \vec{x} = A^T \vec{y}$ it must be that.

$$A^T \vec{y} - A^T A \vec{x}_{ls} = 0$$

$$A^T (\vec{y} - A \vec{x}_{ls}) = 0$$

$$A^T (\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y})) = 0$$

So if $\vec{x}_{ls}$ is to be a solution it would need to be that $\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y}) \in \text{null}(A^T)$ Let's see if it is.

At the very least we know that if we remove the part of $\vec{y}$ that lives in the $\text{col}(A)$ we will get a vector orthogonal to $\text{col}(A)$, which would be a vector in $(\text{col}(A))^\perp$. Thus we could say,

$$\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y}) \in (\text{col}(A))^\perp$$

Ok so if we can show that $(\text{col}(A))^\perp = \text{null}(A^T)$ then we know $\vec{x}_{ls}$ will solve $A^T A \vec{x} = A^T \vec{y}$.

Section 8.5 Least Squares Solutions

Suppose that $\vec{v} \in (\text{col}(A))^\perp$ this directly means that $\vec{v}$ is a vector which is orthogonal to all possible linear combinations of the column vectors of A . So if $A = [\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_m]$ then $\vec{v}$ would need to satisfy,

$$(c_1 \vec{a}_1 + c_2 \vec{a}_2 + \cdots + c_m \vec{a}_m) \cdot \vec{v} = 0$$

$$c_1 \vec{a}_1 \cdot \vec{v} + c_2 \vec{a}_2 \cdot \vec{v} + \cdots + c_m \vec{a}_m \cdot \vec{v} = 0$$

$$c_1 (\vec{a}_1)^T \vec{v} + c_2 (\vec{a}_2)^T \vec{v} + \cdots + c_m (\vec{a}_m)^T \vec{v} = 0$$

$$\begin{bmatrix} c_1 (\vec{a}_1)^T \\ c_2 (\vec{a}_2)^T \\ \vdots \\ c_k (\vec{a}_k)^T \end{bmatrix} \vec{v} = 0$$

So we can say $\vec{v} \in \text{null} \left(\begin{bmatrix} c_1 (\vec{a}_1)^T \\ c_2 (\vec{a}_2)^T \\ \vdots \\ c_k (\vec{a}_k)^T \end{bmatrix} \right)$. Noting that $\{c_1 (\vec{a}_1)^T, \dots, c_m (\vec{a}_m)^T\}$ and $\{(\vec{a}_1)^T, \dots, (\vec{a}_m)^T\}$ span the exact same space it must be that

$$\text{null} \left(\begin{bmatrix} c_1 (\vec{a}_1)^T \\ c_2 (\vec{a}_2)^T \\ \vdots \\ c_k (\vec{a}_k)^T \end{bmatrix} \right) = \text{null} \left(\begin{bmatrix} (\vec{a}_1)^T \\ (\vec{a}_2)^T \\ \vdots \\ (\vec{a}_k)^T \end{bmatrix} \right) = \text{null}(A^T)$$

So, if $\vec{v} \in (\text{col}(A))^\perp$ then $\vec{v} \in \text{null}(A^T)$ and since each space would have the same number of basis vectors, and, one is contained in the other they must have been the same. Infact $(\text{col}(A))^\perp = \text{null}(A^T)$.

So

$$A^T (\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y})) = 0$$

$$A^T (\vec{y} - A\vec{x}_{ls}) = 0$$

$$A^T \vec{y} - A^T A\vec{x}_{ls} = 0$$

$$A^T A\vec{x}_{ls} = A^T \vec{y}$$

So we can find our least squares solution by just solving $A^T A\vec{x} = A^T \vec{y}$!!

Section 8.5 Least Squares Solutions

Find the least squares solution(s) for the given system. (If there are an infinite number of solutions use t as your parameter.)

$$-x_1 - x_2 = 4$$

$$2x_1 + 4x_2 = -1$$

$$-4x_1 + 2x_2 = 2$$

$$(x_1, x_2) = \left(3a, -\frac{29}{44}, -\frac{7}{44} \right)$$

Section 8.5 Least Squares Solutions

Exam 2 Score Booster for this score booster you have two options.

Option 1: Analyzing Real Data

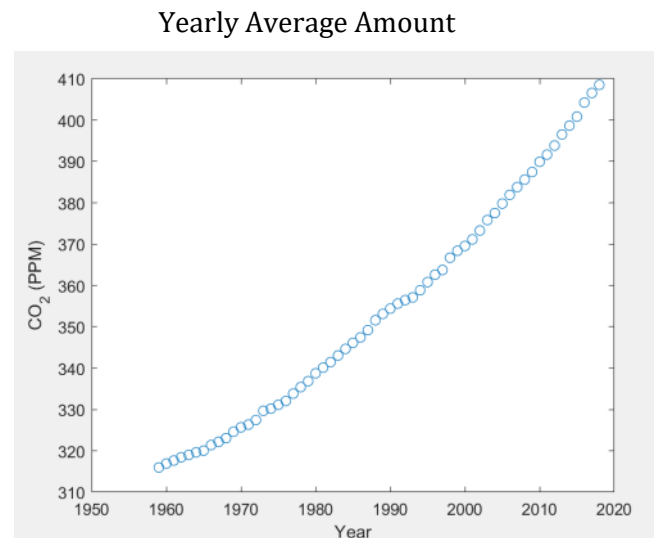
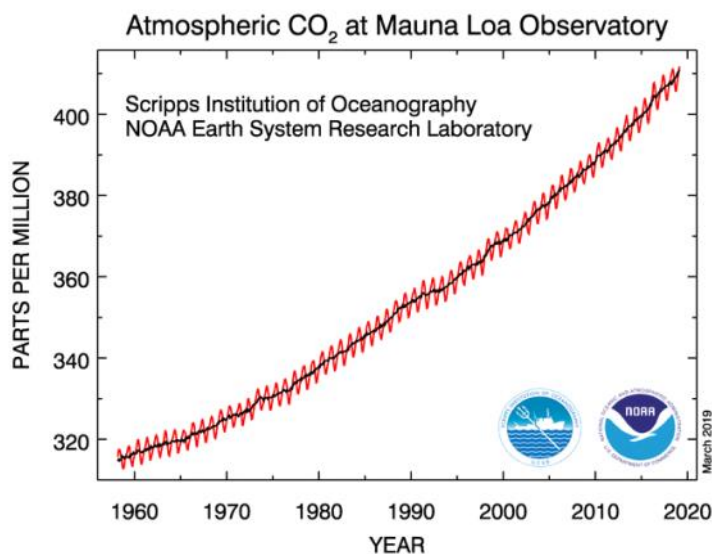
Since 1958 The Mauna Loa Observatory has been monitoring the amount of CO₂ in the earth Atmosphere. You can see this data in the plots below, clearly it is increasing, and geological terms it is increasing rapidly! Note the wiggles in the left figure, they represent seasonality. In the winter the Northern Hemisphere plants stop trapping as much Co₂ and the amount goes up a bit, in the summer they trap lots and it goes down a bit. Neat! One very interesting question we might like to answer is how is Co₂ increasing? Is it linear, exponential, quadratic? In the plots below there is a slight curve in the data, but it almost looks like a nice line as well. Your job will be to find least squares solutions to the data for the following function types:

$$y = ax^m$$

$$y = ax^2 + bx + c$$

$$y = ax + b$$

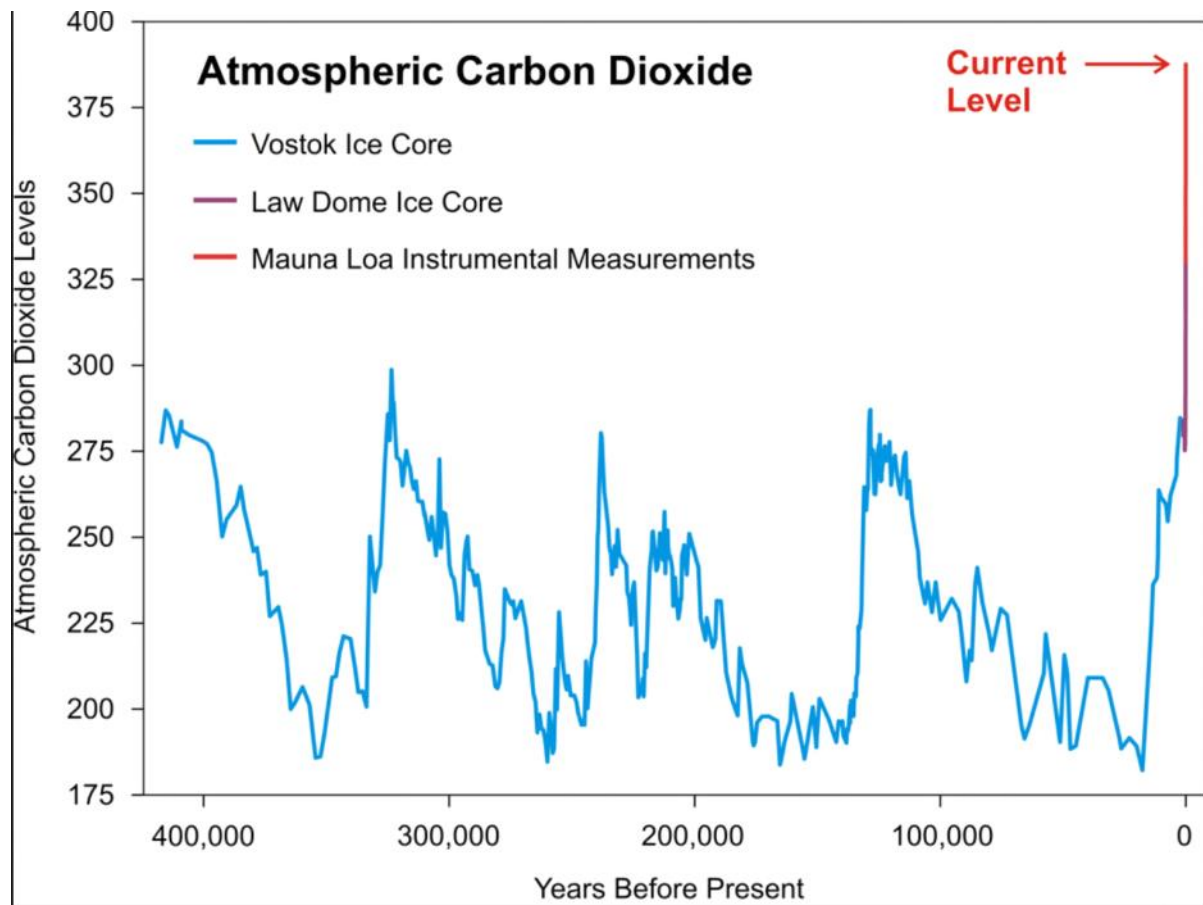
For the data found in the Co2yearlyaverage.txt file on Sakai in the resources folder. Note you won't need the uncertainty listed in the third column of the file for this.



In your analysis take 1959 to be year 1. So your x values will go from 1-60 with your y values the corresponding amount of CO₂ that year. For each of the function types above find the least squares best fit. Plot all three on the same axes with the data points themselves and decide which you think fits best. For credit you will need to include the matrices for each linear system you write, the values given from the least squares solutions, your plot as described above, and the code you used to solve the problem. You will need to upload all of this as a PDF to gradescope no later than the Tuesday after the exam.

Section 8.5 Least Squares Solutions

Here is a bonus chart related to the previous problem.



This chart shows global CO_2 levels inferred from glacial ice core data along with our current level of CO_2 . Notably, an increase of 100 ppm naturally takes 5000-10000 years or so, we have done about the same in 100!

Section 8.5 Least Squares Solutions

Option 2: In Experimental science least squares solutions are all over the place! Often we suspect a particular algebraic form models a real phenomenon and want to know what its parameters are. Given that measurements always have error we take lots and often end up doing least squares.

For this option you will conduct a simple experiment to measure the acceleration of gravity on earth.

What to do:

1. Find something that won't suffer too much from air resistance.
2. With a timer drop it from different heights and measure how long it takes to hit the ground from those heights. Make at least 10 measurements.
3. Physics says that the height of a falling object dropped from an initial height h_0 is given by

$$h(t) = -\frac{1}{2}gt^2 + h_0$$

So the amount of time t it would take for the object to get to $h = 0$ from h_0 would be

$$h_0 = \frac{1}{2}gt^2$$

The data you take will be t 's and h_0 's. Take your 10 measurements and create a linear system to find the least squares value of g , the acceleration due to gravity on earth, that gives a best fit to the data.

For credit you will need to turn the following things in PDF form submitted to gradescope no later than the Tuesday after the exam.

1. A picture of you dropping your object
2. The linear system you used to solve the problem
3. A plot of your best fit curve with your data points
4. The code you used to find the least squares solution.



Section 8.5 Least Squares Solutions

Together in Octave!

Measurements of CO_2 in the atmosphere have been taken regularly over the last 50 years at an observatory. In addition to a general upward trend, the CO_2 data also has an annual cyclic behavior. The table below has the monthly measurements (in parts per million) for 2009. (A graphing calculator is recommended. Let $t = 1$ represent January of 2009.)

Month	CO_2		Month	CO_2
1	386.94		7	387.72
2	387.41		8	385.93
3	388.79		9	384.77
4	389.48		10	384.38
5	390.16		11	385.99
6	389.43		12	387.27

Find constants a , b , and c such that the model $y = a + bt + c \sin(t\pi/6)$ best fits the data, where t is time in months. (Round your final values to four decimal places.)

$(a, b, c) = (10a \text{ } 386.9609, 0.0608, 2.1290)$

Use your model to predict the CO_2 level in January 2020. (Round your answer to two decimal places.)

10b 396.11 parts per million