

## Method of Least Squares

An interesting example of how to use the Approximation Theorem is to consider the situation where we have a collection of points, and we want to find the polynomial that comes closest to having these points on them.

- Calculus tackles this problem using equations.
- In linear algebra, we want to use vectors and matrices to find the answer.

And I mentioned the approximation theorem because we are going to set up a vector  $\vec{y}$  of our data points, and have our subspace be the set of all polynomials of the appropriate degree (translated into  $\mathbb{R}^n$ ), so the idea will be that we are looking for the vector/polynomial that is closest to  $\vec{y}$ .

### Example

Find the equation of the form  $y = a + bt$  that best fits the points  $\{(1, 7), (2, 10), (3, 12)\}$ .

The first thing we need to decide is what we mean by best fit. In general, we want to compare the difference between between our data points and our equation, and make that difference as small as possible.

$$\begin{aligned} (1, 7) - (1, a + b) &= (0, 7 - a - b) \\ (2, 10) - (2, a + 2b) &= (0, 10 - a - 2b) \\ (3, 12) - (3, a + 3b) &= (0, 12 - a - 3b) \end{aligned}$$

Since the first component of the difference will always be 0, we can ignore it. So, we want to minimize  $(7 - a - b) + (10 - a - 2b) + (12 - a - 3b)$ , right?

The problem with this calculation is that there could be large negative and positive values cancelling out.

## Method of Least Squares

### Example

We could look at minimizing the absolute value of the differences, but it is actually much easier to simply square all the values to ensure they are positive.

And thus we have the derivation of the name for our technique: **the Method of Least Squares**.

So, we are trying to find values for  $a$  and  $b$  that minimize

$$(7 - a - b)^2 + (10 - a - 2b)^2 + (12 - a - 3b)^2$$

This is the point that we want to switch into vector mode, as we can think of this as the following:

$$\left\| \begin{bmatrix} 7 - a - b \\ 10 - a - 2b \\ 12 - a - 3b \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} 7 \\ 10 \\ 12 \end{bmatrix} - \left( a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right) \right\|^2$$

Trying to start the generalization process, we can set

$$\begin{aligned} \vec{y} &= \begin{bmatrix} 7 \\ 10 \\ 12 \end{bmatrix}, \text{ our vector of data values,} \\ \vec{1} &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \text{ and} \\ \vec{t} &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \text{ our vector of } t \text{ values from our data points.} \end{aligned}$$

## Method of Least Squares

### Example

Then we are trying to find  $a$  and  $b$  that minimize  $\|\vec{y} - (a\vec{1} + b\vec{t})\|^2$ . If we let  $\mathbb{S} = \text{Span}\{\vec{1}, \vec{t}\}$ , then we are looking for  $\vec{s} \in \mathbb{S}$  that minimizes  $\|\vec{y} - \vec{s}\|^2$ .

By the Approximation Theorem, we know that this is  $\text{proj}_{\mathbb{S}}\vec{y}$ . So, should we find an orthogonal basis for  $\mathbb{S}$  and proceed to find  $\text{proj}_{\mathbb{S}}\vec{y}$ ?

No. That would produce an answer in the coordinates for the orthogonal basis when we want our answer to be in coordinates with respect to the set  $\{\vec{1}, \vec{t}\}$ . But the more important reason is that we don't have to.

Instead, consider that if  $a\vec{1} + b\vec{t} = \text{proj}_{\mathbb{S}}\vec{y}$ , then

$$\vec{y} - (a\vec{1} + b\vec{t}) = \vec{y} - \text{proj}_{\mathbb{S}}\vec{y} = \text{perp}_{\mathbb{S}}\vec{y}$$

This value is called the **error vector**  $\vec{e}$ , since it is the vector that measures the difference between our data values  $\vec{y}$  and the equation values  $a\vec{1} + b\vec{t}$ . But I've shown that

$$\vec{e} = \text{perp}_{\mathbb{S}}\vec{y}$$

which means that

$$\vec{e} \in \mathbb{S}^{\perp}$$

Specifically, we get that

$$\vec{1} \cdot \vec{e} = 0 \text{ and } \vec{t} \cdot \vec{e} = 0$$

Recalling that  $\vec{e} = \vec{y} - (a\vec{1} + b\vec{t})$ , we get  $\vec{1} \cdot (\vec{y} - (a\vec{1} + b\vec{t})) = 0$  and  $\vec{t} \cdot (\vec{y} - (a\vec{1} + b\vec{t})) = 0$

## Method of Least Squares

### Example

The next step is to introduce a matrix:

Let  $X = \begin{bmatrix} \vec{1} & \vec{t} \end{bmatrix}$ , and let  $\vec{a} = \begin{bmatrix} a \\ b \end{bmatrix}$ . Then  $a\vec{1} + b\vec{t} = X\vec{a}$ , so  $\vec{y} - (a\vec{1} + b\vec{t}) = \vec{y} - X\vec{a}$ . And we get

$$\begin{aligned} X^T(\vec{y} - X\vec{a}) &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} (\vec{y} - X\vec{a}) \\ &= \begin{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot (\vec{y} - X\vec{a}) \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot (\vec{y} - X\vec{a}) \end{bmatrix} \\ &= \begin{bmatrix} \vec{1} \cdot (\vec{y} - X\vec{a}) \\ \vec{t} \cdot (\vec{y} - X\vec{a}) \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

## Method of Least Squares

### Example

So we've got

$$X^T(\vec{y} - X\vec{a}) = \vec{0}$$

Distributing  $X^T$ , we get

$$X^T\vec{y} - X^T X\vec{a} = \vec{0}$$

Solving for  $\vec{a}$ , we first get

$$X^T X\vec{a} = X^T\vec{y}$$

Now, if  $X^T X$  is invertible, then we get

$$\vec{a} = (X^T X)^{-1} X^T\vec{y}$$

Is  $X^T X$  invertible?

Well, this goes back to us choosing the appropriate  $t$  values, so again, in this course, you can always assume the answer is yes. And this means we have found  $\vec{a} = \begin{bmatrix} a \\ b \end{bmatrix}$ .

Except for actually calculating  $(X^T X)^{-1} X^T\vec{y}$ .

## Method of Least Squares

### Example

First, we compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

Next, we use the matrix inverse algorithm to compute  $(X^T X)^{-1}$ :

$$\left[ \begin{array}{cc|cc} 3 & 6 & 1 & 0 \\ 6 & 14 & 0 & 1 \end{array} \right] \sim \left[ \begin{array}{cc|cc} 3 & 6 & 1 & 0 \\ 0 & 2 & -2 & 1 \end{array} \right] \sim \left[ \begin{array}{cc|cc} 3 & 0 & 7 & -3 \\ 0 & 2 & -2 & 1 \end{array} \right] \sim \left[ \begin{array}{cc|cc} 1 & 0 & 7/3 & -1 \\ 0 & 1 & -1 & 1/2 \end{array} \right]$$

So

$$(X^T X)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix}$$

which means that

$$(X^T X)^{-1} X^T = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 4/3 & 1/3 & -2/3 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

Finally we have

$$(X^T X)^{-1} X^T\vec{y} = \begin{bmatrix} 4/3 & 1/3 & -2/3 \\ -1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 7 \\ 10 \\ 12 \end{bmatrix} = \begin{bmatrix} 14/3 \\ 5/2 \end{bmatrix}$$

## Method of Least Squares

### Example

And so we see that the line that is the best approximation for the data point  $\{(1, 7), (2, 10), (3, 12)\}$  is

$$y = \frac{14}{3} + \frac{5}{2}t$$

Generalizing this example, we see that to find the equation  $a_0 + a_1t + \dots + a_nt^n$  that is the best fit for data points  $\{(t_1, y_1), \dots, (t_k, y_k)\}$ , we want to set

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}, \vec{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}, \vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

(where  $\vec{1}$  is the same size as  $\vec{y}$ )

$$\vec{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_k \end{bmatrix} \rightarrow \vec{t^2} = \begin{bmatrix} t_1^2 \\ \vdots \\ t_k^2 \end{bmatrix} \text{ and so on, getting } \vec{t^n} = \begin{bmatrix} t_1^n \\ \vdots \\ t_k^n \end{bmatrix}$$

Then we will set  $X = \begin{bmatrix} \vec{1} & \vec{t} & \dots & \vec{t^n} \end{bmatrix}$ , and using the same technique as our example, we will find that

$$\vec{a} = (X^T X)^{-1} X^T \vec{y}$$

## Method of Least Squares

### Example

Find the equation of the form  $a + bt + ct^2$  that best fits the data points  $\{(-1, 4), (0, 7), (1, 12)\}$ .

So we set

$$\vec{a} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \vec{y} = \begin{bmatrix} 4 \\ 7 \\ 12 \end{bmatrix}, \vec{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \vec{t} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \text{ and } \vec{t^2} = \begin{bmatrix} (-1)^2 \\ 0^2 \\ 1^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Now, we can create our matrix

$$X = \begin{bmatrix} \vec{1} & \vec{t} & \vec{t^2} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

and we can begin our calculations:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

Next, we need to find  $(X^T X)^{-1}$  using the matrix inverse algorithm:

$$\begin{bmatrix} 3 & 0 & 2 & | & 1 & 0 & 0 \\ 0 & 2 & 0 & | & 0 & 1 & 0 \\ 2 & 0 & 2 & | & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2/3 & | & 1/3 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 1/2 & 0 \\ 2 & 0 & 2 & | & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2/3 & | & 1/3 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 1/2 & 0 \\ 0 & 0 & 2/3 & | & -2/3 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & | & 1 & 0 & -1 \\ 0 & 1 & 0 & | & 0 & 1/2 & 0 \\ 0 & 0 & 1 & | & -1 & 0 & 3/2 \end{bmatrix}$$

## Method of Least Squares

### Example

And so we have that

$$(X^T X)^{-1} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{bmatrix}$$

We can now calculate

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}$$

And at long last we get

$$\vec{a} = (X^T X)^{-1} X^T \vec{y} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \\ 12 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ 1 \end{bmatrix}$$

So the equation  $7 + 4t + t^2$  is the equation of the form  $a + bt + ct^2$  that best fits the data points  $\{(-1, 4), (0, 7), (1, 12)\}$ .

## Method of Least Squares

### Note:

- The matrix  $X$ , and thus the matrix  $(X^T X)^{-1} X^T$ , is only dependent on the form of the equation we are trying to fit and the experimental  $t$  values. But not the  $y$  values. You can imagine a situation where an experiment is run several times, so the  $t$  values stay constant, but the  $y$  values vary. So we could reuse our matrix  $(X^T X)^{-1} X^T$  for each of our  $y$  vectors, which cuts down on work.
- If we want one of the  $a_i$  values in our goal equation to be 0, we need to remove that entry from our list vectors  $\vec{t}^i$  when creating  $X$ . If we don't, our technique may (in fact, probably will) present us with a solution where the  $a_i$  value is not 0. Basically, we want to make sure that our matrix  $X$  is a custom fit for our equation.

## Method of Least Squares

### Example

Find the equation of the form  $at + bt^3$  that best fits the data points  $\{(-2, -10), (-1, -3), (0, 1), (1, 5), (2, 12)\}$ .

For equations of the form  $at + bt^3$ , we will want our matrix  $X$  to be  $\begin{bmatrix} \vec{t} & \vec{t^3} \end{bmatrix}$ , where

$$\vec{t} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \text{ and } \vec{t^3} = \begin{bmatrix} -8 \\ -1 \\ 0 \\ 1 \\ 8 \end{bmatrix}$$

With  $X$  determined, we continue with our calculations as before:

$$X^T X = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -8 & -1 & 0 & 1 & 8 \end{bmatrix} \begin{bmatrix} -2 & -8 \\ -1 & -1 \\ 0 & 0 \\ 1 & 1 \\ 2 & 8 \end{bmatrix} = \begin{bmatrix} 10 & 34 \\ 34 & 130 \end{bmatrix}$$

Next, we need to find  $(X^T X)^{-1}$  using the matrix inverse algorithm:

$$\begin{aligned} \left[ \begin{array}{cc|cc} 10 & 34 & 1 & 0 \\ 34 & 130 & 0 & 1 \end{array} \right] &\sim \left[ \begin{array}{cc|cc} 1 & 34/10 & 1/10 & 0 \\ 34 & 130 & 0 & 1 \end{array} \right] &\sim \left[ \begin{array}{cc|cc} 1 & 34/10 & 1/10 & 0 \\ 0 & 144/10 & -34/10 & 1 \end{array} \right] \\ &\sim \left[ \begin{array}{cc|cc} 1 & 34/10 & 1/10 & 0 \\ 0 & 1 & -34/144 & 10/344 \end{array} \right] &\sim \left[ \begin{array}{cc|cc} 1 & 0 & 130/144 & -34/144 \\ 0 & 1 & -34/144 & 10/144 \end{array} \right] \end{aligned}$$

## Method of Least Squares

And so we have that

$$(X^T X)^{-1} = \begin{bmatrix} 130/144 & -34/144 \\ -34/144 & 10/144 \end{bmatrix} = \frac{1}{72} \begin{bmatrix} 65 & -17 \\ -17 & 5 \end{bmatrix}$$

We can now calculate  $(X^T X)^{-1} X^T$ :

$$\begin{aligned} (X^T X)^{-1} X^T &= \frac{1}{72} \begin{bmatrix} 65 & -17 \\ -17 & 5 \end{bmatrix} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -8 & -1 & 0 & 1 & 8 \end{bmatrix} \\ &= \frac{1}{72} \begin{bmatrix} 6 & -48 & 0 & 48 & -6 \\ -6 & 12 & 0 & -12 & 6 \end{bmatrix} \\ &= \frac{1}{12} \begin{bmatrix} 1 & -8 & 0 & 8 & -1 \\ -1 & 2 & 0 & -2 & 1 \end{bmatrix} \end{aligned}$$

And we get:

$$\vec{a} = (X^T X)^{-1} X^T \vec{y} = \frac{1}{12} \begin{bmatrix} 1 & -8 & 0 & 8 & -1 \\ -1 & 2 & 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} -10 \\ -3 \\ 1 \\ 5 \\ 12 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 42 \\ 6 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 1/2 \end{bmatrix}$$

So the equation  $\frac{7}{2}t + \frac{1}{2}t^3$  is the equation of the form  $at + bt^3$  that best fits the data points  $\{(-2, -10), (-1, -3), (0, 1), (1, 5), (2, 12)\}$ .