

# Least Squares Method

## MATH 3512, BCIT

Matrix Methods and Statistics for Geomatics

October 12, 2018

The **projection**  $u_H$  of a vector  $u$  onto a hyperplane  $H$  is the vector in the hyperplane that is “most similar” to  $u$ . The formal definition for  $u_H$  requires that

- 1  $u$  is in  $H$
- 2  $(u - u_H)$  is orthogonal to all basis vectors of  $H$

**Example 1: Finding a Projection.** Let  $H$  be the line spanned by  $\vec{v} = (-1, 1)^\top$  in  $\mathbb{R}^2$ . What is the projection  $\vec{w}$  of  $\vec{u} = (3, -2)^\top$ ?



Let  $\vec{w} = (w_1, w_2)^\top$ . Then (1)  $\vec{u} - \vec{w}$  is orthogonal to  $\vec{v}$  and (2)  $\vec{w} = \alpha \vec{v}$  for some  $\alpha \in \mathbb{R}$ .

$$\begin{aligned} w_1 - w_2 &= 5 \\ w_1 + w_2 &= 0 \end{aligned} \tag{1}$$

Cramer's rule tells us that  $\vec{w} = (2.5, -2.5)^\top$ .

Let  $u = (u_1, \dots, u_n)^T$  be a vector and  $H$  be a  $k$ -dimensional hyperplane in the vector space  $\mathbb{R}^n$ . Let  $x_1, \dots, x_k$  be a basis for  $H$ . Then it is true for all vectors  $v$  in the hyperplane that

$$\|u - v\| \geq \|u - u_H\| \quad (2)$$

Proof: use the theorem of Pythagoras for

$$\|u - v\|^2 = \|u - u_H\|^2 + \|u_H - v\|^2 \geq \|u - u_H\|^2 \quad (3)$$

The claim follows. It illustrates what I mean when I say that  $u_H$  is the vector in  $H$  that is most similar to  $u$ .

**Example 2: Finding Another Projection.** What is the projection of  $\vec{u} = (5, 2, 10)^\top$  onto the plane  $T$  characterized by  $2x + y + 3z = 0$ ?

First we find two linearly independent vectors in  $H$  to form a basis of  $H$ , for example  $\vec{v}_1 = (1, 1, -1)^\top$  and  $\vec{v}_2 = (0, -3, 1)^\top$ . The conditions

- ①  $u_H \in T$
- ②  $(u - u_H) \perp v_1$
- ③  $(u - u_H) \perp v_2$

give us the system of linear equations

$$\begin{bmatrix} 2 & 1 & 3 \\ -1 & 1 & 1 \\ 0 & 3 & -1 \end{bmatrix} \cdot \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ -4 \end{bmatrix} \quad (4)$$

for which the solution is  $u_H = (\hat{x}, \hat{y}, \hat{z})^\top = (-1, -1, 1)^\top$ .

Let there be two linearly independent vectors  $u$  and  $v$  in  $\mathbb{R}^n$ . Then the formula for the projection  $u_v$  of  $u$  onto the line spanned by  $v$  is

$$u_v = \left( \frac{u \cdot v}{v \cdot v} \right) v \quad (5)$$

To verify the formula, note that  $u_v = av$  for some real number  $a$ . Therefore

$$(u - av) \perp v \quad (6)$$

Isolate  $a$  in the equation  $(u - av) \cdot v = 0$  to yield the formula.

Formula (5) only works when the hyperplane is a line. You can scale up the idea in terms of dimensions by the following theorem.

## Formula for Projection Onto Plane with Orthogonal Basis

Let  $\{u, v\}$  be an orthogonal basis for  $H$ . Then the projection of  $w$  onto  $H$  is the sum of  $w_u$  and  $w_v$ , the projections of  $w$  onto the lines spanned by  $u$  and  $v$ , respectively.

Proof: check the following

- 1  $(w_u + w_v) \in H$  (trivial)
- 2  $(w - (w_u + w_v)) \perp u$  (use the fact that  $u \perp v$ )
- 3  $(w - (w_u + w_v)) \perp v$  (same idea)

# Least Squares Method

Consider the following table of measurements for the length of shoe prints and the height of the person wearing the shoes.

Shoe Print (cm)	Height (cm)
29.7	175.3
29.9	177.8
31.4	185.4
31.8	175.3
27.6	172.7

In the statistics portion of this course, we will learn whether the paired data provide evidence of a linear relationship. In the linear algebra portion, we will learn how to find the line which is closest to the data points in the least squares sense.



# Least Squares Method



# Least Squares Method

## Least Squares Method

If  $L$  is a given line, the **error** for each data point is the vertical distance from that point to the line. The **squared error** is the sum of the squares of the errors. The line that best fits the data in the least squares sense is the line that minimizes the squared error.

You can find the **regression line** using calculus optimization. However, there is also an elegant method using linear algebra.

# Least Squares Method

Let  $L$  be a line with slope  $m$  and  $y$ -intercept  $b$ . Let  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  be a set of paired data. Then the following equations hold:

$$\begin{aligned} y_1 &= mx_1 + b + \epsilon_1 \\ y_2 &= mx_2 + b + \epsilon_2 \\ &\vdots \\ y_n &= mx_n + b + \epsilon_n \end{aligned} \tag{7}$$

where the  $\epsilon_i$  are the errors ( $i = 1, \dots, n$ ). This system is equivalent to the following vector equation,

$$Y = AV + E \tag{8}$$

where  $Y, A, V, E$  are defined on the next slide.

# Least Squares Method

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, V = \begin{bmatrix} m \\ b \end{bmatrix}, E = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$E$  is called the error vector. According to (8), it is

$$E = Y - AV \quad (9)$$

We are trying to choose  $m, b$  so that

$$\|E\|^2 = \|Y - AV\|^2 \quad (10)$$

is minimal.

Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (11)$$

Then  $AV = mX + bB$ . The set  $S = \{AV | m, b \in \mathbb{R}\}$  is a plane in  $n$ -dimensional space. The ordered pair  $(m, b)$  that minimizes the squared error corresponds to the projection  $Y_S$  of  $Y$  onto  $S$ .

# Least Squares Method

Let  $C$  be  $B - B_X$ , where  $B_X$  is the projection of  $B$  onto the line spanned by  $X$ . Then

$$C = B - \left( \frac{B \cdot X}{X \cdot X} \right) X \quad (12)$$

Note that  $X \perp C$ .  $X$  and  $C$  form an orthogonal basis for  $S$ . We have chosen  $C$  by a process called **successive orthogonal selection**. Consequently,

$$Y_S = Y_X + Y_C = \left( \frac{Y \cdot X}{X \cdot X} \right) X + \left( \frac{Y \cdot C}{C \cdot C} \right) C \quad (13)$$

# Least Squares Method

Replace the rightmost  $C$  by  $B - \left(\frac{B \cdot X}{X \cdot X}\right) X$  for

$$Y_S = \left( \frac{Y \cdot X}{X \cdot X} - \left( \frac{Y \cdot C}{C \cdot C} \right) \left( \frac{B \cdot X}{X \cdot X} \right) \right) X + \left( \frac{Y \cdot C}{C \cdot C} \right) B \quad (14)$$

or alternatively

$$m = \frac{Y \cdot X}{X \cdot X} - \left( \frac{Y \cdot C}{C \cdot C} \right) \left( \frac{B \cdot X}{X \cdot X} \right) \quad (15)$$

$$b = \frac{Y \cdot C}{C \cdot C} \quad (16)$$

with

$$C = B - \left( \frac{B \cdot X}{X \cdot X} \right) X \quad (17)$$

# Shoe Print and Height

**Example 3: Shoe Print and Height.** Recall

Shoe Print (cm)	Height (cm)
29.7	175.3
29.9	177.8
31.4	185.4
31.8	175.3
27.6	172.7

Which **regression line** best represents a possible linear relationship between shoe print and height? In this case, the shoe print is the independent variable  $x$ , on the basis of which we are trying to predict the dependent variable  $y$ , the height.

$$X = \begin{bmatrix} 29.7 \\ 29.9 \\ 31.4 \\ 31.8 \\ 27.6 \end{bmatrix}, Y = \begin{bmatrix} 175.3 \\ 177.8 \\ 185.4 \\ 175.3 \\ 172.7 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (18)$$



# Shoe Print and Height

Calculate  $C$  using the formula

$$C = B - \left( \frac{B \cdot X}{X \cdot X} \right) X = \begin{bmatrix} 0.0150340 \\ 0.0084012 \\ -0.0413445 \\ -0.0546101 \\ 0.0846780 \end{bmatrix} \quad (19)$$

Now calculate  $m$  and  $b$

$$m = \frac{Y \cdot X}{X \cdot X} - \left( \frac{Y \cdot C}{C \cdot C} \right) \left( \frac{B \cdot X}{X \cdot X} \right) = 1.7528 \quad (20)$$

$$b = \frac{Y \cdot C}{C \cdot C} = 124.58 \quad (21)$$

# Shoe Print and Height

	A	B	C	D	E	F	
1	29.7	175.3		<b>Regression</b>			
2	29.9	177.8					
3	31.4	185.4		Regression Model	Linear		
4	31.8	175.3		R^2	0.356575848683		
5	27.6	172.7		Standard Error	4.51435351068		
6							
7				Slope	1.752811026478		
8				Intercept	124.5754443235		
9							
10				29.7	176.6339318099		
11				29.9	176.9844940152		
12				31.4	179.613710555		
13				31.8	180.3148349655		
14				27.6	172.9530286543		
15							



**Example 4: Angles at Gray Cliff.** This example is from Oscar S. Adams's *Application of the Theory of Least Squares to the Adjustment of Triangulation* (1915), a “working manual for the computer in the office.” You measure the following angles.

from	to	angle
Boulder	Tower	$65^{\circ}6'29.3''$
Tower	Tyonek	$19^{\circ}46'26.9''$
Tyonek	Round Point	$8^{\circ}39'14.2''$
Round Point	Boulder	$266^{\circ}27'47.9''$

Notice that the angles do not add up to  $360^\circ$ . We are missing  $1.7''$ . How should we adjust these numbers?

Basic assumptions underlying least squares theory in surveying are

- ① mistakes and systematic errors have been eliminated
- ② the number of observations being adjusted is large
- ③ the frequency distribution of the errors is normal

Convert the angles to real numbers

$$\hat{a} = 65.108, \hat{b} = 19.774, \hat{c} = 8.6539, \hat{d} = 266.46 \quad (22)$$

The sum is 359.999527778. Here is a system of equations with measurement errors, exploiting the fact that  $d$  is supposed to be  $360^\circ - (a + b + c)$

$$\begin{aligned} a &= 65.108 + \epsilon_1 \\ b &= 19.774 + \epsilon_2 \\ c &= 8.6539 + \epsilon_3 \\ 360 - (a + b + c) &= 266.46 + \epsilon_4 \end{aligned} \quad (23)$$

The system of equations is equivalent to the following matrix equation.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 65.108 \\ 19.774 \\ 8.6539 \\ -93.537 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \quad (24)$$

In symbols,

$$AV = Y + E \quad (25)$$

Again, we want to minimize

$$\|Y - AV\|^2 = \|E\|^2 \quad (26)$$

The minimization is achieved by projecting  $Y$  onto the hyperplane  $S$

$$a \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \quad (27)$$

These three vectors  $\alpha, \beta, \gamma$  form a basis of  $S$ , but the basis vectors are not orthogonal to each other. We will search for a different basis of  $S$  that is orthonormal by successive orthogonal selection.

Start with a non-zero vector  $b_1$  in  $S$ , for example  $b_1 = \alpha = (1, 0, 0, -1)^T$ . This is our first basis vector. The second basis vector  $b_2$  must fulfill

- ①  $b_2 \cdot b_1 = 0$  (which is equivalent to  $b_2 \perp b_1$ )
- ②  $b_2 \in S$

For example,  $b_2 = (1, -1, -1, 1)^T$  qualifies. Follow the same procedure for  $b_3 = (0, 1, -1, 0)^T$ .



# Angles at Gray Cliff

Let  $Y_i$  be the projection of  $Y$  onto the line spanned by  $b_i$ , for example.

$$Y_1 = \left( \frac{Y \cdot b_1}{b_1 \cdot b_1} \right) b_1 = \begin{bmatrix} 79.32242 \\ 0 \\ 0 \\ -79.32242 \end{bmatrix} \quad (28)$$

Then the projection  $Y_S$  equals  $Y_1 + Y_2 + Y_3 =$

$$\begin{bmatrix} 79.32242 \\ 0 \\ 0 \\ -79.32242 \end{bmatrix} + \begin{bmatrix} -14.214 \\ 14.214 \\ 14.214 \\ -14.214 \end{bmatrix} + \begin{bmatrix} 0 \\ 5.56010 \\ -5.56010 \\ 0 \end{bmatrix} = \begin{bmatrix} 65.1083 \\ 19.7743 \\ 8.6541 \\ -93.5366 \end{bmatrix}$$

The least squares adjusted angles are

$65^{\circ}6'29.7''$ ,  $19^{\circ}46'27.3''$ ,  $8^{\circ}39'14.6''$  compared to the original

$65^{\circ}6'29.3''$ ,  $19^{\circ}46'26.9''$ ,  $8^{\circ}39'14.2''$ .

**Example 5: Abigail, Ben, and Charlie.** Here is an example adapted from Paul R. Wolf's *Survey Measurement Adjustments by Least Squares*. Abigail measures a length  $\overline{XY}$  to be 211.52 units. Ben measures a length  $\overline{YZ}$  to be 220.10 units. Charlie measures the length  $\overline{XZ}$  to be 431.71 units. What lengths should they report to their supervisor?

We will solve this problem three different ways.

- 1 use calculus
- 2 use projection
- 3 use a matrix formula based on projection

# Abigail, Ben, and Charlie [Calculus]

Here are the three equations with errors:

$$\begin{aligned}a &= 211.52 + \epsilon_1 \\b &= 220.10 + \epsilon_2 \\a + b &= 431.71 + \epsilon_3\end{aligned}\tag{29}$$

Define the function

$$\begin{aligned}F(a, b) = \|E\|^2 &= (a - 211.52)^2 + (b - 220.10)^2 + (a + b - 431.71)^2 = \\&= 2a^2 + 2ab + 2b^2 - 1286.46a - 1303.62b + 279558.2445\end{aligned}\tag{30}$$

We want to minimize  $F$ . The partial derivatives are

$$\frac{\partial F}{\partial a} = 4a + 2b - 1286.46\tag{31}$$

$$\frac{\partial F}{\partial b} = 2a + 4b - 1303.62\tag{32}$$

Setting the partial derivatives to zero gives us the following system of linear equations.

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1286.46 \\ 1303.62 \end{bmatrix} \quad (33)$$

The solution is  $a = 211.55$ ,  $b = 220.13$  for Abigail and Ben's least squares adjusted measurements, compared to the original  $\hat{a} = 211.52$ ,  $\hat{b} = 220.10$ . Charlie's measurement is adjusted from 431.71 to 431.68.

Equation (29) translates into the following matrix equation.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 211.52 \\ 220.10 \\ 431.71 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix} \quad (34)$$

or

$$AV = Y + E \text{ and therefore } E = Y - AV \quad (35)$$

# Abigail, Ben, and Charlie [Projection]

Find the projection  $Y_S$  of  $Y$  onto the hyperplane  $S$  defined by  $AV$  with free variables  $a, b$ .  $u = (1, 0, 1)^T$  and  $v = (0, 1, 1)^T$  are not orthogonal. Find the projection  $v_u$  of  $v$  onto the line spanned by  $u$  and define  $w = v - v_u$  (successive orthogonal selection).

$$w = v - v_u = v - \left( \frac{v \cdot u}{u \cdot u} \right) u = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} \quad (36)$$

Consequently,

$$Y_S = Y_u + Y_w = \left( \frac{Y \cdot u}{u \cdot u} \right) u + \left( \frac{Y \cdot w}{w \cdot w} \right) w = \begin{bmatrix} 211.55 \\ 220.13 \\ 431.68 \end{bmatrix} \quad (37)$$

The solution found by using projection agrees with the solution found by using calculus.

On a rainy day, you may not be in the mood to remember the projection procedure. You just want to use a formula. Here it is for Abigail, Ben, and Charlie:

$$V = \begin{bmatrix} a \\ b \end{bmatrix} = (A^T A)^{-1} A^T Y = \begin{bmatrix} 211.55 \\ 220.13 \end{bmatrix} \quad (38)$$

Why does this formula work?

# End of Lesson

Next Lesson: Eigenvalues and Eigenvectors