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 $\vec{v} = (-1, 1)^\intercal$ in $\mathbb{R}^2$. What is the projection $\vec{w}$ of $\vec{u} = (3, -2)^\intercal$?

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

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

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

Projection

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

$$\|u - v\| \geq \|u - u_H\| \tag{2}$$

Proof: use the theorem of Pythagoras for

$$\|u - v\|^2 = \|u - u_H\|^2 + \|u_H - v\|^2 \geq \|u - u_H\|^2 \tag{3}$$

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.

Projection

Example 2: Finding Another Projection. What is the projection of $\vec{u} = (5, 2, 10)^{\mathsf{T}}$ 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 $\vec{v_1} = (1, 1, -1)^{\mathsf{T}}$ and $\vec{v_2} = (0, -3, 1)^{\mathsf{T}}$. 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

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

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

Least Squares Method

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

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

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

$$(u - av) \perp v \tag{6}$$

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 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ 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 ϵ_i are the errors $(i = 1, \ldots, n)$. 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.

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

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

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

We are trying to choose m, b so that

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

is minimal.

Let

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

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.

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

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

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,

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

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

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

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.

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

Calculate C using the formula

$$C = B - \left(\frac{B \cdot X}{X \cdot X} \right) X = \left[\begin{array}{c} 0.0150340 \\ 0.0084012 \\ -0.0413445 \\ -0.0546101 \\ 0.0846780 \end{array} \right] \tag{19}$$

Now calculate m and b

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

$$b = \frac{Y \cdot C}{C \cdot C} = 124.58 \tag{21}$$

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	$65°6'29.3''$
Tower	Tyonek	$19°46'26.9''$
Tyonek	Round Point	$8°39'14.2''$
Round Point	Boulder	$266°27'47.9''$

Notice that the angles do not add up to $360°$. 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

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

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

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

Angles at Gray Cliff

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

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

In symbols,

$$AV = Y + E \tag{25}$$

Again, we want to minimize

$$\|Y - AV\|^2 = \|E\|^2 \tag{26}$$

The minimization is achieved by projecting Y onto the hyperplane S

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

These three vectors α, β, γ 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
$b_1 = \alpha = (1, 0, 0, -1)^\intercal$. 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, $b_2 = (1, -1, -1, 1)^\intercal$ qualifies. Follow the same
procedure for $b_3 = (0, 1, -1, 0)^\intercal$.

Angles at Gray Cliff

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

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

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
$65°6'29.7'', 19°46'27.3'', 8°39'14.6''$ compared to the original
$65°6'29.3'', 19°46'26.9'', 8°39'14.2''$.

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:

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

Define the function

$$
F(a, b) = \|E\|^2 = (a - 211.52)^2 + (b - 220.10)^2 + (a + b - 431.71)^2 =
$$

$$
2a^2 + 2ab + 2b^2 - 1286.46a - 1303.62b + 279558.2445 \tag{30}
$$

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.

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

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.

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

or

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

Abigail, Ben, and Charlie [Projection]

Find the projection Y_S of Y onto the hyperplane S defined by AV with free variables a, b. $u = (1, 0, 1)^\mathsf{T}$ and $v = (0, 1, 1)^\mathsf{T}$ 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).

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

Consequently,

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

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:

$$V_0 = \left[\begin{array}{c} a \\ b \end{array} \right] = (A^\mathsf{T} A)^{-1} A^\mathsf{T} Y = \left[\begin{array}{c} 211.55 \\ 220.13 \end{array} \right] \qquad (38)$$

Why does this formula work?

Let V_0 be the solution to the least squares problem, AV_0 is the projection Y_S of Y onto the hyperplane S defined by AV when the variables in V are still free. $Y - AV_0$ is orthogonal to S. S is the hyperplane $\{AU | U$ is a vector with the right dimensions$\}$. Consequently,

$$(Y - AV_0) \cdot AU = 0 \tag{39}$$

for some U with the right dimensions. Rewrite

$$Y \cdot AU = AV_0 \cdot AU \tag{40}$$

For column vectors u and v, it is generally true that $u \cdot v = u^\mathsf{T} v$. Also recall that for any two matrices A and B

$$(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T} \tag{41}$$

Using these two facts, continue with

$$Y \cdot AU = Y^\mathsf{T} AU = (A^\mathsf{T} Y)^\mathsf{T} U \tag{42}$$

and

$$AV_0 \cdot AU = (AV_0)^\mathsf{T} AU = (A^\mathsf{T} AV_0)^\mathsf{T} U \tag{43}$$

Combine (40), (42) and (43) for

$$(A^\mathsf{T} Y)^\mathsf{T} U = (A^\mathsf{T} A V_0)^\mathsf{T} U \tag{44}$$

Because this is true for all well-dimensioned vectors U, it must be true that

$$(A^\mathsf{T} Y)^\mathsf{T} = (A^\mathsf{T} A V_0)^\mathsf{T} \tag{45}$$

If $A^\mathsf{T} A$ is invertible (this can be shown to be true), it follows that

$$V_0 = (A^\mathsf{T} A)^{-1} A^\mathsf{T} Y \tag{46}$$

Shoe Print and Height

Now that we have a formula, let us return to the problem of shoe print and height.

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

The yave setup $Y - AV = E$ is

$$\begin{bmatrix} 175.3 \\ 177.8 \\ 185.4 \\ 175.3 \\ 172.7 \end{bmatrix} - \begin{bmatrix} 29.7 & 1 \\ 29.9 & 1 \\ 31.4 & 1 \\ 31.8 & 1 \\ 27.6 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \tag{47}$$

Use

$$V_0 = (A^\mathsf{T} A)^{-1} A^\mathsf{T} Y = \begin{bmatrix} 1.7528 \\ 124.5754 \end{bmatrix} \qquad (48)$$

As we already know, the slope m of the regression line is 1.7528 and the y-intercept b is 124.5754.

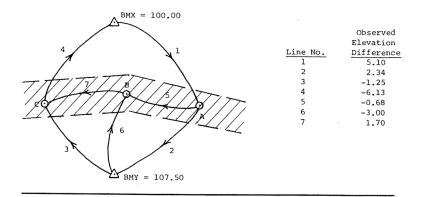**Exercise 1:** Find the quadratic of best fit for the data

x	0	1	2	3	4
y	9.1	2.8	1.2	3.1	8.7

Exercise 2: At 6327 ft (or 6.327 thousand feet), Mario Triola recorded the temperature. Find the best predicted temperature at that altitude based on other measurements, assuming a linear relationship. How does the result compare to the actual recorded value of 48°F?

Altitude	3	10	14	22	28	31	33
Temperature	57	37	24	-5	-30	-41	-54

Exercise 3: You measure the four angles in a quadrilateral
$\hat{a} = 8.490426°, \hat{b} = 182.029154°, \hat{c} = 119.148088°, \hat{d} = 50.32948.$
What are the least squares adjusted measurements?

Exercise 4: Consider the following leveling network.

Line No.	Observed Elevation Difference
1	5.10
2	2.34
3	-1.25
4	-6.13
5	-0.68
6	-3.00
7	1.70

Next Lesson: Eigenvalues and Eigenvectors