Least Squares Method

MATH 3512, BCIT

Matrix Methods and Statistics for Geomatics

October 19, 2018

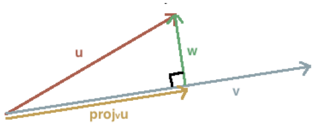
Projection

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

Projection

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 $u = (u_1, \ldots, u_n)^\intercal$ 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.

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

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

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

Projection

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.

Projection

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

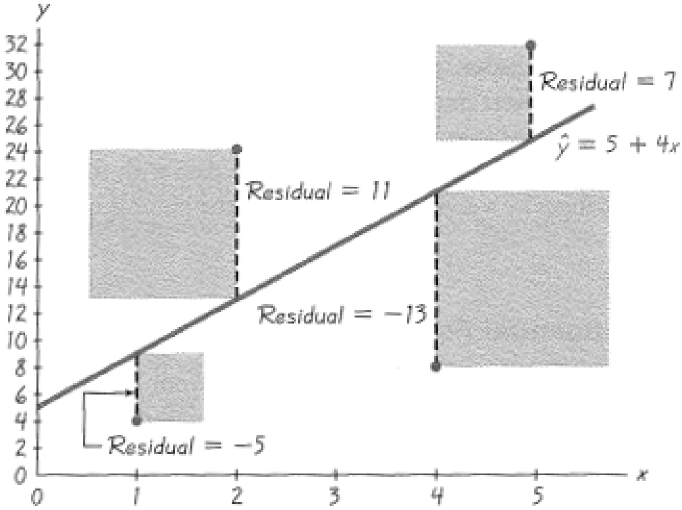
1. $(w_u + w_v) \in H$ (trivial)
2. $(w - (w_u + w_v)) \perp u$ (use the fact that $u \perp v$)
3. $(w - (w_u + w_v)) \perp v$ (same idea)

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:

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

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

$$E = Y - AV \tag{9}$$

We are trying to choose m, b so that

$$\|E\|^2 = \|Y - AV\|^2 \tag{10}$$

is minimal.

Least Squares Method

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.

Least Squares Method

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

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

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,

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

Least Squares Method

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

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

or alternatively

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

$$b = \frac{Y \cdot C}{C \cdot C} \tag{16}$$

with

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

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.

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

Shoe Print and Height

Calculate C using the formula

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

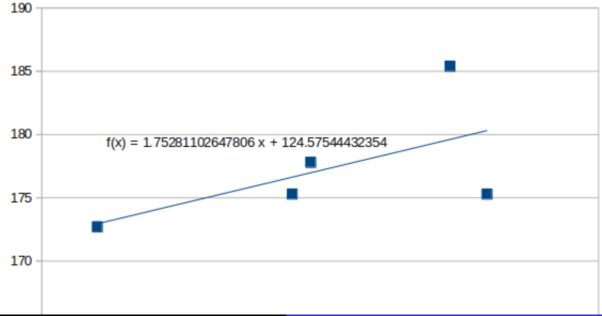
Now calculate m and b

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

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

Shoe Print and Height

	A	B	C	D	E	F
1	29.7	175.3		**Regression**		
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	



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	$266°27'47.9''$

Angles at Gray Cliff

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

Angles at Gray Cliff

Convert the angles to real numbers

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

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

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

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

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

Angles at Gray Cliff

The minimization is achieved by projecting Y onto the hyperplane S

$$a \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \tag{27}$$

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.

Angles at Gray Cliff

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.

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

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:

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

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

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

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

Abigail, Ben, and Charlie [Calculus]

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.

Abigail, Ben, and Charlie [Projection]

Equation (29) translates into the following matrix equation.

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

or

$$AV = Y + E \text{ and therefore } E = Y - AV \tag{35}$$

Abigail, Ben, and Charlie [Projection]

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

Abigail, Ben, and Charlie [Linear Algebra]

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:

$$V_0 = \begin{bmatrix} a \\ b \end{bmatrix} = (A^\intercal A)^{-1} A^\intercal Y = \begin{bmatrix} 211.55 \\ 220.13 \end{bmatrix} \tag{38}$$

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 $u \cdot v = u^\intercal v$. Also recall that for any two matrices A and B

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

Using these two facts, continue with

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

and

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

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

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

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

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

If $A^\intercal A$ is invertible (this can be shown to be true), it follows that

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

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

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

Shoe Print and Height

Use

$$V_0 = (A^\intercal A)^{-1} A^\intercal Y = \begin{bmatrix} 1.7528 \\ 124.5754 \end{bmatrix} \tag{48}$$

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

Least Squares Exercises

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

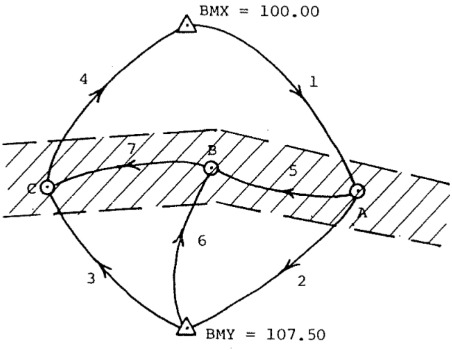
Exercise 2: At 6327 ft (or 6.327 thousand feet), Mario Triola recorded the temperature. Find the best predicted temperature at that altitude based on other measurements, assuming a linear relationship. How does the result compare to the actual recorded value of 48°F?

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

Least Squares Exercises

Exercise 3: You measure the four angles in a quadrilateral $\hat{a} = 8.490426^\circ, \hat{b} = 182.029154^\circ, \hat{c} = 119.148088^\circ, \hat{d} = 50.32948$. 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

The objective is to determine elevations of A, B, and C, which are to serve as temporary project bench marks to control construction of a highway through the crosshatched corridor.

Obviously, it would have been possible to obtain elevations for A, B, and C by beginning at BMX and running a single closed loop consisting of only courses 1, 5, 7, and 4. Alternatively, a single closed loop could have been initiated at BMY and consist of courses 2, 5, 7, and 3.

However, by running all seven courses, redundancy is achieved that enables checks to be made, blunders to be isolated, and precision to be increased. Having run all seven courses, it would be possible to compute the adjusted elevation of B, for example, using several different single closed circuits. Loops 1-5-6, 2-5-6, 3-7-6, and 4-7-6 could each be used, but it is almost certain that each would yield a different elevation for B.

A more logical approach, which will produce only one adjusted value for B—its most probable one—is to use all seven courses in a simultaneous least squares adjustment. In adjusting level networks, the observed difference in elevation for each course is treated as one observation containing a single random error.

This single random error is the total of the individual random errors in backsight and foresight readings for the entire course. In the figure, the arrows indicate the direction of leveling. Thus, for course number 1, leveling proceeded from BMX to A and the observed elevation difference was $+5.10$ feet.

Try to find the equations yourself (before looking at the next slide to check if they are correct) and then calculate the least squares adjusted elevations of A, B, and C.

Least Squares Exercises

The first set of equation for the seven courses is

$$\begin{aligned}
A &= BMX + 5.10 + \epsilon_1 \\
BMY &= A + 2.34 + \epsilon_2 \\
C &= BMY - 1.25 + \epsilon_3 \\
BMX &= C - 6.13 + \epsilon_4 \\
B &= A - 0.68 + \epsilon_5 \\
B &= BMY - 3.00 + \epsilon_6 \\
C &= B + 1.70 + \epsilon_7
\end{aligned} \tag{49}$$

The resulting yave setup is

$$
\begin{aligned}
A &= 105.10 + \epsilon_1 \\
A &= 105.16 + \epsilon_2 \\
C &= 106.25 + \epsilon_3 \\
C &= 106.13 + \epsilon_4 \\
A - B &= 0.68 + \epsilon_5 \\
B &= 104.50 + \epsilon_6 \\
B - C &= -1.70 + \epsilon_7
\end{aligned}
\tag{50}
$$

The adjusted benchmark elevations are
$A = 105.14$, $B = 104.48$, $C = 106.19$.

Linearizing Distance Observation Equations

Example 6: Linearizing Equations. You are trying to measure the coordinates of stations A and B. Your provisional estimate is $(8.3995, 3.0161)$ and $(-2.872, 1.4937)$. Then you observe the length between A and B to be 11.391. How would you report your least squares adjusted coordinates for A and B, given that you weigh equally the errors for A and B's coordinates as well as the distance between them?

Linearizing Distance Observation Equations

Setting up the first four yave equations is simple. The fifth one, however, is non-linear.

$$\begin{aligned} x &= 8.3995 + \epsilon_1 \\ y &= 3.0161 + \epsilon_2 \\ z &= -2.872 + \epsilon_3 \\ w &= 1.4937 + \epsilon_4 \\ \sqrt{(z-x)^2 + (w-y)^2} &= 11.391 + \epsilon_5 \end{aligned} \tag{51}$$

We shall linearize the fifth equation using the Taylor polynomial expansion of the function $G(x, y, z, w) = \sqrt{(z-x)^2 + (w-y)^2}$.

Linearizing Distance Observation Equations

Recall the Taylor polynomial expansion of a function $f : \mathbb{R} \to \mathbb{R}$ about $x = a$,

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots \tag{52}$$

To linearize a function f, approximate the function only using the first two terms,

$$f(x) \approx f(a) + f'(a)(x-a) \tag{53}$$

Linearizing Distance Observation Equations

Without knowing it, we used Taylor polynomials for Newton's method. We were trying to solve $f(x) = 0$ but could not because it gave us a non-linear equation. Instead, we solved

$$f(x) \approx f(a) + f'(a)(x - a) = 0 \tag{54}$$

which is equivalent to

$$x = a - \frac{f(a)}{f'(a)} \tag{55}$$

This x, however, is only an approximation of the true x, and so we repeated the process until we were as close to the x-intercept of f as we needed to be.

Linearizing Distance Observation Equations

The function, however, is a function from $\mathbb{R}^4 \to \mathbb{R}$.

$$G(x, y, z, w) = \sqrt{(z-x)^2 + (w-y)^2} \tag{56}$$

We need to use a generalization of Taylor polynomials for higher dimensions. Let $F : \mathbb{R}^n \to \mathbb{R}^m$. Then

$$F(\vec{x}) = F(\vec{a}) + J(\vec{a})(\vec{x} - \vec{a}) \tag{57}$$

where J is the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} \tag{58}$$

Note that $F_i(\vec{x}) = y_i$ for $F(\vec{x}) = \vec{y}$.

Linearizing Distance Observation Equations

The Jacobian for G is

$$J(x, y, z, w) = \begin{bmatrix} \frac{x-z}{\sqrt{(z-x)^2+(w-y)^2}} \\ \frac{z-x}{\sqrt{(z-x)^2+(w-y)^2}} \\ \frac{y-w}{\sqrt{(z-x)^2+(w-y)^2}} \\ \frac{w-y}{\sqrt{(z-x)^2+(w-y)^2}} \end{bmatrix}^{\mathsf{T}}$$

Therefore

$$G(x, y, z, w) \approx G(\hat{x}, \hat{y}, \hat{z}, \hat{w}) + \begin{bmatrix} \frac{\hat{x}-\hat{z}}{\sqrt{(\hat{z}-\hat{x})^2+(\hat{w}-\hat{y})^2}} \\ \frac{\hat{y}-\hat{w}}{\sqrt{(\hat{z}-\hat{x})^2+(\hat{w}-\hat{y})^2}} \\ \frac{\hat{z}-\hat{x}}{\sqrt{(\hat{z}-\hat{x})^2+(\hat{w}-\hat{y})^2}} \\ \frac{\hat{w}-\hat{y}}{\sqrt{(\hat{z}-\hat{x})^2+(\hat{w}-\hat{y})^2}} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} x - \hat{x} \\ y - \hat{y} \\ z - \hat{z} \\ w - \hat{w} \end{bmatrix}$$

Linearizing Distance Observation Equations

It makes sense to use $\hat{x}, \hat{y}, \hat{z}, \hat{w}$ as our first estimate for x, y, z, w. Therefore, the equation

$$\sqrt{(z-x)^2+(w-y)^2} = 11.391 + \epsilon_5$$

linearizes to (approximately)

$$\sqrt{(-2.8724-8.3995)^2+(1.4937-3.0161)^2}+$$

$$0.99100(x-8.3995)+0.13385(y-3.0161)-$$

$$0.99100(z+2.8724)-0.13385(w-1.4937) = 11.391+\epsilon_5$$

or, equivalently,

$$0.99100x+0.13385y-0.99100z-0.13385w+0.000018071$$

Finally, we have a linear yave setup.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.99100 & 0.13385 & -0.99100 & -0.13385 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 8.3995 \\ 3.0161 \\ -2.872 \\ 1.4937 \\ 11.391 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \tag{59}$$

Linearizing Distance Observation Equations

Use the formula

$$V_0 = (A^\intercal A)^{-1} A^\intercal Y \tag{60}$$

for

$$V_0 = (8.4052, 3.0169, -2.8777, 1.4929)^\intercal \tag{61}$$

The least squares adjusted coordinates are $(8.4052, 3.0169)$ and $(-2.8777, 1.4929)$, compared to the original $(8.3995, 3.0161)$ and $(-2.8720, 1.4937)$.

Linearizing Angle Observation Equations

Next: do this for angles (partial derivative of arctan). Then solve system of non-linear equations. Tie this in with Shields, example 2 on page 296.

Linearizing Angle Observation Equations

Example 7: Angle Observations. There are three points whose coordinates with measurement errors are

$$\begin{aligned} I &= (595.74, 537.76) \\ J &= (800.92, 658.44) \\ K &= (302.96, 168.88) \end{aligned} \tag{62}$$

From station I, you observe an angle of $158°49'21''$ instead of the expected $158°54'5.9107''$ between $\vec{IJ}$ and $\vec{IK}$. How should you least squares adjust the coordinates of I, J, K in light of your angle measurement? (Note that it is unnatural to give equal weight to the errors in coordinate measurements and angle measurements: this can be addressed by weight factors, but I will skip this step here for simplicity.)

Linearizing Angle Observation Equations

Label the coordinates of I to be x and y; the coordinates of J to be z and w; the coordinates of K to be u and v. The first six equations are as usual, for example

$$x = 595.74 + \epsilon_1 \tag{63}$$

The seventh equation is non-linear.

$$\pi + \arctan\left(\frac{w-y}{z-x}\right) - \arctan\left(\frac{v-y}{u-x}\right) = 158.82 + \epsilon_7 \tag{64}$$

Using the Jacobian of G, the seventh equation is linearized on the next slide. The non-linear function is

$$G(x, y, z, w, u, v) = \pi + \arctan\left(\frac{w-y}{z-x}\right) - \arctan\left(\frac{v-y}{u-x}\right) \tag{65}$$

Linearizing Angle Observation Equations

The Jacobian is

$$\begin{bmatrix} \frac{w-y}{||\vec{IJ}||^2} - \frac{v-y}{||\vec{IK}||^2} \\ \frac{z-x}{||\vec{IJ}||^2} - \frac{u-x}{||\vec{IK}||^2} \\ \frac{w-y}{||\vec{IJ}||^2} \\ \frac{z-x}{||\vec{IJ}||^2} \\ \frac{v-y}{||\vec{IK}||^2} \\ \frac{u-x}{||\vec{IK}||^2} \end{bmatrix}^{\mathsf{T}} \tag{66}$$

for the linearization

$$G(x, y, z, w, u, v) \approx \pi + 0.00379297940350206x +$$

$$0.00494115229670866y + 0.00212980385748917z -$$

$$0.00166317554601289u - 0.00132006217838228v +$$

$$0.00362109011832638w - 9.71163063930754$$

Linearizing Angle Observation Equations

Let the linearization
$G(x, y, z, w, u, v) \approx ax + by + cz + dw + ev + fu - g$. Define

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ a & b & c & d & e & f \end{bmatrix}, Y = \begin{bmatrix} 595.74 \\ 537.76 \\ 800.92 \\ 658.44 \\ 302.96 \\ 168.88 \\ 158.82 + g \end{bmatrix} \tag{67}$$

$(A^\intercal A)^{-1} A^\intercal Y$ yields the results on the next slide.

Linearizing Angle Observation Equations

Compare the original measurements to the least squares adjusted values.

variable	original	adjusted	adjusted again
x	595.74	596.330374912751	596.330444508634
y	537.76	538.529087317840	538.529645106905
z	800.92	801.251502661307	801.251871588615
w	658.44	659.003620451168	659.004496897710
u	302.96	303.218872251445	303.218572920019
v	168.88	169.085466866672	169.085148209195

Because we arrived at these values by an approximation (using a linearization instead of the non-linear function), we should repeat the process using the new numbers as we did with Newton's method. In this case, however, the difference between the first iteration and the second iteration is below the sensitivity of our measuring instruments.

Systems of Non-Linear Equations

Solve

$$\begin{array}{ccccc} \cos x & - & y & = & 0 \\ x & - & y^2 & = & 0 \end{array} \tag{68}$$

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a function.

$$F(x, y) = \left(\cos x - y, x - y^2\right) \tag{69}$$

Then the Jacobian is

$$J = \begin{bmatrix} -\sin x & -1 \\ 1 & -2y \end{bmatrix} \tag{70}$$

The linearization of F is

$$F(x,y) \approx F(x_0,y_0) + \begin{bmatrix} -\sin x_0 & -1 \\ 1 & -2y_0 \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \qquad (71)$$

Setting it to zero translates into the matrix equation

$$\begin{bmatrix} -\sin x_0 & -1 \\ 1 & -2y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\cos x_0 - x_0 \sin x_0 \\ -y_0^2 \end{bmatrix} \qquad (72)$$

Taking $(x_0, y_0) = (0.6, 0.8)$ as our first approximation, this yields

$$\begin{bmatrix} -0.565 & -1 \\ 1 & -1.6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1.164 \\ -0.64 \end{bmatrix} \qquad (73)$$

for an approximation of our system of linear equations. The solution is $(x_1, y_1) = (0.64231, 0.80144)$.

Now plug $(x_1, y_1) = (0.64231, 0.80144)$ into

$$\begin{bmatrix} -\sin x_1 & -1 \\ 1 & -2y_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\cos x_1 - x_1 \sin x_1 \\ -y_1^2 \end{bmatrix} \tag{74}$$

for

$$\begin{bmatrix} -0.59905 & -1 \\ 1 & -1.6029 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1.1855 \\ -0.6423 \end{bmatrix} \tag{75}$$

The solution is $(x_2, y_2) = (0.64171, 0.80107)$. With five significant digits, (x_3, y_3) is already indistinguishable from (x_2, y_2). The solution set for the system of non-linear equations is

$$S = \{(x, y) \in \mathbb{R}^2 | x \approx 0.64171, y \approx 0.80107\} \tag{76}$$

Example 8: Square Root of a Matrix. Find a matrix C such that

$$C^2 = A \text{ where } A = \begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix} \tag{77}$$

In the next lesson, we will learn how to solve this problem using eigenvalues. For now, we are faced with a system of non-linear equations given

$$C = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \tag{78}$$

Systems of Non-Linear Equations

$$\begin{array}{ccccc} x_1^2 & + & x_2x_3 & = & 3 \\ x_1x_2 & + & x_2x_4 & = & -1 \\ x_1x_3 & + & x_3x_4 & = & 2 \\ x_2x_3 & + & x_4^2 & = & 0 \end{array} \tag{79}$$

The idea is to linearize these equations and then use Newton's method. Let $\vec{x} = (x_1, x_2, x_3, x_4)^\intercal$. Then

$$F(\vec{x}) =$$

$$(x_1^2 + x_2x_3, x_1x_2 + x_2x_4, x_1x_3 + x_3x_4, x_2x_3 + x_4^2) \tag{80}$$

Systems of Non-Linear Equations

The Jacobian matrix of F is

$$J = \begin{bmatrix} 2x_1 & x_3 & x_2 & 0 \\ x_2 & x_1 + x_4 & 0 & x_2 \\ x_3 & 0 & x_1 + x_4 & x_3 \\ 0 & x_3 & x_2 & 2x_4 \end{bmatrix} \tag{81}$$

Let $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)^\intercal$ be a first approximation of a solution for the system of non-linear equations. Then we can linearize (see next slide).

Systems of Non-Linear Equations

$$F_1(\vec{x}) \approx$$

$$F_1(\hat{x}) + 2x_1(x_1 - \hat{x}_1) + x_3(x_2 - \hat{x}_2) + x_2(x_3 - \hat{x}_3) + 0(x_4 - \hat{x}_4)$$

$$F_2(\vec{x}) \approx$$

$$F_2(\hat{x}) + x_2(x_1 - \hat{x}_1) + (x_1 + x_4)(x_2 - \hat{x}_2) + 0(x_3 - \hat{x}_3) + x_2(x_4 - \hat{x}_4)$$

$$F_3(\vec{x}) \approx$$

$$F_3(\hat{x}) + x_3(x_1 - \hat{x}_1) + 0(x_2 - \hat{x}_2) + (x_1 + x_4)(x_3 - \hat{x}_3) + x_3(x_4 - \hat{x}_4)$$

$$F_4(\vec{x}) \approx$$

$$F_4(\hat{x}) + 0(x_1 - \hat{x}_1) + x_3(x_2 - \hat{x}_2) + x_2(x_3 - \hat{x}_3) + 2x_4(x_4 - \hat{x}_4)$$

Systems of Non-Linear Equations

Now rewrite as a system of linear equations. Remember that $\hat{x}_i$ are simply numbers.

$$\begin{bmatrix} 2\hat{x}_1 & \hat{x}_3 & \hat{x}_2 & 0 \\ \hat{x}_2 & \hat{x}_1 + \hat{x}_4 & 0 & \hat{x}_2 \\ \hat{x}_3 & 0 & \hat{x}_1 + \hat{x}_4 & \hat{x}_3 \\ 0 & \hat{x}_3 & \hat{x}_2 & 2\hat{x}_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 + \hat{x}_1^2 + \hat{x}_2\hat{x}_3 \\ -1 + \hat{x}_1\hat{x}_2 + \hat{x}_2\hat{x}_4 \\ 2 + \hat{x}_1\hat{x}_3 + \hat{x}_3\hat{x}_4 \\ \hat{x}_2\hat{x}_3 + \hat{x}_4^2 \end{bmatrix}$$

Systems of Non-Linear Equations

Let $\hat{x}_i = 1$ for $i = 1, 2, 3, 4$. Then the system looks as follows.

$$\begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 4 \\ 2 \end{bmatrix} \tag{82}$$

The determinant is zero. We have accidentally chosen a critical point as our initial estimate. Let's try again on the next slide.

Systems of Non-Linear Equations

Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4) = (2, 0, 1, 1)$. Then the system looks as follows.

$$\begin{bmatrix} 4 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 1 & 0 & 3 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ 5 \\ 1 \end{bmatrix} \tag{83}$$

The solution is $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4) = (2, 0, 1, 1)$.

End of Lesson

Next Lesson: Eigenvalues and Eigenvectors