Least Squares Method

MATH 3512, BCIT

Matrix Methods and Statistics for Geomatics

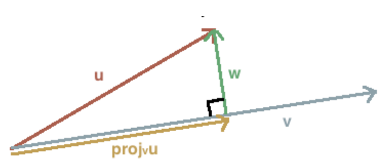
October 19, 2018

The projection u_H of a vector u onto a hyperplane H is the vecto
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

ection

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



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

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

amer'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

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

Proof: use the theorem of Pythagoras for

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

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

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

rst we find two linearly independent vectors in H to form a basis
H, for example $\vec{v_1} = (1, 1, -1)^{\mathsf{T}}$ and $\vec{v_2} = (0, -3, 1)^{\mathsf{T}}$. The
nditions

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

ve us the system of linear equations

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

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

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

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

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

$$(u - av) \perp v \qquad ($$

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

ection

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

rmula for Projection Onto Plane with Orthogonal Basis

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

oof: check the following

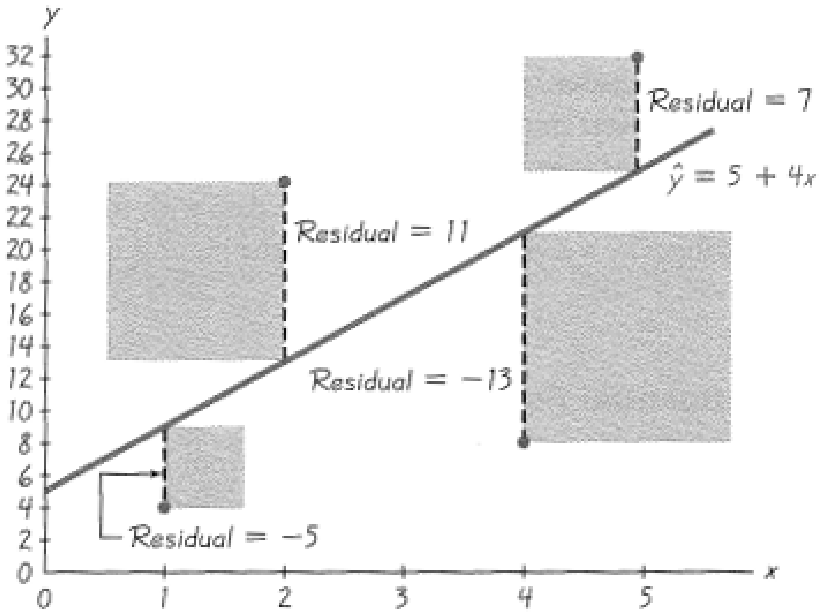
1. $(w_u + w_v) \in H$ (trivial)
2. $(w - (w_u + w_v)) \perp u$ (use the fact that $u \perp v$)
3. $(w - (w_u + w_v)) \perp v$ (same idea)

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 closes
to the data points in the least squares sense.



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 sur of the squares of the errors. The line that best fits the data in th 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.

t Squares Method

t L be a line with slope m and y-intercept b. Let $_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of paired data. Then the llowing equations hold:

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

here the ϵ_i are the errors $(i = 1, \ldots, n)$. This system is equivalent the following vector equation,

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

here 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

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

We are trying to choose m, b so that

$$\|E\|^2 = \|Y - AV\|^2 \quad (1$$

is minimal.

t

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

hen $AV = mX + bB$. The set $S = \{AV | m, b \in \mathbb{R}\}$ is a plane in
dimensional space. The ordered pair (m, b) that minimizes the
uared 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

$$C = B - \left(\frac{B \cdot X}{X \cdot X}\right) X \quad (1$$

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

$$Y_S = Y_X + Y_C = \left(\frac{Y \cdot X}{X \cdot X}\right) X + \left(\frac{Y \cdot C}{C \cdot C}\right) C \quad (1$$

t Squares Method

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

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

alternatively

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

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

th

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

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 relationshi 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.

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

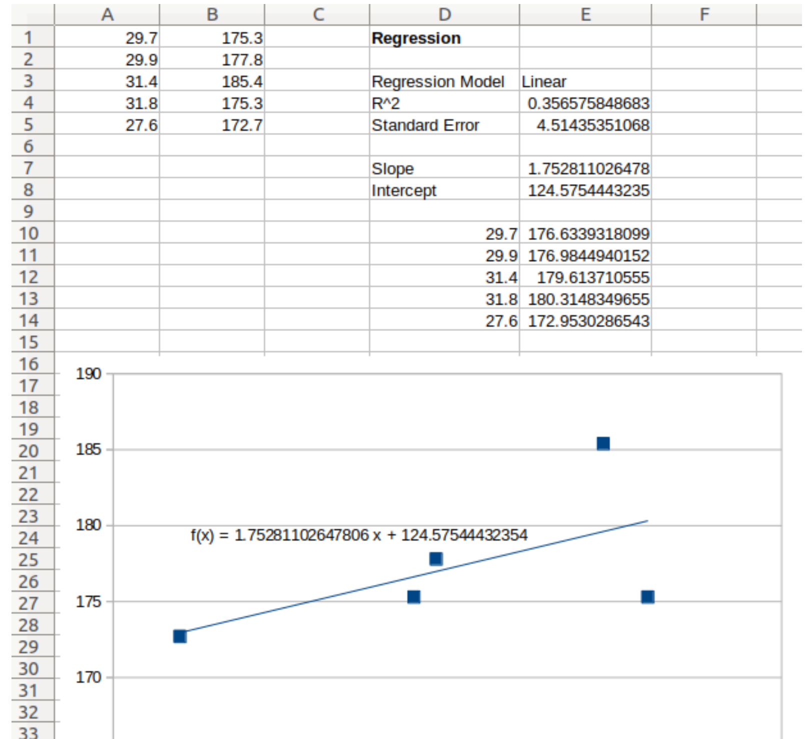
lculate C using the formula

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

ow calculate m and b

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

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



	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	
15						

les at Gray Cliff

ample 4: Angles at Gray Cliff. This example is from scar S. Adams's *Application of the Theory of Least Squares to e Adjustment of Triangulation* (1915), a "working manual for the mputer 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	$8°39'14.2''$
ound 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 ar

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

onvert the angles to real numbers

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

he sum is 359.999527778. Here is a system of equations with easurement errors, exploiting the fact that d is supposed to be $0^\circ - (a + b + c)$

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

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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 65.108 \\ 19.774 \\ 8.6539 \\ -93.537 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \quad (2$$

In symbols,

$$AV = Y + E \quad (2$$

Again, we want to minimize

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

les at Gray Cliff

he minimization is achieved by projecting Y onto the hyperplane

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

hese three vectors α, β, γ form a basis of S, but the basis vectors e not orthogonal to each other. We will search for a different sis 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 secon 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 $b_3 = (0, 1, -1, 0)^\intercal$.

t Y_i be the projection of Y onto the line spanned by b_i, for
ample.

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

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

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

he least squares adjusted angles are
$°6'29.7'', 19°46'27.3'', 8°39'14.6''$ compared to the original
$°6'29.3'', 19°46'26.9'', 8°39'14.2''$.

Example 5: Abigail, Ben, and Charlie. Here is an example adapted from Paul R. Wolf's *Survey Measurement Adjustments* *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 repo 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

ail, Ben, and Charlie [Calculus]

ere are the three equations with errors:

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

efine the function

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

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

e 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 syste of linear equations.

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1286.46 \\ 1303.62 \end{bmatrix} \quad (3$$

The solution is $a = 211.55, b = 220.13$ for Abigail and Ben's leas 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.

uation (29) translates into the following matrix equation.

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

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

Find the projection Y_S of Y onto the hyperplane S defined by A 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 and define $w = v - v_u$ (successive orthogonal selection).

$$w = v - v_u = v - \left(\frac{v \cdot u}{u \cdot u}\right) u = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix}$$ (3

Consequently,

$$Y_S = Y_u + Y_w = \left(\frac{Y \cdot u}{u \cdot u}\right) u + \left(\frac{Y \cdot w}{w \cdot w}\right) w = \begin{bmatrix} 211.55 \\ 220.13 \\ 431.68 \end{bmatrix}$$ (3

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

ail, Ben, and Charlie [Linear Algebra]

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

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

hy does this formula work?

Abigail, Ben, and Charlie [Linear Algebra]

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 t variables in V are still free. $Y - AV_0$ is orthogonal to S. S is the hyperplane $\{AU | U \text{ is a vector with the right dimensions}\}$. Consequently,

$$(Y - AV_0) \cdot AU = 0$$ (3

for some U with the right dimensions. Rewrite

$$Y \cdot AU = AV_0 \cdot AU$$ (4

r column vectors u and v, it is generally true that $u \cdot v = u^\intercal v$.
so recall that for any two matrices A and B

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

sing these two facts, continue with

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

d

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

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

$$(A^\intercal Y)^\intercal U = (A^\intercal AV_0)^\intercal U \quad (4$$

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

$$(A^\intercal Y)^\intercal = (A^\intercal AV_0)^\intercal \quad (4$$

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

$$V_0 = (A^\intercal A)^{-1} A^\intercal Y \quad (4$$

e Print and Height

ow that we have a formula, let us return to the problem of shoe
int 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

he 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

$$V_0 = (A^\intercal A)^{-1} A^\intercal Y = \begin{bmatrix} 1.7528 \\ 124.5754 \end{bmatrix} \quad (4$$

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

xercise 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 a that altitude based on other measurements, assuming a linear relationship. How does the result compare to the actual recorded value of 48°F?

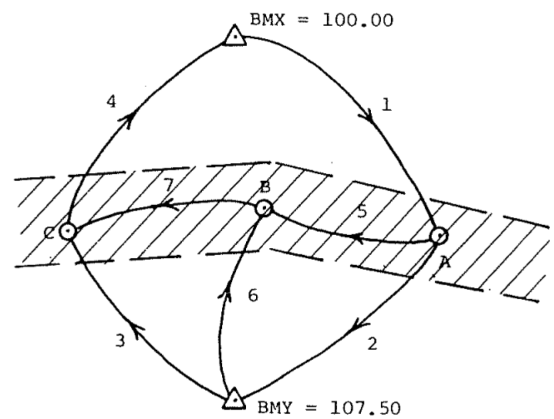
Altitude	3	10	14	22	28	31	33
Temperature	57	37	24	-5	-30	-41	-54

t Squares Exercises

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

Least Squares Exercises

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

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

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

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

t Squares Exercises

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

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

Least Squares Exercises

The first set of equation for the seven courses is

$$
\begin{aligned}
A &= BMX + 5.10 + \epsilon_1 \\
BMY &= A + 2.34 + \epsilon_2 \\
C &= BMY - 1.25 + \epsilon_3 \\
BMX &= C - 6.13 + \epsilon_4 \\
B &= A - 0.68 + \epsilon_5 \\
B &= BMY - 3.00 + \epsilon_6 \\
C &= B + 1.70 + \epsilon_7
\end{aligned}
\qquad (4
$$

he resulting yave setup is

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

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

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 yo 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?

arizing Distance Observation Equations

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

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

e shall linearize the fifth equation using the Taylor polynomial pansion 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 $x = a$,

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots \qquad (5$$

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

$$f(x) \approx f(a) + f'(a)(x-a) \qquad (5$$

ithout knowing it, we used Taylor polynomials for Newton's ethod. We were trying to solve $f(x) = 0$ but could not because gave us a non-linear equation. Instead, we solved

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

hich is equivalent to

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

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

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

$$G(x, y, z, w) = \sqrt{(z - x)^2 + (w - y)^2} \qquad (5$$

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

$$F(\vec{x}) = F(\vec{a}) + J(\vec{a})(\vec{x} - \vec{a}) \qquad (5$$

where J is the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} \qquad (5$$

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

arizing Distance Observation Equations

he Jacobian for G is

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

herefore

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

Matrix Methods and Statistics for Geomatics Least Squares Method

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,

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

nally, we have a linear yave setup.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.99100 & 0.13385 & -0.99100 & -0.13385 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} =$$

$$\begin{bmatrix} 8.3995 \\ 3.0161 \\ -2.872 \\ 1.4937 \\ 11.391 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \quad (59)$$

Use the formula

$$V_0 = (A^\intercal A)^{-1} A^\intercal Y \quad (6$$

for

$$V_0 = (8.4052, 3.0169, -2.8777, 1.4929)^\intercal \quad (6$$

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

arizing Angle Observation Equations

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

Linearizing Angle Observation Equations

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

$$\begin{aligned} I &= (595.74, 537.76) \\ J &= (800.92, 658.44) \\ K &= (302.96, 168.88) \end{aligned} \quad (6$$

From station I, you observe an angle of $158°49'21''$ instead of th expected $158°54'5.9107''$ between $\vec{IJ}$ and $\vec{IK}$. How should you lea 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.)

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

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

ne seventh equation is non-linear.

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

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

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

The Jacobian is

$$\begin{bmatrix} \frac{w-y}{\|\vec{IJ}\|^2} - \frac{v-y}{\|\vec{IK}\|^2} \\ \frac{z-x}{\|\vec{IJ}\|^2} - \frac{u-x}{\|\vec{IK}\|^2} \\ \frac{w-y}{\|\vec{IJ}\|^2} \\ \frac{z-x}{\|\vec{IJ}\|^2} \\ \frac{v-y}{\|\vec{IK}\|^2} \\ \frac{u-x}{\|\vec{IK}\|^2} \end{bmatrix}^\intercal \quad (6$$

for the linearization

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

$$0.00494115229670866y + 0.00212980385748917z -$$

$$0.00166317554601289u - 0.00132006217838228v +$$

$$0.00362109011832638w - 9.71163063930754$$

arizing Angle Observation Equations

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

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

$^\intercal A)^{-1} A^\intercal Y$ 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.

lve

$$\begin{array}{rcrcl} \cos x & - & y & = & 0 \\ x & - & y^2 & = & 0 \end{array} \quad (68)$$

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

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

en the Jacobian is

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

The linearization of F is

$$F(x,y) \approx F(x_0,y_0) + \begin{bmatrix} -\sin x_0 & -1 \\ 1 & -2y_0 \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \quad (7$$

Setting it to zero translates into the matrix equation

$$\begin{bmatrix} -\sin x_0 & -1 \\ 1 & -2y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\cos x_0 - x_0 \sin x_0 \\ -y_0^2 \end{bmatrix} \quad (7$$

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

$$\begin{bmatrix} -0.565 & -1 \\ 1 & -1.6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1.164 \\ -0.64 \end{bmatrix} \quad (7$$

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

ems of Non-Linear Equations

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

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

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

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

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

Systems of Non-Linear Equations

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

$$C^2 = A \text{ where } A = \begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix} \quad (7$$

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

$$C = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \quad (7$$

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

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

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

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

The Jacobian matrix of F is

$$J = \begin{bmatrix} 2x_1 & x_3 & x_2 & 0 \\ x_2 & x_1 + x_4 & 0 & x_2 \\ x_3 & 0 & x_1 + x_4 & x_3 \\ 0 & x_3 & x_2 & 2x_4 \end{bmatrix} \quad (8$$

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

ems of Non-Linear Equations

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

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

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

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

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

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

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

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

Systems of Non-Linear Equations

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

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

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

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

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

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

$$\begin{bmatrix} 4 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 1 & 0 & 3 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ 5 \\ 1 \end{bmatrix} \quad (8$$

The solution is $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4) = (\frac{11}{6}, -\frac{1}{3}, \frac{5}{6}, \frac{2}{3})$.

ems of Non-Linear Equations

ne more application of this procedure gives us the approximate
lution $x_1 = 1.8278, x_2 = -0.4111, x_3 = 0.8278, x_4 = 0.5889$,
hich is pretty close to the true solution that we will learn how to
d using eigenvalues:

$$C = \begin{bmatrix} 2\sqrt{2} - 1 & -\sqrt{2} + 1 \\ 2\sqrt{2} - 2 & -\sqrt{2} + 2 \end{bmatrix} \approx \begin{bmatrix} 1.8284 & -0.4142 \\ 0.8284 & 0.5858 \end{bmatrix} \tag{84}$$

End of Lesson

Next Lesson: Eigenvalues and Eigenvectors