LINEAR ALGEBRA

Week 12

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Solution:

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} =$$

$$A^T \vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} =$$

The normal equations $A^T A\vec{x} = A^T\vec{b}$ become:

Theorem

Theorem (Unique Solutions for Least Squares)

Let A be any $m \times n$ matrix. These statements are equivalent.

1. The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b} \in \mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

And, if these statements hold, the least square solution is

$$\widehat{x} = (A^T A)^{-1} A^T \vec{b}.$$

Useful heuristic: $A^T A$ plays the role of 'length-squared' of the matrix A. (See the sections on symmetric matrices and singular value decomposition.)

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Hint: the columns of A are orthogonal.

Theorem (Least Squares and QR)

Let $m \times n$ matrix A have a QR decomposition. Then for each $\vec{b} \in \mathbb{R}^m$ the equation $A\vec{x} = \vec{b}$ has the unique least squares solution

$$R\hat{x} = Q^T\vec{b}.$$

(Remember, R is upper triangular, so the equation above is solved by back-substitution.)

THEOREM 15

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \qquad (6)$$

Example 3. Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

Solution. The QR decomposition of A is

$$A = QR = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

$$Q^T\vec{b} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} \\ -6 \\ 4 \end{bmatrix}$$

And then we solve by backwards substitution $R\vec{x} = Q^T\vec{b}$

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \\ -6 \\ 4 \end{bmatrix}$$

6.5 EXERCISES

In Exercises 1–4, find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ by (a) constructing the normal equations for $\hat{\mathbf{x}}$ and (b) solving for $\hat{\mathbf{x}}$.

1. $A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$

2. $A = \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}$

In Exercises 5 and 6, describe all least-squares solutions of the equation $A\mathbf{x} = \mathbf{b}$.

5. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 5 \\ 4 \end{bmatrix}$

7. Compute the least-squares error associated with the least-squares solution found in Exercise 3.

8. Compute the least-squares error associated with the least-squares solution found in Exercise 4.

In Exercises 9–12, find (a) the orthogonal projection of $\mathbf{b}$ onto Col A and (b) a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

9. $A = \begin{bmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ -2 \\ -3 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

11. $A = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -5 & 1 \\ 6 & 1 & 0 \\ 1 & -1 & -5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 9 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}$

13. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \\ 3 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 11 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Could $\mathbf{u}$ possibly be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

14. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

In Exercises 15 and 16, use the factorization $A = QR$ to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

15. $A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}$

16. $A = \begin{bmatrix} 1 & -1 \\ 1 & 4 \\ 1 & -1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$

In Exercises 17 and 18, A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$. Mark each statement True or False. Justify each answer.

17. a. The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

b. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that $\|\mathbf{b} - A\mathbf{x}\| \le \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

d. Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

e. If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

18. a. If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

b. The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the point in the column space of A closest to $\mathbf{b}$.

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a list of weights that, when applied to the columns of A, produces the orthogonal projection of $\mathbf{b}$ onto Col A.

d. If $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$, then $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

e. The normal equations always provide a reliable method for computing least-squares solutions.

f. If A has a QR factorization, say $A = QR$, then the best way to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

19. Let A be an $m \times n$ matrix. Use the steps below to show that a vector $\mathbf{x}$ in $\mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$ if and only if $A^TA\mathbf{x} = \mathbf{0}$. This will show that Nul A = Nul A^TA.

a. Show that if $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = \mathbf{0}$.

b. Suppose $A^TA\mathbf{x} = \mathbf{0}$. Explain why $\mathbf{x}^TA^TA\mathbf{x} = 0$, and use this to show that $A\mathbf{x} = \mathbf{0}$.

20. Let A be an $m \times n$ matrix such that A^TA is invertible. Show that the columns of A are linearly independent. [*Careful:* You may not assume that A is invertible; it may not even be square.]

21. Let A be an $m \times n$ matrix whose columns are linearly independent. [*Careful: A* need not be square.]

a. Use Exercise 19 to show that A^TA is an invertible matrix.

b. Explain why A must have at least as many rows as columns.

c. Determine the rank of A.

22. Use Exercise 19 to show that rank A^TA = rank A. [*Hint:* How many columns does A^TA have? How is this connected with the rank of A^TA?]

23. Suppose A is $m \times n$ with linearly independent columns and $\mathbf{b}$ is in $\mathbb{R}^m$. Use the normal equations to produce a formula for $\hat{\mathbf{b}}$, the projection of $\mathbf{b}$ onto Col A. [*Hint:* Find $\hat{\mathbf{x}}$ first. The formula does not require an orthogonal basis for Col A.]

24. Find a formula for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthonormal.

25. Describe all least-squares solutions of the system

$$x + y = 2$$
$$x + y = 4$$

26. [M] Example 3 in Section 4.8 displayed a low-pass linear filter that changed a signal $\{y_k\}$ into $\{y_{k+1}\}$ and changed a higher-frequency signal $\{w_k\}$ into the zero signal, where $y_k = \cos(\pi k/4)$ and $w_k = \cos(3\pi k/4)$. The following calculations will design a filter with approximately those properties. The filter equation is

$$a_0 y_{k+2} + a_1 y_{k+1} + a_2 y_k = z_k \quad \text{for all } k \qquad (8)$$

Because the signals are periodic, with period 8, it suffices to study equation (8) for $k = 0, \ldots, 7$. The action on the two signals described above translates into two sets of eight equations, shown below:

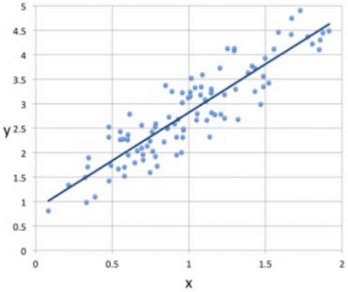
$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} \begin{array}{c} \begin{array}{ccc} y_{k+2} & y_{k+1} & y_k \end{array} \\ \begin{bmatrix} 0 & .7 & 1 \\ -.7 & 0 & .7 \\ -1 & -.7 & 0 \\ -.7 & -1 & -.7 \\ 0 & -.7 & -1 \\ .7 & 0 & -.7 \\ 1 & .7 & 0 \\ .7 & 1 & .7 \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{array}{c} y_{k+1} \\ \begin{bmatrix} .7 \\ 0 \\ -.7 \\ -1 \\ -.7 \\ 0 \\ .7 \\ 1 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} \begin{array}{c} \begin{array}{ccc} w_{k+2} & w_{k+1} & w_k \end{array} \\ \begin{bmatrix} 0 & -.7 & 1 \\ .7 & 0 & -.7 \\ -1 & .7 & 0 \\ .7 & -1 & .7 \\ 0 & .7 & -1 \\ -.7 & 0 & .7 \\ 1 & -.7 & 0 \\ -.7 & 1 & -.7 \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Write an equation $A\mathbf{x} = \mathbf{b}$, where A is a 16×3 matrix formed from the two coefficient matrices above and where $\mathbf{b}$ in $\mathbb{R}^{16}$ is formed from the two right sides of the equations. Find a_0, a_1, and a_2 given by the least-squares solution of $A\mathbf{x} = \mathbf{b}$. (The .7 in the data above was used as an approximation for $\sqrt{2}/2$, to illustrate how a typical computation in an applied problem might proceed. If .707 were used instead, the resulting filter coefficients would agree to at least seven decimal places with $\sqrt{2}/4$, $1/2$, and $\sqrt{2}/4$, the values produced by exact arithmetic calculations.)

WEB

Chapter 6 : Orthogonality and Least Squares
6.6 : Applications to Linear Models



Topics and Objectives

Topics

1. Least Squares Lines
2. Linear and more complicated models

Learning Objectives

For the topics covered in this section, students are expected to be able to do the following.

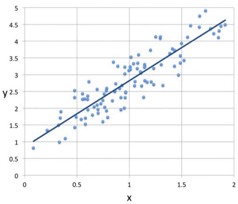
1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. Apply least-squares to fit polynomials and other curves to data.

Motivating Question

Compute the equation of the line $y = \beta_0 + \beta_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

Chapter 6 : Orthogonality and Least Squares
6.6 : Applications to Linear Models



Topics and Objectives

Topics

1. Least Squares Lines
2. Linear and more complicated models

Learning Objectives
For the topics covered in this section, students are expected to be able to do the following.

1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. Apply least-squares to fit polynomials and other curves to data.

Motivating Question
Compute the equation of the line $y = \beta_0 + \beta_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

Course Schedule

Cancellations due to inclement weather will likely result in cancelling review lectures and possibly moving through course

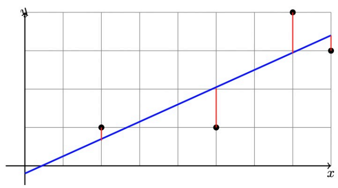
Week	Dates	Mon Lecture	Tue Studio	Wed Lecture	Thu Studio	Fri Lecture
1	8/21 – 8/25	1.1	WS1.1	1.2	WS1.2	1.3
2	8/28 – 9/1	1.4	WS1.3,1.4	1.5	WS1.5	1.7
3	9/4 – 9/8	Break	WS1.7	1.8	WS1.8	1.9
4	9/11 – 9/15	2.1	WS1.9,2.1	**Exam 1**, Review	Cancelled	2.2
5	9/18 – 9/22	2.3,2.4	WS2.2,2.3	2.5	WS2.4,2.5	2.8
6	9/25 – 9/29	2.9	WS2.8,2.9	3.1,3.2	WS3.1,3.2	3.3
7	10/2 – 10/6	4.9	WS3.3,4.9	5.1,5.2	WS5.1,5.2	5.2
8	10/9 – 10/13	Break	Break	**Exam 2**, Review	Cancelled	5.3
9	10/16 – 10/20	5.3	WS5.3	5.5	WS5.5	6.1
10	10/23 – 10/27	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	10/30 – 11/3	6.4	WS6.3,6.4	6.4,6.5	WS6.4,6.5	6.5
12	11/6 – 11/10	6.6	WS6.5,6.6	**Exam 3**, Review	Cancelled	PageRank
13	11/13 – 11/17	7.1	WSPageRank	7.2	WS7.1,7.2	7.3
14	11/20 – 11/24	7.3,7.4	WS7.2,7.3	Break	Break	Break
15	11/27 – 12/1	7.4	WS7.3,7.4	7.4	WS7.4	7.4
16	12/4 – 12/8	Last lecture	Last Studio	Reading Period		
17	12/11 – 12/15	Final Exams: **MATH 1554 Common Final Exam Tuesday, December 12th at 6pm**				

The Least Squares Line

Graph below gives an approximate linear relationship between x and y.

1. Black circles are data.
2. Blue line is the **least squares** line.
3. Lengths of red lines are the ___________.

The least squares line minimizes the sum of squares of the ___________.



Example 1 Compute the least squares line $y = \beta_0 + \beta_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

We want to solve

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

This is a least-squares problem : $X\vec{\beta} = \vec{y}$.

The normal equations are

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & & \\ & & & & & \end{bmatrix} \begin{bmatrix} 1 & \\ 1 & \\ 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & \\ & & & & \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x = \frac{-5}{21} + \frac{19}{42} x$$

As we may have guessed, β_0 is negative, and β_1 is positive.

Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

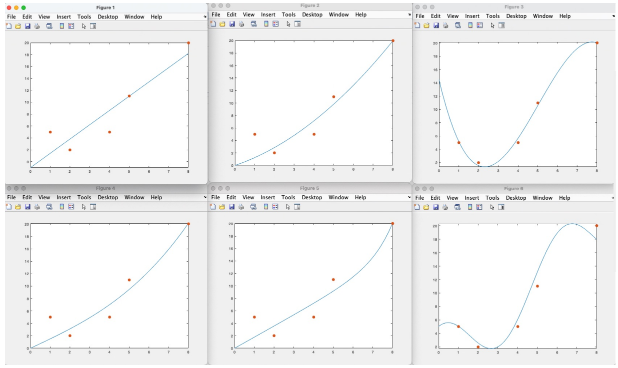
$$y = \beta_0 + \beta_1 f_1(x) + \beta_1 f_2(x) + \cdots + \beta_k f_k(x).$$

where the functions f_j are known. Should have only a few functions!
Keep in mind this is a **linear problem in the β variables.**

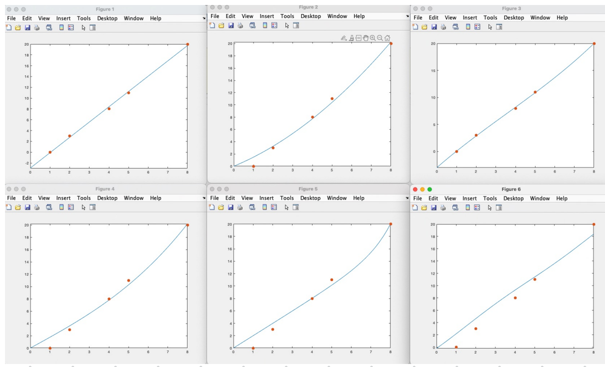
https://piazza.com/class/lldv6iezpqh7jy/post/776

```
clc
format shortG

%% dataÅ
data=[1 5 ; 2 2 ; 4 5 ; 5 11 ; 8 20];
%data=[1 0 ; 2 3 ; 4 8 ; 5 11 ; 8 20];
%data=[1 1 ; 2 1 ; 4 8 ; 5 11 ; 8 20];
% x-values in column 1
x=data(:,1);
% y-values in column 2
y=data(:,2);
% plot points with 28pt dots
%plotdata=plot(x,y,'.','markersize',28);

%% first model: linear y=alpha*x+beta
A=[x(1) 1 ; x(2) 1; x(3) 1 ; x(4) 1 ; x(5) 1];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f1=@(x) alpha.*x+beta;
errorLine=sqrt((A*xhat-b)'*(A*xhat-b))

%% second model: quadratic y=alpha*x+beta*x^2
A=[x(1) x(1)^2 ; x(2) x(2)^2 ; x(3) x(3)^2 ; x(4) x(4)^2 ; x(5) x(5)^2];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f2=@(x) alpha.*x+beta.*x.^2;
errorQuad=sqrt((A*xhat-b)'*(A*xhat-b))

%% third model: general cubic y=a0+a1*x+a2*x^2+a3*x^3
A=[1 x(1) x(1)^2 x(1)^3 ; 1 x(2) x(2)^2 x(2)^3 ; 1 x(3) x(3)^2 x(3)^3 ; 1 x(4) x(4)^2 x(4)^3 ; 1 x(5)
x(5)^2 x(5)^3];
b=y;
xhat=inv(A'*A)*A'*b;
a0=xhat(1);
a1=xhat(2);
a2=xhat(3);
a3=xhat(4);
f3=@(x) a0+a1.*x+a2.*x.^2+a3.*x.^3;
errorGeneralCubic=sqrt((A*xhat-b)'*(A*xhat-b))

%% fourth model: truncated cubic y=a1*x+a3*x^3
A=[x(1) x(1)^3 ; x(2) x(2)^3 ; x(3) x(3)^3 ; x(4) x(4)^3 ; x(5) x(5)^3];
b=y;
xhat=inv(A'*A)*A'*b;
a1=xhat(1);
a3=xhat(2);
f4=@(x) a1.*x+a3.*x.^3;
errorTruncatedCubic=sqrt((A*xhat-b)'*(A*xhat-b))

%% fifth model: linear+exponential y=alpha*x+beta*exp(x)
A=[x(1) exp(x(1)) ; x(2) exp(x(2)) ; x(3) exp(x(3)) ; x(4) exp(x(4)) ; x(5) exp(x(5))];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f5=@(x) alpha.*x+beta.*exp(x);
errorExp=sqrt((A*xhat-b)'*(A*xhat-b))

%% sixth model: linear+cosine y=alpha*x+beta*cos(x)
A=[x(1) cos(x(1)) ; x(2) cos(x(2)) ; x(3) cos(x(3)) ; x(4) cos(x(4)) ; x(5) cos(x(5))];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f6=@(x) alpha.*x+beta.*cos(x);
errorCos=sqrt((A*xhat-b)'*(A*xhat-b))

%% draw figure
figure(1)
fplot(f1,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(2)
fplot(f2,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(3)
fplot(f3,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(4)
fplot(f4,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(5)
fplot(f5,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(6)
fplot(f6,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off
```

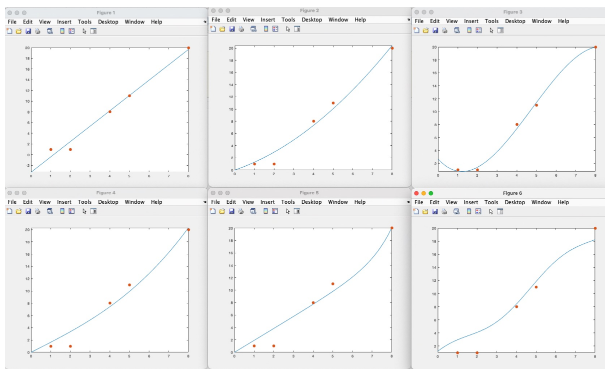
```
data =

     1     5
     2     2
     4     5
     5    11
     8    20

data =

     1     0
     2     3
     4     8
     5    11
     8    20

data =

     1     1
     2     1
     4     8
     5    11
     8    20
```




```
A =

     1     1
     2     1
     4     1
     5     1
     8     1

xhat =

       2.4667
      -1.0667

errorLine =

       3.5024
```

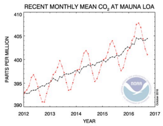



```
A =

     1     1
     2     4
     4    16
     5    25
     8    64

xhat =

       1.3882
      0.13705

errorQuad =

       2.4685
```




```
A =

     1     1
     2     8
     4    64
     5   125
     8   512

xhat =

       1.7102
     0.012391

errorCubic =

       2.3115
```

Least Squares Fitting for Other Curves



Black line is yearly CO_2 levels, and the monthly is the red line. To capture seasonality, would need a curve

$$\text{daily } CO_2 = \beta_0 + \beta_1 t + \beta_2 \sin(2\pi \tfrac{t}{12}) + \beta_3 \cos(2\pi \tfrac{t}{12})$$

Above, t is time, measured in months.

WolframAlpha and Mathematica Syntax

Least squares problems can be computed with WolframAlpha, Mathematica, and many other software.

WolframAlpha

$$\texttt{linear fit } \{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\}$$

Mathematica

$$\texttt{LeastSquares}[\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \dots, \{x_n, x_n, y_n\}\}]$$

Almost any spreadsheet program does this as a function as well.

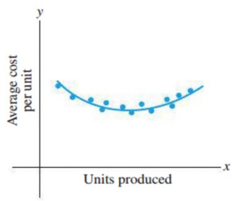


FIGURE 3
Average cost curve.

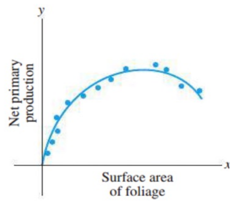


FIGURE 4
Production of nutrients.

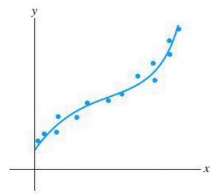


FIGURE 5
Data points along a cubic curve.

Theorem

Theorem (Unique Solutions for Least Squares)

Let A be any $m \times n$ matrix. These statements are equivalent.

1. The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b} \in \mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

And, if these statements hold, the least square solution is

$$\widehat{x} = (A^T A)^{-1} A^T \vec{b}.$$

Useful heuristic: $A^T A$ plays the role of 'length-squared' of the matrix A. (See the sections on symmetric matrices and singular value decomposition.)

Theorem (Least Squares and QR)

Let $m \times n$ matrix A have a QR decomposition. Then for each $\vec{b} \in \mathbb{R}^m$ the equation $A\vec{x} = \vec{b}$ has the unique least squares solution

$$R\widehat{x} = Q^T \vec{b}.$$

(Remember, R is upper triangular, so the equation above is solved by back-substitution.)

6.6 EXERCISES

In Exercises 1–4, find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the given data points.

1. $(0, 1), (1, 1), (2, 2), (3, 2)$
2. $(1, 0), (2, 1), (4, 2), (5, 3)$
3. $(-1, 0), (0, 1), (1, 2), (2, 4)$
4. $(2, 3), (3, 2), (5, 1), (6, 0)$
5. Let X be the design matrix used to find the least-squares line to fit data $(x_1, y_1), \dots, (x_n, y_n)$. Use a theorem in Section 6.5 to show that the normal equations have a unique solution if and only if the data include at least two data points with different x-coordinates.
6. Let X be the design matrix in Example 2 corresponding to a least-squares fit of a parabola to data $(x_1, y_1), \dots, (x_n, y_n)$. Suppose x_1, x_2, and x_3 are distinct. Explain why there is only one parabola that fits the data best, in a least-squares sense. (See Exercise 5.)
7. A certain experiment produces the data $(1, 1.8)$, $(2, 2.7)$, $(3, 3.4)$, $(4, 3.8)$, $(5, 3.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

   $$y = \beta_1 x + \beta_2 x^2$$

   Such a function might arise, for example, as the revenue from the sale of x units of a product, when the amount offered for sale affects the price to be set for the product.

   a. Give the design matrix, the observation vector, and the unknown parameter vector.

   b. **[M]** Find the associated least-squares curve for the data.
8. A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level x, has the form $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. There is no constant term because fixed costs are not included.

   a. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data $(x_1, y_1), \dots, (x_n, y_n)$.

   b. **[M]** Find the least-squares curve of the form above to fit the data $(4, 1.58)$, $(6, 2.08)$, $(8, 2.5)$, $(10, 2.8)$, $(12, 3.1)$, $(14, 3.4)$, $(16, 3.8)$, and $(18, 4.32)$, with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.
9. A certain experiment produces the data $(1, 7.9)$, $(2, 5.4)$, and $(3, -.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

   $$y = A \cos x + B \sin x$$
10. Suppose radioactive substances A and B have decay constants of .02 and .07, respectively. If a mixture of these two substances at time $t = 0$ contains M_A grams of A and M_B grams of B, then a model for the total amount y of the mixture present at time t is

    $$y = M_A e^{-.02t} + M_B e^{-.07t} \tag{6}$$

    Suppose the initial amounts M_A and M_B are unknown, but a scientist is able to measure the total amounts present at several times and records the following points (t_i, y_i): $(10, 21.34)$, $(11, 20.68)$, $(12, 20.05)$, $(14, 18.87)$, and $(15, 18.30)$.

    a. Describe a linear model that can be used to estimate M_A and M_B.

    b. **[M]** Find the least-squares curve based on (6).

Halley's Comet last appeared in 1986 and will reappear in 2061.

11. **[M]** According to Kepler's first law, a comet should have an elliptic, parabolic, or hyperbolic orbit (with gravitational attractions from the planets ignored). In suitable polar coordinates, the position (r, ϑ) of a comet satisfies an equation of the form

    $$r = \beta + e(r \cdot \cos \vartheta)$$

    where β is a constant and e is the *eccentricity* of the orbit, with $0 \le e < 1$ for an ellipse, $e = 1$ for a parabola, and $e > 1$ for a hyperbola. Suppose observations of a newly discovered comet provide the data below. Determine the type of orbit, and predict where the comet will be when $\vartheta = 4.6$ (radians).[3]

ϑ	.88	1.10	1.42	1.77	2.14
r	3.00	2.30	1.65	1.25	1.01

12. **[M]** A healthy child's systolic blood pressure p (in millimeters of mercury) and weight w (in pounds) are approximately related by the equation

    $$\beta_0 + \beta_1 \ln w = p$$

    Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

[3] The basic idea of least-squares fitting of data is due to K. F. Gauss (and, independently, to A. Legendre), whose initial rise to fame occurred in 1801 when he used the method to determine the path of the asteroid *Ceres*. Forty days after the asteroid was discovered, it disappeared behind the sun. Gauss predicted it would appear ten months later and gave its location. The accuracy of the prediction astonished the European scientific community.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

13. [M] To measure the takeoff performance of an airplane, the horizontal position of the plane was measured every second, from $t = 0$ to $t = 12$. The positions (in feet) were: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7, 686.8, and 809.2.

a. Find the least-squares cubic curve $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ for these data.

b. Use the result of part (a) to estimate the velocity of the plane when $t = 4.5$ seconds.

14. Let $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ and $\overline{y} = \frac{1}{n}(y_1 + \cdots + y_n)$. Show that the least-squares line for the data $(x_1, y_1), \ldots, (x_n, y_n)$ must pass through $(\overline{x}, \overline{y})$. That is, show that $\overline{x}$ and $\overline{y}$ satisfy the linear equation $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1\overline{x}$. [*Hint:* Derive this equation from the vector equation $\mathbf{y} = X\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$. Denote the first column of X by $\mathbf{1}$. Use the fact that the residual vector $\boldsymbol{\epsilon}$ is orthogonal to the column space of X and hence is orthogonal to $\mathbf{1}$.]

Given data for a least-squares problem, $(x_1, y_1), \ldots, (x_n, y_n)$, the following abbreviations are helpful:

$$\textstyle\sum x = \sum_{i=1}^n x_i, \quad \sum x^2 = \sum_{i=1}^n x_i^2,$$
$$\textstyle\sum y = \sum_{i=1}^n y_i, \quad \sum xy = \sum_{i=1}^n x_i y_i$$

The normal equations for a least-squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ may be written in the form

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \textstyle\sum x &= \textstyle\sum y \\ \hat{\beta}_0 \textstyle\sum x + \hat{\beta}_1 \sum x^2 &= \textstyle\sum xy \end{aligned} \qquad (7)$$

15. Derive the normal equations (7) from the matrix form given in this section.

16. Use a matrix inverse to solve the system of equations in (7) and thereby obtain formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ that appear in many statistics texts.

17. a. Rewrite the data in Example 1 with new x-coordinates in mean deviation form. Let X be the associated design matrix. Why are the columns of X orthogonal?

b. Write the normal equations for the data in part (a), and solve them to find the least-squares line, $y = \beta_0 + \beta_1 x^*$, where $x^* = x - 5.5$.

18. Suppose the x-coordinates of the data $(x_1, y_1), \ldots, (x_n, y_n)$ are in mean deviation form, so that $\sum x_i = 0$. Show that if X is the design matrix for the least-squares line in this case, then X^TX is a diagonal matrix.

Exercises 19 and 20 involve a design matrix X with two or more columns and a least-squares solution $\hat{\boldsymbol{\beta}}$ of $\mathbf{y} = X\boldsymbol{\beta}$. Consider the following numbers.

(i) $\|X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares of the "regression term." Denote this number by SS(R).

(ii) $\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares for error term. Denote this number by SS(E).

(iii) $\|\mathbf{y}\|^2$—the "total" sum of the squares of the y-values. Denote this number by SS(T).

Every statistics text that discusses regression and the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ introduces these numbers, though terminology and notation vary somewhat. To simplify matters, assume that the mean of the y-values is zero. In this case, SS(T) is proportional to what is called the *variance* of the set of y-values.

19. Justify the equation SS(T) = SS(R) + SS(E). [*Hint:* Use a theorem, and explain why the hypotheses of the theorem are satisfied.] This equation is extremely important in statistics, both in regression theory and in the analysis of variance.

20. Show that $\|X\hat{\boldsymbol{\beta}}\|^2 = \hat{\boldsymbol{\beta}}^T X^T\mathbf{y}$. [*Hint:* Rewrite the left side and use the fact that $\hat{\boldsymbol{\beta}}$ satisfies the normal equations.] This formula for SS(R) is used in statistics. From this and from Exercise 19, obtain the standard formula for SS(E):

$$\text{SS(E)} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T\mathbf{y}$$

Midterm 3 Lecture Review Activity, Math 1554

1. Indicate **true** if the statement is true, otherwise, indicate **false**.

	true	false
a) If S is a two-dimensional subspace of $\mathbb{R}^{50}$, then the dimension of $S^\perp$ is 48.	○	○
b) An eigenspace is a subspace spanned by a single eigenvector.	○	○
c) The $n \times n$ zero matrix can be diagonalized.	○	○
d) A least-squares line that best fits the data points $(0, y_1), (1, y_2), (2, y_3)$ is unique for any values y_1, y_2, y_3.	○	○

2. If possible, give an example of the following.

2.1) A matrix, A, that is in echelon form, and $\dim\left((\text{Row}A)^\perp\right) = 2$, $\dim\left((\text{Col}\,A)^\perp\right) = 1$

2.2) A singular 2×2 matrix whose eigenspace corresponding to eigenvalue $\lambda = 2$ is the line $x_1 = 2x_2$. The other eigenspace of the matrix is the x_2 axis.

2.3) A subspace S, of $\mathbb{R}^4$, that satisfies $\dim(S) = \dim(S^\perp) = 3$.

2.4) A 2×3 matrix, A, that is in RREF. $(\text{Row}\,A)^\perp$ is spanned by $\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$.

3. Circle **possible** if the set of conditions are create a situation that is possible, otherwise, circle **impossible**. For the situations that are possible give an example.

3.1) A is $n \times n$, $A\vec{x} = A\vec{y}$ for a particular $\vec{x} \neq \vec{y}$, $\vec{x}$ and $\vec{y}$ are in $\mathbb{R}^n$, and $\dim((\text{Row } A)^\perp) \neq 0$.

possible **impossible**

3.2) A is $n \times n$, $\lambda \in \mathbb{R}$ is an eigenvalue of A, and $\dim((\text{Col}(A - \lambda I))^\perp) = 0$.

possible **impossible**

3.3) $\text{proj}_{\vec{v}}\vec{u} = \text{proj}_{\vec{u}}\vec{v}$, $\vec{v} \neq \vec{u}$, and $\vec{u} \neq \vec{0}$, $\vec{v} \neq \vec{0}$.

possible **impossible**

4. Consider the matrix A.

$$A = \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Construct a basis for the following subspaces and state the dimension of each space.

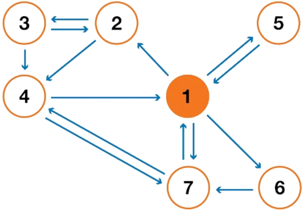
4.1) $(\text{Row}A)^\perp$

4.2) Col A

4.3) $(\text{Col}A)^\perp$

Chapter 10 : Finite-State Markov Chains

10.2 : The Steady-State Vector and Page Rank



Trajectory : 1

Topics and Objectives

Topics

1. Review of Markov chains
2. Theorem describing the steady state of a Markov chain
3. Applying Markov chains to model website usage.
4. Calculating the PageRank of a web.

Learning Objectives

1. Determine whether a stochastic matrix is regular.
2. Apply matrix powers and theorems to characterize the long-term behaviour of a Markov chain.
3. Construct a transition matrix, a Markov Chain, and a Google Matrix for a given web, and compute the PageRank of the web.

Chapter 10 : Finite-State Markov Chains

10.2 : The Steady-State Vector and Page Rank

Topics and Objectives

Topics

1. Review of Markov chains
2. Theorem describing the steady state of a Markov chain
3. Applying Markov chains to model website usage.
4. Calculating the PageRank of a web.

Learning Objectives

1. Determine whether a stochastic matrix is regular.
2. Apply matrix powers and theorems to characterize the long-term behaviour of a Markov chain.
3. Construct a transition matrix, a Markov Chain, and a Google Matrix for a given web, and compute the PageRank of the web.

Course Schedule

Cancellations due to inclement weather will likely result in cancelling review lectures and possibly moving through course material at a faster pace.

Week	Dates	Mon Lecture	Tue Studio	Wed Lecture	Thu Studio	Fri Lecture
1	8/21 – 8/25	1.1	WS1.1	1.2	WS1.2	1.3
2	8/28 – 9/1	1.4	WS1.3,1.4	1.5	WS1.5	1.7
3	9/4 – 9/8	Break	WS1.7	1.8	WS1.8	1.9
4	9/11 – 9/15	2.1	WS1.9,2.1	**Exam 1**, Review	Cancelled	2.2
5	9/18 – 9/22	2.3,2.4	WS2.2,2.3	2.5	WS2.4,2.5	2.8
6	9/25 – 9/29	2.9	WS2.8,2.9	3.1,3.2	WS3.1,3.2	3.3
7	10/2 – 10/6	4.9	WS3.3,4.9	5.1,5.2	WS5.1,5.2	5.2
8	10/9 – 10/13	Break	Break	**Exam 2**, Review	Cancelled	5.3
9	10/16 – 10/20	5.3	WS5.3	5.5	WS5.5	6.1
10	10/23 – 10/27	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	10/30 – 11/3	6.4	WS6.3,6.4	6.4,6.5	WS6.4,6.5	6.5
12	11/6 – 11/10	6.6	WS6.5,6.6	**Exam 3**, Review	Cancelled	PageRank
13	11/13 – 11/17	7.1	WSPageRank	7.2	WS7.1,7.2	7.3
14	11/20 – 11/24	7.3,7.4	WS7.2,7.3	Break	Break	Break
15	11/27 – 12/1	7.4	WS7.3,7.4	7.4	WS7.4	7.4
16	12/4 – 12/8	Last lecture	Last Studio	Reading Period		
17	12/11 – 12/15	Final Exams: **MATH 1554 Common Final Exam Tuesday, December 12th at 6pm**				

Where is Chapter 10?

- The material for this part of the course is covered in Section 10.2
- Chapter 10 is not included in the **print** version of the book, but it is in the **on-line version**.
- If you read 10.2, and I recommend that you do, you will find that it requires an understanding of 10.1.
- You are not required to understand the material in 10.1.

Other sources that you may find helpful are listed below.

1. PageRank Algorithm (Math Explorer's Club, Cornell Univ.) `http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html`
2. Austin, D. *How Google Finds Your Needle in the Web's Haystack.* Available at: `http://www.ams.org/samplings/feature-column/fcarc-pagerank`
3. Bryan, K., Leise, T. *The $25,000,000,000 Eigenvector: The Linear Algebra behind Google.* SIAM Review, 48(3). Available at: `http://userpages.umbc.edu/~kogan/teaching/m430/GooglePageRank.pdf`

Steady State Vectors

Recall the car rental problem from our Section 4.9 lecture.

Problem

A car rental company has 3 rental locations, A, B, and C.

		rented from		
		A	B	C
returned to	A	.8	.1	.2
	B	.2	.6	.3
	C	.0	.3	.5

There are 10 cars at each location today, what happens to the distribution of cars after a long time?

Long Term Behaviour

Can use the transition matrix, P, to find the distribution of cars after 1 week:

$$\vec{x}_1 = P\vec{x}_0$$

The distribution of cars after 2 weeks is:

$$\vec{x}_2 = P\vec{x}_1 =$$

The distribution of cars after n weeks is:

Long Term Behaviour

To investigate the long-term behaviour of a system that has a regular transition matrix P, we could:

1. compute $P^n\vec{x}_0$ for large n.
2. compute the **steady-state vector**, $\vec{q}$, by solving $\vec{q} = P\vec{q}$.

To solve PageRank problems, we will rely on the first approach.

Theorem 1

If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

1. There is a stochastic matrix Π such that

$$\lim_{n\to\infty} P^n = \Pi$$

2. Each column of Π is the same probability vector $\vec{q}$.
3. For any initial probability vector $\vec{x}_0$,

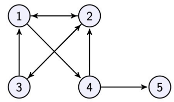
$$\lim_{n\to\infty} P^n \vec{x}_0 = \vec{q}$$

4. P has a unique eigenvector, $\vec{q}$, which has eigenvalue $\lambda = 1$.
5. The eigenvalues of P satisfy $|\lambda| \leq 1$.

We will apply this theorem when solving PageRank problems.

Example 1

Suppose we have 4 web pages that link to each other according to this diagram.



Page 1 has links to pages ________________ .

Page 2 has links to pages ________________ .

If a user on a page in this web is **equally likely** to go to any of the pages that their page links to, construct a Markov chain that represents how users navigate this web.


```
clc
format bank
%% rental car - long term analysis
A=[.8 .1 .2 ; .2 .6 .3 ; 0 .3 .5]
k=10
A^k


%% google PageRank
P0=[0   1/2 1/2 0   1/5 ;
    1/2 0   1/2 1/2 1/5 ;
    0   1/2 0   0   1/5 ;
    1/2 0   0   0   1/5 ;
    0   0   0   1/2 1/5 ];
K0=(1/5)*[1 1 1 1 1 ;
    1 1 1 1 1 ;
    1 1 1 1 1;
    1 1 1 1 1 ;
    1 1 1 1 1];
G0=.85*P0+.15*K0;
P1=[0 0 1 0 1/5
    1/3 0 0 1/2 1/5 ;
    1/3 0 0 1/2 1/5 ;
    1/3 1/2 0 0 1/5 ;
    0 1/2 0 0 1/5];
G1=.85*P1+.15*K0;
P2=[0 1/2 0   1/7 0   0   1/7 ;
    0 0   1/3 1/7 1/2 0   1/7 ;
    1 0   0   1/7 0   1/3 1/7 ;
    0 0   1/3 1/7 0   0   1/7 ;
    0 1/2 0   1/7 0   0   1/7 ;
    0 0   1/3 1/7 1/2   1/3 1/7;
    0 0   0   1/7 0   1/3 1/7];
K2=1/7*[1 1 1 1 1 1 1;
    1 1 1 1 1 1 1;
    1 1 1 1 1 1 1;
    1 1 1 1 1 1 1;
    1 1 1 1 1 1 1;
    1 1 1 1 1 1 1;
    1 1 1 1 1 1 1];
G2=.85*P2+.15*K2;
k=20;
test=[1 1 1 1 1 1 1 ] *G2;
format short
for i=1:k
    i;
    G1^i;
end
```

Transition Matrix, Importance, and PageRank

- The square matrix we constructed in the previous example is a **transition matrix**. It describes how users transition between pages in the web.
- The steady-state vector, $\vec{q}$, for the Markov-chain, can characterize the long-term behavior of users in a given web.
- If $\vec{q}$ is unique, the **importance** of a page in a web is given by its corresponding entry in $\vec{q}$.
- The **PageRank** is the ranking assigned to each page based on its importance. The highest ranked page has PageRank 1, the second PageRank 2, and so on.

Is the transition matrix in Example 1 a regular matrix?

Adjustment 1

Adjustment 1

If a user reaches a page that doesn't link to other pages, then the user will choose any page in the web, with equal probability, and move to that page.

Let's denote this modified transition matrix as P_*. Our transition matrix in Example 1 becomes:

Adjustment 2

Adjustment 2

A user at any page will navigate any page among those that their page links to with equal probability p, and to any page in the web with equal probability $1-p$. The transition matrix becomes

$$G = pP_* + (1-p)K$$

All the elements of the $n \times n$ matrix K are equal to $1/n$.

p is referred to as the **damping factor**, Google is said to use $p = 0.85$.

With adjustments 1 and 2, our the Google matrix is:

Computing Page Rank

- Because G is stochastic, for any initial probability vector $\vec{x}_0$,

$$\lim_{n\to\infty} G^n \vec{x}_0 = \vec{q}$$

- In practice we can compute the page rank for each page in the web by evaluating:

$$G^n \vec{x}_0$$

for large n. The elements of the resulting vector give the page ranks of each page in the web.

On a MATH 1554 exam,

- problems that require a calculator will not be on your exam
- you may construct your G matrix using factions instead of decimal expansions

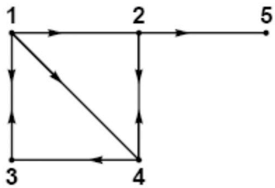
Example 2 (if time permits)

Construct the Google Matrix for the web below (your instructor would provide the web).

There is (of course) Much More to PageRank

The PageRank Algorithm currently used by Google is under constant development, and tailored to individual users.

- When PageRank was devised, in 1996, Yahoo! used humans to provide a "index for the Internet, " which was 10 million pages.
- The PageRank algorithm was produced as a competing method. The patent was awarded to Stanford University, and exclusively licensed to the newly formed Google corporation.
- Brin and Page combined the PageRank algorithm with a webcrawler to provide regular updates to the transition matrix for the web.
- The explosive growth of the web soon overwhelmed human based approaches to searching the internet.



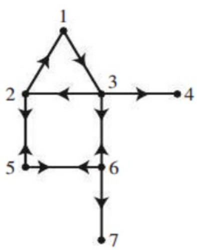


FIGURE 1

A seven-page Web.

DEFINITION A stochastic matrix P is **regular** if some power P^k contains only strictly positive entries.

THEOREM 1 If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

a. There is a stochastic matrix Π such that $\lim_{n\to\infty} P^n = \Pi$.

b. Each column of Π is the same probability vector $\mathbf{q}$.

c. For any initial probability vector $\mathbf{x}_0$, $\lim_{n\to\infty} P^n\mathbf{x}_0 = \mathbf{q}$.

d. The vector $\mathbf{q}$ is the unique probability vector which is an eigenvector of P associated with the eigenvalue 1.

e. All eigenvalues λ of P other than 1 have $|\lambda| < 1$.

10.2 Exercises

In Exercises 1 and 2, consider a Markov chain on $\{1, 2\}$ with the given transition matrix P. In each exercise, use two methods to find the probability that, in the long run, the chain is in state 1. First, raise P to a high power. Then directly compute the steady-state vector.

1. $P = \begin{bmatrix} .2 & .4 \\ .8 & .6 \end{bmatrix}$ **2.** $P = \begin{bmatrix} 1/4 & 2/3 \\ 3/4 & 1/3 \end{bmatrix}$

In Exercises 3 and 4, consider a Markov chain on $\{1, 2, 3\}$ with the given transition matrix P. In each exercise, use two methods to find the probability that, in the long run, the chain is in state 1. First, raise P to a high power. Then directly compute the steady-state vector.

3. $P = \begin{bmatrix} 1/3 & 1/4 & 0 \\ 1/3 & 1/2 & 1 \\ 1/3 & 1/4 & 0 \end{bmatrix}$ **4.** $P = \begin{bmatrix} .1 & .2 & .3 \\ .2 & .3 & .4 \\ .7 & .5 & .3 \end{bmatrix}$

In Exercises 5 and 6, find the matrix to which P^n converges as n increases.

5. $P = \begin{bmatrix} 1/4 & 2/3 \\ 3/4 & 1/3 \end{bmatrix}$ **6.** $P = \begin{bmatrix} 1/4 & 3/5 & 0 \\ 1/4 & 0 & 1/3 \\ 1/2 & 2/5 & 2/3 \end{bmatrix}$

In Exercises 7 and 8, determine whether the given matrix is regular. Explain your answer.

7. $P = \begin{bmatrix} 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \\ 1/3 & 1/2 & 0 \end{bmatrix}$

8. $P = \begin{bmatrix} 1/2 & 0 & 1/3 & 0 \\ 0 & 2/5 & 0 & 3/7 \\ 1/2 & 0 & 2/3 & 0 \\ 0 & 3/5 & 0 & 4/7 \end{bmatrix}$

9. Consider a pair of Ehrenfest urns with a total of 4 molecules divided between them.

a. Find the transition matrix for the Markov chain that models the number of molecules in urn A, and show that this matrix is not regular.

b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

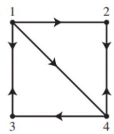
10. Consider a pair of Ehrenfest urns with a total of 5 molecules divided between them.

a. Find the transition matrix for the Markov chain that models the number of molecules in urn A, and show that this matrix is not regular.

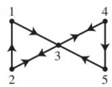
b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

In Exercises 15 and 16, consider a simple random walk on the given directed graph. In the long run, what fraction of the time will the walk be at each of the various states?

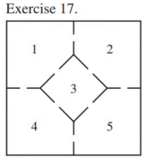
15.



16.



17. Consider the mouse in the following maze from Section 10.1, Exercise 17.



The mouse must move into a different room at each time step and is equally likely to leave the room through any of the available doorways. If you go away from the maze for a while, what is the probability that the mouse will be in room 3 when you return?

11. Consider an unbiased random walk with reflecting boundaries on $\{1, 2, 3, 4\}$.

a. Find the transition matrix for the Markov chain and show that this matrix is not regular.

b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

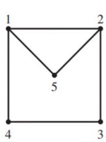
12. Consider a biased random walk with reflecting boundaries on $\{1, 2, 3, 4\}$ with probability $p = .2$ of moving to the left.

a. Find the transition matrix for the Markov chain and show that this matrix is not regular.

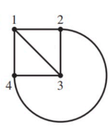
b. Assuming that the steady-state vector may be interpreted as occupation times for this Markov chain, in what state will this chain spend the most steps?

In Exercises 13 and 14, consider a simple random walk on the given graph. In the long run, what fraction of the time will the walk be at each of the various states?

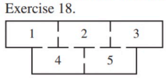
13.



14.

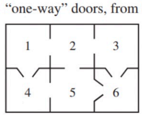


18. Consider the mouse in the following maze from Section 10.1, Exercise 18.



What fraction of the time does it spend in room 3?

19. Consider the mouse in the following maze, which includes "one-way" doors, from Section 10.1, Exercise 19.



Show that

$$\mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

is a steady-state vector for the associated Markov chain, and interpret this result in terms of the mouse's travels through the maze.