Some Applications of Linear Algebra

Shuzo Takahashi*

This document contains several applications of linear algebra. It has been used in the course "Linear Algebra" (Math 442) at University of North Dakota. Excel workbooks accompany some of the applications.

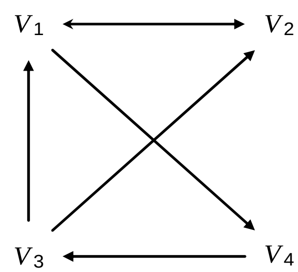
- Communication Networks and Matrix Powers — Excel Workbook
- Linear Transformations and Markov Chains I — Excel Workbook
- Linear Transformations and Markov Chains II
- The Fibonacci Sequence
- Least-Squares Solutions and Least-Squares Curves — Excel Workbook
- Affine Transformations and Fractals — Excel Workbook

* Address: Department of Mathematics, University of North Dakota, Grand Forks, ND 58202-8376. Email: shuzo.takahashi@und.nodak.edu

Communication Networks and Matrix Powers

In this note, we study how matrix powers (A^m) can help to answer certain questions regarding communication networks, such as how many ways there are to get from one station to another involving exactly one station between them, what the smallest number of stations involved in connecting one station to another is, etc.

Example Consider the following communication network with four stations $V_1, V_2, V_3,$ and V_4.



The stations can be intersections of roads, telephone switchboards, human beings, etc., and the communication links could be roads, telephone lines, human communications, etc. Some stations are linked by two-way communication (such as V_1 and V_2), some stations are linked by one-way communication (such as V_1 and V_3), and some stations may not have any direct communication (such as V_2 and V_4). ■

If the communication network in question is as small as the one described above, we can answer the above questions relatively easily by just our eyes. But in practice, communication networks can be huge, and it is impractical to obtain information about them just by looking at them. In this note, we learn how we represent communication networks by matrices and how we can get information about communication networks by matrix powers.

A communication network is mathematically described by a diagraph.

Definition A *diagraph* is a finite collection of *vertices* $V_1, V_2, \cdots, V_n$ together with *directed arcs* joining certain pairs of vertices. A *path* from the vertex V_i to the vertex V_j is a sequence of arcs that connects V_i to V_j. The *length* of a path is its number of arcs. A path of length n is called an *n-path*. The *distance* from the vertex V_i to the vertex V_j is the length of a shortest path from the vertex V_i to the vertex V_j if there is such a path. If there is no path from the vertex V_i to the vertex V_j, the distance is undefined. ■

With this new terminology, our questions above are phrased as how many 2-paths there are from one vertex to another, what the distance from one vertex to another is, etc.

Example (continued) The communication network described above is an example of a diagraph. Stations are vertices and communication links are arcs. For example, $V_1 \to V_2$, $V_2 \to V_1$, $V_3 \to V_1$, etc. are arcs. There is no arc between V_2 and V_4. A sequence of arcs, $V_2 \to V_1 \to V_4$, is a path from V_2 to V_4, and the length of this path is 2. The shortest path from V_3 to V_2 is $V_3 \to V_2$, and hence the distance from V_3 to V_2 is 1. On the other hand, the shortest path from V_2 to V_3 is $V_2 \to V_1 \to V_4 \to V_3$ and hence the distance from V_2 to V_3 is 3. ■

A diagraph is represented by an adjacency matrix.

Definition Consider a diagraph with vertices $V_1, V_2, \cdots, V_n$. The adjacency matrix A of the diagraph is an $n \times n$ matrix whose element A_{ij} in the i^{th} row and j^{th} column is defined by

$$A_{ij} = \begin{cases} 1 & \text{if there is an arc from vertex } V_i \text{ to } V_j \\ 0 & \text{if there is no arc from vertex } V_i \text{ to } V_j \end{cases}$$ ■

Example (continued) The adjacency matrix of the above communication network is

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

For example, $A_{12} = 1$ because there is an arc from V_1 to V_2, and $A_{24} = 0$ because there is no arc from V_2 to V_4. ■

The adjacency matrix describes the network completely. The following theorem answers the questions which we would like to ask.

Theorem Suppose that A is the adjacency matrix of a diagraph with vertices $V_1, V_2, \cdots, V_n$. Let $A_{ij}^{(m)}$ be the element in the i^{th} row and j^{th} column of A^m. Then, we have

$$\text{the number of } m\text{-paths from } V_i \text{ to } V_j = A_{ij}^{(m)}$$

Proof The proof is by mathematical induction on m.

Base Step $m = 2$

Fix i and j. Let k be any number from $1,\ldots,n$. Then, A_{ik} is the number of arcs from V_i to V_k and A_{kj} is the number of arcs from V_k to V_j. Thus, $A_{ik}A_{kj}$ is the number of 2-paths from V_i to V_j, passing through V_k. Thus, the total number of 2-paths from V_i to V_j is $A_{i1}A_{1j} + A_{i2}A_{2j} + \cdots + A_{in}A_{nj}$, which is $A_{ij}^{(2)}$.

Inductive Step Suppose that the statement is true for *m*, i.e., that for all *i* and *j*, the number of *m*-paths from V_i to V_j = $A_{ij}^{(m)}$. Interpreting an (*m* + 1)-path as an *m*-path followed by an arc, we have that $A_{ik}^{(m)}A_{kj}$ is the number of (*m* + 1)-paths from V_i to V_j passing through V_k right before the last vertex V_j. Hence, the total number of (*m* + 1)-paths from V_i to V_j is $A_{i1}^{(m)}A_{1j} + A_{i2}^{(m)}A_{2j} + \cdots + A_{in}^{(m)}A_{nj}$. This last number is the element in the *i*th row and *j*th column of $A^m A$ and hence it is equal to $A_{ij}^{(m+1)}$. ■

Remark In the above proof, we could have started with $m = 1$ as our base step. Then, the statement in the base step is nothing but the definition of $A_{ij}^{(1)} = A_{ij}$. The only reason why we included the $m = 2$ case is to illustrate the basic idea of the proof with the second simplest case. ■

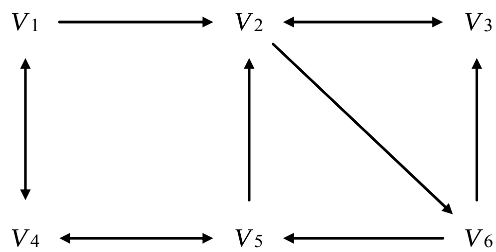
Let's analyze the above example using this theorem.

Example (continued) First, compute $A, A^2, A^3, A^4,$ etc.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad A^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad A^3 = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad A^4 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

For example, we see that $A_{31}^{(4)} = 2$. This means that there are two different 4-paths from V_3 to V_1. By looking at the graph, we can tell that the two 4-paths are $V_3 \to V_1 \to V_4 \to V_3 \to V_1$ and $V_3 \to V_2 \to V_1 \to V_2 \to V_1$. (It is not known how to obtain this information from matrix powers $A, A^2, A^3, A^4,$ etc. There are algorithms that determine such information.) Also, we observe that $A_{23}^{(1)} = A_{23} = 0, A_{23}^{(2)} = 0, A_{23}^{(3)} = 1$. This shows that there is no 1-path (arc) or 2-path from V_2 to V_3, and there is one 3-path from V_2 to V_3. Hence, this 3-path connects V_2 to V_3 for the first time. Thus, the distance from V_2 to V_3 is 3. ■

Exercise A communication network is described as follows:



(a) Define the adjacency matrix A that represents the communication network graph.
(b) Calculate A^2, A^3, A^4, A^5. Write down the results.
(c) Using the results from (b), determine the distance from V_1 to V_3.
(d) Using the results from (b), determine the distance from V_3 to V_1.
(e) Using the results from (b), determine the number of 4-paths from V_5 to V_3.
(f) Using the above graph, describe all 4-paths from V_5 to V_3.
(g) Are there any 4-paths from V_3 to V_5?

Linear Transformations and Markov Chains I

In this note, through an example, we study how linear transformations can be used to analyze population movement between various locations. Locations can be cities and surrounding suburbs, various restaurants, or various rooms in one building.

Let's consider the following situation. In a small town, there are two competing restaurants. Suppose that there are 1000 customers in total and each person eats once every week in one of the two restaurants. 30% of the customers who ate in Restaurant *X* change to Restaurant *Y* the next week, and 20% of the customers who ate in Restaurant *Y* change to Restaurant *X*. We would like to predict the long-term distributions of customers in the two restaurants. More specifically, we may ask the following questions:

(a) If we start with the situation that 800 customers ate in Restaurant *X* and 200 customers ate in Restaurant *Y*, what would be the long-term distributions in the two restaurants.
(b) If we start with the other way around, namely, with the situation that 200 customers ate in Restaurant *X* and 800 customers ate in Restaurant *Y*, is the outcome different?
(c) Does the long-term distributions of customers depend on the initial situation?

Let $s_0 = (x_0, y_0)$ be the pair of initial numbers of customers in Restaurant *X* and Restaurant *Y*, and let $s_n = (x_n, y_n)$ be the pair of numbers of customers in Restaurant *X* and Restaurant *Y* n weeks later. Suppose that we start with the situation that 800 customers ate in Restaurant *X* and 200 customers ate in Restaurant *Y* initially, i.e., $s_0 = (x_0, y_0) = (800, 200)$. The next week, since 70% of 800 customers stay in Restaurant *X* and 20% of 200 customers move from Restaurant *Y* to Restaurant *X*, we have that $x_1 = (0.7)(800) + (0.2)(200) = 600$. Similarly, we have $y_1 = (0.3)(800) + (0.8)(200) = 400$. Thus, $s_1 = (600, 400)$. Generally, once we know $s_n = (x_n, y_n)$, then we have

$$s_{n+1} = (x_{n+1}, y_{n+1}) = (0.7x_n + 0.2y_n, 0.3x_n + 0.8y_n).$$

The pair s_n can be considered as a vector in $\mathbb{R}^2$ and a transition from s_n to s_{n+1} can be described by a linear transformation. Namely, define the linear transformation

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \text{ by } T(x, y) = (0.7x + 0.2y, 0.3x + 0.8y).$$

Then, we have,

$$s_{n+1} = Ts_n.$$

Starting with s_0 and applying T consecutively, we obtain a sequence of vectors:

$$s_0 \xrightarrow{T} s_1 \xrightarrow{T} s_2 \xrightarrow{T} \cdots\cdots \xrightarrow{T} s_n \xrightarrow{T} s_{n+1} \xrightarrow{T} \cdots\cdots.$$

We have

$$s_1 = Ts_0, s_2 = Ts_1 = T(Ts_0) = T^2 s_0, \cdots, \text{ and } s_n = T^n s_0.$$

Thus, our questions are formulated as follows:

(a) What is $\lim_{n\to\infty} s_n = \lim_{n\to\infty} T^n s_0$ when $s_0 = (800, 200)$?

(b) What is $\lim_{n\to\infty} s_n = \lim_{n\to\infty} T^n s_0$ when $s_0 = (200, 800)$?

(c) Does $\lim_{n\to\infty} s_n = \lim_{n\to\infty} T^n s_0$ (if it exists) depend on what we choose for s_0?

Later in the course, we will study how to find such limits precisely. But for now, let's calculate $s_n = T^n s_0$ for $n = 0, 1, 2, 3, \cdots$, and find out what will happen to the sequence $s_1, s_2, \cdots, s_n \cdots$. To do such computation, it is more convenient to represent the linear transformation T by a matrix. The transformation $s_{n+1} = Ts_n$, i.e.,

$$(x_{n+1}, y_{n+1}) = (0.7x_n + 0.2y_n, 0.3x_n + 0.8y_n),$$

can be written as

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}.$$

Thus, $s_n = T^n s_0$ is equivalent to

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

We obtain the data below when we start with $s_0 = (800, 200)$ (for example, using the worksheet "Restaurants" in the Excel workbook "Linear Transformation" accompanying this note). According to the data, the sequence approaches $(400, 600)$ after 9 weeks. Thus, it appears to be that eventually, in each week, 400 and 600 customers eat in Restaurants *X* and *Y*, respectively.

n	x_n	y_n
0	200	800
1	300	700
2	350	650
3	375	625
4	388	613
5	394	606
6	397	603
7	398	602
8	399	601
9	400	600
10	400	600
11	400	600
12	400	600
13	400	600
14	400	600
15	400	600

Remarks

1. The sequence

$$s_0 \xrightarrow{T} s_1 \xrightarrow{T} s_2 \xrightarrow{T} \cdots\cdots \xrightarrow{T} s_n \xrightarrow{T} s_{n+1} \xrightarrow{T} \cdots\cdots$$

   described above can be viewed as a Markov chain. But, since our interest is mainly in the use of linear transformations, we omitted the description of the sequence as a Markov chain to avoid introducing unnecessary terminology. Students who are interested in Linear Algebra and Markov chains are advised to read other books such as Linear Algebra with applications by Gareth Williams.

2. The matrix $\begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}$ is called a stochastic matrix. A stochastic matrix in general is a matrix such that each entry represents a probability and the sum of entries of each column is equal to 1.

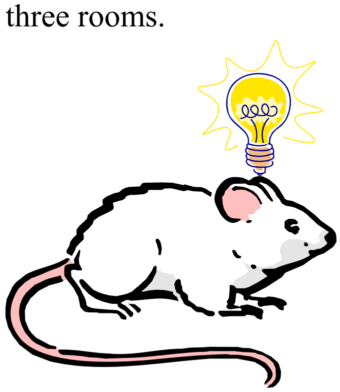
Exercise 1 Finish Parts (b) and (c) of the above example. For (c), choose your own several (four or more) different values for the initial pair $s_0 = (x_0, y_0)$. Provide your data to support your conclusion. (I expect different students to use different values!!!!!)

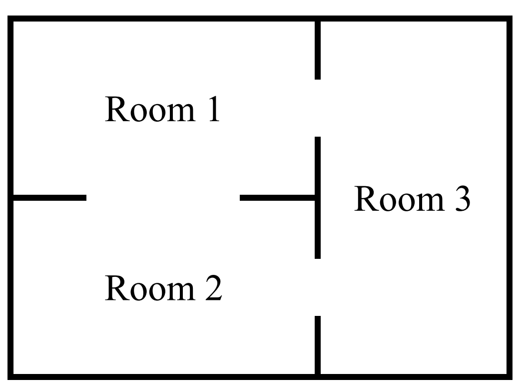
Exercise 2 There are 24 rats randomly distributed in three rooms. Suppose that every minute each rat moves to another room in accordance with the following probabilities:

- The probability of a rat moving from Room 1 to Room 2 is 2/3, and the probability of a rat moving from Room 1 to Room 3 is 1/3.
- The probability of a rat moving from Room 2 to Room 1 is 2/3, and the probability of a rat moving from Room 2 to Room 3 is 1/3.
- The probability of a rat moving from Room 3 to Room 1 is 1/2, and the probability of a rat moving from Room 3 to Room 2 is 1/2.

Answer the following questions:

(1) Write the stochastic matrix that describes the above probability information.
(2) Similarly to Part (c) of Exercise 1, predict the long-term distributions of rats in the three rooms.





Exercise 3 In the year 2000, the number of people living in cities in the United States was 58 million and the number of people living in the surrounding suburbs was 142 million. Suppose that the population flow from cities to suburbs each year after 2000 is described in terms of probabilities as follows:

- The probability of a person staying in the city is 0.96.
- The probability of a person moving to the suburbs is 0.04.
- The probability of a person moving to the city is 0.01.
- The probability of a person staying in the suburbs is 0.99.

Predict the long-term distributions of populations. (We are assuming that the total population living in cities and the surrounding suburbs is unchanged after 2000.)

Linear Transformations and Markov Chains II

In Exercise 1 in the document “Linear Transformations and Markov Chains I,” we observed that the long-term distributions of customers in the two restaurants do not depend on the initial distributions of customers. More precisely, we observed that $\lim_{n\to\infty} T^n(x_0, y_0)$ is always equal to $(400, 600)$ if $x_0 + y_0 = 1000$. That is, the limit does not depend on what we choose for x_0 and y_0. Now, we would like to prove this observation, using the diagonalization technique. Equivalently, we prove that

$$\lim_{n\to\infty} A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 400 \\ 600 \end{bmatrix} \text{ where } A = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \text{ and } x_0 + y_0 = 1000.$$

More generally, we will show that

$$\lim_{n\to\infty} A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = (x_0 + y_0) \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}.$$

(1) Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x, y) = \left(0.7x + 0.2y, 0.3x + 0.8y\right)$. Our goal is to evaluate $\lim_{n\to\infty} T^n(x_0, y_0)$.

(2) The matrix of T relative to the standard basis $\mathfrak{B} = \{(1,0), (0,1)\}$ is $A \overset{def}{=} [T]_{\mathfrak{B}} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}$. So, our goal is to evaluate $\lim_{n\to\infty} A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$.

(3) The characteristic polynomial of T is given by

$$f(x) = |xI - A| = \begin{vmatrix} x-0.7 & -0.2 \\ -0.3 & x-0.8 \end{vmatrix} = \begin{vmatrix} x-1 & x-1 \\ -0.3 & x-0.8 \end{vmatrix} = (x-1)\begin{vmatrix} 1 & 1 \\ -0.3 & x-0.8 \end{vmatrix}$$
$$= (x-1)(x-0.5).$$

Thus, the characteristic values are 1 and 0.5.

(4) We have to determine the characteristic spaces $V_c = N_{T-cI} = N_{A-cI}$ for each $c = 1$ and 0.5.

For $c = 1$, we have

$$(x, y) \in V_1 \Leftrightarrow \left(A - 1\cdot I\right)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -0.3 & 0.2 \\ 0.3 & -0.2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\Leftrightarrow -0.3x + 0.2y = 0 \Leftrightarrow x = (2/3)y.$$

Hence, we have

$$V_1 = \{((2/3)y, y) \mid y \text{ is any real number.}\} = \{(2/3,1)y \mid y \text{ is any real number.}\}$$
$$= \{(2,3)y \mid y \text{ is any real number.}\}.$$

So, V_1 is spanned by $\beta_1 \overset{def}{=} (2,3)$.

For $c = 0.5$, we have

$$(x,y) \in V_{0.5} \Leftrightarrow (A - 0.5 \cdot I)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0.2 & 0.2 \\ 0.3 & 0.3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow y = -x.$$

Hence, we have

$$V_{0.5} = \{(x,-x) \mid x \text{ is any real number.}\} = \{(1,-1)x \mid x \text{ is any real number.}\}.$$

So, $V_{0.5}$ is spanned by $\beta_2 \overset{def}{=} (1,-1)$.

(5) The matrix of T relative to the basis $\mathfrak{B}' = \{\beta_1, \beta_2\}$ is $B \overset{def}{=} [T]_{\mathfrak{B}'} = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$

because we have $\begin{cases} T\beta_1 = 1 \cdot \beta_1 + 0 \cdot \beta_2 \\ T\beta_2 = 0 \cdot \beta_1 + 0.5 \cdot \beta_2 \end{cases}$

(Note that the vectors β_1 and β_2 are linearly independent. So, $\mathfrak{B}'$ is a basis.)

(6) The matrix P such that $B = P^{-1}AP$ and $A = PBP^{-1}$ is given by

$P = \left[[\beta_1]_{\mathfrak{B}}, [\beta_2]_{\mathfrak{B}}\right] = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$. Moreover, its inverse is $P^{-1} = \frac{1}{5}\begin{bmatrix} 1 & 1 \\ 3 & -2 \end{bmatrix}$.

(7) Finally, we can evaluate $\lim_{n\to\infty} A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$.

$$\lim_{n\to\infty} A^n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \lim_{n\to\infty} PB^nP^{-1}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \lim_{n\to\infty} \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 1^n & 0 \\ 0 & 0.5^n \end{bmatrix}\left(\frac{1}{5}\right)\begin{bmatrix} 1 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$
$$= \left(\frac{1}{5}\right)\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \left(\frac{1}{5}\right)\begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$
$$= (x_0 + y_0)\begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}.$$

Exercise Using the diagonalization technique, explain what you observed in Exercise 2 in the document "Linear Transformations and Markov Chains I."
That is, prove that

$$\lim_{n\to\infty} A^n s_0 = \begin{bmatrix} 9 \\ 9 \\ 6 \end{bmatrix} \text{ where } A = \begin{bmatrix} 0 & 2/3 & 1/2 \\ 2/3 & 0 & 1/2 \\ 1/3 & 1/3 & 0 \end{bmatrix},\ s_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, \text{ and } x_0 + y_0 + z_0 = 24\,.$$

More generally, prove that

$$\lim_{n\to\infty} A^n s_0 = (x_0 + y_0 + z_0) \begin{bmatrix} 3/8 \\ 3/8 \\ 1/4 \end{bmatrix} \text{ where } A = \begin{bmatrix} 0 & 2/3 & 1/2 \\ 2/3 & 0 & 1/2 \\ 1/3 & 1/3 & 0 \end{bmatrix} \text{ and } s_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}.$$

The Fibonacci Sequence

The Fibonacci sequence is the sequence of numbers a_n given by the rules $a_0 = 1, a_1 = 1,$ and $a_{n+2} = a_{n+1} + a_n$. Using the diagonalization technique, we will prove Binet's Formula for the Fibonacci sequence.

Binet's Formula

$$a_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$$

Proof

(1) Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x, y) = (y, x + y)$. We have $T(a_n, a_{n+1}) = (a_{n+1}, a_n + a_{n+1}) = (a_{n+1}, a_{n+2})$. Thus, $(a_n, a_{n+1}) = T^n(a_0, a_1) = T^n(1,1)$. Hence, our goal is to find a formula for $T^n(1,1)$.

(2) The matrix of T relative to the standard basis $\mathfrak{B} = \{(1,0),(0,1)\}$ is $A \overset{def}{=} [T]_{\mathfrak{B}} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. So, our goal is to find a formula for $\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(3) The characteristic polynomial of T is given by $f(x) \overset{def}{=} |xI - A| = \begin{vmatrix} x & -1 \\ -1 & x-1 \end{vmatrix} = x(x-1) - (-1)(-1) = x^2 - x - 1$.

Thus, the characteristic values are $c_1 \overset{def}{=} \frac{1+\sqrt{5}}{2}$ and $c_2 \overset{def}{=} \frac{1-\sqrt{5}}{2}$.

(4) We have to determine the characteristic spaces $V_c = N_{T-cI} = N_{A-cI}$ for each $c = c_1$ and c_2. Note that $c_1 c_2 = -1$ and $c_1 + c_2 = 1$.

For $c = c_1$, we have

$$(x, y) \in V_{c_1} \Leftrightarrow (A - c_1 \cdot I)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -c_1 & 1 \\ 1 & 1-c_1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} -c_1 & 1 \\ 1 & c_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -c_1 & 1 \\ -c_1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow y = c_1 x.$$

Hence, we have

$V_{c_1} = \{(x, c_1 x) \mid x \text{ is any real number.}\} = \{(1, c_1)x \mid x \text{ is any real number.}\}$.

So, V_{c_1} is spanned by $\beta_1 \overset{def}{=} (1, c_1)$.

For $c = c_2$, similarly, V_{c_2} is spanned by $\beta_2 \overset{def}{=} (1, c_2)$.

(5) The matrix of T relative to the basis $\mathfrak{B}' = \{\beta_1, \beta_2\}$ is $B \overset{def}{=} [T]_{\mathfrak{B}'} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}$.

(Note that the vectors β_1 and β_2 are linearly independent. So, $\mathfrak{B}'$ is a basis.)

(6) The matrix P such that $B = P^{-1}AP$ and $A = PBP^{-1}$ is given by $P = \begin{bmatrix} 1 & 1 \\ c_1 & c_2 \end{bmatrix}$. Moreover, its inverse is $P^{-1} = \dfrac{1}{\sqrt{5}} \begin{bmatrix} -c_2 & 1 \\ c_1 & -1 \end{bmatrix}$.

(7) Finally, we can find a formula for $\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} = PB^nP^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} c_1^{\,n} & 0 \\ 0 & c_2^{\,n} \end{bmatrix} \left(\frac{1}{\sqrt{5}} \right) \begin{bmatrix} -c_2 & 1 \\ c_1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \left(\frac{1}{\sqrt{5}} \right) \begin{bmatrix} 1 & 1 \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} c_1^{\,n} & 0 \\ 0 & c_2^{\,n} \end{bmatrix} \begin{bmatrix} c_1 \\ -c_2 \end{bmatrix} = \left(\frac{1}{\sqrt{5}} \right) \begin{bmatrix} c_1^{\,n+1} - c_2^{\,n+1} \\ c_1^{\,n+2} - c_2^{\,n+2} \end{bmatrix}.$$

Thus, we have Binet's Formula

$$a_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right].$$

Exercise Consider the sequence of numbers a_n given by the rules $a_0 = 1, a_1 = 1,$ and $a_{n+2} = a_{n+1} + 6a_n$. Find a formula for a_n, similar to Binet's Formula for the Fibonacci sequence.

Least-Squares Solutions and Least-Squares Curves

Consider three points $(1,0),(2,2),$ and $(4,5)$ on the *x*-*y* plane.
If the line $y = a + bx$ passes through the three points, then *a* and *b* must satisfy the following system of linear equations:

$$\begin{cases} a+b=0 \\ a+2b=2 \\ a+4b=5 \end{cases} \text{ or equivalently } \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}.$$

We can see that the system has no solution. Thus, there is no straight line passing through the three points. Our question is to find the line that best fits the three points in a certain sense. In the real world, such a question is important in the situation where the data for two variables are linearly associated.

A least-squares solution for the above system of linear equations is a solution (a,b) such that the distance between the following two vectors is the least:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ a+2b \\ a+4b \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}.$$

The distance is $\sqrt{(a+b-0)^2+(a+2b-2)^2+(a+4b-5)^2}$. (This is where the name "least-squares solution" comes from.) For such a solution (a,b) (if it exists), the line $y = a + bx$ is called the least-squares line for the three points $(1,0),(2,2),$ and $(4,5)$. We consider the least-squares line as the line that best fits the three points. Our question is to find a least-squares solution if it exists.

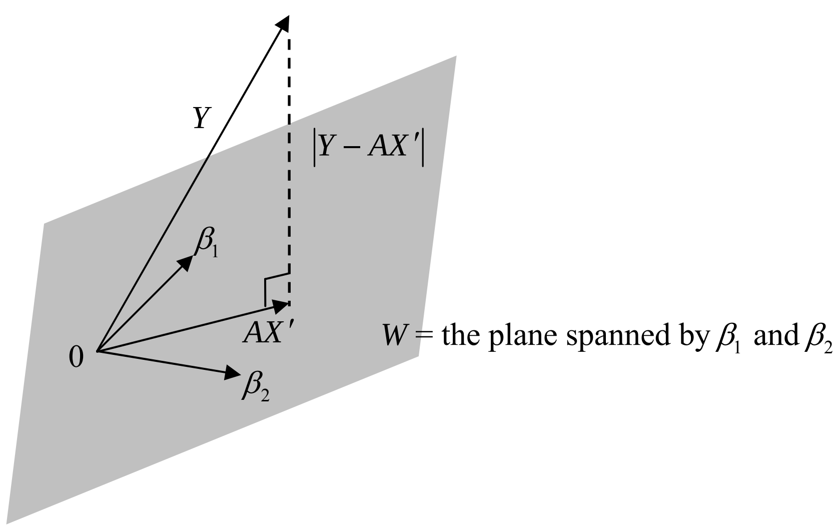
Let's generalize this question. Let A be an $m \times n$ matrix with entries in $\mathbb{R}$. Given a vector Y in $\mathbb{R}^m$, consider the system of linear equations $AX = Y$. This system has a solution if and only if the vector Y is in the range of the matrix A (considering A as a linear transformation). Thus, if Y is not in the range of A, the system does not have any solution. But, we would like to ask whether or not there is a vector X' in $\mathbb{R}^n$ such that the distance between AX' and Y is the smallest. If such a vector exists, it is called a least-squares solution for $AX = Y$. The above example is the case where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \text{ and } Y = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}.$$

Given an $m \times n$ matrix A, consider the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $TX = AX$. The rank of A is defined as the rank of T. Recall that the rank of T is the dimension of the range of T, i.e., $\operatorname{rank}(T)=\dim R_T$. Let $\beta_1,\cdots,\beta_n$ be the column vectors of A; so $A=[\beta_1,\cdots,\beta_n]$. Then, the range of T is the vector space spanned by $\{\beta_1,\cdots,\beta_n\}$. Thus, the condition that the rank of A is n is equivalent to the condition that $\{\beta_1,\cdots,\beta_n\}$ is linearly independent. We have the following theorem:

Theorem Let A be an $m\times n$ matrix with entries in $\mathbb{R}$. Given a vector Y in $\mathbb{R}^m$, consider the system of linear equations $AX=Y$. If the rank of A is n, then the matrix A^tA is invertible and there is a unique least-squares solution for $AX=Y$. Moreover, the least-squares solution is given by $\left(\left(A^tA\right)^{-1}A^t\right)Y$. Note: $\left(A^tA\right)^{-1}A^t$ is called the pseudoinverse of A and it is denoted by $\operatorname{pinv}(A)$.

Proof The invertibility of A^tA will be shown in the lemma below. For the existence and description of the least-squares solution, we consider the case where $m=3$ and $n=2$ only for the visual sake (the proof of the general case is essentially the same as the proof of this special case). We use some ideas from Calc III such as the length of a vector, the dot product, and orthogonal projection. Let β_1 and β_2 be the two column vectors of A. They are vectors in $\mathbb{R}^3$. Then, let W be the range of A, i.e., the vector space spanned by β_1 and β_2. The space W is the plane containing β_1 and β_2. We are looking for a vector AX' in W such that the distance between Y and AX' is the smallest. The distance between Y and AX' is the length $|Y-AX'|$. Thus, we are looking for a vector AX' such that AX' is equal to the orthogonal projection of Y on W. Thus, $Y-AX'$ must be orthogonal to both β_1 and β_2, that is, $\beta_1\cdot(Y-AX')=0$ and $\beta_2\cdot(Y-AX')=0$. (The operation $\cdot$ is the dot product on $\mathbb{R}^3$.)



The last two conditions together are expressed as a single equation $A^t(Y-AX')=0$, which is equivalent to $A^tAX'=A^tY$. Thus, we have $X'=\left(\left(A^tA\right)^{-1}A^t\right)Y$. ■

Lemma Let A be an $m \times n$ matrix with entries in $\mathbb{R}$. If the rank of A is n, then the $n \times n$ matrix $A^t A$ is invertible.

Proof First, considering A as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, we have $\operatorname{rank}(A) + \operatorname{nullity}(A) = n$. Since $\operatorname{rank}(A) = n$, we have $\operatorname{nullity}(A) = 0$. Thus, A is non-singular, i.e., $AX = 0 \Rightarrow X = 0$. Next, consider $A^t A$ as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. We have that $(AX)\cdot(AX) = (AX)^t(AX) = (X^t A^t)(AX) = X^t\left(\left(A^t A\right)X\right)$. So, $\left(A^t A\right)X = 0 \Rightarrow (AX)\cdot(AX) = 0 \Rightarrow AX = 0 \Rightarrow X = 0$. That is, $A^t A$ is non-singular and hence it is invertible since both the domain and range of $A^t A$ have the same dimension. ■

Example 1 Let's go back to the question raised at the beginning of this note, namely, the question of finding the least-squares line $y = a + bx$ for three points $(1,0), (2,2),$ and $(4,5)$. We have

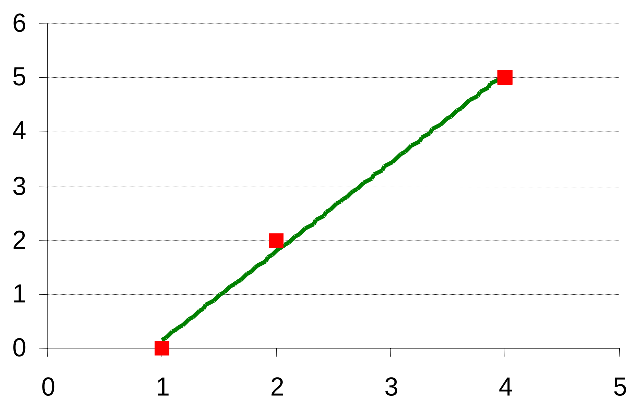
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \text{ and } Y = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}.$$

First, check that the two column vectors of the matrix A are linearly independent. Thus, the rank of A is 2. Then, by the theorem, there is a unique least-squares solution for $AX = Y$ given by $\left(\left(A^t A\right)^{-1} A^t\right)Y$. We have

$$A^t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix},\ A^t A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 7 & 21 \end{bmatrix},\ \left(A^t A\right)^{-1} = \begin{bmatrix} 3 & 7 \\ 7 & 21 \end{bmatrix}^{-1} = \frac{1}{14}\begin{bmatrix} 21 & -7 \\ -7 & 3 \end{bmatrix},$$

and
$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(\left(A^t A\right)^{-1} A^t\right)Y = \left(\frac{1}{14}\begin{bmatrix} 21 & -7 \\ -7 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}\right)\begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} -3/2 \\ 23/14 \end{bmatrix} \approx \begin{bmatrix} -1.5 \\ 1.64 \end{bmatrix}.$$

Thus, the least-squares line for three points (1,0), (2,2), and (4,5) is $y = -1.5 + 1.64x$. The graph of the three points together with the line is given as:



Example 2 We can use the above theorem to find a parabola $y = a + bx + cx^2$ that best fits four points $(1.1, 3.1), (2.2, 3.7), (3.4, 5.1),$ and $(3.9, 6.3)$. If the parabola $y = a + bx + cx^2$ passes through the four points, then *a*, *b*, and *c* must satisfy the following system of linear equations:

$$\begin{cases} a + 1.1b + 1.1^2 c = 3.1 \\ a + 2.2b + 2.2^2 c = 3.7 \\ a + 3.4b + 3.4^2 c = 5.1 \\ a + 3.9b + 3.9^2 c = 6.3 \end{cases} \text{ or equivalently } \begin{bmatrix} 1 & 1.1 & 1.1^2 \\ 1 & 2.2 & 2.2^2 \\ 1 & 3.4 & 3.4^2 \\ 1 & 3.9 & 3.9^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3.1 \\ 3.7 \\ 5.1 \\ 6.3 \end{bmatrix}.$$

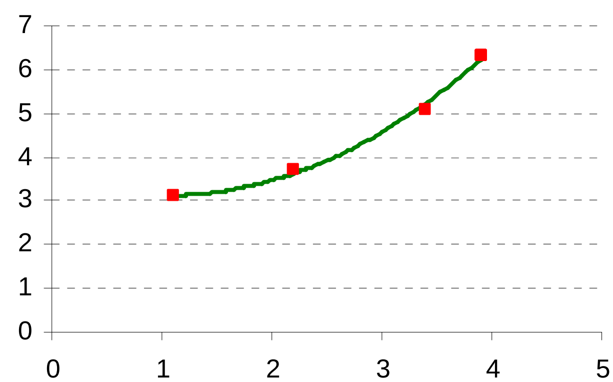
We can see that the system has no solution. Define A and Y by

$$A = \begin{bmatrix} 1 & 1.1 & 1.1^2 \\ 1 & 2.2 & 2.2^2 \\ 1 & 3.4 & 3.4^2 \\ 1 & 3.9 & 3.9^2 \end{bmatrix} \text{ and } Y = \begin{bmatrix} 3.1 \\ 3.7 \\ 5.1 \\ 6.3 \end{bmatrix}.$$

We need to find the least-squares solution $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ for $AX = Y$. By the theorem, we have

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \left(\left(A^t A \right)^{-1} A^t \right) Y \approx \begin{bmatrix} 3.56 \\ -0.82 \\ 0.39 \end{bmatrix}.$$

Thus, the least-squares parabola for four points $(1.1, 3.1), (2.2, 3.7), (3.4, 5.1),$ and $(3.9, 6.3)$ is $y = 3.56 - 0.82x + 0.39x^2$. The graph of the four points together with the parabola is given as:



Exercise 1 Using a pencil and paper, determine the least-squares line for three points $(-1,3)$, $(0,0)$, and $(2,5)$. Show all your work.

Exercise 2 Archaeopteryx is an extinct species that is between birds and dinosaurs. The following are the data[†] for the lengths of the femur and the humerus for five fossil specimens:

Femur (cm)	38	56	59	64	74
Humerus (cm)	41	63	70	72	84

(a) Determine the least-squares line with femur length as the *x*-variable and humerus length as the *y*-variable. In this exercise, you may use computer software (such as the Excel workbook accompanying this note).

(b) Suppose that you found a new specimen of archaeopteryx ☺, but you realized that you only found its femur with length 42 cm but that its humerus was missing ☹. What would be your educated guess for the length of its humerus?

Exercise 3 Determine the least-squares parabola for the following data points: (1,5), (2,2), (3,3), (4,8).

Exercise 4 The following weights (in ounces) are suspended from a spring and the following lengths (in inches) of the spring are obtained. Predict the length of the spring when a weight of 7 ounces is suspended.

Force (ounces)	1	2	3	4
Length (inches)	2.8	4.1	5.2	6.1

[†] These data are from M. A. Houck, et al., "Allometric scaling in the earliest fossil bird, Archaeopteryx lithographica," Science, 247 (1990), pp. 195-198.

Affine Transformations and Fractals

A function $T:\mathbb{R}^2 \to \mathbb{R}^2$ defined as follows is called an affine transformation:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}.$$

(Note: if $e = f = 0$, then $T:\mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation.)

Consider a set S of m affine transformations $T_i:\mathbb{R}^2 \to \mathbb{R}^2$ ($1 \le i \le m$). Suppose that a probability p_i is associated with each affine transformation $T_i:\mathbb{R}^2 \to \mathbb{R}^2$. We have $\sum_{i=1}^{m} p_i = 1$. Randomly choosing n affine transformations from the set S according to associated probabilities, we obtain a sequence of n affine transformations

$$T^{(1)}, T^{(2)}, \cdots, T^{(n-1)}, T^{(n)}.$$

On average, $100p_i\%$ of $T^{(1)}, T^{(2)}, \cdots, T^{(n-1)}, T^{(n)}$ are T_i for each i.

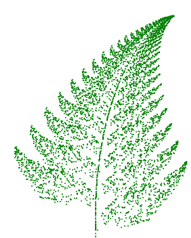
Start with the point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and apply $T^{(1)}, T^{(2)}, \cdots, T^{(n-1)}, T^{(n)}$ consecutively. Then, $n+1$ points $P_0, P_1, P_2, \cdots, P_n$ are generated as follows:

$$P_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, P_1 = T^{(1)}(P_0), P_2 = T^{(2)}(P_1), \cdots, P_n = T^{(n)}(P_{n-1}).$$

A plot of the $n+1$ points produces an interesting geometrical object called a fractal.

For example, consider the following four affine transformations $T_i:\mathbb{R}^2 \to \mathbb{R}^2$ with associated probabilities p_i :

T_i	a	b	c	d	e	f	p_i
1	0.86	0.03	-0.03	0.86	0	1.5	0.83
2	0.2	-0.25	0.21	0.23	0	1.5	0.08
3	-0.15	0.27	0.25	0.26	0	0.45	0.08
4	0	0	0	0.17	0	0	0.01



A plot of 5001 points (generated as explained above) produces a Black Spleenwort fern.

Exercise Regarding the above example, answer the following questions:

(a) Investigate the effects of varying the parameters of the four affine transformations.
(b) Determine the role of each of the four transformations.