MINERVA®

Final Problem Set

Yu-Ang Chang & Taha Bouhoun

CS111B ▯ Spring 2019

Assignment V[1]

Group[2] Final Problem Set

1. The least of your problems(#vectors, #linearsystems, #computationaltools)

In this problem, you will derive and then apply the least squares approximation to a system of linear equations. To this end, suppose that you want to solve the linear system

$$Ax = b$$

but the system is overdetermined, i.e., it has no solution. This often happens when you have more equations than you have unknowns. Your goal is to find a vector $\boldsymbol{x}$ that is as close as possible to being a solution to $Ax = b$. The approach we take avoids the use of calculus and is the approach often used with large systems.

a. Consider the system $Ax = b$ with:

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

i. Why is this system inconsistent?

---

[1] #professionalism and #organization: organizes the paper in a readable and clean style to communicate effectively with the audience.

[2] #Differences: Our team was relying on different approaches at tackling problems, I tend to use Mathematical tools and Yu-Ang was relying on the abstract part of the problems. I think this made us balance the effort needed for this Final Set and take advantage of our differences to finish the assignment (e.g., Yu-Ang is brilliant at recalling theories, and my Mathematica skills helps me to confirm and check the reasoning)

Suppose that we let $\boldsymbol{x}$ be a vector with components (x, y), and from the system A $\boldsymbol{x}$ = b, we can see that:

$$\begin{cases} x + 2y = 4 \\ x + 3y = 5 \end{cases}$$

$$0 \cdot x + 0 \cdot y = 6$$

Because the third equation has no solutions, the system is inconsistent, meaning that a solution in R doesn't exist for the system A $\boldsymbol{x}$ = b.

ii. Are the columns of A independent? Why or why not?

Yes, the columns of the matrix A are independent. A set of vectors are independent when, given vectors $v_1, v_2, .., v_n$, and scalars $c_1, c_2, .., c_n$, the only solution to the system $c_1 v_1 + c_2 v_2 + \ .., c_n v_n = 0$ is when the scalars $c_1, c_2, .. , c_n$ are all 0.

Plugging the columns of A into this concept, we see that:

$$\text{In[20]:=}\ \texttt{Solve}\left[x * \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} == 0,\ \{x, y\}\right]$$

$$\text{Out[20]=}\ \{\{x \to 0,\ y \to 0\}\}$$

where the solution set (x, y) is trivial and equals 0, so the columns are Indeed linearly independent.

We can also check linear independence by finding the reduced row echelon form of A

```
In[23]:= RowReduce[A] // MatrixForm
```

Out[23]//MatrixForm=

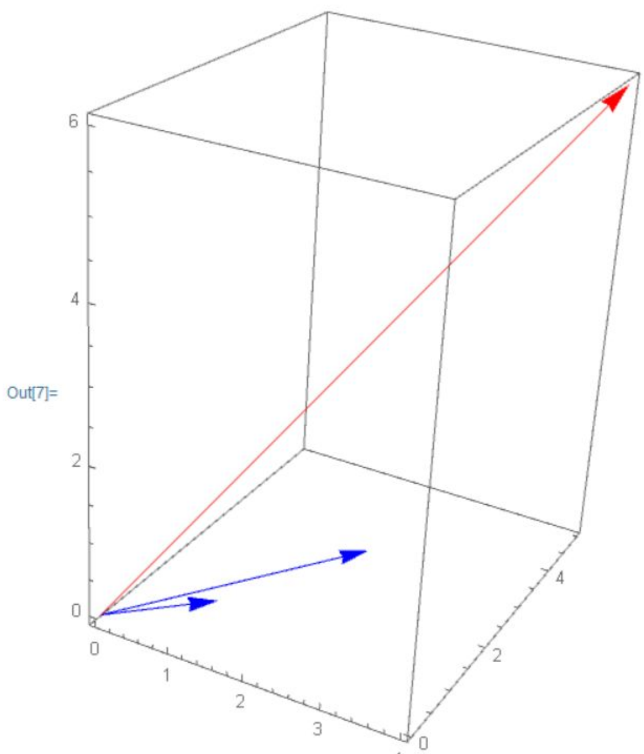
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

There are two pivot columns in the RREF, the same number of columns in A, so we know the columns of A are linearly independent, which means that we can't express one of the columns as the linear combination of another.

iii. Make a 3D plot that includes the vector b and basis vectors for the column space of A. Be sure to label the vectors, and illustrate Col(A).

The column space of A is made up of (1, 1, 0) and (2, 3, 0). Since Col(A) has two linearly independent columns, we know that the columns of A form the basis of Col(A), depicted in the diagram below with the blue vectors. Col(A) is spanned by (1, 1, 0) and (2, 3, 0), so we know that Col(A) is a plane at the bottom where the vertical axis is 0, where x and y-axis can be any value. Any vector in this plane Col(A) can be written as the linear combination of (1, 1, 0) and (2, 3, 0). Vector b is the red vector.

```
In[7]:= Graphics3D[{ Blue,
          Arrow[{{0, 0, 0}, {1, 1, 0}}], Blue,
         Arrow[{{0, 0, 0}, {2, 3, 0}}], Red,
         Arrow[{{0, 0, 0}, {4, 5, 6}}]}, Axes → True]
```



b. Another way to state the goal is that you want to find a vector $\boldsymbol{x}$ such that A $\boldsymbol{x}$ is as close as possible to b. In other words, you want to minimize

$$\| Ax - b \|$$

i. Why should A $\boldsymbol{x}$ be an element of the column space of A?

We can think of this problem in terms of matrix multiplication: A $\boldsymbol{x}$ is equal to the addition of the columns of A multiplied by each entry in the column vector from top to bottom. Hence, because the columns of A form the basis of Col(A), and A $\boldsymbol{x}$ is equal to the linear combination of the columns of A multiplied by the scalars from $\boldsymbol{x}$, so we know that A $\boldsymbol{x}$ should be in the Col(A), which is spanned by the two columns of A.

ii. Why is the description of the goal equivalent to $b - Ax$ being orthogonal to every vector in the column space of A? (In other words, A $\boldsymbol{x}$ is the projection of b onto Col(A).)[3]

Let us think about this problem geometrically: because our goal is to find the vector b that is as close to A $\boldsymbol{x}$ as possible, so it is not completely equal to A $\boldsymbol{x}$, so it can not be found in the solution set, or Col(A). Because of this inequality, vector b has to be composed of two component vectors:

- The component that is orthogonal to the solution set Col(A)
- The component that is within the solution set Col(A)

From the previous question, we know that A $\boldsymbol{x}$ is in Col(A), so if we subtract A $\boldsymbol{x}$ from b, we will subtract off the second component from b, and get the remaining component of b that is orthogonal to the solution set Col(A). Therefore, $b - Ax$, where b is close, but not identical to A $\boldsymbol{x}$, should be the remaining component of b that is orthogonal to the column space of A.

[3] #deduction: using geometric interpretation to deductively reason until the desired results emerge.

iii. Justify that the above is equivalent to saying:

$$A^T(b - Ax) = \overrightarrow{0}$$

The fundamental subspace perpendicular to the column space is the left null space. We showed that $b - Ax$ is perpendicular to Col(A), so $b - Ax$ is in the nullspace.
The definition of left null space is the vector space $N(A^T)$ such that, $A^T N(A^T)$ is 0. Therefore, since $b - Ax$ is $N(A^T)$, so we know that $A^T(b - Ax)$, according to the definition of the left null space, is 0.

c. We need only manipulate the above formula to solve for x

i. Assume for the moment that $(A^T A)$ is invertible. Justify that

$$x = \left(A^T A\right)^{-1} A^T b$$

Starting from the relation from part (b.iii), we can get to the above equation by the following steps:

$$A^T(b - Ax) = \overrightarrow{0}$$

$$\left(A^T\right) b - \left(A^T A\right) x = \overrightarrow{0}$$

$$A^T b = \left(A^T A\right)\ x$$

$$\left(A^T A\right)^{-1} A^T b = \left(A^T A\right)^{-1} \left(A^T A\right)\ x$$

$$\left(A^T A\right)^{-1} A^T b = I\ x$$

$$x = \left(A^T A\right)^{-1} A^T b$$

ii. For the example system in (1), compute x and Ax, noting that $(A^T A)$ is indeed invertible in this case.

Building on the proof from the previous question:[4]

[4] #optimization: the goal was to approximate the solution for the system Ax = b by minimizing (optimizing) the value of ||Ax-b|| using the formula computed in part c.i. Although the basis of A doesn't have a combination that leads us to b, the closest we can get to be is a vector that technically represents the projection of the vector b on the plane.

In[22]:=
```
Inverse[Transpose[A].A].Transpose[A].b // MatrixForm
```

$$X = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

```
A.X // MatrixForm
```

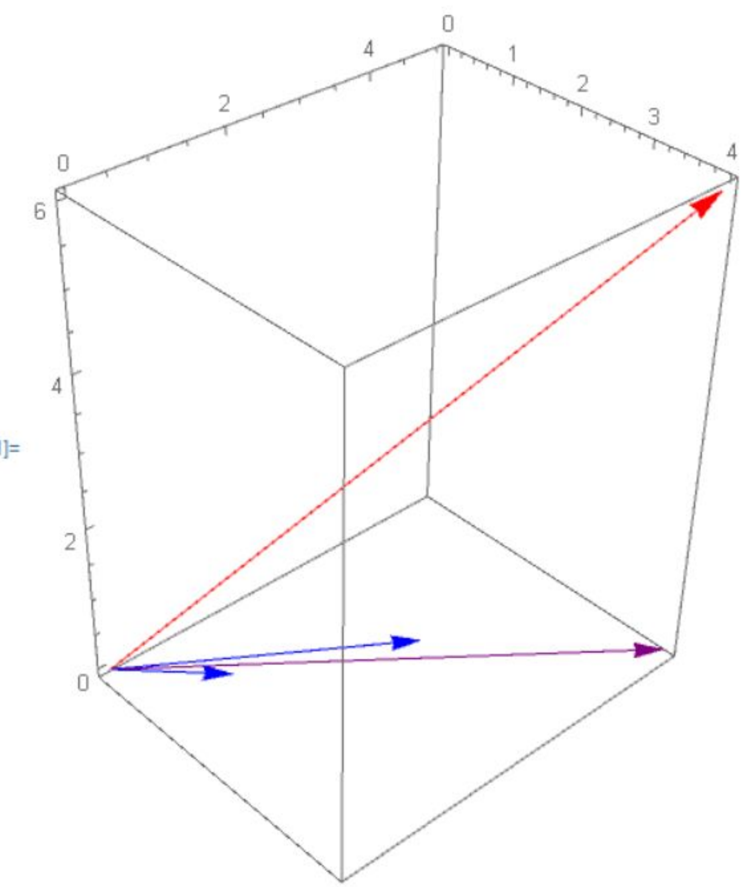
Out[24]//MatrixForm=

$$\begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix}$$

iii. Remake the plot from (1), but include the vector Ax

In[11]:=
```
Graphics3D[{Purple, Arrow[{{0, 0, 0}, {4, 5, 0}}], Blue,
  Arrow[{{0, 0, 0}, {1, 1, 0}}], Blue,
  Arrow[{{0, 0, 0}, {2, 3, 0}}], Red,
  Arrow[{{0, 0, 0}, {4, 5, 6}}]}, Axes → True]
```

Out[11]=



iv. Compute the projections of b onto the lines spanned by each of the columns of A. How does the sum of these vectors relate to A $\boldsymbol{x}$?

Let's first look at the basis set for Col(A) (1, 0, 0) and (0, 1, 0). We can use the vector (t, 0, 0) to represent the lines spanned by (0, t, 0). The formula to calculate the projection of b on a

$$\frac{\overrightarrow{a} \cdot \overrightarrow{b}}{|\overrightarrow{a}|^2} \times \overrightarrow{a}$$

Plug in the numbers, we get

$$\frac{4t}{t^2} \times \overrightarrow{a} = (4, 0, 0)$$

Second, let's look at the second basis (0, 1, 0), and we can use the vector to (0, t, 0) to represent the lines spanned by (0, 1, 0). Plug in the number, we get that the projection is

$$\frac{5t}{t^2} \times \overrightarrow{a} = (0, 5, 0)$$

The sum of these two vectors is (4, 5, 0). Note that this is the vector Ax, which is also (4, 5, 0). Therefore, we know that the sum of the projections of b onto the lines spanned by the basis of Col(A) is identical to A $\boldsymbol{x}$

d. [Optional] The above formula for finding x relies on $A^T A$ being invertible. Prove that this is true whenever the columns of A are independent. Hint: Show that A and $A^T A$ have the same nullspace.

```
In[12]:= A = (1 2
              1 3
              0 0)

Out[12]= {{1, 2}, {1, 3}, {0, 0}}

In[15]:= Transpose[A].A // MatrixForm

Out[15]//MatrixForm=
(2 5
 5 13)

In[13]:= NullSpace[Transpose[A].A] == NullSpace[A]

Out[13]= True
```

e. Regression Example. Suppose you want to predict your average HC/LO score across your next day of classes based on whether you got more or less sleep than your usual 6 hours. You decide to make a regression model $b = C + M\ t$, where b is the output, C and M are the coefficients of your model (what you need to find), and t is the number of hours above (or below for negative values) 6 hours of sleep. You feel lazy and only take three measurements: $b = 2$ when $t = -2$, $b = 2$ when $t = 1$, but $b = 5$ when $t = 3$.

   i. Inputting the above into the formula for our linear model gives us a system of three equations in two unknowns - C and M. Write out this system explicitly, then describe it as the matrix equation $Ax = b$, where:

$$x = \begin{bmatrix} C \\ M \end{bmatrix}$$

The three equations are

$$\begin{cases} C - 2M = 2 \\ C + M = 2 \\ C + 3M = 5 \end{cases}$$

Because b should be a 3x1 column vectors, so A should have the dimension 3x2. Since the coefficient of C is 1 in all three of the equations, we know that the first column of A should be [1] ,[1] ,[1]

The second column of A corresponds to the coefficient of M, which is -2, 1, and 3, so we know that the second column of A should be [-2],[1], [3].

Therefore, A is the matrix:

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

ii. Demonstrate that the system is inconsistent and that the columns of A are independent.

A system is inconsistent when there is no solution. If we subtract the first equation from the second equation, we get $-3M = 0$, which means $M = 0$. However, if $M = 0$, then the third equation is in contradiction with the first two equations. If $M = 0$, C should be 2 in the first two equations, but C should be 5 according to the third equation. Therefore, this system of equations does not have a solution that solves all the three equations in the system, hence an inconsistent system.

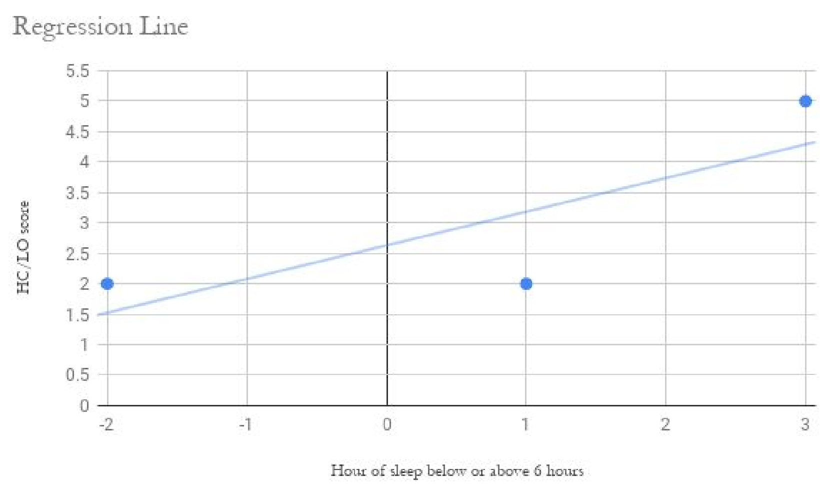
iii. Use the techniques from above to solve for x

The equation $x = (A^T A)^{-1} A^T b$ is what we use to approximate x in the inconsistent system $Ax = b$. Plugging in the matrix into the computational tool, we get that x equals

$$x = \begin{pmatrix} \frac{50}{19} \\ \frac{21}{38} \end{pmatrix}$$

iv. A diagram that includes each of the data points and the regression line.

The data points include two variables, t and b. Because we are observing how b -the scores- respond to variation in t, the hour of sleep, we use b as our y-axis and t as the x-axis. [5] [6]



2. Multinational mayhem(#vectors, #transformations)

(Adapted from G. Strang, Linear Algebra and Its Applications, 4e.) Multinational companies in the Americas, Asia, and Europe have assets of $4 trillion.
At the start, $2 trillion are in the Americas, and $2 trillion are in Europe. Each year 1/2 the American money stays home, and 1/4 goes to each of Asia and Europe. For Asia and Europe, ½ stays home and ½ is sent to the Americas.

a. Find the matrix A such that:

$$\begin{bmatrix} Americas \\ Asia \\ Europe \end{bmatrix}_{year\ k+1} = A \begin{bmatrix} Americas \\ Asia \\ Europe \end{bmatrix}_{year\ k}$$

[5] #dataviz: visualizes the data effectively using Google sheet.

[6] #correlation: draw the regression plot to correlate HC/LO score with the hours of sleep.

The matrix A is:

$$A = \begin{pmatrix} 0.5 & 0.5 & 0.5 \\ 0.25 & 0.5 & 0 \\ 0.25 & 0 & 0.5 \end{pmatrix}$$

The columns add up to 1, so it is a Markov matrix. The first column of A represents the movement of money after a year in the Americas, the second column Asia, and the third column Europe.

b. Find the diagonalization of A

To diagonalize A, we need to find the eigenvalues corresponding to A, because the diagonalization of A, is the matrix where its eigenvalues are on the diagonal, given the equation $AS = SD$, where S is the matrix whose columns are the eigenvectors of A. Using computational tool, we get that the three eigenvalues are 0, ½ and 1. Hence, the diagonalization of A, D would have eigenvalues of A on the diagonal, which is:

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We can check this by seeing if $AS = SD$, or $D = S^{-1}AS$. The eigenvectors associated with the eigenvalues 0, ½, 1 are (-2, 1, 1), (0, -1, 1), (2, 1, 1).

Therefore, S is:

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

The equation $D = S^{-1}AS$ checks out.

c. What is the initial distribution of assets? Call this y0.

The initial distribution of the assets, according to the description, is $2 trillion in the Americas and $2 trillion in Europe, and the unit is trillions of dollars:

$$y_0 = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$$

d. Express the distribution in year k, yk, as a linear combination of the eigenvectors of A and in terms of y0. Hint: First write y0 as a linear combination of the eigenvectors of A, i.e., in terms of the eigenvector basis. Then multiply this linear combination by Ak, but use your diagonalization.

When we write y0 as a linear combination of the eigenvectors of A, we get that [2, 0, 2] is the combination of the following:

$$\begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} = a \cdot \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} + b \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + c \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

From this equation, we get that

$$\begin{cases} -2a + 2c = 2 \\ a - b + c = 0 \\ a + b + c = 2 \end{cases}$$

Hence $b = 1$ and $c = 1$ and $a = 0$

We know that $y_k = A^k y_0$, and y_0 is $1 \times (0,\ -1,\ 1) + \ 1 \times (2,\ 1,\ 1)$

So we get $y_k = A^k(0,\ -1,\ 1) + A^k(2,\ 1,\ 1)$ because $D = S^{-1}AS$

We get that $A = SDS^{-1} \ \rightarrow \ A^k = SD^kS^{-1}$

Hence, we get that the distribution at yk is:

$$SD^kS^{-1} \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + SD^kS^{-1} \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

e. What happens to the distribution of assets in the very long term? How does the above help you determine this?

We can find D^k based on the diagonal matrix:

$$D^k = \begin{pmatrix} (0)^k & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^k & 0 \\ 0 & 0 & (1)^k \end{pmatrix}$$

When k goes to infinity, $(\frac{1}{2})^k$ approaches 0, so D^k becomes

$$D^k = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$A^k = SD^kS^{-1}$

$$A^k = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Plug it into the above linear combination, we get that the long term distribution is [2, 1, 1]. This is correct because it echoes with the eigenvector corresponding to the eigenvalue 1.

3. A vector space by any other name (#vectors, #transformations, #theoretical_tools)[7]

Throughout the course, we have asserted that finite, n-dimensional, abstract vector spaces are "essentially the same as" R^n . For example, we associated each polynomial $ax^2 + bx + c$ in P(2) with the vector <a, b, c> in R^3 . The goal of this problem is to upgrade this assertion to a full-fledged theorem. You'll start by exploring a few definitions.

a. Recall that the range of a map f: X → Y is the set $\{f(x) \mid x \in X\}$ in Y

   i. Let T : P(2) → R be given by $T(p(x)) = p(0)$. What is the range of T?

Following the formula above, we get that range of T is set $\{T(p(2)) \mid p(2) \in P(2)\}$ in R. Since T transforms p(x) to p(0), we know that all the P(2) will be linearly transformed p(0) in R. Hence, the range is the set p(0) in R.

   ii. Let T: V → W be a linear transformation. Show that range(T) is a subspace of W.

We know that range(T), set $\{T(v) \mid v \in V\}$ in W, is a subset of W because the destination of the linear map T is W. To further show that range(T) is a subspace of W, we need to verify the following two characteristics: first, zero vector exists in range(T), and second, range(T) is closed under addition and scalar multiplication. range(T) contains the zero vector because we know that V is a vector space, which means that it must contain a zero vector. If we send the zero vector through T into W, we would get the zero vector in W because T is a linear map.

Hence, we know that range(T) contains the zero vector. We then show that range(T) is closed under addition. Let u, v ∈ V, and send to W using T, we get T(u), T(v). If we add the two vectors, we get $T(u) + T(v)$, and given that T is a linear map, we know that this equals to T(u+v). Because u, v ∈V, we know that u+v should also ∈V. Therefore, we show that this is closed under addition. Finally, let us show that range(T) is closed under scalar multiplication. Let v ∈V, and send to W using T, we get T(v).

---

[7] #emotional_persuasion: we almost give up on this problem because this is too difficult for me at first. However, we used emotional tools, such as Taha's humour and my encouragement, to successfully conquer this gigantic obstacle that is question 3.

Let c be a scalar, and multiply T(v) with c we get cT(v). Given the linearity of T, we get that $c \cdot T(v) = T(c \cdot v)$. Because $v \in V$, cv should also ∈V, we show that range(T) is also closed under multiplication.

iii. If T : $R^m \rightarrow R^n$ given by $T(\bar{x}) = A\bar{x}$ for an n×m matrix A, which of the four fundamental subspaces does range of T correspond to? Use this identification to justify the following statement:
$dim(null(T)) + dim(range(T)) = dim(Rm)$
Note that this is true in general: given a linear transformation
$T : V \rightarrow W : dim(null(T)) + dim(range(T)) = dim(V)$

range(T) is the set $\{Ax \mid x \in R^m\}$ in R^n. Ax can be seen as the linear combination of the column of A multiplied by the values in x from top to bottom, which is the same as column space of A, or the linear map T. Hence, we know that range(T) correspond to the column space of T. Since range(T) is any vector in R^n, we know that dim(range(T)) is n, because there are n vectors that form the basis of range(T). Since dim(R^m) is m, we need to show that dim(null(T)) is (m - n) for the equation
$dim(null(T)) + dim(range(T)) = dim(R^m)$ to hold.
Let x be the set of vectors such that Tx=0 (x is the null space of T). If we want to find x, we can reduce T into RREF, then let the Tx equals to 0. When we do this, we notice that the solutions is written in terms of the linear combination of the values in x corresponding to the non-pivot columns in the RREF of T. So we know that the number of basis to null(T) is the number of non-pivot columns in T, meaning that the dimension of null(T) is the number of non-pivot columns in T. In the matrix A corresponding to the linear map T, we know that the total number of column is m, and the number of rows is n, meaning that there are m-n pivot columns. Hence, we show that the dimension of null(T) is m-n. Finally, we show that $dim(null(T)) + dim(range(T))$, which is m-n+n, is equals to $dim(R^m)$, which is m.

b. Recall that a map f: X → Y is invertible if there exists a map g: Y → X such that $(f \circ g)(x) = x$ for all x ∈ X and $(g \circ f)(y) = y$ for all y ∈ Y . Then $g = f^{-1}$

i. Show that the linear transformation T: $R^2 \rightarrow R^2$ given by $T(x, y) = (x+y, x)$ is invertible.

A linear transformation is invertible if we can use the transformed result to revert back to its original input. For this transformation T, the transformed result $(x+y, x)$ contains enough information about the original input for us to create another linear map to revert back to its input before the transformation, (x,y), so it is invertible. Let the linear mapping of the inverse of T, S, be given by $S(x, y) = (y, x-y)$. When we apply S on the result of T, we get that $S(x+y, x) = (x, y)$, showing that S is indeed the inverse of T.

ii. Let S be the inverse of T from Part (i). Find the matrix A associated with T and the matrix B associated with S. Show $B = A^{-1}$. Is the true for all invertible maps from R^n to R^n? Remember to justify your conclusion.

The matrix A associated with T is $[1, 1][1, 0]$ because the first term of the transformed input is x+y, represented in the first row as $1 \cdot x + 1 \cdot y$, and the second term of the transformed input is x, represented in the second row as $1 \cdot x + 0 \cdot y$.

The matrix B associated with S is $[0, 1][1, -1]$ because the first term of the transformed input is y, represented in the first row as $0 \cdot x + 1 \cdot y$, and the second term of the transformed input is x-y, represented in the second row as $1 \cdot x - 1 \cdot y$.

To find the inverse of A, we swap the top-left and the bottom-right entries, put a negative sign in front of the top-right and the bottom-left entries, then divide it by the determinant of A. Doing so, we get

$$[0, -1][-1, 1]/(0-1) = [0, 1][1, -1]$$

which is equal to B, so we show that B is the inverse of A.

This is true for all of the invertible maps from R^n to R^n. Suppose linear map A and B are the inverse of each other, meaning that $(A \circ B)(x) = x$. When we find the matrix C and D associated with map A and B, we should also observe that $C \times D \times x = x$. Since $C \times D$

multiple by x is still x, we know that $C \times D$ must be the identity matrix. If C multiplied by D equals the identity matrix, we hence know that C and D are the inverse of each other.

c. The Greek word isos means equal and morph means shape. Mathematicians use the term isomorphic to indicate that two objects have the same underlying structure. Two vector spaces are isomorphic if there exists an invertible linear map from one vector space to another. This invertible linear transformation is called an isomorphism.

   i. Let V be the subspace {hx, y, 0i|x, y ∈ R} of R3 (i.e., the plane z = 0). Show that the map T: V → R^2 given by $T(x,\ y,\ 0)\ =\ (x,\ y)$ is an isomorphism from V to R^2.

We can see that in this transformation, where (x, y, 0) is sent to (x, y), all the information from the original vector, in this case, the value of x and y, is kept.
Hence it is possible to send from V to R^2 using T, and send from R^2 to V using the inverse of T, which is given by $T^{-1}(x,\ y) =\ (x,\ y,\ 0)$. Hence, T is an isomorphism.

   ii. Find an isomorphism from the vector space of 2 × 3 matrices M_{2x3} to R^6

A possible isomorphism that keeps the information between the two vector spaces is to let the entries of the 2x3 matrix become the entries of the 6X1 column vector.
For example, $T(\ [a, b, c]\ [d, e,f]\ ) =\ [a]\ [b]\ [c]\ [d]\ [e]\ [f]$ so values do not get lost when transforming from one vector space to another. The inverse of this map T^{-1} is that:

$$T^{-1}(\ [a]\ [b]\ [c]\ [d]\ [e]\ [f]\ ) =\ [a, b, c]\ [d, e,f]$$

d. Now that you have built some intuition for the terms, you can prove the following lemmas that form the foundation for the theorem.

   i. Show that if T: V → W is an isomorphism then the nullspace of T is trivial (i.e., $null(T)\ =\ \{\bar{0}\}$). Use this fact to show range(T) = W. What does this tell you about the relationship between dim(V) and dim(W)?

We know that T is an invertible matrix because it is an isomorphism. Then consider the system Tx=0. Because T is invertible, we can multiply T^{-1} to both side of the equation, $T^{-1}Tx = 0$, so we know that the only solution to x, the nullspace of T, is the zero vector. We can also show this by referring to (a.iii), where we found out that the dim(null(T)) is the number of non-pivot columns in T. Again consider Tx=0, this time think of Tx as the linear combination of columns of T multiplied with the corresponding values in x. We show that x has to be 0 for Tx to be zero, meaning that the columns of T are all linearly independent. Hence, we know that there are no non-pivot columns in T, so dim(null(T)) is zero, meaning that it is trivial. Then, let us refer to the equation:

$$dim(null(T)) + dim(range(T)) = dim(V)$$

We show that for isomorphism dim(null(T)) is 0, so dim(range(T)) equals to dim(V). In this map T, AV=W, where A is the invertible matrix corresponding to T. Because an invertible matrix has to be a square matrix, we know that V and W should have the same dimension.

Therefore, we show that dim(V) equals dim(W), and $dim(range(T)) = dim(V) = dim(W)$.

Since V and W are in the same dimension, we know that every T(v) corresponds to a unique w in W. Therefore, we know that range(T), which is the set T(v) for v belongs to V in W, occupies the whole W. Hence, we show that range(T)= W.

ii. Show that if T: V → W is a linear map with the properties $null(T) = \{\bar{0}\}$ and range(T) = W, then for all w ∈ W there exists a unique v ∈ V such that T(v) = w. This implies that T is invertible, and thus an isomorphism.

For a map T, if null(T)={0}, it means that the only solution of x to Tx = 0 is 0.

This shows that T is scalar, and hence only sends a vector into another vector space with the same dimension. If range(T)=W, it means that the vector space W has completely consist of the transformed vectors v ∈ V. Hence, this means that the T is invertible because 1) the two vector spaces have the same dimension, and 2) the transformation of the vectors between the two vectors spaces are bidirectional, so there exists a T^-1 that can send w ∈ W back to v ∈V. Also, the map has to be linear, so that the above characteristics are respected. Therefore, for T : V → W, if we know that T is a linear map, null(T) ={0}, range(T) = W, then for all w ∈ W there exists a unique v ∈ V such that T(v) = w.

iii. Let $\{v1,\ v2,\ ...,\ vn\}$ be a basis for V and $\{w1,\ w2,\ ...,\ wn\}$ be a basis for W. T : V → W: $T(c1v1\ +\ c2v2\ +\ ...\ +\ cn \times vn)\ =\ c1w1\ +\ c2w2\ +\ ...\ + cn \times wn$ Show that T is an linear map, null(T) = {0}, and range(T) = W. Thus T is an isomorphism from V to W.

For T to be a linear map, it has to suffice the following two rules:

1) $T(u+v) = T(u)+T(v)\ \ \forall\ u, v \in V$

2) $T(k \cdot v) = k \cdot T(v)$

We first show it complies with the first rule

T(u+v)= T(c1(u1+v1) + c2(u2+v2) + ... + cn(un+vn))= c1(w1+w1) + c2(w2+w2) + ... + cn(wn+wn)= 2(c1w1 + c2w2 + ... + cnwn)

Note that this is equal to T(u)+ T(v), which will be

c1w1 + c2w2 + ... + cnwn+ c1w1 + c2w2 + ... + cnwn

So we show the first rule stands. Now let’s look at the second rule.

$$T(k \cdot v) = T(k \cdot c1 \cdot v1\ + k \cdot c2 \cdot v2\ +\ ...\ + k \cdot cn \cdot vn)$$
$$=\ k \cdot c1 \cdot w1\ + k \cdot c2 \cdot w2\ +\ ...\ + k \cdot cn \cdot wn$$

Note that this is the same as kT(v), which is $k(c1 \cdot w1\ +\ c2 \cdot w2\ +\ ...\ +\ cn \cdot wn)$, so we show that T is indeed a linear map. Then, we show that null(T) = {0}. Consider the equation Tx=0. Since we see that T send one linear combination to another, so T must be only a scalar. Hence, the only solution to Tx=0 is when x is 0. Finally, let us show that range(T)= W. Because range(T) is $c1 \cdot w1\ +\ c2 \cdot w2\ +\ ...\ +\ cn \cdot wn$, which is essentially the linear combination of scalars multiplied by the basis of W, so we know that range(T) is W. Therefore, showing that the three rules stand, we know that T is an isomorphism from V to W.

e. Theorem "*Two finite-dimensional vector spaces are isomorphic if and only they both have the same dimension.*"
Well done! Use this theorem to show that the following spaces are isomorphic.

i. P(n) and R^{n+1}

P(n) and R^{n+1} are isomorphic because they have the same dimension. For R^{n+1}, we know that its dimension is n+1. Its basis is formed by n+1 standard unit vectors in R^n. For P(n), because of the involvement of the constant term, so we know that its dimension is n+1. For example, in P(2), the terms that form the basis of P(2) is P(0), the constant term, P(1), and P(2), so the dimension is n+1 or 2+1=3 in this case.

ii. $M_{n \times m}$ and $R^{n\,m}$

The $n \times m$ matrix can be formed by the linear combination of the $n \times m$ matrices with one entry of 1 and the remaining entries of 0. Because there are n times m entries in the $n \times m$ matrix, the basis consists of n times m matrices, so the dimension is $n \times m$.
In $R^{n\,m}$, its basis consists of n times m standard unit vectors, so the dimension is n times m. Since the two vector spaces have the same dimension, we show that they are isomorphic.

iii. P(5) and $M_{2 \times 3}$

As shown in (ii), a $n \times m$ matrix has the dimension of n times m, so $M_{2 \times 3}$ has a dimension of 2 times 3, 6. As shown in (i), a P(n) polynomial has the dimension n+1, so P(5) has a dimension of 5+1, 6. Because the two vector spaces have the same dimension, we show they are isomorphic.

4. PageRank(#linearsystems, #vectors, #transformations, #computationaltools)

In this problem, we will explore a toy model of the Internet and Google's search algorithm, Page Rank. In our toy Internet, there will be a total of 6 webpages, denoted as nodes on a graph, as shown in Fig. 1 below.
Each edge will correspond to a link from one page to another. Note that pages that link back to one another are represented by double arrows (but you may think of them as two different edges). The main idea of the Page Rank algorithm is to find which web pages are "better" by taking into account both the number of links and the quality of these links.

a. If Mark is currently on a webpage, then we will assume that after some time he will click on one of the links and end up on a different webpage. If a page has multiple links, there is an equal probability that he will click on any of those links.
For instance, a user on webpage 1 has a 50%-50% chance of moving either to page 2 or 4 (the two links leaving page 1). We can create a Markov matrix to model this situation: consider that the columns correspond to Mark's current webpage, and the rows to the next webpage he reaches. Write a Markov matrix M that models the probabilities of moving from one page to another on our toy Internet.

The Markov matrix M should be:

$$M = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & 1 & 0 & 0 \end{pmatrix}$$

Since it is a Markov matrix, the column should add up to 1, and our matrix M satisfies this condition. In this matrix, page 1 has the probability of ½ each to end up on page 2 and page 4 next; page 2 has the probability of ⅓ each to end up on page 1, page 4, and page 5 next; page 3 has the probability of ½ each to page 1 and page 3 next, page 4 can only end up on page 6 next; page 5 can only end up on page 4 next; and page 6 and only end up on page 5 next.

b. Assume that Mark can begin his Internet adventure on any of these six webpages, with equal probability. Create a vector u_0 that represents this initial probability distribution. Compute the probability that Mark will be on each of the 6 pages after.

Since Mark can begin his Internet adventure on any of these six webpages with equal probability:

$$U_0 = [1/6,\ 1/6,\ 1/6,\ 1/6,\ 1/6,\ 1/6]$$

i. 1-time step

$$M \cdot u_0 = \begin{pmatrix} \frac{5}{36} \\ \frac{1}{12} \\ 0 \\ \frac{11}{36} \\ \frac{2}{9} \\ \frac{1}{4} \end{pmatrix} = \begin{pmatrix} 0.13889 \\ 0.08333 \\ 0 \\ 0.30556 \\ 0.22222 \\ 0.25 \end{pmatrix}$$

ii. 2-time steps

$$M \cdot M \cdot u_0 = \begin{pmatrix} \frac{1}{36} \\ \frac{5}{72} \\ 0 \\ \frac{23}{72} \\ \frac{5}{18} \\ \frac{11}{36} \end{pmatrix} = \begin{pmatrix} 0.27778 \\ 0.06944 \\ 0 \\ 0.31944 \\ 0.27778 \\ 0.30556 \end{pmatrix}$$

iii. 3-time steps

$$M \cdot M \cdot M \cdot u_0 = \begin{pmatrix} \frac{5}{216} \\ \frac{1}{72} \\ 0 \\ \frac{17}{54} \\ \frac{71}{216} \\ \frac{23}{72} \end{pmatrix} = \begin{pmatrix} 0.0231 \\ 0.0139 \\ 0 \\ 0.3148 \\ 0.3287 \\ 0.3194 \end{pmatrix}$$

iv. 10-time steps

$$M \times M \times \cdot \cdot \times M \cdot u_0 = \begin{pmatrix} \frac{1}{46656} \\ \frac{5}{93312} \\ 0 \\ \frac{32369}{93312} \\ \frac{3865}{11664} \\ \frac{469}{1458} \end{pmatrix} = \begin{pmatrix} 0.00002 \\ 0.00005 \\ 0 \\ 0.34689 \\ 0.33136 \\ 0.32167 \end{pmatrix}$$

b. Do your finding in part (b) match your expectations? Use the graph structure and the eigenvectors of matrix M to justify.

From the graph structure, we can observe that once Mark ends up in either page 4, 5, or 6, he will be “trapped” within these 3 pages between these three pages do not lead to any other pages other than page 4, 5, or 6. Furthermore, pages 4, 5, and 6 form a loop among themselves, meaning that once you’re in page 4, you will go to page 6 next, then page 5, and loop back to page 4 again. This means that once Mark is trapped in either of these 3 pages, the probability that he would stay on either page 4, 5, and 6 should be equally, so ⅓. Therefore, in the long term, we should expect a stable equilibrium distribution to be (0, 0, 0, ⅓, ⅓, ⅓)[8].

This can be supported by the eigenvector corresponding to the eigenvalue of 1, which illustrates the distribution of stable equilibrium distribution. Using Cocalc, we calculate

[8] #systemdynamics: explain how in this system, once the user gets into either page 4, 5, or 6, he/she is trapped there. Explain this system dynamics using mathematical explanation, so I think this is a really good transfer of HC

this eigenvector to be (0,0,0,1,1,1). Divided this vector by 3 to let the columns add up to 1, we get (0, 0, 0, ⅓, ⅓, ⅓), which supports our expectation.

In part b, we observe that as step increases, the probability to end up in page 1 or page 2 decreases to approach 0, while the probability to end up in page 4, 5, or 6, converges towards ⅓, which matches our expectation from the graph structure and the eigenvector.

c. What if we knew for sure that Mark started browsing on webpage 4? Create a new initial vector v0 that represents this initial probability distribution. Compute the probability that Mark will be on each of the pages after:

This vector v0 would be (0, 0, 0, 1, 0, 0) since Mark is going to start browsing at his favorite webpage, page 4.

i. 1-time step

```
In[22]:= M.v0 // MatrixForm
```

$$\begin{pmatrix} 0. \\ 0. \\ 0. \\ 0. \\ 0. \\ 1. \end{pmatrix}$$

ii. 2-time steps

```
In[23]:= M.M.v0 // MatrixForm
```

$$\begin{pmatrix} 0. \\ 0. \\ 0. \\ 0. \\ 1. \\ 0. \end{pmatrix}$$

iii. 3-time steps

In[24]:= `M.M.M.v0 // MatrixForm`

$$\begin{pmatrix} 0. \\ 0. \\ 0. \\ 1. \\ 0. \\ 0. \end{pmatrix}$$

d. How do your findings compare with those in part (b)? Use the graph structure to justify.

Compare to part b, the next page would be constrained to only either page 4, 5, or 6. This is supported by the graph structure because we see that page 4, 5, and 6 form a loop among themselves, so once you go into either of these 3 pages, you would not be able to go to either page 1, 2 or 3. And page 4 first go to page 6 next, then go to page 5, and eventually end up on page 4, which can be observed from the results of part d.

e. As you probably know, Internet browsing is not as deterministic as our toy model. Let's consider another extreme: completely random browsing. That is, after a one-time step, Mark has an equal chance of being on any of the other five pages. Write a Markov matrix R that models the probabilities of moving from one page to another in this completely random browsing mode. Describe the expected long-term probability distribution of random browsing, using the eigenvectors of matrix R to justify.

Since the probability of going to any of the 5 other pages are the same in this scenario, the matrix R would be:

$$R = \begin{pmatrix} 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 \end{pmatrix}$$

The expected long-term probability should be (⅙, ⅙, ⅙, ⅙, ⅙, ⅙) because the long-term probability to end in either of the six pages is the same. This can be justified by finding the eigenvector corresponding to the eigenvalue 1 of this matrix R.

This eigenvector is (1, 1, 1, 1, 1, 1), and divided by 6 to let the probability add up to 1, we get (⅙, ⅙, ⅙, ⅙, ⅙, ⅙), which is the same as our expectation.

f. Finally, consider a mix of deterministic and random browsing. Consider a matrix A which is a weighted sum of the two Markov matrices:

$$A = \frac{2}{3}M + \frac{1}{3}R$$

Verify that A is a Markov matrix and then describe the expected long-term probability distribution of mixed browsing, using the eigenvectors of matrix A to justify.

Matrix A, after adding ⅔ M with ⅓ R is

$$A = \frac{2}{3}M + \frac{1}{3}R = \begin{pmatrix} 0 & \frac{13}{45} & \frac{2}{5} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{2}{5} & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{2}{5} & \frac{13}{45} & \frac{1}{15} & 0 & \frac{11}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{13}{45} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{11}{15} \\ \frac{1}{15} & \frac{1}{15} & \frac{2}{5} & \frac{11}{15} & \frac{1}{15} & 0 \end{pmatrix}$$

It is a Markov matrix, as we can see its columns add up to 1. The eigenvector corresponding to the eigenvalue 1 is:

$$[\ 1,\ 1083/1168,\ 359/584,\ 128939/50224,\ 116269/50224,\ 121109/50224\ ]$$

So we know that this is the long-term distribution.

Adjusting the vector so that it adds up to 1, we get approximately

$$[\ 0.102,\ 0.094,\ 0.063,\ 0.261,\ 0.235,\ 0.245\ ]$$

Which represents the long-term distribution of this mixed matrix.

g. The probability of landing on a given page at the end corresponds to its Page Rank. Given our mixed model (part (g)), in what order would our search engine present these pages?

The higher the probability a given page would be landed on, the higher its PageRank.

Hence, the search engine should present these pages by the rank

1) page 4

2) page 6

3) page 5

4) page 1

5) page 2

6) page 3

Appendix:

Mathematica code for the final problem set:
https://drive.google.com/file/d/1p06S0zX4gmpOsJ6FDH3_gkaBoswuzh3X/view?usp=sharing

P.S: Professor, we come up with a really neat linear algebra pick up line, want to share it with you, so here it goes: “let me be the eigenvalue to your stabilized Markov matrix <3” (the one to your in long-term)

Reference:

Minerva CS111B Faculty. (2019). CS111B course materials. Retrieved from: https://seminar.minerva.kgi.edu/