Learn Linear Algebra – The coding way

Eigen Values and Eigen Vectors

Eigen Values and Vectors

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.[1]

Why should A be a N X N matrix ? Why not a M X N matrix?

What does nontrivial mean here?

Coding Assignment 1

Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

*Are **u** and **v** eigenvectors of A? Prove that visually.*

Save this as EigenValuesVectors_Notebook1

Coding Assignment 2

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 4 \end{bmatrix}$$

Write a piece of code to calculate the eigenvalue and eigenvector of matrix A. You may use Python functions to do calculate them.

How can you verify that the eigenvectors that correspond to distinct eigenvalues are independent? Is there a way to code it in Python? Clue : Reduced Row Echelon.

Save this as EigenValuesVectors_Notebook2

Coding Assignment 2

Formulate the importance of each page as a linear algebra problem.

Verify that the solution to the problem is an EigenVector.

Rescale the Eigen Vector as a probability by making sure sum of all entries in the eigenvector =1.

Save this as EigenValuesVectors_Notebook2

Coding Assignment 3

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix},$$

Write a piece of code to calculate the inverse of this matrix. Do not use any Python function to calculate the inverse.

Calculate the product of the original matrix A and its inverse. Create a Python program to calculate the product. Do not use any Python function.

Save this as EigenValuesVectors_Notebook3

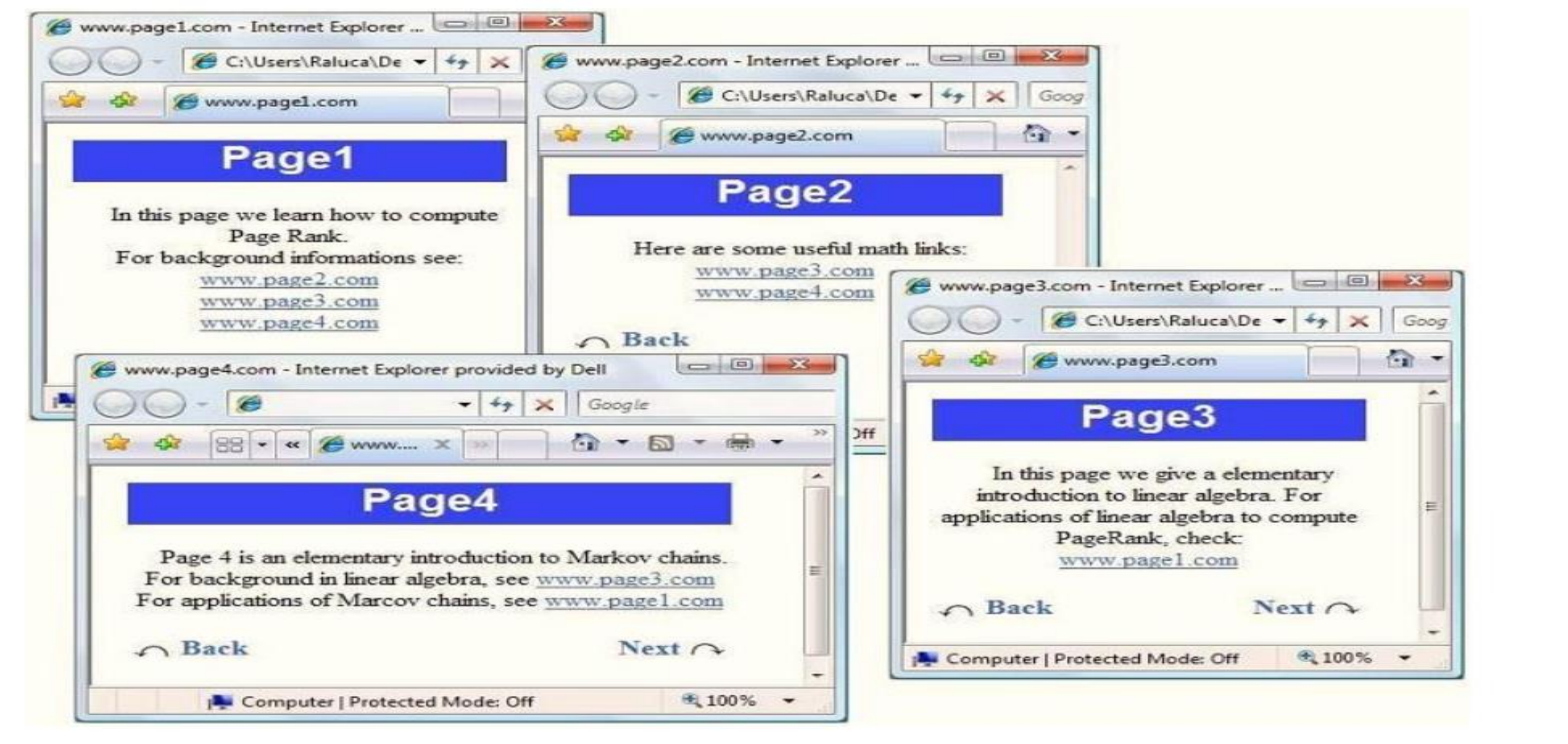
Coding Assignment 4

Modern search engines employ methods of ranking the results to provide the "best" results first that are more elaborate than just plain *text ranking*. One of the most known and influential algorithms for computing the relevance of web pages is the Page Rank algorithm used by the Google search engine. It was invented by Larry Page and Sergey Brin while they were graduate students at Stanford, and it became a Google trademark in 1998. The idea that Page Rank brought up was that, the importance of any web page can be judged by looking at the pages that link to it. If we create a web page i and include a hyperlink to the web page j, this means that we consider j important and relevant for our topic. If there are a lot of pages that link to j, this means that the common belief is that page j is important. If on the other hand, j has only one backlink, but that comes from an authoritative site k, (like www.google.com, www.cnn.com, www.cornell.edu) we say that k transfers its authority to j; in other words, k asserts that j is important. Whether we talk about popularity or authority, we can iteratively assign a rank to each web page, based on the ranks of the pages that point to it.

To this aim, we begin by picturing the Web net as a directed graph, with nodes represented by web pages and edges represented by the links between them.

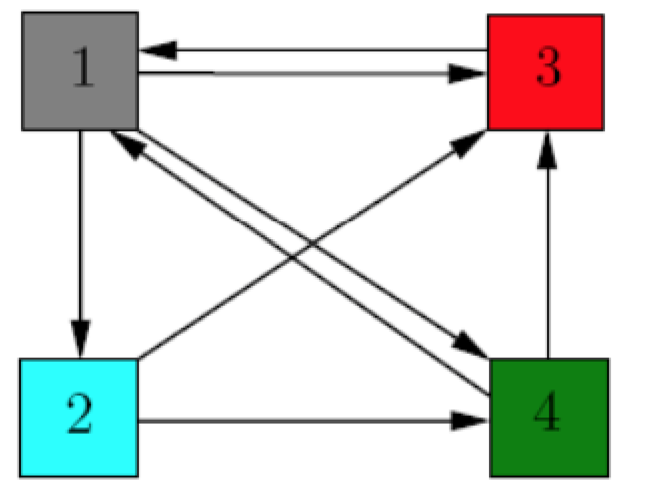
Suppose for instance, that we have a small Internet consisting of just 4 web sites www.page1.com, www.page2.com, www.page3.com, www.page4.com, referencing each other in the manner suggested by the picture:

Coding Assignment 4

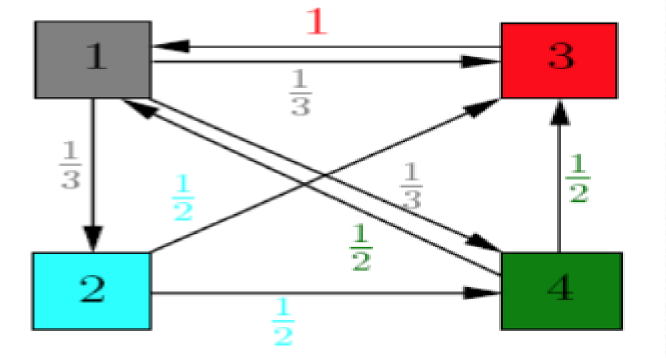


Coding Assignment 4

We "translate" the picture into a directed graph with 4 nodes, one for each web site. When web site i references j, we add a directed edge between node i and node j in the graph. For the purpose of computing their page rank, we ignore any navigational links such as back, next buttons, as we only care about the connections between different web sites. For instance, Page1 links to all of the other pages, so node 1 in the graph will have outgoing edges to all of the other nodes. Page3 has only one link, to Page 1, therefore node 3 will have one outgoing edge to node 1. After analyzing each web page, we get the following graph:



In our model, each page should transfer evenly its importance to the pages that it links to. Node 1 has 3 outgoing edges, so it will pass on $\frac{1}{3}$ of its importance to each of the other 3 nodes. Node 3 has only one outgoing edge, so it will pass on all of its importance to node 1. In general, if a node has k outgoing edges, it will pass on $\frac{1}{k}$ of its importance to each of the nodes that it links to. Let us better visualize the process by assigning weights to each edge.



Coding Assignment 4

Formulate the importance of each page as a linear algebra problem.

Verify that the solution to the problem is an EigenVector.

Rescale the Eigen Vector as a probability by making sure sum of all entries in the eigenvector =1.

Save this as EigenValuesVectors_Notebook4

Coding Assignment 5

	Page 1	Page 2	Page 3	Page 4
Page 1	0	0	1	0.5
Page 2	0.333333	0	0	0
Page 3	0.333333	0.5	0	0.5
Page 4	0.333333	0.5	0	0

The table represents the probability matrix (P), i.e the columns represent the probability of landing on a certain page (represented as rows) given that you are on a certain page (column). For example, the probability of landing on Page 2 given that you are on Page 1 is .33333

Let's represent vector **x** as the probability of landing on a certain page.

Let's say you start on Page 1, and then depending on this matrix, you move to the next page. Repeat this process, say a hundred times. How would the composition of vector **x** look ?

Let's say you start on Page 3, and then depending on this matrix, you move to the next page. Repeat this process, say a hundred times. How would the composition of vector **x** look ?

What is this **x** called ?

Save this as EigenValuesVectors_Notebook5

Coding Assignment 6

$$
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
P = \begin{bmatrix}
0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 1/3 & 1 & 0 & 0 & 0 \\
0 & 1/2 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/3 & 1
\end{bmatrix}
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\end{array}
$$

Consider this transition matrix where 1,2,3,4,5,6,7 are webpages.

Assume a user starts navigating from Page 1 and reaches Page 4. What happens to the user when she/he reaches Page 4?

Do you see any other page that suffers from the same problem?

Coding Assignment 6

ADJUSTMENT 1: If the surfer reaches a dangling node, the surfer will pick any page in the Web with equal probability and will move to that page. In terms of the transition matrix P, if state j is an absorbing state, replace column j of P with the vector

$$\begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

where n is the number of rows (and columns) in P.

In the seven-page example, the transition matrix is now

$$P_* = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 1/2 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 1 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 0 & 0 & 1/7 \\ 0 & 1/2 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 0 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

Coding Assignment 6

There could be another problem, the problem of cycles. If page j is linked only to page i and page i linked only to page j, a random surfer entering either page would be condemned to spend eternity hopping from page i to j and back again. Matrix G is called the ***Google Matrix***!

ADJUSTMENT 2: Let p be a number between 0 and 1. Assume the surfer is now at page j. With probability p the surfer will pick from among all possible links from page j with equal probability and will move to that page. With probability $1 - p$, the surfer will pick *any* page in the Web with equal probability and will move to that page. In terms of the transition matrix P_*, the new transition matrix will be

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

where K is an $n \times n$ matrix all of whose columns are[3]

$$\begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

Derive the Google Matrix for the problem and calculate the steady state vector x.

Save this as EigenValuesVectors_Notebook6

Diagonalization

Coding Assignment 1

$T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $T(\boldsymbol{x}) = D\boldsymbol{x} = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \boldsymbol{x}.$

D is what is called a Diagonal Matrix. Calculate D^2 to understand why diagonal matrices can be useful.

What does T mean here (Clue : Try to interpret Dx) ?

Consider the vectors e1 = $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and e2 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

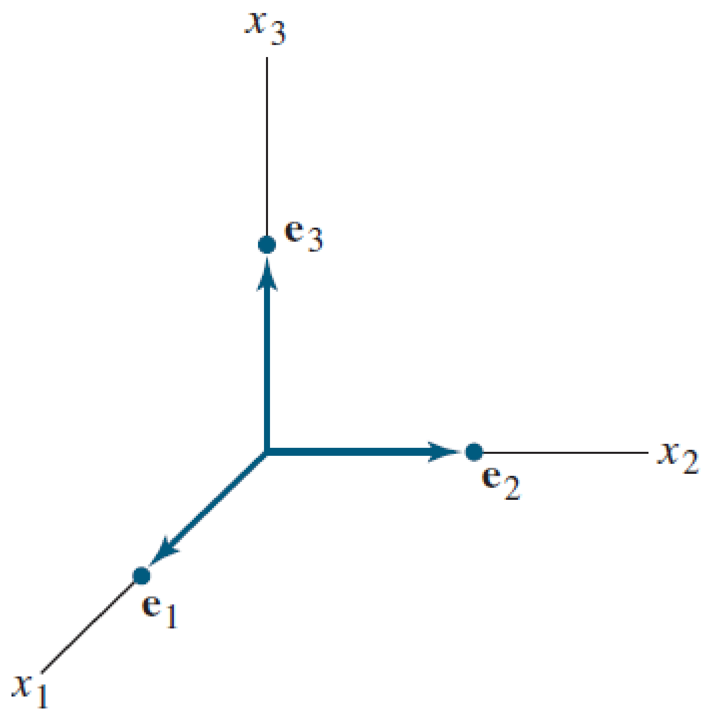
e1 and e2 are basis vecto ... $\mathbb{R}^2$. What ...s basis vectors mean?

Plot e1 and e2 vectors on a coordinate system. The transform e1 and e2 using the Diagonal matrix D. What do you notice?

Try transforming the following vector x $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ using the Diagonal matrix D. Plot it. What do you observe?

Save this as Diagonalization_Notebook1

Basis



$$\mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}.$$

One way to imagine basis, say in R^3 is to think of 3 independent vectors that can span R^3

In the image e1,e2,e3 are of length 1 unit. Can you write down e1,e2 and e3 vectors?

Can you write b vector as a product of a matrix comprising of e1, e2 and e3 vectors and another column vector?

Can you try to interpret b as a linear combination of basis vectors e1, e2 and e3?

Change of Basis

Diagonal matrices are extremely useful. They stretch vectors. But let's say the matrix is not diagonal. Are there some circumstances where we can frame the non-diagonal matrix into a diagonal matrix and use that for computations?

Now suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$, where $T(\boldsymbol{x}) = A\boldsymbol{x}$.

If A is not diagonal, then sometimes we can pick a new basis $\mathcal{B}$ and set up a new coordinate system in which the transformation T is just a rescaling along the new coordinate aces. This process (when we can do it) is called "diagonalization."

The change of coordinates matrix is $P_{\mathcal{B}} = [\boldsymbol{b}_1\ \boldsymbol{b}_2\ \dots \boldsymbol{b}_n]$ and $\begin{cases} P_{\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} = \boldsymbol{x} \\ P_{\mathcal{B}}^{-1}\boldsymbol{x} = [\boldsymbol{x}]_{\mathcal{B}} \end{cases}$

$$\begin{array}{ccc} \boldsymbol{x} & \longrightarrow & A\boldsymbol{x} \\ P_{\mathcal{B}}^{-1} \downarrow & & \uparrow P_{\mathcal{B}} \\ [\boldsymbol{x}]_{\mathcal{B}} & \longrightarrow & D[\boldsymbol{x}]_{\mathcal{B}} \end{array}$$

1) first convert $\boldsymbol{x}$ to $\mathcal{B}$ coordinates (compute $P_{\mathcal{B}}^{-1}\boldsymbol{x}$ to get $[\boldsymbol{x}]_{\mathcal{B}}$)
2) then perform the transformation
   = rescaling the $\mathcal{B}$-coordinates (multiply $DP_{\mathcal{B}}^{-1}\boldsymbol{x}$ to get $D[\boldsymbol{x}]_{\mathcal{B}}$)
3) then convert back to standard coordinates
   to arrive at the original result $A\boldsymbol{x}$ (compute $P_{\mathcal{B}}DP_{\mathcal{B}}^{-1}\boldsymbol{x} = A\boldsymbol{x}$)

This motivates an official definition: →

Suppose A is an $n \times n$ matrix. We say that A is diagonalizable if there is an invertible $n \times n$ matrix P and a diagonal matrix D such that

$$A = PDP^{-1}$$

Coding Assignment 2

Suppose $T(\boldsymbol{x}) = A\boldsymbol{x}$, where $A = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}$

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} = \underset{\displaystyle P}{\underset{\uparrow}{\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}}} \underset{\displaystyle D}{\underset{\uparrow}{\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}}} \underset{\displaystyle P^{-1}}{\underset{\uparrow}{\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}}}$$

P is invertible. Therefore the columns of P are linearly independent. *Why?*

$$\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2\} = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

B is a new basis in R^2. *Why?*

For $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, we can compute $A\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$.

$$A\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \underset{\displaystyle A}{\underset{\uparrow}{\begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}}} \underset{\displaystyle \boldsymbol{x}}{\underset{\uparrow}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}} = \underset{\displaystyle P_{\mathcal{B}}}{\underset{\uparrow}{\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}}} \underset{\displaystyle D}{\underset{\uparrow}{\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}}} \underset{\displaystyle P_{\mathcal{B}}^{-1}}{\underset{\uparrow}{\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}}} \underset{\displaystyle \boldsymbol{x}}{\underset{\uparrow}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

This has more insights.

Coding Assignment 2

$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \mapsto \underbrace{\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}}_{\uparrow} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}}_{\uparrow}.$$

multiplication by $P_{\mathcal{B}}^{-1}$ converts standard coordinates into $\mathcal{B}$-coordinates

$\uparrow$ $[\boldsymbol{x}]_{\mathcal{B}}$

Then

$$\underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}}_{\uparrow} \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 8 \end{bmatrix}}_{\uparrow}$$

the diagonal matrix D stretches the new $\mathcal{B}$-coordinates by factors of 3 and 6 (in the coordinate directions corresponding to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$)

stretched $\mathcal{B}$-coordinates

(cont. $\longrightarrow$)

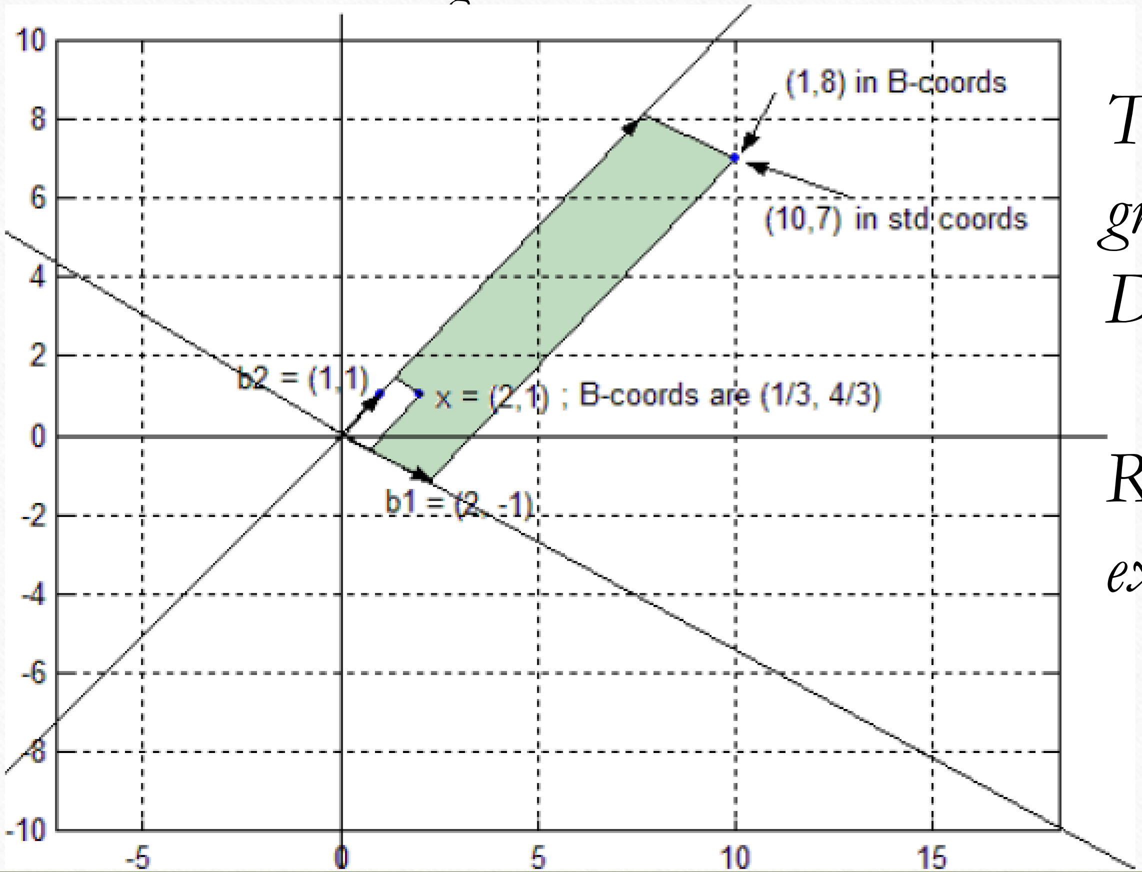
Finally,

$$\underbrace{\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}}_{\uparrow} \begin{bmatrix} 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

multiplication by $P_{\mathcal{B}}$ converts the (stretched) $\mathcal{B}$-coordinates back into standard coordinates

Coding Assignment 2

Draw the axes of the original basis. Draw the axes of the revised basis let's call this basis B (vectors in matrix P)



The coding assignment is to draw this graph. Save the notebook as Diagonalization_Notebook2

Refer the next slide for some useful explanation on the image.

Coding Assignment 2

1) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$; the same geometric point, named in $\mathcal{B}$-coordinates, is $[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}$

2) Rescaling $[\boldsymbol{x}]_{\mathcal{B}}$ by the factors 3 and 6 in the $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ directions gives a new point whose name in $\mathcal{B}$-coordinates is $\begin{bmatrix} 1 \\ 8 \end{bmatrix}$.

3) This new point has a different name in standard coordinates. If we convert the $\mathcal{B}$-coordinates $\begin{bmatrix} 1 \\ 8 \end{bmatrix}$ back to standard coordinates, we get $\begin{bmatrix} 10 \\ 7 \end{bmatrix} = A \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Thus
$$\begin{cases} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \mapsto A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix} & \text{expressed in standard coordinate} \\ & \text{describes the same thing as} \\ \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} \mapsto D \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix} & \text{expressed in } \mathcal{B}\text{-coordinates} \end{cases}$$

Coding Assignment 3

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

Write a piece of code to form P, a matrix of eigenvectors of A (You may use a readymade function). Also use a readymade function to calculate D.

Write a piece of code to calculate P^{-1} (do not use a readymade function).

Multiply PDP^{-1} (do not use a readymade function). Check if it matches with A.

Take the columns of P and prove using a piece of code that they are independent.

Save this as Diagonalization_Notebook3

Matrix Multiplication – A different perspective

Some useful theory for a change!

Rank of a Matrix

Multiplication Ax Using Columns of A

$$\begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$$

Thus Ax is a linear combination of the columns of A. This is fundamental.

The combinations of the columns fill out the **column space** of A.

Independent Columns and Rank of A

All this comes with an understanding of **independence**. The goal is to create a matrix C whose columns come directly from A—but not to include any column that is a combination of previous columns. The columns of C (as many as possible) will be "independent". Here is a natural construction of C from the n columns of A:

If column 1 of A is not all zero, put it into the matrix C.

If column 2 of A is not a multiple of column 1, put it into C.

If column 3 of A is not a combination of columns 1 and 2, put it into C. *Continue*.

At the end C will have r columns ($\boldsymbol{r \leq n}$).

They will be a "basis" for the column space of A.

The left out columns are combinations of those basic columns in C.

Independent Columns and Rank of A

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}$$

How many independent columns are there?
Construct Matrix C.

The rank of a matrix is the dimension of its column space.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} = \boldsymbol{CR}$$

Interpret this as linear combination of columns and linear combination of rows.

The number of *independent columns* equals the number of *independent rows*

The big factorization for data science is the "SVD" of A—when the first factor C has r *orthogonal* columns and the second factor R has r *orthogonal* rows.

More about this later….

Matrix-Matrix Multiplication

"Outer product" $uv^{\mathrm{T}} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix} =$ **"rank one matrix"**

$$AB = \textbf{Sum of Rank One Matrices}$$

$$AB = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & \boldsymbol{b}_1^* & — \\ & \vdots & \\ — & \boldsymbol{b}_n^* & — \end{bmatrix} = \underbrace{\boldsymbol{a}_1 \boldsymbol{b}_1^* + \boldsymbol{a}_2 \boldsymbol{b}_2^* + \cdots + \boldsymbol{a}_n \boldsymbol{b}_n^*}_{\textbf{sum of rank 1 matrices}}.$$

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 5 \end{bmatrix} = $$

Write this as a sum of rank 1 matrices

Why is the outer product approach essential in data science ?
The short answer is : *We are looking for the important part of a matrix A.*

Diagonalization of Symmetric Matrices

Coding Assignment 1

A **symmetric** matrix is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

If A is a symmetric matrix, then the eigenvectors are orthogonal. *What does this mean?*

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}. \quad \lambda = 8: \ \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = 6: \ \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}; \quad \lambda = 3: \ \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

When A is symmetric, it is orthogonally diagonalizable.

$$A = PDP^T = PDP^{-1}$$

Let's do some coding to understand this.

Take the above matrix A, calculate Eigen vectors and values. Check if the Eigenvectors are orthonormal (after converting them to unit vectors).Construct P, D, P^T and P^{-1}

Save this as Diagonalization_SymmetricNotebook1

Coding Assignment 2

Suppose $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$
\begin{aligned}
A = PDP^T &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 \mathbf{u}_1 & \cdots & \lambda_n \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}
\end{aligned}
$$

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A.

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$$

Construct a spectral decomposition of A.

Save this as DiagonalizationSpectralDecomposition_Notebook2

Quadratic Forms

Coding Assignment 1

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic forr**

Is **Q(x)** a scalar or a vector?

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Compute $\mathbf{x}^T A \mathbf{x}$ for the following matrices:

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \qquad A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

Which one was easier to compute?

$$Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2.$$

Express Q(x) in $x^T Ax$ format. Calculate Q(x) where x = [2,
-2]

Change of Variable in a Quadratic Form

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\mathbf{x} = P\mathbf{y}, \qquad \text{or equivalently,} \qquad \mathbf{y} = P^{-1}\mathbf{x}$$

What is the basis of vector **y**?

Coding Assignment 1

Let's apply the change of variable to the quadratic form x^TAx.

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T A P\mathbf{y} = \mathbf{y}^T(P^T A P)\mathbf{y}$$

A is a symmetric matrix. So it can be orthogonally diagonalizable. i.e.

$$A = PDP^T$$

So what is P^TAP ??

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product term.

Coding Assignment 1

$Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2.$ $\quad \mathbf{y}^T(P^TAP)\mathbf{y} \quad \mathbf{y}^TD\mathbf{y}.$

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix} \qquad \mathbf{x} = P\mathbf{y}, \qquad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Construct P matrix and D matrix. Verify that D is a diagonal matrix.

Compute Q(x) for x = [-2,
2]

For the same x compute y^TDy

Verify that Q(x) is the same after change of variables.

Save this as QuadraticForms_Notebook1

Constrained Optimization

Constrained Optimization

Engineers, economists, scientists, and mathematicians often need to find the maximum or minimum value of a quadratic form Q(**x**) for **x** in some specified set. Typically the problem can be arranged so that **x** varies over a set of unit vectors. Such a problem has an elegant solution.

The requirement that a vector $\mathbf{x}$ in $\mathbb{R}^n$ be a unit vector can be stated in several equivalent ways:

$$\|\mathbf{x}\| = 1, \qquad \|\mathbf{x}\|^2 = 1, \qquad \mathbf{x}^T\mathbf{x} = 1$$

and

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$$

When a quadratic form Q has no cross-product terms, it is easy to find the maximum and minimum of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

EXAMPLE 1 Find the maximum and minimum values of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

Coding Assignment 1

$$m = \min\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}, \quad M = \max\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}$$

Let A be a symmetric matrix, and define m and M as in (2). Then M is the greatest eigenvalue λ_1 of A and m is the least eigenvalue of A. The value of $\mathbf{x}^T A\mathbf{x}$ is M when $\mathbf{x}$ is a unit eigenvector $\mathbf{u}_1$ corresponding to M. The value of $\mathbf{x}^T A\mathbf{x}$ is m when $\mathbf{x}$ is a unit eigenvector corresponding to m.

(2) Refers to definition of m and M.

EXAMPLE 3 Let $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$. Find the maximum value of the quadratic form $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$, and find a unit vector at which this maximum value is attained.

Save this as ConstrainedOptimization_Notebook1

Coding Assignment 2

Let A, λ_1, and $\mathbf{u}_1$ be as in Theorem 6. Then the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

is the second greatest eigenvalue, λ_2, and this maximum is attained when $\mathbf{x}$ is an eigenvector $\mathbf{u}_2$ corresponding to λ_2.

Theorem 6 refers to the theorem in the previous slide.

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}.$$

EXAMPLE 5 Let A be the matrix in Example 3 and let $\mathbf{u}_1$ be a unit eigenvector corresponding to the greatest eigenvalue of A. Find the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the conditions

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

Save this as ConstrainedOptimization_Notebook2

Constrained Optimization

Let A be a symmetric $n \times n$ matrix with an orthogonal diagonalization $A = PDP^{-1}$, where the entries on the diagonal of D are arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and where the columns of P are corresponding unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Then for $k = 2, \ldots, n$, the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0, \quad \ldots, \quad \mathbf{x}^T\mathbf{u}_{k-1} = 0$$

is the eigenvalue λ_k, and this maximum is attained at $\mathbf{x} = \mathbf{u}_k$.

Try interpreting this.

Singular Value Decomposition

Coding Assignment 1

Diagonalization is very useful but not all matrices can be factorized as $A = PDP^{-1}$. However a factorization $A = QDP^{-1}$ is possible for any m x n matrix.

This special types of factorization called the SVD, is the most useful matrix factorization in linear algebra applications.

Let's consider a non-symmetric matrix A. Find a unit vector **x** at which the length $||A\mathbf{x}||$ is maximized and compute this length

If $||A\mathbf{x}||$ is maximized for a given **x,** $||A\mathbf{x}||^2$ is also maximized.

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T(A^T A)\mathbf{x}$$

Is A^TA , a symmetric matrix? *Is $||\mathbf{x}|| = 1$? Why?*

Think about the x that would maximize $||A\mathbf{x}||^2$. Clue : Think about constrained optimization

Coding Assignment 1

Now that we have understood the theory, let's solve a problem.

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix},$$

Find a unit vector **x** **(**let's call this **v1)** at which the length ||A**x**|| is maximized. Compute the maximum length.

Save this as SVD_Notebook1

Coding Assignment 2

The Singular Values of an $m \times n$ Matrix

Let A be an $m \times n$ matrix. Then A^TA is symmetric and can be orthogonally diagonalized. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA, and let $\lambda_1, \dots, \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \le i \le n$,

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^TA\mathbf{v}_i = \mathbf{v}_i^T A^TA\mathbf{v}_i \\ &= \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) \qquad \text{Since } \mathbf{v}_i \text{ is an eigenvector of } A^TA \\ &= \lambda_i \qquad \text{Since } \mathbf{v}_i \text{ is a unit vector} \end{aligned} \tag{2}$$

So the eigenvalues of A^TA are all nonnegative. By renumbering, if necessary, we may assume that the eigenvalues are arranged so that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$$

The **singular values** of A are the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \dots, \sigma_n$, and they are arranged in decreasing order. That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \le i \le n$. By equation (2), *the singular values of A are the lengths of the vectors* $A\mathbf{v}_1, \dots, A\mathbf{v}_n$.

Coding Assignment 2

Let's expand a bit more on the previous assignment.(Coding Assignment 1)

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix},$$

Calculate the Singular Values of A. Identify the unit vector x (let's call this **v2**) that maximizes||A**x**||, subject to ||**x**|| =1 and is orthogonal to **v1.**

Are A**v1**, A**v2,** orthogonal?

Are A**v1**, A**v2** in the column space of A? Why?,

Are A**v1**,A**v2** a basis for Column Space of A?

What is the rank of A?

Save this as SVD_Notebook2