LINEAR ALGEBRA IN MACHINE LEARNING, ROBOTICS AND COMPUTER VISION

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

MASTER OF SCIENCE

IN

MATHEMATICS

SUBMITTED BY:

Chitrankan Rathore
(2K22/MSCMAT/07)

Preeti Yadav
(2K22/MSCMAT/31)

UNDER THE SUPERVISION OF
Mr. JAMKHONGAM TOUTHANG

DEPARTMENT OF APPLIED MATHEMATICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur,Main Bawana Road, Delhi – 110042

Delhi Technological University
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

We, **Chitrankan Rathore**, Roll No. 2K22/MSCMAT/07, and **Preeti Yadav**, Roll No. 2K22/MSCMAT/31, students of Master of Science in Mathematics, hereby declare that the dissertation titled *"Linear Algebra in Machine Learning, Robotics and Computer Vision"*, submitted to the Department of Applied Mathematics, Delhi Technological University, Delhi, in partial fulfillment of the requirements for the degree of Master of Science, is our original work under the supervision of **Mr. Jamkhongam Touthang**. Proper citations have been given wherever necessary, and this work has not been submitted previously for any degree, diploma, associateship, or any other similar title or recognition.

PREETI YADAV **CHITRANKAN RATHORE**

This is to certify that the students have incorporated all the corrections suggested by the examiners in the dissertation and the statement made by the candidates is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE BY THE SUPERVISOR

Certified that **CHITRANKAN RATHORE** (2K22/MSCMAT/07) and **PREETI YADAV** (2K22/MSCMAT/31) have carried out their search work presented in this dissertation entitled"**LINEAR ALGEBRA IN MACHINE LEARNING, ROBOTICS AND COMPUTER VISION**"for the award of the degree of **Master of Science** from Department of Applied Mathematics, Delhi Technological University, Delhi under my supervision .The dissertation embodies results of original work, and studies are carried out by the students themselves and the contents of this dissertation do not form the basis for the award of any other degree to the candidates or to anybody else from this or any other University.

Place: Delhi
Date:

Mr. JAMKHONGAM TOUTHANG
SUPERVISOR
Assistant Professor
Delhi Technological University

ACKNOWLEDGEMENT

Throughout the process of writing this dissertation, we received substantial assistance and encouragement. First and foremost, we are deeply grateful to Prof. Jamkhongam Touthang for his significant contributions to the development of the research topics and methodology. His insightful feedback motivated us to refine our ideas and enhance the quality of our work. His guidance and provision of necessary resources were crucial in helping us navigate the right path and complete this dissertation successfully.

Additionally, we would be delighted to express our sincere thanks to God and to our parents for their wise counsel and empathetic support. Their unwavering encouragement and support were indispensable throughout this journey.

CHITRANKAN RATHORE
PREETI YADAV
M.Sc Mathematics
DTU, New Delhi

Abstract

Linear algebra is fundamental to machine learning, robotics, and computer vision, providing the mathematical foundation for data representation, model training, optimization, transformations, and complex computations. Its extensive applications, ranging from basic data preprocessing to advanced algorithm development and system design, make it indispensable for developing and implementing technologies in these fields.Wavelets and Haar matrices plays a crucial role in compressing as well as processing audio and video signals.Various different methods are also used to deal with the problem of curve interpolation.We have also discussed about vector norms which is used to evaluate model's error or reduce model's complexity.The versatility of Singular Value Decomposition(SVD) in data compresssion, particularly in image processing, solve problems like least square optimization, dimensionality reduction(PCA),pattern recognition and approximation also cannot be neglected.

Keywords: Hadamard matrices, Haar bases, Euclidean spaces,Singular Value Decomposition, Spectral theorem.

Contents

Chapter 1

Introduction

1.1 Vector Spaces, Linear Maps and Bases

1.1.1 Linear Combination and Linear Independence

For the [1] linear optimization problems, we frequently encounter the systems of linear equations. For instance, let the problem of solving the system of linear equations with 3 variables: $x_1, x_2, x_3 \epsilon \mathbb{R}$ where

$$\begin{aligned} 2x_1 - x_2 + 2x_3 &= 1 \\ x_1 + x_2 + 2x_3 &= 3 \\ x_1 - 2x_2 - x_3 &= 2 \end{aligned}$$

Easiest method to solve this kind of problem is to introduce the vectors: a, b, c and u given by

$$a = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}, \quad c = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

and to represent our linear system of given equations as a linear combination of a, b, c ; i.e.

$$x_1 a + x_2 b + x_3 c = u$$

where [1] a, b, c are the vectors and $x_{i's}$ are scalars.

Hence, the vectors are said to be linearly independent as there is no triple $(x_1, x_2, x_3) \neq 0$ i.e.

$$x_1 a + x_2 b + x_3 c = 0_3$$

Here, 0_3 is a zero vector : $0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

Thus, every vector z ϵ $\mathbb{R}^{3\times1}$ could be represented uniquely in the form of a linear combination:

$$z = x_1a + x_2b + x_3c$$

then this linear combination [1] can be written in the matrix form also as

$$\begin{pmatrix} 2 & -1 & 2 \\ 1 & 1 & 2 \\ 1 & -2 & -1 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

has a unique solution with the values:

$$x_1 = 1.47,\ x_2 = -0.87,\ x_3 = 1.2$$

Determinant of a, b, c : $\begin{vmatrix} 2 & -1 & 2 \\ 1 & 1 & 2 \\ 1 & -2 & -1 \end{vmatrix} = -3 \neq 0$

This means if the determinant of a, b, c is non-zero then these vectors are **linearly independent.**

1.1.2 Linear Maps, Linear Forms and Dual Space

Definition: *A linear map* between any 2 vector spaces [2] E and F be a function f : E→F fulfilling the following properties:

$$\text{f(x+y) = f(x) + f(y) , for all x,y } \epsilon \text{ E;}$$

$$\text{f}(\lambda x) = \lambda \text{ f(x) for all } \lambda \epsilon \mathbb{R}, \text{ x } \epsilon \text{ E.}$$

When we putting x and y be 0 in the 1st condition and getting f(0) = 0.*The basic property of the linear maps is that they can transform linear combinations into linear combinations.* Let any finite family $(u_i)_{i\epsilon I}$ of the vectors in E and $(\lambda_i)_{i\epsilon I}$ of scalars in $\mathbb{R}$, we get:

$$f(\textstyle\sum_{i\epsilon I} \lambda_i u_i) = \sum_{i\epsilon I} \lambda_i f(u_i)$$

Definition: Given any [1] linear map f : E→F, we can explain its *image (or range)* Im f=f(E) as :

$$\text{Im f} = \{y\epsilon F | (\exists x \epsilon E)(y = f(x))\},$$

and its *Kernel (or nullspace)* Ker f = $f^{-1}(0)$ as:

$$\text{Ker f} = \{x \epsilon E | f(x) = 0\}.$$

Note: Given any linear map f : E→F, the *rank (or rk(f))* of f is the dimension of the image Im f of f.

Definition: Given any vector space E with the vector space Hom(E,K) of a linear maps from E to K (field) is said to be its *dual space (or dual)* of E.

The space Hom(E,K) is also represent as E^* and the linear maps in E^* are known as *the linear forms (or covectors).*

[2] [1] The dual space E^{**} of the space E^* is said to be the *bidual of E.*

Theorem: *(Existence of dual bases)* Let E be any vector space of the dimension n which holds the mentioned condition:

For each basis $(u_1, ..., u_n)$ *of vector space E and the family of coordinate forms* $(u_1^*, ..., u_n^*)$ *is a basis of* E^* is known as the dual basis of $(u_1, ..., u_n)$ [1] .

1.2 Matrices with Linear Maps

1.2.1 Matrix Representation of Linear Maps

Let any 2 [2] vector spaces E and F with the basis $(u_j)_{j \epsilon J}$ of E, each linear map $f : E \to F$ is distinctive set on the family of $(f(u_j))_{j \epsilon J}$ of images under f of the vectors in the basis $(u_j)_{j \epsilon J}$.

Therefore, similarly [2] a basis $(v_i)_{i \epsilon I}$ of F, then each vector $f(u_j)$ will be represented uniquely as -

$$f(u_j) = \sum_{i \epsilon I} a_{ij} v_i,$$

where $j \epsilon J$, for family of scalars $(a_{ij})_{i \epsilon I}$.

Hence, w.r.t. two bases $(u_j)_{j \epsilon J}$ and $(v_i)_{i \epsilon I}$ of E and F respectively, the linear map f will be entirely set on by [1] $I \times J$ - *matrix* i.e.

$$M(f) = (a_{ij})_{i \epsilon I, j \epsilon J}.$$

Let I and J be finites then say $|I| = c$ and $|J| = d$. Then, f which is a linear map is aimed by the matrix M(f) whose entries in the j-th column are the constituents of the vector $f(u_j)$ over the basis $(v_1, v_2, ..., v_c)$ i.e.

$$\mathrm{M(f)} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{c1} & a_{c2} & \dots & a_{cd} \end{pmatrix}$$

whose enteries [1] on row i and column j is $a_{ij} (1 \leq i \leq c, 1 \leq j \leq d)$

Then, we prove that E and F having finite dimensions. As we can simply represent linear maps by the [2] matrices and the composition of linear maps corresponds to the matrix multiplication.

[2] Let E and F be 2 vector spaces with finite bases $(u_1, u_2, ..., u_d)$ and $(v_1, v_2, ..., v_c)$ respectively. From previous, we proven that each vector x ϵ E and y ϵ F can be written uniquely as-

$$x = x_1u_1 + x_2u_2 + ... + x_du_d$$
$$y = y_1v_1 + y_2v_2 + ... + y_cv_c$$

Let $f : E \to$ F be a linear mapping between vector spaces E and F. Thus, for every $x = x_1u_1 + ... + x_du_d$ in E (using linearity) we get -

$$f(x) = x_1f(u_1) + ... + x_df(u_d)$$

Letting $f(u_j) = a_{1j}v_1 + ... + a_{cj}v_c$

$$\text{i.e. } \mathrm{f(u}_j) = \sum_{i=\ 1}^{c}\ a_{ij}v_i$$

for all j, 1 $\leq$ j $\leq$ d.

We also expressed this by [2] the coefficients $a_{1j}, a_{2j}, ..., a_{cj}$ of the $f(u_j)$ over the basis $(v_1, v_2, ..., v_c)$ as j-th column of the matrix M(f) as:

$$\begin{pmatrix} \\ v_1 \\ v_2 \\ \vdots \\ v_c \end{pmatrix} \begin{pmatrix} f(u_1) & f(u_2) & \dots & f(u_d) \\ a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{c1} & a_{c2} & \dots & a_{cd} \end{pmatrix}$$

Now, we will putting the RHS of every [2] $f(u_j)$ into the expression f(x) and getting-

$$\text{f(x)} = \text{x}_1(\sum_{i=\ 1}^{c}\ a_{i1}v_i) + ... + x_d(\sum_{i=\ 1}^{c}\ a_{id}v_i)$$

$$= (\sum_{j=\ 1}^{d}\ a_{1j}x_j)v_1 + ... + (\sum_{j=\ 1}^{d}\ a_{cj}x_j)v_c$$

Now, letting: f(x) = y = $y_1v_1 + ... + y_cv_c$ which gives:-

$$\text{y}_i = \sum_{j=\ 1}^{d}\ a_{ij}x_j$$

for all i, $1 \leq i \leq c$

Taking a case where c=2 and d=3:

$\text{f}(\text{u}_1) = a_{11}v_1 + a_{21}v_2$
$\text{f}(\text{u}_2) = a_{12}v_1 + a_{22}v_2$
$\text{f}(\text{u}_3) = a_{13}v_1 + a_{23}v_2$

which is expressed in the matrix.

$$\begin{pmatrix} \\ v_1 \\ v_2 \end{pmatrix} \begin{pmatrix} f(u_1) & f(u_2) & f(u_3) \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

and for any $x = x_1u_1 + x_2u_2 + x_3u_2$, we have-

$$f(x) = f(x_1u_1 + x_2u_2 + x_3u_3) = x_1f(u_1) + x_2f(u_2) + x_3f(u_3)$$
$$= x_1(a_{11}v_1 + a_{21}v_2) + x_2(a_{12}v_1 + a_{22}v_2) + x_3(a_{13}v_1 + a_{23}v_2)$$
$$= (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)v_1 + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3)v_2$$

Since, $y = y_1v_1 + y_2v_2$

We have:

$$y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3$$

This truely agree with the matrix equation:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Now we will easily formalize the presentation of the linear maps by matrices.

Definition: Given [1] E and F be any 2 vector spaces with their bases $(u_1, ..., u_d)$ and $(v_1, ..., v_c)$ respectively where every vector x ϵ E convey in the basis $(u_1, ..., u_d)$ as the linear combination: $x = x_1u_1 + ... + x_du_d$ is presented by the column matrix-

$$\mathrm{M(x)} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

[1] and likewise for the every vector y ϵ F which is indicated in the basis $(v_1, ..., v_c)$.

1.2.2 Linear Maps and Matrix Multiplication with their compositions

Definition: Let [3] $U = (u_1, ..., u_d)$ and $V = (v_1, ..., v_c)$ for the bases of E and F respectively and denoted by $M_{u,v}(f)$, the matrix of f w.r.t the bases of U and V respectively. Therefore, use x_U for the co-ordinates of $M(x) = (u_1, ..., u_d)$ of x ϵ E w.r.t basis U and then write the y_V for the corresponding co-ordinates $M(y) = (v_1, ..., v_c)$ of y ϵ F w.r.t basis V. Hence,

$$\mathrm{y = f(x)}$$

is showed in a matrix form by:

$$y_V = M_{U,V}(f)x_U$$

Propositions: 1) Let any 3 matrices be A [2] ϵ $M_{m,n}(K)$, B ϵ $M_{n,p}(K)$ and C ϵ $M_{p,q}(K)$, we get:

$$(AB)C = A(BC);$$

i.e. a matrix multiplication (which is associative).

Proof: For each [1] m $\times$ n matrix A $= (a_{ij})$ which explains the function $f_A : K^n \to K^m$ given as:

$$\mathrm{f}_A(x) = Ax,$$

for each x ϵ K^n. This is instantly proved [2] that f_A is a linear and a matrix $M(f_A)$ represents the f_A over the cannonical bases in K^n and K^m is equal to A.

Hence, $M(f_A o f_B)$ = [3] $M(f_A)M(f_B)$ = AB,

here we get -

$$M((f_A o f_B) o f_C) = M(f_A o f_B)M(f_C) = \text{(AB)C}$$

$$\text{and } M(f_A o (f_B o f_C)) = M(f_A)M(f_B o f_C) = \text{[1] A(BC)}$$

hence, the composition of functions is associative, we have

$$(f_A o f_B) o f_C = f_A o (f_B o f_C)$$

which implies that :

$$(\mathrm{AB})\mathrm{C} = \mathrm{A}(\mathrm{BC})$$

2) Let any 2 [2] matrices A, B ϵ $M_{m,n}(K)$ and C, D ϵ $M_{n,p}(K)$, for all $\lambda \epsilon K$, we get:

$$\begin{gathered}(\mathrm{A+B})\mathrm{C}=\mathrm{AC}+\mathrm{BC} \\ \mathrm{A}(\mathrm{C+D})=\mathrm{AC}+\mathrm{AD} \\ (\lambda A)C = \lambda(AC) \\ \mathrm{A}(\lambda C) = \lambda(AC)\end{gathered}$$

then, the matrix multiplication . : $M_{m,n}(K) \times M_{n,p}(K) \to M_{m,p}(K)$ *is bilinear.* [2]

3) Any 3 vector spaces A, B, C with their [5] respective bases $(u_1, ..., u_p)$, $(v_1, ..., v_n)$ and $(w_1, ..., w_m)$, a mapping M: $Hom(A, B) \to M_{n,p}$ that will connects the matrix M(g) to a linear map $g : A \to B$ serves the following conditions for every x ϵ A, $g, h : A \to B$ and $f : B \to C$:

$$\begin{gathered}\mathrm{M}(\mathrm{g}(\mathrm{x})) = \mathrm{M}(\mathrm{g})\mathrm{M}(\mathrm{x}) \\ \mathrm{M}(\mathrm{g+h})=\mathrm{M}(\mathrm{g})+\mathrm{M}(\mathrm{h}) \\ \mathrm{M}(\lambda g) = \lambda M(g) \\ \mathrm{M}(\mathrm{f\ o\ g})= \mathrm{M}(\mathrm{f})\mathrm{M}(\mathrm{g})\end{gathered}$$

where M(x) is the column vector connected with the vector x and M(g(x)) is the column vector which is connected with the g(x).

1.2.3 The Change of Basic Matrix

Proposition: Let A be any vector space with a basis $(u_1, ..., u_d)$. For each family $(v_1, ..., v_d)$, let $P = (a_{ij})$ be a matrix explained as $v_j = \sum_{i=\ 1}^{d}\ a_{ij}u_i$. The matrix P is an invertible iff $(v_1, ..., v_d)$ is also a basis of E.

Definition: Let any vector space E with dimension d, for some 2 [1] bases $(u_1, ..., u_d)$ and $(v_1, ..., v_d)$ of E, assume $P = (a_{ij})$ be an invertible matrix explained as:

$$v_j = \sum_{i=\ 1}^{d}\ a_{ij}u_i$$

NOTE: [2] Change of the basis matrix from U to V is indicated by $P_{V,U}$ such that

$$P_{U,V} = P^{-1}{}_{V,U}$$

1.2.4 Effect of Change of the Bases on Matrices

Proposition: Let A and B be 2 vector spaces where let $U = (u_1, u_2, ..., u_n)$ and $U' = (u'_1, u'_2, ..., u'_n)$ be the 2 bases of A and then, let $V = (v_1, v_2, ..., v_m)$ and $V' = (v'_1, v'_2, ..., v'_m)$ be the 2 bases of B.

Assume $P = P_{U'U}$ [1] and Q = $P_{V'V}$ be a change of basis of matrix from U to U' and V to V' respectively.

Now, for some linear map $f : A \to B$, let $M(f) = M_{U,V}(f)$ and $M'(f) = M_{U'V'(f)}$ be the matrix related to f with respect to the bases U, V and U', V' respectively. We get:

$$M'(f) = Q^{-1}M(f)P$$

or simply-

$$M_{U'V'}(\mathrm{f}) = \mathrm{P}^{-1}{}_{V'V}\ M_{U,V}(\mathrm{f})\ P_{U'U}$$

$$= P_{V,V}'\ M_{U,V}(\mathrm{f})\ P_{U'U}$$

Definition: Two n × n matrices E and F are known as similar if and only if there is some invertible matrix P so that-

$$\mathrm{F} = P^{-1}EP$$

Chapter 2

Determinants,Matrices with their Analysis

2.1 Haar bases and Determinants

2.1.1 Haar Bases, Wavelets and Hadamard Matrices

Wavelets have a significant role in compressing as well as processing audio and video signals.In this process long signals are compressed into small ones without losing enough information due to which when we listen or see it, we find no difference. The functional matrices in computer science are Haar matrices & Hadamard matrices.

2.1.2 Signal Compression through Haar Wavelets

Lets discuss it with the help of an example. first consider haar wavelet from R^4.Consider four vectors which are pairwise orthogonal.Thus we can recall that they are independent as well.

$$U_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, U_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, U_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, U_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

Now $U = u_1, u_2, u_3, u_4$ is the Haar basis and also let us assume $W = w_1, w_2, w_3, w_4$ is basis of R^4 which is canonical. Matrix change ,i.e, from W to v is determined by-

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

Also the inverse that can be calculated easily -

$$u^{-1} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Take an arbitrary vector $b = (6, 4, 9, 3)$ over the basis W that becomes $d = (d_1, d_2, d_3, d_4)$ over the Haar basis V

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \\ 9 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \\ 2 \\ 1 \end{pmatrix}$$

We were given a signal b, first of all we tranformed it into its coefficients $d = (d_1, d_2, d_3, d_4)$ over Haar basis by computing $d = V^{-1}b$.

$$d_1 = \frac{(b_1 + b_2 + b_3 + _4)}{4}$$

This is the calculated mean value of the signal given.d_1 is the image background part or of sound.d_2 gives rough details of b ,d_3 corresponds to [2] details in the 1st part of b and d_4 yields the details in 2nd half of b.

An excellent compression is that in which we set the values of some coefficients equal to 0,denoting the compressed signal by $\hat{b}$ and then also the crucial information is not lost due to which the reconstructed signal equally effective as the original one was. Thus the reconstructed signal $\hat{b}$ is given by $\hat{b} = V\hat{d}$ Such compression is used in modern day video conferences.

2.1.3 Characterisitcs of Haar Wavelets

Haar wavelet is used for its scaling property.Such method can be used for any signal of length 2^n. Now consider $n = 3$.The Haar basis is given by the matrix-

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

The columnn of this matrix are orthogonal and also

$$U^{-1} = diag(1/8, 1/8, 1/4, 1/4, 1/2, 1/2, 1/2, 1/2, V^T)$$

Here we can see a pattern developing .It seems like Haar basis vector generated all the other basis vectors excluding v_1 which is there to determine the average.In general,

$$u_2 = (1, \cdots, 1, -1, \cdots, -1)$$

Similarly other vectors can be deduced by shifting and scaling of Haar basis.

2.1.4 Construction of Haar Matrices using Kronecker Product

This method for construction of Haar matrix V_n of dimension 2^n will clear why we are emphasising so much on pairwise orthogonal matrices.If we divide V_n into 2 matrix,i.e, $2^k \times 2^{k-1}$, so $2nd$ matrix having last 2^{n-1} column of V_k has quite ordinary structure, which is made up of vector,

$$(1, -1, 0, \cdots, 0)$$

$2^{k-1} - 1$ shifted copies of it. The matrix manipulation can be done by product operation as well which is known as Kronecker product.

2.1.5 Analysis of Multiresolution Signal using Haar Bases

One more marvellous feature of Haar bases is multiresolution signal analysis.Let us consider we are given a signal s, if $s = (s_1, s_2, \cdots s_{2^n})$ is Haar coefficient vector.Here coefficient with lesser value represents rough information about s, coefficient with more value gives minute details about the signal,e.g,if s is an audio signal corresponding to Diljit Dosanjh concert, then s_1 corresponds to background noise, s_2 corresponds to bass, s_3 corresponds to first cello, s_4 corresponds to second cello, and the rest coefficients corresponds to the other instrument sounds created during the concert.The multiresolution is used where we set some coefficients equal to zero,i.e, lesser coefficients represent the signal.

2.1.6 Digital Images using Haar Transform

One more characteristics of Haar transform is we can apply it to matrices with no extra effort, due to which compression of digital image is possible.Before that let us deal with the issue of normalization first. We have noticed that the Haar basis have orthogonal column vectors, but these column does not have length 1 as final result.

2.1.7 Hadamard Matrices

One of the known family of matrices which is like Haar matrices are Hadamard , the only difference is they have entries as $+1$ and -1 A real $m \times m$ matrix is Hadamard if $v_{jk} = +1, -1$ for

each j, k such that $1 \leq j,\ k \leq m$ and if

$$V^T V = mI_m$$

. [2]So the Hadamard columns are pairwise orthogonal because H is square matrix. According to the above equation H is invertible. Some examples of Hadamard matrix are as follows:-

$$V_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, V_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

A basic question that comes to our mind is that to determine the positive integer value of m using which a matrix of Hadamard type having dimension m exists, which still remains an open question.The speculation of Hadamard is that for each +ve value having the form $m = 4p$, we have a Hadamard matrix with dimension m. Sylvester presented a collection of [1] Hadamard matrices and proved the existence of Hadamard matrices having dimension 2^n for all $n \leq 1$ by utilizing the theorem mentioned below-

Proposition- Let V be a Hadamard matrix of n dimension, then we have a block matrix

$\begin{pmatrix} V & V \\ V & -V \end{pmatrix}$ of dimension $2m$, which is a Hadamard matrix as well.

If we begin with V_2, we get a infinite dimension collection of Hadamard matrices which are symmetric in nature, commonly known as Sylvester Hadamard matrices that can be written as H_{2^m}. The examples of such matrix are V_2,V_4 etc which are shown above. There are various applications of Hadamard matrices like error correcting codes, numerical and linear algebra and signal processing as well.

2.2 Determinants

*D*eterminant can be defined as alternating multi linear maps having value 1 on unit matrix.Irrespective of the approach we follow , we need certain preliminaries about permutation on a finite set It can be defined as
$G : M_n(K) \to K$, i.e. mapping of n columns in domain matrix is alternating in which $G(I_n) = 1$ for identity matrix I_n.

2.2.1 Inverse Matrices and Determinants

Let us consider a matrix $Q \in M_n(K)$ where K is a commutative ring then $Q = (c_{jk})$ is a matrix defined as $c_{jk} = (-1)^{(j+k)} det(R_{kj})$ where c_{jk} is cofactor of a_{kj}. Each matrix Q_{kj} is called a minor of matrix Q.It is nothing but the transpose of cofactors of elements of Q. Here,we have a

result deduced by using the above fact which is stated below: Let the commutative ring be K then for some matrix R we have

$$R\tilde{R} = \tilde{R}R = det(R)I_n$$

$$\Rightarrow RR^T = R^TR = Det(R)I_n$$

from this, R can be concluded as invertible if Det (R) is invertible if so $R^{-1} = (Det(R))^{-1}\tilde{R}$

2.2.2 Determinant of Linear Mapping

Let us consider P, a vector space with finite dimension m , whose basis can be represented as $(u_1, ..., u_m)$ then for each map $f : P \to P$, [2] if $M(f)$ is a matrix w.r.t. given basis $(u_1, ..., u_m)$, then we have $det(f) = det(M(f))$ [4]. Additionally if we consider $(x_1, ..., x_n)$ as [2] some other basis of P and if Q is any change of basis, then the matrix of f w.r.t. basis $(x_1, ..., x_n)$ is $Q^-1M(f)Q$

we know that $det(BC) = det(B).det(C)$.

Using this rule, we have-

$$det(Q^{-1}M(f)Q) = det(Q^{-1}).det(M(f)).det(Q)$$

$$\Rightarrow det(Q^{-1}).det(Q).det(M(f))$$

$$\Rightarrow det(M(f))$$

So here we can conclude that det(f) is independent to basis of P.
A very crucial result deduced from this is stated below:
For [1] any vector space Q of finite dimension n, a linear map $f : Q \to Q$ is invertible if $det(f)$ is non zero.

2.2.3 The Cayley-Hamilton Theorem

We are all well known with the concept of characteristic polynomial and here we consider any commutative ring K .

[1] This theorem states that for each $m \times m$ matrix $B \in M_m(K)$, the characteristic polynomial $Q_B(X)$ of B is the determinant such that-

$$Q_B(X) = Det(XI - B)$$

.

2.2.4 Permanents

Permanent is a multilinear symmetric form.They don't have any natural geometric interpretation like determinants have.Infact, the computation of permanents is very typical as compared to determinants.They have many conjuctional interpretations and one of them is in terms of perfect matching of bipartite graphs.

Now we need to understand bipartite graphs first in order to understand permanents.

A bipartite graph is a undirected graph $H = (U, E)$ [2] whose set of nodes can be partitioned into two non empty disjoint sets U_1 and U_2 such that every edge $e \in E$ has one end point in U_1 and the other in U_2

2.3 LU, Cholesky, Echelon Form, Gaussian Elimination

Curve interpolation is another important problem faced in robotics as well as computer graphics.A way to tackle this problem is through usage of cubic splines to find the solution.They [1] consist of cubic Bezier curves,they are used as it gives more flexibility.A Bezier curve is specified by 4 control points (c_0, c_1, c_2, c_3) and is yielded in paramteric form using the given equation

$$B(x) = (1-x)^3 c_0 + 3(1-x)^2 x c_1 + 3(1-x)x^2 c_2 + x^3 c_3$$

Here we have $B(0) = c_0, B(1) = c_3$ and for $x \in [0, 1]$,the point $B(x) \in$ convex hull of c_0, c_1, c_2, c_3 and the polynomials involved are known as Bernstein polynomials of degree 3.The location of control points rigorously affects curve shape .The cases of self intersection of curve can also be seen. In interpolation problems we find curves traverse to some points(given), which satisfies certain conditions additionally.

A bezier spline curve F which consists of, say m curve fragment known as bezier curve namely $C_1, C_2, ...C_m where (m \geqslant 2)$.Also we need some smoothness between any two junction points,i.e, $C^0 - continuity$ and $C^2 - continuity$ are necessary which actually [1] ensures that the tangents as well as curvatures agree. In order to solve cubic spline we come across system of linear equations which needs to be solved . So now we will discuss the most efficient methods to solve such linear systems.

2.3.1 Gauss Elimination

[1] Let C be $n \times n$ invertible matrix and $D \in R^n$ be n dimensional.We are aiming to solve the linear system $Cx = D$.As we know C is invertible thus the system yields a unique solution $x = C^{-1}D$.The following facts can be stated from this which are as follows:

(1) We can avoid to compute C^{-1} explicitly.because it would lead to solving n linear systems.

(2) We don't need to solve linear systems by computing determinants as this specific methodology needs a certain number of addition and respective multiplications also that are in proportion to $(n+1)!$(resp $(n+2)!$.

The primary idea of gauss elimination method is-if C is [2] upper triangular which means $c_{ij} = 0$ for $1 \leq j < i \leq n$ (resp lower triangular which means that $b_{ij} = 0$ for $1 \leq i < j \leq n$, then computing x is trivial.

$$\begin{pmatrix} b_{11} & b_{12}\cdots & b_{1n-2} & b_{1n-1} & b_{1n} \\ 0 & b_{22}\cdots & b_{2n-2} & b_{2n-1} & b_{2n} \\ 0 & 0\ddots & \vdots & \vdots & \vdots \\ & \ddots & \vdots & \vdots & \vdots \\ 0 & 0\cdots & 0 & b_{m-1m-1} & b_{m-1m} \\ 0 & 0\cdots & 0 & 0 & b_{nn} \end{pmatrix}$$

Let the above matrix be C. $Det(C) = b_{11}b_{22}\cdots b_{nn}$ which means $b_{ii} \neq 0$ [2] for $i = 1, 2, \cdots n$ and now we can solve the system $Cx = D$ using bottom up approach by method of backward substitution.Also for lower triangular matrix we solve it by forward substitution from top down approach.

Theorem: Let b [2] be a $n \times n$ matrix(invertible or non-invertible), then there is a invertible matrix M such that $U = MA$ is a upper triangular.The pivots are all nonzero iff A is invertible [1] .

2.3.2 LU-Factorization

Suppose C is an invertible matrix then if we can write C in the form $C = LU$, where U is upper-triangular L is lower-triangular s.t. $L_{ii} = 1$ for $i = 1, \cdots n$. We have a $n \times n$ matrix $C = (c_{ij})$ for any k s.t. $1 \leq k \leq n$,let D(1:k,1:k) be submatrix of C whose entries are (c_{ij}), where $1 \leq i, j \leq k$.e.g. if C is a 4×4 matrix

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{pmatrix} \text{ then } D(1:2, 1:2) = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$$

Theorem-Let C be a $m \times m$ matrix then C has LU-factorization $C = LU$ if and only if $D(1:k, 1:k)$ [2] is invertible for $k = 1, \cdots n$. Additionally, when C has LU-factorization, we have

$$det(D(1:k, 1:k)) = \pi_1\pi_2\pi_3\cdots\pi_k \text{ for } k = 1, 2\cdots, n$$

where π_k is obtained pivot after k-1 no. of elimination steps. So we can get kth pivot by

$$\pi_k = \begin{cases} c_{11} = det(D(1:1,1:1)) & \text{for} \quad k = 1 \\ \frac{det(D(1:k,1:k))}{det(D(1:k-1,1:k-1))} & \text{for} \quad k = 2, \cdots n \end{cases}$$

Corollary- Suppose C as [2] an invertible $n \times n$ matrix. [1] If every matrix C(1:k,1:k) is invertible for $k = 1, \cdots n$ then Gauss elimination method do not require any pivoting to get the LU- factorization $C = LU$.

2.3.3 SPD Matrices and Cholesky Decomposition

Positive-Definite A $m \times m$ real symmetric matrix B is positive-definite iff $x^T Bx > 0$ for each $x \epsilon R^n$ with $x \neq 0$ Also B is said to be symmetric +ve definite iff its each eigen value is strictly positive. The following facts can be deduced using this information-

(1) The matrix B will be invertible.

(2) For each $m \times m$ real as well as invertible matrix X , we have the matrix $X^T BX$ as real, symmetric and +ve [1] definite iff B is real symmetric positive definite.

(3) We have $\beta_{ii} > 0$ [1] for $i = 1 \cdots n$ (Notice that for $x = e_i$, the ith canonical basis vector of R_n we get, ${e_i}^T Be_i = \beta ii > 0$

(4) The set $m \times m$ real symmetric and +ve definite matrix will be convex, which means if B and c are two $m \times m$ symmetric and +ve definite matrix then for any arbitrary $\lambda \in R$ where $0 \leq \lambda \leq 1$ the matrix $(1 - \lambda B) + \lambda C$ is also +ve definite and symmetric. For some $x \in R^n$ where x is non-zero, we get $x^T Bx > 0$ and $x^T Cx > 0$, so

$x^T((1-\lambda)B + \lambda C)x = (1-\lambda)x^T Bx + \lambda x^T Cx > 0$ as $0 \leq \lambda \leq 1 \implies$ (1-λ) ≥ 0 and also $\lambda \geq 0$. It should be noted that λ and $(1 - \lambda)$ cannot be 0 .
(5) The set of $m \times m$ +ve definite and real symmetric matrix is a conical shape which means that if is positive and real B is symmetric as well as positive definite then B is also symmetric and positive definite.

Cholesky Factorization- Consider a real, symmetric as well as positive definite matrix B then - $\exists$ a lower triangular matrix A such that $B = AA^T$.Also A must be chosen in such a way that its diagonal entries are strictly positive so that A is unique. We can verify the unique characteristic of Cholesky decomposition by LU-decomposition.

Theorem- Let X be a real symmetric matrix the following conditions implies the same thing-

(a) X is +ve definite.

(b) Each principal minor of X is positive,i.e, $Det(X(1:k,1:k)) = det(Q)^2 > 0 \implies X$ is positive definite.

(c) X has LU-factorization and each pivots is non-negative.

(d) X has $LDL^T - factorization$ and each pivots is non-negative.

Chapter 3

Vector norms and Euclidean Spaces

3.1 Vector and Matrix Norms

It outlines the definition and key properties of a norm on a vector space E over either the real numbers ($\mathbb{R}$) or complex numbers ($\mathbb{C}$). A norm is a function that assigns a non-negative real number to each vector and satisfies three conditions: positivity, homogeneity and the triangle inequality.

It establishes a relation between these 3 properties, particularly focusing on the triangle inequality. It concludes that in a normed vector space E, the absolute difference between the norms of two vectors x and y is always less than or equal to the norm of their difference: $|\|x\| - \|y\|| \leq \|x - y\|$ for all $x, y \in E$. This inequality denotes the closeness of vectors in terms of their norms within the vector space.

In the context where E represents either the vector space of complex numbers ($\mathbb{C}^n$) or real numbers ($\mathbb{R}^n$), it can be stated that for any real number $p \geq 1$, the p-norm qualifies as a valid norm [2].

3.1.1 The Normed Vector Spaces

[1]In order to define the closeness of 2 vectors or 2 matrices and also to define the order of convergence of sequence of vectors or matrices, we can use norm.

Definition: Let F be a vector space over the field K, where K is either the field $\mathbb{C}$ of complex numbers or the field $\mathbb{R}$ of real numbers. A **norm** on F is a function $\|\cdot\| : F \to \mathbb{R}_+$ that assigns a non-negative real number $\|u\|$ to any vector $u \in F$. This function must satisfy [1]the following conditions for all $x, y \in F$ and $\lambda \in K$ [1]:

1. **Positivity:**
$$\|x\| \geq 0 \quad \text{and} \quad \|x\| = 0 \iff x = 0$$
[2]

2. **Homogeneity (or Scaling):**
$$\|\lambda x\| = |\lambda| \|x\|$$

3. **Triangle Inequality:**
$$\|x + y\| \leq \|x\| + \|y\|$$

A vector space F equipped with a norm $||\ ||$ is called a **normed vector space**.

[1]**Examples:** 1) Let F= $\mathbb{R}$ and $||b|| = |b|$, is the absolute value of b.

2) Let F= $\mathbb{C}$ and $||k|| = |k|$, is the modulus of k.

3) [1]Let F = R^n (orC^n), [1]then there are 3 standard norms for every $(x_1, ..., x_n)$ ϵ F. we have a norm which is defined as :

$$||x||_1 = |x_1| + ... + |x_n|$$

then we have the Euclidean norm, defined such that:
$$||x||_2 = (|x_1|^2 + ... + |x_n|^2)^{\ 1/2}$$

and the l^p-norm (for p$\geq$ 1)*by* :
$$||x||_p = (\ |x_1|^p + ... + |x_n|^p)^{\ 1/p}$$

at last, we define the sup-norm as:
$$||x||_\infty = \max \{|x_i| | 1 \leq i \leq n\}$$

Corollary: (Holder's Inequalities) For any of real numbers p, q $\geq$ 1 and

$$\tfrac{1}{p} + \tfrac{1}{q} = 1,$$

with q = $+\infty$ if p = 1 and p = $+\infty$ if q=1, we have the inequalities

$$\textstyle\sum_{i=\ 1}^{n} |u_i v_i| \leq (\sum_{i=\ 1}^{n} |u_i|^p)^{1/p} \ (\sum_{i=\ 1}^{n} |v_i|^q)^{1/q}$$ [2]

and

$$| < u, v > | \leq ||u||_p \ ||v||_q$$

Now, for p = 2 , *we have a standard Cauchy-Schwarz inequality.*

For l^p-norm , the triangle inequality will be known as Minkowski's inequality i.e.

$$(\sum_{i=\ 1}^{n}\ (|u_i + v_i|)^p)^{1/p} \leq (\sum_{i=\ 1}^{n}\ |u_i|^p)^{1/p}\ + (\sum_{i=\ 1}^{n}\ |v_i|^q)^{1/q}$$

Remark: For $0 < p < 1$ [2], the function $x \mapsto \|x\|_p$ does not qualify as a norm because it does not satisfy the triangle inequality. To illustrate this, [1] consider the vectors $x = (2, 0)$ and $y = (0, 2)$. [1]Their sum is $x + y = (2, 2)$, and we have: [1]

$$\|x\|_p = 2, \quad \|y\|_p = 2.$$

Thus, we get: [1]

$$\|x + y\|_p = 2^{\frac{p+1}{p}}.$$

Therefore, [1]

$$\|x + y\|_p = 2^{\frac{p+1}{p}} \quad \text{and} \quad \|x\|_p + \|y\|_p = 4.$$

For $0 < p < 1$, it holds that $2p < p + 1$, which implies $\frac{p+1}{p} > 2$, leading to $2^{\frac{p+1}{p}} > 4$. This demonstrates that the inequality $\|x + y\|_p \leq \|x\|_p + \|y\|_p$ does not hold.

Definition: In a [2] real or complex vector space F, two norms $\|\|_u$ and $\|\|_v$ are considered **equivalent** if there exist positive constants c_1 and c_2 such that:

$$\|a\|_u \leq c_1\|a\|_v \quad \text{and} \quad \|a\|_v \leq c_2\|a\|_u, \quad \text{for all } a \in F.$$

3.1.2 The Matrix Norms

Definition: A norm $\|\|$ defined on the set of square matrices $n \times n$ within $M_n(K)$, where K can be either $\mathbb{R}$ or $\mathbb{C}$, is termed a matrix norm. [1] It possesses an additional property known as submultiplicativity, [1] denoted as $\|AB\| \leq \|A\|\|B\|$ for every matrices $A, B \in M_n(K)$ [1]. A matrix norm that satisfies this property is referred to as a submultiplicative matrix norm.
For any matrix $A = (a_{ij})$ in $M_{m,n}(\mathbb{C})$ [2]:

1. The conjugate of A, denoted as $\overline{A}$, is defined element-wise as $\overline{A_{ij}} = \overline{a_{ij}}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$.

2. The transpose of A, denoted as A^T, is obtained by interchanging the rows and columns, resulting in a $n \times m$ matrix with elements $(A^T)_{ij} = a_{ji}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$ [1].

3. The adjoint of A, denoted as A^*, is the transpose of the conjugate, i.e., $A^* = \overline{(A^T)} = \overline{(A)}^T$. For real matrices, A^* equals the transpose, $A^* = A^T$.

4. A is termed Hermitian if its adjoint equals itself, i.e., $A^* = A$.

5. A is considered symmetric if its transpose equals itself, i.e., $A^T = A$.

6. A is termed normal if it commutes with its adjoint, i.e., $AA^T = A^TA$.

7. A unitary matrix $U \in M_n(\mathbb{C})$ satisfies $UU^* = U^*U = I$.

8. An orthogonal matrix $Q \in M_n(\mathbb{R})$ satisfies $QQ^T = Q^TQ = I$.

9. [2]The trace tr(A) of a matrix A is the sum of its diagonal elements, given by tr$(A) = a_{11} + \ldots + a_{nn}$ [2].

To demonstrate linearity of the trace of a matrix A, we need to verify the following two properties:

1. tr$(\lambda A) = \lambda$tr(A)
2. tr$(A + B) =$ tr$(A) +$ tr(B)

Definition: Any square matrix A ϵ $M_n(\mathbb{C})$, a complex number $\lambda \epsilon \mathbb{C}$ is said to be an eigenvalue of A if there is some non-zero vector u ϵ $\mathbb{C}^n$,

$$\text{Au} = \lambda u$$

[2] If λ is an eigenvalue of a matrix A, then the non-zero vectors u ϵ $\mathbb{C}^n$ such that $Au = \lambda u$ are called the eigenvectors of A associated w.r.t. λ, along with the zero-vector these eigenvectors form a subspace of $\mathbb{C}^n$ denoted by $E_\lambda(A)$ and called the eigenspace associated with λ [1] .

Definition: Any square n $\times$ n matrix A ϵ $M_n(\mathbb{C})$, the polynomial

$$\det(\lambda \text{I - A}) = \lambda^n \text{ - tr(A)}\lambda^{n-1} + \ldots\ldots + (-1)^n det(A)$$

[1] is called *the characteristic polynomial of A*. Here, n (not necessarily distinct) roots $\lambda_1, \ldots\ldots, \lambda_n$ of the characteristic polynomial are all eigenvalues of A and constitute the spectrum of matrix A.

We let : $\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$
be the largest modulus of eigenvalues of A which is called the *spectral radius of A*.

Note: $tr(A) = \lambda_1 + \lambda_2 + \ldots\ldots + \lambda_n$ and det(A)= $\lambda_1\lambda_2\ldots\ldots\lambda_n$

Definition: The *Frobenius norm (denoted by* $||\ ||_F$*)* is defined that for every sqaure n $\times$ n matrix $A \epsilon M_n(\mathbb{C})$,

$$||A||_F = (\sum_{i,j=\ 1}^{n} |a_{ij}|^2)^{1/2} = \sqrt{tr(AA^*)} = \sqrt{tr(A^*A)}$$

The *Frobenius norm* is a matrix norm on $M_n(\mathbb{C})$ which satisfies the following properties:

1- $||\ ||_F$ is matrix norm i.e. $||AB||_F \leq ||A||_F\ ||B||_F$, for all A,B ϵ $M_n(\mathbb{C})$

2- $||\ ||_F$ is unitarily invariant which means that for all the unitary matrices U,V: we have -

$$||A||_F = ||UA||_F = ||AV||_F = ||UAV||_F$$

3- $\sqrt{\rho(A^*A)} \leq ||A||_F \leq \sqrt{n}\ \sqrt{\rho(A^*A)}$, for all A ϵ $M_n(\mathbb{C})$

3.1.3 The Subordinate Norms

Definition: If [2] any norm (|| ||) on $\mathbb{C}^n$, we will define the function $||\ ||_{op}$ on $M_n(\mathbb{C})$ by:

$$||A||_{op} = \sup_{x\epsilon\mathbb{C}^n, x\neq 0} \frac{||Ax||}{||x||}$$

$$= \sup_{x\epsilon\mathbb{C}^n, ||x||=1} ||Ax||$$

The function A$\rightarrow ||A||_{op}$ is said to be a subordinate matrix norm or operator norm induced by the || ||.

We can easily check that the fuction A$\rightarrow ||A||_{op}$ is indeed a norm as by definition it will satisfy the property:

$$||Ax|| \leq ||A||_{op}\ ||x||\ \text{, for all x } \epsilon\ \mathbb{C}^n.$$

If the [2] norm $||\ ||_{op}$ on $M_n(\mathbb{C})$ satisfying this above property then it is said to be subordinate to the vector norm || || on $\mathbb{C}^n$. [2] As a result of above inequality, we have:

$$||ABx|| \leq ||A||_{op}\ ||Bx|| \leq ||A||_{op}\ ||B||_{op}\ ||x||,$$

for all x $\epsilon\ \mathbb{C}^n$ which implies that [2]

$$||AB||_{op} \leq ||A||_{op}\ ||B||_{op} \text{ for all A,B } \epsilon\ M_n(\mathbb{C}).$$

[2]showing that A$\rightarrow ||A||_{op}$ is a matrix norm (*which is submultiplicative*).

[2]The operator norm is also defined by :

$$||A||_{op}= \inf \{\ \lambda\ \epsilon\ \mathrm{R}\ |\ ||Ax|| \leq \lambda\ ||x||, \text{ for all x } \epsilon\ \mathbb{C}^n\ \}.$$

Hence, the function x$\rightarrow ||Ax||$ is continous as

$$||Ay||\ \text{-}\ ||Ax|| \leq ||Ay - Ax|| \leq C_A||x - y||$$

and the unit sphere $S^{n-1} = \{x\ \epsilon\ \mathbb{C}^n\ |\ ||x|| = 1\ \}$ *is compact,* there is some x $\epsilon\mathbb{C}^n$

$$\text{and } ||Ax|| = ||A||_{op}$$

Equivalently, there is some of x $\epsilon\ \mathbb{C}^n$ such that x $\neq 0$ and

$$||Ax|| = ||A||_{op}\ ||x||\ [1].$$

The definition of an operator norm also implies that: $||I||_{op} = 1$

Definition: [1] If K = $\mathbb{R}$ or $\mathbb{C}$, for any norms (|| ||) on $M_{m,n}(K)$ and for any 2 norms $||\ ||_a$ and $||\ ||_b$ on K^n and K^m respectively, we also say that the norm || || is subordinate to the norms $||\ ||_a$ and $||\ ||_b$ if

$$||Ax||_b \leq ||A||\ ||x||_a \text{ , for all A } \epsilon\ M_{m,n}(\text{K}) \text{ and x } \epsilon\ \text{k}^n.$$

Definition: For A = $(a_{ij})\ \epsilon\ M_n(C)$, [1]the norm $||A||_2$ is said to be a *spectral norm.*

Here, using the property : [1]

$$\sqrt{\rho(A^*A)} \leq ||A||_F \leq \sqrt{n}\ \sqrt{\rho(A^*A)}, \text{ for all A } \epsilon\ M_n(C)$$

we can say that $||A||_2 \leq ||A||_F \leq \sqrt{n}\ ||A||_2$ which proves that the Frobenius norm will be an upper bound on the spectral norm and much easier to compute than the spectral norm [1].

3.1.4 Inequalities Involving Subordinate Norms

Proposition: [2]Let || || be any matrix norm and let A $\epsilon\ M_n(\mathbb{C})$ such that $||A||<1$

1) If the norm (|| ||) is a subordinate matrix norm, then the matrix I + A is invertible and also defined as:

$$||(I+A)^{-1}|| \leq \frac{1}{1-||A||}$$

Proof: We see that (I+A)v = 0 which implies that Av = -v so,

$$||v|| = ||Av||$$

using, $||Av|| \leq ||A||\ ||v||$

for every subordinate norm. Hence, $||A|| < 1$, if v $\neq$ 0, thus

$$||Av|| < ||v||$$

which contradicts $||v|| = ||Av||$.

Therefore, [3] we have v = 0 which proves that the I + A is injective and bijective i.e. invertible here.

Then, [3] we have: $(I+A)^{-1} + A(I+A)^{-1} = (I+A)(I+A)^{-1} = \text{I}$

so we get , [3] $(I+A)^{-1} = \text{I - } A(I+A)^{-1}$

which shows : [3]

$$||(I+A)^{-1}|| \leq 1 + ||A||\ ||(I+A)^{-1}||$$

which finally proves that : [3]

$$||(I+A)^{-1}|| \leq \frac{1}{1-||A||}$$

2) If the matrix of the form I + A will be singular, then the $||A|| \geq 1$ for every matrix norm (not necessarily a subordinate norm).

Proof: If the matrix I+A will be singular, then -1 will be an eigenvalue of A and using preposition ([2] Any matrix norm || || on a $M_n(\mathbb{C})$ and for any square n X n matrix Aϵ $M_n(\mathbb{C})$ such that: $\rho(A) \leq ||A||$)

$$\text{which implies that, } 1 \leq \rho(A) \leq ||A||$$

This [5] result is used to deal with the convergence of the sequences of powers of the matrices,

Proposition: For any matrix A ϵ $M_n(\mathbb{C})$ and for every $\epsilon > 0$, there will be some subordinate matrix norm such that

$$||A|| \leq \rho(A) + \epsilon$$

3.2 The Dual Space with their Duality

3.2.1 Linear Forms and The Dual Space E^*

Definition: Let E be a vector space, the Hom(E,K) of linear maps from E to the field is called *the dual space(or dual) of E.* The space (E,K) denoted by E^* are called *the linear forms*, or covectors. [2]

The dual space E^{**} of the space E^* is called the *bidual of E* [1].

Definition: Given any vector space E and any basis $(u_i)_{i\epsilon I}$ for E, for every i ϵ I, there is a unique linear form u_i^* such that

$$u_i^*(u_j) = 1 \text{ if i=j and 0 if i}\neq\text{j}$$

for every j ϵ I. The linear form u_i^* is called the *coordinate form of index i* w.r.t. the basis $(u_i)_{i\epsilon I}$.

Proposition: [2] Let $(u_1, u_2, ..., u_n)$ and $(v_1, v_2, ..., v_n)$ be any two bases of E and let P = (a_{ij}) be the change of basis matrix from $(u_1, u_2, ..., u_n)$ to $(v_1, v_2, ..., v_n)$ so that

$$v_j = \sum_{i=1}^{n} a_{ij} u_i$$

Then, the change of the basis from dual basis $(u_1^*, ..., u_n^*)$ to the dual basis $(v_1^*, ..., v_n^*)$ is $(P^{-1})^T$ and for any linear form ψ, the new coordinates $\psi_j^{'}$ of ψ are expressed in terms of the old coordinates ψ_i of ψ using the matrix P^T i.e.

$$(\psi_1^{'},.....,\psi_n^{'}) = (\psi_1,.....,\psi_n)\text{P}$$

3.2.2 Duality and Pairing Between E and E^*

Here a linear form $u^* \epsilon E^*$ and a vector v ϵ E, the result $u^*(v)$ of applying u^* to v is also denoted by $< u^*, v >$. This defines a binary operation $< -, - >: E^* \times \mathrm{E} \to \mathrm{K}$ satisfying the following properties:

$$< u_1^* + u_2^*, v > = < u_1^*, v > + < u_2^*, v >$$
$$< u^*, v_1 + v_2 > = < u^*, v_1 > + < u^*, v_2 >$$
$$< \lambda u^*, v > = \lambda < u^*, v >$$
$$< u^*, \lambda v > = \lambda < u^*, v >$$

These above identities means that $< -, - >$ [2] is a *bilinear map*, as it is linear in each argument. It is also often called the *cannonical pairing* between E^* and E.

With the help of above identities, given any fixed vector v ϵ E, the map $eval_v: E^* \to \mathrm{K}$ (i.e. evaluation at v) defined such that:

$$eval_v(u^*) = < u^*.v > = u^*(v) \text{ for every } u^* \epsilon E^*$$

is a linear map from E^* to K,i.e. $eval_v$ is a linear form in E^{**}.

Again using the above identities, the map $eval_E : \mathrm{E} \to E^{**}$ defined such that

$$eval_E(v) = eval_v \text{ for every v } \epsilon \text{ E},$$

is a linear map. We see that

$$eval_E(v)(u^*) = eval_v(u^*) = < u^*, v > = u^*(v), \text{ for all v } \epsilon \text{ E and all } u^* \epsilon E^*$$

Definition: Given any [2] vector space E and its dual E^*, we also say that a vector v ϵ E and a linear form $u^* \epsilon E^*$ are *orthogonal* iff $< u^*, v > = 0$.

Theorem: [3]*(Duality Theorem)* Let E be a vector space of dimension n and the following properties hold:

1) For every basis $(u_1, ..., u_n)$ of E, the family of coordinates forms $(u_1^*, ..., u_n^*)$ is a basis of E^* (i.e. the dual basis of $(u_1, ..., u_n)$).

2) [3] For every subspace V of E, we have $V^{00} = \mathrm{V}$ [3].

3) [1]For every pair of subspaces V and W of E such that E = V $\bigoplus$ W, with V of dimension m, for every basis $(u_1, ..., u_n)$ of E such that $(u_1, ..., u_m)$ is a basis of V and $(u_{m+1}, ..., u_n)$ is a basis of W, [3]the family $(u_1^*, ..., u_m^*)$ is a basis of the orthogonal W^0 of W in E^*, so that [3]

$$\dim(\mathrm{W}) + \dim(W^0) = \dim(\mathrm{E}).$$

Further, [3]we have $W^{00} = \mathrm{W}$.

4) [3]For every subspace U of E^*, we have
[3]

$$\dim(\mathrm{U}) + \dim(U^0) = \dim(\mathrm{E}).$$

where U^0 is the orthogonal of U in E and $U^{00} = \mathrm{U}$. [3]

3.2.3 Transpose of Matrix and Linear Map

Definition: Given a linear map f : E $\to$ F, the transpose $f^T : F^* \to E^*$ of f is the linear map defined such that

$$f^T(v^*) = v^* o f, \text{ for every } v^* \epsilon F^*.$$

Equivalently, the linear map $f^T : F^* \to E^*$ is defined such that

$$< v^*, f(u) >=< f^T(v^*), u >$$

for all $u \in E$ and all $v^* \in F^*$ [2].

It is very easy to verify that the following properties hold:

$$(f + g)^T = f^T + g^T$$

$$(\mathrm{g\ o\ f})^T = f^T \text{ o } g^T$$

$$id_E^T = id_E$$

The property $(\mathrm{g\ o\ f})^T = f^T$ o g^T implies the following proposition:

[2] If f : E $\to$ F is any linear map, then the following properties hold:

1) If f is *injective*, then f^T is *surjective.*

2) If f is *surjective*, then f^T is *injective* [1].

Theorem: Given a linear map f: E $\to$ F, the following properties hold [1].

a) The dual $(Imf)^*$ of Im f is *isomorphic* to Im $f^T = f^T(F^*)$,i.e.

$$(Imf)^* \cong Imf^T \text{ [1].}$$

b) If F is *finite dimensional*, then rk(f) = $rk(f^T)$ [1].

Proposition: [2] If f : E → F is any linear map, then the following identities hold:

$$Imf^T = (Ker(f))^0$$

$$Ker(f^T) = (Imf)^0$$

$$Imf = (Ker(f^T))^0$$

$$Ker(f) = (Imf^T)^0$$

3.2.4 4 Fundamental Subspaces

Given any linear map $f : E \to F$ (where E and F are finite dimensional) [2].

Here, we using the preposition where any linear map $f : E \to F$, for any subspace V of E, we have

$$f(V)^0 = (f^T)^{-1}\ (V^0) = \{\ w^*\ \epsilon\ F^*\ |\ f^t(w^*)\ \epsilon\ V^0\ \}$$

As a result,

$$Kerf^T = (Imf)^0 \text{ and } Kerf = (Imf^T)^0.$$

[1]This preposition revealed that the four spaces are

$$\text{Im f, } Imf^T, \text{ [1] Ker f, } Kerf^T$$

play a special role. They are often called the *fundamental subspaces* associated with f [1].

3.3 Euclidean Spaces

The concept of Euclidean geometry stands out for its capacity to introduce crucial geometric measurements like angles, distances, and orthogonality into the realm of vector spaces. While the framework of vector spaces in linear algebra forms a fundamental basis, it lacks these essential geometric aspects. Euclidean geometry enriches this framework by incorporating a metric structure, enabling the measurement of angles, distances, and relationships between vectors.

Euclidean geometry encompasses several fundamental concepts:

1. **Inner Products and Euclidean Spaces:** Inner products extend the concept of the dot product to general vector spaces, allowing the definition of angles, lengths, and orthogonality.
2. **Rigid Motions and Transformations:** Rigid motions are transformations that maintain the distances between points. These transformations, such as rotations and reflections, hold significant importance in Euclidean geometry.

3. **Orthogonality:** Vectors in Euclidean spaces are deemed orthogonal if their inner product equals zero. This concept extends to subspaces, bases, and the creation of orthonormal bases, where vectors are both mutually orthogonal and normalized.

4. **Orthonormal Bases:** Bases in Euclidean spaces where vectors are not only linearly independent but also normalized (having unit length) and orthogonal to one another.

5. **Gram-Schmidt Procedure:** A method used to generate an orthonormal basis from a given set of vectors, ensuring independence and orthogonality among them.

6. **QR-Decomposition:** A matrix factorization into an orthogonal matrix and an upper triangular matrix, often derived using the Gram-Schmidt procedure [3].

7. [3] **Linear Isometries (Orthogonal Transformations):** Transformations that preserve inner products and distances, thus maintaining the Euclidean structure of the space.

The applications of Euclidean geometry span various disciplines such as physics, computer graphics, signal processing, and more. Its ability to model and comprehend geometric relationships in both physical and abstract spaces makes it invaluable in understanding real-world phenomena.

A comprehensive study of Euclidean geometry often involves exploring properties invariant under rigid motions, analyzing the behavior of linear transformations on Euclidean spaces, and applying these concepts in practical contexts.

Initially, we establish a Euclidean form within [2] a vector space. [1]Formally, a Euclidean form on a vector space E is established by a symmetric bilinear form possessing specific additional conditions [1]. Remember that any [2] bilinear form $\phi : E \times E \to \mathbb{R}$ is considered definite, for each $u \in E$, $u \neq 0$ suggest that $\phi(u, u) \neq 0$, and positive , for each $u \in E$, $\phi(u, u) \geq 0$.

Definition- A Euclidean space refers to a real vector space E endowed with any symmetric bilinear form $\phi : E \times E \to \mathbb{R}$ i.e. positive definite [1]. More directly, the bilinear form $\phi : E \times E \to \mathbb{R}$ agrees the following conditions:

$$\begin{aligned}
&\phi(u_1 + u_2, v) = \phi(u_1, v) + \phi(u_2, v),\\
&\phi(u, v_1 + v_2) = \phi(u, v_1) + \phi(u, v_2),\\
&\phi(\lambda u, v) = \lambda\phi(u, v),\\
&\phi(u, \lambda v) = \lambda\phi(u, v),\\
&\phi(u, v) = \phi(v, u),\\
&u \neq 0 \text{ implies that } \phi(u, u) > 0.
\end{aligned}$$

[2] The real number $\phi(u, v)$ is also referred to as the inner product (or scalar product) of u and v [2]. We explain the associated quadratic form with ϕ as the function $\Phi : E \to \mathbb{R}^+$, so that

$\Phi(u) = \phi(u, u)$, for each $u \in E$ [2]. Hence, [2] ϕ is bilinear, we deduce $\phi(0, 0) = 0$. Additionally, due to its positive definiteness, we establish the more robust certainity that $\phi(u, u) = 0$ if and only if $u = 0$, which translates to $\Phi(u) = 0$ if and only if $u = 0$.
Let any inner product $\phi : E \times E \to \mathbb{R}$ on a vector space E, alternative notations for $\phi(u, v)$ include $u \cdot v$, $\langle u, v\rangle$, or $(u|v)$, while $\Phi(u)$ can be denoted by $\|u\|$.

It illustrates the standard instance of a Euclidean space as $\mathbb{R}^n$ equipped with the inner product $\cdot$. This inner product is defined by:

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1y_1 + x_2y_2 + \ldots + x_ny_n.$$

In this context, E^n represents the Euclidean space where vectors are n-dimensional tuples in $\mathbb{R}^n$.

Example: Consider [2]the set $C[a, b]$ representing continuous functions $f : [a, b] \to \mathbb{R}$. It can be simply verified that $C[a, b]$ constitutes an infinite-dimensional vector space. For any 2 functions f and g within $C[a, b]$, explain the inner product $\langle f, g\rangle$ as follows:

$$\langle f, g\rangle = \int_a^b f(t)g(t)\, dt.$$

Demonstrating that $\langle -, -\rangle$ forms an inner product on $C[a, b]$ is a simple exercise. By considering $a = -\pi$ and $b = \pi$ (or $a = 0$ and $b = 2\pi$, which yields similar outcomes), one can put [2] $\langle \sin(px), \sin(qx)\rangle$, $\langle \sin(px), \cos(qx)\rangle$, and $\langle \cos(px), \cos(qx)\rangle$ for each natural numbers p and q greater than or equal to 1. These calculations form the basis for Fourier analysis.

Example: Let any vector space $E = M_n(\mathbb{R})$ of $n \times n$ real matrices. [1] By considering a matrix $A \in M_n(\mathbb{R})$ as a "tall" column vector attained by arranging its columns consecutively, the inner product of 2 matrices A and $B \in M_n(\mathbb{R})$ can be defined as: [1]

$$\langle A, B\rangle = \sum_{i,j=1}^{n} a_{ij}b_{ij},$$

or equivalently: [1]

$$\langle A, B\rangle = \mathrm{tr}(A^T B) = \mathrm{tr}(B^T A).$$

Hence, [2] this may be interpreted as the Euclidean product within $\mathbb{R}^{n^2}$, it establishes an inner product on $M_n(\mathbb{R})$. The relative norm $\|A\|_F$ is determined by [1]:

$$\|A\|_F = \sqrt{\mathrm{tr}(A^T A)}.$$

Proposition: [1] The bilinear form $\phi(u, v)$ is related to the quadratic form $\Phi(u)$ by the expression:

$$\phi(u, v) = \frac{1}{2}[\Phi(u + v) - \Phi(u) - \Phi(v)]$$

This relationship defines ϕ as the polar form of Φ.

Proposition: [2] In a Euclidean space E equipped with inner product ϕ and corresponding quadratic form Φ, the following inequalities hold for any $u, v \in E$:

- The Cauchy–Schwarz inequality: $|\phi(u,v)|^2 \leq \Phi(u)\Phi(v)$, where equality holds iff u and v are linearly dependent.
- The Minkowski inequality: $\sqrt{\Phi(u+v)} \leq \sqrt{\Phi(u)} + \sqrt{\Phi(v)}$, with equality if and only if u and v are linearly dependent [1]. Moreover, when $u \neq 0$ and $v \neq 0$, then $u = \lambda v$ for any $\lambda > 0$ [1].

3.3.1 Orthogonality and Duality in Euclidean Spaces

In spaces with inner products, orthogonality defines non-zero vectors that are mutually orthogonal, forming linearly independent sets termed orthogonal families. Finite-dimensional spaces always allow the discovery of orthogonal bases, crucial for efficient coordinate computations in Fourier series. For finite-dimensional E, [1] its inner product makes a natural isomorphism between E and its dual E^*, enabling an intrinsic definition of linear map adjoints. Orthonormalizing bases, particularly in finite dimensions, is feasible via two methods: duality and the constructive Gram–Schmidt orthonormalization procedure.

DEFINITION OF ORTHOGONALITY- In the realm of Euclidean spaces, the concept of vector orthogonality—where two vectors u and v are perpendicular if their dot product $u \cdot v$ equals zero—is fundamental. When we consider a collection $(u_i)_{i\in I}$ of vectors within such a space, we describe this group as orthogonal if every pair of distinct vectors u_i and u_j, where i and j belong to the set I, yields a dot product of zero, implying perpendicularity between them.

Now, if this collection of vectors $(u_i)_{i\in I}$ is termed orthonormal, it encompasses not just orthogonality but also the property that each vector has a norm (magnitude) of 1, denoted by $\|u_i\| = u_i \cdot u_i = 1$ for every vector u_i in the set.

Additionally, for any subset F of vectors within the Euclidean space E, the orthogonal complement $F^\perp$ denotes a set that contains all vectors $v \in E$, satisfying the condition that they are perpendicular to every vector in subset F, i.e., $u \cdot v = 0$ for all $u \in F$.

It's crucial to understand that while the condition $u \cdot v = 0$ for every $v \in E$ implies orthogonality, it doesn't mandatory that the vector u must be the zero vector. Instead, it signifies that u is perpendicular to all vectors within the space E.
Proposition: For given Euclidean space E, and any family $(u_i)_{i\in I}$ of non-null vectors in E, if $(u_i)_{i\in I}$ is orthogonal, then it is said to be linearly independent.

Proposition- For a given Euclidean space E, any arbitrary $u, v \in E$ are orthogonal vectors iff $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

Definition- Given [2] some vector $u \in E$, consider $\phi_u : E \to \mathbb{R}$ be the map defined so that $\phi_u(v) = u \cdot v$, for every $v \in E$.

Hence, the inner product is bilinear and the map ϕ_u is a linear form in E^*. [1]Therefore, we have a map $[\cdot] : E \to E^*$, explained so that $[u] = \phi_u$.

3.3.2 Orthogonal Transformations or Linear Isometries

Here, we explore linear transformations linking [1]Euclidean spaces that uphold the Euclidean norm. These transformations, often referred to as rigid motions, bear significant importance in the realm of geometry.

Definition- For any 2 non-trivial [1] Euclidean spaces E and F with equivalent finite dimensions n, a function $f : E \to F$ is termed an orthogonal transformation or a linear isometry if it satisfies two conditions: it is linear and conserves the Euclidean norm, i.e., $||f(u)|| = ||u||$ [2] for every $u \in E$.

Remarks:

1. [1]A linear isometry is commonly known as a linear mapping f for which $||f(v) - f(u)|| = ||v - u||$ holds true for every $u, v \in E$ [1]. Hence, the function f is linear, both definitions are equivalent [1]. The latter definition particularly emphasizes the preservation of vector distances [1].

2. Occasionally, a linear map fulfilling the criteria in above definition is termed a metric map, and a linear isometry is specifically defined as a bijective metric map.

3. [1]The dimension assumption is necessary to prove that property (3) implies property (1) when f is not assumed to be linear. However, the proof demonstrates that property (1) implies injectivity of f [1].

4. The implication that property (3) leads to property (1) holds under the assumption of surjectivity for f, even in scenarios where the vector space E has infinite dimensions.

Proposition -Let any 2 non-trivial [2] Euclidean spaces E and F of the same finite dimension n, for each function $f : E \to F$, if

$$\|f(v) - f(u)\| = \|v - u\| \quad \text{for all } u, v \in E,$$

then f be an affine map, and its related linear map g is an isometry.

3.3.3 Rodrigues Formula-

A skew-symmetric matrix A refers to a square matrix $(n \times n)$ where its transpose equals the negative of the matrix itself, denoted as $A^T = -A$. For a 3×3 skew-symmetric matrix represented as:

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$$

The exponential of a skew-symmetric matrix A can be expressed using the power series expansion:

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \frac{A^4}{4!} + \dots$$

For a 3×3 matrix, A^2 is a symmetric matrix, and A^3 will be skew-symmetric again. The computation of the exponential of a skew-symmetric matrix involves a series expansion, leading to an interesting property:

1. For any skew-symmetric matrix A, the result e^A is an orthogonal matrix [1].

2. Eigenvalues of skew-symmetric matrix are purely imaginary or zero.

There might be additional variables, like θ or matrix B, introduced possibly to simplify the expression for e^A. However, further context or equations are necessary to provide a specific formula without ambiguity or incomplete data.

3.3.4 Some Applications of Euclidean Geometry

Within matrix analysis, symmetric matrices hold significance due to their possession of real eigenvalues and their capacity to be diagonalized by orthogonal matrices. This property allows the representation of a symmetric matrix A while $A = PDP^T$, where D is a diagonal matrix and P is orthogonal.

Though not universally applicable, several decompositions involving orthogonal matrices are notably practical. The QR-decomposition, for instance, expresses any real matrix A while $A = QR$, where Q is orthogonal and R is upper triangular. Techniques like Gram-Schmidt orthonormalization or Householder matrices facilitate obtaining this decomposition.

Another essential factorization is the polar decomposition, representing a real matrix A while [2] $A = QS$, where Q and S is orthogonal and symmetric positive semidefinite respectively. This separation of stretching from rotation holds significance in fields like continuum mechanics and robotics.

[1]The singular value decomposition (SVD) states that a real matrix A can be decomposed as $A = VDU^T$, where U and V are orthogonal matrices, and D be a diagonal matrix with non-negative entries [1]. The SVD leads to the concept of pseudo-inverse, widely used in engineering applications, particularly in solving least squares problems.

Additionally, [1] the method of least squares, introduced by Gauss and Legendre in the 1800s, is a significant application of Euclidean geometry. It addresses inconsistent linear systems by minimizing the Euclidean norm $\|Ax - b\|_2$, identifying a unique solution x^+ that minimizes $\|Ax - b\|_2$. This solution x^+ satisfies the square system $A^T Ax = A^T b$.

These applications illustrate the versatile and fundamental role of Euclidean geometry across mathematical domains and practical problem-solving scenarios.

3.3.5 QR-Decomposition for Invertible Matrices

The process of Gram-Schmidt is a technique used to orthogonalize a set of vectors within an inner product space. It begins with a set of linearly independent vectors and transforms them into an orthogonal (or orthonormal) set of vectors that still span the same subspace.

Given a matrix A, applying the Gram-Schmidt process allows us to achieve a QR-decomposition. This decomposition expresses A as the product of an orthogonal matrix Q and R be an upper triangular matrix.

Here's how the Gram-Schmidt process leads to the QR-decomposition:

1. **Orthogonalization:** Initially, the process takes a set of linearly independent columns from matrix A, denoted as $\{a_1, a_2, \ldots, a_n\}$. The Gram-Schmidt process works on these columns to produce an orthogonal set of vectors $\{q_1, q_2, \ldots, q_n\}$.
2. **Normalization:** After obtaining the orthogonal vectors, the next step involves normalizing them to create an orthonormal set $\{q_1, q_2, \ldots, q_n\}$. This normalization step entails dividing each orthogonal vector by its magnitude, resulting in unit vectors.
3. **Matrix Construction:** These orthonormal vectors serve as the columns of matrix Q. Simultaneously, the coefficients derived during the orthogonalization process form an upper triangular matrix R. Essentially, R encapsulates how the original vectors were combined to obtain the orthonormal vectors.

Therefore, by implementing the Gram-Schmidt process, one can derive an orthogonal matrix Q and R be an upper triangular matrix in a method that satisfies $A = QR$ [1].

Proposition: The proposition states that for any invertible real $n \times n$ matrix A, there exists Q be an orthogonal matrix and R be an upper triangular matrix with the positive diagonal entries such that $A = QR$ [2].

This proposition aligns with the concept of QR-decomposition derived from the Gram-Schmidt process. It implies that for an invertible matrix, it's always possible to express it as the product of an orthogonal matrix and an upper triangular matrix with positive diagonal entries.

This decomposition method holds significance in various numerical algorithms, aiding in solving linear systems, least squares problems, eigenvalue computations, and other applications across mathematics and engineering domains.

Proposition (Hadamard): Let some real $n \times n$ matrix $A = (a_{ij})$, we get [2]

$$|\det(A)| \leq \prod_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

and

$$|\det(A)| \leq \prod_{j=1}^{n} \left(\sum_{i=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Moreover, equality holds if and only if either A has a zero row in the left inequality or a zero column in the right inequality, or A is orthogonal [2].

3.4 Arbitrary Matrices QR-Decomposition

3.4.1 Orthogonal Reflections

Householder matrices are crucial tools used in various numerical techniques, notably in solving linear equations, least squares problems, eigenvalue computations, and transforming symmetric matrices into tridiagonal [2] forms. They represent reflections across hyperplanes and play a pivotal role in these computational methods.

To derive the QR-decomposition of arbitrary matrices using Householder matrices [1], a simple geometric lemma can be proven. This lemma offers a direct method to express matrices in terms of these reflection matrices.

One essential application of Householder matrices lies in isometry, particularly in representing orthogonal symmetries. Before delving into these applications, it's beneficial to revisit the concept of projections.

Consider a vector space E, comprising subspaces F and G, which together [2] form a direct sum $E = F \oplus G$. Each vector u in E can be uniquely decomposed as $u = v + w$, where v belongs to F and w belongs to G. This decomposition allows the definition of two projection mappings, denoted as $p_F : E \to F$ and $p_G : E \to G$. In earlier sections, these mappings were denoted as π_1 and π_2; however, for the current context, using p_F and p_G proves more convenient.

It can be demonstrated straightforwardly that p_F and p_G are linear transformations satisfying certain properties: $p_F^2 = p_F$, $p_G^2 = p_G$, $p_F \circ p_G = p_G \circ p_F = 0$, and $p_F + p_G = \mathrm{id}$ (identity transformation) [1]. These properties are fundamental in understanding and utilizing projections within the direct sum of subspaces F and G.

Definition- Let E be any vector space and some 2 subspaces F and G that will form a direct sum $E = F \oplus G$, the symmetry (or the reflection) w.r.t. [1] F and parallel to G is the linear map $s : E \to E$ explained as

$$s(u) = 2p_F(u) - u,$$

for each $u \in E$.

As $p_F + p_G = \mathrm{id}$, also gives

$$s(u) = p_F(u) - p_G(u)$$

and

$$s(u) = u - 2p_G(u),$$

$s^2 = \mathrm{id}$, s is the identity on F, and $s = -\mathrm{id}$ on G.

Now, we let that E is an Euclidean space of finite dimension.
In Euclidean space E of dimension n, if 2 subspaces F and G form a direct sum $E = F \oplus G$ and [2] are orthogonal ($F = G^{\perp}$), the orthogonal symmetry or reflection w.r.t. [1]F and parallel to G is denoted as $s : E \to E$. This symmetry, s, is defined as: [1]

$$s(u) = 2p_F(u) - u = p_F(u) - p_G(u),$$

for all $u \in E$ [1]. Therefore, F represents a hyperplane, s is termed a hyperplane symmetry or reflection about F [1]. If G is a plane ($\dim(F) = n - 2$), s is referred to as a flip about F.

Applying the bilinearity property of the inner product to any pair of vectors $u, v \in E$, it's straightforward to establish the identity:

$$\|u + v\|^2 - \|u - v\|^2 = 4(u \cdot v). \quad (*)$$

In the specific case where $u \cdot v = 0$, it results in $\|u + v\| = \|u - v\|$. Expressing u as $u = p_F(u) + p_G(u)$ and $s(u)$ as $s(u) = p_F(u) - p_G(u)$, given the orthogonality of F and G, we deduce that $p_F(u) \cdot p_G(v) = 0$. Utilizing identity $(*)$, we derive:

$$\|s(u)\| = \|p_F(u) - p_G(u)\| = \|p_F(u) + p_G(u)\| = \|u\|,$$

illustrating [2] that s acts as an isometry.

[1] It becomes feasible to determine an orthonormal basis $(e_1, \ldots, e_n)$ for E, encompassing both an orthonormal basis for F and G. Assuming F possesses dimension p and G has dimension $n - p$, regarding the orthonormal basis $(e_1, \ldots, e_n)$, the symmetry operator s can be defined by a matrix in the form: [2]

$$\begin{pmatrix} I_p & 0 \\ 0 & -I_{n-p} \end{pmatrix}.$$

3.4.2 Using Householder Matrices for QR-Decomposition

Proposition- Given E be any non-trivial [3] Euclidean space of dimension n. For some orthonormal basis $(e_1, \ldots, e_n)$ and some n-tuple of vectors $(v_1, \ldots, v_n)$, there exists a sequence of n isometries $h_1, \ldots, h_n$ so that h_i is a hyperplane reflection or the identity [2]. If $(r_1, \ldots, r_n)$ are the vectors defined as [1]

$$r_j = h_n \circ \ldots \circ h_2 \circ h_1(v_j),$$

then each r_j can be expressed as [2] a linear combination of the vectors $(e_1, \ldots, e_j)$ for $1 \leq j \leq n$. Similarly, the matrix R whose columns are the components of r_j over the basis $(e_1, \ldots, e_n)$ is an upper triangular matrix. [1]Moreover, the h_i can be selected so that the diagonal entries of R are non-negative [2].

Remarks:

(1) [2]As each h_i is a hyperplane reflection or the identity, $\rho = h_n \circ \ldots \circ h_2 \circ h_1$ represents an isometry [2].

(2) Allowing negative diagonal entries in R enables the omission of the last isometry h_n [2].

Chapter 4

Hermitian Spaces and Spectral theorems

4.1 Hermitian Spaces

4.1.1 Hermitian Forms and Spaces, Pre-Hilbert Spaces, Sesquilinear Forms

We specialized the primary outcomes of the Euclidean geometry over the complex numbers. For instance, linear maps always have complex eigenvalues but may not have real eigenvalues. Further, some of very necessary [1] classes of linear maps can be easily diagonalized as long as they are easily expanded to the complexification of vector space (real). This case is basically for the normal and orthogonal matrices.

Take some complex number p ϵ C, if p = c + id, where c, d ϵ R. [3] Then, let c and d be the real and imaginary part respectively of p. Hence, the *conjugate* of p will be

$$\bar{p} = c - id$$

and the *absolute value* (or modulus or length) of p is $|p|$.

Recollect that $|p|^2$ = p$\bar{p}$=c^2 + d^2.

Here are countless pure cases where a map ψ : E $\times$ X $\rightarrow$ C is linear in its 1st argument with only semilinear in its 2nd argument, i.e. $\psi(c, \beta d) = \bar{\beta}\psi(c, d)$ as in opposition to the $\psi(c, \beta d) = \beta\psi(c, d)$.

Take an instance of the natural inner product to assign along the function f : R$\rightarrow$C, mostly for the [1] Fourier series is:

$$< f, h > = \int_{-\pi}^{\pi} f(x)\overline{h(x)}dx$$

which is a semilinear in function h. Hence, we overlook a consequence from the real case of an Euclidean space to the complex one and check our proofs that it will not rely on the linearity in second argument.

Definition: Given any 2 vector spaces A, B over the field C which is complex, a function $f : A \to B$ is a *semilinear* if

$$\begin{aligned} \text{f(c+d)} &= \text{f(c)} + \text{f(d)}, \\ \text{f}(\lambda c) &= \bar{\lambda} f(d). \end{aligned}$$

for every c, d ϵ A and every $\lambda \epsilon C$.

Then, it is very simple way to verify that the [1] function f : A ϵ C is *semilinear* iff

$$\text{f} : \bar{A} \epsilon C$$

Now, we easily define the sesquilinear forms and the Hermitian forms.

Definition: Let any complex vector space A, a function ψ: A $\times$ A $\to$ C is known as [2] *sesquilinear form* if it has linear in its 1st argument with semilinear in its 2nd argument as :

$$\begin{aligned} \psi(a_1 + a_2, b) &= \psi(a_1, b) + \psi(a_2, b), \\ \psi(a, b_1 + b_2) &= \psi(a, b_1) + \psi(a, b_2), \\ \psi(\lambda a, b) &= \lambda \psi(a, b), \\ \psi(a, \beta b) &= \bar{\beta} \psi(, b). \end{aligned}$$

for all a, b, a_1, a_2, b_1, b_2 ϵ A and all λ , β ϵ C. A function ψ: $A\times$ A $\to$ C is known as [1] *Hermitian form* if it is a sesquilinear and if

$$\psi(a, b) = \overline{\psi(a, b)}$$

for every a, b ϵ A.

It is so obvious that $\psi(0, b)$ [1] $= \psi(a, 0) = 0$. We [1] also note that if ψ: A $\times$ A $\to$ C is a sesquilinear and if ψ: A $\times$ A $\to$ C is a Hermitian, we have

$$\psi(\lambda a + \beta b, \lambda a + \beta b) = |\lambda|^2 \psi(a, a) + 2Re(\lambda \bar{\beta} \psi(a, b)) + |\beta|^2 \psi(b, b)$$

Definition: Given any sesquilinear form ψ: A $\times$ A $\to$ C, the fuction $\Phi : A \to C$ so that $\Phi(a) = \psi(a, a)$ for every a ϵ A is known as *the quadratic form* associated with the ψ.

The basic instance for the [1] Hermitian form on C^n is the map ψ explained so that

$$\psi((a_1, ..., a_n), (b_1, ..., b_n)) = a_1 \overline{b_1} + a_2 \overline{b_2} + ... + a_n \overline{b_n}$$

Proposition: Given any complex vector space A, the after conditions will hold:

(1) A sesquilinear form ψ: A $\times$ A $\rightarrow$ C is a *Hermitian form*iff $\psi(a,a)\epsilon R$ for all a ϵ A.

(2) If ψ: A $\times$ A $\rightarrow$ C is a *sesquilinear form*, then

$4\psi(a,b) = \psi(a+b,a+b)$-$\psi(a-b,a-b)$+$i\psi(a+ib,a+ib)$-$i\psi(a-ib,a-ib)$
and
2 $\psi(a,b) = (1+i)(\psi(a,a)+\psi(b,b))$ - $\psi(a-b,a-b)$ - i$\psi(a-ib,a-ib)$

These are called the **polarization identities.**

The quantity $\psi(a,b)$ is generally called the *Hermitian product* of a and b.

Any given pre-Hilbert space $< A,\psi >$ in case of the Euclidean space, we also signify $\psi(a,b)$ by

$$\textbf{a.b} \text{ either } < a,b >$$

and $\sqrt{\Phi(a)}as||a||$.

Example: C^n be a complex vector space under the Hermitian form

$$\psi((x_1,....,x_n),(y_1,...,y_n)) = x_1\overline{y_1} + x_2\overline{y_2} + ... + x_n\overline{y_n}$$
is the Hermitian space.

Definition: [2] The matrix G related with the Hermitian product is known as *Gram matrix* of the Hermitian space w.r.t. the basis $(e_1,....,e_2)$.

4.1.2 Duality, Orthogonality and Adjoint of Linear Map

We will conducting with Hermitian space and indicate the Hermitian inner product by $a.b$ or $< a,b >$. All the ideas of orthogonal [1] family of vectors, orthogonality and orthogonal complement of a set of vectors remains unchanged from the Euclidean case.

Take an example of the set $C[-\pi,\pi]$ of the continuous function f : $[-\pi,\pi] \rightarrow C$ is a Hermitian space under the product

$$< f,h > = \int_{-\pi}^{\pi} f(x)\overline{h(x)}dx$$

and the family of the $(e^{ikx})_{k\epsilon Z}$ is orthogonal.

In the [1] Euclidean spaces of finite dimensions, the Hermitian product activates a cannonical bijection i.e. not dependent on the choice of bases linking the vector spaces A and A^*.

Let any Hermitian space A, for some vector a ϵ A, assume ψ_a^l : A $\rightarrow C$ be the map explained so that

$$\psi_a^l(\text{b}) = \overline{a.b}, \text{ for every b } \epsilon \text{ A}$$

Like, for any vector b ϵ A, ψ_b^r : A $\rightarrow C$ be the map defined as

$$\psi_b^r(\text{a}) = \text{a.b , for all a } \epsilon \text{ A}$$

Hence, [2] Hermitian product is linear in its 1st argument a with map ψ_b^r is a linear form in A^* and as it is semilinear in its 2nd argument b with map ψ_a^l which is also linear form in A^*. Therefore, we get 2 maps $b^l : A \rightarrow A^*$ and $b^r : A \rightarrow A^*$ explained so that

$$b^l(a) = \psi_a^l \text{ and } b^r(b) = \psi_b^r$$

Theorem: Given any Hermitian space A, the map $b : A \rightarrow A^*$ defined so that

$$b(a) = \psi_a^l = \psi_a^r \text{ for every a } \epsilon \text{ A}$$

is *semilinear as well as injective.* Hence, *A is also of finite dimension* then, the map $b : \bar{A} \rightarrow A^*$ is a *cannonical isomorphism* [1].

Given any Hermitian space A of the finite dimensions, for each linear map $f : A \rightarrow A$, the unique linear maps $f^* : A \rightarrow A$ so that

$$f^*(a).b = a.f(b), \text{ for every a,b } \epsilon \text{ A}$$

is said to be [1] the *adjoint of f* (w.r.t. the Hermitian product).

4.1.3 The Unitary Matrices and Group

As a result [2] of the Gram-Schmidt orthonormalization process we get the QR-decomposition for an invertible matrices. [1] The matrix of adjoint of the linear map in the Hermition is not provided by its transpose but by its conjugate of the original matrix.

Definition: Let any complex m $\times$ n matrix E, the *transpose* E^T of E is the n $\times$ m matrix where $E^T = ({e_i}_j^T)$ explained as:

$${e_i}_j^T = e_{ji}$$

and the *conjugate* $\overline{E}$ of E is the m $\times$ n matrix where $\overline{E} = (f_{ij})$ explained as:

$$(f_{ij}) = \bar{e}_{ij}$$

for every i, j, $1 \leq i \leq m$, $1 \leq i \leq n$.

The *adjoint* E^* of E is the matrix explained as:

$$E^* = \overline{(E^T)} = \overline{(E)}^T$$

Definition: Any complex n x n matrix is said to be a *unitary matrix* if

$$EE^* = E^*E = I_n$$

Definition: Let any [1] Hermitian space A of dimensions n, the set of the isometries $f : A \to A$ forms a subgroup of **GL(A,C)** denotes **U(A) or U(n)** when A = C^n then it is known as the *unitary group (of A).*

4.1.4 Hermitian Reflections with QR-Decomposition

If there is any n × n complex singular matrix say A, which is not surely unique but there is some QR-decomposition A = QR with any unitary matrix Q which is the product of [1] an upper triangular matrix be R and Householder reflections.

Definition: Given E be any Hermitian space of the finite dimensions. Let random hyperplane H, for some non-null vector u which is [2] orthogonal to H such that E = H ⊕ G, where G = Cu, a *Hermitian reflection about H* of angle θ is a linear map of the form $\rho_{H,\theta}E \to E$, explained so that

$$\rho_{H,\theta}E(w) = p_H(w) + e^{i\theta}p_G(w)$$

for some unit complex number $e^{i\theta} \neq 1$ ($or\theta \neq k2\pi$). For any non-zero vector u ϵ E, we represent it by $\rho_{u,\theta}$ [2] the Hermitian reflection present by $\rho_{H,\theta}$ where H be the hyperplane orthogonal to the u.

Hence, [2] w = $\rho_H(w)$ + $p_G(w)$, the Hermitian reflection $\rho_{u,\theta}$ is also indicates as

$$\rho_{u,\theta}(w) = \text{w} + (e^{i\theta} - 1)p_G(w)$$

also

$$\rho_{u,\theta}(w) = \text{w} + (e^{i\theta} - 1)\frac{(w.u)}{||u||^2}u$$

Note: The example of a basic hyperplane reflection is attained when $e^{i\theta}$ = -1, i.e. $\theta = \pi$.

Now, the situation arises:

$$\rho_{u,\pi}(w) = \text{w - } 2\frac{(w.u)}{||u||^2}u$$

and the matrix of this type of reflections is a *Householder matrix* except that u may be a complex vector.

We will easily examine that $\rho_{u,\theta}$ [1] is an isometry and the inverse of $\rho_{u,\theta}$ is $\rho_{u,-}\theta$. If we choose an orthogonal basis $(e_1, e_2, ..., e_n)$ so that $(e_1, e_2, ..., e_{n-1})$ is an orthonormal basis of H and the matrix of $\rho_{u,\theta}$ is

$$\begin{pmatrix} \mathrm{I}_{n-1} & 0 \\ 0 & \mathrm{e}^{i\theta} \end{pmatrix}$$

We now conclude that any 2 different [1] vectors u and v so that $||u|| = ||v||$, not always a hyperplane reflection mapping u to v as this can be ended by using 2 *Hermition reflections.*

4.2 Eigenvalues and Eigenvectors

4.2.1 Eigenvalues and Eigenvectors of Linear Map

Let any vector space B which is a finite-dimensional , assume the defined funtion g : B $\rightarrow$ B be any linear map and there exists a basis $(e_1, ..., e_m)$ of B w.r.t. g is written by a *diagonal matrix*

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_m \end{pmatrix}$$

then work of g on B is extremely basic in the each direction i.e. e_i and we get:

$$g(e_i) = \lambda_i e_i.$$

Here, we conclude the function g as a transformation which shrinks or stretches the space across the direction of the basis $(e_1, ..., e_m)$ (if B be a real vector space). The above concept gives the certainity [2] that there exits an invertible matrix P and also a diagonal matrix D so that the matrix B can be factored as

$$\mathrm{B} = PDP^{-1}$$

If it is satisfy then, we can say that the function g (or B) is *diagonalizable*, $\lambda_i's$ are said to be an eigenvalues of g and $e_i's$ are the eigenvectors of g.

Every symmetric matrix can be diagonalized.

Definition: Let [1] any vector space B of the finite dimension m and a linear map g : B $\rightarrow$ B, a scalar $\lambda \epsilon$ K is known as an *eigenvalue (or the characteristic value or the proper value) of g* if there exists some non-zero vector v ϵ B so that

$$g(v) = \lambda v$$

Therefore, λ be an eigenvalue of g if $Ker(\lambda id - g) \neq 0$(i.e. non-trivial) iff $\lambda id - g$ is not an invertible (Here, we used the truth that B is a finite-dimensional; a linear map from B to itself is injective if and only if it is invertible).

A vector v ϵ B is said to be an *eigenvector* of g if $v \neq 0$ and if there exists any $\lambda \epsilon K$ so that

$$g(v) = \lambda v$$

where the λ, a scalar is the eigenvalue and v is said to be an *eigenvector related with* λ.

Let any eigenvalue $\lambda \epsilon$ K, the non-trivial subspace $ker(\lambda id - g)$ contain every eigenvectors related [1] with λ jointly with the zero vector and the subspace is usually represented by $B_\lambda(g)$ or $B(\lambda, g)$ or B_λ and known as the *eigenspace related with* λ *(or proper subspace related with* λ*).*

Definition: Let B be any vector space of the dimension m, for some linear map f : $B \to B$, the polynomial $P_g(X) = \S_g(X) = det(Xid - g)$ is said to be a *[2] characteristic polynomial of g.* For some square matrix E, the polynomial $P_E(X) = \S_E(X) = det(XI - E)$ is said to be a *characteristic polynomial of E.*

Definition: Let E be any m $\times$ m matrix over the K-field and let all the roots of the characteristic polynomial $\S_E(X) = det(XI - E)$ of E belongs to the field K, which measures that we may express

$$det(XI - E) = (X - \lambda_1)^{k_1}.....(X - \lambda_n)^{k_n}$$

where $\lambda_1, ..., \lambda_n \epsilon K$ are the different roots of det(XI - E) and $k_1 + ... + k_n = m$. The integers k_i is known as the *algebraic multiplicity of the eigenvalue* λ_i **(denoted by** $alg(\lambda_i)$**)** [2] and the dimension of the eigenspace $E_{\lambda_i} = ker(\lambda_i I - E)$ is known as the *geometric multiplicity of* λ_i (denoted by $geo(\lambda_i)$)

4.2.2 Reduction to an Upper Triangular Form

As we can say that not each linear map on the complex vector space can be diagonalized . Hence, we have to study about the "triangularize" which process to find the basis over which the matrix has 0 entries below the main diagonal [1]

Let B be a square *upper triangular matrix* if it has the below standard shape:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m-1} & a_{1m} \\ 0 & a_{22} & a_{23} & \dots & a_{2m-1} & a_{2m} \\ 0 & 0 & a_{33} & \dots & a_{3m-1} & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{(m-1)m-1} & a_{m-1m} \\ 0 & 0 & 0 & \dots & 0 & a_{mm} \end{pmatrix}$$

i.e. $a_{ij} = 0$ whenever $j < i, 1 \leq i, j \geq m$.

Theorem: Let some [2] finite dimensional vector space over a K-field , for some linear map g:$E \to E$, there exists a basis $(u_1, ..., u_m)$ w.r.t. which g is written by an upper triangular matrix (in $M_n(L)$) if and only if all eigenvalues of g belongs to L. Similarly, for each m $\times$ m matrix $B \epsilon M_m(L)$, there exists [2] an invertible matrix P and D be an *upper triangular matrix* (both in $M_m(L)$) so that

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

if and only if all the eigenvalues of $B \epsilon L$.

Theorem: *(Schur decomposition)* Let [1] any linear map g : $E \to E$ over a complex Hermitian space E, there exists an orthonoraml basis $(u_1, ..., u_m)$ w.r.t. which g is written by an *upper triangular matrix*. Similarly, for each m $\times$ m matrix $B \epsilon M_m(\mathbb{C})$, there exists an U be [3] a *unitary matrix* and D be an *upper triangular matrix* so that

$$B = UDU^*$$

If B is a real matrix with all its real eigenvalues then, there exists Q be an *orthogonal matrix* and D be an *upper triangular matrix* so that

$$B = QDQ^T$$

4.2.3 Eigenvalues Location

Let B be an m x m real (or complex) matrix , it will help us to know that [2] where the eigenvalues of B are detected in the complex plane $\mathbb{C}$. The Gershgorin disc also give any of important fact about this.

Definition: Let B be any complex m $\times$ m matrix , for i = 1,....,m and assume

$$R_i'(B) = \sum_{j=1}^{m} |a_{ij}|$$

and

$$\text{G(B)}= \bigcup_{i=1}^{n} \{z \epsilon C \mid |z - a_{ij}| \leq R_i'(B)\}$$

Each disc $\{\text{z } \epsilon \text{ C} \mid |z - a_{ij}| \leq R_i'(B)\}$ is known as *Gershgorin disc* and G(B) be their union is said to be a *Gershgorin domain.*

Theorem: (Gershgorin's disc theorem) Let B be any complex m × m matrix and every eigenvalues of B ϵ G(B). Therefore, the through conditions hold:

1) Let any strictly row diagonally dominant matrix B, i.e.

$$|a_{ij}| > \sum_{j=1, j \neq i}^{m} |a_{ij}| \text{,for i = 1,.....,m}$$

then B is invertible [1] .

2) If B is a strictly row diagonally dominant and if $a_{ii} > 0$ for i= 1,...,m then every eigenvalue of B has a strictly positive real part [1] .

4.3 Euclidean and Hermitian Spaces: Spectral Theorems

In this section, we will try to prove that here we have some nice normal forms for the symmetric, orthogonal, normal and skew-symmetric [2] matrices. [1] The spectral theorem for the symmetric matrices defines that the eigenvalues are real and they can be easily diagonalized over an orthonormal basis. For the Hermitian matrices, the spectral theorem defines that they also have real eigenvalues but they can be diagonalized over a complex orthonoraml basis.

[3] The spectral result for the symmetric matrices can be helpful to prove 2 characterizations of the eigenvalues of a symmetric matrix in labels of the *Rayleigh ratio*, where 1st characterization is the *Rayleigh-Ritz theorem* and the 2nd one is the *Courant-Fischer theorem.* [2] [1] Normal real matrices can be a block diagonalized over an orthonoraml basis with the blocks having a size of at most 2 and for the skew-symmetric , orthogonal matrix there are some refinements of this normal form.

Here, all the vector spaces are finite-dimensional real (or complex) vector spaces [1] .

4.3.1 Eigenvalues and Eigenvectors of Normal Linear Maps

Definition: Given any A be an [1] Euclidean or Hermitian space , a linear map g : $A \to A$ is said to be *normal* if

$$\text{g } o \text{ } g^* = g^* \text{ } o \text{ g}$$

A linear map g : $A \to A$ **is said to be *self-adjoint* if** $g = g^*$ **and *orthogonal* if g** o $g^* = g^*$ o **g = id.**

It is obvious that skew-self-adjoint, [2] self-adjoint or orthogonal linear map is a normal linear map.

Firstly, we have to show that every normal map g : $A \to A$ then there exists an orthonoraml [1] basis (w.r.t. $< -, - >$) so that the matrix of g over this assumed basis has an especially nice form which is a block diagonal matrix where the blocks are either 1D matrices (that is single entries) or 2D matrices of the form

$$\begin{pmatrix} \lambda & \beta \\ \text{-}\beta & \lambda \end{pmatrix}$$

Hence, this normal form can be lately refined if g is skew-self-adjoint, orthogonal or self-adjoint, as firstly, we have to prove that when g is normal then g^* and g have the same kernel.

Now, there are some following prepositions:

1) Let A be any Euclidean space , if g : $A \to A$ is a *normal linear map*, then

$$\text{Ker g} = \text{Ker } g^*.$$

2) Let any [1] Hermitian space A, for some normal linear map g : $A \to A$, we get

$$\text{Ker(g)} \bigcap \text{Im(g)} = (0).$$

3) Given any Hermitian space A, for any linear map g : $A \to A$, a vector v is an *eigenvector of g for the eigenvalue* λ (in $\mathbb{C}$) iff for the eigenvalue $\bar{\lambda}$, v is an eigenvector of [1] g^*.

4) Let any Hermitian space A, for some linear map g : $A \to A$, if eigenvectors of g be a and b related with the eigenvalues λ and β (in $\mathbb{C}$) where $\lambda \neq \beta$, then

$$< a, b >= 0$$

5) Let any Hermitian space A, every [1] eigenvalues of any *self-adjoint linear map g : $A \to A$ are real.*

Definition: Given A be a real vector space and $E_{\mathbb{C}}$ be the form of A × A under the addition operation

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2),$$

and [3] consider multiplication by a complex scalar z = u + iv be explained as

$$\text{(u + iv).(a,b) = (ua - vb, va + ub)}$$

Here, the space $E_{\mathbb{C}}$ is said to be the *complexification of E.*

4.3.2 Spectral Theorem: Normal Linear Maps

Theorem: *(Spectral theorem for self-adjoint linear maps on a Euclidean space)* Let any Euclidean space A of dimension n, for each self-adjoint linear map f:$A \to A$, there exists an orthonormal basis $(e_1, ..., e_n)$ of eigenvectors of f so that the matrix of f with respect to this basis is a diagonal matrix [2]

$$\begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

with $\lambda_i \, \epsilon \, \mathbb{R}$.

Theorem: *(Main Spectral Theorem)* Let any Euclidean space A of dimension n, for each normal linear map f:$A \to A$, there exists an orthonormal basis $(e_1, ..., e_n)$ so that the matrix of f with respect to this basis is a *block diagonal matrix of the form* [1]

$$\begin{pmatrix} E_1 & & \dots & \\ & E_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & E_p \end{pmatrix}$$

so that every block E_j is either a 1D matrix i.e. a real scalar or the 2D matrix of the form

$$E_j = \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}$$

where $\lambda_j, \mu_j \, \epsilon \, \mathbb{R}$, with $\mu_j > 0$

Theorem: *[2] (Spectral theorem for normal linear maps on a Hermitian space)* Let any Hermitian space A of dimension n, for each normal linear map f : $A \to A$ there exists an orthonormal basis $(e_1, ..., e_n)$ of eigenvectors of f so that the matrix of f with respect to this basis is a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_j \, \epsilon \, \mathbb{C}$.

4.4 Computing Eigenvectors and Eigenvalues

After solving the problems of the [1] linear system, the new problem arises of computing the eigenvectors and eigenvalues of a real (or complex) matrix. Therefore, methods which are performing better exists only for the symmetric matrices which are special types of matrices.

4.4.1 The Basic of QR Algorithm

Given C [1] be any n × n matrix which both diagonalizable and invertible. Then, the basic QR Algorithm made up of 2 very easy steps. Initially $C_1 = C$, we make a sequences of the matrices $(C_L), (Q_L), (R_L) and (P_L)$ as follows:

$$\text{Factor } C_1 = Q_1 R_1$$

$$\text{Set } C_2 = R_1 Q_1$$

$$\text{Factor } C_2 = Q_2 R_2$$

$$\text{Set } C_3 = R_2 Q_2$$

$$\vdots$$

$$\text{Factor } A_L = Q_L R_L$$

$$\text{Set } A_{L+1} = R_L Q_L$$

$$\vdots$$

Therefore, the C_{L+1} is contained from the QR-factorization of C_L as $C_L = Q_L R_L$ by swappings the Q_L and R_L. Also, explain the P_L as

$$P_L = Q_1 Q_2 .... Q_L.$$

Hence, $C_L = Q_L R_L$ then we get $R_L = Q_L^* C_L$ and using $C_{L+1} = R_L Q_L$, we get

$$C_{L+1} = Q_L^* C_L Q_L$$

The clear induction proves that

$$C_{L+1} = Q_L^*....Q_1^* C_1 Q_1....Q_L = P_L^* C P_L$$

$$\text{i.e. } C_{L+1} = P_L^* C P_L$$

Hence, C_{L+1} and C are alike as they also have the similar eigenvalues.

4.4.2 A Hessenberg Matrices

Definition: A real (or complex) n × n matrix H is [1] an (upper) *Hessenberg matrix* if it will almost a triangular excluding that it may have an extra non-zero diagonal below the main diagonal. Practically, $h_{jl} = 0$ [1] for every (j,l) so that $j - l \geq 2$.

For example: Let 5 × 5 [2] Hessenberg matrix.

$$H = \begin{pmatrix} * & * & * & * & * \\ h_{21} & * & * & * & * \\ 0 & h_{32} & * & * & * \\ 0 & 0 & h_{43} & * & * \\ 0 & 0 & 0 & h_{54} & * \end{pmatrix}$$

Definition: [1] An upper Hessenberg n × n matrix H is said to be *unreduced* if $h_{i+1i} \neq 0$ for i = 1,...,n-1.

The Hessenberg matrix H is known as *reduced when it is not unreduced*

.

4.4.3 Arnoldi Iteration for Krylov Subspaces

Let the dimension of the square real (or complex) matrix A be n. Let n × n [2] matrix A has been lowered to an upper Hessenberg form H as A = UHU^*. For some $m \leq n$, let the $(m+1) \times m$ be an [2] upper left block

$$\widetilde{H}_n = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1m} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2m} \\ 0 & h_{32} & h_{33} & \dots & h_{3m} \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & h_{mm-1} & h_{mm} \\ 0 & \dots & 0 & 0 & h_{m+1m} \end{pmatrix}$$

of H and then m × m upper [2] Hessenberg matrix H_m acquired by removing the last row of the $\widetilde{H}_m$;

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1m} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2m} \\ 0 & h_{32} & h_{33} & \dots & h_{3m} \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & h_{mm-1} & h_{mm} \end{pmatrix}$$

If we indicate the U_m the n × n matrix consisting of the 1st m columns of U, indicated $u_1, ..., u_m$, then the matrix consisting of 1st n columns of matrix UH = AU can be written as [1]

$$AU_m = U_{m-1}\widetilde{H}_m.$$

It also accompany [1] that the nth column of this matrix can be written as

$$Au_m = h_{1m}u_1 + ... + h_{mm}u_m + h_{m+1m}u_{m+1}.$$

Hence, $(u_1, ..., u_m)$ forms an orthonormal basis and we can deduce the above equation as:

$$< u_j, Au_m >= u_j^* Au_m = h_{jm}, \text{ j = 1,...,m.}$$

From the recent above 2 equations we can prove that U_{m+1} and $\widetilde{H}_m$ can be calculated constantly by using the below algorithm due to Arnoldi, i.e. *Arnoldi iteration:*

Let any random non-zero vector b ϵ $\mathbb{C}^n$ and [2] $u_1 = b/||b||$;

for m = 1, 2, 3,...**do**
z := Au_m; **for** j = 1 **to** n **do** $h_{jm} := u_j^* z$;
z := z - $h_{jm}u_j$
endfor
$h_{m+1m} := ||z||$;
if $h_{m+1m} = 0$ **quit**
$u_{m+1} = z/h_{m+1m}$

We will stop at $h_{m+1m=0}$ (*breakdown* of the *Arnoldi iteration.*)

4.4.4 Lanczos Iteration for the GMRES and the Hermitian Case;

Let E be an n × n invertible matrix and a non-zero vector, b in $\mathbb{C}^n$. Suppose $x_0 = E^{-1}b$, is an unique solution of Ex = b. We can easily show that x_0 ϵ $\mathcal{K}_m(E, b)$ for few $m \leq n$. Also, there exists a distinctive monic polynomaial [2] p(z) of the minimal degree $s \leq n$ so that p(E)b = 0, so x_0 ϵ $\mathcal{K}_s(E, b)$. As it builds sight to look for the solution of Ex = b in the Krylov spaces of dimensions $n \leq s$.

We use the concept of finding the approximation $x_n \epsilon \mathcal{K}_m(E, b)$ [2] of x_0 so that $r_m = b - Ex_m$ is minimized i.e. $||r_m||_2 = ||b - Ex_m||_2$ is also minimized over $x_n \epsilon \mathcal{K}_n(E, b)$.

[1] This minimized [1] problem can be expressed as:

$$\text{minimize } ||r_m||_2 = ||Ex_m - b||_2,\ x_m \epsilon \mathcal{K}_m(E, b).$$

This is the least-squares problem where r_m is said to be *residual* and this method consist in the minimizing $||r_m||_2$ is said to be *GMRES* for the generalized minimial residuals.

If A be an n $\times$ n [1] symmetric or Hermitian matrix therefore, Arnoldi's method is much easier and capable. So, in the Hermitian (also symmetric) case we can easily see the upper Hessenberg matrix H_m [2] and thus tridiagonal. Therefore, the eigenvalues of A and H_m are real and we can write

$$H_m = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{m-1} \\ & & & \beta_{m-1} & \alpha_m \end{pmatrix}$$

Three term recurrence

$$Au_m = \beta_{m-1}u_{m-1} + \alpha_m u_m + \beta_m u_{m+1}$$

We also have an Arnoldi's algorithm i.e. $\alpha_m = u_m^* A U_m$ which turn the mention algorithm [1] as *Lanczos iteration (or Lanczos algorithm).*

[1] The inner loop on j from 1 to n has been removed and changed by a single assignment. Let any random non-zero vector b $\epsilon \mathbb{C}^n$ and [2] $u_1 = b/||b||$;

for m 1, 2, ... **do**
z := Au_m;
$\alpha_m := u_m^* z$;
z := $z - \beta_{m-1}u_{m-1} - \alpha_m u_m$
$\beta_m := ||z||$;
if $\beta_m = 0$ **quit**
$u_{m+1} = z/\beta_m$

We will stop at $\beta_m = 0$ (*breakdown of the Lanczos iteration*)

Chapter 5

Spectral Graphs ,SVD and its applications

5.1 Graphs and Graph Laplacians

Here we will discuss about implementation of linear algebra in graph theory.We can define graph in terms of many different kind of matrices like Incidence and Adjacency.

By incidence matrix we mean a matrix that shows the relationship between two classes of objects and by adjacency matrix we meant a connection matrix containing rows and columns used to represent a simple labelled graph.Certain connectivity properties are involved by properties of these matrices.

One of the significant matrix is the Graph Laplacian which is related to undirected graph. The former is symmteric, +ve definite and its corresponding eigen values involve certain characteristic of the graph underlying [1].

The graph Laplacian is related to an undirected graph [1].For both of them we have a degree matrix, say P, the incidence matrix,say Q and the adjacency matrix A.Also let us define a weighted graph. [1] It is actually a pair (U, W), where U is finite and W is $m \times m$ symmetric matrix having non-negative entry and zero at diagonals. [1]

Degree matrix ,i.e.,P is the diagonal matrix. $\forall$ node $u_i \in U$ the degree $d(u_i)$ is total sum of weights of the edges adjoining to u_i.

$$d_i = d(u_i) = \sum_{j=1}^{n} w_{ij}$$

[2] The earliest way to define [3] unnormalized graph Laplacian L of an undirected graph is by

displaying that for some random orientation of a graph H

$$QQ^T = P - A = L$$

is referred as an invariant.Additionally it can a be defined as of weighted graph $H = (U, W)$ as $L = P - W$.

Graph drawing is a fascinating field where the objective is to represent a given graph in a low-dimensional space, typically in Euclidean space $\mathbb{R}^n$ where $n \ll m$, the no. of vertices in graph. Then our objective is to find a drawing that minimizes this energy function. However, the trivial solution, where all vertices are mapped to the same point (e.g., the origin), results in zero energy. Therefore, additional constraints or regularization techniques are often employed to prevent this trivial solution and encourage meaningful representations of the graph structure. As there are infinitely many graphs possible, so we must be having a criterion to decide which graph is better.Taking inspiration from a physical model which has edges as springs,it is absolutely correct to acknowledge the representation to be better when the springs are not extended or they are least extended. [2] Also the graph having minimum energy function are considered as Good graphs.The most simple representation corresponding to zero matrix is good choice in order to rule out trivial solutions we need to impose a few extra constraints.

Now we will discuss some of the graph drawings-

5.1.1 Directed and Undirected Graph

A directed graph, often denoted as $H = (U, E)$, comprises a set of vertices $U = \{u_1, \ldots, u_m\}$ representing nodes, and a set of edges $E \subseteq U \times U$ consisting of ordered pairs (t, u) where t and u are distinct vertices in U. These ordered pairs denote directional relationships between vertices. For any given edge $e = (t, u)$, the vertex t is termed the source, denoted by $s(e) = t$, while u is referred to as the target, denoted by $t(e) = u$.
In the [2] realm of graph theory, specifically pertaining to a graph $H = (U, E)$ comprising a collection of vertices U and edges E, we define the concept of vertex degree.

For any given vertex U belonging to the set of vertices U, its degree, denoted $d(u)$, signifies the count of edges connected to that particular vertex. This encompasses both incoming and outgoing edges, thereby encapsulating the total number of edges incident upon the vertex.

Expressed formally, the degree of vertex u is determined by cardinality of a set containing vertices $u \in V$ for which either the edge (t, u) or (u, t) is present in the edge set E.

To succinctly represent the degree of each vertex within a graph, we employ the degree matrix $P(G)$. This matrix is structured as a diagonal matrix, whose every diagonal entry corresponds to

degree of the respective vertices. Thus, if there are m vertices in the graph, $P(G)$ emerges as an $m \times m$ matrix, with every diagonal element representing the degree of associated vertices.

$$P(G) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_m \end{pmatrix}$$

Incidence matrix-

Consider a directed graph $H = (U, E)$, where $U = \{u_1, \ldots, u_m\}$ represents a set of m nodes and $E = \{e_1, \ldots, e_n\}$ denotes the set of n edges. The incidence matrix $Q(H)$ for H is constructed as an $m \times n$ matrix. Its entries a_{ij} are defined as follows:

- If the source vertex of edge e_j is u_i, then $a_{ij} = +1$.
- If the target vertex of edge e_j is u_i, then $a_{ij} = -1$.
- else, $a_{ij} = 0$.

For instance, consider a directed graph G_1 with the corresponding incidence matrix Q. Each column of Q has precisely 2 non-zero entries, i.e.,+1 and -1, representing the connection between nodes and edges:

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

Each row corresponds to a node, and each column corresponds to an edge. The $+1$ and -1 entries denote the direction of the edge relative to the node,while 0 indicates no connection.
Furthermore, in [1] a directed graph with m nodes $u_1, \ldots, u_m$ [1] and n edges $e_1, \ldots, e_n$, a vector $x \in \mathbb{R}^m$ can be viewed as a function $x : U \to \mathbb{R}$, assigning the value x_i to node u_i [1]. In this interpretation, $\mathbb{R}^m$ is considered as $\mathbb{R}^U$. [2]

5.1.2 Undirected graph-

In the realm of graph theory, a graph, often termed as an undirected graph, is characterized as a pair $H = (U, E)$, where $U = \{u_1, \ldots, u_m\}$ denotes a set comprising nodes or vertices, and E signifies a set containing [2] pairs of elements selected from U. These pairs, denoted as $\{t, u\}$, consist of two distinct vertices t and u belonging to U, thereby ensuring the absence of self-loops.
Remark:

It is necessary to notice that each edge within the graph G is represented by a subset $\{u, v\}$, where u and v are distinct vertices. This condition eliminates the possibility of self-loops, ensuring that each edge connects precisely two distinct vertices. Consequently, for any given pair of nodes $\{u, v\}$, there exists at most one edge connecting u and v [2]. These graphs are commonly referred to as simple graphs, reflecting the straightforwardness of their edge structure.

5.1.3 Weighted Graphs and Laplacian Matrices in Graph Theory

In graph theory, a weighted graph is described by the pair $H = (U, W)$, where $U = \{u_1, \ldots, u_m\}$ is the set of vertices, W is the weight matrix, a symmetric matrix with specific properties. The entries w_{jk} of W satisfy $w_{jk} \geq 0 \; \forall \; j, k \in \{1, , m\}$ and $w_{jj} = 0$ for $j = 1, \ldots, m$. An edge $\{u_j, u_k\}$ exists if and only if $w_{jk} > 0$. The undirected graph (U, E), where $E = \{\{u_j, u_k\} \mid w_{jk} > 0\}$, is known as underlying graph of H.

Laplacian Matrices of Graphs

For an undirected graph H with any chosen orientation σ, let Q_σ be an incidence matrix of directed graph H_σ, A the adjacency matrix, and P the degree matrix [2] with $P_{ii} = p(u_i)$. It follows that $Q_\sigma Q_\sigma^T \geq P - A$. Therefore, the Laplacian matrix $L = Q_\sigma Q_\sigma^T$ is invariant under any orientation σ of H, and $P - A$ is both symmetric as well as positive semi-definite, indicating its eigenvalues to be real and positive [1].

5.1.4 Unnormalized graph-

For some weighted graph $H = (U, W)$ with $U = \{u_1, \ldots, u_m\}$, unnormalized Laplacian matrix $L(H)$ can be defined as:

$$L(H) = P(H) - W,$$

where $P(H)$ is the degree matrix of H, represented as $\text{diag}(p_1, \ldots, p_m)$. Each diagonal entry p_i is calculated as summation of weights of all edges connected to vertex u_i. Specifically,

$$p_i = \sum_{j=1}^{m} w_{ij}.$$

Definition- In the context of a weighted graph $H = (U, W)$ with $U = \{u_1, \ldots, u_m\}$, and with edges $\{e_1, \ldots, e_n\}$ representing the connections in the underlying graph of H [2] (where $\{u_i, u_j\}$ is considered an edge if $w_{ij} > 0$), the incidence matrix Q_σ of any oriented graph H_σ (obtained by assigning a direction to the underlying graph of H) is a $m \times n$ matrix. The entries q_{ij} are defined as follows: [2]

$$q_{ij} = \begin{cases} +\sqrt{w_{ij}} & \text{if } s(e_j) = u_i, \\ -\sqrt{w_{ij}} & \text{if } t(e_j) = u_i, \\ 0 & \text{otherwise.} \end{cases}$$

For instance, considering the weight matrix

$$W = \begin{pmatrix} 0 & 3 & 6 & 3 \\ 3 & 0 & 0 & 3 \\ 6 & 0 & 0 & 3 \\ 3 & 3 & 3 & 0 \end{pmatrix}$$

the incidence matrix Q to orientation of underlying graph of W, where an edge (i, j) is oriented positively if $i < j$, [2] :

$$Q = \begin{pmatrix} 1.7121 & 2.4595 & 1.7221 & 0 & 0 \\ -1.7321 & 0 & 0 & 1.7321 & 0 \\ 0 & -2.4495 & 0 & 0 & 1.7321 \\ 0 & 0 & -1.7321 & -1.7321 & -1.7321 \end{pmatrix}$$

It should be verified by the reader that $BB^T = D - W$.

5.1.5 Normalized Laplacian Matrices of Graphs-

Definition: In the context of [1] a weighted graph $H = (U, W)$, [1] a vertex $u \in U$ is considered isolated when it is not connected with some other vertex [1]. This implies that each row of the weight matrix W has at least one that is strictly greater than 0 [1]. If H has 0 isolated vertices, then the degree matrix P has positive entries, making it invertible [1]. Hence, the matrix $P^{-1/2}$ is well-defined, given by $P^{-1/2} = \text{diag}(p_1^{-1/2}, \ldots, p_m^{-1/2})$, and similarly for any real exponent α [2].

Definition: For any weighted directed graph $H = (U, W)$ with 0 isolated vertices, and [2] where $U = \{u_1, \ldots, u_m\}$, the (normalized) graph Laplacians L_{Sym} and L_{rw} of H are defined as follows:

$$L_{\text{Sym}} = P^{-1/2} L P^{-1/2} = I - P^{-1/2} W P^{-1/2}$$

$$L_{\text{rw}} = P^{-1} L = I - P^{-1} W$$

5.2 Spectral Graph Drawing-

When dealing with undirected graphs, it's common to visualize them by assigning points in either 2D or 3D space to each vertex and connecting points with line segments whenever there's an edge between the corresponding vertices. This process utilizes the graph Laplacian and can yield effective visual representations.

5.2.1 Graph Drawing and Energy Minimization

Consider a graph $H = (U, E)$ with m vertices. A graph drawing involves assigning a point $\eta(v_i)$ in $\mathbb{R}^n$ to every vertex $u_i \in U$. We then create a matrix S, where each row represents the coordinates of the corresponding vertex in $\mathbb{R}^n$.

For practical purposes, we aim for n to be less than or equal to m, often preferring $n = 2$ or 3.

A drawing is considered balanced if the sum of each column in the matrix S equals zero, denoted by $1^T \cdot S = 0$. If a drawing isn't balanced, we can make it so by adjusting the positions accordingly.

Additionally, it's beneficial for the columns of S to be linearly independent. Therefore, we typically assume $n \leq m$.

It's important to note that a graph drawing need not be one-to-one, meaning different vertices might be represented by the same point. This concept is termed graph immersion to differentiate from injective graph embeddings used in differential geometry.

When discussing graph drawing, it's intuitive to evaluate the effectiveness of a representation based on how much the connections (edges) are stretched. This concept can be formalized by encircling energy of a drawing, denoted as $E(S)$, which sums up the squared distances between connected vertices:

$$E(S) = \sum_{\{u_i,u_j\}\in E} \|\eta(u_i) - \eta(u_j)\|^2,$$

where $\eta(u_i)$ denotes the position of the i^{th} vertex represented as the i^{th} row of the drawing matrix S, and $\|\eta(u_i) - \eta(u_j)\|^2$ represents the squared Euclidean length between $\eta(u_i)$ and $\eta(u_j)$.

Therefore, we aim for "good drawings" that reduces this energy function E to its minimum. However trivial solution of a zero matrix is the optimal representation, so we need to introduce additional constraints to avoid this.

[2]In a broader scenario where the connections (edges) have varying strengths, [2] we introduce a weight matrix which is symmetric $W = (w_{ij})$ with $w_{ij} \geq 0$. [2]This adjustment modifies the energy function as follows:

$$E(S) = \sum_{\{u_i,u_j\}\in E} w_{ij}\|\eta(u_i) - \eta(u_j)\|^2.$$

[1]Interestingly, this function can be expressed using the Laplacian $L = P - W$, where P represents the diagonal matrix containing the degrees of the vertices.

Proposition- [1] Let $H = (U, W)$ denote a weighted graph with $|U| = m$ vertices and W representing an $m \times m$ symmetric matrix. [1]Suppose S is the matrix corresponding to a graph drawing η of H in $\mathbb{R}^n$, where S is an $m \times n$ matrix. [1] If $L = P - W$ represents the unnormalized Laplacian matrix associated with W, [1]then the energy $E(S)$ of the drawing S can be expressed as:

$$E(S) = \mathrm{tr}(S^T LS).$$

The symmetry of $S^T LS$ implies it possesses real eigenvalues. Given that L is positive semidefinite, $S^T LS$ inherits this property, making its trace equivalent to sum of its non-zero eigenvalues, which corresponds to the energy $E(S)$ of the graph drawing.

Suppose S represents the matrix of a graph drawing in $\mathbb{S}^n$ [1]. For some invertible $n \times n$ matrix M [2], associating $\eta(u_i)M$ with u_i generates another valid graph drawing of H. Both representations contain the same information, emphasizing that a graph drawing is essentially defined by the column space of S. Thus, assuming orthogonal columns with unit length in S is a reasonable choice, satisfying the condition $S^T S = I$.

Definition: A graph drawing is labeled as orthogonal if its corresponding matrix S adheres to $S^T S = I$. This criterion also eliminates trivial representations.

Following this guideline allows us to ascertain minimum energy orthogonal balanced graph drawings, especially for connected graphs. Recall $L_1 = 0$.

Theorem. Consider a weighted graph $H = (U, W)$ with $|U| = m$. Let $L = P - W$ represent the unnormalized Laplacian of H, where P is the degree matrix. Suppose the eigenvalues of L are ordered as

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_m.$$

Then, the minimum energy required for any balanced orthogonal graph drawing of H in $\mathbb{S}^n$ equals $\lambda_2 + \cdots + \lambda_{n+1}$, indicating that the dimension n is less than m [1].

[2]To achieve this minimal energy, we can construct [2] an $m \times n$ matrix R using any unit eigenvectors $u_2, \ldots, u_{n+1}$ associated with eigenvalues $\lambda_2 \leq \cdots \leq \lambda_{n+1}$. [1] This matrix R defines a balanced orthogonal graph drawing with minimal energy, satisfying the condition $S^T S = I$.

[2]Due to the fact that the nullspace of L is spanned by 1, including u_1 (which belongs to the kernel of L) as one of the vectors in S would lead to all points representing vertices of G having the same first coordinate [3]. This would confine the drawing to a hyperplane in $\mathbb{S}^n$, which is undesirable, especially when $n = 2$ as all vertices would [2] then lie on a single line. Therefore, we exclude the first eigenvector u_1.

It's worth noting that for any orthogonal matrix Q of size $n \times n$, the energy of the graph drawing remains unchanged when we apply Q^T to the rows of S. This implies that the matrix SQ also results in a minimum energy orthogonal graph drawing.

To summarize, if the second smallest eigenvalue $\lambda_2 > 0$, [1] a simple method for drawing a graph in $\mathbb{S}^2$ is as follows:

1. Calculate two smallest non-zero eigenvalues $\lambda_2 \leq \lambda_3$ of Laplacian graph LL, considering that $\lambda_3 = \lambda_2$ if λ_2 is an eigenvalue which is multiple in nature [2].

2. Obtain 2 eigenvectors(unit) u_2 and u_3 corresponding to λ_2 and λ_3, and construct the matrix $S = [u_2\ u_3]$ of size $m \times 2$, with u_2 and u_3 as its columns.

3. Position each vertex u_i at the coordinates represented by the i^{th} row of [2] S, i.e., (S_{i1}, S_{i2}).

While such method usually yields satisfactory results, it's important to note that identical rows in S may result in non-unique assignments of images to distinct nodes, although this is infrequent in practice.

5.2.2 Algorithm used in Graphical Drawings-

Consider a scenario where you have a random graph with four nodes. To represent the connections between these nodes, we create an adjacency matrix, denoted as A. Let's denote this matrix as follows:

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

To analyze such graph further, we utilize a computational approach. First, we compute the Laplacian matrix L, which is derived from the adjacency matrix A. L is defined as the difference between the degree matrix P and the adjacency matrix A.

Next, we find the eigenvectors and eigenvalues of the Laplacian matrix. These eigenvectors and eigenvalues provide valuable insights into the structure and properties of the graph.

With the obtained eigenvectors, we visualize the graph using a spectral layout technique. This technique leverages the eigenvalues to determine the positions of the nodes in a visually meaningful way.

Through this computational process, we gain a deeper understanding of the graph's connectivity and structure, which can be crucial for various applications in network analysis and graph theory.

5.3 Singular value decomposition and Polar form-

5.3.1 Propeties of f^*of

[2]In this section, we delve into the properties of linear maps within Euclidean spaces, exploring their potential for diagonalization. We introduce (SVD),the powerful concept in linear algebra that enables to reduce any linear map provided we're open to employing two orthonormal bases.

A crucial observation lies in the self-adjoint nature of $f^* \circ f$ and $f \circ f^*$. This self-adjoint property ensures that these compositions can be diagonalized, as well as guaranteeing that their eigenvalues are real.

The central proposition we tackle is the nonnegativity of the eigenvalues of $f^* \circ f$ and $f \circ f^*$. This proposition is pivotal, and its proof is built upon the self-adjointness of these compositions.

The essence of the proof lies in demonstrating that for any eigenvalue λ_i of $f^* \circ f$, and corresponding eigenvector u_i, the associated λ_i is non-negative. Similarly, for any eigenvalue μ_i of $f \circ f^*$, with eigenvector v_i, the corresponding μ_i is also nonnegative.

This affirmation is established by scrutinizing the inner products of $f(v_i)$ and $f^*(u_i)$, respectively, and leveraging the properties of self-adjointness. Through this analytical lens, we unveil the non-negativity of the eigenvalues, underscoring a fundamental property of linear maps.

5.3.2 Singular Values of a Linear Map:

Consider a [1]linear map $f : E \to F$, the square roots $\sigma_i > 0$ of +ve eigen values of $f^* \circ f$ (and $f \circ f^*$) are termed as the singular values of f.

Positive Semidefinite and Positive Definite Linear Maps: A self-adjoint linear map $f : E \to E$ is labeled positive semidefinite (or positive) if its eigenvalues are nonnegative [1]. Additionally, if f is invertible, it's termed as positive definite, signifying that each eigenvalue of f is strictly positive.

Given any linear map $f : E \to F$ [1], it's now apparent that both $f^* \circ f$ and $f \circ f^*$ qualify as positive semidefinite self-adjoint linear maps [1]. This observation leads to an intriguing realization: every linear map can be decomposed in two significant ways: [1]

1. **The Polar Form**
2. **The Singular Value Decomposition (SVD)**

Proposition: For any linear map $f : E \to F$ between 2 Euclidean spaces E and F, where E has dimension n and F has dimension m [2], the following relationships hold:

1. Ker f = Ker $(f^* \circ f)$,
2. Ker f^* = Ker $(f \circ f^*)$,
3. Ker $f = (\text{Im } f^*)^{\perp}$,
4. Ker $f^* = (\text{Im } f)^{\perp}$,
5. $\dim(\text{Im } f) = \dim(\text{Im } f^*)$,
6. The ranks of $f, f^*, f^* \circ f,$ and $f \circ f^*$ are identical.

[1]

5.3.3 Singular Value Decomposition for Square Matrices

Now, our objective is to establish the existence of a Singular Value Decomposition (SVD) for every square matrix. To achieve more robust results, we can initially explore the polar form, leveraging its uniqueness properties to derive the SVD. This approach allows us to delve deeper into the structure of linear transformations and unveil significant insights.

Theorem: (Singular Value Decomposition) For any real square matrix A, there exists 2 orthogonal matrices U and V along a diagonal matrix D such that $A = VDU^T$ [1]. The diagonal matrix D takes the form:

$$D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix},$$

where $\sigma_1, \ldots, \sigma_r$ represent the singular values of A [2]. These singular values are the positive square roots of the nonzero eigenvalues of both A^TA and AA^T, while $\sigma_{r+1} = \ldots = \sigma_n = 0$ [2]. Furthermore, the columns of U correspond to the eigenvectors of A^TA, while the columns of V correspond to the eigenvectors of AA^T. [2]

5.3.4 Single valued decomposition:

A triplet (U, D, V) satisfying $A = VDU^T$, where U and V are orthogonal matrices, and D is a diagonal matrix with nonnegative entries (i.e., positive semidefinite), is termed as a Singular Value Decomposition (SVD) of A.

MATLAB command to compute SVD $A = VDU^T$ of a matrix A is given by:

$$[V, D, U] = \text{svd}(A).$$

5.3.5 Polar form for square matrices-

Another concept closely linked to Singular Value Decomposition (SVD) is the polar form of matrices [1].

Polar decomposition of A : A pair (R, S) such that $A = RS$, where R is orthogonal and S is symmetric as well as +ve semi-definite, is termed as polar decomposition of A [1].

5.3.6 Weyl's Inequalities:

[1] For any complex square matrix A, [1]where $\lambda_1, \ldots, \lambda_n$ denote its [1] eigenvalues and $\sigma_1, \ldots, \sigma_n$ represent its singular values [1], both ordered such that $|\lambda_1| \geq \ldots \geq |\lambda_n|$ and $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$, the following hold:

1. Product of absolute values of the eigenvalues equals the product of the singular values: $|\lambda_1| \cdot \ldots \cdot |\lambda_n| = \sigma_1 \cdot \ldots \cdot \sigma_n$.
2. For each $k = 1, \ldots, n-1$, the product of absolute values of first k eigenvalues is less than or equal to the product of the first k singular values: $|\lambda_1| \cdot \ldots \cdot |\lambda_k| \leq \sigma_1 \cdot \ldots \cdot \sigma_k$.

These inequalities provide valuable insights into the relationships between the eigenvalues and singular values of a matrix. [1]

5.3.7 Singular Value Decomposition for Rectangular Matrices-

Theorem:(Singular Value Decomposition) For any real $m \times n$ matrix A, there exist two orthogonal matrices U $(n \times n)$ and V $(m \times m)$ along with a diagonal $m \times n$ matrix D, such that $A = VDU^T$. The matrix D takes one of the following forms: [1]

$$D = \begin{pmatrix} \sigma_1 & & & & 0 & \cdots & 0 \\ & \sigma_2 & & & 0 & \cdots & 0 \\ & & \ddots & & \vdots & \ddots & \vdots \\ & & & \sigma_r & 0 & \cdots & 0 \end{pmatrix},$$

or [1]

$$D = \begin{pmatrix} \sigma_1 & & 0 & \cdots & 0 \\ & \sigma_2 & & \ddots & \vdots \\ & & \ddots & & 0 \\ & \vdots & & \sigma_m & 0 \end{pmatrix},$$

where $\sigma_1, \ldots, \sigma_r$ denote the singular values of A [1]. These values are the positive square roots of the nonzero eigenvalues of [1] A^TA and AA^T, and $\sigma_{r+1} = \ldots = \sigma_p = 0$, where $p = \min(m, n)$ [1]. The columns of U correspond to the eigenvectors of A^TA, while the columns of V correspond to the eigenvectors of [2]AA^T.

The polar form is essential in continum mechanics, aiding in the separation of stretching and rotation during deformations, a fundamental requirement in deformation analysis. The real eigenvalues $\sigma_1, \sigma_2, \ldots, \sigma_r$ of S serve as indicators of stretch or compression factors. The ability to diagonalize S using an orthogonal matrix implies a natural alignment of axes, known as the principal axes. [1]

5.4 Applications of Singular Value Decomposition (SVD)

1. **Pseudo-inverse and Least Squares Method**: SVD is crucial for solving overdetermined systems of linear equations $(Ax = b)$ using method of least squares. When the matrix A has more equations (rows) than unknowns (columns) $(m > n)$, a direct solution x satisfying $Ax = b$ may not exist. Instead, SVD helps compute the pseudo-inverse A^+ of A which minimizes the residual vector $b - Ax$ in the Euclidean norm sense, providing the "best" approximate solution.

2. **Data Compression**: SVD is used for data compression by reducing the dimensions of a dataset while retaining important information. By retaining only the most significant singular values and their corresponding vectors, SVD enables efficient representation of data in a lower-dimensional space.

3. **Principal Component Analysis (PCA)**: PCA is a statistical technique used to identify patterns and reduce dimensionality of high-dimensional data. SVD plays a crucial role in PCA by computing the principal components (eigenvectors) of the covariance matrix of the data. These principal components capture the maximum variance of the data and facilitate data visualization and feature extraction.

4. **Best Affine Approximation**: SVD is also applied in determining the best affine approximation of a dataset. This involves finding a linear transformation (represented by a matrix) that minimizes the average squared distance between the original data points and their transformed counterparts.

5.4.1 Application of SVD and Pseudo-Inverse-

In this section we would discuss various applications of SVD.The first being one being psuedo inverse.It plays a major role in solving system of linear equations by least square method [2]. Some other applications are data compresssing, principal component analysis(PCA) etc.

Least square method is used to solve an overdetermined sytem of linear equations.By overdetermined system we mean the case when no. of equations is more than the no. of unknown variables.

[3]The reason behind this process is that we take repeated measurements for minimizing the errors. [3]It produces an overdetermined and often inconsistent system of linear equations.For example ,Gauss solved a system of 11 equations with 6 unknowns to determine the orbit of asteroid Pallas [3].

5.4.2 Moore-Penrose Pseudo-inverse via Singular Value Decomposition (SVD)

Definition: Given a nonzero $m \times n$ matrix A of rank r, with its singular value decomposition (SVD) $A = UDV^T$ [3], where:

- U is an $m \times m$ orthogonal matrix,
- D is an $m \times n$ diagonal matrix with singular values $\lambda_1, \lambda_2, \ldots, \lambda_r$ (nonzero singular values of A),
- V is an $n \times n$ orthogonal matrix.

The Moore-Penrose pseudo-inverse A^+ is computed as follows:

1. **Construction of D^+ (Pseudo-inverse of D):**

$$D^+ = \begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

where Λ^{-1} is an $r \times r$ diagonal matrix with entries $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_r}$ (reciprocals of nonzero singular values), and the remaining entries are zeros.

2. **Computation of A^+ (Pseudo-inverse of A):**

$$A^+ = UD^+V^T$$

resulting in an $n \times m$ matrix A^+ that behaves like a pseudo-inverse of A.

[2]Given matrix D:

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

[2] The pseudo-inverse D^+:

$$D^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5.4.3 Properties of the Pseudo-inverse-

Property :-*Moore-Penrose Pseudo-Inverse for Matrices with Full Rank
When A has full rank:

- **Case $m \geq n$ (More rows than columns):**

$$A^+ = (A^TA)^{-1}A^T$$

Here, A^T is the transpose of A. This results in $A^+A = (A^TA)^{-1}A^TA = I$, indicating A^+ is a left inverse of A [1].

- **Case $n \geq m$ [2] (More columns than rows):**

$$A^+ = A^T(AA^T)^{-1}$$

This implies $AA^+ = AA^T(AA^T)^{-1} = I$, showing A^+ is a right inverse of A.

In both cases, A^+ is a pseudo-inverse of A that satisfies the properties:

$$AA^+A = A \quad \text{(left inverse property)}$$

$$A^+AA^+ = A^+ \quad \text{(right inverse property)}$$

These expressions demonstrate how the Moore-Penrose pseudo-inverse is computed based on the dimensions of A and its full rank, providing a generalized form of matrix "inversion" for non-square matrices.

Property :-Projection and Relationship between Range and Null Space
The Moore-Penrose pseudo-inverse A^+ can be interpreted as the projection matrix onto the kernel (null space) of A, denoted by $P_{\ker(A)}$, where:

$$A^+A = P_{\ker(A)}$$

This means A^+A projects any vector onto the space of vectors that A maps to zero, effectively identifying the components of the vector that lie in the null space of A.

[1]The relationship between the range range(A) (denoted Im(A)) and the null space ker(A) (denoted ker(A)) of A is important. [1] The range of A represents all possible outputs of A when applied to vectors in its domain. The orthogonal complement of the range of A, denoted range$(A)^\perp$ or Im(A^T), is equal to the null space (kernel) of the transpose of A, i.e.:

$$\text{range}(A)^\perp = \ker(A^T) = \text{Im}(A^T)$$

Therefore, A^+ can also be seen as providing the projection onto the null space ker(A), and A^+A projects onto ker(A), capturing the essence of the decomposition of the vector space based on the action of A.

Conclusion

A framework for comprehending vector spaces, which is crucial to machine learning, is provided by linear algebra. The data is represented by vectors, and it is manipulated and transformed using different concepts of linear algebra. Images can be represented as matrices in linear algebra through the use of image representation, which makes processing and analysis of picture data more effective.

This project has demonstrated the pivotal role of linear algebra in machine learning, computer vision, and robotics. By providing a framework for understanding vector spaces, linear algebra enables the representation and manipulation of data, optimization of functions, and solution of systems of equations. Its applications in computer vision, such as image processing, object detection, and facial recognition, have revolutionized the field. Moreover, linear algebra's contributions to machine learning, including neural network optimization and dimensionality reduction, have significantly improved model performance. Finally, its role in robotics, particularly in motion planning, has enabled the development of more efficient and accurate robotic systems. As the fields of machine learning, computer vision, and robotics continue to evolve, the importance of linear algebra will only continue to grow.

Linear algebra offers a strong foundation for resolving issues in robotics, computer vision, and machine learning. It is utilized to represent and work with data, optimize functions, and resolve equation systems—all crucial abilities for any professional working in these domains.

Machine learning, computer vision, and robotics will all be impacted by linear algebra's continued spurring of innovation, which will open up new avenues and completely transform these industries in the process. Its uses will keep growing, encompassing fields like: Autonomous Systems, Healthcare and Biomedical Imaging and Cybersecurity.

Bibliography

[1] Jean Gallier and Jocelyn Quaintance. Algebra, topology, differential calculus, and optimization theory for computer science and engineering, 2019.

[2] Jean Gallier and Jocelyn Quaintance. Linear algebra for computer vision, robotics, and machine learning. *University of Pennsylvania*, 2019.

[3] Jean Gallier and Jocelyn Quaintance. *Linear Algebra and Optimization with Applications to Machine Learning: Volume I: Linear Algebra for Computer Vision, Robotics, and Machine Learning*. World Scientific, 2020.

[4] Jean Gallier and Stephen S Shatz. Algebraic geometry. 2005.

[5] Victor P Havin and Nikolai K Nikolski. *Linear and Complex Analysis Problem Book 3: Part 1*. Springer, 2006.