Algebraic Foundation Of Semidefinite Programming

A THESIS

Submitted by

Arti Nigam

In partial fulfillment for the award of the Degree of

Master Of Technology

In

Computer Science and Engineering

Under the guidance of

Dr. K. Muralikrishnan

Department of Computer Science

National Institute Of Technology-Calicut

NIT Campus PO, Calicut

Kerela, INDIA 673601

July 2015

Acknowledgement

It is with immense gratitude that I acknowledge my guide Dr. K. Muralikrishnan for introducing me to the topic and for the motivation and support he has given me throughout in connection with the project.
I am deeply grateful to Dr. Abdul Nazeer K. A, Head of the Department, for allowing me to use all the facilities of the Department, which I required for this project.
I would like to thank the members of my evaluation panel, Dr. Sudeep K. S., Dr. Subhashini R, Ms. Nadiya, and Mr. Binu Jasim for the thoughts and advices given to me that helped me improve the outcome of the project.
I would like to thank Dr. Priya Chandran, our Project Coordinator and Professor, for allowing me to do this project on this topic.
Last but not the least I would also like to thank all, especially my family and my friends, without whose support, moral and intellectual, this work would not have been completed.

Arti Nigam

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: **Signature:**

Date : **Name:**

Reg.No:

Certificate From Supervisor

This is to certify that the thesis entitled Algebraic Foundation of Semidefinite Programming submitted by Ms Arti Nigam to the National Institute of Technology Calicut towards partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Science and Engineering is a bonafide record of the work carried out by her under my/our supervision and guidance.

Dr. K. Muralikrishnan

Place: Kozhikode

Date:

Signature of Head of the Department

Abstract

Semidefinite programming is a relatively new field of combinatorial optimization and is becoming a vital tool for improving the performance of algorithm for many combinatorial problems. Other application areas include operational research, convex constrained optimization, control theory etc. Here we present some concepts of linear algebra that are pre-requisites to an understand for semidefinite programming. There is also proof for the Spectral Theorem for Hermitian operator over finite dimension Euclidean Spaces and an introduction to positive semidefinite operators. We then move to semidefinite programming, its formulation and how it is equivalent to vector programming. The thesis concludes with a study of an application of semidefinite programming for the design of an approximation algorithm for the Max-Cut problem.

Contents

List of Figures

Chapter 1

Introduction

This thesis collects the algebraic foundations necessary to an understanding of semidefinite programming.

Semidefinite programming is a field of optimization. It is the generalization of linear programming and quadratic programming and still easy to solve. In the next two chapters we discuss some algebraic notions that form prerequisites to an understanding of semidefinite programming. Next two chapters will have basic linear algebra concepts related to real vector space like inner product, linear transformations, operators, basis transformations, projections, orthogonal projections, Spectral theorem, symmetric positive semidefinite matrices etc.

In chapter 3 we will discuss semidefinite programming, its formulation and its equivalent vector programming formulations. In chapter 5 we study an application of semidefinite programming works to solve the Max-cut Problem. A randomized approximation algorithm for max-cut discovered by Goemans and Williamson [1] is presented and analysed.

1.0.1 Literature Survey

There are number of books available related to linear algebra. In this thesis we have referred to the Algebra by Michael Artin [7].

Semidefinite programming is not a new topic in optimization field but has been recently studied for several combinatorial optimization problems. For literature review on semidefinite programming refer the work on Semidefinite Programming by Lieven Vandenberghe and Stephen Boyd[13] as well as [3], [6] and [8].

Max-cut or maximum cut problem is a well known problem that has been exclusively studied in theoretical computer science. It is one of the Karp's NP-Complete Problems[12]. As it is NP-Hard, so no polynomial time algorithm is known but we can solve it in polynomial time for some special cases like planar graphs. Goemans and Williamson have presented the approximation solution for Max-cut problem [1]. The algorithm is discussed in chapter 6.

Chapter 2

Finite Dimensional Euclidean Spaces

The following notions in the theory of real vector spaces R^n are pre-requisites to an understanding of semi-definite programming. Here are brief definitions, for detailed explanations one can refer to[7].

2.0.2 Real Vector Space

A vector space V over a field of real numbers is a set together with two laws of composition[7]:

- addition: $V \times V \longrightarrow V$, written v, $w \rightsquigarrow v + w$, for v and w in V,
- scalar multiplication by elements of the field: $R \times V \longrightarrow V$, written $c, v \rightsquigarrow cv$, for c in R and v in V.

such that $(V, +)$ is an albelian group with $0 \in V$ as the zero element and satisfies scalar multiplication.

The vector space of n tuples over R denoted by R^n with the standard vector addition and scalar multiplication will be the principal object to study in this report.

Vector Subspace : Let V be vector space then W is a real vector subspace of V if W is a subset of V and,

- for every w_1, $w_2 \in W$ $w_1 + w_2 \in W$
- for every $c \in R$ $cw_1 \in W$.

Ex. $\{\begin{smallmatrix} x \\ y \\ z \end{smallmatrix} : x + y + z = 0\}$ is a subspace of R^3

Linear Independence: An ordered set of vectors $S = (v_1, ..., v_n)$ is (linearly) independent if for any column vector $X = (x_1, x_2, ..., x_n)^T$, there is no linear relation $SX = 0$ except for the trivial one in which $X = 0$, i.e., in which all the coefficients x_i are zero. A set that is not independent is dependent.[7]

Span: The set of all vectors that are linear combinations of $S = (v_1, ..., v_n)$ forms a subspace of V, called the subspace spanned by the set.[7] Let S be an ordered set of vectors of V, and let W be a subspace of V. If $S \subset W$, then Span $S \subset W$.
Ex: The vectors $\left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$ span the plane $z = 0$ in the $x - y - z$ plane in R^3.

Basis: A basis B of a vector space V is a set $(v_1, ..., v_n)$ of vectors that are independent and also spans V.
The ordered set of vectors $(e_1, e_2, ..., e_n)$ where $e_i = [0...1...0]^T$ with 1 appearing at the i^{th} position is the basis for R^n and is called the standard basis of R^n.
Every vector $v \in V$ can be expressed as a linear combination of basis vector of V:
$v = x_1v_1 + x_2v_2 + ... + x_nv_n$

Lemma 2.1. *Any basis of vector space has the same number of elements(i.e same cardinality).*

Proof. See [7]. □

Dimensions: Let V be a vector space and and let the basis of V has n number of elements. This cardinality of basis is called the dimension of the vector space which is represented as $dim(V) = n$

Lemma 2.2. *If $(v_1, v_2, ..., v_n)$ is a basis of V then for every vector $v \in V$ there exists unique scalars $(x_1, x_2, ..., x_n)$ such that $v = x_1v_1 + x_2v_2 + ... + x_nv_n$.*

Proof. Let there exist some other unique scalars $(y_1, y_2, ..., y_n)$ such that
$v = y_1v_1 + y_2v_2 + ... + y_nv_n$
This results into ,
$v = x_1v_1 + x_2v_2 + ... + x_nv_n = y_1v_1 + y_2v_2 + ... + y_nv_n$
i.e. $x_i - y_i = 0 \Longrightarrow x_i = y_i$
Therefore for every vector $v \in V$ there exists unique scalars $(x_1, x_2, ..., x_n)$. □

Coordinate Vector: A coordinate vector of any vector $v \in V$ relative to basis B is the sequence of coordinates i.e representation of v with respect to basis B.
$[v]_B = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$
i.e., if $v = x_1v_1 + x_2v_2 + ... + x_nv_n$ then $[x_1, x_2, ..., x_n]^T$ is the coordinate vector of v with respect to $(v_1, v_2, ..., v_n)$

2.0.3 Inner Product

An inner product on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in R$ i.e $\langle ., .\rangle : V \times V \longrightarrow R$ and has following properties:

- Positivity: $\langle v, v\rangle \geq 0 \; \forall v \in V$
- Definiteness: $\langle v, v\rangle = 0$ iff $v = 0$
- Additivity: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
- Homogenity: $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$
- Symmetry: $\langle u, v\rangle = \langle v, u\rangle$

The dot product of two n tuples $u(u_1, u_2, ..., u_n)$ and $v(v_1, v_2, ..., v_n)$ defined by $\langle u, v\rangle = u^Tv = \sum_{i=1}^{n} u_iv_i$ is called the standard inner product in R^n and will be of the principal interest in this report.

Inner Product Space$(V(R), \langle\rangle)$**:** A (real)vector space with an inner product defined on it is called inner product space. For example, R^n with standard inner product defined above is called **Euclidean space** $(R^n, \langle\rangle)$.

The inner product induces the notions of length(norm) and distance(metric).

Norm(Length): Length or norm of a vector v in an inner product space is denoted be $||v||$ and is defined as $||v|| = +\sqrt{\langle v, v\rangle}$ or $||v||^2 = \langle v, v\rangle$. The norm satisfies following axioms:

- $\forall v \in V : \quad ||v|| \geq 0; \; ||v|| = 0$ iff $v = 0$.
- $\forall \alpha \in R, \forall v \in V : \quad ||\alpha v|| = |\alpha|||v||$.
- $\forall u, v \in V : \quad ||u + v|| \leq ||u|| + ||v||$. ...(Triangle Inequality)

Metric(Distance): The distance between two vectors $u, v \in V$ in an inner product space is defined by $d(u, v) = ||u - v||$. It satisfies following axioms:

- $\forall u, v \in V: \quad d(u, v) \geq 0; d(u, v) = 0$ iff $u = v$.
- $\forall u, v \in V: \quad d(u, v) = d(v, u)$. ...(Symmetry)
- $\forall u, v, w \in V: \quad d(u, w) \leq d(u, v) + d(v, w)$. ...(Triangle Inequality)

In the following, we assume the Euclidean Space $(R^n, \langle\rangle)$ where $\langle\rangle$ refers to the standard inner product.

Orthogonality(Perpendicularity): Two vectors v_i and v_j are orthogonal with an inner product$\langle\rangle$ if their inner product is zero, i.e., $\langle v_i, v_j\rangle = 0$.

Lemma 2.3. *Orthogonal Vectors are (linearly) independent.*

Proof. Let u and v be orthogonal and (linearly) dependent. That means for some scalars $c_1, c_2 \in R$ we have:

$$c_1 u + c_2 v = 0$$

Taking inner product with u on LHS, we get:

$0 = \langle(c_1 u + c_2 v), u\rangle = \langle c_1 u, u\rangle + \langle c_2 v, u\rangle = c_1\langle u, u\rangle + 0 = c_1 = 0$

Similarly taking inner product with v, we get:

$0 = \langle(c_1 u + c_2 v), v\rangle = \langle c_1 u, v\rangle + \langle c_2 v, v\rangle = 0 + c_2\langle v, v\rangle = c_2 = 0$

Therefore there is no relation $c_1 u + c_2 v = 0$ except for the trivial one in which c_1 and c_2 is zero, which contradicts the assumption of orthogonal vectors being linearly dependent.

Therefore orthogonal vectors are (linearly) independent. □

Orthogonal Basis: An orthogonal basis $B = (v_1, ..., v_n)$ of V is a basis whose vectors are mutually orthogonal: $\langle v_i, v_j\rangle = 0$ for all indices i and j with $i \neq j$.

Orthonormal Basis: An orthonormal basis $B = (v_1, ..., v_n)$ of R^n is a basis of orthogonal unit vectors (vectors of length one):

$\langle v_i, v_j\rangle = 0 \ \forall i, j \ i \neq j$

$||v_i|| = 1$ i.e $\langle v_i, v_j\rangle = 1 \ \forall i$.

The standard basis is easily seen to be orthonormal. Start from any basis of Euclidean space we can obtain orthonormal basis with the help of Gram Schmidt Procedure[7].

In all the discussion that follows, we work with only orthonormal bases.

Lemma 2.4. Parseval's Identity: *If $B = (v_1, ..., v_n)$ be orthonormal basis for V and for $v \in V$ we have $v = \sum_{i=1}^{n} x_i v_i$ then $||v||^2 = \sum_{i=1}^{n} x_i^2$.*

Proof. For a vector $v \in V$, we have

$$v = \sum_{i=1}^{n} x_i v_i$$

$$||v||^2 = || \sum_{i=1}^{n} x_i v_i ||^2$$

$$= \langle \sum_{i=1}^{n} x_i v_i, \sum_{j=1}^{n} x_j v_j \rangle$$

$$= \sum_{i=1}^{n} x_i \langle v_i, \sum_{j=1}^{n} x_j v_j \rangle$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j \langle v_i, v_j \rangle$$

As $\langle v_i, v_j \rangle = 0 \ \forall i, j \ i \neq j$ and $||v_i|| = 1$ i.e $\langle v_i, v_i \rangle = 1 \ \forall i$.

$$= \sum_{i=1}^{n} x_i^2$$

□

Lemma 2.5. *If $B = (v_1, v_2, ..., v_n)$ is an orthonormal basis for V. Suppose $v \in V$ have coordinates $(x_1, x_2, ..., x_n)^T$ with respect to B i.e., $v = \sum_{i=1}^{n} x_i v_i$ then $x_j = \langle v, v_j \rangle$*

Proof. We know that $v = x_1 v_1 + x_2 v_2 + ... + x_n v_n$.
Now $\langle v, v_j \rangle = \langle x_1 v_1, v_j \rangle + .. + \langle x_j v_j, vj \rangle .. + \langle x_n v_n, v_j \rangle$
$= x_1 \langle v_1, v_j \rangle + .. + x_j \langle v_j, vj \rangle .. + x_n \langle v_n, v_j \rangle$
$= x_j$
As for orthonormal basis $\langle v_i, v_j \rangle = 0$ if $i \neq j$ and $\langle v_i, v_i \rangle = 1$. □

2.0.4 Linear Transformation and Operators

Linear Transformation: A linear transformation $T : V \longrightarrow W$ from one vector space V to another vector space W is a map that is compatible with addition and scalar multiplication:

$$T(v_1 + v_2) = T(v_1) + T(v_2) \text{ and } T(cv_1) = cT(v_1),$$

for all v_1 and v_2 in V and all c in R.

Matrix of Linear Transformation: We have linear transformation $T : V \longrightarrow W$, ordered bases $B_v = (v_1, ..., v_n)$ of V and $B_w = (w_1, w_2, ..., w_m)$ of W. For each j, $1 \leq j \leq n$ $T(v_i)$ can be expressed in terms of basis of W i.e.,
$T(v_1) = \alpha_{11}w_1 + \alpha_{21}w_2 + ... + \alpha_{m1}w_m$
$T(v_2) = \alpha_{12}w_1 + \alpha_{22}w_2 + ... + \alpha_{m2}w_m$
$\vdots$
$T(v_n) = \alpha_{1n}w_1 + \alpha_{2n}w_2 + ... + \alpha_{mn}w_m$
for some α_{ij}, $1 \leq i \leq m$ and $1 \leq j \leq n$
In short $T(v_j) = \sum_{i=1}^{m} \alpha_{ij}w_i$
The coordinate vector $[T(v_j)]_{B_w} = \begin{pmatrix} \alpha_1j \\ \alpha_2j \\ . \\ . \\ \alpha_nj \end{pmatrix}$
Let $[x]_{B_v} = \begin{bmatrix} x_1 & x_2 & . & . & x_n \end{bmatrix}$ be the coordinate vector for $x \in V$. Then,

$T(x) = T(\sum_{j=1}^{n} x_j v_j)$
$T(x) = \sum_{j=1}^{n} x_j T(v_j)$
$T(x) = \sum_{j=1}^{n} x_j(\sum_{i=1}^{m} \alpha_{ij}w_i)$
$T(x) = \sum_{j=1}^{n}(\sum_{i=1}^{m} \alpha_{ij}x_j)w_i$
Now the coordinate vector of

$$[T(x)]_{B_w} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Let matrix $(\alpha_{ij}) = A$. Therefore,
$[T(x)]_{B_w} = A[x]_{B_v}$
The matrix A is matrix of linear transformation with respect to ordered bases B_v and B_w.

Linear Operator: A linear transformation from a vector space V to itself is called a linear operator.

Basis Transformation: Let $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, .., c_n)$ be two basis of vector space V. $T : R^n \longrightarrow R^n$ is a basis transformation operator from B to C if for each $v \in V$; T transforms the coordinate vector of v wrt basis B to the coordinate vector v wrt basis C.

Matrix of basis change: Let $[b_1, b_2, ..., b_n]$ and $[c_1, c_2, ...c_n]$ be bases of same vector space V. Any vector v with coordinate vector $[x_1, x_2, ..., x_n]$ and $[y_1, y_2, ..., y_n]$ can be represented in terms of linear combination of its basis, so

$v = x_1b_1 + x_2b_2 + ... + x_nb_n = y_1c_1 + y_2c_2 + ... + y_nc_n$

i.e,

$$\begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = \begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$$

As every vector can be represented in terms of basis, therefore:

$c_1 = \alpha_{11}b_1 + \alpha_{21}b_2 + ... + \alpha_{n1}b_n$

$c_2 = \alpha_{12}b_1 + \alpha_{22}b_2 + ... + \alpha_{n2}b_n$

$\vdots$

$c_n = \alpha_{1n}b_1 + \alpha_{2n}b_2 + ... + \alpha_{nn}b_n$

for some α_{ij}, $1 \leq i \leq n$ and $1 \leq j \leq n$

So,

$$\begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} \end{bmatrix}$$

Let $(\alpha_{ij}) = Q$ and $Q_i = \begin{bmatrix} \alpha_{1i} & \alpha_{2i} & .... & \alpha_{ni} \end{bmatrix}^T$

i.e

$$Q = \begin{bmatrix} Q_1 & Q_2 & ..... & Q_n \end{bmatrix}$$

Therefore , $\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = Q \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$ or $Q^{-1} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$

Here the matrix Q is the matrix of basis change from B to C .

Lemma 2.6. *The matrix of basis change Q is invertible.*

Proof. Just like Q is matrix of basis change from basis B to basis C, we have Q^{-1} for basis change from C to B.To prove Q is invertible, let us consider Q be singular.
If Q is singular then there is no Q^{-1} i.e Q is not invertible. Let Q' be the matrix of basis change from basis B to basis C.
Now take $QQ'[v]_B$:
$QQ'[v]_B = Q[v]_C = [v]_B$
Therefore $QQ' = I$ which means $Q' = Q^{-1}$ and hence Q is invertible. $\square$

Effect of Basis Transformation on operators: Any vector representation wrt one basis can have their equivalent representation wrt another basis. Let $(b_1, b_2, ..., b_n)$ and $(c_1, c_2, ...c_n)$ be two bases of vector space V. For any $v \in V$ having coordinates $(x_1, x_2, ..., x_n)$ and $(y_1, y_2, ...y_n)$ respectively in two bases, we have:
$v = x_1b_1 + x_2b_2 + ... + x_nb_n = y_1c_1 + y_2c_2 + ... + y_nc_n$
As we have seen in matrix of basis change, we get
$\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = Q \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$ or $Q^{-1} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$
Now if we have any linear transformation $T : V \longrightarrow W$ with matrix A and D as matrix of linear transformation for basis $(b_1, b_2, ..., b_n)$ and $(c_1, c_2, ..., c_n)$ respectively then we have:

$$A \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = QD \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$$

$$A \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix} = QDQ^{-1} \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$$

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

2.0.5 Hermitian and Unitary Operators

Orthogonal Matrix: A matrix Q such that $Q^TQ = I$ $(or\ Q^T = Q^{-1})$ is called an orthogonal matrix. A matrix Q is orthogonal if and only if its columns $Q_1, ..., Q_n$ are orthonormal with respect to the standard inner product form, i.e., if and only if $Q_i^TQ_i = 1$ and $Q_i^TQ_j = 0$ when $i \neq j$.

Lemma 2.7. *The basis transformation matrix between two orthonormal bases is an orthogonal matrix.*

Proof. Let $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, ..., c_n)$ be two orthonormal bases.
Let $\begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q$.
Let $Q = \begin{bmatrix} Q_1 & Q_2 & .....Q_n \end{bmatrix}$ $Q_i \in R^n$
Since $c_i = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q_i$
$c_j = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q_j$
We have:

$$\langle c_i, c_j \rangle = \langle \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q_i, \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q_j \rangle$$

$\langle c_i, c_j \rangle = \delta_{ij}$, where $\delta_{ij} = 0 \forall i \neq j$ and $\delta_{ij} = 1 \forall i = j$. Therefore,

$$\delta_{ij} = \langle c_i, c_j \rangle = \langle b_1 q_{i1} + b_2 q_{i2} + ... + b_n q_{in}, b_1 q_{j1} + b_2 q_{j2} + ...b_n q_{jn} \rangle$$

$$= q_{i1} q_{j1} + q_{i2} q_{j2} + ... + q_{in} q_{jn}$$

$$= Q_i^T Q_j$$

$$= \delta_{ij}$$

$$\Longrightarrow Q^T Q = I$$

□

Therefore from lemma 2.6 we get,

$$A = QDQ^T$$

Orthogonal Transformation(Unitary Transformation): A linear transformation $T : V \longrightarrow W$ is called a orthogonal transformation if $\forall v, w \in V$

$$\langle Tv, Tw \rangle = \langle v, w \rangle$$

Lemma 2.8. *Matrix of orthogonal transformation(unitary transformation) is orthogonal wrt to any orthonormal bases.*

Proof. Let $B = (b_1, b_2, ...b_n)$ be an orthonormal basis and let A be the matrix of orthogonal transformation T wrt basis B. Let $A = \begin{bmatrix} A_1 & A_2 & ... & A_n \end{bmatrix}$.
Now, $T(b_i) = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} A_i$
and $T(b_j) = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} A_j$
Then

$$\langle T(b_i), T(b_j)\rangle = \langle b_1 a_{i1} + b_2 a_{i2} + ... + b_n a_{in}, b_1 a_{j1} + b_2 a_{j2} + ...b_n a_{jn}\rangle$$

$$= a_{i1}a_{j1} + a_{i2}a_{j2} + ... + a_{in}a_{jn}$$

$$= A_i^T A_j$$

As orthogonal transformation preserves inner product, we have:

$$\langle T(b_i), T(b_j)\rangle = \langle b_i, b_j\rangle = \delta_{ij}$$

$$\Longrightarrow A_i^T A_j = \delta_{ij}$$

$$\Longrightarrow A^T A = I$$

□

Corollary: Let $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, ..., c_n)$ be two orthonormal bases of vector space V such that:
$\begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q$
where Q is the orthogonal matrix of basis change from B to C. Let $T : V \longrightarrow W$ be uniform(orthogonal) transformation with matrix A and D as matrix of uniform transformation for basis B and C respectively. Now using lemma 2.6 and 2.7 we can say if A is orthogonal matrix in orthogonal basis transformation then QAQ^T is also orthogonal.

Symmetric Operator(Hermitian Operator): A transformation $T : V \longrightarrow V$ is a symmetric operator if $\forall u, v \in V$ $\langle u, T(v)\rangle = \langle T(u), v\rangle$.

Lemma 2.9. *Matrix of a symmetric operator with respect to any orthogonal basis must be symmetric.*

Proof. Let $B = (b_1, b_2, ...b_n)$ be an orthonormal basis and let A be the matrix of symmetric transformation T wrt basis B. Let $A = \begin{bmatrix} A_1 & A_2 & ... & A_n \end{bmatrix}$.
Now, $T(b_i) = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} A_i$
and $T(b_j) = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} A_j$
According to the symmetric operator definition:

$$\langle b_i, T(b_j) \rangle = \langle T(b_i), b_j \rangle$$

$$\langle b_i, T(b_j) \rangle = \langle b_i, b_1 a_{j1} + b_2 a_{j2} + ...b_n a_{jn} \rangle = a_{ji}$$

Similarly,

$$\langle T(b_i), b_j \rangle = \langle b_1 a_{i1} + b_2 a_{i2} + ... + b_n a_{in}, b_j \rangle = a_{ij}$$

Therefore $a_{ij} = a_{ji}$ which implies $A = A^T$ □

Corollary: Let $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, ..., c_n)$ be two orthonormal bases of vector space V such that:
$\begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q$
where Q is the orthogonal matrix of basis change from B to C. Let $T : V \longrightarrow W$ be symmetric(hermitian) transformation with matrix A and D as matrix of symmetric operator for basis B and C respectively. Now using 2.8 we can say if A is symmetric matrix then QAQ^T is also symmetric.

2.0.6 Subspaces and Direct Sum

Independent Subspaces: Let V be real vector space and let U and W be subspaces of V.Two subspaces U and W are independent if
$U \cap W = \{0\}$.
Similarly if $V_1, V_2, ..V_k$ are the subspaces of V, then they are independent if $V_i \cap V_j = \{0\}\ \forall i \neq j$

Direct Sum: Let U and W be subspaces of V. We say V is a direct sum of U and W if $V = U + W$ and U and W are linearly independent . We write $V = U \oplus W$. Similarly, we say if $V_1, V_2, .., V_k$ are the subspaces of V, where $V = V_1, V_2, ..V_k$ and they are independent then we can write, $V = V_1 \oplus V_2 \oplus ... \oplus V_k$

Lemma 2.10. *If $V = U \oplus W$, $\forall v \in V$ there exist unique $u \in U$ $w \in W$ such that $v = u + w$.*

Proof. Let $v = u + w$ be not unique, there exist u' and w' such that $v = u' + w'$. Now

$$v = u + w = u' + w'$$

$$\implies u - u' = w' - w = 0$$

As $U \cap W = 0$

$$\implies u = u', w = w'$$

Therefore, there exist unique $u \in U$ $w \in W$ such that $v = u + w$. $\square$

Orthogonal Subspaces: Two subspaces U and W of V are orthogonal if $\forall u \in U, w \in W$ $\langle u, w\rangle = 0$. Note that orthogonal subspaces are necessarily independent.

Orthogonal Complements: Let W be subspace of V with $dim(W) = k$, where k is $0 \leq k \leq n$. We define orthogonal complement as:
$W^{\perp} = \{w' \in V | \langle w', w\rangle = 0 \, \forall w \in W\}$

Lemma 2.11. *$dim(W^{\perp}) = dim(V) - dim(W) = n - k$*

Proof. Let $[b_1, b_2, ..., b_k]$ and $[c_1, c_2, ..., c_m]$ be orthogonal bases for W and $W^{\perp}$. To prove the lemma we need to prove that $(b_1, b_2, ..., b_k, c_1, c_2, ..., c_m)$ spans V i.e., it is the basis of V which in turn will be proved if we can prove that for any $v \in V$ we can represent it a:
$v = \alpha_1 b_1 + ... + \alpha_k b_k + \beta_1 c_1 + ... + \beta_m c_m$

Consider $v' = v - \langle v, b_1\rangle b_1 - ... - \langle v, b_k\rangle - \langle v, c_1\rangle c_1 - ... - \langle v, c_m\rangle$.
ETPT: $v' = 0$ Suppose $v' \neq 0$

$$\langle v', b_i\rangle = \langle v, b_i\rangle - \langle\langle v, b_i\rangle b_i, b_i\rangle$$

$$= \langle v, b_i\rangle - \langle v, b_i\rangle\langle b_i, b_i\rangle$$

$$= 0$$

which results $v' \perp b_i \; \forall \; 1 \leq i \leq k$ i.e., $v' \in W^{\perp}$.
Similarly $\langle v', c_j \rangle = 0$. That means $v' = 0$.
Therefore $v = \alpha b_1 + ... + \alpha b_k + \beta c_1 + ... + \beta c_m$ for $v \in V$
Also $(b_1, b_2, ..., b_k, c_1, c_2, ..., c_m)$ is a basis of V(It spans V). This results into
$dim(V) = dim(W) + dim(W^{\perp})$
The $dim(V) = n$ and $dim(W) = k$, therefore
$dim(W^{\perp}) = dim(V) - dim(W) = n - k$ □

Corollary: If V is a vector space with W as a subspace of V and $W^{\perp}$ be the orthogonal complement then from previous lemma 2.10 we can say $V = W \oplus W^{\perp}$.

2.0.7 Projections

Projection: Let V be a finite dimensional vector space and $P : V \longrightarrow V$ be a linear operator on it. P is a projection if:

$$P^2 = P$$

In the following U is the range i.e image $Img(P)$ and W is the kernel $Ker(P)$ of P.

Theorem 2.12. (Rank-Nullity Theorem) *Let V and V' be vector spaces and let $T : V \longrightarrow V'$ be a linear transformation. Assuming the dimension of V is finite then,*

$$dim(V) = dim(Ker(T)) + dim(Img(T))$$

Its proof can be found in [7](Dimension Formula).

Lemma 2.13. *If P is a projection then $U \cap W = \{0\}$*

Proof. Let $v \in U \cap W$.
As $v \in U$; $\exists u \in V$ such that

$$Pu = v$$
$$P(Pu) = P^2u = Pu = v$$

Also $v \in W$ too, so by definition $Pv = 0$. Hence,

$$P(Pu) = Pv = 0$$

This is contradiction, therefore $U \cap W = 0$ i.e V is a direct sum of U and W.

$\square$

Corollary: By Rank-Nullity Theorem we have,

$$dim(V) = dim(U) + dim(W)$$

$$\text{i.e } V = U \oplus W \longrightarrow \text{direct sum}$$

Lemma 2.14. *If P is a projection operator then $V = Ker(P) \oplus Img(P)$*
Proof by lemma 2.12 and Rank-Nullity Theorem.

Lemma 2.15. *$u \in Img(P)$ for a projection operator P if and only if $Pu = u$*

Proof. $u \in Img(P)$ then $v \in V$ such that,

$$P(v) = u$$
$$P(P(v)) = Pu$$
$$\text{Also, } P(P(v)) = P^2v = Pv = u$$
$$\text{Therefore } Pu = u.$$

Using Lemma 2.12 and 2.14 $\exists u, w$ such that $v = u + w$ where $u \in Img(P)$ and $w \in Ker(P)$
Moreover $y \neq 0$ otherwise $u = v \in Img(P)$ then $P(v) = P(u + w) = u$

$$\begin{aligned} P(u + w) &= P(u) + P(w) \\ &= u + 0 \\ &= u \neq v \\ Pv = v &\Longleftrightarrow v \in Img(R) \end{aligned}$$

$\square$

Lemma 2.16. *$(I - P)$ is also a projection operator.*

Proof. Taking the square of $I - P$, we get:

$$(I - P)^2 = I - 2P + P^2 = I - P$$

Now to prove $Img(P) = Ker(I - P)$, we have,

$$(I - P)u = 0 \iff u = Pu \iff u \in Img(P) \text{ ....(Lemma 3 and 4)}$$

Now to prove $Img(I - P) = Ker(P)$ we have,

$$v \in Img(I - P) \text{ iff } (I - P)v = v$$
$$\text{i.e } Pv = 0 \iff v \in Ker(P)$$
$$Img(I - P) = Ker(P)$$

□

Orthogonal Projection Let V be finite-dimensional vector space and $P : V \longrightarrow V$ be a linear operator on it. P is an orthogonal projection if:

- P is a projection.
- $\langle u, Pw \rangle = \langle Pu, w \rangle \; \forall u, w \in V$

Lemma 2.17. *If P is an orthogonal projection then $Img(P) \perp Ker(P)$*

Proof. Suppose $u \in Img(P)$ and $w \in Ker(P)$
For P to be orthogonal projection we need to prove:

$$\langle u, w \rangle = \langle Pu, w \rangle = \langle u, Pw \rangle = 0$$

We know that any vector v can be represented as :

$$v = \sum_{i=1}^{n} \alpha_i b_i$$

$$\text{Also, } \langle v, b_j \rangle = \alpha_j$$

This makes the representation of vector v as,

$$v = \sum_{i=1}^{n} \langle v, b_i \rangle b_i$$

Now V has orthonormal basis $B = (b_1, b_2, ..., b_n)$ and $V = U \oplus W = Img(P) \oplus Ker(P)$. We can assume that basis for subspace U is $b_1, b_2, ..., b_k$ and for W is $b_{k+1}, ..., b_n$.
Therefore, $u = \sum_{i=1}^{k} \langle v, b_i \rangle b_i$ Similarly $w = \sum_{i=k+1}^{n} \langle v, b_j \rangle b_j$

$$\langle Pu, w \rangle = \langle u, w \rangle = \langle \sum_{i=1}^{k} \langle u, b_i \rangle b_i, \sum_{i=k+1}^{n} \langle v, b_j \rangle b_j \rangle = 0$$

Similarly,

$$\langle u, Pw \rangle = \langle \sum_{i=1}^{k} \langle v, b_i \rangle b_i, 0 \rangle = 0$$

$$\langle u, w \rangle = \langle Pu, w \rangle = \langle u, Pw \rangle = 0$$

□

Corollary: Let P be a projection operator on V. Let A be matrix of P wrt any orthonormal basis B. Then A satisfies following:

- $A^2 = A$
- If P is orthogonal then $A^T = A$

Lemma 2.18. *Let P be a projection on V. Let U be the $Img(P)$ and W be the $Ker(P)$ such that $U = W^{\perp}$. For every $v \in V$, $v = u + w : u \in U, v \in W$. Define $Pv = u$ then P is orthogonal projection with $Img(P) = U$ and $Ker(P) = W$.*

Proof. Let $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$
we have to prove that $\langle Pv_1, v_2 \rangle = \langle v_1, Pv_2 \rangle$
$\langle Pv_1, v_2 \rangle = \langle P(u_1 + w_1), u_2 + w_2 \rangle = \langle u_1, u_2 + w_2 \rangle$
$\qquad = \langle u_1, u_2 \rangle + \langle u_1, w_2 \rangle = \langle u_1, u_2 \rangle + 0 = \langle u_1, u_2 \rangle$
$\langle v_1, Pv_2 \rangle = \langle u_1 + w_1, P(u_2 + w_2) \rangle = \langle u_1 + w_1, u_2 \rangle$
$\qquad = \langle u_1, u_2 \rangle + \langle u_2 + w_1 \rangle = \langle u_1, u_2 \rangle + 0 = \langle u_1, u_2 \rangle$
From above two lines we get $\langle Pv_1, v_2 \rangle = \langle v_1, Pv_2 \rangle$, therefore P is hermitian operator which satisfies definition of orthogonal projection. □

Direction: A unit vector $d \in V$ is called a direction.

Projection onto a direction: The projection onto direction d is defined as: $P_d(v) = \langle v, d \rangle d$

Lemma 2.19. *Let d be any direction and v be any vector in V as shown in figure 2.1 then:*

$$v - \langle v, d \rangle d \perp d$$

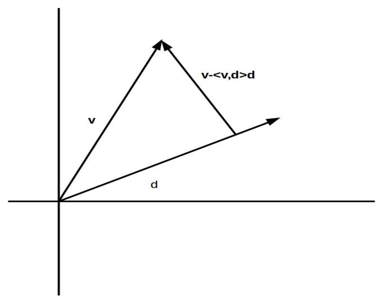

FIGURE 2.1: Projection

Proof. Take the inner product of $v - \langle v, d \rangle d$ and d

$\langle v - \langle v, d \rangle d, d \rangle = \langle v, d \rangle - \langle v, d \rangle \langle d, d \rangle$

$=\langle v, d \rangle - \langle v, d \rangle$

$=0$

There inner product is zero which proves they are perpendicular. $\square$

That means $v = \langle v, d \rangle d + (v - \langle v, d \rangle d)$

Projection of v i.e $P_d(v) = \langle v, d \rangle .d$

Lemma 2.20. *Let d be the direction in V. Let $B = (b_1, b_2, ..., b_n)$ be the basis of V. We can represent d as:$d = x_1b_1 + x_2b_2 + ... + x_nb_n$, where $\begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix}^T$ is the coordinate vector of d.*

Matrix of projection wrt basis B can be represented as:$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix}$

Proof. Let $B = (b_1, b_2, ..., b_n)$ be the orthonormal basis of V. Let $d = x_1b_1 + x_2b_2 + ... + x_nb_n$ be a direction. Let $v = \sum_{i=1}^{n} \alpha_i b_i$
Now

$$P_d(v) = \langle \sum_{i=1}^{n} \alpha_i b_i, \sum_{j=1}^{n} x_j b_j \rangle.(x_1b_1 + x_2b_2 + ... + x_nb_n)$$

Let A_d be the matrix of projection in the direction d then from above line we get:

$$A_d = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix}$$

$\square$

Lemma 2.21. *P_d is an orthogonal projection.*

Proof. From lemma 2.18 and 2.19 we can say P_d is orthogonal projection as
$v = \langle v, d \rangle d + (v - \langle v, d \rangle d)$
where $\langle v, d \rangle d \in Img(P)$ and $(v - \langle v, d \rangle d) \in Ker(P)$. $\square$

Note that, matrix A_d satisfies following properties:

- $A_d^2 = A_d$
- $A_d = A_d^T$

Projection onto a subspace: Let U be the subspace of V. For any $v \in V$, let $v = u + u^{\perp}$ be the unique expression for v such that $u \in U$ and $u^{\perp} \in U^{\perp}$. We define:

$$P_U(v) = u$$

as projection onto a subspace U.
P_U is a projection because, $P_U(v + v') = P_U(v) + P_U(v')$ and $P_U(\alpha v) = \alpha P_U(v)$
$P_U^2 = P_U = P_U^T$ If $b_1, b_2, .., b_k$ be an orthonormal basis of U and $b_{k+1}, ..., b_n$ be an orthonormal extension of U to V then,

$$P_U(v) = \sum_{i=1}^{k} \langle v, b_i \rangle .b_i$$

Chapter 3

Hermitian Operators in Euclidean Spaces

In the following we study Hermitian operators over a finite dimensional real vector space V having orthonormal basis $B = (b_1, b_2, ..., b_n)$

3.0.8 Eigenvalues and Eigenvectors

Eigenvalues of a operator: Let $T : V \longrightarrow V$ be a operator, then $\lambda \in R$ is a real eigenvalue of T if there exist $v \neq 0, v \in V$ such that:
$T(v) = \lambda v$

Eigenvector of a operator: Let $T : V \longrightarrow V$ be a operator, then $v \neq 0, v \in V$ is an eigenvector if there exist a scalar(eigenvalue) $\lambda \in R$ such that:
$T(v) = \lambda v$

Lemma 3.1. *Let $T : V \longrightarrow V$ be a operator. λ is an eigenvalue of T iff λ is a root of the characteristic equation $det(T - xI) = 0$*

Proof. For λ to be eigenvalue there exist $v \neq 0$ such that:

$$Tv = \lambda v$$

$$\Longrightarrow (T - \lambda I)v = 0$$

where $v \neq 0$ which implies $T - \lambda I$ is singular i.e., $det(T - \lambda I) = 0$
Therefore λ is a root of the characteristic equation $det(T - xI) = 0$ □

As $det(T - xI)$ is a polynomial of degree n, we have:
Corollary: $T : V \longrightarrow V$ has at most n distinct eigenvalues.
This is because $det(T-xI)$ is a real polynomial with degree n and a real polynomial with degree n has at most n roots over R. (Refer[7].)

Eigenvalues and eigenvectors of a matrix: A square matrix A has an eigenvalue λ with the corresponding non-zero eigenvector $x \in R^n$ if $Ax = \lambda x$.

Let $\lambda_1, ..., \lambda_k$ be eigenvalues of matrix A with multiplicities $d_1, ..., d_k$. We have relation $\sum d_{i=1}^{k} = n$

Lemma 3.2. *Eigenvalues of a matrix does not depend on basis.*

Proof. Let A be a matrix. After basis change we get matrix QAQ^{-1}, where Q is a matrix of basis change from basis B to basis C. Now,

$$det(QAQ^{-1} - xI) = det(QAQ^{-1} - xQQ^{-1})$$

$$= det(Q(A - xI)Q^{-1})$$

$$= det(Q)det(A - xI)det(Q^{-1})$$

$$= det(A - xI)$$

As $det(Q^{-1}) = \frac{1}{det(Q)}$. Therefore characteristic equation doesn't change with basis change. □

Corollary: The eigenvalue of a matrix is essentially the eigenvalue of the operator represented by the matrix in the standard basis.

Change in eigenvectors with basis change: For any two orthonormal bases $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, ..., c_n)$ of vector space V we have relation.
$\begin{bmatrix} c_1 & c_2 & ..... & c_n \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & ..... & b_n \end{bmatrix} Q$
where Q is the matrix of basis change from B to C.
Let T be a linear transformation having matrix A and D wrt basis B and C.

Let $v_1, v_2, ..., v_k$ are eigenvectors of matrix A and $v'_1, v'_2, ..., v'_k$ be eigenvectors of matrix D. As the eigenvectors are orthogonal we can easily make them orthonormal vectors(Gram Schmidt Procedure) which are orthonormal bases for V. Let $x_1, x_2, ..., x_k \in R^n$ and $y_1, y_2, ..., y_k \in R^n$ be coordinate vectors of eigenvectors of A and D respectively, then we have :
$v = x_1v_1 + x_2v_2 + ... + x_nv_n = y_1v'_1 + y_2v'_2 + ... + y_nv'_n$
i.e,
$\begin{bmatrix} v_1 & v_2 & ..... & v_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{bmatrix} v'_1 & v'_2 & ..... & v'_n \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$
Using the matrix of basis change from B to C, we get
$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = Q \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$ or $\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = Q^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$

Lemma 3.3. *Eigenvalues of real symmetric matrix must be real.*

Proof. See [7]. □

Lemma 3.4. *The eigenvectors corresponding to distinct eigenvalues of a hermitian operator(matrix) are orthogonal to each other.*

Proof. For a symmetric(hermitian) operator $T : V \longrightarrow V$ we have $\langle T(u), v\rangle = \langle u, T(v)\rangle$ for any vector $u, v \in R^n$.
Suppose u and v are eigenvectors of T with corresponding eigenvalues λ_1 and λ_2, we have:

$$\begin{aligned} \lambda_1\langle u, v\rangle &= \langle \lambda_1 u, v\rangle \\ &= \langle T(u), v\rangle \\ &= \langle u, T(v)\rangle \\ &= \langle u, \lambda_2 v\rangle \\ &= \lambda_2\langle u, v\rangle \end{aligned}$$

Which results into $\lambda_1 - \lambda_2\langle u, v\rangle = 0$.
Since the eigenvalues are distinct, $\lambda_1 - \lambda_2 \neq 0$, therefore $\langle u, v\rangle = 0$
i.e $u \perp v$
This is true for any two eigenvectors corresponding to distinct eigenvalues of symmetric matrix. □

Eigenspace: Let $T : V \longrightarrow V$ be a linear operator. Let $\lambda_1, \lambda_2, ..., \lambda_k$ be the eigenvalues of T where $k \leq n$. Then we define the eigenspace of λ_i as the set:
$E_{\lambda_i} = \{v \in V | Tv = \lambda_i v\}$

Lemma 3.5. *The set $E_{\lambda_i} = \{v \in V | Tv = \lambda_i v\}$ is a subspace of V.*

Proof. This follows as $\forall u, v \in E_{\lambda_i}$ and $c \in R$ we have:
$T(u + v) = \lambda_i(u + v)$ and
$T(cu) = c\lambda u$.
This shows that E_{λ_i} is a subspace of V □

Now we have a subspace E of V, so we also have an orthogonal complement $E_{\lambda_i}^{\perp}$ which can be defined as: $E_{\lambda_i}^{\perp} = \{u \in V | \langle u, v \rangle = 0 \, \forall v \in E_{\lambda_i}\}$
Using Lemma 2.10 we can say that $dim(E_\lambda) + dim(E_\lambda^{\perp}) = n$.

Suppose d_i be $dim(E_{\lambda_i})$ then $dim(E_{\lambda_i}^{\perp}) = n - d_i$ (from lemma 2.10).

Lemma 3.6. *Eigenspaces corresponds to distinct eigenvalues.*

Proof. Let $T : V \longrightarrow V$ be a linear operator and $\lambda_1, ..., \lambda_k$ be distinct eigenvalues of operator and E_{λ_i} is corresponding eigenspace associated to λ_i. Let E_{λ_i} may have two eigenvalues λ_i and λ_j.
Now for any vector $v \in E_{\lambda_i}$ we have:
$T(v) = \lambda_i v = \lambda_j v$
$\implies (\lambda_i - \lambda_j)v = 0$
$\implies v = 0$.
as eigenvalues are distinct. Therefore any eigenspace can have only one eigenvalue. □

Now we turn to Hermitian operators. The following is a fundamental property of Hermitian operators:

Theorem 3.7. *Let $T : V \longrightarrow V$ be Hermitian operator.Let E_{λ_i} be the eigenspace corresponding to λ_i and $E_{\lambda_i}^{\perp}$ be the orthogonal component. λ_i and $E_{\lambda_i}^{\perp}$ are T invariant. i.e.,*

$$\forall u \in E_{\lambda_i} \; Tu \in E_{\lambda_i}$$
$$\forall w \in E_{\lambda_i}^{\perp} \; Tw \in E_{\lambda_i}^{\perp}$$

Proof. Let $u \in E_{\lambda_i}$ then we need to prove $Tu \in E_{\lambda_i}$
ETPT: $T(Tu) = \lambda(Tu)$
Now, $T(Tu) = T(\lambda u) = \lambda^2 u$.
Similarly $\lambda(Tu) = \lambda(\lambda u) = \lambda^2 u$
Therefore $T(Tu) = \lambda(Tu)$ which means $Tu \in E_{\lambda_i}$

Let $w \in E_{\lambda_i}^{\perp}$ then we need to prove $Tw \in E_{\lambda_i}^{\perp}$
$w \in E_{\lambda_i}^{\perp}$ that means $\langle u, w\rangle = 0 \ \forall u \in E_{\lambda_i}$
ETPT: $\langle u, Tw\rangle = 0 \ \forall u \in E_{\lambda_i}$
Now, $\langle u, Tw\rangle = \langle Tu, w\rangle = \lambda\langle u, w\rangle = 0$ i.e., $\langle u, Tw\rangle = 0$
That means $Tw \in E_{\lambda_i}^{\perp}$ □

Lemma 3.8. *Eigenspaces of a linear operator $T : V \longrightarrow V$ intersects only at the origin (i.e., independent subspaces)*

Proof. Let $T : V \longrightarrow V$ be a Hermitian operator with distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_k$, where $k \leq n$ and corresponding subspaces $E_{\lambda_1}, E_{\lambda_2}, ..., E_{\lambda_k}$.
ETPT: $E_{\lambda_i} \cap E_{\lambda_j} = \{0\}$ for $i \neq j$. Let $u \in E_{\lambda_i} \cap E_{\lambda_j}$. Then,
$Tu = \lambda_i u$ as $u \in E_{\lambda_i}$
$Tu = \lambda_j u$ as $u \in E_{\lambda_j}$
$\Longrightarrow u = 0$ as $\lambda_i \neq \lambda_j$
Therefore $E_{\lambda_i} \cap E_{\lambda_j} = \{0\}$ for $i \neq j$ which means they are linearly independent. □

The following is a central decomposition theorem:

Theorem 3.9. *Let V be a real vector space, let $T : V \longrightarrow V$ be Hermitian operator with distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_k$, where $k \leq n$ then:*

1. $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus .... \oplus E_{\lambda_k}$.
2. *Each E_{λ_i} is T invariant.*
3. $\forall u \in E_{\lambda_i}, \forall v \in E_{\lambda_j}$ $\langle u, v\rangle = 0$ *if* $i \neq j$

Proof. Let $T : V \longrightarrow V$ be a Hermitian operator with distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_k$, where $k \leq n$. We have already proved in theorem 3.7 that (E_{λ_i}) and $(E_{\lambda_i}^{\perp})$ are invariant and $V = E_{\lambda_i} \oplus E_{\lambda_i}^{\perp}$.
Now $T_{E_{\lambda_i}^{\perp}}$ is Hermitian as $E_{\lambda_i^{\perp}}$ is a subspace of V which is invariant. Moreover $T_{E_{\lambda_i}^{\perp}}$

has eigenvalues $\lambda_1, ..\lambda_{i-1}, \lambda_{i+1}, ..\lambda_k$ with the same eigenspaces $E_{\lambda_1}, ..E_{\lambda_{i-1}}, E_{\lambda_{i+1}}, ..E_{\lambda_k}$. Now these eigenspaces are independent(Lemma 3.7) Therefore, by induction on $dim(E_{\lambda_i})$, we have:

$$E_{\lambda_i^\perp} = E_{\lambda_1} \oplus .. \oplus E_{\lambda_{i-1}} \oplus E_{\lambda_{i+1}} \oplus .. \oplus E_{\lambda_k}$$

and using lemma 2.10 we have:

$$V = E_{\lambda_i} \oplus E_{\lambda_i}^\perp$$

Therefore,

$$V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus ... \oplus E_{\lambda_k}$$

We have already seen in theorem 3.7 that E_{λ_i} is invariant which can be proved for each eigenspace..

Where as in lemma 3.4 we have already proved that eigenvectors corresponding to distinct eigenvalues are orthogonal to each other. Therefore, when we say $\forall u \in E_{\lambda_i}, \forall v \in E_{\lambda_j}$ we mean the eigenvectors corresponding to λ_i and λ_j which are orthogonal to each other. So $\langle u, v\rangle = 0$ if $i \neq j$. □

Lemma 3.10. *Let $T : V \longrightarrow V$ be a Hermitian operator. Let P be a projection on V and P_{λ_i} be projection onto subspace E_{λ_i}, then:*

$$P_{\lambda_1} + P_{\lambda_2} + ... + P_{\lambda_n} = I$$

Proof. T is a Hermitian operator. Let $\lambda_1, \lambda_2, ..., \lambda_k$ be distinct eigenvalues of T and E_{λ_i} be the respective eigenspace. Using theorem 3.9 we can say:

$$V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus ... \oplus E_{\lambda_k}$$

Let E_{λ_i} have the $dim(E_{\lambda_i}) = d_i$ and let $b_1^i, ..., b_{d_i}^i$ be the orthonormal basis of subspace E_{λ_i}
Let P_{λ_i} be projection onto subspace E_{λ_i} be:
$P_{\lambda_i}(v) = \sum_{j=1}^{d_i} \langle v, b_j^i\rangle b_j^i$ Now,

$$P_{\lambda_i}(v) = \sum_{j=1}^{d_i} \langle v, b_j^i\rangle b_j^i$$

Taking the sum of all projections onto subspace we have:

$$(P_{\lambda_1} + P_{\lambda_2} + ... + P_{\lambda_k})v = \sum_{l=1}^{k} \sum_{j=1}^{d_i} \langle v, b_j^i \rangle b_j^i$$

As we have $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus ... \oplus E_{\lambda_k}$, we can say that orthonormal basis of V is $B = (b_1, b_2, ..., b_n) = (b_1^1, .., b_{d_1}^1, ...b_1^k, .., b_{d_k}^k) = (B_1, B_2, ..., B_k)$
Also we have $v = \sum_{i=1}^{n} \alpha_i b_i$

$\langle v, b_j \rangle = \alpha_j$

so, $v = \sum_{i=1}^{n} \langle v, b_i \rangle b_i$

Combining above facts we get:

$$(P_{\lambda_1} + P_{\lambda_2} + ... + P_{\lambda_k})v = v_1 + v_2 + ... + v_k$$

where v_i is a vector wrt basis B_i. Now $v = v_1 + v_2 + ..., v_k$

$$(P_{\lambda_1} + P_{\lambda_2} + ... + P_{\lambda_k})v = v$$

$$(P_{\lambda_1} + P_{\lambda_2} + ... + P_{\lambda_k}) = I$$

□

3.0.9 Spectral Theorem

Theorem 3.11. Spectral Theorem(Projection Version) *Let $T : V \longrightarrow V$ be a Hermitian operator and $\lambda_1, \lambda_2, ..., \lambda_k$ be eigenvalues of the operator and P_{λ_i} be the projection onto subspace E_{λ_i}. Then,*

$$T = \lambda_1 P_{\lambda_1} + \lambda_2 P_{\lambda_2} + ... + \lambda_k P_{\lambda_k}$$

Proof. We have already proven in lemma 3.9 $\sum_{i=1}^{k} P_{\lambda_i} = I$ Now take $v = v_1 + ... + v_k$ with $v_i \in E_{\lambda_i}$ apply T, we get:

$$T(v) = T(v_1) + ... + T(v_k)$$

$$= \lambda_1 v_1 + \lambda_2 v_2 + ... + \lambda_k v_k$$

Now,

$$\sum_{i=1}^{k} \lambda_i P_i v = \sum_{i=1}^{k} \lambda_i P_i (v_1 + v_2 + ... + v_k)$$

$$= \lambda_1 v_1 + \lambda_2 v_2 + ... + \lambda_k v_k$$

$$= T(v)$$

Therefore,

$$T = \lambda_1 P_{\lambda_1} + \lambda_2 P_{\lambda_2} + ... + \lambda_k P_{\lambda_k}$$

□

Theorem 3.12. Spectral Theorem(Matrix Version) *Every $n \times n$ symmetric matrix can be decomposed as:*

$$A = QDQ^T,$$

where A is a $n \times n$ symmetric matrix, D is a diagonal matrix whose entries are eigenvalues of matrix A and Q is orthogonal matrix associated with eigenvectors of A.[6]
or:
According to spectral theorem for any $n \times n$ symmetric matrix A, [6]there exist exactly n, possibly not distinct eigenvalues $\lambda_1, ...\lambda_n$ and their associated eigenvectors $u_1, ...u_k$ where $k \leq n$ is number of non-zero eigenvalues and x_i is the coordinate vector of u_i such that

$$A = \sum_{i=1}^{k} \lambda_i x_i x_i^T = QDQ^T.$$

Proof. Let $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be (not necessarily distinct) eigenvalues of symmetric matrix A. Given any Hermitian operator(matrix) we can choose coordinate vectors $x_1, x_2, ...x_k \in R^n$ where $k \leq n$ such that they are associated to eigenvectors of A which are orthonormal. Let $\begin{bmatrix} x_1 & x_2 & ..... & x_k \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & ..... & e_k \end{bmatrix} Q$ where $e_1, e_2, .., e_k$ is the standard basis. Clearly
$Q = \begin{bmatrix} Q_1 & Q_2 & ..... & Q_k \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & ..... & x_k \end{bmatrix}$. Moreover, since eigenvectors are orthonormal i.e., $x_i's$ are orthonormal we have,
$Q^T = Q^{-1}$

Let $v \in R^n$ such that,

$$v = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_k x_k$$

$$Av = A(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_k x_k)$$

As $Ax_i = \lambda_i x_i$, we have:

$$Av = \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + ... + \alpha_k \lambda_k x_k$$

Also $v = \sum_{i=1}^{k} \alpha_i x_i \implies \alpha_i = \langle v, x_i \rangle$. Therefore we have:

$$Av = \sum_{i=1}^{k} \lambda_i \langle v, x_i \rangle x_i$$

$$Av = \sum_{i=1}^{k} \lambda_i x_i \langle v, x_i \rangle$$

$$Av = \sum_{i=1}^{k} \lambda_i x_i \langle x_i, v \rangle$$

$$Av = \sum_{i=1}^{k} \lambda_i x_i x_i^T v$$

$$A = \sum_{i=1}^{k} \lambda_i x_i x_i^T$$

We have $\sum_{i=1}^{k} \lambda_i x_i x_i^T$ which can written as:

$$\sum_{i=1}^{k} \lambda_i x_i x_i^T = \begin{bmatrix} x_1 & x_2 & ..... & x_k \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_k \end{bmatrix} \begin{bmatrix} x_1 & x_2 & ..... & x_k \end{bmatrix}^T$$

Without the loss of generality, we can have n eigenvalues(need not to be distinct) $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ such that we can write above equation as:

$$\sum_{i=1}^{n} \lambda_i x_i x_i^T = \begin{bmatrix} x_1 & x_2 & ..... & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & ..... & x_n \end{bmatrix}^T$$

As $\begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_k & \\ & & & 0 \end{bmatrix}$
where $k \leq n$ (k denotes the rank of matrix.)

$$= QDQ^T$$

$$= A$$

where A is a $n \times n$ symmetric matrix, D is a diagonal matrix whose entries are eigenvalues of matrix A and Q is orthogonal matrix associated with eigenvectors of A □

3.0.10 Positive Semidefinite Matrices

A $n \times n$ symmetric matrix A is a positive semidefinite matrix iff

$$y^T Ay \geq 0$$
$$\forall y \in R^n$$

We use the notation $A \succeq 0$ to indicate a symmetric positive semidefinite matrix(psd). Given $n \times n$ symmetric matrix A, the following lines are equivalent:

1. A symmetric matrix A is positive semidefinite.
2. All eigen values of A are non-negative.
3. $\exists V \in R^{k \times n}$ $k \leq n$ such that $A = V^T V$.
4. $A = \sum_{i=1}^{k} \lambda_i v_i v_i^T$, for $v_i \in R^n$. $v_i \perp v_j = 0$ if $i \neq j$ and $\forall i$ $||v_i|| = 1$.

Proof. Following is the proof for above statements[8],
$1 \Rightarrow 2$.
If $A \succeq 0$ then all eigen values of A are non-negative.
Any vector v is eigen vector of A if it satisfies $Av = \lambda v$, where λ is eigen value corresponding to v. Symmetric matrices have real-valued eigen values. Now we have equation :

$$Av = \lambda v$$

Multiplying both sides v^T we get,

$$v^T A v = v^T \lambda v = \lambda v^T v$$

By definition of positive semidefinite $v^T A v \geq 0$. Therefore, R.H.S of the equation $\lambda v^T v \geq 0$, also $v^T v \geq 0$. Thus $\lambda \geq 0$
$2 \Rightarrow 3$.
If all eigen values of A are non-negative then $\exists V \in R^{k \times n}$ $k \leq n$ such that $A = V^T V$. It is already been proved in spectral theorem that for a symmetric matrix A there exist (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ and have corresponding eigenvectors $u_1, u_2, ..., u_n$.

Let Λ be the diagonal matrix that contains eigenvalues $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ of A.
$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_k & \\ & & & 0 \end{bmatrix}$ where k is number of non zero eigenvalues and $k \leq n$
Let U be the orthogonal matrix associated with the eigenvectors $\begin{bmatrix} u_1 & u_2 & .... & u_n \end{bmatrix}$ of A. From theorem 3.12 we have

$$A = U \Lambda U^T$$

$$\text{i.e } A = \begin{bmatrix} u_1 & u_2 & .... & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix}$$

If $k < n$ then we can write it as: $A = \begin{bmatrix} u_1 & u_2 & .... & u_k \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_k & \\ & & & 0 \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{bmatrix}$

Since Λ has all diagonal values and we can decompose it into $\Lambda^{\frac{1}{2}} (\Lambda^{\frac{1}{2}})^T$. Using point (2) $\Lambda^{\frac{1}{2}}$ must have all entries real Thus,

$$A = U \Lambda^{\frac{1}{2}} (\Lambda^{\frac{1}{2}})^T U^T$$

where $\Lambda_{ij}^{\frac{1}{2}} = +\sqrt{\Lambda_{ij}}$

$$A = (U\Lambda^{\frac{1}{2}})(U\Lambda^{\frac{1}{2}})^T$$

Let $V^T = (U\Lambda^{\frac{1}{2}})$. Therefore we get the equation,

$$A = V^T V$$

$3 \Rightarrow 1$.
If $\exists V \in R^{k \times n}$ $k \leq n$ such that $A = V^T V$ then $A \succeq 0$.

$$y \in R^n, y^T A y = y^T V^T V y = (Vy)^T (Vy) = \langle Vy, Vy \rangle \geq 0.$$

$1 \Rightarrow 4$.
If a symmetric matrix A is psd then we have $A = \sum_{i=1}^{k} \lambda_i v_i v_i^T$, for $v_i \in R^n$. $v_i \perp v_j = 0$ if $i \neq j$ and $\forall i$ $||v_i|| = 1$
Using spectral theorem we can say any symmetric matrix A can be represented as:

$$A = QDQ^T = \sum_{i=1}^{k} \lambda_i x_i x_i^T$$

where $x_1, x_2, ..., x_k$ are orthogonal eigenvectors of A.

$4 \Rightarrow 1$.
Similarly according to Spectral theorem for any $n \times n$ symmetric matrix A, [6]there exist exactly n, possibly not distinct eigenvalues $\lambda_1, ...\lambda_n$ and their associated eigenvectors $u_1, ...u_k$ where $k \leq n$ are non-zero eigenvalues and x_i is the coordinate vector of u_i such that

$$A = \sum_{i=1}^{k} \lambda_i x_i x_i^T.$$

□

Chapter 4

Semidefinite Programming

Semidefinite programming is a relatively new field of optimization and is becoming a tool for improving the performance guarantees of many problem. Some of its application areas are operational research, convex constrained optimization, combinatorial optimization, control theory etc. All linear programs as well as strict quadratic programs can be represented as SDPs. One of the best example of application of semidefinite programming is in MAXCUT problem done by Goemans and Williamson [1]. They provided a .87856-approximation algorithm for the problem. There are other interesting applications of SDP in various fields.

4.1 Semidefinite Programming

4.1.1 SDP Formulation

Let S_n denote the set of symmetric $n \times n$ matrices. Let $X \in S_n$ and C is a constant matrix for objective function. $C(X)$ is a linear function of X which is defined as Frobenius inner product and denoted as:

$$C(X) = Tr(CX) = C \bullet X = \sum_{i,j} c_{ij}x_{ij} \text{ where } C = (c_{ij}) \text{ and } X = (x_{ij})$$

A semidefinite program Z with the objective function $C(X)$ and m linear equations that X must satisfy, can be formulated[2] [3] as :

$$\text{Minimize or Maximize } \sum_{i,j} c_{ij}x_{ij} \qquad ..(Z)$$

$$\begin{aligned} &\text{subject to } \sum_{i,j} a_{ijk}x_{ij} = b_k && k = 1.....,m \\ &x_{ij} = x_{ji} \quad \forall i,j && \text{....Symmetry Constraint} \\ &X = (x_{ij}) \succeq 0 && \text{....PSD Constraint} \end{aligned}$$

The variable x_{ij} denotes the element of matrix $X \in R^{n\times n}$ Let us see an example, we have $n = 3$, $m = 3$,

$$A_1 = \begin{pmatrix} 1 & 3 & 2 \\ 3 & 4 & 0 \\ 2 & 0 & 7 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 4 \\ 0 & 4 & 3 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 7 \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 3 & 7 \\ 3 & 8 & 0 \\ 7 & 0 & 5 \end{pmatrix}$$

$b_1 = 17, b_2 = 15, b_3 = 11$

X will be 3×3 symmetric matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$

$$\sum_{i,j} c_{ij}x_{ij} = 2x_{11} + 3x_{12} + 7x_{13} + 3x_{21} + 8x_{22} + 0x_{23} + 7x_{31} + 0x_{32} + 5x_{33}$$

$$= 2x_{11} + 6x_{12} + 14x_{13} + 8x_{22} + 0x_{23} + 5x_{33}$$

SDP: minimize $2x_{11} + 6x_{12} + 14x_{13} + 8x_{22} + 0x_{23} + 5x_{33}$

subject to

$x_{11} + 6x_{12} + 4x_{13} + 4x_{22} + 0x_{23} + 7x_{33} = 17$

$0x_{11} + 2x_{12} + 0x_{13} + 4x_{22} + 8x_{23} + 3x_{33} = 15$

$2x_{11} + 0x_{12} + 2x_{13} + 3x_{22} + 4x_{23} + 7x_{33} = 11$

$x_{ij} = x_{ji}$

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \succeq 0$$

4.2 SDP as Vector Programming

SDP is often experienced in the form of vector programs. Given a SDP, the equivalent vector program will have the variables which are vectors $v_i \in R^n$, where $dim(n)$ is number of vectors.
In a vector program there is an objective function and constraints that are linear in the inner product of these vectors. Typically we require that the vectors are unit vectors. A vector program V is formulated as:

$$\begin{gathered} \text{Minimize or Maximize } \textstyle\sum_{i,j} c_{ij}(v_i.v_j) \qquad ..(V) \\ \text{subject to } \textstyle\sum_{i,j} a_{ijk}(v_i.v_j) = b_k \ \forall k \\ v_i \in R^n \end{gathered}$$

Lemma 4.1. *The SDP (Z) and above given vector program(V) are equivalent[3].*

Proof. Semidefinite programming and vector programming are considered to be equivalent because from one form we get another one. This can be done by taking the solution of semidefinite programming $X \in R^{n\times n}$, $X \succeq 0$ and computing $X = V^TV$ for some column vector V in polynomial time(Cholesky Decomposition[10]). $X = V^TV$ can be solved in polynomial time with some small error that can be ignored. Taking v_i, i^{th} column of V is the solution for vector programs.
Therefore, $x_{ij} = v_i.v_j = v_i^T v_j$
Similarly for given $v_i \in R^n$ we can construct V and then X=V^TV, which is solution for SDP. □

Here is the vector program for the above example:
minimize $2v_1.v_1 + 6v_1.v_2 + 14v_1.v_3 + 8v_2.v_2 + 0v_2.v_3 + 5v_3.v_3$
subject to
$v_1.v_1 + 6v_1.v_2 + 4v_1.v_3 + 4v_2.v_2 + 0v_2.v_3 + 7v_3.v_3 = 17$
$0v_1.v_1 + 2v_1.v_2 + 0v_1.v_3 + 4v_2.v_2 + 8v_2.v_3 + 3v_3.v_3 = 15$
$2v_1.v_1 + 0v_1.v_2 + 2v_1.v_3 + 3v_2.v_2 + 4v_2.v_3 + 7v_3.v_3 = 11$
$v_i \in R^n$

Chapter 5

Max-cut Problem

Given an undirected graph $G(V, E)$ having non negative weight on all edges $(i, j) \in E$, the maxcut problem can be defined as "finding a partition of vertices into two disjoint subsets $(S, \bar{S})$ such that sum of the weights of the edges that crosses the cut is maximized."

IP formulation:

$\forall i \in V \; x_i = \begin{cases} 1, & \text{if } i \in S \\ 0, & \text{otherwise} \end{cases}$

$z_{ij} = \begin{cases} 1, & \text{if edge (i,j) crosses the cut} \\ 0, & \text{otherwise} \end{cases}$

$$\max \textstyle\sum_{(i,j) \in E} w_{ij} z_{ij}$$

$$\begin{gathered} \text{subject to } z_{ij} \leq x_i + x_j \quad \forall (i, j) \in E \\ z_{ij} \leq 2 - (x_i + x_j) \quad \forall (i, j) \in E \\ x_i \in \{0, 1\} \quad \forall i \in V \\ z_{ij} \in \{0, 1\} \quad \forall (i, j) \in E \end{gathered}$$

But this formulation doesn't provides any better solution. If we find LP optimal solution the integrality gap is still 1/2. So we need to reformulate it.

Quadratic Programming Formulation:

The present formulation also represents the max-cut problem.

$\forall i \in V y_i = \begin{cases} -1, & \text{if } i \in S \\ 1, & \text{otherwise} \end{cases}$

$$\max \tfrac{1}{2} \textstyle\sum_{(i,j)\in E} w_{ij}(1 - y_i y_j)$$

$$\begin{aligned} \text{subject to } y_i.y_i = 1 \quad \forall i \in V \\ y_i \in -1, 1 \quad \forall i \in V \end{aligned}$$

The above formulation represents the maxcut, but it's NP Complete. We consider further relaxation and try to apply vector programming relaxation.

Vector Programming Formulation:

Vector programming relaxations can be made by relaxing some of the constraints and extending the objective function to the larger space i.e ,

$$\max \tfrac{1}{2} \textstyle\sum_{(i,j)\in E} w_{ij}(1 - v_i.v_j)$$

$$\begin{aligned} \text{subject to } v_i.v_i = 1 \quad \forall i \in V \\ v_i \in R^n \quad \forall i \in V \end{aligned}$$

The above obtained formulation gives all possible solution of previous OPT of max-cut, therefore making it more general or extended version of max-cut. We can say that the optimal solution obtained $Z^* \geq OPT$ [4] because if we set vectors v_i to $(y_i, 0, ..., 0)$, we can get the quadratic program from the above vector program which is due to vector program being a generalization of the quadratic program.

As the vector program is equivalent to SDP we can solve the problem. Goemans and Williamson[1] have presented the solution for this with the expected value of atleast 0.878 times of OPT. Here we are trying to reformulate it and try to solve it using SDP and then present the Goemans and Williamson algorithm.

5.0.1 Solving The Formulation

The quadratic formulation of problem statement can be reformulated as:

$$Z = \frac{1}{2}\sum_{(i,j)\in E} w_{ij}(1 - y_i y_j)$$

$$= \frac{1}{4}(\sum_i \sum_j w_{ij} - \sum_i \sum_j w_{ij} y_i y_j)$$

$$= \frac{1}{4}(\sum_i \sum_j w_{ij} y_i^2 - \sum_i \sum_j w_{ij} y_i y_j)$$

$$= \frac{1}{4}(\sum_i (\sum_j (w_{ij})) y_i^2 - \sum_i \sum_j w_{ij} y_i y_j)$$

$$= \frac{1}{4}(y^T Diag(W_e) y - y^T W y)$$

Let $L = Diag(W_e) - W$

$$Z = \frac{1}{4}(y^T L y)$$

Also,
$y^T L y = trace(y^T L y) = trace(y^T (L y)) = trace(L y y^T) = trace(L X)$
where $X = y y^T$

Therefore the formulation Z is:

$$\max \frac{1}{4} trace(LX)$$

$$\begin{gathered}\text{subject to } X_{ii} = 1\\ rank(X) = 1\\ X \succeq 0\end{gathered}$$

The above formulation is equivalent to previous one($Z = \frac{1}{4}(y^T L y)$). As matrix X is rank one constrained and using spectral theorem(Theorem 3.11) we can say that for any such matrix X having rank 1 we have one and only one decompostion i.e $X = y y^T$ as rank 1 matrix has only one eigenvalue. Therefore we can say

$Z = \frac{1}{4}trace(LX) \longleftrightarrow \frac{1}{4}(y^T Ly)$
The above formulation has rank-1 constraint and can be termed as rank constrained form. By relaxing this rank constraint i.e removing it we obtain convex problem infact the SDP. This SDP is the relaxation of the original problem, that is, it is a new problem obtained by removing constraint. The SDP Z^* is:

$$\max \frac{1}{4}trace(LX)$$

$$\text{subject to } X_{ii} = 1$$
$$X \succeq 0$$

Using any semidefinite algorithm, one can obtain for $\varepsilon > 0$, a solution of value greater than $Z^* - \varepsilon$ in time polynomial in the input size and $log 1/\varepsilon$. There are various algorithms to solve semidefinite programming like ellipsoid algorithm, interior point algorithm and other polynomial time algorithms for convex programming.

5.0.2 Goemans-Williamson Algorithm

Goemans and Williamson in their paper gave rounding procedure for finding an approximate solution to the max-cut problem. Following is the random hyperplane rounding algorithm[5]:

- Solve the model Z. Let X^* be the optimal solution.
- Compute the Cholesky Decomposition for X^* i.e $X^* = V^T V$ where $v_i, i = 1...n$ is normalized column of V.
- Rounding procedure:
  Set $S = \emptyset$
  - Uniformly generate a random vector r on the unit n-sphere.
  - For $i = 1...n$, if $v_i^T.r \geq 0$ then $i \in S$ else $i \in \bar{S}$.
  - Find the weight after obtaining the cut.

The random vector $r = (r_1, r_2, .., r_n)$, each component is picked from $N(0, 1)$, the Normal Distribution with mean 0 and variance 1. The random unit vector

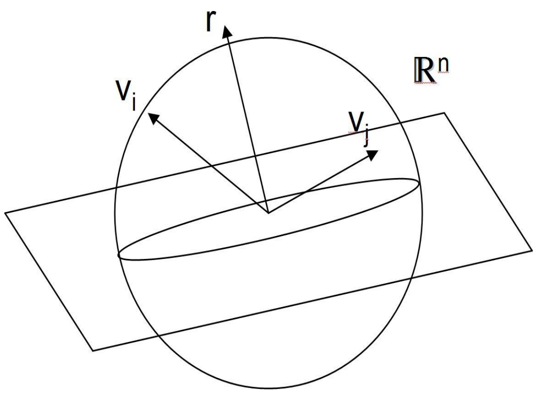

FIGURE 5.1: Random Hyperplane Rounding[5]

is equivalent to the random hyperplane with normal r containing the origin. As $v_i.v_i = 1$, therefore all the vectors lie in unit sphere. The random hyperplane splits the sphere into two half S and $\bar{S}$. This approximation algorithm has guaranteed performance of 0.878; however the solution obtained by this algorithm is better than 0.878 in many cases.

5.0.3 Analysis of the Algorithm

The GW Algorithm gives performance of 0.878. Practically it gives better than this. Let us analyze the algorithm.
The expected value of the max-cut depends on the fact that any two vector should have opposite sign. Let $v_1, v_2, .., v_n$ be the vectors, so the expected value of cut depends on the probability that two vectors are separated by the random hyperplane,i.e.:

$$E[W]= \sum_{(i,j)\in E} w_{ij}.Pr[(v_i.r \geq 0 \text{ and } v_j.r < 0) \text{ or } (v_j.r \geq 0 \text{ and } v_i.r < 0)]$$

Also we have,

$$Pr[(v_i.r \geq 0 \text{ and } v_j.r < 0) \text{ or } (v_j.r \geq 0 \text{ and } v_i.r < 0)] = Pr[sgn(v_i.r) \neq sgn(v_j.r)]$$

By combining above two equations we get,

$$E[W]= \sum_{(i,j)\in E} w_{ij}.Pr[sgn(v_i.r) \neq sgn(v_j.r)],$$

Now let us find the probability.

Lemma 5.1. $Pr[sgn(v_i.r) \neq sgn(v_j.r)]{=}\frac{1}{\pi}arccos(v_i.v_j) \geq 0.878\frac{1}{2}(1 - v_i.v_j)$

Proof. Assuming certain facts,
Fact 1: As r is uniform random vector over the unit sphere, the projection of r onto a plane is uniformly distributed on a unit circle.
Fact 2: The projection of r onto two vectors are independent and normally distributed.
To compute the probability consider a plane containing vectors v_i, v_j having angle θ between them. The random vector r has two components i.e $r = r' + r_\perp$. r' is the component of r that lies on the plane where as $r_\perp$ is orthogonal to the plane. Therefore $v_i.r = v_i.(r' + r_\perp) = v_i.r'$ as $v_i.r_\perp = 0$.
Similarly $v_j.r = v_j.r'$.

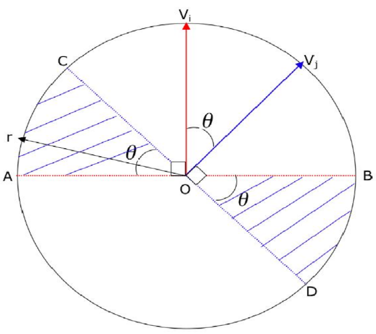

FIGURE 5.2: Figure for proof[5]

From the figure 2 we get if r lies in the arc AOC or arc BOD then $sgn(v_i.r) \neq sgn(v_j.r)$. We can also find that the angle in both the arc is θ. As a result,

$$Pr[sgn(v_i.r) \neq sgn(v_j.r)]{=}\frac{2\theta}{2\pi}$$

where $\theta = arccos(v_i.v_j)$
Therefore,

$$Pr[sgn(v_i.r) \neq sgn(v_j.r)]{=}\frac{arccos(v_i.v_j)}{\pi}$$

$$E[W]{=}\ \frac{1}{\pi}\sum_{(i,j)\in E} w_{ij} arccos(v_i.v_j)$$

Also,

$$min_{-1\leq x\leq 1}\frac{\frac{1}{\pi}arccos(v_i.v_j)}{\frac{1}{2}(1-v_i.v_j)} \geq 0.878$$

$$=> \frac{1}{\pi}arccos(v_i.v_j) \geq 0.878\frac{1}{2}(1 - v_i.v_j)$$

$$\text{i.e } Pr[sgn(v_i.r) \neq sgn(v_j.r)] \geq 0.878\frac{1}{2}(1 - v_i.v_j)$$

□

Lemma 5.2. $E[W] \geq 0.878\frac{1}{2}\sum_{(i,j)\in E} w_{ij}(1 - v_i.v_j)$

Proof. We expected value is:

$$E[W] = \sum_{(i,j)\in E} w_{ij}.Pr[sgn(v_i.r) \neq sgn(v_j.r)]$$

$$= \frac{1}{\pi}\sum_{(i,j)\in E} w_{ij} arccos(v_i.v_j)$$

$$\geq 0.878\frac{1}{2}\sum_{(i,j)\in E} w_{ij}(1 - v_i.v_j)$$

$$= 0.878 Z_{SDP}$$

$$\geq 0.878 Z_{OPT}$$

□

Chapter 6

Conclusion

With the given prerequisites in the thesis for semidefinite programming, one can understand the semidefinite programming. We have only given one example of semidefinite programming used in combinatorial problem(Max-cut). Other combinatorial problems like MAX-2SAT[1], k-colorability[4] etc. have also been studied using semidefinite programming, providing better result than previous solutions. Not many combinatorial problems have been studied till now but the field is still in its growing phase for combinatorial problems.

One can have chances of finding better solutions for existing combinatorial problems using semidefinite programming. Also one can study to find out whether there is any possiblity of having better complexity for existing solutions of semidefinite programming or any other approach to solve problems in semidefinite programming.

There lies lot of possibility in semidefinite programming for finding better approximation algorithms for combinatorial problems. Hence it is important tool which can be used in combinatorial optimizations.