An-Najah National University
Faculty of Graduate Studies

On Singular Value Decomposition of Rectangular Matrices

By

Shireen Najeh Issa Odeh

Supervisor
Dr."Mohammad Othman" Omran

This thesis is submitted in Partial Fulfillment of the Requirements for the Degree of Master in Science in Mathematics , Faculty of Graduate Studies, at An-Najah National University, Nablus, Palestine.

2009

On Singular Value Decomposition of Rectangular Matrices

By

Shireen Najeh Odeh

This thesis was defended successfully on 29/10/2009 and approved by

Committee Members		Signature
1- Dr. "Mohammad Othman" Omran	Supervisor	…………
2- Dr. Samir Matar	Internal Examiner	…………
3- Dr. Ali Abed Almohesen	External Examiner	…………

Dedication

To my parents "Najeh and Nahla"

Acknowledgements

There are many people who have directly or indirectly made it possible for me to continue on this journey.

I am deeply grateful to my supervisor Dr. "Mohammad Othman" Omran who without his support, kind supervision, helpful suggestions and valuable remarks, my work would have been more difficult.

Also, my great thanks are due to my family for their support, encouragement and great efforts for me.

I would also sincerely like to thank my teachers and friends for helping me during working on this thesis.

الإقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان

On Singular Value Decomposition of Rectangular Matrices

حول تحليل القيمة المنفردة للمصفوفات المستطيلة

اقر بأن ما اشتملت عليه هذه الرسالة إنما هي نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وأن هذه الرسالة ككل، أو أي جزء منها لم يقدم من قبل لنيل أية درجة علمية أو بحث علمي أو بحثي لدى أية مؤسسة تعليمية أو بحثية أخرى.

Declaration

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualifications.

Student's name:	اسم الطالب:
Signature:	التوقيع:
Date:	التاريخ:

List of Contents

On Singular Value Decomposition of Rectangular Matrices

Prepared by
Shireen Najeh Odeh

Supervisor
Dr."Mohammad Othman" Omran

Abstract

The singular value decomposition of matrices stands as one of the most important concepts in mathematics, because of its variety of applications in mathematics, statistics, biology and many other areas of science.

In this thesis, we present the singular value decomposition and its relation to the spectral decomposition . We also investigate the singular value decomposition of a matrix together with some of its applications. Some of these applications include the Moore-Penrose psuedoinverse, the effective rank of matrices and image compression.

Introduction

The singular value decomposition (SVD) plays an important role in matrix theory. While some decompositions are restricted to real square matrices, the (SVD) can be applied to any rectangular matrix *A*. Through this decomposition, we study some of the properties of *A* such as: the rank, the norm and the basis of the four fundamental subspaces related to *A*. It also has many applications. In numerical analysis, the SVD provides a measure of the effective rank of a given matrix. In statistics, the SVD is a particularly useful tool for finding least-squares solutions and approximations. It has many applications in: signal processing, biology, statistical analysis and mathematical modeling.

In this thesis, two types of the (SVD) of *A* are defined and computed: the full and reduced; this decomposition is compared with the well known spectral decomposition of *A* (whenever exists); some properties of the matrix via its (SVD) are studied. The (SVD) is used to compute the Moore-Penrose pseudo inverse that can be used in solving a system of linear equations and can give the optimal solution of the least squares problem when solving an overdetermined system; we also use the (SVD) to compute the best low rank approximation according to either the Euclidean or the spectral norm.

In the first chapter, some preliminary definitions are presented, as well as basic results and properties of matrices; some special matrices (unitary, normal and Hermitian) are reviewd. Eigenvalues, norms and computations of the condition number are also studied.

In the second chapter, the diagonalization of matrices is studied and that includes Schur's theorem and the spectral decomposition.

In the third chapter, the singular value decomposition (SVD) is defined and its relation to the spectral decomposition is studied. Some properties of the original matrix are studied via its SVD and the geometric interpretation of SVD is also introduced.

In the fourth chapter, the SVD is used to compute the Moore-Penrose pseudoinverse which is used for solving linear systems of equations.

In the fifth chapter, the SVD is used in other applications such as, low rank approximation with respect to a given norm, image compression and finding the affective rank of a matrix.

History

The singular value decomposition has a long history. It was originally developed in the nineteenth century by differential geometers and algebraists who wanted to determine, for given matrices $A=[a_{ij}]$ and $B=[b_{ij}]\in M_n(R)$, whether the two bilinear forms:

$\Phi_A(x,y)=\sum_{i,j=1}^{n} a_{ij}x_i y_j$ and $\Phi_B(x,y)=\sum_{i,j=1}^{n} b_{ij}x_i y_j$,

could be made equal for every $x=[x_i]$ & $y=[y_i]\in R^n$, under independent real orthogonal transformation of the two spaces it acts on; i.e, does there exist $Q_1,Q_2\in M_n(R)$ such that $\Phi_B(Q_1x,Q_2y)=\Phi_A(x,y)$

for all $x,y\in R^n$?

This problem could be approached by finding a canonical form to which any such bilinear form can be reduced by orthogonal substitution, or by finding a complete set of invariants for a bilinear form under orthogonal substitutions.

The Italian differential geometer Eugenio Beltrami discovered in 1873 that for each real matrix $A\in M_n(R)$, there are always $Q_1,Q_2\in M_n(R)$ such that

$$(*)\qquad Q_1^T A Q_2=\Sigma=\begin{bmatrix} s_1(A) & & & \\ & s_2(A) & & \\ & & . & \\ & & & s_n(A)\end{bmatrix},$$

where $s_1^2(A) \geq s_2^2(A) \geq \ldots \geq s_n^2(A) \geq 0$ are the eigenvalues of AA^T as well as $A^T A$; he also found that the columns of Q_1 are eigenvectors of AA^T and the columns of Q_2 are eigenvectors of $A^T A$.

Independently, in 1874, the French algebraist Camille Jordan came to the same canonical form but from a different point of view. He found that the eigenvalues of the 2*n*-*by*-2*n* real symmetric matrix $\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ are paired by sign and that its *n* largest eigenvalues are the desired coefficients $s_1(A),\ldots,s_n(A)$ of the canonical form $\Phi_{Q_1^T A Q_2}(x,h) = s_1(A)x_1h_1 + \ldots + s_n(A)x_nh_n$.

In 1889/90 unaware of Beltrami and Jordan, James Joseph Sylvester gave a third proof to (*) for real square matrices and he called the s's the canonical multipliers of the matrix *A*.

In 1902 L-Autone proved that every non singular complex matrix $A \in M_n$ can be written as $A = UP$, where $U \in M_n$ is unitary and $P \in M_n$ is positive definite. In 1913/15 he returned to these ideas and used the similarity of AA^* and $A^* A$ to show that any square complex matrix $A \in M_n$ can be written as $A = U\Sigma V^*$ where $U, V \in M_n$ are unitary and $\Sigma \in M_n$ is a nonnegative diagonal matrix. He also discovered that if *A* is nonsingular Hermitian then *A* can be written as $U\Sigma U^T$ for some unitary *U*, and a nonnegative diagonal matrix Σ. In 1910 Emile Picard call the numbers s_k's singular values.

In 1939 Eckart and Young gave the first complete proof of the singular value decomposition for rectangular complex matrix and they didn't give any name to the numbers s_k's.

The existence proof of the singular value decomposition opens many ways for the mathematician to search for inequalities, properties and applications to this decomposition.

During 1949-50, a remarkable series of papers in the Proceeding of the National Academy of Science (U.S.) established all of the basic inequalities involving singular values and eigenvalues. One of these papers is "Inequalities Between the Two Kinds of Eigenvalues of a Linear Transformation", established by Weyl.

In 1950 Poyla gave an alternative proof of a key lemma in Weyl's 1949 paper(also, established by U.S.).

In 1954, A. Horn proved that Weyl's 1949 inequalites were sufficient for the existence of a matrix with prescribed singular values and eigenvalues, and in this paper he used the expression "singular values" in the context of matrices.

In 1954/55 practical methods for computing the SVD date back to Kogbetliantz .

Hestenes in 1958 resembling closely the Jacobi eigenvalue algorithm, used plane rotations or Givens rotations, i.e, $\begin{bmatrix} \cos q & -\sin q \\ \sin q & \cos q \end{bmatrix}$. However, these were replaced by the method of Gene Golub and William Kahan (the reduction to bidiagonal form) published in 1965, which uses Householder transformations or reflections; they introduce the SVD into numerical analysis.

It is a fact that the QR algorithm for the singular values of bidiagonal matrices was first derived by Golub in 1968 without reference to the QR algorithm, which has been the workhorse for two decade.

Recently in 1990, Demmel and Kahan have proposed an interesting alternative for 1968's Golub algorithm.

In the last 30 years, the singular value decomposition has become a popular numerical tool in statistical data analysis, signal processing, system identification and control system analysis and design. [12 &23]

Chapter One

Basic Concepts in Matrix Analysis

In this chapter we review some preliminary concepts and definitions in matrix analysis and present some basic properties related to these definitions.

Remark: In our thesis, we denote by **R** the set of real numbers and by **C** the set of all complex numbers $\mathbf{C} = \{x + iy,\ x, y \in \mathbf{R}\}$.

1.1 Vector Spaces Over C

Definition 1.1

A complex vector space V is a nonempty set of elements (called vectors) together with two operations: vector addition $\oplus$ and scalar multiplication Θ satisfying the following properties:

For all u, v and $w \in V$, c and $d \in \mathbf{C}$, then:

(1) a - $u \oplus v \in V$ (i.e. V is closed under vector addition $\oplus$)

b- $u \oplus v = v \oplus u$.

c- $u \oplus (v \oplus w) = (u \oplus v) \oplus w$.

d- there is an element $0 \in V$ such that $u + 0 = 0 + u = u$ (0 is called the additive identity).

e- $\exists -u \in V$ such that $u \oplus -u = 0$. ($-u$ is called the additive inverse).

(2) a- $c \,\Theta\, u \in V$ (i.e. V is closed under scalar multiplication Θ)

b- $c \,\Theta\, (u \oplus v) = c\Theta u \oplus c\Theta v$.

c- $(c + d) \,\Theta\, u = c \,\Theta\, u \oplus d \,\Theta\, u$.

d- $c \,\Theta\, (d \,\Theta\, u) = (cd) \,\Theta\, u = d \,\Theta (c\Theta u)$.

e- $1 \Theta u = u$.

Note: a real vector space has the same definition as a complex vector space except that the constants are real numbers.

Example 1

$\mathbf{C}^n$ with the usual addition and scalar multiplication is a complex vector space:

For any $u = \{u_1, u_2, \ldots, u_n\}$, $v = \{v_1, v_2, \ldots, v_n\}$ and $w = \{w_1, w_2, \ldots, w_n\} \in \mathbf{C}^n$

u_i, v_i and $w_i \in C, i = 1,2,\ldots,n$ and for any c and $d \in \mathbf{C}$ then :

(1) a- $u + v = \{u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n\} \in \mathbf{C}^n$, so $\mathbf{C}^n$ is closed under addition.

b- $u + v = \{u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n\} = \{v_1 + u_1, v_2 + u_2, \ldots, v_n + u_n\} = v + u$

c- $u + (v + w) = \{u_1 + (v_1 + w_1), u_2 + (v_2 + w_2), \ldots, u_n + (v_n + w_n)\}$

$$= \{(u_1 + v_1) + w_1, (u_2 + v_2) + w_2, \ldots, (u_n + v_n) + w_n\} = (u + v) + w$$

d- $(0,0,\ldots,0) \in \mathbf{C}^n$ is the additive identity.

e- The additive inverse for u is $-u = \{-u_1, -u_2, \ldots, -u_n\} \in \mathbf{C}^n$

(2) a- $c\,u = \{cu_1, cu_2, \ldots, cu_n\} \in \mathbf{C}^n$, so $\mathbf{C}^n$ is closed under scalar multiplication.

b- $c\,(u + v) = c\{u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n\} = \{c(u_1 + v_1), c(u_2 + v_2), \ldots, c(u_n + v_n)\}$

$= \{cu_1 + cv_1, cu_2 + cv_2, \ldots, cu_n + cv_n\} = \{cu_1, cu_2, \ldots, cu_n\} + \{cv_1, cu_2, \ldots, cv_n\}$

$= c\,u + c\,v$

c- (c+d)u =cu+du

d- c(du)=(cd)u

e- 1u=u

Definition 1.2

A subspace U of a vector space V over **C** is a nonempty subset of V which is by itself a vector space over **C** with respect to the operations on V.

Example 2

$U=\{(a,b,0)^T : a,b \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$ which is a real vector space.

Theorem 1.1

A nonempty subset U of V is a subspace of V if U is closed under the same operations $\oplus$ and Θ on V.

Definition 1.3

A set of vectors $\{\mathbf{x_1, x_2, ..., x_k}\}$ in a vector space V is said to be linearly dependent over **C** if there exists coefficients $a_1, a_2, ..., a_k \in \mathbf{C}$ not all zero, such that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_k\mathbf{x}_k = 0$.

A set which is not linearly dependent is said to be linearly independent.

Theorem 1.2

Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}$ be linearly independent vectors in $\mathbf{C}^n$ and P an n by n nonsingular matrix then $P^{-1}\mathbf{x}_1, P^{-1}\mathbf{x}_2, ..., P^{-1}\mathbf{x}_n$ are also linearly independent vectors in $\mathbf{C}^n$.

Definition 1.4

Let $A \in M_{m,n}(\mathbf{C})$. The rank of A denoted by r(A), is the number of linearly independent columns or rows of A, where $M_{m,n}(\mathbf{C})$ denotes all m by n matrices with entries from **C**.

Note: rank A = rank $A^* A$ (where A^* is the conjugate transpose of A i.e., if $A = [a_{ij}], a_{ij} \in \mathbf{C}$, then $A^* = [\overline{a_{ji}}]$), where " ‾ " denotes the conjugate.

Definition 1.5

A subset $S = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ of a vector space V is said to span V if every element $v \in V$ can be represented as a linear combination of the elements of S (i.e. if $v \in V, \exists\, a_1, a_2, ... a_n \in \mathbf{C}$ such that $v = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n$) . We then write V =Span S.

Definition 1.6

A nonempty subset S of a vector space V is said to be a basis of V, if it's both linearly independent and if it spans V. The number of elements of elements of S (a basis) is called the dimension of V, denoted by dim V.

Remark

Most of our work in the thesis will be over finite dimensional vector spaces, unless otherwise stated.

Note: If the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ form a basis for a vector space V, then they must be distinct and nonzero, so we write them as a set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$.

Definition 1.7

Let $\mathbf{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}, \mathbf{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\in \mathbf{C}^n$. The dot product of **x** and **y** is defined as

$$x.y = x^* y = \begin{bmatrix}\overline{x_1} & \overline{x_2} & \cdots & \overline{x_n}\end{bmatrix}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} = \overline{x_1}y_1 + \overline{x_2}y_2 + \ldots + \overline{x_n}y_n .$$

The length of x is denoted by $\| x \| = \sqrt{x.x}$.

Definition 1.8

Two vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbf{C}^n$ are said to be orthogonal if **x.y** = 0. Two subspaces U and V of a vector space are said to be orthogonal if $\{\mathbf{u}.\mathbf{v} = 0, \forall \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$.

Definition 1.9

Two vectors ***x*** and ***y*** are said to be orthonormal if ***x*** and ***y*** are orthogonal and of unit length each.

Definition 1.10

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \subseteq \mathbf{C}^n$ is said to be orthogonal if they are pairwise orthogonal. If in addition each x_i has a unit length, (i.e. $\| \mathbf{x}_i \| = 1, i = 1,2,\ldots,k$) then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k\}$ is an orthonormal set.

Definition 1.11

Let V be a subspace of C^n. A subspace $U \subset C^n$ is said to be the orthogonal complement of V in C^n if every vector $\mathbf{u} \in U$ is orthogonal to every $\mathbf{v} \in V$ and every if vector $\mathbf{v} \in V$ is orthogonal to every $\mathbf{u} \in U$. V is also called the orthogonal complement of U.

Example 3

$$W_1 = \text{Span}\left\{\begin{bmatrix}1\\2\\4\end{bmatrix}, \begin{bmatrix}5\\0\\7\end{bmatrix}\right\} = \text{span}\{u_1,u_2\} \text{ and } W_2 = \text{span}\left\{\begin{bmatrix}-1.4\\-1.3\\1\end{bmatrix}\right\} = \text{span } \{t_1\}$$

are orthogonal complements of each other in R^n, since $u_1.\ t_1=0$, $u_2.t_1=0$. Then, every element in W_1 is orthogonal to every element in W_2 .

Definition 1.12

Let $A \in M_{m,n}(\mathbf{C})$ and suppose r(A)=rank(A)=r. Then there are four fundamental subspaces related to A:

1- The Range of $A = \{\mathbf{y} \in \mathbf{C}^m : \exists\, \mathbf{x} \in \mathbf{C}^n, A\mathbf{x} = \mathbf{y}\}$. It is also called the column space of A, and dim (range A)= r .

2- The Null space of $A^* = \{\mathbf{y} \in \mathbf{C}^m : A^*\mathbf{y} = 0\}$ and its dimension = m-r. It is the orthogonal complement of range A.

3- Range $A^* = \{\mathbf{y} \in \mathbf{C}^n : \exists \mathbf{x} \in C^m, A^*\mathbf{x} = \mathbf{y}\}$ with dimension r . It is also called rowspace of A.

4- The Null Space of $A = \{\mathbf{y} \in \mathbf{C}^n : A\,\mathbf{y} = 0\}$ with dimension=n-r. It is the orthogonal complement of range A^*.

Example 4

Let

13

$$A=\begin{bmatrix}1 & 5 & 3 & 7\\ 2 & 0 & -4 & -6\\ 4 & 7 & -1 & 2\end{bmatrix},$$

Compute the four fundamental subspaces associated with *A*.

Solution:

By elementary row operations, the reduced row echelon form of *A* (written RREF(*A*)) is

$$RREF(A)=B=\begin{bmatrix}1 & 0 & -2 & -3\\ 0 & 1 & 1 & 2\\ 0 & 0 & 0 & 0\end{bmatrix}$$

Then

1- rank(*A*)=2

2-The basis of the column space of *A* is $S=\left\{\begin{bmatrix}1\\2\\4\end{bmatrix},\begin{bmatrix}5\\0\\7\end{bmatrix}\right\}$

3- Now to find the basis of the null space of A^*, find A^*,

$$A^*=\begin{bmatrix}1 & 2 & 4\\ 5 & 0 & 7\\ 3 & -4 & -1\\ 7 & -6 & 2\end{bmatrix}.$$

Again, by elementary row operations on A^*, we obtain

$$RREF(A^*)=C=\begin{bmatrix}1 & 0 & \frac{7}{5}\\ 0 & 1 & \frac{13}{10}\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

Solving the linear system $C\,\mathbf{x}=0$, we get

$S'=\left\{\begin{bmatrix}-1.4\\-1.3\\1\end{bmatrix}\right\}$ is the basis of the null space of A^*.

S and S' are orthogonal complements of each other in $\mathbf{R}^3$.

4- $T = \{[1 \quad 0 \quad -2 \quad -3] \;,\; [0 \quad 1 \quad 1 \quad 2]\}$ is a basis for the row space of A.

5- To find the basis of the null space of A, solve the linear system $B\mathbf{x} = 0$ to get

$$T' = \left\{ \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix} , \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$ is a basis for the null space of A**.**

Theorem 1.3

An orthonormal set of vectors is linearly independent.

1.2 Gram-Schmidt Orthonormalization Process

In our work we require linearly independent sets to be orthonormal. We can convert a linearly independent set into orthonormal set in many ways. One simple way to obtain an orthogonal set from linearly independent set is the Gram-Schmidt process.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a set of linearly independent vectors in a complex vector space V then the following are the steps of the Gram-Schmidt Orthonormalization Process.

1- Let $\mathbf{v}_1 = \mathbf{u}_1$

2- Compute the vectors $\mathbf{v}_i = \mathbf{u}_i - \sum_{k=1}^{i-1} \left(\frac{\mathbf{u}_i \cdot \mathbf{v}_k}{\|\mathbf{v}_k\|^2}\right)\mathbf{v}_k$ $\quad i=2,3,\ldots,n$. The vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ form an orthogonal set.

3- Let $\mathbf{w}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$, $\quad 1 \le i \le n$. Then $T = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is an orthonormal set of vectors.

Example 5

Obtain an orthonormal set from $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

First, these vectors are linearly independent.

1- Let $\mathbf{v}_1=(1,0,2,1)^T$

2- Compute $\mathbf{v}_i = \mathbf{u}_i - \sum_{k=1}^{i-1}(\frac{\mathbf{u}_i \cdot \mathbf{v}_k}{\|\mathbf{v}_k\|^2})\mathbf{v}_k \quad i=2,3$

$\mathbf{v}_2 = \mathbf{u}_2 - (\frac{\mathbf{u}_2.\mathbf{v}_1}{\|\mathbf{v}_1\|^2})\mathbf{v}_1 = (2,2,3,1)^T - \frac{9}{6}(1,0,2,1)^T = (\frac{1}{2},2,0,\ -\frac{1}{2})^T$ or $(1,4,0,-1)^T$

$\mathbf{v}_3 = \mathbf{u}_3 - (\frac{\mathbf{u}_3.\mathbf{v}_1}{\|\mathbf{v}_1\|^2})\mathbf{v}_1 - (\frac{u_3.v_2}{\|v_2\|^2})v_2$

$\mathbf{v_3} = (\ 1,0,1,0)^T - \frac{1}{2}(1,0,2,1)^T - \frac{1}{18}(1,4,0,1)^T = (\frac{4}{9},\frac{-2}{9},0,\ \frac{-4}{9})^T$ or $(4,-2,0,-4)^T$

you can see that $\mathbf{v}_1$,$\mathbf{v}_2$ and $\mathbf{v}_3$ are pairwise orthogonal.

3- Compute $\mathbf{w}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$, we obtain $\mathbf{w}_1=(\frac{1}{\sqrt{6}},0,\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}})^T$, $\mathbf{w}_2=(\frac{1}{3\sqrt{2}},\frac{4}{3\sqrt{2}},0,\frac{-1}{3\sqrt{2}})^T$ and $\mathbf{w}_3= (\frac{2}{3},\frac{-1}{3},0,\frac{-2}{3})^T$ respectively.

Note

This process may be applied to any finite or countable set of vectors (not necessary linearly independent). In this case at least one of the v_i's will equal zero, and the set $\{\mathbf{v_1},\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ will not be orthonormal.

Example 6

Obtain an orthogonal set from $T = \left\{ \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 3 \end{bmatrix} \right\}$

Solution:

Note that $\begin{bmatrix}-3\\2\\3\end{bmatrix}=1\begin{bmatrix}3\\2\\-1\end{bmatrix}+2\begin{bmatrix}-3\\0\\2\end{bmatrix}$, so T is linearly dependent; thus we can't convert them to an orthonormal set, but orthogonal only:

1- Let $\mathbf{v}_1 = (3, 2, -1)^T$

2- Compute $\mathbf{v}_i = \mathbf{u}_i - \sum_{k=1}^{i-1}(\frac{\mathbf{u}_i \cdot \mathbf{v}_k}{\|\mathbf{v}_k\|^2})\mathbf{v}_k \quad i=2,3$

$$\mathbf{v}_2 = (-\frac{9}{14},\frac{22}{14},\frac{17}{14})^T \text{ and } \mathbf{v}_3 = (0,0,0)^T$$

These vectors are orthogonal but $\mathbf{v}_3$ is not a unit vector .

1.3 Some Special Matrices

1.3.1 Unitary Matrices

Definition 1.13

A matrix $U \in M_n(\mathbf{C})$ is said to be unitary if U^*U=I .If U is real then U is called orthogonal.

We have some important theorems:

Theorem 1.4

Let $U \in M_n(\mathbf{C})$. The following are equivalent:

1- U is unitary.

2- U is non singular and $U^*=U^{-1}$ (where U^{-1} denotes the inverse of U).

3- UU^* = I.

4- U^* is unitary.

5- the columns of U form an orthonormal set in $\mathbf{C}^n$.

6- the rows of U form an orthonormal set in $\mathbf{C}^n$.

7- U preserves length, i.e., if $\mathbf{y} = U\mathbf{x}$, with $\mathbf{x} \in \mathbf{C}^n$, then $\|\mathbf{y}\| = \|\mathbf{x}\|$.

It is easy to prove this theorem and we prove only the last statement.

Suppose U is unitary then $\| \mathbf{y} \|^2 = \mathbf{y}^* \mathbf{y} = (U\mathbf{x})^* (U\mathbf{x}) = \mathbf{x}^* U^* U\mathbf{x} = \mathbf{x}^* \mathbf{x} = \| \mathbf{x} \|^2$

Example 7

Consider the unitary matrix $U = \begin{bmatrix} \frac{-1034}{1053} & \frac{125}{661} \\ \frac{-125i}{661} & \frac{1034i}{1053} \end{bmatrix}$ and the vector $\mathbf{x} = \begin{bmatrix} 1053 \\ 661 \end{bmatrix}$. Then, $\| \mathbf{x} \| = 1545730$

We have: $Ux = \begin{bmatrix} -909 \\ \frac{-50044i}{59} \end{bmatrix}$ with $\| U\mathbf{x} \| = 1545730$

Theorem 1.5

If U is unitary then $\det(U) = \pm 1$.

Theorem 1.6

If U and $V \in M_n(\mathbf{C})$ are unitary so is the product UV.

Special cases of unitary matrices are the permutation matrices. Define as follows:

Definition 1.14

A square matrix P is a permutation matrix if its columns are a permutation of the columns of I.

Example 8

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

is a permutation matrix.

1.3.2 Normal Matrices

Definition 1.15

A matrix $A \in M_n(\mathbf{C})$ is said to be normal if $AA^* = A^*A$

It is obvious that unitary and diagonal matrices are normal.

Example 9

Let

$$A = \begin{bmatrix} 2 & 3i \\ -3i & -5 \end{bmatrix}$$

Then A is normal since

$$AA^* = A^*A = \begin{bmatrix} 13 & -9i \\ 9i & 34 \end{bmatrix}$$

1.3.3 Hermitian Matrices

Definition 1.16

A matrix $A \in M_n(\mathbf{C})$ is said to be Hermitian if $A^* = A$. If A is real then A is said to be symmetric. It is skew Hermitian if $A^* = -A$.

It is obvious that Hermitian and skew Hermitian matrices are normal.

Theorem 1.7

Let $A \in M_n(\mathbf{C})$ be Hermitian. Then:

1- $A^T, \overline{A}, A^*$ are all Hermitian and if in addition, A is non singular then A^{-1} is also Hermitian.

2- $x^* Ax$ is real for all $x \in \mathbf{C}^n$.

Theorem 1.8

Let $A \in M_{m,n}(\mathbf{C})$ then AA^* and $A^* A$ are Hermitian matrices.

Remark

The main diagonal entries of a Hermitian matrix are all real.

Definition 1.17

A Hermitian matrix $A \in M_n(\mathbf{C})$ is said to be positive definite if $x^* Ax > 0$ for all nonzero $x \in \mathbf{C}^n$. " It's positive semi definite if $x^* Ax \geq 0$ "(note that since A is Hermitian then $x^* Ax$ is real).

Remark

If $A \in M_{m,n}(\mathbf{C})$ then AA^* and $A^* A$ are positive semi definite. If A has linearly independent columns then $A^* A$ is positive definite.

Proof

$\mathbf{x}^*(A^* A)\mathbf{x} = (A\mathbf{x})^* A\mathbf{x} = \| A\mathbf{x} \|^2 \geq 0, \forall \mathbf{x} \neq 0$. So A is positive semi definite. If in addition A has independent columns then $\| A\mathbf{x} \| > 0$, for all $\mathbf{x} \neq 0$

$\mathbf{x}^*(AA^*)\mathbf{x} = (A^*\mathbf{x})^* A^*\mathbf{x} = \| A^*\mathbf{x} \|^2 > 0, \forall \mathbf{x} \neq 0$

Example 10

Let $A=\begin{bmatrix}1&1\\1&2\\0&0\end{bmatrix}$, then $A^*=\begin{bmatrix}1&1&0\\1&2&0\end{bmatrix}$.

Therefor $A^*\mathbf{x}=0$ will give $\mathbf{x}=\begin{bmatrix}0\\0\\1\end{bmatrix}$

Now, $AA^*=\begin{bmatrix}2&3&0\\3&5&0\\0&0&0\end{bmatrix}$

$$\begin{bmatrix}0&0&1\end{bmatrix}\begin{bmatrix}2&3&0\\3&5&0\\0&0&0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}=\begin{bmatrix}0&0&0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}=0, \text{ although } \mathbf{x}\neq 0$$

But $A^*A=\begin{bmatrix}2&3\\3&5\end{bmatrix}$ and $\begin{bmatrix}a&b\end{bmatrix}\begin{bmatrix}2&3\\3&5\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=2a^2+6ab+5b^2=2(a+\frac{3}{2}b)^2+\frac{b^2}{2}\neq 0$ except if both *a* and *b* =0.

1.4 Eigenvalues and Eigenvectors

Definition 1.18

Let $A\in M_n(\mathbf{C})$. The number $l\in\mathbf{C}$ is called an eigenvalue of *A*, if there exists a non zero vector $\mathbf{x}\in\mathbf{C}^n$ such that $A\mathbf{x}=l\mathbf{x}$. In this case, **x** is called eigenvector of *A* associated with the eigenvalue l. The set of all eigenvalues of $A\in M_n(\mathbf{C})$ is called the spectrum of *A* and is denoted by $\Lambda(A)$.

Theorem 1.9

If **x** and **y** are eigenvectors of $A \in M_n(\mathbf{C})$ associated with the eigenvalue l then:

1) If $a\ \mathbf{x} + b\ \mathbf{y}$ is not the zero vector, then $a\ \mathbf{x} + b\ \mathbf{y}$ is also an eigenvector of A associated with the eigenvalue l .

2) If A is also Hermitian then all its eigenvalues are real.

3) If A is also positive definite then its eigenvalues are positive.

Example 11

Let $A = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$. Then A is Hermitian and note that

$$A \begin{bmatrix} -(1+i) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} -(1+i) \\ 1 \end{bmatrix}$$

so $x_1 = \begin{bmatrix} -(1+i) \\ 1 \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue $l = 0$

also $x_2 = \begin{bmatrix} 1+i \\ 2 \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue $l = 3$.

Thus $\Lambda(A) = \{0,3\}$

Definition 1.19

Let $A \in M_n(\mathbf{C})$. Then $f_t(A) = \det(tI_n - A) = \det \begin{bmatrix} t-a_{11} & -a_{12} & . & . & -a_{1n} \\ -a_{21} & t-a_{22} & . & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n1} & -a_{n2} & \cdot & \cdot & t-a_{nn} \end{bmatrix}$

is called the characteristic polynomial of A .

Theorem 1.10

The eigenvalues of A are the roots of the characterstic polynomial of A.

Thus to find the eigenvalues of a given matrix A we must find the roots of the characteristic polynomial. Then the corresponding eigenvectors are obtained by substituting the values of l in the system of equations $(l\mathbf{I}_n - A)\mathbf{x} = 0$ and solving the resulting system. So, the eigenvectors of A corresponding to l span the null space of $(l\mathrm{I}_n - A)$.

The characteristic polynomial can be written as the product of n factors each of the form $(l - l_i)$ where l_i is a root of the characteristic polynomial so we write $f_t(l) = (l - l_1)^{k_1}(l - l_2)^{k_2}\ldots(l - l_r)^{k_r}$, where $l_i, i = 1,2,\ldots,r$ are the distinct eigenvalues of A, and k_i are integers whose sum is n and which is called the algebraic multiplicity of $l_i, i = 1,2,\ldots,r$. Each eigenvalue has also a geometric multiplicity which is defined as the dimension of the subspace spanned by its eigenvectors. An eigenvalue is simple if its algebraic multiplicity is one in this case the algebraic and geometric multiplicities are equal.

It is easy to show that the algebraic multiplicity of each eigenvector is greater than or equal to the geometric multiplicity.

Example 12

Let $A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$

Its characteristic polynomial is

$$f(l)=\begin{bmatrix} l-1 & -i \\ -i & l-1 \end{bmatrix}=(l-1)^2-i^2=(l-1-i)(l-1+i).$$

Hence the eigenvalues are $l_1=1+i$ *and* $l_2=1-i$,

Using the first eigenvalue $l_1=1+i$ and substituting in $(l\mathbf{I}_2-A)\mathbf{x}=0$ gives

$$\begin{bmatrix} i & -i \\ -i & i \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

the second row gives $\mathbf{x}_1=\mathbf{x}_2$. So, $(1, 1)^T$ is an eigenvector corresponding to $l_1=1+i$. The same argument with $l_2=1-i$ gives the corresponding eigenvector $(1,-1)^T$.

Definition 1.20

The set of eigenvectors corresponding to an eigenvalue l togother with the zero vector form a subspace of $\mathbf{C}^n$ known as the eigenspace of l .

Definition 1.21

Let $A=\lfloor a_{ij}\rfloor\in M_n(\mathbf{C})$. Then the trace of A is defined as $A=\sum_{i=1}^{n} a_{ii}$

Theorem 1.11

Let $A\in M_n(\mathbf{C})$. Then the eigenvalues of A^* are the complex conjugate of the eigenvalues of A. i.e., if $l\in\Lambda(A)$ then $\overline{l}\in\Lambda(A^*)$

It follows from this theorem that for any matrix $A\in M_{m,n}(\mathbf{C})$, A^*A and AA^* have the same non zero eigenvalues.

Theorem 1.12

Let $A\in M_n(\mathbf{C})$ be Hermitian. Then it has asset of n orthogonal eigenvectors.

Example 13

Consider the Hermitian matrix $A=\begin{bmatrix}1 & 1+i\\ 1-i & 2\end{bmatrix}$ in example 11, which has eigenvectors $\mathbf{x}_1=\begin{bmatrix}-(1+i)\\ 1\end{bmatrix}$ and $\mathbf{x}_2=\begin{bmatrix}1+i\\ 2\end{bmatrix}$, to show that these vectors are linearly independent , we arrange them as columns in a matrix then transform it into row echelon form:

$$\begin{bmatrix}-1-i & 1+i\\ 1 & 2\end{bmatrix}\xrightarrow{\text{R.E.F}}\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}$$

Since each column has a leading one , so these two vectors are linearly independent.

1.5 Norm of Vectors and Matrices

One way to measure the size of vectors and matrices is to study the norm, so what is the norm?

Definition 1.22

Let V be a vector space over the field of complex numbers, a function $\|\,.\,\|: V\rightarrow \mathbf{R}$ is a vector norm if for all $\mathbf{x}, \mathbf{y} \in V$

1- $\|\mathbf{x}\| \geq 0$

2- $\|\mathbf{x}\| = 0$ iff $x = 0$

3- $\|c\,\mathbf{x}\| = |c|\,\|\mathbf{x}\|$ for all scalars $c\in \mathbf{C}$

4- $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Note: A function that satisfies axioms 1, 3 and 4 of definition 1.22 is called semi norm.

Example 14

1- The Euclidean norm (or ℓ_2 norm) on $\mathbf{C}^n$ is

$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}.\mathbf{x}} = \sqrt{\mathbf{x}^*\mathbf{x}} = (|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2)^{\frac{1}{2}}$ where $\mathbf{x} = (x_1, x_2, ..., x_n)^T$

2- The sum norm (or ℓ_1 norm) on $\mathbf{C}^n$ is

$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$ where $\mathbf{x} = (x_1, x_2, ..., x_n)^T$.

Definition 1.23

A norm is said to be unitarily invariant if $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbf{C}^n$ and all unitary matrices $U \in M_n(\mathbf{C})$

Example 15

The ℓ_2 norm is unitarily invariant.

Definition 1.24

A function $|||\ .\ |||\ : M_{m,n}(\mathbf{C}) \to \mathbf{R}$ is a matrix norm if for all $A, B \in M_{m,n}(\mathbf{C})$

1- $|||A||| \geq 0$

2- $|||A|||=0$ iff $A=0$

3- $|||cA||| = |c|\ |||A|||$, for all complex scalars c

4- $|||A+B||| \leq |||A|||+|||B|||$ triangle inequality

5- $|||AB||| \leq |||A|||\ |||B|||$ (if $n=m$) sub multiplicative

Remark

By 5 of definition 1.24, for any nonzero matrix A for which $A^2 = A$, we have that $|||A||| \geq 1$; this is because $|||A||| = |||A^2||| = |||AA||| \leq |||A|||\,|||A||| = |||A|||^2$. In particular $|||\,I_n\,||| \geq 1$ for any matrix norm, so if A is invertible then $|||\mathbf{I}_n||| = |||AA^{-1}||| \leq |||A|||\,|||A^{-1}||| \Rightarrow |||A^{-1}||| \geq \frac{1}{|||A|||}$ for any matrix norm.

Example 16

The Euclidean norm (ℓ_2 or Frobenius norm) on $M_{m,n}(\mathbf{C})$ is defined as

$$|||A|||_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}.$$

Note that in the vector norms the Euclidean norm is denoted by $||\mathbf{x}||_2$ while in the matrix norms the ℓ_2 norm is denoted by $|||A|||_F$.

Example 17

The spectral norm $|||.|||_2$ is defined on $M_{m,n}(\mathbf{C})$ by

$|||A|||_2 = \max\{\sqrt{l} : l$ is an eigenvalue of $AA^*\}$

$|||A|||_2$ is defined since AA^* is positive semi definite and so all its eigenvalues are non negative.

Note

$||A\mathbf{x}||_2 \leq |||A|||_2\,||\mathbf{x}||_2$, where $A \in M_{m,n}(\mathbf{C})$ and $\mathbf{x} \in \mathbf{C}^n$.

Definition 1.25

The inner product of A and $B \in M_{m,n}(\mathbf{C})$ is defined as $A \cdot B = tr(A^* B) = \sum_{j=1}^{n} \sum_{i=1}^{m} \overline{a_{ij}} b_{ij}$. We then write the Frobenius norm as $||| A |||_F = \sqrt{A \cdot A}$,

Theorem 1.13

Let $\| . \|_a$ be a given norm on C^n and let $||| . |||_a$ be the matrix norm on $M_{m,n}(\mathbf{C})$. Then $||| A |||_a = \max \left\{ \frac{\| Ax \|_a}{\| x \|_a}, x \neq 0 \right\}$

Definition 1.26

A matrix norm is said to be unitarily invariant if $||| UAV ||| = ||| A |||$ for all $A \in M_{m,n}(\mathbf{C})$ and all unitary matrices U and V.

For instance both the Spectral and Frobenius norms are unitary invariant i.e., $||| A |||_2 = ||| UAV |||_2$ *and* $||| A |||_F = ||| UAV |||_F$ for all unitary matrices U and V.

1.6 Condition number

It's a measure for singularity defined as follow:

Definition 1.27

The condition number of a square matrix A with respect to a given matrix norm is defined as:

$$\operatorname{cond}(A) = \begin{cases} |||A|||\,|||A^{-1}||| & , A \text{ } is \text{ nonsinular} \\ \infty & , A \text{ } is \text{ singular} \end{cases}$$

The following theorem gives us some important properties of the condition number:

Theorem 1.14

1. For any matrix A, $\operatorname{cond}(A) \geq 1$.

2. For the identity matrix, $\operatorname{cond}(\mathrm{I}) = 1$.

4.5. For any matrix A and nonzero scalar a, $\operatorname{cond}(aA) = \operatorname{cond}(A)$

Most of the material of this chapter can be found in [11] and [15].

Chapter Two
Similarity and Unitarily Diagonalization

We start this chapter with the definition of matrix diagonalization.

2.1 Diagonalization

Definition 2.1

A matrix $B \in M_n(\mathbf{C})$ is said to be similar to a matrix $A \in M_n(\mathbf{C})$ if there exists a non singular matrix *P* such that $A = P^{-1}BP$. We say *B* is similar to *A* via *P*. We also call *P* the matrix of similarity between *A* and *B*.

If *B* is similar to *A* then *A* is similar to *B*. So we can simply say *A* and *B* are similar. If *P* is unitary then *A* and *B* are said to be unitarily similar.

Example 1

The matrix $B = \begin{bmatrix} -8 & 3 \\ -18 & 7 \end{bmatrix}$ is similar to $A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$ since:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} = \underbrace{\begin{bmatrix} -2 & 1 \\ 3 & -1 \end{bmatrix}}_{P^{-1}} \underbrace{\begin{bmatrix} -8 & 3 \\ -18 & 7 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}}_{P}$$

Definition 2.2

A matrix $B \in M_n(\mathbf{C})$ is said to be diagonalizable (or can be diagonalized) if it is similar to an n by n diagonal matrix. We then write $B = P^{-1}DP$, where D is diagonal.

Example 2

The matrix $B = \begin{bmatrix} -8 & 3 \\ -18 & 7 \end{bmatrix}$ in Example 1 is diagonalizable since it is similar to the diagonal matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$.

Theorem 2.1

If a matrix $A \in M_n(\mathbf{C})$ is diagonalizable then it has n linearly independent eigenvectors.

Proof

Let $A \in M_n(\mathbf{C})$ be diagonalizable. Then there exists a non singular matrix $P \in M_n(\mathbf{C})$ and a diagonal matrix $\Lambda \in M_n(\mathbf{C})$ such that $A = P\Lambda P^{-1}$ or $AP = P\Lambda$

Let $P = [x_1 \quad x_2 \quad . \quad . \quad x_n]$ with $x_i \in \mathbf{C}^n$ and $\Lambda = \mathrm{diag}(d_{ii})$ where $d_{ii} \in \mathbf{C}$, then

$$A\,[x_1 \quad . \quad . \quad x_n] = [x_1 \quad . \quad . \quad x_n] \begin{bmatrix} d_{11} & & & \\ & . & & \\ & & . & \\ & & & d_{nn} \end{bmatrix}$$

Comparing the left hand side with the right hand side column by column we have $Ax_i = d_{ii}x_i$. Since P is non singular then its columns are linearly independent and none of them is zero. Then, by definition of the eigenvectors, the columns of P are linearly independent eigenvectors of A corresponding to the eigenvalues d_{ii}. ■

Definition 2.3

A matrix $A \in M_n(\mathbf{C})$ is unitarily diagonalizable if it is diagonalizable via a unitary matrix.

Example 3

Let $A = \begin{bmatrix} -\frac{1}{2} & \frac{5i}{2} \\ -\frac{5i}{2} & -\frac{1}{2} \end{bmatrix}$ and let $P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$. Then P is unitary and

$$P\Lambda P^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{-1}{2} & \frac{5i}{2} \\ \frac{-5i}{2} & \frac{-1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

So, A is unitarily diagonalizable.

Definition 2.4

An eigenvalue l of $A \in M_n(\mathbf{C})$ is called defective if its geometric multiplicity is less than its algebraic multiplicity. A matrix A is defective if it has a defective eigenvalue. Otherwise A is non defective.

Note

If an eigenvalue is simple then it is non defective.

Example 4

The matrix $A=\begin{bmatrix}1 & -1 & -1\\ 1 & -1 & 0\\ 1 & 0 & -1\end{bmatrix}$ is non defective since its eigenvalues are –1, *i* and –*i* and they are all simple.

Theorem 2.2

Similar matrices share the same eigenvalues with the same algebraic and geometric multiplicities.

Proof

Suppose A and B are similar, so there exists a non singular matrix P such that $B = P^{-1}AP$. Then

$$P_t(B)\det(lI - B) = \det(lI - P^{-1}AP) = \det(P^{-1}lP - P^{-1}AP) = \det(P^{-1}(lI - A)P)$$

$$= \det(P^{-1})\det(lI - A)\det(P) = \det(lI - A) = P_t(A)$$

This means that both A and B have the same characteristic polynomial and so they have the same eigenvalues (roots) with the same algebraic multiplicity. If $\mathbf{x}$ is an eigenvector of A associated to l, then $A\mathbf{x} = l\mathbf{x}$ and $PBP^{-1}\mathbf{x} = l\mathbf{x}$ which gives $B(P^{-1}\mathbf{x}) = P^{-1}l\mathbf{x} = l(P^{-1}\mathbf{x})$, i.e., $P^{-1}\mathbf{x}$ is an eigenvector of B associated to l; hence, by theorem 1.2, A and B have the same geometric multiplicity. ∎

Note

The zero matrix $0 \in M_n(\mathbf{C})$ is non defective since it has only one eigenvalue (zero) with algebraic multiplicity n and $e_1, e_2, \ldots, e_n$ are its eigenvectors. So the geometric multiplicity of the zero eigenvalue is n; so it's non defective.

Remark

Any diagonal matrix is non defective.

Example 5

Let

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Then $\Lambda(A) = \{2,3\}$ with algebraic multiplicities 3 and 2, respectively. To find the geometric multiplicity for 2, we solve $(2\mathbf{I}_5 - A)\mathbf{x} = 0$ which gives

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The solution of this equations are $x_1 = x_4 = 0$ and x_2, x_3 and x_5 are free variables so we can choose the eigenvectors associated to 2 to be e_2, e_3 and e_5. So, the geometric multiplicity for 2 is 3.

In the same way we show that the geometric multiplicity for 3 is 2. So A is non defective.

Theorem 2.3

Let $A \in M_n(\mathbf{C})$. Then A is non defective if and only if it is diagonalizable.

Proof

Suppose A is non defective then it has n linearly independent eigenvectors, $x_1, x_2, ..., x_n$. Since $A\mathbf{x}_i = l_i\mathbf{x}_i$ then $A[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] = [l_1\mathbf{x}_1, l_2\mathbf{x}_2, ..., l_n\mathbf{x}_n]$ and hence $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]\begin{bmatrix} l_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & l_n \end{bmatrix} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]D$. So, we have $AP = PD$ and hence

$P^{-1}AP = D$, where $D = diag\{l_1, ..., l_n\} = \begin{bmatrix} l_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & l_n \end{bmatrix}$.

Conversely,

Suppose $P^{-1}AP = D$, where $P = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ and $D = \begin{bmatrix} d_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & d_n \end{bmatrix}$. Then A is similar to D and by Theorem 2.2 A have the same eigenvalues as D, namely $d_1, ..., d_n$, with the same algebraic and geometric multiplicities. But D is diagonal and thus by the remark above it is non defective and so is A. ■

Not all matrices are diagonalizable, see the following example:

Example 6

The matrix $A = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ is not diagonalizable. To see this, note that -1 is an eigenvalue of A with algebraic multiplicity =2, but the eigenvectors associated to -1 are $(r,0)^T$. Hence, the geometric multiplicity of -1 is 1 and so it is defective and so A is not diagonalizable.

2.2 Schur's Theorem

In the previous section we showed that not all square matrices are diagonalizable. In this section we prove that all square matrices are unitarily similar to an upper triangular matrix.

Theorem 2.4(Schur's Theorem):

Given $A \in M_n(\mathbf{C})$ with eigenvalues $l_1, l_2, \ldots, l_n$, there is a unitary matrix $U \in M_n(\mathbf{C})$ such that $U^* AU = T = [t_{ij}]$, where T is upper triangular, with $t_{ii} = l_i, i = 1,2,\ldots,n$. [22]

Proof

Let $\mathbf{x_1}$ be a unit eigenvector associated to l_1, so $A\mathbf{x_1} = l\mathbf{x_1}$. Since $\mathbf{x_1}$ is not zero we may use Gram Schmidt orthonormalization process to extend

T_1 ={$\mathbf{x_1}$} to an orthonormal basis{ $\mathbf{x_1}, \mathbf{z_2}, \mathbf{z_3}, \ldots, \mathbf{z_n}$ } of $\mathbf{C}^n$.

Then for the unitary matrix $U_1 = [\mathbf{x_1} \quad \mathbf{z_2} \quad . \quad . \quad \mathbf{z_n}]$

$$B = U_1^* A U_1 = \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{z}_2^* \\ . \\ . \\ \mathbf{z}_n^* \end{bmatrix} A[\mathbf{x}_1 \quad \mathbf{z}_2 \quad . \quad . \quad \mathbf{z}_n]$$

$$= \begin{bmatrix} l_1 & * \\ \mathbf{0} & A_1 \end{bmatrix}$$

The matrix $A_1 \in M_{n-1}(\mathbf{C})$, has eigenvalues $l_2, l_3, \ldots, l_n$. We find a normalized eigenvector $\mathbf{x}_2 \in C^{n-1}$ of A_1 corresponding to l_2, and then extend $T_2 = \{\mathbf{x_2}\}$ to an orthonormal basis$\{\mathbf{x_2}, \mathbf{w_3}, \ldots, \mathbf{w_n}\}$of C^{n-1}.

Determine a unitary matrix$U_2 \in M_{n-1}(\mathbf{C})$, where

$$U_2^* A_1 U_2 = \begin{bmatrix} l_2 & * \\ \mathbf{0} & A_2 \end{bmatrix}$$

Let

$$V_2 = \begin{bmatrix} \mathrm{I}_1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix}$$

where I_1 is the 1by 1 identity matrix. Then V_2 and U_1V_2 are unitary and

$$V_2^* B V_2 = V_2^*(U_1^* A U_1)V_2 = \left[\begin{array}{cc|c} l_1 & * & \\ 0 & l_2 & * \\ \hline & \mathbf{0} & A_2 \\ & & \end{array}\right]$$

Countinue this reduction to produce unitary matrices $U_i \in M_{n-i+1}(\mathbf{C}),\ i = 1,2,\ldots,n-1$ & $V_i \in M_n(\mathbf{C}), i = 2,3,\ldots,n-1$, then the matrix $U = U_1 V_2 V_3 \ldots V_{n-1}$ is unitary and $U^* AU = \begin{bmatrix} l_1 & & t_{ij} \\ & . & \\ \mathbf{0} & & . \\ & & & l_n \end{bmatrix}$ yields the desired form. ■

Example 7

Let

$$A=\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}$$

The eigenvalues of A are 0, 1 and 3 and the normalized eigenvector corresponding to 0 is $x_1=(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}})^T$. Expand $T_1=\{x_1\}$ to a basis of $\mathbf{C}^3$, we obtain $\{x_1,\ \ e_1,\ \ e_2\}=\left\{\begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{bmatrix},\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$

Applying Gramschmidt orthonormalization process to these vectors to get $\left\{\begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{bmatrix},\begin{bmatrix}\frac{2}{\sqrt{6}}\\ \frac{-1}{\sqrt{6}}\\ \frac{-1}{\sqrt{6}}\end{bmatrix},\begin{bmatrix}0\\ \frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}\end{bmatrix}\right\}$, then the unitary matrix $U_1=\begin{bmatrix}\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0\\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{2}}\end{bmatrix}$ is such that $U_1^*AU_1=$

$$\frac{1}{\sqrt{6}}\begin{bmatrix}\sqrt{2} & \sqrt{2} & \sqrt{2}\\ 2 & -1 & -1\\ 0 & \sqrt{3} & -\sqrt{3}\end{bmatrix}\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}(\frac{1}{\sqrt{6}}\begin{bmatrix}\sqrt{2} & 2 & 0\\ \sqrt{2} & -1 & \sqrt{3}\\ \sqrt{2} & -1 & -\sqrt{3}\end{bmatrix})=\begin{bmatrix}0 & 0 & 0\\ 0 & \frac{3}{2} & \frac{-\sqrt{3}}{2}\\ 0 & \frac{-\sqrt{3}}{2} & \frac{5}{2}\end{bmatrix}$$

$=\begin{bmatrix}0 & 0 & 0\\ 0 & A_1 & \\ 0 & & \end{bmatrix}$ where $A_1=\begin{bmatrix}\frac{3}{2} & \frac{-\sqrt{3}}{2}\\ \frac{-\sqrt{3}}{2} & \frac{5}{2}\end{bmatrix}\in M_2(\mathbf{C})$ and its eigenvalues are 1 and 3. The normalized eigenvector associated with 1 is $\mathbf{x}_2=(\frac{\sqrt{3}}{2},\frac{1}{2})^T$. Let $T_2=\{\mathbf{x}_2\}$ and repeat the same steps as above to find $U_2=\frac{1}{2}\begin{bmatrix}\sqrt{3} & 1\\ 1 & -\sqrt{3}\end{bmatrix}$

So, $U_2^* A_1 U_2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ and $V_2 = \begin{bmatrix} I_1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix}$.

Let $U = U_1 V_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & 2 & 0 \\ \sqrt{2} & -1 & \sqrt{3} \\ \sqrt{2} & -1 & -\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

Then $T = U^* A U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ is upper triangular matrix with $t_{ii} = l_{ii}$.

Note

Neither U nor T in the theorem is unique.

Schur's Theorem says that every square matrix is similar to an upper triangular matrix via a unitary matrix. In the next section, we consider similarity of a matrix to a diagonal matrix via a unitary matrix in what is known by the spectral decomposition theorem.

2.3 Spectral Decomposition

Lemma 1

Any upper triangular normal matrix must be diagonal.

Proof

Let $T = [t_{ij}] \in M_n(\mathbf{C})$ be upper triangular and normal. Then t_{ij} =0 for $i>j$ and $T^*T = TT^*$. Comparing the diagonal entries of both sides, we obtain

$$(T^*T)ii = \sum_{k=1}^{n} \overline{t_{ki}} t_{ki} = (TT^*)_{ii} = \sum_{k=1}^{n} t_{ik} \overline{t_{ik}} \quad , i = 1,2,\ldots,n$$

By using the fact that T is upper triangular we have:

$$(T^*T)_{11} = \sum_{k=1}^{n} \overline{t_{k1}} t_{k1} = \sum_{k=1}^{n} |t_{k1}|^2 = |t_{11}|^2 + \sum_{k=2}^{n} |t_{k1}|^2 = |t_{11}|^2 + 0 = |t_{11}|^2 \quad (1)$$

$$(TT^*)_{11} = \sum_{k=1}^{n} t_{1k} \overline{t_{1k}} = \sum_{k=1}^{n} |t_{1k}|^2 = |t_{11}|^2 + \sum_{k=2}^{n} |t_{1k}|^2 \quad (2)$$

The equality of (1) and (2) gives:

$$\sum_{k=2}^{n} |t_{1k}|^2 = 0 \Rightarrow t_{1k} = 0, k > 1 \quad (3)$$

Since t_{k2}=0, k >2 (T is upper triangular) and from (3), $|t_{12}|^2 = 0$, we have:

$$(T^*T)_{22} = \sum_{k=1}^{n} |t_{k2}|^2 = |t_{12}|^2 + |t_{22}|^2 + \sum_{k=3}^{n} |t_{k2}|^2 = |t_{22}|^2 \quad (4)$$

And by (3)

$$(TT^*)_{22} = \sum_{k=1}^{n} |t_{2k}|^2 = |t_{21}|^2 + |t_{22}|^2 + \sum_{k=3}^{n} |t_{2k}|^2 = |t_{22}|^2 + \sum_{k=3}^{n} |t_{2k}|^2 \quad (5)$$

The equality of (4) and (5) gives:

$$\sum_{k=3}^{n} |t_{2k}|^2 = 0 \text{ and hence } t_{2k} = 0 \ , k > 2 .$$

Continuing in the same way, we obtain t_{ij} =0 for all $j > i$, i = 1,2, . . .,$n-1$.

So T is a diagonal matrix . ■

Lemma 2

Let $A \in M_n(\mathbf{C})$ be similar to a matrix T via a unitary matrix U. Then A is normal if and only if T is normal.

Proof:

Let A be normal and $T = U^*AU$, where U is unitary. Then $T^*T = U^*A^*UU^*AU = U^*A^*AU = U^*AA^*U = U^*AUU^*A^*U = TT^*$ so, T is normal.

Conversely,

If T is normal than A^*A=$AA^*$where $A = UTU^*$. ■

Theorem 2.5 (Spectral Theorem for Normal Matrices)

Let $A \in M_n(\mathbf{C})$. Then A is normal if and only if it's unitarily diagonalizable. [11], [13]

Proof

Let $A \in M_n(\mathbf{C})$ with eigenvalues $l_1, l_2, ..., l_n$. Then by Schur's theorem there exists upper triangular matrix $T = U^* AU$,where U is a unitary matrix. Since A is normal then by Lemma 2, T is normal and by Lemma 1, T is diagonal $=D$. So, A is unitarily diagonalizable.

Conversely, let $A \in M_n(\mathbf{C})$ be unitarily diagonalizable then $A = UDU^*$, for some diagonal D and unitary U. Since D is diagonal then D is normal and hence by Lemma 2, A is normal. ■

By this theorem, only the normal matrices are unitarily diagonalizable.

Example 8

Let $A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$. Then $A^* = \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix}$ and $AA^* = A^*A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Hence,

A is normal.

We now show A is unitarily diagonalizable.

The eigenvalues of A are $1+i$ and $1-i$, with corresponding eigenvectors $(1,1)^T$ and $(1,-1)^T$, respectively. [see Example 12 in Chapter 1].

The eigenvectors $(1,1)^T$ and $(1,-1)^T$ are orthogonal and so we have the unitary matrix $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ and $U^*AU = \begin{bmatrix} 1+i & \\ & 1-i \end{bmatrix}$. So A is unitarily diagonalizable.

Corollary 1

Let $A = UDU^*$ be the spectral decomposition of A. Then $A^k = UD^kU^*$, where k is any non negative integer (with A^0=identity matrix) .

proof

$$A^k = \underbrace{(UDU^*)(UDU^*)\ldots(UDU^*)}_{k-times}$$

$$= U\underbrace{DD\ldots D}_{k-times}U^* = UD^kU^* .$$

So, if A is normal then $A^2, A^3, \ldots, A^k$ are all normal. ■

Chapter Three

The Singular Value Decomposition

This chapter is the main topic of our thesis "The Singular Value Decomposition (SVD)". We introduce its definition, investigate its proof of existence and clarify its relation to the spectral decomposition. We also investigate some of its properties.

3.1 Definition and Computation of the SVD

The spectral decomposition of a matrix A as shown in Chapter 2 exists only for normal matrices. We generalize this decomposition to any matrix using the SVD.

In this section we restrict our attention to the definition of SVD, the way of computing it and to the proof of its existence.

Definition 3.1

Let $A \in M_{m,n}(\mathbf{C})$. A non negative real number s is said to be a singular value for A if there exists two unit length vectors $u \in \mathbf{C}^m$ and $v \in \mathbf{C}^n$ such that $Av = su$ and $A^*u = sv$. see [11, 6]

Example 1

$\sqrt{12}$ is a singular value of $A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$ since:

$$A\begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \sqrt{12}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } A^*\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \sqrt{12}\begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

Theorem 3.1 (The Singular Value Decomposition)

Let $A \in M_{m,n}(\mathbf{C})$, with $m \geq n$ and rank $A = r$. Then A can be written as $A = U\Sigma V^*$, where $U \in M_m(\mathbf{C})$ and $V \in M_n(\mathbf{C})$ are unitary, $\Sigma = \begin{bmatrix} S \\ \mathbf{0} \end{bmatrix} \in M_{m,n}(\mathbf{C})$, $S = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_r, \sigma_{r+1}, .., \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$, $\mathbf{0} \in M_{m-n,n}(\mathbf{C})$ and s_i's are the square roots of the eigenvalues of A^*A. [21, 14]

Proof

Let $A \in M_{m,n}(\mathbf{C})$ with $m \geq n$. Then $A^*A \in M_n(\mathbf{C})$ is Hermitian so it is normal. Let $l_1, l_2, ..., l_n$ be the eigenvalues of A^*A, with associated orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, i.e., $A^*A\mathbf{v}_i = l_i \mathbf{v}_i \quad i = 1,2,...,n$

Then $AA^*A\mathbf{v}_i = A\, l_i \mathbf{v}_i = l_i A\mathbf{v}_i \;\; i = 1,2,...,n$. So, l_i is also an eigenvalue of $A\,A^*$ with associated eignvector $A\mathbf{v}_i$.

Now, $0 \leq \| A\mathbf{v}_i \|_2^2 = (A\mathbf{v}_i)^*(A\mathbf{v}_i) = \mathbf{v}_i^* A^* A\mathbf{v}_i = \mathbf{v}_i^* l_i \mathbf{v}_i = l_i \mathbf{v}_i^* \mathbf{v}_i = l_i \| \mathbf{v}_i \|_2^2 = l_i$.

So, $\| Av_i \|_2 = \sqrt{l_i}$, which we denote by s_i .

Since rank A = r = rank A^*A =number of non zero eigenvalues of A^*A, then $s_i \neq 0, \forall\, i \leq r$. Define $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{s_i}, i = 1,2,...,r$. Then $\mathbf{u_i}$ is a unit vector since $\| \mathbf{u}_i \|_2 = \dfrac{\| A\mathbf{v}_i \|_2}{| s_i |} = \dfrac{s_i}{s_i} = 1$. In addition, $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ are orthonormal since $\mathbf{u}_i^* \mathbf{u}_j = (\dfrac{A\mathbf{v}_i}{s_i})^* \dfrac{A\mathbf{v}_j}{s_j} = \dfrac{\mathbf{v}_i^* A^* A\mathbf{v}_j}{s_i s_j} = \dfrac{\mathbf{v}_i^* l_j \mathbf{v}_j}{s_i s_j} = \dfrac{l_j \mathbf{v}_i^* \mathbf{v}_j}{s_i s_j} = 0$, for $i \neq j$.

Now, $\mathbf{u}_i^* A\mathbf{v}_j = (\dfrac{A\mathbf{v}_i}{s_i})^* A\mathbf{v}_j = \dfrac{\mathbf{v}_i^* A^* A\mathbf{v}_j}{s_i} = \dfrac{\mathbf{v}_i^* l_j \mathbf{v}_j}{s_i} = \dfrac{l_j}{s_i} \mathbf{v}_i^* \mathbf{v}_j = \begin{cases} 0 & i \neq j \\ s_i & i = j \end{cases}$ $\quad i,j = 1,2,\ldots,r$

since $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r$ are orthonormal.

Writing all these equations for i,j=1, 2, . . .,r in matrix form gives:

$$\begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \cdot \\ \cdot \\ \mathbf{u}_r^* \end{bmatrix} A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad . \quad . \quad \mathbf{v}_r] = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & . & & \\ & & & . & \\ & & & & \sigma_r \end{bmatrix}$$

The orthonormal vectors $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$ form an m dimensional subspace of $\mathbf{C}^m$, and can expand to a basis $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r, \mathbf{u}_{r+1}, \mathbf{u}_{r+2}, ..., \mathbf{u}_m$ of $\mathbf{C}^m$.

Since $s_j = 0$ for all $j > r$ then, $\mathbf{u}_i^* A\mathbf{v}_j = 0$ (because $\| A\mathbf{v}_j \|_2 = s_j = 0$ and so, $A\mathbf{v}_j = 0$) for $i > r$ and $j \le r$ then $A\mathbf{v}_j = s_j \mathbf{u}_j$ and so $\mathbf{u}_i^* A\mathbf{v}_j = \mathbf{u}_i^* s_j \mathbf{u}_j = 0$, $i > r \ge j$

So, we have $U^* AV =$

$$\begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \cdot \\ \cdot \\ \mathbf{u}_m^* \end{bmatrix} A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad . \quad . \quad \mathbf{v}_n] = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & . & & \\ & & & . & \\ & & & & \sigma_n \\ & & \mathbf{0} & & \end{bmatrix} = \Sigma = \begin{bmatrix} S \\ \mathbf{0} \end{bmatrix},$$

where $S = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r, \sigma_{r+1}, .., \sigma_n)$, $\sigma_i = 0, i = r+1, ..., n$ and $\mathbf{0} \in M_{m-n,n}(\mathbf{C})$.

Defining $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad . \quad . \quad \mathbf{u}_m] \in M_m(C)$ and $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad . \quad . \quad \mathbf{v}_n] \in M_n(c)$, then both U and V are unitary and $U^* AV = \Sigma$, hence $A = U\Sigma V^*$ as required. ∎

Notes on the proof of this theorem:

1. This decomposition can be applied to all rectangular complex matrices, and if $m \le n$ we compute the SVD of A^*.
2. A and Σ are $m \times n$ matrices.
3. The square roots of the eigenvalues of $A^* A$ are the singular values of A. To see this, since $A = U\Sigma V^*$, then we have $AV = U\Sigma$ in which we get

$A\mathbf{v}_i = s_i\mathbf{u}_i$, $i = 1,2,...,n$. Similarly, $A^*U = V\Sigma^*$. Hence $A^*\mathbf{u}_i = \sigma_i\mathbf{v}_i, i = 1,2,...,n$ and $A^*\mathbf{u} = 0$, $i > n$.

4. Rank $A = r$ = number of nonzero singular values of A.
5. The matrix Σ is uniquely determined.
6. The columns of U are orthonormal eigenvectors of AA^* and are called left singular vectors and the columns of V are orthonormal eigenvectors of A^*A and are called right singular vectors.

Example 2

Let $A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$. Then $A^* = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$ and $A^*A = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$.

The eigenvalues of A^*A are 10,12 and their associated eigenvectors are $(1,-1)^T$ and $(1,1)^T$, respectively. Since these eigenvectors are orthogonal, define $V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$. So V is a unitary matrix whose columns are orthonormal eigenvectors of A^*A.

Also , we find the eigenvalues of $AA^* = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$ either by calculations or directly since they are the same as the eigenvalues of A^*A plus 3-2=1 zero eigenvalues. So we have 12, 10 and 0 as simple eigenvalues of the symmetric matrix AA^* and hence their associated eigenvectors $(1,2,1)^T$, $(2,-1,0)^T$ and $(1,2,-5)^T$, respectively, are orthogonal. Dividing each eigenvector by its length and by ordering these unit eigenvectors in a matrix in decreasing order according to the associated eigenvalues we obtain the unitary matrix U as

$$U = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix}.$$

Now, to find Σ we take the positive roots of the nonzero eigenvalues and populate them on the diagonal of Σ in a decreasing order. $\sigma_1 = \sqrt{12}$ and $\sigma_2 = \sqrt{10}$. So, we have

$$S = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix}, \ \Sigma = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix} \text{ and}$$

$$A = U\Sigma V^* = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}.$$

Example 3

Let $B = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}$.

Since $m = 2 < 3 = n$, we then consider $A = B^* = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 2 & 2 \end{bmatrix}$.

Now, $A^*A = \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}$ and its eigenvalues are 18 and 0 (so the singular values of the matrix A is $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$), with associated orthonormal eigenvectors $\mathbf{v_1} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$ and $\mathbf{v_2} = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$

On the other hand, the matrix

$$AA^* = \begin{bmatrix} 2 & 4 & 4 \\ 4 & 8 & 8 \\ 4 & 8 & 8 \end{bmatrix}$$

has eigenvalues 18, 0 and 0 with associated eigenvectors $(1,2,2)^T$, $(2,0,-1)^T$ and $(2,-1,0)^T$, respectively. These eigenvectors are not orthonormal, so by Gram Schmidt orthonormalization process we obtain $\mathbf{u}_1 = (\frac{1}{3},\frac{2}{3},\frac{2}{3})^T$ $\mathbf{u}_2 = (\frac{2}{\sqrt{5}},0,\frac{-1}{\sqrt{5}})^T$ and $\mathbf{u}_3 = (\frac{2}{3\sqrt{5}},\frac{-5}{3\sqrt{5}},\frac{4}{3\sqrt{5}})^T$ as orthonormal eigenvectors of AA^*.

Let $V = [\mathbf{v}_1\ \ \mathbf{v}_2]$ and $U=[\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$ then

$$B^* = A = \begin{bmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{-5}{3\sqrt{5}} \\ \frac{2}{3} & \frac{-1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

And hence

$$B = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & 0 & \frac{-1}{\sqrt{5}} \\ \frac{2}{3\sqrt{5}} & \frac{-5}{3\sqrt{5}} & \frac{4}{3\sqrt{5}} \end{bmatrix}.$$

In the following example we use Matlab to determine the SVD of A. Note that in Matlab, D is used instead of Σ.

Example 4 (with Matlab)

A =

1	4	0	0
2	2	1	0
1	1	1	0
1	-2	-1	1
0	1	1	-2

>> [*U*,*D*,*V*]=svd(*A*)

U =

-1502/2109	-420/1811	933/1723	-571/1799	-385/1814
-735/1537	-587/1526	-2355/4586	175/2171	539/907
-735/2812	-125/881	-686/1667	905/1923	-1309/1814
193/548	-634/1027	-763/2991	-325/538	-231/907
-697/2602	2969/4712	-5867/12844	-507/916	-231/1814

D =

2635/477	0	0	0
0	1561/584	0	0
0	0	1679/880	0
0	0	0	914/1093
0	0	0	0

$V =$

-423/1481	-928/1409	-468/775	-1301/3756
-353/387	17/1756	451/1102	91/4496
-461/1873	202/749	-512/867	457/635
741/4610	-727/1035	322/933	347/577

Theorem 3.2

Let $A \in M_{m,n}(\mathbf{C})$, $m \geq n$ with singular values $\{s_1, s_2, ..., s_n\}$ of A. Then A^*, $\overline{A}$ and A^T have the same singular values of A. [11, 12]

Proof

Consider the SVD of $A = U\Sigma V^*$ where $\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}$ and S = diag(s_1, s_2, . . . , s_n).

Then $A^* = V\Sigma^* U^*$, where Σ^* = $[S \quad 0]$, V and U^* are unitary. So A and A^* have the same singular values $\{s_1, s_2, ..., s_n\}$.

Similarly, (since $\overline{U}$ and U^T are unitary for any unitary matrix U) we show that $\overline{A}$ and A^T have the same singular values of A. ■

Theorem 3.3

If A is real then U and V can be chosen to be real. [4]

Proof

Since A is real then $A^T A$ and AA^T are symmetric and both have real eigenvalues. We then choose the eigenvectors of $A^T A$ and AA^T to be real. ■

The SVD decomposition that has been found in the examples above is the full SVD. Some applications require a faster and more economical SVD. We now consider reduced versions of the SVD: thin and compact SVD.

1- Thin SVD

With $A \in M_{m,n}(\mathbf{C})$, with $m \geq n$, the full SVD of A is: $A=U\Sigma V^*$, where $\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}$. We can write U as [U_n U_{m-n}] where $U_n \in M_{m,n}, U_{m-n} \in M_{m,m-n}$

Noting that the elements of U_{m-n} will multiply the zero elements of Σ, then write A as $A=U_nSV^*$. This version of SVD is called the thin SVD, where V is unitary but U_n is no more unitary, but it has the property that $U_n^*U_n=I_n$(i.e., only has orthogonal columns).

So, in the thin SVD we only need to calculate the first n columns of U, and obviously it is faster than the full SVD especially when $m >> n$.

It is easy to see that if A is square, then the full and the thin SVD are the same.

Example 5

Consider $A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$ as in Example 1 its full SVD was given by

$$A= \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

Since 3-2=1 so we delete the last column of *U*, with $S = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix}$ we get the thin SVD of *A*

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

Example 6(with Matlab):

A =

1	2	-4	8
5	0	3	6
3	7	4	8
11	-2	0	0
0	-1	0	0

>> [*U*,*S*,*V*]=svd(*A*,0)

U =

-442/1061	-828/2591	-827/977	-365/4127
-327/656	625/4652	273/2518	405/491
-529/790	-27189/74770	893/1765	-216/577
-693/1928	621/719	-337/2939	-616/1865
135/6851	491/12201	-169/3312	434/1735

$S =$

4239/271	0	0	0
0	5713/515	0	0
0	0	239/37	0
0	0	0	1041/337

$V =$

-2833/4995	1214/1537	-109/14718	-1399/5998
-618/2005	-879/1969	383/1162	-1302/1685
-374/2333	237/11593	1029/1159	3617/8395
-1259/1686	-499/1188	-765/2383	1231/3050

Note

$$A = U_n S V^* = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad . \quad . \quad \mathbf{u}_n] \begin{bmatrix} s_1 & 0 & . & . & 0 \\ 0 & s_2 & 0 & . & . \\ . & 0 & . & 0 & . \\ . & . & 0 & . & 0 \\ 0 & . & . & 0 & s_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^* \\ \mathbf{v}_2^* \\ . \\ . \\ \mathbf{v}_n^* \end{bmatrix}$$

Where $\mathbf{u}_{\mathrm{i}}$ and $\mathbf{v}_{\mathrm{i}}$, i=1,2, . . .,n . are the columns of U and V respectively .

Then the outer product sum is defined as:

$$A = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] \begin{bmatrix} s_1 \mathbf{v}_1^* \\ s_2 \mathbf{v}_2^* \\ . \\ . \\ s_n \mathbf{v}_n^* \end{bmatrix} = \sum_{i=1}^{n} s_i \mathbf{u}_i \mathbf{v}_i^* = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^* \text{, since } s_k = 0,\ k = r+1, \ldots, n \quad [6]$$

2- Compact SVD

If $A \in M_{m,n}(\mathbf{C})$ with $m \geq n$ and rank r. Then the compact SVD of A is $A = U_r S_r V_r^*$, where $S_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & . & \\ & & & \sigma_r \end{bmatrix}$, U_r has only the first r columns of U and V_r has only the first r columns of V corresponding to the non zero singular values of A.

This is the second type of the reduced SVD, which is the same as the thin SVD if A has full rank, but if $r \ll n$ then this decomposition will be faster than both the full and the thin SVD.

The compact SVD is sometimes called the economy version of the SVD. And it can be calculated by matlab with the order [U,S,V]=svds(A,(rank(A))).

Example 7

Consider $A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$ in Example 5, which has full rank =2, then the compact SVD is the same as the thin SVD which means that it has the same decomposition as in Example 5.

3.2 The Singular Value Decomposition Versus the Spectral Decomposition

In this section. Let $A \in M_n(\mathbf{C})$ be normal , and consider its spectral decomposition $A = UDU^*$ and its SVD $A = V\Sigma W^*$ with $\Sigma = \mathrm{diag}(s_1, s_2, \ldots, s_n)$, $D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and U, V and $W \in M_n(\mathbf{C})$ are unitary .

Note

We have shown that if A is real then V and W can be chosen to be real. But even if A is real then the spectral decomposition may not be real.

But what is the relation between the spectral decomposition and the SVD, and when are they equal?

Theorem 3.4

If $A \in M_n(\mathbf{C})$ is normal . Then its SVD is given by $A = U\,|D|\,(U\overline{E})^*$ where $|D|$ = diag $(|l_1|, |l_2|, \ldots, |l_n|)$, E = diag $(e^{ij_1}, e^{ij_2}, \ldots, e^{ij_n})$ where $l_k = |l_k|\, e^{ij_k}, j_k \in (-p, p)$, are the eigenvalues of A. [11]

Proof

Since A is normal then its spectral decomposition is $A = UDU^*$. For each eigenvalue l of A, $l_k = |l_k|\, e^{ij_k}, j_k \in (-p, p)$.

So D = diag$(|l_1|\, e^{ij_1}, |l_2|\, e^{ij_2}, \ldots, |l_n|\, e^{ij_n})$ and hence

D = diag$(|l_1|, |l_2|, \ldots, |l_n|)$ diag $(e^{ij_1}, e^{ij_2}, \ldots, e^{ij_n})$.

Let $|D|$ = diag $(|l_1|, |l_2|, \ldots, |l_n|)$ and E = diag $(e^{ij_1}, e^{ij_2}, \ldots, e^{ij_n})$.

Then $A = U\,|D|\,EU^* = U\,|D|\,(U\overline{E})^*$ is the SVD of A with $\Sigma = |D|$, $V = U\overline{E}$ and the columns of U are orthonormal eigenvectors of A. ■

Remark

If $A \in M_n(\mathbf{C})$ is normal then we have $s_i = |l_i|, i = 1, 2, \ldots, n$.

Example 8

Consider the matrix $A=\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$ of Example 8 in Chapter 2.

We know that $A=\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$ is normal and its spectral decomposition is:

$$A=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1+i & 0\\ 0 & 1-i\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}.$$

Since $(l_1, l_2)=(1+i, 1-i)=(\sqrt{2}e^{i\frac{p}{4}}, \sqrt{2}e^{-i\frac{p}{4}})$, then $|D|=\textbf{diag}(\sqrt{2},\sqrt{2})$ and

$E=\textbf{diag}\,(e^{i\frac{p}{4}}, e^{-i\frac{p}{4}})=\textbf{diag}\,(\frac{1}{\sqrt{2}}(1+i), \frac{1}{\sqrt{2}}(1-i))$.

So,

$$A=U\,|D|\,(U\overline{E})^*=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}\end{bmatrix}\left(\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}(1-i) & 0\\ 0 & \frac{1}{\sqrt{2}}(1+i)\end{bmatrix}\right)^*$$

$$=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}\end{bmatrix}\begin{bmatrix}\frac{1+i}{2} & \frac{1+i}{2}\\ \frac{1-i}{2} & \frac{-1+i}{2}\end{bmatrix}=U\Sigma V^*$$ is its SVD.

Note

If we define $\text{sign}(l_i)=\text{sgn}(l_i)=\begin{cases}1 & \text{if } l_i \geq 0\\ -1 & \text{if } l_i < 0\end{cases}$, and if in addition of normal, A is Hermitian then all its eigenvalues are real so, $E=\overline{E}$ and hence $e^{ij_k}=\pm 1$. Hence, E = diag(sgn(l_1) , sgn(l_2) , . . . , sgn(l_n)) where sgn(0)=1 and $A=U\,|D|\,EU^*=U\,|D|\,(UE)^*$ is the singular value decomposition of the Hermitian matrix A.

If A is also positive definite then all the eigenvalues are real and positive and $E=I$ and in this case the spectral decomposition and the SVD are the same.

Example 9

Find the SVD of

$$A=\begin{bmatrix} 5 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 3 \end{bmatrix}.$$

Solution:

The eigenvalues of A are 287/4705, 687/329 and 3211/409 are all positive real numbers then the matrix is positive definite so the spectral decomposition and the SVD are the same.

The spectral decomposition of A is $A = UDU^*$ where

$$U=\begin{bmatrix} -580/787 & -539/912 & -1247/3802 \\ 1247/3802 & -580/787 & 539/912 \\ -539/912 & 1247/3802 & 580/787 \end{bmatrix} \text{and } D=\begin{bmatrix} 3211/409 & 0 & 0 \\ 0 & 687/329 & 0 \\ 0 & 0 & 287/4705 \end{bmatrix}.$$

Remark

If $A = U\Sigma V^*$ is the SVD of A then:

$$AA^* = U\Sigma V^* V\Sigma^* U^* = U\Sigma\Sigma^* U^*$$
$$A^*A = V\Sigma^* U^* U\Sigma V^* = V\Sigma^*\Sigma V^*$$

are the spectral decomposition of AA^* and A^*A, respectively.

3.3 Matrix Properties via SVD

The power of SVD comes from all the information that can be gleaned from it. In this section we show these information.

For the next theorems (see [1]), let $A \in M_{m,n}(\mathbf{C})$ with $m \geq n$, rank $A=r$ with singular values $s_1, s_2, \ldots, s_n$.

Theorem 3.5

Given the SVD of A as $A = U\Sigma V^*$ then:

1. The singular vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ form an orthonormal basis for Range A .
2. The singular vectors $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n$ form an orthonormal basis for Null A
3. The singular vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ form an orthonormal basis for Range A^*.
4. The singular vectors $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \ldots, \mathbf{u}_m$ form an orthonormal basis for Null A^*.

Proof

By Definition 3.1 write $A = U\Sigma V^*$ as $A\mathbf{v}_i = \sigma_i \mathbf{u}_i, i = 1,2,\ldots,n$.

(1) If $\sigma_i \neq 0_i$, then $\frac{A\mathbf{v}_i}{\sigma_i} = \mathbf{u}_i$ and so $\mathbf{u_i}$ is in the range of A. Since rank $A = r$, then there exist r non zero singular values and associated orthonormal eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ that span Range(A). Since they are orthonormal then they are linearly independent, so, they form an orthonormal basis of the range of A.

(2) For $i = r+1, \ldots, n$, we have $\sigma_i = 0$, , then $A\mathbf{v}_i=0$, $i = r+1, \ldots, n$ and $\boldsymbol{v_i}$ is in the null space of A. Since the $\mathbf{v}_i$'s are orthonormal then they are linearly

independent; since there exists *n-r* zero singular values, where r is the rank of *A*, then $\mathbf{v}_{r+1},\mathbf{v}_{r+2},...,\mathbf{v}_n$ form the basis of the null space of *A*.

To prove 3 and 4, we use $A^*\mathbf{u}_i = s_i\mathbf{v}_i$ and the same idea as above. ■

Example 10

Let $A=\begin{bmatrix}1 & 5 & 3 & 7\\ 2 & 0 & -4 & -6\\ 4 & 7 & -1 & 2\end{bmatrix}$, compute the four fundamental subspaces associated with *A*.

Solution:

The SVD of *A* is $A=U\Sigma V^*$ where

$$U=\begin{bmatrix} -421/558 & 283/2942 & -509/784\\ 660/1523 & -1175/1754 & -548/909\\ -452/917 & -120/163 & 454/979\end{bmatrix}$$

$$\Sigma=\begin{bmatrix} 3571/295 & 0 & 0 & 0\\ 0 & 4772/599 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$V^*=\begin{bmatrix} -443/2884 & -287/481 & -607/2097 & -1153/1574\\ -531/1010 & -3119/5318 & 166/357 & 344/851\\ 379/459 & -382/891 & 817/2237 & 273/8597\\ 1009/7478 & -528/1549 & -825/1096 & 812/1485\end{bmatrix}$$

since r(*A*)=2, we have:

1- The first two columns of *U* form an orthonormal basis for range *A*. so

$$\text{range } A=\text{span}\left\{\begin{bmatrix} -421/558\\ 660/1523\\ -452/917\end{bmatrix}, \begin{bmatrix} 283/2942\\ -1175/1754\\ -120/163\end{bmatrix}\right\}.$$

2- The last two columns of *V* form an orthonormal basis for null *A*. so

$$\text{Null } A = \text{span}\left\{\begin{bmatrix} 379/459 \\ -382/891 \\ 817/2237 \\ 273/8597 \end{bmatrix}, \begin{bmatrix} 1009/7478 \\ -528/1549 \\ -825/1096 \\ 812/1485 \end{bmatrix}\right\}$$

3- The first two columns of V form an orthonormal basis for range A^*. so

$$\text{range } A^* = \text{span}\left\{\begin{bmatrix} -443/2884 \\ -287/481 \\ -607/2097 \\ -1153/1574 \end{bmatrix}, \begin{bmatrix} -531/1010 \\ -3119/5318 \\ 166/357 \\ 344/851 \end{bmatrix}\right\}.$$

4- The last column of V forms an orthonormal basis for null A^*. so

$$\text{null } A^* = \text{span}\left\{\begin{bmatrix} -509/784 \\ -548/909 \\ 454/979 \end{bmatrix}\right\}.$$

Theorem 3.6

If $A \in M_n(\mathbf{C})$ with non zero singular values and with SVD $A = U\Sigma V^*$ then A^{-1} exists and $A^{-1} = V\Sigma^{-1}U^* = \sum \frac{1}{s_i} v_i u_i^*$, where $\Sigma^{-1} = \text{diag}(\frac{1}{s_1}, \frac{1}{s_1}, \ldots, \frac{1}{s_n})$. [8, 12] ■

So, the singular values of A^{-1} are the reciprocal of the singular values of A.

Theorem 3.7

Let $A = U\Sigma V^*$ then

1- $\|| A \||_2 = s_1$

2- $\|| A^{-1} \||_2 = \frac{1}{s_n}$ where A is square and non singular.

3- $\|| A \||_F = \sqrt{s_1^2 + s_2^2 + \mathbf{L} + s_n^2}$

Proof

1- By definition of spectral norm $\||A\||_2$ = max $\{\sqrt{l}, l$ is an eigenvalues of $AA^*\}$. So $\|| A \||_2 = s_1$.

2- By theorem 3.6, write $A^{-1} = V\Sigma^{-1}U^*$, then since the spectral norm is unitary invariant we have $||| A^{-1} |||_2 = ||| V\Sigma^{-1}U^* |||_2 = ||| \Sigma^{-1} |||_2 = \frac{1}{s_n}$.

3- $||| A |||_F = ||| U\Sigma V^* |||_F = ||| \Sigma |||_F = \sqrt{s_1^2 + s_2^2 + \ldots + s_n^2}$. (Since the Frobenius norm is unitary invariant). [22] ■

Theorem 3.8

The condition number of a non zero matrix A with respect to the spectral norm is $A = \frac{s_1}{s_n}$ [9]

Proof

If A is singular then $s_n = 0$ and the condition number is ∞,

So suppose A is invertible then condition number is $||| A |||_2 ||| A^{-1} |||_2 = \frac{s_1}{s_n}$ ■

Theorem 3.9

If $A \in M_n(\mathbf{C})$ then $|\det A| = \prod_{i=1}^{n} s_i$ [12]

Proof

Let $A = U\Sigma V^*$ be its SVD where $\Sigma \in M_n(\mathbf{C})$, then

$$|\det(A)| = |\det(U\Sigma V^*)| = |(\det(U))(\det(\Sigma))(\det(V^*))| = |\det(\Sigma)| = \det(\Sigma) = \prod_{i=1}^{n} s_i$$

Since the determinant of a unitary matrix is ± 1 and Σ is diagonal. ■

Theorem 3.10

For any $A \in M_n(\mathbf{C})$ and unitary matrix *W*, the matrices *A*, *AW* and *WA* have the same singular values. [11, 12]

Proof

Suppose $WA = U\Sigma V^*$ is the SVD of *WA*. Then $A = W^* U\Sigma V^* = H\Sigma V^*$, where $H=W^*U$ is unitary. So, *A* and *WA* have the same singular values.

Similarly, *A* and *AW* have the same singular values. ■

Theorem 3.11

A matrix $A \in M_n(\mathbf{C})$ is unitary if and only if all its singular values are equal to one.

Proof

Suppose *A* is unitary then $AA^* = I$. The singular values of *A* are the positive square roots of the eigenvalues of $AA^* = I$ which are all equal to 1.

Conversely, suppose that the singular values of *A* are all equal to 1, then $A = UIV^*$ is the SVD of *A*, where *U* and *V* are unitary. Hence $A=UV^*$ and *A* is unitary ■

3.4 Geometric Interpretation

The SVD provides us with a nice geometric interpretation of the action of a matrix**;** the image of the unit sphere under any *m*-by-*n* matrix is a hyperellipse (*m*-dimentional generalization of an ellipse).

One way to understand this is to consider the unit sphere in $\mathbf{R}^n$**.** So**,**

Suppose $\mathbf{x}$ lies on this unit sphere in $\mathbf{R}^n$. Then $\mathbf{x}$ can be written as $\mathbf{x} = x_1\mathbf{v} + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$, with $\sum_{i=1}^{n} x_i^2 = 1$ and where $\mathbf{v}_i$'s are orthonormal basis in $\mathbf{R}^n$. Let $A = U\Sigma V^*$ be the SVD of *A*. Then the image of $\mathbf{x}$ under *A* is

$$A\mathbf{x} = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^* x_i \mathbf{v}_i = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^* \mathbf{v}_i x_i = \sum_{i=1}^{r} \sigma_i x_i \mathbf{u}_i = \sum_{i=1}^{r} y_i \mathbf{u}_i$$, where $y_i = \sigma_i x_i$ and

r is the rank of A.

So, the image of a unit sphere is $y_1\mathbf{u}_1 + y_2\mathbf{u}_2 + \cdots + y_r\mathbf{u}_r$, where

$$\frac{y_1^2}{s_1^2} + \frac{y_2^2}{s_2^2} + \cdots + \frac{y_r^2}{s_r^2} = \sum_1^r x_i^2 \leq 1$$

If A has full column rank, then $n=r$ and so the inequality is actually an equality; otherwise, some of the x_i are missing on the right, and the sum can be anything from 0 to 1. This shows that A maps the unit sphere of $\mathbf{R}^n$ to a k-dimensional ellipsoid with semi-axes in the directions $\mathbf{u}_i$ and with the magnitudes s_i.

Example 11

Consider a matrix $A \in \mathbf{R}^3$ with rank 2, this matrix will affect the unit sphere in R^3 as figure (1) illustrates.

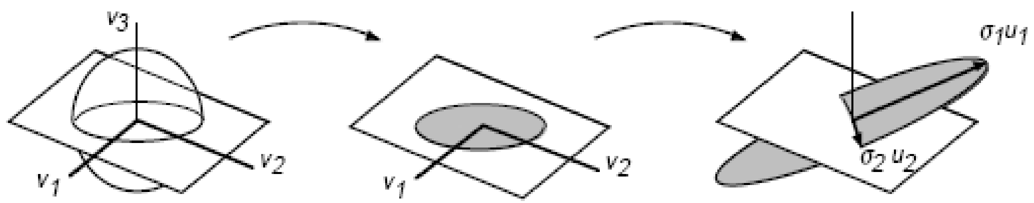

Figure (1)

Chapter Four

Applications of the Singular Value Decomposition

In mathematics, and particularly in linear algebra, the inverse of a matrix takes a big area in solving a set of linear equations. But since the inverse is not defined except for some square matrices, this fact pushed Moore (1912) and Penrose (1955) to establish -independently- a generalization of the inverse to rectangular matrices.

In this chapter we describe the Moore-Penrose Pseudo inverse, how to compute it, study some of its properties and more important how to use it in solving a system of linear equations or gives a least square solution (whether it is overdetermined or underdetermined system).

4.1 Moore-Penrose Pseudoinverse

This is the first application of the SVD and it is defined as below.

Definition 4.1

Let $A \in M_{m,n}(\mathbf{C})$. The Moore-Penrose pseudoinverse of A is defined as the matrix $A^+ \in M_{n,m}(\mathbf{C})$ satisfying the following four criteria:

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $(AA^+)^* = (AA^+)$
4. $(A^+A)^* = (A^+A)$

Remark

If $A=[a]$ is a 1×1 complex number then A^+ is defined by:

$$A^+ = \begin{cases} \frac{1}{a} & if, a \neq 0 \\ 0 & if, a = 0 \end{cases}$$

Example 1

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$. Then B is the Moore-Penrose pseudoinverse of A since it satisfies four criteria in Definition 4.1:

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

1- $(AB)A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = A$

2- $B(AB) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} = B$

3- $(AB)^* = AB = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ and

4- $(BA)^* = BA = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

So, $B = A^+$.

Before proving the existence of A^+ we prove a remark and a lemma.

Remark

If $A = \lfloor a_{ij} \rfloor$ and $B = \lfloor b_{ij} \rfloor$ are diagonal matrices and if *C*=*AB*, then *C* is diagonal with $c_{ij} = a_{ij} b_{ij}$.

Proof

$c_{ij} = (AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$.

In the summation, if $k \neq i$, then $a_{ik} = 0$ and if $k \neq j$, then $b_{kj} = 0$. Hence, $c_{ij} = a_{ii} b_{ij} + a_{ij} b_{jj}$. If $i = j$, then $c_{ij} = a_{ii} b_{ii}$. If $i \neq j$, then $a_{ij} = b_{ij} = 0$ and hence *C*=*AB* is diagonal and we have $c_{ij} = a_{ij} b_{ij}$. ■

Lemma 1

If *D* is a diagonal matrix, then its pseudoinverse D^{+} is given by

$$(D^{+})_{ij} = \begin{cases} D_{ij}^{+} & i = j \\ 0 & i \neq j \end{cases}$$ [7]

Proof

1. $(DD^{+}D)_{ij} = D_{ij}(D^{+}D)_{ij} = D_{ij} D^{+}{}_{ij} D_{ij} = D_{ij} D_{ij}{}^{+} D_{ij} = D_{ij} \Rightarrow DD^{+}D = D$

2. $(D^{+}DD^{+})_{ij} = D^{+}{}_{ij}(DD^{+})_{ij} = D_{ij}{}^{+} D_{ij} D_{ij}{}^{+} = D_{ij}{}^{+} \Rightarrow D^{+}DD^{+} = D^{+}$

3. $(DD^{+})_{ij}^{*} = \overline{(DD^{+})_{ji}} = \overline{D_{ji}(D^{+})_{ji}} = \overline{D_{ji} D_{ji}^{+}} = D_{ji} D_{ji}^{+} = D_{ij} D_{ij}^{+} = (DD^{+})_{ij}$
   $\Rightarrow (DD^{+})^{*} = DD^{+}$

4. Similar to 3 ■

Theorem 4.1

If $A \in M_{m,n}(\mathbf{C})$, then the Moore-Penrose pseudoinverse of *A* exists and is unique. [3 & 14]

Proof

Let $A \in M_{m,n}(\mathbf{C})$ with $m \geq n$.Use the thin SVD to write $A = U_n S V^*$, where V is unitary, $U_n^* U_n = I_n$ and $S = \text{diag}(s_1, s_2, \ldots s_n)$.

Define $B = VS^+U_n^*$. Then $A^+ = B$ shown below:

1. $ABA = U_n SV^* VS^+ U_n^* U_n SV^* = U_n SS^+ SV^* = U_n SV^* = A$

2. $BAB = VS^+ U_n^* U_n SV^* VS^+ U^* = VS^+ SS^+ U_n^* = VS^+ U_n^* = B$

3. $(AB)^* = (U_n SV^* VS^+ U_n^*)^* = (U_n SS^+ U_n^*)^* = U_n (SS^+)^* U_n^*$

$$= U_n (SS^+) U_n = U_n SV^* VS^+ U_n^* = AB$$

4. $(BA)^* = (VS^+ U_n^* U_n SV^*)^* = (VS^+ SV^*)^* = V(S^+ S)^* V^*$

$$= V(S^+ S) V^* = VS^+ U_n^* U_n SV^* = BA$$

So, $A^+ = VS^+U_n^*$ satisfies pseudoinverse conditions.

We now show the uniqueness.

Suppose $C \in M_{n,m}(\mathbf{C})$ be another pseudoinverse of A then:

1. $ABA=A$	1- $ACA=A$
2. $BAB=B$	2- $CAC=C$
3. $(AB)^*=AB$	3- $(AC)^*=AC$
4. $(BA)^*= BA$	4- $(CA)^*=CA$

As a first step we show $AB=AC$

$AB=(AB)^*=B^*A^*= B^*(ACA)^*= B^* A^*C^*A^*=(AB)^* (AC)^*= ABAC=AC$

In the same way we can show that $BA=CA$

Now, $B=BAB=BAC=CAC=C$. ∎

This theorem shows how to compute the Moore Penrose psuedoinvers of A, i.e., $A^+ = VS^+U_n^+$, where $A = U_n SV^*$.

Note

The pseudoinverse is a generalization of the inverse, i.e., if the matrix is invertible then $A^+ = A^{-1}$. [18]

Lemma 2

Let $A \in M_{m,n}(\mathbf{C})$. Then $(A^*)^+ = (A^+)^*, (A^T)^+ = (A^+)^T$ *and* $(\overline{A})^+ = \overline{(A^+)}$. [3, 21]

Proof

We want to show that $(A^+)^*$ is the pseudoinverse of A^*. So, we examine the four conditions:

1. $A^*(A^+)^*A^* = (AA^+A)^* = A^*$
2. $(A^+)^*A^*(A^+)^* = (A^+AA^+)^* = (A^+)^*$
3. $((A^+)^*A^*)^* = AA^+ = (AA^+)^* = (A^+)^*A^*$
4. $(A^*(A^+)^*)^* = A^+A = (A^+A)^* = A^*(A^+)^*$.

In the same way we can prove that:

$(A^T)^+ = (A^+)^T$ and $(\overline{A})^+ = \overline{(A^+)}$ ■

According to this lemma, we can state and prove the following theorem.

Theorem 4.2 (Identity Transformation)

1. $A^+ = A^+(A^+)^*A^*$
2. $A^+ = A^*(A^+)^*A^+$
3. $A = (A^+)^*A^*A$
4. $A = AA^*(A^+)^*$
5. $A^* = A^*AA^+$
6. $A^* = A^+AA^*$

Proof

1. $A^+(A^+)^*A^* = A^+(AA^+)^* = A^+AA^+ = A^+$

2. $A^*(A^+)^*A^+ = (A^+A)^*A^+ = A^+AA^+ = A^+$

3. $(A^+)^*A^*A = (AA^+)^*A = AA^+A = A$

4. $AA^*(A^+)^* = A(A^+A)^* = AA^+A = A$

5. This is the conjugate transpose of 3.

6. This is the conjugate transpose of 4. ■

We now state:

Theorem 4.3

Let $A \in M_{m,n}(\mathbf{C})$ and $k \in \mathbf{C}$ be nonzero. Then

1- $(kA)^+ = \frac{1}{k}A^+$.

2- $(A^+)^+ = A$. [4]

To prove $(kA)^+ = \frac{1}{k}A^+$, one can show that $\frac{1}{k}A^+$ satisfies the pseudoinverse conditions of kA. To prove $(A^+)^+ = A$, one also shows that A satisfies the pseudoinverse conditions of A^+.

Lemma 2 and Theorem 4.3 give us some properties of A^+ which are true for A^{-1}, but it is not true that all properties of the inverse also hold for the psuedoinverse. For example $(AB)^+ \neq B^+A^+$ in general.

Example 2

Let $A = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then $A^+ = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $B^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

So, $(AB)^+ = [1]^+ = 1$, whereas $B^+A^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{2}$.

But, when does $(AB)^+ = B^+A^+$ hold?

There are more than one case in which $(AB)^+ = B^+A^+$ holds.

Theorem 4.4:

If $(A^*A)^+$ is the psuedoinverse of A^*A, then $A^+ = (A^*A)^+A^*$.

Similarly, if $(AA^*)^+$ is the psuedoinverse of AA^*, then $A^+ = A^*(AA^*)^+$. [3]

To prove this theorem we have to prove some lemma:

Lemma 3

Let $A \in M_{m,n}(\mathbf{C})$ such that $A^*A = 0$. Then $A = 0$.

Proof

Fix $j = 1,...,n$. Then, $(A^*A)_{jj} = \sum_{i=1}^{m} a^*{}_{ji} a_{ij} = \sum_{i=1}^{m} \overline{a_{ij}} a_{ij} = \sum_{i=1}^{m} |a_{ij}|^2 = 0$ and so,

$|a_{ij}| = 0,\ i = 1,...,m$. Hence, $a_{ij} = 0,\ i = 1,...,m$. Since *j* is chosen arbitrary, then this is true for all *j*, hence $a_{ij} = 0,\ i = 1,...,m,\ j = 1,...,n$. Thus *A*=0. ■

Lemma 4

If $BAA^* = CAA^*$ *then* $BA = CA$ **,**and if $A^*AB = A^*AC \Rightarrow AB = AC$

Proof:

If *B*=*C* then *BA*=*CA*. So, suppose $B \neq C$. Then

$BAA^* = CAA^* \Rightarrow BAA^* - CAA^* = 0 \Rightarrow (BA - CA)A^* = 0 \Rightarrow (BA - CA)A^*(B - C)^* = 0$

$(BA - CA)(BA - CA)^* = 0 \Rightarrow$ (by lemma 3) $BA - CA = 0 \Rightarrow BA = CA$. ■

Proof of Theorem 4.4

To prove this theorem we have to show that $C=(A^*A)^+A^*$ is the pseudoinverse of A by showing the four criteria:

Since $(A^*A)^+$ is the psuedoinverse of A^*A, then:

1- $(A^*A)(A^*A)^+(A^*A)=A^*A$. Thus by Lemma 4, $A((A^*A)^+A^*)A=A$.

.So, $ACA=A$.

2- $(A^*A)^+(A^*A)(A^*A)^+=(A^*A)^+$. Multiplying the equation by A^* from the right side gives $((A^*A)^+A^*)A((A^*A)^+A^*)=(A^*A)^+A^*$. So, $CAC=C$.

3- $(AC)^*=(A(A^*A)^+A^*)^*=A((A^*A)^+)^*A^* = A(A^*A)^+A^*=AC$

4- $(CA)^*=((A^*A)^+A^*A)^*=((A^*A)^+(A^*A))^*=(A^*A)^+(A^*A)=CA$

So, C satisfies the the psuedoinverse conditions. So, $C=A^+$ ■

Theorem 4.5

Let $A\in M_{m,k}(\mathbf{C})$ and $B\in M_{k,n}(\mathbf{C})$.

1- If A has orthonormal columns ($A^*A=I_k$), then $(AB)^+=B^+A^+$.

2- If B has orthonormal rows ($BB^*=I_k$) then $(AB)^+=B^+A^+$.

Proof

We only prove the first case. By, Theorem 4.4 we have

$(AB)^+=((AB)^*AB)^+(AB)^*=(B^*A^*AB)^+B^*A^*=(B^*B)^+B^*A^*=B^+A^*$ (applying Theorem 4.4 on B).

Now, since $A^*A=I_k$, then $A^+=(A^*A)^+A^*=A^*$. Hence $(AB)^+=B^+A^+$. ■

Note

If A has linearly independent columns, then $(A^*A)^{-1}$ exists and then $A^+ = (A^*A)^{-1}A^*$. If A has linearly independent rows then $A^+ = A^*(AA^*)^{-1}$.

According to this theorem the pseudoinverse of any vector $\mathbf{v}$ can be written as $\mathbf{v}^+ = (\mathbf{v}^*\mathbf{v})^+\mathbf{v}^* = \begin{cases} \dfrac{\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}} & if\ \ \mathbf{v} \neq 0 \\ 0 & if\ \ \mathbf{v} = 0 \end{cases}$.

4.2 Computing the Moore-Penrose Pseudoinverse

In proving Theorem 4.1, we gave a method for finding the pseudoinverse of any $A \in M_{m,n}(\mathbf{C})$, and in this section we give some numerical examples to illustrate.

Note: the steps of finding the pseudoinverse is summarized in two steps:

1- if A has inverse then $A^+ = A^{-1}$

2- if A is not invertible write the thin SVD of $A = U_n S V^*$,then $A^+ = VS^+U_n^*$

Example 3

Let $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2i & 0 \\ 0 & 0 & 0 \end{bmatrix}$ which is a singular diagonal matrix, so by Lemma 1 of Section 4.1,

$$A^{+} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{-i}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Example 4

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. It is obvious that *A* is singular and positive semi definite.

Since *A* is positive semi definite then its SVD and spectral decomposition are the same.

$$AA^{*} = A^{*}A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

The eigenvalues of *A* are 2 and 0 corresponding to eigenvectors $(1,1)^{T}$ and $(1,-1)^{T}$, respectively. We now have:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = V \text{ and } S = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Hence $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ is the SVD of *A*.

So, $A^{+} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$.

Example 5

Let $A = \begin{bmatrix} 2 & 3i & 4 \\ -1 & 0 & 2 \\ 0 & 0 & 2i \\ 3i & -4 & i \\ 1 & 2 & 0 \end{bmatrix}$.

Then $A^*A = \begin{bmatrix} 15 & 2+18i & 9 \\ 2-18i & 29 & -16i \\ 9 & 16i & 25 \end{bmatrix}$

The eigenvalues of A^*A are 10384/195, 2910/217 and 4600/1967.

The normalized eigenvectors corresponding to these eigenvalues are

$$\begin{bmatrix} -1229/2672 \\ -1351/22308 \;+ 1040i/1487 \\ -471/868 \;-\; 167i/486 \end{bmatrix}, \begin{bmatrix} 1395/3376 \\ 490/3853 - 1110i/3131 \\ -551/680 - 365i/2079 \end{bmatrix} \& \begin{bmatrix} -503/640 \\ 293/2864 - 1080i/1813 \\ -127/1171 - 212i/2935 \end{bmatrix}$$

These normalized eigenvectors will be the columns of the unitary matrix *V*.

To obtain the unitary matrix *U*, consider the normalized eigenvectors of

$$AA^* = \begin{bmatrix} 29 & 6 & -8i & -22i & 2+6i \\ 6 & 5 & -4i & i & -1 \\ 8i & 4i & 4 & 2 & 0 \\ 22i & -i & 2 & 26 & -8+3i \\ 2-6i & -1 & 0 & -8-3i & 5 \end{bmatrix}.$$

The eigenvalues of AA^* are 10384/195, 2910/217, 4600/1967 and zero of multiplicity 2. The normalized eigenvectors corresponding to the nonzero eigenvalues are

$$\begin{bmatrix} -3521/4952 - 225i/5149 \\ -285/3326 - 108i/11489 \\ 108/11489 - 743i/4996 \\ 374/9869 - 478i/739 \\ -283/3554 + 355i/1852 \end{bmatrix}, \begin{bmatrix} -452/1225 - 520i/5937 \\ -356/641 - 254i/2649 \\ 254/2649 - 1009i/2280 \\ -410/4507 + 1856i/3679 \\ 383/2101 - 1026i/5299 \end{bmatrix} \& \begin{bmatrix} -231/1616 + 49i/4166 \\ 1603/4308 - 531i/5621 \\ 531/5621 - 279i/1967 \\ -500/2269 - 557i/10204 \\ -915/2407 - 1735i/2227 \end{bmatrix}$$

These normalized eigenvectors will be the columns of the matrix U_3.

So, $A = U_3 S V^*$ is the thin SVD of A where

$$S = \begin{bmatrix} 3853/528 & 0 & 0 \\ 0 & 4171/1139 & 0 \\ 0 & 0 & 1072/701 \end{bmatrix} \text{and } S^+ = \begin{bmatrix} 528/3853 & 0 & 0 \\ 0 & 1139/4171 & 0 \\ 0 & 0 & 701/1072 \end{bmatrix}.$$

Thus, the Moore Penrose psuedoinverse is $A^+ = VS^+U_3^*$

$$\begin{bmatrix} 64/835 + 11i/835 & -83/334 - 32i/835 & -32/835 - 27i/835 & 84/835 - 21i/167 & 369/1670 - 306i/835 \\ -14/835 + 21i/835 & 43/835 - 15i/167 & 78/835 - 18i/835 & -18/167 + 66i/835 & 269/835 + 53i/835 \\ 124/835 + i/167 & 187/1670 - 16i/835 & -107i/835 & 12/835 + 62i/835 & 63/1670 - 62i/835 \end{bmatrix}$$

The Moore Penrose pseudoinverse of a matrix $A \in M_{m,n}(\mathbf{C})$ also can be computed using Matlab with the function pinv(A).

Example 6

A =

$$\begin{matrix} 0 + 2.0000i & 5.0000 & 0 + 4.0000i & -3.0000 \\ 2.0000 & 0 & 0 & -1.0000 \\ 0 - 2.0000i & 3.0000 & 4.0000 & 0 + 5.0000i \end{matrix}$$

>> pinv(A)

$-0.0051 + .0162i \quad 0.5000 + 0.0185i \quad 0.0134 - 0.0467i$

$0.1553+.0887i$	$0.2440-0.0610i$	$0.0647-0.1479i$
$-0.0841-0.0564i$	$-0.1460+0.2957i$	$0.1081+0.0841i$
$-0.0102+0.0323i$	$0.0000+0.0370i$	$0.0268-0.0933i$

Theorem 4.6 [4]

Let $A\in M_{m,n}(\mathbf{C})$, then r(A)=r(A^+).

Proof

The proof follows directly from the fact that $A^+=VS^+U_n^*$

$r(A^+)=r(VS^+U_n^*)=r(S^+)=r(S)=r(A)$

Theorem 4.7

Let $A\in M_{m,n}(\mathbf{C})$, with a zero last column. Then A^+ is the pseudoinverse of the first $n-1$ columns with a zero last row.

Proof

Write $A=[B\quad 0]$, where $B\in M_{m,n-1}(\mathbf{C})$ and $\mathbf{0}$ is an m-dimensional zero column.

We show that $A^+=\begin{bmatrix}B^+\\0^+\end{bmatrix}$, where B^+ is the pseudoinverse of B and $\mathbf{0}^+$ is an n-dimensional zero row.

Let $C=\begin{bmatrix}B^+\\\mathbf{0}^+\end{bmatrix}$. We show that C satisfies the pseudoinverse conditions of A.

1- $ACA=[B\quad 0]\begin{bmatrix}B^+\\0^+\end{bmatrix}[B\quad 0]=\left[BB^+\right][B\quad 0]=\left[BB^+B\quad 0\right]=[B\quad 0]=A$

2- $CAC=\begin{bmatrix}B^+\\0^+\end{bmatrix}[B\quad 0]\begin{bmatrix}B^+\\0^+\end{bmatrix}=\begin{bmatrix}B^+B & 0\\0^+ & 0^+0\end{bmatrix}\begin{bmatrix}B^+\\0^+\end{bmatrix}=\begin{bmatrix}B^+BB^+\\0^+\end{bmatrix}=\begin{bmatrix}B^+\\0^+\end{bmatrix}=C$

3- $(AC)^*=\left([B\quad 0]\begin{bmatrix}B^+\\0^+\end{bmatrix}\right)^*=\left(\left[BB^+\right]\right)^*=BB^+=[B\quad 0]\begin{bmatrix}B^+\\0^+\end{bmatrix}=AC$

4- $$(CA)^* = \left(\begin{bmatrix} B^+ \\ 0^+ \end{bmatrix}\begin{bmatrix} B & 0 \end{bmatrix}\right)^* = \left(\begin{bmatrix} B^+B & 0 \\ 0^+ & 0^+0 \end{bmatrix}\right)^* = \begin{bmatrix} (B^+B)^* & 0 \\ 0^+ & 0^+0 \end{bmatrix}$$
$$= \begin{bmatrix} B^+B & 0 \\ 0^+ & 0^+0 \end{bmatrix} = \begin{bmatrix} B^+ \\ 0^+ \end{bmatrix}\begin{bmatrix} B & 0 \end{bmatrix} = CA$$

So, $C=A^+$.

Note

The same idea can be applied to $A \in M_{m,n}(\mathbf{C})$ with last zero row.

Example 7

Find the Moore Penrose pseudoinverse of

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

Write $A = \begin{bmatrix} & & 0 \\ & B & 0 \\ & & 0 \end{bmatrix}$ where $B = \begin{bmatrix} 1 & -1 & 3 \\ 1 & -2 & -3 \\ 5 & 4 & -1 \end{bmatrix}$

Then

$$B^+ = B^{-1} = \begin{bmatrix} 1/5 & 11/70 & 9/70 \\ -1/5 & -8/35 & 3/35 \\ 1/5 & -9/70 & -1/70 \end{bmatrix}$$

so

$$A^+ = \begin{bmatrix} 1/5 & 11/70 & 9/70 \\ -1/5 & -8/35 & 3/35 \\ 1/5 & -9/70 & -1/70 \\ 0 & 0 & 0 \end{bmatrix}.$$

Example 8

Find the Moore Penrose pseudo inverse of

$$A = \begin{bmatrix} 3.0000 & 1.0000i & 1.0000i \\ 2.0000 & -2.0000 & 1.0000 \\ 0 & 0 & 0 \end{bmatrix}$$

By Matlab the Moore Penrose pseudoinverse of A is

$$A^{+} = \begin{bmatrix} 0.2778 - 0.1111i & 0.0741 + 0.1667i & 0 \\ 0.2222 - 0.0556i & -0.3519 + 0.1111i & 0 \\ -0.1111 + 0.1111i & 0.1481 - 0.1111i & 0 \end{bmatrix}$$

Corollary 1

Let $A \in M_{m,n}(\mathbf{C})$, with a jth zero column, then the pseudoinverse is just the pseudoinverse of the others $n-1$ columns with a jth zero row .

Proof

Let $A = \begin{bmatrix} a_1 & .... & a_{j-1} & \mathbf{0} & a_{j+1} & .... & a_n \end{bmatrix}$ where each $a_k \in C^m$ and $\mathbf{0}$ is an m-dimensional zero column.

Post multiply A by the permutation matrix P that permutes cyclically the *jth* through the *nth* columns, in which the zero column (the *jth)* is transformed to the last column, i.e.,

$$AP = \begin{bmatrix} a_1 & .... & a_{j-1} & a_{j+1} & .... & a_n & \mathbf{0} \end{bmatrix}.$$

Then by Theorem 4.7, we have:

$$(AP)^{+} = \begin{bmatrix} a_1 & .... & a_{j-1} & a_{j+1} & .... & a_n & \mathbf{0} \end{bmatrix}^{+} = \begin{bmatrix} \begin{bmatrix} a_1 & .... & a_{j-1} & a_{j+1} & .... & a_n \end{bmatrix}^{+} \\ \mathbf{0}^{+} \end{bmatrix}$$

By Theorem 4.5 and since the permutation matrix has orthonormal rows ; we have $(AP)^{+} = P^{+}A^{+} = P^{-1}A^{+}$

$$A^{+} = PP^{-1}A^{+} = P(AP)^{+} = P\begin{bmatrix} \begin{bmatrix} a_1 & .... & a_{j-1} & a_{j+1} & .... & a_n \end{bmatrix}^{+} \\ \mathbf{0}^{+} \end{bmatrix} = P\begin{bmatrix} b_1 \\ . \\ . \\ . \\ b_{n-1} \\ \mathbf{0}^{+} \end{bmatrix} = \begin{bmatrix} b_1 \\ . \\ b_{j-1} \\ \mathbf{0}^{+} \\ b_j \\ . \\ b_{n-1} \end{bmatrix}$$

Note

The same idea can be applied to $A \in M_{m,n}(\mathbf{C})$ with a jth zero row.

Example 9

Let $A = \begin{bmatrix} 7 & 5 \\ 0 & 0 \\ 4 & 3 \end{bmatrix}$

Delete the second row, we then have

$$B = \begin{bmatrix} 7 & 5 \\ 4 & 3 \end{bmatrix}.$$

B has inverse $B^{-1} = \begin{bmatrix} 3 & -5 \\ -4 & 7 \end{bmatrix}$.

So, $A^{+} = \begin{bmatrix} 3 & 0 & -5 \\ -4 & 0 & 7 \end{bmatrix}$

Remark

Let $A \in M_{m,n}(\mathbf{C})$ with some zero rows or columns, then we can delete them and find the psuedoinverse of the resulting matrix then replace the zero rows with zero columns and vise versa and the consequential matrix will be the pseudoinverse of A.

Example 10

Find the pseudoinverse of A where $A = \begin{bmatrix} 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 6 & 0 \end{bmatrix}$

Solution:

By corollary 1of Theorem 4.7, we can delete the second and forth columns and the second and third rows.

The resulting matrix is $\begin{bmatrix} 3 & 2 \\ -1 & 6 \end{bmatrix}$ which is invertible and its inverse is

$$\begin{bmatrix} \frac{3}{10} & \frac{-1}{10} \\ \frac{1}{20} & \frac{3}{20} \end{bmatrix}$$

Now

$$A^{+} = \begin{bmatrix} \frac{3}{10} & 0 & 0 & \frac{-1}{10} \\ 0 & 0 & 0 & 0 \\ \frac{1}{20} & 0 & 0 & \frac{3}{20} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

one can examine the four criteria of the pseudoinverse.

Example 11

$A =$

3.0000	0	2.0000	0	6.0000
0	0	0	0	0

-1.0000	0	$0+2.0000i$	0	3.0000
$0+1.0000i$	0	$0-1.0000i$	0	3.0000

>> pinv(A)

0.1902 + 0.0878i	0.0000 + 0.0000i	-0.1854 + 0.0683i	-0.1951 - 0.2439i
-0.0000 - 0.0000i	-0.0000 - 0.0000i	-0.0000 - 0.0000i	0.0000 + 0.0000i
0.0927 - 0.0341i	-0.0000 - 0.0000i	-0.0390 - 0.2488i	-0.1463 + 0.3171i
0	0	0	0
0.0407 - 0.0325i	-0.0000 + 0.0000i	0.1057 + 0.0488i	0.1463 + 0.0163i

4.3 Linear Least Squares Problem

Let $A \in M_{m,n}(\mathbf{C})$. While solving the system $A\mathbf{x}=\mathbf{b}$ we have three cases :

1. Number of equations = the number of unknowns.
2. Overdetermined system: number of equations > than the number of unknowns.
3. Underdetermined system: number of equations < than the number of unknowns.

The solution of the above system exists only if **b** lies on the column space of *A*. If not, we can find a vector **x** that makes $A\mathbf{x}$ as close as possible to **b**. This occurs sometimes if we have a matrix $A \in M_{m,n}(\mathbf{C})$, with $m \geq n$ and the solution of the system $A\mathbf{x}=\mathbf{b}$ doesn't exist. So we can find what is called the least squares solution.

Definition 4.2

The optimal solution $\tilde{\mathbf{x}}$ of the system $A\mathbf{x}=\mathbf{b}$, where $A\in M_{m,n}(\mathbf{C})$ with $m\geq n$ which satisfies $\|\mathbf{b}-A\tilde{\mathbf{x}}\|_2$ is minimized overall vectors $\mathbf{x}\in\mathbf{C}^n$, is called the linear least squares solution. [16]

The least squares problem can be solved using the SVD.

Theorem 4.8

Let the matrix $A\in M_{m,n}(\mathbf{C})$, with $m\geq n$ has full column rank and its thin SVD $A=U_nSV^*$. Then the least squares problem has the unique solution $\tilde{\mathbf{x}}=VS^{-1}U_n^*$. [5, 6]

Proof:

By the SVD write A as $A=U\Sigma V^*=\begin{bmatrix}U_n & U_{m-n}\end{bmatrix}\begin{bmatrix}S\\0\end{bmatrix}V^*$, where $S=\text{diag}(s_1,s_2,\ldots s_n)$ and $s_1\geq s_2\geq\cdots\geq s_n>0$.

Using the fact that the norm is invariant under unitary transformations, we have:

$$\|\mathbf{b}-A\mathbf{x}\|_2^2=\|\mathbf{b}-U\begin{bmatrix}S\\0\end{bmatrix}V^*\mathbf{x}\|_2^2=\|U^*\|_2^2\|\mathbf{b}-U\begin{bmatrix}S\\0\end{bmatrix}V^*\mathbf{x}\|_2^2=\|U^*\mathbf{b}-\begin{bmatrix}S\\0\end{bmatrix}V^*\mathbf{x}\|_2^2$$

$$=\|\begin{bmatrix}U_n^*\\U_{m-n}^*\end{bmatrix}\mathbf{b}-\begin{bmatrix}SV^*\mathbf{x}\\0\end{bmatrix}\|_2^2=\|\begin{bmatrix}U_n^*\mathbf{b}-SV^*\mathbf{x}\\U_{m-n}^*\mathbf{b}\end{bmatrix}\|_2^2=\|U_n^*\mathbf{b}-SV^*\mathbf{x}\|_2^2+\|U_{m-n}^*\mathbf{b}\|_2^2$$

The vector $\tilde{\mathbf{x}}$ that minimize $\|b-A\mathbf{x}\|_2^2$ is $\tilde{\mathbf{x}}$ that sets the first sum to zero, i.e., $U_n^*\mathbf{b}-SV^*\tilde{\mathbf{x}}=0$, hence which equal $\tilde{\mathbf{x}}=VS^{-1}U_n^*\mathbf{b}=A^+\mathbf{b}$.

Since $S^{-1}=\text{diag}(\frac{1}{s_1},\frac{1}{s_2},\ldots,\frac{1}{s_n})$ then $\tilde{\mathbf{x}}$ can be written as $\tilde{\mathbf{x}}=\sum_{i=1}^{n}\frac{u_i^*b}{s_i}v_i$ [9]

Note that if $m=n$ then $\tilde{\mathbf{x}}=A^{-1}\mathbf{b}$.

Example 12

Find the least squares solution to the following system of linear equations:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ -5 \end{bmatrix}.$$

The least square solution is the vector $\mathbf{x} = A^{+}\mathbf{b}$. Using SVD to find A^{+} give us $A^{+} = \begin{bmatrix} 1/3 & 2/3 & -1/3 \\ 1/3 & -1/3 & 2/3 \end{bmatrix}$

So, $\mathbf{x} = A^{+}\mathbf{b} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$

But what if the rank is deficient?

Theorem 4.9

Let $A \in M_{m,n}(\mathbf{C})$, with $m \geq n$ and rank A =r <n then there is a set of (n-r) vectors that minimize $\| \mathbf{b} - A\mathbf{x} \|_2^2$.

Proof

Since rank A= r then there are (n-r) vectors that spans the null space of A and let $\mathbf{z}$ be one of them. Then $A\mathbf{z}$=0, So if $\tilde{\mathbf{x}}$ minimizes $\| \mathbf{b} - A\mathbf{x} \|_2^2$ then so does $\tilde{\mathbf{x}}$ + $\mathbf{z}$. ■

The above theorem shows that if A is rank deficient, then the least squares solution is not even unique, and the theorem below explains what kind of solution is it.

Definition 4.3

Let $A \in M_{m,n}(C), \mathbf{b} \in \mathbf{C}^m$ and let $L = \{\mathbf{x} \in C^n : \| \mathbf{b} - A\mathbf{x} \|_2 = \min\}$. A minimum norm solution is any $\tilde{\mathbf{x}} \in L$ such that $\|\tilde{\mathbf{x}}\|_2$ is minimum.

Theorem 4.10

Let $A \in M_{m,n}(\mathbf{C})$, $m \geq n$, $A = U\Sigma V^*$ as described before and r(A) $= r < n$. Then the minimum norm solution of the least square solutions occurs at $\tilde{\mathbf{x}} = V_r S_r^{-1} U_r^* \mathbf{b}$, where V_r and U_r contain the first r-columns of V and U, respectively and $S_r = \operatorname{diag}(s_1, s_2, \ldots, s_r)$.

Proof

A can be written as

$$A = U\Sigma V^* = [U_r \quad U_{m-r}] \begin{bmatrix} S_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^* \\ V_{n-r}^* \end{bmatrix} = U_r S_r V_r^* .$$

Then the error norm can now be written

$$\| \text{error} \|_2^2 = \| A\mathbf{x} - \mathbf{b} \|_2^2 = \left\| [U_r \quad U_{m-r}] \begin{bmatrix} S_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^* \\ V_{n-r}^* \end{bmatrix} \mathbf{x} - \mathbf{b} \right\|_2^2 .$$

Putting

$$\mathbf{y} = V^* \mathbf{x} = \begin{bmatrix} V_r^* \mathbf{x} \\ V_{n-r}^* \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix}, \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} U_r^* \mathbf{b} \\ U_{n-r}^* \mathbf{b} \end{bmatrix}$$

and since the norm is unitarily invariant, then the error becomes

$$\| \text{error} \|_2^2 = \left\| \begin{bmatrix} S_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} - \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} \right\|_2^2 = \| S_1 \mathbf{y_1} - \mathbf{b_1} \|_2^2 + \| \mathbf{b_2} \|_2^2 .$$

Thus, we can minimize the error if we choose $y_1 = S_r^{-1} \mathbf{b_1}$ and so

$$\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} = \begin{bmatrix} S_r^{-1} \mathbf{b_1} \\ \mathbf{y_2} \end{bmatrix},$$

where $\boldsymbol{y_2}$ is arbitrary(it is in the null space since the columns of V_{n-r} span the null space of A), so the solution of the least squares is not unique so we now have:

$$\|\tilde{\mathbf{x}}\|_2^2 = \|\mathbf{y}\|_2^2 = \|\mathbf{y}_1\|_2^2 + \|\mathbf{y}_2\|_2^2.$$

Therefore, we can obtain the solution of the minimum norm by letting $\boldsymbol{y}_2 = \mathbf{0}$; in this case $\tilde{\mathbf{x}} = V_r\mathbf{y}_1 = V_r S_r^{-1}\mathbf{b}_1 = V_r S_r^{-1} U_r^{*}\mathbf{b}_1$. ■

Example 13

Let

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

Note that ***b*** lies on the column space of *A*, so it has a solution but it is not unique since r(*A*) =1. One of the solutions is $(5,1)^T$ another is $(11,3)^T$ but the solution with the minimum norm is $\mathbf{x} = A^{+}b$

$$\tilde{\mathbf{x}} = \begin{bmatrix} 0.05 & 0 & -0.05 \\ -0.15 & 0 & 0.15 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/5 \\ -3/5 \end{bmatrix}$$

Example 14

Let

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

b doesn't lie in the column space of *A*, so this system doesn't have a solution but we have $\tilde{\mathbf{x}} = A^{+}\mathbf{b} = \begin{bmatrix} 1/20 \\ -3/20 \end{bmatrix}$ as the least squares solution.

The above two theorems illustrate how to solve an overdetermined system of linear equations using SVD, which can also solve an undetermined system of linear equations.

For the next theorem, let $A \in M_{m,n}(\mathbf{C})$, where $m<n$ and suppose A have full row rank and let $A = U\Sigma V^*$ be its SVD as described before. Define $\tilde{L} = \{\mathbf{x} \in C^n : A\mathbf{x} = \mathbf{b}\}$. We now have

Theorem 4.11

Let A and $\tilde{L}$ be as described above.

(a) The system $A\mathbf{x}=\mathbf{b}$ always has a solution but it is not unique.

(b) Of all solutions to $A\mathbf{x}=\mathbf{b}$, the solution that has minimum norm is unique and it is given by

(*) $$\tilde{\mathbf{x}}=V_m S^{-1}U^*\mathbf{b}$$

Proof:

(a) Given $A=U\Sigma V^*$, we can write

$A=U[S \quad 0]\begin{bmatrix} V_m^* \\ V_{n-m}^* \end{bmatrix}, V_m \in M_m(\mathbf{C})$, where S is an $m\times m$ diagonal containing the nonzero singular values of A, so S it is nonsingular.

Then $A\mathbf{x}=U[S \quad 0]\begin{bmatrix} V_m^*\mathbf{x} \\ V_{n-m}^*\mathbf{x} \end{bmatrix}=U[S \quad 0]\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix}=US\mathbf{y}_1=\mathbf{b}$.

So, we have $USV_m^*\mathbf{x}=\mathbf{b}$, so $\tilde{\mathbf{x}}=V_m S^{-1}U^*\mathbf{b}$ is a solution to $A\mathbf{x}=\mathbf{b}$.

Since $m<n$, then the nullspace of A is nonempty and is spanned by the vectors $v_{m+1},\ldots,v_n$ by Theorem 3.5. Let $\mathbf{z}$ be an element in the nullspace of A, then $\tilde{\mathbf{x}}+\mathbf{z}$ is also a solution to $A\mathbf{x}=\mathbf{b}$ and this proves (a).

(b) As seen in part (a), $\tilde{\mathbf{x}}+\mathbf{z}$ represents all solutions to $A\mathbf{x}=\mathbf{b}$. Since $\tilde{\mathbf{x}}=V_m S^{-1}U^*\mathbf{b}$, then $\tilde{\mathbf{x}}$ is a linear combination of the vectors $v_1,\ldots,v_m$; on the other hand $\mathbf{z}$ is a linear combination of the vectors $v_{m+1},\ldots,v_n$ and since the columns of V are orthonormal, then $\tilde{\mathbf{x}}$ & $\mathbf{z}$ are orthogonal. Hence, $\|\mathbf{x}+\mathbf{z}\|_2^2=\|\mathbf{x}\|_2^2+\|\mathbf{z}\|_2^2$ which is minimized only if of $\mathbf{z}=0$. So, the minimum norm solution to $A\mathbf{x}=\mathbf{b}$ is as described by (*). ■

Example 15:

Let

$$A = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 4 & 0 \end{bmatrix} \text{and } b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Then the minimum norm solution for the system $A\mathbf{x}=b$ is

$$\mathbf{x} = A^{+}b = \begin{bmatrix} 23/65 \\ 63/130 \\ -1/26 \end{bmatrix}$$

The rank deficient may or may not have a solution depending on the right hand side, and it can be treated as in Theorem 4.10

Summary: For any system of linear equations $A\mathbf{x} = b$, the solution if exists is $A^{+}b$; if there are more than one solution, then the minimum norm solution is also $A^{+}b$; if we have no solution to the overdetermined system then, again, the minimum linear least squares solution is $A^{+}b$.

Chapter Five

More Applications of the Singular Value Decomposition

In this chapter we study more applications of SVD such as low rank approximation of matrices, determining the effective rank of matrices and study image compression as an application of SVD in image processing.

5.1 Low Rank Approximation of Matrices

In science and engineering problems such as image processing, data compression and effective rank, one would like to approximate a given matrix by a lower rank matrix according to a given norm. One easy way to do this is simply the truncated SVD.

We will consider low-rank approximation of a matrix A according to spectral norm and Frobenius norm. First we give the theorem for the spectral norm.

Recall that:

For any matrix $A \in M_{m,n}(\mathbf{C})$, the Frobenius norm is defined as $|||A|||_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}}$. Whereas the spectral norm is defined as $|||A|||_2 = \max\{\sqrt{\lambda} : \lambda$ is an eigenvalues of $AA^*\}$.

Theorem 5.1:

Suppose $A \in M_{m,n}(\mathbf{C})$ with rank $r > k$ has SVD $A = U\Sigma V^*$ where $U \in M_m(\mathbf{C})$ and $V \in M_n(\mathbf{C})$ are unitary matrices and Σ contains the singular values of A on its diagonal arranged in decreasing order. Then

$$\min_{\text{rank}(Z)=k} ||| A - Z |||_2 = \sigma_{k+1}$$

This minimum occurs when

$$Z = A_k = U_k \Sigma_k V_k^*$$

where $U_k = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k) \in M_{m,k}(C)$, $V_k = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k) \in M_{n,k}(C)$ and

$\Sigma_k = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_k)$ [5]

Proof

We first show $||| A - A_k |||_2 = s_{k+1}$, write $A = \sum_{i=1}^{n} s_i \mathbf{u}_i \mathbf{v}_i^*$, and define $A_k = \sum_{i=1}^{k} s_i u_i v_i^*$. Then:

$||| A - A_k |||_2 = ||| \sum_{i=k+1}^{n} s_i u_i v_i^* |||_2 = s_{k+1}$. ($\sigma_i$'s are in decreasing order).

It remains to show that there is no rank k matrix closer to A than A_k. Suppose there exists $B \in M_{m,n}(\mathbf{C})$ with r(B)=k such that $||| A - B |||_2 < ||| A - A_k |||_2 = s_{k+1}$. Since the dimension of the null space of B = n-k, the space spanned by the set $T = \{v_1, v_2, ..., v_{k+1}\}$ has dimension k+1 and since the sum of the two dimensions is n+1, then there exists a unit vector w in their intersection. i.e., $B\mathbf{w} = 0$ and $\mathbf{w} = \sum_{i=1}^{k+1} c_i v_i$ with $\| \mathbf{w} \|_2^2 = \sum_{i=1}^{k+1} c_i^2 = 1$.

Thus, we now have:

$$V^*\mathbf{w}=\begin{bmatrix}\mathbf{v}_1^*\\ .\\ \mathbf{v}_{k+1}^*\\ .\\ \mathbf{v}_n^*\end{bmatrix}[c_1v_1+\cdots+c_{k+1}v_{k+1}]=\begin{bmatrix}c_1\\ .\\ c_{k+1}\\ 0\\ .\\ 0\end{bmatrix}$$

$$\|\Sigma V^*\mathbf{w}\|_2^2=\left\|\begin{bmatrix}c_1s_1\\ .\\ c_{k+1}s_{k+1}\\ 0\\ .\\ 0\end{bmatrix}\right\|_2^2\geq\left\|\begin{bmatrix}c_1s_{k+1}\\ .\\ c_{k+1}s_{k+1}\\ 0\\ .\\ 0\end{bmatrix}\right\|_2^2\geq s_{k+1}^2\left\|\begin{bmatrix}c_1\\ .\\ c_{k+1}\\ 0\\ .\\ 0\end{bmatrix}\right\|_2^2=s_{k+1}^2.$$

We now get:

$\||A-B\||_2^2\geq\|(A-B)\mathbf{w}\|_2^2=\|A\mathbf{w}\|_2^2=\|U\Sigma V^*\mathbf{w}\|_2^2=\|\Sigma V^*\mathbf{w}\|_2^2\geq s_{k+1}^2$, a contradiction. ■

Example 1

Let

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

Find a rank 1 approximation of A with respect to the spectral norm.

Solution:

We first find the SVD of A,

$$A=U\Sigma V^*=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}3&0\\0&2\\0&0\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix}.$$

The matrix A is of full rank, we want to approximate it with a rank 1 matrix, so we find:

$$U_1=\begin{bmatrix}0\\1\\0\end{bmatrix},\ \Sigma_1=[3]\ \text{and}\ V_1^*=[0\quad -1].$$

Then

$$A_1 = U_1\Sigma_1 V_1^* = \begin{bmatrix} 0 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix} \text{with } ||| A - A_1 |||_2 = ||| \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} |||_2 = s_2 = 2 .$$

Example 2

Let

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

Find a rank 2 approximation of A with respect to the spectral norm.

Solution:

Using Matlab, the SVD of A is $U\Sigma V^*$ where

$$U = \begin{bmatrix} -1153/1729 & 665/1223 & 587/1152 \\ -298/431 & -277/1410 & -2028/2917 \\ -515/1853 & -297/364 & 437/862 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 5139/892 & 0 & 0 & 0 & 0 \\ 0 & 2500/531 & 0 & 0 & 0 \\ 0 & 0 & 803/494 & 0 & 0 \end{bmatrix} \text{and}$$

$$V^* = \begin{bmatrix} -322/797 & -1929/2620 & 111/5761 & 397/1654 & -307/631 \\ -203/1437 & -135/326 & -506/1095 & 347/4158 & 1411/1840 \\ -199/865 & 837/2995 & 911/2936 & 704/823 & 1120/5533 \\ -392/4479 & 473/1378 & -3101/3746 & 815/3273 & -575/1612 \\ -767/882 & 1306/4341 & 153/2272 & -428/1137 & 335/3997 \end{bmatrix}.$$

This matrix is of rank 3, so we find $A_2 = U_2\Sigma_2 V_2^*$

$$U = \begin{bmatrix} -1153/1729 & 665/1223 \\ -298/431 & -277/1410 \\ -515/1853 & -297/364 \end{bmatrix}, \ \Sigma_2 = \begin{bmatrix} 5139/892 & 0 \\ 0 & 2500/531 \end{bmatrix} \text{and}$$

$$V^* = \begin{bmatrix} -322/797 & -1929/2620 & 111/5761 & 397/1654 & -307/631 \\ -203/1437 & -135/326 & -506/1095 & 347/4158 & 1411/1840 \end{bmatrix}.$$

The problem $\min_{rank(Z)=2} ||| A-Z |||_2$ has the solution $Z = A_2 = U_2\Sigma_2 V_2^*$.

So, $A_2 = \begin{bmatrix} 756/635 & 2315/1309 & -763/607 & -1490/2103 & 640/167 \\ 3701/2127 & 4672/1409 & 587/1674 & -962/931 & 188/153 \\ 571/480 & 4077/1472 & 689/395 & -2353/3338 & -2533/1169 \end{bmatrix}$

with $||| A - A_2 |||_2 = s_3 = 803/494$.

Example 3

Let

$$A = \begin{bmatrix} 1 & i & -i \\ 2 & 4 & -4 \\ 0 & 1 & -1 \end{bmatrix}$$

Approximate A with a rank 1 matrix(i.e., find A_1).

With Matlab $A = U\Sigma V^*$ where

$$U = \begin{bmatrix} -257/3386\text{-}217i/1026 & 4754/5955 - 371i/951 & -532/1753 + 425i/1631 \\ -1407/1489 + 217i/2052 & 215/1608 + 521i/2671 & 266/1753 - 348i/2671 \\ -217/1026 + 297i/11234 & -371/951 + 223i/4573 & -1264/1457 + 436i/2003 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 4747/753 & 0 & 0 \\ 0 & 1217/1085 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{and}$$

$$V^* = \begin{bmatrix} -517/1658 & -57265/85898 & 57265/85898 \\ 1353/1424 & -177/809 - 200i/7313 & 177/809 + 200i/7313 \\ 0 & 1658/2353 + 162i/2741 & 1658/2353 + 162i/2741 \end{bmatrix}.$$

Since r(A)=2, we find $A_1 = U_1 \Sigma_1 V_1^*$, where

$$U_1 = \begin{bmatrix} -257/3386 - 217i/1026 \\ -1407/1489 + 217i/2052 \\ -217/1026 + 297i/11234 \end{bmatrix}, \ \Sigma_1 = [\ 4747/753\ ],$$

$V_1^* = [\ -517/1658 \quad -57265/85898 \quad 57265/85898\ ]$ and

$$A_1 = \begin{bmatrix} 131/878 + 955i/2297 & 554/2665 + 365i/393 & -554/2665 - 365i/393 \\ 730/393 - 554i/2665 & 1651/410 + 277i/5330 & -1651/410 - 277i/5330 \\ 955/2297 - 277i/5330 & 1207/1337 & -1207/1337 \end{bmatrix}$$

and computing $|||\, A - A_1 \,|||_2 = s_2 = 1217/1085$.

Fact

For A and $B \in M_{m,n}(\mathbf{C})$, $tr(A^* B) = tr(BA^*)$.

Lemma 1

Let $A \in M_{m,n}(\mathbf{C})$, with $A = U\Sigma V^*$, then the matrices

$$\mathbf{u}_i \mathbf{v}_j^*, \quad i = 1,2,\ldots,m, \ j = 1,2,\ldots,n$$

form an orthonormal basis for $M_{m,n}(\mathbf{C})$.

proof

For $i \neq k$, $\mathbf{u}_i$ and $\mathbf{u}_k$ are orthonormal and by Fact above, we have

$tr((\mathbf{u}_i \mathbf{v}_j^*)(\mathbf{u}_k \mathbf{v}_l^*)) = tr(\mathbf{v}_j \mathbf{u}_i^* \mathbf{u}_k \mathbf{v}_l^*) = 0.$

Similarly, for $l \neq j$, $tr((\mathbf{u}_i \mathbf{v}_j^*)(\mathbf{u}_k \mathbf{v}_l^*)) = tr(\mathbf{v}_j \mathbf{u}_i^* \mathbf{u}_k \mathbf{v}_l^*) = 0.$

For $i = k$ & $l = j$, $(\mathbf{u}_i \mathbf{v}_j^*).(\mathbf{u}_i \mathbf{v}_j^*) = tr(\mathbf{v}_j \mathbf{u}_i^* \mathbf{u}_i \mathbf{v}_j^*) = tr(\mathbf{v}_j \mathbf{v}_j^*) = 1$.

So the set of matrices $\mathbf{u}_i \mathbf{v}_j^*$, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ are orthonormal and so linearly independent. Since we have *mn* such matrices then they form a basis for $M_{m,n}(\mathbf{C})$. ■

Theorem 5.2

Suppose $A \in M_{m,n}(\mathbf{C})$ with rank $r > k$ has SVD $A = U\Sigma V^*$ where $U \in M_m(\mathbf{C})$ **and** $V \in M_n(\mathbf{C})$ are unitary matrices and Σ contains the singular values of A on its diagonal arranged in decreasing order. Then

$$\min_{r(Z)=k} |||A - Z|||_F = \sqrt{\sum_{i=k+1}^{\min(m,n)} \sigma_i^2}$$

And it occurs for

$$Z = A_k = U_k \Sigma_k V_k^*$$

where $U_k = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k) \in M_{m,k}(C), V_k = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k) \in M_{n,k}(C)$ and $\Sigma_k = \operatorname{diag}(\sigma_1, \sigma_2, ..., \sigma_k)$. [6]

Proof

Let $Z \in M_{m,n}(\mathbf{C})$. By Lemma 1 above, we write $Z = \sum_{i,j} m_{ij} \mathbf{u}_i \mathbf{v}_j^*$ for some coefficients m_{ij} and for an orthonormal basis $M_{m,n}(\mathbf{C})$ of the form

$$\mathbf{u}_i \mathbf{v}_j^*, \quad i = 1,2,...,m,\ j = 1,2,...,n.$$

Now we have

$$|||A - Z|||_F^2 = \left\| \sum_{i,j} s_{ij} \mathbf{u}_i \mathbf{v}_j^* - \sum_{i,j} m_{ij} \mathbf{u}_i \mathbf{v}_j^* \right\|_F^2 = \left\| \sum_{i,j} (s_{ij} - m_{ij}) \mathbf{u}_i \mathbf{v}_j^* \right\|_F^2 = \sum_{i,j} (s_{ij} - m_{ij})^2$$

$= \sum_{i=j} (s_{ii} - m_{ii})^2 + \sum_{i \neq j} m_{ij}^2$. For minimum choose the second term equal to zero.

We then have

$$Z = \sum_i m_{ii} \mathbf{u}_i \mathbf{v}_i^*$$

Since the rank of Z is equal to the number of terms in this sum, we see that the constraint $r(Z) = k$ implies that we should have exactly k nonzero terms in the sum. To minimize the objective function, we then choose

$m_{ii} = s_{ii}$, $i = 1,2,...,k$. So, $Z = \sum_{i=1}^{k} s_i \mathbf{u}_i \mathbf{v}_i^*$ which gives the desired result. ■

Example 4

Consider the matrix $A = \begin{bmatrix} 1 & 2 & -1 & 0 & 4 \\ 2 & 3 & 0 & -2 & 1 \\ 1 & 3 & 2 & 0 & -2 \end{bmatrix}$ in Example 5.2.Then the problem $\min_{r(Z)=2} ||| A - Z |||_F$ has the same solution A_2 with

$$\|A - A_2\|_F = \sqrt{\sum_{i=3}^{3} s_i^2} = \mathbf{803/494}\ .$$

Example 5

Consider the matrix $A = \begin{bmatrix} 1 & i & -i \\ 2 & 4 & -4 \\ 0 & 1 & -1 \end{bmatrix}$ in Example 5.3. Then the problem $\min_{r(Z)=1} ||| A - Z |||_F$ has the same solution A_1 with

$$\|A - A_1\|_F = \sqrt{\sum_{i=2}^{3} s_i^2} = \mathbf{1}217/1085\,\mathbf{.}$$

5.2 Image Compression Using the Singular Value Decomposition

Another approximation of SVD that will use the low rank approximation is the image compression, which we define below.

A computer represents an image on a display device with a set of evenly spaced coloured dots called pixels (picture elements). If these pixels are close enough together they approximate a continuous image.

Each pixel represents a section of a picture, but how is all the relevant information associated? In a simple case of black and white pictures each pixel can be one of two states 0 or 1. This can be represented with elements of Z_2 (Integers modulo 2). If the picture is a greyscale then every pixel represents the intensity using a real number from 0 (black) to ∞. However any picture can be viewed as a matrix; the picture can be partitioned into tiny squares. Each square represents an entry in that matrix. If the picture is partitioned into *m* by *n* matrix, then we need *mn* spaces in the computer to store it. But we can compress (reduce the number of spaces in which we store the picture) some images using SVD.

Consider an image which is partitioned into *m* by *n* matrix with rank *r*, then we need $r(m+n+1)$ spaces for storage by SVD instead of *mn*, this is because any matrix can be written with the SVD as $\sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^* = U_r \Sigma_r V_r^*$. That computes $r(m+n+1)$ entries.

We can further approximate the matrix by leaving off more singular terms of the matrix *A*. Since the singular values are arranged in decreasing order, the last terms will have the least effect on the overall image. Doing this reduces the amount of space required to store the image on a computer. The following example illustrates this.

Example 6

Consider the black and white image of a tree in Figure 2 which we are going to compress to a computer using SVD. It is partitioned - for example - into (12)(12)=144 tiny squares (see Figure 3). Each square has a blackness level, and we assign the number 1 to any at least half black and 0 otherwise.

The storage of this image needs 144 spaces. And this number increases if we partition this photo in more than 144 squares.

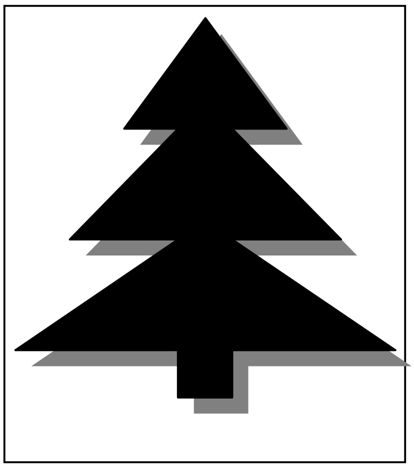

Figure (2) The original image

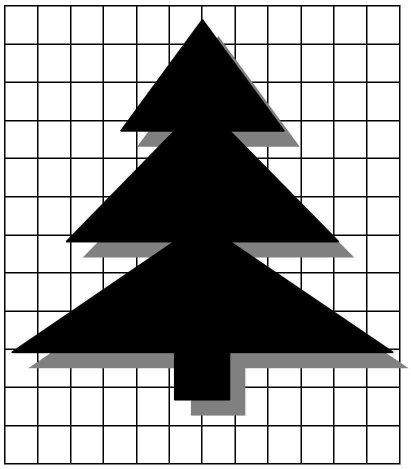

Figure (3)

The matrix that represents Figure (3) is

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

Using matlab the SVD of A is $A = U\Sigma V^*$ where

U=

Columns 1 through 9

0	0	0	0	0	0	0	0	1.0000
-0.1721	-0.2350	-0.4472	0.4638	-0.0000	-0.0000	-0.0000	-0.0000	0
-0.3238	-0.2777	-0.0000	-0.2609	0.8614	-0.0849	0.0280	-0.0000	0
-0.3238	-0.2777	-0.0000	-0.2609	-0.2172	0.8111	0.2118	-0.0000	0
-0.3238	-0.2777	-0.0000	-0.2609	-0.3628	-0.5517	0.5604	-0.0000	0
-0.3982	0.1341	0.4472	0.3514	-0.0000	0.0000	-0.0000	-0.7071	0
-0.3238	-0.2777	-0.0000	-0.2609	-0.2815	-0.1746	-0.8002	0.0000	0
-0.3982	0.1341	0.4472	0.3514	-0.0000	0.0000	0.0000	0.7071	0
-0.4521	0.7383	-0.4472	-0.2249	0.0000	-0.0000	-0.0000	-0.0000	0
-0.1721	-0.2350	-0.4472	0.4638	-0.0000	-0.0000	-0.0000	-0.0000	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Columns 10 through 12

0	0	0
-0.7071	0	0
0	0	0
0	0	0
-0.0000	0	0
-0.0000	0	0
0.0000	0	0
0.0000	0	0
0	0	0
0.7071	0	0

0	1.0000	0
0	0	1.0000

Σ =

5.7923	0	0	0	0	0	0	0	0	0	0	0
0	2.2109	0	0	0	0	0	0	0	0	0	0
0	0	1.4142	0	0	0	0	0	0	0	0	0
0	0	0	1.2494	0	0	0	0	0	0	0	0
0	0	0	0	0.0000	0	0	0	0	0	0	0
0	0	0	0	0	0.0000	0	0	0	0	0	0
0	0	0	0	0	0	0.0000	0	0	0	0	0
0	0	0	0	0	0	0	0.0000	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

V=

Columns 1 through 9

0	0	0	0	0	0	0	0	1.0000
-0.0780	0.3339	-0.3162	-0.1800	0.6260	-0.2813	-0.3093	-0.4241	0
-0.0780	0.3339	-0.3162	-0.1800	-0.7721	-0.1665	-0.2371	-0.2619	0
-0.2155	0.4553	0.3162	0.3825	0.0087	0.6238	-0.1205	-0.3073	0
-0.4391	-0.0472	0.3162	-0.4527	-0.0236	-0.0852	0.5787	-0.3930	0
-0.4985	-0.2598	-0.3162	0.2898	0.0000	-0.0000	0.0000	0.0960	0

-0.4985	-0.2598	-0.3162	0.2898	-0.0000	0.0000	-0.0000	-0.0960	0
-0.4391	-0.0472	0.3162	-0.4527	0.0236	0.0852	-0.5787	0.3930	0
-0.2155	0.4553	0.3162	0.3825	-0.0087	-0.6238	0.1205	0.3073	0
-0.0780	0.3339	-0.3162	-0.1800	0.0731	0.2239	0.2732	0.3430	0
-0.0780	0.3339	-0.3162	-0.1800	0.0731	0.2239	0.2732	0.3430	0
0	0	0	0	0	0	0	0	0

Columns 10 through 12

0	0	0
-0.0581	0	0
-0.0359	-0.0000	0
-0.0421	0.0000	0
-0.0539	0.0000	0
-0.7006	-0.0000	0
0.7006	0.0000	0
0.0539	-0.0000	0
0.0421	0.0000	0
0.0470	-0.7071	0
0.0470	0.7071	0
0	0	1.0000

Since we have 4 nonzero singular values then r(A) = 4<12 ; so, using the compact SVD we can write A as $A = U_4 \Sigma_4 V_4^*$ where

U_4 =

0	0	0	0
-0.1721	-0.2350	-0.4472	0.4638

-0.3238	-0.2777	-0.0000	-0.2609
-0.3238	-0.2777	-0.0000	-0.2609
-0.3238	-0.2777	-0.0000	-0.2609
-0.3982	0.1341	0.4472	0.3514
-0.3238	-0.2777	-0.0000	-0.2609
-0.3982	0.1341	0.4472	0.3514
-0.4521	0.7383	-0.4472	-0.2249
-0.1721	-0.2350	-0.4472	0.4638
0	0	0	0
0	0	0	0

$\Sigma_4 =$

5.7923	0	0	0
0	2.2109	0	0
0	0	1.4142	0
0	0	0	1.2494

and V_4 =

0	0	0	0
-0.0780	0.3339	-0.3162	-0.1800
-0.0780	0.3339	-0.3162	-0.1800
-0.2155	0.4553	0.3162	0.3825
-0.4391	-0.0472	0.3162	-0.4527
-0.4985	-0.2598	-0.3162	0.2898
-0.4985	-0.2598	-0.3162	0.2898
-0.4391	-0.0472	0.3162	-0.4527

-0.2155	0.4553	0.3162	0.3825
-0.0780	0.3339	-0.3162	-0.1800
-0.0780	0.3339	-0.3162	-0.1800
0	0	0	0

If we use rank 1 approximation for *A* we have:

$A_1 = U_1 \Sigma_1 V_1^* = \sum_{i=1}^{1} s_i u_i v_i^*$ (need (1(12+12+1)=25 spaces for storage)

Columns 1 through 9

0	0	0	0	0	0	0	0	0
0	0.0778	0.0778	0.2149	0.4378	0.4971	0.4971	0.4378	0.2149
0	0.1464	0.1464	0.4042	0.8235	0.9349	0.9349	0.8235	0.4042
0	0.1464	0.1464	0.4042	0.8235	0.9349	0.9349	0.8235	0.4042
0	0.1464	0.1464	0.4042	0.8235	0.9349	0.9349	0.8235	0.4042
0	0.1800	0.1800	0.4971	1.0127	1.1498	1.1498	1.0127	0.4971
0	0.1464	0.1464	0.4042	0.8235	0.9349	0.9349	0.8235	0.4042
0	0.1800	0.1800	0.4971	1.0127	1.1498	1.1498	1.0127	0.4971
0	0.2044	0.2044	0.5644	1.1498	1.3055	1.3055	1.1498	0.5644
0	0.0778	0.0778	0.2149	0.4378	0.4971	0.4971	0.4378	0.2149
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Columns 10 through 12

0	0	0
0.0778	0.0778	0
0.1464	0.1464	0
0.1464	0.1464	0
0.1464	0.1464	0
0.1800	0.1800	0
0.1464	0.1464	0
0.1800	0.1800	0
0.2044	0.2044	0
0.0778	0.0778	0
0	0	0
0	0	0

The picture that represents A_1 is

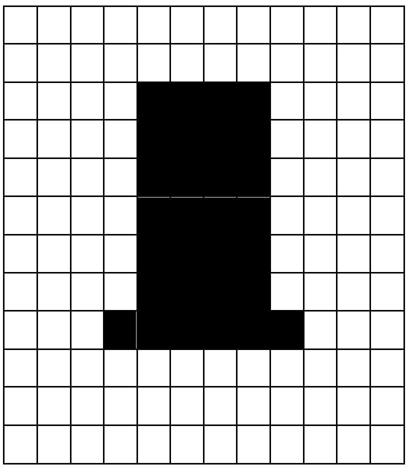

Figure (4)

If we use rank 3 approximation for *A* we have:

$A_3 = U_3 \Sigma_3 V_3^* = \sum_{i=1}^{3} s_i u_i v_i^*$ (need 3(12+12+1)=75 spaces for storage)

Columns 1 through 9

0	0	0	0	0	0	0	0	0
0	0.1043	0.1043	-0.2216	0.2624	0.8321	0.8321	0.2624	-0.2216
0	-0.0587	-0.0587	0.1247	0.8524	1.0944	1.0944	0.8524	0.1247
0	-0.0587	-0.0587	0.1247	0.8524	1.0944	1.0944	0.8524	0.1247
0	-0.0587	-0.0587	0.1247	0.8524	1.0944	1.0944	0.8524	0.1247
0	0.0790	0.0790	0.8321	1.1988	0.8728	0.8728	1.1988	0.8321
0	-0.0587	-0.0587	0.1247	0.8524	1.0944	1.0944	0.8524	0.1247
0	0.0790	0.0790	0.8321	1.1988	0.8728	0.8728	1.1988	0.8321
0	0.9494	0.9494	1.1075	0.8728	1.0814	1.0814	0.8728	1.1075
0	0.1043	0.1043	-0.2216	0.2624	0.8321	0.8321	0.2624	-0.2216
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Columns 10 through 12

0	0	0
0.1043	0.1043	0
-0.0587	-0.0587	0
-0.0587	-0.0587	0
-0.0587	-0.0587	0
0.0790	0.0790	0

-0.0587	-0.0587	0
0.0790	0.0790	0
0.9494	0.9494	0
0.1043	0.1043	0
0	0	0
0	0	0

The picture that represents A_3 is

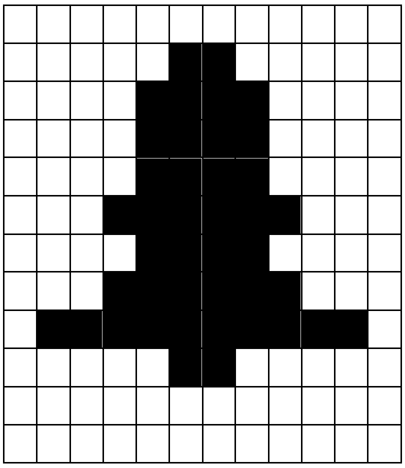

Figure (5)

we can see that bigger rank gives better approximation.

Note

A serious problem with the use of SVD for compression is that starting with a single m by n image matrix A and applying SVD results in a m×m, $n \times n$, and a diagonal $m \times n$ matrix. If $m = n$, originally m^2 values are required for picture, and after SVD we have $2m^2$+ m(values for an exact representation of picture where the rank of this picture = m. Therefore the approximation must be at most rank $m^2/(2m+1)$ in order to have compressed the image at all, otherwise the storage requirements increase.

5.3 Determination of the Effective Rank

The SVD can be used to determine both the numerical (effective) and the actual rank of a matrix. This is done by counting the number of singular values that are above a certain tolerance, t . The tolerance t =0 is used for the actual rank and some small number determined by the user according to the application at hand for the numerical rank(i.e., t >0 for numerical rank) (e.g., $\tau = \varepsilon \| A \|_2 = \varepsilon\sigma_1$, where ε is machine precision). The numerical rank of a matrix is now defined as the number of singular values $s > t$, $r(A)_\tau = \{k : \sigma_k(A) > \tau, \sigma_{k+1}(A) \le \tau\}$.

Example 5.4

Let $A = \begin{bmatrix} 1.0000 & 0.1000 & -2.1000 & 0 & -1.0000 \\ 0.9000 & -2.0000 & 1.7000 & -3.5000 & 0 \\ 1.2000 & 0 & 0 & -0.1100 & 3.0000 \\ 0 & 0.1000 & -0.2000 & 0.2500 & -0.1700 \end{bmatrix}$.

Find the effective rank where the tolerance t =0.5. By Matlab, the SVD of A is given by $A = U\Sigma V^*$, where

U=

0.2137	-0.2520	-0.9421	-0.0570
-0.9583	-0.2293	-0.1597	0.0600
-0.1740	0.9396	-0.2928	0.0330

0.0758	-0.0317	-0.0346	0.9960

$\Sigma =$

4.5791	0	0	0	0
0	3.2551	0	0	0
0	0	2.3230	0	0
0	0	0	0.0629	0

and $V^* =$

-0.1873	0.4249	-0.4571	0.7408	-0.1635
0.2056	0.1322	0.0448	0.2123	0.9451
-0.6187	0.0954	0.7378	0.2507	0.0299
0.5817	-0.4159	0.3596	0.5609	-0.2114
0.4488	0.7873	0.3397	-0.1695	-0.1858

Since σ_1, σ_2 and $\sigma_3 > \tau$, but $\sigma_4 = 0.0629 < \tau$, then the matrix we want is A_3 and we find

$U_3 =$

0.2137	-0.2520	-0.9421
-0.9583	-0.2293	-0.1597
-0.1740	0.9396	-0.2928
0.0758	-0.0317	-0.0346

$\Sigma_3 =$

4.5791	0	0

0	3.2551	0
0	0	2.3230

$V_3^* =$

-0.1873	0.4249	-0.4571	0.7408	-0.1635
0.2056	0.1322	0.0448	0.2123	0.9451
-0.6187	0.0954	0.7378	0.2507	0.0299

$$\text{and } A_3 = U_3 \Sigma_3 V_3^* = \begin{bmatrix} 1.0021 & 0.0985 & -2.0987 & 0.0020 & -1.0008 \\ 0.8978 & -1.9984 & 1.6986 & -3.5021 & 0.0008 \\ 1.1988 & 0.0009 & -0.0007 & -0.1112 & 3.0004 \\ -0.0364 & 0.1261 & -0.2225 & 0.2149 & -0.1568 \end{bmatrix}.$$

جامعة النجاح الوطنية
كلية الدراسات العليا

حول تحليل القيمة المنفردة للمصفوفات المستطيلة

إعداد
شيرين ناجح عيسى عودة

إشراف
د. "محمد عثمان" عمران

قدمت هذه الاطروحة استكمالا لمتطلبات درجة الماجستير في الرياضـيات بكليـة الدراسات العليا في جامعة النجاح الوطنية في نابلس. فلسطين.

2009

حول تحليل القيمة المنفردة للمصفوفات المستطيلة

اعداد

شيرين ناجح عيسى عودة

اشراف

الدكتور"محمد عثمان" عمران

الملخص

يعتبر تحليل القيمة المنفردة المصفوفات واحدا من أهـم مفـاهيم الرياضـيات وذلـك لارتباطه بعدد كبير من التطبيقات في الرياضيات والإحصاء والأحياء والعديد مـن المجـالات العلمية الأخرى.

نقوم في هذه الرسالة بتقديم تحليل القيمة المنفردة للمصفوفات ومقارنتها بالتحليل الطيفي، كما ونقوم بعرض لمجموعة من التطبيقات والتي تحوي النظيـر المزيـف ل(مـور وبنـروز) والرتبة الفعالة للمصفوفات و ضغط الصور.