AFRICAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
(AUST)

Moore-Penrose Pseudoinverse and Applications.

Thesis submitted in partial fulfilment for the award of Master Of Science in Pure and Applied Mathematics.

By

Areji Jonathan Gebechukwu

Supervisor:
Prof. E. H. Zerouali
Faculty of Sciences,
(Mohammed V University, Rabat, Morocco)

Co-Supervisor:
Dr. A. U. Bello
Mathematics Institute,
(African University Of Science and Technology, Abuja)

June 9, 2020

Contents

Abstract
An underlying theorem due to Gauss and Lengendre asserts that for an over determined system, there are solutions that minimize $\|Ax - b\|^2$ which is given by the generalized inverse of the matrix A even when A is singular or rectangular.

Our objective is to prove algebraic analogs of this result for arbitrary operators on complex Hilbert spaces and its generalization for the Moore-Penrose Inverse. We employ the generalized inverse matrix of Moore-Penrose to study the existence and uniqueness of the solutions for over- and under-determined linear systems, in harmony with the least squares method.

Chapter One

1. Introduction

1.1.

The concept of Moore-Penrose pseudoinverse originated from findings relating to generalized inverses which served as tool for studying astronomy, geodesy and other physical problems reduced to over determined system of linear equations modelled as

$$Ax = b, \tag{1}$$

where $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ with $m > n$. The reason why more equations than unknowns arise is that repeated measurement are taken to minimize errors.This produces over determined system which are often times inconsistent. We are interested in finding $x \in \mathbb{C}^n$ satisfying (1) and it is trivial to note that if $m = n$ and A is non-singular, then $x = A^{-1}b$ is the unique solution.

However, in wide applications with large computation as oftenly observed in areas like control theory, image processing, portfolio analysis, data management and so on, the matrix A is singular or nearly singular matrices. Thus we consider the alternative problem of solving the minimization problem

$$\min \|Ax - b\|, \tag{2}$$

whose solution eventually gives the vectors that are best approximants to the solution of (1) in terms of Euclidean norm. In general Gauss and Legendre (1810) discovered [12] that for an over determined system, there is a unique solution of minimum norm that minimizes $\|Ax - b\|^2$.

The concept of generalised inverse [1] was first introduced by Fredholm (1903), he called a particular generalized inverse as pseudoinverse which served as integral operator. Generalized inverse of differential operators already explicit in Hilbert's discussion of 1904 of the generalized green functions were consequently studied by authors like Myller (1906), Elliot(1928), Reid (1931).

E.H Moore [4] introduced the study of general reciprocal of a singular matrix. He writes the objective as thus:" The effectiveness of the reciprocal of a nonsingular finite matrix in the study of properties of such matrices makes it desirable to define if possible an analogous matrix to be associated with each finite matrix A even if A is not square or, if square, is not necessarily nonsingular". He established uniqueness, its main properties and applications to linear equations.

The striking analogies between the theories for linear equations in n–dimensional Euclidean space, for Fredholm integral equations in the space of continuous functions defined on a finite real interval, and for linear equations in Hilbert space of infinitely many dimensions, led Moore to lay down his well–known principle. Although this meet a lot of criticism because of some ambiguity. Penrose [9] redefined the Moore inverse in slightly different ways which gained more acceptance. C.R Rao [11] discussed methods of of computation for a singular matrix and applied it to solve the normal equation with singular matrix in the least square theory and also to express the variance of estimators.

The generalized inverse is a great tool in solving linear dependent and ill posed problems proposed as unbalanced system of linear equations. It has the ability to find the solution of square and non-square matrices even when they are singular. The computation of the so called minimizing vector is essential equivalent to determining the Moore-Penrose Pseudoinverse of the associated matrix which is the unique generalized inverse of A. This thesis is organized as follows:

- In Chapter One we introduce generalized inverses and revisit earliest discussions on this concept.
- Chapter Two is dedicated to basic definitions of all concept in this text, properties and auxiliary results and complete proofs.
- In Chapter Three we motivate, and analyze the Moore-Penrose Pseudoinverse. We will include the main results as regards uniqueness and existence of the Moore-penrose, a survey of Tikhonov's regularization Theorem, Single value Decomposition and spectral relationships.
- Finally chapter Four we consider some applications to least squares problems. Also Some applications to portfolio selection is investigated with examples.

Chapter Two

2. Definitions and Auxiliary Results

2.1. PRELIMINARIES

This chapter's purpose is to introduce relevant definitions and notations and furthermore to discuss some basic results that will be needed later on.

2.2. REVIEW OF HILBERT SPACES

Definition (Inner product)**.** *Let* $(\mathbb{C}, +, .)$ *be a vector space. A function on* $\mathbb{C} \times \mathbb{C}$ *is an inner product if:*

1. $\langle u, u \rangle \geq 0 \quad and \langle u, u \rangle = 0 \iff u = 0$ *for every* $u \in \mathbb{C}$.
2. $\langle u, (\alpha_1 v_1 + \alpha_2 v_2) \rangle = \alpha_1 \langle u, v_1 \rangle + \alpha_2 \langle u, v_2 \rangle$, *for every* $u, v_1, v_2 \in \mathbb{C}$ *and* $\alpha_1, \alpha_2 \in \mathbb{F}$ *(scalar field)* .
3. $\overline{\langle u, v \rangle} = \langle v, u \rangle, \ \forall \quad u, v \in \mathbb{C}$.

The pair $(\mathbb{C}, \langle ., . \rangle)$ is an inner product space. The norm associated to the inner product is a given as $\|u\| = \sqrt{\langle u, u \rangle} \quad \forall \quad u \in \mathbb{C}$. We note that as a consequence of the Cauchy-Schwartz inequality, the inner product is continuous and its associated norm satisfies the parallelogram law given as

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

With respect to the norm above, a sequence $(x_n)_{n=1}^{\infty} \subset \mathbb{C}$ is convergent to a point $x \in \mathbb{C}$ if and only if

$$\langle x_n - x, x_n - x \rangle^{\frac{1}{2}} := \|x_n - x\| \longrightarrow 0 \quad as \quad n \longrightarrow \infty.$$

Similarly, a sequence $(x_n)_{n=1}^{\infty} \in \mathbb{C}$ is Cauchy if and only if

$$\langle x_n - x_m, x_n - x_m \rangle^{\frac{1}{2}} := \|x_n - x_m\| \longrightarrow 0, \text{ as } n, m \longrightarrow \infty.$$

Consequently, an inner product space is called complete if every Cauchy sequence in $\mathbb{C}$ converges to a point of $\mathbb{C}$. A complete inner product space is called a Hilbert Space. We will denote in the sequel $\mathcal{H}$ for a Hilbert space.

Theorem 1 (Best Approximant Theorem:). *Let A be a closed and convex subset of a Hilbert space $\mathcal{H}$. Then $\forall \quad x \in \mathcal{H}$, there exists a unique $y^* \in A$ such that $\|x - y^*\| = \inf_{y \in A} \|x - y\|$.*

Orthogonal Compliments Let A be a subset of a Hilbert space. Then the orthogonal complement of A denoted by $A^{\perp}$ is the set

$$A^{\perp} = \{y \in H : \langle x, y \rangle = 0, \ \forall \quad x \in A\}.$$

It is trivial to check that $A^{\perp}$ is a closed linear subspace of $\mathcal{H}$.

Theorem 2 (Orthogonal Decompostion Theorem:). *Let A be a closed linear subspace of a Hilbert space $\mathcal{H}$. Then $\forall \quad x \in \mathcal{H}$, is written uniquely in the form $x = y + z$, where $y \in A$ and $z \in A^{\perp}$. The vector y is the best approximant of x in A.*

2.3. REVIEW OF MATRICES

Definition. *Given $A \in \mathrm{Mat}(\mathbb{C}, m, n)$, the Transpose of A denoted by $A^T \in \mathrm{Mat}(\mathbb{C}, n, m)$ is the matrix whose elements are given by*

$$(A^T)_{ij} = A_{ji}, \ \forall\, 1 \leq i \leq n, \ \forall\, 1 \leq j \leq m.$$

Definition. *Let* $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. *The adjoint of* A *is the unique matrix* $A^* \in \mathrm{Mat}(\mathbb{C}, n, m)$ *such that*

$$\langle u, Av \rangle = \langle A^* u, v \rangle$$

or equivalently, $\forall u \in \mathbb{C}^m, v \in \mathbb{C}^n$ *and* $(A^*_{ij}) = \overline{A_{ji}}$

The following properties are satisfied by the adjoint:

- $A^{**} = A$
- $(\alpha_1 A_1 + A_2)^* = \bar{\alpha_1} A_1^* + A_2^*$
- $(AB)^* = B^* A^*$

Definition. *A square matrix* $A \in \mathrm{Mat}(\mathbb{C}, n)$ *is self-Adjoint if* $A = A^*$.

All self-adjoint matrices are diagonalizable, have real eigenvalue since for any nonzero eigenvector u associated with the eigenvalue λ, we have

$$\begin{aligned} \lambda =& \lambda\langle u, u \rangle = \langle u, \lambda u \rangle = \langle u, Au \rangle = \langle Au, u \rangle \\ &= \langle \lambda u, u \rangle = \overline{\lambda}\langle u, u \rangle = \overline{\lambda}. \end{aligned}$$

Definition. *A square matrix* $A \in \mathrm{Mat}(\mathbb{C}, n)$ *is normal if*

$$AA^* = A^* A$$

.

Definition. *A square matrix* $A \in \mathrm{Mat}(\mathbb{C}, n)$ *is Unitary or Orthogonal if* $AA^* = A^*A = I$.

Definition. $A \in \mathrm{Mat}(\mathbb{C}, n)$ *is a projector if* $A^2 = A$ *and orthogonal projector if it is self-adjoint projector i.e.* $A^2 = A$ *and* $A^* = A$

Definition. *Let* $A \in \mathrm{Mat}(\mathbb{C}, n)$*, the spectrum of* A *denoted by* $\sigma(A)$ *is the set of all eigenvalues of* A*. i.e* $\lambda \in \mathbb{C}$*, such that* $(A - \lambda I)$ *is singular.*

The characteristics polynomial of A defined as $P_A = det(A - \lambda I)$.This is a polynomial of degree n. In particular $\sigma(A)$ is precisely the set of the roots of P_A. For $\lambda \in \sigma(A)$ an eigenvalue, it is trivial to see that the set of eigenvector $\cup\{0\}$ is a linear subspace. Multiplicity of roots of P_A is called algebraic multiplicity and the dimension of subspace generated by the associated eigenvectors is called geometric multiplicity of the eigenvalues of A. We note that the algebraic multiplicity is greater than or equal to the geometric multiplicity.

Proposition 2.1. *Let* $A \in \mathrm{Mat}(\mathbb{C}, n)$ *be arbitrary and let* $B \in \mathrm{Mat}(\mathbb{C}, n)$ *be invertible. Then, there exists* $M > 0$ *such that,* $A + \mu B$ *is invertible for all* $\mu \in \mathbb{C}$ *with* $0 < |\mu| < M$*.*

Definition. *Let* $A, B \in \mathrm{Mat}(\mathbb{C}, n)$*. Then* A *is similar to* B *if, there exists* $P \in \mathrm{Mat}(\mathbb{C}, n)$ *invertible, such that* $A = PBP^{-1}$*.*

Proposition 2.2. *Let* $A, B \in \mathrm{Mat}(\mathbb{C}, n)$ *be similiar. Then*

(*i*) $P_A = P_B$*;*

(*ii*) $\sigma(A) = \sigma(B)$ *as well as the geometric multiplicities of the eigenvalues;*

(*iii*) $P_{AB} = P_{BA}$*;*

(*iv*) $\sigma(AB) = \sigma(BA)$*.*

Proof. (*i*) Since A, B are similiar, Then $B = P^{-1}AP$. So $P_A = \det(A - \lambda I) = \det[P^{-1}(A - \lambda I)P] = \det(B - \lambda I) = P_B$

(*ii*) Let (λ, v) is an eigenpair of A. Then v solves $Av = PBP^{-1}v = \lambda v$. Premultiplying each side by P^{-1}, it follows that $u := P^{-1}v$ solves $Bu = \lambda u$.
Moreover, because P^{-1} is non-singular, the equation $u := P^{-1}v$ has the trivial solution $v = 0$ if $u = 0$. Hence $u \neq 0$, implies that (λ, u) is an eigenpair of B. Also by a similar argument if (λ, u) is an eigenpair of B, then (λ, Pu) is an eigenpair of A

(*iii*) If A, B or both are non-singular, Then AB and BA are similar since $AB = A(BA)A^{-1}$ and $BA = B(AB)B^{-1}$. Therefore from (i), it follows that $P_{AB} = P_{BA}$.
Suppose neither A nor B is invertible. Assume without loss of generality that A is non singular, then by Proposition 2.1, there exists $M > 0$, such that $A + \mu I$ is invertible, for $0 < |\mu| < M$. Therefore for such μ and by (ii) we have $P_{(A+\mu I)B} = P_{B(A+\mu I)}$. The corresponding coefficient of the polynomials are in μ and remains continuous. Hence equality is maintained as $\mu \to 0$. This eventually results to $P_{AB} = P_{BA}$

$\square$

Proposition 2.3. *Let* $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ *and* $B \in \mathrm{Mat}(\mathbb{C}, n, m)$. *Since* $AB \in \mathrm{Mat}(\mathbb{C}, m)$ *and* $BA \in \mathrm{Mat}(\mathbb{C}, n)$, *we have that* $x^n P_{AB} = x^m P_{BA}$. *Therefore* $\sigma(AB)\backslash\{0\} = \sigma(BA)\backslash\{0\}$.

Definition. *A matrix* $A \in \mathrm{Mat}(\mathbb{C}n)$ *is diagonalizable if it is similiar to a diagonal matrix D such that* $D = P^{-1}AP$.

A necessary and sufficient condition for $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ to be diagonalizable is that A has n-linearly independent eigenvectors. That is it has n- dimensional subspace generated by its eigenvectors.

The spectral theorem is a fundamental result of functional analysis and its versions for bounded and unbounded self - adjoint operators in Hilbert spaces play crucial roles in probability and Quantum Physics. Its simplest version for square matrices is given below:

Theorem 3. *Let $A \in \mathrm{Mat}(\mathbb{C}, n)$. Then, A is diagonalizable if and only if, there exists $r : 1 \leq r \leq n$, scalars $(\alpha_a)_{a=1}^r \in \mathbb{C}$ and non-zero distinct projectors $(E_a)_{a=1}^r \in A \in \mathrm{Mat}(\mathbb{C}, n)$ such that*

$$A = \sum_{a=1}^{r} \alpha_a E_a \tag{3}$$

and

$$I = \sum_{a=1}^{r} E_a. \tag{4}$$

with $E_i E_j = \delta_{ij} E_j$. (3) is called the spectral decomposition of the matrix A, α_a are distinct eigenvalues of A, E_a are called spectral projectors of A.

We shall subsequently see that E_a can be expressed in terms of polynomials.

Proof. $\Longrightarrow$) Suppose A is diagonalizable, then there exists $P \in \mathrm{Mat}(\mathbb{C}, n)$ such that $P^{-1}AP = D = \{\lambda_1, ..., \lambda_n\}$, where $(\lambda_i)_i^n$ are the eigenvalues of A. Let $(\alpha_i)_i^r : 1 \leq r \leq n$, be the set of distinct eigenvalues. Then

$$D = \sum_{a-1}^{r} \alpha_a K_a.$$

where $K_a \in \mathrm{Mat}(\mathbb{C}, n)$ is a diagonal matrix with 0 *or* 1 as diagonal elements so that

$$(K_a)_{ij} = \begin{cases} 1, & i = j \quad and \quad (D)_{ii} = \alpha_a \\ 0, & i = j \quad and \quad (D)_{ii} \neq \alpha_a \\ 0, & i \neq j. \end{cases}$$

We have that

$$\sum_{a=1}^{r} K_a = I \tag{5}$$

and

$$K_a K_b = \delta_{ab} K_a. \tag{6}$$

Since $A = PDP^{-1}$, we have $A = \sum_{a=1}^{r} \alpha_a E_a, E_a = PK_aP^{-1}$. Then from (3), we have that $I = \sum_{a=1}^{r} E_a$ and from (4), we have that

$$E_iE_j = PK_iP^{-1}PK_jP^1 = \delta_{ij}PK_iP^{-1} = \delta_{ij}E_i.$$

$\Longleftarrow$) Suppose that A has a representation like (3) and $E_a's$ maintains the properties mentioned, then for any vector x and $k \in \{1, ..r\}$ we have by (3) that

$$AE_K x = \sum_{j=1}^{r} \alpha_j E_j E_k x = \alpha_k E_k x.$$

Hence $E_k x$ is either zero or an eigenvalue of A. Therefore the subspace S generated by all the vectors $\{E_k x, x \in \mathbb{C}^n, k = 1, ..., r\}$ is a subspace of the space W generated by all the eigenvalues of A. Also from (4), we have that $x = \sum_{i=1}^{r} E_k x$ and so $\mathbb{C}^n = S \subset W$. Hence $W = \mathbb{C}^n$ and by since W is n-dimensional, A is diagonalizable. $\square$

Theorem 4. *Let $A \in \mathrm{Mat}(\mathbb{C}, n)$ be diagonalizable. Then for any polynomial $p, p(A) = \sum_{a=1}^{r} p(\alpha_a) E_a$.*

Proof. Since A is diagonalizable, the $A = \sum_{a=1}^{r} \alpha_a E_a$. We see that

$$A^2 = \sum_{a,b=1}^{r} \alpha_a \alpha_b E_a E_b = \sum_{a,b=1}^{r} \alpha_a \alpha_b \delta_{a,b} E_b = \sum_{a=1}^{r} (\alpha_a)^2 E_a.$$

With the convention $A^0 = I$ and by induction we have $A^m = \sum_{a=1}^{r} (\alpha_a)^m E_a, m \in \mathbb{N}$. One sees that for A non- singular and diagonalizable, $A^{-1} = \sum_{a=1}^{r} \left(\frac{1}{\alpha_a}\right) E_a$ since $\alpha_i \neq 0$, for all i. $\square$

Proposition 2.4. *Let $A \in \mathrm{Mat}(\mathbb{C}, n)$ be non- zero and diagonalizable. So we have that $A = \sum_{a=1}^{r} \alpha_a E_a$. Given the polynomial $p_j, j = 1..., r$ defined as*

$$p_j(x) = \prod_{i=1, i \neq j}^{r} \left(\frac{x - \alpha_i}{\alpha_j - \alpha_i} \right). \tag{7}$$

Then

$$E_j = p_j(A) = \prod_{k=1, k \neq j}^{r} \left(\frac{1}{\alpha_j - \alpha_k} \right) \prod_{i=1, i \neq j}^{r} (A - \alpha_i I), \tag{8}$$

for all $j = 1, \cdots, r$.

Proof. By definition of p_j, we have that $p_j(\alpha_k) = \delta_{j,k}$. Therefore by theorem 4, we have that $p_j(A) = \sum_{k=1}^{r} p_j(\alpha_k) E_k = \sum_{k=1}^{r} \delta_{j,k} E_k = E_j$ $\square$

Proposition 2.5. *The Spetral decomposition of a diagonalizable matrix $A \in \mathrm{Mat}(\mathbb{C}, n)$ is unique.[8].*

Proof. Since A is diagonalizable, $A = \sum_{k=1}^{r} \alpha_k E_k$. Suppose $A = \sum_{k=1}^{s} \beta_k F_k$ be another spectral decomposition of A with

β_k's distinct and $F_k's$ non-vanishing and satsifying the properties of the projectors. Then for a vector $x \neq 0$, we have that $x = \sum_{k=1}^{r} \beta_k F_k x$. This implies that, there exists some non-vanishing vector say $F_{k_0}x$ such that $F_{k_0}x \neq 0$. Then applying A, we have that $AF_{k_0}x = \sum_{k=1}^{r} \beta_k F_k F_{k_0} x = \beta_{k_0} F_{k_0} x$. This shows that β_{k_0} is an eigenvalue of A and $\{\beta_1, ..., \beta_s\} \subset \{\alpha_1, ..., \alpha_r\}$ and so $s \leq r$. Let us order both sets such that $\beta_k = \alpha_k$, $\forall\ 1 \leq k \leq s$. Then

$$A = \sum_{k=1}^{r} \alpha_k E_k = \sum_{k=1}^{s} \alpha_k F_k. \tag{9}$$

By considering the polynomial, $p_j, p_j(\alpha_k) = \delta_{j,k}$, we have from (9) that for all $1 \leq j \leq s$, $p_j(A) = \sum_{k=1}^{r} p(\alpha_k) E_k = \sum_{k=1}^{s} p(\alpha_k) F_k$. Consequently, $E_j = F_j$.

The equality follows because E_j and F_j satisfy the same algebraic relations. Since $I = \sum_{a=1}^{r} E_k = \sum_{a=1}^{s} E_k$ and $E_j = F_j$, for all $1 \leq j \leq s$, so one has that $\sum_{a=s+1}^{r} E_k = 0$

Hence, multiplying by E_l with $s+1 \leq l \leq r$, it follows that $E_l = 0, s+1 \leq l \leq r$. This is only if $r = s$. Hence $E_k's$ are non-vanishing. This completes the proof. $\square$

Chapter Three

3. Generalized inverse and the Moore-Penrose Pseudoinverse

3.1. Introduction

Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. we are interested in finding a solution to the problem $Ax = b$, $x \in \mathbb{C}^n, b \in \mathbb{C}^m$.

If A is square and invertible, Then $x = A^{-1}b$. If otherwise, then either b is not in the range of A in which case we have no solution or we have infinitely many solutions. A solution of a linear system of equations in this case is found in general from the notion of generalized inverse of a matrix. Generalized inverse is of great importance in solving linearly dependent and unbalanced equation with lots of applications to square, non- square, and singular matrices.

Definition. *Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. Then $B \in \mathrm{Mat}(\mathbb{C}, n, m)$ is a generalized inverse of A if*

i $ABA = A$
ii $BAB = B$

If $A \in \mathrm{Mat}(\mathbb{C}, n)$ [1] non-singular, then $B = A^{-1}$ satisfies trivially the definition. If $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ is $1-1$ (has linearly independent columns i.e rank $n \leq m$) then $(A^*A)^{-1}$ exists and $A_l^{-1} = (A^*A)^{-1}A^*$, we find that $A_l^{-1}A = I$. A_l^{-1} is called the left inverse of A. similarly, if A is surjective (has linearly independent rows, $m \leq n$), then $(AA^*)^{-1}$ exists and $A_r^{-1} = A^*(AA^*)^{-1}$, we find that $AA_l^{-1} = I$. A_r^{-1} is called the right inverse of A. It is obvious to see that whenever such inverse exists, a solution of the equation $Ax = b$ exists. A very interesting question that arises is: Can we present a solution $x = Bb$ of the consistent equation $Ax = b$ in the absence of these inverse? if yes, then we call B a generalized inverse. We note there that every matrix $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ has at least one generalized inverse B. Thus in general, from [11] B is not unique and is only unique if more conditions are imposed on it.

Theorem 5. *Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ and B a generalized inverse of A. Then, for fixed $b \in \mathbb{C}^m$,*

1. *$Ax = b$ has a solution if and only if $ABb = b$.*
2. *x is a solution of $Ax = b$ if and only if $x = Bb + (I - BA)z, z \in \mathbb{C}^n$. A particular solution of $ax = b$, for $b \in R(A)$ is obtained from $x = Bb$.*

3.2. Algorithm for the Generalized inverse of solution of Ax=b

Given $Ax = b$ the following outlines a procedure to find a generalized inverse of A.

1. Choose an invertible submatrix G of dimension r.
2. Find $(G^{-1})^T$.
3. Replace the element of the submatrix G in the original matrix A by the elements of $(G^{-1})^T$.
4. Replace all other elements by zero to get a new matrix A^1.
5. then $B = (A^1)^T$ is a generalised inverse.
6. Use $x = Bb + (I - BA)z$ to calculate a solution for $Ax = b$

Example. *Consider* $A = \begin{pmatrix} 1 & 3 & 2 \\ 5 & 2 & 6 \\ 2 & 6 & 4 \end{pmatrix}$ *and let* $b = \begin{pmatrix} 1 & 5 & 2 \end{pmatrix}^T$.

If we take two sub-matrices $G_1 and G_2$ *as follows: Let* $G_1 = \begin{pmatrix} 1 & 3 \\ 5 & 2 \end{pmatrix}$ *and* $G_2 = \begin{pmatrix} 2 & 6 \\ 6 & 4 \end{pmatrix}$. *it follows that* $(G_1^{-1})^T = \frac{1}{13} \begin{pmatrix} -2 & 5 \\ 3 & -1 \end{pmatrix}$ *and* $(G_2^{-1})^T = \frac{1}{14} \begin{pmatrix} -2 & 3 \\ 3 & -1 \end{pmatrix}$.*The corresponding generalized inverses are* $B_1 = \frac{1}{13} \begin{pmatrix} -2 & 3 & 0 \\ 5 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *and* $B_2 = \frac{1}{14} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 3 \\ 0 & 3 & 1 \end{pmatrix}$

The example above validates that the generalized inverse of a matrix is not necessarily unique. It depends on the number of non-singular sub-matrices obtained. consequently it is obvious to state that the solution is also not unique.

3.3. Moore-Penrose Pseudoinverse

Definition. *Given* $A \in \mathrm{Mat}(\mathbb{C}, m, n)$, *then* $A^\dagger \in \mathrm{Mat}(\mathbb{C}, n, m)$ *is called a Moore-Penrose inverse of* A *if and only if*

I $AA^\dagger A = A$

II $A^\dagger A A^\dagger = A^\dagger$.

III $(A^\dagger A)^* = A^\dagger A$ *(i.e.* $A^\dagger A$ *is self-adjoint)*

IV $(AA^\dagger)^* = AA^\dagger$ *(i.e.*$AA^\dagger$ *self-adjoint)*

Remark. A striking difference between the generalized inverse and the Moore-Penrose inverse is that the Penrose inverse is always unique.

It is trivial to see that if $A \in \text{Mat}(\mathbb{C}, n)$ is invertible then $A^\dagger = A^{-1}$ satisfies all defining properties of Moore-Pernrose inverse. It is also evident from the definition that for $A \in \text{Mat}(\mathbb{C}, m, n)$ such that $A_{ij} = 0_{ij}$, the $A^\dagger = 0_{ji}$. $A \in \text{Mat}(\mathbb{C}, 1, 1)$, i.e $A = (z) \in \mathbb{C}$ Then

$$(z)^\dagger = \begin{cases} 0, & z = 0 \\ z^{-1}, & z \neq 0 \end{cases}$$

$A \in \text{Mat}(\mathbb{C}, m, 1)$, a non-zero column vector. Then $A^\dagger = \dfrac{1}{\|A\|^2} A^*$. In general, if $A \in \text{Mat}(\mathbb{C}, m, n)$ and is left or right invertible, the $A^\dagger = (A^*A)^{-1}A^*$ or $A^\dagger = A^*(AA^*)^{-1}$.

Example. *Let* $A = \begin{pmatrix} 2 & 0 & i \\ 0 & i & 1 \end{pmatrix}$*. It is clear that* $A^* = \begin{pmatrix} 2 & 0 \\ 0 & -i \\ -i & 1 \end{pmatrix}$ *and one sees that AA^* is invertible since* $\det AA^* = 9 \neq 0$*. Therefore* $A^\dagger = \dfrac{1}{9}\begin{pmatrix} 4 & -2i \\ 1 & -5i \\ -i & 4 \end{pmatrix}$ *is the Moore-Penrose Pseudoinverse of A. This is indeed true since [I, II, III, IV] can be easily verified.*

3.4. Properties of Moore-Penrose Pseudoinverse

The following are obvious properties of the Moore-Penrose Pseudoinverse:

- $A^{\dagger\dagger} = A$
- $(A^\dagger)^* = (A^*)^\dagger$
- $\overline{(A^\dagger)} = \overline{(A)}^\dagger$

- $(A^\dagger)^* = (A^*)^\dagger$
- $(zA)^\dagger = z^{-1}A^\dagger$

At this juncture, we clearly state that unlike the usual inverse, the Moore-Penrose pseudoinverse is not commutative. Also $AA^\dagger$ and $A^\dagger A$ is not necessarily the identity. In the same vain the reverse law, $(AB)^{-1} = B^{-1}A^{-1}$, for $A \in \text{Mat}(\mathbb{C}, m, n)$ and $B \in \text{Mat}(\mathbb{C}, n, p)$ for $m = n = p$, which is valid for the usual matrix is not necessarily true for the Moore-penrose Pseudoinverse.

Proposition 3.1. *Let* $A \in \text{Mat}(\mathbb{C}, m, n)$ *then*

I a) $A = AA^*(A^\dagger)^* = (A^\dagger)^*A^*A$

b) $A^* = A^*A(A^\dagger) = (A^\dagger)AA^*$

c) $A^\dagger = A^\dagger(A^\dagger)^*A^* = A^*(A^\dagger)^*A^\dagger$

II $(AA^*)^\dagger = (A^*)^\dagger A^\dagger$

Proof. I. Recall that $A^\dagger = A^\dagger AA^\dagger = A^\dagger(AA^\dagger)^* = A^\dagger(A^\dagger)^*A^*$. Similarly $A^\dagger = A^\dagger AA^\dagger = (A^\dagger A)^*A^\dagger = A^*(A^\dagger)^*A^\dagger$ and so equality is established as obtained in $Ia)$. The same argument is valid in proving for A since $A = AA^\dagger A$ thus we have $Ib)$. Replacing A by A^* we have $Ic)$.

II. Let $C = (A^*)^\dagger A^\dagger$. We show that C satisfies the definition for the the Moore-penrose inverse of AA^*. We have

i. for the first one

$$\begin{aligned} AA^* &= A(A^*) = AA^*(A^\dagger)^*A^* \\ &= AA^*(A^\dagger)^*A^\dagger AA^* = AA^*CAA^*. \end{aligned}$$

ii. Also,

$$\begin{aligned} C &= (A^*)^\dagger A^\dagger = (A^*)^\dagger A^\dagger(A^\dagger)^*A^* \\ &= (A^\dagger)^*A^\dagger AA^*(A^\dagger)^*A^\dagger = C(AA^*)C. \end{aligned}$$

iii. $AA^*C = (AA^*)(A^*)^\dagger A^\dagger = (AA^*(A^*)^\dagger)A^\dagger = AA^\dagger$.

iv. Lastly,

$$\begin{aligned} C(AA^*) &= (A^*)^\dagger A^\dagger(AA^*) = (A^*)^\dagger(A^\dagger AA^*) \\ &= (A^*)^\dagger A^\dagger. \end{aligned}$$

Analogously Replacing A by A^*, we have $(A^*A)^\dagger = (A)^\dagger(A^*)^\dagger$

□

3.5. Uniqueness

Theorem 6 (Uniqueness Theorem)**.** *Let* $A \in \mathrm{Mat}(\mathbb{C}, m, n)$*. Then the Moore-Penrose Pseudoinverse,* $A^\dagger \in \mathrm{Mat}(\mathbb{C}, n, m)$ *is unique.[2]*

Proof. Assuming existence, we show the uniqueness of the Moore-penrose pseudoinverse. Let $A^\dagger \in \mathrm{Mat}(\mathbb{C}, n, m)$ be the Moore-Penrose Pseudoinverse. Suppose that there exists $B \in \mathrm{Mat}(\mathbb{C}, n, m)$, another Moore-Penrose Pseudoinverse. We show that $A^\dagger = B$.

Define $M_1 = AB - AA^\dagger = A(B - A^\dagger)$. Clearly M_1 is self-adjoint. It follows that $M_1^2 = (AB - AA^\dagger)A(B - A^\dagger) = (ABA - AA^\dagger A)(B - A^\dagger) = (A - A)(B - A^\dagger) = 0$. This implies that $M_1 = 0$, since for $x \in \mathbb{C}^n$, $\|M_1 x\|^2 = \langle M_1 x, M_1 x\rangle = \langle x, (M_1)^2 x\rangle = 0$. Hence $AB = AA^\dagger$.

Similarly, let $M_2 = BA - A^\dagger A$. Then $BA = A^\dagger A$.
Therefore $A^\dagger = A^\dagger(AA^\dagger) = (A^\dagger A)B = BAB = B$.

□

3.6. Existence

We present two existence theorems for the Moore-penrose Pseudoinverse. Our discussion will be based on the Single Value Decompositon(SVD) and the next is a spectral decomposition approach.

3.7. Single Value Decomposition.

The singular value decomposition (SVD) is a powerful technique in many matrix computations and analysis. Using the SVD of a matrix in computations rather than the original matrix has the advantage of being more robust to numerical error. Additionally the SVD exposes the geometric structure of a matrix, an important aspect of many matrix calculations. A matrix can be described as a transformation from one vector space to another. The components of the SVD quantify the resulting change between the underlying geometry of those vector spaces.

A singular value decomposition of an $m \times n$ matrix A is any factorization of the form $A = UDV^*$ where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix and D is an $m \times n$ diagonal matrix with $s_{ij} = 0$, if $i \neq j$ and $s_{ii} = s_i, s_i > 0$, [5] The columns of U are orthonormal eigenvectors of AA^* , the columns of V are orthonormal eigenvectors of A^*A, and D is a diagonal matrix containing square roots of the eigenvalues of the square matrices A^*A or AA^* in descending order, which are the same values, and the number of singular values is equal to the rank of A. . The quantities s_i are called the singular values of A and the columns of U and V are called the left and right singular vectors respectively.

The SVD can be rewritten as a summation,

$$A = \sum_{i=1}^{r} s_i x_i y_i^*.$$

Where the columns of U are $\{x_1, ...x_m\}$ and V are $\{y_1, ...y_n\}$ and orthonormal.

Theorem 7. *Let $A \in Mat(\mathbb{C}, m, n)$ and has rank r and $A = UDV^*$. Define $X = VD^\dagger U^*$ such that $D^\dagger = diag\{\frac{1}{s_i}, ..., \frac{1}{s_r}, 0, ..., 0\}$. Then $X = A^\dagger$.*

Proof. We observe that from [12] X can also be written as

$$X = \sum_{i=1}^{r} s_i^{-1} y_i x_i^*.$$

Our job is to prove that $X = A^\dagger$ satisfies the defining properties of the Moore-Penrose Pseudoinverse.

i

$$\begin{aligned} AXA &= \sum_{a=1}^{r} s_a x_a y_i^* \sum_{b=1}^{r} s_b^{-1} x_b^* y_b \sum_{c=1}^{r} s_c x_c y_c^* \\ &= \sum_{a=1}^{r} \sum_{b=1}^{r} \sum_{c=1}^{r} \frac{s_a s_c}{s_b} x_a (y_a . y_b)(x_b . x_c) y_c^* \\ &= \sum_{i=1}^{r} s_i x_i y_i^* = A \end{aligned}$$

ii

$$XAX = \sum_{a=1}^{r} s_a^{-1} x_a^* y_a \sum_{b=1}^{r} s_b x_b y_b^* \sum_{a=1}^{r} s_c^{-1} x_c^* y_c$$
$$= \sum_{a=1}^{r} \sum_{b=1}^{r} \sum_{c=1}^{r} \frac{s_b}{s_a s_c} x_a^* (y_a . y_b^*)(x_b . x_c^*) y_c$$
$$= \sum_{i=1}^{r} s_i x_i^* y_i = X$$

iii

$$AX = \sum_{a=1}^{r} s_a x_a y_i^* \sum_{b=1}^{r} s_b^{-1} x_b^* y_b$$
$$= \sum_{a=1}^{r} \sum_{b=1}^{r} \frac{s_a}{s_b} x_a (y_a^* . y_b) x_b^*$$
$$= \sum_{i=1}^{r} x_i x_i^* = I$$

□

XA is also self-adjoint. Hence having satisfied all the conditions, we conclude that $X = A^\dagger$.

Example. *Let* $A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$. *Find the Moore-penrose inverse using SVD.*

Solution.

We are interested in finding U,V orthogonal matrices such that $A = UDV^*$ *after which the may find* $A^\dagger$.

For U, We consider the matrix $AA^* = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}$. *We have the eigenvalues are* $\lambda = 12$, *and* $\lambda = 10$ *with eigenvectors* $v_1 = [1, 1]$ *and* $v_2 = [1, -1]$ *respectively. Arranging in desending order* *of the singular values we have the matrix* $P_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. *So by Gram-schmidt Orthonormalization process to convert to orthonormal matrix we that*

$$u_1 = \frac{v_1}{\|v_1\|} \tag{10}$$

$$w_n = v_n - \sum_{k=1}^{n-1} \langle v_n, u_k \rangle u_k,\ k \geq 2 \tag{11}$$

$$u_n = \frac{w_n}{\|w_n\|} \tag{12}$$

So we have $u_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$ *and* $u_2 = \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]$.
Therefore

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

Now for V, consider $A^*A = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 2 & 4 \\ 2 & 4 & 2 \end{pmatrix}$ *With eigenvalues* $\lambda = 12$, $\lambda = 10$ *and* $\lambda = 0$ *with eigenvectors* $v_1 = [1, 2, 1]$ *and* $v_2 = [2, -1, 0]$ *and* $v_3 = [1, 2, -5]$ *respectively. Arranging in desending order of the singular values we have the matrix*

$$P_2 = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{pmatrix}.$$

Also by Gram-schmidt Orthonormalization process, we that $u_1 = \left[\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right]$, $u_2 = \left[\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0\right]$ *and* $u_3 = \left[\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}}\right]$.

Therefore $V = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix}$

So by SVD, we have that

$$A = UDV^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix}^* .$$

Consequently, we have that

$$A^\dagger = VD^\dagger U^* = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{12}} & 0 \\ 0 & \frac{1}{\sqrt{10}} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} =$$

$$\begin{pmatrix} \frac{17}{60} & \frac{-7}{60} \\ \frac{1}{15} & \frac{4}{15} \\ \frac{1}{12} & \frac{1}{12} \end{pmatrix}.$$

It is easily verifiable that $A^\dagger$ *satisfies all defining properties of the Moore-Penrose inverse.*

3.8. Tikhonov's Regularization Process

We have earlier seen that if $A^{-1}, A_r^{-1}, A_l^{-1}$ exists, then $A^\dagger = A^{-1}, A^\dagger = A_r^{-1}, and \quad A^\dagger = A_l^{-1}$ respectively.
Suppose $(AA^*)^{-1}$ and $(A^*A)^{-1}$ do not exist, Then an alter-

native way to obtain $A^\dagger$ is the provisional replcament of the singular matrices AA^* and A^*A by the non-singular ones $(AA^* + \mu I)$ and $(A^*A + \mu I)$, (Regularization procedure) with $\mu \neq 0$ and $|\mu|$ small enough [2]. This regularization process is guranteed by proposition (2.1).
Thus given the problem $Ax = b$ and applying A^* to get $A^*Ax = A^*b$. We Consider the regularization equation and rather seek a solution $\check{x}$ such that $(A^*A + \mu I)\check{x} = A^*b$. we ask if the lim as $\mu \to 0$ exists. Consequently, the following ensues:

Lemma 3.1. *Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ be arbitrary and $\mu \in \mathbb{C}$ such that $AA^* + \mu I_m$ and $A^*A + \mu I_n$ are non-singular. Then $A^*(AA^* + \mu I_m)^{-1} = (A^*A + \mu I_n)^{-1}A^*$*

Proof. Let $B := A^*(AA^* + \mu I_m)^{-1}$ and $C := (A^*A + \mu I_n)^{-1}A^*$. Then applying A^*A to B, we have

$$\begin{aligned}
A^*AB &= A^*AA^*(AA^* + \mu I_m)^{-1} \\
&= A^*[AA^*](AA^* + \mu I_m)^{-1} \\
&= A^*[AA^* + \mu I_m - \mu I_m](AA^* + \mu I_m)^{-1} \\
&= A^*[I_m - \mu(AA^* + \mu I_m)^{-1}] \\
&= A^* - \mu B.
\end{aligned}$$

So we have $B(A^*A + \mu I_n) = A^*$. This implies that $B = (A^*A + \mu I_n)^{-1}A^* = C$ □

Lemma 3.2. *Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$. Then*

$$\lim_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} = \lim_{\mu \to 0}(A^*A + \mu I_n)^{-1}A^*$$

Proof. We note that A is not identically zero matrix else A^*A and AA^* is zero. Suppose A^*A and AA^* are non-zero Matrices.

Since AA^* is self-adjoint , then it is diagonalizable. Let $\alpha_1, ..., \alpha_r$ be distinct eigenvalues of AA^*, Then $AA^* = \sum_{k=1}^{r} \alpha_k E_k$, and satisfying $I = \sum_{a=1}^{r} E_a$, $E_i E_j = \delta_{ij} E_i$ and $E_k^* = E_k$.
So we have $(AA^* + \mu I_m) = \sum_{k=1}^{r} (\alpha_k + \mu) E_k$ Since $\mu \notin \{\alpha_1, ..., \alpha_r\}$, by theorem (4) we have
$(AA^* + \mu I_m)^{-1} = \sum_{k=1}^{r} \dfrac{1}{(\alpha_k + \mu)} (\alpha_k + \mu) E_k$ and so applying A^* from the left we have
$A^*(AA^* + \mu I_m)^{-1} = \sum_{k=1}^{r} \dfrac{1}{(\alpha_k + \mu)} A^* E_k$ now consider two cases:

Case I: $0 \notin \sigma(AA^*)$.
Then $\lim\limits_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} = \sum_{k=1}^{r} \dfrac{1}{\alpha_k} A^* E_k$.
Case II: $0 \in \sigma(AA^*)$.
Suppose say $\alpha_1 = 0$. then E_1 projects on to the kernel of AA^* and so for $x \in Ker(AA^*)$, $A^* x = 0$ so that $A^* E_1 = 0$ Therefore $A^*(AA^* + \mu I_m)^{-1} = \sum_{k=2}^{r} \dfrac{1}{(\alpha_k + \mu)} A^* E_k$ and so $\lim\limits_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} = \sum_{k=2}^{r} \dfrac{1}{\alpha_k} A^* E_k$. This provides the existence of the limit and by Lemma (3.1), we have that

$$\lim_{\mu \to 0} (A^* A + \mu I_n)^{-1} A^*, \tag{13}$$

exists and are equal. □

The main consequence of the previous discussion is the next theorem which contains a general proof for the existence of the Moore-penrose Pseudoinverse.

Theorem 8 (Tikhonov's Regularization)**.** *Let* $A \in \mathrm{Mat}(\mathbb{C}, m, n)$*.*

$$A^{\dagger} = \lim_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} \tag{14}$$

$$A^{\dagger} = \lim_{\mu \to 0} (A^*A + \mu I_n)^{-1} A^* \tag{15}$$

Proof. Sequel to the lemmas above, we consider these two cases again
Case I: $0 \notin \sigma(AA^*)$.
Then $\lim\limits_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} = \sum_{k=1}^{r} \dfrac{1}{\alpha_k} A^* E_k := B$.
We show that B satisfies the defining conditions of Moore-Penrose Pseudoinverse of A. Now

$$\begin{aligned} AB &= \sum_{k=1}^{r} \frac{1}{\alpha_k} AA^* E_k \\ &= \sum_{k=1}^{r} \frac{1}{\alpha_k} \left(\sum_{i=1}^{r} \alpha_i E_i \right) E_k \\ &= \sum_{k=1}^{r} \sum_{i=1}^{r} \frac{1}{\alpha_k} \alpha_i \delta_{ki} E_k \\ &= \sum_{k=1}^{r} E_k = I \end{aligned}$$

since I is self-adjoint, then AB is self-adjoint.

$$BA = \sum_{k=1}^{r} \frac{1}{\alpha_k} A^* E_k A. \tag{16}$$

We observe that $(A^* E_k A)^* = A^* E_k A$ since $\forall \; k, E_k^* = E_k$.

From $AB = I$, we have that $ABA = A$

$$BAB = \left(\sum_{k=1}^{r} \frac{1}{\alpha_k} A^* E_k A\right) \left(\sum_{i=1}^{r} \frac{1}{\alpha_i} A^* E_i\right)$$
$$= \sum_{k=1}^{r} \sum_{i=1}^{r} \frac{1}{\alpha_k \alpha_i} A^* E_k A A^* E_i.$$

But recall $AA^* E_i = \alpha_i E_i$, so that

$$BAB = \sum_{k=1}^{r} \sum_{i=1}^{r} \frac{1}{\alpha_k} A^* E_k E_i$$
$$= \left(\sum_{k=1}^{r} \frac{1}{\alpha_k} A^* E_k\right) \left(\sum_{i=1}^{r} E_i\right)$$
$$= B$$

CASE II: $0 \in \sigma(AA^*)$.
Then $\lim_{\mu \to 0} A^*(AA^* + \mu I_m)^{-1} = \sum_{k=2}^{r} \frac{1}{\alpha_k} A^* E_k := B$.

$$AB = \sum_{k=2}^{r} \frac{1}{\alpha_k} AA^* E_k = \sum_{k=2}^{r} \frac{1}{\alpha_k} \alpha_k E_k = I - E_{1.} \tag{17}$$

which is self-adjoint.

$$BA = \sum_{k=2}^{r} \frac{1}{\alpha_k} AA^* E_k A \tag{18}$$

which is also self-adjoint.
From (17), we have that $ABA = A - AE_1$.
But $(AE_1)^* = A^* E_1 = 0$. Therefore $AE_1 = 0$ and hence

$ABA = A$.

$$\begin{aligned}
BAB &= \left(\sum_{k=2}^{r} \frac{1}{\alpha_k} A^* E_k A\right) \left(\sum_{i=2}^{r} \frac{1}{\alpha_i} A^* E_i\right) \\
&= \sum_{k=2}^{r} \sum_{i=2}^{r} \frac{1}{\alpha_k \alpha_i} A^* E_k A A^* E_i \\
&= \sum_{k=2}^{r} \sum_{i=2}^{r} \frac{1}{\alpha_k} A^* E_k E_i \\
&= \left(\sum_{k=2}^{r} \frac{1}{\alpha_k} A^* E_k\right) (I_m - E_1) \\
&= B - \left(\sum_{k=2}^{r} \frac{1}{\alpha_k} A^* E_k E_1\right) \\
&= B,
\end{aligned}$$

since for every $k \neq 1$ we have $E_k E_1 = 0$. $\square$

Theorem 9. *Let $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ be non-zero matrix and let $AA^* = \sum_{k=1}^{r} \alpha_k E_k$ be the spectral representation of AA^* where $\{\alpha_1, ..., \alpha_r\} \subset \mathbb{R}$ is the set of distinct eigenvalues of AA^* and E_k are the corresponding self-adjoint spectral projections. Then*

$$A^\dagger = \sum_{a=1, \alpha_a \neq 0}^{r} \frac{1}{\alpha_a} A^* E_a \tag{19}$$

Analogously if $A^* A = \sum_{b=1}^{s} \beta_b F_b$ and $\{\beta_1, ..., \beta_s\} \subset \mathbb{R}$ is the set of distinct eigenvalues of $A^* A$ and F_b the corresponding self-adjoint spectral projections. Then

$$A^\dagger = \sum_{b=1, \beta_b \neq 0}^{s} \frac{1}{\alpha_a} F_a A^* \tag{20}$$

consequently for A non zero, we have

$$A^{\dagger} = \sum_{a=1,\alpha_a\neq 0}^{r} \frac{1}{\alpha_a} \left(\prod_{k=1,k\neq a}^{r} (\alpha_a - \alpha_k)^{-1} \right) A^{*} \left[\prod_{k=1,k\neq a}^{r} (AA^{*} - \alpha_k I_m) \right] \tag{21}$$

$$A^{\dagger} = \sum_{b=1,\beta_b\neq 0}^{s} \frac{1}{\beta_b} \left(\prod_{l=1,l\neq b}^{s} (\beta_b - \beta_l)^{-1} \right) \left[\prod_{l=1,l\neq b}^{s} (A^{*}A - \beta_l I_n) \right] A^{*} \tag{22}$$

Expressions (21) or (22) provide a general algorithm for the computation of the Moore-Penrose pseudoinverse for any non-zero matrix A. Its implementation requires only the determination of the eigenvalues of AA^* or of A^*A and the computation of polynomials on AA^* or A^*A.

Chapter Four

4. Applications of the Moore-Penrose Pseudoinverse

4.1. Application to least squares problem

In general, there is no solution to overdetermined systems. The Moore-Penrose Pseudoinverse solves the problem in least square sense. This is rather interesting because it finds the best approximant to the possible solution of the equation $Ax = b$.
The least square solution of a sysem is a vector $x^{\dagger}$ such that

$$\|Ax^{\dagger} - b\| \leq \|Ax - b\| \forall x \in \mathbb{C}^n$$

Theorem 10. *Every linear system $Ax = b$ for $A \in \mathrm{Mat}(\mathbb{C}, m, n)$ has a unique least square solution of smallest norm.*

Proof. Let b be a point $\in \mathbb{C}^m$ and $U \subseteq \mathbb{C}^m$ be the image of subspace of A. We claim that x minimizes $\|Ax - b\|^2$ if and only Ax is the orthogonal projection p of b onto U i.e $pb = b - Ax$ being orthogonal to U. Since $U^{\perp}$ is orthogonal to U, the space $b + U^{\perp}$ is a unique point p. Thus $\forall y \in U$, py and bp are orthogonal. So that

$$\|by\|^2 = \|bp\|^2 + \|py\|^2$$

Thus p is the unique point in U that minimizes the distance from b to any point in U. To show uniqueness of x for which $\|Ax - b\|^2$ is minimized, we use the fact that $\mathbb{C}^n = Ker A \bigoplus (Ker A)^{\perp}$. Then $x = u + v, u \in Ker A, v \in (Ker A)^{\perp}$. So for $u \in Ker A$, $Au = 0$, and $Ax = p$ if and only $Av = p$ which show that the solution of $Ax = p$ for which x has a minimum norm is in

$(ker A)^{\perp}$. Since $A|_{(Ker A)^{\perp}}$ is injective, there exist a unique x of minimum norm minimizing $\|Ax - b\|^2$. $\square$

Theorem 11. *The least squares solution of smallest norm of the linear system $Ax = b$ is given by $x^{\dagger} = A^{\dagger}b = VD^{\dagger}U^{*}b$*

Proof. Assume A is (rectangular) diagonal matrix D. Since x minimizes $\|Dx - b\|^{*}2$ if and only if Dx is the projection of b onto the image subspace F of D. Then $x^{\dagger} = D^{\dagger}b$. But by single value decomposition, $A = UDV^{*}$, U, V being orthogonal. Since U is isometry,

$$\|Ax - b\| = \|UDV^{*}x - b\| = \|DV^{*}x - U^{*}b\|.$$

if $y = V^{*}x$, then $\|x\| = \|y\|$ since V is isometry and surjective. Therefore $\|Ax - b\|$ is minimized if and only if $\|Dy - U^{*}b\|$ is minimized and we showed that the least square solution is $y^{\dagger} = D^{\dagger}U^{*}b$ and so $x^{\dagger} = VD^{\dagger}U^{*}b = A^{\dagger}b$ which is the optimal solution to the least sqaure problem. $\square$

This theorem allows us to affirm that $A^{\dagger}b$ is either the unique least square solution or is the least square solution of minimum norm.

Example. *Given a problem modelled as*

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 4 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \\ 4 \\ 3 \end{pmatrix}.$$

As earlier stated, the solution is given by the Moore-penrose . we compute as follows:

Since $(A^*A)^{-1}$ *exist, we have*

$$A^\dagger = (A^*A)^{-1}A^* = \frac{1}{120}\begin{pmatrix} -2 & -6 & -6 & -2 \\ 10 & 20 & -10 & 0 \end{pmatrix}.$$

Also, $x_0 = A^\dagger b = \frac{1}{120}\begin{pmatrix} -48 \\ 40 \end{pmatrix}$.

4.2. Application to Finance (Portfolio Selection)

Maximimizing profit with least risk is the objective of every investor thus the construction of portfolios. Introduced by Harry Markowitz in 1952, the portfolio selection problem is really the process of delineating the most efficient portfolios and then selecting the best portfolio from the set. This covariance matrix is used to calculate the portfolio weights. When the number of stocks N is of the same order of magnitude as the number of returns per stock T, the total number of parameters to estimate is of the same order as the total size of the data set. When N is larger than T the covariance matrix is always singular. In this case the problem is that we need the inverse of the covariance matrix and it does not exist. To get around to this problem we can use the generalized inverse or Moore-Penrose inverse. Especially if we replace the inverse of the sample covariance matrix by the pseudo-inverse we can define the portfolio weights w_i.
The problem of portfolio selection is, as defined by Markowitz in [7]

$$min \quad w^{'}\Sigma w \tag{23}$$

subject to $w'1 = 1$ and $w'\mu = q$, where 1 denotes a conformable vector of ones and q is the expected rate of return that is required on the portfolio. Negative elements of w denote short positions. The well-known solution is:

$$w = \frac{1}{AC - B^2}\left[(C - qB)1 + (qA - B)\mu\right] \tag{24}$$

where $A = 1'\Sigma^{-1}1, B = 1'\Sigma^{-1}\mu, C = \mu'\Sigma^{-1}\mu$.
This equation shows that optimal portfolio weights depend on the inverse of the covariance matrix. This sometimes causes difficulty if the covariance matrix estimator is not invertible, close to singular or numerically ill-conditioned, which means that inverting it amplifies estimation error tremendously. One possible trick to get around this problem is to use the Moore-Penrose inverse. Replacing the inverse of the sample covariance matrix by the pseudo-inverse into equation (24) yields well-defined portfolio weights,
We shall make use of this property of the Moore-Penrose Inverse to minimize the risk in portfolio selection. Since the covariance matrix is self-adjoint ($\Sigma = \Sigma^*$), it is well known that $\Sigma^\dagger = \Sigma^{\dagger *}$. An interesting property of self adjoint matrices, is that their Moore-Penrose inverse coincides with two other types of generalized inverses, the Drazin inverse and the Group inverse.
When the matrix Σ is singular, [8] then we propose $\Sigma^\dagger$ as a candidate for the minimizer, in order to achieve the optimal portfolio positions. The proof of Markowitz's problem is performed using the standard Langrage method. The conditions for the Lagrangian give the equation:

$$\Sigma w = \frac{-1}{2}(\lambda_1\mu + \lambda_2 1) \tag{25}$$

When the covariance matrix Σ is singular, then from the fact that Moore-penrose gives the best approximant, then the minimum norm solution (i.e the optimal portfolio positions) of this system of equations is

$$w' = \frac{-1}{2}\left(\lambda_1 \Sigma^{\dagger}\mu + \lambda_2 \Sigma^{\dagger} 1\right) \tag{26}$$

The uniqueness of the solution is due to the uniqueness of the Moore-Penrose Inverse. Using this vector, we have that the optimum portfolio selection is given by

$$w = \frac{1}{AC - B^2}\left[(C - qB)\Sigma^{\dagger} 1 + (qA - B)\Sigma^{\dagger}\mu\right] \tag{27}$$

where $A = 1'\Sigma^{\dagger}1, B = 1'\Sigma^{\dagger}\mu, C = \mu'\Sigma^{\dagger}\mu$.
When the covariance matrix is close to singular, then, if we replace Σ^{-1} with $\Sigma^{\dagger}$, the results coincide. In the case of a close to singular, but ill-conditioned covariance matrix, (a large condition number) the use of the Moore-Penrose Inverse $\Sigma^{\dagger}$ gives marginally better results than Σ^{-1}.

References

[1] Awni Abu-Saman A. Solution of Linear-Dependent Equations by Generalised inverse of a Matrix *International Journal of Science and Emerging Technologies*, vol 4, No 2. 2012.

[2] Barata, J.C.A. and Hussein, M.S. The Moore-Penrose Pseudoinverse: A Tutorial review of the theory *Instituto de Fısica, Universidade de Sao Paulo*, 2011.

[3] Ben-Israel, A. and Granville, T. N.E Generalized Inverses: Theory and Applications *Springer-Verlag*, Berlin 2002.

[4] Ben-Israel, A. The Moore of the Moore-Penrose Inverse *Electronic Journal of Linear Algebra*, pp 150-157, vol 9 2002.

[5] Kirk B. Single Value Decomposition Tutorial notes 2005.

[6] Lopez-Bonilla J., Lopez-Vasquez, R. and Vidal-Beltran, S. Moore-Penrose inverse and solutions to linear systems *World Scientific News*, pp 246-252, 2018.

[7] Markowitz, H. Porfolio Selection *The Journal of Finance*, Vol 7,No 1, (1952), pp 77-91.

[8] Pappas, D. and Kanstantinis, K. Optimal Portfolio Selection with singular Covariance matrix *International mathematical Forum*, No 47, pp 2305-2318, Vol 5, 2010.

[9] Penrose, R. A Generalized inverse of Matrices *Mathematical proceedings of the Cambridge Philosophical Society*, vol 51, pp 406-413 1955.

[10] Peter J. H. Notes on Matrix Algebra, Similiar Matrices and Diagonalization 2018.

[11] Rao, C.R. and Maitra, S.K. Generalized inverse of Matrices and its Applications *Wiley, New York*, 1971.

[12] Ross, M. Moore Penrose Inverse, Lecture notes on Advanced Topics in Linear Algebra, *University of Puget Sound*, 2014.