UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

TRABAJO DE FIN DE GRADO EN MATEMÁTICAS

PSEUDOINVERSA DE MOORE-PENROSE : PROPIEDADES, APLICACIONES Y GENERALIZACIÓN A ESPACIOS VECTORIALES DE DIMENSIÓN INFINITA

Autor: Inés Milán Sanz

Tutor: Fernando Pablos Romo

Salamanca, Julio 2021

UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

TRABAJO DE FIN DE GRADO EN MATEMÁTICAS

PSEUDOINVERSA DE MOORE-PENROSE : PROPIEDADES, APLICACIONES Y GENERALIZACIÓN A ESPACIOS VECTORIALES DE DIMENSIÓN INFINITA.

Autor:

Tutor:

Firmado por
Inés Milán Sanz
el día 6 julio
2021

Firmado por PABLOS ROMO
FERNANDO - 07872052A el
día 06/07/2021 con un
certificado emitido por

Fdo: Inés Milán Sanz

Fdo: Fernando Pablos Romo

Salamanca, Julio 2021

Índice

1. Introducción

El objetivo de este trabajo consiste en obtener resultados que permitan demostrar la existencia y unicidad de la inversa de Moore-Penrose de una matriz con coeficientes complejos.

A partir de las propiedades de dicha inversa se estudiará su aplicación a la resolución de sistemas de ecuaciones lineales no homogéneos e incompatibles por el método de mínimos cuadrados. Finalmente, se extiende la noción y propiedades de inversa de Moore-Penrose a aplicaciones lineales en espacios vectoriales de dimensión infinita.

Este trabajo comienza con una breve sección dedicada a los aspectos de notación y terminología de la que se hará uso durante el desarrollo del mismo. A continuación, se analizan las primeras definiciones de inversas generalizadas, así como algunos resultados fundamentales de las mismas y, para el mejor entendimiento por parte de los lectores, se aportan algunos ejemplos. Además, se introduce una descomposición particular de matrices (*la descomposición centro-nilpotente*) y, paralelamente, se define este concepto para endomorfismos.

Una vez concluida esta sección, el trabajo se centra en una inversa generalizada concreta, *la inversa de Moore-Penrose*.

Se empieza por dar su definición y demostrar la unicidad. A continuación, se enuncian sus propiedades básicas, para algunas de las cuales se detalla su demostración.

Más adelante, se introducen dos algoritmos de obtención de la inversa de Moore-Penrose basados en dos teoremas, los cuales también se incluyen en este apartado. Además, se introducen dos nuevos conceptos de matrices (*matriz escalonada por filas* y *matriz en la forma escalonada de Hermite*), de los que se hará uso en los algoritmos mencionados. Este apartado concluye, para una mejor compensión por parte del lector, con dos ejemplos sobre lo anteriormente mencionado.

Una vez terminado el apartado dedicado a la obtención de la inversa de Moore-Penrose, y previamente al desarrollo de las aplicaciones de la misma, se estudian los resultados de dicha inversa del producto de dos matrices. Además, se cuestiona bajo qué condiciones, la inversa de Moore-Penrose de un producto, sigue las propiedades conocidas para el producto de matrices invertibles.

Como se ha mencionado, la última sección trata algunas de las aplicaciones que tiene la inversa de Moore-Penrose.

El hilo central de esta parte del trabajo es el de la resolución de sistemas de ecuaciones lineales no homogéneos, para los cuales, la matriz de coeficientes no es una matriz cuadrada e inverible.

Se comienza dando la definición de *mínimos cuadrados*, que es un método usado para la estimación de soluciones. Se aporta, además, la expresión de una solución de mínimos cuadrados en relación a la inversa de Moore-Penrose y se desarrollan resultados equivalentes para los cuales la solución coincide con la de mínimos cuadrados.

Después, se realizan dos *hipótesis lineales* para las que se estiman los parámetros que las forman y se calcula cómo de buena ha sido dicha estimación, es decir, la bondad de ajuste.

Todo esto posibilita afrontar el problema que enfrenta Gauss, y el resto de la comunidad científica, que, aunque se desarrolla con más profundidad en la sección dedicada a él, consiste en ajustar una cónica a un conjunto de puntos dados. Para ello, se realizan cuatro versiones del problema real en las que se usan los conceptos estudiados hasta el momento.

La sección con la que se concluye este trabajo está dedicada a la extensión al caso infinito de los sistemas de ecuaciones lineales. Comienza con la definición y propiedades de la inversa de Moore-Penrose para aplicaciones lineales, así como la demostración de su existencia y unicidad. Para poder entender estos resultados, se definen varios conceptos, entre los que se encuentra el *producto interno* y la *inversa generalizada de una ecuación lineal.*

Por último, se detallan los conceptos necesarios par dar la condición necesaria y suficiente bajo la cual, los sistemas de ecuaciones lineales de dimensión infinita, tienen solución denominada *menor solución mínima de norma* $\bar{g}$.

2. Preliminares

Para facilitar el entendimiento por parte del lector de los sucesivos apartados de este trabajo, se aportan las siguientes definiciones previas y notaciones, incluídas en el apartado de 'Terminología'. Así mismo, se aporta a lo largo de esta primera introdución demostraciones que serán útiles para el desarrollo del bloque central, que será el objeto de estudio.

2.1. Terminología.

A lo largo de este trabajo se hará uso de las sigueintes notaciones.
El espacio vectorial de las matrices complejas de orden $m \times n$ viene dado por $Mat_{m\times n}(\mathbb{C})$. I_n denota la matriz identidad de orden n. Si $A \in Mat_{m\times n}(\mathbb{C})$, se usará $rg(A)$ para denotar al máximo número de vectores columna linealmente independientes de A y $R(A)$ para el rango de A, es decir, el espacio vectorial de las columnas de A. Por otra parte, cuando se hable del núclo o kernel de una aplicación que tiene como matriz asociada A, se lo denotará por $N(A)$.
Dentro de la notación utilizada para las matrices, la *traza de una matriz* A a la que se hará referencia por $tr(A)$ es el escalar que se obtiene al sumar los elementos de la diagonal principal de dicha matriz.

Además, se dice que una matriz A es *transpuesta conjugada* de otra y se denotará por A^* si es el resultado de tomar el complejo conjugado de cada elemento de A y transponer, es decir, permutar filas por columnas. Una matriz A se dice que es *hermitiana* si es igual a su conjugada transpuesta, $A = A^*$. Si $A^2 = A$, entonces A recibe el nombre de *proyección* de $\mathbb{C}^n$ sobre $R(A)$. Si, $A^2 = A$ y $A = A^*$, entonces A es una *proyección ortogonal* y se denota por $P_{R(A)}$.

Se usará el término de *morfismo* para hacer referencia a las aplicaciones lineales; así como el de *monomorfismo*, *epimorfismo* e *isomorfismo* cuando el morfismo sea inyectivo, epiyectivo o biyectivo respectivamente y *endomorfismo* cuando el espacio de salida y llegada del morfismo coincida. Además, se usará el nombre de *automorfismo* cuando se tenga un endomorfismo isomorfo.

Dado V un k-espacio vectorial y U, W dos subespacios de V, se dice que dichos subespacios están en posición de *suma directa* y se denota por $V = U \oplus W$ si verifican $U \cap W = 0$ y $U + W = V$.

Dado un morfismo $\varphi : V \longrightarrow V'$ entre k-espacios vectoriales, la *imagen* de φ es el subespacio $Im(\varphi) = \{v' \in V' : \forall v \in V$ de modo que $\varphi(v) = v'\} \subseteq V'$ y el *núcleo* de φ es el subespacio $Ker\varphi = \{v \in V : \varphi(v) = 0\} \subseteq V$.

Se conoce por *matriz nilpotente* a una matriz cuadrada $A \in Mat_{n\times n}$ que verifica $A^n = 0$ para cierto n. El menor número n tal que $A^n = 0$ recibe el nombre de *índice de nilpotencia*. También puede decirse que la matriz nilpotente es de *orden n*. Además, se verifica que si una matriz es nilpotente, su determinante es nulo.

2.2. Inversas generalizadas.

En este primer apartado se va a analizar el concepto de matriz inversa de una matriz cuadrada $A \in Mat_{n\times n}(\mathbb{C})$ y se introducirán lo conceptos de *inversas latalaterales, 1-inversa, 2-inversa e inversa de grupo.*

Con el fin de motivar las definiciones de inversas generalizadas, se recuerda que una matriz cuadrada $A \in Mat(n, \mathbb{C})$ es **no singular** si existe una matriz $B \in Mat(n, \mathbb{C})$ que verifique:

$$AB = BA = I_n.$$

Dicha matriz B se denota por A^{-1} y recibe el nombre de **matriz inversa de A**.

Al considerar la restricción de esta definición, pues es únicamente válida para matrices cuadradas, surge la necesidad de ampliarla, de obtener nuevas bases teóricas que permitan trabajar con el resto de matrices.

Definición 2.2.1. *Dada una matriz $A \in Mat_{m\times n}(\mathbb{C})$, se llama* ***inversa a derecha*** *de A a la matriz $R \in Mat_{n\times m}(\mathbb{C})$ que verifica $A \cdot R = I_m$. Análogamente, la matriz $L \in Mat_{n\times m}(\mathbb{C})$ tal que $L \cdot A = I_n$ recibe el nombre de* ***inversa a izquierda*** *de A.*

Ejemplo 2.2.1. *Se considera la matriz $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \in Mat_{3\times 2}(\mathbb{C})$. La matriz inversa a izquierda de A es $L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \in Mat_{2\times 3}(\mathbb{C})$ que verifica:*

$$L \cdot A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

Nótese que $A \cdot L \neq I_3$.

Análogamente a las nociones dadas para matrices, se introduce la definición de *inversa a derecha e izquierda* para morfismos.

Definición 2.2.2. *Dado el morfismo $\phi : V \longrightarrow V'$, se llama* ***inversa a la izquierda*** *de ϕ al morfismo $\psi : V' \longrightarrow V$ tal que $\psi \circ \phi = I_V$ donde $I_V : V \longrightarrow V$ denota la aplicación identidad en V. De modo análogo se define la* ***inversa a derecha*** *de una aplicación lineal.*

Teorema 2.1. *Dada el morfismo $\phi : V \longrightarrow V'$, se verifica:*

1. *ϕ tiene inversa a la izquierda si y solo si es inyectiva.*

2. *ϕ tiene inversa a la derecha si y solo si es un epiyectiva.*

1. *Si ϕ es un monomorfismo, es decir, una aplicación lineal inyectiva*

Demostración. La demostración a este teorema está disponible en [9]. □

Este teorema da lugar al siguiente lema.

Lema 2.1. *Una matriz $A \in Mat_{m\times n}(\mathbb{C})$ tiene inversa a derecha si y solo si el rango de A es m; y tiene inversa a izquierda si y solo si el rango de A es n.*
Además, una matriz tiene inversa a izquierda y derecha si y solo si es una matriz cuadrada de rango máximo.

Demostración. La demostración está disponible en [9]. □

El hecho de que, de nuevo, no se asegure la existencia de inversas laterales para todas las matrices, permite introducir un nuevo concepto de inversa.

Definición 2.2.3. *Sea $A \in Mat_{m\times n}(\mathbb{C})$. La matriz $X \in Mat_{n\times m}(\mathbb{C})$ que cumple*

$$AXA = A$$

recibe el nombre de ***{1}-inversa*** *o* ***inversa interior*** *de* A*.*

Definición 2.2.4. *Sea* $A \in Mat_{m\times n}(\mathbb{C})$*. Una matriz* $X \in Mat_{n\times m}$ *que satisface*

$$XAX = X$$

se llama ***{2}-inversa*** *o* ***inversa exterior*** *de* A*.*

Las matrices que verifiquen tanto la Definición 2.1.2 y 2.1.3 se llaman **{1,2}-inversa** de A.

Definición 2.2.5. *Sea* $A \in Mat_{n\times n}(\mathbb{C})$*. Se define el* ***índice de A*** *y se denota por* $i(A)$ *al menor entero no negativo que satisface*

$$rg(A^{i(A)}) = rg(A^{i(A)+1}).$$

Observación 2.2.1. *Por convenio, toda matriz invertible tiene índice 0 y toda matriz nula tiene índice 1.*

Ejemplo 2.2.2. *Se considera la matriz* $A = \begin{bmatrix} 5 & -9 & 4 \\ 6 & -11 & -5 \\ -7 & 13 & 6 \end{bmatrix}$ *con* $rg(A) = 2$*.*

Se tiene: $A^2 = \begin{bmatrix} -1 & 2 & 1 \\ -1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix}$ *con* $rg(A^2) = 1$*.*

Y se concluye que $A^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ *con* $rg(A^3) = 0$*. Las potencias de la matriz* A *que sean mayores o iguales a tres trendrán como resultado la matriz nula y por tanto, el rango de todas ellas será cero. Luego, el menor entero positivo* k *para el que se verifica la definición de índice es 3, con lo que* $i(A) = 3$

En el caso de una matriz invertible A, es claro que $AA^{-1} = I$. Conjuntamente con las definiciones de inversa interior y exterior, se tiene la siguiente generalización.

Definición 2.2.6. *Sea* $A \in Mat_{n\times n}(\mathbb{C})$*. Se dice que una matriz es* ***inversa de grupo de*** A *y se denota por* $A^{\#} \in Mat_{n\times n}(\mathbb{C})$ *si verifica:*

- $A^{\#}$ *es una inversa interior de* A*. Es decir,* $AA^{\#}A = A$*.*
- $A^{\#}$ *es una inversa exterior de* A*. Es decir,* $A^{\#}AA^{\#} = A^{\#}$*.*
- $AA^{\#} = A^{\#}A$*.*

Como se demuestra en [7, Teorema 2.1.6.], la inversa de grupo no necesariamente existe, aunque, de ser así, es única.

Una condición necesaria y suficiente para que una matriz cuadrada tenga inversa de grupo es que $i(A) \leq 1$. Para más infomación sobre este resultado, consultar [6] y [2].

2.3. Inversas generalizadas clásicas.

Definición 2.3.1. *En 1958, M. P. Drazin -[5]- prueba que, dada una matriz cuadrada* $A \in Mat_{n\times n}(\mathbb{C})$ *se define la* ***inversa de Drazin*** *como la única matriz cuadrada* $A^D \in Matn \times n(\mathbb{C})$ *que satisface:*

- $A^{k+1}A^D = A^k$ *para* $k = i(A)$
- $A^D A A^D = A^D$
- $A^D A = A A^D$

Además, la inversa de Drazin también cumple:

- $(A^D)^D = A$ *si y solo si* $i(A) \leq 1$
- *Si* $A^2 = A$ *entonces* $A^D = A$

Análogamente a la definición que se ha dado sobre matrices, tanto de índice como de inversa de Drazin, se introducen estos conceptos para endomorfismos. No obstante, para poder dar una defición de manera rigurosa, primero se ve qué es un *endomorfismo de potencia finita* y en qué consiste la *descomposición AST invariante por* ϕ de un k-espacio vectorial V, donde ϕ es un endomorfismo sobre V.

Definición 2.3.2. *Sea* V *un* k*-espacio vectorial y* $\phi : V \longrightarrow V$ *un endomorfismo sobre* V*. Se dice que* ϕ *es* ***de potencia finita*** *si para algún* $n \in \mathbb{N}$ $\phi^n V$ *tiene dimensión finita.*

En 2007 M. Argerami, F. Szechtman y R. Tifenbach -[1]- demostraron que, una condición necesaria y suficiente para que ϕ sea de potencia finita es que V admita una descomposición invariante por ϕ, es decir, $V = U_\phi \oplus W_\phi$ de modo que $\phi\ |_{U_\phi}$ es nilpotente, W_ϕ tiene dimensión finita y $\phi\ |_{W_\phi}\colon W_\phi \longrightarrow W_\phi$ sea un isomorfismo. Esta descomposición es única y recibe el nombre de **descomposición AST de V invariante por** ϕ.

Definición 2.3.3. *Dado* V *un* k*-espacio vectorial arbitrario y* $\phi \in End_k(V)$*, se denomina* ***índice*** *de* ϕ *y se denota por* $i(\phi)$ *al orden de nilpotencia de* $\phi|_{U_\phi}$*.*

En [8, Definition 2.2.] se enuncia que, para espacios vectoriales de dimensión finita, la deficinión de índice de un endomorfismo coincide con la dada para la matriz asociada al endomorfismo en una pareja de bases. Se cumple que $i(\phi) = 0$ si y solo si V es un espacio vectorial de dimensión finita y ϕ es un automorfismo. Además, $i(\phi) = 1 \Leftrightarrow Ker\phi = U_\phi$.

Una vez definidos estos conceptos, se puede definir la inversa de Drazin en endomorfimos de potencia finita.

Definición 2.3.4. *Para cada endomorfismo de potencia finita* ϕ *existe un único endomorfismo de potencia finita al que se denota por* ϕ^D *que cumple:*

- $\phi^{k+1} \circ \phi^D = \phi^k$
- $\phi^D \circ \phi \circ \phi^D = \phi^D$
- $\phi^D \circ \phi = \phi \circ \phi^D$

  donde k *es el índice de* ϕ*.*

*ϕ^D es la **inversa de Drazin** de ϕ y verifica:*

$$\phi^D(v) = \begin{cases} (\phi|_{W_\phi})^{-1} & si \quad v \in W_\phi \\ \\ 0 & si \quad v \in U_\phi \end{cases}$$

Además, ϕ^D satisface las siguiente propiedades:

- *$(\phi^D)^D = \phi$ si y solo si $i(\phi) \leq 1$.*
- *$\phi = \phi^D$ si y solo si $\phi|_{U_\phi} = 0$ y $(\phi|_{W_\phi})^2 = Id|_{W_\phi}$.*
- *$tr_V(\phi + \phi^D) = tr_V(\phi) + tr_V(\phi^D)$ donde $tr_V(\phi)$ denota la traza del endomorfismo de potencia finita ϕ. Esta traza viene definida por la siguiente expresión:*

  $$tr_V(\phi) = tr_{w_\phi}(\phi|_{w_\phi})$$

  y además cumple las siguientes tres propiedades:

  1. *Si V es un espacio vectorial de dimensión finita, entonces $tr_V(\phi)$ coincide con la definición habitual de traza.*
  2. *Si W es un subespacio de V de modo que $\phi W \subseteq W$, entonces:*

     $$tr_V(\phi) = tr_W(\phi) + tr_{V/W}(\phi)$$

  3. *Si ϕ es nilpotente, entonces $tr_V(\phi) = 0$.*
- *Si ϕ es una proyeción de un endomorfismo de potencia finita, entonces $\phi^D = \phi$.*

Se concuye esta sección con las definiciones de *inversa generalizada reflexiva de una matriz*, *inversa generalizada reflexiva de un morfismo* e *invesa de Moore-Penrose* para una matriz.

Definición 2.3.5. *Sea $A \in Mat_{n\times m}(\mathbb{C})$. Se denomina **inversa generalizada reflexiva de la matriz A** y se denota por $A^+ \in Mat_{n\times m}(\mathbb{C})$ a una matriz que verifica:*

- *$AA^+A = A$*
- *A^+AA^+*

Observación 2.3.1. *La inversa generalizada reflexiva de una matriz no tiene por qué ser única.*

Definición 2.3.6. *Sea $\phi : V \longrightarrow V'$ una aplicación lineal entre κ-espacios vectoriales, se define la **inversa generaliada reflexiva de** ϕ a un morfismo $\phi^+ : V' \longrightarrow V$ tal que:*

- *$\phi \circ \phi^+ \circ \phi = \phi$*
- *$\phi^+ \circ \phi \circ \phi^+ = \phi^+$*

Observación 2.3.2. *Al igual que en el caso de las matrices, la inversa generalizada de un morfismo no tiene por qué ser única. Además, para cada invesa generalizada reflexiva ϕ^+ de ϕ, se cumple que, si $v' \in V'$*

$$v' \in Im\phi \Leftrightarrow (\phi \circ \phi^+)(v') = v'$$

Se introduce la noción de *inversa de Moore-Penrose* puesto que se necesita para avanzar con las definiciones. No obstante, es simplemente un primer acercamiento a esta inversa generalizada puesto que, en los capítulos posteriores se profundizará en su definición, propiedades y aplicaciones.

Definición 2.3.7. *Sea $A \in Mat_{m\times n}(\mathbb{C})$, la* ***inversa de Moore-Penrose*** *de A es la matriz que se denota por $A^\dagger$ y es la única matriz X que verifica:*

- $AXA = A$
- $XAX = X$
- $(AX)^* = AX$
- $(XA)^* = XA$

donde X^* es la transpuesta conjugada de la matriz X.

2.4. Descomposición CN

Rescatando el concepto de índice de una matriz, anteriormente introducido, en este apartado se dará la definición y se explicarán algunos resultados interesantes de la descomposición CN, centro-nilpotente, tanto para matrices como para endomoporfismos y la definición de *inversa centro* de una matriz derivada de esta descomposición.

Teorema 2.2. *Sea $A \in Mat_{n\times n}(\mathbb{C})$. A puede escribirse como suma de dos matrices A_1 y A_2 verificando:*

- $R(A_1) = R(A_1^2)$ *o lo que es lo mismo,* $i(A) \leq 1$.
- A_2 *es nilpotente.*
- $A_1 \cdot A_2 = A_2 \cdot A_1 = 0$, *con* $A_1, A_2 \in Mat_{n\times n}(\mathbb{C})$

Esta descomposición es la ***descomposición CN, centro-nilpotente, de A*** *donde A_1 y A_2 son las partes 'centro' y 'nilpotente' de A respectivamente y es única.*

Demostración. La demostración a este teorema se encuentra en [6, Theorem 2.2.21.] □

A partir de este teorema, así como de las nociones de inversa de grupo y de Moore-Penrose, es posible considerar una nueva inversa generalizada, llamada *inversa centro.*

Definición 2.4.1. *Dada una matriz $A \in Mat_{nxn}(\mathbb{C})$ se llama* ***inversa centro de A*** *y se denota por $A^{\circledast} \in Mat_{nxn}(\mathbb{C})$ a la matriz tal que:*

- $A \cdot A^{\circledast} = A \cdot A^\dagger$
- $R(A^{\circledast}) \subseteq R(A)$

Se prueba en [7] que, en caso de existir, dicha matriz es única y verifica los siguientes teoremas:

Teorema 2.3. *Sea $A \in Mat_{nxn}(\mathbb{C})$. Si la matriz A admite inversa centro, entonces el índice de A es menor o igual que 1.*

Demostración. La demostración está disponible en [7, Proposición 3.13.]. □

Teorema 2.4. *Sea $A \in Mat_{nxn}(\mathbb{C})$ con índice 1. Entonces la inversa core $A^{\circledast}$ existe y es única.*

Demostración. La demostración está disponible en [7, Teorema 3.14.]. □

Análogamente a como se ha hecho con anteriores definiciones y, como se adelantaba al principio de este trabajo, se ofrece paralelamente a la descomposición para matrices, su ánalogo para endomorfismos.

Dado V un k-espacio vectorial arbitrario, ψ un endomorfismo de potencia finita de V y se considera la descomposición AST de V dada por $V = U_\psi \oplus W_\psi$.

Lema 2.2. *Con la notación previa, si $\psi \neq 0$ y ψ^D es la inversa de Drazin de ψ, entonces:*

1. $\psi_1 = \psi \circ \psi^D \circ \psi$ *es finito potente e* $i(\psi) \leq 1$;
2. $\psi_2 = \psi - \psi \circ \psi^D \circ \psi$ *es nilpotente.*

Demostración. La demostración se encuentra en [8, Lemma 3.1.]. □

Teorema 2.5. ***(Exitencia y unicidad de la descomposicón CN de endomorfismos de potencia finita).*** *Para cada endomorfismo de potencia finita ψ sobre V con $\psi \neq 0$ existe una única descomposición $\psi = \psi_1 + \psi_2$, donde $\psi_1, \psi_2 \in End_k(V)$ son endomorfismos finito potente que satisfacen:*

- $i(\psi) \leq 1$;
- ψ_2 *es nilpotente;*
- $\psi : 1 \circ \psi_2 = \psi_2 \circ \psi_1 = 0$.

De esta manera, los endomorfismos de potencia finita ψ_1 y ψ_2 son las partes centro y nilpotente de las descomoposición CN de ψ respectivamente.

Demostración. La demostración se encuentra en [8, Theorem 3.2.]. □

3. Existencia y propiedades de la pseudoinversa de Moore-Penrose

A lo largo de esta sección se van a dar las definiciones de Moore y Penrose de la pseudoinversa de una matriz y se va a demostrar la equivalencia entre ambas definiciones.

Además, se van a enunciar y demostrar varias de sus propiedades principales.

Esta sección concluye con el desarrollo de las condiciones bajo las cuales, la inversa de Moore-Penrose del producto de dos matrices, cumple las propiedades conocidas para matrices invertibles.

3.1. Existencia.

Como se ha introducido anteriormente, la inversa de Moore-Penrose o pseudoinversa de Moore-Penrose es una inversa generalizada de matrices, que puede ser aplicada para la resolución de sistemas de ecuaciones lineales.

En 1935, Moore dio la definición que se detalla a continuación.

Definición 3.1.1. ***Definición de inversa generalizada de Moore.*** *Si $A \in Mat_{mxn}(\mathbb{C})$, entonces la inversa generalizada de A es la única matriz $A^\dagger$ tal que:*

I. $AA^\dagger = P_{R(A)}$

II. $A^\dagger A = P_{R(A^\dagger)}$

Años más tarde, en 1955, Penrose ofreció una nueva definición de inversa generalizada.

Definición 3.1.2. ***Definición de inversa generalizada de Penrose.*** *Si $A \in Mat_{mxn}(\mathbb{C})$, entonces $A^\dagger$ es la única matriz en $Mat_{nxm}(\mathbb{C})$ que verifica:*

I. $AA^\dagger A = A$

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

III. $(AA^\dagger)^* = AA^\dagger$

IV. $(A^\dagger A)^* = A^\dagger A$

Esta última definición es la que se conoce como la pseudoinversa de Moore-Penrose y es con la que se trabajará a lo largo de este proyecto. Esta matriz es única, lo que se demuestra en el siguiente teorema.

Teorema 3.1. *Las definiciones de inversa generalizada dadas por Moore y Penrose son equivalentes.*

Demostración. Sea $A \in Mat_{mxn}(\mathbb{C})$ verificando las condiciones de la definición de Moore, es decir, tal que $AA^\dagger = P_{R(A)}, A^\dagger A = P_{R(A^\dagger)}$. Se tiene entonces que verifican las condiciones III, IV de la Definición 3.1.2. pues

- $(AA^\dagger)^* = (P_{R(A)})^* = P_{R(A)} = AA^\dagger$;
- $(A^\dagger A)^* = (P_{R(A^\dagger)})^* = P_{R(A^\dagger)} = A^\dagger A$.

Por otra parte, para probar I y II de la Definición 3.1.2., se observa que

- $AA^\dagger A = P_{R(A)}A = A$;
- $A^\dagger AA^\dagger = P_{R(A^\dagger)}A^\dagger = A^\dagger$.

Sea ahora $A \in Mat_{mxn}(\mathbb{C})$ verificando las condiciones I, II, III, IV de la Definición 3.1.2. de Penrose. Multiplicando II a la izquierda por A se obtiene

$$AA^\dagger AA^\dagger = AA^\dagger \Rightarrow (AA^\dagger)^2 = AA^\dagger.$$

Esto, junto con $(AA^\dagger)^* = AA^\dagger$, demuestra que $AA^\dagger$ es una proyección ortogonal. Falta demostrar que el rango de $AA^\dagger$ coincide con el rango de A para lo cual, se tiene en cuenta la condición I de la Definición 3.1.2 junto con el hecho de que $R(BC) \subseteq R(B)$ para dos matrices B, C. Con esto, se tiene que

$$R(A) = R(AA^\dagger A) \subseteq R(AA^\dagger)$$
$$R(AA^\dagger) \subseteq R(A).$$

Luego, $R(A) = R(AA^\dagger)$ y, por tanto, $AA^\dagger = P_{R(A)}$.

Para probar la segunda condición de la Definición 3.1.1. se procede de modo análogo.

□

Teorema 3.2. *Dada una matriz $A \in Mat_{nxm}(\mathbb{C})$, existe una única matriz $A^\dagger \in Mat_{mxn}(\mathbb{C})$. Es decir, la inversa de More-Penrose de una matriz es única.*

Demostración. Supóngase que existen dos matrices B, C que satisfacen las cuatro propiedades de la definición de Moore-Penrose. Se tiene entonces:

$B = BAB = (BA)^*B = (A^*B^*)B = (ACA)^*B^*B = (A^*C^*A^*)B^*B = (CA)^*(BA)^*B = CABAB = CAB$.

Además

$C = CAC = C(AC)^* = CC^*A^* = CC^*(ABA)^* = CC^*(A^*B^*A^*) = C(AC)^*(AB)^* = C(ACA)B = CAB$.

Con lo que $B = C$. □

3.2. Propiedades

Se darán a continuación algunas de las propiedades básicas de la inversa de Moore-Penrose aunque previamente hay que considerar la siguiente definición.

Definición 3.2.1. *Una matriz unitaria $U \in Mat_{nxn}(\mathbb{C})$ es una matriz compleja cuadrada que satisface:*

$$U^*U = UU^* = I_n.$$

Teorema 3.3. *Sea $A \in Mat_{mxn}(\mathbb{C})$. Se cumple:*

1. $(A^\dagger)^\dagger = A$
2. $(A^\dagger)^* = (A^*)^\dagger$
3. *Si* $\lambda \in \mathbb{C}, (\lambda A)^\dagger = \lambda^\dagger A^\dagger$ *donde* λ *viene definida por* $\lambda^\dagger = \frac{1}{\lambda}$ *si* $\lambda \neq 0$ *y* $\lambda^\dagger = 0$ *si* $\lambda = 0$

4. $A^* = A^*AA^\dagger = A^\dagger AA^*$

5. $(A^*A)^\dagger = A^\dagger(A^*)^\dagger$

6. $A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger$

7. $(UAV)^\dagger = V^*A^\dagger U^*$ *donde* U, V *son matrices unitarias.*

Demostración. Para los enunciados 1, 2, 3, 5 y 7, se trata de una inversa generalizada, luego para su demostración se van a comprobar los cuatro axiomas de la definición de Moore-Penrose.

1. Cumple los axiomas de la definición sin necesidad de realizar ningún cálculo.

2.
   - $A^*(A^\dagger)^*A^* = (AA^\dagger A)^* = A^*$
   - $(A^\dagger)^*A^*(A^\dagger)^* = (A^\dagger AA^\dagger)^* = (A^\dagger)^*$
   - $((A^\dagger)^*A^*)^* = AA^\dagger = (AA^\dagger)^* = (A^\dagger)^*A^*$
   - $(A^*(A^\dagger)^*)^* = A^\dagger A = (A^\dagger A)^* = A^*(A^\dagger)^*$

3. Para $\lambda = 0$ la demostración es trivial. En caso de que $\lambda \neq 0; (\lambda = \frac{1}{\lambda})$ se tiene:
   - $\lambda A\lambda^\dagger A^\dagger \lambda A = \lambda\lambda^\dagger\lambda(AA^\dagger A) = \lambda A$
   - $\lambda^\dagger A^\dagger\lambda A\lambda^\dagger A^\dagger = \lambda^\dagger\lambda\lambda^\dagger(A^\dagger AA^\dagger) = \lambda^\dagger A^\dagger$
   - $(\lambda A\lambda^\dagger A^\dagger)^* = (\lambda\lambda^\dagger AA^\dagger)^* = (AA^\dagger)^* = AA^\dagger)\lambda\lambda^\dagger AA^\dagger = \lambda A\lambda^\dagger A^\dagger$
   - $(\lambda^\dagger A^\dagger\lambda A)^* = (\lambda^\dagger\lambda A^\dagger A)^* = (A^\dagger A)^* = A^\dagger A = \lambda\lambda^\dagger A^\dagger A = \lambda^\dagger A^\dagger\lambda A$

4. Por una parte se tiene que $A^* = (A(A^\dagger A))^* = (A^\dagger A)^*A^* = A^\dagger AA^*$. Para la otra igual se procede de manera similar: $A^* = ((AA^\dagger)A)^* = A^*(AA^\dagger)^* = A^*AA^\dagger$

5.
   - $(A^*A)(A^\dagger(A^*)^\dagger)(A^*A) = (A^*AA^\dagger)(A^*)^\dagger A^*A = A^*(A^\dagger)^*A^*A = (A^\dagger A)^*A^*A = (AA^\dagger A)^*A = A^*A$
   - $(A^\dagger(A^*)^\dagger)(A^*A)(A^\dagger(A^*)^\dagger = A^\dagger(A^*)^\dagger(A^*AA^\dagger)(A^*)^\dagger = A^\dagger(A^\dagger)^*A^*(A^\dagger)^* = A^\dagger(AA^\dagger)^*(A^\dagger)^* = A^\dagger(A^\dagger AA^\dagger)^* = A^\dagger(A^\dagger)^* = A^\dagger(A^*)^\dagger$
   - $(A^*AA^\dagger(A^*)^\dagger)^* = (A^*AA^\dagger(A^*)^*)^\dagger = A^\dagger(A^\dagger)^*A^*A = A^\dagger(AA^\dagger)^*A = A^\dagger AA^\dagger A = A^\dagger A = (A^\dagger A)^* = A^*(A^\dagger)^* = A^*AA^\dagger(A^*)^\dagger$
   - $(A^\dagger(A^*)^\dagger A^*A)^* = (A^\dagger(A^\dagger)^*A^*A)^* = A^*AA^\dagger(A^\dagger)^* = A^*(A^\dagger)^* = (A^\dagger A)^* = A^\dagger A = A^\dagger AA^\dagger A = A^\dagger(AA^\dagger)^*A = A^\dagger(A^*)^\dagger A^*A$

6. $A^\dagger = A^\dagger AA^\dagger = A^\dagger(AA^\dagger)^* = A^\dagger(A^\dagger)^*A^* = A^\dagger(A^*)^\dagger A^* = (A^*A)^\dagger A^*$

7.
   - $UAVV^*A^\dagger U^*UAV = UAA^\dagger AV = UAV$
   - $V^*A^\dagger U^*UAVV^*A^\dagger U^* = V^*A^\dagger AA^\dagger U^* = V^*A^\dagger U^*$
   - $(V^*A^\dagger U^*UAV)^* = V^*A^*U^*U(A^\dagger)^*V = V^*(A^\dagger A)^*V = V^*A^\dagger AV = V^*A^\dagger U^*UAV$
   - $(UAVV^*A^\dagger U^*)^* = (UAA^\dagger U^*)^* = U(AA^\dagger)^*U^* = UAA^\dagger U^* = UAVV^*A^\dagger U^*$

□

Proposición 3.1. *Se verifica que:*

$$\begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix}^{\dagger} = \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix}$$

Demostración. De manera similar a como se ha demostrado el Teorema 3.2, se comprueban las condiciones de la definición para $\begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix}$.

- $$\begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix} \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix} \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix} = \begin{bmatrix} A_1A_1^{\dagger}A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_mA_m^{\dagger}A_m \end{bmatrix} = \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix}$$
- $$\begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix} \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix} \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix} = \begin{bmatrix} A_1^{\dagger}A_1A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger}AA_m^{\dagger} \end{bmatrix} = \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix}$$
- $$\left(\begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix} \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix}\right)^* = \left(\begin{bmatrix} A_1A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_mA_m^{\dagger} \end{bmatrix}\right)^* = \begin{bmatrix} A_1A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_mA_m^{\dagger} \end{bmatrix} =$$
$$\begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix} \begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix}$$
- $$\left(\begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix} \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix}\right)^* = \left(\begin{bmatrix} A_1^{\dagger}A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger}A_m \end{bmatrix}\right)^* = \begin{bmatrix} A_1^{\dagger}A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger}A_m \end{bmatrix}$$
$$\begin{bmatrix} A_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m^{\dagger} \end{bmatrix} \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{bmatrix}$$

$\square$

Como se enuncia en [4, Theorem 1.2.2.].

Teorema 3.4. *Si $A \in Mat_{mxn}(\mathbb{C})$, entonces se verifica:*

1. $R(A) = R(AA^{\dagger}) = R(AA^*)$.
2. $R(A^{\dagger}) = R(A^*) = R(A^{\dagger}A) = R(A^*A)$.
3. $R(I - AA^{\dagger}) = N(AA^{\dagger}) = N(A^*) = N(A^{\dagger}) = R(A)^{\perp}$.
4. $R(I_A^{\dagger}A) = N(A^{\dagger}A) = N(A) = R(A^*)^{\perp}$.

3.3. Cálculo de la pseudoinversa de Moore-Penrose

En esta sección se van a dar dos métodos que suponen un primer acercamiento al cálculo de $A^\dagger$. Estos algoritmos son últiles únicamente para matrices de dimensiones pequeñas.

Previamente al desarrollo de los mismo, se tiene en cuenta la siguiente definición.

Definición 3.3.1. ***Inversa funcional de la inversa generalizada.***

Dada una matriz $A \in Mat_{mxn}(\mathbb{C})$, se define una transformación lineal $A^\dagger : \mathbb{C}^m \longrightarrow \mathbb{C}^n$ por $A^\dagger x = 0$ si $x \in R(A)^\perp$ y $A^\dagger x = (A|_{R(A^\dagger)})^{-1}x$ si $x \in R(A)$.

Teorema 3.5. *Sea $A \in Mat_{mxn}(\mathbb{C})$ de rango r. Si $\{v_1 \dots v_r\}$ es una base de $R(A^*)$ y $\{w_1, \dots w_{n-r}\}$ es una base de $N(A^*)$, entonces*

$$A^\dagger = [v_1|v_2|\dots|v_r|0|\dots|0] \cdot [Av_1|Av_2|\dots|Av_r|w_1|w_2|\dots|w_{n-r}]^{-1}$$

Demostración. Teniendo en cuenta la definición funcional de la inversa generalizada, es claro que

$$A^\dagger \cdot [Av_1|Av_2|\dots|Av_r|w_1|w_2|\dots|w_{n-r}] =$$
$$[A^\dagger \cdot Av_1|A^\dagger \cdot Av_2|\dots|A^\dagger \cdot Av_r|A^\dagger \cdot w_1|A^\dagger \cdot w_2|\dots|A^\dagger \cdot w_{n-r}] = [v_1|\dots|v_r|0|\dots|0].$$

Además, $\{Av_1, Av_2, \dots, Av_r\}$ es una base de $R(A)$. Además, como $R(A)^\perp = N(A^*)$, se deduce que la matriz $[Av_1|Av_2|\dots|Av_r|w_1|w_2|\dots|w_{n-r}]$ debe ser no singular, con lo que se concluye. □

Este teorema, así como el siguiente que se enunciará en esta sección, pueden considerarse como algoritmos para el cálculo de la inversa de More-Penrose. Previamente a detallar dichos algoritmos se introducen los concpetos de *matriz escalonada por filas* y la *forma escalonada de Hermite*, así como propiedades de ambas y la siguiente proposición, útil para futuras demostraciones.

Se prueba en [4, Theorem 1.3.2.] la siguiente proposición.

Proposición 3.2. *Si $A \in Mat_{m\times n}(\mathbb{C})$ entonces existen dos matrices $B \in Mat_{m\times r}(\mathbb{C})$, $C \in Mat_{r\times n}(\mathbb{C})$ de modo que $A = BC$ y $r = rg(A) = rg(B) = rg(C)$.*

Teorema 3.6. *Si $A = BC$ donde $A \in Mat_{m\times n}(\mathbb{C}), B \in Mat_{m\times r}(\mathbb{C}), C \in Mat_{r\times n}(\mathbb{C})$ y $r = rg(A) = rg(B) = rg(C)$, entonces $A^\dagger = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$.*

Demostración. Lo primero a tener encuenta es que tiene sentido hablar de la inversa usual de B^*B y CC^* puesto que el resultado es una matriz cuadrada de dimensión $n \times n$. Sea $X = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$ y se va a probar que satisface las condiciones de la definición de Moore-Penrose:

- $(AX)^* = AX$

$$AX = BCX = BCC^*(CC^*)^{-1}(B^*B)^{-1}B^* = B(B^*B)^{-1}B^*$$

con lo que se tiene que:

$$(AX)^* = (B(B^*B)^{-1}B^*)^* = (B^*)^*((B^*(B^*)^*)^{-1}B^* = B(B^*B)^{-1}B^* = AX$$

- $(XA)^* = XA$

$$XA = XBC = C^*(CC^*)^{-1}(B^*B)^{-1}B^*C = C^*(CC^*)^{-1}C$$

con lo que se tiene:

$$(XA)^* = (C^*(CC^*)^{-1}C)^* = C^*((CC^*)^{-1})^*(C^*)^* = C^*((CC^*)^*)^{-1}C = C^*(CC^*)^{-1}C = XA$$

Nota. Para las siguientes dos propiedades se tendrá en cuenta que $XA = C^*(CC^*)^{-1}C$

- $AXA = A$

$$AXA = BCXA = BCC^*(CC^*)^{-1}C = BC = A$$

- $XAX = X$

$$XAX = (XA)X = (C^*(CC^*)^{-1}C)C^*(CC^*)^{-1}(B^*B)^{-1}B^* = C^*(CC^*)^{-1}(B^*B)^{-1}B^* = X$$

Como X verifica todas las propiedades, se concluye que $X = A^\dagger$ □

Se enuncia en [4, Theorem 1.3.3.] la siguiente proposición.

Proposición 3.3. *Si* $A \in Mat_{m\times n}(\mathbb{C})$ *con* $rg(A) = 1$*, entonces se verifica* $A^\dagger = \frac{1}{\alpha}A^*$*,* $\alpha = trA^*A = \sum_{i,j} |a_{ij}|^2$*.*

Definición 3.3.2. *Una matriz* $E \in Mat_{m\times n}(\mathbb{C})$ *de rango* r *se dice que es* ***escalonada por filas*** *si* E *es de la forma:*

$$E = \begin{bmatrix} C_{r\times n} \\ \dots \\ 0_{(m-r)xn} \end{bmatrix}$$

donde los elementos c_{ij} *de* $C_{r\times n}$ *satisfacen las siguientes condiciones:*

1. $c_{ij} = 0$ *cuando* $i > j$*.*
2. *El primer elemento distinto de cero de cada fila de* C *es 1.*
3. *Si* $c_{ij} = 1$ *es el primer elemento distinto de cero de la fila i-ésima, entonces la columna j-ésima de* C *es el vector unitario* e_i *cuyo único elemento distinto de cero es el que ocupa la posición* i*.*

Ejemplo 3.3.1. $E = \begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$

Algunas de las propiedades que cabe destacar de la forma escalonada por filas de las matrices son las siguientes:
Sea $A \in Mat_{m\times n}$con $rg(A) = r$.

- A siempre se puede reducir a la forma escalonada por filas mediante operaciones elementales con las filas (es decir, siempre existe una matriz $P \in Mat_{mxm}(\mathbb{C})$ de manera que $P \cdot A = E_A$ donde E_A es la forma escalonada por filas).
- Para una matriz A, la forma escalonada por filas E_A obtenida mediante la reducción por filas de A es única.

- Si E_A es la forma escalonada por filas de A y los vectores unitarios de E_A aparecen en las columnas $i_1, i_2, \dots, i_r$ entonces las correspondientes columnas de A son una base de $R(A)$. Esta base recibe el nombre de **columnas distinguidas de** A.
- Se verifica que $N(A) = N(E_A) = N(C)$, donde C es la matriz de orden rxn de la definición.
- Si se consideran las columnas distinguidas de A y se utilizan para formar una matriz $B \in Mat_{m\times r}(\mathbb{C})$ colocadas en el mismo orden en el que aparecen en A; entonces se cumple que: $A = BC$, donde C vuelve a ser la matriz de orden $r \times n$ de la definición.

De manera similar a cómo se ha definido la forma escalonada por filas de una matriz, se tiene la *forma escalonada de Hermite.*

Observación 3.3.1. *La forma escalonada de Hermite es únicamente válida para matrices cuadradas.*

Definición 3.3.3. *Una matriz $H \in Mat_{n\times n}(\mathbb{C})$ se dice que está en la* ***forma escalonada de Hermite*** *si sus elementos h_{ij} verifican:*

1. *H es triangular superior, es decir, $h_{ij} = 0$ para $i > j$.*
2. *Los elementos de la diagonal principal h_{ii} son o 1 o 0.*
3. *Si $h_{ii} = 0$, entonces $h_{ik} = 0$ para todo $k, 1 \leq k \leq n$. Es decir, si el elemento de la diagonal es cero, entonces todos los elementos de la misma fila, son cero.*
4. *Si $h_{ii} = 1$, entonces $h_{ki} = 0$ para todo $k \neq i$. Luego, si el elemento de la diagonal es 1, entonces el resto de elementos de esa columna son cero; o equivalentemente, dicho vector columna es un vector unitario.*

Ejemplo 3.3.2. $H = \begin{bmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$

Así mismo, esta matriz cuenta con una serie de propiedades.
Sea $A \in Mat_{n\times n}(\mathbb{C})$. Entonces:

- Siempre se puede reducir A por filas a la forma de Hermite. Una vez que A se haya reducido por filas, siempre se puede realizar una permutación de las mismas para obtener la forma de Hermite.
- La forma de Hermite H_A de una matriz A, obtenida mediante la reducción por filas, es siempre única.
- $H_A^2 = H_A$, por lo que, H_A es una proyección.
- $N(A) = N(H_A) = R(I - H_A)$. Además, una base de $N(A)$ es el conjunto de las columnas distintas de cero de $I - H_A$

Como ya se introdujo anteriormente, los últimos teoremas (Teorema 3.5 y Teorema 3.6) pueden considerarse como algoritmos paa el cálculo de la inversa generalizada. Se detallan a continuación los pasos a seguir, así como un ejemplo práctico de cada uno de ellos.

1. **Primer algoritmo.** Hay que tener en cuenta que, aunque se enuncie para matrices cuadradas, es ampliable a matrices que no lo sean únicamente añadiendo filas o columnas de ceros hasta hacerlas cuadradas. En estos casos, se tendrá en cuenta que $(A|0)^{\dagger} = (\frac{A^{\dagger}}{0^*}) = (A^{\dagger}|0^*)$.

   Dada $A \in Mat_{n\times n}(\mathbb{C})$, el primer algoritmo sigue los siguientes pasos.

   a) Reducir A^* por filas a su forma de Hermite.

   b) Seleccionar las *columnas distinguidas* de A^*. Nombrar a dichas columnas $v_1, v_2; \cdots, v_r$ y utilizarlas como columnas únicas de una nueva matriz que recibirá el nombre de L.

   c) Formar la matriz $AL = A \cdot L$.

   d) Formar la matriz $I - H_{A^*}$. Seleccionar las columnas distintas de cero de dicha matriz y denotarlas por $w_1, w_2, \cdots, w_{n-r}$.

   e) Utilizar las columnas de AL y las w_i del paso anterior como columnas de una nueva matriz $M = (AL|w_1|w_2|\cdots|w_{n-r})$ y calcular M^{-1}.

   f) Usar las r primeras filas de M^{-1} en el mismo orden en el que aparecen, en una nueva matriz llamada R.

   g) Calcular $A^{\dagger}$ como $A^{\dagger} = L \cdot R$.

2. **Segundo algoritmo.** Sea $A \in Mat_{mxn}(\mathbb{C})$.

   a) Reducir A a su forma escalonada por filas E_A.

   b) Seleccionar las *columnas distinguidas* de A para usarlas como columnas de una nueva matriz B en el mismo orden en el que aparecen en A.

   c) Seleccionar las filas distintas de cero de E_A y utilizarlas como filas de una nueva matriz C en el mismo orden en el que aparecen en E_A.

   d) Calcular $(CC^*)^{-1}$ y $(B^*B)^{-1}$.

   e) Calcular $A^{\dagger}$ como $A^{\dagger} = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$

Se van a aplicar ahora lo dos algoritmos a la misma matriz, de manera que la pseudoinversa de Moore-Penrose obtenida, por ser única, deberá coincidir.

Ejemplo 3.3.3. ***Primer algoritmo.***

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

1. *La matriz conjugada transpuesta de A es la siguiente:*

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

   con la cual se realizan operaciones de filas para obtener su forma escalonada de Hermite:

   $$H_{A^*} = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2. *Las columnas distinguidas de A^* son la primera, tercera y cuarta (r=3), las cuales se seleccionan, se las nombra como v_1, v_2, v_3 y se usan para formar L.*

$$v_1 = \begin{bmatrix}1\\1\\0\\1\end{bmatrix} ; v_2 = \begin{bmatrix}3\\3\\2\\3\end{bmatrix} ; v_3 = \begin{bmatrix}4\\4\\1\\1\end{bmatrix} ; L = \begin{bmatrix}1 & 3 & 4\\1 & 3 & 4\\0 & 2 & 1\\1 & 3 & 1\end{bmatrix}$$

3. *Se calcula el producto de A por L:*

$$AL = \begin{bmatrix}3 & 9 & 9\\6 & 18 & 18\\9 & 31 & 29\\9 & 29 & 34\end{bmatrix}$$

4. *Tras realizar el cálculo de $I - H_{A^*}$ se tiene que la columna distinta de cero, denominada w_1, es la segunda:*

$$I - H_{A^*} = \begin{bmatrix}0 & -2 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\end{bmatrix} ; w_1 = \begin{bmatrix}-2\\1\\0\\0\end{bmatrix}$$

5. *Con lo que se tiene que M es de la forma*

$$M = \begin{bmatrix}3 & 9 & 9 & -2\\6 & 18 & 18 & 1\\9 & 31 & 29 & 0\\9 & 29 & 34 & 0\end{bmatrix} ; M^{-1} = \begin{bmatrix}\frac{71}{120} & \frac{71}{60} & \frac{-5}{8} & \frac{-1}{4}\\ \frac{-1}{8} & \frac{-1}{4} & \frac{7}{24} & \frac{-1}{12}\\ \frac{-1}{20} & \frac{-1}{10} & \frac{-1}{12} & \frac{1}{6}\\ \frac{-2}{5} & \frac{1}{5} & 0 & 0\end{bmatrix}$$

6. *Las r=3 primeras filas de M^{-1} forman la matriz R:*

$$R = \begin{bmatrix}\frac{71}{120} & \frac{71}{60} & \frac{-5}{8} & \frac{-1}{4}\\ \frac{-1}{8} & \frac{-1}{4} & \frac{7}{24} & \frac{-1}{12}\\ \frac{-1}{20} & \frac{-1}{10} & \frac{-1}{12} & \frac{1}{6}\end{bmatrix}$$

7. *Se concluye que $A^\dagger$ es de la forma:*

$$A^\dagger = LR = \begin{bmatrix}\frac{1}{60} & \frac{1}{30} & \frac{-1}{12} & \frac{1}{6}\\ \frac{1}{60} & \frac{1}{30} & \frac{-1}{12} & \frac{1}{6}\\ \frac{-3}{10} & \frac{-3}{5} & \frac{1}{2} & 0\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{-1}{3}\end{bmatrix}$$

Ejemplo 3.3.4. ***Segundo algoritmo***

Sea, de nuevo, $A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 2 & 0 & 2 \\ 3 & 3 & 2 & 3 \\ 4 & 4 & 1 & 1 \end{bmatrix}$

1. *De nuevo mediante el uso de operaciones básicas de filas se obtiene la forma escalonada por filas de la matriz* A*.*

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

2. *Las columnas distinguidas de* A *son la primera, la tercera y la cuarta, las cuales se usan para formar la matriz* B*.*

   $$B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 3 & 2 & 3 \\ 4 & 1 & 1 \end{bmatrix} \text{ ; } B^* = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix}$$

3. *Las filas que configuran la matriz* C *se corresponden con las filas distintas de cero de* E_A*, es decir, primera, segunda y tercera.*

   $$C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ ; } C^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4. *Se calculan* $(B^*B)^{-1}$ *y* $(CC^*)^{-1}$

   $$(CC^*)^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ ;}$$

   $$(B^*B)^{-1} = \begin{bmatrix} 30 & 10 & 18 \\ 10 & 5 & 7 \\ 18 & 7 & 15 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{13}{90} & \frac{-2}{15} & \frac{-1}{9} \\ \frac{-2}{15} & \frac{7}{10} & \frac{-1}{6} \\ \frac{-1}{9} & \frac{-1}{6} & \frac{5}{18} \end{bmatrix}$$

5. *Finalmente se aplica la fórmula para obtener* $A^\dagger$ *y se sustituyen las matrices obtenidas en pasos anteriores.*

   $$A^\dagger = C^*(CC^*)^{-1}(B^*B)^{-1}B^* = \begin{bmatrix} \frac{1}{60} & \frac{1}{30} & \frac{-1}{12} & \frac{1}{6} \\ \frac{1}{60} & \frac{1}{30} & \frac{-1}{12} & \frac{1}{6} \\ \frac{-3}{10} & \frac{-3}{5} & \frac{1}{2} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{-1}{3} \end{bmatrix}$$

Los lectores pueden comprobar que la matriz obtenida en los ejemplos anteriores verifica las propiedades de $A^\dagger$.

3.4. Inversa generalizada de un producto

El objetivo de esta sección es determinar en qué casos, o bajo qué condiciones, se cumple $(AB)^\dagger = B^\dagger A^\dagger$.

Para ello se va a hacer uso de dos teoremas. El primero de ellos ofrece una fórmula para obtener $(AB)^\dagger$ a partir de las proyecciones. El segundo detalla diferentes enunciados equivalentes. Al concluir ambas demostraciones, se llega a la conclusión de la condición que debe darse para que la inversa de Moore-Penrose del productor de dos matrices sea tal que $(AB)^\dagger = B^\dagger A^\dagger$.

Hay otras maneras de enfocar cuándo se verifica la igualdad, pero no se desarrollarán en este trabajo.

Teorema 3.7. *Si $A \in Mat_{m\times n}(\mathbb{C})$ y $B \in Mat_{n\times p}(\mathbb{C})$, entonces $(AB)^\dagger = (P_{R(A^*)}B)^\dagger(AP_{R(B)})^\dagger$.*

Demostración. La demostración de este teorema se divide en dos partes. Una primera consistente en, asumiendo que $(AB)^\dagger = (P_{R(A^*)}B)^\dagger(AP_{R(B)})^\dagger$ satisface la primera condición de la definición de Penrose, dar con una fórmula que se pueda verificar independientemente de dicha condición. La segunda determinar si la fórmula obtenida es reversible, en cuyo caso se habrá probado que la primera condición de la definición de Penrose se verifica.

Sea $X = (AB)^\dagger = (P_{R(A^*)}B)^\dagger(AP_{R(B)})^\dagger$ verificando la primera condición de la definición de Penrose, es decir, $ABXAB = AB$ o, equivalentemente

$$AB(A^\dagger AB)(ABB^\dagger)AB = AB \tag{1}$$

Multiplicando (1) por $A^\dagger$ a la izquierda y por $B^\dagger$ a la derecha, se tiene que

$$A^\dagger AB(A^\dagger AB)(ABB^\dagger)ABB^\dagger = A^\dagger ABB^\dagger \tag{2}$$

Por tanto, teniendo en cuenta que

$$A^\dagger AB(A^\dagger AB)^\dagger = P_{R(A^\dagger AB)} \text{ y}$$
$$(ABB^\dagger)^\dagger ABB^\dagger = P_{R((ABB^\dagger)^\dagger)} = P_{R((ABB^\dagger)^*)} = P_{R((BB^\dagger)^*A^*)} = P_{R(BB^\dagger A^*)}$$

además, se obtiene

$$P_{R(A^\dagger AB)}P_{R(BB^\dagger A^*)} = P_{R(A^*)}P_{R(B)}.$$

Equivalentemente,

$$P_{A^\dagger AR(B)}P_{BB^\dagger R(A^*)} = P_{R(A^*)}P_{R(B)}. \tag{3}$$

Para ver que es cierto, se toma $E_1 = P_{A^\dagger AR(B)}P_{BB^\dagger R(A^*)}$, $E_2 = P_{R(A^*)}P_{R(B)}$ y se prueba que, para todo $u \in \mathbb{C}^n$, $E_1u = E_2u$.

Si $u \in R(B)^\perp$, entonces $E_1u = E_2u = 0$. Si $u \in R(B)$, para saber cuánto vale E_1u, se necesita calcular $P_{BB^\dagger R(A^*)}u$. Como $BB^\dagger R(A^*)$ es un subespacio de $R(A)$, sea $u = u_1 \oplus u_2$ con $u_1 \in BB^\dagger R(A^*)$ y $u_2 \in [BB^\dagger R(A^*)]^\perp \cap R(B)$. Entonces

$$E_1u = P_{A^\dagger AR(B)}P_{BB^\dagger R(A^*)}u = \tag{4}$$

$$= P_{A^\dagger AR(B)}u_1 = AA^\dagger u_1. \tag{5}$$

Ahora, $E_2u = P_{R(A^*)}P_{R(B)}u = P_{R(A^*)}u = P_{R(A^\dagger)}u = A^\dagger Au$ que, por (3) se tiene que $A^\dagger Au = A^\dagger Au_1$, esto es, si $u_1 \in N(A) = R(A^*)^\perp$. Se va a ver que así es.

Supóngase que $v \in R(A^*)$. Por definición, $u_2 \perp BB^\dagger v$. Así,

$$0 = (BB^\dagger v, u_2) = (v, BB^\dagger u_2) = (v, u_2).$$

Por tanto, $u_2 \in R(A^*)^\perp$ como se quería ver.

Haciendo el proceso inverso, (3) es equivalente a (2) y, si (2) se multiplica por A a la izquierda y por B a la derecha se tiene:

$$AA^\dagger AB(A^\dagger AB)(ABB^\dagger)ABB^\dagger B = AA^\dagger ABB^\dagger B$$

que, por las definición de inversa de Moore-Penrose, es (1).

Ahora bien, teniendo en cuenta $ABXAB = AB$, para ver que X satisface la definición de inversa de Moore de $(AB)^\dagger$, se va a multiplicar a la derecha por $(AB)^\dagger$, con lo que

$$ABXAB(AB)^\dagger = AB(AB)^\dagger$$

o, equivalentemente

$$ABXP_{R(AB)} = P_{R(AB)}.$$

Pero $N(X) \subseteq N((ABB^\dagger)^\dagger) = R(ABB^\dagger)^\perp = R(AB)^\perp$.

Así, $XP_{R(AB)} = P_{R(AB)}$ y, por ello,

$$ABX = P_{R(AB)}. \tag{6}$$

Ahora, multiplicando $ABXAB = AB$ por $(AB)^\dagger$ a la izquierda, queda

$$(AB)^\dagger ABXAB = (AB)^\dagger AB$$

o, equivalentemente,

$$P_{R((AB)^\dagger)}XAB = P_{R((AB)^\dagger)}.$$

Pero $N(X) \subseteq R((A^\dagger AB)^\dagger) = R((A^\dagger AB)^*) = R(B^*A^*(A^*)^\dagger) = R(B^*A^*)$, por tanto

$$P_{R(B^*A^*)}X = X$$

$$XAB = P_{R((AB)^\dagger}. \tag{7}$$

Por (6) y (7) se concluye que la definición se cumple. □

Como se enuncia en [4, Corollary 1.4.1.].

Corolario 3.4.1. *Sea $A \in Mat_{m\times n}(\mathbb{C})$ y $B \in Mat_{n\times p}(\mathbb{C})$*

1. *Si $rg(A) = n$, entonces $(AB)^\dagger = B^\dagger(AP_{R(B)})^\dagger$.*

2. *Si $rg(B) = n$, entonces $(AB)^\dagger = (P_{R(A^*)}B)^\dagger A^\dagger$.*

Como se enuncia en [4, Corollary 1.4.2.].

Corolario 3.4.2. *Sea $A \in Mat_{m\times n}(\mathbb{C})$ y $B \in Mat_{n\times p}(\mathbb{C})$, y $rg(A) = rg(B) = n$. Entonces $(AB)^\dagger = B^\dagger A^\dagger$ y $A^\dagger = A^*(AA^*)^{-1}$ mientras que $B^\dagger = (B^*B)^{-1}B^*$.*

Previamente al enunciado y demostración del segundo y último teorema de esta sección, se van a detallar varios razonamientos que serán necesarios para la demostración de dicho teorema.

Sean $A \in Mat_{m\times n}(\mathbb{C}), B \in Mat_{n\times p}(\mathbb{C})$ y supóngase que

$$(AB)^\dagger = B^\dagger A^\dagger. \tag{8}$$

Del Teorema 3.6 se tiene que

$$(AB)^\dagger = (P_{R(A^*)}B)^\dagger (AP_{R(B)})^\dagger = (P_{R(A^\dagger)}B)^\dagger (AP_{R(B)})^\dagger = (A^\dagger AB)^\dagger (ABB^\dagger)^\dagger$$
$$\text{luego } B^\dagger A^\dagger = (A^\dagger AB)^\dagger (ABB^\dagger)^\dagger.$$

Multiplicando a la izquierda por $A^\dagger AB$ y a la derecha por $ABB^\dagger$ se obtiene

$$A^\dagger ABB^\dagger A^\dagger ABB^\dagger = A^\dagger AB(A^\dagger AB)^\dagger (ABB^\dagger)^\dagger ABB^\dagger$$

$$P_{R(A^\dagger)}P_{R(B)}P_{R(A^\dagger)}P_{R(B)} = P_{R(A^\dagger AB)}P_{R((ABB^\dagger)^\dagger)}$$

$$(P_{R(A^\dagger)}P_{R(B)})^2 = P_{R(A^\dagger AB)}P_{R((ABB^\dagger)^*)}$$

$$(P_{R(A^*)}P_{R(B)})^2 = P_{R(A^\dagger AB)}P_{R(BB^\dagger A^*)}. \tag{9}$$

Por (3), (9) puede escribirse como $(P_{R(A^*)}P_{R(B)})^2 = P_{R(A^*)}P_{R(B)}$ y por tanto, $P_{R(A^*)}P_{R(B)}$ es una proyección.

Ahora bien, el producto de dos proyecciones hermitianas es una proyección si y solo si las dos poryecciones hermitianas conmutan. Así, teniendo en cuenta $[X,Y] = XY - YX$:

$$[P_{R(A^*)}, P_{R(B)}] = 0 \tag{10}$$

o, equivalentemente,

$$[A^\dagger A, BB^\dagger] = 0 \tag{11}$$

si $(AB)^\dagger = B^\dagger A^\dagger$.

Como $ABB^\dagger A^\dagger$ es hermitiana, se tiene que $A^*(ABB^\dagger A^\dagger)A$ también lo será. Así,

$$A^*(ABB^\dagger)A^\dagger A = A^\dagger ABB^\dagger A^* A \tag{12}$$

lo que se traduce en $A^*ABB^\dagger = BB^\dagger A^*A$, o bien

$$[A^*A, BB^\dagger] = 0 \tag{13}$$

donde se ha usado (11) y que $A^\dagger AA^* = A^*$.

Ahora, suponiendo de nuevo que $(AB)^\dagger = B^\dagger A^\dagger$ se cumple, $BB^\dagger A^\dagger ABB^*$ será hermitian y, de manera análoga al razonamiento ya hecho

$$[BB^*, A^\dagger A] = 0. \tag{14}$$

Una vez aclarados estos pasos, se tiene el siguiente teorema:

Teorema 3.8. *Sean $A \in Mat_{m\times n}(\mathbb{C})$ y $B \in Mat_{n\times p}(\mathbb{C})$. Los siguientes enunciados son equivalentes:*

1. $(AB)^\dagger = B^\dagger A^\dagger$

2. $BB^*A^\dagger A$ *y* $A^*ABB^\dagger$ *son hermitianas.*

3. $R(A^*)$ *es un subespacio invariante de* BB^* *y* $R(B)$ *es un subespacio invariante de* A^*A.

4. $P_{N(A)}BB^*P_{N(A)}^\perp = 0$ *y* $P_{N(B^*)}A^*AP_{N(B^*)}^\perp = 0$

5. $A^\dagger ABB^*A^* = BB^*A^*$ *y* $BB^\dagger A^*AB = A^*AB$.

Demostración. El enunciado (2) es equivalente a las ecuaciones (13) y (14), por lo que (1) implica (3).

Se va a demostrar en primer lugar que (2) y (5) son equivalentes. Como BB^* y A^*A son hermitianas, todos sus espacios invariantes son reducibles. Así, (2) y (3) son equivalentes. Ahora bien, si C es una matriz y M un subespacio, entonces

$$CP_M^\perp = (I - P_M)CP_M^\perp + P_MCP_M^\perp = P_M^\perp CP_M^\perp + P_MCP_M^\perp.$$

Esto quiere decir que $M^\perp$ es un subespacio invariante si y solo si $P_MAP_M^\perp = 0$. Así, (3) y (4) son equivalentes.

Se prueba ahora que (5) es equivalente a (2). Supóngase que $BB^*A^\dagger A$ es hermitiana. Entonces $BB^*A^\dagger A = A^\dagger ABB^*$. Así, $BB^*(A^\dagger A)A^* = A^\dagger ABB^*A^*$ o

$$BB^*A^* = A^\dagger ABB^*A^* \tag{15}$$

De manera similar, si $A^*ABB^\dagger$ es hermitiana, entonces

$$BB^\dagger A^*AB = A^*AB \tag{16}$$

con lo cual (2) y (5) son equivalentes.

Asumiendo ahora que (15) y (16) se cumplen, multiplicando (16) por $B^\dagger$ a la derecha y por $(A^*)^\dagger$ a la derecha se obtiene la demostración de (3) que es equivalente a (2), luego (2)-(5) son todas equivalentes.

Para probar (1) se tiene en cuenta que $B^\dagger A^\dagger = B^\dagger(BB^\dagger))(A^\dagger A)A^\dagger = B^\dagger(A^\dagger A)(BB^\dagger)A^\dagger$. Como ya se ha probado, $(AB)^\dagger = (A^\dagger AB)^\dagger(ABB^\dagger)^\dagger$, luego este teorema quedaría probado si se cumplen:

$$(A^\dagger AB)^\dagger = B^\dagger(A^\dagger A) \tag{17}$$

y

$$(ABB^\dagger)^\dagger = BB^\dagger A^\dagger \tag{18}$$

parte que se encuentra detallada en [4, Theoreme 1.4.2.]. □

4. Aplicaciones

En esta sección, habiendo sido detalladas las nociones necesarias de la pseudoinversa de Moore-Penrose, así como algunas de sus propiedades básicas y métodos para su obtención, se va introducir su aplicación en la resolución de sistemas de ecuaciones lineales por la aproximación de mínimos cuadrados.

4.1. Aproximación por mínimos cuadrados

El problema a resolver será el de la resolución de del sistema de ecuaciones lineales

$$Ax = b \tag{19}$$

con $A \in Mat_{m\times n}(\mathbb{C}), x \in \mathbb{C}^n, b \in \mathbb{C}^m$.

Si $A \in Mat_{n\times n}(\mathbb{C})$ y es invertible, entonces las soluciones de (19) son de la forma $x = A^{-1}b$. Sin embargo, cuando A no es una matriz cuadrada invertible, la resolución de (19) se complica.

Las inversas generalizadas, y en concreto la que es objeto de estudio en este trabajo, la inversa de Moore-Penrose, permiten analizar si el sistema de ecuaciones lineales no homogéneo (19) tiene solución. En caso afirmativo, posibilitan obtener la expresión de dicha solución y, en caso negativo, proporcionan una solución aproximada.

4.1.1. Introducción

Antes de introducir la aproximación a las soluciones que facilita esta inversa, se va a dar la definición de *norma euclídea de un vector*, así como una de las propiedades que esta tiene y que se usará a lo largo de la sección.

Definición 4.1.1. *Se define la **norma euclídea** de un vector $w = [w_1, \ldots, w_p]^* \in \mathbb{C}^p$ y se denota por $\|\mathbf{w}\|$ a*

$$\|\mathbf{w}\| = (\textstyle\sum_{i=1}^p |\mathbf{w}|^2)^{\frac{1}{2}} = (w^* w)^{\frac{1}{2}} \ .$$

Lema 4.1. *Si $u, v \in \mathbb{C}^p$ y el producto escalar de u, v es $(u, v) = 0$, entonces $\|\mathbf{u}+\mathbf{v}\|^2 = |\mathbf{u}\|^2 + |\mathbf{v}\|^2$.*

Demostración. $\|\mathbf{u}+\mathbf{v}\|^2 = (u+v, u+v) = (u,u)+(u,v)+(v,v) = (u,u)+(v,v) = |\mathbf{u}\|^2 + |\mathbf{v}\|^2$ □

Definición 4.1.2. *Sea $A \in Mat_{m\times n}(\mathbb{C})$ y $b \in \mathbb{C}^m$. Entonces, un vector $u \in \mathbb{C}^n$ se dice que es una **solución de mínimos cuadrados** para $Ax = b$ si $\|\mathbf{Au}-\mathbf{b}\| \leq \|\mathbf{Av}-\mathbf{b}\|$ para todo $v \in \mathbb{C}^n$. Un vector $u \in \mathbb{C}^n$ se dice que es una **mínima solución de mínimos cuadrados** para $Ax = b$ si u es solución de mínimos cuadrados para $Ax = b$ y $\|\mathbf{u}\| < \|\mathbf{w}\|$ para cualquier otra solución de mínimos cuadrados w.*

Teorema 4.1. *Sea $A \in Mat_{m\times n}$ y $b \in \mathbb{C}^m$. Entonces $A^\dagger b$ es la solución de mínimos cuadrados de $Ax = b$.*

Demostración. Se tiene que $\|\mathbf{Ax}-\mathbf{b}\|^2 = \|(\mathbf{AA^\dagger Ax} - \mathbf{AA^\dagger b}) \oplus -\mathbf{P}^\perp_{\mathbf{R(A)}}\mathbf{b}\|^2 =$
$= \|(\mathbf{Ax} - \mathbf{AA^\dagger b}) \oplus -(\mathbf{I} - \mathbf{AA^\dagger})\mathbf{b}\|^2 = \|\mathbf{Ax} - \mathbf{AA^\dagger b}\|^2 + \|(\mathbf{I} - \mathbf{AA^\dagger})\mathbf{b}\|^2$.

Por lo que x será solución de mínimos cuadrados si y solo si x es solución del sistema consistente, $Ax = AA^\dagger b$. Por otra parte, las soluciones de dicho sistema son de la forma

$$x = A^\dagger(AA^\dagger b) \oplus (I - A^\dagger A)h = A^\dagger b \oplus (I - A^\dagger A)h.$$

Como $\|\mathbf{r}\|^2 = \|\mathbf{r}\|^2$, se ve que existe una mínima solución de mínimos cuadrados $x = A^\dagger b$.

□

Corolario 4.1.1. *Sea M es un subespacio de $\mathbb{C}^p$ y P_M la proyección ortogonal de $\mathbb{C}^n$ sobre M. Si $b \in \mathbb{C}$, entonces $P_M b$ es el único vector más cercano a b en M con respecto a la norma euclídea.*

Teorema 4.2. *Sea $A \in Mat_{m\times n}(\mathbb{C})$ y $b \in \mathbb{C}^m$. Entonces las siguientes afirmaciones son equivalentes:*

1. *u es una solución de mínimos cuadrados de $Ax = b$.*
2. *u es una solución de $Ax = AA^\dagger b$.*
3. *u es una solución de $A^*Ax = A^*b$.*
4. *u es de la forma $A^\dagger b + h$ donde $h \in N(A)$.*

Demostración. De la demostración del Teorema 4.1, se sabe que 1, 2 y 4 son equivalentes. Se detalla a continuación la equivalencia entre 1, 2 y 3.

La equivalencia entre 1 y 3 se deduce de que, si u es solución de $Ax = b$, entonces $Au = b$. Multiplicando a la izquierda por A^* se tiene:

$$A^*Au = A^*b$$

luego u es solución de $A^*Ax = A^*b$.

La equivalencia entre 3 y 2, por otra parte, se demuestra teniendo en cuenta que si u es solución de $A^*Ax = A^*b$, entonces $A^*Au = A^*b$. Multiplicando esta última igualdad a la izquierda por $(A^*)^\dagger$ se tiene:

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

luego u es solución de $Ax = AA^\dagger b$. □

Para terminar la sección se tiene en cuenta que el sistema de ecuaciones 3, es decir, $A^*Ax = A^*b$, recibe el nombre de **ecuaciones normales** cuya importancia se resalta en áreas de la estadística y al que se hará referencia en posteriores apartados de este trabajo.

4.1.2. Hipótesis lineal, estimación de los parámetros y bondad de ajuste

A continuación se va a explicar en qué consiste el planteamiento de una hipótesis lineal y se va hacer uso de la noción de *variables linealmente relacionadas* que dan lugar a una ecuación con *parámetros* desconocidos. Posteriormente, se detallará cómo encontrar una estimación de dichos parámetros y, una vez obtenida, determinar cómo de certera es la modelación del problema inicial si se hace uso de dichas estimaciones.

Se considera un problema o experimento con dos variables x e y modelado por una función lineal $f(x) = y = \beta x$, donde β es un parámetro desconocido. Dicho parámetro puede despejarse como $\beta = \frac{y}{x}$ de manera que, para diferentes mediciones de las variables x e y, se obtienen resultados distintos de β.

Ahora bien, si se desea estimar el valor que toma β, es más realista considerar, para cada valor fijo x, diferentes valores y_i de y que satisfagan la ecuación $y_i = \beta x + e_i$ donde e_i es el valor del error que en promedio a la larga se hace 0.

Supóngase ahora que, en dicho experimento, los diferentes resultados de β no solo solo varían por el error e_i, con lo cual, la función $f(x)$ realmente es de la forma $y = \beta x + g(u_1, u_2, \ldots, u_n)$, donde g es una función desconocida. Análogamente al razonamiento anterior, para cada valor fijo de x, la función g toma valores cuyo promedio a la larga es 0. Este tipo de error se llama **error funcional**.

Estas consideraciones dan lugar a la siguiente definición.

Definición 4.1.3. *Cuando se plantea la hipótesis de que la variable y está **linealmente relacionada** con $x_1, x_2, \ldots, x_n$, quiere decir que, para cada conjunto de valores $p_i = (x_1, x_2, \ldots, x_n)$ de $x_1, x_2, \ldots, x_n$, las observaciones y_i de y en p_i se pueden expresar como:*

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}$$

donde

1. *$\beta_1, \beta_2, \ldots, \beta_n$ son constantes desconocidas que reciben el nombre de **parámetros**.*

2. *e_{y_i} es el valor que toma una función de valor real desconocida e_{p_i} tal que e_{p_i} tiene la propiedad de que sus valores tienen promedio cero en todas las observaciones posibles y_i en p_i.*

Una vez se han aclarado estos conceptos, se va a dar solución al problema de estimar los parámetros β_i en dos hipótesis y para ambas se dará la definición de *matriz de diseño*.

Este apartado se concluye con el enunciado de dos teoremas, el primero sobre la solución por mínimos cuadrados, así como su expresión, y el segundo detallando la condición necesaria y suficiente para que, en una situación similar, la estimación de una combinación lineal tenga un único valor.

Las dos hipótesis mencionadas, en las cuales $x_1, x_2, \ldots, x_n$ están linealmente relacionados con y, son:

1. Aquella en la cual aparece el término β_0 y recibe el nombre de **hipótesis de intersección**:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}, \tag{20}$$

2. Aquella en la que no aparece β_0 y que se denomina **hipótesis de no intersección**.

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}, \tag{21}$$

Para la estimación de los parámetros β_i, en ambas situaciones, se procede del siguiente modo: se considera un primer conjunto de valores para las x a los que se denota por $p_1 = (x_{11}, x_{12}, \ldots, x_{1n})$. Después se calcula el valor de la y en p_1 y se le llama y_1. Este procedimiento se repite con m ($m > n$) conjuntos de valores para las x, de manera que se obtienen m valores $y_1, y_2, \ldots, y_m$ para y.

Si se colocan los valores de las x es las filas de una matriz, que recibe el nombre de **matriz de diseño** y se denotará por X para (21) y X_1 para (20), se tendrá:

1. Para el caso de la ecuación (21):

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{bmatrix}$$

donde, si se denota por $y = y(y_1, \ldots, y_m)^T$ al vector formado por los valores que toma y, $b = (\beta_1, \ldots, \beta_n)^T$ el vector de los parámetros desconocidos y $e_y = (e_{y_1}, \ldots, y_{y_m})^T$ el vector desconocido de los errores, se puede escribir (21) como:

$$y = Xb + e_y \tag{22}$$

2. Para el caso de la ecuación (20):

$$X_1 = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}_{m\times(n+1)} = [j|X],\ j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m\times 1}$$

En este caso, se tendrá que el vector de parámetros es $b_1 = [\beta_0, \beta_1, \ldots, \beta_n]^T = [\beta_0|b^T]^T$, donde b es el obtenido para el apartado anterior.

Además, la ecuación asociada a (20) será:

$$y = X_1 b_1 + e_y \tag{23}$$

Para ambas hipótesis, tanto en el caso en el que β_0 está presente, como para el que no, se ha considerado que y está linealmente relacionada con $x_1, x_2, \ldots, x_n$, lo que quiere decir que para el punto $p_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, el 'valor esperado' $E(y_i)$ de la observación y_i en p_i, y no y_i, satisface la ecuación

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in}.$$

A la hora de obtener las estimaciones $\hat{b}$ de b en la hipótesis asociada a (22), una de las formas más efectivas es imponer que $\hat{b}$ sea un vector tal que $X\hat{b}$ esté tan próximo a y como sea posible, o equivalentemente, que e_y esté tan cerca de 0 como sea posible. Esto se traduce en que $\hat{b}$ debe ser una solución por mínimos cuadrados de $Xb = y$.

A esta imposición, debe sumarse la restricción de que $\|\hat{\mathbf{b}}\|$ debe ser la mínima estimación por mínimos cuadrados, de manera que la estimación buscada sea $\hat{b} = X^\dagger y$.

Para cada estimación por mínimos cuadrados $\hat{b}$ de b, el vector $\hat{y}$ de la forma $X\hat{b} = \hat{y}$ es una estimación para el vector de 'valor esperado' $\widehat{E(y)}$, donde

$$E(y) = (E(y_1), \ldots, E(y_n))^T$$

Teorema 4.3. *El vector $\hat{y} = X\hat{b}$ es el mismo para todas las soluciones por mínimos cuadrados $\hat{b}$ de $X\hat{b} = \hat{y}$. Además, para todas las soluciones por mínimos cuadrados $\hat{b}$ de $X\hat{b} = y$, que por el Teorema 4.1.1 se sabe que también es solución de $X\hat{b} = XX^\dagger y$, se verifica que $\hat{y} = X\hat{b} = P_{R(X)}y = XX^\dagger y$ y $r = y - \hat{y} = P^\perp_{R(X)}y = (I - XX^\dagger)y$.*

Demostración. Es una consecuencia trivial del Teorema 4.2. □

Se plantea ahora una nueva situación en la que se desea predecir un valor para la combinación lineal de las β_i en las bases de las observaciones hechas en los puntos $p_1, \ldots, p_n$:

$$y(c^*) = c_1\beta_1 + c_2\beta_2, \cdots, c_n\beta_n = c^*b$$

donde $c^* = (c_1, \ldots, c_n)$.

Se busca predecir el valor $\widehat{y(c^*)}$ para y en el punto C^*. Si se usa $\widehat{y(c^*)} = \widehat{c^*b} = c^*\hat{b}$ es posible tener infinitas estimaciones $\hat{b}$, por lo que $\widehat{y(c^*)}$ podría tomar infinitos valores. Sin embargo, hay algunos casos en los que $c^*\hat{b}$ es invariante entre todas las estimaciones por mínimos cuadrados $\hat{b}$, con lo que $\widehat{y(c^*)}$ tendría un único valor.

Estas ideas se recogen en el siguiente teorema:

Teorema 4.4. *Sea $c \in \mathbb{C}^n$. La transformación linal de la forma $c^*\hat{b}$ es invariante entre todas las soluciones por mínimos cuadrados si y solo si $c \in R(X^*)$; en cuyo caso, $c^*\hat{b} = c^*X^\dagger y$.*

Demostración. Si $\hat{b}$ es una solución por mínimos cuadrados de $X\hat{b} = y$, entonces $\hat{b}$ es de la forma $\hat{b} = X^\dagger y + h$, con $h \in N(X)$.

Multiplicando por c^* a la izquierda se tiene: $c^*\hat{b} = X^\dagger y + h$, con $h \in N(X)$, que es invariante si y solo si $c^*h = 0$ para todo $h \in N(X)$. Es decir, $c \in R(X^*) = N(X)^\perp$. □

A lo largo de esta sección se ha relatado en qué consisten las hipótesis de variables linealmente relacionadas. Si se quiere ver gráficamente, en el supuesto de que se tengan únicamente dos variables, y y x, quiere decir que se asume la existencia de una recta $f(x) = \beta_0 + \beta_1 x$ tal que cada punto $(x_i, E(y_i))$ se encuentra en dicha línea recta, donde $E(y_i)$ es el 'valor esperado' para la observación y_i.

Si en lugar de dos, se tienen n variables independientes, la hipótesis de forma gráfica que se hace es asumir la existencia de una superficie en $\mathbb{C}^{n+1}$, lo que se traduce en un subespacio, que pasa por los puntos $(p_i, E(y_i))$ (donde se ha utilizado la notación descrita en este mismo apartado para denotar a los y_i y p_i valores de las x). Esta superficie recibe el nombre de **plano**.

Rescatando las dos hipótesis con las que se está trabajando, de intersección (cuando β_0 está presente) y la de no intersección (cuando β_0 no está presente), surje la duda de cómo de buena son las estimaciones hechas $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$ para los $\beta_0, \beta_1, \ldots, \beta_n$ (o $\hat{\beta}_1, \ldots, \hat{\beta}_n$ para los $\beta_1, \ldots, \beta_n$ en el caso de la segunda hipótesis).

Si se considera el caso de no intersección, en el cual el conjunto de valores p_i de la las $x's$ daba lugar a los corrrespondientes valores y_i de y, se tenía la ecuación

$$y = Xb + e_y \tag{24}$$

Además, usando los $\hat{\beta}_i$ mencionados para, dado un punto $c^* = [c_1, \ldots, c_n]$, se podía predecir un valor $y(\hat{c^*})$ con lo que se obtenía la **ecuación de estimación**

$$f(c_1, \ldots, c_n) = \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \ldots + \hat{\beta}_n c_n \tag{25}$$

Una medida efectiva de la bondad de ajuste de la estimación hecha es considerar el conjunto de observaciones que se han usado para calcular X y medir 'cómo de cerca' está el vector y del estimado $\hat{y}$. Es decir, cómo de cerca pasa el plano definido por (25) por los puntos (p_i, y_i) en $\mathbb{C}^{n+1}$.

Esto se traduce en $\|\mathbf{y}-\hat{\mathbf{y}}\|$, que, por el Teorema 4.3, se sabe que $\hat{y}$ entre todas las estimaciones de mínimos cuadrados $\hat{b}$ y que $\|\mathbf{y}-\hat{\mathbf{y}}\| = \|\mathbf{r}\| = \|(\mathbf{I}-\mathbf{X}\mathbf{X}^{\dagger})\mathbf{y}\|$.

Observación 4.1.1. *No debe cometerse el error de considerar el error absoluto en lugar del relativo.*

Haciendo una representación gráfica de esta idea, si se considera el ángulo θ formado por $\hat{y} = P_{R(X)}y \in R(X)$ e y, puede tomarse $|\mathbf{cos}\theta|$ para medir la cercanía entre los vectores dados, de manera que

$$|\mathbf{cos}\theta| = \frac{\|\hat{\mathbf{y}}\|}{\|\mathbf{y}\|} = \frac{\|\mathbf{X}\mathbf{X}^{\dagger}\mathbf{y}\|}{\|\mathbf{y}\|}$$

Si se descompone y como $y = \hat{y} + r$, donde $\hat{y} \in R(X)$ y $r \in R(X)^{\perp}$, y se denota R al $cos\theta$, de manera que $0 \leq |\mathbf{R}| \leq 1$, la forma de medir la bondad de ajuste sería:

1. Si $|\mathbf{R}| = 1$, entonces $y \in R(X), r = 0, y = \hat{y}$ y el plano definido por la ecuación $f(x_1,\ldots,x_n) = \hat{\beta}_1 x_1 + \ldots \hat{\beta}_n x_n$ pasa por los puntos (p_i, y_i) y por tanto, se tiene una estimación exacta.

2. Si $|\mathbf{R}| = 0$, entonces $y \perp R(X), y = r, \hat{y} = 0$, con lo que y está muy alejado de $P_{R(X)} = \hat{y}$ y, por tanto, se tiene una mala estimación.

En la práctica, suele usarse $R^2 = cos^2\theta = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2}$ en lugar de $|\mathbf{R}|$

Una forma conocida de expresar R^2, cuando todos los números son reales, es

$$R^2 = \frac{(\sum_{i=1}^{m} y_i\hat{y}_i)^2}{(\sum_{i=1}^{m} y_i^2)(\sum_{i=1}^{m} \hat{y}_i{}^2)}$$

donde y_i denota la entrada i-ésima de $XX^{\dagger}y = \hat{y}$ e y_i es la entrada i-ésima de y.

Esto se deduce de que $\|\hat{\mathbf{y}}\|^2 = [y^*XX^{\dagger}y] = [y^*\hat{y}] = \sum_{i=1}^{m} y_i\hat{y}_i$.

Con lo cual, $R^2 = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \frac{\|\hat{\mathbf{y}}\|^4}{\|\mathbf{y}\|^2\|\hat{\mathbf{y}}\|^2} = \dfrac{(\sum_{i=1}^{m} y_i\hat{y}_i)^2}{(\sum_{i=1}^{m} y_i^2)(\sum_{i=1}^{m} \hat{y}_i{}^2)}$.

Observación 4.1.2. *En ámbitos estadísticos, R recibe el nombre de coeficiente de correlación del producto-momento entre las observaciones $y_i's$ y las predicciones $\widehat{y_i's}$ y R^2 es de coeficiente de determinación.*

Para el caso de las hipótesis en las que aparece β_0, se necesita el siguiente teorema para obtener la bondad de ajuste.

Teorema 4.5. *Sea* $X_1 = [j|X] \in Mat_{m\times(n+1)}$.

El vector $\hat{b} = [\hat{\beta}_0, \ldots, \hat{\beta}_n]^T = [\hat{\beta}_0|\hat{b}^T]^T$, $b^T = [\hat{\beta}_1, , , , \ldots, \hat{\beta}_n]$, *es una solución de mínimos cuadrados de* $X_1b_1 = y$ *si y solo si*

$$\hat{\beta}_0 = \frac{1}{m}j^*(y - X\hat{b}) \tag{26}$$

y $\hat{b}$ *es una solución de mínimos cuadrados de*

$$(I - \frac{1}{m}J)Xb = (I - \frac{1}{m}J)y \tag{27}$$

donde $J = jj^*$ *es una matriz de unos.*

Demostración. Se empezará suponiendo que $\hat{\beta}_0$ cumple la ecuación (24) y que $\hat{b}$ es la solución de mínimos cuadrados de (25) y se usará las ecuaciones normales del Teorema 4.2 para demostrar que $\hat{b_1} = [\hat{\beta}_0, \hat{b}^*]^*$ es solución de mínimos cuadrados de $X_1b_1 = y$. Es decir, se probará que $X^*X_1b_1 = X^*y$.

Por una parte se tiene:

$$X_1^*X_1 = [j^*|X^*][j|X] = \begin{bmatrix} j^*j & \vdots & j^*X \\ \ldots & \vdots & \ldots \\ X^*j & \vdots & X^*X \end{bmatrix} = \begin{bmatrix} m & \vdots & j^*X \\ \ldots & \vdots & \ldots \\ X^*j & \vdots & X^*X \end{bmatrix}$$

donde se ha usado que $j^*j = [1, \ldots, 1]^T_{1\times m}[1, \ldots, 1]_{m\times 1} = \sum_{i=1}^m i = m$.

Por lo tanto,

$$X_1^*X_1\hat{b_1} = X_1^*X_1[\hat{\beta}_0, \ldots, \hat{\beta}_n]^T = \begin{bmatrix} m & \vdots & j^*X \\ \ldots & \vdots & \ldots \\ X^*j & \vdots & X^*X \end{bmatrix} [\tfrac{1}{m}j^*(y - X\hat{b}), \hat{b}]^T =$$

$$\begin{bmatrix} j^*(y - X\hat{b}) + j^*X\hat{b} \\ \ldots \\ \frac{1}{m}X^*jj^*(y - X\hat{b}) + X^*X\hat{b} \end{bmatrix} = \begin{bmatrix} j^*y \\ \ldots \\ \frac{1}{m}X^*Jy - \frac{1}{m}X^*JX\hat{b} + X^*X\hat{b} \end{bmatrix} \tag{28}$$

Ahora, como $\hat{b}$ es solución de mínimos cuadrados de (25), entonces $\hat{b}$ será solución de

$$((I - \tfrac{1}{m}J)X)^*(I - \tfrac{1}{m}J)Xb = ((I - \tfrac{1}{m}J)X)^*(I - \tfrac{1}{m}J)y$$

o equivalentemente:

$$X^*(I - \frac{1}{m}J)^*(I - \frac{1}{m}J)Xb = X^*(I - \frac{1}{m}J)^*(I - \frac{1}{m}J)y \tag{29}$$

donde se ha aplicado el Teorema 4.2.

Como $(I - \frac{1}{m}J) = (I - \frac{1}{m}J)^2 = (I - \frac{1}{m}J)^*$ (es decir, $(I - \frac{1}{m}J)$ es una proyección ortonogonal sobre $N(J)$), la ecuación (25) es equivalente a

$$X^*(I - \tfrac{1}{m}J)X\hat{b} = X^*(I - \tfrac{1}{m}J)y$$

$$(X^* - \tfrac{1}{m}X^*J)X\hat{b} = (X^* - \tfrac{1}{m}X^*J)y$$

$$X^*X\hat{b} - \frac{1}{m}X^*JX\hat{b} = X^*y - \frac{1}{m}X^*Jy$$

o bien

$$(12)\frac{1}{m}X^*Jy - \frac{1}{m}X^*JX\hat{b} + X^*X\hat{b} = X^*y \tag{30}$$

Sustiuyendo en (26) se tiene que $X_1^*X_1\hat{b_1} = \begin{bmatrix} j^*y \\ \dots \\ X^*y \end{bmatrix} = X_1^*y$, lo que prueba que $\hat{b}$ es una solución de mínimos cuadrados de $X_1\hat{b_1} = y$.

Recíprocamente, se supone que $\hat{b_1}$ es una solución de mínimos cuadrados de $X_1\hat{b_1} = y$. De nuevo por el Teorema 4.2, $\hat{b}$ verifica las ecuaciones normales $X_1^*X_1\hat{b_1} = X_1^*y$.

Se sabe que $\hat{b_1} = [\hat{\beta}_0, \hat{b}]^T$ y que $X_1^* = [j^*|X^*]^T$, además, como ya se probó $X_1^*X_1 = \begin{bmatrix} m & \vdots & j^*X \\ \dots & \vdots & \dots \\ X^*j & \vdots & X^*X \end{bmatrix}$, luego $\hat{\beta}_0$ y $\hat{b}$ deben satisfacer

$$\begin{bmatrix} m & \vdots & j^*X \\ \dots & \vdots & \dots \\ X^*j & \vdots & X^*X \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \dots \\ \hat{b} \end{bmatrix} = \begin{bmatrix} j^*y \\ \dots \\ X^*y \end{bmatrix}$$

De donde se obtienen las ecuaciones

$$m\hat{\beta}_0 + j^*X\hat{b} = j^*y \tag{31}$$

$$X^*j\hat{\beta}_0 + X^*X\hat{b} = X^*y \tag{32}$$

De la ecuación (29) se puede despejar $\hat{\beta}_0$ como $\hat{\beta}_0 = \frac{1}{m}j^*(y - X\hat{b})$ que es la ecuación(24) que se busaba demostrar.

Susituyendo ahora en (30) se tiene $\frac{1}{m}X^*jj^*(y - X\hat{b}) + X^*X\hat{b} = X^*y$ o equivalentemente, $\frac{1}{m}X^*Jy - \frac{1}{m}X^*JX\hat{b}) + X^*X\hat{b} = X^*y$ que es la ecuación (28). Por lo tanto, $\hat{b}$ satisface (27), por lo que $\hat{b}$ es una solución de mínimos cuadrados de $(I - \frac{1}{m}J)X\hat{b} = (I - \frac{1}{m}J)y$.

Por todo ello el teorema queda probado.

□

Las ideas hasta ahora expuestas hacen referencia a la hipótesis de no intersección. Tanto para esta, como para la de intersección, se tiene en cuenta el siguiente teorema.

Teorema 4.6. *Para la hipótesis de no intersección*

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}$$

la bondad de ajuste viene dada por la expresión:

$$R^2 = \frac{\|\mathbf{X}\mathbf{X}^\dagger\mathbf{y}\|^2}{\|\mathbf{y}\|^2} = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \frac{(\sum_{i=1}^{m} y_i\hat{y}_i)^2}{(\sum_{i=1}^{m} y_i^2)(\sum_{i=1}^{m} \hat{y_i}^2)}.$$

Para la hipótesis de intersección

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}$$

la bondad de ajuste viene dada por la expresión:

$$R^2 = \frac{\|\mathbf{X_M X_M^\dagger y_M}\|^2}{\|\mathbf{y_M}\|^2} = \frac{\|\mathbf{y\hat{_M}}\|^2}{\|\mathbf{y_M}\|^2} = \frac{(\sum_{i=1}^{m}[y_i - \hat{y}][\hat{y}_i - \hat{y}])^2}{(\sum_{i=1}^{m}[y_i - \hat{y}]^2)(\sum_{i=1}^{m}[\hat{y}_i - \hat{y}]^2)}.$$

donde $X_M = (I - \frac{1}{m}J)X$, $y_M = (I - \frac{1}{m}J)y$ *y las* $\hat{y}_i$ *son las entradas i-ésimas de* $\hat{y} = X_1 X_1^\dagger y = (X_M X_M^\dagger + \frac{1}{m}J)y$. *Además se tiene que* $X_1 = [j|X]$ *y* $J = jj^*, J = [1, \ldots, 1]^*$.

En cada caso, $0 \leq R^2 \leq 1$ *y no tiene unidades. Cuando* $R^2 = 1$*, la estimación es precisa, y cuando* $R = 0$ *la estimación no es precisa.*

Demostración. El primer caso se supone demostrado con las ideas que se han detallado previamente.

Para el segundo, se considera $\tilde{x_j} = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$ y $\tilde{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$. Es decir, $\tilde{x_j}$ es la media de los elementos de la columna j-ésima de X e $\tilde{y}$ es la media de las observaciones y_i para y.

De esta manera se tiene que X_M y y_M son las matrices

$$X_M = \begin{bmatrix} x_{11} - \tilde{x_1} & x_{12} - \tilde{x_2} & \cdots & x_{1n} - \tilde{x_n} \\ x_{21} - \tilde{x_1} & x_{22} - \tilde{x_2} & \cdots & x_{2n} - \tilde{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} - \tilde{x_1} & x_{m2} - \tilde{x_2} & \cdots & x_{mn} - \tilde{x_n} \end{bmatrix}; \, y_M = \begin{bmatrix} y_1 - \tilde{y} \\ y_2 - \tilde{y} \\ \vdots \\ y_m - \tilde{y} \end{bmatrix}$$

Se ha demostrado que, al obtener las soluciones de mínimos cuadrados $\hat{b}$ de $X_1\hat{b} = y$, se están obteniendo las soluciones de mínimos cuadrados $\hat{b}$ de $X_M\hat{b} = y_M$ y, con ello, ajustar la hipótesis de intersección

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + e_{y_i}$$

equivale a hacerlo con la hipótesis de no intersección

$$y_i - \tilde{y} = \beta_1(x_{i1} - \tilde{x_1}) + \beta_2(x_{i2} - \tilde{x_2}) + \ldots + \beta_n(x_{in} - \tilde{x_n}) + e_{y_i - \tilde{y}}.$$

De esta manera, la bondad de ajuste para una hipótesis de intersección es

$$R^2 = \frac{\|\mathbf{X_M X_M^\dagger y_M}\|^2}{\|\mathbf{y_M}\|^2} = \frac{\|\mathbf{y\hat{_M}}\|^2}{\|\mathbf{y_M}\|^2}.$$

En el caso de que todos los números sean reales, R^2 es tal que

$$R^2 = \frac{(\sum_{i=1}^{m}[y_i - \hat{y}][\hat{y}_i - \hat{y}])^2}{(\sum_{i=1}^{m}[y_i - \hat{y}]^2)(\sum_{i=1}^{m}[\hat{y}_i - \hat{y}]^2)} \tag{33}$$

Para probar (33), primero debe comprobarse que

$$X_M X_M^{\dagger} y = (X_1 X_1^{\dagger} - \frac{1}{m} J) y \tag{34}$$

lo que se deduce de que, $X_M^{\dagger} y$ es una solución de mínimos cuadrados de $X_M \hat{b} = y$, luego por el Teorema 4.5 se verifica que

$$s = \begin{bmatrix} \frac{1}{m} j^*(y - X X_M^{\dagger} y) \\ \dots \\ X_M^{\dagger} y \end{bmatrix}$$

es una solución por mínimos cuadrados de $X_1 \hat{b_1} = y$.

Así, todas las soluciones por mínimos cuadrados $\hat{b_1}$ de $X_1 \hat{b_1} = y$ son de la forma $\hat{b_1} = s + h$, $h \in N(X_1)$.

Como $X_1^{\dagger} y$ es una solución de mínimos cuadrados de $X_1 \hat{b_1} = y$, existirá un vector $h_0 \in N(X_1)$ tal que $X_1^{\dagger} = s + h_0$. Por lo tanto

$$X_1 X_1^{\dagger} y = X_1(s + h_0) = X_1 s + X_1 h_0 = X_1 s = \tfrac{1}{m} j j^*(y - X X_M^{\dagger} y) + X X_M^{\dagger} y =$$
$$= \tfrac{1}{m} J y - \tfrac{1}{m} J X X_M^{\dagger} y = \tfrac{1}{m} J y + (I - \tfrac{1}{m} J) X X_M^{\dagger} y =$$
$$= \tfrac{1}{m} J y + X_M X_M^{\dagger} y = (X_M X_M^{\dagger} + \tfrac{1}{m} J) y$$

con lo que queda demostrado (34).

Para concluir, se observa de (34) que, la i-ésima entrada $(\hat{y_M})_i$ de $\hat{y_M} = X_M X_M^{\dagger} y$, viene dada por

$$(\hat{y_M})_i = \tilde{y_i} - \tilde{y} \tag{35}$$

y con ello se obtiene la expresión buscada para R^2:

$$R^2 = \frac{\|\hat{\mathbf{y_M}}\|^2}{\|\mathbf{y_M}\|^2} = \frac{\|\hat{\mathbf{y_M}}\|^4}{\|\mathbf{y_M}\|^2 \|\hat{\mathbf{y_M}}\|^2} = \frac{(y_M^* X_M X_M^{\dagger} y_M)^2}{\|\mathbf{y_M}\|^2 \|\hat{\mathbf{y_M}}\|^2} = \frac{(y_{M^*} \hat{y_M})^2}{\|\mathbf{y_M}\|^2 \|\hat{\mathbf{y_M}}\|^2} \tag{36}$$

□

4.1.3. Aplicación para la adaptación de una curva y aproximaciones polinómicas

Para completar la sección de aplicaciones de pseudoinversa de Moore-Penrose por aproximación de mínimos cuadrados, se va a hacer un acercamiento a unos de los problemas que resolvió Carl Friedrich Gauss. Se hará mediante cuatro versiones del mismo, que suponen una adaptación para la mejor compresión por parte del lector y que incluyen los conceptos introducidos hasta el momento.

En enero de 1801, el astrónomo G.Piazzi pierde la localización de un 'nuevo planeta', actualmente conocido como el asteroide Ceres. Durante el resto de 1801, parte de la comunidad científica centra su atención en la búsqueda de dicho planeta, basándose en las observaciones obtenidas hasta el momento y, sentando como hipótesis de su estudio, una aproximación de órbita cicular. En diciembre de 1801, el científico Carl Friedrich Gauss, informa tanto de la ubicación del planeta, cómo de la posición que este tendría en el tiempo, sin revelar el método usado, que se basaba en una hipótesis de órbita elíptica.

Se incluyen a continuación las adaptaciones mencionadas a dicho problema.

1. **Primera versión.**

   Supóngase que la localización del planeta en cuestión viene dada por las m coordenadas $(x_1, y_1), \ldots, (x_m, y_m)$. Se desea buscar la elipse
   $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ que pase lo más cerca posible de los m puntos dados.
   Como encontrar los parámetros $\beta_1 = \frac{1}{a^2}, \beta_2 = \frac{1}{b^2}$ que satisfagan las m ecuaciones

   $$\beta_1(x_i)^2 + \beta_2(y_i)^2 = 1 \text{ para } i = 1, \ldots, m$$

   no es razonable debido al tamaño del error, se busca una aproximación de la misma que minimice dicho error. Para ello se consideran

   $$e_i = \beta_1(x_i)^2 + \beta_2(y_i)^2 - 1 \text{para} i = 1, 2, \ldots, m. \tag{37}$$

   que, en notación matricial, viene dado por la expresión

   $$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} x_1^2 & y_1^2 \\ x_2^2 & y_2^2 \\ \vdots & \vdots \\ x_m^2 & y_m^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

   es decir $e = Xb - j$.

   Para minimizar dicho error hay varias alternativas, pero, acorde con las deducciones de Gauss, la que da lugar a la mejor aproximación es la que impone

   $$\sum_{i=1}^{m} e_i^2 = \|\mathbf{e}\|^2. \tag{38}$$

   De esta manera se puede reescribir el problema como se detalla a continuación.

2. **Segunda versión.**

   Encontrar un vector $\hat{b} = [\hat{\beta}_1, \hat{\beta}_2]^T$ que sea la solución de mínimos cuadrados de $Xb = j$.

   Se ha demostrado en el Teorema 4.2 que todas las soluciones de mínimos cuadrados son de la forma $\hat{b} = X^\dagger j + h$, con $h \in N(X)$. Suponiendo que los puntos son no colineales, la matriz $X \in Mat_{m\times 2}(\mathbb{C})$ tiene rango 2, con lo que $N(X) = \{0\}$ y la única solución por mínimos cuadrados es $\hat{b} = X^\dagger j = (X^*X)^{-1}X^*j$.

   No obstante, hasta este punto no se ha impuesto ninguna condición sobre la no negatividad de los coeficientes de $X^\dagger j$. Por ello, surje la tercera versión equivalente del problema.

3. **Tercera versión.**

   Encontrar un vector u con coeficientes positivos tal que

   $$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|$$

   para todo v con coeficientes positivos.

4. **Cuarta versión.**

   Dados algunos puntos, encontrar la sección cónica que proporcione la solución más ajustada.

   Dicha sección cónica, descrita por $ax^2 + by^2 + cx + dy + fxy = 1$ se supone no centrada en el orgien.

Se ha visto hasta el momento cómo ajustar una cónica a un conjunto de puntos dados. Una variación de este problema, que se va a desarrollar a continuación, es el de encontrar el polinomio de grado n que mejor que se ajuste a un conjunto de puntos dados. Para esta aproximación también se va a dar la bondad de ajuste.

Sean $(x_1, y_1), \ldots, (x_m, y_m)$ los m puntos dados y

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \ldots + \hat{\beta}_n x^n \tag{39}$$

el polinomio buscado.

Observación 4.1.3. *En la situación mencionada, se considera $m > n+1$ ya que en caso contrario se obtendría el ajuste exacto por interpolación.*

Análogamente a lo visto anteriormente, se toma

$$\sum_{j=0}^{n} \beta_j x_i^j - y_i \tag{40}$$

para cada $i = 1, \ldots, m$.

Así,

$$\begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_1 & x_1^2 & \ldots & x_2^n \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_m & x_m^2 & \ldots & x_m^n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

o, equivalentemente, $e = Xb - y$.

Imponiendo la condición de que $\|\mathbf{e}\|$ sea mínima, el polinomio de grado n más cercano a los puntos dados tiene como coeficientes $\hat{\beta}_0, \ldots, \hat{\beta}_n$, que son las componente de la solución por mínimos cuadrados $\hat{b} = X^\dagger y + h$, con $h \in N(X)$ de $Xb = y$. En el caso de que X tenga rango completo, situación que se da si las $x_i's$ son todas diferentes, se tiene $N(X) = \{0\}$ y la única solución por mínimos cuadrados es $\hat{b} = X^\dagger y = (X^*X)^{-1}X^*y$

Del Teorema 4.6 se tiene que, la bondad de ajuste, al tratarse (40) de una hipótesis de intersección, viene dada por $R^2 = \frac{\|\mathbf{X_M X_M^\dagger y_M}\|^2}{\|\mathbf{y_M}\|^2}$.

Por último se considera una variación de la aproximación polinómica que consiste en, dadas n funciones $g_i(x)$, n funciones lineales l_i de k parámetros desconocidos β_i y m puntos (x_i, y_i); encontrar los valores $\hat{\beta}_i$ de manera que

$$y = l_1(\hat{\beta}_1, , \ldots, \hat{\beta}_k) g_1(x) + \ldots + l_n(\hat{\beta}_1, , \ldots, \hat{\beta}_k) g_n(x)$$

esté tan próximo a los puntos dados como sea posible.

Sea $l_i(\beta_1, \ldots, \beta_k) = w_{i1}\beta_1 + \ldots + w_{ik}\beta_k$ y se define $e_i = \sum_{j=1}^{n} l_j g_j(x_i) - y_i$. Entonces la correspondiente ecuación matricial es

$$e = XWb - y, \text{ donde } e = [e_1, \ldots, e_m]^T, b = [\beta_1, \ldots, \beta_k]^T, y = [y_1, \ldots, y_m]^T$$

$$X = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \ldots & g_n(x_1) \\ g_1(x_2) & g_2(x_2) & \ldots & g_n(x_2) \\ \vdots & \vdots & & \vdots \\ g_1(x_m) & g_2(x_m) & \ldots & g_n(x_m) \end{bmatrix}, W = \begin{bmatrix} w_{11} & \ldots & w_{1k} \\ \vdots & & \vdots \\ w_{n1} & \vdots & w_{nk} \end{bmatrix}$$

Nótese que el problema es equivalente a encontrar los valores $\hat{\beta}_i$ tales que $y = \hat{\beta}_1 L_1(x) + \hat{\beta}_2 L_2(x) + \ldots + \hat{\beta}_k L_k(x)$ esté tan próximo a los puntos dados como sea posible, donde $L_i(x)$ es una combinación lineal de las funciones $g_1(x), \ldots, g_n(x)$.

Para segurar que $\|\mathbf{e}\|$ sea mínima, los parámetros $\hat{\beta}_i$ deben ser las componentes de la solución de mínimos cuadrados $\hat{b}_w$ de $XW\hat{b}_w = y$ que, tal y como se ha demostrado, tiene por valor

$$\hat{b}_w = (P_{R(X^*)}W)^\dagger (XP_{R(W)})^\dagger y + h \text{ con } h \in N(XW).$$

En el caso de que W sea invertible y $N(X) = \{0\}$, con lo que también se tendría $N(XW) = \{0\}$, (40) proporciona la solución por mínimos cuadrados, que es única, dada por

$$\hat{b}_w = W^{-1}X^\dagger y = W^{-1}(X^*X)^{-1}X^*y = (X^*XW)X^*y. \tag{41}$$

5. Extensión al caso infinito

En esta última sección del trabajo se van a estudiar las soluciones de un sistema infinito de ecuaciones lineales a partir de la inversa de Moore-Penrose.

Para ello, primero se va a dar, tanto la definición de *inversa de Moore-Penrose de un endomorfismo entre espacios vectoriales de producto interno de dimensión finita* , como varias de sus propiedades más destacadas y que serán de gran utilidad para el entendimiento y desarrollo de resultados posteriores.

Después, se van a extender estos resultados a las *aplicaciones lineales entre espacios arbitrarios de producto interno*, en concreto de dimensión infinita, de manera que, se obtiene la definición de inversa de Moore-Penrose en estas restricciones, así como condiciones necesarias y suficientes para su existencia y unicidad. Posteriormente, de manera análoga a la situación en la que se trabaja con dimensión finita, se llevará a cabo el cálculo de la misma.

Como se adelantaba, esta sección, y con ello el desarrollo del trabajo, concluye con la extensión a *sistemas infinitos de ecuaciones lineales basados en la inversa de Moore-Penrose de aplicaciones lineales.*

5.1. Inversa de Moore-Penrose de espacios vectoriales finitos de producto interno

Previamente al desarrollo del tema de este apartado, se va a dar la definición de *espacios vectoriales de producto interno*, que se usará a lo largo de la sección 5 en numerosas ocasiones, así como la de invariante por f, para el mejor entendimiento por parte del lector.

Admemás se amplian las nociones sobre matrices, ya desarrolladas, de inversa de Moore-Penrose para endomorfismos para detallar, tanto la definición de la misma, como propiedades básicas.

Definición 5.1.1. *Sea $f \in End_k(V)$, un subespacio $W \subseteq V$ se dice que es **invariante por** f si $f(W) \subseteq W$.*

Definición 5.1.2. *Espacios vectoriales de producto interno.* *Sea k el cuerpo de los números reales o el cuerpo de los números complejos y sea V un k-espacio vectorial.*

*Se define el **producto interno** de V como la aplicación $g : V \times V \longrightarrow k$ que tastisface:*

1. *g es lineal en la primera componente:*

$$g(\lambda v_1 + \mu v_2, v') = \lambda g(v_1, v') + \mu g(v_2, v') \text{ para todo } v_1, v_2, v' \in V$$

2. *$g(v', v) = \overline{g(v, v')}$ para todo $v, v' \in V$, donde $\overline{g(v, v')}$ es la conjugada transpuesta de $g(v, v')$*

3. *g es definida positiva:*

$$g(v, v) \geq 0 \text{ y } g(v, v) = 0 \Leftrightarrow v = 0$$

*Un **espacio de producto interno** es el par (V, g).*

Sea (E, g) un espacio vectorial de dimensión finita de producto interno en $k = \mathbb{R}$ o $k = \mathbb{C}$. Si $f \in End_k(E)$, se tiene que $E = Kerf \oplus [Kerf]^{\perp} = Imf \oplus [Imf]^{\perp}$ y existe un endomorfismo

$$f : [Kerf]^{\perp} \xrightarrow{\sim} Imf$$

Así, la **inversa de Moore-Penrose** de f es la única aplicación lineal $f^{\dagger} \in End_k(E)$ tal que

$$f^{\dagger}(e) = \begin{cases} (f_{[Kerf]^{\perp}})^{-1}(e) & \text{si } e \in Imf \\ 0 & \text{si } e \in [Imf]^{\perp} \end{cases} \tag{42}$$

Definición 5.1.3. *Si $f \in End_k(E)$ y $\mathcal{H}_f = \{H_1, \dots, H_n\}$ es una familia de subespacios de E invariantes por f tales que $E = H_1 \oplus \dots \oplus H_n$, se define el **endomorfismo** $f^{+}_{\mathcal{H}_f} \in End_k(E)$ como la única aplicación lineal tal que*

$$[f^{+}_{\mathcal{H}_f}]_{|H_i} = [f_{|H_i}]^{\dagger} \textit{ para cada } i \in \{1, \cdots, n\}$$

Si se denota $f_i = f_{|H_i}$ para cada $i \in \{1, \dots, n\}$, es claro que para cada vector $e \in E$ tal que $e = h_{j_1} + \dots + h_{j_n}$ con $h_{j_i} \in H_i$, se cumple

$$f^{+}_{\mathcal{H}f}(e) = f^{\dagger}_1(h_{j_1}) + \dots + f^{\dagger}_n(h_{j_n})$$

Además, se observa de la Definición 5.1.3, y de las propiedades de la inversa de Moore-Penrose, que $f^{+}_{\mathcal{H}_f}$ es la inversa generalizada reflexiva de f para cada familia $\mathcal{H}_f$.

Siguiendo con la notación establecida, dado un subespacio $W \subseteq E$ tal que $W \subset H_i$ par cierto $i \in \{1, \dots, n\}$, se denota por

$$W_i^{\perp} = \{v_i \in H_i \text{ tal que } g(w, v_i) = 0 \text{ para todo } w \in W\}.$$

Lema 5.1. *Si $U \subset E$ es un subespacio y $\{U_1, \dots, U_n\}$ son subespacios de U tales que $U = U_1 \oplus \dots \oplus U_n$ con $U_i \subseteq H_i$, entonces*

$$[U_1]_1^{\perp} \oplus \dots \oplus [U_n]_n^{\perp} \subseteq U^{\perp} \textit{ si y solo si } [U_i]_i^{\perp} \subset [\sum_{j \neq i} U_j]^{\perp} \textit{ para todo } i \in \{1, \dots, n\}.$$

Demostración. Supóngase primero que $[U_1]_1^{\perp} \oplus \dots \oplus [U_n]_n^{\perp} \subseteq U^{\perp}$. Entonces $[U_i]_i^{\perp} \subseteq U^{\perp}$ para todo $i \in \{1, \dots, n\}$ y se deduce que $[U_i]_i^{\perp} \subset [\sum_{j \neq i} U_j]^{\perp}$ para todo $i \in \{1, \dots, n\}$.

Recíprocamente, supóngase que $[U_i]_i^{\perp} \subset [\sum_{j \neq i} U_j]^{\perp}$ para todo $i \in \{1, \dots, n\}$, con lo que

$$g(v_1 + \dots + v_n, u_1 + \dots + u_n) = 0$$

donde $v_i \in [U_i]_i^{\perp}$ y $u_i \in U_i$ para todo $i \in \{1, \dots, n\}$ y, por tanto, $[U_1]_1^{\perp} \oplus \dots \oplus [U_n]_n^{\perp} \subseteq U^{\perp}$. Con lo que queda probado el lema. □

Lema 5.2. *Usando la notación previa, se tiene que*

1. $Imf = Imf_1 \oplus \dots \oplus Imf_n$
2. *Si* $\mathcal{H}_f^{\perp} = [Imf_1]_1^{\perp} \oplus \dots \oplus [Imf_n]_n^{\perp}$, *entonces* $E = Imf \oplus \mathcal{H}_f^{\perp}$
3. $Kerf = Kerf_1 \oplus \dots \oplus Kerf_n$
4. *Si* $\tilde{\mathcal{H}}_f^{\perp} = [Kerf_1] \oplus \dots \oplus [Kerf_n]_n^{\perp}$, *entonces*

$$E = Kerf \oplus \tilde{\mathcal{H}}_f^{\perp}$$

5. f induce un isomorfismo entre $\mathcal{H}_f^{\perp}$ y Imf donde

$$[f^{+}_{\mathcal{H}_f}](e) = \begin{cases} (f_{|_{\widetilde{\mathcal{H}_f}^{\perp}}})^{-1}(e) & si\ e \in Imf \\ 0 & si\ e \in \mathcal{H}_f^{\perp} \end{cases} \tag{43}$$

Demostración. La demostración se encuentra en [3, LEMMA 3.3.]. □

Observación 5.1.1. *Con la notación usada hasta el momento, en general $\widetilde{\mathcal{H}}_f^{\perp} \neq [Kerf]^{\perp}$ y $\mathcal{H}_f^{\perp} \neq [Imf]^{\perp}$.*

Observación 5.1.2. *Dado un endomorfismo $f \in End_k(E)$, en general $f^{+}_{\mathcal{H}_f}$ no es la inversa de Moore-Penrose de f.*

En [3, LEMMA 3.7.] se prueba el siguiente lema.

Lema 5.3. *Se veririca que $\mathcal{H}_f^{\perp} = [Imf]^{\perp}$ si y solo si*

$$[Imf_i]_i^{\perp} \subseteq [\sum_{j \neq i} Imf_i]^{\perp} \text{ para todo } i \in \{1, \dots, n\}.$$

Como se enuncia en [3, LEMMA 3.8.].

Lema 5.4. *Se tiene que $\tilde{H}_f^{\perp} = [Kerf]^{\perp}$ si y solo si*

$$[Kerf_i]_i^{\perp} \subseteq [\sum_{j \neq i} Kerf_i]^{\perp} \text{ para cada } i \in \{1, \dots, n\}$$

Proposición 5.1. *Si $f \in End_k(E)$ y $\mathcal{H}_f = \{H_1, \dots, H_n\}$ es una familia de subespacios de E invariantes por f tales que*

$$E = H_1 \oplus \ldots \oplus H_n \text{ y } H_i \subseteq [\sum_{j \neq i} H_j]^{\perp} \text{ para todo } i \in \{1, \dots, n\},$$

entonces $f^{+}_{\mathcal{H}_f} = f^{\dagger}$.

5.2. Inversa de Moore-Penrose de aplicaciones lineales entre espacios arbitrarios de producto inertno

Una vez se conocen las propiedades de la inversa de Moore-Penrose para endomorfismos entre espacios de dimensión finita, se generalizan estas nociones para aplicaciones lineales entre espacios arbitrarios, en concreto, espacios de dimensión infinita.

En este apartado 5.2 se va a dar la definición de *aplicación lineal que admite inversa de Moore-Penrose*, las condiciones necesarias y suficientes de la existencia y unicidad de dicha invesa y se mencionarán algunos resultados consecuentes a dichas condiciones. Algunas de las demostraciones de estas últimas nociones están disponible en [3] para consulta por parte del lector.

De aquí en adelante se considerarán (V, g) y $(W, \bar{g})$ dos k-espacios vectoriales de producto interno con $k = \mathbb{C}$ o $k = \mathbb{R}$.

Definición 5.2.1. *Dad una aplicación lineal $f : V \longrightarrow W$, se dice que f* ***admite inversa de Moore-Penrose*** *cuando $V = Kerf \oplus [Kerf]^{\perp}$ y $W = Imf \oplus [Imf]^{\perp}$.*

Observación 5.2.1. *Se sabe que existen espacios vectoriales de dimensión infinita V y subespacios vectoriales $U \subseteq V$ tales que $V \neq U \oplus U^{\perp}$. En este caso, si $V = U \oplus W$ es claro que la aplicación lineal $f_U \in End_k(E)$ definida por*

$$f_U(v) = \begin{cases} 0 & si\ v \in U \\ v & si\ v \in W \end{cases}$$

no admite inversa de Moore-Penrose.

Previamente a la demostración del teorema de existencia y unicidad para la inversa de Moore-Penrose, se detalla la definición de *aplicación lineal adjunta y auto-adjunta.*

Definición 5.2.2. *Sean (V, g) y $(W, \bar{g})$ son k-espacios vectoriales de producto interno.*

Si $f : V \longrightarrow W$ es una aplicación lineal, el operador lineal $f^* : W \longrightarrow V$ recibe el nombre de ***adjunto de*** *f cuando*

$$g(f^*(w), v) = \bar{g}(w, f(v))$$

para todo $v \in V$ y $w \in W$. Si $f \in End_k(E)$, se dice que f es ***auto-adjunto*** *cuando $f^* = f$.*

Teorema 5.1. *Existencia y unicidad de la inversa de Moore-Penrose*. *Si (V, g) y $(W, \bar{g})$ son k-espacios vectoriales de producto interno, entonces $f : V \longrightarrow W$ es una aplicación lineal que admite inversa de Moor-Penrose si y solo si existe una única aplicación lineal $f^{\dagger} : W \longrightarrow V$ tal que:*

1. *$f^{\dagger}$ es una inversa generalizada reflexiva de f;*

2. *$f^{\dagger} \circ f$ y $f \circ f^{\dagger}$ son auto-adjuntas, es decir:*

   - $g([f^{\dagger} \circ f](v), v') = g(v, [f^{\dagger} \circ f](v'))$
   - $\bar{g}([f \circ f^{\dagger}](w), w') = \bar{g}(w, [f \circ f^{\dagger}](w'))$.

   para todo $v, v' \in V$ y $w, w' \in W$.

Demostración. Se va comenzar probando la existencia de la inversa de Moore-Penrose.

Si f admite inversa de Moore-Penrose, entonces la restricción $f_{|_{[Kerf]^{\perp}}}$ es un isomorfismo entre $[Kerf]^{\perp}$ e Imf y existe una aplicación lineal que satisface

$$f^{\dagger}(w) = \begin{cases} (f_{|_{[Kerf]^{\perp}}})^{-1}(w) & \text{si } w \in Imf \\ 0 & \text{si } w \in [Imf]^{\perp} \end{cases}$$

En este caso, f es única.

Para ver si verifica las condiciones del enunciado, se tiene que, como

$$(f \circ f^{\dagger})(w) = \begin{cases} w & \text{si } w \in Imf \\ 0 & \text{si } w \in [Imf]^{\perp} \end{cases}$$

y $(f^{\dagger} \circ f)(v) = v_1$ con $v = v_1 + v_2$, donde $v_1 \in [Kerf]^{\perp}$ y $v_2 \in Kerf$, es claro que $f^{\dagger}$ es una inversa generalizada reflexiva de f porque

- $(f \circ f^\dagger \circ f)(v) = f(v)$;
- $(f^\dagger \circ f \circ f^\dagger)(w) = f^\dagger(w)$.

Además,

$$\bar{g}([f \circ f^\dagger](w), w') = \bar{g}(w, [f \circ f^\dagger](w')) = \begin{cases} \bar{g}(w, w') & si \quad w, w' \in Imf \\ 0 & si \quad w \in [Imf]^\perp \\ 0 & si \quad w' \in [Imf]^\perp \end{cases}$$

Si $v, v' \in V$ con $v = v_1 + v_2, v' = v'_1 + v'_2$, donde $v_1, v'_1 \in [Kerf]^\perp$ y $v_2, v'_2 \in Kerf$, entonces se tiene que

$$g([f^\dagger \circ f](v), v') = g(v_1, v'_1) = g(v, [f^\dagger \circ f](v'))$$

Por lo que se concluye que $f^\dagger$ satisface las condiciones del teorema.

Para ver la unicidad de la inversa de Moore-Penrose de f, se considera la aplicación lineal $\tilde{f} : W \longrightarrow V$ tal que

1. $\tilde{f}$ es la inversa generalizada reflexiva de f,
2. $g([\tilde{f} \circ f](v), v') = g(v, [\tilde{f} \circ f](v'))$,
3. $\tilde{g}([f \circ \tilde{f}](w), w') = \tilde{g}(w, [f \circ \tilde{f}](w'))$

para todo $v, v' \in V$ y $w, w' \in W$.

Una consecuencia directa de 1. es que $(\tilde{f} \circ f)^2 = \tilde{f} \circ f$. Con ello, $\tilde{f} \circ f$ es una proyección y, como $Imf = Im(f \circ \tilde{f} \circ f) \subseteq Im(f \circ \tilde{f}) \subseteq Imf$, entonces $Im(f \circ \tilde{f}) = Imf$.

Por consiguiente, dada $w \in Imf$, existe $\bar{w} \in W$ tal que $(f \circ \tilde{f})(\bar{w})$, y por tanto

$$(f \circ \tilde{f})(w) = (f \circ \bar{f})^2(\bar{w}) = (f \circ \tilde{f})(\bar{w}) = w.$$

Además, si $w' \in [Imf]^\perp$, se tiene que

$$0 = \bar{g}([f \circ \tilde{f}]^2(w'), w') = \bar{g}([f \circ \tilde{f}](w'), [f \circ \tilde{f}](w')) \Rightarrow [f \circ \tilde{f}](w') = 0$$

Así,

$$(f \circ \tilde{f})(w) = \begin{cases} w & \text{si } w \in Imf \\ 0 & \text{si } w \in [Imf]^\perp \end{cases}$$

y, en particular, $\tilde{f}(w') = 0$ cuando $w' \in [Imf]^\perp$.

Además, se tiene que $(\tilde{f} \circ f)^2 = \tilde{f} \circ f$ y $Im(\tilde{f} \circ f) = Im\tilde{f}$.

Ahora bien, si $v \in Im\tilde{f}, v = [\tilde{f} \circ f](\bar{v})$ y $v' \in Kerf$, entones

$$g(v, v') = g([\tilde{f} \circ f](\bar{v}), v') = g(\bar{v}, 0) = 0$$

y se deduce que $Im\tilde{f} \subseteq [Kerf]^\perp$.

Finalmente, como $f_{|_{[Kerf]^\perp}} : [Kerf]^\perp \xrightarrow{\sim} Imf$ y $(f \circ \tilde{f})_{|_{Imf}} = Id_{|_{Imf}}$, entonces

$$(\tilde{f})_{|_{Imf}} = (f_{|_{[Kerf]^\perp}})^{-1} \Rightarrow \tilde{f} = f^\dagger.$$

Recíprocamente, se supone que existe la inversa de Moore-Penrose $f^\dagger : W \longrightarrow V$ de una aplicación lineal $f^\dagger : W \longrightarrow W$. En base a los argumentos desarrollados hasta ahora, se deduce de manera inmediata que

- $f \circ f^\dagger$ y $f^\dagger \circ f$ son proyecciones,
- $Im(f \circ f^\dagger) = Imf$,
- $[Imf]^\perp \subseteq Ker(f \circ f^\dagger)$,
- $Im(f^\dagger \circ f) \subseteq [Kerf]^\perp$

Además, si $w \notin [Imf]^\perp$, existe $\bar{w} \in W$ tal que

$$0 \neq \bar{g}([f \circ f^\dagger]^2(\bar{w}), w) = \bar{g}([f \circ f^\dagger](\bar{w}), [f \circ f^\dagger](w)),$$

de donde se deduce que $w \notin Ker(f^\dagger \circ f)$ y $[Imf]^\perp = Ker(f \circ f^\dagger)$.

Por otra parte, es claro que $Kerf \subseteq Ker(f^\dagger \circ f)$ y, si $v \in V$ con $f(v) \neq 0$, entonces $v \notin Ker(f^\dagger \circ f)$, porque

$$f(v) = (f \circ f^\dagger \circ f)(v) \neq 0.$$

Por ello, $Kerf = Ker(f^\dagger \circ f)$ y, teniendo en cuenta que si $g \in End_k(V)$ es una proyección, entonces $V = Kerg \oplus Img$, se concluye que

$$V = Ker(f^\dagger \circ f) \oplus Im(f^\dagger \circ f) = Kerf \oplus [Kerf]^\perp$$

y

$$W = Ker(f \circ f^\dagger) \oplus Im(f \circ f^\dagger) = Imf \oplus [Imf]^\perp.$$

Con lo cual, f admite inversa de Moore-Penrose y se finaliza la demostración.

□

Corolario 5.2.1. *Si (V, g) y $(W, \bar{g})$ son k-espacios de producto interno y $f : V \longrightarrow W$ es una aplicación lineal que admite inversa de Moore-Penrose, entonces $f^\dagger$ también admite inversa de Moore-Penrose y $(f^\dagger)^\dagger = f$.*

Demostración. Se deduce del Teorema 5.1 teniendo en cuenta que

- $Imf^\dagger = [Kerf]^\perp$;
- $[Imf^\dagger]^\perp = Kerf$;
- $Kerf^\dagger = [Imf]^\perp$;
- $[Kerf^\dagger]^\perp = Imf$.

□

Como se enuncia en [3, COROLLARY 3.14.].

Corolario 5.2.2. *Si (V, g) y $(W, \bar{g})$ son k-espacios de producto interno y $f : V \longrightarrow W$ es una aplicación lineal que admite inversa de Moore-Penrose, entonces:*

- $f^\dagger \circ f = P_{[Kerf]^\perp}$;
- $f \circ f^\dagger = P_{Imf}$

donde $P_{[Kerf]^\perp}$ y P_{Imf} son las proyecciones inducidas por la descomposición $V = Kerf \oplus [Kerf]^\perp$ y $W = Imf \oplus [Imf]^\perp$ respectivamente.

5.3. Cálculo de la inversa de Moore-Penrose de endomorfismos entre espacios de producto interno arbitarios

La sección previa al desarrollo de los sistemas de ecuaciones lineales de dimensión ininita, se centra en determinar bajo qué condiciones, la inversa de Moore-Penrose de un endomorfismo coincide con el endomorfismo $f^{+}_{\mathcal{H}_f}$ definido en 5.1.

Para llegar al resultado mencionado, se define *inversa generalizada de un endomorfismo asociada a la familia de vectores* $\mathcal{H}_f$ (detallada en capítulos previos) y se aporta el enunciado de varios resultados fundamentales.

Definición 5.3.1. *La única aplicación lineal* $f^{+}_{\mathcal{H}_f} \in End_k(V)$ *tal que* $[f^{+}_{\mathcal{H}_f}]_{|H_i} = f_i^{\dagger}$ *para cada* $i \in I$ *recibe el nombre de* ***inversa generalizada de*** f ***asociada a la familia*** $\mathcal{H}_f$.

Si f admite inversa de Moore-Penrose, se va a estudiar bajo qué condiciones $f^{+}_{\mathcal{H}_f} = f^{\dagger}$. Para ello, se detalla a continuación la generalización del Lema 5.2 a espacios vectoriales de dimensión infinita.

Lema 5.5. *Se tiene que:*

1. $Imf = \bigoplus_{i\in I} Imf_i$;

2. *Si* $\mathcal{H}_f^{\perp} = \bigoplus_{i\in I} [Imf_i]_i^{\perp}$, *entonces* $V = Imf \oplus \mathcal{H}_f^{\perp}$;

3. $Kerf = \bigoplus_{i\in I} Kerf_i$;

4. *Si* $\widetilde{\mathcal{H}_f}^{\perp} = \bigoplus_{i\in I} [Kerf_i]_i^{\perp}$, *entonces* $V = Kerf \oplus \widetilde{\mathcal{H}_f}^{\perp}$

5. f *induce un isomorfismo entre* $\widetilde{H}_f^{\perp}$ *e* Imf

Demostración. Puede encontrarse en [3, LEMMA 3.16.]. □

Como se enuncia en [3, LEMMA 3.17.].

Lema 5.6. *Teniendo en cuenta las suposiciones hechas sobre* V *y* $\mathcal{H}_f = \{U_i\}_{i\in I}$, *si* $U \subset V$ *es un subespacio y* $\{U_i\}_{i\in I}$ *es una familia de subespacios de* V *tal que* $U = \bigoplus_{i\in I}$ *con* $U_i \subseteq H_i$, *entonces*

$$\bigoplus_{i\in I} [U_i]_i^{\perp} \subseteq U^{\perp} \Leftrightarrow [U_i]_i^{\perp} \subset [\sum_{j\neq i} U_j]^{\perp} \text{ para todo } i \in I.$$

Como se enuncia en [3, LEMMA 3.18.].

Lema 5.7. *Si* (V, g) *es un* k*-espacio vectorial de producto interno,* $f \in End_k(V)$, *y* $\mathcal{H}_f = \{H_i\}_{i\in I}$ *con* $V = \bigoplus_{i\in I} H_i$ *y cada subespacio* H_i *es* f*-invariante, entonces* $\mathcal{H}_f^{\perp} = [Imf]^{\perp}$ *si y solo si*

$$[Imf_i]_i^{\perp} \subseteq [\sum_{i\neq j} Imf_i]^{\perp} \text{ para todo } i \in I.$$

Como se enuncia en [3, LEMMA 3.19.].

Lema 5.8. *Si (V, g) es un k-espacio vectorial de producto interno, $f \in End_k(V)$, y $\mathcal{H}_f = \{H_i\}_{i\in I}$ con $V = \bigoplus_{i\in I} H_i$ y cada subespacio H_i es f-invariante, entonces $\tilde{\mathcal{H}}_f^{\perp} = [Kerf]^{\perp}$ si y solo si*

$$[Kerf_i]_i^{\perp} \subseteq [\sum_{i\neq j} Kerf_i]^{\perp} \text{ para todo } i \in I.$$

Lo visto hasta ahora lleva al siguiente teorema.

Teorema 5.2. *Si (V, g) es un k-espacio vectorial de producto interno, $f \in End_k(V)$ que admite inversa de Moore-Penrose, y $\mathcal{H}_f = \{H_i\}_{i\in I}$ y cada subespacio H_i es f-invariante, entonces $f^{+}_{\mathcal{H}_f} = f^{\dagger}$ si y solo si se satisfacen las siguientes condiciones*

1. $[Imf_i]_i^{\perp} \subseteq [\sum_{i\neq j} Imf_i]^{\perp}$ *para todo* $i \in I$.

2. $[Kerf_i]_i^{\perp} \subseteq [\sum_{i\neq j} Kerf_i]^{\perp}$ *para todo* $i \in I$.

Demostración. La demostración es una consecuencia directa del Lema 5.7 y el Lema 5.8. □

Corolario 5.3.1. *Si (V, g) es un k-espacio vectorial de producto interno, $f \in End_k(V)$ y $\mathcal{H}_f = \{H_i\}_{i\in I}$ tal que cada subespacio H_i es f-invariante y $H_i \subseteq [\sum_{j\neq i} H_j]^{\perp}$ para todo $i \in I$, entonces f admite inversa de Moore-Penrose y $f^{+}_{\mathcal{H}_f} = f^{\dagger}$.*

Demostración. La demostración puede consultarse en [3, COROLLARY 3.21.]. □

5.4. Sistemas infinitos de ecuaciones lineales a partir de la pseudoinversa de Moore-Penrose

Como se anunciaba al inicio de esta última sección 5, se concluye el trabajo con la extensión de los resultados obtenidos al caso de sistemas infinitos de ecuaciones lineales a partir de la inversa de Moore-Penrose.

Para ello, se va dar la definición de *sistema lineal*, *mínima solución de norma* $\bar{g}$ (donde $\bar{g}$ determina el espacio de producto interior $(W, \bar{g})$), *menor solución mínima de norma* $\bar{g}$ así como las condiciones para que la invesra de Moore-Penrose sea la menor solución mínima de norma $\bar{g}$ de un sistema lineal.

Definición 5.4.1. *Si V y W son dos k-espacios vectoriales cualesquiera y $f : V \longrightarrow W$ es una aplicación lineal, un **sitema lineal** es la expresión*

$$f(x) = w \text{ con } w \in W.$$

*Este sistema recibe el nombre de **consistente** cuando $w \in Imf$.*

A partir de esta definición se tiene que

- Si v y W son espacios vectoriales de dimensión infinita, fijadas las bases de ambos, el sistema lineal $f(x) = 0$ es equivalente a un sistema infinito de ecuaciones lineales.

- Si un sistema lineal $f(x) = w$ es consistente y $f(v_0) = w$ para cierto $v_0 \in V$, entonces el conjunto de soluciones de este sistema es $v_0 + Kerf$, donde el vector v_0 recibe el nombre de **solución particular del sistema**.

Observación 5.4.1. ***Condición necesaria y suficiente.***
Sean $(V, g), (W, \bar{g})$ dos k-espacios vectoriales de producto interno sobre k, sea $f : V \longrightarrow W$ una aplicación lineal que admite inversa de Moore-Penrose y sea $f^{\dagger}$ su inversa de Moore-Penrose.

Si $f(x) = w$ es un sistema lineal, como $f^{\dagger}$ es la inverse reflexiva generalizada de f, entonces f es consistente si y solo si $(f \circ f^{\dagger})(w) = w$ y, en este caso, el conjunto de soluciones del sistema lineal es $f^{\dagger}(w) + Kerf$.

Definición 5.4.2. *Si (V, g) y $(W, \bar{g})$ son dos k-espacios de producto interior abitrarios, entonces $v' \in V$ recibe el nombre de* ***mínima solución de norma*** *$\bar{g}$ del sistema lineal $f(x) = w$ cuando*

$$\|\mathbf{f}(\mathbf{v}') - \mathbf{w}\|_{\bar{g}} \leq \|\mathbf{f}(\mathbf{v}) - \mathbf{w}\|_{\bar{g}}$$

para todo $v \in V$.

Definición 5.4.3. *Si (V, g) y $(W, \bar{g})$ son dos k-espacios de producto interior abitrarios, entonces $\tilde{v} \in V$ recibe el nombre de* ***menor solución mínima de norma*** *$\bar{g}$ del sistema lineal $f(x) = w$ cuando*

$$\|\tilde{\mathbf{v}}\|_{g} = \|\mathbf{v}'\|_{g}$$

para toda solución de norma $\bar{g}$-mínima $v' \in V$.

Proposición 5.2. *Si (V, g) y $(W, \bar{g})$ son dos k-espacios de producto interior abitrarios, $f : V \longrightarrow W$ es una aplicación lineal que admite inversa de Moore-Penrose y $f(x) = w$ es un sistema lineal, entonces $f^{\dagger}(w)$ es la única menor solución mínima de norma $\bar{g}$ de dicho sistema lineal.*

Demostración. Como $(f \circ f^{\dagger} - Id) \subseteq [Imf]^{\dagger}$, se tiene que

$$\begin{aligned} \|\mathbf{f}(\mathbf{v}) - \mathbf{w}\|_{\bar{g}}^2 &= \|[\mathbf{f}(\mathbf{v}) - \mathbf{f}(\mathbf{f}^{\dagger}(\mathbf{w}))] + [\mathbf{f}(\mathbf{f}^{\dagger}(\mathbf{w})) - \mathbf{w}]\|_{\bar{g}}^2 = \\ &= \|[\mathbf{f}(\mathbf{v}) - \mathbf{f}(\mathbf{f}^{\dagger}(\mathbf{w}))]\|_{\bar{g}}^2 + \|\mathbf{f}(\mathbf{f}^{\dagger}(\mathbf{w})) - \mathbf{w}]\|_{\bar{g}}^2 \end{aligned} \tag{44}$$

para todo $v \in V$. Se deduce así que

$$\|\mathbf{f}(\mathbf{f}^{\dagger}(\mathbf{w})) - \mathbf{w}]\|_{\bar{g}} \leq \|\mathbf{f}(\mathbf{v}) - \mathbf{w}\|_{\bar{g}}$$

para todo $v \in V$, y también que $f^{\dagger}(w)$ es una solución de $\bar{g}$-norma mínima de $f(x) = w$.

Además, de (44) se deduce que $v‘ \in V$ es una solución mínima de norma $\bar{g}$ del sistema lineal si y solo si $f(v‘) - f(f^{\dagger}(w)) = 0$, lo que significa que v' es solución del sistema consistente

$$f(x) - f(f^{\dagger}(w)) = 0$$

Así, para cada solución mínima de norma $\bar{g}$ $v' \in V$, se tiene que

$$f^{\dagger}(w) = v' + h$$

con $h \in Kerf$ y, teniendo en cuenta que $f^{\dagger}(w) \in [Kerf]^{\perp}$, se obtiene que $f^{\dagger}(w)$ es la única menor solución mínima de norma $\bar{g}$ de $f(x) = w$ porque

$$\|\mathbf{f}^{\dagger}(\mathbf{w})\|_g < \|\mathbf{v}'\|_g$$

para todo $v` \neq f^{\dagger}(w)$. $\square$

6. Conclusiones y líneas futuras

En el presente trabajo se ha estudiado en profundidad la inversa de Moore-Penrose y además, se han probado, tanto su existenica y unicidad, como sus propiedades básicas. También se ha desarrollado su aplicación a la resolución de sistemas de ecuaciones lineales no homogéneos y se han extendido las nociones al caso infinito.

El hecho de haber usado esta matriz y no otras invesras generalizadas, se basa en que, como ya se ha mencionado numerosas veces, es una matriz única. Por ello, a la hora de dar ejemplos por diferentes algoritmos, el resultado es el mismo y resulta más fácil, tanto para el lector como para la autora, comprobar que se han seguido corectamente los pasos detallados.

Algunas de las posibles líneas para continuar este trabajo podrían ser, tanto el desarrollo de propiedades de las inversas generalizadas de las que se ha hecho mención, como su relación con la que ha sido objeto de estudio de este trabajo. Además, las diferentes versiones sobre el problema de Gauss, podrían tener una adaptación más cercana a la realidad.

7. Agradecimientos

En primer lugar, quería agradecer a mi tutor de Trabajo de Fin de Grado, Fernando Pablos Romo, por la implicación en el desarrollo de este trabajo, por el detalle con el que ha leído y corregido las entregas realizadas y por el tema propuesto, que resultaba desconocido en un principio para mí, pero del que ha sido muy interesante aprender.

En segundo lugar, quería agradecer a mis compañeros de carrera y amigos que me ha dado esta etapa, tanto en mis años en Salamanca, como a los que he conocido durante el año de SICUE en La Laguna, el apoyo que me han dado siempre para seguir.

En tercer lugar, muchísimas gracias a mis amigas y compañeras de piso, a todo mi grupo de amigos, Víctor, Miguel, Lucía, Andrea y los demás, que han sido un pilar fundamental estos años. Gracias, en especial, a mi pareja Miguel y mi amiga Esther por la fuerza que necesitaba, y me han dado, durante este último año de carrera.

Por último, le estoy muy agradecida a mi familia, en especial a mis padres, Juan y Olga, y a mi hermano Óscar, por el amor y cariño con el que siempre me han tratado, y por la confianza que han tenido en mí.

Referencias

[1] Argerami M, Szechtman F, Tifenbach R. *On Tate´s trace.* Linear Multilinear Algebra. 2007;55(4):323-326.

[2] Ben-Israel A, Greville TNE.*Generalized inverses: theory and applications*. 2nd ed. Berlin:Springer; 2003.

[3] Cabezas Sánchez, V.; Pablos Romo, F., *Moore-Penrose Inverse of Some Linear Maps on Infinite-Dimensional Vector Spaces*, Electron. J. Linear Algebra 36 (2020), 570 - 586.

[4] Campbell, S. L.; Meyer, Jr., C. D.; *Generalized Inverses of Linear Transformations*, Dover, (1991). ISBN 978-0-486-66693-8.

[5] Drazin MP. Pseudo-inverses in associative rings and semigroups. Am Math Monthly. 1958;65(7):506-514.

[6] Mitra SK, Bhimasankaram P, Malik SB. *Matrix partial orders, shorted operators and applications.* Singapore: World Scientific Prublishing Company; 2010. (Series in algebra; vol.10). Disponible en https://doi.org/10.1142/7170.

[7] Orquera V. (2019). *Sobre la inversa generalizada core y algunos resultados relacionados.* (Trabajo de Fin de Máster). Universitat Politécnica de Valencia. Valencia. Recuperado de https://riunet.upv.es/bitstream/handle/10251/129508/Orquera %20- %20Sobre %20la %20 inversa %20generalizada %20core %20y %20algunos %20resultados %20relacionados.pdf? sequence=1.

[8] Pablos Romo, F., *On the Drazin Inverse of Finite Potent Endomorphisms*, Linear and Multilinear Algebra 67(10), (2019) 2135-2146.

[9] Rossignoli R.(2012). *Álgebra Lineal: Aplicaciones a la Física.* Recuperado de:

http://www.fisica.unlp.edu.ar/Members/rossigno/Curso2012.pdf.