Publicaciones Electrónicas
Sociedad Matemática Mexicana

La inversa generalizada de Moore-Penrose y aplicaciones

Néstor Thome

www.smm.org.mx

Serie: Textos. Vol. 21 (2019)
ISBN:

La inversa generalizada de Moore-Penrose y aplicaciones

Néstor Thome

Catedrático de Universidad
Departamento de Matemática Aplicada
Instituto Universitario de Matemática Multidisciplinar
Universitat Politècnica de València, España
njthome@mat.upv.es

Índice general

Prólogo

Este libro fue elaborado a partir de unas notas que sirvieron como material básico para un curso impartido durante el congreso MACI 2019, Matemática Aplicada, Computacional e Industrial celebrado en la Universidad Nacional de Río Cuarto de Argentina.

La idea era impartir un curso dirigido básicamente a alumnos de los últimos años de la Licenciatura en Matemática, Licenciatura en Física, Ingenierías, etc.

Se trataba de un curso que pudiese servir como continuación de asignaturas (como Álgebra Lineal, Análisis Matricial, Álgebra y Geometría, Teoría de Matrices, etc.) que los alumnos hubiesen estudiado en las carreras anteriormente mencionadas, de modo que se pudiese seguir con cierta facilidad y sea autocontenido.

El tema elegido para el curso fue "La inversa generalizada de Moore-Penrose y aplicaciones", que es un tópico que se centra dentro del Análisis Matricial y sus Aplicaciones y tiene vertientes tanto teóricas como computacionales.

Tras realizar un breve repaso de los conocimientos necesarios requeridos y, sobre todo, poner en claro las notaciones utilizadas, el curso comienza con tres definiciones de la inversa de Moore-Penrose y se demuestra que las mismas son equivalentes. Las definiciones proporcionan un enfoque funcional (pseudoinversa), un enfoque geométrico (definición de Moore) y, en tercer lugar, un enfoque algebraico (definición de Penrose). A continuación, se presentan propiedades de la inversa de Moore-Penrose.

Se estudian cuatro métodos para el cálculo de la inversa de Moore-Penrose basados en la pseudoinversa, en la descomposición en valores singulares, en el algoritmo de Gauss-Jordan y, por último, el método numérico de Greville.

También se presentan algunas aplicaciones. La primera de las aplicaciones es la más clásica: la resolución aproximada de sistemas de ecuaciones lineales por el método de los mínimos cuadrados. Una aplicación menos conocida corresponde al estudio de las clases de matrices EP y bi-dagger, que son extensiones de la clase conocida de matrices normales. Por último, se presenta una breve introducción a los órdenes parciales matriciales, considerando el orden parcial estrella. El libro termina con un anexo, en el que se demuestra un resultado técnico necesario para realizar la caracterización de este orden parcial, que tiene importancia por sí mismo dentro del Análisis

Matricial.

Se ilustran las definiciones y los resultados con ejemplos sencillos de manera que ayuden a una rápida asimilación de los mismos. Además, una serie de ejercicios tanto numéricos como algunos de carácter más teórico han sido intercalados durante el desarrollo del contenido para que se comprenda su utilización de manera directa, inmediata y que sirvan para completar la teoría correspondiente.

Se presentan también algunos aspectos cronológicos y aplicaciones de las inversas generalizadas, en general, tanto de las clásicas como de las más recientes.

Me gustaría expresar mi agradecimiento a la Dra. Marina Lattanzi (Universidad Nacional de La Pampa, Argentina) y a la Dra. Laura Rueda (Universidad Nacional del Sur, Argentina) por sus sugerencias y comentarios sobre una versión preliminar de las notas preliminares que me permitieron realizar mejoras sustanciales, así como también al Dr. David Ferreyra y al Dr. Fabián Levis (Universidad Nacional de Río Cuarto, Argentina) por las discusiones mantenidas mediante las cuales los intercambios de ideas han enriquecido mis puntos de vista sobre el Análisis Matricial y sus Aplicaciones.

Para sugerencias o comentarios, no duden en ponerse en contacto escribiéndome al correo electrónico `njthome@mat.upv.es`.

El autor.

Capítulo 1

Introducción, resultados previos y notaciones

El objetivo de este capítulo es introducir los resultados necesarios para estudiar el concepto de inversa de Moore-Penrose de una matriz (que se realizará a partir del capítulo 2), establecer las notaciones necesarias para poder seguir los capítulos posteriores y recordar muy brevemente algunos conceptos conocidos de Análisis Matricial. Para más detalles sobre conceptos básicos, el lector puede consultar, por ejemplo, el libro de C.D. Meyer [25].

1.1. Antecedentes, propósito y alcance

Con la intención de introducir el tema estudiado en este libro a partir de sus antecedentes, se recordará la definición de matriz inversa (en sentido ordinario) de una matriz cuadrada, se indicarán condiciones que las caracterizan considerando diferentes puntos de vista y se mostrarán las ideas que llevan a su extensión a situaciones más generales.

Es conocido que una matriz A de tamaño $n \times n$ (es decir, cuadrada) a coeficientes complejos es **invertible** si existe una matriz compleja X del mismo tamaño que satisfaga:

$$AX = XA = I_n, \tag{1.1}$$

siendo I_n la matriz identidad de tamaño $n \times n$. En caso contrario, se dice que A es **singular** o **no invertible**. En caso de existir dicha matriz X, es única y se denota por A^{-1}.

Ejemplo 1.1. *Es fácil comprobar que las matrices A y X cumplen las condiciones de (1.1):*

$$A = \begin{bmatrix} 3 & 2 & 0 \\ -1 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \qquad \textit{y} \qquad X = \begin{bmatrix} 1 & 2 & -2 \\ -1 & -3 & 3 \\ 2 & 6 & -5 \end{bmatrix}.$$

El siguiente resultado es el teorema fundamental de caracterización de matrices invertibles donde se aprecian los diferentes aspectos mediante los cuales se puede garantizar la existencia de dicha inversa.

Teorema 1.1. *Se consideran la matriz $A \in \mathbb{C}^{n\times n}$ y la aplicación lineal $T : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ definida por $x \longmapsto Ax$, cuya matriz respecto de la base canónica de $\mathbb{C}^n$ es A. Las siguientes condiciones son equivalentes:*

- *A es invertible.*
- *El determinante de la matriz A es no nulo.*
- *A tiene inversa a izquierda (es decir, existe $B \in \mathbb{C}^{n\times n}$ tal que $BA = I_n$).*
- *A tiene inversa a derecha (es decir, existe $C \in \mathbb{C}^{n\times n}$ tal que $AC = I_n$).*
- *El sistema lineal homogéneo $Ax = 0$ tiene sólo la solución $x = 0$.*
- *Para cada vector $b \in \mathbb{C}^n$, el sistema lineal $Ax = b$ tiene solución única.*
- *La forma escalonada reducida por filas (columnas) de A es I_n (es decir, A tiene n pivotes no nulos).*
- *A es producto (finito) de matrices elementales.*
- *Las filas (columnas) de A son vectores linealmente independientes en $\mathbb{C}^n$.*
- *Las filas (columnas) de A generan todo el espacio vectorial $\mathbb{C}^n$.*
- *Las filas (columnas) de A forman una base de $\mathbb{C}^n$.*
- *El núcleo de A tiene dimensión 0.*
- *El núcleo de A contiene únicamente al vector nulo.*
- *El rango de A es n.*
- *El espacio imagen de A es $\mathbb{C}^n$.*
- *La aplicación lineal T es un isomorfismo.*
- *0 no es un valor propio de A.*
- *La matriz de Gramm A^*A (o la matriz AA^*) es invertible.*
- *0 no es un valor singular de A.*

Es claro que si existe la matriz X que satisface (1.1) entonces se cumplen las condiciones:

- $AXA = A$ (que se obtiene multiplicando a derecha la igualdad $AX = I_n$ por A),
- $XAX = X$ (que se obtiene multiplicando a izquierda la igualdad $AX = I_n$ por X),
- AX es hermítica (por ser la matriz identidad),
- XA es hermítica (por ser la matriz identidad),

y, además, se tiene que:

- AX es un proyector ortogonal (por ser la matriz identidad),
- XA es un proyector ortogonal (por ser la matriz identidad).

Sin embargo, al considerar por ejemplo las matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{y} \qquad X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

se observa que todas las propiedades anteriores se cumplen a pesar de que A es singular.

A partir de estas observaciones, E.H. Moore [28] y R. Penrose [32] introdujeron dos nociones de matriz que mantienen la esencia de la inversa, ahora en un sentido generalizado. Con estas extensiones será también posible definir un concepto de inversa para matrices (cuadradas) singulares e incluso para matrices rectangulares. El propósito de este libro es estudiar estos hechos en profundidad estableciendo conexiones entre ellos y aplicaciones de los mismos.

A pesar de que la inversa introducida por Moore tenía un marcado componente geométrico en contraposición al claro carácter algebraico de la inversa de Penrose, se prueba que ambas definiciones son equivalentes. A partir de este punto se han podido estudiar numerosas propiedades de la inversa de Moore-Penrose y aplicaciones de la misma en diferentes áreas.

El alcance de este libro se pone de manifiesto a continuación donde se especifica su organización.

Después de repasar los conceptos necesarios (en lo que queda del capítulo 1), el capítulo 2 está dedicado a introducir detalladamente las definiciones de Moore y de Penrose y de probar su equivalencia. El capítulo 3 está dedicado al estudio detallado del problema de los mínimos cuadrados, que tanta utilidad presenta no sólo en áreas teóricas sino también aplicadas. El capítulo 4 proporciona una serie de métodos que pueden ser utilizados para el cálculo de la inversa de Moore-Penrose. En el capítulo 5 se presenta otra aplicación: matrices EP y matrices bi-dagger. Se estudian estas clases como generalizaciones de las clases conocidas de matrices normales utilizando la inversa de Moore-Penrose para su análisis. En el capítulo 6 se introduce el orden parcial estrella con la intención de mostrar al lector una aplicación a un área aparentemente distante como puede parecer el del Álgebra de la Lógica. Finalmente, en el capítulo 7 se comentan aspectos cronológicos y aplicaciones que permiten ver cómo se ha producido el desarrollo de esta inversa generalizada y se presentan algunas inversas generalizadas más recientes que se estudian en la actualidad. El libro finaliza con un anexo en el que se prueba un resultado sobre diagonalización simultánea y otro sobre descomposición en valores singulares simultánea para pares de matrices.

1.2. Resultados previos y notaciones

Se representa por $\mathbb{C}^{m\times n}$ al conjunto de todas las matrices a coeficientes complejos de tamaño $m \times n$ y por $\mathbb{C}_r^{m\times n}$ las matrices de $\mathbb{C}^{m\times n}$ de rango r.

Sea $A \in \mathbb{C}^{m\times n}$. El conjunto

$$\mathcal{R}(A) = \{Ax : x \in \mathbb{C}^n\}$$

se llama **espacio imagen** de A. El conjunto

$$\mathcal{N}(A) = \{x \in \mathbb{C}^n : Ax = 0\}$$

se llama **espacio nulo** o **núcleo** de A . La matriz cuyas columnas son las respectivas filas de A se denomina **matriz traspuesta** de A y se denota por A^T; y si, además de trasponer A, se toma el conjugado de cada elemento de A, se denomina **matriz traspuesta conjugada** de A y se denota por A^*.

Ejemplo 1.2. *Se considera la matriz*

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

El núcleo de A está formado por el conjunto de vectores de la forma $\begin{bmatrix} (-1-i)z \\ \frac{i}{2}z \\ z \end{bmatrix}$*, para $z \in \mathbb{C}$ arbitrario. El espacio imagen de A es $\mathbb{C}^2$ puesto que los vectores* $\begin{bmatrix} i \\ 0 \end{bmatrix}$ *y* $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ *son linealmente independientes en $\mathbb{C}^2$. La traspuesta conjugada de A es la matriz*

$$A^* = \begin{bmatrix} -i & 0 \\ 0 & 2 \\ -1-i & i \end{bmatrix}.$$

El subespacio generado por las filas de A se denotará por $\mathcal{F}(A)$. El **rango** de la matriz A se denotará por $\mathrm{rg}(A)$ y es el menor número de filas (o columnas) linealmente independientes de A.

La **matriz nula** de tamaño $m \times n$ se representará por $O_{m\times n}$ o bien por O si no hay lugar a confusión, lo mismo que la **matriz identidad** de tamaño $n \times n$ se representará por I_n.

El **espectro** de una matriz cuadrada $A \in \mathbb{C}^{n\times n}$ (es decir, el conjunto de todos los valores propios de A) se representará por $\sigma(A)$.

Una propiedad útil que muchas veces permite intercambiar información entre subespacios y matrices es:

$$AB = O \qquad \Longleftrightarrow \qquad \mathcal{R}(B) \subseteq \mathcal{N}(A),$$

para toda $A \in \mathbb{C}^{m\times n}$ y $B \in \mathbb{C}^{n\times p}$.

Con $S^\perp$ se denota al **subespacio ortogonal** del subconjunto $S \subseteq \mathbb{C}^n$ con respecto al producto escalar (canónico) de $\mathbb{C}^n$ definido por

$$\langle x, y\rangle_{\mathbb{C}^n} = x^* y, \quad \text{para} \quad x, y \in \mathbb{C}^{n\times 1},$$

donde x^* significa tomar traspuesta conjugada en el vector x. Es decir,

$$S^\perp = \{x \in \mathbb{C}^{n\times 1} : \ \langle x, y\rangle_{\mathbb{C}^n} = 0, \quad \text{para todo} \quad y \in \mathbb{C}^{n\times 1}\}.$$

La ortogonalidad de los subespacios S_1 y S_2 de denota por $S_1 \perp S_2$.

El símbolo $\oplus$ representa la **suma directa** de dos subespacios S_1 y S_2 de un mismo espacio vectorial complejo V y cuando estos subespacios son ortogonales se indicará con $\oplus^\perp$ (**suma directa ortogonal**). Se recuerda que $V = S_1 \oplus S_2$ si: (i) para cada vector $v \in V$ existen vectores $s_1 \in S_1$ y $s_2 \in S_2$ tales que $v = s_1 + s_2$ y (ii) $S_1 \cap S_2 = \{0\}$. Si en la suma directa los subespacios S_1 y S_2 son ortogonales, se denotará por $V = S_1 \oplus^\perp S_2$.

Se denota por P_S al **proyector ortogonal** que proyecta $\mathbb{C}^n$ sobre S (paralelamente a $S^\perp$) y se recuerda que, como siempre es posible descomponer

$$\mathbb{C}^n = S \oplus S^\perp,$$

se puede definir P_S como el único operador lineal de $\mathbb{C}^n$ tal que

$$P_S(x) = \left\{ \begin{array}{ll} x, & x \in S \\ 0, & x \in S^\perp \end{array} \right. .$$

En algunos casos se hará abuso de lenguaje utilizando la misma letra para representar tanto una aplicación lineal $A : \mathbb{C}^n \to \mathbb{C}^m$ como a la matriz $A \in \mathbb{C}^{m\times n}$ que la representa con respecto a sus bases canónicas. Se denota por $A_{|S}$ a la **restricción**, que evidentemente también es una aplicación lineal, de la matriz (aplicación lineal) A al subconjunto $S \subseteq \mathbb{C}^n$.

Con Id_S se representa a la **aplicación lineal identidad** sobre el subespacio S, es decir $\mathrm{Id}_S(x) = x$ para todo $x \in S$.

En Álgebra Lineal se estudian sobre $\mathbb{R}^{n\times n}$ las **matrices simétricas** ($A^T = A$), **matrices antisimétricas** ($A^T = -A$), **matrices ortogonales** (si A es invertible y $A^{-1} = A^T$) y **matrices normales reales** ($AA^T = A^TA$). En este trabajo se utilizarán sus versiones complejas. Una matriz $A \in \mathbb{C}^{n\times n}$ se llama:

- **hermítica** si $A^* = A$,
- **antihermítica** si $A^* = -A$,
- **unitaria** si A es invertible y $A^{-1} = A^*$,
- **normal** si $AA^* = A^*A$.

Ejemplo 1.3. *Las siguientes matrices son, respectivamente, hermítica, antihermítica, unitaria y normal:*

$$H = \begin{bmatrix} 1 & -i \\ i & 2 \end{bmatrix}, \qquad A = \begin{bmatrix} i & -i \\ -i & 2i \end{bmatrix}, \qquad U = \begin{bmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ i/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \qquad N = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}.$$

El siguiente resultado se estudia en Álgebra Lineal.

Teorema 1.2. *Sea $A \in \mathbb{C}^{n\times n}$. Entonces A es normal $\iff$ existen una matriz unitaria $U \in \mathbb{C}^{n\times n}$ y una matriz diagonal $D \in \mathbb{C}^{n\times n}$ tales que $A = UDU^*$.*

Se observa que las matrices hermíticas, antihermíticas y unitarias son todas matrices normales. Además, el siguiente resultado caracteriza cuándo son válidas las recíprocas.

Teorema 1.3. *Sea $A \in \mathbb{C}^{n\times n}$ una matriz normal. Entonces:*

(a) A *es hermítica* $\iff$ $\sigma(A) \subseteq \mathbb{R}$,

(b) A *es antihermítica* $\iff$ $\sigma(A) \subseteq i\,\mathbb{R}$,

(c) A *es unitaria* $\iff$ $\sigma(A) \subseteq \{z \in \mathbb{C} : |z| = 1\}$.

Ejemplo 1.4. *Los espectros de las matrices del Ejemplo 1.3 son, respectivamente,*

$$\sigma(H) = \left\{ \frac{3-\sqrt{5}}{2}, \frac{3+\sqrt{5}}{2} \right\}, \qquad \sigma(A) = \left\{ \frac{3-\sqrt{5}}{2}i, \frac{3+\sqrt{5}}{2}i \right\},$$

$$\sigma(U) = \left\{ \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i, \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i \right\}, \qquad \sigma(N) = \{1-i, 1+i\}.$$

Ejemplo 1.5. *El resultado del Teorema 1.2 se observa en el caso particular de la matriz normal* N*, la cual se puede diagonalizar mediante una matriz unitaria* U *como sigue:*

$$N = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 1-i & 0 \\ 0 & 1+i \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{U^*}.$$

Ejercicio 1.1. *Comprobar que los espectros de las matrices del Ejemplo 1.4 cumplen las condiciones del Teorema 1.3.*

En lo que sigue se repasan resultados clásicos del Análisis Matricial.

Teorema 1.4. *Sea* $A \in \mathbb{C}^{m\times n}$*. Entonces*

$$\mathcal{R}(A^*) = [\mathcal{N}(A)]^\perp \qquad \text{y} \qquad \mathcal{N}(A^*) = [\mathcal{R}(A)]^\perp.$$

Ejemplo 1.6. *Se considera de nuevo la matriz* A *del Ejemplo 1.2. Por el Teorema 1.4 se obtiene fácilmente que*

$$\mathcal{N}(A^*) = [\mathcal{R}(A)]^\perp = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

y, a partir de algunos cálculos, se tiene que

$$\mathcal{R}(A^*) = [\mathcal{N}(A)]^\perp = \left\{ \begin{bmatrix} -iz_1 \\ 2z_2 \\ (-1-i)z_1 + iz_2 \end{bmatrix} : \ z_1, z_2 \in \mathbb{C} \right\}.$$

Teorema 1.5. *Sea* $A \in \mathbb{C}^{m\times n}$*. Entonces*

$$\mathbb{C}^n = \mathcal{R}(A^*) \oplus^\perp \mathcal{N}(A) \quad \text{y} \quad \mathbb{C}^m = \mathcal{N}(A^*) \oplus^\perp \mathcal{R}(A).$$

Las matrices de Gramm A^*A y AA^* poseen interesantes propiedades.

Teorema 1.6. *Sea* $A \in \mathbb{C}^{m\times n}$*. Entonces*

(a) $\mathcal{R}(AA^*) = \mathcal{R}(A)$ *y* $\mathcal{R}(A^*A) = \mathcal{R}(A^*)$.

(b) $\mathcal{N}(AA^*) = \mathcal{N}(A^*)$ *y* $\mathcal{N}(A^*A) = \mathcal{N}(A)$.

(c) $\mathrm{rg}(AA^*) = \mathrm{rg}(A^*A) = \mathrm{rg}(A^*) = \mathrm{rg}(A)$.

La idempotencia y la simetría (en el sentido de traspuestas conjugadas) caracterizan el concepto de proyector ortogonal.

Teorema 1.7. *Sea $A \in \mathbb{C}^{n\times n}$. Entonces*

A es un proyector ortogonal de $\mathbb{C}^n$ sobre $\mathcal{R}(A)$, i.e., $A = P_{\mathcal{R}(A)}$ $\iff$ $A^2 = A = A^*$.

Ejemplo 1.7. *Se consideran las matrices*

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad \text{y} \qquad C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Del Teorema 1.7 se tiene que, al ser $A^2 = A \neq A^*$ se cumple que A es proyector oblicuo (es decir, no es ortogonal); mientras que, de $B^2 = B = B^*$, se deduce que B es un proyector ortogonal y es sencillo ver que proyecta $\mathbb{C}^2$ ortogonalmente sobre $\mathcal{R}(B) = \left\{ z \begin{bmatrix} 1 \\ 1 \end{bmatrix} : z \in \mathbb{C} \right\}$. Por último, puesto que $C^2 \neq C$, se deduce que C no es proyector.

Ejercicio 1.2. *Se considera la familia de matrices*

$$A_{\alpha,\beta} = \begin{bmatrix} \frac{1}{2} & \alpha \\ \beta & \frac{1}{2} \end{bmatrix},$$

en función de los parámetros $\alpha, \beta \in \mathbb{C}$. Se pide, en función de los valores de α y β:

(a) calcular el rango de $A_{\alpha,\beta}$ y el de $A^*_{\alpha,\beta}$,

(b) encontrar el núcleo de $A_{\alpha,\beta}$ y el de $A^*_{\alpha,\beta}$,

(c) hallar el espacio imagen de $A_{\alpha,\beta}$ y el de $A^*_{\alpha,\beta}$,

(d) indicar si $A_{\alpha,\beta}$ es hermítica y, si lo es, hallar su espectro,

(e) indicar si $A_{\alpha,\beta}$ es antihermítica y, si lo es, hallar su espectro,

(f) indicar si $A_{\alpha,\beta}$ es unitaria y, si lo es, hallar su espectro,

(g) indicar si $A_{\alpha,\beta}$ es normal,

(h) analizar si $A_{\alpha,\beta}$ es proyector y si es proyector ortogonal.

Capítulo 2

Definiciones equivalentes de la inversa de Moore-Penrose

En esta sección se presentarán tres definiciones para la matriz inversa generalizada de Moore-Penrose y se probará su equivalencia.

Antes de ello se profundizará en el estudio de una aplicación lineal pensando a los espacios de salida y de llegada descompuestos según los cuatro subespacios fundamentales: $\mathcal{N}(A)$, $\mathcal{R}(A)$, $\mathcal{N}(A^*)$, $\mathcal{R}(A^*)$.

Es conocida la importancia del subespacio $\mathcal{R}(T^*)$, ortogonal al núcleo de T, a la hora de calcular la inversa de una aplicación lineal $T : \mathbb{C}^n \to \mathbb{C}^m$. Para los casos en que $m \neq n$ o bien $m = n$ pero la inversa T^{-1} no exista, se presenta el siguiente resultado que permite analizar la aplicación lineal

$$T_{|\mathcal{R}(T^*)} : \mathcal{R}(T^*) \to \mathbb{C}^m, \tag{2.1}$$

restringiendo el dominio de T a un subespacio S complementario con $\mathcal{N}(T)$, es decir, de modo que $S \oplus \mathcal{N}(T) = \mathbb{C}^n$. Una posibilidad para elegir el subespacio S es claramente $S = \mathcal{R}(T^*)$.

Recordar que la **adjunta** de T es la única aplicación lineal $T^* : \mathbb{C}^m \to \mathbb{C}^n$ tal que

$$\langle x, T^*(y)\rangle_{\mathbb{C}^n} = \langle T(x), y\rangle_{\mathbb{C}^m}, \quad \text{para todo } x \in \mathbb{C}^n,\ y \in \mathbb{C}^m$$

y que en las bases canónicas (ortonormales) coincide con la traspuesta conjugada de la matriz de T en las bases canónicas correspondientes, es decir

$$[T^*]_{\mathcal{C}',\mathcal{C}} = ([T]_{\mathcal{C},\mathcal{C}'})^*,$$

siendo $\mathcal{C}$ y $\mathcal{C}'$ las bases canónicas de $\mathbb{C}^n$ y $\mathbb{C}^m$, respectivamente.

El siguiente resultado es fundamental para demostrar el Teorema 2.1.

Lema 2.1. *Sea $T : \mathbb{C}^n \to \mathbb{C}^m$ una aplicación lineal. Entonces $\mathcal{N}(T_{|\mathcal{R}(T^*)}) = \{0\}$.*

Demostración. Si $x \in \mathcal{N}(T_{|\mathcal{R}(T^*)})$ entonces $Tx = 0$ con $x \in \mathcal{R}(T^*) = [\mathcal{N}(T)]^\perp$. Luego, $x \in \mathcal{N}(T) \cap \mathcal{N}(T)^\perp = \{0\}$; es decir, $x = 0$. □

Este lema permite asegurar la existencia de la aplicación lineal $(T_{|\mathcal{R}(T^*)})^{-1}$ si se restringe adecuadamente el espacio imagen de $T_{|\mathcal{R}(T^*)}$ en (2.1).

Nota 2.1. *Sean V y W dos espacios vectoriales sobre $\mathbb{C}$ con V de dimensión finita y sea $f : V \to W$ una aplicación lineal. Se recuerda que el* **teorema de la dimensión** *establece que*

$$\dim(V) = \dim(\mathcal{N}(f)) + \dim(\mathcal{R}(f)).$$

Como consecuencia, si V y W son dos espacios vectoriales sobre $\mathbb{C}$ de dimensión finita tales que $\dim(V) = \dim(W)$ y $f : V \to W$ es una aplicación lineal entonces se tiene que son equivalentes: (a) f es monomorfismo, (b) f es isomorfismo.

Teorema 2.1. *Dada la aplicación lineal $T : \mathbb{C}^n \to \mathbb{C}^m$, la restricción*

$$T_{|\mathcal{R}(T^*)} : \mathcal{R}(T^*) \to \mathcal{R}(T) \tag{2.2}$$

es un isomorfismo de espacios vectoriales.

Demostración. Primera forma: La forma clásica de demostrar que la aplicación (2.2), que se puede denotar por f, es un isomorfismo entre los espacios vectoriales $\mathcal{R}(T^*)$ y $\mathcal{R}(T)$ es como sigue:

- f es lineal.

  Como $T : \mathbb{C}^n \to \mathbb{C}^m$ es una aplicación lineal, su restricción f al subespacio $\mathcal{R}(T^*)$ también lo es.

- f es inyectiva. (Véase el Lema 2.1).

  Sean $x_1, x_2 \in \mathcal{R}(T^*)$ dos vectores tales que $f(x_1) = f(x_2)$. Entonces $Tx_1 = Tx_2$, de donde se tiene que $T(x_1 - x_2) = 0$, es decir, $x_1 - x_2 \in \mathcal{N}(T) = [\mathcal{R}(T^*)]^\perp$. Por lo tanto, $x_1 - x_2 \in \mathcal{R}(T^*) \cap [\mathcal{R}(T^*)]^\perp = \{0\}$, y así, $x_1 = x_2$.

- f es sobreyectiva.

  Para probar que $\mathcal{R}(f) = \mathcal{R}(T_{|\mathcal{R}(T^*)}) = \mathcal{R}(T)$ alcanza con probar que $\mathcal{R}(T) \subseteq \mathcal{R}(f)$ pues la otra implicación es evidente. En efecto, dado $y \in \mathcal{R}(T)$, existe $x \in \mathbb{C}^n$ tal que $y = Tx$. Puesto que $\mathbb{C}^n = \mathcal{R}(T^*) \oplus^\perp \mathcal{N}(T)$, existen vectores únicos $x_r \in \mathcal{R}(T^*)$ y $x_n \in \mathcal{N}(T)$ tales que $x = x_r + x_n$. Luego, $y = Tx = T(x_r + x_n) = Tx_r + Tx_n = Tx_r = T_{|\mathcal{R}(T^*)}x_r \in \mathcal{R}(f)$.

Si bien se acaba de justificar que (2.2) es un isomorfismo, es interesante comentar que una manera más directa de hacerlo es la siguiente.

Segunda forma: Como claramente se cumple que $\mathcal{R}(T_{|\mathcal{R}(T^*)}) \subseteq \mathcal{R}(T)$, la aplicación está bien definida. Ahora, la demostración sigue del hecho conocido que $\mathrm{rg}(T^*) = \mathrm{rg}(T)$, de la Nota 2.1 y del Lema 2.1. $\square$

Ahora es posible decir que la inversa de la aplicación lineal T de (2.2) viene dada por la aplicación lineal

$$(T_{|\mathcal{R}(T^*)})^{-1} : \mathcal{R}(T) \to \mathcal{R}(T^*) \tag{2.3}$$

que, para cada $x \in \mathcal{R}(T)$, se define como

$$(T_{|\mathcal{R}(T^*)})^{-1}(x) = y,$$

donde $y \in \mathcal{R}(T^*)$ es el único vector tal que $T_{|\mathcal{R}(T^*)}(y) = x$.

Ejercicio 2.1. *Realizar un diagrama que permita interpretar los resultados del Teorema 1.5 y el Teorema 2.1 y su relación con la demostración del teorema de la dimensión.*

Resolución de sistemas de ecuaciones lineales: Dadas una matriz $A \in \mathbb{C}^{m\times n}$ y un vector columna $b \in \mathbb{C}^m$, se considera la ecuación matricial

$$Ax = b, \tag{2.4}$$

cuya incógnita es el vector columna $x \in \mathbb{C}^n$.

Si $b \in \mathcal{R}(A)$ entonces la ecuación matricial $Ax = b$ tiene una única solución $x_0 = (A_{|\mathcal{R}(A^*)})^{-1}b$ en $\mathcal{R}(A^*)$. Además, cualquier otra solución x de (2.4) cumple que $A(x - x_0) = Ax - Ax_0 = b - b = 0$, es decir $x \in x_0 + \mathcal{N}(A)$ y, como es claro que todas las de esta forma son soluciones de (2.4), es posible establecer que el conjunto S de todas las soluciones de (2.4) es la variedad lineal afín paralela a $\mathcal{N}(A)$ que pasa por $(A_{|\mathcal{R}(A^*)})^{-1}b$, es decir

$$S = (A_{|\mathcal{R}(A^*)})^{-1}b + \mathcal{N}(A).$$

Para el caso en que la ecuación matricial (2.4) no tenga solución (es decir, $b \notin \mathcal{R}(A)$), se presentará una *inversa generalizada* que permitirá encontrar una solución aproximada en un cierto sentido que se definirá en el capítulo 3.

A continuación se presenta la definición funcional para dicha inversa generalizada.

2.1. Definición funcional y pseudoinversa

Definición 2.1 (Definición funcional). *Sea $T : \mathbb{C}^n \to \mathbb{C}^m$ una aplicación lineal. Se llama* **inversa generalizada funcional** *(o* **aplicación pseudoinversa***) de T a la aplicación*

$$T^\dagger : \mathbb{C}^m \to \mathbb{C}^n$$

definida, para todo $x \in \mathbb{C}^m$, por

$$T^\dagger(x) = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(x)).$$

Es claro que, por el Teorema 2.1, la aplicación $T^\dagger$ está bien definida y, además, $T^\dagger$ es lineal por ser composición de las aplicaciones lineales $(T_{|\mathcal{R}(T^*)})^{-1}$ y $P_{\mathcal{R}(T)}$.

Nota 2.2. *De la Definición 2.1 se sigue que*

$$T^\dagger(x) = \begin{cases} (T_{|\mathcal{R}(T^*)})^{-1}(x), & \text{si } x \in \mathcal{R}(T) \\ 0, & \text{si } x \in [\mathcal{R}(T)]^\perp \end{cases},$$

que se puede utilizar como definición equivalente.

Nota 2.3. *De la construcción de $T^\dagger$ en la Definición 2.1 se observa que $\mathcal{R}(T^\dagger) = \mathcal{R}(T^*)$. Este hecho se puede demostrar como sigue.*

Si $y \in \mathcal{R}(T^\dagger)$ entonces $y = T^\dagger(x) = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(x))$, para algún $x \in \mathbb{C}^m$. Como existe un único par de vectores $u \in \mathcal{R}(T)$ y $v \in [\mathcal{R}(T)]^\perp$ tal que $x = u + v$, se tiene que $y = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(u+v)) = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(u) + P_{\mathcal{R}(T)}(v)) = (T_{|\mathcal{R}(T^*)})^{-1}(u)$, $u \in \mathcal{R}(T)$. Utilizando el isomorfismo establecido en (2.3) se obtiene que $y \in \mathcal{R}(T^*)$. Por lo tanto, $\mathcal{R}(T^\dagger) \subseteq \mathcal{R}(T^*)$.

Para demostrar la otra inclusión, sea $y \in \mathcal{R}(T^*)$. Entonces $x := T(y) \in \mathcal{R}(T)$. Por el isomorfismo establecido en (2.3) se tiene que $y = (T_{|\mathcal{R}(T^*)})^{-1}(x)$. Ahora, $T^\dagger(x) = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(x)) = (T_{|\mathcal{R}(T^*)})^{-1}(x) = y$, de donde $y \in \mathcal{R}(T^\dagger)$, y así $\mathcal{R}(T^*) \subseteq \mathcal{R}(T^\dagger)$.

Definición 2.2 (Definición de matriz pseudoinversa). *Sea $A \in \mathbb{C}^{m\times n}$ y sea $T_A : \mathbb{C}^n \to \mathbb{C}^m$ la aplicación lineal asociada a A con respecto a las bases canónicas $\mathcal{C}$ y $\mathcal{C}'$ de $\mathbb{C}^n$ y $\mathbb{C}^m$, respectivamente, es decir $[T_A]_{\mathcal{C},\mathcal{C}'} = A$. Se llama* **matriz pseudoinversa** *de A a la matriz que representa a la inversa generalizada funcional $T_A^\dagger : \mathbb{C}^m \to \mathbb{C}^n$ de la Definición 2.1 con respecto a las bases canónicas. Es decir, la matriz pseudoinversa de A es*

$$[T_A^\dagger]_{\mathcal{C}',\mathcal{C}} \in \mathbb{C}^{n\times m}.$$

Ejercicio 2.2. *En el diagrama del Ejercicio 2.1 interpretar cómo actúan las definiciones anteriores.*

Ejemplo 2.1. *Calcular la matriz pseudoinversa (o inversa generalizada funcional) de la matriz*

$$A = [T_A]_{\mathcal{C},\mathcal{C}'} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Es fácil ver que $\operatorname{rg}(A) = 2$. *El conjunto* $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ *forma una base de* $\mathcal{R}(T_A^*) = \overline{\mathcal{F}(A)}$. *Luego, por el isomorfismo entre* $\mathcal{R}(T_A^*)$ *y* $\mathcal{R}(T_A)$, *los vectores*

$$T_A\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad T_A\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \end{bmatrix}$$

forman una base de $\mathcal{R}(T_A)$.

Según la definición

$$T_A^\dagger\left(\begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{y} \qquad T_A^\dagger\left(\begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Ahora se debe hallar una base de $[\mathcal{R}(T_A)]^\perp = \mathcal{N}(T_A^*)$. *Al resolver el sistema* $A^*x = 0$ *se encuentra que su solución general se puede escribir como:*

$$x = x_1 \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{con } x_1, x_2 \in \mathbb{C}.$$

Como los dos vectores de la combinación lineal previa son linealmente independientes, forman una base de $[\mathcal{R}(T_A)]^\perp$.

Según la definición,

$$T_A^\dagger\left(\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}\right) = T_A^\dagger\left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Por lo tanto, al unir una base de $\mathcal{R}(T_A)$ con una de $\mathcal{N}(T_A^*)$ se obtiene una base de $\mathbb{C}^4 = \mathcal{R}(T_A) \oplus \mathcal{N}(T_A^*)$ (esto garantiza que la matriz formada por sus columnas es invertible). Además, denotando $Z := [T_A^\dagger]_{\mathcal{C}',\mathcal{C}}$ y operando por bloques se tiene

$$\begin{aligned}
[T_A^\dagger]_{\mathcal{C}',\mathcal{C}} \left[\begin{array}{c|c|r|r} 2 & 1 & 1 & 1 \\ 3 & 3 & -1 & -1 \\ 1 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right] &= \left[\begin{array}{cccc} Z\begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix} & Z\begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \end{bmatrix} & Z\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} & Z\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \end{array}\right] \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Finalmente,

$$\begin{aligned}
[T_A^\dagger]_{\mathcal{C}',\mathcal{C}} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 1 \\ 3 & 3 & -1 & -1 \\ 1 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}^{-1} \\
&= \frac{1}{15} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 7 & 3 & -4 & -4 \\ -5 & 0 & 5 & 5 \\ 3 & -3 & 9 & -6 \\ 3 & -3 & -6 & 9 \end{bmatrix} \\
&= \begin{bmatrix} 7/15 & 1/5 & -4/15 & -4/15 \\ -1/3 & 0 & 1/3 & 1/3 \\ 2/15 & 1/5 & 1/15 & 1/15 \end{bmatrix}.
\end{aligned}$$

2.2. Definición de Moore

La definición de Moore (para matrices) involucra proyectores ortogonales tanto de la matriz dada A como de la matriz X que se está definiendo.

Definición 2.3 (Definición de Moore)**.** *Se* $A \in \mathbb{C}^{m\times n}$. *Se llama* **inversa de Moore** *de A a la única matriz $X \in \mathbb{C}^{n\times m}$ que satisface:*

(a) $AX = P_{\mathcal{R}(A)}$,

(b) $XA = P_{\mathcal{R}(X)}$.

Tanto la existencia como la unicidad señaladas en la definición anterior se justificarán en el Teorema 2.2.

Según esta definición se tiene que:

- Si $x \in \mathcal{R}(A)$ entonces $AXx = x$, es decir, $AX = \mathrm{Id}_{\mathcal{R}(A)}$, con lo que AX actúa como la identidad solamente sobre el subespacio $\mathcal{R}(A)$ (pero no se puede asegurar que $AX = \mathrm{Id}_{\mathbb{C}^m}$).
- Si $x \in \mathcal{R}(X)$ entonces $XAx = x$, es decir, $XA = \mathrm{Id}_{\mathcal{R}(X)}$, con lo que XA actúa como la identidad solamente sobre el subespacio $\mathcal{R}(X)$ (pero no se puede asegurar que $XA = \mathrm{Id}_{\mathbb{C}^n}$).

Nota 2.4. *En el caso especial en que $AX = I_m$ y $XA = I_n$ (es decir que AX actúa como la identidad sobre todo $\mathbb{C}^m$ y XA actúa como la identidad sobre todo $\mathbb{C}^n$) entonces*

$$m = \mathrm{rg}(I_m) = \mathrm{rg}(AX) \leq \mathrm{rg}(A) \leq m, \text{ es decir } \quad \mathrm{rg}(A) = m$$

y

$$n = \mathrm{rg}(I_n) = \mathrm{rg}(XA) \leq \mathrm{rg}(A) \leq n, \text{ es decir } \quad \mathrm{rg}(A) = n.$$

En definitiva, si AX y XA fuesen las identidades sobre los respectivos espacios $\mathbb{C}^m$ y $\mathbb{C}^n$ entonces debería ser $m = n$ y esto correspondería al caso $X = A^{-1}$.

Ejemplo 2.2. *Calcular la inversa de Moore de la matriz*

$$A = \begin{bmatrix} 3 & 0 & 4 \end{bmatrix}.$$

Se trata de hallar la matriz

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{C}^{3\times 1}$$

que cumpla las condiciones (a) y (b) de la Definición 2.3.

Para calcular el proyector ortogonal $P_{\mathcal{R}(A)} \in \mathbb{C}^{1\times 1}$ es necesario hallar los subespacios de la descomposición $\mathbb{C}^1 = \mathcal{R}(A) \oplus [\mathcal{R}(A)]^\perp$ y claramente se tiene que $\mathcal{R}(A) = \{\alpha : \alpha \in \mathbb{C}\}$ y $[\mathcal{R}(A)]^\perp = \{0\}$. Una base de $\mathcal{R}(A)$ está dada por $\{1\}$ y puesto que $1 = 1 + 0$ con $1 \in \mathcal{R}(A)$ y $0 \in [\mathcal{R}(A)]^\perp$ y esta representación es única, se tiene que $P_{\mathcal{R}(A)}1 = 1$. Así, el proyector ortogonal sobre $\mathcal{R}(A)$ es $P_{\mathcal{R}(A)} = \begin{bmatrix} 1 \end{bmatrix}$.

Luego, de

$$AX = P_{\mathcal{R}(A)}$$

se tiene que $3x + 4z = 1$, y por tanto, $z = \frac{1}{4}(1 - 3x)$. Así,

$$X = \begin{bmatrix} x \\ y \\ \frac{1}{4}(1-3x) \end{bmatrix}, \qquad x, y \in \mathbb{C}.$$

Ahora, se verifica fácilmente que

$$XA = \begin{bmatrix} 3x & 0 & 4x \\ 3y & 0 & 4y \\ \frac{3}{4}(1-3x) & 0 & 1-3x \end{bmatrix}$$

es idempotente. Para que XA sea hermítica deben ser $x \in \mathbb{R}$, $3y = 0$ y $\frac{3}{4}(1-3x) = 4x$, es decir $x = \frac{3}{25}$, $y = 0$ y, por tanto,

$$X = \begin{bmatrix} \frac{3}{25} \\ 0 \\ \frac{4}{25} \end{bmatrix}.$$

2.3. Definición de Penrose

Las ecuaciones de Penrose son posiblemente las más conocidas a la hora de definir la inversa generalizada estudiada.

Definición 2.4 (Definición de Penrose)**.** *Sea $A \in \mathbb{C}^{m\times n}$. Se llama* **inversa de Penrose** *de A a la única matriz $Y \in \mathbb{C}^{n\times m}$ que satisface:*

(I) $AYA = A$,

(II) $YAY = Y$,

(III) $(AY)^* = AY$,

(IV) $(YA)^* = YA$.

De nuevo, tanto la existencia como la unicidad señaladas en la definición anterior se justificarán en el Teorema 2.2.

Según esta definición se tiene que:

- $AY(Ax) = Ax$, para todo $x \in \mathbb{C}^n$, es decir, $AY = \mathrm{Id}_{\mathcal{R}(A)}$.
- $YA(Yx) = Yx$, para todo $x \in \mathbb{C}^m$, es decir, $YA = \mathrm{Id}_{\mathcal{R}(Y)}$.

Ejemplo 2.3. *Calcular la inversa de Penrose de la matriz*

$$A = \begin{bmatrix} 3 & 0 & 4 \end{bmatrix}.$$

Como $A \in \mathbb{C}^{1\times 3}$, la inversa de Penrose será del tipo

$$Y = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{C}^{3\times 1}.$$

Se tiene que $AY = \left[\begin{array}{c} 3x+4z \end{array}\right]$. *Está claro que se cumple que* AY *es hermítica si y sólo si* $3x+4z \in \mathbb{R}$. *De* $AYA = A$ *se llega a* $3(3x+4z) = 3$ *y* $4(3x+4z) = 4$, *de donde* $3x+4z = 1$. *Es evidente que bajo esta última condición se cumple que* $YAY = Y$. *Por último, para que se cumpla que*

$$YA = \begin{bmatrix} 3x & 0 & 4x \\ 3y & 0 & 4y \\ 3z & 0 & 4z \end{bmatrix}$$

sea hermítica debe ser $x, z \in \mathbb{R}$, $y = 0$ *y* $3z = 4x$. *Resolviendo el sistema lineal se obtiene* $x = \frac{3}{25}$, $z = \frac{4}{25}$ *y finalmente*

$$Y = \begin{bmatrix} \frac{3}{25} \\ 0 \\ \frac{4}{25} \end{bmatrix}.$$

2.4. Equivalencia entre las definiciones

El hecho importante que se demostrará en el Teorema 2.2 es que estas tres definiciones son equivalentes y, por lo tanto, las matrices $[T_A^\dagger]_{\mathcal{C}',\mathcal{C}}$, X e Y de las definiciones anteriores son todas iguales y esto permite denotarlas a todas por $A^\dagger$ y se la denomina **inversa de Moore-Penrose** de A.

Teorema 2.2. *Sea* $A \in \mathbb{C}^{m\times n}$. *Las definiciones funcional (pseudoinversa), de Moore y de Penrose de la inversa generalizada de* A *son equivalentes.*

Demostración. Para simplificar la notación se recuerda que T es la aplicación lineal de la Definición 2.1 y se llamará $T^\dagger$ a la aplicación lineal $T_A^\dagger$ de la Definición 2.2 y $Z := [T_A^\dagger]_{\mathcal{C}',\mathcal{C}}$ a la matriz pseudoinversa de la Definición 2.2.

Se demostrará en varias etapas.

- <u>La matriz Z de la Definición 2.2 satisface las condiciones (a) y (b) de la Definición 2.3.</u>

  En efecto, decir que la matriz Z satisface la Definición 2.2 es equivalente a decir que la aplicación lineal $T^\dagger$ satisface la Definición 2.1. Entonces, realizando previamente el análisis funcional se tiene que

  $$TT^\dagger(x) = T(T_{|\mathcal{R}(T^*)})^{-1}P_{\mathcal{R}(T)}(x) = T(T_{|\mathcal{R}(T^*)})^{-1}(x) = \mathrm{Id}_{\mathcal{R}(T)}(x) = x \quad \text{si } x \in \mathcal{R}(T)$$

  y, por ser T y $(T_{|\mathcal{R}(T^*)})^{-1}$ aplicaciones lineales, se tiene que

  $$TT^\dagger(x) = T(T_{|\mathcal{R}(T^*)})^{-1}P_{\mathcal{R}(T)}(x) = T(T_{|\mathcal{R}(T^*)})^{-1}(0) = 0 \quad \text{si } x \in [\mathcal{R}(T)]^\perp.$$

De las igualdades demostradas y, al ser $TT^\dagger$ una aplicación lineal, se tiene que $TT^\dagger = P_{\mathcal{R}(T)}$. Ahora, el análisis matricial lleva a

$$AZ = [T]_{\mathcal{C}\mathcal{C}'}[T^\dagger]_{\mathcal{C}',\mathcal{C}} = [TT^\dagger]_{\mathcal{C}',\mathcal{C}'} = [P_{\mathcal{R}(T)}]_{\mathcal{C}',\mathcal{C}'} = P_{\mathcal{R}(A)}$$

(demostrar la última igualdad como ejercicio). Así, Z cumple (a).

De manera semejante, si $x \in \mathcal{R}(T^*)$,

$$T^\dagger T(x) = (T_{|\mathcal{R}(T^*)})^{-1}(P_{\mathcal{R}(T)}(T(x))) = (T_{|\mathcal{R}(T^*)})^{-1}(T(x)) = x$$

y además, si $x \in [\mathcal{R}(T^*)]^\perp = \mathcal{N}(T)$,

$$T^\dagger T(x) = T^\dagger(0) = 0.$$

Como, por construcción, $\mathcal{R}(T^\dagger) = \mathcal{R}(T^*)$ (véase la Nota 2.3), se ha probado que $T^\dagger T = P_{\mathcal{R}(T^\dagger)}$ y, matricialmente, se tiene que

$$ZA = [T^\dagger]_{\mathcal{C}',\mathcal{C}}[T]_{\mathcal{C},\mathcal{C}'} = [T^\dagger T]_{\mathcal{C},\mathcal{C}} = [P_{\mathcal{R}(T^\dagger)}]_{\mathcal{C},\mathcal{C}} = P_{\mathcal{R}(Z)}.$$

Así, Z también satisface (b).

- <u>La matriz X de la Definición 2.3 satisface las condiciones (I)–(IV) de la Definición 2.4.</u>

  En efecto, si X satisface la Definición 2.3 entonces, de (a), se deduce directamente (III) y, de (b), se deduce directamente (IV). Además, de (a), se tiene que $AXA = P_{\mathcal{R}(A)}A = A$ que es (I), y de (b) resulta que $XAX = P_{\mathcal{R}(X)}X = X$, que es (II).

- <u>La matriz Y de la Definición 2.4 satisface las condiciones (a) y (b) de la Definición 2.3.</u>

  En efecto, si Y satisface la Definición 2.4 entonces, de (I), se deduce inmediatamente que $(AY)^2 = AY$, es decir que AY es un proyector y, de (III), se tiene que además $(AY)^* = AY$ con lo que AY es un proyector ortogonal. Para demostrar que $\mathcal{R}(AY) = \mathcal{R}(A)$ se utiliza el apartado (I) como sigue

  $$\mathcal{R}(AY) \subseteq \mathcal{R}(A) = \mathcal{R}(AYA) \subseteq \mathcal{R}(AY).$$

  De este modo, $AY = P_{\mathcal{R}(A)}$, de donde Y satisface (a). De modo semejante, de (I) se deduce rápidamente que $(YA)^2 = YA$, y combinándolo con (IV) se tiene que YA es un proyector ortogonal. Utilizando (II) se tiene que

  $$\mathcal{R}(YA) \subseteq \mathcal{R}(Y) = \mathcal{R}(YAY) \subseteq \mathcal{R}(YA),$$

  con lo que $YA = P_{\mathcal{R}(Y)}$, de donde Y satisface (b).

Ahora bien, sin haber considerado aún la unicidad requerida en las Definiciones 2.3 y 2.4, se ha probado que si una matriz cumple las condiciones (a) y (b) también cumple las condiciones (I)-(IV) y recíprocamente. En los dos puntos restantes se probará dicha unicidad y la recíproca, es decir, si X satisface cualesquiera de los dos conjuntos de condiciones ((a)-(b) o bien (I)-(IV)) entonces X satisface la Definición 2.2.

- Las matrices X e Y de las Definiciones 2.3 y 2.4 satisfacen la Definición 2.2.

  Se probará que si X satisface (a) y (b) y, por tanto, también satisface las condiciones (I)-(IV), entonces X satisface la Definición 2.2. En efecto, se analizarán dos situaciones.

  - $x \in \mathcal{R}(A)^{\perp}$. En este caso, por (a), se tiene $AXx = P_{\mathcal{R}(A)}x = 0$ y entonces, por (II), $Xx = XAXx = X0 = 0$.
  - $x \in \mathcal{R}(A)$. En este caso, $x = Ay$ para algún $y \in \mathbb{C}^n$. Como $\mathbb{C}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^*)$, existen vectores únicos $y_1 \in \mathcal{N}(A)$ e $y_2 \in \mathcal{R}(A^*)$ tales que $y = y_1 + y_2$. Luego, $x = Ay = Ay_1 + Ay_2 = Ay_2$. Así, por (b), $Xx = XAy_2 = P_{\mathcal{R}(X)}y_2 = y_2$ pues $y_2 \in \mathcal{R}(A^*) \subseteq \mathcal{R}(X)$ (ya que de (I) y (IV) se tiene que $A^* = (AXA)^* = (XA)^*A^* = XAA^*$). Finalmente, puesto que $x = Ay_2$ con $y_2 \in \mathcal{R}(A^*)$, se tiene que $y_2 = (A_{|\mathcal{R}(A^*)})^{-1}x$ con lo que $Xx = (A_{|\mathcal{R}(A^*)})^{-1}x$.

  De esta manera, X satisface la Definición 2.2.

- Las matrices X e Y de las Definiciones 2.3 y 2.4 son únicas.

  Una forma de probarlo es mostrando que si X satisface (a) y (b) y, por tanto, también satisface las condiciones (I)-(IV), entonces X satisface la Definición 2.2, que es exactamente lo que se ha probado en el apartado anterior.

Así, la demostración queda completa. □

De este modo, dada una matriz $A \in \mathbb{C}^{m\times n}$, queda demostrada la existencia y unicidad de su matriz inversa de Moore-Penrose $A^{\dagger}$.

Ejercicio 2.3. *Sea $u \in \mathbb{C}^{n\times 1}$ un vector no nulo.*

(a) Demostrar que $u^{\dagger} = \frac{1}{u^*u}u^*$.

(b) Encontrar una condición necesaria y suficiente para que $u^{\dagger} = u^*$.

2.5. Propiedades de la inversa de Moore-Penrose

El siguiente resultado proporciona algunas propiedades que serán de utilidad en lo que sigue.

Lema 2.2. *Sea $A \in \mathbb{C}^{m\times n}$. Entonces*

(a) $A^* = A^*AA^\dagger = A^\dagger AA^*$. *En consecuencia, si* $\mathrm{rg}(A) = n$ *entonces* $A^\dagger = (A^*A)^{-1}A^*$ *y si* $\mathrm{rg}(A) = m$ *entonces* $A^\dagger = A^*(AA^*)^{-1}$.

(b) $\mathcal{R}(I - AA^\dagger) = [\mathcal{R}(A)]^\perp$ *y* $\mathcal{N}(AA^\dagger) = \mathcal{N}(A^\dagger)$.

(c) $\mathcal{R}(A^\dagger) = \mathcal{R}(A^*)$. *En consecuencia,* $\mathrm{rg}(A^\dagger) = \mathrm{rg}(A)$.

(d) $\mathcal{N}(A^\dagger) = \mathcal{N}(A^*)$.

Demostración. (a) De las propiedades de Penrose se tiene que $A^* = (AA^\dagger A)^* = A^*(AA^\dagger)^* = A^*AA^\dagger$. Para probar que $A^\dagger = (A^*A)^{-1}A^*$ basta observar que $\mathrm{rg}(A^*A) = \mathrm{rg}(A)$ y que, por hipótesis, $A^*A \in \mathbb{C}_n^{n\times n}$. De forma semejante, $A^* = (AA^\dagger A)^* = (A^\dagger A)^*A^* = A^\dagger AA^*$. La consecuencia también es inmediata en este caso.

(b) Como $AA^\dagger$ es un proyector ortogonal sobre $\mathcal{R}(A)$ (y, por lo tanto, paralelamente a $[\mathcal{R}(A)]^\perp$) es inmediato que $I - AA^\dagger = P_{[\mathcal{R}(A)]^\perp}$, con lo que $\mathcal{R}(I - AA^\dagger) = [\mathcal{R}(A)]^\perp$.

Por otro lado, es claro que $\mathcal{N}(A^\dagger) \subseteq \mathcal{N}(AA^\dagger)$. Además, si $x \in \mathcal{N}(AA^\dagger)$ entonces $AA^\dagger x = 0$ y premultiplicando por $A^\dagger$ se tiene $A^\dagger x = A^\dagger AA^\dagger x = 0$, es decir, $x \in \mathcal{N}(A^\dagger)$.

(c) De la definición de Penrose se tiene que

$$\mathcal{R}(A^\dagger) = \mathcal{R}(A^\dagger AA^\dagger) \subseteq \mathcal{R}(A^\dagger A) = \mathcal{R}((A^\dagger A)^*) = \mathcal{R}(A^*(A^\dagger)^*) \subseteq \mathcal{R}(A^*).$$

Por (a),

$$\mathcal{R}(A^*) = \mathcal{R}(A^\dagger AA^*) \subseteq \mathcal{R}(A^\dagger).$$

Esta propiedad ha sido probada en la Nota 2.3 a partir de la definición funcional.

(d) Puesto que

$$\mathcal{N}(A^*) \subseteq \mathcal{N}((A^\dagger)^*A^*) = \mathcal{N}((AA^\dagger)^*) = \mathcal{N}(AA^\dagger) \subseteq \mathcal{N}(A^\dagger AA^\dagger) = \mathcal{N}(A^\dagger),$$

y teniendo en cuenta que $\mathrm{rg}(A^*) = \mathrm{rg}(A^\dagger)$ se obtiene que $\dim(\mathcal{N}(A^*)) = \dim(\mathcal{N}(A^\dagger))$ con lo que se llega a $\mathcal{N}(A^\dagger) = \mathcal{N}(A^*)$. □

Ejercicio 2.4. *(a) Una vez que $A^\dagger$ ha sido establecida (por ejemplo mediante las ecuaciones de Penrose) o construida (mediante la Definición 2.2), del Lema 2.2 se tiene que es válida la propiedad $\mathcal{R}(A^\dagger) = \mathcal{R}(A^*)$. Sin embargo, no es posible usar $\mathcal{R}(A^*)$ en lugar de $\mathcal{R}(A^\dagger)$ en la propia definición de Moore. Para constatarlo, considerar las matrices*

$$A = \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix}, \qquad X = \begin{bmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{\alpha} \end{bmatrix}, \qquad \textit{con} \quad \alpha \in \mathbb{C} - \{0\}$$

y comprobar que $AX = XA = \frac{1}{\alpha}A$ *y*

$$AX = P_{\mathcal{R}(A)}, \qquad XA = P_{\mathcal{R}(A^*)},$$

pero, sin embargo,

$$X \neq A^\dagger = \begin{bmatrix} \frac{1}{\alpha} & 0 \\ 0 & 0 \end{bmatrix}.$$

De hecho, $XA \neq P_{\mathcal{R}(X)}$ *y por tanto* $XAX \neq X$. *El siguiente apartado muestra que es posible usar* $\mathcal{R}(A^*)$ *en lugar de* $\mathcal{R}(X)$ *en la Definición de Moore cuando se añade la condición* $XAX = X$.

(b) Sea $A \in \mathbb{C}^{m\times n}$. *Demostrar que*

$$X = A^\dagger \qquad \Longleftrightarrow \qquad AX = P_{\mathcal{R}(A)},\ XA = P_{\mathcal{R}(A^*)}\ \text{ y }\ XAX = X.$$

2.6. Las composiciones entre las aplicaciones T y $T^\dagger$ y entre sus restricciones

Sea $T : \mathbb{C}^n \to \mathbb{C}^m$ una aplicación lineal. Si se consideran las dos composiciones entre las restricciones $T_{|\mathcal{R}(T^*)}$ y $T^\dagger_{|\mathcal{R}(T)}$:

$$T_{|\mathcal{R}(T^*)}T^\dagger_{|\mathcal{R}(T)} : \mathcal{R}(T) \to \mathcal{R}(T) \qquad \text{y} \qquad T^\dagger_{|\mathcal{R}(T)}T_{|\mathcal{R}(T^*)} : \mathcal{R}(T^*) \to \mathcal{R}(T^*), \tag{2.5}$$

y además, $\mathcal{B}$ es una base ortonormal de $\mathcal{R}(T^*)$ y $\mathcal{B}'$ es una base ortonormal de $\mathcal{R}(T)$, de la Definición 2.1 es fácil ver (ejercicio) que la aplicación $T^\dagger_{|\mathcal{R}(T)}$ cumple

$$[T_{|\mathcal{R}(T^*)}]_{\mathcal{B},\mathcal{B}'}[T^\dagger_{|\mathcal{R}(T)}]_{\mathcal{B}',\mathcal{B}} = [T_{|\mathcal{R}(T^*)}T^\dagger_{|\mathcal{R}(T)}]_{\mathcal{B}',\mathcal{B}'} = I_r$$

y

$$[T^\dagger_{|\mathcal{R}(T)}]_{\mathcal{B}',\mathcal{B}}[T_{|\mathcal{R}(T^*)}]_{\mathcal{B},\mathcal{B}'} = [T^\dagger_{|\mathcal{R}(T)}T_{|\mathcal{R}(T^*)}]_{\mathcal{B},\mathcal{B}} = I_r,$$

donde $r = \operatorname{rg}(T)$; lo que de otro modo se expresa diciendo que la matriz inversa de $[T_{|\mathcal{R}(T^*)}]_{\mathcal{B},\mathcal{B}'}$ es $[T^\dagger_{|\mathcal{R}(T)}]_{\mathcal{B}',\mathcal{B}}$, i.e.,

$$[T^\dagger_{|\mathcal{R}(T)}]_{\mathcal{B}',\mathcal{B}} = ([T_{|\mathcal{R}(T^*)}]_{\mathcal{B},\mathcal{B}'})^{-1}. \tag{2.6}$$

Ejercicio 2.5. *Comparar el tamaño de las matrices de la relación (2.6) con el de la matriz de la Definición 2.2.*

Si las bases $\mathcal{B}$ y $\mathcal{B}'$ anteriores se completan a las bases ortonormales $\mathcal{B}_1$ y $\mathcal{B}'_1$ de $\mathbb{C}^n$ y $\mathbb{C}^m$ con bases ortonormales de $\mathcal{N}(T)$ y $\mathcal{N}(T^*)$ (poniendo en las primeras posiciones los vectores de las bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$) respectivamente, se tiene que

$$[T]_{\mathcal{B}_1,\mathcal{B}'_1}[T^\dagger]_{\mathcal{B}'_1,\mathcal{B}_1} = [TT^\dagger]_{\mathcal{B}'_1,\mathcal{B}'_1} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \in \mathbb{C}^{m\times m},$$

que es la matriz del proyector ortogonal de $\mathbb{C}^m$ sobre $\mathcal{R}(T)$ en la base $\mathcal{B}'_1$, y

$$[T^\dagger]_{\mathcal{B}'_1,\mathcal{B}_1}[T]_{\mathcal{B}_1,\mathcal{B}'_1} = [T^\dagger T]_{\mathcal{B}_1,\mathcal{B}_1} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \in \mathbb{C}^{n\times n},$$

que es la matriz del proyector ortogonal de $\mathbb{C}^n$ sobre $\mathcal{R}(T^*)$ en la base $\mathcal{B}_1$.

De todo lo expuesto anteriormente se deduce el siguiente resultado, que es lo que en realidad demostró Moore.

Teorema 2.3. *Si $T : \mathbb{C}^n \to \mathbb{C}^m$ es una aplicación lineal entonces existe una única aplicación lineal*

$$X : \mathcal{R}(T) \to \mathcal{R}(T^*)$$

tal que

$$TX = P_{\mathcal{R}(T)} \qquad \text{y} \qquad XT = P_{\mathcal{R}(T^*)}.$$

Ejercicio 2.6. *Observar la importancia del hecho que la aplicación X sea $X : \mathcal{R}(T) \to \mathcal{R}(T^*)$ en el Teorema 2.3 (y de la imposibilidad de cambiarlo por $X : \mathbb{C}^m \to \mathbb{C}^n$). Comparar con el Ejercicio 2.4.*

En la práctica, se tiene una forma de calcular los proyectores ortogonales sobre $\mathcal{R}(A)$ y sobre $\mathcal{R}(A^*)$ mediante la inversa de Moore-Penrose de una matriz.

Teorema 2.4. *Para cada matriz $A \in \mathbb{C}^{m\times n}$ su inversa de Moore-Penrose $A^\dagger$ cumple que:*

$$P_{\mathcal{R}(A)} = AA^\dagger \qquad \text{y} \qquad P_{\mathcal{R}(A^*)} = A^\dagger A.$$

Ejercicio 2.7. *Dada la aplicación lineal $T : \mathbb{C}^3 \to \mathbb{C}^2$ definida por $T(x) = Ax$ para $x \in \mathbb{C}^3$ siendo*

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

se pide:

(a) Hallar la expresión de T^*,

(b) *Calcular el espacio nulo y el espacio imagen de T y de T^* y sus dimensiones,*

(c) *Expresar los espacios $\mathbb{C}^3$ y $\mathbb{C}^2$ como suma directa de los subespacios fundamentales hallados en el apartado anterior,*

(d) *Determinar $T_{|\mathcal{R}(T^*)} : \mathcal{R}(T^*) \to \mathcal{R}(T)$ y comprobar que es un isomorfismo,*

(e) *Calcular la expresión de la aplicación lineal $T^\dagger$ a partir de la Definición 2.1,*

(f) *Calcular la expresión de $A^\dagger$ a partir de la Definición 2.3 de Moore,*

(g) *Calcular la expresión de $A^\dagger$ a partir de la Definición 2.4 de Penrose,*

(h) *Hallar el espacio nulo y el espacio imagen de $A^\dagger$,*

(i) *Calcular $P_{\mathcal{R}(A)}$ y $P_{\mathcal{R}(A^*)}$ a partir de $A^\dagger$,*

(j) *Comprobar que $A^\dagger = \frac{1}{4}A^*$.*

Capítulo 3

El problema de los mínimos cuadrados

Se recuerdan los siguientes hechos:

- Para un vector $u \in \mathbb{C}^{n\times 1}$, se define su **norma euclídea** como $\|u\|_2 = \sqrt{u^* u}$.
- Dos vectores $u, v \in \mathbb{C}^{n\times 1}$ se llaman **ortogonales** si $u^* v = 0$.
- **Teorema de Pitágoras**: Si $u, v \in \mathbb{C}^{n\times 1}$ son ortogonales entonces $\|u+v\|_2^2 = \|u\|_2^2 + \|v\|_2^2$.

3.1. Planteamiento del problema

PROBLEMA: Encontrar las soluciones $u \in \mathbb{C}^{n\times 1}$ de

$$Ax = b, \qquad \text{donde} \quad A \in \mathbb{C}^{m\times n} \quad \text{y} \quad b \in \mathbb{C}^{m\times 1}.$$

Si el problema no tiene solución, se hallará u de modo que $Au - b$ sea lo más pequeño posible en el sentido que se define a continuación.

Definición 3.1. *Sean $A \in \mathbb{C}^{m\times n}$ y $b \in \mathbb{C}^{m\times 1}$. Un vector $u \in \mathbb{C}^{n\times 1}$ se dice que es*

- *una* **solución de mínimos cuadrados** *de $Ax = b$ si $\|Au - b\|_2 \leq \|Av - b\|_2$ para todo $v \in \mathbb{C}^{n\times 1}$, es decir,*

$$\min_{v\in\mathbb{C}^{n\times 1}} \|Av - b\|_2 = \|Au - b\|_2.$$

- *una* **solución de mínimos cuadrados de norma mínima** *de $Ax = b$ si es una solución de mínimos cuadrados de $Ax = b$ y $\|u\|_2 \leq \|w\|_2$, para toda otra solución de mínimos cuadrados $w \in \mathbb{C}^{n\times 1}$, es decir,*

$$\|u\|_2 = \min\{\|w\|_2 : w \in \mathbb{C}^{n\times 1} \text{ satisface } \min_{v\in\mathbb{C}^{n\times 1}} \|Av - b\|_2 = \|Aw - b\|_2\}.$$

Nota 3.1. *Si $b \in \mathcal{R}(A)$, los conceptos de solución de $Ax = b$ y solución de mínimos cuadrados coinciden.*

Ejemplo 3.1. *Se considera*

$$A = \begin{bmatrix} 1 & 1 \end{bmatrix} \in \mathbb{R}^{1\times 2} \qquad \text{y} \qquad b = 2.$$

Como $b \in \mathcal{R}(A)$, cualquier solución de $Ax = b$ (que es del tipo $\begin{bmatrix} x_1 \\ 2 - x_1 \end{bmatrix}$, $x_1 \in \mathbb{R}$) es una solución de mínimos cuadrados de $Ax = b$ y recíprocamente. Sin embargo, no toda solución de mínimos cuadrados de $Ax = b$ es de norma mínima. Es claro que

$$\left\| \begin{bmatrix} x_1 \\ 2 - x_1 \end{bmatrix} \right\|_2^2 = x_1^2 + (2 - x_1)^2 = 2x_1^2 - 4x_1 + 4 = 2\left(x_1 - 1\right)^2 + 2$$

alcanza su valor mínimo en $x_1 = 1$ y, por tanto, la solución de mínimos cuadrados de norma mínima de $Ax = b$ viene dada por $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Ejercicio 3.1. *Como el ejemplo anterior se ha resuelto en el cuerpo de los números reales, interpretar geométricamente en el plano cartesiano las soluciones de mínimos cuadrados de $Ax = b$ y la de norma mínima.*

Ejercicio 3.2. *Idem a los dos ejemplos anteriores para las matrices:*

(a)

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2} \qquad \text{y} \qquad b = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

(b)

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \in \mathbb{R}^{3\times 2} \qquad \text{y} \qquad b = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix}$$

encontrando las soluciones a partir de métodos de optimización vistos en Análisis Matemático o en Álgebra Lineal.

3.2. Soluciones geométrica y algebraicas

En el siguiente resultado se dan todas (en caso de haber más de una) las soluciones de mínimos cuadrados de una ecuación $Ax = b$ especificando la que proporciona la de norma mínima. Es importante resaltar el protagonismo indiscutible que cobra la inversa de Moore-Penrose en dicha solución.

Teorema 3.1. *Sean* $A \in \mathbb{C}^{m\times n}$ *y* $b \in \mathbb{C}^{m\times 1}$.

(a) El conjunto S *de todas las soluciones de mínimos cuadrados de* $Ax = b$ *es*

$$\begin{aligned} S &= \{x \in \mathbb{C}^{n\times 1} : Ax = P_{\mathcal{R}(A)}b\} && (\textbf{Solución geométrica}) \\ &= \{x \in \mathbb{C}^{n\times 1} : A^*Ax = A^*b\} && (\textbf{Ecuaciones normales}) \\ &= A^\dagger b + \mathcal{N}(A) && (\textbf{Solución algebraica}) \\ &= A^\dagger b + \{(I_n - A^\dagger A)h : h \in \mathbb{C}^{n\times 1}\}. \end{aligned}$$

(b) La (única) solución de mínimos cuadrados de norma mínima de $Ax = b$ *es* $A^\dagger b$.

Demostración. (a) Para $x \in \mathbb{C}^{n\times 1}$ arbitrario se tiene que

$$Ax - b = Ax - AA^\dagger b + AA^\dagger b - b = A(x - A^\dagger b) - (I - AA^\dagger)b \in \mathcal{R}(A) + \mathcal{R}(I - AA^\dagger). \quad (3.1)$$

El Lema 2.2 asegura que $A(x - A^\dagger b) \perp -(I - AA^\dagger)b$ y entonces el teorema de Pitágoras permite establecer que

$$\|Ax - b\|_2^2 = \|A(x - A^\dagger b)\|_2^2 + \| - (I - AA^\dagger)b\|_2^2 \geq \|A(A^\dagger b) - b\|_2^2, \text{ para todo } x \in \mathbb{C}^{n\times 1}. \quad (3.2)$$

Como (3.2) se cumple para todo $x \in \mathbb{C}^{n\times 1}$, el conjunto S de todas soluciones de mínimos cuadrados de $Ax = b$ debe coincidir con el conjunto solución (llamada **solución geométrica**) del sistema

$$Ax = AA^\dagger b = P_{\mathcal{R}(A)}b, \quad (3.3)$$

pues

$$\min_{x\in\mathbb{C}^{n\times 1}} \left(\|Ax - b\|_2^2\right) = \min_{x\in\mathbb{C}^{n\times 1}} \left(\|A(x - A^\dagger b)\|_2^2\right) + \|(I - AA^\dagger)b\|_2^2, \quad (3.4)$$

donde $\|(I - AA^\dagger)b\|_2^2$ está fijo y (3.3) es un sistema compatible.

Teniendo en cuenta (3.1), se tiene que x satisface (3.3) si y sólo si $Ax - b \in [\mathcal{R}(A)]^\perp$, es decir, $(Ay)^*(Ax - b) = 0$, $\forall y \in \mathbb{C}^{n\times 1}$, lo que equivale a $y^*A^*(Ax - b) = 0$, $\forall y \in \mathbb{C}^{n\times 1}$. Particularizando y en los vectores canónicos se tiene que el conjunto S de todas las soluciones de mínimos cuadrados de $Ax = b$ también debe coincidir con el conjunto solución del sistema

$$A^*Ax = A^*b,$$

que son las llamadas **ecuaciones normales**.

Además, de (3.4),

$$\min_{x\in\mathbb{C}^{n\times 1}} \|Ax - b\|_2 = \|A(A^\dagger b) - b\|_2,$$

con lo que $A^\dagger b$ es una solución de mínimos cuadrados de $Ax = b$.

Ahora, recordando que el conjunto S de todas las soluciones del sistema (3.3) es la suma de una solución particular (en este caso es claro que $A^\dagger b$ es una solución particular) más la solución general (o el conjunto solución S_H) del sistema homogéneo asociado $Ax = 0$, se tiene que $S = A^\dagger b + S_H$. Es fácil ver que $\mathcal{N}(A) = \mathcal{N}(A^\dagger A) = \mathcal{R}(I - A^\dagger A)$ (ejercicio), con lo que

$$S = A^\dagger b + \mathcal{N}(A) = A^\dagger b + \{(I_n - A^\dagger A)h : h \in \mathbb{C}^{n\times 1} \text{ arbitrario}\}. \tag{3.5}$$

(b) Del apartado (a), $A^\dagger b$ es una solución de mínimos cuadrados de $Ax = b$. Falta probar que dicha solución de mínimos cuadrados de $Ax = b$ tiene norma mínima y es la única en dichas condiciones. En efecto, es claro que los vectores de S cumplen que

$$A^\dagger b + (I - A^\dagger A)h \in \mathcal{R}(A^\dagger) + \mathcal{R}(I - A^\dagger A), \quad \text{para todo } h \in \mathbb{C}^n.$$

Por ser $\mathcal{R}(I - A^\dagger A) = \mathcal{N}(A) = [\mathcal{R}(A^*)]^\perp = [\mathcal{R}(A^\dagger)]^\perp$, el teorema de Pitágoras asegura que

$$\|A^\dagger b + (I - A^\dagger A)h\|_2^2 = \|A^\dagger b\|_2^2 + \|(I - A^\dagger A)h\|_2^2 \geq \|A^\dagger b\|_2^2, \quad \text{para todo } h \in \mathbb{C}^{n\times 1}.$$

Por lo tanto, $A^\dagger b$ es una solución de mínimos cuadrados de norma mínima de $Ax = b$.

Para ver que es la única en esta situación, si $z \neq A^\dagger b$ fuese otra solución de mínimos cuadrados de norma mínima de $Ax = b$, debería existir $h_z \in \mathbb{C}^{n\times 1}$ tal que $z = A^\dagger b + (I - A^\dagger A)h_z$ con $(I - A^\dagger A)h_z \neq 0$. Luego, $\|z\|_2^2 = \|A^\dagger b\|_2^2 + \|(I - A^\dagger A)h_z\|_2^2 > \|A^\dagger b\|_2^2$, lo que contradice la minimalidad de $\|z\|_2$. □

Ejemplo 3.2. *Sean*

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \in \mathbb{C}^{3\times 3} \qquad \text{y} \qquad b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

En este caso, $b \notin \mathcal{R}(A)$. Primero se calcula (hacerlo como ejercicio) la inversa de Moore-Penrose de A, que es

$$A^\dagger = \frac{1}{9}\begin{bmatrix} 5 & -4 & 1 \\ -4 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Ahora,

$$A^\dagger b = \frac{1}{9}\begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} \qquad \text{e} \qquad I - A^\dagger A = \frac{1}{3}\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

Utilizando (3.5) se tiene que las soluciones de mínimos cuadrados del sistema $Ax = b$ son de la forma

$$\begin{bmatrix} \frac{4}{9} \\ \frac{4}{9} \\ \frac{8}{9} \end{bmatrix} + \frac{1}{3}\begin{bmatrix} h_1 + h_2 - h_3 \\ h_1 + h_2 - h_3 \\ -h_1 - h_2 + h_3 \end{bmatrix}, h_1, h_2, h_3 \in \mathbb{C},$$

y la de norma mínima viene dada por

$$A^\dagger b = \begin{bmatrix} \frac{4}{9} \\ \frac{4}{9} \\ \frac{8}{9} \end{bmatrix}.$$

Ejemplo 3.3. *Sean*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{C}^{3\times 2} \qquad \text{y} \qquad b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

En este caso, $b \notin \mathcal{R}(A)$. Primero se calcula (hacerlo como ejercicio) la inversa de Moore-Penrose de A, que es

$$A^\dagger = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}.$$

Como

$$I - A^\dagger A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

utilizando (3.5), se tiene que la única solución de mínimos cuadrados del sistema $Ax = b$ coincide con la solución de norma mínima y viene dada por

$$A^\dagger b = \begin{bmatrix} \frac{4}{3} \\ \frac{4}{3} \end{bmatrix}.$$

3.3. Algunos comentarios

Definición 3.2. *Sean $A \in \mathbb{C}^{m\times n}$ y $b \in \mathbb{C}^{m\times 1}$. Si el vector $y \in \mathbb{C}^{n\times 1}$ cumple que*

$$\min_{x\in\mathbb{C}^{n\times 1}} \|Ax - b\|_2 = \|Ay - b\|_2,$$

se llama **residuo de mínimos cuadrados** *al vector $Ay - b$.*

Las siguientes son algunas observaciones importantes. Sean $A \in \mathbb{C}^{m\times n}$ y $b \in \mathbb{C}^{m\times 1}$.

- Unicidad del residuo de mínimos cuadrados: Todas las soluciones $y \in \mathbb{C}^{n\times 1}$ de mínimos cuadrados de $Ax = b$ tienen el mismo residuo $Ay - b = -(I_m - AA^\dagger)b$.

  En efecto, si y_1 e y_2 son dos soluciones de mínimos cuadrados de $Ax = b$ entonces $y_1 = A^\dagger b + n_1$ e $y_2 = A^\dagger b + n_2$, con $n_1, n_2 \in \mathcal{N}(A)$. Luego, $Ay_1 = AA^\dagger b = Ay_2$ y, por tanto, $Ay_1 - b = Ay_2 - b = AA^\dagger b - b = -(I_m - AA^\dagger)b$.

  También se podría justificar este punto a partir de la igualdad $Ax - b = -(I_m - AA^\dagger)b$ de (3.1) en el caso en que x sea solución de mínimos cuadrados de $Ax = b$.

- El residuo es ortogonal al espacio imagen de A: Si $y \in \mathbb{C}^{n\times 1}$ es una solución de mínimos cuadrados de $Ax = b$ entonces $Ay - b \in [\mathcal{R}(A)]^{\perp}$.

  Es inmediato pues en la demostración del Teorema 3.1 se probó que si $y \in \mathbb{C}^{n\times 1}$ es una solución de mínimos cuadrados de $Ax = b$ entonces $Ay - b \in [\mathcal{R}(A)]^{\perp}$.

- Matriz A de rango completo: Si $\mathrm{rg}(A) = n$ entonces la única solución de mínimos cuadrados de norma mínima de $Ax = b$ es $(A^*A)^{-1}A^*b$. En este caso, $A^\dagger A = I_n$, es decir, $A^\dagger$ es una inversa a izquierda de A. Si $\mathrm{rg}(A) = m$ entonces esta única solución viene dada por $A^*(AA^*)^{-1}b$. En este caso, $AA^\dagger = I_m$, es decir, $A^\dagger$ es una inversa a derecha de A. Y si A es invertible ($\mathrm{rg}(A) = m = n$) entonces dicha solución es $A^{-1}b$.

  Este hecho es una consecuencia directa del Lema 2.2 y del Teorema 3.1 (b).

- Caracterización de la unicidad de la solución de mínimos cuadrados: El conjunto S de las soluciones de mínimos cuadrados de $Ax = b$ tiene un solo elemento (solución única) si y sólo si $\mathrm{rg}(A) = n$.

  La necesidad se deduce del hecho que si $\mathrm{rg}(A) = n$, del apartado anterior, la única solución de mínimos cuadrados de $Ax = b$ es $(A^*A)^{-1}A^*b$. Para demostrar la suficiencia, se supone que S tiene un solo elemento. Del Teorema 3.1 este elemento debe ser $A^\dagger b$ y además $A^\dagger Ah = h$ para todo $h \in \mathbb{C}^{n\times 1}$. Luego, $A^\dagger A = I_n$, de donde $n = \mathrm{rg}(A^\dagger A) \le \mathrm{rg}(A) \le n$.

Ejercicio 3.3. *Se considera la familia de matrices y el vector siguientes:*

$$A_\alpha = \begin{bmatrix} \alpha & 0 & -i \\ i & 0 & \alpha \end{bmatrix} \qquad \text{y} \qquad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

en función del parámetro $\alpha \in \mathbb{C}$. Se pide, en función del valor de α:

(a) Hallar el rango de A_α,

(b) En función de los valores del rango hallado en el apartado anterior, caracterizar la unicidad de la solución de mínimos cuadrados del sistema $A_\alpha x = b$,

(c) Calcular las soluciones de mínimos cuadrados del sistema $A_\alpha x = b$ de tres formas posibles:

(I) solución geométrica,

(II) ecuaciones normales,

(III) solución algebraica,

(d) Determinar la solución de mínimos cuadrados de norma mínima del sistema $A_\alpha x = b$,

(e) *¿Cuál es el mínimo valor que toma* $\|A_\alpha x - b\|_2$ *para* x *variando en* $\mathbb{C}^{3\times 1}$*?*

(f) *Calcular el/los residuo/s de mínimos cuadrados del sistema* $A_\alpha x = b$ *y la norma 2 de dicho/s residuo/s.*

(g) *Comprobar que el/los residuo/s de mínimos cuadrados del sistema* $A_\alpha x = b$ *es/son ortogonal/es al espacio imagen de* A_α*.*

Capítulo 4

Métodos para calcular la inversa de Moore-Penrose

Se presentan cuatro métodos de cálculo para la inversa de Moore-Penrose.

4.1. Método basado en la pseudoinversa

El método utilizado en el Ejemplo 2.1 es válido en general.

Nota 4.1. *De la Definición 2.2 se observa que la pseudoinversa de la matriz nula es ella misma.*

Proposición 4.1. *Sea $A \in \mathbb{C}^{m\times n}$ una matriz de rango $r > 0$. Si $\{v_1, \cdots, v_r\}$ es una base de $\mathcal{R}(A^*)$ y $\{w_1, \cdots, w_{m-r}\}$ es una base de $\mathcal{N}(A^*)$ entonces*

$$A^\dagger = \begin{bmatrix} v_1 & \cdots & v_r & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} Av_1 & \cdots & Av_r & w_1 & \cdots & w_{m-r} \end{bmatrix}^{-1}. \tag{4.1}$$

Demostración. Como $\{v_1, \cdots, v_r\}$ es una base de $\mathcal{R}(A^*)$, del Teorema 2.1 es posible asegurar que $\{Av_1, \cdots, Av_r\}$ es una base de $\mathcal{R}(A)$. Al ser $\{w_1, \cdots, w_{m-r}\}$ una base de $\mathcal{N}(A^*) = [\mathcal{R}(A)]^\perp$ resulta que $\{Av_1, \cdots, Av_r, w_1, \cdots, w_{m-r}\}$ forma una base de $\mathbb{C}^m$ y, por ende, la matriz cuyas columnas son los vectores de esta base debe ser invertible.

Ahora, de $A^\dagger A = P_{\mathcal{R}(A^*)}$ se tiene que

$$\begin{aligned} A^\dagger \begin{bmatrix} Av_1 & \cdots & Av_r & w_1 & \cdots & w_{m-r} \end{bmatrix} &= \begin{bmatrix} A^\dagger Av_1 & \cdots & A^\dagger Av_r & A^\dagger w_1 & \cdots & A^\dagger w_{m-r} \end{bmatrix} \\ &= \begin{bmatrix} v_1 & \cdots & v_r & 0 & \cdots & 0 \end{bmatrix}. \end{aligned}$$

Despejando se obtiene la fórmula (4.1) que permite calcular $A^\dagger$. $\square$

La Proposición 4.1 permite observar, una vez más, la estrecha relación entre $A^\dagger$ y A^*, del hecho que para calcular $A^\dagger$ alcanza con disponer de una base de $\mathcal{N}(A^*)$ y una de $\mathcal{R}(A^*)$.

Corolario 4.1. *Si en la Proposición 4.1 se considera también una base* $\{b_{r+1}, \cdots, b_n\}$ *de* $\mathcal{N}(A)$, *y se definen las matrices invertibles*

$$Q := \begin{bmatrix} v_1 & \cdots & v_r & b_{r+1} & \cdots & b_n \end{bmatrix} \in \mathbb{C}^{n\times n},$$

y

$$T = \begin{bmatrix} Av_1 & \cdots & Av_r & w_1 & \cdots & w_{m-r} \end{bmatrix},$$

entonces

$$A^\dagger = Q \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}_{n\times m} T^{-1}.$$

Demostración. Basta observar que en la proposición se puede reescribir

$$\begin{bmatrix} v_1 & \cdots & v_r & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_r & b_{r+1} & \cdots & b_n \end{bmatrix} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}_{n\times m}.$$

□

En el corolario anterior se observa que $A^\dagger$ es equivalente a la matriz $\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}_{n\times m}$. En la siguiente sección se mejorará el resultado en el sentido que se conseguirá que la equivalencia sea mediante matrices unitarias.

4.2. Método basado en la descomposición en valores singulares

Es conocido (véase, por ejemplo, [25]) que toda matriz rectangular $A \in \mathbb{C}_r^{m\times n}$, con $r > 0$, admite siempre una descomposición en valores singulares, es decir existen dos matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = U \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{m\times n} V^* \tag{4.2}$$

con $D = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$ una matriz diagonal donde $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

Suponiendo que

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \quad \text{y} \quad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \quad \text{con } U_1 \in \mathbb{C}^{m\times r}, V_1 \in \mathbb{C}^{n\times r}$$

de la relación $AV = U \begin{bmatrix} D & O \\ O & O \end{bmatrix}$, se observan los siguientes hechos:

- $AV_1 = U_1 D$, es decir $A(V_1 D^{-1}) = U_1$,
- $AV_2 = O$,

de donde se desprende que

- las columnas de U_1 forman una base ortonormal de $\mathcal{R}(A)$,
- las columnas de V_1 forman una base ortonormal de $\mathcal{R}(A^*)$ (pues las de V_2 lo son de $\mathcal{N}(A)$, V es unitaria y $\mathbb{C}^n = \mathcal{N}(A) \oplus^\perp \mathcal{R}(A^*)$),
- las columnas de U_2 forman una base ortonormal de $\mathcal{R}(A^*)$ (pues las de U_1 lo son de $\mathcal{R}(A)$, U es unitaria y $\mathbb{C}^m = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A^*)$).

Teorema 4.1. *Sea $A \in \mathbb{C}^{m\times n}$ una matriz de rango $r > 0$ con una descomposición en valores singulares como en (4.2). Entonces su inversa de Moore-Penrose viene dada por*

$$A^\dagger = V \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}_{n\times m} U^*.$$

Demostración. De las observaciones previas y recordando la definición de matriz pseudoinversa y que $\mathcal{N}(A^\dagger) = \mathcal{N}(A^*)$ se tiene que:

$$A^\dagger \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} A^\dagger U_1 & A^\dagger U_2 \end{bmatrix} = \begin{bmatrix} V_1 D^{-1} & O \end{bmatrix} = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}.$$

De

$$A^\dagger U = V \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}$$

se obtiene inmediatamente el resultado. □

Nota 4.2. *En el método basado en la pseudoinversa se calculan bases de los cuatro subespacios fundamentales $\mathcal{R}(A^*)$, $\mathcal{N}(A^*)$, $\mathcal{R}(A)$ y $\mathcal{N}(A)$ para determinar la inversa de Moore-Penrose (véase Corolario 4.1). Sin embargo, aunque dichas bases se elijan ortonormales, en general no pueden utilizarse para calcular la inversa de Moore-Penrose a partir de una descomposición en valores singulares, puesto que no es posible asegurar que satisfagan la condición $AV = U \begin{bmatrix} D & O \\ O & O \end{bmatrix}$. (Como ejercicio, comprobar este hecho con las bases calculadas en el Ejemplo 2.1).*

Ejemplo 4.1. *Calcular la inversa de Moore-Penrose, a partir de una descomposición en valores singulares, de*

$$A = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ \sqrt{2} & -\sqrt{2} \end{bmatrix}.$$

En primer lugar se diagonaliza la matriz A^*A (hermítica y definida no negativa) obteniendo

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

es decir los valores propios son $\lambda_1 = 4$ y $\lambda_2 = 1$ (en realidad es definida positiva) y una base ortonormal de vectores propios de A^*A conforman las columnas de

$$V = \left[\begin{array}{c|c} v_1 & v_2 \end{array}\right] = \left[\begin{array}{c|c} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{array}\right].$$

Ahora se definen $\sigma_1 = \sqrt{\lambda_1} = 2$ y $\sigma_2 = \sqrt{\lambda_2} = 1$ y se calcula

$$\left[\begin{array}{c|c} Av_1 & Av_2 \end{array}\right] = AV = \left[\begin{array}{r|c} 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \\ -2 & 0 \end{array}\right];$$

así

$$Av_1 = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} \qquad \text{y} \qquad Av_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}.$$

Definiendo

$$u_1 = \frac{1}{\sigma_1} Av_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \qquad \text{y} \qquad u_2 = \frac{1}{\sigma_2} Av_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix},$$

y completando con

$$u_3 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix},$$

a una base ortonormal de $\mathbb{C}^3$, se puede definir

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

de modo que una descomposición en valores singulares para A es

$$A = U \begin{bmatrix} D \\ O \end{bmatrix}_{3\times 2} V^*$$

con

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Utilizando el Teorema 4.1 se tiene que

$$A^\dagger = V \begin{bmatrix} D^{-1} & O \end{bmatrix}_{2\times 3} U^* = \begin{bmatrix} 1/2 & 1/2 & \sqrt{2}/4 \\ 1/2 & 1/2 & -\sqrt{2}/4 \end{bmatrix}.$$

4.3. Método basado en el algoritmo de Gauss-Jordan

En el siguiente teorema se verá un método que se basa en la idea del método de Gauss-Jordan para el cálculo de una inversa ordinaria. Esquemáticamente, este método indica que para calcular la inversa de una matriz $A \in \mathbb{C}^{n\times n}$ se debe premultiplicar la matriz por bloques $\begin{bmatrix} A & I_n \end{bmatrix}$ por matrices elementales por filas y, si se llega a una matriz del tipo $\begin{bmatrix} I_n & B \end{bmatrix}$, entonces se puede concluir que A es invertible y $B = A^{-1}$. Es decir, si existe una matriz $B \in \mathbb{C}^{n\times n}$ tal que

$$B \begin{bmatrix} A & I_n \end{bmatrix} = \begin{bmatrix} I_n & B \end{bmatrix}$$

entonces $BA = I_n$ y, por tanto, A es invertible y $A^{-1} = B$.

Ahora se procederá de manera semejante, para el caso en que A no sea invertible (ni siquiera se requiere que sea cuadrada).

Teorema 4.2. *Sea $A \in \mathbb{C}^{m\times n}$ una matriz de rango $r > 0$. Si $\begin{bmatrix} B \\ O \end{bmatrix}$ es la forma escalonada reducida por filas de A^*, con $B \in \mathbb{C}_r^{r\times m}$, y F es el producto de las matrices elementales que lleva A^* a su forma escalonada reducida, es decir,*

$$F \begin{bmatrix} A^* & I_n \end{bmatrix} = \begin{bmatrix} B & F_1 \\ O & F_2 \end{bmatrix}, \tag{4.3}$$

con $F_1 \in \mathbb{C}_r^{r\times n}$ y $F_2 \in \mathbb{C}_{n-r}^{(n-r)\times n}$, entonces $\begin{bmatrix} BA \\ F_2 \end{bmatrix}$ es invertible y

$$A^\dagger = \begin{bmatrix} BA \\ F_2 \end{bmatrix}^{-1} \begin{bmatrix} B \\ O \end{bmatrix}.$$

Demostración. Sean $F \in \mathbb{C}_n^{n\times n}$, $B \in \mathbb{C}_r^{r\times m}$, $F_1 \in \mathbb{C}^{r\times n}$ y $F_2 \in \mathbb{C}^{(n-r)\times n}$ matrices tales que

$$F \begin{bmatrix} A^* & I_n \end{bmatrix} = \begin{bmatrix} B & F_1 \\ O & F_2 \end{bmatrix}.$$

Como $\operatorname{rg}(F) = n$, se tiene que $\operatorname{rg}(F_1) = r$ y $\operatorname{rg}(F_2) = n - r$. Comparando los bloques de acuerdo a sus tamaños, de la igualdad

$$\begin{bmatrix} FA^* & F \end{bmatrix} = \begin{bmatrix} B & F_1 \\ O & F_2 \end{bmatrix}$$

se tiene que

$$FA^* = \begin{bmatrix} B \\ O \end{bmatrix} \qquad \text{y} \qquad F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}.$$

Además,

$$\begin{bmatrix} B \\ O \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} A^* = \begin{bmatrix} F_1A^* \\ F_2A^* \end{bmatrix}$$

con lo que

$$F_1A^* = B \qquad \text{y} \qquad F_2A^* = O.$$

De $F_2A^* = O$ se tiene que $\mathcal{R}(A^\dagger) = \mathcal{R}(A^*) \subseteq \mathcal{N}(F_2)$ y así $F_2A^\dagger = O$.

Por otro lado, $BAA^\dagger = F_1A^*AA^\dagger = F_1A^* = B$ y por tanto

$$\begin{bmatrix} BA \\ F_2 \end{bmatrix} A^\dagger = \begin{bmatrix} BAA^\dagger \\ F_2A^\dagger \end{bmatrix} = \begin{bmatrix} B \\ O \end{bmatrix}, \qquad \text{con} \qquad \begin{bmatrix} BA \\ F_2 \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

De esta última igualdad, se deduce inmediatamente el resultado si se prueba que $\begin{bmatrix} BA \\ F_2 \end{bmatrix}$ es invertible.

Antes, será necesario probar que $\mathcal{N}(A^*) = \mathcal{N}(B)$. En efecto, al ser F invertible, $\mathcal{R}(A) = \mathcal{R}(AF^*) = \mathcal{R}(\begin{bmatrix} B^* & O \end{bmatrix}) = \mathcal{R}(B^*)$ y entonces $\mathcal{N}(A^*) = [\mathcal{R}(A)]^\perp = [\mathcal{R}(B^*)]^\perp = \mathcal{N}(B)$.

Ahora, para probar la invertibilidad de la matriz por bloques mencionada, se considera $x \in \mathcal{N}\left(\begin{bmatrix} BA \\ F_2 \end{bmatrix}\right)$. Entonces

$$0 = \begin{bmatrix} BA \\ F_2 \end{bmatrix} x = \begin{bmatrix} BAx \\ F_2x \end{bmatrix},$$

es decir $BAx = 0$ y $F_2x = 0$. Luego, $x \in \mathcal{N}(F_2) = \mathcal{R}(A^*)$ (pues $\mathcal{R}(A^*) \subseteq \mathcal{N}(F_2)$ y $\dim(\mathcal{N}(F_2)) = n - \operatorname{rg}(F_2) = r = \operatorname{rg}(A^*)$). Además, como $Ax \in \mathcal{N}(B) = \mathcal{N}(A^*)$ se tiene que $A^*Ax = 0$ y premultiplicando por x^* se obtiene $Ax = 0$, es decir, $x \in \mathcal{N}(A) = [\mathcal{R}(A^*)]^\perp$. Como $\mathcal{R}(A^*) \cap [\mathcal{R}(A^*)]^\perp = \{0\}$ se llega a $x = 0$, con lo que $\begin{bmatrix} BA \\ F_2 \end{bmatrix} \in \mathbb{C}_n^{n\times n}$. □

Ejemplo 4.2. *Calcular la inversa de Moore-Penrose de la matriz*

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

Se ha visto que $\mathrm{rg}(A) = 2$.

Aplicando a $\begin{bmatrix} A^* & I_3 \end{bmatrix}$ *primero la operación elemental* $f_3 + (-f_1) \to f_3$ *y posteriormente* $f_3 + (-f_2) \to f_3$ *se obtiene*

$$\left[\begin{array}{cccc|ccc} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|ccc} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|ccc} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right]$$

de donde

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \qquad y \qquad F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline -1 & -1 & 1 \end{array}\right].$$

Además,

$$BA = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 3 & 4 \end{bmatrix}.$$

La inversa de

$$\begin{bmatrix} BA \\ F_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 3 & 4 \\ -1 & -1 & 1 \end{bmatrix} \qquad es \qquad \begin{bmatrix} BA \\ F_2 \end{bmatrix}^{-1} = \frac{1}{15}\begin{bmatrix} 7 & -4 & -5 \\ -5 & 5 & -5 \\ 2 & 1 & 5 \end{bmatrix}.$$

Finalmente,

$$A^\dagger = \begin{bmatrix} BA \\ F_2 \end{bmatrix}^{-1} \begin{bmatrix} B \\ O \end{bmatrix} = \begin{bmatrix} 7/15 & 1/5 & -4/15 & -4/15 \\ -1/3 & 0 & 1/3 & 1/3 \\ 2/15 & 1/5 & 1/15 & 1/15 \end{bmatrix}.$$

4.4. Método de Greville

La idea del método es la siguiente. Sea $A \in \mathbb{C}^{m\times n}$ una matriz particionada por columnas como

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix},$$

con $\mathbf{a}_k \in \mathbb{C}^{m\times 1}$ para $1 \le k \le n$. En cada paso, este método numérico iterativo construye una nueva columna de la inversa buscada. Para ello, si A_k contiene las primeras k columnas de A, con

$1 \leq k \leq n$, en el paso k-ésimo, el método permite construir la matriz $A_k^\dagger$. En un número finito de pasos (exactamente n) se construirá una matriz $A_n^\dagger$, que es la inversa de Moore-Penrose de A. De forma esquemática, la información de la matriz A se utilizará en cada paso de la siguiente manera:

$$\begin{array}{rclcc} A_1 & = & \left[\begin{array}{c} \mathbf{a}_1 \end{array}\right] & \longrightarrow & A_1^\dagger = ? \\ A_2 & = & \left[\begin{array}{cc} A_1 & \mathbf{a}_2 \end{array}\right] & \longrightarrow & A_2^\dagger = ? \\ A_3 & = & \left[\begin{array}{cc} A_2 & \mathbf{a}_3 \end{array}\right] & \longrightarrow & A_3^\dagger = ? \\ & \vdots & & & \vdots \\ A_n & = & \left[\begin{array}{cc} A_{n-1} & \mathbf{a}_n \end{array}\right] & \longrightarrow & A_n^\dagger = ? \end{array}$$

Como $A_n = A$ resulta que $A_n^\dagger = A^\dagger$. Falta ver cómo se construye cada una de las matrices $A_k^\dagger$ para $1 \leq k \leq n$.

Teorema 4.3. *Sea*

$$A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{array}\right] \in \mathbb{C}^{m\times n}$$

una matriz particionada por sus columnas $\mathbf{a}_k \in \mathbb{C}^{m\times 1}$*, para* $1 \leq k \leq n$*, y sean*

$$A_1 := \left[\begin{array}{c} \mathbf{a}_1 \end{array}\right]$$

y si $2 \leq k \leq n$*,*

$$A_k \;:=\; \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right].$$

Si para $k = 2, 3, \ldots, n$ *se definen*

$$\mathbf{d}_k \;:=\; A_{k-1}^\dagger \mathbf{a}_k, \tag{4.4}$$

$$\mathbf{c}_k \;:=\; \mathbf{a}_k - A_{k-1}\mathbf{d}_k, \tag{4.5}$$

$$\mathbf{b}_k^* \;:=\; \begin{cases} \mathbf{c}_k^\dagger, & \mathbf{c}_k \neq 0 \\ (1 + \mathbf{d}_k^*\mathbf{d}_k)^{-1}\mathbf{d}_k^* A_{k-1}^\dagger, & \mathbf{c}_k = 0 \end{cases}, \tag{4.6}$$

entonces

$$A_k^\dagger = \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right]^\dagger = \left[\begin{array}{c} A_{k-1}^\dagger - \mathbf{d}_k\mathbf{b}_k^* \\ \mathbf{b}_k^* \end{array}\right].$$

Demostración. Teniendo en cuenta que, como $A \in \mathbb{C}^{m\times n}$, debe ser $A^\dagger \in \mathbb{C}^{n\times m}$, se supone que la inversa de Moore-Penrose de la matriz $A_k = \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right]$ está particionada como

$$A_k^\dagger = \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right]^\dagger = \left[\begin{array}{c} B_k \\ \mathbf{b}_k^* \end{array}\right],$$

y se tratan de determinar tanto la submatriz B_k como el vector $\mathbf{b}_k^*$, que es la k-ésima fila de $A_k^\dagger$.

Multiplicando por bloques se tiene

$$\begin{aligned} A_k A_k^\dagger &= \begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix} \begin{bmatrix} B_k \\ \mathbf{b}_k^* \end{bmatrix} \\ &= A_{k-1}B_k + \mathbf{a}_k \mathbf{b}_k^*. \end{aligned}$$

Utilizando las propiedades de la inversa de Moore-Penrose del Lema 2.2,

$$\mathcal{N}(A_k A_k^\dagger) = \mathcal{N}(A_k^\dagger) = \mathcal{N}(A_k^*) \subseteq \mathcal{N}(A_{k-1}^*) = \mathcal{N}(A_{k-1}^\dagger),$$

donde la inclusión sigue directamente tomando un elemento en $\mathcal{N}(A_k^*)$ y viendo que, en particular, está en $\mathcal{N}(A_{k-1}^*)$. Antes de continuar es necesario el resultado del siguiente Ejercicio.

Ejercicio 4.1. *Sean L y M dos subespacios complementarios de $\mathbb{C}^n$ o $\mathbb{C}^m$ según corresponda y sea $A \in \mathbb{C}^{m\times n}$. Demostrar que:*

(a) $AP_{L,M} = A$ si y sólo si $M \subseteq \mathcal{N}(A)$.

(b) $P_{L,M}A = A$ si y sólo si $\mathcal{R}(A) \subseteq L$.

Como $A_k A_k^\dagger$ es un proyector sobre $\mathcal{R}(A_k A_k^\dagger)$ paralelo a $\mathcal{N}(A_k A_k^\dagger)$ y $\mathcal{N}(A_k A_k^\dagger) \subseteq \mathcal{N}(A_{k-1}^\dagger)$, del Ejercicio 4.1 se tiene que

$$A_{k-1}^\dagger A_k A_k^\dagger = A_{k-1}^\dagger.$$

Se probará ahora que $\mathcal{R}(B_k) \subseteq \mathcal{R}(A_{k-1}^*)$. En efecto, si $y \in \mathcal{R}(B_k)$ entonces $y = B_k x$ para algún x y por tanto

$$\begin{aligned} \begin{bmatrix} y \\ \mathbf{b}_k^* x \end{bmatrix} &= \begin{bmatrix} B_k x \\ \mathbf{b}_k^* x \end{bmatrix} = \begin{bmatrix} B_k \\ \mathbf{b}_k^* \end{bmatrix} x \in \mathcal{R}\left(\begin{bmatrix} B_k \\ \mathbf{b}_k^* \end{bmatrix} \right) = \mathcal{R}(A_k^\dagger) = \mathcal{R}(A_k^*) \\ &= \mathcal{R}(\begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix})^* = \mathcal{R}\left(\begin{bmatrix} A_{k-1}^* \\ \mathbf{a}_k^* \end{bmatrix} \right), \end{aligned} \tag{4.7}$$

es decir

$$\begin{bmatrix} y \\ \mathbf{b}_k^* x \end{bmatrix} = \begin{bmatrix} A_{k-1}^* \\ \mathbf{a}_k^* \end{bmatrix} z = \begin{bmatrix} A_{k-1}^* z \\ \mathbf{a}_k^* z \end{bmatrix}, \qquad \text{para algún } z,$$

de donde se tiene que $y = A_{k-1}^* z$ para algún z, y así, $y \in \mathcal{R}(A_{k-1}^*)$.

Como $A_{k-1}^\dagger A_{k-1}$ es un proyector sobre $\mathcal{R}(A_{k-1}^\dagger A_{k-1})$ paralelo a $\mathcal{N}(A_{k-1}^\dagger A_{k-1})$ y

$$\mathcal{R}(B_k) \subseteq \mathcal{R}(A_{k-1}^*) = \mathcal{R}(A_{k-1}^\dagger) = \mathcal{R}(A_{k-1}^\dagger A_{k-1}),$$

del Ejercicio 4.1 se tiene que

$$A_{k-1}^\dagger A_{k-1} B_k = B_k.$$

Ahora,

$$\begin{aligned} A_{k-1}^{\dagger} &= A_{k-1}^{\dagger} A_k A_k^{\dagger} \\ &= A_{k-1}^{\dagger}(A_{k-1}B_k + \mathbf{a}_k \mathbf{b}_k^*) \\ &= A_{k-1}^{\dagger} A_{k-1} B_k + A_{k-1}^{\dagger} \mathbf{a}_k \mathbf{b}_k^* \\ &= B_k + A_{k-1}^{\dagger} \mathbf{a}_k \mathbf{b}_k^*, \end{aligned}$$

de donde usando la notación de (4.4) (para $\mathbf{d}_k$) resulta

$$B_k = A_{k-1}^{\dagger} - A_{k-1}^{\dagger} \mathbf{a}_k \mathbf{b}_k^* = A_{k-1}^{\dagger} - \mathbf{d}_k \mathbf{b}_k^*.$$

Hasta ahora se ha conseguido que

$$A_k^{\dagger} = \begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix}^{\dagger} = \begin{bmatrix} B_k \\ \mathbf{b}_k^* \end{bmatrix} = \begin{bmatrix} A_{k-1}^{\dagger} - \mathbf{d}_k \mathbf{b}_k^* \\ \mathbf{b}_k^* \end{bmatrix},$$

quedando por determinar $\mathbf{b}_k^*$.

Ahora se van a distinguir 2 casos, según sea $\mathbf{c}_k \neq 0$ o $\mathbf{c}_k = 0$.

- Caso 1: $\mathbf{c}_k \neq 0$.

  Sustituyendo $B_k = A_{k-1}^{\dagger} - \mathbf{d}_k \mathbf{b}_k^*$ en $A_k A_k^{\dagger} = A_{k-1} B_k + \mathbf{a}_k \mathbf{b}_k^*$ y recordando la expresión para $\mathbf{c}_k$ de (4.5) se tiene

$$\begin{aligned} A_k A_k^{\dagger} &= A_{k-1}(A_{k-1}^{\dagger} - \mathbf{d}_k \mathbf{b}_k^*) + \mathbf{a}_k \mathbf{b}_k^* \\ &= A_{k-1} A_{k-1}^{\dagger} - A_{k-1} \mathbf{d}_k \mathbf{b}_k^* + \mathbf{a}_k \mathbf{b}_k^* \\ &= A_{k-1} A_{k-1}^{\dagger} + (\mathbf{a}_k - A_{k-1} \mathbf{d}_k) \mathbf{b}_k^* \\ &= A_{k-1} A_{k-1}^{\dagger} + \mathbf{c}_k \mathbf{b}_k^*. \end{aligned} \tag{4.8}$$

  Al ser $A_k A_k^{\dagger}$ y $A_{k-1} A_{k-1}^{\dagger}$ matrices hermíticas resulta que $\mathbf{c}_k \mathbf{b}_k^*$ también lo es. Luego, premultiplicando $\mathbf{c}_k \mathbf{b}_k^* = \mathbf{b}_k \mathbf{c}_k^*$ por $\mathbf{c}_k^*$ se tiene

$$\|\mathbf{c}_k\|_2^2 \mathbf{b}_k^* = \mathbf{c}_k^* \mathbf{c}_k \mathbf{b}_k^* = \mathbf{c}_k^* \mathbf{b}_k \mathbf{c}_k^*$$

  y tomando $\delta := \frac{\mathbf{c}_k^* \mathbf{b}_k}{\|\mathbf{c}_k\|_2^2} \in \mathbb{C}$ (puesto que $\mathbf{c}_k \neq 0$), se obtiene

$$\mathbf{b}_k^* = \delta \mathbf{c}_k^*. \tag{4.9}$$

  Usando de nuevo que $\mathbf{c}_k \mathbf{b}_k^*$ es hermítica se deduce que $\overline{\delta} \mathbf{c}_k \mathbf{c}_k^* = \delta \mathbf{c}_k \mathbf{c}_k^*$ de donde se llega a que $\delta \in \mathbb{R}$ pues $\mathbf{c}_k \neq 0$.

Por otro lado, utilizando (4.8), (4.4) y (4.5),

$$\begin{aligned}\left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right] &= A_k \\ &= (A_k A_k^\dagger) A_k \\ &= (A_{k-1}A_{k-1}^\dagger + \mathbf{c}_k \mathbf{b}_k^*) \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right] \\ &= \left[\begin{array}{cc} (A_{k-1}A_{k-1}^\dagger + \mathbf{c}_k \mathbf{b}_k^*) A_{k-1} & (A_{k-1}A_{k-1}^\dagger + \mathbf{c}_k \mathbf{b}_k^*) \mathbf{a}_k \end{array}\right] \\ &= \left[\begin{array}{cc} A_{k-1}A_{k-1}^\dagger A_{k-1} + \mathbf{c}_k \mathbf{b}_k^* A_{k-1} & A_{k-1}A_{k-1}^\dagger \mathbf{a}_k + \mathbf{c}_k \mathbf{b}_k^* \mathbf{a}_k \end{array}\right] \\ &= \left[\begin{array}{cc} A_{k-1} + \mathbf{c}_k \mathbf{b}_k^* A_{k-1} & A_{k-1}\mathbf{d}_k + (\mathbf{b}_k^* \mathbf{a}_k)\mathbf{c}_k \end{array}\right] \\ &= \left[\begin{array}{cc} A_{k-1} + \mathbf{c}_k \mathbf{b}_k^* A_{k-1} & (\mathbf{a}_k - \mathbf{c}_k) + (\mathbf{b}_k^* \mathbf{a}_k)\mathbf{c}_k \end{array}\right]\end{aligned}$$

e igualando bloques se llega a $A_{k-1} = A_{k-1} + \mathbf{c}_k \mathbf{b}_k^* A_{k-1}$ y $\mathbf{a}_k = (\mathbf{a}_k - \mathbf{c}_k) + (\mathbf{b}_k^* \mathbf{a}_k)\mathbf{c}_k$ o equivalentemente (usando que $\mathbf{c}_k \neq 0$) a las expresiones

$$\mathbf{c}_k \mathbf{b}_k^* A_{k-1} = 0$$

y

$$\mathbf{b}_k^* \mathbf{a}_k = 1. \tag{4.10}$$

Teniendo en cuenta que $\mathbf{d}_k = A_{k-1}^\dagger \mathbf{a}_k$ y recordando que $A_{k-1}A_{k-1}^\dagger$ es el proyector ortogonal sobre $\mathcal{R}(A_{k-1})$ paralelamente a $\mathcal{N}(A_{k-1}^*)$ se tiene

$$\begin{aligned}\mathbf{c}_k &= \mathbf{a}_k - A_{k-1}\mathbf{d}_k \\ &= \mathbf{a}_k - A_{k-1}A_{k-1}^\dagger \mathbf{a}_k \\ &= (I - A_{k-1}A_{k-1}^\dagger)\mathbf{a}_k \\ &= (I - P_{\mathcal{R}(A_{k-1})})\mathbf{a}_k \\ &= P_{\mathcal{N}(A_{k-1}^*)}\mathbf{a}_k.\end{aligned}$$

Luego, llamando $P := P_{\mathcal{N}(A_{k-1}^*)}$, y observando que $P^2 = P = P^*$, de (4.10) y (4.9) se consigue

$$\begin{aligned}1 &= \mathbf{b}_k^* \mathbf{a}_k \\ &= \delta \mathbf{c}_k^* \mathbf{a}_k \\ &= \delta \mathbf{a}_k^* P^* \mathbf{a}_k \\ &= \delta \mathbf{a}_k^* P^* P \mathbf{a}_k \\ &= \delta \mathbf{c}_k^* \mathbf{c}_k \\ &= \delta \|\mathbf{c}_k\|_2^2,\end{aligned}$$

de donde $\delta = \frac{1}{\|\mathbf{c}_k\|_2^2}$ y así

$$\mathbf{b}_k^* = \delta \mathbf{c}_k^* = \frac{1}{\|\mathbf{c}_k\|_2^2}\mathbf{c}_k^* = \mathbf{c}_k^\dagger,$$

donde el último paso resulta de aplicar el Ejercicio 2.3. La expresión encontrada para $\mathbf{b}_k^*$ coincide con la indicada en (4.6).

- Caso 2: $\mathbf{c}_k = 0$.

  De $\mathbf{c}_k = \mathbf{a}_k - A_{k-1}\mathbf{d}_k$ se tiene que $\mathbf{a}_k = A_{k-1}\mathbf{d}_k \in \mathcal{R}(A_{k-1})$ y por tanto

$$\mathcal{R}(A_k) = \mathcal{R}\left(\left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right]\right) = \mathcal{R}(A_{k-1}).$$

  De aquí se deduce que

$$\mathcal{N}(A_k^*) = \mathcal{R}(A_k)^\perp = \mathcal{R}(A_{k-1})^\perp = \mathcal{N}(A_{k-1}^*).$$

  Además, $\mathcal{N}(A_k^\dagger) \subseteq \mathcal{N}(\mathbf{b}_k^*)$. En efecto, si $x \in \mathcal{N}(A_k^\dagger) = \mathcal{N}\left(\left[\begin{array}{c} B_k \\ \mathbf{b}_k^* \end{array}\right]\right)$ entonces $0 = \left[\begin{array}{c} B_k \\ \mathbf{b}_k^* \end{array}\right] x = \left[\begin{array}{c} B_k x \\ \mathbf{b}_k^* x \end{array}\right]$, de donde $\mathbf{b}_k^* x = 0$ y así $x \in \mathcal{N}(\mathbf{b}_k^*)$.

  En relación al proyector ortogonal $A_{k-1}A_{k-1}^\dagger$ se cumple que

$$\mathcal{N}(A_{k-1}A_{k-1}^\dagger) = \mathcal{N}(A_{k-1}^\dagger) = \mathcal{N}(A_{k-1}^*) = \mathcal{N}(A_k^*) = \mathcal{N}(A_k^\dagger) \subseteq \mathcal{N}(\mathbf{b}_k^*).$$

  Por el Ejercicio 4.1,

$$\mathbf{b}_k^* A_{k-1} A_{k-1}^\dagger = \mathbf{b}_k^*.$$

  Ahora se calculará $A_k^\dagger A_k$ recordando (4.4) y denotando $\alpha := \mathbf{b}_k^* \mathbf{a}_k$:

$$\begin{aligned}
A_k^\dagger A_k &= \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right]^\dagger \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right] \\
&= \left[\begin{array}{c} A_{k-1}^\dagger - \mathbf{d}_k \mathbf{b}_k^* \\ \mathbf{b}_k^* \end{array}\right] \left[\begin{array}{cc} A_{k-1} & \mathbf{a}_k \end{array}\right] \\
&= \left[\begin{array}{cc} A_{k-1}^\dagger A_{k-1} - \mathbf{d}_k \mathbf{b}_k^* A_{k-1} & A_{k-1}^\dagger \mathbf{a}_k - \mathbf{d}_k \mathbf{b}_k^* \mathbf{a}_k \\ \mathbf{b}_k^* A_{k-1} & \mathbf{b}_k^* \mathbf{a}_k \end{array}\right] \\
&= \left[\begin{array}{cc} A_{k-1}^\dagger A_{k-1} - \mathbf{d}_k \mathbf{b}_k^* A_{k-1} & (1-\alpha)\mathbf{d}_k \\ \mathbf{b}_k^* A_{k-1} & \alpha \end{array}\right].
\end{aligned}$$

  Como la matriz $A_k^\dagger A_k$ es hermítica, debe ser $\alpha \in \mathbb{R}$ y $\mathbf{b}_k^* A_{k-1} = (1-\alpha)\mathbf{d}_k^*$, con lo que

$$\mathbf{b}_k^* = \mathbf{b}_k^* A_{k-1} A_{k-1}^\dagger = (1-\alpha)\mathbf{d}_k^* A_{k-1}^\dagger. \tag{4.11}$$

Postmultiplicando la última igualdad por $\mathbf{a}_k$ resulta:

$$\alpha = \mathbf{b}_k^* \mathbf{a}_k = (1-\alpha)\mathbf{d}_k^* A_{k-1}^\dagger \mathbf{a}_k = (1-\alpha)\mathbf{d}_k^* \mathbf{d}_k.$$

Despejando se llega a

$$1-\alpha = \frac{1}{1+\mathbf{d}_k^*\mathbf{d}_k}$$

pues $\mathbf{d}_k^*\mathbf{d}_k \geq 0$. Sustituyendo en (4.11) finalmente se tiene que:

$$\mathbf{b}_k^* = \frac{1}{1+\mathbf{d}_k^*\mathbf{d}_k}\mathbf{d}_k^* A_{k-1}^\dagger,$$

que coincide con (4.6).

□

Se ilustrará el uso de este método mediante un ejemplo.

Ejemplo 4.3. *Aplicar el método de Greville para encontrar la inversa de Moore-Penrose de la matriz:*

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

En primer lugar se particiona A según sus columnas como

$$A = \left[\begin{array}{c|c|c} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 0 \end{array}\right] = \left[\begin{array}{c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{array}\right]$$

y se consideran las submatrices

$$A_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \qquad y \qquad A = A_3 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 0 \end{bmatrix}.$$

- $\underline{k=2}$. *Se calcula $A_1^\dagger$ mediante la inversa de Moore-Penrose de un vector:*

$$A_1^\dagger = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

y ahora

$$\mathbf{d}_2 = A_1^\dagger \mathbf{a}_2 = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{2}$$

y

$$\mathbf{c}_2 = \mathbf{a}_2 - A_1\mathbf{d}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ 1 \\ \frac{3}{2} \end{bmatrix}.$$

Al ser $\mathbf{c}_2 \neq \mathbf{0}$, *debe ser*

$$\mathbf{b}_2^* = \mathbf{c}_2^\dagger = \frac{2}{11}\begin{bmatrix} -\frac{3}{2} & 1 & \frac{3}{2} \end{bmatrix} = \begin{bmatrix} -\frac{3}{11} & \frac{2}{11} & \frac{3}{11} \end{bmatrix},$$

$$A_1^\dagger - \mathbf{d}_2\mathbf{b}_2^* = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -\frac{3}{11} & \frac{2}{11} & \frac{3}{11} \end{bmatrix} = \begin{bmatrix} \frac{7}{11} & -\frac{1}{11} & \frac{4}{11} \end{bmatrix}.$$

Luego,

$$A_2^\dagger = \begin{bmatrix} \frac{7}{11} & -\frac{1}{11} & \frac{4}{11} \\ -\frac{3}{11} & \frac{2}{11} & \frac{3}{11} \end{bmatrix}.$$

- $\underline{k = 3}$.

$$\mathbf{d}_3 = A_2^\dagger\mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

y

$$\mathbf{c}_3 = \mathbf{a}_3 - A_2\mathbf{d}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Al ser $\mathbf{c}_3 = \mathbf{0}$, *debe ser*

$$\mathbf{b}_3^* = (1 + \mathbf{d}_3\mathbf{d}_3^*)^{-1}\mathbf{d}_3^*A_2^\dagger = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix},$$

$$A_2^\dagger - \mathbf{d}_3\mathbf{b}_3^* = A_2^\dagger.$$

Luego,

$$A^\dagger = A_3^\dagger = \begin{bmatrix} \frac{7}{11} & -\frac{1}{11} & \frac{4}{11} \\ -\frac{3}{11} & \frac{2}{11} & \frac{3}{11} \\ 0 & 0 & 0 \end{bmatrix}.$$

Ejercicio 4.2. *Encontrar la inversa de Moore-Penrose de las siguientes matrices:*

(a) $B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$

(b) $C = \begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix}$,

(c) $D = \begin{bmatrix} 1 & -i \\ 0 & 0 \\ i & 2 \end{bmatrix}$,

(d) $E = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ -1 & 2 & 1 \end{bmatrix}$,

(e) $F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$,

a partir de:

(I) *el método basado en la pseudoinversa,*

(II) *la descomposición en valores singulares,*

(III) *el método basado en el algoritmo de Gauss-Jordan,*

(IV) *el método de Greville.*

Capítulo 5

Ampliación de la clase de matrices normales

Se ha visto la importancia de los proyectores $AA^\dagger$ y $A^\dagger A$ a la hora de trabajar con la inversa de Moore-Penrose. Si bien el primero proyecta ortogonalmente sobre $\mathcal{R}(A)$ y el segundo sobre $\mathcal{R}(A^*)$, cabe la posibilidad de que dichos operadores de proyección coincidan para algunas matrices. En la primera parte de esta sección se estudian dichas matrices y se las llamará EP teniendo en cuenta las iniciales de su nombre en inglés: "Equal Projectors". Se observa que es condición necesaria para que A sea EP que se trate de una matriz cuadrada.

Por otro lado, en el Teorema 2.1 se ha probado que, dada una aplicación lineal $T : \mathbb{C}^n \to \mathbb{C}^m$ tal que $\mathrm{rg}(T) = r$, la aplicación lineal $T_{|\mathcal{R}(T^*)}$ proporciona un isomorfismo entre los espacios vectoriales $\mathcal{R}(T^*)$ y $\mathcal{R}(T)$. Recordando que si un espacio vectorial sobre el cuerpo de los complejos tiene dimensión $s > 0$ entonces es isomorfo a $\mathbb{C}^s$, es posible asegurar que siempre se cumple

$$\mathcal{R}(T^*) \simeq \mathcal{R}(T) \simeq \mathbb{C}^r, \tag{5.1}$$

independientemente del valor de los números naturales n y m. En relación a los núcleos de T y T^* es posible establecer que:

- si $m \neq n$ entonces los subespacios $\mathcal{N}(T)$ y $\mathcal{N}(T^*)$ no son isomorfos (pues $\dim(\mathcal{N}(T^*)) = m - \mathrm{rg}(T^*) = m - \mathrm{rg}(T) = m - n + \dim(\mathcal{N}(T))$),
- si $m = n$ entonces
$$\mathcal{N}(T) \simeq \mathcal{N}(T^*) \simeq \mathbb{C}^{n-r}. \tag{5.2}$$

En el caso en que $m = n$, ¿es posible que los isomorfismos $\mathcal{R}(T^*) \simeq \mathcal{R}(T)$ y $\mathcal{N}(T) \simeq \mathcal{N}(T^*)$ de (5.1) y (5.2) se transformen en igualdades de subespacios? La respuesta es afirmativa y tiene relación con la clase de matrices EP.

5.1. Matrices EP

Definición 5.1. *Sea $A \in \mathbb{C}^{n\times n}$. Se dice que A es una matriz EP si $AA^\dagger = A^\dagger A$.*

Ejemplo 5.1. *Toda matriz invertible es EP. La matriz nula también es EP.*

Ejemplo 5.2. *Se consideran las matrices*

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & -i & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A partir de las inversas de Moore-Penrose de A y B es posible comprobar que A no es EP y B sí lo es.

5.1.1. Caracterizaciones de matrices EP

La siguiente caracterización permitirá comprender mejor la estructura de este tipo especial de matrices.

Teorema 5.1. *Sea $A \in \mathbb{C}_r^{n\times n}$ con $r > 0$. Las siguientes condiciones son equivalentes:*

(a) A *es EP,*

(b) $\mathcal{R}(A) = \mathcal{R}(A^*)$,

(c) $\mathcal{N}(A) = \mathcal{N}(A^*)$,

(d) $\mathbb{C}^n = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A)$,

(e) Existen una matriz invertible $C \in \mathbb{C}^{r\times r}$ y una matriz unitaria $U \in \mathbb{C}^{n\times n}$ tales que

$$A = U \begin{bmatrix} C & O \\ O & O \end{bmatrix} U^*. \tag{5.3}$$

Demostración. (a) $\Longrightarrow$ (b) Si $AA^\dagger = A^\dagger A$ entonces $P_{\mathcal{R}(A)} = P_{\mathcal{R}(A^*)}$. Luego,

$$x \in \mathcal{R}(A) \iff P_{\mathcal{R}(A)}x = x \iff P_{\mathcal{R}(A^*)}x = x \iff x \in \mathcal{R}(A^*).$$

(b) $\iff$ (c) Es evidente del Teorema 1.4 y usando la propiedad $(S^\perp)^\perp = S$ para todo subespacio vectorial S.

(b) $\Longrightarrow$ (d) Es inmediato del Teorema 1.5.

(d) $\Longrightarrow$ (e) Se fijan dos bases ortonormales $\{u_1, \dots, u_r\}$ y $\{u_{r+1}, \dots, u_n\}$ de los subespacios $\mathcal{R}(A)$ y $\mathcal{N}(A)$, respectivamente (observar que $\mathrm{rg}(A) = r > 0$) y se llama $\mathcal{B}$ a la unión de ambas bases, que proporciona una base ortonormal de $\mathbb{C}^n$ por tratarse, por hipótesis, de una suma directa ortogonal. Se colocan ordenadamente los vectores de $\mathcal{B}$ en una matriz $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$, que es unitaria.

Ahora se analiza cómo actúa la matriz A sobre los elementos de la base $\mathcal{B}$. En primer lugar se tiene que $Au_i \in \mathcal{R}(A)$ para $1 \leq i \leq r$ y, por tanto, se pueden escribir como combinación lineal de los vectores de $\{u_1, \ldots, u_r\}$, es decir, existe $C \in \mathbb{C}^{r\times r}$ tal que $AU_1 = U_1C$. Es claro que C contiene los escalares que resultan de escribir los vectores $Au_i \in \mathcal{R}(A)$ para $1 \leq i \leq r$ como combinación lineal de los vectores de la base $\{u_1, \ldots, u_r\}$. En segundo lugar, $Au_i = 0$ para $r + 1 \leq i \leq n$ y, por tanto, $AU_2 = O$. Entonces,

$$AU = A \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} AU_1 & AU_2 \end{bmatrix} = \begin{bmatrix} U_1C & O \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} C & O \\ O & O \end{bmatrix} = U \begin{bmatrix} C & O \\ O & O \end{bmatrix}.$$

Como U es invertible, se observa que $\operatorname{rg}(C) = \operatorname{rg}(A) = r$ con lo que $C \in \mathbb{C}_r^{r\times r}$.

(e) $\Longrightarrow$ (a) A partir de (5.3) se considera la matriz

$$B := U \begin{bmatrix} C^{-1} & O \\ O & O \end{bmatrix} U^*.$$

Multiplicando por bloques es fácil probar que se cumplen las cuatro propiedades de la Definición de Penrose con lo cual, de la unicidad se tiene que $B = A^\dagger$. Ahora, de nuevo multiplicando por bloques es fácil ver que $AA^\dagger = U \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} U^* = A^\dagger A$. □

Ejemplo 5.3. *Se consideran las matrices A y B del Ejemplo 5.2. Ahora es inmediato comprobar que A no es EP y B sí lo es a partir de la caracterización del apartado (b).*

Ejercicio 5.1. *Confrontar los apartados (b), (c) y (d) del Teorema 5.1 en el diagrama dibujado en el Ejercicio 2.1.*

Nota 5.1. *Todo proyector ortogonal $A \in \mathbb{C}^{n\times n}$ satisface $A^2 = A = A^*$. De la igualdad $A^* = A$ se cumple la condición*

$$\mathbb{C}^n = \mathcal{R}(A) \oplus^\perp [\mathcal{R}(A)]^\perp = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A)$$

del Teorema 5.1 y, por tanto, A es una matriz EP.

Sin embargo, la condición $\mathbb{C}^n = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A)$ no implica que A sea un proyector ortogonal, ni siquiera implica que sea un proyector. Basta verlo con $A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$. Es fácil probar que si A es EP y $A^2 = A$ entonces A es un proyector ortogonal (ejercicio).

5.2. Extensión de la clase de matrices normales

Ahora se considera el siguiente problema.

PROBLEMA: ¿Cuál es la clase de matrices más amplia para la cual

$$\mathcal{R}(A^*) = \mathcal{R}(A) \qquad \text{y} \qquad \mathcal{N}(A^*) = \mathcal{N}(A) \tag{5.4}$$

se cumplen?

De nuevo, una primera observación es que las matrices de dicha clase deben ser cuadradas.

Si $A \in \mathbb{C}^{n\times n}$ es hermítica entonces claramente las relaciones (5.4) se satisfacen. Las matrices normales también las cumplen como se demuestra a continuación.

Teorema 5.2. *Sea $A \in \mathbb{C}^{n\times n}$. Si A es normal entonces A es EP.*

Demostración. Si A es normal entonces $AA^* = A^*A$. Aplicando el Teorema 1.6 se tiene que $\mathcal{R}(A) = \mathcal{R}(AA^*) = \mathcal{R}(A^*A) = \mathcal{R}(A^*)$. Por el Teorema 5.1 se tiene que A es EP. □

Nota 5.2. *La recíproca no es cierta como se puede comprobar con la matriz*

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

que es EP y no es normal.

Ahora, ¿son las EP todas las matrices que cumplen estas propiedades?

La respuesta al problema es afirmativa y se encuentra en el Teorema 5.1 que, no sólo caracteriza las matrices EP sino que, además, prueba que la clase de las matrices EP es la más grande que satisface las relaciones $\mathcal{R}(A^*) = \mathcal{R}(A)$ y $\mathcal{N}(A^*) = \mathcal{N}(A)$. Dicho teorema prueba también que con que se cumpla una de las dos condiciones, la otra se deduce inmediatamente.

5.3. Matrices bi-dagger

En esta sección se presentará otra extensión de la clase de matrices normales.

Para ello, primero se observa que, para toda matriz invertible $A \in \mathbb{C}^{n\times n}$, es válida la propiedad

$$(A^s)^{-1} = (A^{-1})^s, \qquad \text{para todo } s \in \mathbb{Z}.$$

Sin embargo, la igualdad $(A^2)^\dagger = (A^\dagger)^2$ no siempre se cumple. Esto último se puede constatar fácilmente con la matriz

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

cuya inversa de Moore-Penrose es

$$A^\dagger = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

En este caso se comprueba que

$$(A^2)^\dagger = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \neq \frac{1}{4}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = (A^\dagger)^2.$$

Definición 5.2. *Sea $A \in \mathbb{C}^{n\times n}$. Se dice que A es una matriz bi-dagger si $(A^2)^\dagger = (A^\dagger)^2$.*

Ejercicio 5.2. *Comprobar que las matrices*

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad y \qquad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

son bi-dagger.

Con relación a las matrices EP se cumple el siguiente resultado.

Teorema 5.3. *Sea $A \in \mathbb{C}^{n\times n}$. Si A es EP entonces A es bi-dagger.*

Demostración. Si A es EP entonces existen una matriz invertible $C \in \mathbb{C}^{r\times r}$ y una matriz unitaria $U \in \mathbb{C}^{n\times n}$ tales que

$$A = U\begin{bmatrix} C & O \\ O & O \end{bmatrix}U^*.$$

Es fácil ver que

$$(A^2)^\dagger = \left(U\begin{bmatrix} C^2 & O \\ O & O \end{bmatrix}U^*\right)^\dagger = U\begin{bmatrix} C^{-2} & O \\ O & O \end{bmatrix}U^*.$$

Por otro lado,

$$(A^\dagger)^2 = \left(U\begin{bmatrix} C^{-1} & O \\ O & O \end{bmatrix}U^*\right)^2 = U\begin{bmatrix} C^{-2} & O \\ O & O \end{bmatrix}U^*.$$

La conclusión es inmediata. □

Nota 5.3. *La recíproca no es cierta como se puede comprobar con la matriz*

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

que es bi-dagger y no es EP.

De esta forma, la clase de matrices bi-dagger proporciona una extensión de las clases de matrices conocidas, más amplia aún que la de las matrices EP.

Ejercicio 5.3. *Representar, mediante un diagrama de Venn, las clases de matrices:*

(a) hermíticas,

(b) anti-hermíticas (es decir, aquellas matrices $A \in \mathbb{C}^n$ tales que $A^* = -A$),

(c) invertibles,

(d) unitarias,

(e) normales,

(f) EP,

(g) bi-dagger,

ordenadas según la inclusión de conjuntos.

Capítulo 6

El orden parcial estrella

En esta sección se estudia un orden parcial definido sobre el conjunto de las matrices rectangulares complejas de tamaño $m \times n$. La inversa de Moore-Penrose jugará un papel importante a la hora de su caracterización.

6.1. Definición

Se comienza mediante la siguiente definición.

Definición 6.1. *Sean $A, B \in \mathbb{C}^{m\times n}$. Se define la relación binaria*

$$A \leq^* B \qquad \Longleftrightarrow \qquad A^*A = A^*B \qquad \text{y} \qquad AA^* = BA^*.$$

Ejemplo 6.1. *Es fácil comprobar que $A \leq^* B$ para las matrices*

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

pero la relación $A \leq^* B_1$ no se cumple para

$$B_1 = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Lema 6.1. *Sean $A, B \in \mathbb{C}^{m\times n}$. Si $A \leq^* B$ entonces $UAV \leq^* UBV$ para todo par de matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$.*

Demostración. Comprobarlo como ejercicio. □

Ejemplo 6.2. *Si las matrices diagonales*

$$D_1 = \operatorname{diag}(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^{n\times n} \quad \text{y} \quad D_2 = \operatorname{diag}(\beta_1, \ldots, \beta_n) \in \mathbb{C}^{n\times n}$$

cumplen $D_1 \leq^* D_2$ entonces se tiene que:

(a) Si $\alpha_i \neq 0$ para algún i entonces $\beta_i = \alpha_i$.

(b) Si $\beta_i = 0$ para algún i entonces $\alpha_i = 0$,

Por tanto, si D_1 tiene todos sus elementos no nulos en las primeras posiciones diagonales y D_2 tiene todos sus elementos nulos en las últimas posiciones diagonales entonces D_1 y D_2 tienen la siguiente forma:

$$D_1 = \operatorname{diag}(D_3, O, O) \qquad \text{y} \qquad D_2 = \operatorname{diag}(D_3, D_4, O),$$

donde D_3 y D_4 son matrices diagonales con elementos no nulos en sus diagonales.

Si D_1 y D_2 tuviesen sus elementos diagonales ordenados arbitrariamente, es posible hallar una matriz de permutación $P \in \mathbb{R}^{n\times n}$ de modo que, aplicando el Lema 6.1,

$$D_1 = P\operatorname{diag}(D_3, O, O)P^* \qquad \text{y} \qquad D_2 = P\operatorname{diag}(D_3, D_4, O)P^*.$$

6.2. Caracterizaciones

El siguiente resultado proporciona caracterizaciones para la relación binaria $\leq^*$. En parte de la demostración se utiliza la siguiente propiedad:

$$AB = O \qquad \Longleftrightarrow \qquad \mathcal{R}(B) \subseteq \mathcal{N}(A),$$

para cualquier par de matrices $A \in \mathbb{C}^{m\times n}$ y $B \in \mathbb{C}^{n\times p}$.

Teorema 6.1. *Sean $A, B \in \mathbb{C}^{m\times n}$. Entonces las siguientes condiciones son equivalentes:*

(a) $A \leq^* B$,

(b) $A^\dagger A = A^\dagger B$ y $AA^\dagger = BA^\dagger$,

(c) $\mathcal{R}(A) \perp \mathcal{R}(B - A)$ y $\mathcal{R}(A^*) \perp \mathcal{R}((B - A)^*)$,

(d) Existen matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = U \begin{bmatrix} D_a & O & O \\ O & O & O \\ O & O & O \end{bmatrix} V \qquad \text{y} \qquad B = U \begin{bmatrix} D_a & O & O \\ O & D_{b-a} & O \\ O & O & O \end{bmatrix} V,$$

donde $D_a \in \mathbb{R}^{a\times a}$ y $D_{b-a} \in \mathbb{R}^{(b-a)\times(b-a)}$ son matrices diagonales definidas positivas.

Demostración. (a) $\Longleftrightarrow$ (b) Se cumple que $A^*A = A^*B$ y $AA^* = BA^*$ si y sólo si

$$A^*(B-A) = O \qquad \text{y} \qquad (B-A)A^* = O.$$

Aplicando el Lema 2.2 se tiene que

$$A^*(B-A) = O \quad \Longleftrightarrow \quad \mathcal{R}(B-A) \subseteq \mathcal{N}(A^*) = \mathcal{N}(A^\dagger) \quad \Longleftrightarrow \quad A^\dagger(B-A) = O$$

y

$$(B-A)A^* = O \quad \Longleftrightarrow \quad \mathcal{R}(A^\dagger) = \mathcal{R}(A^*) \subseteq \mathcal{N}(B-A) \quad \Longleftrightarrow \quad (B-A)A^\dagger = O.$$

Por lo tanto, $A^*A = A^*B$ y $AA^* = BA^*$ es equivalente a $A^\dagger A = A^\dagger B$ y $AA^\dagger = BA^\dagger$.

(a) $\Longleftrightarrow$ (c) Se deduce de la propiedad $M^*N = O \Longleftrightarrow \mathcal{R}(M) \perp \mathcal{R}(N)$.

(a) $\Longrightarrow$ (d) Si $A^*A = A^*B$ y $AA^* = BA^*$ es claro que A^*B y BA^* son matrices hermíticas y definidas no negativas. Por el Teorema A.2.2 del Apéndice A.2, existen matrices unitarias $U_1 \in \mathbb{C}^{m\times m}$ y $V_1 \in \mathbb{C}^{n\times n}$ tales que

$$A = U_1 D_1 V_1 \qquad \text{y} \qquad B = U_1 D_2 V_1$$

donde D_1 y D_2 son matrices diagonales con elementos no negativos en la diagonal (es decir, D_1 y D_2 sólo tienen elementos no nulos en algunas de sus posiciones (i,i) aunque no son necesariamente matrices cuadradas). Por la hipótesis y del Lema 6.1, $D_1 \leq^* D_2$. Ahora, utilizando adecuadamente el Ejemplo 6.2 y escribiendo con detalle los tamaños de las matrices es posible demostrar (ejercicio) que existen dos matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = UD_3V \qquad \text{y} \qquad B = UD_4V,$$

con $D_3 = \operatorname{diag}(D_a, O, O)$ y $D_4 = \operatorname{diag}(D_a, D_{b-a}, O)$ donde $D_a \in \mathbb{R}^{a\times a}$ y $D_{b-a} \in \mathbb{R}^{(b-a)\times(b-a)}$, son matrices diagonales definidas positivas, lo que demuestra (d).

(d) $\Longrightarrow$ (a) Es una sencilla comprobación. $\square$

Ejemplo 6.3. *Es claro que las matrices*

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

cumplen que $A \leq^* B$ por el apartado (d) del Teorema 6.1.

Ejemplo 6.4. *Para las matrices*

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad y \qquad B = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

es claro que no cumple la relación $A \leq^* B$. Basta comprobar el apartado (b) del Teorema 6.1 observando que

$$A^\dagger = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

6.3. La relación binaria es un orden parcial

Teorema 6.2. *La relación binaria $\leq^*$ define un orden parcial sobre $\mathbb{C}^{m\times n}$.*

Demostración. La reflexividad es evidente.

Si $A, B \in \mathbb{C}^{m\times n}$ son tales que $A \leq^* B$ y $B \leq^* A$, entonces en particular se cumplen

$$A^*A = A^*B, \qquad B^*B = B^*A.$$

Luego,

$$(A-B)^*(A-B) = A^*A - A^*B - B^*A + B^*B = O.$$

De aquí, $A = B$, lo que se prueba la antisimetría.

Sean $A, B \in \mathbb{C}^{m\times n}$ tales que $A \leq^* B$ y $B \leq^* C$. Por el Teorema 6.1 se tiene que existen matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = U \begin{bmatrix} D_a & O & O \\ O & O & O \\ O & O & O \end{bmatrix} V \qquad \text{y} \qquad B = U \begin{bmatrix} D_a & O & O \\ O & D_{b-a} & O \\ O & O & O \end{bmatrix} V,$$

con D_a y D_{b-a} matrices diagonales definidas positivas (y, por tanto, invertibles). Sea

$$C = U \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} V,$$

particionada de modo que los productos implicados puedan realizarse. Usando que $B \leq^* C$, y realizando las multiplicaciones de la definición, se llega a

$$C = U \begin{bmatrix} D_a & O & O \\ O & D_{b-a} & O \\ O & O & C_{33} \end{bmatrix} V.$$

Ahora es fácil comprobar por definición que $A \leq^* C$. Por lo tanto, la relación $\leq^*$ es transitiva. $\square$

El orden parcial definido sobre $\mathbb{C}^{m\times n}$ por la relación binaria $\leq^*$ se denomina **orden parcial estrella**. Del Teorema 6.2 y de los apartados (a) y (b) del Teorema 6.1 queda clara la relación entre el nombre del orden parcial en relación a la inversa generalizada de Moore-Penrose utilizada.

Ejercicio 6.1. *Dadas las familias de matrices*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} x & y & z \\ 1 & 0 & 1 \\ t & u & v \end{bmatrix}$$

en función de los parámetros $x, y, z, t, u, v \in \mathbb{C}$, se pide:

(a) Calcular A^* y $A^\dagger$,

(b) Determinar los valores de los parámetros para los cuales $A \leq^* B$ utilizando:

(I) la definición de orden parcial estrella,

(II) el apartado (b) del Teorema 6.1,

(c) Para la matriz B hallada en el apartado anterior, comprobar que se cumplen las relaciones:

$$\mathcal{R}(A) \perp \mathcal{R}(B - A) \qquad \text{y} \qquad \mathcal{R}(A^*) \perp \mathcal{R}((B - A)^*),$$

(d) Encontrar matrices U, V, D_a y D_b en las condiciones del apartado (d) del Teorema 6.1 para la matriz B hallada en el apartado (b),

(e) Encontrar, si es posible, una matriz $C \in \mathbb{C}^{3\times 3}$ que cumpla:

(I) $A \leq^* C$,

(II) que C sea distinta de todas las matrices B encontradas en el apartado (b),

(III) C una matriz invertible.

(f) ¿Cuántas matrices C se podrían obtener en las condiciones del apartado (e)?

Capítulo 7

Algunos aspectos cronológicos y aplicaciones de las inversas generalizadas

A continuación se presenta una breve cronología de la aparición de algunas inversas generalizadas clásicas, algunas aplicaciones y otras más recientes.

El concepto de inversa generalizada fue mencionado por primera vez en 1903 por Fredholm [13] quien formuló una pseudoinversa para un operador integral lineal que no es invertible en el sentido ordinario. Un año después, Hilbert [17] abordó el estudio de inversas generalizadas de operadores diferenciales. Recién en 1920, Moore [28] señaló la existencia de una matriz, única, que actúa como la inversa de una matriz (no invertible) dada y la denominó recíproca general. Sin embargo, su definición no tuvo notoriedad en la comunidad científica debido, fundamentalmente, a la engorrosa notación empleada. La definición dada por Moore fue realizada desde un punto de vista funcional o geométrico mediante dos proyectores. El objetivo que se había propuesto Moore en el estudio de la inversa recíproca lo describe Ben-Israel en [3] del siguiente modo:

> *"The effectiveness of the reciprocal of a nonsingular finite matrix in the study of properties of such matrices makes it desirable to define if possible an analogous matrix to be associated with each finite matrix A even if A is not square or, if square, is not necessarily nonsingular."*

En su trabajo, Moore demostró lo siguiente:

Teorema 7.1. *Sea $A \in \mathbb{C}^{m\times n}$ y $b \in \mathcal{R}(A)$. Entonces la solución general del sistema $Ax = b$ es*

$$A^{\dagger}b + \{y : y \perp \mathcal{R}(A^*)\}.$$

pero no abordó el caso de sistemas lineales incompatibles, hecho que sí estudió posteriormente Penrose. En realidad, el gran objetivo de Moore fue encontrar una ambiciosa unificación que Ben-Israel describe como [3]:

> *"The existence of analogies between central features of various theories implies the existence of a more fundamental general theory embracing the special theories as particular instances and unifying them as to those central features."*

En 1955, Penrose [32] publicó sus resultados partiendo de una definición más bien algebraica pero, sin embargo, equivalente a la dada por Moore, como se probó en el Teorema 2.2. Desde entonces, esta inversa (única) es comúnmente conocida como la inversa de Moore-Penrose. Es interesante remarcar que el trabajo de Penrose fue independiente del de Moore.

Tres años después de la aparición del trabajo de Penrose, Drazin [10] introdujo una nueva inversa generalizada en el contexto de anillos abstractos y, en particular, para matrices cuadradas, que llamó la atención a la comunidad matemática por sus interesantes propiedades espectrales. Si $A \in \mathbb{C}^{n\times n}$ *tiene índice* k *(es decir,* k *es el menor entero no negativo tal que* $\mathrm{rg}(A^k) = \mathrm{rg}(A^{k+1})$*), la inversa de Drazin se define como la única matriz* $A^D \in \mathbb{C}^{n\times n}$ *que satisface las tres ecuaciones matriciales*

$$(1^k)\, A^{k+1}A^D = A^k, \qquad (2)\, A^D A A^D = A^D, \qquad (5)\, AA^D = A^D A.$$

Cuando $k = 1$*, la inversa de Drazin de* A *se llama inversa de grupo de* A *y se denota por* $A^{\#}$*. Por lo tanto, la inversa de grupo es un caso especial de la inversa de Drazin. Sin embargo, la inversa de grupo es utilizada en muchas aplicaciones interesantes y es por eso que se la considera como una entidad con importancia intrínseca. Por mencionar alguna de ellas se cita el trabajo de C.D. Meyer titulado The role of the group generalized inverse in the theory of finite Markov chains (véase la cita [1282] de [4]), donde hizo una importante contribución a la teoría de las inversas generalizadas mostrando la aplicación de la inversa de grupo a las cadenas de Markov.*

Las inversas generalizadas cubren un amplio rango de áreas dentro de las Matemáticas: Teoría de matrices, teoría de operadores, C^**-álgebras, anillos, etc. Los estudios más recientes se centran en: cálculo numérico, leyes del orden inverso, teoría de perturbación, órdenes parciales matriciales, etc. Numerosas aplicaciones incluyen áreas tales como: Estadística, ecuaciones diferenciales, análisis numérico, cadenas de Markov, criptografía, teoría de control, teoría de códigos, recuperación de datos y robótica, por mencionar sólo algunas.*

Muchas propiedades y aplicaciones fueron desarrolladas a partir de la inversa de Moore-Penrose, la inversa de grupo y la inversa de Drazin. Más específicamente, la inversa de Moore-Penrose juega un papel determinante en la resolución aproximada de sistemas de ecuaciones lineales por el método de mínimos cuadrados (véase la sección 3) y en el análisis de problemas de control óptimo. Las $\{1\}$*-inversas generalizadas de una matriz* $A \in \mathbb{C}^{m\times n}$ *dada, son aquellas matrices* $X \in \mathbb{C}^{n\times m}$ *que cumplen la condición* $AXA = A$ *de Penrose. Estas inversas, introducidas por Rao [37], permiten encontrar todas las soluciones de un sistema de ecuaciones lineales y proporcionar una representación de las mismas. Por otra parte, la inversa de grupo se aplica, entre* *otras cosas, para encontrar el estado estacionario en la teoría de cadenas de Markov [6, 19], y la inversa de Drazin permite resolver ecuaciones diferenciales (y en diferencias) matriciales cuyos coeficientes sean matrices singulares. Todas estas inversas generalizadas son casos especiales de la* $\{2\}$*-inversa generalizada (cumplen que* $XAX = X$*) cuyos espacio imagen y espacio nulo son prescritos [7, 38, 41]. Una amplia repercusión debido a su aplicación en la resolución de algunos sistemas lineales con restricciones que surgen en la teoría de redes eléctricas tuvo la conocida como inversa generalizada de Bott-Duffin.*

En la última década han aparecido nuevas inversas generalizadas. La primera de ellas fue introducida en el año 2010 por Baksalary y Trenkler [1]. Para una matriz $A \in \mathbb{C}^{n\times n}$*, la inversa core se define como la única matriz* $A^{\#\!\!\!\bigcirc} \in \mathbb{C}^{n\times n}$ *que verifica las condiciones* $AA^{\#\!\!\!\bigcirc} = P_{\mathcal{R}(A)}$ *y* $\mathcal{R}(A^{\#\!\!\!\bigcirc}) \subseteq \mathcal{R}(A)$*. Sin embargo, esta nueva inversa se limita a matrices cuadradas de índice a lo sumo 1. Esta limitación condujo a que muchos autores intentaran extenderla a matrices de índice arbitrario. La primera de ellas fue introducida en el año 2014 por Baksalary y Trenkler, se define como* $A^{\diamond} = (A^2A^{\dagger})^{\dagger}$ *y se la suele citar como la inversa BT de* A *[2]. En el mismo año, Prasad y Mohana consideraron otra definición alternativa y la llamaron inversa core-EP. La inversa core-EP de* A *se define como* $A^{\dagger\!\!\!\bigcirc} = A^k((A^*)^kA^{k+1})^{\dagger}(A^*)^k$ *[34]. Paralelamente, Malik y Thome definieron, mediante la expresión* $A^{D,\dagger} = A^DAA^{\dagger}$*, la llamada inversa DMP y mediante* $A^{\dagger,D} = A^DAA^{\dagger}$*, la llamada inversa DMP dual de* A *[22], también definidas para matrices cuadradas de índice arbitrario. Todas estas inversas existen para toda matriz dada, son únicas y cada una de ellas está caracterizada por un determinado grupo de ecuaciones matriciales. Posteriormente, la conocida como inversa CMP de* A *fue definida por Mehdipour y Salemi en 2017 para matrices cuadradas de índice arbitrario como* $A^{C,\dagger} = A^{\dagger}AA^DAA^{\dagger}$ *[24].*

Muchos trabajos de investigación han surgido a partir de estas inversas incluyendo sus extensiones a conjuntos más generales como álgebra de operadores y anillos abstractos [21, 29, 30, 31, 35, 42, 43].

Por otro lado, el interés por los órdenes parciales matriciales y los preórdenes matriciales se han desarrollado paralelamente a los avances de la teoría de inversas generalizadas y éstos han sido estudiados por numerosos autores. Utilizando la inversa de Moore-Penrose, Drazin introdujo el orden parcial estrella en 1978 en [11]. Debilitando las condiciones de Drazin, el orden parcial menos, introducido por Hartwig en 1980 en [16] mediante $\{1\}$*-inversas generalizadas, permite analizar la propiedad conocida como substractivity rank, por la que el rango de la diferencia de dos matrices es la diferencia de sus rangos. Otro orden parcial matricial, introducido por S.K. Mitra en 1987 en [26], es el orden sharp sobre el conjunto de matrices de índice a lo sumo 1 a partir de la inversa de grupo. Todos ellos corresponden a órdenes parciales matriciales conocidos como* $\mathcal{G}$*-based (es decir, basados en una inversa generalizada) presentan una definición semejante a:*

$$A \leq^? B \qquad \Longleftrightarrow \qquad A^?A = A^?B \quad \text{y} \quad AA^? = BA^?,$$

donde $\leq^?$ indica alguno de los órdenes parciales mencionados y $A^?$ la correspondiente inversa generalizada. Al intentar utilizar la misma idea con la inversa de Drazin, se obtuvo el pre-orden Drazin. Por otra parte, Malik, Rueda y Thome [23] analizaron en profundidad el orden parcial core, introducido por O.M. Baksalary y G. Trenkler en [1]. Otro orden parcial fue definido y estudiado recientemente por Guterman, Herrero y Thome [14] y está basado en una descomposición matricial espectralmente ortogonal; este orden no es del tipo $\mathcal{G}$-based. Sobre un orden parcial matricial definido, por Hernández, Lattanzi y Thome en 2014, a partir de la inversa de Drazin ponderada se puede consultar [15]. Mosić, Djordjević, Wei y Patrício han realizado numerosas aportaciones a la literatura donde tratan las extensiones a espacios de Hilbert, espacios de Banach y anillos de resultados obtenidos previamente sobre el anillo de las matrices [20, 31, 36]. El caso de los órdenes parciales para las inversas generalizadas clásicas son tratados con detalle en el libro de Mitra, Bhimasankaram y Malik [27]. Otra relación binaria en la teoría de órdenes parciales matriciales estudiada recientemente es el preorden core-EP introducido en [39] para matrices cuadradas y luego extendido a operadores lineales acotados sobre espacios de Hilbert en [29]. Para indicar un último orden parcial, se mencionará el orden parcial G-Drazin introducido en [40] en 2016 y luego generalizado a matrices rectangulares y a operadores en [8] y [30], en 2018 y 2019, respectivamente.

Estas notas finalizan con el siguiente comentario. Adi Ben-Israel en su libro [4, pág. 282] responde a la siguiente pregunta:

> *¿Qué es lo más importante que se puede hacer con las inversas generalizadas que no se podría hacer sin ellas?*

Tras su dilatada carrera en el tema respondió que, probablemente, es el cálculo del proyector $P_{S\cap W}$ sobre una intersección de subespacios S y W de un espacio vectorial V en términos de los proyectores sobre cada subespacio P_S y P_W. Está dado por la siguiente fórmula:

$$P_{S\cap W} = 2P_S(P_S + P_W)^\dagger P_W = 2P_W(P_S + P_W)^\dagger P_S.$$

Una demostración se puede ver en (el Manual de Soluciones de) [25].

Bibliografía

[1] *O.M. Baksalary, G. Trenkler*, Core inverse of matrices, *Linear and Multilinear Algebra, 58, 681–697, 2010.*

[2] *O.M. Baksalary, G. Trenkler*, On a generalized core inverse, *Applied Mathematics and Computation, 236, 450–457, 2014.*

[3] *A. Ben-Israel*, The Moore of the Moore-Penrose inverse, *Electronic Journal of Linear Algebra, 9, 150–157, 2002.*

[4] *A. Ben-Israel, T.N.E Greville*, Generalized Inverses: Theory and Applications, *Segunda Edición, Springer-Verlag, Nueva York, 2003.*

[5] *R. Bott, R.J. Duffin*, On the algebra of networks, *Transactions American Mathematical Society, 74, 99–109, 1953.*

[6] *S.L. Campbell, C.D. Meyer*, Generalized Inverses of Linear Transformations, *SIAM, Filadelfia, 2009.*

[7] *Y. Chen, X. Chen*, Representation and approximation of the outer inverse of a matrix, *Linear Algebra and its Applications, 308, 85–107, 2000.*

[8] *C. Coll, M. Lattanzi, N. Thome*, Weighted G-Drazin inverses and a new pre-order on rectangular matrices, *Applied Mathematics and Computation, 317, 12–24, 2018.*

[9] *J. de Burgos*, Álgebra Lineal y Geometría Cartesiana, *Tercera Edición, McGrawHill, Madrid, 2006.*

[10] *M.P. Drazin*, Pseudo inverses in associative rings and semigroups, *The American Mathematical Monthly, 65, 506–514, 1958.*

[11] *M.P. Drazin*, Natural structures on semigroups with involution, *Bulletin of the American Mathematical Society, 84, 139–141, 1978.*

[12] *C. Eckart, G. Young*, A principal axis transformation for non-hermitian matrices, *Bulletin of the American Mathematical Society, 45, 118–121, 1939.*

[13] *I. Fredholm*, Sur une classe déquations fonctionnelles, *Acta Mathematicae, 27 365–390, 1903.*

[14] *A. Guterman, A. Herrero, N. Thome*, New matrix partial order based on spectrally orthogonal matrix decomposition, *Linear and Multilinear Algebra, 64, 3, 362–374, 2016.*

[15] *A. Hernández, M. Lattanzi, N. Thome*, Weighted binary relations involving the Drazin inverse, *Applied Mathematics and Computation, 253, 215–223, 2015.*

[16] *R.E. Hartwig*, How to order regular elements?, *Mathematica Japonicae, 25, 1–13, 1980.*

[17] *D. Hilbert*, Grundzüge einer algemeinen Theorie der linearen Integralgleichungen, *Götingen Nachrichten, 49–51, 1904.*

[18] *J. Ji*, Gauss-Jordan elimination methods for the Moore-Penrose inverse of a matrix, *Linear Algebra and its Applications, 437, 1835–1844, 2012.*

[19] *S.J. Kirkland, M. Neumann*, Group Inverses of M-Matrices and Their Applications, *Chapman and Hall/CRC, Londres, 2012.*

[20] *L. Lebtahi, P. Patrício, N. Thome*, The diamond partial order in rings, *Linear and Multilinear Algebra, 62, 3, 386–395, 2014.*

[21] *T. Li, J. Chen*, Characterizations of core and dual core inverses in rings with involution, *Linear and Multilinear Algebra, 66, 4, 717–730, 2018.*

[22] *S. Malik, N. Thome*, On a new generalized inverse for matrices of an arbitrary index, *Applied Mathematics and Computation, 226, 575–580, 2014.*

[23] *S. Malik, L. Rueda, N. Thome*, Further properties on the core partial order and other matrix partial orders, *Linear and Multilinear Algebra, 62, 12, 1629–1648, 2014.*

[24] *M. Mehdipour, A. Salemi*, On a new generalized inverse of matrices, *Linear and Multilinear Algebra, 66, 5, 1046–1053, 2017.*

[25] *C.D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, Nueva York, 2010.*

[26] *S.K. Mitra*, On group inverses and the sharp order, *Linear Algebra and its Applications, 92, 17–37, 1987.*

[27] *S.K. Mitra, P. Bhimasankaram, S. Malik*, Matrix partial orders, shorted operators and applications, *World Scientific Publishing Company, 2010.*

[28] *E.H. Moore*, On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society, 26, 394–395, 1920.*

[29] *D. Mosić*, Core-EP pre-order of Hilbert space operators, *Quaestiones Mathematicae, 41, 5, 585–600, 2018.*

[30] *D. Mosić*, Weighted core-EP inverse of an operator between Hilbert spaces, *Linear and Multilinear Algebra, 67, 2, 2019.*

[31] D. Mosić, D.S. Djordjević, Reverse order law for the group inverse in rings, *Applied Mathematics and Compututation, 219, 2526–2534, 2012.*

[32] R. Penrose, A generalized inverse for matrices, *Proceedings of the Cambridge Philosophical, 51, 466–473, 1955.*

[33] R. Penrose, On best approximate solution of linear matrix equations, *Proceedings of the Cambridge Philosophical Society, 52, 17–19, 1956.*

[34] K.M. Prasad, K.S. Mohana, Core EP inverse, *Linear and Multilinear Algebra, 62, 3, 792–802, 2014.*

[35] D.S. Rakić, D.S. Djordjević, Core inverse and core partial order of Hilbert space operators, *Applied Mathematics and Computation, 244, 283–302, 2014.*

[36] D.S. Rakić, D.S. Djordjević, Space pre-order and minus partial order for operators on Banach spaces, *Aequationes Mathematicae, 85, 429–448, 2013.*

[37] C.R. Rao, S.K. Mitra, Generalized Inverse of Matrices and its Applications, *Wiley, Nueva York, 1971.*

[38] X. Sheng, G. Chen, New proofs of two representations and minor of generalized inverse $A^{(2)}_{T,S}$, *Applied Mathematics and Computation, 217, 6309–6314, 2011.*

[39] H. Wang, Core EP Decomposition and its applications, *Linear Algebra and its Applications, 508, 289–300, 2016.*

[40] X. Wang, X. Liu, Partial orders based on core-nilpotent decomposition, *Linear Algebra and its Applications, 488, 235–248, 2016.*

[41] Y. Wei, A characterization and representation of the generalized inverse and its applications, *Linear Algebra and its Applications, 280, 87–96, 1998.*

[42] H. Zhu, On DMP inverses and m-EP elements in rings, *Linear and Multilinear Algebra, doi: 10.1080/03081087.2018.1432546*

[43] H. Zou, J. Chen, P. Patrício, Characterizations of m-EP elements in rings, *Linear and Multilinear Algebra, 66, 6, 1244–1256, 2018.*

Apéndices

A.1. Diagonalización simultánea

Es conocido que toda matriz $A \in \mathbb{C}^{n\times n}$ (cuadrada) es **unitariamente diagonalizable**, *es decir existe una matriz unitaria $U \in \mathbb{C}^{n\times n}$ y una matriz diagonal $D \in \mathbb{C}^{n\times n}$ tales que*

$$A = UDU^*.$$

El siguiente resultado proporciona una condición necesaria y suficiente para que dos matrices normales, que es conocido que son unitariamente diagonalizables, lo sean de forma simultánea.

Teorema A.1.1. *Sean $A, B \in \mathbb{C}^{n\times n}$ dos matrices normales. Las siguientes condiciones son equivalentes:*

(a) Existen una matriz unitaria $U \in \mathbb{C}^{n\times n}$ y dos matrices diagonales $D_A, D_B \in \mathbb{C}^{n\times n}$ tales que

$$A = UD_AU^* \qquad y \qquad B = UD_BU^*,$$

(b) $AB^* = B^*A$.

Demostración. (a) $\Longrightarrow$ (b) Reemplazando A y B por las expresiones de la hipótesis se tiene que:

$$AB^* = (UD_AU^*)(UD_BU^*)^* = UD_AD_B^*U^* = UD_B^*D_AU^* = (UD_BU^*)^*(UD_AU^*) = B^*A.$$

(b) $\Longrightarrow$ (a) Como $AA^* = A^*A$, existe una matriz unitaria $U_1 \in \mathbb{C}^{n\times n}$ tal que

$$A = U_1 \begin{bmatrix} \lambda_1 I_s & O \\ O & D \end{bmatrix} U_1^*,$$

con $\lambda_1 \in \sigma(A) - \sigma(D)$ y $D \in \mathbb{C}^{(n-s)\times(n-s)}$ una matriz diagonal.

Se particiona

$$B = U_1 \begin{bmatrix} W & X \\ Y & Z \end{bmatrix} U_1^*,$$

con $W \in \mathbb{C}^{s\times s}$ y $Z \in \mathbb{C}^{(n-s)\times(n-s)}$. Por hipótesis $AB^* = B^*A$, luego se tiene

$$\begin{bmatrix} \lambda_1 I_s & O \\ O & D \end{bmatrix} \begin{bmatrix} W^* & Y^* \\ X^* & Z^* \end{bmatrix} = \begin{bmatrix} W^* & Y^* \\ X^* & Z^* \end{bmatrix} \begin{bmatrix} \lambda_1 I_s & O \\ O & D \end{bmatrix}$$

lo que equivale a

$$\begin{bmatrix} \lambda_1 W^* & \lambda_1 Y^* \\ DX^* & DZ^* \end{bmatrix} = \begin{bmatrix} \lambda_1 W^* & Y^* D \\ \lambda_1 X^* & Z^* D \end{bmatrix},$$

es decir

$$\text{(1) } \lambda_1 Y^* = Y^* D, \qquad \text{(2) } DX^* = \lambda_1 X^* \qquad \text{y} \qquad \text{(3) } DZ^* = Z^* D.$$

De (2) y teniendo en cuenta que $\lambda_1 \notin \sigma(D)$ debe ser $X = O$. De manera similar, de (1) y puesto que $\lambda_1 \notin \sigma(D)$ debe ser $Y = O$. Luego,

$$B = U_1 \begin{bmatrix} W & O \\ O & Z \end{bmatrix} U_1^*,$$

con $DZ^* = Z^* D$.

Como B es normal, se tiene que

$$B_1 := U_1^* B U_1 = \begin{bmatrix} W & O \\ O & Z \end{bmatrix}$$

también lo es y, en consecuencia, W y Z también son matrices normales. Por lo tanto, W y Z son unitariamente diagonalizables, es decir existen matrices unitarias $Q \in \mathbb{C}^{s\times s}$ y $T \in \mathbb{C}^{(n-s)\times(n-s)}$ y matrices diagonales $D_W \in \mathbb{C}^{s\times s}$ y $D_Z \in \mathbb{C}^{(n-s)\times(n-s)}$ tales que

$$W = QD_W Q^* \qquad \text{y} \qquad Z = TD_Z T^*.$$

Entonces

$$B = U_1 \begin{bmatrix} QD_W Q^* & O \\ O & TD_Z T^* \end{bmatrix} U_1^* = U_1 \begin{bmatrix} Q & O \\ O & T \end{bmatrix} \begin{bmatrix} D_W & O \\ O & D_Z \end{bmatrix} \begin{bmatrix} Q & O \\ O & T \end{bmatrix}^* U_1^*.$$

Denotando

$$R := U_1 \begin{bmatrix} Q & O \\ O & T \end{bmatrix}$$

se tiene que

$$B = R \begin{bmatrix} D_W & O \\ O & D_Z \end{bmatrix} R^*$$

y

$$\begin{aligned} A &= U_1 \begin{bmatrix} \lambda_1 I_s & O \\ O & D \end{bmatrix} U_1^* \\ &= U_1 \begin{bmatrix} Q & O \\ O & T \end{bmatrix} \begin{bmatrix} Q^*(\lambda_1 I_s) Q & O \\ O & T^* D T \end{bmatrix} \begin{bmatrix} Q & O \\ O & T \end{bmatrix}^* U_1^* \\ &= R \begin{bmatrix} \lambda_1 I_s & O \\ O & T^* D T \end{bmatrix} R^*. \end{aligned}$$

Ahora $A_2 := T^*DT$ y $B_2 := D_Z$ son matrices normales de tamaño $(n-s) \times (n-s)$ que cumplen la relación $A_2 B_2^* = B_2^* A_2$ pues $DZ^* = Z^* D$ (ejercicio). De este modo, el problema se ha deflacionado, es decir se ha reducido a uno similar de tamaño menor. Tras realizar un razonamiento similar para todos los valores propios de A se obtiene el resultado. □

Se observa que las matrices diagonales $D_A, D_B \in \mathbb{C}^{n\times n}$ del Teorema A.1.1 no tienen restricción alguna.

El siguiente resultado proporciona una condición necesaria y suficiente para que dos matrices hermíticas, lo sean de forma simultánea. Claramente, es un caso particular del Teorema A.1.1.

Teorema A.1.2. *Sean $A, B \in \mathbb{C}^{n\times n}$ dos matrices hermíticas. Las siguientes condiciones son equivalentes:*

(a) Existen una matriz unitaria $U \in \mathbb{C}^{n\times n}$ y dos matrices diagonales $D_A, D_B \in \mathbb{R}^{n\times n}$ tales que

$$A = UD_AU^* \qquad \text{y} \qquad B = UD_BU^*,$$

(b) $AB = BA$.

Ahora se observa que las matrices diagonales $D_A, D_B \in \mathbb{R}^{n\times n}$ del Teorema A.1.2 deben ser a coeficientes reales.

A.2. Descomposición en valores singulares simultánea

Es conocido que toda matriz rectangular $A \in \mathbb{C}_r^{m\times n}$, con $r > 0$, admite (siempre) una **descomposición en valores singulares**, *es decir existen dos matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que*

$$A = UDV^*$$

con $D \in \mathbb{R}^{m\times n}$ una matriz diagonal (rectangular), donde los primeros r elementos de su diagonal son positivos.

El siguiente resultado proporciona una condición necesaria y suficiente para que un par de matrices rectangulares admitan una descomposición similar a la descomposción en valores singulares de forma simultánea (a falta de la no negatividad de los elementos de ambas matrices diagonales). Es decir, permite realizar una transformación de ejes principales de forma simultánea a un par de matrices no necesariamente hermíticas ni tan siquiera cuadradas.

Teorema A.2.1. *Sean $A, B \in \mathbb{C}^{m\times n}$. Las siguientes condiciones son equivalentes:*

(a) Existen dos matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = UD_AV^* \qquad \text{y} \qquad B = UD_BV^*,$$

donde $D_A, D_B \in \mathbb{R}^{m\times n}$ son matrices diagonales (rectangulares) y D_A tiene elementos no negativos en su diagonal.

(b) AB^* y B^*A son hermíticas.

Demostración. (a) $\Longrightarrow$ (b) Reemplazando las expresiones de A y B por las de la hipótesis se tiene que:

$$AB^* = (UD_AV^*)(UD_BV^*)^* = (UD_AV^*)(VD_B^*U^*) = UD_AD_B^*U^*$$

que es hermítica puesto $(AB^*)^* = U(D_AD_B^*)^*U^*$ y $D_AD_B^*$ lo es (ejercicio). De forma semejante se demuestra que B^*A es hermítica. Se observa que $AB^* \in \mathbb{C}^{m\times m}$, $B^*A \in \mathbb{C}^{n\times n}$, $D_AD_B^* \in \mathbb{C}^{m\times m}$ y $D_B^*D_A \in \mathbb{C}^{n\times n}$.

(b) $\Longrightarrow$ (a) Se considera una descomposición en valores singulares de A, es decir,

$$U_1^*AV_1 = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$$

con $U_1 \in \mathbb{C}^{m\times m}$ y $V_1 \in \mathbb{C}^{n\times n}$ unitarias y $D \in \mathbb{R}^{r\times r}$ una matriz diagonal definida positiva.

Sea

$$U_1^*BV_1 = \begin{bmatrix} G & K \\ L & H \end{bmatrix}$$

una partición acorde a los tamaños de los bloques de $U_1^*AV_1$. Ahora, puesto que

$$AB^* = U_1 \begin{bmatrix} DG^* & DL^* \\ O & O \end{bmatrix} U_1^*$$

es hermítica, se tiene que DG^* es hermítica y $L = O$ pues D es invertible.

Análogamente, usando que B^*A es hermítica, se llega a que G^*D es hermítica y $K = O$.

Sean $G = [g_{ij}]_{i,j=1}^r$ y $D = \mathrm{diag}(d_1, \ldots, d_r)$. Usando que DG^* es hermítica se obtiene

$$d_i\overline{g_{ji}} = d_j g_{ij}. \tag{1}$$

Del hecho que G^*D es hermítica se llega a

$$d_j\overline{g_{ji}} = d_i g_{ij}.$$

(Comprobar los cálculos anteriores como ejercicio). Operando se consigue $(d_j^2 - d_i^2)g_{ij} = 0, \forall i, j$ y teniendo en cuenta que $d_i > 0, \forall i$ se tiene

$$(d_j - d_i)g_{ij} = 0, \forall i, j. \tag{2}$$

Si $d_i = d_j (> 0)$, de (1) se llega a $\overline{g_{ji}} = g_{ij}$. Si $d_i \neq d_j$, de (2) se llega a $g_{ij} = 0$. En ambos casos, $\overline{g_{ji}} = g_{ij}$, para todo i, j, con lo cual G es hermítica[1].

Por lo tanto, como G y D son dos matrices hermíticas y, además, conmutan (ejercicio), del Teorema A.1.2 se tiene que existe una matriz unitaria $P \in \mathbb{C}^{r\times r}$ tal que

$$P^*DP = D_1 \qquad \text{y} \qquad P^*GP = M,$$

con D_1 y M matrices diagonales con elementos reales en la diagonal.

Realizando una descomposición en valores singulares a H se llega a

$$Q^*HR = N$$

con N diagonal con elementos diagonales reales y Q y R unitarias.

Definiendo ahora,

$$U_2 := \begin{bmatrix} P & O \\ O & Q \end{bmatrix}, \qquad V_2 := \begin{bmatrix} P & O \\ O & R \end{bmatrix}, \qquad U := U_1U_2 \qquad \text{y} \qquad V := V_1V_2$$

se obtiene (ejercicio)

$$U^*AV = \begin{bmatrix} D_1 & O \\ O & O \end{bmatrix} =: D_A \qquad \text{y} \qquad U^*BV = \begin{bmatrix} M & O \\ O & N \end{bmatrix} =: D_B.$$

□

[1] Otra forma de demostrar que G es hermítica es mediante una reordenación los elementos d_i de D, de modo que todos los iguales queden agrupados, es decir $D = \mathrm{diag}(d_1I_{r_1}, d_2I_{r_2}, \ldots, d_sI_{r_s})$ siendo $d_i \neq d_j$ para $i \neq j$ con $r_1 + r_2 + \cdots + r_s = r$, particionando la matriz G en bloques de tamaños adecuados y utilizando el mismo truco usado para obtener la diferencia de cuadrados, se puede probar que G es diagonal por bloques donde cada bloque diagonal es hermítico, y por tanto unitariamente diagonalizables.

Por último, el resultado que sigue caracteriza el hecho de que dos matrices del mismo tamaño tengan una descomposición en valores singulares de forma simultánea.

Teorema A.2.2. *Sean $A, B \in \mathbb{C}^{m\times n}$. Las siguientes condiciones son equivalentes:*

(a) Existen dos matrices unitarias $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = UD_1V^* \qquad y \qquad B = UD_2V^*,$$

donde $D_1, D_2 \in \mathbb{R}^{m\times n}$ son matrices diagonales (rectangulares) con elementos no negativos en sus diagonales.

(b) AB^* y B^*A son hermíticas y al menos una de ellas es definida no negativa.

Demostración. (a) $\Longrightarrow$ (b) Esta implicación se prueba con un razonamiento parecido al utilizado en la demostración de (a) $\Longrightarrow$ (b) del Teorema A.2.1. En realidad, se puede probar no sólo que AB^* y B^*A son hermíticas y al menos una de ellas es definida no negativa sino que las dos lo son.

(b) $\Longrightarrow$ (a) Como AB^* y A^*B son hermíticas, del Teorema A.2.1 se tiene que existen dos matrices unitarias $W \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ tales que

$$A = WD_AV^* \qquad \text{y} \qquad B = WD_BV^*. \tag{3}$$

con $D_A, D_B \in \mathbb{R}^{m\times n}$ matrices diagonales (rectangulares) y D_A con elementos no negativos en la diagonal. Por ser, por ejemplo, AB^* definida no negativa, se tiene que

$$AB^* = WD_AV^*(WD_BV^*)^* = WD_AD_B^TW^*$$

y por tanto $D_AD_B^T$ es hermítica y definida no negativa.

Sean $\operatorname{diag}(D_A) = \{a_1, \ldots, a_q\}$ los elementos de la diagonal de D_A y $\operatorname{diag}(D_B) = \{b_1, \ldots, b_q\}$ los elementos de la diagonal de D_B.

Como $a_i \geq 0$ para todo $i = 1, \ldots, q$, se tiene que:

- si $a_i > 0$ entonces $b_i \geq 0$.
- si $a_{i_0} = 0$ para algún i_0, puede ocurrir:
  - $b_{i_0} \geq 0$,
  - $b_{i_0} < 0$. En este caso, se debe cambiar la i_0-ésima columna w_{i_0} de W por $-w_{i_0}$ y b_{i_0} por $-b_{i_0}$ en D_B.

Luego, se pueden construir nuevas matrices $\widetilde{D}_B$ y $\widetilde{W}$ satisfaciendo (3) donde $\widetilde{D}_B$ tiene sus elementos diagonales no negativos. □

Por último, en el caso del Teorema A.2.2 las matrices diagonales (rectangulares) $D_1, D_2 \in \mathbb{R}^{m\times n}$ *tienen elementos no negativos en sus diagonales, lo que muestra cómo han ido cambiando las diagonales desde matrices normales (y hermíticas) a rectangulares (arbitrarias).*

Índice alfabético