Facultad de Ciencias

notas de clase

Álgebra Lineal II [con aplicaciones en Estadística]

José Alfredo Jiménez Moscoso

Biología

Estadística

Farmacia

Física

Geociencias

Instituto de Ciencias Naturales

Matemáticas

Observatorio Astronómico

Química

UNIVERSIDAD NACIONAL DE COLOMBIA
Sede Bogotá

Notas de Clase Álgebra Lineal II Con Aplicaciones en Estadística

José Alfredo Jiménez Moscoso

Facultad de Ciencias
Universidad Nacional de Colombia
Sede Bogotá

A

Mi Esposa

Mi Hija

Mis Padres

Índice General

Prólogo

El álgebra de matrices ha llegado a ser en la actualidad un elemento esencial de los conocimientos matemáticos necesarios para ingenieros y científicos. Además, los conocimientos de los métodos fundamentales del álgebra matricial son necesarios para sociólogos, economistas, estudiantes de pedagogía y de comercio.

A pesar de las diversas aplicaciones del álgebra matricial, en la mayoría de textos de álgebra lineal no se introducen estos temas, por eso en muchos casos no se encuentra un libro que se ajuste a los requerimientos y necesidades de ciertas materias. Estas notas de clase están basadas en el curso de Álgebra Lineal II de la carrera de Estadística, el cual he tenido a mi cargo durante tres años y que he preparado usando diferentes textos. Se resaltan principalmente los libros referenciados en la bibliografía como Apostol (1985), Asmar (1995), Barbolla (1998), Bru (2001) y Searle (1982) los cuales fueron de gran apoyo en la elaboración de estas notas.

Este escrito será una ayuda adicional para aquellos estudiantes que toman varios cursos en los cuales deben tener o les serían provechosos los conocimientos del álgebra de matrices. Aun cuando en estas circunstancias siempre es inadecuado comenzar un curso de teoría de matrices, estas notas le permitirá al lector adquirir la práctica necesaria en el manejo de matrices.

El objetivo principal de estas notas consiste en capacitar al lector a fin de que adquiera la habilidad de usar el álgebra de matrices. Este material busca proporcionar al lector los conceptos de diagonalización y factorización matricial, formas cuadráticas e inversas generalizadas de una manera sencilla; durante su desarrollo se plantean ejemplos y ejercicios relacionados con la teoría.

Estas notas están escritas en forma programada por lo cual le ayudará al lector a alcanzar su principal objetivo. A cada lector le proporciona un medio individual para estudiar el tema expuesto y es muy útil como texto auto-didáctico. Además, le permitirá al lector avanzar a su propio ritmo. De esta manera, las notas pueden usarlas estudiantes con diferentes aptitudes, conocimientos y rapidez de lectura.

Por supuesto que estas notas pueden contener errores, espero que me los comenten en una forma constructiva, que me permita corregirlos. Esta es tal vez la única forma como en un ambiente académico se podrá ir avanzando.

Agradezco la colaboración del Departamento de Matemáticas quien a través de su oficina de publicaciones me permitió la divulgación de este material. También quiero dar las gracias tanto a los colegas que evaluaron este manuscrito como a mis estudiantes del curso de *Álgebra Lineal II* de la carrera de Estadística, por sus sugerencias y comentarios, los cuales fueron muy útiles en la redacción de estos apuntes.

José Alfredo Jiménez M.

Capítulo 1

Preliminares

Este capítulo es una recopilación de conceptos, procedimientos y resultados básicos que, por lo general, forman parte del primer curso de *Álgebra Lineal*. Por consiguiente, una gran parte de estos resultados aparecen sin prueba; además, en algunos casos se consideran temas que el lector debe manejar y que por su importancia son retomados posteriormente.

El propósito fundamental de este material es servir como prerrequisito para los siguientes capítulos y, como ya se mencionó, no se profundizara en los temas considerados en este capítulo. Si el lector tiene amplios conocimientos del contenido de este apartado puede pasar de inmediato al siguiente capítulo, aunque es recomendable que desarrolle las secciones 1.5 y 1.9.

1.1. Matrices

En esta sección se introducen los conceptos y las reglas básicas del álgebra de matrices. Dentro de los diferentes elementos estudiados por el Álgebra Lineal, uno de los más utilizados es el de matriz. Esto se debe a que la Teoría de Matrices ofrece, entre otras, la posibilidad de trabajar cómodamente con modelos de gran dimensión, tanto en número de variables, como de ecuaciones o datos, ya que brinda una notación simple y compacta para designar amplios conjuntos de información.

1.1.1. Conceptos básicos

En esta sección se presenta la definición formal del término “matriz”, las matrices se denotan con letras mayúsculas y con minúsculas los elementos que las constituyen.

Definición 1.1. Una matriz A de tamaño $m \times n$ es un arreglo rectangular de $m \cdot n$ números reales (o complejos[1]) dispuestos en m filas y n columnas, escritos entre corchetes (o paréntesis), como sigue

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix},$$

donde los subíndices indican la "fila" y la "columna" de localización en la matriz de cada número real (o complejo); a los números $a_{11}, a_{12}, \dots, a_{mn}$ se les llama elementos o entradas de la matriz.

Observación

Cuando $m = n$, la matriz recibe el nombre de *cuadrada*; si es $m \neq n$, se denomina *rectangular*. Al conjunto de todas las matrices de tamaño $m \times n$, se le notará por $\mathcal{M}_{mn}$.

Definición 1.2. Matrices iguales

Sean las matrices reales $A = [a_{ij}]$ y $B = [b_{ij}]$, se dice que son iguales, cuando teniendo el mismo tamaño, se verifica que

$$a_{ij} = b_{ij} \qquad \begin{array}{l} \forall \quad i = 1, 2, \dots, m; \\ \forall \quad j = 1, 2, \dots, n. \end{array}$$

[1]Si el lector no está familiarizado con estos números puede consultar el apéndice.

1.1.2. Operaciones con matrices

En esta sección se consideran las operaciones con matrices, además se recuerda que sólo se pueden sumar matrices que tienen el mismo tamaño y para que el producto sea posible es preciso que el número de columnas de la primera matriz coincida con el número de filas de la segunda.

Definición 1.3. Dadas $A = [a_{ij}]$ y $B = [b_{ij}]$, matrices de tamaño $m \times n$, la suma $A+B$, es una matriz $C = [c_{ij}]$ de tamaño $m \times n$, donde

$$c_{ij} = a_{ij} + b_{ij} \qquad \begin{array}{l} \forall \quad i = 1,2,\ldots,m, \\ \forall \quad j = 1,2,\ldots,n. \end{array}$$

Teorema 1.1. Propiedades básicas de la suma

Para todas A, B y C matrices de tamaño $m \times n$, se verifica que

1. Conmutativa: $A+B = B+A$.

2. Asociativa: $(A+B)+C = A+(B+C)$.

3. Existencia de elemento neutro o matriz nula: *Existe una matriz O de tamaño $m \times n$ en donde, todos sus elementos son iguales a cero, tal que $\forall A$ de tamaño $m \times n$, se verifica que*
$$A+O = O+A = A.$$

4. Elemento opuesto: *Para todas matriz A de tamaño $m \times n$, existe una matriz que llamaremos matriz opuesta de A y denotaremos por $-A$, que verifica*
$$A+(-A) = O.$$

La última propiedad permite, dadas dos matrices $A = [a_{ij}]$ y $B = [b_{ij}]$ del mismo tamaño $m \times n$, introducir el concepto de matriz diferencia $A - B$, la cual puede ser definida como sigue
$$A - B = A + (-B).$$

Demostración

Sean $A=[a_{ij}]$, $B=[b_{ij}]$ y $C=[c_{ij}]$.

1. $a_{ij}+b_{ij}=b_{ij}+a_{ij}$.
2. $(a_{ij}+b_{ij})+c_{ij}=a_{ij}+(b_{ij}+c_{ij})$.
3. Es evidente, que $\forall A$ de tamaño $m\times n$, se verifica que
$$a_{ij}+0=0+a_{ij}=a_{ij}.$$
4. Al tomar $-A=[-a_{ij}]$, se verifica que
$$a_{ij}+(-a_{ij})=(-a_{ij})+a_{ij}=0.$$

Definición 1.4. El producto de una matriz $A=[a_{ij}]$ de tamaño $m\times n$, por un escalar $\alpha\in\mathbb{R}$, es una matriz $C=[c_{ij}]$ del mismo tamaño que A, de elementos

$$c_{ij}=\alpha a_{ij} \qquad \begin{array}{ll}\forall & i=1,2,\ldots,m;\\ \forall & j=1,2,\ldots,n,\end{array}$$

esto es, los elementos de C se obtienen multiplicando los elementos correspondientes de A por α.

El resultado de efectuar el producto de una matriz A por un escalar α, se simboliza por αA y se lee *multiplicación de A por* α.

Teorema 1.2. Propiedades de la multiplicación por un escalar

Para todas A y B de tamaño $m\times n$ y $\alpha,\beta\in\mathbb{R}$ se satisface que

a) $\alpha(A+B)=\alpha A+\alpha B$, *b*) $(\alpha+\beta)A=\alpha A+\beta A$,

c) $\alpha(\beta A)=(\alpha\beta)A$, *d*) $1A=A$.

Demostración

Sean $A=[a_{ij}]$ y $B=[b_{ij}]$, entonces

a) $\alpha(A+B)=\alpha[a_{ij}+b_{ij}]=\alpha[a_{ij}]+\alpha[b_{ij}]$.

b) $(\alpha+\beta)A = (\alpha+\beta)[a_{ij}] = \alpha[a_{ij}] + \beta[a_{ij}]$.

c) $\alpha(\beta A) = \alpha(\beta[a_{ij}]) = (\alpha\beta)[a_{ij}]$.

d) $1A = 1[a_{ij}] = [a_{ij}]$.

Definición 1.5. Matriz identidad

Una matriz $A = [a_{ij}]$ de tamaño $n \times n$, cuyos elementos son

$$a_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

se llama matriz identidad y se denota por I_n.

Definición 1.6. Sean $A = [a_{ij}]$ y $B = [b_{jk}]$ de tamaño $m \times n$ y $n \times p$, respectivamente, entonces el producto de las matrices A y B, operación que se denotará por $A.B$, es una matriz C de tamaño $m \times p$, cuyo elemento genérico c_{ik} $(i = 1,2,\ldots,m; k = 1,2,\ldots,p)$, es

$$\begin{aligned} c_{ik} &= a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} \\ &= \sum_{j=1}^{n} a_{ij}b_{jk}. \end{aligned}$$

Teorema 1.3. Propiedades del producto de matrices

Sean A, B, C y D matrices reales tales que $A \in \mathcal{M}_{mn}$, $B, C \in \mathcal{M}_{np}$ y $D \in \mathcal{M}_{pq}$. Entonces se satisface que

1. Asociativa: $A.(B.D) = (A.B).D$.

2. Distributiva:

   a) $A.(B+C) = A.B + A.C$ *b*) $(B+C).D = B.D + C.D$.

3. El producto por una matriz nula O del tamaño adecuado es una matriz nula.

4. En general, esta operación matricial no es conmutativa:

$$A.B \neq B.A.$$

5. Existen matrices I_m e I_n, tales que

$$I_m.A = A \qquad \text{y} \qquad A.I_n = A.$$

Demostración

Queda como ejercicio para el lector.

Definición 1.7. Transpuesta

Si $A = [a_{ij}]$ es una matriz real de tamaño $m \times n$, se llama transpuesta de A a la matriz $B = [b_{ij}]$ de tamaño $n \times m$ cuyo elemento $b_{ij} = a_{ji}$. Se denota por A^t.

Teorema 1.4. Propiedades de la transpuesta

Sean A y B matrices de tamaño $m \times n$, C una matriz de tamaño $n \times p$ y sea $\alpha \in \mathbb{R}$. Entonces

1. $\left(A^t\right)^t = A$.
2. $(A \pm B)^t = A^t \pm B^t$.
3. $(\alpha A)^t = \alpha A^t$.
4. $(AC)^t = C^t A^t$.

Demostración

Queda como ejercicio para el lector.

1.1.3. Operaciones elementales sobre los renglones

Sea A una matriz real de tamaño $m \times n$, entonces las operaciones elementales en las filas de la matriz son:

R_1 Multiplicar cada elemento de la i-ésima fila por un escalar $\alpha \neq 0$.

R_2 Sumar a la i-ésima fila un múltiplo de la k-ésima fila.

R_3 Intercambiar (permutar) dos filas.

y las operaciones elementales en las columnas de la matriz son:

C_1 Multiplicar cada elemento de la j-ésima columna por un escalar $\alpha \neq 0$.

C_2 Sumar a la j-ésima columna un múltiplo de la l-ésima columna.

C_3 Intercambiar (permutar) dos columnas.

Definición 1.8. Matrices elementales

Una matriz $E_{kl}(\alpha)$ de tamaño $m \times m$ se llama una *matriz elemental* si es el resultado de aplicar una operación elemental a la matriz identidad I_m.

Realizar una operación elemental en una fila (o columna) de una matriz A es equivalente a premultiplicar (o postmultiplicar) a A respectivamente, por la matriz elemental adecuada. Esto se tiene de la definición de multiplicación de matrices que nos aclaró el hecho de que premultiplicar (o postmultiplicar) una matriz A por una matriz elemental daba el mismo resultado que aplicar la operación elemental a la fila correspondiente de A.

Notación

La notación que se usará para los tres tipos de operaciones R_1, R_2 y R_3 con matrices elementales, es el siguiente:

- La Matriz elemental tipo R_1, es una matriz $E_{kl}(\alpha) = [\nu_{ij}]$, cuyos elementos son

$$\nu_{ij} \underset{k=l}{=} \begin{cases} 1 & \text{si } i = j \neq k, \\ \alpha & \text{si } i = j = k, \\ 0 & \text{si } i \neq j. \end{cases}$$

  Nótese que es una matriz diagonal.

- La Matriz elemental tipo R_2, es una matriz $E_{kl}(\alpha) = [\nu_{ij}]$

$$\nu_{ij} \underset{k\neq l}{=} \begin{cases} 1 & \text{si } i = j, \\ \alpha & \text{si } i = k, j = l, \\ 0 & \text{en otro caso.} \end{cases}$$

  Esta matriz es triangular superior (inferior) dependiendo de la relación de orden que exista entre r y s. Además, si $k = l$, coincide con la matriz elemental tipo R_1.

- Matriz elemental tipo R_3, es una matriz $E_{kl}(1) = [\nu_{ij}]$

$$\nu_{ij} \underset{\alpha=1}{=} \begin{cases} 1 & \text{si } i = j, i \neq k, i \neq l, \\ 1 & \text{si } i = k, j = l, \\ 1 & \text{si } i = l, j = k, \\ 0 & \text{en otro caso.} \end{cases}$$

Definición 1.9. Matriz escalonada

Se dice que una matriz es escalonada si el número de ceros que precede al primer elemento diferente de cero de una fila, aumenta fila por fila, hasta tener posiblemente filas de sólo ceros.

Definición 1.10. Forma escalonada reducida

Una matriz se dice que es escalonada reducida si verifica las siguientes condiciones

i. Es una matriz escalonada.

ii. El primer elemento no nulo (por la izquierda) de cada fila no nula es un 1 y éste es el único elemento diferente de cero que se encuentra en la respectiva columna.

iii. Las filas nulas, si existen, están en la parte inferior de la matriz.

1.1.4. Traza de una matriz

En esta sección se estudiará una característica de las matrices cuadradas, la cual se expresa a través de un número llamado traza.

Definición 1.11. Traza de una matriz

Sea $A = [a_{ij}]$ una matriz real de tamaño $n \times n$, la suma de los elementos de la diagonal principal se llama traza de A y se denota como $tr(A)$, o sea

$$tr(A) = \sum_{i=1}^{n} a_{ii}. \tag{1.1}$$

Teorema 1.5. Propiedades

1. $tr(I_n) = n$, siendo, I_n la matriz identidad de tamaño $n \times n$.

2. $tr(O) = 0$, siendo, O la matriz nula de tamaño $n \times n$.

3. $tr(A) = tr(A^t)$.

4. $tr(A.A^t) = tr(A^t.A) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$.

5. $tr(\alpha A) = \alpha tr(A)$, con $\alpha \in \mathbb{R}$.

6. Si A y B son del mismo tamaño, $tr(A+B) = tr(A) + tr(B)$.

7. Si son posibles los productos $A.B$ y $B.A$, entonces se verifica

$$tr(A.B) = tr(B.A).$$

8. $tr(A.X) = 0$, Para todas $X \in \mathcal{M}_{nn}$, implica que $A = O$.

Demostración

Queda como ejercicio para el lector.

1.2. Inversa de una matriz

Es sabido que todo número $\alpha \neq 0$ tiene un inverso α^{-1} tal que

$$\alpha\alpha^{-1} = \alpha^{-1}\alpha = 1, \tag{1.2}$$

este hecho permite resolver las ecuaciones del tipo $\alpha x = \beta$, ya que multiplicando por α^{-1} se obtiene $x = \alpha^{-1}\beta$.

En este apartado se define un tipo de matriz que tiene una propiedad análoga en la teoría de matrices, la matriz inversa.

Definición 1.12. Inversa de una matriz

Sea A una matriz real de tamaño $n \times n$, si existe una matriz real B de tamaño $n \times n$, tal que

$$A.B = B.A = I_n. \tag{1.3}$$

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

Notación

Si A tiene inversa, entonces A se llama *matriz no singular o invertible* y si A no tiene inversa, entonces A se llama *matriz singular o no invertible*.

1.2.1. Método de Gauss-Jordan para encontrar la inversa de una matriz

Para encontrar la inversa de una matriz cuadrada A de tamaño $n \times n$, se procede de la siguiente manera

1. Se forma la matriz aumentada $B = (A \,|\, I_n)$ de tamaño $n \times 2n$.
2. Se aplican operaciones elementales entre filas, hasta llevar a B a una matriz escalonada reducida $C = \left(\tilde{A}_1 \,|\, \tilde{A}_2\right)$.
3. Se decide si A es no singular.
   a. Si $\tilde{A}_1 = I_n$, entonces $\tilde{A}_2 = A^{-1}$.
   b. Si $\tilde{A}_1 \neq I_n$, entonces $\tilde{A}_1$ tiene una fila de ceros. En este caso, A es singular, es decir A^{-1} no existe.

Teorema 1.6. Propiedades de la inversa de una matriz

1. Si una matriz A tiene inversa, ésta es única.

2. La inversa de la inversa es la matriz original. En símbolos,

$$\left(A^{-1}\right)^{-1} = A.$$

3. Si A y B son dos matrices invertibles, el producto $A.B$ es invertible y además

$$(A.B)^{-1} = B^{-1}.A^{-1}.$$

4. Si A es una matriz invertible

$$\begin{array}{lcl} a)\ B.A = O & \Rightarrow & B = O. \\ b)\ B.A = C.A & \Rightarrow & B = C. \end{array}$$

5. La inversa de una matriz transpuesta es la transpuesta de la inversa. En símbolos,

$$\left(A^{t}\right)^{-1} = \left(A^{-1}\right)^{t}.$$

Demostración

Queda como ejercicio para el lector.

1.3. Determinantes

Los determinantes permiten caracterizar cuando una matriz cuadrada es invertible. Un determinante de $n-$ésimo orden es una expresión asociada con una matriz $A = [a_{ij}]$ de tamaño $n \times n$, como se explica a continuación empezando con $n = 2$.

Definición 1.13. Sea $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ una matriz de tamaño 2×2. Entonces el determinante de A se define por

$$\det(A) = a_{11}.a_{22} - a_{12}.a_{21}. \tag{1.4}$$

Con frecuencia se denotará el $\det(A)$ por

$$|A| \qquad\qquad \text{o} \qquad\qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

aquí se usan barras (mientras que una matriz tiene corchetes).

Definición 1.14. Sea $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ una matriz de tamaño 3×3. Entonces el determinante de A se puede escribir en términos de los determinantes de matrices 2×2, como sigue:

$$\det(A) = a_{11}.\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}.\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}.\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{1.5}$$

o en la forma explícita siguiente

$$|A| = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).$$

También hay un artificio para memorizar esta fórmula, llamado *esquema de Sarrus*. Se agregan las dos primeras columnas a la derecha de A y se forman los productos de los elementos que van de la izquierda superior a la derecha inferior se les asigna el signo más y a los elementos que van de la izquierda inferior a la derecha superior el signo menos. A continuación se suman todos los productos con signo.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}.$$

Observación
El esquema de Sarrus no se aplica para matrices de tamaño 4×4, 5×5 o para las de más alto orden.

Hasta ahora, hemos evaluado los determinantes para matrices de tamaño 2×2 y 3×3, para evaluar el determinante de una matriz de tamaño $n \times n$, utilizaremos un método en el cual se reduce el problema a la evaluación de los determinantes de matrices de orden $n - 1$. Luego se repite el proceso para estas matrices de de tamaño $(n-1)\times(n-1)$ hasta llegar a las matrices de 2×2. Obsérvese que este procedimiento fue empleado para calcular el determinante en la expresión 1.5, se obtiene eliminando en A la fila y columna que respectivamente indican el primer y segundo subíndice del elemento a_{ij} por el que van multiplicadas, los determinantes de estas submatrices reciben el nombre de menores y cuando se les asocia los signos $+,-,+$ se denominan cofactores o adjuntos. Las definiciones de estos conceptos son

Definición 1.15. Menor y cofactor

Sea A una matriz real de tamaño $n\times n$.

1. Se llama menor complementario (i,j), al determinante de la submatriz de tamaño $(n-1)\times(n-1)$ que resulta de suprimir la $i-$ésima fila y la $j-$ésima columna de A. Se denota por $M_{ij}(A)$.

2. Se dice que
$$C_{ij}(A)=(-1)^{i+j}M_{ij}(A),$$
es el adjunto o cofactor (i,j) de A. Obsérvese que
$$(-1)^{i+j}=\begin{cases} 1 & \text{si } i+j \text{ es par,}\\ -1 & \text{si } i+j \text{ es impar.}\end{cases}$$

Definición 1.16. Matriz de cofactores

La matriz $C=[C_{ij}(A)]$, donde el elemento $C_{ij}(A)$ es el cofactor (i,j) de A, se denomina matriz de cofactores.

Teorema 1.7. Fórmula o expansión de Laplace

Sea $A=[a_{ij}]$ una matriz de tamaño $n\times n$. Entonces el determinante de A, se puede desarrollar usando

i) La expansión de Laplace por la $i-$ésima fila como

$$\det A = \sum_{j=1}^{n} a_{ij}C_{ij}(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}M_{ij}(A).$$

ii) La expansión de Laplace por la $j-$ésima columna como

$$\det A = \sum_{i=1}^{n} a_{ij}C_{ij}(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}M_{ij}(A).$$

Demostración

Queda como ejercicio para el lector.

Teorema 1.8. Propiedades de los determinantes

Dadas A, B y C matrices de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, se verifica que

i) $\det A^t = \det A$.

ii) Si se multiplica sólo una fila (o columna) de la matriz A por un escalar α, entonces el determinante también queda multiplicado por α.

iii) El determinante de la matriz αA es

$$\det(\alpha A) = \alpha^n \det A.$$

iv) Si todos los elementos de una fila (o columna) de A son cero, el valor del determinante es cero.

v) Si las matrices A, B y C difieren exclusivamente en los elementos de la $j-$ésima columna, siendo los elementos de esta columna para la matriz C la suma de los respectivos elementos de la $j-$ésima columna de las matrices A y B, entonces

$$\det C = \det A + \det B,$$

el mismo resultado se cumple cuando las tres matrices difieren de manera análoga en una fila.

vi) Si dos filas (o columnas) cualesquiera de A se intercambian, el valor del determinante se multiplica por -1.

vii) Si dos filas (o columnas) de A son proporcionales o iguales, el valor del determinante es cero.

viii) Si se suma un múltiplo escalar de una fila (o columna) de A a otra fila (o columna) de A, entonces el determinante no cambia.

ix) a) $\det(A.B) = \det A.\det B$.

b) Para cualquier $k \in \mathbb{N}$, $k \neq 0$ $\det\left(A^k\right) = (\det A)^k$.

c) Si A es invertible, entonces $\det\left(A^{-1}\right) = \dfrac{1}{\det A}$.

Demostración

Queda como ejercicio para el lector.

Teorema 1.9. Sea A una matriz de tamaño $n \times n$, A es invertible si y sólo si

$$\det A \neq 0.$$

Demostración

Queda como ejercicio para el lector.

Definición 1.17. Matriz adjunta

Sea A una matriz de tamaño $n \times n$. La matriz transpuesta de la matriz de cofactores $C_{ij}(A)$ es la adjunta de A y se representa por $Adj(A)$

$$Adj(A) = \begin{bmatrix} C_{11}(A) & C_{21}(A) & \dots & C_{n1}(A) \\ C_{12}(A) & C_{22}(A) & \dots & C_{n2}(A) \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n}(A) & C_{2n}(A) & \dots & C_{nn}(A) \end{bmatrix}. \tag{1.6}$$

Teorema 1.10. Sea A una matriz de tamaño $n \times n$, si $\det A \neq 0$, entonces

$$A^{-1} = \frac{1}{\det A} Adj(A).$$

Demostración

Queda como ejercicio para el lector.

Algunas fórmulas útiles para inversas

- Para una matriz invertible de tamaño 2×2 se obtiene

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad A^{-1} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{1.7}$$

- Para una matriz invertible de tamaño 3×3 se tiene

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|} \begin{bmatrix} C_{11}(A) & C_{21}(A) & C_{31}(A) \\ C_{12}(A) & C_{22}(A) & C_{32}(A) \\ C_{13}(A) & C_{23}(A) & C_{33}(A) \end{bmatrix} \tag{1.8}$$

1.4. Matrices especiales

Los tipos de matrices que se analizan a continuación tienen características particulares y se presentan frecuentemente en el desarrollo de la teoría y en las aplicaciones, de modo que han recibido denominaciones especiales.

Definición 1.18. Matrices triangulares

Una matriz cuadrada real $A = [a_{ij}]$ cuyos elementos abajo de la diagonal principal son todos cero, es decir, $a_{ij} = 0$ para $i > j$, se llama matriz triangular superior. De manera análoga, una matriz triangular inferior es una matriz cuadrada real A cuyos elementos arriba de la diagonal principal son cero, es decir, $a_{ij} = 0$ para $i < j$.

Teorema 1.11. Propiedades de las matrices triangulares

Sean $A, B \in \mathcal{M}_{nn}$ matrices triangulares superiores (inferiores) y $\alpha \in \mathbb{R}$, entonces:

i) Las matrices $A + B$ y αA son triangulares superiores (inferiores).

ii) La matriz $A.B$ es también triangular superior (inferior).

iii) El $\det(A)$ es igual al producto de los elementos de la diagonal principal.

iv) La transpuesta de A es triangular inferior (superior).

v) La matriz A es no singular si y sólo si cada uno de los elementos de la diagonal es distinto de cero.

vi) Si A es invertible, entonces A^{-1} es triangular superior (inferior).

Demostración
Queda como ejercicio para el lector.

Definición 1.19. Matrices simétricas

Una matriz cuadrada real $A = [a_{ij}]$ se llama simétrica si la transposición la mantiene invariable, es decir $[a_{ij}] = [a_{ji}]$.

Teorema 1.12. Propiedades de las matrices simétricas

Sean A y B matrices simétricas de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, entonces:

i) $A+B$ y αA son simétricas.

ii) Cuando $A.B = B.A$, entonces $A.B$ es simétrica. Sin embargo, esto no es cierto si A y B no conmutan en el producto.

iii) Si A es invertible entonces su inversa A^{-1} también es simétrica.

iv) Dada una matriz cualquiera C de tamaño $m \times n$,

   a) Si $m = n$, la matriz $\frac{1}{2}(C + C^t)$ es simétrica.

   b) Si $m \neq n$ ó $m = n$, las matrices $(C.C^t)$ y $(C^t.C)$ son simétricas.

Demostración

Queda como ejercicio para el lector.

Un tipo especial de matrices simétricas son las matrices escalares.

Definición 1.20. Matriz escalar

Una matriz real S de tamaño $n \times n$, se llama matriz escalar si resulta de la multiplicación de I_n por un escalar $c \in \mathbb{R}$, es decir

$$S = c.I_n.$$

Cuando todos los elementos s_{ii} de S no son iguales al escalar c, se tiene un nuevo tipo de matrices simétricas, las matrices diagonales.

Definición 1.21. Matrices diagonales

Una matriz cuadrada real $A = [a_{ij}]$ cuyos elementos arriba y abajo de la diagonal principal son todos cero, es decir, $a_{ij} = 0$ Para todas $i \neq j$, se llama matriz diagonal.

Teorema 1.13. Propiedades de las matrices diagonales

Si $D = [d_{ii}]$ es una matriz diagonal de tamaño $n \times n$, entonces:

i) Su producto por otra matriz diagonal también es una matriz diagonal.

ii) El $\det(D)$ es igual al producto de los elementos de la diagonal principal.

iii) D es una matriz no singular si y sólo si todos los elementos de la diagonal son distintos de cero.

iv) Si sus elementos de la diagonal principal $d_{11}, d_{22}, \ldots, d_{nn}$ son todos distintos de cero, D^{-1} es también una matriz diagonal con elementos en la diagonal principal iguales a $1/d_{11}, 1/d_{22}, \ldots, 1/d_{nn}$.

Demostración

i) Sea $C = [c_{ij}]$; entonces, el elemento ij de DC es

$$\sum_{k=1}^{n} d_{ik}c_{kj}$$

pero como D y C son matrices diagonales, entonces $d_{ik} = 0$ si $i \neq k$ y $c_{kj} = 0$ si $k \neq j$. Luego, el término $d_{ik}c_{kj} = 0$ si $i \neq j$. Por lo tanto, el único término que posiblemente es distinto de cero en esta suma es cuando $i = j = k$, es decir el término $d_{jj}c_{jj}$, el cual corresponde al elemento jj de $D.C$.

ii) Se procede por inducción. Si se desarrolla el $\det(D) = |D|$ por la primera columna, se obtiene $|D| = d_{11}|D'|$ donde D' es una submatriz real de tamaño $(n-1) \times (n-1)$ obtenida al borrar la primera fila y la primera columna de D. Ahora obsérvese que D' también es diagonal. Se deja al lector completar los detalles de la prueba.

iii) Dado que una matriz real de tamaño $n \times n$ es no singular si y sólo si su determinante es diferente de cero. Si D es diagonal, de la parte *ii)* se tiene que $|D| = d_{11}d_{22}\ldots d_{nn}$ y éste es distinto de cero si y sólo si cada $d_{ii} \neq 0$.

$iv)$ Esta se sigue inmediatamente a partir de $i)$ y de $iii)$. Si cada $d_{ii} \neq 0$

$$\begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix} \begin{bmatrix} 1/d_{11} & 0 & \dots & 0 \\ 0 & 1/d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/d_{nn} \end{bmatrix} = I_n.$$

Definición 1.22. Matrices antisimétricas

Una matriz cuadrada real $A = [a_{ij}]$ se llama antisimétrica si la transposición da como resultado la negativa de A, es decir $A^t = -A$.

Teorema 1.14. Propiedades de las matrices antisimétricas

Sean A y B matrices antisimétricas de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, entonces:

i) $A + B$ y αA son antisimétricas.

ii) Cuando $AB = BA$, entonces AB es antisimétrica. Sin embargo, esto no es cierto si A y B no conmutan en el producto.

iii) Si A es invertible entonces su inversa A^{-1} también es antisimétrica.

iv) Dada una matriz cualquiera C de tamaño $n \times n$, la matriz $\frac{1}{2}(C - C^t)$ es antisimétrica.

Demostración

Queda como ejercicio para el lector.

Definición 1.23. Matriz ortogonal

Sea $A = [\vec{a}_1 \ \vec{a}_2 \ \dots \ \vec{a}_n]$ una matriz real de tamaño $n \times n$, donde $\vec{a}_i$ es un vector $n \times 1$ que consiste de los elementos de la i-ésima columna de A. Entonces A es ortogonal si y sólo si

$$\vec{a}_i^t \vec{a}_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Teorema 1.15. Propiedades de las matrices ortogonales

Sean A y B matrices ortogonales de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, entonces:

i) A es ortogonal si y sólo si $A^t = A^{-1}$, o equivalentemente, si y sólo si

$$A^t A = I_n. \tag{1.9}$$

ii) AB y BA son ortogonales pero, en general, $A+B$ y αA no lo son.

iii) El valor absoluto del $\det A$ es 1.

iv) La transpuesta de una matriz ortogonal es ortogonal.

v) Dada una matriz antisimétrica C de tamaño $n \times n$, entonces la matriz $A = (I_n - C)(I_n + C)^{-1}$ es ortogonal.

Demostración

i) Por la unicidad de la inversa de una matriz y sus propiedades, se tiene que

$$\begin{aligned} A.A^t = I_n \Rightarrow A^{-1}.A.A^t = A^{-1} \Rightarrow A^t = A^{-1} \\ A^t = A^{-1} \Rightarrow A.A^t = A.A^{-1} \Rightarrow A.A^t = I_n. \end{aligned}$$

ii) Si A y B son ortogonales, entonces $A.B$ también lo es, ya que

$$(A.B)(A.B)^t = A.B.B^t.A^t = A.I_n.A^t = A.A^t = I_n$$

análogamente se prueba para $B.A$

iii) Si A es ortogonal, como $A^t.A = I_n$, se tiene que

$$\det(I_n) = \det\left(A^t.A\right) = \det\left(A^t\right)\det A = 1$$

y como $\det(A^t) = \det A$, se ve que $(\det A)^2 = 1$ y, por lo tanto,

$$\det A = \pm 1.$$

iv) Obsérvese que $A^t.A = I_n$ se puede escribir como $A^t.(A^t)^t = I_n$.

$v)$ Si $C \in \mathcal{M}_{nn}$ es antisimétrica, se tiene que $C^t = -C$ y, por las propiedades de la matriz transpuesta, resulta que

$$\begin{aligned} A^tA &= \left[(I_n - C)(I_n + C)^{-1}\right]^t (I_n - C)(I_n + C)^{-1} \\ &= \left[(I_n + C)^{-1}\right]^t (I_n - C)^t (I_n - C)(I_n + C)^{-1} \\ &= (I_n - C)^{-1}(I_n + C)(I_n - C)(I_n + C)^{-1} = I_n \end{aligned}$$

ya que $(I_n + C)(I_n - C) = (I_n - C)(I_n + C)$. Así pues, A es ortogonal en virtud de la ecuación (1.9).

Definición 1.24.

Una matriz ortogonal A, tal que $\det A = 1$, se llama **matriz ortogonal propia** y si el $\det A = -1$, se denomina **matriz ortogonal impropia**.

Ejemplo 1.1. ¿Es ortogonal la matriz $A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$?

Solución

Al multiplicar a A por la izquierda por A^t, se obtiene

$$A^tA = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Esto muestra, por la ecuación (1.9), que A es ortogonal.

Definición 1.25. Matriz de Permutación

La matriz elemental tipo R_3 de tamaño $n \times n$, se denomina matriz de permutación ya que resulta de intercambiar (permutar) el orden de las filas de la matriz I_n.

Teorema 1.16. Sea P una matriz de permutación, entonces

a) Para cualquier matriz A, se puede obtener PA a partir de A permutando las filas de A exactamente como se permutaron las filas de I_n para obtener P.

b) P es no singular y ortogonal.

Demostración

a) Esto se sigue fácilmente de las definiciones de la multiplicación de matrices y de matrices de permutación.

b) Separe P en sus respectivas filas $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n$, que son tan sólo las filas $\vec{e}_i^t$ de I_n en cierto orden. Entonces P^t tiene como columnas a $\vec{r}_i^t$. La definición de la multiplicación de matrices implica que el elemento (i, j) de PP^t es simplemente $\vec{r}_i\vec{r}_j^t$, y esto es

$$\vec{r}_i\vec{r}_j^t = \begin{cases} 1 & \text{si } i = j; \\ 0 & \text{si } i \neq j, \end{cases}$$

es decir, $PP^t = I_n$. De manera análoga, en términos de las columnas de P se demuestra que $P^tP = I_n$.

1.5. Matrices particionadas

Un rasgo importante en nuestro trabajo con matrices reales ha sido la habilidad de considerar una matriz A como una lista de vectores columna en lugar de simplemente una serie rectangular de números. Este punto de vista ha sido tan útil que nosotros deseamos considerar otras particiones de la matriz A, tomando por reglas la división de A horizontal y verticalmente. En esta sección estudiaremos la forma de particionar una matriz en submatrices que nos permitan realizar las mismas operaciones que definimos anteriormente para las matrices de una manera más sencilla.

1.5.1. Definiciones y operaciones

Si $A = [a_{ij}]$ la matriz que se obtiene después de que algunas filas y/o columnas de A se han eliminado es llamada una **submatriz de** A. Frecuentemente es conveniente particionar una matriz en submatrices y considerarla como una matriz cuyos elementos son estas submatrices.

Definición 1.26. Una matriz A de tamaño $m \times n$ puede particionarse de la siguiente manera

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix} \tag{1.10}$$

donde A_{11} es la submatriz real de tamaño $m_1 \times n_1$ formada por los elementos de A que ocupan las m_1 primeras filas y las n_1 primeras columnas, A_{12} la submatriz real de tamaño $m_1 \times n_2$ formada por los elementos de A que ocupan las m_1 primeras filas y las columnas $n_1+1, \dots, n_1+n_2$ y así sucesivamente. En general A_{ij} es la submatriz real de tamaño $m_i \times n_j$ formada por los elementos de A que ocupan las filas

$$m_1+\ldots+m_{i-1}+1, m_1+\ldots+m_{i-1}+2, \ldots, m_1+\ldots+m_{i-1}+m_i$$

y las columnas

$$n_1+\ldots+n_{j-1}+1, n_1+\ldots+n_{j-1}+2, \ldots, n_1+\ldots+n_{j-1}+n_j,$$

siendo m_i y n_j números naturales tales que

$$\begin{aligned} m_1+m_2+\ldots+m_r &= m, \\ n_1+n_2+\ldots+n_s &= n. \end{aligned}$$

Se indicará la partición diciendo que A es de tamaño

$$(m_1+m_2+\ldots+m_r) \times (n_1+n_2+\ldots+n_s).$$

Ejemplo 1.2. Sea la matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}.$$

Obtenga una partición de A de tamaños

$$(i)\ (2+2)\times(2+2+1) \qquad \text{y} \qquad (ii)\ (1+2+1)\times(2+3).$$

Solución

En el primer caso

$$A = \begin{bmatrix} a_{11} & a_{12} & \vdots & a_{13} & a_{14} & \vdots & a_{15} \\ a_{21} & a_{22} & \vdots & a_{23} & a_{24} & \vdots & a_{25} \\ \cdots & \cdots & . & \cdots & \cdots & . & \cdots \\ a_{31} & a_{32} & \vdots & a_{33} & a_{34} & \vdots & a_{35} \\ a_{41} & a_{42} & \vdots & a_{43} & a_{44} & \vdots & a_{45} \end{bmatrix},$$

la cual puede ser escrita de la forma

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} & \vdots & A_{13} \\ \cdots & . & \cdots & . & \cdots \\ A_{21} & \vdots & A_{22} & \vdots & A_{23} \end{bmatrix}$$

donde

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad A_{12} = \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix}, \qquad A_{13} = \begin{bmatrix} a_{15} \\ a_{25} \end{bmatrix},$$
$$A_{21} = \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}, \qquad A_{22} = \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix}, \qquad A_{23} = \begin{bmatrix} a_{35} \\ a_{45} \end{bmatrix}.$$

En el segundo caso

$$A = \begin{bmatrix} a_{11} & a_{12} & \vdots & a_{13} & a_{14} & a_{15} \\ \cdots & \cdots & . & \cdots & \cdots & \cdots \\ a_{21} & a_{22} & \vdots & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & \vdots & a_{33} & a_{34} & a_{35} \\ \cdots & \cdots & . & \cdots & \cdots & \cdots \\ a_{41} & a_{42} & \vdots & a_{43} & a_{44} & a_{45} \end{bmatrix},$$

la cual puede ser escrita de la forma

$$A = \begin{bmatrix} A'_{11} & \vdots & A'_{12} \\ \cdots & . & \cdots \\ A'_{21} & \vdots & A'_{22} \\ \cdots & . & \cdots \\ A'_{31} & \vdots & A'_{32} \end{bmatrix}$$

donde

$$A'_{11} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}, \qquad A'_{12} = \begin{bmatrix} a_{13} & a_{14} & a_{15} \end{bmatrix},$$
$$A'_{21} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \qquad A'_{22} = \begin{bmatrix} a_{23} & a_{24} & a_{25} \\ a_{33} & a_{34} & a_{35} \end{bmatrix},$$
$$A'_{31} = \begin{bmatrix} a_{41} & a_{42} \end{bmatrix}, \qquad A'_{31} = \begin{bmatrix} a_{43} & a_{44} & a_{45} \end{bmatrix}.$$

Definición 1.27. Submatriz principal

Si A es una matriz cuadrada se llama submatriz principal a toda submatriz de A formada eligiendo los mismos índices para las filas y las columnas.

El hecho de tomar las mismas filas y columnas es equivalente a que los elementos de la diagonal principal de la submatriz han de ser elementos que ya formaban parte de la diagonal principal de la matriz original. Luego si A es simétrica cualquier submatriz principal también es simétrica.

Ejemplo 1.3. Obtenga algunas submatrices principales de la matriz

$$A = \begin{bmatrix} -5 & 3 & 5 & 1 & 0 \\ -4 & 8 & 10 & 0 & -1 \\ 2 & 13 & 11 & 2 & 1 \\ 0 & 1 & 1 & 3 & 2 \\ 3 & 1 & 0 & 7 & 5 \end{bmatrix}$$

Solución

Las siguientes matrices son submatrices principales de A

$$B = \begin{bmatrix} -5 & 3 \\ -4 & 8 \end{bmatrix}, \qquad C = \begin{bmatrix} -5 & 5 \\ 2 & 11 \end{bmatrix},$$

$$D = \begin{bmatrix} 11 & 2 & 1 \\ 1 & 3 & 2 \\ 0 & 7 & 5 \end{bmatrix}, \qquad E = \begin{bmatrix} -5 & 1 & 0 \\ 0 & 3 & 2 \\ 3 & 7 & 5 \end{bmatrix}.$$

La submatriz B se ha formado con las filas y columnas 1 y 2. La submatriz C se ha conformado con las filas y columnas 1 y 3.

En cambio la submatriz

$$F = \begin{bmatrix} -5 & 1 & 0 \\ -4 & 0 & -1 \\ 2 & 2 & 1 \end{bmatrix},$$

no es principal, por que se ha obtenido con las filas 1, 2 y 3 y con las columnas 1, 4 y 5.

Definición 1.28. Submatriz angular

Las submatrices principales que son formadas con las primeras filas y columnas de la matriz A y que denotaremos por $A_{[k]}$, siendo k el orden de la submatriz, se denomina submatriz angular.

Si A es la matriz cuadrada de tamaño $n \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

entonces las submatrices angulares de A son

$$A_{[1]} = \begin{bmatrix} a_{11} \end{bmatrix},\ A_{[2]} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},\ A_{[3]} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},\ \ldots,\ A_{[n]} = A.$$

Así en el Ejemplo 1.3, la submatriz B sería una submatriz angular, concretamente $A_{[2]}$.

Definición 1.29. Dos matrices A y B están particionadas idénticamente si las submatrices resultantes contienen el mismo número de filas y de columnas y si, además, las partes correspondientes tienen el mismo tamaño. Por lo tanto, dos matrices particionadas idénticamente son iguales si y sólo si las submatrices correspondientes son iguales.

Definición 1.30. Suma de matrices particionadas

Sean A y B dos matrices particionadas idénticamente. Entonces la **suma** de A y B tendra igual partición, en este caso cada bloque de $A+B$ es obtenido de los correspondientes bloques de A y de B, en otras palabras

$$\begin{aligned} A+B &= \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1s} \\ A_{21} & A_{22} & \ldots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \ldots & A_{rs} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \ldots & B_{1s} \\ B_{21} & B_{22} & \ldots & B_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ B_{r1} & B_{r2} & \ldots & B_{rs} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \ldots & A_{1s}+B_{1s} \\ A_{21}+B_{21} & A_{22}+B_{22} & \ldots & A_{2s}+B_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1}+B_{r1} & A_{r2}+B_{r2} & \ldots & A_{rs}+B_{rs} \end{bmatrix}, \end{aligned}$$

donde las submatrices A_{ij} y B_{ij} son de tamaño $m_i \times n_j$.

Definición 1.31. Multiplicación por un escalar

Si A es una matriz real de tamaño $m \times n$ particionada y $\alpha \in \mathbb{R}$. Entonces la **multiplicación de un escalar** por A, es una matriz real de tamaño $m \times n$ obtenida de multiplicar cada bloque de A por el número α, en otras palabras,

$$\alpha A = \alpha \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix} = \begin{bmatrix} \alpha A_{11} & \alpha A_{12} & \dots & \alpha A_{1s} \\ \alpha A_{21} & \alpha A_{22} & \dots & \alpha A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha A_{r1} & \alpha A_{r2} & \dots & \alpha A_{rs} \end{bmatrix},$$

donde cada A_{ij} es una submatriz real de tamaño $m_i \times n_j$.

Definición 1.32. Transpuesta

Sea A una matriz real de tamaño $m \times n$ particionada de alguna manera. Entonces la **transpuesta** de A, que se escribe A^t, es una matriz real de tamaño $n \times m$ obtenida de intercambiar los renglones por las columnas en cada uno de los bloques A_{ij}, en otras palabras,

$$\text{si } A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix}, \text{ entonces } A^t = \begin{bmatrix} A_{11}^t & A_{21}^t & \dots & A_{r1}^t \\ A_{12}^t & A_{22}^t & \dots & A_{r2}^t \\ \vdots & \vdots & \ddots & \vdots \\ A_{1s}^t & A_{2s}^t & \dots & A_{rs}^t \end{bmatrix},$$

donde, A_{ij} es la submatriz real de tamaño $m_i \times n_j$.

Ejemplo 1.4. Obtenga la transpuesta de A

$$A = \begin{bmatrix} 8 & 9 & 3 & -5 \\ 20 & 10 & -10 & 8 \\ 21 & -5 & 13 & 5 \end{bmatrix}.$$

Solución

Consideremos la partición $(2+1)\times(2+2)$ de A, es decir,

$$A=\begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & . & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix}=\begin{bmatrix} 8 & 9 & \vdots & 3 & -5 \\ 20 & 10 & \vdots & -10 & 8 \\ \cdots & \cdots & . & \cdots & \cdots \\ 21 & -5 & \vdots & 13 & 5 \end{bmatrix}.$$

Luego

$$A_{11}^t=\begin{bmatrix} 8 & 9 \\ 20 & 10 \end{bmatrix}^t=\begin{bmatrix} 8 & 20 \\ 9 & 10 \end{bmatrix}, \qquad A_{12}^t=\begin{bmatrix} 3 & -5 \\ -10 & 8 \end{bmatrix}^t=\begin{bmatrix} 3 & -10 \\ -5 & 8 \end{bmatrix},$$

$$A_{21}^t=\begin{bmatrix} 21 & -5 \end{bmatrix}^t=\begin{bmatrix} 21 \\ -5 \end{bmatrix}, \qquad A_{21}^t=\begin{bmatrix} 13 & 5 \end{bmatrix}^t=\begin{bmatrix} 13 \\ 5 \end{bmatrix},$$

por consiguiente,

$$\begin{bmatrix} 8 & 9 & \vdots & 3 & -5 \\ 20 & 10 & \vdots & -10 & 8 \\ \cdots & \cdots & . & \cdots & \cdots \\ 21 & -5 & \vdots & 13 & 5 \end{bmatrix}^t=\begin{bmatrix} 8 & 20 & \vdots & 21 \\ 9 & 10 & \vdots & -5 \\ \cdots & \cdots & . & \cdots \\ 3 & -10 & \vdots & 13 \\ -5 & 8 & \vdots & 5 \end{bmatrix}.$$

Teorema 1.17. Multiplicación

Sean A y B matrices particionadas compatibles para el producto, digamos

$$A=\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & a_{r2} & \dots & A_{rs} \end{bmatrix} \qquad \text{y} \qquad B=\begin{bmatrix} B_{11} & B_{12} & \dots & B_{1t} \\ B_{21} & B_{22} & \dots & B_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ B_{s1} & B_{s2} & \dots & B_{st} \end{bmatrix}.$$

Entonces la multiplicación de las dos matrices es

$$A.B = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1t} \\ C_{21} & C_{22} & \dots & C_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r1} & C_{r2} & \dots & C_{rt} \end{bmatrix},$$

donde $C_{ik} = \sum\limits_{j=1}^{s} A_{ij}B_{jk}$.

Demostración

Consideremos $(m_1+m_2+\ldots+m_r)\times(n_1+n_2+\ldots+n_s)$ una partición de A y $(n_1+n_2+\ldots+n_s)\times(p_1+p_2+\ldots+p_r)$ una partición de B. Entonces,

$$\begin{aligned} A_{ij} &= [a_{hl}]_{\left\{\substack{m_1+m_2+\ldots+m_{i-1}+1\leq h\leq m_1+m_2+\ldots+m_{i-1}+m_i \\ n_1+n_2+\ldots+n_{j-1}+1\leq l\leq n_1+n_2+\ldots+n_{j-1}+n_j}\right\}} \\ B_{jk} &= [b_{lq}]_{\left\{\substack{n_1+n_2+\ldots+n_{j-1}+1\leq l\leq n_1+n_2+\ldots+n_{j-1}+n_j \\ p_1+p_2+\ldots+p_{k-1}+1\leq q\leq p_1+p_2+\ldots+p_{k-1}+p_k}\right\}} \\ A_{ij}B_{jk} &= \left[\sum_{l=n_1+\ldots+n_{j-1}+1}^{n_1+\ldots+n_{j-1}+n_j} a_{hl}b_{lq}\right]_{\left\{\substack{m_1+m_2+\ldots+m_{i-1}+1\leq h\leq m_1+m_2+\ldots+m_{i-1}+m_i \\ p_1+p_2+\ldots+p_{k-1}+1\leq q\leq p_1+p_2+\ldots+p_{k-1}+p_k}\right\}} \end{aligned}$$

y por tanto

$$\begin{aligned} C_{ik} &= \sum_{j=1}^{s} A_{ij}B_{jk} \\ &= \left[\sum_{l=1}^{n} a_{hl}b_{lq}\right]_{\left\{\substack{m_1+m_2+\ldots+m_{i-1}+1\leq h\leq m_1+m_2+\ldots+m_{i-1}+m_i \\ p_1+p_2+\ldots+p_{k-1}+1\leq q\leq p_1+p_2+\ldots+p_{k-1}+p_k}\right\}}, \end{aligned}$$

es decir, C_{ik} es el bloque (i,k) correspondiente a la partición

$$(m_1+m_2+\ldots+m_r)\times(p_1+p_2+\ldots+p_t)$$

de $A.B$.

Ejemplo 1.5. Calcular $A.B$, siendo

$$A=\begin{bmatrix} 8 & 9 & 3 & -5\\ 20 & 10 & -10 & 8\\ 21 & -5 & 13 & 5\end{bmatrix} \qquad \text{y} \qquad B=\begin{bmatrix} 2 & 1 & 3 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & 1\\ -1 & 1 & 2 & -1\end{bmatrix}.$$

Solución

Sean las particiones $(2+1)\times(2+2)$ para A y $(2+2)\times(2+2)$ para B

$$A=\begin{bmatrix} A_{11} & \vdots & A_{12}\\ \cdots & . & \cdots\\ A_{21} & \vdots & A_{22}\end{bmatrix}=\begin{bmatrix} 8 & 9 & \vdots & 3 & -5\\ 20 & 10 & \vdots & -10 & 8\\ \cdots & \cdots & . & \cdots & \cdots\\ 21 & -5 & \vdots & 13 & 5\end{bmatrix},$$

$$B=\begin{bmatrix} B_{11} & \vdots & B_{12}\\ \cdots & . & \cdots\\ B_{21} & \vdots & B_{22}\end{bmatrix}=\begin{bmatrix} 2 & 1 & \vdots & 3 & 1\\ 1 & 1 & \vdots & -1 & -1\\ \cdots & \cdots & . & \cdots & \cdots\\ 1 & -1 & \vdots & 1 & 1\\ -1 & 1 & \vdots & 2 & -1\end{bmatrix}.$$

Luego,

$$\begin{aligned}
A_{11}B_{11}+A_{12}B_{21} &= \begin{bmatrix} 25 & 17\\ 50 & 30\end{bmatrix}+\begin{bmatrix} 8 & -8\\ -18 & 18\end{bmatrix}=\begin{bmatrix} 33 & 9\\ 32 & 48\end{bmatrix},\\
A_{11}B_{12}+A_{12}B_{22} &= \begin{bmatrix} 15 & -1\\ 50 & 10\end{bmatrix}+\begin{bmatrix} -7 & 8\\ 6 & -18\end{bmatrix}=\begin{bmatrix} 8 & 7\\ 56 & -8\end{bmatrix},\\
A_{12}B_{11}+A_{22}B_{21} &= \begin{bmatrix} 37 & 16\end{bmatrix}+\begin{bmatrix} 8 & -8\end{bmatrix}=\begin{bmatrix} 45 & 8\end{bmatrix},\\
A_{12}B_{12}+A_{22}B_{22} &= \begin{bmatrix} 68 & 26\end{bmatrix}+\begin{bmatrix} 23 & 8\end{bmatrix}=\begin{bmatrix} 91 & 34\end{bmatrix}.
\end{aligned}$$

Por consiguiente,

$$A.B=\begin{bmatrix} 33 & 9 & \vdots & 8 & 7\\ 32 & 48 & \vdots & 56 & -8\\ \cdots & \cdots & . & \cdots & \cdots\\ 45 & 8 & \vdots & 91 & 34\end{bmatrix}.$$

1.5.2. Determinantes de matrices particionadas

En esta sección se muestran algunos resultados para encontrar el determinante de una matriz cuadrada particionada en bloques.

Teorema 1.18. Sea A una matriz real de tamaño $n \times n$ particionada como

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Si A_{11} y A_{22} son submatrices cuadradas de tamaño $k \times k$ y $(n-k) \times (n-k)$ respectivamente; y $A_{12} = O$ ó $A_{21} = O$, entonces

$$\det(A) = \det(A_{11}) \det(A_{22}).$$

Demostración

Probaremos este resultado por inducción sobre k. Sin pérdida de generalidad, supongamos que $A_{12} = O$ y asumamos que el teorema es válido Para todass las matrices de tamaño $(n-1) \times (n-1)$ de la forma apropiada. Dado que el determinante de una matriz A de tamaño $n \times n$ se puede calcular mediante la Expansión de Laplace por la primera fila, como sigue

$$\begin{aligned} \det(A) = & \sum_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j}(A) \\ = & a_{11} M_{11}(A) - a_{12} M_{12}(A) + \ldots + (-1)^{n+1} a_{1n} M_{1n}(A), \end{aligned} \tag{1.11}$$

donde cada uno de los menores complementarios $M_{1j}(A)$ son de la forma

$$M_{1j}(A) = \begin{vmatrix} \left[A_{11}^{[j]}\right] & O \\ [A_{21}]_j & A_{22} \end{vmatrix}, \qquad j = 1, 2, \ldots, k.$$

Aquí, $\left[A_{11}^{[j]}\right]$ se consigue borrando de A_{11} la primera fila y la j-ésima columna y $[A_{21}]_j$ se obtiene de suprimir de A_{21} la j-ésima columna. Por inducción sobre k,

$$M_{1j}(A) = \det\left(\left[A_{11}^{[j]}\right]\right) \det(A_{22}).$$

Si se reemplaza en (1.11) se tiene que

$$\begin{aligned} \det(A) = & \left[\sum_{j=1}^{k} (-1)^{1+j} a_{1j} \det\left(\left[A_{11}^{[j]}\right]\right)\right] \det(A_{22}) \\ = & \det(A_{11}) \det(A_{22}). \end{aligned}$$

Como se deseaba.

Teorema 1.19. Determinante de una matriz particionada

Sea la matriz cuadrada A tal que

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

con A_{11} y A_{22} submatrices cuadradas. Entonces

1. Si A_{11} es no singular se verifica que

$$\det(A) = \det(A_{11})\det(A_{22} - A_{21}A_{11}^{-1}A_{12}). \tag{1.12}$$

2. Si A_{22} es no singular se cumple que

$$\det(A) = \det(A_{22})\det(A_{11} - A_{12}A_{22}^{-1}A_{21}). \tag{1.13}$$

Demostración

1. Consideremos que la submatriz A_{11} es no singular, luego la matriz A se puede particionar como sigue

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A} \underset{=}{=} \underbrace{\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix}}_{U}.$$

   Fácilmente el lector puede probar que $A = LU$; por otra parte, dado que $\det(LU) = \det(L)\det(U)$ para matrices cuadradas L y U, por el Teorema 1.18 se tiene que

$$\det(A) = \det(LU) = \det(A_{11})\det(A_{22} - A_{21}A_{11}^{-1}A_{12}).$$

2. Queda como ejercicio para el lector.

Ejemplo 1.6. Para una partición $(2+2)\times(2+2)$ de la matriz A, obtenga el determinante de

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

Solución

Luego la matriz A particionada queda

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

donde $A_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ es no singular ya que su determinante es 1. Además

$$A_{21}A_{11}^{-1}A_{12} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ -9 & -5 \end{bmatrix}$$

y

$$A_{22} - A_{21}A_{11}^{-1}A_{12} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} - \begin{bmatrix} 9 & 5 \\ -9 & -5 \end{bmatrix} = \begin{bmatrix} -8 & -4 \\ 11 & 4 \end{bmatrix}$$

con determinante igual a 12, luego por (1.12) se tiene que

$$\det(A) = 1 \cdot 12 = 12.$$

Teorema 1.20. Desigualdad de Fischer

Sea A una matriz cuadrada y definida positiva, particionada como

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

donde A_{11} y A_{22} son submatrices cuadradas, entonces

$$\det A \leq \det(A_{11}) \det(A_{22}).$$

Demostración

Queda como ejercicio para el lector.

1.5.3. Inversas de matrices particionadas

Para determinar la inversa de una matriz cuadrada usualmente se emplea el método de Gauss-Jordan ó el método del determinante y la matriz adjunta. En esta sección se ilustra la manera en que se pueden calcular las inversas de las matrices usando particiones.

Teorema 1.21. Sea A una matriz no singular particionada como

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{1.14}$$

donde A_{11} y A_{22} son submatrices cuadradas no singulares. Entonces,

1. Si $A_{12} = 0$, la inversa de A es
$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix}.$$

2. Si $A_{21} = 0$, la inversa de la matriz dada en (1.14) es
$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

Demostración

Puesto que A^{-1} es la inversa de A, entonces $AA^{-1} = I$.

1. Supóngase que A_{12} es una submatriz nula, entonces
$$\begin{aligned} AA^{-1} &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}A_{11}^{-1} & 0 \\ A_{21}A_{11}^{-1} - A_{22}A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \end{aligned}$$

2. De manera análoga, cuando $A_{21} = 0$ se tiene que
$$\begin{aligned} AA^{-1} &= \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}A_{11}^{-1} & A_{12}A_{22}^{-1} - A_{11}A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \end{aligned}$$

Teorema 1.22. Inversa de una matriz particionada

Sea A una matriz no singular particionada como sigue:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

donde A_{ij} son submatrices de tamaño $n_i \times n_j$ para $i,\ j = 1,2$, donde $n_1 + n_2 = n$ y $0 < n_1 < n$. Denotemos A^{-1} por B y particionemos B como

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

donde B_{ij} son submatrices de tamaño $n_i \times n_j$ para $i,\ j = 1,2$. Si $\det(A_{11}) \neq 0$, y $\det(A_{22}) \neq 0$, se tienen los siguientes resultados:

1. B_{11}^{-1} y B_{22}^{-1} existen;

2. $\left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1}$ y $\left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1}$ existen;

3. A^{-1} puede escribirse como

$$a)\ A^{-1} = \begin{bmatrix} \left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1} & -A_{11}^{-1}A_{12}\left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1} \\ -A_{22}^{-1}A_{21}\left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1} & \left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1} \end{bmatrix}.$$

$$b)\ A^{-1} = \begin{bmatrix} \left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1} & -\left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1} \\ -\left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1} & \left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1} \end{bmatrix}.$$

Demostración

Para probar 2. multiplicamos la matriz A por la izquierda por la matriz no singular

$$A_1^* = \begin{bmatrix} A_{11}^{-1} & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}.$$

Fácilmente el lector puede probar que $A_1^*A = C$, es

$$C = \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix},$$

pero por el Teorema 1.18 se tiene que

$$\det(C) = \det(I)\det(A_{22} - A_{21}A_{11}^{-1}A_{12}) = \det(A_1^*A) = \det(A_1^*)\det(A) \neq 0.$$

En consecuencia,

$$A_{22} - A_{21}A_{11}^{-1}A_{12},$$

es no singular.

Para probar la otra parte de 2., se multiplica B por la izquierda por A_2^*, donde

$$A_2^* = \begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

Por otra parte, del resultado $BA = I$ se tiene que

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

y de estas se obtienen las siguientes cuatro ecuaciones matriciales:

$$B_{11}A_{11} + B_{12}A_{21} = I \tag{1.15a}$$
$$B_{11}A_{12} + B_{12}A_{22} = 0 \tag{1.15b}$$
$$B_{21}A_{11} + B_{22}A_{21} = 0 \tag{1.15c}$$
$$B_{21}A_{12} + B_{22}A_{22} = I. \tag{1.15d}$$

Si se multiplica por la derecha de la ecuación (1.15b) por A_{22}^{-1} se tiene

$$B_{12} = -B_{11}A_{12}A_{22}^{-1}.$$

Si se reemplaza en la ecuación (1.15a) y se factoriza se obtiene

$$B_{11}\left[A_{11} - A_{12}A_{22}^{-1}A_{21}\right] = I, \tag{1.16}$$

así B_{11}^{-1} existe y es igual a $A_{11} - A_{12}A_{22}^{-1}A_{21}$.

De manera análoga, si se utilizan las ecuaciones (1.15c) y (1.15d) se puede probar que

$$B_{22}\left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right] = I \tag{1.17}$$

es decir, B_{22}^{-1} existe y es igual a $A_{22} - A_{21}A_{11}^{-1}A_{12}$. Esto prueba 1.

De reemplazar en B_{12} y B_{21} se sigue la prueba de 3.

Corolario 1.22.1. Sean A y B como en el Teorema 1.22, entonces

- $\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1}$.
- $\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}=A_{22}^{-1}+A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}$.
- $A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}=\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}$.
- $A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}=\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1}$.

Demostración

Como A_{11} es invertible, la matriz A se puede particionar como sigue

$$\underbrace{\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}}_{A}\underset{=}{=}\underbrace{\begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}-A_{21}A_{11}^{-1}A_{12}\end{bmatrix}}_{L}\underbrace{\begin{bmatrix}I & A_{11}^{-1}A_{12}\\ 0 & I\end{bmatrix}}_{U}.$$

Luego, $A^{-1}=B=U^{-1}L^{-1}$ y por el Teorema 1.21 se tiene

$$A^{-1}=\begin{bmatrix}I & -A_{11}^{-1}A_{12}\\ 0 & I\end{bmatrix}\begin{bmatrix}A_{11}^{-1} & 0\\ -\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1} & \left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}\end{bmatrix}$$
$$=\begin{bmatrix}A_{11}^{-1}+A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}\\ -\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1} & \left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}\end{bmatrix}.$$

Como en el Teorema 1.22 se obtuvo A^{-1}, comparando los términos se obtiene

$$\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1}$$
$$A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}=\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}A_{21}A_{11}^{-1}.$$

Por otra parte, dado que A_{22} es no singular, la matriz A también podría particionarse de la siguiente manera

$$\underbrace{\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}}_{A}\underset{=}{=}\underbrace{\begin{bmatrix}A_{11}-A_{12}A_{22}^{-1}A_{21} & A_{12}\\ 0 & A_{22}\end{bmatrix}}_{R}\underbrace{\begin{bmatrix}I & 0\\ A_{22}^{-1}A_{21} & I\end{bmatrix}}_{S},$$

por lo tanto, $A^{-1}=B=S^{-1}R^{-1}$ y en virtud del Teorema 1.21 se tiene que

$$A^{-1}=\begin{bmatrix} I & 0\\ -A_{22}^{-1}A_{21} & I\end{bmatrix}\begin{bmatrix}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1} & -\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}\\ 0 & A_{22}^{-1}\end{bmatrix}$$
$$=\begin{bmatrix}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1} & -\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}\\ -A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}\end{bmatrix}.$$

Aquí se comparan de nuevo los términos con los de la matriz A^{-1}, para obtener

$$\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}=A_{22}^{-1}+A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}$$
$$A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}=\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}A_{12}A_{22}^{-1}$$

y el corolario queda probado.

Ejemplo 1.7.

Obtenga la inversa de la matriz dada en el Ejemplo 1.6.

Solución

En el Ejemplo 1.6 se considero la partición

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\qquad \text{obteniéndose que}\qquad \det(A)=12,$$

es decir, A es invertible. Asumiendo la misma partición se tiene que

$$A_{11}^{-1}=\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}^{-1}=\begin{bmatrix}1 & -1\\ -1 & 2\end{bmatrix},\qquad A_{22}^{-1}=\begin{bmatrix}1 & 1\\ 2 & -1\end{bmatrix}^{-1}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3}\\ \frac{2}{3} & -\frac{1}{3}\end{bmatrix},$$
$$A_{12}A_{22}^{-1}A_{21}=\begin{bmatrix}3 & 1\\ -1 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 2 & -1\end{bmatrix}^{-1}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}.$$

Aunque $A_{12}A_{22}^{-1}A_{21}$ no es invertible, el primer bloque B_{11} de la matriz A^{-1} es

$$\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}=\begin{bmatrix}1 & 2\\ 2 & 0\end{bmatrix}^{-1}=\begin{bmatrix}0 & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{4}\end{bmatrix}.$$

La submatriz $B_{21}=-A_{22}^{-1}A_{21}\left[A_{11}-A_{12}A_{22}^{-1}A_{21}\right]^{-1}$, esta dada por

$$B_{21}=-\begin{bmatrix}1 & 1\\ 2 & -1\end{bmatrix}^{-1}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}0 & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{4}\end{bmatrix}=\begin{bmatrix}0 & 0\\ \frac{1}{2} & -\frac{3}{4}\end{bmatrix}\qquad \text{y}$$
$$A_{21}A_{11}^{-1}A_{12}=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}^{-1}\begin{bmatrix}3 & 1\\ -1 & -1\end{bmatrix}=\begin{bmatrix}9 & 5\\ -9 & -5\end{bmatrix}.$$

A pesar de que $A_{21}A_{11}^{-1}A_{12}$ no es invertible, el bloque B_{22} se obtiene de

$$\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}=\begin{bmatrix}-8 & -4\\ 11 & 4\end{bmatrix}^{-1}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3}\\ -\frac{11}{12} & -\frac{2}{3}\end{bmatrix}$$

y finalmente, se tiene que $B_{12}=-A_{11}^{-1}A_{12}\left[A_{22}-A_{21}A_{11}^{-1}A_{12}\right]^{-1}$ es

$$B_{12}=-\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}^{-1}\begin{bmatrix}3 & 1\\ -1 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{3} & \frac{1}{3}\\ -\frac{11}{12} & -\frac{2}{3}\end{bmatrix}=\begin{bmatrix}\frac{1}{2} & 0\\ -\frac{13}{12} & -\frac{1}{3}\end{bmatrix}.$$

Por lo tanto,

$$\begin{bmatrix}2 & 1 & 3 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & 1\\ -1 & 1 & 2 & -1\end{bmatrix}^{-1}=\begin{bmatrix}0 & \frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{2} & -\frac{1}{4} & -\frac{13}{12} & -\frac{1}{3}\\ 0 & 0 & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & -\frac{3}{4} & -\frac{11}{12} & -\frac{2}{3}\end{bmatrix}.$$

Ejercicios 1.1.

1. Utilizando particiones encuentre el determinante y la inversa de

$$\begin{bmatrix}-5 & 3 & 5 & 1 & 0\\ -4 & 8 & 10 & 0 & -1\\ 2 & 13 & 11 & 2 & 1\\ 0 & 1 & 1 & 3 & 2\\ 3 & 1 & 0 & 7 & 5\end{bmatrix}.$$

2. Si A y C son no singulares. Pruebe que

   i) $\det(I+AB)=\det(I+BA)$.

   ii) $\det(A+CBC^t)=\det(A)\det(I+BC^tA^{-1}C)$.

3. Demuestre que

$$\begin{bmatrix}P & Q\\ R & S\end{bmatrix}\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X & Y\\ Z & W\end{bmatrix}=\begin{bmatrix}P\\ R\end{bmatrix}\begin{bmatrix}X & Y\end{bmatrix}.$$

4. Muestre que la inversa de una matriz no singular particionada es

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} [D - CA^{-1}B]^{-1} \begin{bmatrix} -CA^{-1} & I \end{bmatrix}.$$

1.6. Espacio vectorial

Los conjuntos $\mathbb{R}^2$ (vectores en el plano) y $\mathbb{R}^3$ (vectores en el espacio) junto con las operaciones de suma de vectores y multiplicación por un escalar se llaman espacios vectoriales. Las propiedades algebraicas de un *espacio vectorial* arbitrario son muy semejantes a las de los elementos de $\mathbb{R}^2$ y $\mathbb{R}^3$. En consecuencia, se acostumbra llamar *vectores* también a los elementos de un espacio vectorial arbitrario.

Definición 1.33. Un espacio vectorial real[2] $\mathbb{V}$ es un conjunto no vacío de vectores, dotado de dos operaciones

Suma	**Multiplicación por un escalar**
$\mathbb{V} \times \mathbb{V} \to \mathbb{V}$	$\mathbb{R} \times \mathbb{V} \to \mathbb{V}$
$(\vec{x}, \vec{y}) \to \vec{x} + \vec{y}$	$(\alpha, \vec{x}) \to \alpha\vec{x}$

que satisfacen los diez axiomas enumerados a continuación.

Axiomas de un espacio vectorial

Dado $\mathbb{V}$ un espacio vectorial real, se verifica que:

- Para la suma en $\mathbb{V}$:

  i. *Clausurativa:* Si $\vec{x}, \vec{y} \in \mathbb{V}$, entonces $\vec{x} + \vec{y} \in \mathbb{V}$.

  ii. *Asociativa:* Para todo $\vec{x}, \vec{y}$ y $\vec{z} \in \mathbb{V}$, $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$.

  iii. *Conmutativa:* Si $\vec{x}, \vec{y} \in \mathbb{V}$, entonces $\vec{x} + \vec{y} = \vec{y} + \vec{x}$.

[2] La palabra “real” significa que los escalares que se usan son números reales.

iv. *Existencia de elemento neutro:* Existe un vector de $\mathbb{V}$, denotado por $\vec{0}$, tal que para todo $\vec{x} \in \mathbb{V}$, $\vec{x} + \vec{0} = \vec{0} + \vec{x} = \vec{x}$.

v. *Elemento opuesto:* Si $\vec{x} \in \mathbb{V}$, existe un vector $-\vec{x}$ en $\mathbb{V}$ tal que $\vec{x} + (-\vec{x}) = \vec{0}$.

- Para el producto por un escalar de $\mathbb{R}$:

  vi. *Clausurativa:* Si $\vec{x} \in \mathbb{V}$ y α es un escalar, entonces $\alpha\vec{x} \in \mathbb{V}$.

  vii. *Distributiva respecto de la suma de vectores:* Si $\vec{x}, \vec{y} \in \mathbb{V}$ y α es un escalar, entonces $\alpha(\vec{x} + \vec{y}) = \alpha\vec{x} + \alpha\vec{y}$.

  viii. *Distributiva respecto de la suma de escalares:* Si $\vec{x} \in \mathbb{V}$ y α y β son escalares, entonces $(\alpha + \beta)\vec{x} = \alpha\vec{x} + \beta\vec{x}$.

  ix. *Asociativa respecto a la multiplicación de escalares:* Si $\vec{x} \in \mathbb{V}$ y α y β son escalares, entonces $\alpha(\beta\vec{x}) = (\alpha\beta)\vec{x}$.

  x. *Existencia del elemento unidad:* Para cada vector $\vec{x} \in \mathbb{V}$, $1\vec{x} = \vec{x}$.

Definición 1.34. Sea $\mathbb{V}$ un espacio vectorial y $\mathbb{W}$ un subconjunto no vacío de $\mathbb{V}$. Se dice que $\mathbb{W}$ es un subespacio vectorial de $\mathbb{V}$ si $\mathbb{W}$ dotado de las mismas operaciones definidas en $\mathbb{V}$ es, a su vez, espacio vectorial.

Teorema 1.23. Un subconjunto no vacío $\mathbb{W}$ de un espacio vectorial $\mathbb{V}$ es un subespacio vectorial de $\mathbb{V}$ si cumple que

i) La suma de elementos de $\mathbb{W}$ es un elemento de $\mathbb{W}$.

ii) El producto de un escalar por un elemento de $\mathbb{W}$ pertenece a $\mathbb{W}$.

Una condición equivalente para que $\mathbb{W}$ sea subespacio vectorial es que para todo par de elementos $\vec{v}$ y $\vec{w}$ de $\mathbb{W}$ y cualesquiera α y β de $\mathbb{R}$, se verifique $\alpha\vec{v} + \beta\vec{w}$ pertenece a $\mathbb{W}$.

Demostración

Queda como ejercicio para el lector.

Definición 1.35. Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios de un espacio vectorial real $\mathbb{V}$, entonces se define la suma $\mathbb{U}+\mathbb{W}$ como

$$\mathbb{U}+\mathbb{W}=\{\vec{u}+\vec{w}\mid \vec{u}\in\mathbb{U},\vec{w}\in\mathbb{W}\}.$$

Teorema 1.24. Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios de un espacio vectorial real $\mathbb{V}$, entonces la suma $\mathbb{U}+\mathbb{W}$ es un subespacio de $\mathbb{V}$.

Demostración

Se debe probar que $\mathbb{U}+\mathbb{W}$ satisface las condiciones del teorema 1.23:

i) Si $\vec{u}_1,\vec{u}_2\in\mathbb{U}$ y $\vec{w}_1,\vec{w}_2\in\mathbb{W}$, entonces

$$(\vec{u}_1+\vec{w}_1)+(\vec{u}_2+\vec{w}_2)=\vec{u}_1+\vec{u}_2+\vec{w}_1+\vec{w}_2\in\mathbb{U}+\mathbb{W}.$$

ii) Si $\alpha\in\mathbb{R}$, entonces

$$\alpha(\vec{u}_1+\vec{w}_1)=\alpha\vec{u}_1+\alpha\vec{w}_1\in\mathbb{U}+\mathbb{W}.$$

Finalmente, $\vec{0}+\vec{0}\in\mathbb{U}+\mathbb{W}$. Lo cual prueba que $\mathbb{U}+\mathbb{W}$ es un subespacio.

Definición 1.36. Se dice que $\mathbb{V}$ es una **suma directa** de $\mathbb{U}$ y $\mathbb{W}$ si todo $\vec{v}\in\mathbb{V}$ tiene una representación única de la forma

$$\vec{v}=\vec{u}+\vec{w}.$$

con $\vec{u}\in\mathbb{U}$ y $\vec{v}\in\mathbb{V}$. Esta suma directa se denotará como $\mathbb{V}=\mathbb{U}\bigoplus\mathbb{W}$.

Teorema 1.25. Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios no nulos de un espacio vectorial real $\mathbb{V}$, su suma $\mathbb{U}+\mathbb{W}$ es una suma directa si y sólo si $\mathbb{U}\bigcap\mathbb{W}=\{\vec{0}\}$.

Demostración

Queda como ejercicio para el lector.

Definición 1.37. Combinación lineal

Sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ vectores en un espacio vectorial real $\mathbb{V}$. Un vector $\vec{v}$ en $\mathbb{V}$ es una combinación lineal de $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ si

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$$

para ciertos números reales $c_1, c_2, \ldots, c_n$.

Definición 1.38. Sea $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ un conjunto de vectores en un espacio vectorial $\mathbb{V}$, entonces se dice que S es

1. Linealmente dependiente o ligado si y sólo si existen escalares $c_1, c_2, \ldots, c_n$ no todos nulos, tales que

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n = \vec{0}.$$

2. Linealmente independiente o libre si y sólo si no son ligados. Esto es,

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n = \vec{0}.$$

se cumple sólo para $c_1 = c_2 = \ldots = c_n = 0$

Definición 1.39. Espacio generado

Si $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es un conjunto de vectores en un espacio vectorial $\mathbb{V}$, entonces el conjunto de todos los vectores en $\mathbb{V}$ que son combinaciones lineales de los vectores en S se denomina espacio generado y se denota por

$$gen S \qquad \text{o} \qquad gen\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}.$$

1.6.1. Bases

En esta sección se continúa con el estudio de la estructura de un espacio vectorial $\mathbb{V}$ determinando un conjunto mínimo de vectores de $\mathbb{V}$ que describa completamente a $\mathbb{V}$.

Definición 1.40. Base

Si $\mathbb{V}$ es cualquier espacio vectorial y $\mathfrak{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ es un conjunto finito de vectores en $\mathbb{V}$, entonces $\mathfrak{B}$ se denomina base para $\mathbb{V}$ si es un conjunto generador para $\mathbb{V}$ con el número más pequeño de elementos en un conjunto generador para $\mathbb{V}$.

El teorema principal acerca de las bases es:

Teorema 1.26. Base

Sea $\mathfrak{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ un conjunto de vectores en un espacio vectorial $\mathbb{V}$. El conjunto $\mathfrak{B}$ es una base para $\mathbb{V}$ si y sólo si $\mathfrak{B}$ es linealmente independiente y genera a $\mathbb{V}$.

Demostración

Queda como ejercicio para el lector.

Definición 1.41. Dimensión

Si un espacio vectorial $\mathbb{V}$ tiene una base $\mathfrak{B}$ con n elementos $(n \in \mathbb{N})$, entonces se define a n como la dimensión del espacio vectorial $\mathbb{V}$ y se escribe

$$n = \dim \mathbb{V}.$$

Si $\mathbb{V} = \left\{\vec{0}\right\}$, entonces se tiene que $\dim \mathbb{V} = 0$.

Teorema 1.27. Sea $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ un conjunto de vectores en un espacio vectorial $\mathbb{V}$ de dimensión n. Sea

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix},$$

entonces S es un conjunto de vectores linealmente independiente si y sólo si

$$\det A \neq 0.$$

Demostración

Queda como ejercicio para el lector.

Teorema 1.28. Suponga que $\dim \mathbb{V} = n$. Si $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ es un conjunto de m vectores linealmente independientes en $\mathbb{V}$, entonces $m \leq n$.

Demostración

Queda como ejercicio para el lector.

Teorema 1.29. Cualesquiera n vectores linealmente independientes en un espacio vectorial $\mathbb{V}$ de dimensión n constituyen una base para $\mathbb{V}$.

Demostración

Queda como ejercicio para el lector.

Definición 1.42. Sea $\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base para el espacio vectorial $\mathbb{V}$ de dimensión n. Las coordenadas de cualquier vector $\vec{x} \in \mathbb{V}$, en $\mathfrak{B}$, se relacionan por

$$(\vec{x})_{\mathfrak{B}} = M_{\mathfrak{B}}^{-1} \vec{x}, \tag{1.18}$$

donde $M_{\mathfrak{B}} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}$.

Ejemplo 1.8. Muestre que los vectores $\vec{v}_1^{\,t} = (2, -1)$ y $\vec{v}_2^{\,t} = (1, 5)$ forman una base de $\mathbb{R}^2$ y halle las componentes del vector $\vec{x}^{\,t} = (7, 4)$ con relación a esta base.

Solución

Fórmese $A = [\vec{v}_1 \; \vec{v}_2]$ y calcúlese su determinante

$$\det A = \begin{vmatrix} \vec{v}_1 & \vec{v}_2 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ -1 & 5 \end{vmatrix} = 11 \neq 0.$$

Luego, $S = \{\vec{v}_1, \vec{v}_2\}$ es un conjunto de vectores linealmente independientes y como $\mathbb{R}^2$ tiene dimensión dos, se deduce que forman una base.

Para hallar las componentes de $\vec{x}$ en términos de esta base, se hace

$$\begin{aligned}(\vec{x})_{\mathfrak{B}} =& M_{\mathfrak{B}}^{-1}\vec{x} = A^{-1}\vec{x}\\ =& \begin{bmatrix} 2 & 1\\ -1 & 5\end{bmatrix}^{-1}\begin{bmatrix}7\\4\end{bmatrix} = \frac{1}{11}\begin{bmatrix}5 & -1\\ 1 & 2\end{bmatrix}\begin{bmatrix}7\\4\end{bmatrix} = \frac{1}{11}\begin{bmatrix}31\\15\end{bmatrix}.\end{aligned}$$

Estas son las componentes de $\vec{x}$ relativas a la base $\vec{v}_1, \vec{v}_2$.

Teorema 1.30. Cambio de Base

Sean $\mathfrak{B}_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ y $\mathfrak{B}_2 = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ bases para el espacio vectorial $\mathbb{V}$ de dimensión n. Para cualquier vector $\vec{x} \in \mathbb{V}$, las coordenadas en $\mathfrak{B}_1$, $(\vec{x})_{\mathfrak{B}_1}$ y las coordenadas en $\mathfrak{B}_2$, $(\vec{x})_{\mathfrak{B}_2}$ se relacionan por

$$(\vec{x})_{\mathfrak{B}_2} = M_{\mathfrak{B}_2}^{-1} M_{\mathfrak{B}_1} (\vec{x})_{\mathfrak{B}_1}. \tag{1.19}$$

Demostración

Queda como ejercicio para el lector.

1.6.2. Espacios con producto interno

En esta sección se define una operación que no puede ser clasificada como externa o interna, ya que aunque se opera con los elementos de un espacio vectorial, el resultado es un escalar el cual no pertenece al conjunto sobre el cual se define la operación.

Definición 1.43. Espacio con producto interno

Un espacio vectorial real $\mathbb{V}$ de dimensión finita, se dice que es un espacio con producto interno si a cada par de vectores $\vec{u}, \vec{v} \in \mathbb{V}$, le asigna un número real denotado por $< \vec{u}, \vec{v} >$, tal que

$$< \vec{u}, \vec{v} >= \vec{u} \cdot \vec{v} = \vec{u}^t \vec{v} = \sum_{i=1}^{n} u_i v_i. \tag{1.20}$$

Debido a la notación en (1.20), el producto interno se llama con frecuencia producto escalar o producto punto entre vectores.

Teorema 1.31. Propiedades del producto interno

Sea $\mathbb{V}$ un espacio vectorial real con un producto interno $<,>$. Entonces, para todo $\vec{u},\vec{v},\vec{w}\in\mathbb{V}$ y todo $\alpha\in\mathbb{R}$ se tiene

i) $<\vec{u},\vec{v}>=<\vec{v},\vec{u}>$.

ii) $<(\vec{u}+\vec{v}),\vec{w}>=<\vec{u},\vec{w}>+<\vec{v},\vec{w}>$.

iii) $<\vec{u},(\vec{v}+\vec{w})>=<\vec{u},\vec{v}>+<\vec{u},\vec{w}>$.

iv) $<(\alpha\vec{u}),\vec{v}>=\alpha<\vec{u},\vec{v}>=<\vec{u},(\alpha\vec{v})>$.

v) $<\vec{u},\vec{u}>\geq 0$ y $<\vec{u},\vec{u}>=0$ si y sólo si $\vec{u}=\vec{0}$.

Definición 1.44. Longitud o norma

Sea $\mathbb{V}$ un espacio vectorial real, con un producto interno $<,>$. Una norma en $\mathbb{V}$ es una función de $\mathbb{V}$ en $\mathbb{R}$, tal que a cada $\vec{v}\in\mathbb{V}$, le asigna un número real no negativo, denotado por $\|\vec{v}\|$ y definido como

$$\|\vec{v}\|=\sqrt{<\vec{v},\vec{v}>}. \tag{1.21}$$

Teorema 1.32. Propiedades de la norma

Para todo $\vec{u},\vec{v}\in\mathbb{V}$ y todo $\alpha\in\mathbb{R}$

i) $\|\vec{u}\|\geq 0$.

ii) $\|\vec{u}\|=0$ si y sólo si $\vec{u}=\vec{0}$.

iii) $\|\alpha\vec{u}\|=|\alpha|\|\vec{u}\|$.

iv) $\|<\vec{u},\vec{v}>\|\leq\|\vec{u}\|\|\vec{v}\|$.

$v)$ $\|\vec{u}+\vec{v}\| \leq \|\vec{u}\|+\|\vec{v}\|$.

Demostración
Queda como ejercicio para el lector.

Teorema 1.33. Sea $<,>$ un producto interno en un espacio vectorial $\mathbb{V}$ de dimensión finita. Sean $\vec{u}$ y $\vec{v}$ dos vectores diferentes de cero. Si θ es el ángulo entre ellos, entonces

$$\cos\theta = \frac{<\vec{u},\vec{v}>}{\|\vec{u}\|\,\|\vec{v}\|}. \tag{1.22}$$

Demostración
Queda como ejercicio para el lector.

Definición 1.45. Vectores ortogonales

1. Sea $\mathbb{V}$ un espacio vectorial con un producto interno y sean dos vectores $\vec{u},\vec{v} \in \mathbb{V}$. Se dice que $\vec{u}$ y $\vec{v}$ son ortogonales $(\vec{u} \perp \vec{v})$ si y sólo si

$$<\vec{u},\vec{v}>=0.$$

2. Un conjunto de vectores $S=\{\vec{v}_1,\vec{v}_2,\ldots,\vec{v}_m\}$ de un espacio vectorial $\mathbb{V}$ se dice que es ortogonal si y sólo si los vectores son ortogonales dos a dos, es decir,

$$<\vec{v}_i,\vec{v}_j>=0 \qquad \text{siempre que} \qquad i \neq j.$$

3. El conjunto de vectores S (en 2), se dice que es ortonormal si y sólo si

   $a)$ S es ortogonal. $\qquad$ $b)$ $\|\vec{v}_i\|=1,$ para todo i.

Teorema 1.34. Todo conjunto ortogonal de un espacio vectorial $\mathbb{V}$ es linealmente independiente.

Demostración
Queda como ejercicio para el lector.

Teorema 1.35. Proceso de ortonormalización de Gram-Schmidt

Todo subespacio H de dimensión k de $\mathbb{R}^n$ tiene al menos una base ortogonal y una base ortonormal. Si $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ es cualquier base de H, entonces $\mathcal{B}' = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_k\}$ es una base ortogonal, donde

$$
\begin{aligned}
\vec{w}_1 &= \vec{v}_1. \\
\vec{w}_2 &= \vec{v}_2 - \frac{<\vec{v}_2, \vec{w}_1>}{\|\vec{w}_1\|^2}\vec{w}_1. \\
\vec{w}_3 &= \vec{v}_3 - \frac{<\vec{v}_3, \vec{w}_1>}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{<\vec{v}_3, \vec{w}_2>}{\|\vec{w}_2\|^2}\vec{w}_2. \\
&\vdots \\
\vec{w}_k &= \vec{v}_k - \sum_{i=1}^{k-1} \frac{<\vec{v}_k, \vec{w}_i>}{\|\vec{w}_i\|^2}\vec{w}_i.
\end{aligned}
$$

y

$$
gen\{\vec{v}_1, \dots, \vec{v}_k\} = gen\{\vec{w}_1, \dots, \vec{w}_k\}, \qquad i = 1, \dots, k.
$$

La base ortonormal $\mathcal{B}''$ se obtiene normalizando $\mathcal{B}'$:

$$
\mathcal{B}'' = \left\{ \frac{\vec{w}_1}{\|\vec{w}_1\|}, \dots, \frac{\vec{w}_k}{\|\vec{w}_k\|} \right\}.
$$

Demostración
Queda como ejercicio para el lector.

1.6.3. Complemento ortogonal

Consideremos un subespacio $\mathbb{V} \subseteq \mathbb{R}^n$. Para $\mathbb{V}$ puede haber muchos subespacios de $\mathbb{R}^n$ que son ortogonales a $\mathbb{V}$ (por ejemplo, en $\mathbb{R}^3$, si $\mathbb{V}$ es el eje Z, todas las rectas en el plano XY que pasan por el origen y el propio plano XY, son ortogonales a $\mathbb{V}$). Entre todos los subespacios que son ortogonales a $\mathbb{V}$ hay uno de particular importancia: aquel subespacio $\mathbb{V}^*$ tal que $\mathbb{R}^n = \mathbb{V} \oplus \mathbb{V}^*$.

Definición 1.46. Complemento Ortogonal

Sea H un subespacio del espacio con producto interno $\mathbb{V}$. Entonces el complemento ortogonal de H, denotado por $H^{\perp}$, está dado por

$$H^{\perp} = \{\vec{u} \in \mathbb{V} : \left\langle \vec{u}, \vec{h} \right\rangle = 0 \text{ Para toda } \vec{h} \in H\}.$$

1.6.4. Subespacios asociados a una matriz

Hay cuatro subespacios asociados a una matriz los cuales se consideran a continuación.

Definición 1.47. Espacio Nulo y Nulidad de una matriz

Sea A una matriz de tamaño $m \times n$. Entonces el conjunto

$$N_A = \{\vec{u} \in \mathbb{R}^n | A\vec{u} = \vec{0}\}, \tag{1.23}$$

se llama el espacio nulo de A y $\nu(A) = \dim N_A$ se denomina nulidad de A.

El espacio nulo de una matriz también se conoce como *núcleo*.

Definición 1.48. Imagen de una matriz

Sea A una matriz de tamaño $m \times n$. Entonces la imagen de A, denotada por Im_A, está dada por

$$Im_A = \{\vec{v} \in \mathbb{R}^m | A\vec{u} = \vec{v} \text{ para algún } \vec{u} \in \mathbb{R}^n\}. \tag{1.24}$$

Definición 1.49. Espacio de los renglones (o de las filas) de una matriz

Sea A una matriz de tamaño $m \times n$, sean $\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n\}$ los renglones (o filas) de A. Entonces se define

$$R(A) = \text{espacio de los renglones de } A = gen\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n\}. \tag{1.25}$$

Definición 1.50. Espacio de las columnas de una matriz

Sea A una matriz de tamaño $m \times n$, sean $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ las columnas de A. Entonces se define

$$C(A) = \text{espacio de las columnas de } A = gen\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}. \tag{1.26}$$

Definición 1.51. Rango de una matriz

Sea $A = [a_{ij}]$ una matriz real de tamaño $m \times n$, se llama rango de A, denotado por $\rho(A)$, al número máximo de vectores fila (o columna) linealmente independiente, o también, a la dimensión del subespacio generado por los vectores fila (o columna) de A.

Teorema 1.36. Propiedades del rango

Sea A una matriz real de tamaño $m \times n$, entonces se cumple que

(i) $0 \leq \rho(A) \leq \text{mín}\{m, n\}$.

(ii) Si $\rho(A) = m < n$, se dice que A tiene *rango completo fila*.

(iii) Si $\rho(A) = n < m$, se dice que A tiene *rango completo columna*.

(iv) Si $\rho(A) = r \leq \text{mín}\{m, n\}$, entonces existen matrices K y L de rango r y tamaños $m \times r$ y $r \times n$, respectivamente, tales que $A = KL$.

(v) $\rho(I_n) = n$, con I_n la matriz identidad de tamaño $n \times n$.

(vi) $\rho(O) = 0$, siendo, O la matriz nula de tamaño $n \times n$.

(vii) $\rho(A) = \rho(A^t)$.

$(viii)$ $\rho(A.B) \leq \text{mín}\{\rho(A), \rho(B)\}$.

(ix) Si A es diagonal, entonces $\rho(A)$, es el número de elementos no nulos en su diagonal.

(x) Si A es no singular, entonces $\rho(A.B) = \rho(B)$ y $\rho(B.A) = \rho(B)$.

Demostración

Queda como ejercicio para el lector.

1.7. Sistemas de ecuaciones lineales

En esta sección se mencionan las relaciones lineales, pues gran parte del Álgebra Lineal estudia y desarrolla relaciones lineales las cuales son una generalización de la ecuación de una recta.

Definición 1.52. Un sistema de ecuaciones se dice que es lineal si todas las ecuaciones que lo componen son lineales en los escalares desconocidos o incógnitas $x_1, x_2, \ldots, x_n$. Es decir, son de la forma

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_j x_j + \ldots + \alpha_n x_n = \beta,$$

donde α_i, $i = 1, 2, \ldots, n$ y β habitualmente son números reales, números complejos o funciones. Así pues, en una ecuación lineal no pueden aparecer productos o potencias de las incógnitas $x_1, x_2, \ldots, x_j, \ldots, x_n$.

Definición 1.53. Se llama sistema lineal de m ecuaciones con n incógnitas al conjunto de m igualdades

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad\qquad\qquad &= \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mj}x_j + \ldots + a_{mn}x_n &= b_m,
\end{aligned} \tag{1.27}$$

donde a_{ij}, $i = 1,2,\ldots,m$ $j = 1,2,\ldots,n$ son los coeficientes del sistema; b_k, $k = 1,2,\ldots,m$ son los términos independientes y $x_1, x_2, \ldots, x_j, \ldots, x_n$ son las incógnitas del sistema.

El sistema de ecuaciones (1.27) puede escribirse en forma matricial como sigue

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1j} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2j} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ldots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mj} & \ldots & a_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}}_{\vec{X}} \underset{=}{=} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_j \\ \vdots \\ b_n \end{bmatrix}}_{\vec{b}}, \tag{1.28}$$

donde A es la matriz de coeficientes de tamaño $m \times n$, $\vec{X} \in \mathbb{R}^n$ es el vector columna de las incógnitas y $\vec{b} \in \mathbb{R}^m$, es el vector columna de los términos independientes.

El sistema (1.28) se dice que es *homogéneo* cuando el vector $\vec{b}$ de términos independientes es nulo. Es decir,

$$A\vec{X} = \vec{0}.$$

Se conoce como *sistema lineal no homogéneo general* al sistema de la forma

$$A\vec{X} = \vec{b} \qquad \text{con} \qquad \vec{b} \neq \vec{0}.$$

Definición 1.54. Sistemas consistentes e inconsistentes

Un sistema de ecuaciones lineales

$$A\vec{X} = \vec{b} \tag{1.29}$$

con A una matriz de tamaño $m \times n$, $\vec{X} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, es

- *Consistente* si admite al menos una solución.

- *Inconsistente* si no tiene solución.

En caso de existir solución para el sistema 1.29, éste es

- *Consistente determinado* si la solución es única.

- *Consistente indeterminado* cuando hay infinitas. soluciones

1.7.1. Método de eliminación de Gauss

Para resolver un sistema de m ecuaciones con n incógnitas de la forma

$$A\vec{X} = \vec{b}$$

se procede de la siguiente manera

1. Se forma la matriz aumentada $B = \left(A \,|\, \vec{b}\right)$.
2. Se aplican operaciones elementales entre filas, hasta llevar a B a una matriz escalonada reducida C.
3. Se halla el sistema de ecuaciones que representa la matriz C.
4. En este último sistema, se encuentra la solución.

Por último en esta sección se presenta la aplicación de determinantes, bien conocida, para resolver sistemas de ecuaciones lineales $A\vec{X} = \vec{b}$, donde A es una matriz invertible de tamaño $n \times n$. Sean

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} \qquad \text{y} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Para $j = 1, 2, \dots, n$, denotemos por B_j la matriz que resulta de sustituir la columna $j-$ésima de A por la columna $\vec{b}$,

$$B_j = \begin{bmatrix} a_{11} & \dots & a_{1,j-1} & b_1 & a_{1,j+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,j-1} & b_2 & a_{2,j+1} & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & b_n & a_{n,j+1} & \dots & a_{nn} \end{bmatrix}. \tag{1.30}$$

Teorema 1.37. Regla de Cramer

Consideremos el sistema $A\vec{X} = \vec{b}$, donde A es invertible. Entonces, la solución del sistema viene dada por

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} \det B_1 \\ \det B_2 \\ \vdots \\ \det B_n \end{bmatrix}, \tag{1.31}$$

donde las matrices $B_1, B_2, \ldots, B_n$, están definidas en (1.30)

Demostración

Queda como ejercicio para el lector.

1.8. Transformaciones lineales

Definición 1.55. Sean $\mathbb{V}$ y $\mathbb{W}$ espacios vectoriales reales. Sea $T : \mathbb{V} \to \mathbb{W}$ una función de $\mathbb{V}$ en $\mathbb{W}$. Se dice que T es una transformación lineal de $\mathbb{V}$ en $\mathbb{W}$ si y sólo si para cualquier $\vec{u}$, $\vec{v}$ vectores de $\mathbb{V}$ y α escalar, se tiene que

$$T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v}) \qquad \text{y} \qquad T(\alpha\vec{u}) = \alpha T(\vec{u}). \tag{1.32}$$

Teorema 1.38. Sea T una transformación lineal de $\mathbb{V}$ en $\mathbb{W}$. Entonces,

1. $T\left(\vec{0}\right) = \vec{0}$.
2. $T(-\vec{u}) = -T(\vec{u})$, para todo $\vec{u} \in \mathbb{V}$.
3. Si $\vec{v} = \sum_{i=1}^{n} \alpha_i \vec{u}_i$, con $\vec{u}_i \in \mathbb{V}$ y α_i escalares, entonces

$$T\left(\sum_{i=1}^{n} \alpha_i \vec{u}_i\right) = \sum_{i=1}^{n} \alpha_i T(\vec{u}_i).$$

4. Si el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$, es linealmente dependiente en $\mathbb{V}$, entonces el conjunto $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_m)\}$ es linealmente dependiente en $\mathbb{W}$.

Demostración
Queda como ejercicio para el lector.

1.8.1. Representación matricial de una transformación lineal

En este apartado se verá que para toda transformación lineal $T : \mathbb{V} \to \mathbb{W}$ existe una matriz A de tamaño $m \times n$ con $m = \dim \mathbb{W}$ y $n = \dim \mathbb{V}$, tal que

$$T\vec{x} = A\vec{x} \qquad \text{para todo } \vec{x} \in \mathbb{V}.$$

Este hecho es sumamente útil ya que permite determinar de manera fácil el núcleo, la imagen, la nulidad y el rango de una transformación lineal.

Teorema 1.39. Sean $\mathbb{V}$ y $\mathbb{W}$ espacios vectoriales reales de dimensiones n y m, respectivamente. Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces existe una matriz única de tamaño $m \times n$, A_T tal que

$$T\vec{x} = A_T\vec{x} \qquad \text{para todo } \vec{x} \in \mathbb{V}.$$

Demostración
Queda como ejercicio para el lector.

Definición 1.56. Matriz de transformación

La matriz A_T en el Teorema 1.39 se llama matriz de transformación correspondiente a T o representación matricial de T.

Teorema 1.40. Sea A_T la matriz de transformación correspondiente a la transformación lineal T. Entonces

i. $Im_T = Im_{A_T} = C(A_T)$.

ii. $\rho(T) = \rho(A_T)$.

iii. $N_T = N_{A_T}$.

iv. $\nu(T) = \nu(A_T)$.

Demostración

Queda como ejercicio para el lector.

Definición 1.57. Transformación uno a uno

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces T es uno a uno, si

$$T\vec{v}_1 = T\vec{v}_2 \qquad \text{implica que } \vec{v}_1 = \vec{v}_2. \tag{1.33}$$

Es decir, T es uno a uno (escrito $1-1$) si y sólo si todo vector $\vec{w}$ en la imagen de T es la imagen de exactamente un vector en $\mathbb{V}$.

Definición 1.58. Transformación sobre

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces T es sobre, si para todo $\vec{w} \in \mathbb{W}$ existe al menos una $\vec{v} \in \mathbb{V}$ tal que $T\vec{v} = \vec{w}$. Es decir, T es sobre si y sólo si $Im_T = \mathbb{W}$.

Definición 1.59. Isomorfismo

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces T es un isomorfismo si T es uno a uno y sobre.

1.9. Matrices con entradas complejas

En esta sección se desarrollará algunas de las propiedades de las matrices cuyos elementos son números complejos. Toda la aritmética y los teoremas que se han expuesto se aplican a matrices complejas. Estas matrices tienen importantes aplicaciones, por ejemplo, en la mecánica cuántica.

1.9.1. Definición y propiedades básicas

Definición 1.60. Matriz compleja

Una matriz A de tamaño $m\times n$ se dice que es una matriz compleja si sus elementos son números complejos.

Definición 1.61. Matriz conjugada

Sea $A=[a_{ij}]$ una matriz compleja, se llama matriz conjugada de A a la matriz $\overline{A}=[\overline{a}_{ij}]$, donde $\overline{a}_{ij}$ es el conjugado complejo de a_{ij}.

Ejemplo 1.9. Determine la matriz conjugada de la matriz compleja

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

Solución

$$\overline{A}=\begin{bmatrix} \overline{2-3i} & \overline{2+i} \\ \overline{2-i} & \overline{2i} \end{bmatrix}=\begin{bmatrix} 2+3i & 2-i \\ 2+i & -2i \end{bmatrix}.$$

Teorema 1.41. Propiedades de la conjugada compleja

Sean A y B matrices de componentes complejas y sea $\alpha\in\mathbb{C}$. Entonces:

1. $\overline{A+B}=\overline{A}+\overline{B}$.
2. $\overline{\overline{A}}=A$.
3. $\overline{A^t}=\overline{A}^t$.
4. $\overline{\alpha A}=\overline{\alpha}\overline{A}$.
5. $\overline{AB}=\overline{A}\,\overline{B}$.

Demostración

Sean $A=[a_{ij}]$ y $B=[b_{ij}]$, entonces:

1. $\overline{A+B}=[\overline{a_{ij}+b_{ij}}]=[\overline{a_{ij}}+\overline{b_{ij}}]=\overline{A}+\overline{B}$.
2. $\overline{\overline{A}}=[\overline{\overline{a_{ij}}}]=[a_{ij}]=A$.
3. Queda como ejercicio para el lector.
4. $\overline{\alpha A}=[\overline{\alpha a_{ij}}]=[\overline{\alpha}\,\overline{a_{ij}}]=\overline{\alpha}\overline{A}$.

5. Definamos $C = AB$, luego el conjugado del elemento c_{ik}, es

$$\begin{aligned}\overline{c_{ik}} =& \overline{a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}}\\ =& \overline{\sum_{j=1}^{n} a_{ij}b_{jk}} = \sum_{j=1}^{n} \overline{a_{ij}b_{jk}} = \sum_{j=1}^{n} \overline{a_{ij}}\,\overline{b_{jk}}\\ =& \overline{a_{i1}}\,\overline{b_{1k}} + \overline{a_{i2}}\,\overline{b_{2k}} + \ldots + \overline{a_{in}}\,\overline{b_{nk}}.\end{aligned}$$

Definición 1.62. Transpuesta conjugada

La transpuesta conjugada de una matriz compleja A, denotada por A^H, se define como

$$A^H = \overline{A^t} \tag{1.34}$$

donde los elementos de $\overline{A}$ son los conjugados complejos de los elementos correspondientes de A.

Ejemplo 1.10. Determine A^H para la matriz

$$A = \begin{bmatrix} 4+3i & 2+i \\ 2-i & 6i \\ -1 & 1+3i \end{bmatrix}$$

Solución

$$\overline{A} = \begin{bmatrix} \overline{4+3i} & \overline{2+i} \\ \overline{2-i} & \overline{6i} \\ \overline{-1} & \overline{1+3i} \end{bmatrix} = \begin{bmatrix} 4-3i & 2-i \\ 2+i & -6i \\ -1 & 1-3i \end{bmatrix}$$

$$A^H = \overline{A}^t = \begin{bmatrix} 4-3i & 2+i & -1 \\ 2-i & -6i & 1-3i \end{bmatrix}$$

Teorema 1.42. Propiedades de la transpuesta conjugada

Si A y B son matrices complejas y $\alpha \in \mathbb{C}$, entonces se cumplen las siguientes propiedades

1. $(A^H)^H = A$.
2. $(A+B)^H = A^H + B^H$.
3. $(\alpha A)^H = \overline{\alpha}\, A^H$.
4. $(A\,B)^H = B^H\, A^H$.

Demostración

Queda como ejercicio para el lector.

Teorema 1.43. Sea A una matriz compleja de tamaño $n \times n$, entonces

$$\det(\overline{A}) = \overline{\det(A)}.$$

Demostración

La prueba se hará por inducción sobre n. Sea $A = [a]$ una matriz de tamaño 1×1, entonces es claro que

$$\overline{\det(A)} = \overline{a} = \det(\overline{A}).$$

Ahora, supongamos que el Teorema es cierto para matrices de tamaño $(n-1) \times (n-1)$.

Sea $A = [a_{ij}]$ una matriz de tamaño $n \times n$, si se calcula el $\det(A)$ por la k-ésima fila, se tiene que

$$\det(A) = \sum_{j=1}^{n} a_{kj} C_{kj}(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} M_{kj}(A),$$

donde $M_{kj}(A)$ es el menor complementario (k, j). Por la hipótesis de inducción, se verifica que

$$\overline{M_{kj}(A)} = M_{kj}(\overline{A}).$$

Por lo tanto,

$$\begin{aligned}
\overline{\det(A)} &= \overline{\sum_{j=1}^{n} (-1)^{k+j} a_{kj} M_{kj}(A)} \\
&= \sum_{j=1}^{n} (-1)^{k+j} \overline{a_{kj} M_{kj}(A)} \\
&= \sum_{j=1}^{n} (-1)^{k+j} \overline{a}_{kj} \overline{M_{kj}(A)} \\
&= \sum_{j=1}^{n} (-1)^{k+j} \overline{a}_{kj} M_{kj}(\overline{A}) \\
&= \det(\overline{A}).
\end{aligned}$$

1.9.2. Espacios vectoriales complejos

Definición 1.63. Espacios vectoriales complejos

Un espacio vectorial complejo se define exactamente como un espacio vectorial real (definión 1.33), excepto que los escalares en los axiomas *vi*. a *ix*. pueden ser números complejos. Los términos espacio vectorial complejo y espacio vectorial real destacan el conjunto del cual se eligen los escalares.

Los conceptos de combinaciones lineales, conjuntos generadores, dependencia lineal, independencia lineal y base no cambian para los espacios vectoriales complejos, excepto que utilizamos escalares complejos.

Definición 1.64. Producto interno en $\mathbb{C}^n$

Sean $\vec{u},\vec{v} \in \mathbb{C}^n$, se define el producto punto $\vec{u}\cdot\vec{v}$ como

$$\vec{u}\cdot\vec{v} = \vec{u}^H\vec{u} = \sum_{i=1}^{n} \overline{u_i}v_i,$$

donde u_i es el i-ésimo elemento de $\vec{u}$ y v_i es el i-ésimo elemento de $\vec{v}$.

Teorema 1.44. Propiedades del producto punto en $\mathbb{C}^n$

Para todo $\vec{u},\vec{v},\vec{w} \in \mathbb{C}^n$ y todo $\alpha \in \mathbb{C}$:

i) $\vec{u}\cdot\vec{u} \geq 0$ y $\vec{u}\cdot\vec{u} = 0$ si y sólo si $\vec{u} = \vec{0}$ en $\mathbb{C}^n$.

ii) $\vec{u}\cdot\vec{u} = \overline{\vec{v}\cdot\vec{u}}$.

iii) $(\vec{u}+\vec{v})\cdot\vec{w} = \vec{v}\cdot\vec{w}+\vec{v}\cdot\vec{w}$.

iv) $(\alpha\vec{u})\cdot\vec{v} = \overline{\alpha}\,(\vec{u}\cdot\vec{v})$.

Definición 1.65. Partes real e imaginaria de un vector complejo

El complejo conjugado de un vector complejo $\vec{u} \in \mathbb{C}^n$ es el vector $\overline{\vec{u}} \in \mathbb{C}^n$, cuyas componentes son los complejos conjugados de las componentes de $\vec{u}$. Las *partes real e imaginaria de un vector complejo* $\vec{u}$ son respectivamente los vectores $Re(\vec{u}) \in \mathbb{R}^n$ y $Im(\vec{u}) \in \mathbb{R}^n$ formados a partir de las partes reales e imaginarias de cada una de las componentes de $\vec{u}$.

Ejemplo 1.11. Determine las partes real e imaginaria y el correspondiente vector conjugado, del vector $\vec{u}^t = (-i, 1+i, 1)$

Solución
Como

$$\vec{u} = \begin{bmatrix} -i \\ 1+i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + i \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

entonces,

$$Re(\vec{u}) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad Im(\vec{u}) = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

Luego, el vector conjugado es

$$\overline{\vec{u}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - i \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} i \\ 1-i \\ 1 \end{bmatrix}.$$

1.9.3. Solución de sistemas lineales con entradas complejas

Los resultados y las técnicas para resolver sistemas lineales, presentados en la sección 1.7, se traducen de manera directa a los sistemas lineales con coeficientes complejos. En este apartado, como transformar un sistema lineal de $n \times n$ con coeficientes complejos en un sistema lineal $2n \times 2n$ con coeficientes reales.

Consideremos el sistema

$$A\vec{X} = \vec{b}, \tag{1.35}$$

donde A es una matriz compleja de tamaño $n \times n$, $\vec{X}, \vec{b} \in \mathbb{C}^n$. Entonces, el sistema dado en (1.35) se puede escribir como

$$\begin{bmatrix} Re(A) & -Im(A) \\ Im(A) & Re(A) \end{bmatrix} \begin{bmatrix} Re(\vec{x}) \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} Re\left(\vec{b}\right) \\ Im\left(\vec{b}\right) \end{bmatrix},$$

donde $Re(\cdot)$ y $Im(\cdot)$ denotan las partes real e imaginaria, respectivamente. Este nuevo sistema lineal con coeficientes reales es de $2n$ ecuaciones con $2n$ incógnitas. Si se emplean los resultados de la sección 1.5 se tiene que el sistema dado en (1.35) tiene única solución si y sólo si

$$\det\{Re(A)\} \neq 0 \qquad \text{y} \qquad \det\left\{Re(A) + Im(A)\,[Re(A)]^{-1}\,Im(A)\right\} \neq 0.$$

En cuyo caso la solución está dada por

$$\begin{bmatrix} Re(\vec{x}) \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} I_n & [Re(A)]^{-1}\,Im(A) \\ -[Re(A)]^{-1}\,Im(A) & I_n \end{bmatrix} \begin{bmatrix} C^{-1}Re\left(\vec{b}\right) \\ C^{-1}Im\left(\vec{b}\right) \end{bmatrix},$$

donde I_n es la matriz identidad de tamaño $n \times n$ y $C = Re(A) + Im(A)\,[Re(A)]^{-1}\,Im(A)$.

Ejemplo 1.12. Determine una solución del sistema de ecuaciones lineales

$$\begin{aligned} (2+i)x_1 + (1+i)x_2 =& 3+6i \\ (3-i)x_1 + (2-2i)x_2 =& 7-i. \end{aligned} \tag{1.36}$$

Solución

Al representar matricialmente (1.36) se llega a

$$\begin{bmatrix} 2+i & 1+i \\ 3-i & 2-2i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3+6i \\ 7-i \end{bmatrix}. \tag{1.37}$$

Luego, la matriz A se puede expresar como

$$\begin{aligned} A = \begin{bmatrix} 2+i & 1+i \\ 3-i & 2-2i \end{bmatrix} &= \underbrace{\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}} + i \underbrace{\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}} \\ &= Re(A) + i\;\; Im(A). \end{aligned}$$

De manera análoga, el vector de constantes es

$$\begin{bmatrix} 3+6i \\ 7-i \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} + i \begin{bmatrix} 6 \\ -1 \end{bmatrix}$$

Si $x_1 = a_1 + ib_1$ y $x_2 = a_2 + ib_2$, entonces (1.36) se puede reescribir como

$$\begin{bmatrix} Re(A) & \vdots & -Im(A) \\ \cdots & . & \cdots \\ Im(A) & \vdots & Re(A) \end{bmatrix} \begin{bmatrix} Re(\vec{x}) \\ \cdots \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} Re\left(\vec{b}\right) \\ \cdots \\ Im\left(\vec{b}\right) \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & \vdots & -1 & -1 \\ 3 & 2 & \vdots & 1 & 2 \\ \cdots & \cdots & . & \cdots & \cdots \\ 1 & 1 & \vdots & 2 & 1 \\ -1 & -2 & \vdots & 3 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ \cdots \\ 6 \\ -1 \end{bmatrix}.$$

Para hallar C, el lector puede realizar las respectivas operaciones y llegar a

$$C = \begin{bmatrix} 0 & -2 \\ 10 & 12 \end{bmatrix} \qquad \text{y} \qquad \det C = 20 \neq 0.$$

Por otra parte, el determinante de $Re(A) = 1 \neq 0$. Luego, el sistema tiene única solución y está dada por

$$\begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \vdots & 3 & 4 \\ 0 & 1 & \vdots & -5 & -7 \\ \cdots & \cdots & . & \cdots & \cdots \\ -3 & -4 & \vdots & 1 & 0 \\ 5 & 7 & \vdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{2} \\ -\frac{3}{2} \\ \cdots \\ \frac{7}{2} \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ \cdots \\ 2 \\ -1 \end{bmatrix}.$$

Por lo tanto, la solución del sistema lineal dado es

$$x_1 = 1 + 2i \qquad \text{y} \qquad x_2 = 2 - i.$$

Capítulo 2

Vectores característicos y valores característicos

En una gran variedad de aplicaciones dada una transformación lineal $T : \mathbb{V} \to \mathbb{V}$, resulta útil encontrar un vector $\vec{v}$ en $\mathbb{V}$ tal que $T\vec{v}$ y $\vec{v}$ sean paralelos. Esto es, se busca un vector $\vec{v}$ y un escalar λ tal que

$$T\vec{v} = \lambda\vec{v} \tag{2.1}$$

tenga una solución $\vec{v} \neq \vec{0}$. En este caso λ se denomina *valor característico* de T y $\vec{v}$ se llama *vector característico* de T correspondiente al valor característico λ. Si $\dim(\mathbb{V}) = n$, el problema de determinar los respectivos valores característicos de T puede resolverse con la ayuda de los determinantes. Nótese que la ecuación (2.1) puede escribirse en la forma

$$(T - \lambda I)\vec{v} = \vec{0}$$

donde I es la transformación identidad. Si denotamos $T_\lambda = T - \lambda I$, entonces λ es un valor característico si y sólo si la ecuación

$$T_\lambda(\vec{v}) = \vec{0} \tag{2.2}$$

tiene una solución $\vec{v}$ no nula, en cuyo caso T_λ no es invertible. Pues una solución no nula de (2.2) existe si y sólo si la matriz de T_λ es singular. Si A_T es una representación matricial de T, entonces $A_T - \lambda I$ es una representación matricial para T_λ. Por esta razón en este capítulo se estudiarán algunas de las propiedades de los valores y vectores característicos de las matrices de tamaño $n \times n$.

2.1. Valores propios y vectores propios

Definición 2.1. Valor Característico y Vector Característico

Un vector característico de una matriz A de tamaño $n \times n$ es un vector $\vec{v}$ diferente de cero, que cumple:

$$A\vec{v} = \lambda\vec{v} \tag{2.3}$$

para algún escalar λ. El escalar λ es llamado valor característico de A si existe una solución no trivial $\vec{v}$ del sistema $A\vec{v} = \lambda\vec{v}$; también $\vec{v}$ se denomina vector característico correspondiente a λ.

Nota
Los valores característicos se llaman también *autovalores, eigenvalores o valores propios* y los vectores característicos, *vectores propios, eigenvectores o autovectores*.

Teorema 2.1. Sea $\vec{v}$ un vector propio de una matriz A asociado al valor propio λ. Sea $\alpha \neq 0$, entonces $\alpha\vec{v}$ también es un vector propio de A correspondiente al valor propio λ.

Demostración
Se debe probar que $\alpha\vec{v}$ satisface la ecuación (2.3). Utilizando el hecho de que $A\vec{v} = \lambda\vec{v}$, se tiene que

$$A(\alpha\vec{v}) = \alpha(A\vec{v}) = \alpha(\lambda\vec{v}) = \lambda(\alpha\vec{v})$$

lo cual completa la prueba.

Ejemplo 2.1. Los vectores $\vec{u}^t = (-1, 1)$ y $\vec{v}^t = (2, 1)$ son vectores propios de la siguiente matriz

$$A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}.$$

Solución

Se tiene que

$$A\vec{u} = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1\vec{u}.$$

$$A\vec{v} = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 14 \end{bmatrix} \neq \lambda \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Así, $\vec{u}$ es un vector propio correspondiente al valor propio $\lambda = -1$; pero $\vec{v}$ no es un vector propio de A, porque $A\vec{v}$ no es un múltiplo de $\vec{v}$, es decir, no existe un escalar λ tal que $2\lambda = 4$ y $\lambda = 14$ se verifiquen simultáneamente.

Ejemplo 2.2. Considere la matriz A dada en el Ejemplo 2.1, muestre que 6 es un valor propio de A y encuentre el vector propio correspondiente.

Solución

El escalar 6 es un valor propio de A si y sólo si la ecuación

$$A\vec{v} = 6\vec{v}, \tag{2.4}$$

tiene una solución no trivial. Pero (2.4) es equivalente a $A\vec{v} - 6\vec{v} = \vec{0}$, o

$$(A - 6I)\vec{v} = \vec{0}. \tag{2.5}$$

Para resolver esta ecuación homogénea, se forma la matriz

$$A - 6I = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} - \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} -5 & 2 \\ 5 & -2 \end{bmatrix}.$$

Las columnas de $A - 6I$ son obviamente linealmente dependientes, es decir, (2.4) tiene solución no trivial; luego 6 es un valor propio de A. Para encontrar el vector propio correspondiente, se realizan operaciones por fila

$$\left[\begin{array}{cc|c} -5 & 2 & 0 \\ 5 & -2 & 0 \end{array}\right] \underset{F_2+F_1}{\sim} \left[\begin{array}{cc|c} -5 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

La solución general tiene la forma $\begin{bmatrix} 1 \\ \frac{5}{2} \end{bmatrix} x$. Según el Teorema 2.1, cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda = 6$.

Teorema 2.2. Sea A una matriz real de tamaño $n \times n$ e I_n la matriz identidad de tamaño $n \times n$, entonces la función p definida por la ecuación

$$\begin{aligned} p_A(\lambda) = \det(A - \lambda I_n) = (-\lambda)^n + tr_1(A)\,(-\lambda)^{n-1} + tr_2(A)\,(-\lambda)^{n-2} + \dots \\ + tr_{n-1}(A)\,(-\lambda) + tr_n(A), \end{aligned} \tag{2.6}$$

donde $tr_k(A)$ denota la suma de los menores principales de orden k del $\det(A)$, es un polinomio en λ de grado n y el término independiente es $p_A(0) = \det(A)$.

Demostración

Vamos a demostrar que p_A es un polinomio de grado n únicamente para el caso $n \leq 3$. La demostración para el caso general puede hacerse por inducción. Para $n = 1$ el determinante es el polinomio lineal $p_A(\lambda) = a_{11} - \lambda$.

Para $n = 2$ se tiene que

$$\begin{aligned} \det(A - \lambda I_2) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 - tr_1(A)\lambda + \det(A). \end{aligned}$$

Obsérvese que el polinomio obtenido es de segundo grado en λ. Para $n = 3$ tenemos

$$\begin{aligned} \det(A - \lambda I_3) &= \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - [a_{32}a_{23}(a_{11} - \lambda) + a_{13}a_{31}(a_{22} - \lambda) + a_{21}a_{12}(a_{33} - \lambda)] \\ &= -\lambda^3 + tr_1(A)\lambda^2 - \left(\sum_{i=1}^{3} M_{ii}(A)\right)\lambda + \det(A) \end{aligned}$$

Nótese que en este caso se obtiene un polinomio de tercer grado, siendo el término de mayor grado $-\lambda^3$.

La afirmación $p_A(0) = \det(A)$ resulta inmediata de la definición de p_A.

Teorema 2.3. Sea A una matriz real de tamaño $n \times n$. Entonces λ es un valor propio de A si y sólo si

$$p_A(\lambda) = \det(A - \lambda I) = 0 \tag{2.7}$$

Demostración

Supóngase que λ es un valor propio de A. Entonces existe un elemento $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda\vec{v}$, de donde $A\vec{v} - \lambda\vec{v} = \vec{0}$ o $(A - \lambda I)\vec{v} = \vec{0}$. Por lo tanto, $A - \lambda I$ tiene un núcleo no nulo y $A - \lambda I$ no es invertible, es decir $\det(A - \lambda I) = 0$.

Recíprocamente, supóngase que $\det(A - \lambda I) = 0$, es decir $A - \lambda I$ no es invertible. Entonces $A - \lambda I$ debe tener un núcleo no nulo, lo que significa que existe un elemento $\vec{v} \neq \vec{0}$ tal que $(A - \lambda I)\vec{v} = \vec{0}$. Por lo tanto, $A\vec{v} - \lambda\vec{v} = \vec{0}$ o $A\vec{v} = \lambda\vec{v}$; así, λ es un valor propio de A.

Definición 2.2. Ecuación y polinomio característico

La ecuación (2.6) se llama el ***polinomio característico*** de A. La ecuación (2.7) se llama la ***ecuación característica*** de A.

Definición 2.3. Multiplicidad algebraica

Sea λ_k un valor propio de una matriz A de tamaño $n \times n$. Entonces la *multiplicidad algebraica* de λ_k es el número de veces que λ_k aparece como raíz del polinomio característico de A; es decir, es igual a su multiplicidad como raíz de la ecuación característica.

Ejemplo 2.3. Encuentre el polinomio y ecuación característica de

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

Solución

Fórmese $A - \lambda I$ y calcúlese su determinante

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} 4-\lambda & -1 & 6 \\ 2 & 1-\lambda & 6 \\ 2 & -1 & 8-\lambda \end{vmatrix} \\
&= (4-\lambda)\begin{vmatrix} 1-\lambda & 6 \\ -1 & 8-\lambda \end{vmatrix} - (-1)\begin{vmatrix} 2 & 6 \\ 2 & 8-\lambda \end{vmatrix} + 6\begin{vmatrix} 2 & 1-\lambda \\ 2 & -1 \end{vmatrix} \\
&= (4-\lambda)[(1-\lambda)(8-\lambda)+6] + [2(8-\lambda)-12] + 6[-2-2(1-\lambda)].
\end{aligned}$$

simplificando el producto, se obtiene el polinomio característico

$$p_A(\lambda) = -\lambda^3 + 13\lambda^2 - 40\lambda + 36.$$

Los resultados del Teorema 2.2 se cumplen, ya que la $tr_1(A) = 13$, $\sum_{i=1}^{3} M_{ii}(A) = 40$ y el $\det(A) = 36$; factorizando p_A, se llega a

$$p_A(\lambda) = -(\lambda - 2)^2(\lambda - 9).$$

En este caso, la ecuación característica es

$$(\lambda - 2)^2(\lambda - 9) = 0.$$

Nótese que el valor propio 2 tiene *multiplicidad algebraica* 2 porque $(\lambda - 2)$ aparece dos veces como factor del polinomio característico.

Teorema 2.4. Sea A una matriz real de tamaño $n \times n$ con valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_m$. Para $k = 1, 2, \ldots, m$, sea

$$\mathcal{B}_k = \{\vec{v} \in \mathbb{C}^n : A\vec{v} = \lambda_k \vec{v}\} \tag{2.8}$$

Entonces para cada k, $\mathcal{B}_k$ es un subespacio de $\mathbb{C}^n$.

Demostración

Como $A\vec{0} = \lambda_k \vec{0}$ para todo k, $\vec{0} \in \mathcal{B}_k$. Si λ_k no es un valor propio, no existen vectores $\vec{v} \neq \vec{0}$ excepto $\vec{0}$ que satisface $A\vec{v} = \lambda_k \vec{v}$; en este caso $\mathcal{B}_k$ es el *subespacio trivial*.

Ahora supongamos que λ_k es un valor propio; entonces existe un $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda_k \vec{v}$, en otras palabras $(A - \lambda_k I)\vec{v} = \vec{0}$. De esta manera $\mathcal{B}_k = \{\vec{v} \in \mathbb{C}^n : (A - \lambda_k I)\vec{v} = \vec{0}\}$; es el *espacio solución* del sistema homogéneo $(A - \lambda_k I)\vec{v} = \vec{0}$, el cual es un subespacio de $\mathbb{C}^n$.

Definición 2.4. Espacio propio

Sea λ_k un valor propio de A. El subespacio $\mathcal{B}_k$ se denomina ***espacio propio*** de A correspondiente al valor propio λ_k.

Teorema 2.5. Sea A una matriz real de tamaño $n \times n$. Entonces los espacios propios correspondientes a valores propios distintos de A tienen al vector nulo en común. Es decir, $\mathcal{B}_k \cap \mathcal{B}_r = \left\{\vec{0}\right\}$ si $\lambda_k \neq \lambda_r$.

Demostración

Supongamos que existe un vector propio de A tal que $\vec{v} \in \mathcal{B}_k$ y $\vec{v} \in \mathcal{B}_r$, entonces por (2.8) se tiene que

$$A\vec{v} = \lambda_k \vec{v} \qquad \text{y} \qquad A\vec{v} = \lambda_r \vec{v}.$$

Luego,

$$(\lambda_k - \lambda_r)\vec{v} = \{\vec{0}\}$$

pero como $\lambda_k \neq \lambda_r$, se concluye que $\vec{v} = \vec{0}$; por lo tanto, $\mathcal{B}_k \cap \mathcal{B}_r = \left\{\vec{0}\right\}$ y la prueba queda completa.

Teorema 2.6. Si $\vec{v}_1, \ldots, \vec{v}_m$ son vectores propios correspondientes a valores propios distintos $\lambda_1, \ldots, \lambda_m$ de una matriz A de tamaño $n \times n$, entonces el conjunto $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es linealmente independiente.

Demostración

La demostración es por inducción sobre m. El resultado es trivial cuando $m = 1$. Entonces supongamos, que se ha demostrado para cualquier conjunto $m = k$ de vectores propios. Sean $\vec{v}_1, \ldots, \vec{v}_{k+1}$, $k+1$ vectores propios pertenecientes a valores propios distintos y supongamos que existen escalares c_i tales que

$$\sum_{i=1}^{k+1} c_i \vec{v}_i = \vec{0}. \tag{2.9}$$

Si se multiplica por A a ambos lados de (2.9) y se utiliza el hecho de que $A\vec{v}_i = \lambda_i \vec{v}_i$, se obtiene

$$\sum_{i=1}^{k+1} c_i \lambda_i \vec{v}_i = \vec{0}. \tag{2.10}$$

Al restar de (2.10) el producto de (2.9) por λ_{k+1}, se obtiene la ecuación

$$\sum_{i=1}^{k} c_i (\lambda_i - \lambda_{k+1}) \vec{v}_i = \vec{0}.$$

Como los vectores $\vec{v}_1, \ldots, \vec{v}_k$ son linealmente independientes por hipótesis de inducción, se debe tener que $c_i(\lambda_i - \lambda_{k+1}) = 0$ para cada $i = 1, 2, \ldots, k$. Además, dado que los valores propios son distintos se tiene que $\lambda_i \neq \lambda_{k+1}$ para $i \neq k+1$, así que $c_i = 0$ para $i = 1, 2, \ldots, k$ y de (2.9) se tiene que c_{k+1} es también 0; por lo tanto, el conjunto de vectores propios $\vec{v}_1, \ldots, \vec{v}_{k+1}$, es también linealmente independiente.

Teorema 2.7. Los vectores no nulos tomados de espacios propios distintos son linealmente independientes. Es decir, los espacios propios $\mathcal{B}_1, \ldots, \mathcal{B}_m$ (correspondientes a los valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_m$) cumplen que

$$\text{“}Si\ \vec{u}_1 + \ldots + \vec{u}_m = \vec{0}\ con\ \vec{u}_k \in \mathcal{B}_k\ entonces\ \vec{u}_1 = \ldots = \vec{u}_m = \vec{0}\text{”}$$

Demostración

Supongamos que $\vec{u}_1 + \ldots + \vec{u}_m = \vec{0}$ con $\vec{u}_k \in \mathcal{B}_k$, esto es, $A\vec{u}_k = \lambda_k \vec{u}_k$. Si algunos $\vec{u}_k$ no fueran $\vec{0}$, ellos serían vectores propios de A, correspondientes a valores propios distintos y entonces el que la suma de ellos sea $\vec{0}$ contradice el Teorema 2.6.

Definición 2.5. Multiplicidad geométrica

Sea λ_k un valor propio de una matriz A de tamaño $n \times n$. Entonces la *multiplicidad geométrica* de λ_k es el número máximo de vectores propios de A linealmente independientes que tienen un valor propio igual a λ_k; es decir, es igual a la dimensión del espacio propio correspondiente a λ_k (lo cual es la nulidad de la matriz $A - \lambda_k I$). En consecuencia,

$$\text{multiplicidad geométrica de } \lambda_k = \dim(\mathcal{B}_k) = \nu\,(A - \lambda_k I_n)\,.$$

Nota

La multiplicidad geométrica de un valor propio nunca es cero. Esto se establece de la definición 2.1, la cual expresa que si λ es un valor propio, entonces existe un vector propio *diferente de cero* correspondiente a λ.

A continuación se presenta un procedimiento para calcular los valores propios y vectores propios de una matriz A de tamaño $n \times n$.

Determinación de valores propios y vectores propios
i) Encuentre $p_A(\lambda) = \det(A - \lambda I)$. ii) Halle las raíces $\lambda_1, \lambda_2, \ldots, \lambda_m$ de $p_A(\lambda) = 0$. iii) Resuelva el sistema homogéneo $(A - \lambda_k I)\vec{v} = \vec{0}$, correspondiente a cada valor propio λ_k.

Ejemplo 2.4. Encuentre los vectores propios y espacios propios asociados a la matriz dada en el Ejemplo 2.3.

Solución

En el Ejemplo 2.3, se obtuvo que la ecuación característica era

$$(\lambda - 2)^2(\lambda - 9) = 0.$$

De esta manera los valores propios de A son $\lambda_1 = 9$ y $\lambda_2 = 2$ (con multiplicidad algebraica 2).

Para $\lambda_1 = 9$ tenemos

$$(A - 9I)\vec{v} = \begin{bmatrix} -5 & -1 & 6 \\ 2 & -8 & 6 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para determinar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c} -5 & -1 & 6 & 0 \\ 2 & -8 & 6 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right] \underset{F_1 \leftrightarrow \frac{1}{2}F_2}{\sim} \left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ -5 & -1 & 6 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right] \underset{\substack{F_2+5F_1 \\ F_3-2F_1}}{\sim} \left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ 0 & -21 & 21 & 0 \\ 0 & 7 & -7 & 0 \end{array}\right]$$

$$\underset{\substack{F_2+3F_3=F_2' \\ F_2' \leftrightarrow \frac{1}{7}F_3}}{\sim} \left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \underset{F_1+4F_2}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

La solución general corresponde a $x = z, y = z$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} z$. Cada vector de esta forma con $z \neq 0$ es un vector propio correspondiente a $\lambda_1 = 9$. Por lo que $\mathcal{B}_1 = gen\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$.

Para $\lambda_2 = 2$ tenemos

$$(A-2I)\vec{v} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\begin{bmatrix} 2 & -1 & 6 & | & 0 \\ 2 & -1 & 6 & | & 0 \\ 2 & -1 & 6 & | & 0 \end{bmatrix} \underset{\substack{F_2-F_1 \\ F_3-F_1}}{\sim} \begin{bmatrix} 2 & -1 & 6 & | & 0 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.$$

La solución general corresponde a $y = 2x + 6z$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix} z$. Cada vector de esta forma con $x, z \neq 0$ es un vector propio correspondiente a $\lambda_2 = 2$. Por lo que $\mathcal{B}_2 = gen \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix} \right\}$.

Teorema 2.8. Sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ valores propios distintos de la matriz A. Si para cada $k = 1, 2, \ldots, m$, S_k es un conjunto linealmente independiente de vectores propios de A correspondientes a λ_k, entonces

$$S = S_1 \cup S_2 \cup \ldots \cup S_m$$

es todavía un conjunto de vectores linealmente independiente.

Demostración

Sea $S_k = \{\vec{x}_{k1}, \ldots, \vec{x}_{kr_k}\}$ un conjunto linealmente independiente, de vectores propios de la matriz A, correspondiente al valor propio λ_k (para cada $k \in \{1, 2, \ldots, m\}$).

Para probar que $S = S_1 \cup S_2 \cup \ldots \cup S_m$ es linealmente independiente, consideremos una combinación lineal de los vectores en esa unión tal que

$$\underbrace{(a_{11}\vec{x}_{11} + \ldots + a_{1r_1}\vec{x}_{1r_1})}_{\text{en } \mathcal{B}_1} + \ldots + \underbrace{(a_{m1}\vec{x}_{m1} + \ldots + a_{mr_m}\vec{x}_{mr_m})}_{\text{en } \mathcal{B}_m} = \vec{0}.$$

Por el Teorema 2.7, se puede afirmar que cada suma entre paréntesis es $\vec{0}$. Entonces, como los vectores que participan en cada una de esas sumas son linealmente independientes, se concluye que los coeficientes tienen que ser nulos, con lo cual se prueba que S es linealmente independiente.

La prueba del siguiente enunciado no es difícil si se explican algunos otros resultados. Su demostración se realiza en la siguiente sección.

Teorema 2.9. Sea λ_k un valor propio de una matriz real A de tamaño $n \times n$, con multiplicidad algebraica r. Entonces,

$$\textit{multiplicidad geométrica de } \lambda_k \leq \textit{multiplicidad algebraica de } \lambda_k.$$

Teorema 2.10. Si A es una matriz no singular de tamaño $n \times n$, con valores propios no nulos $\lambda_1, \lambda_2, \ldots, \lambda_n$ y vectores propios $\vec{v}_1, v_2, \ldots, \vec{v}_n$. Su matriz inversa A^{-1} tiene

$i)$ Los valores propios de la forma $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$.

$ii)$ Los mismos vectores propios de A.

Demostración

$i)$ Si los valores propios de A son diferentes de cero, entonces

$$\begin{aligned}
p_A(\lambda) =& \det(A - \lambda I_n) = \det\left[A\left(I_n - \lambda A^{-1}\right)\right] \\
=& \det(A)\det\left[-\lambda\left(A^{-1} - \frac{1}{\lambda}I_n\right)\right] \\
=& (-\lambda)^n \det(A)\det\left(A^{-1} - \frac{1}{\lambda}I_n\right) \\
=& (-1)^n \lambda^n \det(A) p_{A^{-1}}\left(\frac{1}{\lambda}\right).
\end{aligned}$$

Luego, se deduce que $1/\lambda$ es un valor propio de A^{-1} por cada valor λ de A.

$ii)$ Si $A\vec{u} = \lambda\vec{u}$, entonces premultiplicando por A^{-1}, se tiene

$$\begin{aligned}
A^{-1}(A\vec{u}) =& A^{-1}(\lambda\vec{u}) \\
\vec{u} =& \lambda A^{-1}\vec{u} \\
\frac{1}{\lambda}\vec{u} =& A^{-1}\vec{u}.
\end{aligned}$$

Por lo tanto, $\vec{u}$ es también un vector propio de A^{-1}.

Ejercicios 2.1.

1. Para las siguientes matrices calcule los valores propios y los espacios propios

$a.\ \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}.\quad b.\ \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}.\quad c.\ \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}.\quad d.\ \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$

$e.\ \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}.\quad f.\ \begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}.\quad g.\ \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}.\quad h.\ \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$

2. Sea la matriz $A = \begin{bmatrix} 1 & k \\ 1 & 1 \end{bmatrix}$, con k una constante arbitraria. ¿Para qué valores de k la matriz A tiene dos valores propios reales distintos?.

3. Si A es una matriz diagonal de tamaño $n \times n$, muestre que sus valores propios son las entradas de su diagonal.

4. Si A es una matriz triangular de tamaño $n \times n$, muestre que sus valores propios son las entradas de su diagonal principal.

5. Si A es una matriz real de tamaño $n \times n$, muestre que es invertible si y sólo si el número 0 no es un valor propio de A.

6. Si A es una matriz real de tamaño $n \times n$ con la propiedad de que la suma de los elementos de sus filas es igual siempre a un número s, muestre que s es un valor propio de A.

7. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios de A, demuestre que

   $a)$ La matriz A^t tiene los mismos valores propios.

b) La matriz kA tiene valores propios $k\lambda_1, k\lambda_2, \ldots, k\lambda_n$.

c) La matriz A^k (donde k es un entero positivo) tiene valores propios $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$.

2.2. Matrices semejantes y diagonalización

En la sección anterior, se desarrolló parte del vocabulario y de las propiedades de los valores propios y vectores propios. En este apartado, continuaremos estudiando los valores propios, debido a que estos números son cruciales en muchas consideraciones incluyendo la representación de matrices en formas en las cuales es más fácil de trabajar la resolución de problemas.

Definición 2.6. Matrices congruentes

Dos matrices reales A y B de tamaño $n \times n$ son congruentes, si existe una matriz P no singular de componentes reales de tamaño $n \times n$ tal que

$$A = P^t BP. \tag{2.11}$$

Ejemplo 2.5. Determine si las siguientes matrices son congruentes

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix}.$$

Solución

Veamos si existe una matriz P, tal que $B = P^t AP$, en particular P puede ser una matriz triangular superior, en otras palabras,

$$\begin{aligned} \begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} &= \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}^t \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \\ &= \begin{bmatrix} a^2 & a(b+4d) \\ a(b+4d) & b^2+8bd+d^2 \end{bmatrix}. \end{aligned}$$

En consecuencia,

$$a^2 = 1, \qquad a(b+4d) = 1 \qquad \text{y} \qquad b^2 + 8bd + d^2 = -14.$$

Si se despeja b de la segunda ecuación y se reemplaza en la tercera, se tiene que

$$\begin{aligned} \left(\frac{1}{a} - 4d\right)^2 + 8d\left(\frac{1}{a} - 4d\right) + d^2 = \frac{1}{a^2} - 15d^2 \\ 1 - 15d^2 = -14. \end{aligned}$$

Luego, $d^2 = 1$; por lo tanto, la matriz P puede ser

$$\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \qquad \text{ó} \qquad \begin{bmatrix} -1 & -5 \\ 0 & 1 \end{bmatrix} \qquad \text{ó} \qquad \begin{bmatrix} 1 & 5 \\ 0 & -1 \end{bmatrix}.$$

Teorema 2.11. Dos matrices reales A y B de tamaño $n \times n$ son congruentes si y sólo si tienen el mismo rango.

Demostración

Como A y B son matrices congruentes de tamaño $n \times n$, existe una matriz no singular P tal que $B = P^t AP$, entonces

$$\begin{aligned} \rho(B) &= \rho\left[\left(P^t A\right) P\right] = \rho\left(P^t A\right) && \textit{puesto que } P \textit{ es no singular,} \\ &= \rho(A) && \textit{por ser } P^t \textit{ también no singular.} \end{aligned}$$

Aquí, se utilizo la propiedad (x) dada en el Teorema 1.36.

Teorema 2.12. La congruencia de matrices de tamaño $n \times n$ cumple las propiedades de relación de equivalencia, es decir, es

a) Reflexiva: *A es congruente a A.*

b) Simétrica: *Si A es congruente a B entonces B es congruente a A.*

c) Transitiva: *Si A es congruente a B y B es congruente a C, entonces A es congruente a C.*

Demostración

Queda como ejercicio para el lector.

Para las matrices cuadradas, además del concepto de congruencia, se tiene otro de mayor utilidad o generalidad el de similaridad.

Definición 2.7. Matrices semejantes

Una matriz A de tamaño $n \times n$ es semejante (o similar) a una matriz B de tamaño $n \times n$ si existe una matriz no singular P de tamaño $n \times n$ tal que

$$B = P^{-1}AP. \tag{2.12}$$

De manera análoga se dice que A y B son semejantes si y sólo si existe una matriz no singular P tal que

$$PB = AP. \tag{2.13}$$

Ejemplo 2.6. Sean $A = \begin{bmatrix} -1 & 0 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} -3 & 5 \\ -2 & 4 \end{bmatrix}$ y $P = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$. Determine si A y B son semejantes.

Solución

Se realizan los productos AP y PB

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

$$PB = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -3 & 5 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 3 \end{bmatrix}.$$

Así, $AP = PB$. Como $\det(P) = 1 \neq 0$, entonces, P es no singular. Y por la ecuación (2.13), se tiene que A y B son semejantes.

Teorema 2.13. Las matrices semejantes tienen el mismo polinomio característico y, por tanto, los mismos valores propios.

Demostración

Como A y B son matrices semejantes de tamaño $n \times n$, $B = P^{-1}AP$. Entonces

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}[AP - \lambda P] = P^{-1}[A - \lambda I]P.$$

Por consiguiente,

$$\begin{aligned} \det(B-\lambda I) &= \det[P^{-1}(A-\lambda I)P] = \det(P^{-1})\det(A-\lambda I)\det(P) \\ &= \det(P^{-1})\det(P)\det(A-\lambda I) = \det(A-\lambda I). \end{aligned}$$

Esto significa que A y B tienen la misma ecuación característica y como los valores propios son raíces de la ecuación característica, entonces tienen los mismos valores propios.

Ejemplo 2.7. Para las matrices A y B dadas en el Ejemplo 2.6, muestre que tienen el mismo polinomio característico.

Solución

Tenemos que

$$\det(A-\lambda I) = \begin{vmatrix} -1-\lambda & 0 \\ 1 & 2-\lambda \end{vmatrix} = (-1-\lambda)(2-\lambda) = \lambda^2 - \lambda - 2.$$

y

$$\det(B-\lambda I) = \begin{vmatrix} -3-\lambda & 5 \\ -2 & 4-\lambda \end{vmatrix} = (-3-\lambda)(4-\lambda) + 10 = (\lambda^2 - \lambda - 12) + 10.$$

Como $\det(A-\lambda I) = \det(B-\lambda I)$, las matrices A y B tienen el mismo polinomio característico y por lo tanto los mismos valores propios.

Teorema 2.14. La semejanza de matrices de tamaño $n \times n$ cumple las propiedades de relación de equivalencia, es decir, es

a) Reflexiva: *A es semejante a A.*

b) Simétrica: *Si A es semejante a B entonces B es semejante a A.*

c) Transitiva: *Si A es semejante a B y B es semejante a C, entonces A es semejante a C.*

Demostración

Queda como ejercicio para el lector.

Teorema 2.15. Si B es una matriz semejante a A con $B = P^{-1}AP$, entonces $\vec{v}$ es un vector propio de A asociado con el valor propio λ si y sólo si $P^{-1}\vec{v}$ es un vector propio de B asociado con el valor propio λ.

Demostración

Si $\vec{v}$ es un vector propio de A, se tiene que

$$\begin{aligned} A\vec{v} &= \lambda\vec{v} \\ \left(PBP^{-1}\right)\vec{v} &= \lambda\vec{v} && \text{puesto que } B \text{ es semejante a } A, \\ B\left(P^{-1}\vec{v}\right) &= \lambda P^{-1}\vec{v} && \text{puesto que } P \text{ es no singular.} \end{aligned}$$

lo cual completa la prueba.

Ejemplo 2.8. Para cada una de las matrices A y B dadas en el Ejemplo 2.6 determine sus vectores propios

Solución

Para la matriz A se tiene que los valores propios son $\lambda_1 = -1$ y $\lambda_2 = 2$ y sus correspondientes vectores propios son $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ y $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, respectivamente.

Para la matriz B se tiene que los valores propios son $\lambda_1 = -1$ y $\lambda_2 = 2$ y sus correspondientes vectores propios son $\begin{bmatrix} -5 \\ -2 \end{bmatrix}$ y $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, respectivamente.

El lector puede verificar que los vectores propios de B son iguales a los vectores propios de A premultiplicados por la inversa de la matriz P dada en el Ejemplo 2.6.

Con lo que se ha estudiado hasta ahora en esta sección, se puede presentar una demostración del Teorema 2.9.

Demostración del Teorema 2.9

Sea $\mathcal{B}_k = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ base del espacio propio correspondiente al valor propio λ_k, donde m es la multiplicidad geométrica de λ_k. Se extiende $\mathcal{B}_k$ hasta completar una base de $\mathbb{R}^n$, digamos

$$\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\}.$$

En esta base, la matriz A está particionada como

$$[A]_{\mathfrak{B}} = \begin{bmatrix} \lambda_k I_m & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

Luego, A y $[A]_{\mathfrak{B}}$ son matrices semejantes, es decir, tienen el mismo polinomio característico y los mismos valores propios con idénticas multiplicidades algebraicas. Así que el polinomio característico de A es

$$p_A(\lambda) = p_{[A]_{\mathfrak{B}}}(\lambda) = (\lambda_k - \lambda)^m p_D(\lambda)$$

por lo tanto, λ_k aparece como raíz de $p_A(\lambda)$ por lo menos m veces; por consiguiente, la multiplicidad algebraica de λ_k es mayor o igual a m.

Definición 2.8. Matriz diagonalizable

Una matriz A de tamaño $n \times n$ es diagonalizable si existe una matriz diagonal D tal que A es semejante a D.

Este resultado es muy importante ya que las matrices diagonales poseen muchas propiedades que permiten trabajar fácilmente con ellas, véase Teorema 1.13.

Teorema 2.16. Una matriz A de tamaño $n \times n$ es diagonalizable si y sólo si A tiene n vectores propios linealmente independientes. En tal caso, si $A = PDP^{-1}$ donde D es diagonal, entonces los elementos de la diagonal de D son los valores propios de A y las columnas de P son los vectores propios correspondientes.

Demostración

Primero se supone que A es diagonalizable. Entonces existe una matriz no singular P tal que $P^{-1}AP = D$ es diagonal. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los elementos de la diagonal principal de D y sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ los vectores columna de la matriz P, entonces

$$\begin{aligned} PD &= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \ldots & \lambda_n\vec{v}_n \end{bmatrix}, \end{aligned}$$

pero como $AP = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \dots & A\vec{v}_n \end{bmatrix}$ y por otra parte $P^{-1}AP = D$, se tiene que $AP = PD$, lo cual implica

$$\begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \dots & A\vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \dots & \lambda_n\vec{v}_n \end{bmatrix}.$$

En otras palabras, $A\vec{v}_i = \lambda_i\vec{v}_i$ para todo vector columna $\vec{v}_i$. Esto significa que los vectores columna $\vec{v}_i$ de P son vectores propios de A. Además, como P es una matriz no singular, entonces sus vectores columna son linealmente independientes. Así, A tiene n vectores propios linealmente independientes.

Recíprocamente, suponga que A tiene n vectores propios linealmente independientes $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ con valores propios asociados $\lambda_1, \lambda_2, \dots, \lambda_n$. Sea P la matriz cuyas columnas son estos n vectores propios. Es decir, $P = [\vec{v}_1 \ \vec{v}_2 \ \dots \ \vec{v}_n]$. Como todo $\vec{v}_i$ es un vector propio de A, entonces se tiene que $A\vec{v}_i = \lambda_i\vec{v}_i$ y

$$AP = A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \dots & \lambda_n\vec{v}_n \end{bmatrix}$$

nótese que la matriz del lado derecho de esta ecuación puede escribirse como el siguiente producto de matrices

$$AP = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = PD$$

Por último, dado que los vectores $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ son linealmente independientes, entonces P es no singular y se puede escribir la ecuación $AP = PD$ como $P^{-1}AP = D$, lo cual significa que A es diagonalizable.

Corolario 2.16.1. Si una matriz A de tamaño $n \times n$ tiene n valores propios distintos, entonces A es diagonalizable.

Demostración

Sean $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ los vectores propios correspondientes a los n valores propios distintos de la matriz A. Entonces, por el Teorema 2.6, se tiene que el conjunto $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ es linealmente independientes. Luego por el Teorema 2.16, A es diagonalizable.

A continuación se presenta un procedimiento para diagonalizar una matriz A de tamaño $n \times n$.

Procedimiento para diagonalizar una matriz cuadrada
Sea A una matriz real de tamaño $n \times n$. i) Determine n vectores propios $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ de A, con valores propios correspondientes $\lambda_1, \lambda_2, \ldots, \lambda_n$. Si no existen n vectores propios linealmente independientes, entonces A no es diagonalizable. ii) Obtenga P como la matriz cuyas columnas son los vectores propios obtenidos en el paso i. Es decir, $$P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}.$$ iii) La matriz diagonal $D = P^{-1}AP$ tendrá los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ en su diagonal principal (y ceros en el resto). La posición de los vectores propios en la matriz P determina la posición en que aparecen los valores propios sobre la diagonal de D.

Ejemplo 2.9. Determine si la matriz dada a continuación es diagonalizable

$$A = \begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}.$$

Solución

La ecuación característica asociada a la matriz A es

$$-\lambda^3 + 4\lambda^2 - 4\lambda = -\lambda(\lambda - 2)^2 = 0.$$

Luego, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = 2$ (de multiplicidad algebraica 2). El vector propio correspondiente a $\lambda_1 = 0$ es $\vec{v}_1 = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$ y los correspondientes a $\lambda_2 = 2$ son $\vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$. Entonces,

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

y

$$P^{-1}AP = \frac{1}{2}\begin{bmatrix} 1 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} -1 & -2 & 3 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$
$$= \frac{1}{2}\begin{bmatrix} 1 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} 0 & -4 & 6 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Por lo tanto, A es diagonalizable.

Ejemplo 2.10. Una matriz no diagonalizable

Determine si la siguiente matriz es diagonalizable

$$A = \begin{bmatrix} 3 & -2 \\ 8 & -5 \end{bmatrix}.$$

Solución

La ecuación característica de A es $\lambda^2 + 2\lambda + 1 = (\lambda + 1)^2 = 0$, luego $\lambda = -1$ es un valor propio de multiplicidad algebraica 2. Entonces,

$$(A - \lambda I)\vec{v} = (A + I)\vec{v} = \begin{bmatrix} 4 & -2 \\ 8 & -4 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto lleva al vector propio $\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Por lo tanto, A no contiene dos vectores propios linealmente independientes y entonces, se concluye que la matriz A no es diagonalizable.

Teorema 2.17. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ valores propios distintos de una matriz A de tamaño $n \times n$, entonces

$$tr_1(A) = \sum_{i=1}^{n} \lambda_i \qquad \text{y} \qquad \det(A) = \prod_{i=1}^{n} \lambda_i.$$

Demostración

Como A es diagonalizable, entonces $A = PDP^{-1}$; luego,

$$tr_1(A) = tr_1[P(DP^{-1})] = tr_1[(DP^{-1})P] = tr_1[D(P^{-1}P)] = tr_1(D) = \sum_{i=1}^{n} \lambda_i.$$

Por otra parte,

$$\det(A) = |A| = |P(DP^{-1})| = |P||DP^{-1}| = |D||P^{-1}||P| = |D| = \prod_{i=1}^{n} \lambda_i.$$

Ejercicios 2.2.

1. Para las siguientes matrices determine (en caso de ser posible) una matriz P no singular tal que $P^{-1}AP$ sea diagonal

$$a. \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \quad b. \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \quad c. \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}. \quad d. \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$$

$$f. \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}. \quad g. \begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}. \quad h. \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}. \quad i. \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

2. Muestre que las trazas de matrices semejantes son iguales.

3. Si A y B son semejantes demuestre que tienen el mismo determinante.

4. Sean A una matriz diagonalizable de tamaño $n \times n$ y P una matriz no singular de tamaño $n \times n$ tales que $B = P^{-1}AP$ sea la forma diagonal de A. Pruebe que

   a) $B^k = P^{-1}A^kP$, donde k es un entero positivo.

   b) $A^k = PB^kP^{-1}$, donde k es un entero positivo.

5. Sea $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, demuestre que A es diagonalizable si $-4bc < (a-d)^2$ y no diagonalizable si $-4bc > (a-d)^2$.

2.3. Valores propios complejos

Puesto que la ecuación característica de una matriz real de tamaño $n \times n$ es un polinomio de grado n, por el *Teorema Fundamental del Álgebra* se sabe que cualquier polinomio de grado n con coeficientes reales (o complejos) tiene exactamente n raíces (contando multiplicidades). En las secciones anteriores, desarrollamos la teoría para valores propios y vectores propios reales. En esta sección, estudiaremos los valores propios y vectores propios complejos.

Definición 2.9. Sea A una matriz real de tamaño $n \times n$. El número complejo λ es un valor propio de A si existe un vector no nulo $\vec{v} \in \mathbb{C}^n$ tal que

$$A\vec{v} = \lambda\vec{v} \tag{2.14}$$

Todo vector $\vec{v}$ no nulo que satisfaga (2.14) es un vector propio de A asociado al valor propio λ.

Ejemplo 2.11. Sea $A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$. Determine los valores propios y vectores propios de A.

Solución

Primero se calcula el determinante de la matriz $A - \lambda I$

$$\begin{aligned} \det(A - \lambda I) = & \begin{vmatrix} 1-\lambda & -1 \\ 2 & 1-\lambda \end{vmatrix} \\ = & (1-\lambda)^2 + 2 = [(1-\lambda) + i\sqrt{2}][(1-\lambda) - i\sqrt{2}]. \end{aligned}$$

De esta manera los valores propios de A son los complejos, a saber

$$\lambda_1 = 1 + i\sqrt{2} \qquad \text{y} \qquad \lambda_2 = 1 - i\sqrt{2}.$$

Para $\lambda_1 = 1 + i\sqrt{2}$ tenemos

$$(A - (1+i\sqrt{2})I)\vec{v} = \begin{bmatrix} -i\sqrt{2} & -1 \\ 2 & -i\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\begin{bmatrix} -i\sqrt{2} & -1 & | & 0 \\ 2 & -i\sqrt{2} & | & 0 \end{bmatrix} \underset{i\sqrt{2}F_1}{\sim} \begin{bmatrix} 2 & -i\sqrt{2} & | & 0 \\ 2 & -i\sqrt{2} & | & 0 \end{bmatrix} \underset{F_2-F_1}{\sim} \begin{bmatrix} 2 & -i\sqrt{2} & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}.$$

La solución general corresponde a $2x = i\sqrt{2}y$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_1 = 1 + i\sqrt{2}$. Por lo tanto, $\mathcal{B}_1 = gen\left\{\begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix}\right\}$.

Para $\lambda_2 = 1 - i\sqrt{2}$ tenemos

$$(A - (1 - i\sqrt{2})I)\vec{v} = \begin{bmatrix} i\sqrt{2} & -1 \\ 2 & i\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\begin{bmatrix} i\sqrt{2} & -1 & | & 0 \\ 2 & i\sqrt{2} & | & 0 \end{bmatrix} \underset{-i\sqrt{2}F_1}{\sim} \begin{bmatrix} 2 & i\sqrt{2} & | & 0 \\ 2 & i\sqrt{2} & | & 0 \end{bmatrix} \underset{F_2-F_1}{\sim} \begin{bmatrix} 2 & i\sqrt{2} & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}.$$

La solución general corresponde a $2x = -i\sqrt{2}y$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_2 = 1 - i\sqrt{2}$. Por lo tanto, $\mathcal{B}_2 = gen\left\{\begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix}\right\}$.

Teorema 2.18.

Sea A una matriz real de tamaño $n \times n$. Entonces,

i) Los valores propios de A cuando son complejos ocurren en pares conjugados,

ii) Los vectores propios correspondientes a valores propios complejos, son conjugados complejos entre sí.

Demostración

i) Si A es una matriz real de tamaño $n \times n$, su polinomio característico se puede reescribir como

$$p_A(\lambda) = \det(A - \lambda I_n) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$$

donde cada $c_i \in \mathbb{R}$. Por lo tanto,

$$\overline{p_A(\lambda)} = \overline{\det(A-\lambda I_n)} = \det\left(\overline{A-\lambda I_n}\right) = \det\left(A-\overline{\lambda} I_n\right) = p_A\left(\overline{\lambda}\right).$$

Si λ_0 es una raíz de $p_A(\lambda)$, entonces

$$\overline{p_A(\lambda_0)} = p_A\left(\overline{\lambda_0}\right) = 0.$$

En consecuencia $\overline{\lambda_0}$ es también un valor propio de A.

ii) Si λ es un valor propio complejo de A con un vector propio correspondiente $\vec{u} \in \mathbb{C}^n$, entonces en virtud del Teorema 1.41 se tiene

$$\begin{aligned} A\vec{u} &= \lambda\vec{u} \\ \overline{A\vec{u}} &= \overline{\lambda\vec{u}} \\ A\,\overline{\vec{u}} &= \overline{\lambda}\,\overline{\vec{u}} \end{aligned}$$

Luego, $\overline{\vec{v}}$ es también vector propio de A pero asociado al valor propio $\overline{\lambda}$.

Ejemplo 2.12. Considere la matriz A del Ejemplo 2.11 y aplique el Teorema 2.18.

Solución

Para la matriz A cuyas componentes son reales, se obtuvo un valor propio $\lambda_1 = 1+i\sqrt{2}$, con vector propio asociado $\vec{u}_1 = \begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix}$ y para el otro valor propio $\lambda_2 = 1-i\sqrt{2}$, un vector propio asociado $\vec{u}_2 = \begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix}$ claramente se nota que $\lambda_2 = \overline{\lambda}_1$ y que $\vec{u}_2 = \overline{\vec{u}}_1$.

Teorema 2.19. Sea A una matriz real de tamaño 2×2 con un valor propio complejo $\lambda = a+bi$ $(b \neq 0)$ y vector propio correspondiente $\vec{u} \in \mathbb{C}^2$. Entonces

$$\begin{aligned} ARe(\vec{u}) &= aRe(\vec{u}) - bIm(\vec{u}) \\ AIm(\vec{u}) &= bRe(\vec{u}) + aIm(\vec{u}) \end{aligned} \tag{2.15}$$

además, $Re(\vec{u})$ y $Im(\vec{u})$ son vectores linealmente independientes.

Demostración

Sea $\vec{u} \in \mathbb{C}^2$ un vector propio de A, por lo tanto,

$$\begin{aligned} A\vec{u} =& \lambda\vec{u} \\ A[Re(\vec{u}) + i\, Im(\vec{u})] =& (a+bi)[Re(\vec{u}) + i\, Im(\vec{u})] \\ A\, Re(\vec{u}) + iA\, Im(\vec{u}) =& [a\, Re(\vec{u}) - b\, Im(\vec{u})] + i[b\, Re(\vec{u}) + a\, Im(\vec{u})]. \end{aligned}$$

Al igualar las partes real e imaginaria se llega al sistema de ecuaciones (2.15).

Además, por la definición 2.1 se tiene que $\vec{u}$ es no nulo; luego, si $Im(\vec{u}) = \vec{0}$, entonces $Re(\vec{u}) \neq \vec{0}$ y de la segunda ecuación de (2.15), se tiene que $b\, Re(\vec{u}) = \vec{0}$, es decir, $b = 0$, lo cual contradice la suposición de que $b \neq 0$; por lo tanto, $Im(\vec{u}) \neq \vec{0}$.

Veamos ahora que $Re(\vec{u})$ y $Im(\vec{u})$ son vectores linealmente independientes (por contradicción); Supongamos que $Re(\vec{u}) = \alpha\; Im(\vec{u})$, si se reemplaza en (2.15), se tiene que

$$\begin{aligned} A\,\alpha\, Im(\vec{u}) =& a\,\alpha\, Im(\vec{u}) - b\, Im(\vec{u}) \\ A\, Im(\vec{u}) =& b\,\alpha\, Im(\vec{u}) + a\, Im(\vec{u}). \end{aligned}$$

Si se resuelve dicho sistema de ecuaciones, se obtiene que

$$(\alpha^2 + 1)b\, Im(\vec{u}) = \vec{0}.$$

Como $b \neq 0$ y $Im(\vec{u}) \neq \vec{0}$; entonces, $\alpha = \pm i$, luego $Re(\vec{u}) \in \mathbb{C}^2$, lo cual es contradictorio ya que $Re(\vec{u})$ y $Im(\vec{u}) \in \mathbb{R}^2$.

El corolario que se enuncia a continuación muestra que una matriz con componentes reales cuyos valores propios son complejos no es diagonalizable.

Corolario 2.19.1. Sea A una matriz real de tamaño 2×2 con un valor propio complejo $\lambda = a + bi$ $(b \neq 0)$ y vector propio asociado $\vec{u} \in \mathbb{C}^2$. Entonces,

$$A = PRP^{-1} \tag{2.16}$$

donde,

$$P = \begin{bmatrix} Re(\vec{u}) & Im(\vec{u}) \end{bmatrix} \qquad \text{y} \qquad R = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Demostración

Expresando el sistema de ecuaciones propuesto en (2.15) en forma matricial se tiene que

$$A\left[Re(\vec{u}) \quad Im(\vec{u})\right] = \left[Re(\vec{u}) \quad Im(\vec{u})\right]\begin{bmatrix} a & b \\ -b & a \end{bmatrix},$$

es decir, $AP = PR$. Pero en el Teorema 2.19 se demostró que $Re(\vec{u})$ y $Im(\vec{u})$, eran vectores linealmente independientes; luego, P es no singular. Por lo tanto,

$$A = PRP^{-1}.$$

Ejemplo 2.13. Expresar de la forma PRP^{-1} la matriz

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

Solución

En el Ejemplo 2.11, se encontraron los dos valores propios $\lambda_1 = 1 + i\ \sqrt{2}$ y $\lambda_2 = 1 - i\ \sqrt{2}$ y los respectivos vectores propios $\vec{v}_1 = \begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix}$. Estableciendo

$$P = \begin{bmatrix} 1 & 0 \\ 0 & -\sqrt{2} \end{bmatrix} \qquad \text{y} \qquad R = \begin{bmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & 1 \end{bmatrix}$$

el lector puede verificar fácilmente que $A = PRP^{-1}$.

Ejemplo 2.14. Encuentre las matrices P y R, de tal manera que se pueda expresar la siguiente matriz A como PRP^{-1}

$$A = \begin{bmatrix} 29 & 9 & -31 \\ 20 & 70 & -5 \\ 66 & -164 & 51 \end{bmatrix}.$$

Solución

La ecuación característica de A esta dada por

$$\begin{aligned} \det(A - \lambda I) &= -\lambda^3 + 150\lambda^2 - 8125\lambda + 312500 \\ &= -(\lambda - 100)(\lambda - 25 + 50\ i)(\lambda - 25 - 50\ i) = 0. \end{aligned}$$

Por lo tanto, los valores propios son $\lambda_1 = 100$, $\lambda_2 = 25 - 50\,i$ y $\lambda_3 = \overline{\lambda_2}$.

Para $\lambda_1 = 100$ se resuelve $(A - \lambda_1 I)\vec{v} = \vec{0}$ y se obtiene el vector propio asociado $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$. Para $\lambda_2 = 25 - 50\,i$ se resuelve $(A - \lambda_2 I)\vec{v} = \vec{0}$ y se obtiene el vector propio complejo asociado $\vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ -2 \end{bmatrix} i$.

Por consiguiente, estableciendo

$$P = \begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix} \qquad \text{y} \qquad R = \begin{bmatrix} 25 & -50 & 0 \\ 50 & 25 & 0 \\ 0 & 0 & 100 \end{bmatrix}$$

se obtiene

$$\begin{aligned} PRP^{-1} &= -\frac{1}{25} \begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix} \begin{bmatrix} 25 & -50 & 0 \\ 50 & 25 & 0 \\ 0 & 0 & 100 \end{bmatrix} \begin{bmatrix} 2 & -8 & -3 \\ 7 & -3 & 2 \\ -2 & -17 & 3 \end{bmatrix} \\ &= -\frac{1}{25} \begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix} \begin{bmatrix} -300 & -50 & -175 \\ 275 & -475 & -100 \\ -200 & -1700 & 300 \end{bmatrix} \\ &= -\frac{1}{25} \begin{bmatrix} -725 & -225 & 775 \\ -500 & -1750 & 125 \\ -1650 & 4100 & -1275 \end{bmatrix} = \begin{bmatrix} 29 & 9 & -31 \\ 20 & 70 & -5 \\ 66 & -164 & 51 \end{bmatrix}. \end{aligned}$$

En este ejemplo se ilustra la manera de expresar A como PRP^{-1} cuando sus valores propios no son todos complejos.

Ejercicios 2.3.

1. Exprese cada una de las matrices dadas como PRP^{-1}

$$a. \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}. \qquad b. \begin{bmatrix} 1 & -1 \\ 3 & -1 \end{bmatrix}. \qquad c. \begin{bmatrix} -1 & 3 \\ -1 & 1 \end{bmatrix}.$$

$$d. \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \qquad f. \begin{bmatrix} 2 & 2 & -3 \\ -3 & 0 & 3 \\ 4 & 1 & -5 \end{bmatrix}. \qquad g. \begin{bmatrix} -5 & 2 & 4 \\ -3 & 0 & 3 \\ -3 & 1 & 2 \end{bmatrix}.$$

2. Suponga que A es una matriz real de tamaño 3×3, tal que $\det A = 50$, $tr_1(A) = 8$ y un valor propio es 2. Encuentre los valores propios.

3. Sea A una matriz real de tamaño $n \times n$ y sea $\vec{x} \in \mathbb{C}^n$. Muestre que

$$Re(A\vec{x}) = ARe(\vec{x}) \qquad \text{y} \qquad Im(A\vec{x}) = AIm(\vec{x}).$$

2.4. Diagonalización de matrices simétricas

Como se vio en la sección anterior una matriz A de tamaño $n \times n$ puede tener valores propios complejos, aún en el caso de que todos los elementos de A sean reales. En este apartado, se desarrollará la teoría de valores propios para matrices simétricas reales.

Definición 2.10. Matrices congruentes ortogonalmente

Dos matrices simétricas reales A y B de tamaño $n \times n$ son congruentes ortogonalmente, si existe una matriz P ortogonal de tamaño $n \times n$ tal que

$$A = P^t BP. \tag{2.17}$$

Definición 2.11. Matrices semejantes ortogonalmente

Una matriz simétrica A de tamaño $n \times n$ es semejante ortogonalmente a una matriz simétrica B de tamaño $n \times n$, si existe una matriz P ortogonal de tamaño $n \times n$ tal que

$$A = P^t BP. \tag{2.18}$$

Teorema 2.20. Dos matrices simétricas reales A y B son congruentes ortogonalmente si y sólo si A y B son semejantes ortogonalmente.

Demostración

Si A y B son matrices simétricas congruentes de tamaño $n \times n$, entonces

$$B = P^t AP$$

pero como $P^t P = I_n$, se tiene que $P^t = P^{-1}$. Por lo tanto, las matrices A y B son semejantes ortogonalmente.

Teorema 2.21. Sea A una matriz simétrica real de tamaño $n \times n$. Entonces los valores propios de A son reales.

Demostración

Sea $\vec{v} \in \mathbb{C}^n$ un vector propio asociado al valor propio λ de A; entonces,

$$A\vec{v} = \lambda\vec{v} \tag{2.19}$$

por el Teorema 2.18, se tiene que $\overline{\vec{v}}$ es también vector propio de A pero asociado al valor propio $\overline{\lambda}$.

Si se multiplica (2.19), por la izquierda por $\overline{\vec{v}}^t$, se obtiene

$$\overline{\vec{v}}^t A\vec{v} = \overline{\vec{v}}^t \lambda\vec{v} = \lambda\overline{\vec{v}}^t \vec{v}. \tag{2.20}$$

Luego, la conjugada de $\overline{\vec{v}}^t A\vec{v}$, es

$$\overline{\overline{\vec{v}}^t A\vec{v}} = \overline{\overline{\vec{v}}^t}\ \overline{A}\ \overline{\vec{v}} = \overline{\overline{\vec{v}^t}}\ \overline{\lambda}\ \overline{\vec{v}} = \overline{\lambda}\vec{v}^t\overline{\vec{v}}, \tag{2.21}$$

donde hemos utilizado el Teorema 1.41. Por otra parte, como A es real se tiene que $A = \overline{A}$. Por lo tanto, la ecuación (2.21) es igual a

$$\overline{\overline{\vec{v}}^t A\vec{v}} = \overline{\overline{\vec{v}}^t}\ A\ \overline{\vec{v}} = \vec{v}^t\ A^t\ \overline{\vec{v}} = (A\vec{v})^t\ \overline{\vec{v}} = \lambda\vec{v}^t\overline{\vec{v}} \tag{2.22}$$

aquí se utilizo el hecho de que $A^t = A$ ya que A es simétrica. Si se igualan (2.21) y (2.22) se tiene

$$\overline{\lambda}\ \vec{v}^t\ \overline{\vec{v}} = \lambda\ \vec{v}^t\ \overline{\vec{v}} \tag{2.23}$$

pero $\vec{v}^t\ \overline{\vec{v}} = ||\vec{v}||^2 \neq 0$, ya que $\vec{v}$ es un vector propio. Entonces se puede dividir ambos lados de (2.23) entre $\vec{v}^t\ \overline{\vec{v}}$ para obtener

$$\overline{\lambda} = \lambda$$

lo cual se cumple sólo si λ es real.

Definición 2.12. Matriz diagonalizable ortogonalmente

Una matriz A de tamaño $n \times n$ es diagonalizable ortogonalmente si existe una matriz ortogonal Q tal que

$$Q^t A Q = D \tag{2.24}$$

sea diagonal.

Teorema 2.22. Si A es una matriz simétrica real de tamaño $n \times n$, entonces existe una matriz ortogonal Q tal que

$$Q^{-1} A Q$$

es una matriz diagonal.

Demostración

Sea $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios de A. Puesto que λ_1 es real, existe un vector propio unitario $\vec{u}_1 \in \mathbb{R}^n$ de A correspondiente a λ_1. Denotemos por V el complemento ortogonal a $\vec{u}_1$ de dimensión $n-1$. Sea $\{\vec{u}_2, \vec{u}_3, \ldots, \vec{u}_n\}$ una base ortonormal de V. Luego, cada vector $\vec{X}$ de V tiene la forma

$$\vec{X} = a_2\, \vec{u}_2 + a_3\, \vec{u}_3 + \ldots + a_n\, \vec{u}_n$$

y el producto punto entre $A\vec{X}$ y $\vec{u}_1$ es

$$(A\vec{X}) \cdot \vec{u}_1 = (A\vec{X})^t \vec{u}_1 = \vec{X}^t A^t \vec{u}_1 = \vec{X}^t (A\vec{u}_1) = \vec{X}^t (\lambda_1 \vec{u}_1) = \lambda_1 \vec{X}^t \vec{u}_1 = 0.$$

Puesto que cada vector de la base de V es ortogonal a $\vec{u}_1$. La matriz de cambio de base, de la *base canónica* de $\mathbb{R}^n$ a la base $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ es la matriz ortogonal S cuyas columnas son los elementos de los vectores $\vec{u}_i$. Luego,

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix}, \end{aligned}$$

por lo tanto,

$$S^{-1} A S = S^{-1} \begin{bmatrix} \lambda_1 \vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix}.$$

Pero como S es ortogonal se tiene que $S^{-1} = S^t$, por consiguiente,

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_n \end{bmatrix}^t \begin{bmatrix} \lambda_1\vec{u}_1 & A\vec{u}_2 & \dots & A\vec{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \vec{u}_1^t\lambda_1\vec{u}_1 & \vec{u}_1^tA\vec{u}_2 & \dots & \vec{u}_1^tA\vec{u}_n \\ \vec{u}_2^t\lambda_1\vec{u}_1 & \vec{u}_2^tA\vec{u}_2 & \dots & \vec{u}_2^tA\vec{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_n^t\lambda_1\vec{u}_1 & \vec{u}_n^tA\vec{u}_2 & \dots & \vec{u}_n^tA\vec{u}_n \end{bmatrix} \end{aligned}$$

se puede probar fácilmente que $S^{-1}AS$ es simétrica, ya que

$$(S^{-1}AS)^t = (S^tAS)^t = S^tAS = S^{-1}AS.$$

Por consiguiente,

$$S^{-1}AS = \left[\begin{array}{c|c} \lambda_1 & 0 \quad \dots \quad 0 \\ 0 & \\ \vdots & \left[\quad A_1 \quad\right] \\ 0 & \end{array}\right]$$

donde A_1 es simétrica de tamaño $(n-1)\times(n-1)$.

La prueba es ahora completada por inducción, si R^* es una matriz ortogonal de tamaño $(n-1)\times(n-1)$ tal que $R^*A_1(R^*)^{-1} = diag\{\lambda_2,\lambda_3,\dots,\lambda_n\}$. Entonces, la matriz

$$R = \left[\begin{array}{c|c} 1 & 0 \quad \dots \quad 0 \\ 0 & \\ \vdots & \left[\quad R^* \quad\right] \\ 0 & \end{array}\right]$$

es una matriz ortogonal y

$$R^{-1}S^{-1}ASR = (SR)^{-1}A(SR) = diag\{\lambda_1,\lambda_2,\dots,\lambda_n\}.$$

La matriz $SR = Q$ es el producto de dos matrices ortogonales; por lo tanto, es también una matriz ortogonal. Así, $Q^{-1}AQ$ es una matriz diagonal y nuestra prueba queda completa.

Teorema 2.23. Sea A una matriz simétrica real de tamaño $n\times n$. Entonces los vectores propios asociados con valores propios distintos de A son ortogonales. Esto es, los espacios propios de una matriz simétrica son ortogonales.

Demostración

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios que corresponden a valores propios distintos, digamos, λ_1 y λ_2. Para demostrar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned}
\lambda_1\vec{v}_1 \cdot \vec{v}_2 =&(\lambda_1\vec{v}_1)^t\vec{v}_2 = (A\vec{v}_1)^t\vec{v}_2 && \textit{puesto que } \vec{v}_1 \textit{ es un vector propio}\\
=&(\vec{v}_1^t A^t)\vec{v}_2 = \vec{v}_1^t(A\vec{v}_2) && \textit{puesto que } A \textit{ es simétrica}\\
=&\vec{v}_1^t(\lambda_2\vec{v}_2) && \textit{puesto que } \vec{v}_2 \textit{ es un vector propio}\\
=&\lambda_2\vec{v}_1^t\vec{v}_2 = \lambda_2\vec{v}_1 \cdot \vec{v}_2.
\end{aligned}$$

Por lo tanto, $(\lambda_1 - \lambda_2)\vec{v}_1 \cdot \vec{v}_2 = 0$. Pero $\lambda_1 - \lambda_2 \neq 0$, así que $\vec{v}_1 \cdot \vec{v}_2 = 0$

Teorema 2.24. Sea λ_k un valor propio de multiplicidad algebraica igual a p asociado a una matriz A simétrica real de tamaño $n \times n$. Entonces A tiene exactamente p vectores propios mutuamente ortogonales asociados al valor propio λ_k.

Demostración

Por el Teorema 2.22, existe una matriz Q tal que $Q^{-1}AQ$ es una matriz diagonal en la cual λ_k aparece exactamente p veces en la diagonal principal. Por otra parte, se tiene que, $Q^{-1}AQ - \lambda_k I_n = Q^{-1}(A - \lambda_k I_n)Q$ tiene rango $n - p$. Pero como Q y Q^{-1} son no singulares, $A - \lambda_k I_n$ también tiene rango $n - p$. Por lo tanto, el espacio solución del sistema de ecuaciones

$$(A - \lambda_k I_n)\vec{v} = \vec{0} \qquad \text{con} \qquad \vec{v} \in \mathbb{R}^n$$

tiene dimensión $n - (n - p) = p$ y por consiguiente, existen exactamente p vectores unitarios mutuamente ortogonales de $\mathbb{R}^n$.

Teorema 2.25. Una matriz A simétrica real de tamaño $n \times n$, tiene n vectores propios unitarios mutuamente ortogonales.

Demostración

Si D es la matriz diagonal Q^tAQ, se tiene

$$AQ = QD \tag{2.25}$$

donde $D = diag\{\lambda_1, \ldots, \lambda_n\}$; al igualar los vectores columna de cada miembro de (2.25), se obtiene

$$A\,\vec{v}_1 = \lambda_1\,\vec{v}_1,\; A\,\vec{v}_2 = \lambda_2\,\vec{v}_2,\; \ldots, A\,\vec{v}_n = \lambda_n\,\vec{v}_n$$

donde $\vec{v}_1,\ \vec{v}_2,\ \ldots,\ \vec{v}_n$ son los vectores columna de Q; se deduce que las columnas de Q son vectores propios de A y que son vectores unitarios mutuamente ortogonales, por ser Q ortogonal.

Teorema 2.26. Sea A una matriz real de tamaño $n \times n$. Entonces A es diagonalizable ortogonalmente si y sólo si A es simétrica.

Demostración

Sea A diagonalizable ortogonalmente, entonces por la Definición 2.12, existe una matriz ortogonal Q tal que $Q^tAQ = D$. Si se multiplica esta ecuación por la izquierda por Q y por la derecha por Q^t y se usa el hecho de que $Q^tQ = QQ^t = I_n$, se obtiene

$$A = QDQ^t.$$

Luego

$$A^t = (QDQ^t)^t = (Q^t)^t D^t Q^t = QDQ^t = A.$$

Así, A es simétrica.

Recíprocamente, suponga que A es simétrica. Entonces por los Teoremas 2.23 y 2.25, A es diagonalizable ortogonalmente con la matriz Q cuyas columnas son los vectores propios dados en el Teorema 2.25, y el teorema queda demostrado.

Procedimiento para encontrar una matriz diagonalizante Q

i) Encuentre una base para cada espacio propio de A.

ii) Halle una base ortonormal para cada espacio propio de A usando el proceso de Gram-Schmidt o algún otro.

iii) Obtenga Q como la matriz cuyas columnas son los vectores propios ortonormales obtenidos en el paso ii).

Ejemplo 2.15. Encuentre una matriz Q que diagonalice ortogonalmente a

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

Solución

En este caso, los valores propios de A son $\lambda_1 = 1$ (de multiplicidad algebraica 2) y $\lambda_2 = 7$. Los vectores propios linealmente independientes correspondientes a $\lambda_1 = 1$ son $\vec{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ y el correspondiente a $\lambda_2 = 7$ es $\vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

Para encontrar Q, se aplica el proceso de Gram-Schmidt a $\{\vec{v}_1, \vec{v}_2\}$, una base para $\mathcal{B}_1$. Como $\|\vec{v}_1\| = \sqrt{2}$, se hace $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$. Después,

$$\begin{aligned} \vec{v}_2' =& \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} - \frac{-2}{\sqrt{2}} \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} \\ =& \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Entonces $\|\vec{v}_2'\| = \sqrt{3}$ y $\vec{u}_2 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$. Se puede verificar que la nueva base de $\mathcal{B}_1$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por último, se tiene que $\|\vec{v}_3\| = \sqrt{6}$ luego $\vec{u}_3 = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$. También se verifica que la base obtenida para $\mathbb{R}^3$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_3 = 0$ y $\vec{u}_2 \cdot \vec{u}_3 = 0$. Por lo tanto,

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}, \qquad Q^t = \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}$$

y

$$\begin{aligned} Q^tAQ =& \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \\ =& \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 7/\sqrt{6} \\ 0 & -1/\sqrt{3} & 14/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 7/\sqrt{6} \end{bmatrix} \end{aligned}$$

Así,

$$Q^tAQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

Teorema 2.27. Descomposición Espectral para Matrices Simétricas

Sea A una matriz simétrica real de tamaño $n \times n$ con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \lambda_1 \vec{v}_1 \vec{v}_1^t + \lambda_2 \vec{v}_2 \vec{v}_2^t + \ldots + \lambda_n \vec{v}_n \vec{v}_n^t, \tag{2.26}$$

donde $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ son los vectores propios normalizados de A.

Demostración

Por el Teorema 2.22, existe una matriz Q tal que $Q^{-1}AQ$ es una matriz diagonal. Entonces,

$$\begin{aligned} A =& QDQ^{-1} = QDQ^t \\ =& \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^t \\ \vec{u}_2^t \\ \vdots \\ \vec{u}_n^t \end{bmatrix} \\ =& \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{u}_1^t \\ \lambda_2 \vec{u}_2^t \\ \vdots \\ \lambda_n \vec{u}_n^t \end{bmatrix} \\ =& \lambda_1 \vec{v}_1 \vec{v}_1^t + \lambda_2 \vec{v}_2 \vec{v}_2^t + \ldots + \lambda_n \vec{v}_n \vec{v}_n^t \end{aligned}$$

esto prueba el teorema.

Ejemplo 2.16. Ilustrar el teorema de descomposición espectral para la matriz dada en el Ejemplo 2.15.

Solución

Del Ejemplo 2.15 se tiene que los valores propios asociados a la matriz A son $\lambda_1 = 7,\ \lambda_2 = \lambda_3 = 1$ y los respectivos vectores propios normalizados de A eran

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Entonces

$$\begin{aligned}
\sum_{i=1}^{3}\lambda_i\vec{v}_i\vec{v}_i^t &= \frac{7}{6}\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 1\end{bmatrix}+\frac{1}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix}\begin{bmatrix}1 & -1 & 1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}\begin{bmatrix}1 & 0 & -1\end{bmatrix}\\
&=\frac{7}{6}\begin{bmatrix}1 & 2 & 1\\2 & 4 & 2\\1 & 2 & 1\end{bmatrix}+\frac{1}{3}\begin{bmatrix}1 & -1 & 1\\-1 & 1 & -1\\1 & -1 & 1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1 & 0 & -1\\0 & 0 & 0\\-1 & 0 & 1\end{bmatrix}\\
&=\frac{1}{6}\begin{bmatrix}7 & 14 & 7\\14 & 28 & 14\\7 & 14 & 7\end{bmatrix}+\frac{1}{6}\begin{bmatrix}5 & -2 & -1\\-2 & 2 & -2\\-1 & -2 & 5\end{bmatrix}=\frac{1}{6}\begin{bmatrix}12 & 12 & 6\\12 & 30 & 12\\6 & 12 & 12\end{bmatrix},
\end{aligned}$$

la cual coincide con la matriz A dada en el Ejemplo 2.15.

Teorema 2.28. Teorema espectral para matrices simétricas

Sea A una matriz simétrica real de tamaño $n\times n$, entonces las siguientes afirmaciones son equivalentes

i) A tiene n valores propios reales, contando multiplicidades.

ii) Si λ es un valor propio de A con multiplicidad algebraica k, entonces el espacio propio para λ es k-dimensional.

iii) Los vectores propios correspondientes a valores propios distintos son ortogonales, es decir los espacios propios son mutuamente ortogonales.

iv) A es diagonalizable ortogonalmente.

Demostración

Queda como ejercicio para el lector.

Ejercicios 2.4.

1. Determine si las matrices dadas a continuación son diagonalizables ortogonalmente

$$a.\begin{bmatrix}1 & 1\\ 1 & 3\end{bmatrix}. \qquad b.\begin{bmatrix}1 & 3\\ 3 & -1\end{bmatrix}. \qquad c.\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}.$$

$$d.\begin{bmatrix}1 & 3 & -2\\ 3 & 1 & 2\\ -2 & 2 & 1\end{bmatrix}. \qquad f.\begin{bmatrix}3 & -1 & 4\\ -1 & 1 & -5\\ 4 & -5 & 2\end{bmatrix}.$$

2. Si las matrices A y B son ortogonalmente semejantes y B es ortogonalmente semejante a una matriz C, muestre que A y C son también ortogonalmente semejantes.

3. Sea A una matriz ortogonal de tamaño $n\times n$, muestre que si λ es un valor propio de A, entonces $\lambda=\pm1$.

4. Muestre que si A es ortogonal de tamaño $n\times n$ y $\vec{x}$ y $\vec{y}$ son vectores en $\mathbb{R}^n$, entonces $(A\vec{x})\cdot(A\vec{y})=\vec{x}\cdot\vec{y}$.

2.5. Vectores propios generalizados

En las secciones anteriores hemos considerado matrices en las cuales la multiplicidad algebraica de cada valor propio es igual a su multiplicidad geométrica. En este apartado consideraremos matrices que violan esta condición, es decir la multiplicidad algebraica de cada valor propio es diferente de su multiplicidad geométrica y se obtendrá un nuevo concepto de vector propio asociado a la matriz.

Definición 2.13. Vector propio generalizado

Sea A una matriz real de tamaño $n \times n$ con un valor propio λ_j cuya multiplicidad algebraica es diferente de su multiplicidad geométrica. Un vector $\vec{v} \in \mathbb{C}^n$ se llama ***vector propio generalizado*** de A si cumple que

$$(A - \lambda_j I_n)^k \vec{v} = \vec{0} \tag{2.27}$$

para algún k entero positivo. El mínimo entero k para el cual (2.27) se satisface recibe el nombre de ***índice*** del vector propio generalizado $\vec{v}$.

Nota
Los vectores propios son vectores propios generalizados de índice igual a 1.

Ejemplo 2.17. El vector $\vec{v}^t = (-\frac{1}{7}, 0)$ es un vector propio generalizado asociado al valor propio $\lambda = -5$ de la siguiente matriz

$$A = \begin{bmatrix} -12 & 7 \\ -7 & 2 \end{bmatrix}.$$

Solución

Veamos que $\vec{v}$, cumple (2.27) para algún valor entero k. Para $k = 1$

$$(A - (-5)I)\vec{v} = \begin{bmatrix} -7 & 7 \\ -7 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{7} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \vec{v}_1.$$

Obsérvese que $\vec{v}_1$ es el vector propio correspondiente a λ. Para $k = 2$, se tiene que

$$(A - (-5)I)^2 \vec{v} = (A - (-5)I)\vec{v}_1 = \begin{bmatrix} -7 & 7 \\ -7 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Luego, $\vec{v}$ es un vector propio generalizado de índice $k = 2$.

Definición 2.14. Espacio propio generalizado

Sea λ_j un valor propio de la matriz real $A \in \mathcal{M}_{nn}$. El subespacio

$$\mathcal{V}_j = \{\vec{v} \in \mathbb{C}^n : (A - \lambda_j I)^k \vec{v} = \vec{0}, \text{para cierto entero positivo } k\} \tag{2.28}$$

se denomina ***espacio propio generalizado*** de A asociado con el valor propio λ_j. En otras palabras, $\mathcal{V}_j =$ espacio nulo$\{(A-\lambda_j I)^k\}$.

Teorema 2.29. Sea A una matriz real de tamaño 2×2 con un único valor propio real λ de multiplicidad algebraica distinta de la multiplicidad geométrica. Entonces existe un ***vector propio generalizado*** $\vec{w}$ que satisface la ecuación

$$(A-\lambda I)\vec{w} = \vec{v}, \tag{2.29}$$

donde $\vec{v}$ es un vector propio correspondiente a λ.

Demostración

Sea $\vec{x} \in \mathbb{C}^2$ un vector fijo, tal que $\vec{x} \neq \alpha\vec{v}$; luego, $\vec{x}$ no es un vector propio de A. Sea

$$\vec{y} = (A-\lambda I)\vec{x}. \tag{2.30}$$

Demostremos que $\vec{y}$ es un vector propio de A; en otras palabras, que $\vec{y} = \beta\vec{v}$. Como $\vec{v}$ y $\vec{x}$ son linealmente independientes y $\vec{y} \in \mathbb{C}^2$; entonces existen constantes c_1 y c_2 tales que

$$\vec{y} = c_1\,\vec{v} + c_2\,\vec{x}. \tag{2.31}$$

Debemos mostrar que $c_2 = 0$. Si se reemplaza (2.30) en (2.31) se tiene que

$$\begin{aligned}(A-\lambda I)\vec{x} =& c_1\,\vec{v} + c_2\,\vec{x}\\ [A-(\lambda+c_2)I]\vec{x} =& c_1\,\vec{v}.\end{aligned}$$

Si se supone que $c_2 \neq 0$, entonces, $\lambda + c_2$ no es un valor propio de A (pues el único valor propio asociado a A es λ). Por lo tanto, $det[A-(\lambda+c_2)I] \neq 0$.

Sea $B = A-(\lambda+c_2)I$; entonces B es no singular. Así, $\vec{x}$ es igual a

$$\vec{x} = B^{-1}\,c_1\,\vec{v} = c_1\,B^{-1}\,\vec{v}. \tag{2.32}$$

Al multiplicar a ambos lados de (2.32) por λ, se obtiene

$$\lambda\,\vec{x} = \lambda\,B^{-1}\,c_1\,\vec{v} = c_1\,B^{-1}\,(\lambda\,\vec{v}) = c_1\,B^{-1}\,(A\,\vec{v}).$$

Pero $A = B + (\lambda+c_2)I$, de manera que

$$\begin{aligned}\lambda\,\vec{x} =& c_1\,B^{-1}\,[B+(\lambda+c_2)I]\,\vec{v}\\ =& c_1\,[I+(\lambda+c_2)B^{-1}]\,\vec{v}\\ =& c_1\,\vec{v} + (\lambda+c_2)\,[c_1\,B^{-1}\,\vec{v}]\\ =& c_1\,\vec{v} + (\lambda+c_2)\,\vec{x}.\end{aligned}$$

La cual se obtiene usando el hecho de que $\vec{x} = c_1\ B^{-1}\ \vec{v}$. Por lo tanto,

$$\begin{aligned}\lambda\,\vec{x} &= c_1\ \vec{v} + \lambda\,\vec{x} + c_2\ \vec{x}\\ \vec{0} &= c_1\ \vec{v} + c_2\ \vec{x}.\end{aligned}$$

Pero como $\vec{x} \neq \alpha\vec{v}$, se debe tener que $c_1 = c_2 = 0$; lo cual contradice la suposición de que $c_2 \neq 0$. Luego, $c_2 = 0$ y sustituyendo en (2.31) se tiene que $\vec{y} = c_1\vec{v}$.

Ahora, debemos mostrar que $c_1 \neq 0$, en otras palabras se debe mostrar que $\vec{y} \neq \vec{0}$; pues si $\vec{y} = \vec{0}$ y se reemplaza en (2.31) se tendría que $\vec{x}$ es un vector propio de A; lo cual contradice la suposición de que $\vec{x} \neq \alpha\vec{v}$. Por lo tanto, $c_1 \neq 0$; luego $\vec{y}$ es un múltiplo no nulo de $\vec{v}$ y por el Teorema 2.1, es un vector propio de A.

Por último definamos $\vec{w} = \frac{1}{c_1}\vec{x}$, entonces

$$(A-\lambda I)\vec{w} = \frac{1}{c_1}(A-\lambda I)\vec{x} = \frac{1}{c_1}\vec{y} = \vec{v}.$$

Esto prueba el teorema.

Ejemplo 2.18. Sea $A = \begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}$. Determine sus vectores propios generalizados.

Solución

La ecuación característica de A es

$$\lambda^2 - 6\lambda + 9 = (\lambda-3)^2 = 0.$$

Luego, $\lambda = 3$ es el único valor propio (de multiplicidad algebraica 2). Entonces,

$$(A-\lambda I)\vec{v} = (A-3I)\vec{v} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto conduce a $x_1 = x_2$. Estableciendo $x_2 = 1$, se obtiene sólo un vector propio linealmente independientes: $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_2$ se calcula $(A-3I)\vec{v}_2 = \vec{v}_1$ y se obtiene

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

La solución general corresponde a $x_1 - x_2 = 1$; luego $x_1 = 1 + x_2$. Por lo tanto, si $x_2 = 0$ se tiene $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Ejemplo 2.19. Encuentre los vectores propios generalizados de la matriz

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

Solución

La ecuación característica de A es

$$-\lambda^3+6\lambda^2-12\lambda+8=-(\lambda-2)^3=0$$

Luego, $\lambda=2$ es el único valor propio (de multiplicidad algebraica 3). Entonces,

$$(A-\lambda I)\vec{v}=(A-2I)\vec{v}=\begin{bmatrix}-1 & 1 & 1\\ 2 & -1 & -1\\ -3 & 2 & 2\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}.$$

Por lo tanto, el vector propio correspondiente se obtiene operando por filas

$$\left[\begin{array}{ccc|c}-1 & 1 & 1 & 0\\ 2 & -1 & -1 & 0\\ -3 & 2 & 2 & 0\end{array}\right]\underset{\substack{F_2+2F_1\\ F_3+F_2}}{\sim}\left[\begin{array}{ccc|c}-1 & 1 & 1 & 0\\ 0 & 1 & 1 & 0\\ -1 & 1 & 1 & 0\end{array}\right]\underset{\substack{F_3-F_1\\ F_1-F_2}}{\sim}\left[\begin{array}{ccc|c}-1 & 0 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{array}\right].$$

Esto conduce a $x=0$ y $y=-z$. Estableciendo $z=-1$, se obtiene el vector propio: $\vec{v}_1=\begin{bmatrix}0\\ 1\\ -1\end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_2$ se calcula $(A-2I)\vec{v}_2=\vec{v}_1$ y se tiene que

$$\begin{bmatrix}-1 & 1 & 1\\ 2 & -1 & -1\\ -3 & 2 & 2\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}0\\ 1\\ -1\end{bmatrix},$$

si se realizan operaciones por filas se obtiene

$$\left[\begin{array}{ccc|c}-1 & 1 & 1 & 0\\ 2 & -1 & -1 & 1\\ -3 & 2 & 2 & -1\end{array}\right]\underset{\substack{F_2+2F_1\\ F_3+F_2}}{\sim}\left[\begin{array}{ccc|c}-1 & 1 & 1 & 0\\ 0 & 1 & 1 & 1\\ -1 & 1 & 1 & 0\end{array}\right]\underset{\substack{F_3-F_1\\ F_1-F_2}}{\sim}\left[\begin{array}{ccc|c}-1 & 0 & 0 & -1\\ 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 0\end{array}\right].$$

Es decir, $x = 1$ y $y = 1 - z$. Si se hace $z = 0$, se obtiene el vector propio generalizado: $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Para encontrar el segundo vector propio generalizado $\vec{v}_3$ se calcula

$$\begin{aligned}(A-2I)\vec{v}_3 &= \vec{v}_2 \\ \begin{bmatrix} -1 & 1 & 1 \\ 2 & -1 & -1 \\ -3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.\end{aligned}$$

Al resolver el sistema por Gauss-Jordan, se tiene que

$$\left[\begin{array}{ccc|c} -1 & 1 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ -3 & 2 & 2 & 0 \end{array}\right] \underset{\substack{F_2+2F_1 \\ F_3+F_2}}{\sim} \left[\begin{array}{ccc|c} -1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 3 \\ -1 & 1 & 1 & 1 \end{array}\right] \underset{\substack{F_3-F_1 \\ F_1-F_2}}{\sim} \left[\begin{array}{ccc|c} -1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Luego, $x = 2$ y $y = 3 - z$. Si $z = 0$, se obtiene el vector propio generalizado:

$$\vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Ejemplo 2.20. Encuentre los vectores propios generalizados de la matriz

$$B = \begin{bmatrix} 1 & 18 & 7 \\ -1 & 13 & 4 \\ 1 & -25 & -8 \end{bmatrix}.$$

Solución

La ecuación característica de B es

$$-\lambda^3 + 6\lambda^2 - 12\lambda + 8 = 0.$$

Luego, $\lambda = 2$ es el único valor propio (de multiplicidad algebraica tres). Entonces,

$$(B-\lambda I)\vec{v} = (B-2I)\vec{v} = \begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

El vector propio correspondiente, se obtiene operando por filas

$$\left[\begin{array}{rrr|r} -1 & 18 & 7 & 0 \\ -1 & 11 & 4 & 0 \\ 1 & -25 & -10 & 0 \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{rrr|r} -1 & 18 & 7 & 0 \\ 0 & -7 & -3 & 0 \\ 0 & -7 & -3 & 0 \end{array}\right]$$
$$\underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{rrr|r} -3 & 5 & 0 & 0 \\ 0 & 7 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Esto conduce a $x = \frac{5}{3}y$ y $z = -\frac{7}{3}y$. Estableciendo $y = 3$, se obtiene el vector propio: $\vec{v}_1 = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_2$ se calcula $(B-2I)\vec{v}_2 = \vec{v}_1$ y se tiene que

$$\begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}.$$

Si se realizan operaciones por filas se obtiene

$$\left[\begin{array}{rrr|r} -1 & 18 & 7 & 5 \\ -1 & 11 & 4 & 3 \\ 1 & -25 & -10 & -7 \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{rrr|r} -1 & 18 & 7 & 5 \\ 0 & -7 & -3 & -2 \\ 0 & -7 & -3 & -2 \end{array}\right]$$
$$\underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{rrr|r} -3 & 5 & 0 & 1 \\ 0 & 7 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Es decir, $x = -\frac{1}{3} + \frac{5}{3}y$ y $z = \frac{2}{3} - \frac{7}{3}y$. Si se hace $y = 0$, se obtiene el vector propio generalizado: $\vec{v}_2 = \frac{1}{3}\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$. Para encontrar el segundo vector propio generalizado $\vec{v}_3$ se calcula

$$(B-2I)\vec{v}_3 = \vec{v}_2$$
$$\begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}.$$

Al resolver el sistema por Gauss-Jordan, se tiene que

$$\left[\begin{array}{ccc|c} -1 & 18 & 7 & -\frac{1}{3} \\ -1 & 11 & 4 & 0 \\ 1 & -25 & -10 & \frac{2}{3} \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{ccc|c} -1 & 18 & 7 & -\frac{1}{3} \\ 0 & -7 & -3 & \frac{1}{3} \\ 0 & -7 & -3 & \frac{1}{3} \end{array}\right]$$
$$\underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{ccc|c} -3 & 5 & 0 & \frac{4}{3} \\ 0 & 7 & 3 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Luego, $x = -\frac{4}{9} + \frac{5}{3}y$ y $z = -\frac{1}{9} - \frac{7}{3}y$. Tomando $y = 0$, se obtiene el vector propio generalizado: $\vec{v}_3 = -\frac{1}{9}\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$.

Teorema 2.30. Sea $\vec{v} \neq \vec{0}$ un vector propio generalizado con índice k de una matriz real A de tamaño $n \times n$. Entonces,

$$\left\{\vec{v}, (A-\lambda_j I_n)\vec{v}, \ldots, (A-\lambda_j I_n)^{k-1}\vec{v}\right\} \tag{2.33}$$

es un conjunto de vectores linealmente independientes.

Demostración

Supongamos que $\vec{v} \in \mathbb{C}^n$ es de índice k y que los vectores dados en (2.33) son linealmente dependientes, entonces existen constantes $c_n \neq 0$, tales que

$$\sum_{n=0}^{k-1} c_n \, (A-\lambda_j I_n)^n \vec{v} = \vec{0}, \tag{2.34}$$

donde $(A-\lambda_j I_n)^0 = I_n$. Sea $A_j = A - \lambda_j I_n$ y consideremos el polinomio[1] $f(t) = c_0 + c_1 t + \ldots + c_r t^r$ diferente de cero de grado $r \leq k-1$. De este modo, (2.34) se puede expresar de la manera siguiente

$$f(A_j)\,\vec{v} = \vec{0}.$$

Sea $g(t) = t^k$, así que $g(A_j)\,\vec{v} = \vec{0}$. Si $h(t) = t^d$ con $d \leq k-1$, es el máximo común divisor de $f(t)$ y $g(t)$, usando el Algoritmo de Euclides, el polinomio $h(t)$ se puede escribir como

$$h(t) = h_1(t) f(t) + h_2(t) g(t),$$

[1]Si el lector no está familiarizado con este concepto puede ver Lang (1976), Cap. 9

donde $h_1(t)$ y $h_2(t)$ son polinomios distintos de cero. Luego,

$$h(A_j)\vec{v} = \vec{0}.$$

Por lo tanto, d sería el índice de $\vec{v}$; lo cual contradice la hipótesis de que k es el índice de $\vec{v}$ y se concluye la prueba.

Teorema 2.31. Los vectores no nulos tomados de espacios propios generalizados distintos son linealmente independientes.

Demostración

Sea $A_j = A - \lambda_j I_n$ y $\mathcal{V}_j =$ espacio nulo $\left\{A_j^{k_j}\right\}$ para algún entero k_j, $j = 1, 2, \ldots, r$. Sean $\vec{v}_1, \ldots, \vec{v}_r$, con $\vec{v}_j \in \mathcal{V}_j$ y supongamos que

$$\vec{w} = \sum_{j=1}^{r} \vec{v}_j = \vec{0}. \tag{2.35}$$

Se debe mostrar que cada $\vec{v}_j = \vec{0}$. Si se multiplica ambos lados de (2.35) por la matriz $C = A_2^{k_2} A_3^{k_3} \ldots A_r^{k_r}$ y se utilizan los hechos de que $A_j^{k_j}\vec{v}_j = \vec{0}$ y que $A_i^{k_i} A_j^{k_j} = A_j^{k_j} A_i^{k_i}$, lo cual se tiene ya que

$$\begin{aligned} A_i A_j &= (A - \lambda_i I_n)(A - \lambda_j I_n) = A^2 - \lambda_i A - \lambda_j A + \lambda_i \lambda_j I_n \\ &= A^2 - \lambda_j A - \lambda_i A + \lambda_i \lambda_j I_n = (A - \lambda_j I_n)(A - \lambda_i I_n) = A_j A_i. \end{aligned}$$

Se obtiene que

$$C\vec{w} = C\vec{v}_1 = \vec{0}. \tag{2.36}$$

Por lo tanto, $\vec{v}_1 = \vec{0}$. De manera análoga, todos los $\vec{v}_j$ restantes tienen que desaparecer.

De acuerdo con la definición 2.7, las matrices reales cuadradas A y B se dicen que son *semejantes*, si existe una matriz P no singular, tal que

$$A = PBP^{-1}.$$

En ocasiones, además de establecer el hecho en sí de la semejanza, se requiere encontrar la matriz P de la transformación que satisface que $A = PBP^{-1}$. En estos momentos, se puede construir la matriz P utilizando los vectores propios generalizados de la siguiente manera

$$P = SR^{-1} \tag{2.37}$$

donde las columnas de la matriz S son los vectores propios generalizados de la matriz A y las columnas de la matriz R son los vectores propios generalizados de la matriz B.

Ejemplo 2.21. Determine si las matrices dadas en los ejemplos 2.19 y 2.20 son semejantes.

Solución

Como las dos matrices tienen el mismo polinomio característico entonces son semejantes, encontremos la matriz P. En el Ejemplo 2.19 se obtuvieron los siguientes vectores propios generalizados para la matriz A

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Del Ejemplo 2.20 se tiene que los vectores propios generalizados de la matriz B, son

$$\vec{v}_1 = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} -\frac{4}{9} \\ 0 \\ -\frac{1}{9} \end{bmatrix}.$$

Por lo tanto,

$$\begin{aligned} P &= \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & -\frac{1}{3} & -\frac{4}{9} \\ 3 & 0 & 0 \\ -7 & \frac{2}{3} & -\frac{1}{9} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{11}{3} & \frac{4}{3} \\ -2 & 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{13}{3} & \frac{17}{3} & -\frac{2}{3} \\ -\frac{19}{3} & 7 & -\frac{5}{3} \\ 0 & -\frac{1}{3} & 0 \end{bmatrix}. \end{aligned}$$

El lector puede verificar que $A = PBP^{-1}$.

Teorema 2.32. Todo vector en $\mathbb{R}^n$ es una combinación lineal de vectores de los espacios propios generalizados $\mathcal{V}_j$.

Demostración

Supóngase que $\mathbb{V}$ es el subespacio de $\mathbb{R}^n$ formado por los vectores de la forma $\vec{u}_1 + \ldots + \vec{u}_r$, donde $\vec{u}_j \in \mathcal{V}_j$. Se necesita probar que $\mathbb{V} = \mathbb{R}^n$. Supongamos que $\mathbb{V}$ es un subespacio adecuado. Entonces se escoge una base $\{\vec{v}_1 + \ldots + \vec{v}_s\}$ de $\mathbb{V}$ y se extiende este conjunto a una base $\mathcal{B}$ de $\mathbb{R}^n$. En esta base la matriz $[A]_{\mathcal{B}}$ está particionada como sigue

$$[A]_{\mathcal{B}} = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix}$$

donde A_{22} es una matriz de tamaño $(n-s)\times(n-s)$. Los valores propios de A_{22} son valores propios de A. Como todos los valores propios distintos y vectores propios de A son considerados en $\mathbb{V}$ (es decir, en A_{11}), se tiene una contradicción. Por lo tanto, $\mathbb{V}=\mathbb{R}^n$, como se quería demostrar.

Teorema 2.33. Sea $\mathfrak{V}_{\mathrm{j}}$ una base para el espacio propio generalizado $\mathcal{V}_j$ y sea $\mathfrak{V}$ la unión de los conjuntos $\mathfrak{V}_{\mathrm{j}}$. Entonces, $\mathfrak{V}$ es una base para $\mathbb{R}^n$.

Demostración

Vamos a demostrar que los vectores en $\mathfrak{V}$ generan $\mathbb{R}^n$. Por el Teorema 2.32 se tiene que todo vector en $\mathbb{R}^n$ es una combinación lineal de vectores en $\mathcal{V}_j$. Pero cada vector en $\mathcal{V}_j$ es una combinación lineal de vectores en $\mathfrak{V}_{\mathrm{j}}$. Por lo tanto, los vectores en $\mathfrak{V}$ generan $\mathbb{R}^n$.

Ahora, demostremos que los vectores en $\mathfrak{V}$ son linealmente independientes. Supongamos que una combinación lineal de vectores en $\mathfrak{V}$ suma $\vec{0}$. Es posible escribir esta suma como

$$\vec{v}_1+\ldots+\vec{v}_r=\vec{0}$$

donde $\vec{v}_j$ es la combinación lineal de vectores en $\mathfrak{V}_{\mathrm{j}}$. El Teorema 2.31 indica que cada $\vec{v}_j=\vec{0}$. Como $\mathfrak{V}_{\mathrm{j}}$ es una base para $\mathcal{V}_j$, se deduce que los coeficientes de las combinaciones lineales $\vec{v}_j$ deben ser todos cero. Por lo tanto, los vectores en $\mathfrak{V}$ son linealmente independientes.

Por el Teorema 1.26 se deduce que $\mathfrak{V}$ es una base de $\mathbb{R}^n$.

Ejercicios 2.5.

1. Encuentre la matriz P que satisface que $A=PBP^{-1}$ para

$$A=\begin{bmatrix}3&1&-1\\2&2&-1\\2&2&0\end{bmatrix}\qquad \text{y} \qquad B=\begin{bmatrix}42&130&25\\-8&-24&-5\\-23&-73&-13\end{bmatrix}.$$

2. Determine los vectores propios generalizados para las siguientes matrices

$$a.\ \begin{bmatrix}-12&7\\-7&2\end{bmatrix}. \qquad b.\ \begin{bmatrix}-10&-7\\7&4\end{bmatrix}. \qquad c.\ \begin{bmatrix}4&-1\\1&2\end{bmatrix}.$$

$$d.\begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix}. \qquad e.\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ -2 & -3 & -2 \end{bmatrix}.$$

2.6. Métodos iterativos para estimar valores propios y vectores propios

Hasta ahora para encontrar los valores propios de una matriz $A = [a_{ij}]$ se resuelve la ecuación característica asociada. En muchos problemas prácticos, obtener las raíces correspondientes no es sencillo. Es más, en algunos problemas estadísticos sólo se necesita el valor propio con el valor absoluto más grande. En esta sección se trataran algunos procedimientos para calcular valores aproximados de los valores propios de una matriz.

Definición 2.15. Valor propio dominante y vector propio dominante

La matriz A de tamaño $n \times n$ tiene un valor propio dominante si su valor absoluto es mayor que los valores absolutos de los valores propios restantes. El vector propio asociado al valor propio dominante se denomina vector propio dominante.

Ejemplo 2.22. Determine el valor propio dominante para la matriz

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

Solución

Los valores propios asociados a A son -1 y 6. Por lo tanto, el valor propio dominante es 6.

Ejemplo 2.23. Determine el valor propio dominante de la matriz

$$A = \begin{bmatrix} 4 & 2 & -2 \\ 8 & 10 & -14 \\ 8 & 8 & -12 \end{bmatrix}.$$

Solución

Los valores propios asociados a la matriz A son 2, -4 y 4. Por lo tanto, no hay valor propio dominante.

2.6.1. Método de la potencia

El método de potencias para aproximar valores propios es iterativo. Primero se supone que la matriz A tiene un valor propio dominante con vectores propios dominantes. Luego se elige un vector diferente de cero $\vec{w}_1 \in \mathbb{R}^n$. Por último, se forma la sucesión definida por

$$\begin{aligned} \vec{w}_2 =& A\vec{w}_1 \\ \vec{w}_3 =& A\vec{w}_2 = A(A\vec{w}_1) = A^2\vec{w}_1 \\ \vec{w}_4 =& A\vec{w}_3 = A(A^2\vec{w}_1) = A^3\vec{w}_1 \\ &\vdots \\ \vec{w}_{k+1} =& A\vec{w}_k = A(A^{k-1}\vec{w}_1) = A^k\vec{w}_1 \end{aligned}$$

A medida que k crece, $A^k\vec{w}_1$ se hace paralelo al vector propio dominante de A.

Teorema 2.34. Sea A una matriz real diagonalizable de tamaño $n \times n$ con valores propios reales $\lambda_1, \lambda_2, \ldots, \lambda_n$, tales que

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|.$$

Entonces existe un vector $\vec{w}_1$ diferente de cero de $\mathbb{R}^n$ tal que la sucesión de vectores definida por

$$A\vec{w}_1, A^2\vec{w}_1, A^3\vec{w}_1, \ldots, A^k\vec{w}_1, \ldots$$

se aproxima al vector propio dominante de A cuando k aumenta.

Demostración

Como A es diagonalizable entonces existe una base de $\mathbb{R}^n$ formada por los n vectores propios $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, asociados a los valores propios λ_i, $i = 1, 2, \ldots, n$, respectivamente.

Sea $\vec{w}_1$ cualquier vector distinto de cero de $\mathbb{R}^n$, de forma que

$$\vec{w}_1 = \sum_{i=1}^{n} c_i \vec{v}_i \qquad \text{con } c_i \in \mathbb{R}. \tag{2.38}$$

Definamos el siguiente proceso iterativo

$$\vec{w}_{k+1} = A\,\vec{w}_k, \qquad k = 1, 2, \ldots$$

Nótese que $\vec{w}_{k+1} = A^k\,\vec{w}_1,\ k = 1, 2, \ldots$ Por lo tanto, se tiene que

$$\begin{aligned}
\vec{w}_2 = A\,\vec{w}_1 &= A\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) \\
&= \sum_{i=1}^{n} c_i \lambda_i \vec{v}_i \\
&= \lambda_1 \left(c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \frac{\lambda_i}{\lambda_1} \vec{v}_i\right), \\
\vec{w}_3 = A\,\vec{w}_2 &= A\left[\lambda_1 \left(c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \frac{\lambda_i}{\lambda_1} \vec{v}_i\right)\right] \\
&= \lambda_1^2 \left(c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^2 \vec{v}_i\right),
\end{aligned}$$

en general, por recurrencia, se obtiene

$$\vec{w}_{k+1} = A\,\vec{w}_k = \lambda_1^k \left(c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \vec{v}_i\right).$$

Luego, con base en la hipótesis original de que λ_1 es mayor, en valor absoluto, que los demás valores propios se concluye que cuando k tiende a infinito cada una de las fracciones $\left(\frac{\lambda_i}{\lambda_1}\right)^k$ para $i > 1$, tiende a cero, pues $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$. Esto implica que la aproximación

$$\vec{w}_{k+1} \cong \lambda_1^k c_1 \vec{v}_1, \qquad c_1 \neq 0,$$

mejora a medida que k es suficientemente grande. Como $\vec{v}_1$ es el vector propio asociado a λ_1 entonces es dominante; luego, cualquier múltiplo escalar de $\vec{v}_1$ también es un vector propio dominante. Así se ha demostrado que $A^k\vec{w}_1$ se aproxima arbitrariamente al vector propio dominante cuando k crece.

Como las componentes de $A^k\vec{w}_1$ pueden ser números demasiados grandes al aumentar k, lo cual conduciría a un error de redondeo. Este problema se evita multiplicando $A^k\vec{w}_1$ por un escalar adecuado en cada iteración.

A continuación se presenta un procedimiento para obtener el valor propio dominante de una matriz A.

Cálculo del valor propio dominante de A

i) Seleccione un vector arbitrario diferente de cero $\vec{w}_1$, cuya entrada más grande sea 1.

ii) Para $k = 1, 2, \ldots,$

   a) Calcule $A\vec{w}_k$.
   b) Sea μ_k la componente de $A\vec{w}_k$ con valor absoluto más grande.
   c) Evalúe $\vec{w}_{k+1} = \frac{1}{\mu_k}A\vec{w}_k$

iii) Para casi todas las escogencias de $\vec{w}_1$, la sucesión $\{\mu_k\}$ se aproxima al valor propio dominante y la sucesión $\{\vec{w}_k\}$ se aproxima al correspondiente vector propio.

Con esta metodología no hay reglas eficaces y rápidas para determinar cuántas iteraciones se deben realizar. Además, si se escoge el vector $\vec{w}_1$ de manera que en 2.38 el coeficiente c_1 sea cero el método falla.

Ejemplo 2.24. Ilustrar el método de la potencia para la matriz

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \qquad \text{comenzando con} \qquad \vec{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Solución

Si se aplica el método de la potencia de tal forma que en los resultados de cada no se utilicen cifras decimales y de esa manera nos evitemos el redondeo, se obtiene la tabla siguiente

k	1	2	3	4	5	6	7
$\vec{w}_k$	$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}\frac{3}{7}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{17}{33}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{83}{167}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{417}{833}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{2083}{4167}\\1\end{bmatrix}$	$\begin{bmatrix}0,5\\1\end{bmatrix}$
$A\vec{w}_k$	$\begin{bmatrix}3\\7\end{bmatrix}$	$\begin{bmatrix}\frac{17}{7}\\\frac{33}{7}\end{bmatrix}$	$\begin{bmatrix}\frac{83}{33}\\\frac{167}{33}\end{bmatrix}$	$\begin{bmatrix}\frac{417}{167}\\\frac{833}{167}\end{bmatrix}$	$\begin{bmatrix}\frac{2083}{833}\\\frac{4167}{833}\end{bmatrix}$	$\begin{bmatrix}\frac{10417}{4167}\\\frac{20833}{4167}\end{bmatrix}$	$\begin{bmatrix}2,5\\5,0\end{bmatrix}$
λ_k	-	7	$\frac{33}{7}$	$\frac{167}{33}$	$\frac{833}{167}$	$\frac{4167}{833}$	4.9995

Luego, el valor propio dominante es aproximadamente 4,9995 y el correspondiente vector propio es $\begin{bmatrix}0,5\\1\end{bmatrix}$.
Las respuestas exactas son

$$\lambda_1 = 5 \qquad \text{y} \qquad \vec{v}_1 = \begin{bmatrix}0,5\\1\end{bmatrix}.$$

Definición 2.16. Cociente de Rayleigh

Sea A una matriz real diagonalizable de tamaño $n \times n$ se llama cociente de Rayleigh de A a la función real definida para cada $\vec{x} \neq \vec{0}$ como

$$r_A(\vec{x}) = \frac{\vec{x}^t A \vec{x}}{\vec{x}^t \vec{x}}. \tag{2.39}$$

Aunque el cociente de Rayleight depende de la matriz, el subíndice A de r se omite si no hay confusión

Teorema 2.35. Sea A una matriz real diagonalizable de tamaño $n \times n$. Sea $\vec{w}_1 \in \mathbb{R}^n$ cualquier vector no nulo. Considérese los cocientes de Rayleigh

$$r(\vec{w}_k) = \frac{\vec{w}_k^t A \vec{w}_k}{\vec{w}_k^t \vec{w}_k} \qquad \text{para } k = 1, 2, \ldots, m$$

donde m es la cantidad deseada de iteraciones. El último cociente $r(\vec{w}_m)$ es una aproximación del valor propio dominante λ de A y, si se hace $r(\vec{w}_k) = \lambda + \varepsilon$, de modo que ε es el error de $r(\vec{w}_k)$, entonces,

$$|\varepsilon| \leq \sqrt{\frac{\vec{y}_k^t\,\vec{y}_k}{\vec{w}_k^t\,\vec{w}_k} - r^2(\vec{w}_k)}, \tag{2.40}$$

donde $\vec{y}_k = A\vec{w}_k$.

Demostración

Si se reescribe el cociente de Rayleigh, se obtiene que

$$\vec{w}_k^t A\vec{w}_k = r(\vec{w}_k)\,\vec{w}_k^t\,\vec{w}_k \qquad\qquad k = 1, 2, \ldots, m$$

y dado que $\vec{y}_k = A\vec{w}_k$, se tiene que $\vec{w}_k^t\vec{y}_k = r(\vec{w}_k)\,\vec{w}_k^t\vec{w}_k$. Por lo tanto

$$\begin{aligned}
[\vec{y}_k - r(\vec{w}_k)\,\vec{w}_k]^t\,[\vec{y}_k - r(\vec{w}_k)\,\vec{w}_k] &= \vec{y}_k^t\vec{y}_k - 2r(\vec{w}_k)\,\vec{w}_k^t\vec{y}_k + r^2(\vec{w}_k)\,\vec{w}_k^t\vec{w}_k\\
&= \vec{y}_k^t\vec{y}_k - r^2(\vec{w}_k)\,\vec{w}_k^t\vec{w}_k\\
&= \left(\frac{\vec{y}_k^t\vec{y}_k}{\vec{w}_k^t\vec{w}_k} - r^2(\vec{w}_k)\right)\vec{w}_k^t\vec{w}_k = \delta^2\vec{w}_k^t\vec{w}_k.
\end{aligned} \tag{2.41}$$

Como A es una matriz diagonalizable, por el Teorema 2.16, tiene n vectores propios linealmente independientes $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ correspondientes a los valores propios $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, respectivamente y mediante el proceso de Gram-Schmidt se ortonormalizan estos vectores para obtener $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{R}^n$. Entonces, $\vec{w}_k$ tiene una representación de la forma

$$\vec{w}_k = \sum_{i=1}^{n} a_i\vec{v}_i \qquad\qquad \text{con } a_i \in \mathbb{R}.$$

y como los $\vec{v}_i$ son vectores unitarios ortogonales, entonces

$$\vec{w}_k^t\vec{w}_k = \sum_{i=1}^{n} a_i^2. \tag{2.42}$$

Ahora bien,

$$\vec{y}_k = A\vec{w}_k = A\left(\sum_{i=1}^{n} a_i\vec{v}_i\right) = \sum_{i=1}^{n} a_i\lambda_i.$$

Luego,

$$\vec{y}_k - r(\vec{w}_k)\vec{w}_k = \sum_{i=1}^{n} a_i(\lambda_i - r(\vec{w}_k))\vec{v}_i.$$

Si se reemplaza en (2.41), se obtiene que

$$\delta^2 \vec{w}_k^t \vec{w}_k = \sum_{i=1}^{n} a_i^2 (\lambda_i - r(\vec{w}_k))^2.$$

Si se sustituye cada $(\lambda_i - r(\vec{w}_k))^2$ por el menor de estos términos y se aplica (2.42), se tiene que

$$\delta^2 \vec{w}_k^t \vec{w}_k \geq (\lambda_c - r(\vec{w}_k))^2 \sum_{i=1}^{n} a_i^2 = (\lambda_c - r(\vec{w}_k))^2 \vec{w}_k^t \vec{w}_k,$$

donde λ_c es un valor propio al cual $r(\vec{w}_k)$ está próximo. De esto se llega a (2.40) y queda demostrado el teorema.

Método de los cocientes de Rayleigh
Sea A una matriz diagonalizable de tamaño $n \times n$, con un valor propio dominante. Sea m la cantidad deseada de iteraciones. i) Seleccione un vector arbitrario diferente de cero $\vec{w}_0$. ii) Para $k = 0, 1, \ldots, m-1$ a) Calcule $\vec{z}_k = \dfrac{\vec{w}_k}{\lVert\vec{w}_k\rVert}$. b) Sea $\vec{w}_{k+1} = A\vec{z}_k$. c) Evalúe $r(\vec{z}_k) = \vec{z}_k^t \vec{w}_{k+1}$ iii) Los cocientes de Rayleigh $\{r(\vec{z}_k)\}$ se aproximan al valor propio dominante y la sucesión $\{\vec{z}_k\}$ se aproxima al correspondiente vector propio unitario.

Para matrices simétricas, este método es muy eficiente y requiere menos iteraciones para lograr la misma exactitud.

Ejemplo 2.25. Ilustrar el método de los cocientes de Rayleigh para la matriz dada en el Ejemplo 2.24

Solución

Aplicando el método de los cocientes de Rayleigh se obtiene la tabla siguiente

k	0	1	2	3	4	5	6
$\vec{z}_k$	$\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$	$\begin{bmatrix}\frac{3}{\sqrt{58}}\\ \frac{7}{\sqrt{58}}\end{bmatrix}$	$\begin{bmatrix}\frac{17}{\sqrt{1378}}\\ \frac{33}{\sqrt{1378}}\end{bmatrix}$	$\begin{bmatrix}\frac{83}{\sqrt{34778}}\\ \frac{167}{\sqrt{34778}}\end{bmatrix}$	$\begin{bmatrix}0{,}448\\ 0{,}894\end{bmatrix}$	$\begin{bmatrix}0{,}447\\ 0{,}894\end{bmatrix}$	$\begin{bmatrix}0{,}447\\ 0{,}894\end{bmatrix}$
$A\vec{z}_k$	$\begin{bmatrix}\frac{3}{\sqrt{2}}\\ \frac{7}{\sqrt{2}}\end{bmatrix}$	$\begin{bmatrix}\frac{17}{\sqrt{58}}\\ \frac{33}{\sqrt{58}}\end{bmatrix}$	$\begin{bmatrix}\frac{83}{\sqrt{1378}}\\ \frac{167}{\sqrt{1378}}\end{bmatrix}$	$\begin{bmatrix}\frac{417}{\sqrt{34778}}\\ \frac{833}{\sqrt{34778}}\end{bmatrix}$	$\begin{bmatrix}2{,}236\\ 4{,}473\end{bmatrix}$	$\begin{bmatrix}2{,}236\\ 4{,}472\end{bmatrix}$	$\begin{bmatrix}2{,}236\\ 4{,}472\end{bmatrix}$
$r(\vec{w}_k)$	$-$	$\frac{141}{29}$	$\frac{3461}{689}$	$\frac{86861}{17389}$	$\frac{2169861}{433889}$	4.9998	5

Así, el valor propio dominante es aproximadamente $\lambda = 5$ y el correspondiente vector propio unitario es $\begin{bmatrix}0{,}447\\ 0{,}894\end{bmatrix}$.

Ejercicios 2.6.

Determine los vectores propios dominantes con los métodos descritos en esta sección para las siguientes matrices

$a.\ \begin{bmatrix}1 & 1\\ 1 & 3\end{bmatrix}$. $\quad b.\ \begin{bmatrix}1 & 3\\ 3 & -1\end{bmatrix}$. $\quad c.\ \begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$.

$d.\ \begin{bmatrix}1 & 3 & -2\\ 3 & 1 & 2\\ -2 & 2 & 1\end{bmatrix}$. $\quad e.\ \begin{bmatrix}3 & -1 & 4\\ -1 & 1 & -5\\ 4 & -5 & 2\end{bmatrix}$.

Capítulo 3

Descomposición de matrices

Una factorización de una matriz A es una ecuación que expresa a A como un producto de dos o más matrices. Por ejemplo, los teoremas de diagonalización dados en 2.16 y 2.22 son algunos casos de descomposición de una matriz. Estas descomposiciones son de interés especial cuando algunos de los factores son matrices ortogonales; la razón es que las transformaciones ortogonales preservan normas y ángulos. Desafortunadamente, como sabemos, no todas las matrices pueden ser factorizadas como $A = PDP^{-1}$ con D diagonal. Sin embargo, para cualquier matriz A es posible obtener una de las factorizaciones que se presentan en este capítulo; las cuales son importantes desde el punto de vista teórico, práctico y numérico.

3.1. Triangularización de una matriz

En esta sección nos centraremos en el estudio de diversas factorizaciones de una matriz A como producto de matrices triangulares.

Teorema 3.1. Sea A una matriz real de tamaño $m \times n$, entonces existe una matriz L triangular inferior no singular de tamaño $m \times m$, tal que

$$A = LS, \tag{3.1}$$

donde S es la matriz escalonada de A de tamaño $m \times n$, obtenida sin intercambio de filas.

Demostración

Sea $A = [a_{kl}]$, empecemos con la primera fila no nula de A. Asumamos, sin pérdida de generalidad que la primera fila de A es no nula. Sea a_{1j} el primer elemento no nulo en la primera fila. Tomemos cualquier a_{ij} con $2 \leq i \leq m$; si $a_{ij} = 0$, no se hace nada; si $a_{ij} \neq 0$, se multiplica la primer fila por $-a_{ij}/a_{1j}$ y se suma a la i-ésima fila. Esta operación hace los elementos (i, j) cero. Dicha operación es equivalente a premultiplicar a A por la matriz elemental $E_{i1}(-a_{ij}/a_{1j})$ la cual es una matriz triangular inferior. Así hemos usado el elemento $(1, j)$, es decir a_{1j}, como un pivote para eliminar todos los otros elementos de la j-ésima columna. La matriz resultante o matriz reducida es obtenida matemáticamente premultiplicando a A sucesivamente por un número finito de matrices triangulares inferiores cuyo producto es también una matriz triangular inferior. Ahora continuemos con la matriz reducida, tomemos la segunda fila. Si todos sus elementos son iguales a cero, se pasa a la tercera fila. Si no se encuentra cualquier vector fila no nulo entre la segunda, tercera, $\ldots, m$-ésima fila, el proceso termina. La matriz reducida es claramente una forma escalonada. En otro caso, localice el primer vector no nulo entre las $m-1$ filas de la matriz reducida empezando desde la segunda. Repita el proceso de eliminar todas las entradas debajo del primer elemento no nulo (pivote) del vector escogido no nulo. Repita este proceso hasta que no pueda encontrar ningún otro vector no nulo en la matriz reducida. La matriz reducida es claramente una forma escalonada. La matriz S es simplemente el producto de todas las matrices triangulares inferiores empleadas durante el proceso. Claramente, S es una matriz triangular inferior no singular. Esto completa la prueba.

Definición 3.1. Descomposición *LS*

Una factorización como la indicada en (3.1), es decir, como el producto de una matriz no singular triangular inferior L y una forma escalonada S, si existe, se llama **descomposición** *LS* de la matriz A.

Ejemplo 3.1. Hallar una factorización *LS* de la matriz

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

Solución Se procede en dos columnas como sigue:

Reducción de A a S		Creación de L a partir de I
$A = \begin{bmatrix} 2 & 4 & -2 & 0 \\ 1 & 1 & -3 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$		$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	Dividir la fila 1 por 2	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 1 & -3 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	Sumar la fila 1 multiplicada por -1 a la fila 2	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & -1 & -2 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	Sumar la fila 1 multiplicada por -2 a la fila 3	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & -1 & -2 & 1 \\ 0 & -1 & -2 & 4 \end{bmatrix}$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$
	Multiplicar la fila 2 por -1	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & -2 & 4 \end{bmatrix}$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$
	Sumar una vez la fila 2 a la fila 3	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 1 \end{bmatrix}$
	Dividir la fila 3 por 3	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = S$		$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 3 \end{bmatrix} = L$

El lector puede verificar que $A = LS$.

Teorema 3.2. Sea A una matriz real de tamaño $m \times n$ cuya forma escalonada puede hallarse sin intercambio de filas. Entonces, A tiene una única factorización LS si y sólo si $\rho(A) = m$.

Demostración

Supongamos que $\rho(A) = m$ y que $A = LS$ y $A = MV$, siendo L y M matrices triangulares inferiores, S y V en forma escalonada.

Multiplicando por la izquierda, la igualdad

$$LS = MV$$

por M^{-1}, se obtiene que

$$M^{-1}LS = V. \tag{3.2}$$

Sea $N = M^{-1}L$. Si se logra probar que $N = I$, se tendrá que $L = M$ y sustituyendo en (3.2) se obtendrá $S = V$.

Se procede por inducción sobre m. Si $m = 1$, entonces

$$A = \begin{bmatrix} 0 & \ldots & a_{1,r+1} & \ldots & a_{1,n} \end{bmatrix}$$

y la única factorización posible es

$$A = [a_{1,r+1}] \begin{bmatrix} 0 & \ldots & 1 & \ldots & \frac{a_{1,n}}{a_{1,r+1}} \end{bmatrix}.$$

Supongamos que el resultado es cierto para matrices con $m-1$ filas. Si la primera columna $\vec{s}_1$, de S fuera nula, la primera columna $\vec{v}_1$ de V también lo sería ya que

$$\vec{v}_1 = N\vec{s}_1 = N\vec{0} = \vec{0}$$

y recíprocamente, por ser N no singular. Se puede por lo tanto suponer, sin pérdida de generalidad, que las primeras columnas de S y V son no nulas. Entonces la primera columna tiene un uno principal en la primera fila. Nótese que N es triangular inferior por ser producto de matrices triangulares inferiores. Si se reescribe el producto $NS = V$ particionando N en submatrices $[1 + (m-1)] \times [1 + (m-1)]$ y las matrices S y V en submatrices $[1 + (m-1)] \times [1 + (n-1)]$, se tiene que

$$NS = \begin{bmatrix} a & O \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} 1 & S_{12} \\ O & S_{22} \end{bmatrix} = \begin{bmatrix} 1 & V_{12} \\ O & V_{22} \end{bmatrix},$$

de donde

$$\begin{aligned} a &= 1, & aS_{12} &= V_{12}, \\ N_{21} &= O, & N_{22}S_{22} &= V_{22}. \end{aligned}$$

Ahora bien N_{22} es una submatriz real de tamaño $(m-1)\times(m-1)$, que es triangular inferior y no singular, S_{22} y V_{22} son submatrices en forma escalonada. Por la hipótesis de inducción

$$N_{22} = I_{m-1}$$

y por lo tanto

$$N = \begin{bmatrix} 1 & O \\ O & I_{m-1} \end{bmatrix} = I_m.$$

Si se supone que $r = \rho(A) < m$ aplicando el algoritmo se obtiene la factorización

$$A = \begin{bmatrix} L_{11} & O \\ L_{21} & I_{m-r} \end{bmatrix} \begin{bmatrix} S_{11} \\ O \end{bmatrix}.$$

Pero es obvio que si L_{22} es una submatriz triangular inferior no singular de tamaño $(m-r)\times(m-r)$ cualquiera, también se puede escribir

$$A = \begin{bmatrix} L_{11} & O \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} S_{11} \\ O \end{bmatrix}$$

y la factorización no es única.

Corolario 3.2.1. Sea A una matriz no singular de tamaño $n\times n$ cuya forma escalonada puede hallarse sin intercambio de filas. Entonces A tiene una única factorización LS.

Demostración

Supongamos que $A = L_1S_1$ y $A = L_2S_2$ son dos de dichas factorizaciones. Nótese que tanto L_1^{-1} como L_2^{-1} también son triangulares inferiores y S_1^{-1} y S_2^{-1} son triangulares superiores, que además tienen unos en la diagonal principal por ser A no singular.

Ahora bien, de $L_1S_1 = L_2S_2$ se obtiene que

$$L_2^{-1}L_1 = S_2S_1^{-1}.$$

Vemos fácilmente que $L_2^{-1}L_1$ es triangular inferior, por ser producto de triangulares inferiores y $S_2S_1^{-1}$ es triangular superior, por ser producto de triangulares superiores. Como son iguales se concluye que el producto debe ser diagonal. Además S_2 y S_1^{-1} tienen unos en la diagonal principal y por lo tanto, $S_2S_1^{-1}$ también tiene unos en la diagonal principal.

En definitiva $L_2^{-1}L_1 = S_2S_1^{-1} = I_n$ de donde se deduce la unicidad.

Teorema 3.3. Si no ocurren intercambios de filas durante la reducción de una matriz A de tamaño $m \times n$ a una matriz escalonada S, entonces A puede ser factorizada como

$$A = LDU, \tag{3.3}$$

en donde L es triangular inferior de tamaño $m \times m$, D es la matriz de pivotes de tamaño $m \times m$ y U es una matriz escalonada.

Demostración

Sea

$$S = E_k E_{k-1} \dots E_2 E_1 A$$

la forma escalonada de A, donde E_i, $i = 1, 2, \dots, k$ son matrices elementales de tipo R_1 y R_2, puesto que no se han realizado intercambio de filas. Entonces resulta que

$$A = E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1} S = LS.$$

Después de que se determina $A = LS$, se continúa factorizando S como

$$S = DD^{-1}S = DU,$$

en donde D es la matriz diagonal de pivotes cuyo elemento diagonal en la p-ésima fila es 1 si la p-ésima fila de S es 0 y es a si a es el primer elemento no nulo de la p-ésima fila de S. La matriz escalonada $U = D^{-1}S$. Entonces se puede reescribir el producto $A = LS$ como

$$A = LDU = MU,$$

en donde $M = LD$. Esto prueba el teorema.

Ejemplo 3.2.

Hallar la factorización LDU de la matriz dada en el Ejemplo 3.1

Solución

Del Ejemplo 3.1, se tiene que

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

Los pivotes eran 2, -1 y 3. Luego su factorización LDU es

$$\begin{bmatrix}1 & 0 & 0\\ \frac{1}{2} & 1 & 0\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}2 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 3\end{bmatrix}\begin{bmatrix}1 & 2 & -1 & 0\\ 0 & 1 & 2 & -1\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

Teorema 3.4. Factorización única

Sea A una matriz no singular de tamaño $n \times n$. Una factorización de la forma (3.3) está determinada de manera única, si

1. L es triangular inferior con los elementos en la diagonal iguales a 1,

2. U es triangular superior con los elementos en la diagonal iguales a 1,

3. D es diagonal sin ceros en su diagonal principal.

Demostración

Supongamos que $A = L_1D_1U_1$ y $A = L_2D_2U_2$ son dos factorizaciones de A distintas. Nótese que tanto L_1^{-1} como L_2^{-1} también son triangulares inferiores y U_1^{-1} y U_2^{-1} son triangulares superiores, que además tienen unos en la diagonal principal por ser A no singular.

Ahora bien, de $L_1D_1U_1 = L_2D_2U_2$ se obtiene que

$$U_1U_2^{-1} = D_1^{-1}L_1^{-1}L_2D_2.$$

El lado izquierdo es un producto de dos matrices triangulares superiores con elementos en la diagonal principal iguales a uno. Dicho producto debe ser otra matriz del mismo tipo. Por otra parte, el lado derecho es una matriz triangular inferior. Esto obliga a que ambos lados sean precisamente la matriz identidad: la única matriz que al mismo tiempo es triangular superior con diagonal unitaria y también triangular superior. Así, $U_1U_2^{-1} = I_n$, y después de multiplicar por U_2 se tiene que $U_1 = U_2$.

Análogamente $L_1 = L_2$ y, finalmente, $D_1 = D_2$.

Teorema 3.5. Si A es una matriz simétrica y si puede factorizarse como $A = LDU$ sin intercambios de filas que destruyan la simetría, entonces la triangular superior U es la transpuesta de la triangular inferior L. En otras palabras, toda matriz simétrica tiene una factorización simétrica $A = LDL^t$.

Demostración

Supongamos que A puede factorizarse como $A = LDU$, tomando la transpuesta, se tiene que

$$A^t = (LDU)^t = U^t D^t L^t = U^t D L^t.$$

Como A es simétrica, es igual a A^t, así resulta que tenemos dos factorizaciones de A en triangular inferior por diagonal por triangular superior. (L^t es triangular superior con unos en la diagonal, exactamente como U). De acuerdo con el Teorema 3.4, esta factorización es única; por lo tanto, L^t debe ser idéntica a U, lo cual completa la prueba.

Ejemplo 3.3. Hallar una factorización LDU para la matriz

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 12 & 18 \\ 5 & 18 & 30 \end{bmatrix}.$$

Solución

Se procede en dos columnas como en el Ejemplo 3.1

Reducción de A a S		Creación de L a partir de I
$A = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 12 & 18 \\ 5 & 18 & 30 \end{bmatrix}$		$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	Sumar la fila 1 multiplicada por -3 a la fila 2	
$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 5 & 18 & 30 \end{bmatrix}$		$\sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
	Sumar la fila 1 multiplicada por -5 a la fila 3	
$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 3 & 5 \end{bmatrix}$		$\sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix}$
	Sumar la fila 2 multiplicada por -1 a la fila 3	

$$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 0 & 2 \end{bmatrix} = S \qquad\qquad \sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 1 & 1 \end{bmatrix} = L$$

Por lo tanto, se tiene que $A = LS$, factorizando a S se llega a

$$S = \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Nótese que $L^t = U$.

Teorema 3.6. Descomposición Triangular LU

Sea A una matriz real de tamaño $n \times n$ tal que todos sus menores principales son no nulos. Entonces A puede ser factorizada como

$$A = LU, \tag{3.4}$$

donde L es una matriz triangular inferior y U es una matriz no singular triangular superior, cada una de tamaño $n \times n$.

Demostración

Queda como ejercicio para el lector.

Teorema 3.7. Descomposición de Schur

Si A es una matriz de componentes reales de tamaño $n \times n$ con valores propios reales $\lambda_1, \lambda_2, \dots, \lambda_n$, entonces existe una matriz ortogonal Q tal que

$$Q^t A Q = T = \begin{bmatrix} t_{ij} \end{bmatrix}, \tag{3.5}$$

donde T es una matriz triangular superior con entradas diagonales $t_{ii} = \lambda_i$, $i = 1, 2, \dots, n$. (Es decir: toda matriz cuadrada real que sólo tiene valores propios reales, es ortogonalmente semejante a una matriz triangular superior).

Demostración

La demostración es por inducción sobre n. Si $n = 1$, A es una matriz real de tamaño 1×1 que es triangular. La matriz ortogonal es $Q = [1]$.

Supongamos que toda matriz real de tamaño $(n-1)\times(n-1)$ es triangularizable por una matriz ortogonal. Sea A una matriz real de tamaño $n \times n$ que sólo tiene valores propios reales. Sea $\vec{v}_1 \in \mathbb{R}^n$ un vector propio unitario asociado al valor propio λ_1. Denotemos por W el complemento ortogonal a $\vec{v}_1$ de dimensión $n-1$. Sea $\{\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ una base ortonormal de W. Luego, cada vector $\vec{X}$ de W tiene la forma

$$\vec{X} = a_2\,\vec{v}_2 + a_3\,\vec{v}_3 + \ldots + a_n\,\vec{v}_n.$$

La matriz de cambio de base de la *base canónica* de $\mathbb{R}^n$ a la base $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es la matriz S cuyas columnas son los elementos de los vectores $\vec{v}_i$. Luego,

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$S^{-1}AS = S^{-1}\begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}.$$

Pero como S es ortogonal se tiene que $S^{-1} = S^t$, por consiguiente

$$S^tAS = \begin{bmatrix} \lambda_1 & x_1^t \;\ldots\; x_{n-1}^t \\ 0 & \\ \vdots & \begin{bmatrix} & A_1 & \end{bmatrix} \\ 0 & \end{bmatrix},$$

donde A_1 es una matriz real de tamaño $(n-1)\times(n-1)$.

La prueba es ahora completada por inducción, si R_1 es una matriz ortogonal de tamaño $(n-1)\times(n-1)$ tal que $(R_1)^tA_1R_1 = T_1$, con T_1 triangular superior, por la hipótesis de inducción. Entonces, la matriz

$$R = \begin{bmatrix} 1 & 0 \;\ldots\; 0 \\ 0 & \\ \vdots & \begin{bmatrix} & R_1 & \end{bmatrix} \\ 0 & \end{bmatrix},$$

es una matriz ortogonal y

$$\begin{aligned} (SR)^tA(SR) = R^t\left(S^tAS\right)R &= \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1^t \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{x}^t \\ \vec{0} & A_1^t \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{x}^tR_1 \\ \vec{0} & T_1 \end{bmatrix}, \end{aligned}$$

donde $\vec{x}^t = (x_1{}^t, x_2{}^t, \ldots, x_{n-1}^t)$. La matriz $SR = Q$ es el producto de dos matrices ortogonales; por lo tanto, es también una matriz ortogonal. Así, Q^tAQ es una matriz triangular superior y nuestra prueba queda completa.

Ejemplo 3.4. Dada la matriz

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

Encuentre la triangularización de A.

Solución

El polinomio característico de A es

$$p_A(\lambda) = (2-\lambda)^3.$$

Entonces el único valor propio es $\lambda = 2$ (de multiplicidad algebraica 3). Los vectores propios correspondientes a $\lambda = 2$ son $\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ y $\vec{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. La idea básica del Teorema de Schur consiste en construir una base de $\mathbb{R}^3$ con el mayor número posible de vectores propios.

Si tomamos por ejemplo, $\vec{u}_3 = \vec{u}_1 \times \vec{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$, el conjunto $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es una base de $\mathbb{R}^3$. Mediante el algoritmo de Gram-Schmidt obtenemos la base ortonormal $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, donde

$$\vec{v}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

La matriz ortogonal Q es

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

La matriz triangular es

$$T = Q^t AQ = \begin{bmatrix} \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ 0 & 0 & 1 \\ \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{5}\sqrt{5} & 0 & \frac{2}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} & 0 & -\frac{1}{5}\sqrt{5} \\ 0 & 1 & 0 \end{bmatrix}$$

$$T = \begin{bmatrix} 2 & 0 & -5 \\ 0 & 2 & -\sqrt{5} \\ 0 & 0 & 2 \end{bmatrix}.$$

Obsérvese que los elementos de la diagonal principal de la matriz T son los valores propios de la matriz A.

3.2. Factorización QR

Esta factorización se usa ampliamente en los programas de computadora para resolver sistemas lineales, para determinar aproximaciones por mínimos cuadrados y para determinar los valores propios de una matriz.

Teorema 3.8. Factorización QR

Sea A una matriz real de tamaño $m \times n$ con $\rho(A) = n$. Entonces existen una matriz Q de tamaño $m \times n$ cuyas columnas son ortonormales y una matriz no singular R de tamaño $n \times n$ triangular superior tales que

$$A = QR. \tag{3.6}$$

Demostración

Como $\rho(A) = n$, entonces sus columnas son linealmente independientes; sean $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ las columnas de A, las cuales constituyen una base para el espacio generado por las columnas de A y mediante el proceso de Gram-Schmidt se ortonormalizan estos vectores para obtener $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ una base ortonormal para el espacio generado por las columnas de A. Sea

$$Q = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_n \end{bmatrix}.$$

Al expresar cada $\vec{x}_i$ como una combinación lineal de $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, se tiene

$$\vec{x}_i = r_{1i}\vec{v}_1 + r_{2i}\vec{v}_2 + \dots + r_{ni}\vec{v}_n, \qquad i = 1, 2, \dots, n.$$

Este sistema de ecuaciones escrito en forma matricial queda

$$\underbrace{\begin{bmatrix}\vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_n\end{bmatrix}}_{A} = \underbrace{\begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_n\end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}}_{\begin{bmatrix}\vec{r}_1 & \vec{r}_2 & \dots & \vec{r}_n\end{bmatrix}},$$

donde

$$\vec{r}_k = \begin{bmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{nk} \end{bmatrix}, \qquad\qquad k = 1, 2, \dots, n.$$

Por otra parte, como $\vec{v}_j$ es ortogonal a $gen\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ para $j > k$, es ortogonal a $\vec{x}_k$. Por lo tanto, $r_{jk} = 0$ para $j > k$, ya que

$$r_{jk} = \vec{v}_j^t \vec{x}_k = \vec{v}_j \cdot \vec{x}_k.$$

Sea $R = \begin{bmatrix}\vec{r}_1 & \vec{r}_2 & \dots & \vec{r}_n\end{bmatrix}$, entonces

$$A = QR = Q \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1n} \\ 0 & r_{22} & r_{23} & \dots & r_{2n} \\ 0 & 0 & r_{33} & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_{nn} \end{bmatrix}.$$

Ahora, mostremos que R es no singular. Consideremos el sistema lineal $R\vec{b} = \vec{0}$ y multipliquemos por Q a la izquierda, es decir,

$$\underbrace{QR}_{A}\,\vec{b} = \underbrace{Q\vec{0}}_{\vec{0}}.$$

Pero como las columnas de A son linealmente independientes, el sistema homogéneo $A\vec{b} = \vec{0}$ sólo tiene la solución trivial. Por lo tanto, R es no singular.

Nota

Para el caso en que A sea una matriz real de tamaño $m \times n$ con $\rho(A) = m$, entonces se puede encontrar una factorización de manera análoga a (3.6) de la forma

$$A = LQ, \tag{3.7}$$

donde L es una matriz real de tamaño $m \times m$ triangular inferior y no singular, Q es una matriz real de tamaño $m \times n$ cuyas filas son ortonormales.

Ejemplo 3.5. Encuentre una factorización QR de

$$A = \begin{bmatrix} 4 & 25 \\ 0 & 0 \\ 3 & -25 \end{bmatrix}.$$

Solución

Denotemos las columnas de A por

$$\vec{x}_1 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} \qquad \text{y} \qquad \vec{x}_2 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix}.$$

Si se aplica el algoritmo de Gram-Schmidt al conjunto $\{\vec{x}_1, \vec{x}_2\}$, base para el espacio generado por las columnas de A. Como $\|\vec{x}_1\| = 5$, se hace $\vec{v}_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|} = \begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix}$. Después,

$$\begin{aligned} \vec{v}_2' =& \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\vec{v}_1 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix} - 5 \begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix} \\ =& \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ 0 \\ -28 \end{bmatrix}. \end{aligned}$$

Entonces $\left\|\vec{v}_2'\right\| = 35$ y $\vec{v}_2 = \begin{bmatrix} 3/5 \\ 0 \\ -4/5 \end{bmatrix}$. Se puede verificar que $\{\vec{v}_1, \vec{v}_2\}$ es una nueva base ortonormal para el espacio generado por las columnas de A, observando que $\vec{v}_1 \cdot \vec{v}_2 = 0$. Entonces formamos la matriz

$$Q = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & 3 \\ 0 & 0 \\ 3 & -4 \end{bmatrix}.$$

Para encontrar R, se despeja esta matriz de $A = QR$, de la siguiente manera

$$Q^t A = Q^t\,(QR) = IR = R.$$

De esta manera, la matriz R es

$$R = \frac{1}{5}\begin{bmatrix} 4 & 0 & 3 \\ 3 & 0 & -4 \end{bmatrix}\begin{bmatrix} 4 & 25 \\ 0 & 0 \\ 3 & -25 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 0 & 35 \end{bmatrix}.$$

El lector puede verificar que $A = QR$.

Si la matriz A es cuadrada, entonces se puede enunciar el Teorema 3.8 de la siguiente manera

Teorema 3.9. Toda matriz cuadrada real A puede expresarse en la forma

$$A = QR \tag{3.8}$$

donde Q es una matriz ortogonal propia y R es triangular superior, con $r_{ii} > 0$, $i = 1, 2, \ldots, n-1$.

Demostración

Queda como ejercicio para el lector.

Corolario 3.9.1. Si A es ortogonal y el $\det A = 1$, entonces en (3.8), $R = I_n$. Si $\det A = -1$, entonces los elementos de $R = [r_{ij}]$, cumplen que

$$r_{ij} = \begin{cases} 1 & \text{si} \quad i = j, \\ 0 & \text{si} \quad i \neq j, \end{cases}$$

excepto $r_{nn} = -1$.

Demostración

Si $A = QR$, entonces

$$\begin{aligned} \det A = & \det(QR) = \det R, && \text{ya que } \det Q = 1 \\ = & r_{nn}\left(\prod_{i=1}^{n-1} r_{ii}\right) \end{aligned}$$

por lo tanto, se tiene que $r_{nn} = \pm 1$ ya que $\det A = \pm 1$ y la prueba del corolario se completa.

Corolario 3.9.2. Si A es no singular entonces la representación dada en (3.8) es única.

Demostración

Supongamos que A es no singular y consideremos dos factorizaciones distintas

$$A = QR \qquad \text{y} \qquad A = Q^{'}R^{'},$$

con Q, $Q^{'}$ ambas ortogonales propias y R, $R^{'}$ triangulares superiores. Entonces

$$\begin{aligned} I =& Q^{t}Q^{'}R^{'}R^{-1} = \left(Q^{t}Q^{'}\right)\left(R^{'}R^{-1}\right) \\ =& \widetilde{Q}\widetilde{R}. \end{aligned}$$

Aquí, la matriz ortogonal I está representada como el producto de una matriz ortogonal propia $\widetilde{Q}$ y una triangular superior $\widetilde{R}$. Por lo tanto, de acuerdo con el Corolario 3.9.1, $\widetilde{R} = I$ y, $\widetilde{Q} = I$. Luego, $R^{'} = R$ y $Q^{'} = Q$, de este modo el corolario está probado.

Ejercicios 3.1.

Para cada una de las siguientes matrices determine (en caso de ser posible) las factorizaciones LU y QR y la descomposición de Schur

1. $\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$.
2. $\begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}$.
3. $\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$.
4. $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$.
5. $\begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$.
6. $\begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}$.
7. $\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$.
8. $\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}$.

3.3. Raíces cuadradas

Si se considera un número $a \in \mathbb{R}$, como $(-a)^2 = a^2$, es evidente que para cualquier $a > 0$ tiene dos raíces cuadradas, una positiva y otra negativa; mientras que cuando $a < 0$ sus raíces cuadradas son dos imaginarios puros, una raíz es la conjugada de la otra. En general, si $a \in \mathbb{C}$, también a tiene dos raíces cuadradas distintas. En esta sección se extiende el concepto de raíz cuadrada, para estudiar la raíz cuadrada de una matriz de tamaño $n \times n$, tema poco trabajado en la mayoría de textos de Álgebra Lineal.

Definición 3.2. Sea A una matriz real de tamaño $n \times n$, una matriz X de tamaño $n \times n$ se llama raíz cuadrada de A si cumple que

$$X^2 = A. \tag{3.9}$$

La matriz X puede tener algunos elementos complejos.

Teorema 3.10. Sea $D = [d_{ii}]$ una matriz real diagonal de tamaño $n \times n$, entonces una raíz cuadrada de D es

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{d_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{d_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_{nn}} \end{bmatrix}. \tag{3.10}$$

Demostración

Queda como ejercicio para el lector.

Observación

Como cada elemento d_{ii}, tiene dos raíces cuadradas $\sqrt{d_{ii}}$ y $-\sqrt{d_{ii}}$, entonces en la matriz 3.10, se puede reemplazar por la otra raíz del elemento d_{ii} y se obtiene una nueva raíz cuadrada para D.

Teorema 3.11. Sea A una matriz real de tamaño $n \times n$, diagonalizable y sea P una matriz no singular tal que la matriz $D = P^{-1}AP$ es diagonal. Entonces una raíz cuadrada de A es

$$A^{\frac{1}{2}} = PD^{\frac{1}{2}}P^{-1} \tag{3.11}$$

donde $D^{\frac{1}{2}}$ es definida como en 3.10.

Demostración

La demostración consiste en un cálculo directo,

$$\left(A^{\frac{1}{2}}\right)^2 = \left(PD^{\frac{1}{2}}P^{-1}\right)\left(PD^{\frac{1}{2}}P^{-1}\right) = P\left(D^{\frac{1}{2}}\right)^2 P^{-1} = PDP^{-1} = A.$$

Así, queda el teorema probado.

Cuando son iguales la multiplicidad algebraica y la multiplicidad geométrica de los valores propios de una matriz A, se tiene que A es semejante a una matriz diagonal de D cuyos elementos son los valores propios de A. Por lo tanto, si A es diagonalizable, dado que cada valor propio tiene dos raíces cuadradas, entonces el número de raíces cuadradas de la matriz A es igual a 2^n. Si todos los valores propios de A son nulos, entonces A no tiene raíz cuadrada.

Ejemplo 3.6. Determine las raíces cuadradas de la siguiente matriz

$$A = \begin{bmatrix} 12 & 1 & -2 & -11 \\ 7 & 11 & -5 & 12 \\ -10 & -3 & 16 & -1 \\ -3 & 4 & 7 & 15 \end{bmatrix}.$$

Solución

Para la matriz A se tiene que el polinomio característico es

$$p_A(\lambda) = \lambda^4 - 54\lambda^3 + 969\lambda^2 - 6676\lambda + 14400$$

Entonces, sus valores propios son

$$\lambda_1 = 4, \qquad \lambda_2 = 9, \qquad \lambda_3 = 16, \qquad \lambda_4 = 25.$$

La matriz A se puede expresar como

$$A = \begin{bmatrix} \frac{2}{3} & -\frac{23}{4} & \frac{115}{144} & -\frac{115}{252} \\ -1 & \frac{23}{4} & -\frac{115}{48} & 0 \\ \frac{1}{3} & -\frac{23}{4} & \frac{115}{72} & \frac{115}{252} \\ \frac{1}{3} & 0 & -\frac{115}{144} & \frac{115}{252} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \\ -\frac{4}{23} & -\frac{4}{23} & -\frac{4}{23} & 0 \\ -\frac{48}{115} & -\frac{48}{115} & 0 & -\frac{48}{115} \\ -\frac{84}{115} & 0 & \frac{84}{115} & \frac{84}{115} \end{bmatrix}.$$

Como A tiene 4 valores propios distintos no nulos, entonces posee $2^4 = 16$ raíces cuadradas, al tomar todas las raíces positivas de los valores propios

$$A^{\frac{1}{2}} = \begin{bmatrix} \frac{2}{3} & -\frac{23}{4} & \frac{115}{144} & -\frac{115}{252} \\ -1 & \frac{23}{4} & -\frac{115}{48} & 0 \\ \frac{1}{3} & -\frac{23}{4} & \frac{115}{72} & \frac{115}{252} \\ \frac{1}{3} & 0 & -\frac{115}{144} & \frac{115}{252} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \\ -\frac{4}{23} & -\frac{4}{23} & -\frac{4}{23} & 0 \\ -\frac{48}{115} & -\frac{48}{115} & 0 & -\frac{48}{115} \\ -\frac{84}{115} & 0 & \frac{84}{115} & \frac{84}{115} \end{bmatrix}$$

$$=\frac{1}{3} \begin{bmatrix} 10 & 1 & 0 & -5 \\ 3 & 9 & -3 & 6 \\ -4 & -1 & 12 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix}.$$

Para obtener las otras raíces cuadradas de la matriz A, se modifican los elementos de $D^{\frac{1}{2}}$ por las raíces negativas de los valores propios, como se muestra a continuación

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} 10 & 9 & 8 & -13 \\ 3 & -3 & -15 & 18 \\ -4 & 3 & 16 & -5 \\ -1 & 6 & 7 & 7 \end{bmatrix} \qquad \text{cuando} \qquad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} -8 & -17 & -18 & -5 \\ 21 & 27 & 15 & 6 \\ -22 & -19 & -6 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix} \qquad \text{tomando} \qquad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

$$A^{\frac{1}{2}} = \begin{bmatrix} 6 & 3 & 0 & 1 \\ -7 & -5 & -1 & -6 \\ 4 & 5 & 4 & 5 \\ -3 & -2 & 1 & 1 \end{bmatrix} \qquad \text{asumiendo} \qquad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} 0 & 1 & 10 & 5 \\ 3 & 9 & -3 & 6 \\ 6 & -1 & 2 & -11 \\ 9 & 2 & -7 & 1 \end{bmatrix} \qquad \text{con} \qquad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}.$$

Nótese que en la matriz $D^{\frac{1}{2}}$ se ha modificado sólo un valor propio, ahora se consideran las raíces cuadradas de A cuando se cambian dos valores propios.

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} -8 & -9 & -10 & -13 \\ 21 & 15 & 3 & 18 \\ -22 & -15 & -2 & -5 \\ -1 & 6 & 7 & 7 \end{bmatrix} \qquad \text{cuando} \qquad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

$$A^{\frac{1}{2}} = \frac{1}{3}\begin{bmatrix} 18 & 17 & 8 & -5 \\ -21 & -27 & -15 & -6 \\ 12 & 19 & 16 & 11 \\ -9 & -2 & 7 & -1 \end{bmatrix} \qquad \text{tomando} \qquad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

$$A^{\frac{1}{2}} = \begin{bmatrix} 0 & 3 & 6 & -1 \\ 1 & -1 & -5 & 6 \\ 2 & 1 & 2 & -5 \\ 3 & 2 & -1 & -1 \end{bmatrix} \qquad \text{asumiendo} \qquad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}.$$

Se puede verificar que estas 8 matrices y sus respectivas matrices negativas son las 16 raíces cuadradas de A.

Hasta este momento hemos considerado las raíces cuadradas de matrices diagonalizables, pero como todas las matrices no son diagonalizables a continuación se muestran algunos métodos para obtener las raíces cuadradas de una matriz.

Teorema 3.12. Si A es una matriz real de tamaño 2×2 con al menos un valor propio no nulo, entonces su raíz cuadrada es

$$A^{\frac{1}{2}} = \frac{1}{\sqrt{\lambda_1} + \sqrt{\lambda_2}} \left[A + \left(\sqrt{\lambda_1}\sqrt{\lambda_2} \right) I \right], \tag{3.12}$$

donde $\lambda_i = \frac{1}{2}\left(tr(A) \pm \sqrt{tr^2(A) - 4\det(A)} \right)$.

Demostración

Sea $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, luego, si la matriz dada en (3.12), es la raíz cuadrada de A, entonces $A = \left(A^{\frac{1}{2}}\right)\left(A^{\frac{1}{2}}\right)$

$$\begin{aligned} \left(A^{\frac{1}{2}}\right)^2 &= \left(\frac{1}{\sqrt{\lambda_1} + \sqrt{\lambda_2}} \right)^2 \begin{bmatrix} a + \sqrt{\det(A)} & b \\ c & d + \sqrt{\det(A)} \end{bmatrix}^2 \\ &= \frac{1}{\lambda_1 + \lambda_2 + 2\sqrt{\lambda_1 \lambda_2}} \begin{bmatrix} a^2 + ad + 2a\sqrt{\det(A)} & ba + bd + 2b\sqrt{\det(A)} \\ ca + cd + 2c\sqrt{\det(A)} & da + d^2 + 2d\sqrt{\det(A)} \end{bmatrix} \end{aligned}$$

De este modo,

$$\left(A^{\frac{1}{2}}\right)^2 = \frac{(a+d) + 2\sqrt{\det(A)}}{\lambda_1 + \lambda_2 + 2\sqrt{\lambda_1 \lambda_2}} \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Pero por el Teorema 2.17, se tiene que

$$tr(A) = \lambda_1 + \lambda_2 \qquad \text{y} \qquad \det(A) = \lambda_1 \lambda_2.$$

Por lo tanto,

$$\left(A^{\frac{1}{2}}\right)^2 = A.$$

Ejemplo 3.7. Determine para cada una de las siguientes matrices una raíz cuadrada

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 4 & 2 \\ -2 & 4 \end{bmatrix}.$$

Solución

- Para la matriz A se tiene que

  $$\lambda_1 = \lambda_2 = 2,$$

  como A tiene un único valor propio no nulo, entonces posee $2^1 = 2$ raíces cuadradas. Una raíz cuadrada, es

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

  y multiplicando por -1 se obtiene la otra raíz. El lector puede verificar que $\left(\pm A^{\frac{1}{2}}\right)^2 = A$.

- Para la matriz B se tiene que

  $$\lambda_1 = 4 + 2i, \qquad \lambda_2 = 4 - 2i.$$

  Como los valores propios de A son complejos, entonces sus raíces son

  $$\sqrt{\lambda_1} = \sqrt{4+2i} = (-1)^k \left[\sqrt{2+\sqrt{5}} + \sqrt{2-\sqrt{5}}\right], \qquad k = 0, 1$$
  $$\sqrt{\lambda_2} = \sqrt{4-2i} = (-1)^k \left[\sqrt{2+\sqrt{5}} - \sqrt{2-\sqrt{5}}\right], \qquad k = 0, 1$$

  Si se considera el caso $k = 0$ en ambas raíces, se tiene que

  $$\sqrt{4+2i} + \sqrt{4-2i} = 2\sqrt{2+\sqrt{5}}, \qquad \sqrt{4+2i}\sqrt{4-2i} = 2\sqrt{5}.$$

Luego, la matriz B posee $2^2 = 4$ raíces cuadradas. Una raíz cuadrada de B es

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

Nótese que $-B^{\frac{1}{2}}$ también es raíz, lo cual se puede verificar ya que

$$\begin{aligned} \left(\pm B^{\frac{1}{2}}\right)^2 =& (\pm 1)^2 \begin{bmatrix} \sqrt{\sqrt{5}+2} & \sqrt{\sqrt{5}-2} \\ -\sqrt{\sqrt{5}-2} & \sqrt{\sqrt{5}+2} \end{bmatrix} \begin{bmatrix} \sqrt{\sqrt{5}+2} & \sqrt{\sqrt{5}-2} \\ -\sqrt{\sqrt{5}-2} & \sqrt{\sqrt{5}+2} \end{bmatrix} \\ =& \begin{bmatrix} 4 & 2 \\ -2 & 4 \end{bmatrix}. \end{aligned}$$

Ahora se considera $k = 1$ en la raíz cuadrada de uno de los valores propios, en este caso, se tiene que

$$\sqrt{4+2i} - \sqrt{4-2i} = 2\sqrt{2-\sqrt{5}}, \qquad \sqrt{4+2i}\left(-\sqrt{4-2i}\right) = -2\sqrt{5}.$$

Por lo tanto, otra raíz cuadrada de B es

$$B^{\frac{1}{2}} = \frac{1}{2\sqrt{2-\sqrt{5}}} \begin{bmatrix} 4-2\sqrt{5} & 2 \\ -2 & 4-2\sqrt{5} \end{bmatrix} = i \begin{bmatrix} \sqrt{\sqrt{5}-2} & -\sqrt{\sqrt{5}+2} \\ \sqrt{\sqrt{5}+2} & \sqrt{\sqrt{5}-2} \end{bmatrix}.$$

En este caso, $-B^{\frac{1}{2}}$ también es raíz, el lector puede veficarlo.

Teorema 3.13. Si A es una matriz de componentes reales de tamaño 3×3 con al menos un valor propio no nulo, entonces sus raíces cuadradas son

$$A^{\frac{1}{2}} = [A + \alpha I]^{-1} \left[\beta A + \left(\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\lambda_3}\right) I\right], \tag{3.13}$$

donde $\alpha = \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left(\sqrt{\lambda_i}\sqrt{\lambda_j}\right)$ y $\beta = \sum_{k=1}^{3} \sqrt{\lambda_k}$

Demostración

Supongamos que A es una matriz cuadrada con valores propios reales. Por el Teorema 3.7 es semejante a una matriz triangular superior T, luego, puede expresarse como

$$A = QTQ^t, \tag{3.14}$$

donde Q es una matriz ortogonal y

$$T=\begin{bmatrix}\lambda_1 & a & b\\ 0 & \lambda_2 & c\\ 0 & 0 & \lambda_3\end{bmatrix}. \tag{3.15}$$

Con $\lambda_1,\ \lambda_2,\ \lambda_3$ los valores propios de A. Al reemplazar (3.14) en (3.13), se tiene que

$$\begin{aligned}
[A+\alpha I]^{-1}\left[\beta A+\sqrt{\det A}I\right] &= \left[QTQ^t+\alpha I\right]^{-1}\left[\beta\left(QTQ^t\right)+\sqrt{\det\left(QTQ^t\right)}I\right]\\
&= \left[Q\left(T+\alpha I\right)Q^t\right]^{-1}\left[Q\left(\beta T+\sqrt{\det T}I\right)Q^t\right]\\
&=Q\left[T+\alpha I\right]^{-1}Q^tQ\left[\beta T+\sqrt{\det T}I\right]Q^t\\
&=Q\left[T+\alpha I\right]^{-1}\left[\beta T+\sqrt{\det T}I\right]Q^t.
\end{aligned}$$

Es decir, al utilizar la descomposición de Schur, se llega a que

$$A^{\frac{1}{2}}=QT^{\frac{1}{2}}Q^t.$$

Luego, se debe demostrar que $T^{\frac{1}{2}}=[T+\alpha I]^{-1}\left[\beta T+\sqrt{\det T}I\right]$. Para ello, se calcula α y β, como sigue

$$\alpha=\sqrt{\lambda_1}\sqrt{\lambda_2}+\sqrt{\lambda_1}\sqrt{\lambda_3}+\sqrt{\lambda_2}\sqrt{\lambda_3}\qquad \text{y}\qquad \beta=\sqrt{\lambda_1}+\sqrt{\lambda_2}+\sqrt{\lambda_3}.$$

Por lo tanto,

$$\begin{aligned}
\beta T+\sqrt{\det T}I &= \left(\sqrt{\lambda_1}+\sqrt{\lambda_3}+\sqrt{\lambda_2}\right)\begin{bmatrix}\lambda_1 & a & b\\ 0 & \lambda_2 & c\\ 0 & 0 & \lambda_3\end{bmatrix}+\sqrt{\lambda_1\lambda_2\lambda_3}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\\
&=\xi\begin{bmatrix}\dfrac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}+\sqrt{\lambda_3}} & \frac{a\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}+\sqrt{\lambda_2}\right)}{\xi} & \frac{b\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}+\sqrt{\lambda_2}\right)}{\xi}\\ 0 & \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\sqrt{\lambda_3}} & \frac{c\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}+\sqrt{\lambda_2}\right)}{\xi}\\ 0 & 0 & \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}\end{bmatrix},
\end{aligned}\tag{3.16}$$

donde $\xi=\left(\sqrt{\lambda_1}+\sqrt{\lambda_2}\right)\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}\right)\left(\sqrt{\lambda_2}+\sqrt{\lambda_3}\right)$. Por otra parte,

$$(T+\alpha I)^{-1}=\left\{\begin{bmatrix}\lambda_1 & a & b\\ 0 & \lambda_2 & c\\ 0 & 0 & \lambda_3\end{bmatrix}+\left(\sqrt{\lambda_1\lambda_2}+\sqrt{\lambda_1\lambda_3}+\sqrt{\lambda_2\lambda_3}\right)\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right\}^{-1}=$$

$$\begin{bmatrix}\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}\right)\left(\sqrt{\lambda_1}+\sqrt{\lambda_2}\right) & a & b\\ 0 & \left(\sqrt{\lambda_2}+\sqrt{\lambda_3}\right)\left(\sqrt{\lambda_1}+\sqrt{\lambda_2}\right) & c\\ 0 & 0 & \left(\sqrt{\lambda_2}+\sqrt{\lambda_3}\right)\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}\right)\end{bmatrix}^{-1},$$

la cual en términos de ξ, se puede expresar como

$$(T+\alpha I)^{-1} = \left\{ \xi \begin{bmatrix} \dfrac{1}{\sqrt{\lambda_2}+\sqrt{\lambda_3}} & \dfrac{a}{\xi} & \frac{b}{\xi} \\ 0 & \dfrac{1}{\sqrt{\lambda_1}+\sqrt{\lambda_3}} & \frac{c}{\xi} \\ 0 & 0 & \dfrac{1}{\sqrt{\lambda_1}+\sqrt{\lambda_2}} \end{bmatrix} \right\}^{-1} =$$

$$\frac{1}{\xi} \begin{bmatrix} \sqrt{\lambda_2}+\sqrt{\lambda_3} & -\dfrac{a}{\sqrt{\lambda_1}+\sqrt{\lambda_2}} & \dfrac{ac\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}\right)-b\xi}{\xi\left(\sqrt{\lambda_1}+\sqrt{\lambda_3}\right)} \\ 0 & \sqrt{\lambda_1}+\sqrt{\lambda_3} & -\frac{c}{\sqrt{\lambda_2}+\sqrt{\lambda_3}} \\ 0 & 0 & \sqrt{\lambda_1}+\sqrt{\lambda_2} \end{bmatrix}. \qquad (3.17)$$

Al multiplicar por la derecha la matriz dada en (3.17) por (3.16), se obtiene

$$(T+\alpha I)^{-1}\left(\beta T+\sqrt{\det T}I\right) = \begin{bmatrix} \sqrt{\lambda_1} & \frac{a}{\sqrt{\lambda_1}+\sqrt{\lambda_2}} & \frac{b\left(\sqrt{\lambda_1}+\sqrt{\lambda_2}\right)\left(\sqrt{\lambda_2}+\sqrt{\lambda_3}\right)-ac}{\xi} \\ 0 & \sqrt{\lambda_2} & \frac{c}{\sqrt{\lambda_2}+\sqrt{\lambda_3}} \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}.$$

El lector puede verificar que esta última matriz es una raíz cuadrada de la matriz dada en (3.15) y de esta manera concluir la prueba.

Ejemplo 3.8. Determine las raíces cuadradas para la matriz dada a continuación

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

Solución

El polinomio característico de la matriz A es

$$p_A(\lambda) = -\lambda^3+5\lambda^2-8\lambda+4.$$

De esta suma, se tiene que

$$\lambda_1 = 1, \qquad \lambda_2 = 2, \qquad \lambda_3 = 2.$$

Como A tiene 2 valores propios distintos no nulos, entonces posee $2^2 = 4$ raíces cuadradas. Si se consideran las raíces positivas de los valores propios se tiene que

$$\alpha = 2+2\sqrt{2} \qquad \text{y} \qquad \beta = 1+2\sqrt{2}.$$

Por lo tanto, una raíz cuadrada de A es

$$
\begin{aligned}
A^{\frac{1}{2}} &= \left[A+\left(2+2\sqrt{2}\right)I\right]^{-1}\left[\left(1+2\sqrt{2}\right)A+\sqrt{4}I\right]\\
&= \begin{bmatrix} 5+2\sqrt{2} & -1 & 2\\ -1 & 4+2\sqrt{2} & -1\\ -1 & 1 & 2+2\sqrt{2}\end{bmatrix}^{-1}\begin{bmatrix} 5+6\sqrt{2} & -1-2\sqrt{2} & 2+4\sqrt{2}\\ -1-2\sqrt{2} & 4+4\sqrt{2} & -1-2\sqrt{2}\\ -1-2\sqrt{2} & 1+2\sqrt{2} & 2\end{bmatrix}\\
&= \frac{1}{8}\begin{bmatrix} 1 & 20-14\sqrt{2} & 16\sqrt{2}-23\\ 3-2\sqrt{2} & 4-2\sqrt{2} & 3-2\sqrt{2}\\ 3-2\sqrt{2} & 14\sqrt{2}-20 & 27-18\sqrt{2}\end{bmatrix}\begin{bmatrix} 5+6\sqrt{2} & -1-2\sqrt{2} & 2+4\sqrt{2}\\ -1-2\sqrt{2} & 4+4\sqrt{2} & -1-2\sqrt{2}\\ -1-2\sqrt{2} & 1+2\sqrt{2} & 2\end{bmatrix}\\
&= \frac{1}{4}\begin{bmatrix} 5\sqrt{2} & 4-4\sqrt{2} & -4+5\sqrt{2}\\ -\sqrt{2} & 4\sqrt{2} & -\sqrt{2}\\ -\sqrt{2} & -4+4\sqrt{2} & 4-\sqrt{2}\end{bmatrix}
\end{aligned}
$$

Nótese que $-A^{\frac{1}{2}}$ también es raíz, verifiquemos que en efecto son raíces cuadradas

$$
\begin{aligned}
\left(\pm A^{\frac{1}{2}}\right)^2 &= \begin{bmatrix} \frac{5}{4}\sqrt{2} & 1-\sqrt{2} & -1+\frac{5}{4}\sqrt{2}\\ -\frac{1}{4}\sqrt{2} & \sqrt{2} & -\frac{1}{4}\sqrt{2}\\ -\frac{1}{4}\sqrt{2} & \sqrt{2}-1 & 1-\frac{1}{4}\sqrt{2}\end{bmatrix}\begin{bmatrix} \frac{5}{4}\sqrt{2} & 1-\sqrt{2} & -1+\frac{5}{4}\sqrt{2}\\ -\frac{1}{4}\sqrt{2} & \sqrt{2} & -\frac{1}{4}\sqrt{2}\\ -\frac{1}{4}\sqrt{2} & \sqrt{2}-1 & 1-\frac{1}{4}\sqrt{2}\end{bmatrix}\\
&= \begin{bmatrix} 3 & -1 & 2\\ -1 & 2 & -1\\ -1 & 1 & 0\end{bmatrix}.
\end{aligned}
$$

Si se considera la raíz negativa del valor propio distinto se tiene que

$$\alpha = 2-2\sqrt{2} \qquad \text{y} \qquad \beta = -1+2\sqrt{2}.$$

Por lo tanto, otra raíz cuadrada de A es

$$
\begin{aligned}
A^{\frac{1}{2}} &= \left[A+\left(2-2\sqrt{2}\right)I\right]^{-1}\left[\left(-1+2\sqrt{2}\right)A-\sqrt{4}I\right]\\
&= \begin{bmatrix} 5-2\sqrt{2} & -1 & 2\\ -1 & 4+2\sqrt{2} & -1\\ -1 & 1 & 2-2\sqrt{2}\end{bmatrix}^{-1}\begin{bmatrix} -5+6\sqrt{2} & 1-2\sqrt{2} & -2+4\sqrt{2}\\ 1-2\sqrt{2} & -4+4\sqrt{2} & 1-2\sqrt{2}\\ 1-2\sqrt{2} & -1+2\sqrt{2} & -2\end{bmatrix}\\
&= \frac{1}{8}\begin{bmatrix} 1 & 20+14\sqrt{2} & -23-16\sqrt{2}\\ 3+2\sqrt{2} & 4+2\sqrt{2} & 3+2\sqrt{2}\\ 3+2\sqrt{2} & -20-14\sqrt{2} & 27+18\sqrt{2}\end{bmatrix}\begin{bmatrix} -5+6\sqrt{2} & 1-2\sqrt{2} & -2+4\sqrt{2}\\ 1-2\sqrt{2} & -4+4\sqrt{2} & 1-2\sqrt{2}\\ 1-2\sqrt{2} & -1+2\sqrt{2} & -2\end{bmatrix}\\
&= \frac{1}{4}\begin{bmatrix} 5\sqrt{2} & -4-4\sqrt{2} & 4+5\sqrt{2}\\ -\sqrt{2} & 4\sqrt{2} & -\sqrt{2}\\ -\sqrt{2} & 4+4\sqrt{2} & -4-\sqrt{2}\end{bmatrix}.
\end{aligned}
$$

El lector, puede verificar que esta matriz y su respectiva matriz negativa también son raíces cuadradas de A.

Teorema 3.14. Si A es una matriz de componentes reales de tamaño 4×4 con al menos un valor propio no nulo, entonces sus raíces cuadradas son

$$A^{\frac{1}{2}} = [\alpha A+\beta I]^{-1}\left[A^2+\gamma A+\left(\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\lambda_3}\sqrt{\lambda_4}\right)I\right] \tag{3.18}$$

donde $\alpha=\sum_{k=1}^{4}\sqrt{\lambda_k}$, $\beta=\sum_{k>j>i}\sqrt{\lambda_i}\sqrt{\lambda_j}\sqrt{\lambda_k}$ y $\gamma=\sum_{j>i}\sqrt{\lambda_i}\sqrt{\lambda_j}$

Demostración

Queda como ejercicio para el lector.

Ejemplo 3.9. Determine mediante el método descrito en el Teorema anterior una raíz cuadrada para la matriz dada en el Ejemplo 3.6.

Solución

Como en el Ejemplo 3.6, se obtuvieron los valores propios de A, se tiene que

$$\begin{aligned}
\alpha &= \sqrt{4}+\sqrt{9}+\sqrt{16}+\sqrt{25}=14,\\
\beta &= \sqrt{4}\sqrt{9}\sqrt{16}+\sqrt{4}\sqrt{9}\sqrt{25}+\sqrt{4}\sqrt{16}\sqrt{25}+\sqrt{9}\sqrt{16}\sqrt{25}=154,\\
\gamma &= \sqrt{4}\sqrt{9}+\sqrt{4}\sqrt{16}+\sqrt{4}\sqrt{25}+\sqrt{9}\sqrt{16}+\sqrt{9}\sqrt{25}+\sqrt{16}\sqrt{25}=71.
\end{aligned}$$

Por lo tanto, la matriz $\alpha A+\beta I$, es

$$\begin{bmatrix} 322 & 14 & -28 & -154\\ 98 & 308 & -70 & 168\\ -140 & -42 & 378 & -14\\ -42 & 56 & 98 & 364 \end{bmatrix}.$$

Por otra parte, la matriz $A^2+\gamma A+\sqrt{\det A}I$, es

$$\begin{bmatrix} 1176 & 56 & -280 & -1064\\ 672 & 1092 & -420 & 1092\\ -1008 & -308 & 1540 & -28\\ -336 & 364 & 700 & 1484 \end{bmatrix}.$$

Luego, la raíz cuadrada de A es

$$A^{\frac{1}{2}} = \frac{1}{22680}\begin{bmatrix} 76 & -11 & -6 & 37 \\ -21 & 87 & 27 & -48 \\ 26 & 5 & 60 & 11 \\ 5 & -16 & -21 & 71 \end{bmatrix}\begin{bmatrix} 1176 & 56 & -280 & -1064 \\ 672 & 1092 & -420 & 1092 \\ -1008 & -308 & 1540 & -28 \\ -336 & 364 & 700 & 1484 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 10 & 1 & 0 & -5 \\ 3 & 9 & -3 & 6 \\ -4 & -1 & 12 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix},$$

la cual coincide con una de las obtenidas en el Ejemplo 3.6.

Teorema 3.15. Si A es una matriz real de tamaño $n \times n$ $(n \geq 5)$, con una descomposición de la forma $A = PBP^{-1}$, entonces sus raíces cuadradas se calculan de la siguiente manera

$$A^{\frac{1}{2}} = PB^{\frac{1}{2}}P^{-1} = P\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_k \end{bmatrix}^{\frac{1}{2}} P^{-1}, \tag{3.19}$$

en donde cada submatriz B_t es de tamaño 1×1, 2×2, 3×3 ó 4×4, de tal manera que se le pueda calcular a cada bloque una raíz cuadrada como las dadas en (3.12), (3.13) ó (3.18) respectivamente.

Demostración

Queda como ejercicio para el lector.

Ejemplo 3.10. Determine una raíz cuadrada para la siguiente matriz

$$A = \begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix}.$$

Solución

Para la matriz A se tiene que el polinomio característico es

$$p_A(\lambda) = -\lambda^3 + 8\lambda^2 - 20\lambda,$$

luego, se tiene que

$$\lambda_1 = 0, \qquad \lambda_2 = 4 + 2i, \qquad \lambda_3 = 4 - 2i,$$

como A tiene valores propios complejos, usando el método de factorización dado en (2.16), se tiene

$$\begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 \\ -2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix}^{-1}.$$

Luego, la raíz cuadrada de A es

$$\begin{aligned} A^{\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 \\ -2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} 4 & 2 \\ -2 & 4 \end{bmatrix}^{\frac{1}{2}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}. \end{aligned}$$

Si se usa una de las raíces cuadradas encontradas en el Ejemplo 3.7, se tiene que

$$\begin{aligned} A^{\frac{1}{2}} &= \frac{1}{4\sqrt{2+\sqrt{5}}} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} 4+2\sqrt{5} & 2 \\ -2 & 4+2\sqrt{5} \end{bmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \\ &= \frac{1}{2\sqrt{2+\sqrt{5}}} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 2\sqrt{5}+1 & 2\sqrt{5}+3 & 1 \\ -8-3\sqrt{5} & -4-\sqrt{5} & \sqrt{5}+2 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{2\sqrt{2+\sqrt{5}}} \begin{bmatrix} 8+3\sqrt{5} & 4+\sqrt{5} & -2-\sqrt{5} \\ -7-\sqrt{5} & \sqrt{5}-1 & 3+\sqrt{5} \\ 1+2\sqrt{5} & 3+2\sqrt{5} & 1 \end{bmatrix} \end{aligned}$$

Se puede fácilmente verificar que $\left(A^{\frac{1}{2}}\right)^2 = A$

$$\left(A^{\frac{1}{2}}\right)^2 = \left(\frac{1}{2\sqrt{2+\sqrt{5}}}\right)^2 \begin{bmatrix} 8+3\sqrt{5} & 4+\sqrt{5} & -2-\sqrt{5} \\ -7-\sqrt{5} & \sqrt{5}-1 & 3+\sqrt{5} \\ 1+2\sqrt{5} & 3+2\sqrt{5} & 1 \end{bmatrix}^2$$

$$
\begin{aligned}
\left(A^{\frac{1}{2}}\right)^2 =& \frac{1}{4\left(2+\sqrt{5}\right)} \begin{bmatrix} 64+32\sqrt{5} & 32+16\sqrt{5} & -16-8\sqrt{5} \\ -56-28\sqrt{5} & -8-4\sqrt{5} & 24+12\sqrt{5} \\ 8+4\sqrt{5} & 24+12\sqrt{5} & 8+4\sqrt{5} \end{bmatrix} \\
=& \begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix}
\end{aligned}
$$

3.3.1. Raíces cuadradas de matrices simétricas

De manera análoga a la sección 2.4, en la cual se desarrollo el tema de diagonalización para *matrices simétricas*, en este apartado se presenta por separado la parte concerniente a raíces cuadradas para matrices simétricas.

Teorema 3.16. Toda matriz simétrica A de tamaño $n \times n$ tiene valores propios positivos si y sólo si existe una matriz simétrica B de tamaño $n \times n$, tal que

$$
A = B^2 \tag{3.20}
$$

La matriz B se denomina una *raíz cuadrada* de A.

Demostración

Si los valores propios de A son positivos entonces $\det(A) > 0$ y por ser simétrica se puede factorizar de la forma

$$
\begin{aligned}
A =& QDQ^t = Q\left(D^{\frac{1}{2}}\right)^2 Q^t \\
=& \left(QD^{\frac{1}{2}}Q^t\right)\left(QD^{\frac{1}{2}}Q^t\right) \\
=& B^tB
\end{aligned}
$$

nótese que B es una matriz simétrica y de rango n (como A), por lo tanto, $B^tB = B^2$. La matriz $D^{\frac{1}{2}} = diag\left\{\sqrt{\lambda_i}\right\}$ semejante a B está definida como en (3.10).

Ejemplo 3.11. Determine una raíz cuadrada para la matriz dada en 2.15

Solución

Haciendo referencia al Ejemplo 2.15, se tiene que

$$
Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \qquad \text{y} \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.
$$

Por lo tanto,

$$\begin{aligned}B=&\begin{bmatrix}-1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6}\\ 0 & -1/\sqrt{3} & 2/\sqrt{6}\\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \sqrt{7}\end{bmatrix}\begin{bmatrix}-1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6}\end{bmatrix}\\ =&\begin{bmatrix}-1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6}\\ 0 & -1/\sqrt{3} & 2/\sqrt{6}\\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6}\end{bmatrix}\begin{bmatrix}-1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3}\\ \sqrt{7}/\sqrt{6} & 2\sqrt{7}/\sqrt{6} & \sqrt{7}/\sqrt{6}\end{bmatrix}.\end{aligned}$$

Es decir,

$$B=\frac{1}{6}\begin{bmatrix}5+\sqrt{7} & -2+2\sqrt{7} & -1+\sqrt{7}\\ -2+2\sqrt{7} & 2+4\sqrt{7} & -2+2\sqrt{7}\\ -1+\sqrt{7} & -2+2\sqrt{7} & 5+\sqrt{7}\end{bmatrix}.$$

El lector puede verificar que $A=B^2$.

De acuerdo con la definición 2.6, las matrices reales cuadradas A y B se dicen que son *congruentes*, si existe una matriz P no singular, tal que

$$A=P^tBP.$$

En ocasiones, además de establecer el hecho en sí de la congruencia, se requiere encontrar la matriz P de la transformación, misma que ha de satisfacer $A=P^tBP$. En estos momentos, se puede construir la matriz P para matrices simétricas no singulares A y B utilizando la descomposición LDU de cada una y una de las raíces cuadradas de las D; como sigue

$$P^t=L_1D_1^{\frac{1}{2}}D_2^{-\frac{1}{2}}L_2^{-1} \tag{3.21}$$

donde las descomposiciones LDU para las matrices A y B son $L_1D_1L_1^t$ y $L_2D_2L_2^t$, respectivamente.

Ejemplo 3.12. Determine si las matrices dadas en los ejemplos 2.15 y 3.3 son congruentes.

Solución

La factorización LDL^t de la matriz dada en el Ejemplo 2.15, es

$$\begin{bmatrix}2 & 2 & 1\\ 2 & 5 & 2\\ 1 & 2 & 2\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ \frac{1}{2} & \frac{1}{3} & 1\end{bmatrix}\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & \frac{7}{6}\end{bmatrix}\begin{bmatrix}1 & 1 & \frac{1}{2}\\ 0 & 1 & \frac{1}{3}\\ 0 & 0 & 1\end{bmatrix}=L_1D_1L_1^t.$$

En el Ejemplo 3.3, se obtuvo que

$$\begin{bmatrix}1 & 3 & 5\\ 3 & 12 & 18\\ 5 & 18 & 30\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 3 & 1 & 0\\ 5 & 3 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}1 & 3 & 5\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}=L_2D_2L_2^t.$$

Por lo tanto,

$$\begin{aligned}P^t &= \begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ \frac{1}{2} & \frac{1}{3} & 1\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0 & 0\\ 0 & \sqrt{3} & 0\\ 0 & 0 & \sqrt{\frac{7}{6}}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \sqrt{\frac{1}{3}} & 0\\ 0 & 0 & \sqrt{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ -3 & 1 & 0\\ -2 & -1 & 1\end{bmatrix}\\
&=\begin{bmatrix}\sqrt{2} & 0 & 0\\ \sqrt{2}-3 & 1 & 0\\ \frac{1}{2}\sqrt{2}-\frac{1}{3}\sqrt{21}-1 & \frac{1}{3}-\frac{1}{6}\sqrt{21} & \frac{1}{6}\sqrt{21}\end{bmatrix}.\end{aligned}$$

El lector puede verificar que $A = P^tBP$.

Teorema 3.17. Una matriz simétrica A de tamaño $n \times n$ tiene todos sus valores propios positivos ($\lambda_i > 0$) si y sólo si

$$A = P^tP, \tag{3.22}$$

donde P es no singular.

Demostración

Si A es simétrica y todos sus valores propios son positivos, entonces puede escribirse en la forma

$$A = QDQ^t = \left(QD^{\frac{1}{2}}\right)\left(D^{\frac{1}{2}}Q^t\right) = P^tP$$

con $P = D^{\frac{1}{2}}Q^t$ y $D^{\frac{1}{2}} = diag\left\{\sqrt{\lambda_i}\right\}$ definida como en (3.10).

3.3.2. Descomposición de Cholesky

Entre los tipos de factorizaciones para la matriz A, existe una descomposición especial para aquellas matrices cuadradas cuyos valores propios son todos positivos conocida como *Descomposición de Cholesky*, la cual es considerada en este apartado.

Teorema 3.18. Descomposición de Cholesky

Si A es una matriz simétrica de tamaño $n \times n$ con todos sus valores propios positivos, entonces existe una matriz L triangular inferior tal que

$$A = LL^t, \tag{3.23}$$

donde todos los elementos en la diagonal principal de L son positivos.

Demostración

Por el Teorema 3.5 la matriz A se puede expresar como

$$\begin{aligned} A =& LDL^t = \left(LD^{\frac{1}{2}}\right)\left(D^{\frac{1}{2}}L^t\right) \\ =& \left(LD^{\frac{1}{2}}\right)\left(LD^{\frac{1}{2}}\right)^t, \end{aligned}$$

donde $D^{\frac{1}{2}}$ está definida como en (3.10) y la prueba queda completa.

Procedimiento para encontrar los elementos de $R = D^{\frac{1}{2}}L^t$

- Para $i = 1$, se tiene

$$r_{1j} = \begin{cases} \sqrt{a_{11}} & j = 1, \\ r_{11}^{-1}a_{1j} & j > 1. \end{cases}$$

- Cuando $i > 1$, se obtiene

$$r_{ij} = \begin{cases} 0 & i > j, \\ \sqrt{a_{ii} - \sum\limits_{k=1}^{i-1} r_{ki}^2} & i = j, \\ r_{ii}^{-1}\left(a_{ij} - \sum\limits_{k=1}^{i-1} r_{ki}r_{kj}\right) & i < j. \end{cases}$$

El procedimiento exige que estos elementos se calculen por filas, de izquierda a derecha y de arriba a abajo.

Si no se puede obtener la descomposición de Cholesky de una matriz (por ejemplo, cuando al realizar el procedimiento presentado arriba surge una raíz cuadrada de un número negativo) esto es indicio de que la matriz simétrica no tiene todos sus valores propios positivos.

Ejemplo 3.13. Encuentre la descomposición de Cholesky para la matriz simétrica dada en el Ejemplo 3.3.

Solución

Usando el procedimiento descrito anteriormente, se tiene que

$$\begin{aligned} r_{11} &= \sqrt{a_{11}} = 1, & r_{12} &= r_{11}^{-1} a_{12} = a_{12} = 3, \\ r_{13} &= r_{11}^{-1} a_{13} = a_{13} = 5, & r_{22} &= \sqrt{a_{22} - r_{12}^2} = \sqrt{3}, \\ r_{23} &= r_{22}^{-1} \left(a_{23} - r_{12} r_{13}\right) = \sqrt{3}, & r_{33} &= \sqrt{a_{33} - r_{13}^2 - r_{23}^2} = \sqrt{2}. \end{aligned}$$

Luego,

$$R = \begin{bmatrix} 1 & 3 & 5 \\ 0 & \sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{2} \end{bmatrix}.$$

El lector puede verificar que $A = R^t R$.

Ejercicios 3.2.

1. Para cada una de las siguientes matrices determine (en caso de ser posible) una raíz cuadrada

$$a. \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \qquad b. \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \qquad c. \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}.$$

$$d. \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}. \qquad e. \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}. \qquad f. \begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}.$$

$$g. \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}. \qquad h. \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

2. Determine la descomposición de Cholesky para las siguientes matrices

$$a.\begin{bmatrix}1 & 1\\ 1 & 3\end{bmatrix}. \qquad b.\begin{bmatrix}1 & 3\\ 3 & -1\end{bmatrix}. \qquad c.\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}.$$

$$d.\begin{bmatrix}1 & 3 & -2\\ 3 & 1 & 2\\ -2 & 2 & 1\end{bmatrix}. \qquad e.\begin{bmatrix}3 & -1 & 4\\ -1 & 1 & -5\\ 4 & -5 & 2\end{bmatrix}.$$

3.4. Polinomio mínimo

El polinomio característico de una matriz es un instrumento para calcular sus valores propios. En esta sección se estudia el polinomio mínimo de matrices, el cual resulta muy útil para establecer criterios sobre la posibilidad de reducir matrices a formas canónicas simples.

Definición 3.3. Polinomios de matrices

Si A es una matriz real de tamaño $n \times n$, el polinomio $p_n(A)$ denota la matriz que se genera si se reemplaza cada aparición de x en $p_n(x)$ por la matriz A:

$$p_n(A) = a_n A^n + a_{n-1}A^{n-1} + \ldots + a_1 A + a_0 A^0,$$

donde $a_i \in \mathbb{R}$ $(i = 0, 1, \ldots, n)$ y $A^0 = I_n$.

En consecuencia, se dice que A satisface el polinomio $p_n(x)$ si $p_n(A) = O$.

Ejemplo 3.14. El polinomio $p_2(x) = x^2 - 3x - 28$ es satisfecho por la matriz

$$A = \begin{bmatrix}1 & -5\\ -6 & 2\end{bmatrix}.$$

Solución

Si la matriz A satisface dicho polinomio se debe verificar que

$$P_2(A) = A^2 - 3A - 28I = O.$$

Como

$$A^2 = \begin{bmatrix} 31 & -15 \\ -18 & 34 \end{bmatrix},$$

luego

$$A^2 - 3A - 28I = \begin{bmatrix} 31 & -15 \\ -18 & 34 \end{bmatrix} - 3\begin{bmatrix} 1 & -5 \\ -6 & 2 \end{bmatrix} - 28\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Teorema 3.19. Teorema de Cayley-Hamilton

Sean A una matriz de tamaño $n \times n$ y

$$p_A(\lambda) = \det(A - \lambda I_n) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0 = 0$$

su polinomio característico. Entonces $p_A(A) = O$. Es decir, A satisface la ecuación

$$A^n + c_{n-1}A^{n-1} + \ldots + c_1 A + c_0 A^0 = O.$$

El teorema es verdadero para cualquier matriz, sin embargo la prueba que se presenta en estas notas es únicamente para matrices diagonalizables. La demostración para el caso general puede verse en Apostol (1985), pág. 249.

Demostración

Supongamos que A es diagonalizable. Como $p_A(\lambda)$ es una ecuación escalar, al multiplicarlo por cualquier vector $\vec{v} \in \mathbb{R}^n$, se tiene

$$\begin{aligned} p_A(\lambda)\vec{v} &= \left(\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0\right)\vec{v} = \vec{0} \\ &= \lambda^n\vec{v} + c_{n-1}\lambda^{n-1}\vec{v} + \ldots + c_1\lambda\vec{v} + c_0\vec{v} = \vec{0}. \end{aligned}$$

Si $\vec{v}$ es un vector propio correspondiente al valor propio λ, se cumple

$$p_A(\lambda)\vec{v} = A^n\vec{v} + c_{n-1}A^{n-1}\vec{v} + \ldots + c_1 A\vec{v} + c_0 I_n\vec{v} = \vec{0}. \tag{3.24}$$

Esto se cumple para todos los vectores propios de A. Pero como A es diagonalizable tiene n-vectores propios linealmente independientes, luego, cualquier otro vector de $\mathbb{R}^n$ puede ser expresado como combinación lineal de estos. Por lo tanto (3.24) se cumple para todo vector de $\mathbb{R}^n$. De aquí,

$$A^n + c_{n-1}A^{n-1} + \ldots + c_1 A + c_0 I_n = O,$$

es decir, A satisface su propia ecuación característica.

Sea A una matriz real de tamaño $n \times n$ y definamos

$$S = \{p_n(x) | p_n(A) = 0\}, \qquad n \geq 1.$$

Entonces, en S se puede escoger un polinomio no nulo $q(x)$ que tenga grado mínimo, además se puede suponer que el coeficiente de $q(x)$ correspondiente al término de mayor grado es 1, es decir, que $q(x)$ es mónico. Con estas condiciones para $q(x)$ se puede demostrar que cualquier otro polinomio $p(x)$ de S es múltiplo de $q(x)$. Esto implica que si en S existiera otro polinomio $r(x)$ mónico del mismo grado de $q(x)$, es decir de grado mínimo, entonces $r(x) = q(x)$.

Definición 3.4. El polinomio mínimo de una matriz A es el polinomio no nulo de menor grado que es satisfecho por A. Se denotará por $m_A(x)$.

Como la multiplicidad algebraica de los valores propios de una matriz es a veces distinta de uno, el polinomio característico $p_A(x)$ no es necesariamente el polinomio de grado mínimo satisfecho por A.

Teorema 3.20. El polinomio mínimo de una matriz cuadrada A es único, cuando se impone el coeficiente del término de mayor exponente en la indeterminada que sea igual a la unidad.

Demostración

La prueba se realiza por contradicción, supongamos que $m_A(x)$ y $m_A'(x)$ son polinomios mínimos de A. Por la definición 3.4, ambos tienen el mismo grado. Al considerar los coeficientes dominantes respectivos iguales a la unidad, el polinomio $d(x) = m_A(x) - m_A'(x)$ tiene grado menor que los polinomios mínimos y se anula también para A, necesariamente $d(x) = 0$. Luego, $m_A(x) = m_A'(x)$.

Teorema 3.21. Si A satisface un polinomio $p_n(x)$, entonces $p_n(x)$ es divisible por $m_A(x)$, polinomio mínimo de A.

Demostración

Sea $m_A(x)$ el polinomio mínimo de A, el grado de $p_n(x)$ es mayor que el de $m_A(x)$. Por el algoritmo de la división, sabemos que se pueden encontrar polinomios $q(x)$ y $r(x)$, tales que

$$p_n(x) = m_A(x)q(x) + r(x),$$

con $grad\,[r(x)] < grad\,[m_A(x)]$. Entonces, $r(x) = p_n(x) - m_A(x)q(x)$ y, como $p_n(A) = O$, $m_A(A) = O$. Se tiene que $r(A) = O$. Luego $r(x)$ es el polinomio mínimo de A, lo cual contradice la hipótesis. Como $grad\,[r(x)] < grad\,[m_A(x)]$, se debe tener que $r(x) = 0$. Por lo tanto, $m_A(x)$ es un factor de $p_n(x)$.

Teorema 3.22. Matrices semejantes tienen el mismo polinomio mínimo.

Demostración

Sean A y B matrices semejantes, luego existe una matriz P tal que $A = PBP^{-1}$. Sea $m_A(x)$ el polinomio mínimo de A. Entonces

$$\begin{aligned}
m_A(A) =& c_mA^m + c_{m-1}A^{m-1} + \ldots + c_kA^k + \ldots + c_1A + c_0I = O\\
=& c_m\left(PBP^{-1}\right)^m + c_{m-1}\left(PBP^{-1}\right)^{m-1} + \ldots + c_k\left(PBP^{-1}\right)^k + \ldots +\\
& c_1\left(PBP^{-1}\right) + c_0I = O\\
=& c_m\left(PB^mP^{-1}\right) + c_{m-1}\left(PB^{m-1}P^{-1}\right) + \ldots + c_k\left(PB^kP^{-1}\right) + \ldots +\\
& c_1\left(PBP^{-1}\right) + c_0\left(PP^{-1}\right) = O\\
=& P\left(c_mB^m + c_{m-1}B^{m-1} + \ldots + c_kB^k + \ldots + c_1B + c_0I\right)P^{-1} = O\\
=& P\left(m_A(B)\right)P^{-1} = O.
\end{aligned}$$

Esto implica que $m_A(B) = O$, es decir, $m_A(x)$ es también el polinomio mínimo de B, pues por el Teorema 3.21 si hubiese un divisor que se anulase para B, también se anularía para A, lo que contradice la hipótesis de que $m_A(x)$ es el polinomio mínimo de A. Luego los polinomios mínimos de A y B son iguales.

El siguiente teorema establece una importante relación entre el polinomio característico y el polinomio mínimo.

Teorema 3.23. Si A satisface un polinomio $p_n(x)$, entonces todo valor propio de A es también raíz de $p_n(x)$. Por consiguiente, todo valor propio de A es una raíz del polinomio mínimo de A.

Demostración

Supóngase que A satisface $p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 x^0$. Si λ es un valor propio de A, entonces se puede encontrar un vector propio $\vec{v}$ tal que $A\vec{v} = \lambda\vec{v}$. Así, $A^2\vec{v} = A(A\vec{v}) = A(\lambda\vec{v}) = \lambda(A\vec{v}) = \lambda(\lambda\vec{v}) = \lambda^2\vec{v}$.
Si se continua de esta manera, se origina

$$A^k\vec{v} = \lambda^k\vec{v}, \qquad \text{para todo } k > 0.$$

Pero dado que $p_n(A) = 0$, se tiene

$$\begin{aligned}\vec{0} =& p_n(A)\vec{v} = \left(a_n A^n + a_{n-1}A^{n-1} + \ldots + a_1 A + a_0 I\right)\vec{v} \\ =& (a_n A^n)\vec{v} + \left(a_{n-1}A^{n-1}\right)\vec{v} + \ldots + (a_1 A)\vec{v} + (a_0 I)\vec{v} \\ =& a_n\lambda^n\vec{v} + a_{n-1}\lambda^{n-1}\vec{v} + \ldots + a_1\lambda\vec{v} + a_0\vec{v} \\ =& \left(a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0\lambda^0\right)\vec{v} = p_n(\lambda)\vec{v}\end{aligned}$$

como $p_n(\lambda)$ es un escalar y $p_n(\lambda)\vec{v} = \vec{0}$ con $\vec{v} \neq \vec{0}$, se concluye que $p_n(\lambda) = 0$. Por ello, λ es una raíz de $p_n(x)$.

Puesto que A satisface su polinomio mínimo, todo valor propio de A es raíz de dicho polinomio mínimo.

Ejemplo 3.15. Encuentre el polinomio minimal que satisface la matriz dada en el

Ejemplo 2.9.

Solución

En el Ejemplo 2.9, se obtuvo que los valores propios de A eran $\lambda_1 = 0$, $\lambda_2 = 2$ (de multiplicidad algebraica 2). Como la multiplicidad algebraica del valor propio λ_2 resulto ser igual a su multiplicidad geométrica, el polinomio

$$p(x) = (x - 0)(x - 2) = x^2 - 2x$$

es satisfecho por la matriz A. Veamos que

$$p(A) = A^2 - 2A = O.$$

Luego

$$\begin{aligned}A^2 - 2A =& \begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}^2 - 2\begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix} \\ =& \begin{bmatrix} 6 & 4 & -6 \\ -6 & -8 & 18 \\ -2 & -4 & 10 \end{bmatrix} - \begin{bmatrix} 6 & 4 & -6 \\ -6 & -8 & 18 \\ -2 & -4 & 10 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.\end{aligned}$$

Ejemplo 3.16. Encuentre el polinomio minimal que satisface la matriz dada en el Ejemplo 2.15.

Solución

En el Ejemplo 2.15, se obtuvo que los valores propios de A eran $\lambda_1 = 1$ (de multiplicidad algebraica 2) y $\lambda_2 = 7$. Como la multiplicidad algebraica del valor propio λ_2 resulto ser igual a su multiplicidad geométrica, el polinomio minimal es

$$m_A(x) = (x-1)(x-7) = x^2 - 8x + 7.$$

El lector puede verificar que la matriz A satisface este polinomio.

3.5. Forma canónica de Jordan

Ahora se considera una versión completa del teorema de Hamilton− Cayley, el cual ya se había estudiado antes. Esta nueva versión será usada para la forma canónica de Jordan que estudiaremos ahora. La forma de Jordan utiliza todo el material estudiado en los capítulos precedentes.

Definición 3.5. Bloque de Jordan

Una matriz triangular superior de tamaño $r \times r$, $\mathcal{J}_r(\lambda)$, es un *bloque elemental de Jordan* si se verifica que

i) Todos sus elementos en la diagonal principal son iguales a λ.

ii) Todos sus elementos en la primera sobrediagonal son iguales a 1.

iii) Todos los demás elementos son iguales a 0.

De este modo, $\mathcal{J}_r(\lambda)$ es de la forma

$$\mathcal{J}_r(\lambda) = \begin{bmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{bmatrix} = \lambda I_r + N_r \qquad \text{con} \qquad N_r = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix},$$

en donde la matriz N_r es nilpotente, es decir, $N_r^m = 0$ para algún $m \geq 1$.

Como una matriz de Jordan está constituida por bloques elementales, su definición es la siguiente

Definición 3.6. Matriz de Jordan

Una matriz $\mathcal{J}$ de tamaño $n \times n$, de la forma

$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_{r_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{J}_{r_t} \end{bmatrix},$$

en donde, $\mathcal{J}_{r_1}, \mathcal{J}_{r_2}, \dots, \mathcal{J}_{r_t}$ son bloques elementales de Jordan de órdenes $r_1 \geq r_2 \geq \cdots \geq r_t$ con $t \geq 1$ se denomina *matriz de Jordan*.

Teorema 3.24. Sea A una matriz real de tamaño $n \times n$. Entonces existe una matriz P no singular tal que

$$P^{-1}AP = \begin{bmatrix} \mathcal{J}_{r_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{J}_{r_\mu} \end{bmatrix} = \mathcal{J}, \tag{3.25}$$

en donde cada $\mathcal{J}_{r_k}$ es un bloque de Jordan de tamaño $r_k \times r_k$ y el subíndice $\mu = \mu_1 + \mu_2 + \ldots + \mu_s$ es igual a la suma de las multiplicidades geométricas de los valores propios distintos de A. Un mismo valor propio λ_k puede estar en distintos bloques de Jordan $\mathcal{J}_{r_k}$, pero el número total de bloques con ese valor propio es igual a su multiplicidad geométrica μ_k, mientras que el número total de elementos en la diagonal principal con ese valor propio es igual a su multiplicidad algebraica m_k. Los números r_k y el número total de bloques quedan determinados de manera única por la matriz A.

Demostración

Para identificar los pasos a seguir en la demostración, supongamos que A es una matriz real de tamaño 2×2 que tiene un único valor propio λ. Sea $\vec{u}_1$ el único vector propio correspondiente a λ. Entonces A no es diagonalizable. Veamos que

$$P^{-1}AP = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

Como la multiplicidad algebraica es diferente a la multiplicidad geométrica se encuentra un vector propio generalizado $\vec{u}_2$. Por el Teorema 2.30 los vectores $\vec{u}_1$ y $\vec{u}_2$ son linealmente independientes, luego $P = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}$ es no singular. Por lo tanto,

$$AP = A\begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} = \begin{bmatrix} A\vec{u}_1 & A\vec{u}_2 \end{bmatrix} = \begin{bmatrix} \lambda\vec{u}_1 & A\vec{u}_2 \end{bmatrix}.$$

Pero de la ecuación (2.29), se tiene que $A\vec{u}_2 = \lambda\vec{u}_2 + \vec{u}_1$ de manera que

$$AP = \begin{bmatrix} \lambda\vec{u}_1 & \lambda\vec{u}_2 + \vec{u}_1 \end{bmatrix}.$$

Por otra parte,

$$\begin{aligned} P\mathcal{J} &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} \lambda\vec{u}_1 & \vec{u}_1 + \lambda\vec{u}_2 \end{bmatrix}. \end{aligned}$$

Por lo tanto, $AP = P\mathcal{J}$, lo que significa que $P^{-1}AP = \mathcal{J}$. Luego el teorema es válido para matrices de tamaño 2×2.

Para probar el teorema para el caso general, se escribe P en términos de sus columnas como

$$P = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix},$$

donde algunos $\vec{u}_j$ son vectores propios generalizados. Consideremos $\vec{u}_i$ como elemento de una *hilera de vectores* encabezados por algún vector propio $\vec{u}_{i-1}$ y descritos por

$$A\vec{u}_{i-1} = \lambda_{i-1}\vec{u}_{i-1} \qquad \text{y} \qquad A\vec{u}_i = \lambda_{i-1}\vec{u}_i + \alpha_i\vec{u}_{i-1}. \tag{3.26}$$

En otras palabras,

$$A\begin{bmatrix}\vec{u}_{i-1} & \vec{u}_i\end{bmatrix} = \begin{bmatrix}\vec{u}_{i-1} & \vec{u}_i\end{bmatrix}\begin{bmatrix}\lambda_{i-1} & \alpha_i \\ 0 & \lambda_{i-1}\end{bmatrix},$$

donde λ_{i-1} es el valor propio en el bloque de Jordan que afecta a $\vec{u}_i$ y α_i es igual a 0 o a 1. Luego, la clave para encontrar la forma de Jordan de A se convierte en la búsqueda de las hileras de vectores definidas en (3.26). Además, nótese que cada hilera produce un solo bloque en la matriz $\mathcal{J}$. Esencialmente, se tiene que mostrar de qué manera se pueden construir estas hileras para cada matriz $A \in \mathcal{M}_{nn}$.

Para ello, se procede por inducción matemática, partiendo del hecho de que cada matriz de tamaño 1×1 está ya en su forma de Jordan. La prueba consiste en suponer que se logra la construcción para todas las matrices de orden menor que n (ésta es la “hipótesis de inducción”) y después se aplican tres pasos para obtener la forma de Jordan de una matriz de tamaño $n \times n$. Los pasos que se aplican son

i) Se supone que A es singular, entonces su espacio columna tiene dimensión $r < n$. En lo que respecta solamente a este espacio pequeño, la hipótesis de inducción garantiza que una forma de Jordan es posible, luego, deben haber r vectores linealmente independientes $\vec{v}_i$ en el espacio columna tales que

$$A\vec{v}_i = \lambda_i\vec{v}_i \qquad \text{o} \qquad A\vec{v}_i = \lambda_i\vec{v}_i + \vec{v}_{i-1}. \tag{3.27}$$

ii) Se asume que el espacio nulo y el espacio columna de A tienen una intersección de dimensión p. Luego, cada vector del espacio nulo es también un vector propio correspondiente al valor propio $\lambda = 0$. Por lo tanto, se tienen p hileras en el paso *i*) que comienzan a partir de este valor propio y nos interesan los vectores $\vec{v}_i$ que están al final de dichas hileras. Puesto que cada uno de estos p vectores está en el espacio columna, éstos se pueden expresar como una combinación de las columnas de A, es decir

$$\vec{v}_i = A\vec{w}_i, \qquad \text{para algún} \qquad \vec{w}_i.$$

iii) Se considera que el espacio nulo tiene dimensión $n-r$. Entonces, independientemente de su intersección $p-$dimensional con el espacio columna, debe contener $n-r-p$ vectores básicos adicionales $\vec{y}_i$ fuera de esa intersección.

Juntemos estos pasos para obtener el teorema de Jordan.

Los r vectores $\vec{v}_i$, los p vectores $\vec{w}_i$ y los $n-r-p$ vectores $\vec{y}_i$ forman las hileras de Jordan para la matriz A y estos vectores son linealmente independientes. Si estos vectores conforman las columnas de la matriz P, entonces P es no singular y $P^{-1}AP = \mathcal{J}$ está en la forma de Jordan.

Si se quiere renumerar estos vectores como $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ y hacerlos corresponder con las condiciones de la ecuación (3.26), entonces cada $\vec{w}_i$ debería insertarse inmediatamente después del $\vec{v}_i$ del cual proviene; esto completa una hilera en la cual $\lambda_i = 0$. Los $\vec{y}_i$ vienen al final, cada uno solo en su propia hilera; nuevamente el valor propio es cero, ya que los $\vec{y}_i$ están en el espacio nulo. Los bloques que tienen valores propios diferentes de cero se terminaron en el paso $i)$, los bloques con valores propios iguales a cero crecen en una fila y una columna en el paso $ii)$, y el paso $iii)$ contribuye con cualquiera de los bloques de tamaño 1×1 $\mathcal{J}_i = [0]$.

En esta construcción, el único punto técnico es verificar la independencia de toda la colección $\vec{v}_i$, $\vec{w}_i$ y $\vec{y}_i$. Suponemos, por lo tanto, que alguna combinación es cero

$$\sum_{i=1}^{r} \alpha_i \vec{v}_i + \sum_{i=1}^{p} \beta_i \vec{w}_i + \sum_{i=1}^{n-r-p} \gamma_i \vec{y}_i = \vec{0}. \tag{3.28}$$

Si se premultiplica por A y se usan las relaciones dadas en (3.27) para $\vec{v}_i$

$$\sum_{i=1}^{r} \alpha_i \begin{bmatrix} \lambda_i \vec{v}_i \\ \text{o} \\ \lambda_i \vec{v}_i + \vec{v}_{i-1} \end{bmatrix} + \sum_{i=1}^{p} \beta_i A \vec{w}_i = \vec{0}, \tag{3.29}$$

dado que los $A\vec{w}_i$ son los $\vec{v}_i$ especiales al final de las hileras correspondientes a $\lambda_i = 0$, no pueden aparecer en la primera suma. Además (3.29) es un tipo de combinación de los $\vec{v}_i$, que son independientes por la hipótesis de inducción (proporcionan la forma de Jordan en el espacio columna). Se concluye que cada β_i debe ser cero. Si se reemplaza en (3.28), se llega a

$$\sum_{i=1}^{r} \alpha_i \vec{v}_i = -\sum_{i=1}^{n-r-p} \gamma_i \vec{y}_i.$$

Como el lado izquierdo está en el espacio columna y los $\vec{y}_i$ son independientes de ese espacio, cada γ_i debe ser cero. Por lo tanto,

$$\sum_{i=1}^{r} \alpha_i \vec{v}_i = \vec{0}$$

y de la independencia de los $\vec{v}_i$ se tiene que $\alpha_i = 0$.

Si la matriz A inicial no es singular, entonces se pueden aplicar los tres pasos a $\widetilde{A} = A - cI_n$. (Si se elige la constante c de manera que $\widetilde{A}$ sea singular y que pueda ser cualquiera de los valores propios de A). El algoritmo pone $\widetilde{A}$ en su forma de Jordan $P^{-1}\widetilde{A}P = \widetilde{\mathcal{J}}$, al producir las hileras $\vec{u}_i$ de las $\vec{v}_i$, $\vec{w}_i$ y $\vec{y}_i$. Entonces la forma de Jordan de A utiliza las mismas hileras y la misma P

$$P^{-1}AP = P^{-1}\left(\widetilde{A} + cI_n\right)P = \widetilde{\mathcal{J}} + P^{-1}cP = \widetilde{\mathcal{J}} + cI_n = \mathcal{J}.$$

Esto completa la demostración de que cada A es similar a alguna matriz de Jordan $\mathcal{J}$. Excepto por el reordenamiento de los bloques, *es similar a solo una* $\mathcal{J}$. En este sentido A tiene una forma de Jordan única.

Definición 3.7. Forma Canónica de Jordan

La matriz $\mathcal{J}$ dada en el Teorema 3.24 se denomina *forma canónica de Jordan* de A.

Ejemplo 3.17.

Encuentre una matriz no singular P tal que la matriz A dada a continuación sea semejante a una matriz de Jordan.

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

Solución

En el Ejemplo 2.19 se obtuvieron los siguientes vectores propios generalizados para la matriz A

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Por lo tanto,

$$P = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix} 0 & 0 & -1 \\ 3 & -2 & -2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Al realizar el producto $P^{-1}AP$ se obtiene

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

Además en la diagonal de la matriz de Jordan aparecen los valores propios de A.

Ejemplo 3.18.

Encuentre una matriz P no singular tal que $P^{-1}AP$ sea una matriz de Jordan, para

$$A = \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix}.$$

Solución

La ecuación característica de A es

$$-\lambda^3 + 3\lambda^2 - 3\lambda + 1 = 0,$$

como $\lambda = 1$ es el único vector propio, se tiene que su multiplicidad algebraica es tres. Entonces,

$$(A - \lambda I)\vec{v} = (A - I)\vec{v} = \begin{bmatrix} 4 & -3 & -2 \\ 8 & -6 & -4 \\ -4 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Esto conduce a $4x - 3y = 2z$. Estableciendo $y = 2$, se obtienen los vectores propios linealmente independientes: $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_3$ se calcula

$$(A - I)\vec{v}_3 = c_1\vec{v}_1 + c_2\vec{v}_2$$

$$\begin{bmatrix} 4 & -3 & -2 \\ 8 & -6 & -4 \\ -4 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}.$$

Al realizar las operaciones por filas se obtiene

$$\begin{bmatrix} 4 & -3 & -2 & | & c_1 \\ 8 & -6 & -4 & | & 2c_2 \\ -4 & 3 & 2 & | & 2c_1-3c_2 \end{bmatrix} \underset{\substack{F_2-2F_1 \\ F_3+F_2}}{\sim} \begin{bmatrix} 4 & -3 & -2 & | & c_1 \\ 0 & 0 & 0 & | & 2c_2-2c_1 \\ 0 & 0 & 0 & | & 3c_1-3c_2 \end{bmatrix}.$$

Para poder tener infinitas soluciones $c_1 = c_2$; es decir, $4x-3y-2z = c_1$. Por lo tanto, $z = 2x - \frac{3}{2}y - \frac{1}{2}c_1$. Si se hace $x = y = 0$ y $c_1 = 1$, se obtiene el vector propio generalizado: $\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{2} \end{bmatrix}$. Por consiguiente,

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & -1 & -\frac{1}{2} \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 4 & -3 & -2 \end{bmatrix}.$$

Obsérvese que la segunda columna de P es una combinación lineal de los dos vectores propios que constituyen la base del espacio propio asociado al valor propio $\lambda = 1$. Si se realiza el producto $P^{-1}AP$ se llega a

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 4 & -3 & -2 \end{bmatrix} \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & -1 & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

El lector puede notar que sobre la diagonal de la matriz de Jordan se encuentra el valor propio de la matriz A.

Ejercicios 3.3.

Calcular el polinomio mínimo y la descomposición de Jordan de las siguientes matrices:

$$a. \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}. \qquad b. \begin{bmatrix} 1 & -1 \\ 3 & -1 \end{bmatrix}. \qquad c. \begin{bmatrix} -1 & 3 \\ -1 & 1 \end{bmatrix}.$$

$$d. \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \qquad f. \begin{bmatrix} 2 & 2 & -3 \\ -3 & 0 & 3 \\ 4 & 1 & -5 \end{bmatrix}. \qquad g. \begin{bmatrix} -5 & 2 & 4 \\ -3 & 0 & 3 \\ -3 & 1 & 2 \end{bmatrix}.$$

3.6. Descomposición en valores singulares

Una factorización especial para cualquier matriz A de tamaño $m \times n$ es la *descomposición en valores singulares* (SVD, por sus siglas en inglés), la cual es una de las factorizaciones de matrices más útiles en Álgebra Lineal aplicada.

Definición 3.8. Valores singulares de matrices cuadradas

Los valores singulares de una matriz real A de tamaño $n \times n$ son las raíces cuadradas de los valores propios asociados a la matriz simétrica A^tA (listados con sus multiplicidades algebraicas). Estos valores se denotan por $\sigma_1, \sigma_2, \ldots, \sigma_n$, y se colocan en orden decreciente:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0,$$

donde $\sigma_i = \sqrt{\lambda_i}$ para $1 \leq i \leq n$.

Ejemplo 3.19. Determine los valores singulares de la matriz

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

Solución

La matriz A^tA es

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 1 & 1 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 5 & 3 \\ 5 & 11 & -3 \\ 3 & -3 & 9 \end{bmatrix}.$$

En este caso, la ecuación característica es:

$$\det\left(A^tA - \lambda I\right) = -\lambda^3 + 31\lambda^2 - 276\lambda + 576 = 0.$$

Entonces, los valores propios de A^tA son $\lambda_1 = 16, \lambda_2 = 12$ y $\lambda_3 = 3$. Por lo tanto, los valores singulares de la matriz A son $\sigma_1 = 4, \sigma_2 = 2\sqrt{3}$ y $\sigma_3 = \sqrt{3}$.

Cuando A es una matriz real de tamaño $n \times n$, sabemos que las matrices A^tA y AA^t tienen los mismos valores propios con las mismas multiplicidades algebraicas. Por lo tanto en la definición 3.8 se puede cambiar A^tA por AA^t. Mientras que si A es una matriz real de tamaño $m \times n$, con $m \neq n$, las matrices A^tA y AA^t tendrán n y m valores propios, respectivamente. Por consiguiente, cuando la matriz no sea cuadrada los valores singulares de la misma se definen de la siguiente manera.

Definición 3.9. Valores Singulares de Matrices Rectangulares

Sea A una matriz real de tamaño $m \times n$ ($m \neq n$), los valores singulares son las raíces cuadradas de los valores propios comunes a las matrices simétricas A^tA y AA^t.

Ejemplo 3.20. Encuentre los valores singulares de la matriz

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

Solución

La matriz A^tA es

$$\begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 0 \\ 0 & 3 \end{bmatrix},$$

cuyos valores propios son $\lambda_1 = 14$ y $\lambda_2 = 3$. La matriz AA^t es

$$\begin{bmatrix} 1 & -1 \\ 3 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -3 \\ 2 & 10 & -5 \\ -3 & -5 & 5 \end{bmatrix}.$$

En este caso, los valores propios de AA^t son $\lambda_1 = 14, \lambda_2 = 3$ y $\lambda_3 = 0$. Por lo tanto, los valores singulares de la matriz A son $\sigma_1 = \sqrt{14}$ y $\sigma_2 = \sqrt{3}$.

Teorema 3.25. Sea A una matriz real de tamaño $m \times n$ que tiene r valores singulares no nulos $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ con $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$, entonces el rango de A es r.

Demostración

Sea $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{R}^n$ formada por los vectores propios asociados a A^tA, ordenados de tal forma que los valores propios correspondientes a A^tA satisfacen que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Entonces,

$$\begin{aligned} A\vec{v}_i \cdot A\vec{v}_j =& (A\vec{v}_i)^t A\vec{v}_j = \vec{v}_i^t (A^t A\vec{v}_j) = \vec{v}_i^t \lambda_j \vec{v}_j \\ =& \sigma_j^2 (\vec{v}_i \cdot \vec{v}_j) = \begin{cases} 0 & \text{si } i \neq j, \\ \sigma_j^2 & \text{si } i = j. \end{cases} \end{aligned} \tag{3.30}$$

Luego $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_n\}$ es un conjunto ortogonal. Sea r el número de valores singulares no nulos de A, esto es, r es el número de valores propios no nulos de A^tA. De la expresión (3.30), se tiene que $A\vec{v}_i \neq \vec{0}$ si y sólo si $1 \leq i \leq r$. Entonces $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ son vectores linealmente independientes, los cuales claramente pertenecen al espacio columna de A ($Col(A)$). Además, para cualquier $\vec{y} \in Col(A)$ - digamos, $\vec{y} = A\vec{x}$ - se puede escribir $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$, y

$$\begin{aligned} \vec{y} = A\vec{x} &= c_1 A\vec{v}_1 + \ldots + c_r A\vec{v}_r + c_{r+1} A\vec{v}_{r+1} + \ldots + c_n A\vec{v}_n \\ &= c_1 A\vec{v}_1 + \ldots + c_r A\vec{v}_r + \vec{0} + \ldots + \vec{0} \end{aligned}$$

así que $\vec{y}$ está en el espacio generado por $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$, lo cual muestra que $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ es una base (ortogonal) para $Col(A)$. Por lo tanto, el $\rho(A) = r$.

3.6.1. Descomposición en valores singulares

La descomposición de A involucra una matriz "diagonal" S de tamaño $m \times n$ particionada como sigue

$$S = \begin{bmatrix} D & \vdots & O \\ \ldots & \cdot & \ldots \\ O & \vdots & O \end{bmatrix}, \qquad \text{con} \qquad D = \begin{bmatrix} \sigma_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_r \end{bmatrix}, \tag{3.31}$$

donde los σ_i, para $i = 1, 2, \ldots, r$ son los valores singulares no nulos de A y r no excede el más pequeño de m y n. (Si r es igual a m ó n ó a ambos, algunas o todas de las matrices nulas desaparecen).

Teorema 3.26. Descomposición en Valores Singulares

Sea A una matriz real de tamaño $m \times n$ con rango r. Entonces existen matrices ortogonales U y V de tamaño $m \times m$ y $n \times n$ respectivamente, tales que

$$A = USV^t, \tag{3.32}$$

donde S tiene la forma dada en la ecuación (3.31).

Demostración

Sean λ_i y $\vec{v}_i$ como en la prueba del Teorema 3.25. Entonces $\sigma_i = \sqrt{\lambda_i} = \|A\vec{v}_i\| > 0$ para $1 \leq i \leq r$, $r = \rho(A) \leq \text{mín}\{m,n\}$ y $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ es una base ortogonal para $Col(A)$. Si se normalizan cada uno de los vectores $A\vec{v}_i$, se puede definir

$$\vec{u}_i = \frac{1}{\|A\vec{v}_i\|} A\vec{v}_i = \frac{1}{\sigma_i} A\vec{v}_i, \qquad\qquad i = 1, 2, \ldots, r.$$

Luego el conjunto de vectores $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ es una base ortonormal para $Col(A)$, la cual se puede extender hasta obtenerse una base ortonormal $\{\vec{u}_1, \ldots, \vec{u}_r, \vec{u}_{r+1}, \ldots, \vec{u}_m\}$ de $\mathbb{R}^m$. A partir de la definición de los vectores $\vec{u}_i$, se puede escribir

$$A\vec{v}_i = \begin{cases} \sigma_i \vec{u}_i & \text{para } i = 1, 2, \ldots, r \\ 0\vec{u}_i & \text{para } i = r+1, r+2, \ldots, m \\ \vec{0} & \text{para } i = m+1, m+2, \ldots, n. \end{cases}$$

En forma matricial se expresa de la siguiente manera

$$\begin{aligned} AV &= \begin{bmatrix} A\vec{v}_1 & \ldots & A\vec{v}_r & A\vec{v}_{r+1} & \ldots & A\vec{v}_m & A\vec{v}_{m+1} & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\vec{u}_1 & \ldots & \sigma_r\vec{u}_r & 0\vec{u}_{r+1} & \ldots & 0\vec{u}_m & \vec{0} & \ldots & \vec{0} \end{bmatrix} \\ &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_r & \vec{u}_{r+1} & \ldots & \vec{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_r & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$AV = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_r & \vec{u}_{r+1} & \ldots & \vec{u}_m \end{bmatrix} \begin{bmatrix} D & \vdots & O \\ \cdots & . & \cdots \\ O & \vdots & O \end{bmatrix} = US \tag{3.33}$$

nótese que las columnas de la matriz ortogonal V (de tamaño $n \times n$), son los vectores propios ortonormalizados de la matriz A^tA. Las columnas de la matriz ortogonal U (de tamaño $m \times m$), son los vectores propios ortonormalizados de la matriz AA^t y la matriz S esta definida como en (3.31). Si se multiplica por el lado derecho de la ecuación (3.33) por V^{-1} ($V^{-1} = V^t$), se tiene que

$$A = USV^t.$$

Esto finaliza la demostración del teorema.

Ejemplo 3.21. Encuentre la descomposición en valores singulares de la matriz dada en el Ejemplo 3.19.

Solución

Del Ejemplo 3.19 se tiene que los valores singulares asociados a la matriz A son $\sigma_1^2 = 16,\ \sigma_2^2 = 12$ y $\sigma_3^2 = 3$. Al calcular los respectivos vectores propios unitarios de A^tA, se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-1\\2\end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix}-1\\1\\1\end{bmatrix}.$$

Por otra parte, la matriz AA^t es

$$AA^t = \begin{bmatrix}14 & 2 & 0\\2 & 14 & 0\\0 & 0 & 3\end{bmatrix}$$

y sus respectivos vectores propios unitarios son

$$\vec{u}_1 = \begin{bmatrix}\sqrt{2}/2\\\sqrt{2}/2\\0\end{bmatrix}, \qquad \vec{u}_2 = \begin{bmatrix}-\sqrt{2}/2\\\sqrt{2}/2\\0\end{bmatrix} \qquad \text{y} \qquad \vec{u}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Finalmente, si $U = \begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \vec{u}_3\end{bmatrix}$, $V = \begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \vec{v}_3\end{bmatrix}$ y $S = \begin{bmatrix}4 & 0 & 0\\0 & 2\sqrt{3} & 0\\0 & 0 & \sqrt{3}\end{bmatrix}$. Entonces

$$\begin{aligned}A &= \begin{bmatrix}\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & 0\\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}4 & 0 & 0\\0 & 2\sqrt{3} & 0\\0 & 0 & \sqrt{3}\end{bmatrix}\begin{bmatrix}\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0\\ \frac{1}{6}\sqrt{6} & -\frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{6}\\ -\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3}\end{bmatrix}\\ &= \begin{bmatrix}1 & 3 & -2\\3 & 1 & 2\\-1 & 1 & 1\end{bmatrix}.\end{aligned}$$

Ejemplo 3.22. Encuentre la descomposición en valores singulares de la matriz dada en el Ejemplo 3.20.

Solución

Haciendo referencia al Ejemplo 3.20 se tiene que los valores singulares asociados a la matriz A son $\sigma_1^2 = 14$ y $\sigma_3^2 = 3$. Al calcular los respectivos vectores propios unitarios de A^tA, se obtiene:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Por otra parte, los respectivos vectores propios unitarios de AA^t, son

$$\vec{u}_1 = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \qquad \vec{u}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_3 = \frac{1}{\sqrt{42}} \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}.$$

Si $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \end{bmatrix}$, $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ y $S = \begin{bmatrix} \sqrt{14} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix}$. Entonces

$$A = \begin{bmatrix} \frac{1}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{42}} \\ -\frac{2}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix} \begin{bmatrix} \sqrt{14} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 1 \\ -2 & 1 \end{bmatrix}.$$

3.6.2. Descomposición polar

Una consecuencia interesante y útil de la descomposición en valores singulares para una matriz *cuadrada* A es la *descomposición polar* de A.

Teorema 3.27. Descomposición Polar

Sea A una matriz real de tamaño $n \times n$ con rango r. Entonces existe una matriz simétrica P de tamaño $n \times n$ con valores propios no negativos y una matriz ortogonal Q de tamaño $n \times n$, tales que

$$A = PQ. \tag{3.34}$$

Demostración

Si A es una matriz real de tamaño $n \times n$, en una descomposición de valores singulares las matrices U, S y V son también de tamaño $n \times n$. En este caso se puede escribir la ecuación (3.32) como

$$A = USV^t = US(U^tU)V^t = (USU^t)UV^t = PQ$$

donde $P = USU^t$ es una matriz simétrica y $Q = UV^t$ es una matriz ortogonal. Se deja como ejercicio la comprobación de que P tiene valores propios no negativos.

Ejemplo 3.23. Encuentre una descomposición polar para la matriz dada en el Ejemplo 3.19.

Solución

En el Ejemplo 3.21 se obtuvo la descomposición de A en valores singulares mediante las matrices U, S y V

$$U = \frac{1}{2}\begin{bmatrix} \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad S = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \quad \text{y} \quad V = \frac{1}{6}\begin{bmatrix} 3\sqrt{2} & \sqrt{6} & -2\sqrt{3} \\ 3\sqrt{2} & -\sqrt{6} & 2\sqrt{3} \\ 0 & 2\sqrt{6} & 2\sqrt{3} \end{bmatrix}.$$

Si se definen

$$\begin{aligned} P = USU^t &= \left(\frac{1}{2}\right)^2 \begin{bmatrix} \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & 0 \\ -\sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2+\sqrt{3} & 2-\sqrt{3} & 0 \\ 2-\sqrt{3} & 2+\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \end{aligned}$$

y

$$\begin{aligned} Q = UV^t &= \frac{\sqrt{2}}{12} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 3\sqrt{2} & 0 \\ \sqrt{6} & -\sqrt{6} & 2\sqrt{6} \\ -2\sqrt{3} & 2\sqrt{3} & 2\sqrt{3} \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 3-\sqrt{3} & 3+\sqrt{3} & -2\sqrt{3} \\ 3+\sqrt{3} & 3-\sqrt{3} & 2\sqrt{3} \\ -2\sqrt{3} & 2\sqrt{3} & 2\sqrt{3} \end{bmatrix}. \end{aligned}$$

Entonces

$$A = \frac{1}{6}\begin{bmatrix} 2+\sqrt{3} & 2-\sqrt{3} & 0 \\ 2-\sqrt{3} & 2+\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 3-\sqrt{3} & 3+\sqrt{3} & -2\sqrt{3} \\ 3+\sqrt{3} & 3-\sqrt{3} & 2\sqrt{3} \\ -2\sqrt{3} & 2\sqrt{3} & 2\sqrt{3} \end{bmatrix} = \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Ejercicios 3.4.

1. Para cada una de las matrices dadas a continuación, encuentre una descomposición en valores singulares

$$a.\begin{bmatrix}1 & 1 & -4\\ 1 & 3 & -1\end{bmatrix}. \qquad b.\begin{bmatrix}1 & 3\\ 3 & -1\end{bmatrix}. \qquad c.\begin{bmatrix}a & -b\\ b & a\end{bmatrix}, \qquad a,b\in\mathbb{R}.$$

$$d.\begin{bmatrix}1 & -1 & 0\\ -1 & 1 & -1\end{bmatrix}. \qquad e.\begin{bmatrix}1 & 3\\ 3 & 1\\ -2 & 2\end{bmatrix}. \qquad f.\begin{bmatrix}3 & -1\\ -1 & 1\\ 4 & -5\end{bmatrix}.$$

2. Si A es una matriz real de tamaño $n\times n$ con valores singulares todos iguales a 1, muestre que A es ortogonal.

3. Si A es una matriz real de tamaño $n\times n$ ¿Cuál es el producto de sus valores singulares $\sigma_1.\sigma_2.\ldots.\sigma_n$?.

4. Si A es una matriz real de tamaño 2×2 y $\vec{u}\in\mathbb{R}^2$ es unitario, muestre que

$$\sigma_2 \leq \|A\vec{u}\| \leq \sigma_1,$$

donde $\sigma_1,\ \sigma_2$ son los valores singulares de A.

5. Si A es una matriz real de tamaño $m\times n$ y $\vec{v}\in\mathbb{R}^n$, muestre que

$$\sigma_n\|\vec{v}\| \leq \|A\vec{v}\| \leq \sigma_1\|\vec{v}\|,$$

donde $\sigma_1,\ \sigma_n$ son los valores singulares más grande y más pequeño de la matriz A, respectivamente.

Capítulo 4

Matrices complejas

En la sección 2.3 se consideraron matrices de componentes reales en las cuales los valores propios y vectores propios eran complejos. En este capítulo se desarrollará la teoría correspondiente a valores propios y vectores propios pero para matrices de componentes complejas y el objetivo principal es estudiar algunas factorizaciones para este tipo de matrices, de manera análoga a como vimos en el capítulo 3.

4.1. Clases especiales de matrices complejas

Los tipos especiales de matrices cuadradas complejas que se analizan a continuación son las hermitianas, antihermitianas y unitarias, por tener características particulares y por ser muy útiles en ingeniería y en especial en física atómica. Estas matrices generalizan las tres clases de matrices reales especiales simétricas, antisimétricas y ortogonales, respectivamente.

4.1.1. Matrices hermitianas

Recordemos que una matriz simétrica $A = [a_{ij}]$ con componentes reales es una matriz que tiene la propiedad de que $A = A^t$. Las matrices hermitianas (o hermíticas) son las análogas para el caso en el cual las componentes de la matriz son números complejos.

Definición 4.1. Matriz Hermitiana

Se dice que una matriz A de tamaño $n \times n$ es **hermitiana** si

$$A = A^H. \tag{4.1}$$

Ejemplo 4.1. Sea A la matriz de componentes complejas

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

comprobar que A es una matriz hermitiana.

Solución

$$\overline{A} = \begin{bmatrix} \overline{3} & \overline{4-5i} & \overline{3+2i} \\ \overline{4+5i} & \overline{1} & \overline{7+6i} \\ \overline{3-2i} & \overline{7-6i} & \overline{2} \end{bmatrix} = \begin{bmatrix} 3 & 4+5i & 3-2i \\ 4-5i & 1 & 7-6i \\ 3+2i & 7+6i & 2 \end{bmatrix}.$$

$$A^H = \overline{A}^t = \begin{bmatrix} 3 & 4-5i & 3+2i \\ 4+5i & 1 & 7+6i \\ 3-2i & 7-6i & 2 \end{bmatrix} = A.$$

Nótese que los elementos de la diagonal principal de una matriz hermitiana son números reales, ya que tienen que coincidir con sus conjugados.

Teorema 4.1. Sea A una matriz hermitiana, entonces para todos los vectores $\vec{x} \in \mathbb{C}^n$, $\vec{x}^H A\vec{x}$ es real.

Demostración

La demostración consiste en un cálculo directo

$$(\vec{x}^H A\vec{x})^H = \vec{x}^H A^H (\vec{x}^H)^H = \vec{x}^H A\vec{x}$$

pero como $\vec{x}^H A\vec{x}$ es una matriz hermitiana de tamaño 1×1 se concluye que es un número real.

Teorema 4.2. Si A es una matriz hermitiana, entonces sus valores propios son reales.

Demostración

Supongamos que λ es un valor propio y que $\vec{x}$ es un vector propio correspondiente. Es decir,

$$A\vec{x} = \lambda\vec{x}.$$

Si se premultiplica por $\vec{x}^H$, se obtiene

$$\vec{x}^H A\vec{x} = \lambda\vec{x}^H\vec{x}.$$

Pero por el Teorema 4.1, el lado izquierdo es real y la expresión del lado derecho $\vec{x}^H\vec{x} = |\vec{x}|^2 \neq 0$. Se concluye que, λ debe ser real.

Teorema 4.3. Sea A una matriz hermitiana de tamaño $n\times n$. Entonces los vectores propios correspondientes a valores propios distintos de A son ortogonales.

Demostración

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios asociados a valores propios distintos, digamos, λ_1 y λ_2. Es decir,

$$\begin{aligned} A\vec{v}_1 &= \lambda_1\vec{v}_1 && \text{y} && A\vec{v}_2 = \lambda_2\vec{v}_2 \\ \vec{v}_2^H A\vec{v}_1 &= \lambda_1\vec{v}_2^H\vec{v}_1 && \text{y} && \vec{v}_1^H A\vec{v}_2 = \lambda_2\vec{v}_1^H\vec{v}_2. \end{aligned}$$

Al tomar la transpuesta conjugada de la primera expresión, se tiene

$$\begin{aligned} (\vec{v}_2^H A\vec{v}_1)^H &= (\lambda_1\vec{v}_2^H\vec{v}_1)^H \\ \vec{v}_1^H A\vec{v}_2 &= \lambda_1\vec{v}_1^H\vec{v}_2. \end{aligned}$$

En la última expresión se usaron los hechos de que $A^H = A$ y λ_1 es real. Luego se tiene que

$$\lambda_2\vec{v}_1^H\vec{v}_2 = \lambda_1\vec{v}_1^H\vec{v}_2$$

Por lo tanto $(\lambda_1-\lambda_2)\vec{v}_1^H\vec{v}_2 = 0$. Pero $\lambda_1-\lambda_2 \neq 0$, así que $\vec{v}_1^H\vec{v}_2 = 0$. Esto es, $\vec{v}_1$ y $\vec{v}_2$ son ortogonales.

Teorema 4.4. Sea $A = [a_{ij}]$ una matriz con componentes complejas de tamaño $n\times n$, entonces

$$tr\left(AA^H\right) = 0 \qquad \text{si y sólo si} \qquad A = O. \tag{4.2}$$

En realidad, $tr\left(AA^H\right) > 0$ si $A \neq O$.

Demostración

Si $A = [a_{ij}]$, entonces $A^H = [b_{ij}]$, en donde $b_{ij} = \overline{a}_{ji}$. Si se define $C = AA^H$, se tiene

$$\begin{aligned} c_{ik} &= a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} \\ &= \sum_{j=1}^{n} a_{ij}b_{jk} = \sum_{j=1}^{n} a_{ij}\overline{a}_{kj}. \end{aligned}$$

En particular, las componentes de la diagonal de C están dadas por

$$c_{ii} = \sum_{j=1}^{n} a_{ij}\overline{a}_{ij} = \sum_{j=1}^{n} \left|a_{ij}\right|^2.$$

Por lo tanto,

$$tr(C) = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left|a_{ij}\right|^2 \right).$$

Como $\left|a_{ij}\right|^2 \geq 0$, la única forma en que esta suma puede ser cero es si cada $a_{ij} = 0$ para todo i y j. Esto significa que $A = O$.

4.1.2. Matrices anti-hermitianas

Como se ha visto, una matriz antisimétrica A es una matriz real que tiene la propiedad de que $A^t = -A$. Las matrices anti-hermitianas constituyen el análogo para el caso complejo.

Definición 4.2. Matriz anti-hermitiana

Se dice que una matriz A de tamaño $n \times n$ es **anti-hermitiana** si

$$A^H = -A. \tag{4.3}$$

Teorema 4.5. Si A es anti-hermitiana, entonces para todos los vectores complejos $\vec{z}$, $\vec{z}^H A \vec{z}$ es cero o imaginario puro[1].

[1]Un imaginario puro es un número complejo de la forma αi con α real

Demostración

La demostración consiste en un cálculo directo

$$(\vec{z}^H A\vec{z})^H = \vec{z}^H A^H (\vec{z}^H)^H = -\vec{z}^H A\vec{z}.$$

Si se expresa $\vec{z}^H A\vec{z} = \alpha + i\beta$; entonces la ecuación anterior se puede escribir como

$$\alpha - i\beta = -(\alpha + i\beta),$$

luego se debe tener que $\alpha = -\alpha$, así que $\alpha = 0$. Por lo tanto, se concluye que $\vec{z}^H A\vec{z}$ es un imaginario puro.

Teorema 4.6. Los valores propios de una matriz anti-hermitiana deben ser cero o imaginarios puros.

Demostración

Supongamos que λ es un valor propio y que $\vec{z}$ es el vector propio correspondiente. Es decir,

$$A\vec{z} = \lambda\vec{z}.$$

Si se premultiplica por $\vec{z}^H$, se obtiene

$$\vec{z}^H A\vec{z} = \lambda\vec{z}^H\vec{z}.$$

Pero por el Teorema 4.5, el lado izquierdo es cero o imaginario puro y el lado derecho $\vec{z}^H\vec{z} = |\vec{z}|^2$ es real y distinto de cero. Por lo tanto,

$$\lambda = \frac{\vec{z}^H A\vec{z}}{\vec{z}^H\vec{z}}$$

de donde λ debe ser cero o imaginario puro.

Teorema 4.7. Sea A una matriz anti-hermitiana de tamaño $n \times n$. Entonces los vectores propios asociados con valores propios distintos de A son ortogonales.

Demostración

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios de A que corresponden a valores propios distintos, digamos, λ_1 y λ_2. Para demostrar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned}
\overline{\lambda_1}\vec{v}_1 \cdot \vec{v}_2 =& (\lambda_1\vec{v}_1)^H\vec{v}_2 = (A\vec{v}_1)^H\vec{v}_2 && \textit{puesto que } \vec{v}_1 \textit{es un vector propio}\\
=& (\vec{v}_1^H A^H)\vec{v}_2 = -\vec{v}_1^H(A\vec{v}_2) && \textit{puesto que Aes anti-hermitiana}\\
=& -\vec{v}_1^H(\lambda_2\vec{v}_2) && \textit{puesto que } \vec{v}_2 \textit{es un vector propio}\\
=& -\lambda_2\vec{v}_1^H\vec{v}_2 = -\lambda_2\vec{v}_1 \cdot \vec{v}_2.
\end{aligned}$$

Por lo tanto, se tiene que $(\overline{\lambda_1}+\lambda_2)\vec{v}_1\cdot\vec{v}_2=0$. Pero como $\overline{\lambda_1}=-\lambda_1$ por ser imaginario puro, entonces $\overline{\lambda_1}+\lambda_2\neq 0$, así que $\vec{v}_1\cdot\vec{v}_2=0$.

Teorema 4.8. Si B es una matriz anti-hermitiana, entonces la matriz $A=iB$ es hermitiana. Análogamente, si A es hermitiana, entonces $B=iA$ es anti-hermitiana.

Demostración

Sea $B^H=-B$ y definamos $A=iB$, entonces

$$A^H=(iB)^H=(i)^H\,B^H=(-i)(-B)=A.$$

Esto prueba que A es hermitiana, de la misma manera se puede probar que $B=iA$ es anti-hermitiana cuando A es hermitiana.

4.1.3. Matrices unitarias

Recordemos que una matriz ortogonal A es una matriz real que tiene la propiedad de que $A^t=A^{-1}$. Las matrices unitarias es el análogo para el caso complejo.

Definición 4.3. Matriz unitaria

Una matriz cuadrada U de componentes complejas se dice que es una *matriz unitaria* si $U^HU=I$. En consecuencia, U es no singular y se tiene $U^{-1}=U^H$.

Teorema 4.9. Si U es una matriz unitaria, entonces sus valores propios son de módulo igual a 1.

Demostración

Sea λ un valor propio de U con vector propio asociado $\vec{v}$, es decir

$$U\vec{v}=\lambda\vec{v}. \tag{4.4}$$

Luego, tomando la transpuesta conjugada se tiene

$$\begin{aligned}(\overline{U}\,\overline{\vec{v}})^t &= \overline{\lambda}\,\overline{\vec{v}}^t\\ \overline{\vec{v}}^t\,\overline{U}^t &= \overline{\lambda}\,\vec{v}^H\\ \vec{v}^H\,U^H &= \overline{\lambda}\,\vec{v}^H.\end{aligned}$$

Pero como U es unitaria se tiene que

$$\vec{v}^H\, U^{-1} = \overline{\lambda}\, \vec{v}^H. \tag{4.5}$$

Si se multiplica por la derecha ambos lados de (4.5) por $U\vec{v}$ se obtiene

$$\begin{aligned} \vec{v}^H\, U^{-1}(U\vec{v}) &= \overline{\lambda}\, \vec{v}^H(U\vec{v}) \\ \vec{v}^H(U^{-1}U)\vec{v} &= \overline{\lambda}\, \vec{v}^H(\lambda\vec{v}) && \text{por (4.4)} \\ \vec{v}^H\vec{v} &= \overline{\lambda}\, \lambda\, \vec{v}^H\vec{v} \end{aligned}$$

pero como $\vec{v}^H\vec{v} \neq 0$; se concluye que $\overline{\lambda}\,\lambda = 1$. O sea, $|\lambda|^2 = 1$, así que $|\lambda| = 1$.

Teorema 4.10. Sea U una matriz unitaria de tamaño $n \times n$. Entonces los vectores propios asociados con valores propios distintos de U son ortogonales.

Demostración

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios de U que corresponden a valores propios distintos, digamos, λ_1 y λ_2. Para demostrar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned} (U\,\vec{v}_1)^H(U\,\vec{v}_2) &= (\lambda_1\,\vec{v}_1)^H(\lambda_2\,\vec{v}_2) && \textit{puesto que } \lambda_1,\ \lambda_2 \textit{ son valores propios} \\ (\vec{v}_1^H\,U^H)(U\vec{v}_2) &= \overline{\lambda}_1\,\vec{v}_1^H\lambda_2\,\vec{v}_2 && \textit{puesto que } U \textit{ es unitaria} \\ \vec{v}_1^H\,\vec{v}_2 &= \overline{\lambda}_1\,\lambda_2\,\vec{v}_1^H\,\vec{v}_2. \end{aligned}$$

Por lo tanto $\left(1 - \overline{\lambda}_1\,\lambda_2\right)\vec{v}_1 \cdot \vec{v}_2 = 0$. Pero como λ_1 es distinto a λ_2, entonces $\overline{\lambda}_1\,\lambda_2 \neq 1$, así que $\vec{v}_1 \cdot \vec{v}_2 = 0$.

4.1.4. Matrices normales

Definición 4.4. Matriz normal

Se dice que la matriz de componentes complejas N de tamaño $n \times n$ es **normal** si conmuta con N^H, es decir

$$NN^H = N^HN.$$

Ejemplo 4.2. Comprobar que las matrices complejas diagonales son normales.

Solución

Sea D la siguiente matriz diagonal de tamaño $n \times n$

$$D = diag\{\lambda_1, \lambda_2, \dots, \lambda_n\},$$

entonces

$$\begin{aligned} DD^H =& diag\{\lambda_1, \lambda_2, \dots, \lambda_n\} diag\{\overline{\lambda}_1, \overline{\lambda}_2, \dots, \overline{\lambda}_n\} \\ =& diag\{|\lambda_1|^2, |\lambda_2|^2, \dots, |\lambda_n|^2\} \\ =& diag\{\overline{\lambda}_1, \overline{\lambda}_2, \dots, \overline{\lambda}_n\} diag\{\lambda_1, \lambda_2, \dots, \lambda_n\} \\ =& D^H D \end{aligned}$$

por lo tanto, D es una matriz normal.

Teorema 4.11. Las matrices hermitianas, las anti-hermitianas y las unitarias son matrices normales.

Demostración

Supongamos que A es hermitiana, entonces

$$A^H A = AA = A^2 \qquad \text{y} \qquad AA^H = AA = A^2$$

luego $AA^H = A^H A$. Las demás quedan como ejercicio para el lector.

4.2. Factorizaciones

En esta sección se explica como se puede expresar una matriz A de componentes complejas como el producto de dos o más matrices.

Definición 4.5. Matrices complejas semejantes

Una matriz de componentes complejas A de tamaño $n \times n$ es semejante a una matriz de componentes complejas B de tamaño $n \times n$ si existe una matriz de componentes complejas no singular P de tamaño $n \times n$ tal que

$$B = P^{-1} AP. \tag{4.6}$$

De manera análoga se dice que A y B son semejantes si y sólo si existe una matriz de componentes complejas no singular P tal que

$$PB = AP. \tag{4.7}$$

Teorema 4.12. Las matrices complejas semejantes tienen el mismo polinomio característico y por tanto, los mismos valores propios.

Demostración

Como A y B son matrices complejas semejantes de tamaño $n \times n$, $B = P^{-1}AP$. Entonces

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}[AP - \lambda P] = P^{-1}[A - \lambda I]P.$$

Por consiguiente,

$$\begin{aligned} \det(B - \lambda I) &= \det[P^{-1}(A - \lambda I)P] = \det(P^{-1})\det(A - \lambda I)\det(P) \\ &= \det(P^{-1})\det(P)\det(A - \lambda I) = \det(A - \lambda I). \end{aligned}$$

Esto significa que A y B tienen la misma ecuación característica y como los valores propios son raíces de la ecuación característica, entonces tienen los mismos valores propios.

Definición 4.6. Matrices congruentes hermitianas

Dos matrices hermitianas A y B de tamaño $n \times n$ son congruentes hermitianas, si existe una matriz P no singular de componentes complejas de tamaño $n \times n$ tal que

$$A = P^H BP. \tag{4.8}$$

Teorema 4.13. Teorema de Schur

Sea A una matriz compleja de tamaño $n \times n$. Entonces A es semejante a una matriz triangular superior T, mediante una matriz unitaria U, es decir

$$T = U^H AU.$$

En este caso se dice que A es **triangularizable** por una matriz unitaria U.

Demostración

La demostración es por inducción sobre n. Si $n = 1$, A es una matriz de tamaño 1×1 que es triangular. La matriz unitaria es $U = [1]$.

Supongamos que toda matriz de componentes complejas de tamaño $(n-1) \times (n-1)$ es triangularizable por una matriz unitaria. Sea A una matriz de tamaño $n \times n$. Sabemos que su polinomio característico tiene al menos una raíz compleja λ_1. Sea $\vec{v}_1 \in \mathbb{C}^n$ un vector propio normalizado asociado al valor propio λ_1. Denotemos por W el complemento ortogonal a $\vec{v}_1$ de dimensión $n-1$. Sea $\{\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ una base ortonormal de W. Luego, cada vector $\vec{X}$ de W tiene la forma

$$\vec{X} = a_2\,\vec{v}_2 + a_3\,\vec{v}_3 + \ldots + a_n\,\vec{v}_n.$$

La matriz de cambio de base, de la *base canónica* de $\mathbb{C}^n$ a la base $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es la matriz S cuyas columnas son los elementos de los vectores $\vec{v}_i$. Luego,

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$S^{-1}AS = S^{-1}\begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}.$$

Pero como S es unitaria se tiene que $S^{-1} = S^H$, por consiguiente

$$S^HAS = \begin{bmatrix} \lambda_1 & \overline{z_1} \ \ldots \ \overline{z_{n-1}} \\ 0 & \\ \vdots & \begin{bmatrix} \quad A_1 \quad \end{bmatrix} \\ 0 & \end{bmatrix},$$

donde A_1 es una matriz de tamaño $(n-1) \times (n-1)$.

La prueba ahora se completa por inducción, si R_1 es una matriz unitaria de tamaño $(n-1) \times (n-1)$ tal que $(R_1)^H A_1 R_1 = T_1$, con T_1 triangular superior, por la hipótesis de inducción. Entonces, la matriz

$$R = \begin{bmatrix} 1 & 0 \ \ldots \ 0 \\ 0 & \\ \vdots & \begin{bmatrix} \quad R_1 \quad \end{bmatrix} \\ 0 & \end{bmatrix},$$

es una matriz unitaria y

$$
\begin{aligned}
(SR)^H A(SR) =& R^H \left(S^H AS\right) R \\
=& \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{x}^H \\ \vec{0} & A_1^H \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1 \end{bmatrix} \\
=& \begin{bmatrix} \lambda_1 & \vec{x}^H R_1 \\ \vec{0} & T_1 \end{bmatrix},
\end{aligned}
$$

donde $\vec{x}^H = (\overline{z_1}, \overline{z_2}, \dots, \overline{z_{n-1}})$. La matriz $SR = U$ es el producto de dos matrices unitarias; por lo tanto, es también una matriz unitaria. Así, $U^H AU$ es una matriz triangular superior y nuestra prueba queda completa.

El siguiente resultado es una consecuencia directa del teorema anterior.

Teorema 4.14. Sea A una matriz de componentes complejas de tamaño $n \times n$. Los valores propios de A son los elementos de la diagonal de la matriz triangular superior T semejante a A por una matriz unitaria.

Demostración

Como A y T son semejantes por el Teorema 4.12, tienen el mismo polinomio característico. Por otra parte, como T es triangular se tiene

$$p_A(\lambda) = p_T(\lambda) = (t_{11} - \lambda)(t_{22} - \lambda)\dots(t_{nn} - \lambda),$$

donde $t_{11}, t_{22}, \dots, t_{nn}$ son los elementos de la diagonal de T. Así pues los valores propios de A son los elementos de la diagonal de T.

Ejemplo 4.3. Dada la matriz de componentes complejas

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

encuentre una matriz T que sea la triangularización de A.

Solución

El polinomio característico de A es

$$p_A(\lambda) = (1-\lambda)(i-\lambda)((1-i)-\lambda).$$

Luego los vectores propios son $\lambda_1 = 1$, $\lambda_2 = i$ y $\lambda_3 = 1-i$. El vector propio correspondiente a $\lambda_1 = 1$ es $\vec{v}_1 = \begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix}$; para $\lambda_2 = i$ es $\vec{v}_2 = \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix}$ y por último, para $\lambda_3 = 1-i$ es $\vec{v}_3 = \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix}$.

Para determinar U, se aplica el proceso de Gram-Schmidt a $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, para encontrar una base ortonormal para $\mathbb{C}^3$. Como $|\vec{v}_1| = 5$, se hace $\vec{u}_1 = \frac{\vec{v}_1}{|\vec{v}_1|} = \frac{1}{5}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix}$. Así pues,

$$\begin{aligned}
\vec{v}_2^{\,\prime} &= \vec{v}_2 - \overline{(\vec{v}_2^H \vec{u}_1)}\vec{u}_1 = \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix} - \frac{(-3-i)}{5}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1+3i \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.
\end{aligned}$$

Entonces $|\vec{v}_2^{\,\prime}| = 2$ y $\vec{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Se puede verificar que $\vec{u}_1^H \vec{u}_2 = 0$. Ahora,

$$\begin{aligned}
\vec{v}_3^{\,\prime} &= \vec{v}_3 - \overline{(\vec{v}_3^H \vec{u}_1)}\vec{u}_1 - \overline{(\vec{v}_3^H \vec{u}_2)}\vec{u}_2 \\
&= \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix} - \frac{(-15+20i)}{25}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix} - 2i\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix} + \begin{bmatrix} 0 \\ 5 \\ -2i \end{bmatrix} = \begin{bmatrix} 1-2i \\ 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

Por último, $|\vec{v}_3| = \sqrt{5}$ luego $\vec{u}_3 = \begin{bmatrix} (1-2i)/\sqrt{5} \\ 0 \\ 0 \end{bmatrix}$. También se verifica que la base obtenida para $\mathbb{C}^3$ es ortonormal observando que $\vec{u}_1^H \vec{u}_3 = 0$ y $\vec{u}_2^H \vec{u}_3 = 0$. Por lo tanto, la matriz unitaria U es

$$U = \begin{bmatrix} 0 & 0 & (1-2i)/\sqrt{5} \\ (3+4i)/5 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

y la matriz triangular es

$$T = U^H AU = \begin{bmatrix} 0 & \dfrac{3-4i}{5} & 0 \\ 0 & 0 & 1 \\ \dfrac{1+2i}{\sqrt{5}} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1-i & 0 & 0 \\ i & 1 & 2+i \\ 2i & 0 & i \end{bmatrix} \begin{bmatrix} 0 & 0 & \dfrac{1-2i}{\sqrt{5}} \\ \dfrac{3+4i}{5} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2-i & \left(\frac{2}{5} - \frac{1}{5}i\right)\sqrt{5} \\ 0 & i & \left(\frac{4}{5} + \frac{2}{5}i\right)\sqrt{5} \\ 0 & 0 & 1-i \end{bmatrix}.$$

Nótese que los elementos de la diagonal principal de la matriz T son los valores propios de la matriz A.

Teorema 4.15. Sea A una matriz hermítica de tamaño $n \times n$, entonces existe una matriz unitaria U tal que

$$U^{-1}AU$$

es una matriz diagonal.

Demostración

Como A es una matriz compleja, por el teorema de Schur, A se puede triangularizar mediante una matriz unitaria U, es decir

$$T = U^H AU,$$

donde T es una matriz triangular superior. Al tomar la transpuesta conjugada y usando que $A^H = A$ se tiene que

$$T^H = (U^H AU)^H = U^H A^H (U^H)^H = U^H AU = T,$$

como T^H es una matriz triangular inferior, luego T es una matriz diagonal. En consecuencia A es semejante, mediante una matriz unitaria U, a una matriz diagonal T.

Ejemplo 4.4. Considere la matriz de componentes complejas

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

Comprobar que es diagonalizable mediante una matriz unitaria.

Solución

El polinomio característico de A es

$$p_A(\lambda) = \lambda(3-\lambda)(2+\lambda)$$

En este caso, los valores propios de A son $\lambda_1 = 0$, $\lambda_2 = 3$ y $\lambda_3 = -2$. Los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 2 \\ -i \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix}$, respectivamente.

Para encontrar la matriz unitaria U, se ortonormaliza el conjunto $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. Como $|\vec{v}_1| = \sqrt{6}$, se hace $\vec{u}_1 = \frac{\vec{v}_1}{|\vec{v}_1|} = \begin{bmatrix} 2/\sqrt{6} \\ -i/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$. Por otra parte, se tiene que $|\vec{v}_2| = \sqrt{3}$, entonces $\vec{u}_2 = \begin{bmatrix} 1/\sqrt{3} \\ i/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}$. Por último, $|\vec{v}_3| = \sqrt{2}$ de manera que $\vec{u}_3 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -i/\sqrt{2} \end{bmatrix}$. Se puede verificar que la base obtenida para $\mathbb{C}^3$ es ortonormal observando que $\vec{u}_1^H \vec{u}_2 = 0$, $\vec{u}_1^H \vec{u}_3 = 0$ y $\vec{u}_2^H \vec{u}_3 = 0$. Por lo tanto,

$$U = \begin{bmatrix} 2/\sqrt{6} & 1/\sqrt{3} & 0 \\ -i/\sqrt{6} & i/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{6} & -1/\sqrt{3} & -i/\sqrt{2} \end{bmatrix}.$$

Como el $\det(U) = 1$, se tiene que

$$D = U^H A U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Luego la matriz A es diagonalizable por una matriz unitaria.

Teorema 4.16. Descomposición Espectral para Hermitianas

Sea A una matriz hermitiana de tamaño $n \times n$ con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \lambda_1 \vec{v}_1 \vec{v}_1^H + \lambda_2 \vec{v}_2 \vec{v}_2^H + \ldots + \lambda_n \vec{v}_n \vec{v}_n^H, \tag{4.9}$$

donde $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ son los vectores propios normalizados de A.

Demostración

Por el Teorema 4.15, existe una matriz U tal que $U^{-1}AU = T$, donde T es una matriz diagonal. Entonces,

$$\begin{aligned}
A =&UTU^{-1} = UTU^{H}\\
=&\begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}\begin{bmatrix}\vec{u}_1^H\\ \vec{u}_2^H\\ \vdots\\ \vec{u}_n^H\end{bmatrix}\\
=&\begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1\vec{u}_1^H\\ \lambda_2\vec{u}_2^H\\ \vdots\\ \lambda_n\vec{u}_n^H\end{bmatrix}\\
=&\lambda_1\vec{v}_1\vec{v}_1^H + \lambda_2\vec{v}_2\vec{v}_2^H + \dots + \lambda_n\vec{v}_n\vec{v}_n^H.
\end{aligned}$$

Esto prueba el teorema.

Ejemplo 4.5. Ilustrar el Teorema de Descomposición Espectral para la matriz dada en el Ejemplo 4.4.

Solución

Del Ejemplo 4.4 se tiene que los valores propios asociados a la matriz A son $\lambda_1 = 0,\ \lambda_2 = 3$ y $\lambda_3 = -2$. Los respectivos vectores propios normalizados de A eran

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}2\\ -i\\ 1\end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\ i\\ -1\end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\ 1\\ -i\end{bmatrix}.$$

Entonces

$$\begin{aligned}
\sum_{i=1}^{3}\lambda_i\vec{v}_i\vec{v}_i^H =&\frac{3}{3}\begin{bmatrix}1\\ i\\ -1\end{bmatrix}\begin{bmatrix}1 & -i & -1\end{bmatrix} + \frac{-2}{2}\begin{bmatrix}0\\ 1\\ -i\end{bmatrix}\begin{bmatrix}0 & 1 & i\end{bmatrix}\\
=&\begin{bmatrix}1 & -i & -1\\ i & 1 & -i\\ -1 & i & 1\end{bmatrix} - \begin{bmatrix}0 & 0 & 0\\ 0 & 1 & i\\ 0 & -i & 1\end{bmatrix} = \begin{bmatrix}1 & -i & -1\\ i & 0 & -2i\\ -1 & 2i & 0\end{bmatrix},
\end{aligned}$$

la cual coincide con la matriz A dada en el Ejemplo 4.4.

Teorema 4.17. Sea A una matriz hermitiana y no singular, si A puede ser factorizada como

$$A = LDU,$$

en donde L es unitaria triangular inferior, D es diagonal y U es unitaria triangular superior. Entonces, $L = U^H$.

Demostración

Puesto que A puede factorizarse como $A = LDU$, tomando la transpuesta conjugada, se tiene que

$$A^H = (LDU)^t = U^H D^H L^H = U^H D L^H.$$

Como A es hermítica, es igual a A^H, por lo tanto

$$\begin{aligned} LDU =& U^H D L^H \\ D =& L^{-1} \left(U^H D L^H \right) U^{-1} \\ D =& (UL)^H D (UL)^H, \end{aligned}$$

luego, $L = U^{-1} = U^H$, lo cual completa la prueba.

Teorema 4.18. Si N es una matriz normal, entonces la matriz $T = U^H N U$ (U unitaria) es también normal.

Demostración

Sea N una matriz normal y definamos $T = U^H N U$ multiplicando por T^H, se obtiene que

$$\begin{aligned} TT^H =& (U^H N U)(U^H N U)^H \\ =& U^H N (U U^H) N^H U = U^H N N^H U \\ =& U^H N^H N U && \textit{puesto queN es normal} \\ =& (U^H N^H U)(U^H N U) = T^H T, \end{aligned}$$

como $TT^H = T^H T$ se ha demostrado que T es normal.

Teorema 4.19. Sea A una matriz de componentes complejas de tamaño $n \times n$. La matriz A es normal si y sólo si es diagonalizable mediante una matriz unitaria.

Demostración

Supongamos que A es normal. Por el Teorema de Schur, A es semejante a una matriz triangular superior T, mediante una matriz unitaria U. Es decir $T = U^H AU$. Pero como

$$\begin{aligned} TT^H =&(U^H AU)(U^H AU)^H \\ =&U^H AA^H U = U^H A^H AU \\ =&(U^H A^H U)(U^H AU) \\ =&T^H T, \end{aligned}$$

la matriz T resulta ser normal. El lector puede verificar que la matriz T es diagonal.

Recíprocamente, supóngase que A es diagonalizable por una matriz unitaria U, es decir $U^H AU = D$, donde D es una matriz diagonal. Como D es normal (Ejemplo 4.2), se verifica que

$$\begin{aligned} AA^H =&(UDU^H)(UDU^H)^H \\ =&UDD^H U^H = UD^H DU^H \\ =&(UDU^H)(UDU^H) \\ =&A^H A. \end{aligned}$$

Por lo tanto, la matriz A es normal.

4.2.1. Forma canónica de Jordan

Si A es una matriz real de tamaño $n \times n$, en la sección 3.5 se vio que se podía encontrar una matriz no singular P de tamaño $n \times n$, tal que

$$\mathcal{J} = P^{-1}AP.$$

En esta sección se explica para matrices de componentes complejas la forma canónica de Jordan.

Teorema 4.20. Sea A una matriz compleja de tamaño $n \times n$. Entonces existe una matriz P no singular tal que

$$P^{-1}AP = \begin{bmatrix} \mathcal{J}_{n_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathcal{J}_{n_k}(\lambda_k) \end{bmatrix} = \mathcal{J}, \tag{4.10}$$

en donde cada $\mathcal{J}_{n_i}(\lambda_i)$ es un bloque de Jordan de tamaño $n_i \times n_i$ y $n_1+n_2+\ldots+n_k = n$. Los valores propios λ_i, $i = 1,2,\ldots,k$ no son necesariamente distintos. El número total de bloques quedan determinados unívocamente por la matriz A.

Demostración

Queda como ejercicio para el lector.

Ejemplo 4.6.

Encuentre una matriz no singular P tal que $P^{-1}AP$ sea una matriz de Jordan, para la siguiente matriz

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

Solución

La ecuación característica de A es

$$-\lambda^3 + (3+3i)\lambda^2 - 6i\lambda - (2-2i) = 0.$$

Luego $\lambda = 1+i$ es el único valor propio (de multiplicidad algebraica tres). Entonces

$$(A-\lambda I)\vec{v} = [A-(1+i)I]\,\vec{v} = \begin{bmatrix} 1 & 1 & 0 \\ -2 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Esto conduce a $x+y=0$ y $-x+z=0$. Tomando $x \in \mathbb{R}$, se obtiene el vector propio: $\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_2$ se calcula

$$[A-(1+i)I]\,\vec{v}_2 = \vec{v}_1$$

$$\begin{bmatrix} 1 & 1 & 0 \\ -2 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

Si se realizan operaciones por filas se obtiene el vector propio generalizado: $\vec{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$ y de manera análoga el vector propio generalizado: $\vec{v}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$. Por consiguiente,

$$P = \begin{bmatrix}1 & 0 & 0\\-1 & 1 & 0\\1 & 0 & 1\end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix}1 & 0 & 0\\1 & 1 & 0\\-1 & 0 & 1\end{bmatrix}.$$

Al efectuar el producto $P^{-1}AP$ se llega a

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

El lector puede notar que sobre la diagonal de la matriz de Jordan se encuentra el valor propio de la matriz A.

4.2.2. Descomposición en valores singulares

Si A es una matriz real de tamaño $m \times n$, hemos visto en la Sección 3.6 que se pueden encontrar dos matrices ortogonales U y V de tamaños $m \times m$ y $n \times n$, respectivamente, tales que

$$A = USV^t.$$

En esta sección se describe la descomposición en valores singulares y la descomposición polar para matrices de componentes complejas.

Teorema 4.21. Descomposición en Valores Singulares

Sea A una matriz compleja de tamaño $m \times n$ con rango r. Entonces existen matrices unitarias U y V de tamaño $m \times m$ y $n \times n$, respectivamente, tales que

$$A = USV^H, \tag{4.11}$$

donde S es la matriz particionada de tamaño $m \times n$, dada por

$$S = \begin{bmatrix}D_r & \vdots & O\\ \cdots & . & \cdots\\ O & \vdots & O\end{bmatrix}, \qquad \text{donde} \qquad D_r = \begin{bmatrix}\sigma_1 & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \sigma_r\end{bmatrix}, \tag{4.12}$$

siendo σ_i, para $i = 1, 2, \ldots, r$, los valores singulares no nulos de A.

Demostración

Queda como ejercicio para el lector.

4.2.3. Descomposición polar

Ahora se estudiara la *descomposición polar* para matrices de componentes complejas. El nombre de descomposición polar se debe a la representación polar de un número complejo $z = \rho e^{i\theta}$. La analogía entre esta representación de los números complejos y la descomposición (3.34) de una matriz es debida a que los valores propios de la matriz P son números reales no negativos y los de la matriz Q son números complejos unitarios.

Teorema 4.22. Descomposición Polar

Sea A una matriz compleja de tamaño $n \times n$ con rango r. Entonces, existe una matriz hermitiana P de tamaño $n \times n$ con valores propios no negativos y una matriz unitaria Q de tamaño $n \times n$, tales que

$$A = PQ. \tag{4.13}$$

Demostración

Si A es una matriz de tamaño $n \times n$, en una descomposición de valores singulares las matrices U, S y V son también de tamaño $n \times n$. En este caso se puede escribir la ecuación (4.11) como

$$A = USV^H = US(U^H U)V^H = (USU^H)UV^H = PQ,$$

donde $P = USU^H$ es una matriz hermítica. Como

$$\begin{aligned} P^H =&(USU^H)^H = US^H U^H \\ =&USU^H = P && \text{(por ser } S \text{ simétrica)} \end{aligned}$$

y la matriz $Q = UV^H$ es unitaria, puesto que

$$Q^{-1} = (UV^H)^{-1} = (V^H)^{-1}U^{-1} = VU^H = Q^t.$$

En la última ecuación se usaron los hechos de que U y V eran matrices unitarias.

Se deja como ejercicio la comprobación de que P tiene valores propios no negativos.

Ejemplo 4.7. Encuentre para la siguiente matriz de componentes complejas

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

su descomposición en valores singulares y su descomposición polar.

Solución

La matriz $A^H A$ es

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

En este caso, los valores propios de $A^H A$ son $\lambda_1 = 4$ y $\lambda_2 = 1$. Por lo tanto, los valores singulares asociados a la matriz A son $\sigma_1^2 = 4$ y $\sigma_2^2 = 1$. Al calcular los respectivos vectores propios normalizados se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1+i \\ -2 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ 1-i \end{bmatrix}.$$

Por otra parte, la matriz AA^H es

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

y sus respectivos vectores propios normalizados son

$$\vec{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1+i \\ 2 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ 1+i \end{bmatrix}.$$

Finalmente, si $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}$, $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ y $S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, entonces la SVD de A es

$$A = \left(\frac{1}{\sqrt{6}} \right)^2 \begin{bmatrix} 1+i & 2 \\ -2 & 1-i \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1+i \\ 1+i & 2 \end{bmatrix} = \begin{bmatrix} 1+i & -i \\ 0 & 1-i \end{bmatrix}.$$

El lector puede obtener la descomposición polar de A.

Las propiedades de las matrices complejas descritas en este capítulo son comparables a las propiedades de las matrices reales analizadas anteriormente. En el siguiente resumen se indica la correspondencia entre las matrices complejas unitarias y hermitianas con las matrices reales ortogonales y simétricas.

<table>
<tr><th colspan="2">Comparación entre Matrices Reales y Matrices Complejas</th></tr>
<tr>
<td>Sea $A=[a_{ij}]$, con $a_{ij}\in\mathbb{R}$.
1. Toda matriz simétrica tiene valores propios reales
2. Si A es una matriz simétrica, los vectores propios correspondientes a valores propios distintos son ortogonales.
3. Descomposición de Schur
Si A es una matriz de tamaño $n\times n$ con valores propios reales, existe una matriz ortogonal Q tal que
$$Q^tAQ$$
es una matriz triangular superior.
4. Teorema Espectral
Si A es una matriz simétrica, existe una matriz ortogonal Q tal que
$$Q^tAQ$$
es diagonal.
5. Descomposición en valores singulares
Existen matrices ortogonales U y V de tamaños $m\times m$ y $n\times n$, respectivamente, tales que
$$A=USV^t$$
donde S esta dada por (3.31).
6. Descomposición Polar
Existe una matriz simétrica P de tamaño $n\times n$ con valores propios no negativos y una matriz ortogonal Q de tamaño $n\times n$, tal que
$$A=PQ$$
donde $P=USU^t$ y $Q=UV^t$.</td>
<td>Sea $A=[a_{ij}]$, con $a_{ij}\in\mathbb{C}$.
1. Toda matriz hermitiana tiene valores propios reales
2. Si A es una matriz hermitiana, los vectores propios correspondientes a valores propios distintos son ortogonales.
3. Descomposición de Schur
Si A es una matriz de tamaño $n\times n$, existe una matriz unitaria U tal que
$$U^HAU$$
es una matriz triangular superior.
4. Teorema Espectral
Si A es una matriz hermitiana, existe una matriz unitaria U tal que
$$U^HAU$$
es diagonal.
5. Descomposición en valores singulares
Existen matrices unitarias U y V de tamaños $m\times m$ y $n\times n$, respectivamente, tales que
$$A=USV^H$$
donde S esta dada por (4.12).
6. Descomposición Polar
Existe una matriz hermitiana P de tamaño $n\times n$ con valores propios no negativos y una matriz unitaria Q de tamaño $n\times n$, tal que
$$A=PQ$$
donde $P=USU^H$ y $Q=UV^H$.</td>
</tr>
</table>

Ejercicios 4.1.

1. Determine para cada una de las siguientes matrices una matriz unitaria U tal que $U^H AU$ sea diagonal

   $a.\ \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}.$ $\quad b.\ \begin{bmatrix} 1 & 2+3i \\ 2-3i & -1 \end{bmatrix}.$ $\quad c.\ \begin{bmatrix} 2 & 1+2i \\ 1-2i & -2 \end{bmatrix}.$

   $d.\ \begin{bmatrix} 1 & i & 2+i \\ -i & 2 & 1-i \\ 2-i & 1+i & 2 \end{bmatrix}.$ $\quad e.\ \begin{bmatrix} 1 & 1+i & 2i \\ 1-i & 4 & 2-3i \\ -2i & 2+3i & 7 \end{bmatrix}.$

2. Sea $\mathbb{V} = \left\{A \in M_{22} : a_{ij} \in \mathbb{C}\right\}$; determine si cada uno de los siguientes subconjuntos $\mathbb{H}$ son o no subespacios de $\mathbb{V}$.

   a) $\mathbb{H} = \{A \in \mathbb{V} : a_{ii} = 0\}$.

   b) $\mathbb{H} = \{A \in \mathbb{V} : a_{ii}$ son imaginarios puros$\}$.

   c) $\mathbb{H} = \{A \in \mathbb{V} : A = A^t\}$.

3. Sea $\mathbb{H} = \left\{ A \in M_{22} : A = \begin{bmatrix} w & -\bar{z} \\ \bar{z} & w \end{bmatrix}; w, z \in \mathbb{C} \right\}.$

   a) Demuestre que $\mathbb{H}$ es cerrado para la suma y la multiplicación.

   b) ¿Cuáles matrices en $\mathbb{H}$ son no singulares?

   c) Si una matriz en $\mathbb{H}$ es no singular, entonces la inversa está en $\mathbb{H}$.

   d) Encuentre dos matrices A y B en $\mathbb{H}$ tal que $AB \neq BA$.

4. Sea A una matriz de tamaño $n \times n$ con componentes complejas y sea $\vec{x} \in \mathbb{C}^n$ un vector propio correspondiente al valor propio $\lambda \in \mathbb{C}$. Muestre que para cada escalar complejo no nulo α, el vector $\alpha\vec{x}$ es un vector propio de A.

5. Si A es una matriz normal, pruebe que A y A^H son diagonalizables por la misma matriz unitaria.

6. Si A es una matriz normal, pruebe que λ es un valor propio de A si y sólo si $\overline{\lambda}$ es un valor propio de A^H.

Capítulo 5

Formas bilineales

En este capítulo estudiaremos las formas bilineales sobre espacios de dimensión finita. Se introduce la representación matricial de una forma bilineal y se establece el isomorfismo entre el espacio de las formas y el espacio de las matrices de tamaño $n \times n$.

5.1. Formas bilineales

Definición 5.1. Sean $\mathbb{U}$, $\mathbb{V}$, $\mathbb{W}$ espacios vectoriales reales. Una aplicación $g : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$ se llama **bilineal** si satisface las siguientes propiedades:

BI 1. Para todo $\vec{u}_1, \vec{u}_2 \in \mathbb{U}$ y $\vec{v} \in \mathbb{V}$ se tiene que

$$g(\vec{u}_1 + \vec{u}_2, \vec{v}) = g(\vec{u}_1, \vec{v}) + g(\vec{u}_2, \vec{v})$$

y para todo $\vec{u} \in \mathbb{U}$ y $\vec{v}_1, \vec{v}_2 \in \mathbb{V}$ se tiene que

$$g(\vec{u}, \vec{v}_1 + \vec{v}_2) = g(\vec{u}, \vec{v}_1) + g(\vec{u}, \vec{v}_2).$$

BI 2. Para todo $\alpha \in \mathbb{R}, \vec{u} \in \mathbb{U}$ y $\vec{v} \in \mathbb{V}$ se tiene que

$$g(\alpha\vec{u}, \vec{v}) = \alpha g(\vec{u}, \vec{v}) = g(\vec{u}, \alpha\vec{v}).$$

Ejemplo 5.1. Sea $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ definida por

$$g(\vec{X},\vec{Y}) = \vec{X}^t A\vec{Y},$$

donde $\vec{X} \in \mathbb{R}^m$, $\vec{Y} \in \mathbb{R}^n$ y A es una matriz real de tamaño $m \times n$. Verifique si la aplicación g es bilineal.

Solución

Para todo $\vec{X}_1,\ \vec{X}_2 \in \mathbb{R}^m$ *y* $\vec{Y} \in \mathbb{R}^n$ *se tiene que*

$$\begin{aligned} g(\vec{X}_1+\vec{X}_2,\vec{Y}) =&(\vec{X}_1+\vec{X}_2)^t A\vec{Y} = (\vec{X}_1^t+\vec{X}_2^t)A\vec{Y}\\ =&\vec{X}_1^t A\vec{Y}+\vec{X}_2^t A\vec{Y} = g(\vec{X}_1,\vec{Y})+g(\vec{X}_2,\vec{Y}). \end{aligned}$$

Para todo $\alpha \in \mathbb{R}$, $\vec{X} \in \mathbb{R}^m$ *y* $\vec{Y} \in \mathbb{R}^n$ *se tiene que*

$$\begin{aligned} g(\alpha\vec{X},\vec{Y}) =&(\alpha\vec{X})^t A\vec{Y} = (\alpha^t\vec{X}^t)A\vec{Y}\\ =&\alpha\vec{X}^t A\vec{Y} = \alpha g(\vec{X},\vec{Y}). \end{aligned}$$

Así, la aplicación g es lineal cuando $\vec{Y} \in \mathbb{R}^n$, permanece fijo. De manera análoga, se puede demostrar que g es una transformación lineal cuando la componente $\vec{X} \in \mathbb{R}^m$, se mantiene fija. Por lo tanto, g es una aplicación bilineal.

Teorema 5.1. Sea $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ una aplicación bilineal. Entonces existe una matriz única de tamaño $m \times n$, A tal que

$$g(\vec{X},\vec{Y}) = g_A(\vec{X},\vec{Y}) = \vec{X}^t A\vec{Y}. \tag{5.1}$$

El conjunto de aplicaciones bilineales de $\mathbb{R}^m \times \mathbb{R}^n$ en $\mathbb{R}$ es un espacio vectorial denotado por $Bil(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ y la asociación

$$A \mapsto g_A$$

es un isomorfismo entre $Bil(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ y el espacio de las matrices reales de tamaño $m \times n$.

Demostración

Sean $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m$ los vectores unitarios estándar para $\mathbb{R}^m$ y sean $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ los vectores unitarios estándar para $\mathbb{R}^n$. Luego se puede expresar cualquier $\vec{X} \in \mathbb{R}^m$ y cualquier $\vec{Y} \in \mathbb{R}^n$ de la siguiente manera:

$$\vec{X} = \sum_{i=1}^{m} x_i \vec{e}_i \qquad \text{y} \qquad \vec{Y} = \sum_{j=1}^{n} y_j \vec{u}_j.$$

Entonces, se tiene que

$$g(\vec{X}, \vec{Y}) = g\left(\sum_{i=1}^{m} x_i \vec{e}_i, \sum_{j=1}^{n} y_j \vec{u}_j \right).$$

Como g es una transformación lineal en la primer componente se llega a

$$g(\vec{X}, \vec{Y}) = \sum_{i=1}^{m} x_i \left[g\left(\vec{e}_i, \sum_{j=1}^{n} y_j \vec{u}_j \right) \right]$$

usando el hecho de que g es lineal en la segunda componente se halla que

$$g(\vec{X}, \vec{Y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j \; g(\vec{e}_i, \vec{u}_j).$$

Sea

$$a_{ij} = g(\vec{e}_i, \vec{u}_j)$$

así,

$$g(\vec{X}, \vec{Y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j,$$

que es precisamente la expresión del producto

$$\vec{X}^t A \vec{Y} \qquad \text{con} \qquad A = [a_{ij}].$$

Esto prueba que $g = g_A$ para las a_{ij} escogidas anteriormente.

Ahora se puede demostrar que A es única. Suponga que $g(\vec{X}, \vec{Y}) = \vec{X}^t A \vec{Y}$ y que $g(\vec{X}, \vec{Y}) = \vec{X}^t B \vec{Y}$ para todo $\vec{X} \in \mathbb{R}^m$ y $\vec{Y} \in \mathbb{R}^n$. Entonces

$$\vec{X}^t A \vec{Y} = \vec{X}^t B \vec{Y}$$

o estableciendo $C = A - B$ se tiene que

$$\vec{X}^t C \vec{Y} = 0, \qquad \text{para todo } \vec{X} \in \mathbb{R}^m \text{ y } \vec{Y} \in \mathbb{R}^n.$$

En particular, si $\vec{X} = \vec{e}_i$ e $\vec{Y} = \vec{u}_j$, se tiene que

$$0 = \vec{e}_i^{\,t} C \vec{u}_j = c_{ij}.$$

Así que $c_{ij} = 0$ para todo $i, \ j$, y por lo tanto, $C = \mathbf{0}$, la matriz cero de tamaño $m \times n$. Esto muestra que $A = B$.

Queda como ejercicio para el lector demostrar la parte referente al isomorfismo entre el espacio de las matrices y las aplicaciones bilineales.

Definición 5.2. La matriz A en el Teorema 5.1 se llama **representación matricial** de la aplicación bilineal g_A.

Definición 5.3. Forma Bilineal

Si en la definición 5.1, los espacios $\mathbb{U}$ y $\mathbb{V}$ son iguales al mismo espacio $\mathbb{V}$ y el espacio $\mathbb{W} = \mathbb{R}$, de tal manera que g aplique a $\mathbb{V} \times \mathbb{V}$ en $\mathbb{R}$, entonces se dice que g es una *forma bilineal* sobre $\mathbb{V}$.

Teorema 5.2. Sean $g_1, \ g_2 : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ dos formas bilineales distintas sobre $\mathbb{V}$. Entonces

a) $g_1 + g_2$, es una forma bilineal.

b) αg_1 es también una forma bilineal; donde $\alpha \in \mathbb{R}$

Demostración

Queda como ejercicio para el lector.

Definición 5.4. Rango de una Forma Bilineal

El rango de una forma bilineal g sobre $\mathbb{V}$, escrito $\rho(g)$, se define como el rango de la matriz que representa a g. Se dice que g es degenerada o no degenerada según si $\rho(g) < \dim(V)$ o $\rho(g) = \dim(V)$.

Definición 5.5. Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma bilineal sobre $\mathbb{V}$; entonces g es **simétrica**, si para todo $\vec{v}, \ \vec{w} \in V$, se cumple que

$$g(\vec{v}, \vec{w}) = g(\vec{w}, \vec{v}). \tag{5.2}$$

Teorema 5.3. Una matriz real A de tamaño $n \times n$ representa una forma bilineal simétrica si y sólo si es una matriz simétrica.

Demostración

Supóngase que A es simétrica. Como para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$, la matriz $\vec{X}^t A\vec{Y}$ es una matriz de 1×1, es decir, un elemento de $\mathbb{R}$, entonces es igual a su propia traspuesta. Por lo tanto,

$$\vec{X}^t A\vec{Y} = (\vec{X}^t A\vec{Y})^t = \vec{Y}^t A^t (\vec{X}^t)^t = \vec{Y}^t A\vec{X}.$$

Así que A representa una forma bilineal simétrica.

Recíprocamente, supóngase que A representa una forma bilineal simétrica; es decir,

$$g_A(\vec{X},\vec{Y}) = g_A(\vec{Y},\vec{X}). \tag{5.3}$$

Para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$. Como

$$g_A(\vec{Y},\vec{X}) = \vec{Y}^t A\vec{X} = (\vec{Y}^t A\vec{X})^t = \vec{X}^t A^t (\vec{Y}^t)^t = \vec{X}^t A^t \vec{Y}. \tag{5.4}$$

Si se comparan las expresiones (5.3) y (5.4), se tiene que

$$g_A(\vec{X},\vec{Y}) = \vec{X}^t A\vec{Y} = \vec{X}^t A^t \vec{Y}, \tag{5.5}$$

como (5.5) se cumple para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$, se concluye que $A = A^t$. Es decir, A es simétrica.

Definición 5.6. Inercia

Sea A una matriz simétrica de tamaño $n \times n$. La inercia de A es la terna ordenada de números

$$In(A) = (pos, neg, nul) \tag{5.6}$$

donde $pos,\ neg$ y nul son los números de valores propios de A positivos, negativos y nulos, respectivamente (contando todas las multiplicidades algebraicas).

Nótese que $\rho(A) = pos + neg$.

Definición 5.7. Signatura

La diferencia entre el número de valores propios positivos y el número de valores propios negativos se le denomina signatura de la matriz A. En otras palabras, si $In(A) = (i, j, k)$ se llama signatura de la matriz A a la cantidad

$$sig(A) = i - j.$$

Ejemplo 5.2.

Determinar la inercia y signatura de la matriz dada en el Ejemplo 2.15.

Solución

De los resultados del Ejemplo 2.15 se tiene que los valores propios de la matriz A eran $\lambda_1 = 1$ de multiplicidad algebraica 2 y $\lambda_2 = 7$. Luego,

$$In(A) = (3,0,0) \qquad \text{y} \qquad Sig(A) = 3.$$

Ejercicios 5.1.

1. Determine cuáles de las siguientes funciones $F : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ son aplicaciones bilineales

   $a.\ F(\vec{X},\vec{Y}) = 4x_1^2 + 4x_2y_1 + y_1^2.$ $\quad b.\ F(\vec{X},\vec{Y}) = 9x_1^2 - 24x_1y_2 + 16y_2^2.$

   $c.\ F(\vec{X},\vec{Y}) = x_1^2 + 8x_1y_2 + 16y_2^2.$ $\quad d.\ F(\vec{X},\vec{Y}) = x_1^2 + 2x_1y_1 + y_1^2.$

   $e.\ F(\vec{X},\vec{Y}) = x_1y_2 - x_2y_1.$ $\quad f.\ F(\vec{X},\vec{Y}) = x_1^2 + 2x_2y_1 + y_2^2.$

2. Escriba cada una de las siguientes formas bilineales $F : \mathbb{R}^3 \to \mathbb{R}$ en forma matricial

   $a.\ F(x,y,z) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz.$

   $b.\ F(x,y,z) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz.$

   $c.\ F(x,y,z) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz.$

3. Sea $\mathbb{V} = M_{23}$. Para $A,\ B \in \mathbb{V}$, defina $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ como

$$g(A,B) = tr(A^t B)$$

demuestre que es una aplicación bilineal en $\mathbb{V} \times \mathbb{V}$.

4. Sea $\mathbb{V} = M_{nn}$, demuestre que la función $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ dada por

$$g(A,B) = ntr(AB) - tr(A)tr(B)$$

es una aplicación bilineal en $\mathbb{V} \times \mathbb{V}$.

5. Sea A una matriz simétrica de tamaño $n \times n$ asociada a la aplicación bilineal $g\left(\vec{X},\vec{Y}\right) = \vec{X}^H A\vec{Y}$ con $\vec{X},\vec{Y} \in \mathbb{C}$. Demuestre que $g\left(\vec{X},\vec{Y}\right)$ es real.

6. Si $\mu_x = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ y $\mu_y = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$, represéntese la forma bilineal

$$COV\left(\vec{X},\vec{Y}\right) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

en la forma $\vec{X}^t A\vec{Y}$, con A simétrica. ¿Cuál es el rango de A?.

5.2. Formas cuadráticas

Cuando se considera el cuadrado de la norma de un vector $\vec{X} \in \mathbb{R}^n$, se obtiene la expresión $\|\vec{X}\|^2 = \vec{X}^t\vec{X}$; tales sumas y expresiones en forma general se denominan *Formas Cuadráticas*. Ellas surgen frecuentemente en una gran variedad de aplicaciones. Por ejemplo, se pueden usar formas cuadráticas en Ingeniería (para optimización), en Economía (en el análisis de funciones de costo y utilidad), en Física (para el estudio de energías cinéticas y potenciales) y en Estadística (en el análisis de varianza). En esta sección se estudiaran algunos temas relacionados con estas formas, utilizando la teoría de las matrices simétricas analizada anteriormente.

Definición 5.8. Sea $\mathbb{V}$ un espacio vectorial real de dimensión finita. Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma bilineal simétrica sobre $\mathbb{V}$. Entonces una **forma cuadrática** determinada por g es una función $F : \mathbb{V} \to \mathbb{R}$, tal que

$$F(\vec{v}) = g_A(\vec{v}, \vec{v}) = \vec{v}^t A \vec{v}. \tag{5.7}$$

La matriz A es llamada la **matriz de la forma cuadrática**

Ejemplo 5.3. Sea $\mathbb{V} = \mathbb{R}^n$, y $F(\vec{v}) = v_1^2 + v_2^2 + \ldots + v_n^2$. Escriba esta forma cuadrática como $\vec{v}^t A \vec{v}$.

Solución

Vamos a determinar la matriz $A = (a_{ij})$ de la forma bilineal simétrica g, de tal forma que

$$F(\vec{v}) = \vec{v}^t A \vec{v} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} v_i v_j. \tag{5.8}$$

Es decir, queremos encontrar los valores de a_{ij}, de manera que

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} v_i v_j = \sum_{i=1}^{n} v_i^2.$$

Como la matriz A es simétrica, $a_{ij} = a_{ji}$, por lo tanto, la forma cuadrática dada en (5.8) se puede expresar como

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} v_i v_j = \sum_{i=1}^{n} a_{ii} v_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} v_i v_j. \tag{5.9}$$

Si se comparan términos se establecen las siguientes relaciones

$$\sum_{i=1}^{n} a_{ii} v_i^2 = \sum_{i=1}^{n} v_i^2 \qquad \text{y} \qquad 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} v_i v_j = 0.$$

Pero como en la función $F(\vec{v})$ no aparecen términos de la forma $v_i v_j$, entonces

$$a_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Luego, $A = I_n$; y por lo tanto, $F(\vec{v})$ se puede expresar como $\vec{v}^t I_n \vec{v}$.

Ejemplo 5.4. Sea $\mathbb{V}=\mathbb{R}^3$, y $F(\vec{X})=2x_1^2+5x_2^2+2x_3^2+4x_1x_2+2x_1x_3+4x_2x_3$. Exprese esta forma cuadrática como $\vec{X}^t A\vec{X}$.

Solución

Utilizando el resultado obtenido en (5.9) para $n=3$, se tiene que

$$\vec{X}^t A\vec{X}=\sum_{i=1}^{3}a_{ii}x_i^2+2\sum_{i=1}^{2}\sum_{j=i+1}^{3}a_{ij}x_ix_j.$$

Si se desarrolla esta suma y se comparan los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

$$A=\begin{bmatrix}2&2&1\\2&5&2\\1&2&2\end{bmatrix},$$

la cual permite expresar a $F(\vec{X})$ de la forma $\vec{X}^t A\vec{X}$.

Ejemplo 5.5. Si $\mu_x=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$, represéntese la forma cuadrática

$$(n-1)s_x^2=\sum_{i=1}^{n}(x_i-\mu_x)^2$$

en la forma $\vec{X}^t A\vec{X}$, con A simétrica, ¿cuál es el rango de A?.

Solución

Sea $\vec{X}^t=(x_1,x_2,\ldots,x_n)^t$ el vector que representa las $n-$observaciones. Si en el Ejemplo 5.3 se reemplaza cada v_i por $x_i-\mu_x$, se tiene que

$$F(\vec{v})=\sum_{i=1}^{n}(x_i-\mu_x)^2=\vec{v}^t I_n\vec{v} \tag{5.10}$$

pero $\vec{v}$ se puede reescribir como

$$\vec{v}=\begin{bmatrix}x_1-\mu_x\\x_2-\mu_x\\\vdots\\x_j-\mu_x\\\vdots\\x_n-\mu_x\end{bmatrix}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_j\\\vdots\\x_n\end{bmatrix}-\mu_x\begin{bmatrix}1\\1\\\vdots\\1\\\vdots\\1\end{bmatrix}=\vec{X}-\mathbf{1}\mu_x, \tag{5.11}$$

donde $\mathbf{1}$ es un vector columna de unos de tamaño $n \times 1$. Luego μ_x se puede expresar como sigue

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}^t\vec{X}. \tag{5.12}$$

Si se reemplaza (5.12) en (5.11) se obtiene

$$\vec{v} = \vec{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^t\vec{X} = \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}. \tag{5.13}$$

Al sustituir en (5.10) se tiene que

$$\begin{aligned}
F\left[\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}\right] =& \left[\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}\right]^t I_n \left[\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}\right]\\
=&\vec{X}^t\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}\\
=&\vec{X}^t\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{1}{n^2}\mathbf{1}\underbrace{\mathbf{1}^t\mathbf{1}}\mathbf{1}^t\right)\vec{X}\\
=&\vec{X}^t\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}.
\end{aligned}$$

Aquí se uso que $\mathbf{1}^t\mathbf{1} = n$, denotando $J_n = \mathbf{1}\mathbf{1}^t$, se llega finalmente a que

$$(n-1)\,s_x^2 = \vec{X}^t\left(I_n - \frac{1}{n}J_n\right)\vec{X} = \vec{X}^t\left(I_n - \overline{J}_n\right)\vec{X}. \tag{5.14}$$

Luego, la matriz asociada a la forma es

$$A = I_n - \overline{J}_n \qquad \text{y} \qquad \rho(A) = n-1.$$

Cabe notar, que en estas notas la matriz J_n siempre será la matriz con todos sus elementos iguales a uno de tamaño $n \times n$, definida anteriormente.

Definición 5.9.

Dos formas cuadráticas $\vec{X}^tA\vec{X}$ y $\vec{Y}^tB\vec{Y}$ se dice que son equivalentes si existe una matriz no singular P tal que $B = P^tAP$. Aún más, las formas son equivalentes ortogonalmente si P se puede escoger ortogonal, equivalente-real si P se puede escoger con elementos reales y equivalente-compleja, o simplemente equivalente, si P tiene elementos complejos.

Teorema 5.4. Dos formas cuadráticas $\vec{X}^t A\vec{X}$ y $\vec{Y}^t B\vec{Y}$ son equivalentes si y sólo si las matrices simétricas A y B son congruentes.

Demostración

Si en la forma cuadrática $\vec{X}^t A\vec{X}$ se hace el cambio de variable $\vec{X} = P\vec{Y}$, donde P es una matriz no singular, se obtiene la forma

$$\vec{X}^t A\vec{X} = \vec{Y}^t P^t AP\vec{Y} = \vec{Y}^t B\vec{Y}.$$

Recíprocamente, sea $[a_{ij}] = A$ y $[b_{ij}] = P^t AP$, donde P es una matriz real no singular, las dos matrices simétricas asociadas con las formas cuadráticas

$$\sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij}x_i x_j \qquad \text{y} \qquad \sum_{i=1}^{n} b_{ii}y_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} b_{ij}y_i y_j.$$

El cambio de variable $\vec{X} = P\vec{Y}$, cambia la primera forma cuadrática a la segunda.

Ejemplo 5.6. Muestre que las formas cuadráticas

$$F_1(\vec{X}) = x_1^2 + x_2^2 + 8x_1x_2 \qquad \text{y} \qquad F_2(\vec{Y}) = y_1^2 - 14y_2^2 + 2y_1y_2$$

son equivalentes.

Solución

Utilizando el resultado obtenido en (5.9), para $n = 2$, se tiene que

$$\vec{X}^t A\vec{X} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2.$$

Si se comparan los a_{ij} con los coeficientes de la función $F_1(\vec{X})$, se obtiene la matriz

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix},$$

la cual permite expresar a $F_1(\vec{X})$ de la forma $\vec{X}^t A\vec{X}$. Para la forma cuadrática $F_2(\vec{Y})$, se tiene la matriz

$$B = \begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix}.$$

En el Ejemplo 2.5 se mostró que A y B eran congruentes. Por lo tanto, $F_1(\vec{X})$ es equivalente a la forma cuadrática $F_2(\vec{Y})$.

Teorema 5.5. Dos formas cuadráticas $\vec{X}^t A\vec{X}$ y $\vec{Y}^t B\vec{Y}$ son equivalentes ortogonalmente si y sólo si A y B tienen los mismos valores propios y estos ocurren con la misma multiplicidad.

Demostración

Si A y B tiene valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ y D es una matriz diagonal con $\lambda_1, \lambda_2, \ldots, \lambda_n$ como elementos de su diagonal, entonces existen matrices ortogonales P y Q tal que

$$Q^t AQ = P^t BP = D.$$

Por consiguiente, $B = (P^t)^{-1}[Q^t AQ]P^{-1} = (QP^{-1})^t A(QP^{-1})$ y como QP^{-1} es ortogonal, $\vec{Y}^t B\vec{Y}$ es ortogonalmente equivalente a $\vec{X}^t A\vec{X}$.

Recíprocamente, si las dos formas son ortogonalmente equivalentes, B es similar a A (por que $P^{-1} = P^t$) y A y B tienen los mismos valores propios con las mismas multiplicidades.

Ejercicios 5.2.

1. Exprese las siguientes formas cuadráticas como $\vec{X}^t A\vec{X}$

$$a.\ F(x,y,z) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz.$$

$$b.\ F(x,y,z) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz.$$

$$c.\ F(x,y,z) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz.$$

2. Determine si las siguientes formas cuadráticas son equivalentes

$$a.\ F_1(\vec{X}) = 4x^2 + 4xy + y^2 \qquad \text{y} \qquad F_2(\vec{Y}) = 16u^2 - 24uv + 9v^2.$$

$$c.\ F_1(\vec{X}) = x^2 + 8xy + 16y^2 \qquad \text{y} \qquad F_2(\vec{Y}) = u^2 + 2uv + v^2.$$

3. Demuestre que la forma cuadrática $F : \mathbb{R}^2 \to \mathbb{R}$ dada por

$$F(x,y) = ax^2 + 2bxy + cy^2$$

tiene rango 2 si y sólo si $ac - b^2 \neq 0$.

4. Determine los valores de α para los cuales la matriz asociada a la forma $F : \mathbb{R}^3 \to \mathbb{R}$ dada por

$$F(x,y,z) = 5x^2 + y^2 + \alpha z^2 + 4xy - 2xz - 2yz$$

tiene valores propios positivos.

5. Sea A una matriz simétrica de tamaño $n \times n$ asociada a la forma cuadrática $F(\vec{X}) = \vec{X}^H A\vec{X}$ con $\vec{X} \in \mathbb{C}$. Demuestre que $F(\vec{X})$ es real.

5.3. Diagonalización de una forma cuadrática

Sea $F(\vec{X}) = \vec{X}^t A\vec{X}$ cualquier forma cuadrática con n variables. Para simplificarla, se pasa de las variables $x_1, x_2, \ldots, x_n$ a las variables $y_1, y_2, \ldots, y_n$ y se supone que las variables anteriores están relacionadas con las nuevas mediante la fórmula $\vec{X} = P\vec{Y}$, donde P es una matriz no singular. Entonces

$$\begin{aligned}
\vec{X}^t A\vec{X} =& (P\vec{Y})^t A(P\vec{Y}) && \textit{puesto que } \vec{X} = P\vec{Y}\\
=& (\vec{Y}^t\ P^t)A(P\vec{Y})\\
=& \vec{Y}^t (P^t AP)\vec{Y} = \vec{Y}^t B\vec{Y} && \textit{donde } B \textit{ es congruente a } A.
\end{aligned}$$

Así, $F(\vec{X})$ es equivalente a una forma cuadrática $F(\vec{Y})$, cuya matriz es B. En las nuevas variables no hay términos mixtos, cuando la matriz B sea triangular. A este proceso se le llama *diagonalización de una forma cuadrática*.

5.3.1. Diagonalización por completación de cuadrados

Un procedimiento para diagonalizar una forma cuadrática es la generalización de la técnica familiar de completar cuadrados, aprendido en el álgebra elemental. El método que se va a estudiar a continuación consiste en obtener una expresión canónica para $F(\vec{X}) = \vec{X}^t A\vec{X}$ en términos de los menores de la matriz asociada. Para facilitar la comprensión se comenzará aplicando este método a las formas cuadráticas de dos y tres variables

Caso I Si se considera una forma cuadrática en dos variables

$$F(\vec{X}) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2, \tag{5.15}$$

entonces $F(\vec{X})$ se puede expresar como

$$\vec{X}^t A\vec{X} = [x_1, x_2]\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22}\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix}.$$

Si cualquiera a_{11} o a_{22} es no nulo, sin pérdida de generalidad se puede asumir que a_{11} es distinto de cero. Entonces (5.15) se puede escribir como

$$\begin{aligned} F(\vec{X}) =& a_{11}\left[x_1^2 + 2\frac{a_{12}}{a_{11}}x_1x_2 + \left(\frac{a_{12}}{a_{11}}\right)^2 x_2^2 - \left(\frac{a_{12}}{a_{11}}\right)^2 x_2^2 + \frac{a_{22}}{a_{11}}x_2^2\right] \\ =& a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \left[\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}}\right)^2\right]x_2^2\right\} \\ =& a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{1}{a_{11}^2}\left[a_{11}a_{22} - a_{12}^2\right]x_2^2\right\}. \end{aligned} \tag{5.16}$$

Si se definen

$$y_1 = x_1 + \frac{a_{12}}{a_{11}}x_2 \qquad \text{y} \qquad y_2 = x_2$$

se tiene que

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2\end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} \\ 0 & 1\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix}.$$

Entonces (5.16), en términos de las nuevas variables queda

$$F(\vec{Y}) = a_{11}y_1^2 + \frac{\det A}{a_{11}}y_2^2 \tag{5.17}$$

y la forma (5.15) ha sido diagonalizada. La transformación de variables es no singular ($\det P = 1$), pero no es ortogonal. Además, los coeficientes de y_1^2, y_2^2 en (5.17) no son, en general, los valores propios de la matriz asociada a la forma cuadrática A.

El lector puede verificar que cuando $a_{22} \neq 0$. Entonces (5.15) se puede escribir como

$$F(\vec{X}) = a_{22}\left\{\left(x_2 + \frac{a_{12}}{a_{22}}x_1\right)^2 + \frac{1}{a_{22}^2}\left[a_{11}a_{22} - a_{12}^2\right]x_1^2\right\}. \tag{5.18}$$

Si se define

$$y_1 = x_2 + \frac{a_{12}}{a_{22}} x_1 \qquad \text{y} \qquad y_2 = x_1$$

dicha transformación de variables, se puede expresar como

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P\vec{X} = \begin{bmatrix} \frac{a_{12}}{a_{11}} & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Esta transformación de variables es no singular ($\det P = -1$), pero no es ortogonal. Al reemplazar en (5.18), se obtiene

$$F(\vec{Y}) = a_{22} y_1^2 + \frac{\det A}{a_{22}} y_2^2. \tag{5.19}$$

En el caso de que a_{11}, a_{22} ambas desaparezcan, el procedimiento anterior no se puede trabajar. Cuando $a_{11} = a_{22} = 0$, la expresión (5.15) se vuelve

$$F(\vec{X}) = 2a_{12} x_1 x_2. \tag{5.20}$$

Ahora se hace la transformación

$$x_1 = y_1 + y_2 \qquad \text{y} \qquad x_2 = y_1 - y_2.$$

La cual se puede expresar matricialmente como

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P\vec{Y} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Esta es una transformación no singular la cual reduce (5.20) a

$$F(\vec{Y}) = 2a_{12}\left(y_1^2 - y_2^2\right). \tag{5.21}$$

En este caso también la forma ha sido diagonalizada.

Caso II Si se considera una forma cuadrática en tres variables

$$F(\vec{X}) = \sum_{i=1}^{3} a_{ii} x_i^2 + 2\sum_{i=1}^{2} \sum_{j=i+1}^{3} a_{ij} x_i x_j. \tag{5.22}$$

$F(\vec{X})$ se puede expresar como

$$\vec{X}^t A \vec{X} = [x_1, x_2, x_3] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Si cualquiera a_{11}, a_{22} o a_{33} es no nulo, sin pérdida de generalidad se puede suponer que $a_{11} \neq 0$. Entonces (5.22) se puede escribir como

$$\begin{aligned}
F(\vec{X}) = a_{11} &\left[x_1^2 + 2x_1 \sum_{k=2}^{3} \frac{a_{1k}}{a_{11}} x_k + \left(\sum_{k=2}^{3} \frac{a_{1k}}{a_{11}} x_k \right)^2 - \left(\sum_{k=2}^{3} \frac{a_{1k}}{a_{11}} x_k \right)^2 + \right. \\
&\left. \sum_{k=2}^{3} \frac{a_{kk}}{a_{11}} x_k^2 + 2\frac{a_{23}}{a_{11}} x_2 x_3 \right] \\
= a_{11} &\left\{ \left(x_1 + \sum_{k=2}^{3} \frac{a_{1k}}{a_{11}} x_k \right)^2 + \left[\sum_{k=2}^{3} \left(\frac{a_{kk}}{a_{11}} - \left(\frac{a_{1k}}{a_{11}} \right)^2 \right) x_k^2 \right] + \right. \\
&\left. 2\left(\frac{a_{23}}{a_{11}} - \frac{a_{12}a_{13}}{a_{11}^2} \right) x_2 x_3 \right\}.
\end{aligned}$$

Ahora, se asume que $M_{33}(A) \neq 0$

$$\begin{aligned}
F(\vec{X}) = a_{11} &\left\{ \left(x_1 + \sum_{k=2}^{3} \frac{a_{1k}}{a_{11}} x_k \right)^2 + \left(\frac{a_{33}}{a_{11}} - \left(\frac{a_{13}}{a_{11}} \right)^2 \right) x_3^2 \right. \\
&+ \left(\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}} \right)^2 \right) \left[x_2 + \left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} \right) x_3 \right]^2 \\
&\left. - \left(\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}} \right)^2 \right) \left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} \right)^2 x_3^2 \right\} \\
= a_{11} &\left\{ \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3 \right)^2 + \right. \\
&\left(\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}} \right)^2 \right) \left[x_2 + \left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} \right) x_3 \right]^2 + \\
&\left. \frac{1}{a_{11}^2} \left[\left(a_{33}a_{11} - a_{13}^2 \right) - \frac{\left(a_{11}a_{23} - a_{12}a_{13} \right)^2}{a_{11}a_{22} - a_{12}^2} \right] x_3^2 \right\} \\
= a_{11} &\left\{ \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3 \right)^2 + \frac{1}{a_{11}^2} \left[\frac{a_{11} \det A}{M_{33}(A)} \right] x_3^2 + \right. \\
&\left. \frac{1}{a_{11}^2} M_{33}(A) \left[x_2 + \left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} \right) x_3 \right]^2 \right\}. \qquad (5.23)
\end{aligned}$$

Con la sustitución

$$
\begin{aligned}
y_1 =& x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3, \\
y_2 =& x_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} x_3, \\
y_3 =& x_3,
\end{aligned}
$$

se tiene que

$$
\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} \\ 0 & 1 & \frac{a_{11}a_{23}-a_{12}a_{13}}{a_{11}a_{22}-a_{12}^2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.
$$

Entonces (5.23), en términos de las nuevas variables queda

$$
F(\vec{Y}) = a_{11} y_1^2 + \frac{M_{33}(A)}{a_{11}}\, y_2^2 + \frac{\det A}{M_{33}(A)}\, y_3^2 \tag{5.24}
$$

y la forma (5.22) ha sido diagonalizada. La transformación de variables es no singular ($\det P = 1$), pero no es ortogonal. Los coeficientes de y_1^2, y_2^2, y_3^2 en (5.24) no son, en general, los valores propios de A.

El lector puede verificar que cuando $M_{22}(A) \neq 0$. Entonces (5.22) se puede escribir como

$$
\begin{aligned}
F(\vec{X}) = a_{11} \Bigg\{ \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3 \right)^2 + \frac{1}{a_{11}} \left[\frac{\det A}{M_{22}(A)} \right] x_2^2 + \\
\frac{1}{a_{11}^2} M_{22}(A) \left[x_3 + \left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{33} - a_{13}^2} \right) x_2 \right]^2 \Bigg\}.
\end{aligned} \tag{5.25}
$$

Si se definen

$$
\begin{aligned}
y_1 =& x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3, \\
y_2 =& x_3 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{33} - a_{13}^2} x_2, \\
y_3 =& x_2,
\end{aligned}
$$

dicha transformación de variables, se puede expresar como

$$
\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} \\ 0 & \frac{a_{11}a_{23}-a_{12}a_{13}}{a_{11}a_{22}-a_{12}^2} & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.
$$

Esta transformación de variables es no singular ($\det P = -1$), pero no es ortogonal. Al reemplazar en (5.25), se obtiene

$$F(\vec{Y}) = a_{11}y_1^2 + \frac{M_{22}(A)}{a_{11}}\, y_2^2 + \frac{\det A}{M_{22}(A)}\, y_3^2. \tag{5.26}$$

El procedimiento descrito puede generalizarse para diagonalizar cualquier forma cuadrática con n variables de la siguiente manera:

Teorema 5.6. Método de Reducción de Lagrange

Sea $F(\vec{X})$ una forma cuadrática en $\mathbb{R}^n$ con matriz asociada $A = [a_{ij}]$:

$$F(\vec{X}) = \vec{X}^t A\vec{X} = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij}x_i x_j.$$

Entonces, existe una matriz triangular superior T con elementos en la diagonal iguales a 1, tal que el cambio de coordenadas $\vec{X} = T\vec{Y}$ transforma a $\vec{X}^t A\vec{X}$ en

$$F(\vec{X}) = \vec{Y}^t T^t A T\vec{Y} = \Delta_1 y_1^2 + \frac{\Delta_2}{\Delta_1}\, y_2^2 + \frac{\Delta_3}{\Delta_2}\, y_3^2 + \ldots + \frac{\Delta_n}{\Delta_{n-1}}\, y_n^2, \tag{5.27}$$

donde $\Delta_i \neq 0$, $i = 1, 2, \ldots, n-1$, son los determinantes de las submatrices angulares[1] $A_{[i]}$ de A. ($\Delta_n = \det A$).

Demostración

El resultado se prueba por inducción. Es claro que para $n = 1$, se cumple trivialmente. Supóngase que es cierto para $n-1$, es decir, existe una matriz de tamaño $(n-1)\times(n-1)$,

$$T_{n-1} = \begin{bmatrix} 1 & t_{12} & t_{13} & \ldots & t_{1n-1} \\ 0 & 1 & t_{23} & \ldots & t_{2n-1} \\ 0 & 0 & 1 & \ldots & t_{3n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

tal que si $\vec{X}_{n-1} = (x_1, x_2, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ y $A_{[n-1]}$ es la submatriz angular de A de orden $n-1$, se tiene que

$$\vec{X}_{n-1}^t A_{[n-1]}\vec{X}_{n-1} = \Delta_1 y_1^2 + \frac{\Delta_2}{\Delta_1}\, y_2^2 + \frac{\Delta_3}{\Delta_2}\, y_3^2 + \ldots + \frac{\Delta_{n-1}}{\Delta_{n-2}}\, y_{n-1}^2. \tag{5.28}$$

[1] Véase definición 1.28

Ahora bien, dado

$$\vec{U} = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{n-1n} \end{bmatrix}$$

la matriz A se puede escribir como

$$A = \begin{bmatrix} A_{[n-1]} & \vdots & \vec{U} \\ \cdots & . & \cdots \\ \vec{U}^t & \vdots & a_{nn} \end{bmatrix}.$$

Si para cualquier $\vec{c} \in \mathbb{R}^{n-1}$ se considera la matriz triangular

$$T = \begin{bmatrix} T_{n-1} & \vdots & \vec{c} \\ \cdots & . & \cdots \\ \vec{0}^t & \vdots & 1 \end{bmatrix},$$

entonces

$$F(\vec{X}) = \vec{X}^t A\vec{X} = \vec{Y}^t T^t AT\vec{Y} = \vec{Y}^t \begin{bmatrix} B_{11} & \vdots & B_{12} \\ \cdots & . & \cdots \\ B_{12}^t & \vdots & B_{22} \end{bmatrix} \vec{Y}, \tag{5.29}$$

donde:

$B_{11} = T_{n-1}^t A_{[n-1]} T_{n-1}$	es una matriz de tamaño $(n-1)\times(n-1)$
$B_{12} = T_{n-1}^t A_{[n-1]}\vec{c} + T_{n-1}^t \vec{U}$	es una matriz de tamaño $(n-1)\times 1$
$B_{22} = \vec{c}^t A_{[n-1]}\vec{c} + 2\vec{c}^t\vec{U} + a_{nn}$	es un escalar.

Si se tiene en cuenta (5.28) resulta

$$[y_1, y_2, \ldots, y_{n-1}]\, B_{11} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix} = [y_1, y_2, \ldots, y_{n-1}] \begin{bmatrix} \Delta_1 & 0 & \ldots & 0 \\ 0 & \frac{\Delta_2}{\Delta_1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{\Delta_{n-1}}{\Delta_{n-2}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}$$

y como B_{22} es un escalar, para completar la prueba y obtener (5.27), bastará encontrar un $\vec{c} \in \mathbb{R}^n$ para el vector columna B_{12} sea nulo. Para ello, dado que

$$B_{12} = T_{n-1}^t \left(A_{[n-1]}\vec{c} + \vec{U}\right) \qquad \text{y} \qquad T_{n-1} \text{ es no singular}$$

$\vec{c}$ debe ser tal que

$$A_{[n-1]}\vec{c}+\vec{U}=\vec{0}.$$

Como $\det\left(A_{[n-1]}\right)\neq 0$, existe un único $\vec{c}=-A_{[n-1]}^{-1}\vec{U}$ para el cual B_{12} es un vector columna nulo. Si se reemplaza en B_{22} se tiene que

$$B_{22}=a_{nn}-\vec{U}^tA_{[n-1]}^{-1}\vec{U}$$

y usando el hecho de que $\det(A)=\det\left(A_{[n-1]}\right)\left(a_{nn}-\vec{U}^tA_{[n-1]}^{-1}\vec{U}\right)$ [2] se obtiene

$$B_{22}=\frac{\det(A)}{\det\left(A_{[n-1]}\right)}=\frac{\Delta_n}{\Delta_{n-1}}.$$

De la expresión (5.29) resulta

$$F(\vec{X})=\vec{Y}^tT^tAT\vec{Y}=\Delta_1y_1^2+\frac{\Delta_2}{\Delta_1}\,y_2^2+\frac{\Delta_3}{\Delta_2}\,y_3^2+\ldots+\frac{\Delta_n}{\Delta_{n-1}}\,y_n^2.$$

Esto completa la prueba.

Ejemplo 5.7. Sea $F(\vec{X})=3x_1^2+3x_2^2+2x_1x_2$. Encuentre una diagonalización por el método descrito.

Solución

Utilizando el resultado obtenido en (5.9); para $n=2$, se tiene que

$$\vec{X}^tA\vec{X}=a_{11}x_1^2+a_{22}x_2^2+2a_{12}x_1x_2.$$

Al comparar los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

$$A=\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$$

luego, la $F(\vec{X})$, se puede expresar como

$$\vec{X}^tA\vec{X}=[x_1,x_2]\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \tag{5.30}$$

En este caso, ya que $\Delta_1=a_{11}=3\neq 0$ y $\Delta_2=\det A=8$, la forma (5.30) se puede diagonalizar de la siguiente manera

$$F(\vec{Y})=3y_1^2+\frac{8}{3}y_2^2$$

[2] Véase Teorema 1.19

donde

$$y_1 = x_1 + \frac{1}{3}x_2 \qquad \text{e} \qquad y_2 = x_2.$$

En forma matricial, se obtiene

$$\vec{Y} = \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{X} \qquad \text{o} \qquad \vec{X} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{Y}.$$

Aquí, se usó el hecho de que la transformación tenía determinante 1 y por lo tanto, era no singular. De este modo

$$\begin{aligned} F(\vec{X}) =& \vec{X}^t \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \vec{X} = \vec{Y}^t \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix}^t \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{Y} \\ =& \vec{Y}^t \begin{bmatrix} 1 & 0 \\ -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & \frac{8}{3} \end{bmatrix} \vec{Y} = \vec{Y}^t \begin{bmatrix} 3 & 0 \\ 0 & \frac{8}{3} \end{bmatrix} \vec{Y}. \end{aligned}$$

Ejemplo 5.8. Considere la forma cuadrática dada en el Ejemplo 5.4 y determine una forma diagonal equivalente.

Solución

Haciendo referencia al Ejemplo 5.4, la $F(\vec{X})$ se puede escribir como

$$\vec{X}^t A \vec{X} = [x_1, x_2, x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{5.31}$$

Dado que

$$\Delta_1 = a_{11} = 2 \neq 0, \qquad \Delta_2 = M_{33}(A) = 6 \qquad \text{y} \qquad \Delta_3 = \det A = 7$$

se tiene que

$$F(\vec{Y}) = 2y_1^2 + 3y_2^2 + \frac{7}{6}\, y_3^2,$$

donde

$$y_1 = x_1 + x_2 + \frac{1}{2}x_3, \qquad y_2 = x_2 + \frac{1}{3}x_3 \qquad \text{e} \qquad y_3 = x_3$$

En forma matricial, se tiene que

$$\vec{Y} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X} \qquad \text{o} \qquad \vec{X} = \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{Y}.$$

En la transformación anterior, se empleó el hecho de que la transformación tenía determinante 1 y por lo tanto, era no singular. De este modo

$$\begin{aligned}
F(\vec{X}) =& \vec{X}^t \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \vec{X} = \vec{Y}^t \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}^t \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{Y} \\
=& \vec{Y}^t \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -\frac{1}{6} & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 2 & 3 & 0 \\ 1 & 1 & \frac{7}{6} \end{bmatrix} \vec{Y} = \vec{Y}^t \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \vec{Y}.
\end{aligned}$$

Teorema 5.7.

Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A. Sea L una matriz no singular triangular inferior tal que A se pueda factorizar como LDL^t. Entonces el cambio de coordenadas

$$\vec{Y} = L^t \vec{X} \tag{5.32}$$

transforma a $\vec{X}^t A \vec{X}$ en $\vec{Y}^t D \vec{Y}$.

Demostración

La matriz A asociada a la forma, se puede factorizar como

$$A = LDU.$$

Como A es simétrica, por el Teorema 3.5, $U = L^t$. Por lo tanto,

$$\begin{aligned}
\vec{X}^t A \vec{X} =& \vec{X}^t \left(LDL^t\right) \vec{X} && \textit{puesto que } A = LDL^t \\
=& \left(\vec{X}^t L\right) D \left(L^t \vec{X}\right) \\
=& \left(L^t \vec{X}\right)^t D \left(L^t \vec{X}\right) = \vec{Y}^t D \vec{Y} && \textit{puesto que } \vec{Y} = L^t \vec{X}.
\end{aligned}$$

Así, queda probado el teorema.

A continuación se presenta una versión de este método de diagonalización.

Procedimiento para diagonalizar una forma cuadrática

$i)$ Halle la matriz de coeficientes simétrica A asociada a $F(\vec{X})$.

$ii)$ Obtenga la descomposición LDL^t de A, sin efectuar intercambios de filas que destruyan la simetría y con elementos en $D = diag\{d_{11}, d_{22}, \ldots, d_{nn}\}$ no necesariamente distintos de cero. Forme L de manera que $\det(L) = 1$.

$iii)$ Transforme a $F(\vec{X})$ en $d_{11}y_1^2 + d_{22}y_2^2 + \ldots + d_{nn}y_n^2$, bajo el cambio de coordenadas $\vec{Y} = L^t\vec{X}$.

Ejemplo 5.9. Considere la forma cuadrática dada en el Ejemplo 5.7 y determine una forma diagonal equivalente por el método descrito.

Solución

La factorización LDL^t de la matriz asociada a la forma cuadrática es

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{8}{3} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{bmatrix}.$$

Luego, la $F(\vec{X})$ se puede expresar como

$$\vec{X}^t A \vec{X} = [x_1, x_2] \begin{bmatrix} 1 & 0 \\ \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{8}{3} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{5.33}$$

Como L es no singular. Se hace,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

$F(\vec{X})$ se puede escribir en términos de las variables y_1, y_2 como

$$F(\vec{Y}) = 3y_1^2 + \frac{8}{3}y_2^2.$$

Ejemplo 5.10. Considere la forma cuadrática dada en el Ejemplo 5.4 y determine una forma diagonal equivalente.

Solución

La factorización LDL^t de la matriz asociada a la forma cuadrática es

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

de modo que

$$F(\vec{X}) = \vec{X}^t \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \vec{X} = \vec{X}^t \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X}.$$

Nótese que L es no singular. Por lo tanto, se puede hacer el cambio de variable

$$\vec{Y} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X},$$

se puede escribir $F(\vec{X})$ en términos de las variables y_1, y_2, y_3 como

$$F(\vec{Y}) = 2y_1^2 + 3y_2^2 + \frac{7}{6} y_3^2.$$

5.3.2. Diagonalización por transformación ortogonal

En la diagonalización de la forma cuadrática $F(\vec{X}) = \vec{X}^t A\vec{X}$ a la forma $F(\vec{Y}) = \vec{Y}^t B\vec{Y}$, por el método de completación de cuadrados, sólo se exigió que la matriz P fuera no singular. Cuando la matriz A asociada a la forma cuadrática sea simétrica, entonces, se puede diagonalizar a $F(\vec{X})$ mediante una matriz P ortogonal en cuyo caso los elementos sobre la diagonal de la matriz B son los valores propios de la matriz A.

Teorema 5.8. Teorema de los ejes principales

Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A con valores propios (no necesariamente distintos) $\lambda_1, \lambda_2, \ldots, \lambda_n$. Sea Q una matriz ortogonal propia que diagonaliza a A. Entonces el cambio de coordenadas

$$\vec{X} = Q\vec{Y} \tag{5.34}$$

transforma a $\vec{X}^t A\vec{X}$ en $\vec{Y}^t D\vec{Y}$, donde $D = Q^t AQ = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

Demostración

La demostración consiste en un cálculo directo

$$\begin{aligned} \vec{X}^t A\vec{X} =& (Q\vec{Y})^t A(Q\vec{Y}) && \textit{puesto que } \vec{X} = Q\vec{Y} \\ =& (\vec{Y}^t\ Q^t)A(Q\vec{Y}) \\ =& \vec{Y}^t (Q^t AQ)\vec{Y} = \vec{Y}^t D\vec{Y} && \textit{puesto que } Q \textit{ diagonaliza a } A \end{aligned}$$

así, queda el teorema probado.

El Teorema 5.8 se llama *Teorema de los ejes principales* porque define nuevos ejes (los ejes principales) con respecto a los cuales la forma cuadrática tiene una expresión paticularmente simple.

A continuación se presenta una versión de este método de diagonalización.

Procedimiento para diagonalizar una forma cuadrática

i) Halle la matriz de coeficientes simétrica A asociada a $F(\vec{X})$.

ii) Encuentre los valores propios (no necesariamente distintos), $\lambda_1, \lambda_2, \ldots, \lambda_n$ de A.

iii) Obtenga una base ortonormal para $\mathbb{R}^n$ formada por los vectores propios normalizados de A.

iv) Forme la matriz Q cuyas columnas sean los vectores de la base hallada en el paso *iii*) en el orden correspondiente al listado de los valores propios en el paso *ii*). La transformación $\vec{X} = Q\vec{Y}$ es una **rotación** si $\det(Q) = 1$.

v) Transforme a $F(\vec{X})$ en $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2$.

Nota
Si Q es una matriz ortogonal impropia, es decir, $\det Q = -1$, se debe cambiar de signo todas las componentes de un sólo vector columna (o intercambiar dos vectores columnas de Q).

Ejemplo 5.11. Determine los ejes principales de la forma cuadrática dada en el Ejemplo 5.7.

Solución
En el Ejemplo 5.7, se obtuvo que la $F(\vec{X})$, se puede escribir como

$$\vec{X}^t A \vec{X} = [x_1, x_2] \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{5.35}$$

En este caso, los valores propios de A son $\lambda_1 = 2$ y $\lambda_2 = 4$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, respectivamente.

Para encontrar Q, como $\|\vec{v}_1\| = \sqrt{2}$, se hace $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ y dado que $\|\vec{v}_2\| = \sqrt{2}$ se tiene $\vec{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. Se puede verificar que la base obtenida para $\mathbb{R}^2$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por lo tanto,

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

Como el $\det(Q) = -1$, intercambiamos las columnas y se tiene

$$Q = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \qquad \text{y} \qquad Q^t A Q = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Si se definen los ejes principales como sigue

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = Q^t \vec{X} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

entonces, (5.35) se puede escribir en términos de las nuevas variables y_1, y_2 como $\vec{Y}^t D \vec{Y}$, o sea,

$$4y_1^2 + 2y_2^2. \tag{5.36}$$

Ejemplo 5.12. Considere la forma cuadrática dada en el Ejemplo 5.4 y determine sus ejes principales.

Solución

Haciendo referencia al Ejemplo 5.4, la $F(\vec{X})$, se puede escribir como

$$\vec{X}^t A \vec{X} = [x_1, x_2, x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{5.37}$$

Del Ejemplo 3.11, se tiene que

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \qquad \text{y} \qquad Q^t A Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

Por lo tanto, (5.37) se puede escribir en términos de las nuevas variables y_1, y_2, y_3 como $\vec{Y}^t D \vec{Y}$, o sea,

$$y_1^2 + y_2^2 + 7y_3^2, \tag{5.38}$$

donde los ejes principales, se obtienen como sigue

$$\vec{Y}=\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=Q^t\vec{X}=\begin{bmatrix}-1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}.$$

5.4. Ley de la inercia para formas cuadráticas

Sea $F:\mathbb{R}^n\to\mathbb{R}$ una forma cuadrática. Si $\rho(A)=r$, entonces toda matriz simétrica que represente a F también tiene rango r. En particular todas las formas diagonales a las que F sea semejante, mediante una transformación lineal real invertible de variables, tendrán exactamente r coeficientes no nulos. Además, todas las formas diagonales a las que reduzcamos F tienen el mismo número de coeficientes positivos y el mismo número de coeficientes negativos, como se afirma en el resultado obtenido por *Sylvester*.

Teorema 5.9. Ley de la inercia de Sylvester

Sean A una matriz simétrica de tamaño $n\times n$ y P una matriz no singular del mismo tamaño, entonces

$$In(A)=In\left(P^TAP\right).$$

Demostración

Sea Q una matriz ortogonal tal que

$$Q^tAQ=D=\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix},$$

de forma que $\lambda_1,\lambda_2,\dots,\lambda_i$ son positivos, $\lambda_{i+1},\lambda_{i+2},\dots,\lambda_{i+j}$ son negativos y el resto nulos.

Sean $\widehat{A}$ la matriz simétrica P^tAP y W una matriz ortogonal tal que

$$W^t\widehat{A}W=\widehat{D}=\begin{bmatrix}\widehat{\lambda}_1 & 0 & \dots & 0\\ 0 & \widehat{\lambda}_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \widehat{\lambda}_n\end{bmatrix},$$

de forma que $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_{\widehat{i}}$ son positivos, $\widehat{\lambda}_{\widehat{i}+1}, \widehat{\lambda}_{\widehat{i}+2}, \ldots, \widehat{\lambda}_{\widehat{i}+\widehat{j}}$ son negativos y el resto nulos.

Se prueba por contradicción que $i = \widehat{i}$. Supongamos que $\widehat{i} > i$.

Sean $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_i$ las primeras i columnas de la matriz Q. Sean $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n$ las filas de la matriz $R = W^t P^{-1}$. Se forma una matriz B de tamaño $n \times n$, cuyas primeras filas sean las primeras i columnas de Q traspuestas y el resto sean las filas $\widehat{i}+1, \widehat{i}+2, \ldots, n$ de R, es decir,

$$B = \begin{bmatrix} \vec{q}_1^t \\ \vec{q}_2^t \\ \vdots \\ \vec{q}_i^t \\ \vec{r}_{\widehat{i}+1} \\ \vec{r}_{\widehat{i}+2} \\ \vdots \\ \vec{r}_n \end{bmatrix}$$

es una matriz real de tamaño $(i+n-\widehat{i}) \times n$, donde $i+n-\widehat{i} < n$. Por consiguiente, el sistema lineal homogéneo cuya matriz de coeficientes es B tiene solución distinta de la trivial y por tanto existe un vector $\vec{u} \neq \vec{0}$, tal que $B\vec{u} = \vec{0}$. En otras palabras, el producto de cualquier fila de la matriz B por el vector $\vec{u}$ es cero, es decir

$$\begin{aligned} \vec{q}_k^t \vec{u} &= 0 \qquad && \text{para } k = 1, 2, \ldots, i \\ \vec{r}_k^t \vec{u} &= 0 \qquad && \text{para } k = \widehat{i}+1, \widehat{i}+2, \ldots, n. \end{aligned}$$

Al evaluar $\vec{u}^t A \vec{u}$, se tiene que

$$\vec{u}^t A \vec{u} = \vec{u}^t Q D Q^t \vec{u}$$

$$\begin{aligned} \vec{u}^t A \vec{u} &= \begin{bmatrix} 0 & \ldots & 0 & v_{i+1} & \ldots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix} \\ &= \sum_{s=i+1}^{n} \lambda_s v_s^2 < 0, \end{aligned} \tag{5.39}$$

donde, $v_s = \vec{q}_s^t \vec{u}$ para $s = i+1, i+2, \ldots, n$. Por otra parte,

$$\vec{u}^t A\vec{u} = \vec{u}^t (P^t)^{-1} \widehat{A} P^{-1} \vec{u} = \vec{u}^t (P^t)^{-1} W \widehat{D} W^t P^{-1} \vec{u} = \vec{u}^t R^t \widehat{D} R\vec{u}.$$

Si se denota $w_k = \vec{r}_k \vec{u}$ para $k = 1, 2, \ldots, \widehat{i}$, se tiene

$$\begin{aligned} \vec{u}^t A\vec{u} =& \vec{u}^t R^t \widehat{D} R\vec{u} \\ =& \begin{bmatrix} w_1 & \ldots & w_{\widehat{i}} & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} \widehat{\lambda}_1 & 0 & \ldots & 0 \\ 0 & \widehat{\lambda}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \widehat{\lambda}_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_{\widehat{i}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sum_{s=1}^{\widehat{i}} \widehat{\lambda}_s w_s^2 > 0, \end{aligned}$$

lo que contradice (5.39).

Análogamente se demuestra que la hipótesis $\widehat{i} < i$ conduce a una contradicción. Por lo tanto, se debe tener que $\widehat{i} = i$.

Con $\widehat{j}$ y j se procede de la misma forma.

Teorema 5.10. Teorema de Euler

Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A. El valor de $F\left(\vec{X}\right) = \vec{X}^t A\vec{X}$ en un vector unitario $\vec{U}$ es

$$F\left(\vec{U}\right) = \sum_{j=1}^{n} \lambda_j \cos^2 \theta_j \tag{5.40}$$

donde los λ_j son los valores propios de la matriz simétrica A y los ángulos θ_j son los ángulos entre $\vec{X}$ y los vectores propios ortonormalizados $\vec{q}_j$ correspondientes a los λ_j respectivamente.

Demostración

Sean $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n$ los vectores propios ortonormalizados de la matriz A y sea $\vec{U}$ un vector unitario arbitrario. Supongamos que θ_j representa el ángulo entre $\vec{U}$ y $\vec{q}_j$, así que

$$\cos\theta_j = \vec{q}_j \cdot \vec{U} = \vec{q}_j^t \vec{U}.$$

Si se forma $Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \dots & \vec{q}_n \end{bmatrix}$ y se obtiene $\vec{Y}$ por la transformación $\vec{U} = Q\vec{Y}$ o $\vec{Y} = Q^t\vec{U}$. Entonces,

$$\vec{Y} = Q^t\vec{U} = \begin{bmatrix} \vec{q}_1^t \\ \vec{q}_2^t \\ \vdots \\ \vec{q}_j^t \\ \vdots \\ \vec{q}_n^t \end{bmatrix} \vec{U} = \begin{bmatrix} \cos\theta_1 \\ \cos\theta_2 \\ \vdots \\ \cos\theta_j \\ \vdots \\ \cos\theta_n \end{bmatrix}.$$

Por lo tanto,

$$F\left(\vec{U}\right) = \vec{U}^tA\vec{U} = \vec{Y}^tQ^tAQ\vec{Y} = \vec{Y}^tD\vec{Y} = \sum_{j=1}^{n} \lambda_j \cos^2\theta_j$$

y el teorema queda probado.

Ejercicios 5.3.

1. Diagonalícense cada una de las siguientes formas cuadráticas por completación de cuadrados y mediante transformación ortogonal

   a. $F(x,y) = 4x^2 + 4xy + y^2$.

   b. $F(x,y) = 9x^2 - 24xy + 16y^2$.

   c. $F(x,y) = x^2 + 8xy + 16y^2$.

   d. $F(x,y) = x^2 + 2xy + y^2$

   e. $F(x,y,z) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz$.

   f. $F(x,y,z) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz$.

   g. $F(x,y,z) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz$.

2. Hallar una condición necesaria y suficiente en a, b y c tal que la forma cuadrática $ax^2 + by^2 + cxy$ se pueda expresar como ku^2.

5.5. Aplicaciones a la geometría analítica

En esta sección se pretende poner al alcance de los lectores un algoritmo proporcionado por el método de "valores propios y vectores propios", para tratar con más generalidad, agilidad y libertad algunos objetos de la geometría analítica de no fácil manipulación por los métodos tradicionales usados para el estudio de las ecuaciones cuadráticas.

Definición 5.10. Ecuación Cuadrática

Una ecuación en las variables x y y de la forma

$$ax^2+by^2+cxy+dx+ey+f=0 \tag{5.41}$$

donde $a,\ b,\dots,f$ son números reales con al menos uno de los números $a,\ b,\ c,$ distinto de cero, se denomina **ecuación cuadrática de segundo grado**. Esta se puede escribir en forma matricial como

$$\vec{X}^tA\vec{X}+K.\vec{X}+f=0. \tag{5.42}$$

En esta notación, la expresión $\vec{X}^tA\vec{X}$ es la **forma cuadrática asociada** y la matriz K de tamaño 1×2, es $K=\begin{bmatrix} d & e \end{bmatrix}$.

Definición 5.11. Tipos de ecuación cuadrática

Las curvas representadas por la ecuación cuadrática de segundo grado dada en (5.41) se pueden clasificar según la posición en la cual estén con respecto a un sistema coordenado cartesiano **X**, así:

1. Estándar o canónica, si tiene su centro en el origen.

2. Trasladada, si tiene su centro en un punto diferente del origen.

3. Rotada, si su posición con respecto al sistema **X** no es canónica ni tampoco trasladada, pero es posible encontrar un sistema **Y**, con el mismo origen del

sistema **X** y tal que los ejes coordenados de **Y** forman con los ejes coordenados del sistema **X** un ángulo agudo θ, con respecto al cual la curva está en posición canónica

Ejemplo 5.13. Dada la ecuación cuadrática

$$3x^2+2xy+3y^2-9=0, \tag{5.43}$$

elimine el término cruzado xy utilizando el Teorema 5.8, escriba la ecuación en términos de las nuevas variables e identifique la sección cónica obtenida.

Solución

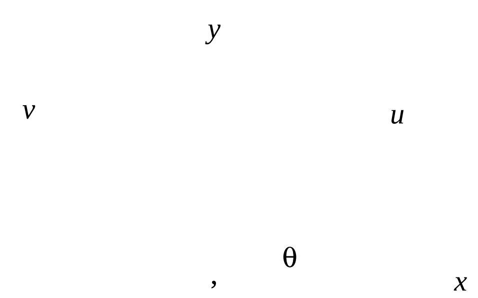
En el Ejemplo 5.11, se vio que la forma cuadrática asociada $3x^2+2xy+3y^2$ se puede expresar como

$$4u^2+2v^2.$$

Luego, la ecuación cuadrática dada en (5.43), se puede escribir como

$$4u^2+2v^2=9,$$

la cual es la ecuación de una elipse estándar. Por lo tanto (5.43), es la ecuación de una elipse estándar rotada. Vea la siguiente figura



Ejemplo 5.14. Dada la ecuación cuadrática

$$2x^2 + 4xy - y^2 - 2x + 3y - 6 = 0, \tag{5.44}$$

elimine el término cruzado xy utilizando el Teorema 5.8, escriba la ecuación en términos de las nuevas variables e identifique la sección cónica obtenida.

Solución

La forma cuadrática asociada $2x^2 + 4xy - y^2$ se puede expresar como

$$\vec{X}^t A \vec{X} = [x, y] \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.45}$$

En este caso, los valores propios de A son $\lambda_1 = -2$ y $\lambda_2 = 3$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, respectivamente.

Para encontrar Q se usa la expresión (5.6), dada en la definición de inercia, la cual establece primero los valores propios positivos y luego los negativos. De esta manera, como $\|\vec{v}_2\| = \sqrt{5}$, se hace $\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$ y dado que $\|\vec{v}_1\| = \sqrt{5}$ se tiene $\vec{u}_1 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. Se puede verificar que la base obtenida para $\mathbb{R}^2$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por lo tanto,

$$Q = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \qquad \text{y} \qquad Q^t A Q = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}.$$

Dado que el $\det(Q) = 1$, se define

$$\vec{Y} = \begin{bmatrix} u \\ v \end{bmatrix} = Q^t \vec{X} = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.46}$$

Entonces, (5.45) se puede escribir en términos de las nuevas variables u, v como $\vec{Y}^t D \vec{Y}$, o sea,

$$3u^2 - 2v^2. \tag{5.47}$$

Si se expresa ahora toda la ecuación dada en (5.44), en la forma matricial dada en (5.42) queda de la siguiente manera:

$$\vec{X}^t A \vec{X} + K\vec{X} - 6 = 0,$$

donde $K=\begin{bmatrix}-2 & 3\end{bmatrix}$ y al hacer el cambio de variable propuesto en (5.46), se tiene

$$K\vec{X}=KQ\vec{Y}=\begin{bmatrix}-2 & 3\end{bmatrix}\begin{bmatrix}2/\sqrt{5} & -1/\sqrt{5}\\ 1/\sqrt{5} & 2/\sqrt{5}\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}=\frac{1}{\sqrt{5}}\begin{bmatrix}-1 & -8\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}.$$

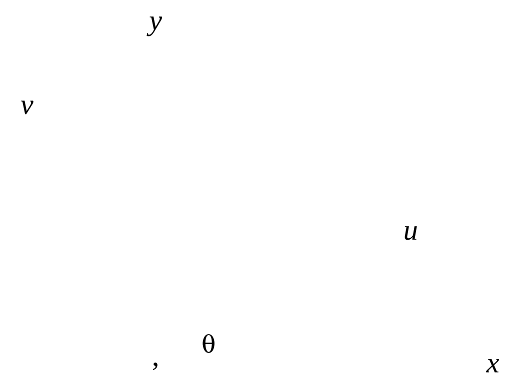
Luego, la ecuación cuadrática dada en (5.44) se puede reescribir como

$$3u^2-2v^2-\frac{1}{\sqrt{5}}u-\frac{8}{\sqrt{5}}v=6.$$

Naturalmente esta no es la ecuación de una cónica en posición canónica, pero si está trasladada porque al completar los cuadrados se obtiene

$$3\left[u^2-\frac{1}{3\sqrt{5}}u+\frac{1}{180}\right]-2\left[v^2+\frac{4}{\sqrt{5}}v+\frac{4}{5}\right]=6-\frac{8}{5}+\frac{1}{60}$$
$$3\left[u-\frac{1}{6\sqrt{5}}\right]^2-2\left[v+\frac{2}{\sqrt{5}}\right]^2=\frac{53}{12},$$

la cual es la ecuación de una **hipérbola** con centro en $\left(\frac{1}{6\sqrt{5}},-\frac{2}{\sqrt{5}}\right)$. Por lo tanto, la ecuación (5.44) es una hipérbola rotada y trasladada. La gráfica es



5.5.1. Rotación de ejes en $\mathbb{R}^2$

Como ya hemos señalado una transformación $\vec{Y} = Q\vec{X}$, donde Q es ortogonal, se llama *transformación ortogonal*. Ahora examinemos la interpretación geométrica de estas transformaciones en $\mathbb{R}^2$.

Teorema 5.11. Rotación de ejes en $\mathbb{R}^2$

Sean $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ una base del sistema coordenado **X** y $\mathcal{B}' = \left\{\vec{e}_1^{\,\prime}, \vec{e}_2^{\,\prime}\right\}$ la base correspondiente al sistema **Y**. Entonces si

$$\vec{e}_1^{\,\prime} = a_1\vec{e}_1 + a_2\vec{e}_2 \qquad \text{y} \qquad \vec{e}_2^{\,\prime} = b_1\vec{e}_1 + b_2\vec{e}_2$$

las coordenadas (x_1, x_2) de un punto cualquiera S en el sistema **X** y las coordenadas (y_1, y_2) del mismo punto en el sistema **Y**, están relacionadas como sigue

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

donde, P es la matriz de transición (o matriz de cambio de base) de $\mathcal{B}$ a $\mathcal{B}'$.

Demostración

Supongamos que los sistemas coordenados **X** y **Y** (en $\mathbb{R}^2$) tienen el mismo origen O. Sea $\overrightarrow{OS}$ el vector formado desde el origen del sistema coordenado **X** al punto $S = (x_1, x_2)$, entonces

$$\overrightarrow{OS} = x_1\vec{e}_1 + x_2\vec{e}_2.$$

Por otra parte, el vector $\overrightarrow{OS}$ formado desde el origen del sistema coordenado **Y** al punto $S = (y_1, y_2)$, es

$$\begin{aligned} \overrightarrow{OS} &= y_1\vec{e}_1^{\,\prime} + y_2\vec{e}_2^{\,\prime} \\ &= y_1\left(a_1\vec{e}_1 + a_2\vec{e}_2\right) + y_2\left(b_1\vec{e}_1 + b_2\vec{e}_2\right) \\ &= \underbrace{(y_1a_1 + y_2b_1)}\vec{e}_1 + \underbrace{(y_1a_2 + y_2b_2)}\vec{e}_2 \\ &= \qquad x_1 \qquad \vec{e}_1 + \qquad x_2 \qquad \vec{e}_2, \end{aligned}$$

aquí se uso el hecho de que la representación de OS como combinación lineal de $\vec{e}_1$ y $\vec{e}_2$ es única. Luego,

$$x_1 = a_1y_1 + a_2y_2 \qquad \text{y} \qquad x_2 = b_1y_1 + b_2y_2$$

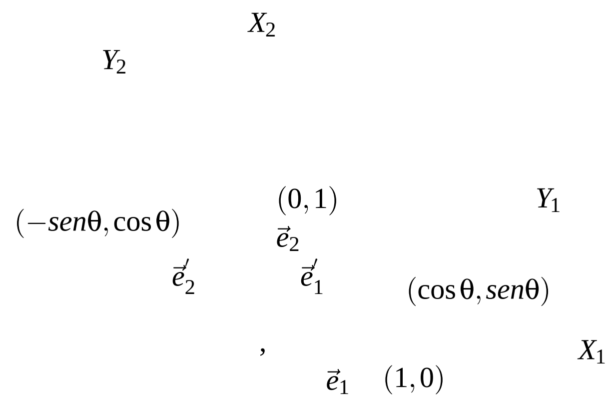
y el teorema queda probado.

Cambio de la dirección de los ejes conservando el mismo origen

Consideremos que el sistema coordenado **X** (en $\mathbb{R}^2$), es rectangular y tiene unidades iguales sobre ambos ejes. Esto significa que $\vec{e}_1$ y $\vec{e}_2$, que son los vectores base (unitarios), son perpendiculares entre sí.

Supongamos que los ejes coordenados **Y** se obtienen haciendo girar el sistema **X** un ángulo θ alrededor del origen, en sentido contrario al de las manecillas del reloj, conservando la ortogonalidad. Los vectores base $\vec{e}_1^{\,\prime}$ y $\vec{e}_2^{\,\prime}$ del sistema **Y**, forman también una base ortonormal y están dados por:

$$\begin{aligned}\vec{e}_1^{\,\prime} =&[\cos\theta, \operatorname{sen}\theta] = \cos\theta\vec{e}_1 + \operatorname{sen}\theta\vec{e}_2.\\ \vec{e}_2^{\,\prime} =&[-\operatorname{sen}\theta, \cos\theta] = -\operatorname{sen}\theta\vec{e}_1 + \cos\theta\vec{e}_2.\end{aligned}$$



Se deduce por el Teorema 5.11, que las coordenadas de un punto en ambos sistemas están relacionadas por

$$x_1 = y_1\cos\theta - y_2\operatorname{sen}\theta \qquad \text{y} \qquad x_2 = y_1\operatorname{sen}\theta + y_2\cos\theta, \tag{5.48}$$

que son las ecuaciones de una rotación levógira de ejes, cuando el ángulo girado es θ. Si se denota por A_θ, la matriz ortogonal

$$A_\theta = \begin{bmatrix}\cos\theta & -\operatorname{sen}\theta\\ \operatorname{sen}\theta & \cos\theta\end{bmatrix} \tag{5.49}$$

entonces (5.48), se puede expresar matricialmente como

$$\begin{bmatrix}x_1\\x_2\end{bmatrix} = A_\theta\begin{bmatrix}y_1\\y_2\end{bmatrix}$$

la cual es una transformación ortogonal propia, puesto que el $\det A_\theta = 1$ y representa cualquier rotación de ejes en $\mathbb{R}^2$.

A continuación se presenta un teorema para obtener el ángulo de rotación θ de una ecuación cuadrática de segundo grado.

Teorema 5.12. Sea la ecuación general de segundo grado

$$ax^2+by^2+cxy+dx+ey+f=0. \tag{5.50}$$

Entonces existe un número único θ, llamado **ángulo de rotación**, tal que

$$\theta=\frac{1}{2}\ \arctan\left(\frac{c}{a-b}\right), \qquad \theta\in\left[0,\frac{\pi}{2}\right). \tag{5.51}$$

Demostración

Sea $A=\begin{bmatrix} a & \frac{c}{2}\\ \frac{c}{2} & b\end{bmatrix}$ la matriz simétrica que representa la forma cuadrática asociada a la ecuación (5.50). Supongamos que λ es un valor propio de A con vector propio unitario correspondiente $\vec{X}=\begin{bmatrix} cos\theta\\ sen\theta\end{bmatrix}$, luego

$$A\vec{X}=\lambda\vec{X}.$$

Si se premultiplica por $\vec{X}^t$ se obtiene

$$\vec{X}^tA\vec{X}=\lambda\vec{X}^t\vec{X}=\lambda.$$

Al sustituir $\vec{X}$ y A se tiene

$$\begin{aligned}\lambda=&\begin{bmatrix}\cos\theta & \operatorname{sen}\theta\end{bmatrix}\begin{bmatrix} a & \frac{c}{2}\\ \frac{c}{2} & b\end{bmatrix}\begin{bmatrix}\cos\theta\\ \operatorname{sen}\theta\end{bmatrix}\\ =&a\ \cos^2\theta+c\ \operatorname{sen}\theta\cos\theta+b\ sen^2\theta\\ =&(a-b)\ \cos^2\theta+c\ \operatorname{sen}\theta\cos\theta+b.\end{aligned}$$

Si se multiplica por 2 y se utilizan las siguientes identidades trigonométricas

$$2\cos^2\theta=\cos(2\theta)+1 \qquad\qquad \text{y} \qquad\qquad \operatorname{sen}(2\theta)=2\operatorname{sen}\theta\cos\theta$$

se llega a

$$2\lambda=(a-b)\ \cos(2\theta)+c\ \operatorname{sen}(2\theta)+(a+b).$$

Pero como λ es valor propio de A se tiene que

$$2\lambda = tr(A) \pm \sqrt{tr^2(A) - 4\det(A)}.$$

Al igualar estas dos expresiones, se obtiene

$$(a-b)\ cos(2\theta) + c\ sen(2\theta) + (a+b) = tr(A) \pm \sqrt{tr^2(A) - 4\det(A)},$$

pero $tr(A) = a + b$, por consiguiente

$$\begin{aligned}
(a-b)\ cos(2\theta) + c\ sen(2\theta) &= \pm\sqrt{tr^2(A) - 4\det(A)} \\
\underbrace{\pm\frac{a-b}{\sqrt{tr^2(A)-4\det(A)}}}_{u}\ cos(2\theta) + \underbrace{\frac{\pm c}{\sqrt{tr^2(A)-4\det(A)}}}_{v}\ sen(2\theta) &= 1 \\
u\ cos(2\theta) + v\ sen(2\theta) &= 1.
\end{aligned}$$

Luego,

$$\begin{aligned}
(v\ sen(2\theta))^2 =& (1 - u\ cos(2\theta))^2 \\
v^2\ sen^2(2\theta) =& 1 - 2u\ cos(2\theta) + u^2\ cos^2(2\theta) \\
v^2(1 - cos^2(2\theta)) =& 1 - 2u\ cos(2\theta) + u^2\ cos^2(2\theta).
\end{aligned}$$

El lector puede probar fácilmente que $u^2 + v^2 = 1$, de manera que

$$\begin{aligned}
cos^2(2\theta) - 2u\ cos(2\theta) + (1 - v^2) &= 0 \\
[cos(2\theta) - u]^2 &= 0 \\
\therefore\ u &= cos(2\theta).
\end{aligned}$$

Por lo tanto, $v = sen(2\theta)$ y así entonces

$$\frac{v}{u} = \tan(2\theta) = \frac{c}{a-b}. \tag{5.52}$$

Al aplicar arctan a ambos lados de (5.52) se obtiene la ecuación (5.51).

Nota

Si se usa la siguiente identidad trigonométrica

$$\tan(2\theta) = \frac{2\tan\theta}{1 - \tan^2\theta},$$

se tiene que

$$\begin{aligned}\frac{2\tan\theta}{1-\tan^2\theta} =& \frac{c}{a-b}\\ 2(a-b)\tan\theta =& c\left(1-\tan^2\theta\right).\end{aligned}$$

Al resolver esta ecuación cuadrática, se llega a

$$\tan\theta = \frac{b-a}{c} \pm \frac{\sqrt{(a-b)^2+c^2}}{c}. \tag{5.53}$$

Ejemplo 5.15. Determine el ángulo de rotación de la ecuación cuadrática dada en el Ejemplo 5.13.

Solución

La matriz simétrica asociada a la forma cuadrática es

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

La ecuación (5.52) no es aplicable ya que como $a = b$ se dividiría por 0, luego se usa (5.53) y se tiene que

$$\tan\theta = 1. \tag{5.54}$$

Cualquier solución de (5.54) sirve a nuestro propósito; si se escoge la solución para la cual $0 < \theta < 90°$, entonces $sen(\theta) = \frac{1}{\sqrt{2}}$ y $\cos(\theta) = \frac{1}{\sqrt{2}}$, es decir, el ángulo de rotación θ vale, aproximadamente $\theta = 45°$ y, construyendo la matriz de rotación 5.49, se tiene que

$$A_\theta = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix},$$

la cual coincide con la matriz ortogonal dada en el Ejemplo 5.11.

Ejemplo 5.16. Determine el ángulo de rotación de la ecuación cuadrática dada en el Ejemplo 5.14.

Solución

La matriz simétrica asociada a la forma cuadrática es

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

Por la ecuación (5.52) se tiene que

$$\tan(2\theta) = \frac{4}{2-(-1)} = \frac{4}{3}. \tag{5.55}$$

Cualquier solución de (5.55) sirve a nuestro propósito. Si se escoge la solución para la cual $0 < 2\theta < 90^\circ$, entonces $sen(2\theta) = \frac{4}{5}$ y $\cos(2\theta) = \frac{3}{5}$ y como θ es un ángulo agudo

$$sen\theta = \sqrt{\frac{1-\cos 2\theta}{2}} = \frac{1}{\sqrt{5}}, \qquad \cos\theta = \sqrt{\frac{1+\cos 2\theta}{2}} = \frac{2}{\sqrt{5}}.$$

Es decir, el ángulo de rotación θ vale aproximadamente $\theta = 26^\circ 33^{'} 54{,}18^{"}$ y, construyendo la matriz de rotación 5.49, se tiene que

$$A_\theta = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix},$$

la cual coincide con la matriz ortogonal dada en el Ejemplo 5.14

5.5.2. Clasificación de las ecuaciones cuadráticas

Para la ecuación general de segundo grado

$$ax^2 + by^2 + cxy + dx + ey + f = 0, \tag{5.56}$$

se definen las cantidades

$$\omega = a+b, \qquad \mu = \det \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix} \qquad \text{y} \qquad \nu = \det \begin{bmatrix} a & \frac{c}{2} & \frac{d}{2} \\ \frac{c}{2} & b & \frac{e}{2} \\ \frac{d}{2} & \frac{e}{2} & f \end{bmatrix}.$$

Entonces la ecuación cuadrática (5.56) representa los siguientes

$$\begin{array}{c}\textbf{LUGARES}\\ \textbf{GEOMÉTRICOS}\end{array}\begin{cases} \mu \neq 0 & \begin{cases} \mu > 0 & \begin{cases} \nu \neq 0 & \begin{cases} \omega\nu < 0 & \text{Una elipse} \\ \omega\nu > 0 & \text{Ninguno} \end{cases} \\ \nu = 0 & \text{Un punto} \end{cases} \\ \mu < 0 & \begin{cases} \nu \neq 0 & \text{Una hipérbola} \\ \nu = 0 & \text{Dos rectas que se cortan} \end{cases} \end{cases} \\ \mu = 0 & \begin{cases} \nu \neq 0 & \text{Una parábola} \\ \nu = 0 & \begin{array}{l}\text{Dos rectas paralelas,} \\ \text{o una recta, o ninguno.}\end{array} \end{cases} \end{cases}$$

El cuadro anterior se acostumbra a interpretar como sigue

$$\textbf{LUGARES GEOMÉTRICOS}\begin{cases} \textbf{Curva con Centro } (\mu\neq 0) \begin{cases} \textbf{De tipo elíptico } (\mu>0) \begin{cases}\text{Una elipse } (\omega\nu<0)\\ \text{Caso degenerado } (\omega\nu\geq 0)\end{cases} \\ \textbf{De tipo hiperbólico } (\mu<0)\begin{cases}\text{Una hipérbola } (\nu\neq 0)\\ \text{Caso degenerado } (\nu=0)\end{cases}\end{cases}\\ \textbf{Curva sin Centro } (\mu=0)\begin{cases}\text{Una parábola } (\nu\neq 0)\\ \text{Caso degenerado } (\nu=0).\end{cases}\end{cases}$$

Las ecuaciones cuadráticas también pueden ser clasificadas de acuerdo a la inercia de la matriz A asociada a la forma cuadrática, como sigue:

Identificación de las ecuaciones cuadráticas	
Inercia	**Nombre de la cónica**
$In(A)=(2,0,0)$	Elipse
$In(A)=(1,1,0)$	Hipérbola
$In(A)=(1,0,1)$	Parábola

Las formas cuadráticas también pueden usarse para analizar ecuaciones de superficies cuádricas en el espacio.

Definición 5.12. Superficie Cuádrica

Una ecuación de segundo grado en x, y, y z de la forma

$$ax^2+by^2+cz^2+dxy+exz+fyz+gx+hy+iz+j=0, \tag{5.57}$$

donde $a,\ b,\dots,j$ son números reales y $|a|+|b|+|c|+|d|+|e|+|f|\neq 0$, se denomina **superficie cuádrica**. Esta se puede escribir en forma matricial como

$$\vec{X}^t A\vec{X}+K\vec{X}+j=0. \tag{5.58}$$

En esta notación, la expresión $\vec{X}^t A\vec{X}$ es la **forma cuadrática asociada** y la matriz K de tamaño 1×3, es $K=[g\ \ h\ \ i]$.

Ejemplo 5.17. Considere la ecuación cuadrática

$$2x^2+5y^2+2z^2+4xy+2xz+4yz-36=0. \tag{5.59}$$

Determine la superficie cuádrica obtenida al eliminar los términos de productos cruzados.

Solución

Haciendo referencia al Ejemplo 5.12, dicha ecuación cuadrática se puede escribir como

$$\vec{X}^t A\vec{X}=36, \tag{5.60}$$

donde $A=\begin{bmatrix}2&2&1\\2&5&2\\1&2&2\end{bmatrix}$. En este caso, (5.59) se puede escribir en términos de las nuevas variables $u,\ v,\ w$ como $\vec{Y}^t D\vec{Y}=36$, o sea,

$$u^2+v^2+7w^2=36. \tag{5.61}$$

En $\mathbb{R}^3$ la superficie definida por (5.61) se llama **elipsoide**.

Ejemplo 5.18. Considere la ecuación cuadrática

$$7x^2+7y^2+10z^2-2xy-4xz+4yz-12x+12y+60z-24=0. \tag{5.62}$$

Determine la superficie cuádrica obtenida al eliminar los términos de productos cruzados.

Solución

La forma cuadrática asociada $7x^2+7y^2+10z^2-2xy-4xz+4yz$, se puede escribir como

$$\vec{X}^t A\vec{X}=\begin{bmatrix}x&y&z\end{bmatrix}\begin{bmatrix}7&-1&-2\\-1&7&2\\-2&2&10\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} \tag{5.63}$$

En este caso, los valores propios de A son $\lambda_1=12$ y $\lambda_2=6$ (de multiplicidad algebraica 2). El vector propio correspondiente a $\lambda_1=12$ es $\vec{v}_1=\begin{bmatrix}-1\\1\\2\end{bmatrix}$ y los correspondientes a $\lambda_2=6$ son $\vec{v}_2=\begin{bmatrix}1\\1\\0\end{bmatrix}$ y $\vec{v}_3=\begin{bmatrix}2\\0\\1\end{bmatrix}$.

Para encontrar Q, como $\|\vec{v}_1\| = \sqrt{6}$ se hace $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}$. Después, se aplica el proceso de ortonormalización de Gram-Schmidt a $\{\vec{v}_2, \vec{v}_3\}$, para obtener una base ortonormal del espacio propio $\mathcal{B}_2$. Puesto que $\|\vec{v}_2\| = \sqrt{2}$, se tiene $\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$. Por último,

$$\begin{aligned} \vec{v}_3' = & \vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \\ = & \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Entonces $\|\vec{v}_3'\| = \sqrt{3}$ y $\vec{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$. Se puede verificar que la nueva base de $\mathcal{B}_2$ es ortonormal observando que $\vec{u}_2 \cdot \vec{u}_3 = 0$. También se puede verificar que la base obtenida para $\mathbb{R}^3$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_3 = 0$ y $\vec{u}_2 \cdot \vec{u}_3 = 0$. Por lo tanto,

$$Q = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}.$$

Como $\det(Q) = -1$, se multiplica la segunda columna por -1 y se tiene

$$Q = \begin{bmatrix} -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix} \qquad \text{y} \qquad Q^tAQ = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Por lo tanto, (5.63) se puede escribir en términos de las nuevas variables u, v, w como $\vec{Y}^t D\vec{Y}$, o sea,

$$12u^2 + 6v^2 + 6w^2 \tag{5.64}$$

donde,

$$\vec{Y} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = Q^t\vec{X} = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{5.65}$$

Si se expresa toda la ecuación dada en (5.62), en forma matricial queda

$$\vec{X}^t A\vec{X} + K\vec{X} - 24 = 0,$$

donde $K = [-12 \quad 12 \quad 60]$. Al hacer $\vec{X} = Q\vec{Y}$ el cambio de variable propuesto en (5.65), se tiene

$$\begin{aligned} K\vec{X} =& KQ\vec{Y} = [-12 \;\; 12 \;\; 60] \begin{bmatrix} -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\ =& \left[24\sqrt{6} \quad 0 \quad 12\sqrt{3}\right] \begin{bmatrix} u \\ v \\ w \end{bmatrix}. \end{aligned}$$

Luego, la ecuación cuadrática dada en (5.62), se puede escribir como

$$12u^2 + 6v^2 + 6w^2 + 24\sqrt{6}u + 12\sqrt{3}w = 24.$$

Al dividir por 6 y completar los cuadrados se obtiene

$$2\left[u^2 + 2\sqrt{6}u + 6\right] + v^2 + \left[w^2 + 2\sqrt{3}w + 3\right] = 4 + 12 + 3$$

o bien,

$$2\left[u + \sqrt{6}\right]^2 + v^2 + \left[w + \sqrt{3}\right]^2 = 19$$

Esta ecuación en $\mathbb{R}^3$, representa un **elipsoide** con centro en $\left(-\sqrt{6}, 0, -\sqrt{3}\right)$.

5.5.3. Rotación de ejes en $\mathbb{R}^3$

La interpretación geométrica dada a las transformaciones ortogonales en $\mathbb{R}^2$, puede generalizarse para $\mathbb{R}^3$. Esto es, si $\vec{X} = A\vec{Y}$ representa una transformación ortogonal de coordenadas en $\mathbb{R}^3$, entonces las columnas de A están dadas por los cosenos directores de los nuevos ejes de referencia con respecto al viejo sistema de referencia.

Definición 5.13. Rotación de ejes en $\mathbb{R}^3$

Una rotación de ejes en $\mathbb{R}^3$ es una transformación ortogonal propia que permite pasar a una nueva base a partir de un movimiento rígido y continuo de los vectores base del sistema primitivo, conservando el origen fijo y preservando la ortogonalidad.

Teorema 5.13. Rotación de ejes en $\mathbb{R}^3$

Sean $\mathcal{B}=\{\vec{e}_1,\vec{e}_2,\vec{e}_3\}$ una base del sistema coordenado **X** y $\mathcal{B}^{'}=\left\{\vec{e}^{'}_1,\vec{e}^{'}_2,\vec{e}^{'}_3\right\}$ la base correspondiente al sistema **Y**. Entonces si

$$\begin{aligned}\vec{e}^{'}_1 =&a_1\vec{e}_1+a_2\vec{e}_2+a_3\vec{e}_3\\ \vec{e}^{'}_2 =&b_1\vec{e}_1+b_2\vec{e}_2+b_3\vec{e}_3\\ \vec{e}^{'}_2 =&c_1\vec{e}_1+c_2\vec{e}_2+c_3\vec{e}_3\end{aligned}$$

las coordenadas (x_1,x_2,x_3) de un punto cualquiera S en el sistema **X** y las coordenadas (y_1,y_2,y_3) del mismo punto en el sistema **Y**, están relacionadas como sigue

$$\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=P\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$$

donde, P es la matriz de transición (o matriz de cambio de base) de $\mathcal{B}$ a $\mathcal{B}^{'}$.

Demostración

Queda como ejercicio para el lector.

Cambio de la dirección de los ejes conservando el mismo origen

Consideremos que el sistema coordenado **X** (en $\mathbb{R}^3$) es rectangular y tiene unidades iguales sobre sus tres ejes, lo que significa que $\vec{e}_1$, $\vec{e}_2$ y $\vec{e}_3$, que son los vectores base estándar (unitarios), son perpendiculares entre sí.

Sea $\vec{e}^{'}_1$, $\vec{e}^{'}_2$ y $\vec{e}^{'}_3$ la base ortonormal del sistema **Y**. Denotemos los productos puntos entre las dos bases por

$$a_{ij}=\vec{e}_i\cdot\vec{e}^{'}_j.$$

Por ejemplo, $a_{1j}=\vec{e}_1\cdot\vec{e}^{'}_j$, $a_{2j}=\vec{e}_2\cdot\vec{e}^{'}_j$ y $a_{3j}=\vec{e}_3\cdot\vec{e}^{'}_j$ son las tres componentes de $\vec{e}^{'}_j$ con respecto a la base anterior $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ y podemos poner

$$\vec{e}^{'}_j =a_{1j}\vec{e}_1+a_{2j}\vec{e}_2+a_{3j}\vec{e}_3, \qquad\qquad j=1,2,3.$$

Como $\vec{e}_j^{\,\prime}$ es también vector unitario, se tiene que

$$a_{ij} = \cos\sphericalangle(\vec{e}_i, \vec{e}_j^{\,\prime}) \tag{5.66}$$

y

$$a_{1j}^2 + a_{2j}^2 + a_{3j}^2 = 1, \qquad j = 1, 2, 3.$$

Además, como $\vec{e}_1^{\,\prime}$, $\vec{e}_2^{\,\prime}$, $\vec{e}_3^{\,\prime}$ son ortogonales por pares, es decir $\vec{e}_i^{\,\prime} \cdot \vec{e}_j^{\,\prime} = 0,\ i \neq j$, se llega a,

$$a_{1j}a_{1k} + a_{2j}a_{2k} + a_{3j}a_{3k} = 0, \qquad 1 \leq j < k \leq 3.$$

Fórmulas de Euler

Euler estableció unas fórmulas que permiten fijar la posición del segundo sistema coordenado con relación al primero, empleando únicamente tres constantes.

Sean X_1, X_2, X_3 los ejes del sistema coordenado $\mathbf{X}$ y representemos por Y_1, Y_2, Y_3 los ejes del sistema de referencia móvil. Se desea definir una transformación ortogonal $\mathbf{Y} = A\mathbf{X}$, que exprese las coordenadas (y_1, y_2, y_3) de un punto arbitrario en términos de sus coordenadas iniciales (x_1, x_2, x_3). Para formular analíticamente esta transformación, se pasará del sistema "viejo" $\mathbf{X}$ al "nuevo" $\mathbf{Y}$ mediante tres cambios sucesivos. En cada cambio se asumirá que se tiene en común con el sistema precedente un eje y el plano opuesto. De esta manera no se exigirán más que las fórmulas correspondientes al cambio de ejes situados en el mismo plano.

1°. Se obtienen unos nuevos ejes W_1, W_2, $W_3 \equiv X_3$ haciendo girar el plano que contiene los vectores $\vec{e}_1$ y $\vec{e}_2$ del sistema $\mathbf{X}$ un ángulo φ alrededor del eje X_3, en sentido contrario al de las manecillas del reloj, conservando la ortogonalidad. Por el Teorema 5.13, se tiene que las coordenadas de un punto en ambos sistemas están relacionadas por

$$\begin{aligned} x_1 &= w_1 \cos\varphi - w_2 \operatorname{sen}\varphi, \\ x_2 &= w_1 \operatorname{sen}\varphi + w_2 \cos\varphi, \\ x_3 &= w_3. \end{aligned} \tag{5.67}$$

Al expresar dicha rotación en forma matricial, se obtiene

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi & 0 \\ -\operatorname{sen}\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A_\varphi \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

2°. Se generan otros nuevos ejes $Z_1 \equiv W_1$, Z_2, Z_3 haciendo girar el plano determinado por los ejes W_2 y W_3 un ángulo θ alrededor del eje W_1, conservando la ortogonalidad; lo que hará tomar al eje X_3 la posición Y_3 y al W_2 la Z_2. Por lo tanto, se deduce por el Teorema 5.13, que las coordenadas de transformación de un punto, serán

$$\begin{aligned} w_1 &= z_1, \\ w_2 &= z_2 \cos\theta - z_3 \operatorname{sen}\theta, \\ w_3 &= z_2 \operatorname{sen}\theta + z_3 \cos\theta. \end{aligned} \tag{5.68}$$

Si se expresa (5.68) en forma matricial, se obtiene

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \operatorname{sen}\theta \\ 0 & -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = A_\theta \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

3°. Finalmente, se gira alrededor del eje $Z_3 \equiv Y_3$ el plano que contiene a los dos ejes Z_2 y Z_3 hasta que forman un ángulo ψ, así quedará W_1 en Y_1 y Z_2 en Y_2 y por el Teorema 5.13, las coordenadas de un punto en ambos sistemas están relacionadas por

$$\begin{aligned} z_1 &= y_1 \cos\psi - y_2 \operatorname{sen}\psi, \\ z_2 &= y_1 \operatorname{sen}\psi + y_2 \cos\psi, \\ z_3 &= y_3. \end{aligned} \tag{5.69}$$

Al expresar (5.69) en forma matricial, se obtiene

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \cos\psi & \operatorname{sen}\psi & 0 \\ -\operatorname{sen}\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = A_\psi \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

La eliminación de los sistemas coordenados **W** y **Z** en las ecuaciones (5.67), (5.68) y (5.69) dará la transformación del producto $\mathbf{Y} = A\mathbf{X}$, mediante la siguiente matriz de rotación en $\mathbb{R}^3$

$$A_\psi A_\theta A_\varphi = \begin{bmatrix} \cos\psi & \operatorname{sen}\psi & 0 \\ -\operatorname{sen}\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \operatorname{sen}\theta \\ 0 & -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi & 0 \\ -\operatorname{sen}\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} \cos\psi\cos\varphi - \operatorname{sen}\psi\operatorname{sen}\varphi\cos\theta & \operatorname{sen}\psi\cos\varphi\cos\theta + \operatorname{sen}\varphi\cos\psi & \operatorname{sen}\psi\operatorname{sen}\theta \\ -\operatorname{sen}\psi\cos\varphi - \operatorname{sen}\varphi\cos\psi\cos\theta & \cos\psi\cos\varphi\cos\theta - \operatorname{sen}\psi\operatorname{sen}\varphi & \cos\psi\operatorname{sen}\theta \\ \operatorname{sen}\varphi\operatorname{sen}\theta & -\cos\varphi\operatorname{sen}\theta & \cos\theta \end{bmatrix},$$

los ángulos $\varphi,\ \theta,\ \psi$, deben escogerse en el intervalo $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ y se conocen como *ángulos eulerianos*. Nótese que el determinante de esta matriz es:

$$\det(A_\psi A_\theta A_\varphi) = \det(A_\psi)\det(A_\theta)\det(A_\varphi) = 1.$$

Luego, esta matriz es ortogonal.

La comparación de las componentes de la matriz $A_\psi A_\theta A_\varphi$ con las expresiones dadas en (5.66) permiten obtener esos nueve cosenos en función de las constantes $\varphi,\ \theta,\ \psi$.

Observaciones

Para determinar los ángulos eulerianos $\varphi,\ \theta,\ \psi$ se comparan las componentes de $Q^t = [c_{ij}]$ con las de $A_\psi A_\theta A_\varphi$ (donde las columnas de Q son los vectores propios normalizados de la matriz simétrica asociada a la forma cuadrática) y se tiene en cuenta que

1. Si $c_{33} \neq 1$, entonces los ángulos eulerianos se pueden determinar de la siguiente manera
$$\cos\theta = c_{33}, \qquad \tan\psi = \frac{c_{13}}{c_{23}} \qquad \text{y} \qquad \tan\varphi = -\frac{c_{31}}{c_{32}}.$$

2. Cuando $c_{33} = 1$, entonces $\theta = 0$ y la matriz de rotación $A_\psi A_\theta A_\varphi$, tiene la forma
$$A_\psi A_\theta A_\varphi = \begin{bmatrix} \cos(\psi+\varphi) & \operatorname{sen}(\psi+\varphi) & 0 \\ -\operatorname{sen}(\psi+\varphi) & \cos(\psi+\varphi) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$
en este caso,
$$\tan(\psi+\varphi) = -\frac{c_{21}}{c_{11}}.$$
Aquí, los ángulos ψ y φ se escogen arbitrariamente.

3. Si la suma de los cuadrados de los elementos de la diagonal principal de la matriz de rotación $A_\psi A_\theta A_\varphi$ es igual a 1. Entonces el ángulo θ, satisface que
$$\cos\theta = \frac{\tan^2\varphi + \tan^2\psi}{\sqrt{(\sec^2\varphi + \sec^2\psi)(\sec^2\varphi\tan^2\psi + \sec^2\psi\tan^2\varphi) \mp 2\tan\varphi\tan\psi}}.$$
En este caso, dado que la tangente de $\frac{\pi}{2}$ no está definida, cuando uno de los ángulos φ ó ψ sea igual a $\pm\frac{\pi}{2}$, se tiene que
$$\cos\theta = \begin{cases} \dfrac{1}{\sqrt{\tan^2\psi + \sec^2\psi}} & \text{si } \varphi \to \left(\pm\frac{\pi}{2}\right) \\ \dfrac{1}{\sqrt{\tan^2\varphi + \sec^2\varphi}} & \text{si } \psi \to \left(\pm\frac{\pi}{2}\right) \end{cases}$$

y si uno de los ángulos φ ó ψ es igual a cero, entonces

$$\cos\theta = \begin{cases} \dfrac{|\tan\psi|}{\sqrt{1+\sec^2\psi}} & \text{si } \varphi = 0 \\ \dfrac{|\tan\varphi|}{\sqrt{1+\sec^2\varphi}} & \text{si } \psi = 0. \end{cases}$$

Ejemplo 5.19. Determine los ángulos eulerianos de la ecuación cuadrática dada en el Ejemplo 5.17.

Solución

La matriz ortogonal asociada a la forma cuadrática era

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}.$$

Al comparar las componentes de Q^t con las de $A_\psi A_\theta A_\varphi$ se tiene que

$$\cos\theta = \frac{1}{\sqrt{6}}, \qquad \tan\psi = \sqrt{\frac{3}{2}} \qquad \text{y} \qquad \tan\varphi = -\frac{1}{2}. \tag{5.70}$$

Se debe escoger una solución de (5.70) para la cual los ángulos eulerianos se encuentren dentro del intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. En este caso, los ángulos θ, ψ y φ valen, aproximadamente $65^\circ 54' 18{,}57''$; $50^\circ 46' 6{,}53''$ y $-26^\circ 33' 54{,}18''$.

Ejemplo 5.20. Determine los ángulos eulerianos de la ecuación cuadrática dada en el Ejemplo 5.18.

Solución

La matriz ortogonal asociada a la forma cuadrática era

$$Q = \begin{bmatrix} -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}.$$

Si se comparan las componentes de Q^t con las de $A_\psi A_\theta A_\varphi$ se tiene que

$$\cos\theta = \frac{1}{\sqrt{3}}, \qquad \tan\psi = \textit{indefinida} \qquad \text{y} \qquad \tan\varphi = 1. \tag{5.71}$$

Se debe escoger una solución de (5.71) para la cual los ángulos eulerianos se encuentren dentro del intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, en este caso, los ángulos θ, ψ y φ valen, aproximadamente $54^\circ 44' 8{,}2''$; 90° y 45°

5.5.4. Clasificación de las superficies cuádricas

Para la ecuación general de segundo grado

$$ax^2+by^2+cz^2+dxy+exz+fyz+gx+hy+iz+j=0 \tag{5.72}$$

se definen las siguientes cantidades

$$A=\begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{e}{2} & \frac{f}{2} & c \end{bmatrix}, \qquad \begin{aligned} \omega &= tr(A), \\ \mu &= \sum_{i=1}^{3} M_{ii}(A), \\ \nu &= \det A, \end{aligned} \qquad \rho=\det\begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} & \frac{g}{2} \\ \frac{d}{2} & b & \frac{f}{2} & \frac{h}{2} \\ \frac{e}{2} & \frac{f}{2} & c & \frac{i}{2} \\ \frac{g}{2} & \frac{h}{2} & \frac{i}{2} & j \end{bmatrix}.$$

Además, si $\lambda_1,\lambda_2,\lambda_3$ son los valores propios de A, entonces la ecuación cuádrica (5.72) representa los lugares geométricos (L.G.) siguientes

$$\text{L.G.}\begin{cases} \nu\neq 0 & \begin{cases} \begin{matrix}\frac{\rho}{\nu}\neq 0 \\ \left(\frac{\rho}{\nu}<0\right)\end{matrix} & \begin{cases} \lambda_1>0,\lambda_2>0,\lambda_3>0 & \textbf{Un elipsoide} \\ \lambda_1>0,\lambda_2>0,\lambda_3<0 & \textbf{Un hiperboloide de una hoja} \\ \lambda_1>0,\lambda_2<0,\lambda_3<0 & \textbf{Un hiperboloide de dos hojas} \\ \lambda_1<0,\lambda_2<0,\lambda_3<0 & \textbf{Conjunto vacío} \end{cases} \\ \frac{\rho}{\nu}=0 & \begin{cases} \lambda_1,\lambda_2,\lambda_3 \text{ tienen el mismo signo.} & \textbf{Un punto} \\ \lambda_1,\lambda_2,\lambda_3 \text{ tienen signos distintos.} & \textbf{Un cono} \end{cases} \end{cases} \\ \begin{matrix}\nu=0 \\ (\lambda_3=0)\end{matrix} & \begin{cases} I=0 & \begin{cases} \frac{\rho}{\nu}\neq 0 & \begin{cases} \lambda_1>0,\lambda_2>0 & \textbf{Un cilindro elíptico} \\ \lambda_1>0,\lambda_2<0 & \textbf{Un cilindro hiperbólico} \\ \lambda_1<0,\lambda_2<0 & \textbf{Conjunto vacío} \end{cases} \\ \frac{\rho}{\nu}=0 & \begin{cases} \lambda_1\lambda_2>0 & \textbf{Una recta} \\ \lambda_1\lambda_2<0 & \textbf{Dos planos que se cortan} \end{cases} \end{cases} \\ I'\neq 0 & \begin{cases} \lambda_1\lambda_2>0 & \textbf{Un paraboloide elíptico} \\ \lambda_1\lambda_2<0 & \textbf{Un paraboloide hiperbólico} \end{cases} \end{cases} \\ \begin{matrix}\nu=0 \\ (\lambda_1\neq 0)\end{matrix} & \begin{cases} I'=0 \text{ y } H'=0 & \begin{cases} \frac{\rho}{\nu}\lambda_1>0 & \textbf{Conjunto vacío} \\ \frac{\rho}{\nu}\lambda_1<0 & \textbf{Dos planos paralelos} \end{cases} \\ \frac{\rho}{\nu}\neq 0 \text{ y/o } H'\neq 0 & \textbf{Un cilindro parabólico.} \end{cases} \end{cases}$$

También, puede ser clasificada teniendo en cuenta la inercia de la matriz simétrica A como sigue:

Identificación de las superficies cuádricas

Inercia	**Nombre de la superficie**
$In(A)=(3,0,0)$	Elipsoide
$In(A)=(2,1,0)$	Hiperboloide de una hoja
$In(A)=(1,2,0)$	Hiperboloide de dos hojas
$In(A)=(2,0,1)$	Paraboloide elíptico
$In(A)=(1,1,1)$	Paraboloide hiperbólico
$In(A)=(1,0,2)$	Cilindro parabólico

Ejercicios 5.4.

1. Determine la sección cónica y el ángulo de rotación para

   a. $4x^2+4xy+y^2=9$. *b*. $36x^2+9y^2+4z^2-36=0$.

   c. $x^2+8xy+16y^2-4x+16y=-7$. *d*. $x^2+2xy+y^2+2x+2y=-1$.

   e. $4x^2-20xy+25y^2+4x-10y=-1$. *f*. $4x^2-4xy+y^2-6x+3y=4$.

   g. $9x^2-24xy+16y^2-20x+110y=50$.

   h. $6x^2+3y^2-2z^2+12x-18y-8z=-7$.

   i. $7x^2+7y^2+10z^2-2xy-4xz+4yz-12x+12y+60z=24$.

   j. $2x^2+2y^2+5z^2-4xy+2xz-2yz+10x-26y-2z=0$.

2. Sea A la representación matricial de la ecuación cuadrática (5.50) con $f\neq 0$. Sean λ_1 y λ_2 los valores propios de A. Demuestre que la curva que describe (5.50) es

   a. Una hipérbola si $\lambda_1\lambda_2<0$. *b*. Un par de rectas si $\lambda_1\lambda_2=0$.

   c. Un circulo, elipse o sección cónica degenerada si $\lambda_1\lambda_2>0$.

5.6. Clasificación de las formas cuadráticas

En esta sección se clasifican las formas cuadráticas según sus valores posibles. Una forma cuadrática $F:\mathbb{R}^n\to\mathbb{R}$ es una función de valor real con dominio en $\mathbb{R}^n$. Luego, se pueden distinguir varias clases importantes de formas cuadráticas de acuerdo a los valores que estas asumen para diferentes $\vec{X}$, dichos números reales pueden ser mayores que, menores que o iguales a 0. Obviamente si el vector $\vec{X}=\vec{0}$ el valor siempre será 0, por lo tanto no se tendra en cuenta este vector. Por otra parte, si la matriz A es nula $F(\vec{X})$ siempre dará el valor cero.

Definición 5.14.

Dada $F(\vec{X}) = \vec{X}^t A\vec{X}$ con $A \neq O$ simétrica, se dice que es

1. **Definida positiva** si $F(\vec{X}) > 0$ para todo $\vec{X}$ distinto de cero en $\mathbb{R}^n$,

2. **Semidefinida positiva** si $F(\vec{X}) \geq 0$ para todo $\vec{X}$ en $\mathbb{R}^n$,

3. **Definida negativa** si $F(\vec{X}) < 0$ para todo $\vec{X}$ distinto de cero en $\mathbb{R}^n$,

4. **Semidefinida negativa** si $F(\vec{X}) \leq 0$ para todo $\vec{X}$ en $\mathbb{R}^n$,

5. **Indefinida** si $F(\vec{X})$ asume ambos valores positivos y negativos.

La matriz simétrica asociada A, se denomina **definida positiva, semidefinida positiva, definida negativa, semidefinida negativa o indefinida** según sea la forma cuadrática $F(\vec{X})$ que define.

Teorema 5.14. Sea A una matriz simétrica de tamaño $n \times n$. Entonces la forma cuadrática $F(\vec{X}) = \vec{X}^t A\vec{X}$ es

1. Definida positiva si y sólo si todos los valores propios de A son positivos.

2. Definida negativa si y sólo si todos los valores propios son negativos.

3. Indefinida si y sólo si A tiene valores propios positivos y negativos.

Demostración

1. Sea A definida positiva, sea λ un valor propio de A; sea $\vec{X}$ un vector propio de A asociado a λ. Calculemos
$$\vec{X}^t A\vec{X} = \vec{X}^t \lambda\vec{X} = \lambda\vec{X}^t\vec{X}.$$
Por consiguiente, $\lambda = \frac{\vec{X}^t A\vec{X}}{\vec{X}^t\vec{X}}$ es positivo puesto que esta es una razón de dos números positivos.

Las otras quedan como ejercicio para el lector.

Ejemplo 5.21. Sea $F(\vec{X}) = 3x_1^2 + 3x_2^2 + 2x_1x_2$. Determine que clase de forma cuadrática es.

Solución

En el Ejemplo 5.11 se obtuvo que los valores propios de la matriz asociada a la forma $F(\vec{X})$ eran 2 y 4. Por lo tanto, dicha forma cuadrática es definida positiva.

Ejemplo 5.22. Dada $F(\vec{X}) = 2x_1^2 + 4x_1x_2 - x_2^2$, determine la clase de forma cuadrática que es.

Solución

En el Ejemplo 5.14, se obtuvo que los valores propios de la matriz asociada a la forma $F(\vec{X})$ eran -2 y 3. Por lo tanto, dicha forma cuadrática es indefinida.

Teorema 5.15. Sea $F : V \to \mathbb{R}$ una forma cuadrática en el espacio vectorial $\mathbb{V}$ de dimensión finita n. La forma $F(\vec{X})$ se dice que es definida positiva si y sólo si $A = [a_{ij}]$ la matriz asociada a la forma $F(\vec{X})$ tiene la propiedad de que todos los determinantes de sus submatrices angulares son positivos.

Demostración

La demostración se hará por inducción sobre n. Para $n = 1$ la forma cuadrática $F(\vec{X})$ esta dada por

$$F(\vec{X}) = [x][a][x] = ax^2,$$

en donde $A = [a]$. El teorema afirma en este caso que la forma $F(\vec{X})$ es definida positiva si y sólo si $a > 0$, lo cual es claro.

Por el Teorema 5.8 la matriz $A = [a]$ se transforma en la matriz Q^tAQ, en donde Q es una matriz cuadrada de tamaño 1×1 no singular, esto es, $Q = [q], q \neq 0$. De esta manera,

$$Q^tAQ = [q][a][q] = aq^2,$$

en tal caso siendo $a > 0$ y $q \neq 0$, se tiene que $aq^2 > 0$, así que la afirmación del teorema no depende de la base considerada en $\mathbb{V}$, para el caso $n = 1$.

Supongamos entonces que el teorema es válido para $n = k - 1$ y veamos si se cumple para $n = k$.

Se demostrará primero que si la forma $F(\vec{X})$ es definida positiva, entonces los determinantes $\Delta_1, \Delta_2, \ldots \Delta_k$ son positivos. Dado que la forma $F(\vec{X})$ se puede escribir como

$$F(\vec{X}) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} x_i x_j = \sum_{i=1}^{k} a_{ii} x_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_{ij} x_i x_j,$$

esta se puede reescribir de la siguiente manera

$$F(\vec{X}) = \sum_{i=1}^{k-1} a_{ii} x_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_{ij} x_i x_j + a_{kk} x_k^2. \tag{5.73}$$

Sea $\mathbb{W}$ un subespacio de $\mathbb{V}$ de dimensión $k-1$ y considérese $F^* : W \to \mathbb{R}$ una forma cuadrática en $\mathbb{W}$ tal que

$$F^*(\vec{X}) = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} a_{ij} x_i x_j.$$

En efecto, la forma $F^*(\vec{X})$ es definida positiva. Supóngase lo contrario, entonces existe un vector $\vec{X}^* \in W$, digamos $\vec{X}^* = (x_1, x_2, \ldots, x_{k-1})$ tal que $F^*(\vec{X}^*) \leq 0$. Formemos el vector $\vec{X} \in V$ como $\vec{X} = (x_1, x_2, \ldots, x_{k-1}, 0)$ y evaluemos $F(\vec{X})$, según la expresión (5.73) se obtiene que $F(\vec{X}) \leq 0$, lo cual contradice la hipótesis de que la forma $F(\vec{X})$ es definida positiva.

Por lo tanto, según la hipótesis de inducción los determinantes de las submatrices angulares de la matriz de la forma $F^*(\vec{X})$ son positivos. Éstos son: $\Delta_1, \Delta_2, \ldots \Delta_{k-1}$. Falta probar que δ_k es también positivo.

Pero por el Teorema 5.8 la matriz A asociada a la forma $F(\vec{X})$ mediante el empleo de una matriz no singular Q se transforma en la matriz

$$D = Q^t A Q,$$

al tomar el determinante de las matrices en esta última expresión se obtiene

$$\det(D) = \det(Q^t A Q) = \det(Q^t) \det(A) \det(Q) = \det(A) [\det(Q)]^2.$$

Pero como $\det(Q) \neq 0$ y $\det(D) = \prod_{i=1}^{k} \lambda_i$, se tiene que

$$\det(A) = \frac{1}{[\det(Q)]^2} \det(D) = \frac{1}{[\det(Q)]^2} \prod_{i=1}^{k} \lambda_i > 0^3.$$

[3]Véase Teorema 5.14

Se ha probado así que $\Delta_k = \det(A) > 0$, como se quería.

Se deja como ejercicio para el lector demostrar que si los determinantes $\Delta_1, \Delta_2, \ldots \Delta_k$ de alguna matriz (arbitraria pero fija) asociada a la forma cuadrática $F(\vec{X})$ son positivos, entonces la forma $F(\vec{X})$ es definida positiva.

Ejemplo 5.23. Determine si la siguiente forma cuadrática es definida positiva

$$F(x,y,z) = 2x^2 + 5y^2 + 2z^2 + 4xy + 2xz + 4yz.$$

Solución

Según el Ejemplo 5.17, la matriz asociada a la forma cuadrática es

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

Luego los determinantes de las submatrices angulares son

$$\Delta_1 = \det(A_{[1]}) = \begin{vmatrix} 2 \end{vmatrix} = 2 > 0, \qquad \Delta_2 = \det(A_{[2]}) = \begin{vmatrix} 2 & 2 \\ 2 & 5 \end{vmatrix} = 6 > 0 \qquad \text{y}$$

$$\Delta_3 = \det(A_{[3]}) = \begin{vmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{vmatrix} = 7 > 0$$

Como los tres determinantes $\Delta_1, \Delta_2, \Delta_3$ son positivos, se concluye, por el Teorema 5.15, que la forma cuadrática $F(\vec{X})$ es definida positiva.

Corolario 5.15.1. La forma $F : V \to \mathbb{R}$ es definida negativa si y sólo si $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots$ Es decir, si los determinantes $\Delta_1, \Delta_2, \ldots \Delta_n$ alternan sus signos, comenzando con $\Delta_1 < 0$.

Demostración

Es claro que la forma $F : V \to \mathbb{R}$ dada por

$$F(\vec{X}) = \vec{X}^t A \vec{X}$$

es definida negativa si y sólo si la forma $-F : V \to \mathbb{R}$

$$-F(\vec{X}) = \vec{X}^t (-A) \vec{X}$$

es definida positiva. Según el Teorema 5.15 se debe tener entonces, que los determinantes

$$\Delta_1 = \left|-a_{11}\right|,\ \Delta_2 = \begin{vmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{vmatrix},\ \Delta_3 = \begin{vmatrix} -a_{11} & -a_{12} & -a_{13} \\ -a_{21} & -a_{22} & -a_{23} \\ -a_{31} & -a_{32} & -a_{33} \end{vmatrix},\ \ldots,$$

$\Delta_n = (-1)^n \det(A)$ deben ser positivos. Es decir, que

$$\begin{aligned}
\Delta_1 =& \det[-a_{11}] = -\det[a_{11}] > 0 \\
\Delta_2 =& \begin{vmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{vmatrix} = (-1)^2 \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0 \\
\Delta_3 =& \begin{vmatrix} -a_{11} & -a_{12} & -a_{13} \\ -a_{21} & -a_{22} & -a_{23} \\ -a_{31} & -a_{32} & -a_{33} \end{vmatrix} = (-1)^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0
\end{aligned}$$

y así sucesivamente, lo que prueba el corolario.

Teorema 5.16. La suma de dos cualesquiera matrices definidas positivas del mismo tamaño es definida positiva. Más generalmente, cualquier combinación lineal no negativa de matrices semidefinidas positivas es semidefinida positiva.

Demostración

Sean A y B matrices semidefinidas positivas, sean $\alpha,\ \beta \geq 0$. Si se denota $C = \alpha A + \beta B$ y se calcula

$$\vec{X}^t C \vec{X} = \vec{X}^t (\alpha A + \beta B)\vec{X} = \alpha(\vec{X}^t A \vec{X}) + \beta(\vec{X}^t B \vec{X}) \geq 0, \qquad \forall \vec{X} \in \mathbb{C}^n.$$

El caso de más de dos sumandos es tratado en el mismo sentido. Si los coeficientes α y β son positivos y si A y B son matrices definidas positivas y además $\vec{X} \neq \vec{0}$, entonces cada término en la suma es positivo. Así una combinación lineal de matrices definidas positivas es definida positiva.

Teorema 5.17. Si A es una matriz simétrica definida positiva de tamaño $n \times n$, entonces cualquier submatriz principal de A es definida positiva.

Demostración

Sea K un subconjunto propio de $\{1, 2, \ldots, n\}$ y denotemos por $A_{(K)}$ la matriz resultante de eliminar las filas y columnas complementarias a las indicadas por K

de la matriz A. Entonces $A_{(K)}$ es una submatriz principal de A. Nótese que todas las submatrices se pueden obtener de esta manera; el número $\det[A_{(K)}]$ es un menor de A. Sea $\vec{X} \in \mathbb{C}^n$ un vector no nulo con entradas arbitrarias en las componentes indicadas por K y cero en las otras entradas. Denotando por $\vec{X}_{(K)}$ el vector que se obtiene de eliminar las componentes nulas de $\vec{X}$ y obsérvese que

$$\vec{X}^H_{(K)} A_{(K)} \vec{X}_{(K)} = \vec{X}^H A \vec{X} > 0.$$

Puesto que $\vec{X}_{(K)} \neq \vec{0}$ es arbitrario, esto significa que $A_{(K)}$ es definida positiva.

Teorema 5.18. La traza, el determinante y todos los menores principales de una matriz simétrica definida positiva son positivos.

Demostración

Sea $A \in M_{n\times n}$ una matriz simétrica definida positiva; luego, por el Teorema 2.17 se sabe que la traza y el determinante son respectivamente la suma y el producto de los valores propios, los cuales por el Teorema 5.14 son todos positivos. La otra parte del teorema se obtiene del Teorema 5.17.

Ejemplo 5.24. Determine si la siguiente forma cuadrática es definida positiva verificando si la matriz asociada a la forma cumple las condiciones del teorema anterior

$$F(x,y,z) = 2x^2 + 5y^2 + 2z^2 + 4xy + 2xz + 4yz.$$

Solución

Según el Ejemplo 5.17, la matriz asociada a la forma cuadrática es

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

La $tr(A) = 9 > 0$, $\det(A) = 7 > 0$ y los determinantes de algunos menores son

$$\det(M_{33}) = \begin{vmatrix} 2 & 2 \\ 2 & 5 \end{vmatrix} = 6, \qquad \det(M_{22}) = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3.$$

Por lo tanto, la forma cuadrática dada si es definida positiva.

Teorema 5.19. Sea $A \in M_{m\times m}$ definida positiva y $P \in M_{m\times n}$, entonces P^tAP es semidefinida positiva. Además, $\rho\left(P^tAP\right) = \rho(P)$, así que P^tAP es definida positiva si y sólo si P tiene rango n.

Demostración

Es claro que P^tAP es simétrica. Para cualquier $\vec{X} \in \mathbb{C}^n$ se tiene que

$$\vec{X}^tP^tAP\vec{X} = \vec{Y}^tA\vec{Y} \geq 0$$

donde $\vec{Y} = P\vec{X}$ y la desigualdad se sigue porque A es definida positiva. Así que P^tAP es semidefinida positiva. Obsérvese que $\vec{X}^tP^tAP\vec{X} > 0$ si y sólo si $P\vec{X} \neq \vec{0}$ ya que A es definida positiva.

Supongamos que $P\vec{X} = \vec{0}$, entonces obviamente $P^tAP\vec{X} = \vec{0}$. Recíprocamente, si $P^tAP\vec{X} = \vec{0}$, entonces $\vec{X}^tP^tAP\vec{X} = 0$ y usando el hecho de que A es definida positiva se concluye que $P\vec{X} = \vec{0}$. Por lo tanto, $P^tAP\vec{X} = \vec{0}$ si y sólo si $P\vec{X} = \vec{0}$ lo cual significa que P^tAP y P tienen el mismo espacio nulo (y por lo tanto también tienen el mismo rango).

Ejercicios 5.5.

1. Muestre que las entradas de la diagonal de una matriz simétrica definida positiva son números reales positivos.

2. Muestre que los valores propios, traza, determinante y menores principales de una matriz semidefinida positiva son todos no negativos.

3. Muestre que si $A \in M_{2\times 2}$ es definida positiva, entonces $a_{11}a_{22} > |a_{12}|^2$.

4. Si A es de tamaño $m \times n$ de rango $n < m$, muestre que A^tA es definida positiva y que AA^t es semidefinida positiva.

5. Si A es de tamaño $m \times n$ de rango $k < \text{mín}\{m,n\}$, muestre que A^tA y AA^t son semidefinidas positivas.

Capítulo 6

Formas hermíticas

En el capítulo anterior se desarrolló la teoría para formas cuadráticas con matriz asociada simétrica real. En este capítulo se consideran formas cuadráticas pero con matriz asociada compleja. Se estudia el caso *complejo* independientemente del caso *real*, ya que si se asume $\vec{X} \in \mathbb{C}^2$ y se obtiene la expresión $\|\vec{X}\|^2 = \vec{X}^t\vec{X}$, de manera análoga al producto escalar estándar de $\mathbb{R}^2$, se llega a resultados ilógicos. Por ejemplo, para el vector *no nulo* $\vec{X} = (a, bi)$, se tiene que

$$\vec{X}^t\vec{X} = a^2 + b^2 i^2 = a^2 - b^2$$

este producto puede ser cero si $a = b$ ó $a = -b$, hecho que contradice la propiedad (v) del producto escalar estándar en $\mathbb{R}^n$ (ver Capítulo 1). Este hecho induce a la redefinición de formas cuadráticas para el caso complejo.

6.1. Forma hermítica

Definición 6.1. Forma sesquilineal

Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre el cuerpo complejo $\mathbb{C}$.

Una forma sesquilineal es una función $g : \mathbb{V} \times \mathbb{V} \to \mathbb{C}$ tal que

$$(i)\ g(\alpha\vec{u}_1 + \beta\vec{u}_2, \vec{v}) = \overline{\alpha}g(\vec{u}_1, \vec{v}) + \overline{\beta}g(\vec{u}_2, \vec{v})$$

$$(ii)\ g(\vec{u}, \alpha\vec{v}_1 + \beta\vec{v}_2) = \alpha g(\vec{u}, \vec{v}_1) + \beta g(\vec{u}, \vec{v}_2)$$

donde $\alpha, \beta \in \mathbb{C}$ y $\vec{u}_i, \vec{v}_j \in \mathbb{V}$.

Como en la definición 5.1, la condición (ii) se interpreta como que g es lineal en la segunda variable. Por otra parte, expresamos la condición (i) diciendo que g es *lineal conjugada* en la primera variable. En el resto de esta sección se omitirá el adjetivo "sesquilineal", salvo que sea importante tenerlo en cuenta.

Definición 6.2. Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{C}$ una forma sobre $\mathbb{V}$, entonces g es **hermítica**, si para todo $\vec{v},\ \vec{w} \in \mathbb{V}$, se cumple que

$$g(\vec{v}, \vec{w}) = \overline{g(\vec{w}, \vec{v})}. \tag{6.1}$$

Ejemplo 6.1. Sea $g : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{R}$ definida por

$$g\left(\vec{X}, \vec{Y}\right) = \vec{X}^H A \vec{Y}$$

donde $\vec{X}$ y $\vec{Y} \in \mathbb{C}^n$ y A es una matriz hermitiana. Verifique si la aplicación g define una forma hermítica sobre $\mathbb{C}^n$.

Solución

(i) Para todo $\vec{X}_1,\ \vec{X}_2$ y $\vec{Y} \in \mathbb{C}^n$ se tiene que

$$\begin{aligned} g\left(\vec{X}_1 + \vec{X}_2, \vec{Y}\right) = \left(\vec{X}_1 + \vec{X}_2\right)^H A\vec{Y} = \left(\vec{X}_1^H + \vec{X}_2^H\right) A\vec{Y} \\ = \vec{X}_1^H A\vec{Y} + \vec{X}_2^H A\vec{Y} = g\left(\vec{X}_1, \vec{Y}\right) + g\left(\vec{X}_2, \vec{Y}\right). \end{aligned}$$

(ii) Para todo $\beta \in \mathbb{C}$, $\vec{X}$ y $\vec{Y} \in \mathbb{C}^n$ se tiene que

$$\begin{aligned} g\left(\vec{X}, \beta\vec{Y}\right) =& \vec{X}^H A\left(\beta\vec{Y}\right) = \beta\vec{X}^H A\vec{Y} \\ =& \beta\vec{X}^H A\vec{Y} = \beta g\left(\vec{X}, \vec{Y}\right) \end{aligned}$$

Así, la aplicación g es lineal en la segunda variable. Además,

$$\overline{g\left(\vec{X}, \vec{Y}\right)} = \overline{\left(\vec{X}^H A\vec{Y}\right)} = \overline{\left(\vec{X}^H A\vec{Y}\right)}^t = \vec{Y}^H A^H \vec{X} = \vec{Y}^H A\vec{X} = g\left(\vec{Y}, \vec{X}\right)$$

Por lo tanto, g es una forma hermítica sobre $\mathbb{C}^n$.

Teorema 6.1. Sean $\mathbb{V}$ un espacio vectorial complejo y g una forma sesquilineal sobre $\mathbb{V}$ tal que $g(\vec{u},\vec{u})$ sea real para todo $\vec{u}\in\mathbb{V}$. Entonces g es hermítica.

Demostración

Sean $\vec{u},\vec{v}\in\mathbb{V}$ y g una forma sesquilineal sobre $\mathbb{V}$ tal que $g(\vec{u},\vec{u})$ sea real para todo $\vec{u}\in\mathbb{V}$. Se debe probar que $g(\vec{u},\vec{v})=\overline{g(\vec{v},\vec{u})}$. En efecto,

$$g(\vec{u}+\vec{v},\vec{u}+\vec{v})=g(\vec{u},\vec{u})+g(\vec{u},\vec{v})+g(\vec{v},\vec{u})+g(\vec{v},\vec{v}).$$

Como por hipótesis $g(\vec{u}+\vec{v},\vec{u}+\vec{v}),\ g(\vec{u},\vec{u})$ y $g(\vec{v},\vec{v})$ son reales, el número $g(\vec{u},\vec{v})+g(\vec{v},\vec{u})$ es real. De manera análoga se tiene

$$g(\vec{u}+i\vec{v},\vec{u}+i\vec{v})=g(\vec{u},\vec{u})+ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u})+g(\vec{v},\vec{v}).$$

Por el mismo razonamiento anterior, vemos que $ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u})$ es real. Al concluir que estos dos números son reales, se pueden igualar sus complejos conjugados y se obtiene

$$g(\vec{u},\vec{v})+g(\vec{v},\vec{u})=\overline{g(\vec{u},\vec{v})}+\overline{g(\vec{v},\vec{u})} \tag{6.2}$$

$$ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u})=-i\overline{g(\vec{u},\vec{v})}+i\overline{g(\vec{v},\vec{u})}. \tag{6.3}$$

Al multiplicar (6.3) por $(-i)$ y sumarle (6.2), se llega a

$$2g(\vec{u},\vec{v})=2\overline{g(\vec{v},\vec{u})}.$$

Teorema 6.2. Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre los números complejos. Sea g una forma hermitiana sobre $\mathbb{V}$. Entonces, existe una matriz única hermitiana A, tal que para todo $\vec{u},\vec{v}\in\mathbb{V}$,

$$g_A(\vec{u},\vec{v})=\vec{u}^H A\vec{v}. \tag{6.4}$$

Demostración

La prueba es completamente análoga a la del Teorema 5.1 y se deja como ejercicio al lector.

Teorema 6.3. Identidad de polarización

Sea g una forma hermitiana sobre un espacio vectorial complejo $\mathbb{V}$, entonces para todo $\vec{u},\vec{v}\in\mathbb{V}$ se cumple que

$$g_A(\vec{u}+\vec{v},\vec{u}+\vec{v})-g_A(\vec{u}-\vec{v},\vec{u}-\vec{v})=2\left[g_A(\vec{u},\vec{v})+g_A(\vec{v},\vec{u})\right]. \tag{6.5}$$

Demostración

La verificación de esta identidad se hace en forma trivial, solo desarrollando el miembro izquierdo que aparece en (6.5).

Teorema 6.4. Sea $\mathbb{V}$ como antes. Si g es una forma hermitiana tal que $g_A(\vec{v},\vec{v}) = 0$ para todo $\vec{v} \in \mathbb{V}$, entonces $A = O$

Demostración

Por el Teorema 6.3, para todo $\vec{u},\vec{v} \in \mathbb{V}$, se tiene que:

$$g_A(\vec{u}+\vec{v},\vec{u}+\vec{v}) - g_A(\vec{u}-\vec{v},\vec{u}-\vec{v}) = 2\left[g_A(\vec{u},\vec{v}) + g_A(\vec{v},\vec{u})\right].$$

Luego, si g es tal que $g_A(\vec{v},\vec{v}) = 0$ para todo $\vec{v} \in \mathbb{V}$, el miembro izquierdo de la identidad de polarización, es igual a 0, de donde se obtiene que

$$g_A(\vec{u},\vec{v}) + g_A(\vec{v},\vec{u}) = 0, \tag{6.6}$$

para todo $\vec{u},\vec{v} \in \mathbb{V}$. Si se reemplaza $\vec{u}$ por $i\vec{u}$, entonces se tiene que

$$g_A(i\vec{u},\vec{v}) + g_A(\vec{v},i\vec{u}) = -ig_A(\vec{u},\vec{v}) + ig_A(\vec{v},\vec{u}) = 0.$$

Así,

$$-g_A(\vec{u},\vec{v}) + g_A(\vec{v},\vec{u}) = 0. \tag{6.7}$$

Si se restan las relaciones (6.6) y (6.7), se obtiene

$$2g_A(\vec{u},\vec{v}) = 0.$$

Por lo tanto $g_A(\vec{u},\vec{v}) = 0$. De donde $A = O$. Como se quería demostrar.

Teorema 6.5. Una matriz compleja A de tamaño $n \times n$ representa una forma hermítica si y sólo si es una matriz hermitiana.

Demostración

Supóngase que A es hermitiana. Como para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$, la matriz $\vec{X}^H A\vec{Y}$ es una matriz de 1×1, es decir, un elemento de $\mathbb{R}$, entonces es igual a su propia traspuesta conjugada. Por lo tanto,

$$\overline{\vec{X}^H A\vec{Y}} = \overline{\left(\vec{X}^H A\vec{Y}\right)^t} = \overline{\vec{Y}^t A^t \overline{\vec{X}}} = \vec{Y}^H A^H \vec{X} = \vec{Y}^H A\vec{X}$$

así que A representa una forma hermitiana.

Recíprocamente, supóngase que A representa una forma hermítica; es decir,

$$g_A(\vec{X},\vec{Y}) = \overline{g_A(\vec{Y},\vec{X})}, \tag{6.8}$$

para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$. Como

$$\overline{g_A(\vec{Y},\vec{X})} = \overline{\vec{Y}^H A\vec{X}} = \overline{\left(\vec{Y}^H A\vec{X}\right)^t} = \vec{X}^H A^H \vec{Y}, \tag{6.9}$$

Al comparar (6.8) y (6.9), se tiene que

$$g_A(\vec{X},\vec{Y}) = \vec{X}^t A\vec{Y} = \vec{X}^t A^H \vec{Y}. \tag{6.10}$$

Como (6.10) se cumple para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$, se concluye que $A = A^H$, es decir, A es hermitiana.

6.2. Forma cuadrática compleja

En esta sección se estudian las formas cuadráticas $F(\vec{X}) = \vec{X}^H A\vec{X}$, en donde A es una matriz compleja de tamaño $n \times n$ y la variable $\vec{X}$ se escoge en $\mathbb{C}^n$. Como en la práctica, generalmente uno sólo se preocupa de las formas cuadráticas $F(\vec{X}) = \vec{X}^H A\vec{X}$ que toman únicamente valores reales, en este apartado se asumirá que la matriz A asociada a la forma es hermitiana. Cabe notar que en los casos en que $F(\vec{X})$ es compleja, por lo general sólo se puede estudiar la parte real de $F(\vec{X})$.

Definición 6.3. Forma cuadrática compleja

Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre los números complejos. Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma hermítica sobre $\mathbb{V}$. Entonces una *forma cuadrática hermítica o forma cuadrática compleja* determinada por g es una función $F : \mathbb{V} \to \mathbb{R}$, tal que

$$F(\vec{v}) = g_A(\vec{v},\vec{v}) = \vec{v}^H A\vec{v}. \tag{6.11}$$

La matriz A se llama la **representación matricial de la forma cuadrática compleja.**

Ejemplo 6.2. ***Producto hermítico canónico***

Sea $V=\mathbb{C}^n$ y considere la forma cuadrática compleja determinada por el producto escalar sobre $\mathbb{C}^n$,

$$F(\vec{v})=|z_1|^2+|z_2|^2+\ldots+|z_n|^2,$$

donde $|z_i|^2=\overline{z_i}z_i$. Exprese esta forma cuadrática compleja como $\vec{v}^H A\vec{v}$.

Solución

Vamos a determinar la matriz compleja $A=(a_{ij})$ de la forma hermítica g, de tal forma que

$$F(\vec{v})=\vec{v}^H A\vec{v}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\overline{z_i}z_j. \tag{6.12}$$

Es decir, se quiere encontrar los valores de a_{ij}, de manera que

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\overline{z_i}z_j=\sum_{i=1}^{n}|z_i|^2.$$

Como la matriz A es hermitiana, $\overline{a_{ij}}=a_{ji}$. Por lo tanto, la forma cuadrática compleja dada en (6.12) se puede expresar como

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\overline{z_i}z_j=\sum_{i=1}^{n}a_{ii}|z_i|^2+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}a_{ij}\overline{z_i}z_j+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\overline{a_{ij}}\,z_i\overline{z_j}, \tag{6.13}$$

si se comparan términos se establecen las siguientes relaciones

$$\sum_{i=1}^{n}a_{ii}|z_i|^2=\sum_{i=1}^{n}|z_i|^2 \qquad \text{y} \qquad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}a_{ij}\overline{z_i}z_j+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\overline{a_{ij}}\,z_i\overline{z_j}=0.$$

Pero como en la función $F(\vec{v})$ no aparecen términos de la forma $\overline{z_i}z_j$, entonces

$$a_{ij}=\begin{cases}1 & \text{si } i=j\\ 0 & \text{si } i\neq j.\end{cases}$$

Luego, $A=I_n$ y por lo tanto, $F(\vec{v})$ se puede expresar como $\vec{v}^H I_n\vec{v}$.

Ejemplo 6.3. Sea $\mathbb{V}=\mathbb{C}^3$ y $F(\vec{X})=|x_1|^2-i\overline{x_1}x_2+i\overline{x_2}x_1-\overline{x_1}x_3-\overline{x_3}x_1-2i\overline{x_2}x_3+2i\overline{x_3}x_2$. Exprese esta forma hermítica como $\vec{X}^H A\vec{X}$.

Solución

Si se utiliza el resultado obtenido en (6.13) para $n = 3$, se tiene que

$$\vec{X}^H A\vec{X} = \sum_{i=1}^{3} a_{ii}|x_i|^2 + \sum_{i=1}^{2}\sum_{j=i+1}^{3} a_{ij}\overline{x_i}x_j + \sum_{i=1}^{2}\sum_{j=i+1}^{3} \overline{a_{ij}}x_i\overline{x_j}.$$

Al resolver esta suma y comparar los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

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

la cual permite expresar a $F(\vec{X})$ de la forma $\vec{X}^H A\vec{X}$.

6.3. Diagonalización de una forma hermítica

Teorema 6.6.

Sea $F(\vec{X})$ una forma cuadrática compleja asociada a una matriz hermítica A. Sea L una matriz compleja triangular inferior tal que A se pueda factorizar como LDL^H. Entonces el cambio de coordenadas

$$\vec{Z} = L^H\vec{X}, \tag{6.14}$$

transforma a $\vec{X}^H A\vec{X}$ en $\vec{Z}^H D\vec{Z}$.

Demostración

La matriz A asociada a la forma, se puede factorizar como

$$A = LDU.$$

Como A es hermítica, por el Teorema 3.5, $U = L^H$. Por lo tanto,

$$\begin{aligned} \vec{X}^H A\vec{X} =& \vec{X}^H \left(LDL^H\right)\vec{X} && \textit{puesto que } A = LDL^H \\ =& \left(\vec{X}^H L\right) D \left(L^H\vec{X}\right) \\ =& \left(L^H\vec{X}\right)^H D \left(L^H\vec{X}\right) = \vec{Z}^H D\vec{Z} && \textit{puesto que } \vec{Z} = L^H\vec{X}. \end{aligned}$$

Así, queda probado el teorema.

A continuación se presenta una versión de este método de diagonalización.

Procedimiento para diagonalizar una forma hermítica

$i)$ Halle la matriz de coeficientes hermítica A asociada a $F(\vec{X})$.

$ii)$ Obtenga la descomposición LDL^H de A, sin efectuar intercambios de filas que destruyan el hecho de que $\overline{a_{ij}} = a_{ji}$ y con elementos en $D = diag\{d_{11}, d_{22}, \ldots, d_{nn}\}$, tales que $d_{ii} \in \mathbb{R}$ no necesariamente distintos de cero. Además, $\det(L) = 1$.

$iii)$ Transforme a $F(\vec{X})$ en $d_{11}|z_1|^2 + d_{22}|z_2|^2 + \ldots + d_{nn}|z_n|^2$, bajo el cambio de coordenadas $\vec{Z} = L^H\vec{X}$.

Ejemplo 6.4. Considere la ecuación cuadrática compleja

$$|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 - \overline{x_1}x_3 - \overline{x_3}x_1 - 2i\overline{x_2}x_3 + 2i\overline{x_3}x_2 = 9 \tag{6.15}$$

encuentre una diagonalización para esta forma hermítica, usando el método descrito anteriormente.

Solución

En el Ejemplo 6.3, se obtuvo que la forma cuadrática hermítica asociada

$$|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 - \overline{x_1}x_3 - \overline{x_3}x_1 - 2i\overline{x_2}x_3 + 2i\overline{x_3}x_2,$$

se puede expresar matricialmente como

$$\vec{X}^H A\vec{X} = [\overline{x_1} \;\; \overline{x_2} \;\; \overline{x_3}] \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{6.16}$$

La factorización LDL^H de la matriz asociada a la forma hermítica es

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

de modo que

$$F(\vec{X}) = \vec{X}^H \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} \vec{X} = \vec{X}^H \begin{bmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ -1 & -i & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix} \vec{X}.$$

Si se hace

$$\vec{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = L^H \vec{X} = \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix} \vec{X}.$$

Nótese que $\det(L) = 1$. Por lo tanto, el cambio de variables

$$z_1 = x_1 - ix_2 - x_3, \qquad z_2 = x_2 + ix_3 \qquad \text{y} \qquad z_3 = x_3,$$

permite reescribir a $F(\vec{X})$ de la siguiente manera,

$$F(\vec{Z}) = |z_1|^2 - |z_2|^2 = 9.$$

Teorema 6.7. Teorema de los ejes principales

Sea $F(\vec{X})$ una forma hermítica asociada a una matriz hermitiana A con valores propios (no necesariamente distintos) $\lambda_1, \lambda_2, \ldots, \lambda_n$. Sea U una matriz unitaria que diagonaliza a A. Entonces el cambio de coordenadas

$$\vec{X} = U\vec{Z} \tag{6.17}$$

transforma a $\vec{X}^H A\vec{X}$ en $\vec{Z}^H U\vec{Z}$, donde $D = U^H AU = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

Demostración

La demostración consiste en un cálculo directo

$$\begin{aligned} \vec{X}^H A\vec{X} &= (U\vec{Z})^H A(U\vec{Z}) && \textit{puesto que } \vec{X} = U\vec{Z} \\ &= (\vec{Z}^H\, U^H) A(U\vec{Z}) \\ &= \vec{Z}^H (U^H AU)\vec{Z} = \vec{Z}^H D\vec{Z} && \textit{puesto que } U \textit{ diagonaliza a } A \end{aligned}$$

A continuación se presentan los pasos a seguir para determinar la diagonalización de una forma hermitiana mediante este método.

Procedimiento para diagonalizar una forma hermítica

i) Halle la matriz de coeficientes hermitiana A asociada a $F(\vec{X})$.

ii) Encuentre los valores propios (no necesariamente distintos), $\lambda_1, \lambda_2, \ldots, \lambda_n$ de A.

iii) Encuentre una base ortonormal para $\mathbb{C}^n$ formada por los vectores propios normalizados de A.

iv) Forme la matriz U cuyas columnas sean los vectores de la base hallada en el paso *iii*) en el orden correspondiente al listado de los valores propios en el paso *ii*). La transformación $\vec{X} = U\vec{Z}$ es una "**rotación**" si $|\det(U)| = 1$.

v) Transforme a $F(\vec{X})$ en $\lambda_1|z_1|^2 + \lambda_2|z_2|^2 + \ldots + \lambda_n|z_n|^2$.

Ejemplo 6.5. Considere la ecuación cuadrática compleja dada en el Ejemplo 6.4.

Determine la "superficie" cuadrática obtenida al eliminar los términos de productos

cruzados.

Solución

Haciendo referencia al Ejemplo 4.4, se tiene que la matriz A asociada a la forma cuadrática compleja es diagonalizable mediante la matriz unitaria

$$U = \begin{bmatrix} 2/\sqrt{6} & 1/\sqrt{3} & 0 \\ -i/\sqrt{6} & i/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{6} & -1/\sqrt{3} & -i/\sqrt{2} \end{bmatrix}.$$

Luego,

$$U^H AU = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Por consiguiente, (6.16) se puede escribir en términos de las nuevas variables z_1, z_2, z_3 como $\vec{Z}^H D\vec{Z}$, es decir,

$$3|z_2|^2 - 2|z_3|^2 = 9, \tag{6.18}$$

donde

$$\vec{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = U^H \vec{X} = \begin{bmatrix} 2/\sqrt{6} & i/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{3} & -i/\sqrt{3} & -1/\sqrt{3} \\ 0 & 1/\sqrt{2} & i/\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{6.19}$$

reescribiendo (6.18), se obtiene

$$\frac{1}{3}|z_2|^2 - \frac{1}{9/2}|z_3|^2 = 1,$$

lo cual corresponde a la ecuación de una “hipérbola” sobre los ejes z_2 y z_3.

Definición 6.4. Forma Polar de una forma hermitiana

Dada F una forma cuadrática compleja, se puede obtener una forma hermítica g de F de acuerdo con la siguiente identidad llamada la *forma polar* de g:

$$g(\vec{u},\vec{v}) = \frac{1}{4}\left[F(\vec{u}+\vec{v}) - F(\vec{u}-\vec{v})\right] + \frac{\imath}{4}\left[F(\vec{u}+\imath\vec{v}) - F(\vec{u}-\imath\vec{v})\right] \tag{6.20}$$

6.4. Clasificación de formas cuadráticas complejas

Definición 6.5. Una forma cuadrática compleja $F(\vec{X}) = \vec{X}^H A\vec{X}$ con $A \neq O$, es

1. **Definida positiva** si $F(\vec{X}) > 0$ para todo $\vec{X}$ distinto de cero en $\mathbb{C}^n$,
2. **Definida negativa** si $F(\vec{X}) < 0$ para todo $\vec{X}$ distinto de cero en $\mathbb{C}^n$,
3. **Indefinida** si $F(\vec{X})$ asume ambos valores positivos y negativos,
4. **Semidefinida positiva** si $F(\vec{X}) \geq 0$ para todo $\vec{X}$ en $\mathbb{C}^n$,
5. **Semidefinida negativa** si $F(\vec{X}) \leq 0$ para todo $\vec{X}$ en $\mathbb{C}^n$.

La matriz hermitiana asociada A, se denomina **definida positiva, semidefinida positiva, definida negativa, semidefinida negativa o indefinida** según sea la forma cuadrática compleja $F(\vec{X})$ que define.

Ejemplo 6.6. Verifique si la forma hermítica dada en el Ejemplo 6.2, es definida positiva.

Solución

La forma $F(\vec{v})$ dada en el Ejemplo 6.2, es definida positiva ya que, para todo $\vec{u} \neq 0$,

$$F(\vec{v}) = \vec{v}^H \vec{v} = \sum_{i=1}^{n} \overline{z_i} z_i = \sum_{i=1}^{n} |z_i|^2 > 0$$

donde $\vec{v}^H = \overline{(z_1, z_2, \ldots, z_n)^t}$ y $z_i \in \mathbb{C}$.

6.5. Orden parcial entre matrices

Dadas dos matrices, además de combinarlas haciendo operaciones entre ellas (suma, resta, multiplicación), las podemos comparar para ordenarlas o clasificarlas. Una comparación que surgió en secciones anteriores fue ver si eran semejantes. En esta sección se hablará de un orden "parcial" entre matrices semidefinidas positivas.

Definición 6.6. Orden entre matrices

Sean A y B matrices hermitianas de tamaño $n \times n$. Se escribe $A \succcurlyeq B$ si la matriz $A - B$ es semidefinida positiva. Similarmente, $A \succ B$ significa que la matriz $A - B$ es definida positiva.

Teorema 6.8. Si A, B son matrices hermitianas de tamaño $n \times n$, entonces

$$A \succcurlyeq B \qquad \text{implica que} \qquad T^H A T \succcurlyeq T^H B T$$

para toda $T \in M_{nn}$.

Demostración

Si $A - B$ es semidefinida positiva, entonces $\vec{Y}^H (A - B) \vec{Y} \geq 0$ para todo $\vec{Y} \in \mathbb{C}^n$. Así, $\vec{X}^H \left(T^H A T - T^H B T\right) \vec{X} = \left(T\vec{X}\right)^H (A - B) \left(T\vec{X}\right) \geq 0$ para todo $\vec{X} \in \mathbb{C}^n$ lo cual, a su vez, significa que $T^H A T \succcurlyeq T^H B T$.

Corolario 6.8.1. Si A, B son matrices de tamaño $n \times n$ definidas positivas, entonces

a) $A \succcurlyeq B$ si y sólo si $B^{-1} \succcurlyeq A^{-1}$.

b) Si $A \succcurlyeq B$, entonces $\det A \geq \det B$ y $tr(A) \geq tr(B)$ y

c) Si $A \succcurlyeq B$, entonces $\lambda_k(A) \geq \lambda_k(B)$ para todo $k = 1, 2, \ldots, n$ si los respectivos valores propios se colocan en el mismo orden (creciente o decreciente).

Ejemplo 6.7. Dadas las siguientes matrices

$$A = \begin{bmatrix} 7 & 1 & 2 \\ 1 & 7 & 2 \\ 2 & 2 & 10 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

muestre que A es más positiva que B.

Solución

Primero se obtiene la matriz $C = A - B$,

$$C = \begin{bmatrix} 5 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 8 \end{bmatrix}$$

los valores propios de esta matriz son

$$\lambda_1 = 5, \qquad \lambda_2 = 5 + \sqrt{11}, \qquad \lambda_3 = 5 - \sqrt{11}$$

como todos los valores propios son positivos entonces C es definida positiva y por lo tanto, A es más positiva que B.

Ejercicios 6.1.

1. Reduzca las siguientes formas complejas a una forma diagonal

   a) $2|x_1|^2 + (1-i)\overline{x_1}x_2 + (1+i)\overline{x_2}x_1 + 3|x_2|^2 = 4$

   b) $|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 + \overline{x_1}x_3 + \overline{x_3}x_1 - i\overline{x_2}x_3 + i\overline{x_3}x_2 - 2|x_3|^2 = 5$

2. Considere las matrices asociadas a cada una de las formas cuadráticas complejas del ejercicio 1. ¿Determine qué tipo de matrices son? (definida positiva o definida negativa).

3. Sea A una matriz hermítica cuadrada de tamaño 3×3. Supóngase que $\Delta_1 > 0$, $\Delta_2 < 0$ y $\Delta_3 > 0$. Muestre que A tiene un valor propio positivo y dos negativos.

4. Sea A cualquier matriz compleja no singular. Muestre que $B = A^H A$ es hermítica y definida positiva.

5. Muestre que si A es una matriz hermítica cuadrada de tamaño $n\times n$ definida positiva con valores propios $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ y B es una submatriz principal de A de tamaño $k\times k$, entonces
$$\prod_{j=1}^{k} \lambda_{n-j+1} \leq \det B \leq \prod_{j=1}^{k} \lambda_j.$$

6. Sean A y B matrices hermíticas cuadradas de tamaño $n\times n$ con valores propios $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ y $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ respectivamente. Sean $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ los valores propios de $A+B$. Demuestre que
$$\text{máx}\{\lambda_k+\mu_n, \lambda_n+\mu_k\} \leq \sigma_k \leq \text{máx}\{\lambda_k+\mu_1, \lambda_1+\mu_k\} \qquad 1\leq k\leq n.$$

Capítulo 7

Normas matriciales

En este capítulo se intenta medir la sensibilidad o la "vulnerabilidad" de la solución de un sistema no singular de ecuaciones lineales $A\vec{X} = \vec{b}$. En otras palabras, se quiere medir que tan grande es el efecto en $\vec{X} = A^{-1}\vec{b}$ si se cambian ligeramente las componentes de A y $\vec{b}$. Es decir, debemos encontrar una manera de medir el cambio ΔA y definir la "longitud" de una matriz, pues para vectores ya sabemos como obtener su longitud ahora necesitamos un concepto análogo para matrices.

7.1. Definición y resultados básicos

Definición 7.1. Norma de una matriz

Sea $\mathcal{M}_{nn}$ el espacio de las matrices de tamaño $n \times n$ con componentes reales (complejas). Una norma de matriz $\|\cdot\|$ de $\mathcal{M}_{nn}$ en $\mathbb{R}$ es una función que satisface para toda $A,\ B \in \mathcal{M}_{nn}$ los cinco axiomas siguientes:

(1)	$\|A\| \geq 0$	No negativa
(2)	$\|A\| = 0$ si y sólo si $A = 0$	Positiva
(3)	$\|cA\| = \|c\|\|A\|$ para todo escalar c	Homogénea

(4)	$\|A+B\| \leq \|A\|+\|B\|$	Desigualdad triangular
(5)	$\|AB\| \leq \|A\|\ \|B\|$	Submultiplicativa.

Teorema 7.1. Sea $\|\cdot\|$ cualquier norma matricial, entonces

1. $\|I_n\| \geq 1$, donde I_n es la matriz identidad de tamaño $n \times n$.
2. $\|A^{-1}\| \geq \dfrac{\|I_n\|}{\|A\|}$, para cualquier matriz $A \in \mathcal{M}_{nn}$ no singular.
3. $\left\|A^k\right\| \leq \|A\|^k$ para cualquier matriz $A \in \mathcal{M}_{nn}$ y todo $k \geq 2$.

Demostración

1. Queda como ejercicio para el lector.
2. Puesto que $AA^{-1} = I_n$, entonces
$$\|I_n\| = \|AA^{-1}\| \leq \|A\|\ \|A^{-1}\|.$$
Pero como $\|A\| > 0$, se tiene que
$$\|A^{-1}\| \geq \frac{\|I_n\|}{\|A\|}.$$
3. La demostración es por inducción sobre k. El resultado es trivial para $k = 2$ puesto que por la propiedad submultiplicativa
$$\|A^2\| \leq \|A\|\ \|A\| = \|A\|^2.$$
Supongamos que se ha demostrado para cualquier $k = m$, es decir
$$\|A^m\| \leq \|A\|^m.$$
Luego, $\|A^{m+1}\| = \|A^m A\|$ y por la propiedad submultiplicativa se tiene
$$\|A^m A\| \leq \|A^m\|\|A\| \leq \|A\|^m\|A\| = \|A\|^{m+1}.$$

7.2. Tipos de normas matriciales

Las siguientes son algunas de las normas que uno puede introducir en el espacio de matrices $\mathcal{M}_{nn}$ análogas a las normas de los espacios vectoriales.

Teorema 7.2. Norma L_1

Para cualquier matriz $A=[a_{ij}]$, la función $\|\cdot\|:\mathcal{M}_{nn}\to\mathbb{R}$ definida por

$$\|A\|_1=\sum_{i,j=1}^{n}|a_{ij}|,$$

es una norma de matriz

Demostración

Los axiomas de $(1)-(3)$ se satisfacen fácilmente de la definición de valor absoluto, se demostrará, por lo tanto, que se cumplen los axiomas (4) y (5)

$$\begin{aligned}(4)\ \|A+B\|_1=&\sum_{i,j=1}^{n}|a_{ij}+b_{ij}|\leq\sum_{i,j=1}^{n}(|a_{ij}|+|b_{ij}|)\\ =&\sum_{i,j=1}^{n}|a_{ij}|+\sum_{i,j=1}^{n}|b_{ij}|=\|A\|_1+\|B\|_1.\end{aligned}$$

Por lo tanto, se cumple el axioma (4).

$$\begin{aligned}(5)\ \|AB\|_1=&\sum_{i,j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}\,b_{kj}\right|\leq\sum_{i,j,k=1}^{n}|a_{ik}\,b_{jk}|\leq\sum_{i,j,k,m=1}^{n}|a_{ik}\,b_{mj}|\\ =&\left(\sum_{i,k=1}^{n}|a_{ik}|\right)\left(\sum_{j,m=1}^{n}|b_{mj}|\right)=\|A\|_1\,\|B\|_1.\end{aligned}$$

En la verificación de este axioma, la primera desigualdad se obtiene de la generalización de la desigualdad triangular y la segunda de los términos adicionales a la suma. Por consiguiente, $\|A\|_1$ si es una norma matricial.

Ejemplo 7.1. Norma Euclideana (L_2)

Determine si la norma L_2 definida por

$$\|A\|_2=\left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}},$$

es una norma de matriz.

Solución

Fácilmente se puede probar que los axiomas de $(1)-(3)$ se satisfacen. Por lo tanto, veamos si se cumplen los axiomas (4) y (5)

$$
\begin{aligned}
4.\ \|A+B\|_2^2 &= \sum_{i,j=1}^{n} |a_{ij}+b_{ij}|^2 \leq \sum_{i,j=1}^{n} \left(|a_{ij}|^2+2|a_{ij}|\ |b_{ij}|+|b_{ij}|^2\right)\\
&= \sum_{i,j=1}^{n} |a_{ij}|^2 + 2\sum_{i,j=1}^{n} |a_{ij}\ b_{ij}| + \sum_{i,j=1}^{n} |b_{ij}|^2\\
&\leq \left[\left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}} + \left(\sum_{i,j=1}^{n} |b_{ij}|^2\right)^{\frac{1}{2}}\right]^2\\
&= \left(\|A\|_2+\|B\|_2\right)^2.
\end{aligned}
$$

Luego, se cumple el axioma (4).

$$
\begin{aligned}
5.\ \|AB\|_2^2 &= \sum_{i,j=1}^{n} \left|\sum_{k=1}^{n} a_{ik}\ b_{kj}\right|^2 \leq \sum_{i,j=1}^{n} \left(\sum_{k=1}^{n} |a_{ik}|^2\right)\left(\sum_{m=1}^{n} |b_{mj}|^2\right)\\
&= \left(\sum_{i,k=1}^{n} |a_{ik}|^2\right)\left(\sum_{j,m=1}^{n} |b_{mj}|^2\right)\\
&= \|A\|_2^2\ \|B\|_2^2.
\end{aligned}
$$

Esta desigualdad es justo la *desigualdad de Cauchy-Schwarz*. Por consiguiente, $\|\cdot\|_2$ si es una norma.

Ejemplo 7.2. Norma L_∞

Determine si la función $\|\cdot\|:\mathcal{M}_{nn}\to\mathbb{R}$ definida por

$$\|A\|_\infty = n \max_{1\leq i,j\leq n} |a_{ij}|,$$

es una norma de matriz.

Solución

Los axiomas de $(1)-(3)$ se satisfacen fácilmente de la definición de máximo. Se probara si se cumplen los axiomas (4) y (5)

$$
\begin{aligned}
4.\ \|A+B\|_\infty =& n \max_{1\le i,j\le n} |a_{ij}+b_{ij}| \le\ n \max_{1\le i,j\le n} (|a_{ij}|+|b_{ij}|)\\
=& n \max_{1\le i,j\le n} |a_{ij}| + n \max_{1\le i,j\le n} |b_{ij}|\\
=& n\|A\|_\infty + n\|B\|_\infty.
\end{aligned}
$$

Por lo tanto, se cumple el axioma (4).

$$
\begin{aligned}
5.\ \|AB\|_\infty =& n \max_{1\le i,j\le n} \left|\sum_{k=1}^{n} a_{ik}\ b_{kj}\right| \le n \max_{1\le i,j\le n} \sum_{k=1}^{n} |a_{ik}\ b_{kj}|\\
\le& n \max_{1\le i,j\le n} \sum_{k=1}^{n} \|A\|_\infty\ \|B\|_\infty = n\|A\|_\infty\ n\|B\|_\infty\\
=& \|A\|\ \|B\|.
\end{aligned}
$$

Por consiguiente, $\|A\|_\infty$ si es una norma.

Definición 7.2. Norma matricial inducida

Sea $\|\cdot\|$ una norma vectorial sobre $\mathbb{C}^n$. Se define $\|\cdot\|_{in}$ sobre $\mathcal{M}_{nn}$ por

$$
\|A\|_{in} = \max_{\|\vec{x}\|=1} \|A\vec{x}\| = \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|}. \tag{7.1}
$$

Las letras "*in*" en la norma es la abreviación de la frase "norma inducida".

Teorema 7.3. Norma Espectral

La norma espectral $\|\cdot\|_S$ se define sobre $\mathcal{M}_{nn}$ por

$$
\|A\|_S = \sigma_1 = \text{máx}\{\sigma_i\ :\ \sigma_i \text{ es un valor singular de } A\}. \tag{7.2}
$$

Demostración

De la ecuación (7.1) se tiene que

$$
\|A\|_{in}^2 = \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} = \max_{\vec{x}\neq\vec{0}} \left\{ \frac{(A\vec{x})^t(A\vec{x})}{\vec{x}^t\vec{x}} \right\}.
$$

Luego, si $A^tA\vec{x} = \sigma^2\vec{x}$ se obtiene

$$\|A\|_{in}^2 = \max_{\vec{x}\neq\vec{0}}\left\{\frac{\vec{x}^t(A^tA\vec{x})}{\vec{x}^t\vec{x}}\right\} = \max_{\vec{x}\neq\vec{0}}\left\{\sigma^2\frac{\vec{x}^t\vec{x}}{\vec{x}^t\vec{x}}\right\} = \sigma^2_{\text{máx}}$$

como A^tA es una matriz simétrica, sus valores propios son reales.

Definición 7.3. Radio espectral

El radio espectral $r(A)$ de una matriz $A \in \mathcal{M}_{nn}$ es definido por la cantidad

$$r(A) = \text{máx}\{|\lambda| : \lambda \text{ es un valor propio de } A\}.$$

Ejemplo 7.3. Obtenga la norma espectral y el radio espectral de la matriz

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

Solución

Como la matriz A es simétrica, sus valores singulares y sus valores propios son iguales, es decir, $\sigma_1 = \lambda_1 = 4$ y $\sigma_2 = \lambda_2 = 2$. Por lo tanto,

$$\|A\|_S = 4 \qquad \text{y} \qquad r(A) = 4.$$

Teorema 7.4. Sea $A \in \mathcal{M}_{nn}$ y $\|\cdot\|$ cualquier norma de matriz, entonces

$$r(A) \leq \|A\|.$$

Demostración

Supóngase que $A\vec{x} = \lambda\vec{x}$, $\vec{x} \neq \vec{0}$ y que $|\lambda| = r(A)$. Sea $X \in \mathcal{M}_{nn}$ la matriz cuyas columnas son todas iguales a $\vec{x}$, entonces $AX = \lambda X$. Luego, si $\|\cdot\|$ es cualquier norma de matriz, se tiene que

$$\begin{aligned} \|AX\| &\leq \|A\|\ \|X\| \\ \|\lambda X\| &\leq \|A\|\ \|X\| \\ |\lambda|\ \|X\| &\leq \|A\|\ \|X\|. \end{aligned}$$

Por consiguiente, $|\lambda| = r(A) \leq \|A\|$.

Ejercicios 7.1.

1. Calcule para cada una de las siguientes matrices la norma espectral y el radio espectral

$$a. \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \qquad b. \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \qquad c. \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}.$$

$$d. \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}. \qquad e. \begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}.$$

2. Muestre que A y A^t tienen el mismo radio espectral y la misma norma espectral.

3. Si A es una matriz simétrica de tamaño $n \times n$, muestre que su norma espectral coincide con su radio espectral.

4. Si A es una matriz hermitiana de tamaño $n \times n$, muestre que la norma espectral y el radio espectral son iguales.

Teorema 7.5. Lema de Banach

Sea A una matriz real de tamaño $n \times n$ y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Suponiendo que $\|A\| < 1$, entonces $I_n - A$ es no singular y

$$\frac{1}{1+\|A\|} \leq \left\|(I_n - A)^{-1}\right\| \leq \frac{1}{1-\|A\|}.$$

Demostración

La matriz $I_n - A$ es no singular si y sólo si la única solución del sistema homogéneo $(I_n - A)\vec{x} = \vec{0}$ es $\vec{x} = \vec{0}$. Suponga entonces que $(I_n - A)\vec{x} = \vec{0}$ de modo que $\vec{x} = A\vec{x}$. Entonces se tiene que

$$\|\vec{x}\| = \|A\vec{x}\| \leq \|A\|\,\|\vec{x}\|.$$

Pero como $\|A\| < 1$, entonces hay una contradicción a menos que $\vec{x} = \vec{0}$, como se ha tratado probar. Así que $(I_n - A)^{-1}$ existe, se denotará con R. Luego

$$I_n = R(I_n - A) = R - RA. \tag{7.3}$$

Por lo tanto,

$$1 = \|I_n\| = \|R(I_n - A)\| \leq \|R\| \|I_n + (-A)\| \leq \|R\| (1 + \|A\|),$$

de manera que $\|R\| \geq 1/(1 + \|A\|)$ como se afirmó. De la expresión (7.3), $R = I_n + RA$, así que

$$\|R\| = \|I_n + RA\| \leq 1 + \|RA\| \leq 1 + \|R\| \|A\|.$$

Por consiguiente, $\|R\| \leq 1/(1 - \|A\|)$, con lo que se completa la prueba.

Ejemplo 7.4. Para la siguiente matriz

$$A = \frac{1}{10} \begin{bmatrix} 11 & -6 \\ 8 & 9 \end{bmatrix}.$$

Determine las cotas superior e inferior del Lema de Banach.

Solución

La matriz A se puede escribir como $A = I - B$, en donde

$$B = \frac{1}{10} \begin{bmatrix} -1 & 6 \\ -8 & 1 \end{bmatrix}.$$

Como se puede emplear cualquier norma matricial, usando la norma espectral se tiene que

$$\|B\|_S = \frac{7 + \sqrt{2}}{10} < 1.$$

El Lema de Banach dice que $A = I - B$ es no singular y por lo tanto

$$\frac{10}{17 + \sqrt{2}} \leq \left\|A^{-1}\right\|_S \leq \frac{10}{3 - \sqrt{2}}.$$

El *Lema de Banach* dice que matrices suficientemente “cercanas” a I_n son no singulares. El teorema siguiente es una generalización de este hecho.

Teorema 7.6. Inversas Perturbadas

Sean A y B matrices de tamaño $n \times n$ siendo A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Defínase $\alpha = \left\|A^{-1}B\right\|$ o $\alpha = \left\|BA^{-1}\right\|$. Si $\alpha < 1$ (es decir en especial si $\|B\| < 1/\left\|A^{-1}\right\|$) entonces $A-B$ también es no singular y

$$\frac{\left\|A^{-1}\right\|}{1+\alpha} \leq \left\|(A-B)^{-1}\right\| \leq \frac{\left\|A^{-1}\right\|}{1-\alpha}.$$

Demostración

Supongamos que $\left\|A^{-1}B\right\| < 1$. El otro caso es semejante. Como A^{-1} existe, se puede escribir

$$A-B \qquad \text{como} \qquad A\left(I_n - A^{-1}B\right) = A\left(I_n - R\right),$$

donde $R = A^{-1}B$. Por hipótesis $\|R\| = \alpha < 1$, de modo que al aplicar el *Lema de Banach*, se obtiene que $I_n - R$ es no singular, como lo es A. Luego

$$A\left(I_n - R\right) = A - B, \tag{7.4}$$

es también no singular y

$$(A-B)^{-1} = \left[A\left(I_n - R\right)\right]^{-1} = \left(I_n - R\right)^{-1} A^{-1}.$$

Por lo tanto,

$$\left\|(A-B)^{-1}\right\| \leq \left\|\left(I_n - R\right)^{-1}\right\| \left\|A^{-1}\right\| \leq \frac{\left\|A^{-1}\right\|}{1-\alpha}.$$

Por el *Lema de Banach*, ésta es la cota superior que se deseaba. Para obtener la cota inferior, se reescribe (7.4) como

$$A^{-1} = \left(I_n - R\right)(A-B)^{-1},$$

de lo cual se deduce que

$$\left\|A^{-1}\right\| \leq \left\|I_n - R\right\| \left\|(A-B)^{-1}\right\| \leq (1+\alpha)\left\|(A-B)^{-1}\right\|.$$

Al dividir por $(1+\alpha)$, se obtiene la cota inferior que se buscaba.

7.3. Condición de ecuaciones lineales

El concepto de *condición* es importante en todas las matemáticas aplicadas. Si "pequeños cambios en los datos" de un problema producen cambios razonablemente pequeños en la solución del mismo, se dice que el problema está *bien planteado*. Si "pequeños cambios en los datos" de algún problema ocasionan cambios inaceptablemente grandes en la solución, se dice que el problema está *mal planteado*. La razón de la importancia de este concepto debería ser evidente: En los problemas aplicados, casi siempre los datos son inexactos por errores de medición y de modelamiento y es crucial conocer los efectos que tienen sobre la solución del problema las inexactitudes en los datos.

Definición 7.4. Sistema de ecuaciones de mal comportamiento

Un sistema de ecuaciones lineales

$$A\vec{X} = \vec{b} \tag{7.5}$$

con A una matriz de tamaño $n \times n$, $\vec{X} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^n$, se dice que es un sistema de mal comportamiento si las n columnas de la matriz son casi linealmente dependientes o, en otras palabras, si la matriz de los coeficientes es casi singular. Esto significa que un cambio pequeño en algunos elementos de A produce una matriz singular.

Definición 7.5. Número de Condición

Sea A una matriz no singular real de tamaño $n \times n$, el **número de condición** se define como

$$\kappa(A) = \|A\|_S \|A^{-1}\|_S = \frac{\sigma_{max}}{\sigma_{min}}, \tag{7.6}$$

donde, σ_{max} y σ_{min} son respectivamente los valores singulares más grande y más pequeño, asociados a A.

Teorema 7.7. Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. La sensibilidad de la solución de un sistema de ecuaciones lineales

$$A\vec{x} = \vec{b}$$

con respecto a la perturbación ΔA en A se relaciona directamente con el número de condición. En otras palabras, si

$$\vec{y}, \qquad \text{resuelve a} \qquad (A+\Delta A)\,\vec{y} = \vec{b}.$$

Entonces el cambio en la solución satisface que

$$\frac{\|\vec{y}-\vec{x}\|}{\|\vec{x}\|} \leq \alpha\kappa(A),$$

en donde $\alpha = \|\Delta A\| \,/\, \|A\|$ es el error relativo en A.

Demostración

Puesto que la solución del sistema perturbado es $\vec{y}$, entonces

$$\begin{aligned}
(A+\Delta A)\,\vec{y} - \vec{b} &= \vec{0} \\
A\vec{y} + \Delta A\vec{y} - A\vec{x} &= \vec{0} \\
A\,(\vec{y}-\vec{x}) &= -\,\Delta A\vec{y} \\
\vec{y}-\vec{x} &= A^{-1}\,(-\Delta A\vec{y})
\end{aligned}$$

Luego, si $\|\cdot\|$ es cuarquier norma matricial, se tiene que

$$\begin{aligned}
\|\vec{y}-\vec{x}\| = \left\|-A^{-1}\Delta A\vec{y}\right\| &\leq \left\|A^{-1}\Delta A\right\| \|\vec{y}\| \\
&\leq \left\|A^{-1}\right\| \|\Delta A\| \, \|\vec{y}\|\,.
\end{aligned}$$

Como se quería.

Ejemplo 7.5. Resuelva el sistema de ecuaciones

$$\begin{aligned}
(1+\varepsilon)\,x_1 + x_2 &= 2 \qquad\qquad \varepsilon \in \mathbb{R} \\
x_1 + x_2 &= 1
\end{aligned}$$

y en las fórmulas que obtenga para x_1 y x_2 sustituya

(i) $\varepsilon = 0.01, 0.02$ $\qquad\qquad$ (ii) $\varepsilon = 2.01, 2.04$

compare los cambios en porcentaje del coeficiente de x_1, en la primera ecuación, en los casos (i) y (ii), con los cambios en porcentaje de la correspondiente solución de x_1.

Solución

Aplicando el método de eliminación de Gauss se obtiene

$$\left[\begin{array}{cc|c} 1+\varepsilon & 1 & 2 \\ 1 & 1 & 1 \end{array}\right] \underset{F_1-F_2}{\sim} \left[\begin{array}{cc|c} \varepsilon & 0 & 1 \\ 1 & 1 & 1 \end{array}\right].$$

Esto conduce a $x_1 = \frac{1}{\varepsilon}$ y $x_2 = 1 - \frac{1}{\varepsilon}$, $(\varepsilon \neq 0)$

	(i)		(ii)	
ε	0.01	0.02	2.01	2.04
Solución	$(100, -99)$	$(50, -49)$	$\left(\frac{100}{201}, \frac{101}{201}\right)$	$\left(\frac{25}{51}, \frac{26}{51}\right)$

El cambio en porcentaje del coeficiente de x_1 es 1% con una cifra decimal en cada caso.
La solución de x_1 en (i) varía en un 50% y en (ii) en un 2,41%. Esto indica que cuando ε es pequeño las "ecuaciones son de mal comportamiento".

Un modo sencillo de probar si un sistema de ecuaciones es de mal comportamiento consiste, precisamente, en proceder como lo hicimos en el ejemplo anterior, esto es, efectuar un pequeño cambio en algunos coeficientes para ver qué efectos se producen en la solución, pero esto es difícil de hacer cuando se trata de un sistema de ecuaciones muy grande. Existe un método que nos da una indicación de cuando se presenta el mal comportamiento, usando la definición 7.5.

El número de condición nos da una regla práctica para determinar si un sistema de ecuaciones es de mal comportamiento

$$\begin{array}{lrl} \text{Si} & 0 \leq \kappa(A) \leq 100 & \text{siempre el sistema es bien condicionado,} \\ & 100 < \kappa(A) \leq 1000 & \text{a veces el sistema es mal condicionado,} \\ & 1000 < \kappa(A) & \text{siempre el sistema es mal condicionado.} \end{array}$$

Ejemplo 7.6. Determine el número de condición de la matriz asociada al sistema de ecuaciones del Ejemplo anterior.

Solución

Como la matriz asociada al sistema es simétrica, entonces su norma espectral coincide con su radio espectral, por lo tanto se necesitan los valores propios de la matriz A. En este caso, el polinomio característico de A es

$$p_A(\lambda) = \lambda^2 - (2+\varepsilon)\lambda + \varepsilon,$$

de donde los valores propios son

$$\lambda_1 = 1 + \frac{1}{2}\varepsilon + \frac{1}{2}\sqrt{4+\varepsilon^2} \qquad \text{y} \qquad \lambda_2 = 1 + \frac{1}{2}\varepsilon - \frac{1}{2}\sqrt{4+\varepsilon^2}.$$

Luego, el número de condición de la matriz A es

$$\kappa(A) = \frac{\left(1 + \frac{1}{2}\varepsilon + \sqrt{1 + \left(\frac{1}{2}\varepsilon\right)^2}\right)^2}{\varepsilon}.$$

Para los valores de ε dados en el Ejemplo 7.5, se tiene que

	(i)		(ii)	
ε	0.01	0.02	2.01	2.04
$\kappa(A)$	402.01	202.02	5.8285	5.8292

Ejemplo 7.7. Suponga que la matriz de covarianza de un experimento con tres variables x_1, x_2 y x_3 es

$$S = \begin{bmatrix} \frac{2}{5} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{3}{10} & \frac{1}{5} \\ \frac{1}{10} & \frac{1}{5} & \frac{3}{10} \end{bmatrix}.$$

Encuentre el número de condición

Solución

Como S es simétrica entonces su norma espectral coincide con su radio espectral, por lo tanto se necesitan los valores propios de la matriz S. En este caso, el polinomio característico de S es

$$p_S(\lambda) = -\lambda^3 + \lambda^2 - \frac{27}{100}\lambda + \frac{9}{500},$$

de donde los valores propios son

$$\lambda_1 = \frac{3}{5}, \qquad \lambda_2 = \frac{3}{10} \qquad \text{y} \qquad \lambda_3 = \frac{1}{10}.$$

Luego, el número de condición es

$$\kappa(A) = \frac{\lambda_1}{\lambda_3} = \frac{3/5}{1/10} = 6.$$

Como $\kappa(A)$ es pequeño (< 100), significa que cambios pequeños en los datos producen cambios razonablemente pequeños en la estimación de la matriz de covarianza.

Teorema 7.8. Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. La sensibilidad de la solución de un sistema de ecuaciones lineales

$$A\vec{x} = \vec{b}$$

con respecto a la perturbación $\Delta\vec{b}$ en $\vec{b}$ se relaciona directamente con el número de condición. En otras palabras, si

$$A\vec{y} = \vec{b} + \Delta\vec{b} \qquad \text{con} \qquad \frac{\left\|\Delta\vec{b}\right\|}{\left\|\vec{b}\right\|} \leq \alpha,$$

entonces el cambio en la solución satisface que

$$\frac{\|\vec{y}-\vec{x}\|}{\|\vec{x}\|} \leq \alpha\kappa(A),$$

en donde $\alpha = \|\Delta A\| / \|A\|$ es el error relativo en A.

Demostración

Puesto que la solución del sistema perturbado es $\vec{y}$, entonces

$$\begin{aligned} A\vec{y} &= \vec{b} + \Delta\vec{b} \\ A\vec{y} - A\vec{x} - \Delta\vec{b} &= \vec{0} \\ A(\vec{y}-\vec{x}) &= \Delta\vec{b} \\ \vec{y}-\vec{x} &= A^{-1}\Delta\vec{b} \end{aligned}$$

Luego, si $\|\cdot\|$ es cuarquier norma matricial, se tiene que

$$\begin{aligned}\|\vec{y}-\vec{x}\| = \left\|A^{-1}\Delta\vec{b}\right\| &\leq \left\|A^{-1}\right\| \left\|\Delta\vec{b}\right\| \\ &\leq \alpha \left\|A^{-1}\right\| \left\|\vec{b}\right\|.\end{aligned}$$

Pero como $\vec{b} = A\vec{x}$ se tiene que $\left\|\vec{b}\right\| \leq \|A\| \|\vec{x}\|$. Entonces,

$$\|\vec{y}-\vec{x}\| \leq \alpha\kappa(A) \|\vec{x}\|.$$

Así, se completa la prueba.

Definición 7.6. Índice de Condición

Sea A una matriz real de tamaño $n \times n$, el **índice de condición** se define como

$$IC(A) = \sqrt{\kappa(A)}, \tag{7.7}$$

donde $\kappa(A)$ es el número de condición de A.

Teorema 7.9. Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Suponga que $\vec{x}$ resuelve a $A\vec{x} = \vec{b}$ mientras que

$$\vec{y} = \vec{x} + \Delta\vec{x} \qquad \text{resuelve a} \qquad (A+\Delta A)\vec{y} = \vec{b} + \Delta\vec{b},$$

para ciertas perturbaciones ΔA y $\Delta\vec{x}$ en los datos. Suponga que la perturbación ΔA es lo suficientemente pequeña como para que $\alpha < 1$, en donde $\alpha = \left\|(\Delta A)A^{-1}\right\|$ o $\alpha = \left\|A^{-1}(\Delta A)\right\|$. Entonces el cambio $\Delta\vec{x}$ en la solución satisface

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq M\kappa(A)\left(\frac{\|\Delta\vec{b}\|}{\|\vec{b}\|} + \frac{\|\Delta A\|}{\|A\|}\right),$$

en donde $M = 1/(1-\alpha)$ y $\kappa(A)$ es el número de condición de A.

Demostración

Si $\alpha < 1$, el Teorema 7.6 implica que $A+\Delta A$ es no singular y da una cota para la norma de su inversa. Como $A+\Delta A$ es no singular, la solución $\vec{y}$ al problema perturbado existe. De hecho, $\Delta\vec{x}$ misma resuelve

$$(A+\Delta A)\,\Delta\vec{x} = \vec{b}+\Delta\vec{b}-A\vec{x}-\Delta A\vec{x} = \Delta\vec{b}-\Delta A\vec{x}.$$

Así que

$$\Delta\vec{x} = (A+\Delta A)^{-1}\left(\Delta\vec{b}-\Delta A\vec{x}\right).$$

Al aplicar la cota superior dada en el Teorema 7.6, siendo $B = -\Delta A$ en esta ecuación para $\Delta\vec{x}$ se obtiene que

$$\begin{aligned}\|\Delta\vec{x}\| &\leq M\left\|A^{-1}\right\|\left\|\Delta\vec{b}-\Delta A\vec{x}\right\| \\ &\leq M\left\|A^{-1}\right\|\left(\left\|\Delta\vec{b}\right\|+\|\Delta A\|\,\|\vec{x}\|\right).\end{aligned}$$

Por lo tanto,

$$\begin{aligned}\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} &\leq M\left\|A^{-1}\right\|\left(\frac{\left\|\Delta\vec{b}\right\|}{\|\vec{x}\|}+\|\Delta A\|\right) \\ &\leq M\left\|A^{-1}\right\|\left(\frac{\left\|\Delta\vec{b}\right\|}{\left\|\vec{b}\right\|}\|A\|+\|\Delta A\|\right).\end{aligned}$$

Como $\vec{b} = A\vec{x}$, esto implica que $\left\|\vec{b}\right\| \leq \|A\|\,\|\vec{x}\|$. Al simplificar, se tiene que

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq M\,\kappa(A)\left(\frac{\left\|\Delta\vec{b}\right\|}{\left\|\vec{b}\right\|}+\frac{\|\Delta A\|}{\|A\|}\right),$$

lo cual completa la demostración.

Ejemplo 7.8. Considere el sistema de ecuaciones

$$\begin{aligned} x_1+6x_2 &= 0 \\ 6x_1+46x_2 &= 20 \end{aligned} \tag{7.8}$$

que tiene la solución exacta $x_1 = -12$, $x_2 = 2$. ¿Está el sistema bien condicionado o mal condicionado?.

Solución

Para determinar si el sistema de ecuaciones dado en (7.8) es estable, se cálcula el número de condición mediante la ecuación dada en (7.6). Si se expresa matricialmente (7.8), se tiene que

$$\begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \end{bmatrix}. \tag{7.9}$$

Como la matriz del sistema es simétrica entonces su norma espectral coincide con su radio espectral, por lo tanto, se necesitan los valores propios de la matriz A. En este caso, el polinomio característico es

$$p_A(\lambda) = \lambda^2 - 47\lambda + 10,$$

de donde los valores propios son

$$\lambda_{max} = \frac{47}{2} + \frac{3}{2}\sqrt{241} \qquad \text{y} \qquad \lambda_{min} = \frac{47}{2} - \frac{3}{2}\sqrt{241}.$$

Luego el número de condición es

$$\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{1}{10}\left(\frac{47}{2} + \frac{3}{2}\sqrt{241}\right)^2 = \frac{1}{20}\left(2189 + 141\sqrt{241}\right) \approx 218{,}9.$$

En este caso, como $\kappa(A)$ es muy grande (> 100) se dice que el sistema no es estable.

Ejercicios 7.2.

1. Encuentre para cada una de las siguientes matrices su número de condición y una matriz singular cercana.

$$a.\ \begin{bmatrix} 0{,}89 & 0{,}53 \\ 0{,}47 & 0{,}28 \end{bmatrix}. \qquad b.\ \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{k} \end{bmatrix}. \qquad c.\ \begin{bmatrix} 1{,}1 & 2{,}1 & 3{,}1 \\ 1{,}0 & -1{,}0 & 2{,}0 \\ 0{,}2 & 3{,}3 & 1{,}4 \end{bmatrix}.$$

2. Si A es una matriz simétrica real de tamaño $n \times n$, muestre que su número de condición es

$$\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}},$$

donde, λ_{max} y λ_{min} son los valores propios asociados a A más grande y más pequeño, respectivamente.

3. Sea A la matriz
$$A = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \qquad \text{por lo tanto,} \qquad A^{-1} = \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix}.$$
Si se emplea la norma $\|\cdot\|_1$ o la norma $\|\cdot\|_\infty$, se obtiene que
$$\|A\| = \|A^{-1}\| = 1+k \qquad \text{para} \qquad k \geq 0.$$
Luego el número de condición $\kappa(A) = (1+k)^2$, es grande para k grande. Sin embargo, si se considera el sistema de ecuaciones $A\vec{x} = \vec{b}$, siendo
$$\vec{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{se tiene que} \qquad \vec{x} = \begin{bmatrix} 1-k \\ 1 \end{bmatrix}.$$
Mientras que si sólo se altera a $\vec{b}$ por medio de δ_1, δ_2 $(\neq 0)$ a
$$\vec{b} + \Delta\vec{b} = \begin{bmatrix} 1+\delta_1 \\ 1+\delta_2 \end{bmatrix} \qquad \text{entonces} \qquad \Delta\vec{x} = \begin{bmatrix} \delta_1 - k\delta_2 \\ \delta_2 \end{bmatrix}.$$
Encuentre una cota para $\|\Delta\vec{x}\|/\|\vec{x}\|$ en términos de $\|\Delta\vec{b}\|/\|\vec{b}\|$ mediante la norma 1 o la norma ∞ para demostrar que este problema está bien condicionado, no obstante que el número de condición de A es grande.

Capítulo 8

Matrices idempotentes

En este capítulo se enuncian algunos teoremas concernientes a un tipo especial de matriz, la *matriz idempotente*. En muchas aplicaciones estadísticas[1] se incluyen este tipo de matrices, por ello se dedica este capítulo de manera exclusiva al tratamiento de dichas matrices.

8.1. Definición y propiedades

En el capítulo 5 cuando trabajamos las formas cuadráticas una condición que colocamos fue que la matriz asociada fuera simétrica. En el estudio de análisis de varianza, la matriz asociada a la forma cuadrática además de ser simétrica resulta ser idempotente (Véase Ejemplo 5.5 y problema 6 de los Ejercicios 5.1).

Definición 8.1. Matriz idempotente

Una matriz cuadrada A se dice que es idempotente si cumple que

$$A = A^2.$$

Teorema 8.1. Los valores propios de una matriz simétrica e idempotente son cero o uno.

[1] Véase capítulo 10

Demostración

Si A es idempotente y si λ es un valor propio de A, existe un vector $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda\vec{v}$. Si se premultiplica ambos lados por A, se tiene que

$$A^2\vec{v} = \lambda A\vec{v} = \lambda^2\vec{v}.$$

Como $A^2\vec{v} = A\vec{v} = \lambda\vec{v}$, se obtiene que

$$\begin{aligned}\lambda\vec{v} &= \lambda^2\vec{v}\\ (\lambda^2 - \lambda)\vec{v} &= \vec{0}.\end{aligned}$$

Pero $\vec{v} \neq \vec{0}$, así que $\lambda^2 - \lambda$ debe ser cero. Luego $\lambda = 0$ o $\lambda = 1$.

El recíproco del Teorema 8.1 no es cierto, véase el siguiente ejemplo.

Ejemplo 8.1. Dada la matriz

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

obtenga los valores propios y verifique si $A^2 = A$.

Solución

En este caso, los valores propios de A son $\lambda_1 = 1$ de multiplicidad algebraica 2 y $\lambda_2 = 0$. Pero $A^2 \neq A$ ya que

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

Por lo tanto, si los valores propios de una matriz son 0 y 1, no implica que la matriz sea idempotente.

Teorema 8.2. Si A es una matriz simétrica, idempotente y no singular, entonces $A = I_n$.

Demostración

Si A es idempotente, entonces $AA = A$. Multiplicando ambos lados por A^{-1}, se tiene lo que se quería demostrar.

Teorema 8.3. Sea A una matriz simétrica e idempotente de tamaño $n \times n$, con rango r, entonces existe una matriz ortogonal Q de tamaño $n \times n$ y una matriz R^* de tamaño $n \times n$ tal que

$$A = QR^* \qquad \text{y} \qquad R^*Q = \begin{bmatrix} I_r & \vdots & \mathbf{0} \\ \dots & \cdot & \dots \\ \mathbf{0} & \vdots & \mathbf{0} \end{bmatrix},$$

donde I_r es la matriz identidad de tamaño $r \times r$.

Demostración

Por el teorema 3.26, la matriz A se puede factorizar como sigue

$$A = USV^t,$$

donde U y V son matrices ortogonales de tamaño $n \times n$ y S es una matriz diagonal de tamaño $n \times n$ con r elementos iguales a uno y los elementos restantes $n - r$ de la diagonal iguales a cero.

Puesto que $A^2 = A$, se tiene que

$$USV^tUSV^t = USV^t,$$

de lo cuál se obtiene que

$$SV^tUS = S \qquad \text{o} \qquad SV^tU = I_r.$$

Tomando $R^* = SV^t$ y $Q = U$, se llega a

$$A = QR^* \qquad \text{con} \qquad R^*Q = I_r.$$

Teorema 8.4. Toda matriz cuadrada real A de tamaño $n \times n$ que pueda expresarse en la forma

$$A = QQ^t, \tag{8.1}$$

donde Q es una matriz real de tamaño $n \times m$ $(m < n)$ con columnas ortonormales en $\mathbb{R}^m$, satisface lo siguiente

(1) A es simétrica e idempotente.

(2) $A(I_n - A) = (I_n - A)A = \mathbf{0}$.

(3) $(I_n - A)Q = \mathbf{0}$.

Demostración

(1) Si $A = QQ^t$, entonces

$$\begin{aligned} A^t &= \left(QQ^t\right)^t = \left(Q^t\right)^t Q^t = QQ^t = A \\ A^2 &= \left(QQ^t\right)\left(QQ^t\right) = Q\left(Q^tQ\right)Q^t = QI_mQ^t = QQ^t = A. \end{aligned}$$

(2) $A(I_n - A) = A - A^2 = \mathbf{0}$, e igual para el otro caso.

(3) $(I_n - A)Q = Q - (QQ^t)Q = Q - QI_m = \mathbf{0}$.

Ejemplo 8.2. Encuentre una factorización QR de

$$A = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}.$$

Solución

Denotemos las columnas de A por

$$\vec{x}_1 = -\frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \vec{x}_2 = -\frac{1}{4}\begin{bmatrix} 1 \\ -3 \\ 1 \\ 1 \end{bmatrix}, \qquad \vec{x}_3 = -\frac{1}{4}\begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix}.$$

Apliquemos el algoritmo de Gram-Schmidt al conjunto $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ el cual es una base para el espacio generado por las columnas de A. Como $\|\vec{x}_1\| = \frac{1}{2}\sqrt{3}$, se hace

$\vec{v}_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|} = -\frac{1}{2\sqrt{3}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Por otra parte,

$$\vec{v}_2' = \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\vec{v}_1 = \begin{bmatrix} -\frac{1}{4} \\ \frac{3}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix} + \frac{1}{3}\begin{bmatrix} \frac{3}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}.$$

Entonces $\left\|\vec{v}_2'\right\| = \frac{1}{3}\sqrt{6}$ y $\vec{v}_2 = -\frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$. Por último,

$$\vec{v}_3' = \vec{x}_3 - (\vec{x}_3 \cdot \vec{v}_1)\vec{v}_1 - (\vec{x}_3 \cdot \vec{v}_2)\vec{v}_2 = \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \\ \frac{3}{4} \\ -\frac{1}{4} \end{bmatrix} + \frac{1}{3}\begin{bmatrix} \frac{3}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix},$$

entonces $\left\|\vec{v}_3'\right\| = \frac{1}{2}\sqrt{2}$ y $\vec{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$. Se puede verificar que es una nueva base ortonormal para el espacio generado por las columnas de A, observando que $\vec{v}_1 \cdot \vec{v}_2 = 0$, $\vec{v}_1 \cdot \vec{v}_3 = 0$ y $\vec{v}_2 \cdot \vec{v}_3 = 0$. Entonces formamos la matriz

$$Q = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} \frac{3}{2}\sqrt{2} & 0 & 0 \\ -\frac{1}{2}\sqrt{2} & 2 & 0 \\ -\frac{1}{2}\sqrt{2} & -1 & \sqrt{3} \\ -\frac{1}{2}\sqrt{2} & -1 & -\sqrt{3} \end{bmatrix}.$$

Luego la matriz R es

$$\begin{aligned} R = Q^t A &= \frac{1}{\sqrt{6}}\begin{bmatrix} \frac{3}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ 0 & 2 & -1 & -1 \\ 0 & 0 & \sqrt{3} & -\sqrt{3} \end{bmatrix}\begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix} \\ &= \frac{1}{\sqrt{6}}\begin{bmatrix} \frac{3}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ 0 & 2 & -1 & -1 \\ 0 & 0 & \sqrt{3} & -\sqrt{3} \end{bmatrix} \end{aligned}$$

nótese que $R = Q^t$, por lo tanto, la matriz A es idempotente.

Teorema 8.5. Si A es simétrica e idempotente y P es ortogonal, entonces P^tAP es idempotente.

Demostración

Si P es ortogonal, entonces

$$(P^tAP)(P^tAP) = P^tA(PP^t)AP = P^t(AA)P = P^tAP.$$

Teorema 8.6.

Sea A una matriz simétrica e idempotente de tamaño $n \times n$, con rango r, entonces existe una matriz ortogonal Q tal que $Q^tAQ = D_r$, donde D_r es una matriz diagonal con r elementos iguales a uno y los elementos restantes $n - r$ de la diagonal iguales a cero.

Demostración

Este se sigue inmediatamente del teorema 2.26.

Teorema 8.7. Si A es una matriz idempotente de tamaño $n \times n$, entonces su forma de Jordan $\mathcal{J} = P^{-1}AP$ satisface que $\mathcal{J}^2 = \mathcal{J}$.

Demostración

Queda como ejercicio para el lector.

Teorema 8.8. Si $A = [a_{ij}]$ es una matriz simétrica e idempotente y si el i-ésimo elemento de la diagonal es cero, entonces los elementos de la i-ésima fila y la i-ésima columna son todos idénticamente cero.

Demostración

Puesto que $A = A^2$, nosotros tenemos que el i-ésimo elemento de la diagonal de A es

$$a_{ii} = \sum_{j=1}^{n} a_{ij}a_{ji}.$$

Pero como A es simétrica $a_{ij} = a_{ji}$,

$$a_{ii} = \sum_{j=1}^{n} a_{ij}^2.$$

Luego si $a_{ii} = 0$, entonces $a_{ij} = 0$ (para $j = 1, 2, \ldots, n$), esto es, los elementos de la i-ésima fila son todos cero. Y dado que $A = A^t$ se tiene que los elementos de la i-ésima columna son también todos cero.

Teorema 8.9. El producto de dos matrices simétricas e idempotentes es idempotente si el producto de las dos matrices es conmutativo.

Demostración

Si $AB = BA$, entonces

$$(AB)(AB) = (AB)(BA) = A(BB)A = A(BA) = A(AB) = (AA)B = AB.$$

Teorema 8.10. Sea A una matriz real de tamaño $m \times n$ $(m > n)$ con $\rho(A) = n$, entonces la matriz $C = A\left(A^tA\right)^{-1}A^t$ es simétrica e idempotente, al igual que $\widetilde{C} = I_n - C$.

Demostración

La matriz $C = A\left(A^tA\right)^{-1}A^t$ es simétrica, ya que

$$C^t = \left[A\left(A^tA\right)^{-1}A^t\right]^t = \left(A^t\right)^t\left[\left(A^tA\right)^{-1}\right]^t A^t = A\left[\left(A^tA\right)^t\right]^{-1}A^t = C.$$

Además es idempotente pues

$$C^2 = \left[A\left(A^tA\right)^{-1}A^t\right]\left[A\left(A^tA\right)^{-1}A^t\right] = AI_k\left(A^tA\right)^{-1}A^t = C.$$

Nótese que la matriz A^tA es no singular pues A es de rango completo columna y $\rho\left(A^tA\right) = \rho\left(A\right)$, el lector puede probar fácilmente que $\widetilde{C} = I_n - C$ es también simétrica e idempotente.

Teorema 8.11. Sea A una matriz simétrica e idempotente de tamaño $n \times n$, entonces

$$\rho(A) = tr(A).$$

Demostración

Por el teorema 8.6, existe una matriz ortogonal Q tal que $A = QD_rQ^t$. Luego, se tiene que

$$tr(A) = tr(QD_rQ^t) = tr(D_rQ^tQ) = tr(D_r) = r = \rho(A).$$

Teorema 8.12. Todas las matrices simétricas idempotentes de rango incompleto son semidefinidas positivas.

Demostración

Queda como ejercicio al lector.

Teorema 8.13. Sea A una matriz real de tamaño $n \times n$, entonces A es simétrica e idempotente si y sólo si

$$\rho(A) + \rho(I_n - A) = n.$$

Demostración

Supongamos que
$$\rho(A) + \rho(I_n - A) = n$$
y sea $R(A)$ el espacio de los renglones de A. Veamos que

$$\mathbb{R}^n = R(A) \oplus R(I_n - A).$$

Obsérvese que

$$\begin{aligned} n =& dim(\mathbb{R}^n) = dim\{R(A) + R(I_n - A)\} \\ =& dim\{R(A)\} + dim\{R(I_n - A)\} - dim\{R(A) \cap R(I_n - A)\} \\ =& \rho(A) + \rho(I_n - A) - dim\{R(A) \cap R(I_n - A)\} \\ =& n - dim\{R(A) \cap R(I_n - A)\}. \end{aligned}$$

Esto implica que $dim\{R(A) \cap R(I_n - A)\} = 0$, de lo cuál se tiene que

$$R(A) \cap R(I_n - A) = \{\vec{0}\}.$$

Por consiguiente, $\mathbb{R}^n = R(A) \oplus R(I_n - A)$, esto exige que

$$A(I_n - A) = \mathbf{0}.$$

Supongamos que no es así, entonces existen vectores no nulos $\vec{u}$ y $\vec{v}$ en $\mathbb{R}^n$, tales que

$$A\,(I_n - A)\,\vec{u} = \vec{v}.$$

Luego $\vec{v} \in R(A)$, pero como $A\,(I_n - A) = (I_n - A)\,A$, se tiene que

$$(I_n - A)\,A\vec{u} = \vec{v}.$$

Esto implica que $\vec{v} \in R(I_n - A)$ y se llega a una contradicción. Por lo tanto, $A\,(I_n - A) = \mathbf{0}$ o $A^2 = A$. Esto completa la prueba.

Ejemplo 8.3. Determine el rango de la matriz asociada a la forma cuadrática del Ejemplo 5.5.

Solución

La matriz asociada a la forma cuadrática dada en el Ejemplo 5.5 era $I_n - \overline{J}_n$. Veamos si es simétrica e idempotente:

$$\begin{aligned}
\left(I_n - \overline{J}_n\right)^t &= \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)^t = \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right) \\
\left(I_n - \overline{J}_n\right)^2 &= \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right) \\
&= I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{1}{n^2}\mathbf{1}\underbrace{\mathbf{1}^t\mathbf{1}}\mathbf{1}^t = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t = I_n - \overline{J}_n.
\end{aligned}$$

Luego, por el teorema anterior se tiene que

$$\rho\left(I_n - \overline{J}_n\right) = n - \rho\left(\overline{J}_n\right) = n - 1,$$

pues la matriz $\overline{J}_n$ tiene únicamente una fila linealmente independiente.

Teorema 8.14. Sean A_1 y A_2 dos matrices cuadradas del mismo tamaño y $A = A_1 + A_2$, entonces las siguientes condiciones son equivalentes

(1) A es simétrica e idempotente y $\rho(A) = \rho(A_1) + \rho(A_2)$.

(2) A_1 y A_2 son simétricas e idempotentes y $A_1A_2 = A_2A_1 = 0$.

Demostración

Supongamos que (2) es verdadero, entonces

$$\begin{aligned} A^2 =&(A_1+A_2)(A_1+A_2)\\ =&A_1^2+A_2^2+A_1A_2+A_2A_1=A_1+A_2. \end{aligned}$$

Puesto que A, A_1 y A_2 son idempotentes,

$$\begin{aligned} \rho(A)=&tr(A)=tr(A_1+A_2)=tr(A_1)+tr(A_2)\\ =&\rho(A_1)+\rho(A_2). \end{aligned}$$

Ahora, supongamos que (1) es verdadero, por el Teorema 8.13

$$\begin{aligned} n=&\rho(A)+\rho(I_n-A)=\rho(A_1)+\rho(A_2)+\rho(I_n-A)\\ \geq&\rho(A_1)+\rho[A_2+(I_n-A)]=\rho(A_1)+\rho(I_n-A_1)\\ \geq&\rho[A_1+(I_n-A_1)]=\rho(I_n)=n. \end{aligned}$$

Por consiguiente, $\rho(A_1)+\rho(I_n-A_1)=n$ y de nuevo por el Teorema 8.13, se tiene que A_1 es idempotente; de manera análoga se puede mostrar que A_2 es idempotente. Ahora demostremos que $A_1A_2=A_2A_1=0$. Dado que A, A_1 y A_2 son idempotentes y $A=A_1+A_2$, multiplicando ambos lados por A se obtiene que

$$\begin{aligned} A=&A^2=(A_1+A_2)(A_1+A_2)\\ =&A_1^2+A_2^2+A_1A_2+A_2A_1=(A_1+A_2)+A_1A_2+A_2A_1\\ =&A+A_1A_2+A_2A_1. \end{aligned}$$

Esto implica que

$$A_1A_2+A_2A_1=\mathbf{0}, \qquad \text{es decir,} \qquad A_1A_2=-A_2A_1.$$

Por otra parte, el hecho de que $\rho(A)=\rho(A_1)+\rho(A_2)$ implica que

$$R(A_1)\cap R(A_2)=\{\vec{0}\}.$$

Este hecho unido con $A_1A_2=-A_2A_1$ da $A_1A_2=0$.

Corolario 8.14.1. Sean A_1,A_2 dos matrices de tamaño $n\times n$ tal que

$$A_1+A_2=I_n$$

entonces las condiciones dadas en el Teorema 8.14 se cumplen.

Demostración

Queda como ejercicio para el lector.

Una generalización del Teorema 8.14 en el que se involucran más de dos matrices es el siguiente

Teorema 8.15. Sean $A_1, A_2, \ldots, A_m$ una colección de m matrices de tamaño $n \times n$ y $A = \sum_{i=1}^{m} A_i$. Considere las siguientes condiciones:

(1) Cada A_i es simétrica e idempotente.

(2) $A_i A_j = 0$ para toda $i \neq j$ y $\rho\left(A_i^2\right) = \rho\left(A_i\right)$ para toda i.

(3) A es simétrica e idempotente.

(4) $\rho(A) = \sum_{i=1}^{m} \rho\left(A_i\right)$.

Entonces cualquiera dos de las condiciones (1), (2) y (3) implica la validez de la condición (4). Además, las condiciones (3) y (4) implica la validez del resto de las condiciones.

Demostración

Suponga que (1) y (2) son dadas. Como $A = \sum_{i=1}^{m} A_i$ es claro que es idempotente. Puesto que A y $A_1, A_2, \ldots, A_m$ son todas idempotentes,

$$\rho(A) = tr(A) = \sum_{i=1}^{m} tr\left(A_i\right) = \sum_{i=1}^{m} \rho\left(A_i\right).$$

Así, la condición (4) es verdadera.

Suponga que (2) y (3) son dadas. El cómputo de A^2 produce

$$A^2 = \sum_{i=1}^{m} A_i^2, \qquad \text{para} \qquad 1 \leq i \leq m.$$

Nótese que

$$AA_i = A_i A = A_i^2 \qquad \text{y} \qquad A^2 A_i = A_i A^2 = A_i^3$$

como A es idempotente, se tiene que $A_i^2 = A_i^3$, lo cual implica que $A_i^2 (I_n - A_i) = \mathbf{0}$. La condición $\rho(A_i) = \rho\left(A_i^2\right)$ es equivalente a la siguiente afirmación

$$dim\{R(A_i)\} = dim\left\{R\left(A_i^2\right)\right\}.$$

Puesto que $R\left(A_i^2\right) \subset R(A_i)$, se tiene que $R(A_i) = R\left(A_i^2\right)$. Por consiguiente, existe una matriz D no singular tal que $A_i = DA_i^2$. Por lo tanto, $A_i^2 (I_n - A) = \mathbf{0}$ implica que $A_i (I_n - A) = \mathbf{0}$ de lo cual se concluye que A_i es idempotente; así, la condición (1) es verdadera y se sigue la (4).

Supongamos que (3) y (4) son válidas. Para $i \neq j$, sea $B = A_i + A_j$ y $C = A - B$, por (4)

$$\begin{aligned}\sum_{i=1}^{m} \rho(A_i) =& \rho(A) = \rho(B + C)\\ \leq& \rho(B) + \rho(C) \leq \sum_{i=1}^{m} \rho(A_i).\end{aligned}$$

De esto, se tiene que $\rho(A) = \rho(B) + \rho(C)$, por otra parte,

$$\begin{aligned}n =& \rho(I_n) = \rho(B + I_n - B) \leq \rho(B) + \rho(I_n - B)\\ =& \rho(B) + \rho(I_n - A + C) \leq \rho(B) + \rho(I_n - A) + \rho(C)\\ =& \rho(A) + \rho(I_n - A) = n.\end{aligned}$$

Por lo tanto, $\rho(B) + \rho(I_n - B) = n$ y por el Teorema 8.13, B es idempotente. Así se tiene que $A_i + A_j$ es idempotente y $\rho(B) = \rho(A_i) + \rho(A_j)$. Por el Teorema 8.14, $A_iA_j = \mathbf{0}$ y A_i y A_j son idempotentes. Así, (2) y (3) se obtienen de una vez.

Suponga que (1) y (2) se cumplen. Es obvio que (4) se sigue aprovechando la conexión entre rango y traza para matrices idempotentes. Por lo tanto, se tiene que (4) es válido, (3) se sigue ahora de lo que se ha establecido anteriormente. Esto completa la prueba.

Corolario 8.15.1. Sean $A_1, A_2, \ldots, A_m$ una colección de matrices simétricas e idempotentes de tamaño $n \times n$ tal que

$$\sum_{i=1}^{m} A_i = I_n$$

entonces las condiciones dadas en el Teorema 8.15 se cumplen. En este caso, las condiciones (1) y (2) son equivalentes.

Demostración
Queda como ejercicio para el lector.

Teorema 8.16. Sean $A_1, A_2, \ldots, A_t$ una colección de matrices simétricas e idempotentes de tamaño $n \times n$. Una condición necesaria y suficiente para que exista una matriz P ortogonal tal que $P^tA_1P, P^tA_2P, \ldots, P^tA_tP$ sean todas diagonales es que $A_iA_j = A_jA_i$ para toda i y j.

Demostración
Queda como ejercicio para el lector.

Ejercicios 8.1.

1. Obtenga condiciones para los elementos de las matrices idempotentes de tamaño 2×2. ¿Se puede generalizar a cualquier dimensión?.

2. Muestre que si A es idempotente, entonces A^t es idempotente.

3. Sea X una matriz de tamaño $m \times n$ $(m > n)$ y rango n. Demuestre que la matriz $H = X(X^tX)^{-1}X^t$ es una matriz simétrica e idempotente. Obtenga la inversa de $I_n - H$.

4. Suponga que $KA = 0$ y K es idempotente. Defina $G = (A - K)^{-1}$. Pruebe que

$$(i)\ AG = I - K, \qquad (ii)\ AGA = A \qquad \text{y} \qquad (iii) \qquad AGK = 0.$$

Capítulo 9

Inversa generalizada de matrices

El concepto de inversa generalizada tiene sus principios en la teoría de ecuaciones lineales simultáneas (sistemas de m ecuaciones lineales con n incógnitas). La solución de un conjunto de ecuaciones lineales consistente

$$A\vec{x} = \vec{b}, \tag{9.1}$$

donde A es de tamaño $m \times n$ con rango $r \leq min(m,n)$, puede asumir dos formas diferentes. Si $m = n = r$, el sistema (9.1) tiene solución única $\vec{x} = A^{-1}\vec{b}$. Sin embargo, cuando A es una matriz rectangular o singular, una representación simple de una solución en términos de A es más difícil. En este capítulo, se trataran estos sistemas de ecuaciones usando las inversas generalizadas de matrices. Dichas matrices las estudiaremos como una aplicación de las descomposiciones de matrices.

9.1. Definición y propiedades básicas

En esta sección, se analizaran las inversas generalizadas de matrices rectangulares o singulares. Este tipo de inversas las estudiaremos como una aplicación de los valores propios, considerando los dos casos valores propios reales o complejos.

Definición 9.1. Inversa Generalizada

Para cualquier matriz A cuadrada o rectangular, se dice que G es una inversa generalizada de A, si satisface las siguientes condiciones:

$$(i)\ AGA = A, \qquad (iii)\ AG \text{ es simétrica} \qquad \text{y} \tag{9.2}$$
$$(ii)\ GAG = G, \qquad (iv)\ GA \text{ es simétrica.}$$

Nota

La inversa generalizada de A se llama también *seudoinversa* de A.

Teorema 9.1.

Si A es una matriz no singular, entonces $G = A^{-1}$.

Demostración

Queda como ejercicio al lector.

Notación

La notación y nomenclatura que se se usará en este capítulo, para los cuatro tipos de inversa generalizada introducido en (9.2) es el siguiente:

Condiciones que satisface	**Nombre**	**Abreviación**	**Notación**
(i)	Inv. Gen. condicionada	g_1-Inversa	A^{g_1} o A^c
(i) y (ii)	Inv. Gen. reflexiva	g_2-Inversa	A^{g_2} o A^r
$(i),(ii)$ y (iii)	Inv. Gen. normalizada	g_3-Inversa	A^{g_3} o A^n
$(i),(ii)$ y (iv)	Inv. Gen. normalizada	g_3^*-Inversa	$A^{g_3^*}$ o A^{n^*}
$(i),(ii),(iii)$y(iv)	La Inversa Generalizada	La g-Inversa	A^g o A^-

como veremos el término “normalizada” significa de norma mínima.

En la Definición 9.1 no se establece que toda matriz tenga inversa generalizada y además que ésta sea única. Por supuesto que así es, como lo establece el siguiente teorema.

Teorema 9.2.

Sea A una matriz cuadrada o rectangular, entonces:

$i)$ Siempre existe G. $\qquad$ $ii)$ G es única.

Demostración

i) Si A es la matriz nula de tamaño $m \times n$, es claro que la $g-$inversa de A es la matriz nula de tamaño $n \times m$.

Si se supone que $\rho(A) = r > 0$, entonces por la propiedad (iv) del rango de una matriz (ver Capítulo 1) se tiene que existen K y L de tamaño $m \times r$ y $r \times n$, respectivamente, ambas con rango r tales que

$$A = KL.$$

Entonces la matriz dada por

$$A^g = L^t \left(LL^t\right)^{-1} \left(K^tK\right)^{-1} K^t \tag{9.3}$$

es una $g-$inversa de A, pues basta sustituir en (9.2) para obtener que

$$\begin{aligned} AA^gA =& KLL^t \left(LL^t\right)^{-1} \left(K^tK\right)^{-1} K^tKL = KL = A. \\ A^gAA^g =& L^t \left(LL^t\right)^{-1} \left(K^tK\right)^{-1} K^t = A^g. \\ AA^g =& KLL^t \left(LL^t\right)^{-1} \left(K^tK\right)^{-1} K^t = K \left(K^tK\right)^{-1} K^t \qquad \text{es simétrica.} \\ A^gA =& L^t \left(LL^t\right)^{-1} \left(K^tK\right)^{-1} K^tKL = L^t \left(LL^t\right)^{-1} L \qquad \text{es simétrica.} \end{aligned}$$

Así pues, siempre existe una $g-$inversa de una matriz A.

ii) Para probar la unicidad se procede por reducción al absurdo y se supone que existen dos matrices A^g y B^g de tamaño $n \times m$ que son inversas generalizadas de A.

Por ser A^g una $g-$inversa de A se tiene que

$$AA^gA = A. \tag{9.4}$$

Al postmultiplicar por B^g se obtiene

$$AA^gAB^g = AB^g$$

y, dada la simetría de AB^g y AA^g, resulta

$$AB^g = (AB^g)^t = [(AA^g)(AB^g)]^t = (AB^gA)A^g = AA^g. \tag{9.5}$$

De manera análoga premultiplicando a (9.4) por B^g se llega a

$$B^gA = (B^gA)^t = [(B^gA)(A^gA)]^t = A^g(AB^gA) = A^gA. \tag{9.6}$$

Por último, si se premultiplica en (9.5) por B^g, se tiene

$$B^gAB^g = B^gAA^g$$

y, de acuerdo con (9.6) y la definición 9.1, resulta

$$B^g = B^gAB^g = (B^gA)A^g = A^gAA^g = A^g.$$

Es decir, la $g-$inversa de una matriz es única.

Ejemplo 9.1.

Dada la matriz $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}$, determine el tipo de inversa generalizada que es $G = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}$.

Solución

Veamos que condiciones cumple G de las dadas en (9.2):

$$AG = \frac{1}{3}\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \tag{9.7}$$

entonces

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

Por lo tanto, la matriz G es A^{g_1}. Observemos si cumple la segunda condición

$$GA = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 & 3 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}, \tag{9.8}$$

luego

$$GAG = \frac{1}{9}\begin{bmatrix} 6 & -3 \\ 3 & -6 \\ 0 & 0 \end{bmatrix} = G.$$

Así, G es una matriz A^{g_2} y de la expresión (9.7) se tiene finalmente que G es una A^{g_3}. No alcanza ser A^g ya que no cumple la cuarta condición, lo cual se verifica en (9.8).

9.2. Propiedades de las inversas generalizadas

Algunas de las propiedades más importantes de la inversa generalizada se resumen en el siguiente teorema.

Teorema 9.3. Sea A una matriz de tamaño $m \times n$ con rango $r \leq \text{mín}(m,n)$ y A^g una matriz de tamaño $n \times m$. Entonces:

a) $(A^g)^g = A$.

b) $(A^t)^g = (A^g)^t$.

c) $A = AA^t\,(A^g)^t = (A^g)^t\,A^tA$.

d) $(\alpha A)^g = \alpha^g A^g$, donde $\alpha \neq 0$ es cualquier escalar con $\alpha^g = \alpha^{-1}$.

d) $(A^tA)^g = A^g\,(A^g)^t$.

e) $A^g = A^g\,(A^g)^t\,A^t = A^t\,(A^g)^t\,A^g$.

f) $A^g = (A^tA)^g A^t = A^t\,(AA^t)^g$.

g) Las matrices AA^g, A^gA, $I_m - AA^g$ e $I_n - A^gA$ son todas idempotentes con rangos iguales a r, r, $m-r$ y $n-r$ respectivamente.

h) $\rho(A^g) = \rho(A)$.

Demostración

En esta demostración se utilizan las condiciones dadas en (9.2):

a) Se tiene inmediatamente de las condiciones.

b) Supongamos que la g-inversa de A^t es $(A^t)^g$, si se transpone la primera condición de la g-inversa de la matriz A, se tiene

$$\begin{aligned} [AA^gA]^t =& A^t \\ A^t(A^g)^tA^t =& A^t. \end{aligned}$$

Pero por la definición 9.1 la g-inversa es única, entonces $(A^t)^g = (A^g)^t$.

c) - f) Quedan como ejercicio para el lector.

g) Para verificar si las matrices son idempotentes se eleva cada una de ellas al cuadrado:

$$\begin{aligned}(AA^g)^2 =&(AA^g)(AA^g) = (AA^gA)A^g = AA^g,\\ (A^gA)^2 =&(A^gA)(A^gA) = A^g(AA^gA) = A^gA.\end{aligned}$$

De manera análoga se demuestra que $I - AA^g$ e $I - A^gA$ son idempotentes. Puesto que el rango de un producto no puede exceder el rango más pequeño de los factores:

$$\begin{aligned}r = \rho(A) =&\rho(AA^gA) \leq \rho(AA^g) \leq \rho(A) = r,\\ r = \rho(A) =&\rho(AA^gA) \leq \rho(A^gA) \leq \rho(A) = r.\end{aligned}$$

Por el teorema del emparedado se tienen las igualdades. Para demostrar que $\rho(I_m - AA^g) = m - r$ y $\rho(I_n - A^gA) = n - r$ se usa el hecho de que el rango de una matriz idempotente es igual a su traza. Luego,

$$\begin{aligned}\rho(I_m - AA^g) =&tr(I_m - AA^g) = m - tr(AA^g) = m - \rho(AA^g) = m - r.\\ \rho(I_n - A^gA) =&tr(I_n - A^gA) = n - tr(A^gA) = n - \rho(A^gA) = n - r.\end{aligned}$$

h) Por la parte a), si $AA^gA = A$, entonces

$$\rho(A) = \rho(AA^gA) \leq \rho(AA^g) \leq \rho(A^g).$$

Por otra parte, la condición $A^gAA^g = A^g$, implica que

$$\rho(A^g) = \rho(A^gAA^g) \leq \rho(AA^g) \leq \rho(A).$$

Así, $\rho(A^g) = \rho(A)$.

Teorema 9.4. Si A es una matriz simétrica, entonces A^g es simétrica.

Demostración

La prueba se sigue de la parte $b)$ del Teorema 9.3, es decir:

$$A^g = \left(A^t\right)^g = (A^g)^t.$$

Corolario 9.4.1. Si A es simétrica e idempotente entonces

$$A^g = A.$$

Demostración

Queda como ejercicio al lector.

Ejemplo 9.2. Determine la inversa generalizada de la matriz asociada a la forma cuadrática del Ejemplo 5.5

Solución

En el Ejemplo 8.3, se mostró que $I_n - \overline{J}_n$ era simétrica e idempotente. Luego por el Corolario 9.4.1, se tiene que

$$\left(I_n - \overline{J}_n\right)^g = I_n - \overline{J}_n.$$

Teorema 9.5. Si A y A^{g_3} son simétricas, entonces $A^{g_3} = A^g$.

Demostración

Puesto que A^{g_3} es simétrica,

$$(A^{g_3}A)^t = AA^{g_3} = (AA^{g_3})^t = A^{g_3}A$$

y la cuarta condición dada en (9.2) se satisface.

9.3. Métodos para calcular inversas generalizadas

En esta sección se ilustra algunos de los métodos para hallar la g-inversa. Se estudiarán sólo los métodos que utilizan las distintas factorizaciones de la matriz A dadas en estas notas. Aunque en esta sección se consideran únicamente matrices reales, cuando el lector necesite emplear alguno de los métodos desarrollados aquí para matrices complejas, simplemente puede realizar los cambios adecuados en cada método. Por ejemplo, en vez de utilizar A^t se usa A^H y si en el método se emplea una matriz ortogonal pues se cambia por una matriz unitaria.

Teorema 9.6. Sea A una matriz real de tamaño $m \times n$ con rango $r \leq \text{mín}(m,n)$ particionada como

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \ldots & . & \ldots \\ A_{21} & \vdots & A_{22} \end{bmatrix},$$

donde A_{11} ó A_{22} es una submatriz real de tamaño $r \times r$; entonces

1. Si A_{11} es no singular y $A_{22} = A_{21}A_{11}^{-1}A_{12}$, una g_2-inversa de A es la matriz A^{g_2} de tamaño $n \times m$, dada por:

$$A^{g_2} = \begin{bmatrix} A_{11}^{-1} & \vdots & 0_{r\times m_1} \\ \dots & . & \dots \\ 0_{n_1\times r} & \vdots & 0_{n_1\times m_1} \end{bmatrix}. \tag{9.9}$$

con $m_1 = m - r$ y $n_1 = n - r$.

2. Si A_{22} es no singular y $A_{11} = A_{12}A_{22}^{-1}A_{21}$, una g_2-inversa de A es la matriz A^{g_2} de tamaño $n \times m$, dada por:

$$A^{g_2} = \begin{bmatrix} 0_{n_1\times m_1} & \vdots & 0_{n_1\times r} \\ \dots & . & \dots \\ 0_{r\times m_1} & \vdots & A_{22}^{-1} \end{bmatrix}. \tag{9.10}$$

Demostración

Como la partición de las matrices expuestas son consistentes para el producto, efectúe los productos $AA^{g_2}A$ y $A^{g_2}AA^{g_2}$. Obsérvese que se obtiene respectivamente A y A^{g_2}.

Ejemplo 9.3.

Dada la matriz $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}$, obtenga una g_2-inversa.

Solución

Para la partición $1 \times (2+1)$ de la matriz dada se tiene

$$A = \begin{bmatrix} 2 & -1 & \vdots & 3 \\ 1 & -2 & \vdots & 3 \end{bmatrix} = (A_1 \vdots A_2).$$

Se calcula la inversa de A_1 y se obtiene

$$A_1^{-1} = -\frac{1}{3}\begin{bmatrix} -2 & 1 \\ -1 & 2 \end{bmatrix}.$$

Luego, una g_2-inversa es la matriz

$$A^{g_2} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}.$$

Nótese que ésta es igual a la dada en el Ejemplo 9.1.

Corolario 9.6.1. Sea A una matriz "diagonal" de tamaño $m \times n$ y rango $r \leq \text{mín}(m,n)$ particionada como sigue

$$A = \begin{bmatrix} D & \vdots & 0_{r\times n_1} \\ \dots & \cdot & \dots \\ 0_{m_1\times r} & \vdots & 0_{m_1\times n_1} \end{bmatrix}, \qquad \text{donde} \qquad D = \begin{bmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_r \end{bmatrix}.$$

con $m_1 = m - r$ y $n_1 = n - r$. Entonces la inversa generalizada de A está dada por

$$G = \begin{bmatrix} D^{-1} & \vdots & 0_{r\times m_1} \\ \dots & & \dots \\ 0_{n_1\times r} & \vdots & 0_{n_1\times m_1} \end{bmatrix}. \tag{9.11}$$

Demostración

Queda como ejercicio para el lector.

Teorema 9.7. Sea A una matriz simétrica de tamaño $n \times n$ y rango r, $(r < n)$. Entonces la inversa generalizada de A está dada por

$$G = P\Lambda^g P^{-1}, \tag{9.12}$$

donde P es una matriz real de tamaño $n \times n$ cuyas columnas son los vectores propios asociados a A, particionada como

$$P = [S \vdots T] = [\vec{v}_1 \ \dots \ \vec{v}_r \vdots \vec{v}_{r+1} \ \dots \ \vec{v}_n]. \tag{9.13}$$

Aquí, la submatriz S es de tamaño $n \times r$, sus columnas corresponden a los vectores propios asociados a los valores propios distintos de cero de la matriz A y la submatriz T es de tamaño $n \times (n-r)$, cuyas columnas corresponden a los vectores propios asociados a los valores propios nulos de A.

$$\Lambda^g = \begin{bmatrix} D^{-1} & 0 \\ 0 & \mathbf{0} \end{bmatrix}, \tag{9.14}$$

con D la submatriz real de tamaño $r \times r$ que tiene en la diagonal los valores propios distintos de cero asociados a A, $\mathbf{0}$ la submatriz real de tamaño $(n-r) \times (n-r)$ en cuya diagonal están los valores propios nulos de A.

Demostración

Puesto que A tiene n vectores propios $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ que corresponden a los valores propios (no necesariamente diferentes) $\lambda_1, \lambda_2, \ldots, \lambda_n$, dichos vectores resultan ser linealmente independientes y por lo tanto la matriz P dada en (9.13) es no singular.

Por consiguiente, la matriz A se puede expresar como $A = P\Lambda P^{-1}$, donde $\Lambda = \begin{bmatrix} D & 0 \\ 0 & \mathbf{0} \end{bmatrix}$.

Veamos si $G = P\Lambda^g P^{-1}$ cumple la primer condición de la definición 9.1

$$\begin{aligned} AGA =& A\left(P\Lambda^g P^{-1}\right)A = \left(P\Lambda P^{-1}\right)\left(P\Lambda^g P^{-1}\right)\left(P\Lambda P^{-1}\right) \\ =& P\Lambda\Lambda^g\Lambda P^{-1} = P\Lambda P^{-1} = A. \end{aligned}$$

Luego G es una matriz A^{g_1}. Observemos si es A^{g_2}:

$$\begin{aligned} GAG =& \left(P\Lambda^g P^{-1}\right)A\left(P\Lambda^g P^{-1}\right) = \left(P\Lambda^g P^{-1}\right)\left(P\Lambda P^{-1}\right)\left(P\Lambda^g P^{-1}\right) \\ =& P\Lambda^g\Lambda\Lambda^g P^{-1} = P\Lambda^g P^{-1} = G. \end{aligned}$$

Ahora, verifiquemos si G es una matriz A^{g_3}:

$$AG = \left[A\left(P\Lambda^g P^{-1}\right)\right] = \left[\left(P\Lambda P^{-1}\right)\left(P\Lambda^g P^{-1}\right)\right] = \left[P\left(\Lambda\Lambda^g\right)P^{-1}\right]. \tag{9.15}$$

Pero como $A = A^t$, por el Teorema 2.26, la matriz A es semejante a una matriz Q ortogonal. Si se ortonormalizan las columnas de la matriz P se tiene que $P^{-1} = P^t$ y por lo tanto

$$(AG)^t = [P(\Lambda\Lambda^g)P^{-1}]^t = \left\{P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}P^t\right\}^t = P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}P^t = AG.$$

También G es A^{g_3}. Observemos si cumple la cuarta condición dada en (9.2)

$$GA = [(P\Lambda^g P^{-1})A] = [(P\Lambda^g P^{-1})(P\Lambda P^{-1})] = [P(\Lambda^g\Lambda)P^{-1}]. \tag{9.16}$$

Usando de nuevo el hecho de que A es diagonalizable ortogonalmente, se tiene

$$(GA)^t = [P(\Lambda^g\Lambda)P^{-1}]^t = \left\{P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}P^t\right\}^t = P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}P^t = GA$$

Así, G es la g-inversa de A y el teorema queda demostrado.

Ejemplo 9.4. Inversa Generalizada de una Matriz Simétrica

Sea $A = \begin{bmatrix} 5 & -4 & 9 \\ -4 & 5 & -9 \\ 9 & -9 & 18 \end{bmatrix}$, obtenga la g-inversa.

Solución

En este caso, la ecuación característica es:

$$\det(A - \lambda I) = -\lambda^3 + 28\lambda^2 - 27\lambda = 0.$$

Entonces, los valores propios de A son $\lambda_1 = 1$, $\lambda_2 = 27$ y $\lambda_3 = 0$.

Para $\lambda_1 = 1$, se tiene el vector propio correspondiente $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

Si $\lambda_2 = 27$, se obtiene el vector propio asociado $\vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

y para $\lambda_3 = 0$, se llega al vector propio $\vec{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$.

Estableciendo

$$P = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = -\frac{1}{6}\begin{bmatrix} -3 & -3 & 0 \\ -1 & 1 & -2 \\ 2 & -2 & -2 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^g = \frac{1}{27}\begin{bmatrix} 27 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$G=-\frac{1}{162}\begin{bmatrix}1 & 1 & -1\\ 1 & -1 & 1\\ 0 & 2 & 1\end{bmatrix}\begin{bmatrix}27 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-3 & -3 & 0\\ -1 & 1 & -2\\ 2 & -2 & -2\end{bmatrix}.$$

Después de realizar la multiplicación de las matrices queda

$$G=\frac{1}{81}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix},$$

de manera que

$$AG=\frac{1}{81}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix},$$
$$GA=\frac{1}{81}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}.$$

Así, $AG = GA$, además los productos dan como resultado matrices simétricas. Por otra parte,

$$AGA=\frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix}=\frac{1}{3}\begin{bmatrix}15 & -12 & 27\\ -12 & 15 & -27\\ 27 & -27 & 54\end{bmatrix}=A,$$
$$GAG=\frac{1}{243}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix}=\frac{1}{243}\begin{bmatrix}123 & 120 & 3\\ 120 & 123 & -3\\ 3 & -3 & 6\end{bmatrix}=G.$$

Corolario 9.7.1. Sea A una matriz singular de tamaño $n\times n$ con valores propios (reales o complejos) distintos de cero $\lambda_1,\lambda_2,\ldots,\lambda_r$ $(r=\rho(A))$. Entonces una g_2-inversa de A es la matriz definida de la siguiente forma:

$$A^{g_2}=P\Lambda^g P^{-1}, \tag{9.17}$$

donde, P^1 y Λ^g estan definidas de manera análoga a (9.13) y (9.14), y

$$D = \begin{cases} diag(\lambda_1, \lambda_2, \ldots, \lambda_r) & \text{si m.a}(\lambda_i) = \text{m.g}(\lambda_i) \\ \mathcal{J} & \text{si m.a}(\lambda_i) \neq \text{m.g}(\lambda_i) \\ R & \text{si posee valores propios complejos,} \end{cases}$$

en donde $\mathcal{J}$ es la matriz de Jordan dada en (3.25). En este caso, $AG = GA$.

Demostración

En el Teorema 9.7, se demostró que G era A^{g_2}. Para demostrar que $AG = GA$, de (9.15) y (9.16) se tiene que

$$AG = [P(\Lambda\Lambda^g)P^{-1}] \qquad \text{y} \qquad GA = [P(\Lambda^g\Lambda)P^{-1}].$$

Dado que $\Lambda\Lambda^g = \Lambda^g\Lambda = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, el corolario queda demostrado.

Ejemplo 9.5. Inv. gen. de una matriz con valores propios reales

Sea $A = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix}$, obtenga una g_2-inversa.

Solución

En este caso, los valores propios de A son $\lambda_1 = 7$, $\lambda_2 = 3$ y $\lambda_3 = 0$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -7 \\ -3 \\ 5 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}$, respectivamente. Estableciendo

$$P = \begin{bmatrix} 1 & -7 & 1 \\ 1 & -3 & -6 \\ 1 & 5 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = \frac{1}{84}\begin{bmatrix} 27 & 12 & 45 \\ -7 & 0 & 7 \\ 8 & -12 & 4 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^g = \frac{1}{21}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

[1] Cuando la multiplicidad algebraica de un λ_i sea mayor que su multiplicidad geométrica, algunas de las columnas de P serán vectores propios generalizados

Se obtiene

$$G = \frac{1}{1764}\begin{bmatrix}1 & -7 & 1\\ 1 & -3 & -6\\ 1 & 5 & 1\end{bmatrix}\begin{bmatrix}3 & 0 & 0\\ 0 & 7 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}27 & 12 & 45\\ -7 & 0 & 7\\ 8 & -12 & 4\end{bmatrix}.$$

Después de realizar el producto entre matrices, se llega a

$$G = \frac{1}{441}\begin{bmatrix}106 & 9 & -52\\ 57 & 9 & -3\\ -41 & 9 & 95\end{bmatrix}$$

de manera que

$$AG = \frac{1}{441}\begin{bmatrix}4 & 1 & 2\\ 3 & 1 & 3\\ 1 & 1 & 5\end{bmatrix}\begin{bmatrix}106 & 9 & -52\\ 57 & 9 & -3\\ -41 & 9 & 95\end{bmatrix} = \frac{1}{21}\begin{bmatrix}19 & 3 & -1\\ 12 & 3 & 6\\ -2 & 3 & 20\end{bmatrix},$$

$$GA = \frac{1}{441}\begin{bmatrix}106 & 9 & -52\\ 57 & 9 & -3\\ -41 & 9 & 95\end{bmatrix}\begin{bmatrix}4 & 1 & 2\\ 3 & 1 & 3\\ 1 & 1 & 5\end{bmatrix} = \frac{1}{21}\begin{bmatrix}19 & 3 & -1\\ 12 & 3 & 6\\ -2 & 3 & 20\end{bmatrix}.$$

Entonces $AG = GA$, pero los productos no dan como resultado matrices simétricas. Por otra parte,

$$AGA = \frac{1}{21}\begin{bmatrix}19 & 3 & -1\\ 12 & 3 & 6\\ -2 & 3 & 20\end{bmatrix}\begin{bmatrix}4 & 1 & 2\\ 3 & 1 & 3\\ 1 & 1 & 5\end{bmatrix} = \frac{1}{21}\begin{bmatrix}84 & 21 & 42\\ 63 & 21 & 63\\ 21 & 21 & 105\end{bmatrix} = A, \quad \text{y}$$

$$GAG = \frac{1}{9261}\begin{bmatrix}19 & 3 & -1\\ 12 & 3 & 6\\ -2 & 3 & 20\end{bmatrix}\begin{bmatrix}106 & 9 & -52\\ 57 & 9 & -3\\ -41 & 9 & 95\end{bmatrix} = \frac{1}{9261}\begin{bmatrix}2226 & 189 & -1092\\ 1197 & 189 & -63\\ -861 & 189 & 1995\end{bmatrix} = G$$

Así, la matriz G cumple los requisitos (i) y (ii) dados en (9.2), pero no con las condiciones (iii) y (iv), pues AG y GA no son matrices simétricas.

Corolario 9.7.2. Sea A una matriz real de tamaño $m \times n$ $(n < m)$ y rango r, $(r \leq n)$.

Entonces la g-inversa de A está dada por

$$G = (A^tA)^gA^t \tag{9.18}$$

donde $(A^tA)^g = P\Lambda^gP^{-1}$ es la matriz definida en (9.12). Si $r = n$, entonces

$$G = (A^tA)^{-1}A^t \qquad \text{y} \qquad GA = I_n.$$

Demostración

En el Teorema 9.7, se demostró que la g-inversa de matrices simétricas cumplen las condiciones establecidas en la definición 9.1. Entonces, $(A^tA)^g$ las cumple. Veamos si la expresión dada en (9.18) verifica las condiciones dadas en (9.2):

$$AG = A(A^tA)^gA^t.$$

Pero por el Teorema 9.3 se tiene que $A = (A^t)^gA^tA$, luego,

$$\begin{aligned} AGA =&[(A^t)^gA^tA](A^tA)^gA^tA = (A^t)^g(A^tA)(A^tA)^g(A^tA)\\ =&(A^t)^g(A^tA) = A. \end{aligned}$$

Por otra parte,

$$GA = (A^tA)^gA^tA.$$

Entonces,

$$\begin{aligned} GAG =&(A^tA)^gA^tA(A^tA)^gA^t = (A^tA)^g(A^tA)(A^tA)^gA^t\\ =&(A^tA)^gA^t = G. \end{aligned}$$

Ahora, observemos si AG y GA son simétricas,

$$\begin{aligned} (AG)^t =&[A(A^tA)^gA^t]^t = A[(A^tA)^g]^tA^t\\ =&A[(A^tA)^t]^gA^t = AG \end{aligned}$$

y

$$\begin{aligned} (GA)^t =&[(A^tA)^gA^tA]^t = [(A^tA)^g(A^tA)]^t\\ =&(A^tA)^g(A^tA) = GA \end{aligned}$$

la última expresión se tiene debido a que $(A^tA)^g$ es una g-inversa de (A^tA).

Corolario 9.7.3. Sea A una matriz real de tamaño $m\times n$ $(m<n)$ y rango r, $(r \leq n)$.

Entonces la g-inversa de A es

$$G = A^t(AA^t)^g \tag{9.19}$$

donde $(AA^t)^g = P\Lambda^gP^{-1}$ es la matriz definida en (9.12). Si $r = m$, entonces

$$G =A^t(AA^t)^{-1} \qquad \text{y} \qquad AG =I_m.$$

Demostración

Queda como ejercicio para el lector.

Ejemplo 9.6. Considérese la matriz dada en el Ejemplo 9.1, obtenga la g-inversa.

Solución

Como $\rho(A) = 2$, el producto de AA^t da como resultado:

$$AA^t = \begin{bmatrix} 14 & 13 \\ 13 & 14 \end{bmatrix} \qquad \text{y} \qquad \left(AA^t\right)^{-1} = \frac{1}{27}\begin{bmatrix} 14 & -13 \\ -13 & 14 \end{bmatrix}.$$

Luego, la inversa generalizada es

$$A^g = A^t\left(AA^t\right)^{-1} = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix}.$$

la cual es diferente a la A^{g_2} dada en el Ejemplo 9.1.

En el caso de que no se establezca primero el rango de la matriz A, se puede realizar el producto de A^tA el cual da como resultado

$$A^tA = \begin{bmatrix} 5 & -4 & 9 \\ -4 & 5 & -9 \\ 9 & -9 & 18 \end{bmatrix}.$$

En el Ejemplo 9.4 se obtuvo que la g-inversa para ésta matriz era

$$(A^tA)^g = \frac{1}{81}\begin{bmatrix} 41 & 40 & 1 \\ 40 & 41 & -1 \\ 1 & -1 & 2 \end{bmatrix}.$$

Por lo tanto, la g-inversa de la matriz A, es:

$$A^g = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix},$$

la cual coincide con la obtenida anteriormente.

Ejemplo 9.7. Determine una g-inversa para la matriz

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

Solución

El producto de AA^t da como resultado:

$$B = AA^t = \begin{bmatrix} 14 & 7 & 7 \\ 7 & 6 & 1 \\ 7 & 1 & 6 \end{bmatrix}.$$

En este caso, los valores propios de B son $\lambda_1 = 21$, $\lambda_2 = 5$ y $\lambda_3 = 0$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, respectivamente. Estableciendo

$$P = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = \frac{1}{6} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -3 & 3 \\ -2 & 2 & 2 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 21 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^g = \frac{1}{105} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 21 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$(AA^t)^g = \frac{1}{630} \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 21 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -3 & 3 \\ -2 & 2 & 2 \end{bmatrix}.$$

Después de multiplicar las matrices queda

$$(AA^t)^g = \frac{1}{315} \begin{bmatrix} 10 & 5 & 5 \\ 5 & 34 & -29 \\ 5 & -29 & 34 \end{bmatrix}.$$

Por lo tanto, la g-inversa de la matriz A es

$$A^g = \frac{1}{105} \begin{bmatrix} 5 & -29 & 34 \\ 10 & 5 & 5 \\ 15 & 18 & -3 \end{bmatrix}.$$

El lector puede verificar que esta matriz cumple las condiciones dadas en (9.2).

Teorema 9.8. Supóngase que $A = LU$ es una descomposición de la matriz A de tamaño $m \times n$, de rango $r \leq \text{mín}(m,n)$. Entonces la inversa generalizada de A está dada por

$$G = \widetilde{U}^t \left(\widetilde{U}\,\widetilde{U}^t\right)^{-1} \left(\widetilde{L}^t \widetilde{L}\right)^{-1} \widetilde{L}^t, \tag{9.20}$$

donde $\widetilde{U}$ es una matriz de tamaño $n \times r$ de rango r, obtenida de eliminar las filas nulas de U y la matriz $\widetilde{L}$ de tamaño $m \times r$, también de rango r, es obtenida eliminando las columnas que multiplican a las filas nulas de U.

Demostración

Queda como ejercicio para el lector.

Ejemplo 9.8. Considérese la transpuesta de la matriz dada en el Ejemplo 9.1 y utilice el Teorema 9.8 para hallar la g-inversa.

Solución

Si se transpone la matriz dada en el Ejemplo 9.1, la factorización LU es

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

Al eliminar la última fila de U y la última columna de L se obtiene

$$\begin{bmatrix} 2 & 1 \\ -1 & -2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \\ \frac{3}{2} & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -\frac{3}{2} \end{bmatrix} = \widetilde{L}\,\widetilde{U}.$$

Luego,

$$\widetilde{U}\,\widetilde{U}^t = \begin{bmatrix} 2 & 1 \\ 0 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & -\frac{3}{2} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 20 & -6 \\ -6 & 9 \end{bmatrix}$$

y por lo tanto,

$$\widetilde{U}^t \left(\widetilde{U}\,\widetilde{U}^t\right)^{-1} = \begin{bmatrix} 2 & 0 \\ 1 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{9} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 3 & 2 \\ 0 & -4 \end{bmatrix}.$$

Por otra parte,

$$\widetilde{L}^t\widetilde{L} = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \\ \frac{3}{2} & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 7 & -4 \\ -4 & 4 \end{bmatrix}.$$

De donde,

$$\left(\widetilde{L}^t\widetilde{L}\right)^{-1}\widetilde{L}^t = \frac{1}{6}\begin{bmatrix} 4 & 4 \\ 4 & 7 \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -1 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 4 & 2 & 2 \\ 4 & 5 & -1 \end{bmatrix}.$$

Siguiendo el procedimiento dado en (9.20), se tiene que

$$G = \frac{1}{36}\begin{bmatrix} 3 & 2 \\ 0 & -4 \end{bmatrix}\begin{bmatrix} 4 & 2 & 2 \\ 4 & 5 & -1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & 4 & 1 \\ -4 & -5 & 1 \end{bmatrix},$$

la cual coincide con la transpuesta obtenida en el Ejemplo 9.6.

Ejemplo 9.9. Considérese la matriz dada en el Ejemplo 9.5, obtenga la factorización LU de A y utilice el Teorema 9.8 para hallar la g-inversa.

Solución

La factorización LU de la matriz dada en el Ejemplo 9.5 es

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

Al eliminar la última fila de U y la última columna de L se obtiene

$$\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{3}{4} & 1 \\ \frac{1}{4} & 3 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 0 & \frac{1}{4} & \frac{3}{2} \end{bmatrix} = \widetilde{L}\,\widetilde{U}.$$

Luego,

$$\widetilde{U}\,\widetilde{U}^t = \begin{bmatrix} 4 & 1 & 2 \\ 0 & \frac{1}{4} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 1 & \frac{1}{4} \\ 2 & \frac{3}{2} \end{bmatrix} = \frac{1}{16}\begin{bmatrix} 336 & 52 \\ 52 & 37 \end{bmatrix}$$

y por lo tanto,

$$\widetilde{U}^t\left(\widetilde{U}\,\widetilde{U}^t\right)^{-1} = \frac{1}{608}\begin{bmatrix} 4 & 0 \\ 1 & \frac{1}{4} \\ 2 & \frac{3}{2} \end{bmatrix}\begin{bmatrix} 37 & -52 \\ -52 & 336 \end{bmatrix} = \frac{1}{152}\begin{bmatrix} 37 & -52 \\ 6 & 8 \\ -1 & 100 \end{bmatrix}.$$

Por otra parte,

$$\widetilde{L}^t\widetilde{L} = \begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{4} \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{3}{4} & 1 \\ \frac{1}{4} & 3 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 13 & 12 \\ 12 & 80 \end{bmatrix},$$

de donde,

$$\left(\widetilde{L}^t\widetilde{L}\right)^{-1}\widetilde{L}^t = \frac{1}{112}\begin{bmatrix} 80 & -12 \\ -12 & 13 \end{bmatrix}\begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{4} \\ 0 & 1 & 3 \end{bmatrix} = \frac{1}{28}\begin{bmatrix} 20 & 12 & -4 \\ -3 & 1 & 9 \end{bmatrix}.$$

Siguiendo el procedimiento dado en (9.20), se tiene que

$$G = \frac{1}{4256}\begin{bmatrix} 37 & -52 \\ 6 & 8 \\ -1 & 100 \end{bmatrix}\begin{bmatrix} 20 & 12 & -4 \\ -3 & 1 & 9 \end{bmatrix} = \frac{8}{4256}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix}.$$

Nótese que esta matriz no coincide con la obtenida en el Ejemplo 9.5, por otra parte

$$AG = \frac{1}{532}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix},$$

$$GA = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{38}\begin{bmatrix} 37 & 6 & -1 \\ 6 & 2 & 6 \\ -1 & 6 & 37 \end{bmatrix}.$$

En este caso AG y GA dan como resultado matrices simétricas. Además,

$$AGA = \frac{1}{14}\begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 56 & 14 & 28 \\ 42 & 14 & 42 \\ 14 & 14 & 70 \end{bmatrix} = A, \quad \text{y}$$

$$GAG = \frac{1}{20216}\begin{bmatrix} 37 & 6 & -1 \\ 6 & 2 & 6 \\ -1 & 6 & 37 \end{bmatrix}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = G.$$

Así, la matriz G cumple todos los requisitos dados en (9.2).

Teorema 9.9. Supóngase que $A = QR$ es una descomposición de la matriz A de tamaño $m \times n$, de rango $r \leq \text{mín}(m,n)$ de modo que Q tiene columnas ortonormales y R es triangular superior de rango r. Entonces la inversa generalizada de A está dada por

$$G = R^t\left(RR^t\right)^g Q^t. \tag{9.21}$$

Demostración

Queda como ejercicio para el lector.

Ejemplo 9.10. Considérese la matriz dada en el Ejemplo 9.1, obtenga la factorización QR de A y utilice el Teorema 9.9 para hallar la g-inversa.

Solución

La factorización QR de la matriz dada en el Ejemplo 9.1 es

$$\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{4}{5}\sqrt{5} & \frac{9}{5}\sqrt{5} \\ 0 & \frac{3}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix}.$$

Luego,

$$RR^t = \begin{bmatrix} \sqrt{5} & -\frac{4}{5}\sqrt{5} & \frac{9}{5}\sqrt{5} \\ 0 & \frac{3}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ -\frac{4}{5}\sqrt{5} & \frac{3}{5}\sqrt{5} \\ \frac{9}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 122 & -39 \\ -39 & 18 \end{bmatrix}$$

y por lo tanto,

$$\left(RR^t\right)^g = \frac{1}{135} \begin{bmatrix} 18 & 39 \\ 39 & 122 \end{bmatrix}.$$

Por otra parte,

$$R^t\left(RR^t\right)^g = \frac{1}{135} \begin{bmatrix} \sqrt{5} & 0 \\ -\frac{4}{5}\sqrt{5} & \frac{3}{5}\sqrt{5} \\ \frac{9}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} 18 & 39 \\ 39 & 122 \end{bmatrix} = \frac{1}{135} \begin{bmatrix} 18\sqrt{5} & 39\sqrt{5} \\ 9\sqrt{5} & 42\sqrt{5} \\ 9\sqrt{5} & -3\sqrt{5} \end{bmatrix}.$$

Empleando el procedimiento dado en (9.21), se tiene que

$$G = \frac{1}{135} \begin{bmatrix} 18\sqrt{5} & 39\sqrt{5} \\ 9\sqrt{5} & 42\sqrt{5} \\ 9\sqrt{5} & -3\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \end{bmatrix} = \frac{15}{135} \begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.6.

Ejemplo 9.11. Inversa generalizada de una matriz cuadrada

Obtenga la g-inversa, usando el procedimiento dado en (9.21) para la matriz

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

Solución

La factorización QR de la matriz dada es

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

Luego,

$$RR^t = \begin{bmatrix} \sqrt{3} & -\sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{6} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ -\sqrt{3} & 0 & 0 \\ \sqrt{3} & \sqrt{6} & 0 \end{bmatrix} = \begin{bmatrix} 9 & 3\sqrt{2} & 0 \\ 3\sqrt{2} & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

y por lo tanto,

$$\left(RR^t\right)^g = \frac{1}{12} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -\sqrt{2} & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Por otra parte,

$$\begin{aligned} R^t\left(RR^t\right)^g &= \frac{1}{12} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ -\sqrt{3} & 0 & 0 \\ \sqrt{3} & \sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -\sqrt{2} & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{12} \begin{bmatrix} 2\sqrt{3} & -\sqrt{6} & 0 \\ -2\sqrt{3} & \sqrt{6} & 0 \\ 0 & 2\sqrt{6} & 0 \end{bmatrix}. \end{aligned}$$

Mediante el procedimiento dado en (9.21), se tiene que

$$G = \frac{1}{12} \begin{bmatrix} 2\sqrt{3} & -\sqrt{6} & 0 \\ -2\sqrt{3} & \sqrt{6} & 0 \\ 0 & 2\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} & -\frac{1}{3}\sqrt{3} \\ -\frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{6} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix},$$

de manera que

$$\begin{aligned} AG &= \frac{1}{12} \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}, \\ GA &= \frac{1}{12} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}. \end{aligned}$$

De este modo $AG = GA$. Además,

$$AGA = \frac{1}{12}\begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}\begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \frac{1}{12}\begin{bmatrix} -12 & 12 & -24 \\ 12 & -12 & 24 \\ -12 & 12 & 12 \end{bmatrix} = A, \quad \text{y}$$

$$GAG = \frac{1}{144}\begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}\begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} = G.$$

Así, la matriz G cumple todos los requisitos dados en (9.2).

Teorema 9.10. Sea A una matriz real de tamaño $n \times n$ con valores singulares distintos de cero $\sigma_1, \sigma_2, \ldots, \sigma_r$ ($r = \rho(A)$). Entonces la inversa generalizada de A está dada por

$$G = VS^gU^t, \tag{9.22}$$

donde U y V son matrices ortogonales de tamaño $n \times n$ y S^g es la inversa generalizada de la matriz dada en (3.31).

Demostración

Por el Teorema 3.25, la matriz A se puede expresar como $A = USV^t$, por consiguiente la g-inversa es

$$G = VS^gU^t.$$

Veamos si G cumple la primer condición de la Definición 9.1

$$\begin{aligned} AGA =& A(VS^gU^t)A = (USV^t)(VS^gU^t)(USV^t) \\ =& US(V^tV)S^g(U^tU)SV^t = USV^t = A. \end{aligned}$$

Aquí se utilizaron los hechos de que U y V son matrices ortogonales y de que S^g es una inversa generalizada de S. Luego G es una matriz A^{g_1}, miremos si es A^{g_2}

$$\begin{aligned} GAG =& (VS^gU^t)A(VS^gU^t) = (VS^gU^t)(USV^t)(VS^gU^t) \\ =& V(S^gSS^g)U^t = VS^gU^t = G. \end{aligned}$$

Ahora, observemos si G es una matriz A^{g_3}

$$AG = \left[A(VS^gU^t)\right] = \left[(USV^t)(VS^gU^t)\right] = \left[U(SS^g)U^t\right]. \tag{9.23}$$

Como

$$SS^g = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \tag{9.24}$$

la matriz AG es simétrica, por lo tanto G también es A^{g_3}. Observemos si cumple la cuarta condición dada en (9.2)

$$GA = \left[(VS^gU^t)A\right] = \left[(VS^gU^t)(USV^t)\right] = \left[V(S^gS)V^t\right]. \tag{9.25}$$

Usando de nuevo el hecho dado en (9.24), se tiene

$$(GA)^t = GA.$$

Así, G es la g-inversa de A y el teorema queda demostrado.

Ejemplo 9.12. Considérese la matriz dada en el Ejemplo 9.11, obtenga la g-inversa.

Solución

En este caso, los valores singulares de A son $\sigma_1^2 = 12$, $\sigma_2^2 = 3$ y $\sigma_3^2 = 0$. Al calcular los respectivos vectores propios normalizados de A^tA, se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Por otra parte, los respectivos vectores propios normalizados de la matriz AA^t son

$$\vec{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \vec{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Por lo tanto, si se establece

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}, \qquad V^t = \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

y

$$S = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \text{luego} \qquad S^g = \frac{1}{6} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$G = \frac{1}{6}\begin{bmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Después de realizar la multiplicación de las matrices queda

$$G = \frac{1}{12}\begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.11.

Corolario 9.10.1. Sea A una matriz real de tamaño $m \times n$ con valores singulares distintos de cero $\sigma_1, \sigma_2, \ldots, \sigma_r$ $(r = \rho(A))$. Entonces la g-inversa de A es la matriz definida de la siguiente forma:

$$G = VS^gU^t, \tag{9.26}$$

donde U y V son matrices ortogonales de tamaño $m \times m$ y $n \times n$, respectivamente y S^g es la inversa generalizada de la matriz dada en (3.31).

Demostración

Queda como ejercicio para el lector.

Ejemplo 9.13. Considérese la matriz dada en el Ejemplo 9.1, obtenga la g-inversa.

Solución

En este caso, los valores singulares de A son $\sigma_1^2 = 27$, $\sigma_2^2 = 1$. Al calcular los respectivos vectores propios normalizados de A^tA, se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Por otra parte, los respectivos vectores propios normalizados de la matriz AA^t son

$$\vec{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Por lo tanto, si se establece

$$U=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}, \qquad V^t=\begin{bmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{bmatrix}$$

y

$$S=\begin{bmatrix}\sqrt{27} & 0 & 0\\ 0 & 1 & 0\end{bmatrix}, \qquad \text{luego} \qquad S^g=\frac{1}{3\sqrt{3}}\begin{bmatrix}1 & 0\\ 0 & 3\sqrt{3}\\ 0 & 0\end{bmatrix}.$$

Así se obtiene

$$G=\frac{1}{3\sqrt{3}}\begin{bmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 3\sqrt{3}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}.$$

Después de multiplicar las matrices queda

$$G=\frac{1}{9}\begin{bmatrix}5 & -4\\ 4 & -5\\ 1 & 1\end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.6.

Teorema 9.11. Método de Penrose

Sea A una matriz real de tamaño $m\times n$ y rango $r\leq \text{mín}(m,n)$, particionada de la siguiente forma

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}, \tag{9.27}$$

con A_{11} una submatriz no singular de tamaño $r\times r$ y $A_{22}=A_{21}A_{11}^{-1}A_{12}$. Entonces la inversa generalizada de A es

$$G=\begin{bmatrix}A_{11}^tPA_{11}^t & A_{11}^tPA_{21}^t\\ A_{12}^tPA_{11}^t & A_{12}^tPA_{21}^t\end{bmatrix}, \tag{9.28}$$

donde $P=(A_{11}A_{11}^t+A_{12}A_{12}^t)^{-1}A_{11}\,(A_{11}^tA_{11}+A_{21}^tA_{21})^{-1}$.

Demostración

Como A_{11} es no singular, la matriz A se puede particionar como sigue

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A_{m\times n}} = \underbrace{\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}}_{=L_{m\times r}} \underbrace{\begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}}_{U_{r\times n}}.$$

En esta partición se usa el hecho de que $A_{22} = A_{21}A_{11}^{-1}A_{12}$. Luego,

$$A^g = \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g.$$

Por el Corolario 9.7.3 se tiene que

$$\begin{aligned} \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g &= \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^t \left\{ \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix} \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^t \right\}^{-1} \\ &= \begin{bmatrix} I_r \\ A_{12}^t (A_{11}^t)^{-1} \end{bmatrix} \begin{bmatrix} I_r + A_{11}^{-1}A_{12}A_{12}^t (A_{11}^t)^{-1} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \left(I_r + A_{11}^{-1}A_{12}A_{12}^t (A_{11}^t)^{-1} \right)^{-1} \\ A_{12}^t (A_{11}^t)^{-1} \left(I_r + A_{11}^{-1}A_{12}A_{12}^t (A_{11}^t)^{-1} \right)^{-1} \end{bmatrix}. \end{aligned}$$

Nótese que

$$\begin{aligned} \left[I_r + A_{11}^{-1}A_{12}A_{12}^t \left(A_{11}^t\right)^{-1} \right]^{-1} &= \left[A_{11}^{-1} \left(A_{11} + A_{12}A_{12}^t \left(A_{11}^t\right)^{-1} \right) \right]^{-1} \\ &= \left[A_{11} + A_{12}A_{12}^t \left(A_{11}^t\right)^{-1} \right]^{-1} A_{11} \\ &= \left[\left(A_{11}A_{11}^t + A_{12}A_{12}^t\right) \left(A_{11}^t\right)^{-1} \right]^{-1} A_{11} \\ &= A_{11}^t \left(A_{11}A_{11}^t + A_{12}A_{12}^t\right)^{-1} A_{11}. \end{aligned}$$

Por lo tanto,

$$\begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g = \begin{bmatrix} A_{11}^t (A_{11}A_{11}^t + A_{12}A_{12}^t)^{-1} A_{11} \\ A_{12}^t (A_{11}A_{11}^t + A_{12}A_{12}^t)^{-1} A_{11} \end{bmatrix}.$$

Por otra parte, por el Corolario 9.7.2 se tiene que

$$\begin{aligned} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g &= \left\{ \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^t \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \right\}^{-1} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^t \\ &= \begin{bmatrix} A_{11}^t A_{11} + A_{21}^t A_{21} \end{bmatrix}^{-1} \begin{bmatrix} A_{11}^t & A_{21}^t \end{bmatrix} \\ &= \begin{bmatrix} (A_{11}^t A_{11} + A_{21}^t A_{21})^{-1} A_{11}^t & (A_{11}^t A_{11} + A_{21}^t A_{21})^{-1} A_{21}^t \end{bmatrix}. \end{aligned}$$

Si se realizan los respectivos productos se llega a que la A^g es

$$A^g = \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g = \begin{bmatrix} A_{11}^t P A_{11}^t & A_{11}^t P A_{21}^t \\ A_{12}^t P A_{11}^t & A_{12}^t P A_{21}^t \end{bmatrix}, \tag{9.29}$$

con $P = (A_{11}A_{11}^t + A_{12}A_{12}^t)^{-1} A_{11} (A_{11}^t A_{11} + A_{21}^t A_{21})^{-1}$ y el teorema queda demostrado.

Corolario 9.11.1. Sea A una matriz cualquiera de rango r, particionada como en 9.27, con A_{22} una submatriz no singular de tamaño $r \times r$ y la submatriz $A_{11} = A_{12}A_{22}^{-1}A_{21}$. Entonces la inversa generalizada de A es

$$G = \begin{bmatrix} A_{21}^t P A_{12}^t & A_{21}^t P A_{22}^t \\ A_{22}^t P A_{12}^t & A_{22}^t P A_{22}^t \end{bmatrix}, \tag{9.30}$$

donde $P = (A_{22}A_{22}^t + A_{21}A_{21}^t)^{-1} A_{22} (A_{22}^t A_{22} + A_{12}^t A_{12})^{-1}$.

Demostración

Queda como ejercicio para el lector.

Antes de dar ejemplos, se proporciona el siguiente procedimiento que recoge el Método de Penrose para determinar la inversa generalizada

Procedimiento para determinar la inversa generalizada

Realice una partición de la matriz A como sigue

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

de tal manera que una de las submatrices A_{11} ó A_{22} sea cuadrada y con rango igual al de la matriz A.

I. Si A_{11} es la submatriz no singular

a) Verifique que $A_{22} = A_{21}A_{11}^{-1}A_{12}$.

b) Obtenga

$$P = \left(A_{11}A_{11}^t + A_{12}A_{12}^t\right)^{-1} A_{11} \left(A_{11}^t A_{11} + A_{21}^t A_{21}\right)^{-1}.$$

c) Forme la matriz

$$\begin{bmatrix} A_{11}^t P A_{11}^t & A_{11}^t P A_{21}^t \\ A_{12}^t P A_{11}^t & A_{12}^t P A_{21}^t \end{bmatrix}$$

este resultado es la inversa generalizada de A.

II. Si A_{22} es la submatriz no singular

a) Verifique que $A_{11} = A_{12}A_{22}^{-1}A_{21}$.

b) Obtenga

$$P = \left(A_{22}A_{22}^t + A_{21}A_{21}^t\right)^{-1} A_{22} \left(A_{22}^t A_{22} + A_{12}^t A_{12}\right)^{-1}.$$

c) Forme la matriz

$$\begin{bmatrix} A_{21}^t P A_{12}^t & A_{21}^t P A_{22}^t \\ A_{22}^t P A_{12}^t & A_{22}^t P A_{22}^t \end{bmatrix}$$

este resultado es la inversa generalizada de A.

Ejemplo 9.14. Calcule utilizando el método descrito anteriormente la inversa generalizada de la matriz dada en el Ejemplo 9.5.

Solución

Para la partición $(2+1)\times(2+1)$ de la matriz dada se tiene:

$$A = \begin{bmatrix} 4 & 1 & \vdots & 2 \\ 3 & 1 & \vdots & 3 \\ \cdots & \cdots & . & \cdots \\ 1 & 1 & \vdots & 5 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Luego A_{11} es no singular ya que su determinante es 1. El lector puede verificar que $A_{22} = A_{21}A_{11}^{-1}A_{12}$. Además

$$A_{11}A_{11}^t + A_{12}A_{12}^t = \begin{bmatrix} 17 & 13 \\ 13 & 10 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} = \begin{bmatrix} 21 & 19 \\ 19 & 19 \end{bmatrix}$$

y

$$A_{11}^t A_{11} + A_{21}^t A_{21} = \begin{bmatrix} 25 & 7 \\ 7 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 26 & 8 \\ 8 & 3 \end{bmatrix}.$$

Por lo tanto, P se obtiene como

$$P = \begin{bmatrix} 21 & 19 \\ 19 & 19 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 26 & 8 \\ 8 & 3 \end{bmatrix}^{-1} = \frac{1}{532}\begin{bmatrix} 57 & -152 \\ -55 & 156 \end{bmatrix}.$$

El lector puede realizar los otros productos y llegar a que

$$G = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix}.$$

La cual coincide con la obtenida en el Ejemplo 9.9.

Ejemplo 9.15. Considérese la matriz dada en el Ejemplo 9.11. Obtenga la g-inversa mediante el método descrito anteriormente.

Solución

Realizando una partición $(1+2)\times(1+2)$ a la matriz dada se tiene:

$$A=\begin{bmatrix}-1 & \vdots & 1 & -2\\ \cdots & . & \cdots & \cdots\\ 1 & \vdots & -1 & 2\\ -1 & \vdots & 1 & 1\end{bmatrix}=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}.$$

Luego A_{22} es no singular ya que su determinante es -3. El lector puede verificar que $A_{11}=A_{12}A_{22}^{-1}A_{21}$. También

$$A_{22}A_{22}^t+A_{21}A_{21}^t=\begin{bmatrix}5 & 1\\ 1 & 2\end{bmatrix}+\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}=\begin{bmatrix}6 & 0\\ 0 & 3\end{bmatrix}$$

y

$$A_{11}^tA_{11}+A_{21}^tA_{21}=\begin{bmatrix}2 & -1\\ -1 & 5\end{bmatrix}+\begin{bmatrix}1 & -2\\ -2 & 4\end{bmatrix}=\begin{bmatrix}3 & -3\\ -3 & 9\end{bmatrix}.$$

Por lo tanto, P se obtiene como

$$P=\begin{bmatrix}6 & 0\\ 0 & 3\end{bmatrix}^{-1}\begin{bmatrix}-1 & 2\\ 1 & 1\end{bmatrix}\begin{bmatrix}3 & -3\\ -3 & 9\end{bmatrix}^{-1}=\frac{1}{36}\begin{bmatrix}-1 & 1\\ 8 & 4\end{bmatrix}.$$

El lector puede realizar los otros productos y llegar a que

$$G=\frac{1}{12}\begin{bmatrix}-1 & 1 & -4\\ 1 & -1 & 4\\ -2 & 2 & 4\end{bmatrix}.$$

Nótese que esta matriz coincide con la obtenida en el Ejemplo 9.11.

9.4. Vectores y valores propios

Si A es una matriz no singular, por el Teorema 2.10 se sabe que los valores propios de A^{-1} son los recíprocos de los valores propios de A y los correspondientes vectores propios son los mismos. En esta sección se muestran las relaciones entre los valores y vectores propios de una matriz cuadrada y los asociados a la g-inversa.

Teorema 9.12. Sea G la g-inversa de A y λ un valor propio distinto de cero de A con vector propio correspondiente $\vec{v} \neq 0$. Entonces una condición suficiente para que $G\vec{v} = \lambda^{-1}\vec{v}$ es que $AG = GA$.

Demostración

Si $A\vec{v} = \lambda\vec{v}$, entonces premultiplicando por AG, se obtiene que

$$\begin{aligned}\underbrace{AGA}\vec{v} =& \lambda AG\vec{v}\\ A\ \ \vec{v} =& \lambda AG\vec{v}\\ \lambda\vec{v} =& \lambda AG\vec{v}.\end{aligned}$$

Como $\lambda \neq 0$, entonces

$$\begin{aligned}\vec{v} =& AG\vec{v} && \text{si } AG = GA\\ =& GA\vec{v} = \lambda G\vec{v}.\end{aligned}$$

Es decir, $G\vec{v} = \lambda^{-1}\vec{v}$.

Ejemplo 9.16. Determine los vectores y valores propios de la matriz g_2-inversa, obtenida en el Ejemplo 9.5.

Solución

En el Ejemplo 9.5, se obtuvo que

$$G = \frac{1}{441}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix}.$$

El polinomio característico de G es

$$p_G(\lambda) = -\frac{1}{21}\lambda(7\lambda - 1)(3\lambda - 1)$$

luego, los valores propios de G son $\lambda_1 = 0$, $\lambda_2 = \frac{1}{7}$ y $\lambda_3 = \frac{1}{3}$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} -7 \\ -3 \\ 5 \end{bmatrix}$, respectivamente.

Teorema 9.13. Si A es simétrica, los valores propios no nulos de A y A^{g_3} son recíprocos.

Demostración

Si $A\vec{v} = \lambda\vec{v}$, entonces premultiplicando por AA^{g_3} da

$$\begin{aligned}\underbrace{AA^{g_3}A}\vec{v} =& \lambda AA^{g_3}\vec{v}\\ A\ \vec{v} =& \lambda\left(AA^{g_3}\right)^t\vec{v}\\ \lambda\vec{v} =& \lambda\left(A^{g_3}\right)^t A\vec{v}.\end{aligned}$$

Como $\lambda \neq 0$, entonces

$$\left(A^{g_3}\right)^t\vec{v} = \lambda^{-1}\vec{v}.$$

El resultado se sigue, puesto que una matriz y su transpuesta tienen los mismos valores propios.

Ejercicios 9.1.

1. En los siguientes problemas determine la inversa generalizada con los métodos descritos en esta sección

   $a)$ $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. $b)$ $\begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & -2 \end{bmatrix}$. $c)$ $\begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 6 \end{bmatrix}$.

   $d)$ $\begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & -1 & -3 \end{bmatrix}$. $e)$ $\begin{bmatrix} 3 & -1 & 4 & 1 \\ -1 & 1 & -5 & -1 \\ 4 & -5 & 2 & 0 \end{bmatrix}$.

2. Encuentre una inversa generalizada para cada una de las siguientes matrices:

   $a)$ PAQ cuando P y Q son no singulares.

   $b)$ GA cuando G es una inversa generalizada de A.

   $c)$ kA cuando k es un escalar.

   $d)$ PAP^t cuando P es ortogonal y A es idempotente.

3. Sean A y X matrices simétricas tal que $AX = 0$, si X es idempotente y $A+X$ es no singular, pruebe que $(A+X)^{-1}$ es una inversa generalizada para A y X.

4. Para X particionada como $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ con X_1 de rango completo columna, pruebe que $X(X^tX)^gX^t = X_1(X_1^tX_1)^gX_1^t$.

9.5. Aplicaciones a la solución de sistemas de ecuaciones lineales

Entre las múltiples aplicaciones que tiene el concepto de inversa generalizada cabe señalar el papel que desempeña en el análisis y "solución" de sistemas lineales tanto consistentes como inconsistentes. En el caso de sistemas consistentes, dado que las soluciones existen permiten caracterizarlas y para sistemas inconsistentes posibilitan hallar soluciones aproximadas, ya que por definición este tipo de sistemas carece de soluciones. En esta sección se analiza con ayuda de la g-inversa de la matriz A, cuando (9.1) es consistente y cómo son sus soluciones.

Teorema 9.14. El sistema de ecuaciones lineales dado en (9.1) es consistente si y sólo si se verifica que

$$AA^{g_1}\vec{b} = \vec{b}. \tag{9.31}$$

Demostración

Si para A^{g_1}, se cumple que

$$AA^{g_1}\vec{b} = \vec{b}$$

entonces el sistema (9.1) es consistente, puesto que al menos $\vec{x}' = A^{g_1}\vec{b}$ es solución del mismo.

Recíprocamente, la condición es necesaria pues si el sistema, es consistente, existe $\vec{x}_0 \in \mathbb{R}^n$ tal que

$$A\vec{x}_0 = \vec{b}.$$

Ahora bien, dado que A^{g_1} siempre existe, premultiplicando la expresión anterior por AA^{g_1} se obtiene

$$\begin{aligned}\underbrace{AA^{g_1}A}\vec{x}_0 &= AA^{g_1}\vec{b}\\ \underbrace{A\quad\vec{x}_0} &= AA^{g_1}\vec{b}\\ \vec{b}\quad &= AA^{g_1}\vec{b}\end{aligned}$$

lo cual prueba el teorema.

Teorema 9.15. Dado un sistema consistente

$$A\vec{x} = \vec{b},$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se verifica que

i) Para todo $\vec{d} \in \mathbb{R}^n$,

$$\vec{x}_0 = A^{g_1}\vec{b} + (I_n - A^{g_1}A)\,\vec{d} \tag{9.32}$$

es solución del sistema.

ii) Si $\vec{x}'$ es una solución cualquiera del sistema, existe $\vec{d} \in \mathbb{R}^n$ tal que $\vec{x}'$ puede expresarse en la forma dada en (9.32).

Demostración

i) Sea $\vec{x}_0$ una solución del sistema, es decir

$$A\vec{x}_0 = \vec{b}.$$

Entonces por el Teorema 9.14 y la definición de A^{g_1}, se tiene que

$$A\vec{x}_0 = AA^{g_1}\vec{b} + A\,(I_n - A^{g_1}A)\,\vec{d} = \vec{b} + A\vec{d} - A\vec{d}.$$

ii) Si $\vec{x}'$ es una solución cualquiera del sistema se verifica que

$$\vec{b} - A\vec{x}' = \vec{0}.$$

Si se premultiplica por A^{g_1} y se suma $\vec{x}'$ a ambos lados, se tiene que

$$\begin{aligned}A^{g_1}\vec{b} - A^{g_1}A\vec{x}' + \vec{x}' &= \vec{x}'\\ A^{g_1}\vec{b} + (I_n - A^{g_1}A)\,\vec{x}' &= \vec{x}'.\end{aligned}$$

Luego, tomando en el lado izquierdo $\vec{d} = \vec{x}'$, se obtiene lo que se deseaba.

Teorema 9.16. Dado el sistema consistente

$$A\vec{x} = \vec{b},$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se verifica que existe solución única $\vec{x}'$ si y sólo si $A^g A = I_n$ siendo A^g la $g-$inversa de la matriz A.

Demostración

Queda como ejercicio para el lector.

Ejemplo 9.17. Determine una solución del sistema de ecuaciones lineales

$$\begin{aligned} x_1 + 6x_2 + 11x_3 =& 0 \\ 6x_1 + 46x_2 + 86x_3 =& 20 \\ 11x_1 + 86x_2 + 161x_3 =& 40. \end{aligned} \tag{9.33}$$

Solución

Si se reescribe (9.33) se llega a

$$\begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \\ 40 \end{bmatrix}. \tag{9.34}$$

El sistema dado en (9.33), se puede resolver usando inversa generalizada y usando cualquiera de los métodos descritos en este capítulo. De esta manera, se tiene que

$$A^g = \frac{1}{180} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix}.$$

Veamos si el sistema de ecuaciones (9.33) es consistente determinado,

$$\begin{aligned} A^g A =& \frac{1}{180} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix} \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \\ =& \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \neq I_3. \end{aligned}$$

Por el Teorema 9.16 el sistema es consistente indeterminado, luego una solución de (9.33) es

$$\begin{aligned}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} &= \frac{1}{180}\begin{bmatrix}517 & 190 & -137\\190 & 70 & -50\\-137 & -50 & 37\end{bmatrix}\begin{bmatrix}0\\20\\40\end{bmatrix}\\ &= \frac{1}{9}\begin{bmatrix}517 & 190 & -137\\190 & 70 & -50\\-137 & -50 & 37\end{bmatrix}\begin{bmatrix}0\\1\\2\end{bmatrix} = -\frac{1}{3}\begin{bmatrix}28\\10\\-8\end{bmatrix}.\end{aligned} \tag{9.35}$$

Teniendo en cuenta el Teorema 9.15, para todo $\vec{d} = (\alpha,\beta,\gamma) \in \mathbb{R}^3$ el vector

$$\vec{x}_0 = A^g\vec{b} + (I_n - A^gA)\,\vec{d} = \begin{bmatrix}-\frac{28}{3}+\frac{1}{6}\alpha-\frac{1}{3}\beta+\frac{1}{6}\gamma\\ -\frac{10}{3}-\frac{1}{3}\alpha+\frac{2}{3}\beta-\frac{1}{3}\gamma\\ \frac{8}{3}+\frac{1}{6}\alpha-\frac{1}{3}\beta+\frac{1}{6}\gamma\end{bmatrix},$$

es una solución del sistema, como el lector puede comprobar fácilmente sustituyendo $\vec{x}_0$ en (9.33).

Teorema 9.17. Dado el sistema de ecuaciones lineales

$$A\vec{x} = \vec{b}$$

en donde A es una matriz de tamaño $m \times n$ con rango r, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$. Considérese la función residuo

$$\boldsymbol{r}\,(\vec{x}) = A\vec{x} - \vec{b}.$$

para cualquier $\vec{x}$. Entonces $\vec{x}' = A^g\vec{b}$ es una *solución aproximada mínimo cuadrática* (*LS*, por sus siglas en inglés) del sistema si y sólo si mínimiza a $\|\boldsymbol{r}\,(\vec{x})\|$.

Demostración

Si se reescribe la norma euclídea de $\boldsymbol{r}\,(\vec{x})$ se obtiene que

$$\begin{aligned}\|\boldsymbol{r}\,(\vec{x})\| = \underbrace{\boldsymbol{r}\,(\vec{x})^t\,\boldsymbol{r}\,(\vec{x})}_{F\,(\vec{x})} &= \left(A\vec{x}-\vec{b}\right)^t\left(A\vec{x}-\vec{b}\right) = \left(\vec{x}^tA^t-\vec{b}^t\right)\left(A\vec{x}-\vec{b}\right)\\ &= \vec{x}^tA^tA\vec{x} - \vec{b}^tA\vec{x} - \vec{x}^tA^t\vec{b} + \vec{b}^t\vec{b}.\end{aligned}$$

para determinar la $\vec{x}$ que minimiza esta suma de cuadrados de los residuos, se calculan las derivadas parciales[2] de $F(\vec{x})$ con respecto a $\vec{x}$

$$\frac{\partial F(\vec{x})}{\partial \vec{x}} = 2A^tA\vec{x} - 2A^t\vec{b} = \vec{0}. \tag{9.36}$$

Al despejar $\vec{x}$, se obtiene un mínimo global de $F(\vec{x})$ pues

$$\frac{\partial^2 F(\vec{x})}{\partial \vec{x}^2} = 2A^tA,$$

la cual es una matriz definida positiva si A es de rango completo, o semidefinida positiva en caso contrario y por ello, en ambas situaciones, $F(\vec{x})$ es una función convexa.

Luego, si se sustituye $\vec{x}'$ en (9.36) se tiene

$$A^tAA^g\vec{b} - A^t\vec{b} = \vec{0} \qquad \text{o} \qquad \left(A^tAA^g - A^t\right)\vec{b} = \vec{0}$$

pero como $\vec{b}$ es cualquier vector de $\mathbb{R}^m$ esto equivale a la condición

$$A^tAA^g = A^t,$$

Esto se sigue inmediatamente del Teorema 9.3 parte c).

Definición 9.2. Dado el sistema lineal $A\vec{x} = \vec{b}$ y la función residuo $\boldsymbol{r}(\vec{x}) = A\vec{x} - \vec{b}$, se dice que

i) $\vec{x}'$ es una *solución aproximada mínimo cuadrática* (LS, por sus siglas en inglés) del sistema si y sólo si para todo $\vec{x} \in \mathbb{R}^n$ se verifica que

$$\boldsymbol{r}\left(\vec{x}'\right)^t \boldsymbol{r}\left(\vec{x}'\right) \leq \boldsymbol{r}(\vec{x})^t \boldsymbol{r}(\vec{x}).$$

ii) Una solución aproximada mínimo cuadrática $\vec{x}'$ es de *norma mínima* ($MNLS$, por sus siglas en inglés), si y sólo si para todo $\vec{x} \in \mathbb{R}^n$ se cumple que

$$\boldsymbol{r}\left(\vec{x}'\right)^t \boldsymbol{r}\left(\vec{x}'\right) = \boldsymbol{r}(\vec{x})^t \boldsymbol{r}(\vec{x}).$$

[2] Si el lector desea consultar técnicas de derivación matricial puede ver Barbolla (1998), cap. 5

En el siguiente resultado se recogen dos características de las soluciones mínimo cuadráticas LS para sistemas inconsistentes a partir de la g_3-Inversa de la matriz de coeficientes del sistema.

Teorema 9.18. Dado el sistema de ecuaciones lineales inconsistente

$$A\vec{x} = \vec{b}$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se verifica que

(i) $\vec{x}^{\prime} = G\vec{b}$ es una solución LS del sistema si y sólo si G es una A^{g_3}

(ii) La solución $\vec{x}_0$ dada en (9.32) es una solución LS del sistema si y sólo si $\vec{x}^{\prime}$ es solución del sistema consistente

$$A\vec{x} = AA^{g_3}\vec{b}$$

Demostración

(i) Por el Teorema 9.17, una solución LS del sistema es de la forma

$$\vec{x} = A^{g}\vec{b}.$$

Falta entonces comprobar qué condiciones de las dadas en (9.2), debe cumplir G si $\vec{x}^{\prime} = G\vec{b}$ es solución LS del sistema y por tanto, solución de (9.36). Así sustituyendo $\vec{x}^{\prime}$ en (9.36) se tiene

$$A^tAG\vec{b} - A^t\vec{b} = \vec{0}$$

y como $\vec{b}$ es cualquier vector de $\mathbb{R}^m$ esto equivale a la condición

$$A^tAG = A^t,$$

que se verifica si y sólo si G es una g_3-Inversa de A, ya que

$$AG \text{ es simétrica} \qquad \text{y} \qquad A^tG^tA^t = A^t.$$

(ii) Si el sistema
$$A\vec{x} = AA^{g_3}\vec{b}$$
es consistente, entonces por el Teorema 9.15 sus soluciones son de la forma
$$\vec{x}' = A^{g_1}\left(AA^{g_3}\vec{b}\right) + (I_n - A^{g_1}A)\,\vec{d},$$
para cualquier $\vec{d} \in \mathbb{R}^m$ o también
$$\vec{x}' = A^{g_3}\left(AA^{g_3}\vec{b}\right) + (I_n - A^{g_3}A)\,\vec{d} = A^{g_3}\vec{b} + (I_n - A^{g_3}A)\,\vec{d}$$
dado que cualquier g_3-Inversa de A es a su vez g_1-Inversa.

La demostración concluye si se muestra que $\vec{x}'$ es una solución LS del sistema $A\vec{x} = \vec{b}$. Para ello, razonando como en (i), $\vec{x}'$ debe ser solución de (9.36), lo cual es válido, pues de acuerdo con la definición de la g_3-Inversa
$$A^tA\left[A^{g_3}\vec{b} + (I_n - A^{g_3}A)\,\vec{d}\right] - A^t\vec{b} = A^tAA^{g_3}\vec{b} - A^t\vec{b}$$
y como $AA^{g_3} = (A^{g_3})^t A^t$ dada la simetría de AA^{g_3} se obtiene finalmente que
$$A^t\,(A^{g_3})^t A^t\vec{b} - A^t\vec{b} = \vec{0}.$$

Teorema 9.19. Dado el sistema inconsistente
$$A\vec{x} = \vec{b},$$
con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, su solución $MNLS$ es única y está dada por
$$\vec{x}' = A^g\vec{b}. \tag{9.37}$$

Demostración

Si A^g es la $g-$Inversa de A, también es g_3-Inversa de A y, por ello, en virtud del teorema anterior, está garantizado que
$$\vec{x}' = A^g\vec{b}$$
es solución LS del sistema.

Bastará, por tanto, probar que $\vec{x}'$ es única y de mínima norma.

- $\vec{x}^{\prime}$ es *MNLS*.

  En efecto, por ser $\vec{x}^{\prime}$ *LS* es solución del sistema
$$A\vec{x} = AA^g\vec{b}$$
  o equivalentemente
$$A^gA\vec{x} = A^gAA^g\vec{b} = A^g\vec{b}. \tag{9.38}$$
  Por lo tanto, cualquier solución *LS* es de la forma
$$\vec{x} = A^g\vec{b} + (I_n - A^gA)\,\vec{x}$$
  el cuadrado de su norma euclídea es
$$\vec{x}^t\vec{x} = \vec{b}^t\,(A^g)^t\,A^g\vec{b} + (\vec{x} - A^gA\vec{x})^t\,(\vec{x} - A^gA\vec{x})\,.$$
  Si se sustituye (9.38) en esta expresión resulta
$$\begin{aligned}\vec{x}^t\vec{x} =& \vec{b}^t\,(A^g)^t\,A^g\vec{b} + \left(\vec{x} - A^g\vec{b}\right)^t\left(\vec{x} - A^g\vec{b}\right)\\ =& \left\|A^g\vec{b}\right\|^2 + \left\|\vec{x} - A^g\vec{b}\right\|^2,\end{aligned}$$
  de donde
$$\vec{x}^t\vec{x} \geq \left(\vec{x}^{\prime}\right)^t\left(\vec{x}^{\prime}\right) = \left\|A^g\vec{b}\right\|^2$$
  cuando $\vec{x}^{\prime} = A^g\vec{b}$.

- $\vec{x}^{\prime}$ es única

  Supóngase que $\vec{x}_0 \in \mathbb{R}^n$ es también solución *MNLS* del sistema. Entonces $\vec{x}_0$ cumple que
$$(\vec{x}_0)^t\,(\vec{x}_0) = \left\|A^g\vec{b}\right\|^2 + \left\|\vec{x}_0 - A^g\vec{b}\right\|^2 = \left(\vec{x}^{\prime}\right)^t\left(\vec{x}^{\prime}\right) = \left\|A^g\vec{b}\right\|^2.$$
  Por lo tanto,
$$\left\|\vec{x}_0 - A^g\vec{b}\right\|^2 = 0.$$
  Es decir,
$$\vec{x}_0 - A^g\vec{b} = \vec{0} \qquad\qquad \Leftrightarrow \qquad\qquad \vec{x}_0 = A^g\vec{b}.$$
  En consecuencia,
$$\vec{x}_0 = \vec{x}^{\prime}.$$

Esto es lo que se quería demostrar.

Ejercicios 9.2.

1. Encuentre una solución para cada uno de los sistemas de ecuaciones lineales dados a continuación

$a)$
$$\begin{aligned} 3x_1 - 2x_2 - x_3 &= 1 \\ -x_1 + 2x_2 + 2x_3 &= 2 \\ x_1 + 2x_2 + 3x_3 &= 4. \end{aligned}$$

$b)$
$$\begin{aligned} x_1 - x_2 - x_3 &= 2 \\ 2x_1 + x_2 + 2x_3 &= 4 \\ x_1 - 4x_2 - 5x_3 &= 2. \end{aligned}$$

$c)$
$$\begin{aligned} 2x_1 + 6x_2 - 4x_3 + 2x_4 &= 4 \\ x_1 \qquad - x_3 + x_4 &= 5 \\ -3x_1 + 2x_2 - 2x_3 \qquad &= -2. \end{aligned}$$

$d)$
$$\begin{aligned} x_1 - 2x_2 + x_3 + x_4 &= 2 \\ 3x_1 \qquad + 2x_3 - 2x_4 &= -8 \\ 4x_2 - x_3 - x_4 &= 1 \\ 5x_1 \qquad + 3x_3 - x_4 &= -3. \end{aligned}$$

2. Sea X una matriz de tamaño $m \times n$ $(m > n)$ y rango $r < n$. Sea G una inversa generalizada de X^tX defina:

$$\vec{b} = GX^t\vec{Y}, \qquad s^2 = (\vec{Y} - X\vec{b})^t(\vec{Y} - X\vec{b})$$
$$\vec{b}_0 = \vec{b} - GQ(Q^tGQ)^{-1}(Q^t\vec{b} - \vec{m}), \qquad \text{con} \quad Q = \left(GX^tXG^t\right)^t X.$$

Pruebe que

$a)$ $s^2 = \vec{Y}^t\vec{Y} - \vec{b}^tX^t\vec{Y}$.

$b)$ $Q^t\vec{b}_0 = \vec{m}$.

$c)$ $(\vec{Y} - X\vec{b}_0)^t(\vec{Y} - X\vec{b}_0) = s^2 + (Q^t\vec{b} - \vec{m})^t(Q^tGQ)^{-1}(Q^t\vec{b} - \vec{m})$.

Capítulo 10

Aplicaciones

En este capítulo se recopila algunos desarrollos teóricos de la Estadística que el lector que esté interesado en profundizar puede consultar textos de Modelos Lineales, Estadística Multivariada (o cursar las asignaturas correspondientes). El propósito de este capítulo es ilustrar la utilidad de la mayor cantidad posible de los conceptos tratados en este escrito y por eso en esta parte de aplicaciones se omiten tanto conceptos básicos del área de la Estadística, como aquellos temas avanzados que el lector aprenderá posteriormente.

10.1. Matrices estocásticas

Las matrices estocásticas corresponden a un tipo especial de matrices definidas positivas y se usan con frecuencia en el estudio de fenómenos aleatorios, en Teoría de la Probabilidad y Estadística.

Definición 10.1. Una matriz $A = [a_{ij}]$ de tamaño $n \times n$ se dice que es estocástica por filas (columnas) si todos sus elementos son números reales no negativos y la suma de los elementos de cada una de sus filas (columnas) es igual a 1. Es decir

$$0 \leq a_{ij} \leq 1 \qquad\qquad i,j = 1,2,\ldots,n$$

y además

$$\sum_{j=1}^{n} a_{ij} = 1 \qquad\qquad \text{Si } A \text{ es estocástica por filas.}$$

$$\sum_{i=1}^{n} a_{ij} = 1 \qquad\qquad \text{Si } A \text{ es estocástica por columnas.}$$

Se dice que A es doblemente estocástica cuando es estocástica tanto por filas como por columnas.

Teorema 10.1.

Si A y B son estocásticas (doblemente estocásticas) se verifica

$i)$ AB es estocástica (doblemente estocástica).

$ii)$ Para todo $k \in \mathbb{N}$, A^k es estocástica (doblemente estocástica).

$iii)$ Cuando A es doblemente estocástica, entonces A^t también lo es.

Teorema 10.2.

Si A es una matriz estocástica por filas (columnas) entonces $\lambda = 1$ es uno de sus valores propios.

Demostración

Sea A una matriz de tamaño $n \times n$ tal que A es una matriz estocástica por columnas. Basta probar que

$$\det(A - I) = 0$$

para ello veamos que las filas de la matriz $A - I$ no son linealmente independientes. Si $B = A - I$, consideremos la suma vectorial de las filas de la matriz B

$$\vec{B}_1^t + \vec{B}_2^t + \ldots \vec{B}_n^t = \begin{bmatrix} a_{11}-1 \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix} + \begin{bmatrix} a_{21} \\ a_{22}-1 \\ \vdots \\ a_{2n} \end{bmatrix} + \ldots + \begin{bmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nn}-1 \end{bmatrix}.$$

Es decir,

$$\vec{B}_1^t+\vec{B}_2^t+\ldots\vec{B}_n^t=\begin{bmatrix}a_{11}+a_{21}+\ldots+a_{n1}-1\\a_{12}+a_{22}+\ldots+a_{n2}-1\\\vdots\\a_{1n}+a_{2n}+\ldots+a_{nn}-1\end{bmatrix} \tag{10.1}$$

como A es una matriz estocástica por columnas, las entradas de cada columna de A suman uno y por lo tanto

$$\sum_{i=1}^{n} a_{ij} = 1 \qquad\qquad \forall j$$

luego (10.1), se transforma en $\vec{B}_1^t+\vec{B}_2^t+\ldots+\vec{B}_n^t=\vec{0}$, es decir, se encontro una combinación lineal no trivial de las filas de $B=A-I$ que producen el vector cero de $\mathbb{R}^n$. Por lo tanto, las filas de $A-I$ son linealmente dependientes, luego la matriz $A-I$ es singular, es decir, $\det(A-I)=0$ y entonces $\lambda=1$ es un valor propio de A.

Definición 10.2. Matriz regular

Una matriz estocástica A se dice regular si todas las componentes de al menos una de sus potencias A^k (k entero positivo) son estrictamente positivas (mayores que cero).

Definición 10.3. Cadena de Markov

Una cadena de Markov o proceso de Markov es un proceso en el cual la probabilidad de que el sistema esté en un estado particular en un periodo dado de observación depende solamente de su estado en el periodo de observación inmediatamente anterior.

Definición 10.4. Probabilidad de transición

Se define la probabilidad de transición p_{ij} $(i,j=1,2,\ldots,n)$, como la probabilidad de que el sistema pase del estado j al estado i en la siguiente observación.

Definición 10.5. Matriz de transición

A cada cadena de Markov se le puede asignar una única matriz de transición P, cuyos elementos son las probabilidades p_{ij}. Esta matriz es cuadrada y su dimensión depende del número posible de estados, la matriz P resulta ser estocástica.

Definición 10.6. Vector de probabilidad

Un vector de probabilidad es un vector columna, con entradas no negativas, en el que la suma de sus elementos es igual a la unidad. Se dice que los vectores de probabilidad $X^{(n)}$ para $n = 0, 1, \ldots$ son los vectores de estado de un proceso de Markov, si la componente de orden i, $p_i^{(n)}$ de $X^{(n)}$, es la probabilidad de que el sistema esté en el estado i cuando se hace la observación n.

Teorema 10.3. Si P es la matriz de transición de un proceso de Markov y $X^{(n)}$ es el vector columna de la observación n, se tendrá que:

$$X^{(n)} = \begin{cases} PX^{(n-1)} & \text{Si } P \text{ es estocástica por columnas} \\ P^t X^{(n-1)} & \text{Si } P \text{ es estocástica por filas} \end{cases} \tag{10.2}$$

La ecuación (10.2) implica

$$\begin{aligned} X^{(1)} =& PX^{(0)} \\ X^{(2)} =& PX^{(1)} = P\left(PX^{(0)}\right) = P^2 X^{(0)} \\ X^{(3)} =& PX^{(2)} = P\left(P^2 X^{(0)}\right) = P^3 X^{(0)} \end{aligned}$$

y, en general,

$$X^{(n)} = P^n X^{(0)}. \tag{10.3}$$

Así, la matriz de transición y el vector de estados inicial $X^{(0)}$ determinan completamente los demás vectores de estado.

Definición 10.7. Un proceso de Markov es regular si su matriz de transición es una matriz estocástica regular.

Teorema 10.4.

Si P es una matriz de transición regular de tamaño $n \times n$. Entonces, cuando $n \to \infty$, P^n tiende a una matriz R de tamaño $n \times n$, de la forma

$$R = [\vec{v} \ \ \vec{v} \ \ \ldots \ \ \vec{v}]$$

donde $\vec{v}$ es un vector de probabilidad de tamaño $n \times 1$, con todos sus elementos mayores que cero.

Demostración

El lector puede consultarla en Kemeny (1976).

Teorema 10.5. Si P es una matriz de transición regular de tamaño $n \times n$ y R y $\vec{v}$ son como en el Teorema 10.4, entonces

(i) Para cualquier vector $X^{(0)}$ de probabilidad inicial, $P^n X^{(0)}$ tiende a $\vec{v}$ cuando aumenta n, esto es,

$$\lim_{n\to\infty}\left(P^n X^{(0)}\right) = \vec{v}.$$

Es decir, todo proceso regular de Markov tiene un vector estacionario $\vec{v}$.

(ii) El vector estacionario $\vec{v}$ es el único vector de probabilidad que satisface la ecuación

$$P\vec{v} = \vec{v}, \qquad \text{o} \qquad (P - I)\vec{v} = \vec{0}.$$

Luego, $\vec{v}$ es un vector propio de P asociado al valor propio $\lambda = 1$.

Ejemplo 10.1. Una empresa de investigación de mercado estudia un grupo de consumidores de café, los cuales compran una lata del grano cada semana. Las personas que actualmente toman la marca A, la comprarán de nuevo la próxima semana con una probabilidad de 0,50, cambiarán a la marca B con una probabilidad de $0{,}25$ y comprarán la marca D con una probabilidad de $0{,}25$. De las personas que ahora consumen la marca B, preferirán la próxima semana la marca A, B o D con probabilidades de $0{,}60, 0{,}30, 0{,}10$ respectivamente. Ahora, de las personas que en la actualidad compran la marca D adquirirán la próxima semana la marca A, B o D con probabilidades de $0{,}30, 0{,}40, 0{,}30$. Suponga que al iniciar el estudio, la marca A tiene el 20% del mercado, la marca B tiene el 20% y la otra marca el 60%. ¿A la larga cual será el porcentaje del mercado que tendrán las marcas A, B y D?.

Solución

Si se aborda el problema por medio de las Cadenas de Markov.

$$P = \begin{array}{c} \begin{array}{ccc} A & B & D \end{array} \\ \begin{bmatrix} 0{,}50 & 0{,}60 & 0{,}40 \\ 0{,}25 & 0{,}30 & 0{,}30 \\ 0{,}25 & 0{,}10 & 0{,}30 \end{bmatrix} \end{array} \begin{array}{c} A \\ B \\ D \end{array}$$

Como P es estocástica por columnas, al calcular los valores propios de P se obtienen $\lambda_1 = 1$, $\lambda_2 = \frac{1}{20} - \frac{1}{20}i\sqrt{3}$ y $\lambda_3 = \frac{1}{20} + \frac{1}{20}i\sqrt{3}$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + i \begin{bmatrix} \sqrt{3} \\ 0 \\ -\sqrt{3} \end{bmatrix}$, respectivamente.

Esto implica que la matriz P no es diagonalizable, sin embargo

$$P^n = \begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & -\frac{1}{20}\sqrt{3} \\ 0 & \frac{1}{20}\sqrt{3} & \frac{1}{20} \end{bmatrix}^n \begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix}^{-1}.$$

Luego,

$$\lim_{n\to\infty} P^n = \frac{1}{273} \begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & \frac{1}{20}\sqrt{3} \\ 0 & -\frac{1}{20}\sqrt{3} & \frac{1}{20} \end{bmatrix}^n \begin{bmatrix} 3 & 3 & 3 \\ -75 & 198 & -75 \\ 20\sqrt{3} & 20\sqrt{3} & -71\sqrt{3} \end{bmatrix}.$$

Aquí, las potencias de la forma $3n$ cumplen que

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & -\frac{1}{20}\sqrt{3} \\ 0 & \frac{1}{20}\sqrt{3} & \frac{1}{20} \end{bmatrix}^{3n} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{(-1)^n}{10^{3n}} & 0 \\ 0 & 0 & \frac{(-1)^n}{10^{3n}} \end{bmatrix},$$

y dado que $\left(\frac{-1}{10^3}\right)^n$ tiende a cero cuando $n \to \infty$,

$$\lim_{n\to\infty} P^{3n} = \frac{1}{91}\begin{bmatrix} 46 & 46 & 46 \\ 25 & 25 & 25 \\ 20 & 20 & 20 \end{bmatrix}.$$

Nótese que $\lim\limits_{n\to\infty} P^n$ da como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$; para convertir el vector propio $\vec{v}_1 = \begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}$, en un vector de probabilidad, se divide cada una de sus componentes por la suma de todos sus elementos, es decir

$$\vec{v} = \frac{1}{91}\vec{v}_1 = \frac{1}{91}\begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}.$$

Por lo tanto, el vector de estados a largo plazo es

$$\lim_{n\to\infty} X^{(n+1)} = \lim_{n\to\infty} P^n X^{(0)} = \frac{1}{91}\begin{bmatrix} 46 & 46 & 46 \\ 25 & 25 & 25 \\ 20 & 20 & 20 \end{bmatrix}\begin{bmatrix} \frac{1}{5} \\ \frac{1}{5} \\ \frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{46}{91} \\ \frac{25}{91} \\ \frac{20}{91} \end{bmatrix}$$

Entonces, se puede decir que a largo plazo, la marca A tendrá el control de cerca del $\frac{46}{91} \approx 50{,}55\,\%$ del mercado, la marca B el $\frac{25}{91} \approx 27{,}47\,\%$ del mercado y la otra marca el $\frac{20}{91} \approx 21{,}98\,\%$ del mercado.

Ejemplo 10.2. Supóngase que el clima en cierta ciudad es, bueno, regular o malo. Si el clima es bueno hoy, será bueno mañana con una probabilidad de $0{,}60$, será regular con una probabilidad de $0{,}20$ y será malo con una probabilidad de $0{,}20$. Si el clima es regular hoy, será bueno, regular o malo con probabilidades de $0{,}25, 0{,}50, 0{,}25$ respectivamente. Ahora si el clima es malo hoy mañana será bueno, regular o malo con probabilidades de $0{,}25, 0{,}25, 0{,}50$. ¿A la larga cual será el porcentaje de días buenos, regulares, malos?

Solución

Si se aborda el problema por medio de las Cadenas de Markov con

$$P = \begin{array}{c} \\ B \\ R \\ M \end{array} \begin{array}{c} \begin{array}{ccc} B & R & M \end{array} \\ \begin{bmatrix} 0{,}60 & 0{,}20 & 0{,}20 \\ 0{,}25 & 0{,}50 & 0{,}25 \\ 0{,}25 & 0{,}25 & 0{,}50 \end{bmatrix} \end{array}.$$

Como P es estocástica por filas, entonces se transpone

$$P^t = \begin{bmatrix} \frac{3}{5} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

En este caso, los valores propios de P^t son $\lambda_1 = 1$, $\lambda_2 = \frac{7}{20}$ y $\lambda_3 = \frac{1}{4}$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, respectivamente.

Luego, la diagonalización de P^t da como resultado

$$\left(P^t\right)^n = \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{7}{20} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}^n \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix}^{-1}.$$

Dado que $\left(\frac{7}{20}\right)^n$ y $\left(\frac{1}{4}\right)^n$ tienden a cero cuando $n \to \infty$,

$$\lim_{n\to\infty} \left(P^t\right)^n = \frac{1}{26} \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 2 \\ -8 & 5 & 5 \\ 0 & -13 & 13 \end{bmatrix} = \frac{1}{13} \begin{bmatrix} 5 & 5 & 5 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

Nótese que $\lim\limits_{n\to\infty} \left(P^t\right)^n$ da como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$, para convertir el vector propio $\vec{v}_1 = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, en un vector de probabilidad, se divide cada una de sus componentes por la suma de todos sus elementos, es decir

$$\vec{v} = \frac{1}{13}\vec{v}_1 = \frac{1}{13} \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}.$$

Entonces, se puede decir que a la larga, el clima será bueno el $\frac{5}{13} \approx 38{,}46\,\%$ de los días, regular el $\frac{4}{13} \approx 30{,}77\,\%$ de los días y malo el $\frac{4}{13} \approx 30{,}77\,\%$ de los días.

10.2. Modelos Genéticos

La relación entre las Matemáticas y la Biología forma parte de un problema antiguo en la historia de las ciencias. Esta sección se ha dedicado a una de las aplicaciones de métodos de modelación matemática en Genética. Para desarrollar este tipo de aplicación únicamente es requisito conocer el proceso para diagonalizar una matriz desde el punto de vista matemático ya que también se tendrán que manejar términos propios utilizados dentro del desarrollo de la Genética. Para una mejor comprensión se darán a conocer algunas nociones básicas de este tema, se comienza con una reseña histórica muy breve de cómo inicio la Genética y de cómo desde el principio estuvo muy ligada a la Estadística, además de una explicación sencilla de que es la Genética en cuanto a lo que al lector le interesa.

La ciencia de la Genética nació en 1900, cuando varios investigadores de la reproducción de las plantas descubrieron el trabajo del monje austríaco Gregor Mendel, que aunque fue publicado en 1865 había sido ignorado en la práctica. Mendel, quien trabajó con la planta del guisante (o chícharo), describió los patrones de la herencia en función de siete pares de rasgos contrastantes que aparecían en siete variedades diferentes de esta planta. Observó que los caracteres se heredaban como unidades separadas y cada una de ellas lo hacía de forma independiente con respecto a las otras. Señaló que cada progenitor tiene pares de unidades, pero que sólo aporta una unidad de cada pareja a su descendiente. Más tarde, las unidades descritas por Mendel recibieron el nombre de genes. Su publicación era el resultado de cerca de 10 años de observaciones minuciosas, las cuales expresó matemáticamente mediante las leyes de la probabilidad y así predijo los resultados de los cruces genéticos, datos que se escriben en fracciones o porcentajes.

La observación obtenida al cruzar dos plantas puras con diferentes caracteres llevó a Mendel a deducir que existía un rasgo más fuerte, al que llamo *dominante* y al rasgo más débil o que aparentemente desaparece, le dio el nombre de *recesivo*. Estos dos conceptos de rasgo dominante y recesivo, aunque muy fáciles de comprender, son de vital importancia a la hora de desarrollar la parte matemática de esta rama.

En la historia de la Biología este ha sido uno de los experimentos más extensos que ha realizado un solo autor. La recepción que tuvo esta publicación fue prácticamente nula entre la comunidad científica de su época. Casi cuatro décadas después, las leyes de Mendel fueron redescubiertas. A partir de entonces comenzó el desarrollo impetuoso de la Genética. Aún cuando sea un simplismo, el haber ignorado por parte de la comunidad científica las leyes de Mendel, ha costado cuarenta años de retraso a la Biotecnología moderna.

Un gene particular puede ocurrir en varias formas o *alelos*. Para simplificar, consideraremos un gene con dos alelos, los genetistas denotan los caracteres dominantes con letras mayúsculas y los caracteres recesivos, con minúsculas. De esta manera los alelos serán A y a.

10.2.1. Herencia autosómica

En esta sección se considera la herencia como *autosómica*, esto quiere decir que un individuo hereda un gene de cada uno de los genes de sus padres formando así su propio par. Hasta donde se sabe, es el azar el que determina cuál de los dos genes de un progenitor pasa a su descendiente. Si fuera posible clasificar los individuos de una población de una especie dada en cuanto a los genotipos AA, Aa y aa (téngase en cuenta que el genotipo Aa es igual que el aA) sería posible determinar las proporciones de los alelos en la población. Esto no sería factible si, por ejemplo, no se pudieran distinguir AA de Aa.

Para $n = 0, 1, 2, \ldots,$ se establece que

p_n = Proporción del genotipo AA que hay en la generación de orden n,

q_n = Proporción del genotipo Aa que hay en la generación de orden n,

r_n = Proporción del genotipo aa que hay en la generación de orden n.

Si se supone que se pueden determinar esas proporciones, nótese que se debe tener

$$p_n + q_n + r_n = 1. \tag{10.4}$$

Entonces, las proporciones u y v de los dos alelos A y a en la población satisfacen las ecuaciones

$$u = p_n + \frac{1}{2}q_n \qquad \text{y} \qquad v = \frac{1}{2}q_n + r_n. \tag{10.5}$$

Aquí, se usó el hecho de que los alelos A y a constituyen el $100\,\%$ del genotipo AA (con proporción p_n) y el $50\,\%$ del genotipo Aa y similarmente para los alelos. Si se supone que los genotipos ocurren en las mismas proporciones entre los machos y las hembras, entonces u y v representan (en toda la población) las probabilidades de que el gene sea A o a, respectivamente.

Ejemplo 10.3. En una población, la distribución de genotipos en la $n-$ésima generación es de $50\,\%$ de AA, $30\,\%$ de Aa y $20\,\%$ de aa. ¿Qué proporciones de los genes en esta población son A y a ?

Solución

En este ejemplo, $p_n = 0{,}50$, $q_n = 0{,}30$, $r_n = 0{,}20$. Por lo tanto,

$$u = 0{,}50 + \frac{1}{2}(0{,}30) = 0{,}65 \qquad \text{y} \qquad v = 0{,}15 + 20 = 0{,}35.$$

Es decir, que de la "población" de genes el 65 % es de A y el 35 % es de a.

Con frecuencia es interesante el problema inverso al de la determinación de las proporciones de los genotipos cuando se conocen las proporciones de los alelos. En general, este problema no tiene solución única. El sistema de ecuaciones dado en (10.5) se reduce a una ecuación de dos incógnitas, $u = p_n + (1/2)q_n$. Para obtener una segunda ecuación independiente, supondremos apareamiento aleatorio. Esto quiere decir que la probabilidad de que un individuo dado se aparee con otro individuo no depende del genotipo de este último. En muchos casos, ésta es una suposición correcta. En otros no, por ejemplo, se sabe que la gente alta tiende a casarse con gente alta y por lo tanto la característica de la estatura en los humanos no se puede analizar de esta manera. Por otro lado, se ha demostrado que la suposición de apareo aleatorio se aplica a la característica de los tipos de sangre humana. La mayoría de los individuos escogen su cónyuge sin preocuparse por su tipo de sangre.

Igual que antes, supóngase que u y v son las proporciones de los alelos A y a entre los machos y entre las hembras. Entonces, si suponemos que la población es grande, la probabilidad de que la descendencia reciba el alelo A de los dos padres es u^2. De manera similar, las probabilidades de los genotipos AA y aa son $2uv$ y v^2, respectivamente. El término $2uv$ viene del hecho de que los alelos Aa y aA son el mismo, hecho que ya se había enunciado. Este resultado conduce al siguiente teorema, descubierto de manera independiente por Hardy y Weinberg en 1908.

Teorema 10.6. Ley de Hardy–Weinberg

Supóngase que, en una gran población de padres, los alelos A y a de un gene en particular se presentan en las proporciones u y $v = 1 - u$. Suponiendo que estas proporciones son las mismas para los machos y para las hembras y, además, que el apareo es aleatorio, la primera y todas las generaciones sucesivas se compondrán de los tres genotipos, AA, Aa y aa en las proporciones u^2, $2uv$ y v^2.

Demostración

Como se ha visto, un individuo de la primera generación es de genotipo AA si sus dos padres contribuyen con los alelos A. Como la probabilidad es u de que cualquiera de los padres contribuya con un alelo A, la probabilidad del genotipo AA en la descendencia inmediata es de v^2. De manera semejante, las probabilidades de los genotipos Aa y aa son de $2uv$ y v^2. Esto implica que las proporciones p_1 y q_1 de los alelos A y a en la primera generación están dadas por

$$p_1 = u^2 + \frac{1}{2}(2uv) = u(u+v) = u$$

y

$$q_1 = \frac{1}{2}(2uv) + v^2 = v(u+v) = v.$$

Por lo tanto, las proporciones de los dos alelos no se afectan por la generación inicial. Esto continúa de generación en generación. Concluimos que, después de la generación inicial, las proporciones de los tres genotipos AA, Aa y aa permanecen constantes en u^2, $2uv$ y v^2.

Ejemplo 10.4. El color de la flor de chícharo está controlado por un par de genes. Los tres genotipos AA, Aa y aa se caracterizan por sus flores de color rojas, color rosa y blancas, respectivamente. Si se cultiva un campo al azar con 60 % de flores rojas y 40 % de flores blancas. ¿Qué proporciones de los tres genotipos estarán presentes en la cuarta generación?.

Solución

En este ejemplo, $u = 0{,}6$ y $v = 0{,}4$. Por la *Ley de Hardy–Weinberg*, las proporciones de flores rojas, rosadas y blancas en la primera generación y en todas las subsecuentes son de u^2, $2uv$ y v^2, o sea de 0,36, 0,48 y 0,16, respectivamente. Nótese que la suposición de cultivo aleatorio equivale a la suposición de polinización aleatoria.

La *ley de Hardy–Weinberg* sólo es válida cuando el apareamiento es aleatorio y cuando los tres genotipos son igualmente probables. En ciertos casos, es bastante difícil verificar que el apareo es aleatorio. Sin embargo, si las proporciones de los genotipos permanecen constantes durante varias generaciones y si satisfacen la ley de Hardy-Weinberg, esto se puede tomar como una fuerte evidencia de que el apareamiento es aleatorio. Así, el conocimiento de que el apareo es aleatorio para los tipos de sangre humana, así como para muchas características de las plantas y animales, se dedujo de observaciones de las proporciones de genotipos en cuanto cumplen esta ley.

10.2.2. Los cuadros de Punnett

Un cuadro de Punnett es una gráfica que muestra todas las combinaciones posibles de genes resultantes del cruce de dos organismos (de quienes los genes son conocidos). Se nombran cuadros de Punnett por el genetista inglés, Reginald Punnett. Él descubrió algunos principios básicos de la Genética, incluso la unión del sexo y determinación del sexo. Además, trabajó con las características del color de las plumas de los pollos de manera separada para pollos machos y hembras.

Para ilustrar como se construye un cuadro de Punnett, se debe tener en cuenta que si uno de los padres es del genotipo *Aa*, entonces es igualmente probable que el descendiente herede, de este progenitor el alelo *A* o el alelo *a*. Por otra parte, si uno de los padres es de genotipo *aa* y el otro es de *Aa*, el descendiente recibirá siempre un alelo *a* del progenitor de genotipo *aa* y un alelo *A* o *a*, con igual probabilidad, del progenitor del genotipo *Aa*. Así, el descendiente tiene la misma probabilidad de ser de genotipo *AA* ó *Aa*. En la siguiente tabla se ubican las probabilidades de los posibles genotipos de los descendientes para todas las combinaciones posibles de los genotipos de los padres.

Genotipos de	los progenitores					
los hijos	*AA-AA*	*AA-Aa*	*AA-aa*	*Aa-Aa*	*Aa-aa*	*aa-aa*
AA	1	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
Aa	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0
aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	1

Tabla 10.1: Probabilidades de los posibles genotipos

Situaciones en el que apareamiento no es aleatorio, se presentan frecuentemente en experimentos biológicos controlados. Un ejemplo evidente se da en la cría de caballos de carreras, donde un ganador probado tiene gran demanda como semental. El ejemplo siguiente muestra una de las situaciones de apareamiento controlado.

Ejemplo 10.5. Un agricultor tiene una gran población de plantas con cierta distribución de los tres posibles genotipos, *AA*, *Aa* y *aa*. Este hombre desea iniciar un programa de cultivos en el que todas las plantas de la población sean fecundadas por una planta del genotipo *AA*. Se quiere obtener la fórmula de la distribución de los tres posibles genotipos de la población, después de un cierto número de generaciones.

Solución

Sean p_n, q_n y r_n las proporciones de los tres genotipos en la generación n. Luego, para $n = 0, 1, 2, \ldots,$ se tiene que

$$p_n = p_{n-1} + \frac{1}{2} q_{n-1}, \qquad q_n = \frac{1}{2} q_{n-1} + r_{n-1}, \qquad r_n = 0. \tag{10.6}$$

Estas ecuaciones determinan la distribución de los genotipos en cada generación a partir de la distribución en la generación anterior y se lograron establecer por medio de la Tabla 10.1. El sistema (10.6) se puede expresar en notación matricial como

$$X^{(n)} = PX^{(n-1)} \qquad n = 1, 2, \ldots \tag{10.7}$$

donde

$$X^{(n)} = \begin{bmatrix} p_n \\ q_n \\ r_n \end{bmatrix}, \qquad X^{(n-1)} = \begin{bmatrix} p_{n-1} \\ q_{n-1} \\ r_{n-1} \end{bmatrix} \qquad \text{y} \qquad P = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Nótese que las columnas de la matriz P son iguales a las tres primeras columnas dadas en la tabla 10.1

La ecuación (10.7) implica

$$\begin{aligned} X^{(1)} &= PX^{(0)} \\ X^{(2)} &= PX^{(1)} = P\left(PX^{(0)}\right) = P^2 X^{(0)} \\ X^{(3)} &= PX^{(2)} = P\left(P^2 X^{(0)}\right) = P^3 X^{(0)} \end{aligned}$$

y, en general,

$$X^{(n+1)} = P^n X^{(0)}. \tag{10.8}$$

Así las proporciones de los genotipos futuros están completamente determinados por el vector $X^{(0)}$ de las proporciones iniciales y por la matriz P.

Ahora, es fácil comprobar que los valores propios de P son $\lambda_1 = 1, \lambda_2 = \frac{1}{2}$ y $\lambda_3 = 0$, con vectores propios correspondientes

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{bmatrix}.$$

Luego, P será diagonalizable por la matriz

$$C = \begin{bmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \qquad \text{ya que} \qquad C^{-1}PC = D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Por último, como $P = CDC^{-1}$, se tiene que

$$P^n = (CDC^{-1})^n.$$

Este hecho no es desconocido para nosotros, pues

$$P^n = CD^nC^{-1}.$$

Determinar C^{-1} no tiene mayor inconveniente, después de un breve cálculo se llega a

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

y como ya se hallo C y C^{-1} y además sabemos que la matriz D es la matriz diagonal que contiene los valores propios asociados a P, se tiene que

$$D^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Como $\lim\limits_{n\to\infty} \left(\frac{1}{2}\right)^n \to 0$, se ve que D^n tiende a la matriz

$$\lim_{n\to\infty} D^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

De donde se tiene que

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Nótese que como P es estocástica por columnas, entonces $\lim_{n\to\infty} P^n$ dio como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$.

Por otra parte,

$$\lim_{n\to\infty} X^{(n)} = \lim_{n\to\infty} P^n X^{(0)} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_0 \\ q_0 \\ r_0 \end{bmatrix} = \begin{bmatrix} p_0 + q_0 + r_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

ya que $p_0 + q_0 + r_0 = 1$. Así queda demostrado que a largo plazo, todas las plantas serán genotipo AA.

10.3. Modelo de regresión lineal

El problema central de la inferencia estadística en una distribución bivariada es determinar la relación entre las variables y conocer de qué manera los cambios en una variable afectan la otra.

La variable que es la base de estimación es convencionalmente llamada *variable independiente* que se designa por X y la variable cuyo valor es estimado se llama *variable dependiente* la cual se designa por Y. La selección de las variables dependiente e independiente se hacen de acuerdo con lo conocido y con lo que se desee estimar. En este caso de dependencias entre variables, Y es una variable aleatoria pero X no lo es.

La naturaleza de la relación entre variables se establece a través del *análisis de regresión*. Esta es una técnica con la cual se establece la relación funcional entre las variables, de modo que permite predecir el valor que toma una variable en función del valor determinado de la otra. La regresión es generalmente clasificada en dos tipos: regresión simple y regresión múltiple o general.

La regresión simple se refiere al estudio de las relaciones entre dos variables de las cuales una es independiente (X) y la otra dependiente (Y).

La regresión múltiple comprende tres o más variables, una de las cuales es la variable dependiente que debe ser estimada con base en los valores de las otras variables que son las independientes.

Definición 10.8. Modelo

Es una relación entre dos o más variables cuantitativas, de tal forma que se pueda predecir una variable en función de otra u otras. En este punto es necesario distinguir entre dos tipos de relaciones:

1. Una relación *determinística o funcional* es de la forma

$$Y = f(X),$$

   donde X es la variable independiente y Y es la variable dependiente.

2. Una relación *estocástica o estadística*, no es una relación perfecta, es decir, no proporciona valores únicos de Y para valores determinados de X pero puede describirse con precisión en términos probabilísticos.

En el análisis de regresión se consideran relaciones del segundo tipo, no del primero.

Ejemplo 10.6. La relación entre la variable aleatoria Y y la variable no aleatoria X puede ser expresada por

$$Y = \beta_0 \exp(\beta_1 X) + \varepsilon. \tag{10.9}$$

La ecuación (10.9) significa que para un valor dado de la variable X, el correspondiente de Y es la suma del valor $\beta_0 \exp(\beta_1 X)$ más una cantidad ε. Los parámetros son β_0 y β_1 y ε es la diferencia entre Y y el valor esperado de Y condicionada a un valor de X, es decir

$$\varepsilon = Y - E(Y|X).$$

Definición 10.9. Modelo Lineal

Es una ecuación matemática que involucra variables aleatorias ligadas por parámetros y que es "lineal en los parámetros" y en algunas ocasiones en las variables aleatorias. La frase lineal en los parámetros significa que ningún parámetro en el modelo aparece como un exponente o es multiplicado (o dividido) por cualquier otro parámetro.

Ejemplo 10.7. Cuáles de los siguientes modelos son lineales

$$(i)\ Y = \beta_0 + \beta_1 \frac{1}{X}. \qquad\qquad (ii)\ Y = \beta_0 + \beta_1 \sqrt{X}.$$

$$(iii)\ Y = \beta_0^2 + \beta_1 X.$$

Solución

Los modelos dados en (i) y (ii) son lineales en los parámetros y el modelo dado en (iii) no es lineal en los parámetros ya que β_0 no tiene exponente uno.

Definición 10.10. Modelo de Regresión Lineal Simple

El modelo de la forma

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{10.10}$$

donde

y_i : Es el valor de la variable respuesta en el $i-$ésimo ensayo.

β_0, β_1 :Son los coeficientes (o parámetros) de regresión que corresponden al intercepto y a la pendiente, respectivamente.

x_i : Es un valor conocido, el valor de la variable independiente en el $i-$ésimo ensayo.

ε_i : Es la componente aleatoria, se conoce como error o perturbación.

Se dice que es un modelo de regresión lineal simple. El nombre se debe al hecho de que es lineal tanto en los parámetros como en la variable independiente y simple porque sólo se tiene una variable independiente.

La ecuación (10.10) es una expresión abreviada para el siguiente conjunto de n

ecuaciones simultáneas

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots &= \quad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n. \end{aligned} \tag{10.11}$$

El sistema de ecuaciones (10.11) puede escribirse en forma matricial como sigue

$$\begin{array}{ccccccc} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & = & \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} & + & \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ \vec{Y} & = & \mathbf{X} & \vec{\beta} & + & \vec{\varepsilon}. \end{array} \tag{10.12}$$

Esta representación permite formular la siguiente definición

Definición 10.11. Modelo Lineal General

Un modelo de la forma

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.13}$$

donde

$\vec{Y}$: Es un vector columna de tamaño $n \times 1$, de observaciones sobre la variable dependiente Y.

$\mathbf{X}$: Es una matriz de tamaño $n \times p,\ p \leq n$ de cantidades conocidas fijas, donde la primer columna es de unos y tiene rango igual a $k \leq p$.

$\vec{\beta}$: Es un vector columna de tamaño $p \times 1$, de parámetros desconocidos $\beta_0, \beta_1, \ldots, \beta_{p-1}$.

$\vec{\varepsilon}$: Es un vector aleatorio o perturbado de tamaño $n \times 1$.

Se dice que es un Modelo Lineal General. Este modelo es de rango completo si el rango de $\mathbf{X}$ es igual a p.

10.3.1. Métodos de estimación de parámetros del modelo

Para el modelo dado en (10.13), existen varios métodos de estimación de parámetros, entre estos tenemos:

- Mínimos Cuadrados Ordinarios (M.C.O.)
- Mínimos Cuadrados generalizados o Ponderados (M.C.P.)
- Máxima Verosimilitud (M.V.)

En este capítulo se desarrollará una parte del método de M.C.O., el lector que este interesado en complementar dicho método y en los otros métodos puede revisar el texto Searle (1971).

Método de mínimos cuadrados ordinarios

El método de mínimos cuadrados ordinarios se atribuye a Carl Friedrich Gauss. Bajo ciertos supuestos, este método tiene algunas propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Supuestos para su aplicación:

1. $E\left[\vec{\varepsilon}\right] = \vec{0}$.
2. $E\left[\vec{\varepsilon}\vec{\varepsilon}^{t}\right] = \sigma^2 I_n$.
3. La matriz **X**, es no estocástica, es decir, consta de números fijos.
4. El rango de **X** es $\rho(X) = p$.
5. $\vec{\varepsilon}$ tiene una distribución normal multivariada, es decir, $\vec{\varepsilon} \sim N\left(\vec{0}, \sigma^2 I_n\right)$.

El supuesto 1 significa que el valor esperado del vector de perturbaciones (desviaciones) $\vec{\varepsilon}$, es decir, de cada uno de sus elementos, es cero. Más explícitamente, $E\left[\vec{\varepsilon}\right] = \vec{0}$ significa

$$E\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix} = \begin{bmatrix}E(\varepsilon_1)\\ E(\varepsilon_2)\\ \vdots\\ E(\varepsilon_n)\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{10.14}$$

El supuesto 2 establece que las perturbaciones ε_i y ε_j no están correlacionadas y además que la varianza de ε_i para cada X_i (esto es, la varianza condicional de ε_i)

es algún número positivo constante igual a σ^2, es decir, representa el supuesto de **homoscedasticidad**, o *igual* (homo) *dispersión* (cedasticidad), o *igual varianza*. Más explícitamente, $E\left[\vec{\varepsilon}\vec{\varepsilon}^t\right] = \sigma^2 I_n$ significa

$$E\left[\vec{\varepsilon}\vec{\varepsilon}^t\right] = E\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}\begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n\end{bmatrix} = E\begin{bmatrix}\varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n\\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_n\\ \vdots & \vdots & \ddots & \vdots\\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \dots & \varepsilon_n^2\end{bmatrix}.$$

Al aplicar el operador de valor esperado E a cada elemento de la matriz anterior, se obtiene

$$E\left[\vec{\varepsilon}\vec{\varepsilon}^t\right] = \begin{bmatrix}E\left(\varepsilon_1^2\right) & E\left(\varepsilon_1\varepsilon_2\right) & \dots & E\left(\varepsilon_1\varepsilon_n\right)\\ E\left(\varepsilon_2\varepsilon_1\right) & E\left(\varepsilon_2^2\right) & \dots & E\left(\varepsilon_2\varepsilon_n\right)\\ \vdots & \vdots & \ddots & \vdots\\ E\left(\varepsilon_n\varepsilon_1\right) & E\left(\varepsilon_n\varepsilon_2\right) & \dots & E\left(\varepsilon_n^2\right)\end{bmatrix}. \tag{10.15}$$

La matriz dada en (10.15) se denomina **matriz de varianza-covarianza** de las perturbaciones ε_i. Los elementos sobre la diagonal principal son las varianzas y los elementos por fuera de la diagonal principal son las covarianzas. Por definición

$$Var\left(\varepsilon_i\right) = E\left[\left(\varepsilon_i - E\left(\varepsilon_i\right)\right)^2\right] \qquad Cov\left(\varepsilon_i, \varepsilon_j\right) = E\left[\left(\varepsilon_i - E\left(\varepsilon_i\right)\right)\left(\varepsilon_j - E\left(\varepsilon_j\right)\right)\right],$$

pero debido a los supuestos $E\left(\varepsilon_i\right) = 0$ para cada i y $E\left(\varepsilon_i\varepsilon_j\right) = 0$ si $i \neq j$, la matriz (10.15) se reduce a

$$E\left[\vec{\varepsilon}\vec{\varepsilon}^t\right] = \begin{bmatrix}\sigma^2 & 0 & \dots & 0\\ 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \sigma^2\end{bmatrix}. \tag{10.16}$$

El supuesto 3 estipula que la matriz $\mathbf{X}$, de tamaño $n \times p$ es no-estocástica, es decir, consiste en un conjunto de números fijos.

El supuesto 4 establece que la matriz $\mathbf{X}$ tiene rango columna completo igual a p, el número de columnas en la matriz. Esto significa que las columnas de la matriz $\mathbf{X}$ son linealmente independientes, es decir, no hay **relación lineal exacta** entre las variables X. En otras palabras, no hay *multicolinealidad*.

Forma operativa

Se minimiza la Suma de Cuadrados del Error

$$SCE = \vec{\varepsilon}^t\vec{\varepsilon} = \left(\vec{Y} - \mathbf{X}\vec{\beta}\right)^t\left(\vec{Y} - \mathbf{X}\vec{\beta}\right), \tag{10.17}$$

con respecto a $\vec{\beta}$. Las derivadas parciales de $\vec{\varepsilon}^t\vec{\varepsilon}$ con respecto a $\vec{\beta}$ dan

$$\frac{\delta(\varepsilon^t\varepsilon)}{\delta\vec{\beta}} = 2\mathbf{X}^t\mathbf{X}\hat{\beta} - 2\mathbf{X}^t\vec{Y}.$$

Si se iguala al vector nulo, se llega a las ecuaciones normales de la teoría de *M.C.O.*

$$\mathbf{X}^t\mathbf{X}\hat{\beta} = \mathbf{X}^t\vec{Y}. \tag{10.18}$$

Cuando $\mathbf{X}^t\mathbf{X}$ sea no singular, se puede premultiplicar (10.18) por $(\mathbf{X}^t\mathbf{X})^{-1}$, luego

$$MCO(\beta) = \widehat{\beta} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\vec{Y} = \mathbf{C}\vec{Y}, \tag{10.19}$$

donde $\mathbf{C} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$. Otras estimaciones para el modelo (10.13) mediante el método de *M.C.O.* son

$$\begin{aligned}
\widehat{Y} &= \mathbf{X}\hat{\beta} = \left[\mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\right]\vec{Y} = H\vec{Y}\\
\widehat{\varepsilon} &= \vec{Y} - \hat{Y} = \vec{Y} - H\vec{Y} = (I_n - H)\vec{Y}\\
SCE &= \widehat{\varepsilon}'\widehat{\varepsilon} = \left[(I_n - H)\vec{Y}\right]'(I_n - H)\vec{Y} = \vec{Y}'(I_n - H)\vec{Y} && (10.20)\\
SCT &= \vec{Y}'(I_n - \bar{J}_n)\vec{Y}\\
SCR &= SCT - SCE = \vec{Y}'\left[(I_n - \bar{J}_n) - (I_n - H)\right]\vec{Y} = \vec{Y}'(H - \bar{J}_n)\vec{Y} && (10.21)
\end{aligned}$$

La penúltima expresión se tiene del Ejemplo 5.5.

Obsérvese que la matriz $H = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ la cual en la literatura estadística se conoce como *"Matriz Hat"*, determina muchos de los resultados de las estimaciones por *M.C.O.* Por ejemplo, cuando premultiplica al vector de respuestas $\vec{Y}$ se obtienen las predicciones de la variable dependiente, por eso en algunos textos de Estadística la denominan *Matriz de Predicción* y a la matriz $I_n - H$ la llaman *Matriz Residual*, puesto que al anteponérsele a la variable dependiente $\vec{Y}$ se obtienen los respectivos residuales.

Propiedades de las componentes de la matriz H

La matriz $H = [h_{ij}]$ de tamaño $n \times n$, cumple que

a) $h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j\neq i} h_{ij}^2$, ya que *H es simétrica e idempotente.*

b) $0 < h_{ii} \leq 1$, si $i = 1, 2, \ldots, n$.

c) $-0{,}5 \leq h_{ij} \leq 0{,}5$, para $i \neq j$.

d) $(1 - h_{ii})(1 - h_{jj}) - h_{ij}^2 \geq 0$.

e) $h_{ii}h_{jj} - h_{ij}^2 \geq 0$.

f) Si $h_{ii} = 1$, entonces $h_{ij} = 0$, para todo $j \neq i$.

Si la matriz X de tamaño $n\times r$ es de rango r, entonces

$g)$ $\sum_{i=1}^{n} h_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}^2 = r = tr(H)$.

$h)$ $\sum_{i=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ij} = 1$,

Además, como $h_{ij} = x_i\left(\mathbf{X}^t\mathbf{X}\right)^{-1}x_j'$, h_{ii} está determinada por la localización de x_i en el espacio $\mathbf{X}$. Es decir, un pequeño (grande) valor de h_{ii} indica que x_i se encuentra cerca (lejos) de la masa de los otros puntos.

Ejemplo 10.8. Ajuste el modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \tag{10.22}$$

al conjunto de datos hipotéticos de la siguiente tabla

$\mathbf{x_1}$	**y**	$\mathbf{x_1}$	**y**	$\mathbf{x_1}$	**y**
1	-10	5	-2	9	6
2	-8	6	0	10	8
3	-6	7	2	11	10
4	-4	8	4		

.

Solución

Si se expresa el modelo dado en (10.22) en la forma matricial dada en (10.13) se tiene

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.23}$$

donde $\mathbf{X} = [X_0\ X_1]$ y $X_0 = \mathbf{1}$. Al calcular $(\mathbf{X}^t\mathbf{X})$ se tiene que

$$\left(X^tX\right) = \begin{bmatrix} 11 & 66 \\ 66 & 506 \end{bmatrix} = 11\begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix}.$$

Si se reescribe (10.18) se llega a

$$11\begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 220 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \end{bmatrix}. \tag{10.24}$$

En el Ejemplo 7.8, se obtuvó que el $\kappa(A) \approx 218.9$, luego, existe multicolinealidad moderada, es decir, variaciones muy pequeñas en la varianza de la variable regresora X_1 produce cambios drásticos en las estimaciones de los parámetros.

Por lo tanto, la solución de (10.24) está dada por

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \frac{1}{10}\begin{bmatrix}46 & -6\\ -6 & 1\end{bmatrix}\begin{bmatrix}0\\ 20\end{bmatrix} = \frac{1}{10}\begin{bmatrix}-120\\ 20\end{bmatrix} = \begin{bmatrix}-12\\ 2\end{bmatrix}. \tag{10.25}$$

10.4. Multicolinealidad

La *multicolinealidad* se refiere a la existencia de más de una relación lineal exacta. Inicialmente, este término significó la existencia de una relación “perfecta” o exacta entre algunas o todas las variables explicativas de un modelo de regresión. Para la regresión con p variables que incluye las variables explicativas $X_0, X_1, \ldots, X_p$ (donde $X_0 = 1$ para todas las observaciones que den cabida al término intercepto), se dice que existe una relación lineal exacta si se satisface la siguiente condición

$$\alpha_0 X_0 + \alpha_1 X_1 + \ldots + \alpha_p X_p = \vec{0}, \tag{10.26}$$

donde $\alpha_0, \alpha_1, \ldots, \alpha_p$ son constantes tales que no todas ellas son simultáneamente iguales a cero.

Sin embargo, el término multicolinealidad se utiliza también para el caso en el cual las variables $X_0, X_1, \ldots, X_p$ estan intercorrelacionadas pero no en forma perfecta, de la siguiente manera:

$$\alpha_0 X_0 + \alpha_1 X_1 + \ldots + \alpha_p X_p + \nu_i = \vec{0}, \tag{10.27}$$

donde ν_i es un término de error estocástico.

En los textos de econometría se emplea como medida para detectar la multicolinealidad el índice de condición, de la siguiente manera

$$\begin{array}{lll}\text{Si} & 0 \leq IC\left(\mathbf{X}^t\mathbf{X}\right) \leq 10 & \text{no existe multicolinealidad,}\\ & 10 < IC\left(\mathbf{X}^t\mathbf{X}\right) \leq 30 & \text{existe multicolinealidad entre moderada y fuerte,}\\ & 30 < \kappa(A) & \text{existe multicolinealidad severa.}\end{array}$$

10.4.1. Soluciones al problema de la multicolinealidad

Regresión por Componentes Principales

Una solución que muchas veces se sugiere para el problema de la multicolinealidad es la regresión por componentes principales. Supongamos que se tiene un conjunto de p variables explicativas, $X_1, X_2, \ldots, X_p$. Entonces se construyen funciones lineales de estas variables

$$Z_i = a_{i1}X_1^* + a_{i2}X_2^* + \ldots + a_{ip}X_p^*, \qquad i = 1, 2, \ldots, p, \tag{10.28}$$

con $X_i^* = \frac{X_i - \overline{X}_i}{S_{X_i}}$, de tal manera que un grupo m $(m < p)$ de las variables $Z_1, Z_2, \ldots, Z_p$ contengan aproximadamente la misma información que $X_1, X_2, \ldots, X_p$. Las variables $Z_1, Z_2, \ldots, Z_p$ se buscan de manera que

(i) $Cov(Z_i, Z_j) = 0$, $i, j = 1, 2, \ldots, p$, $i \neq j$.

(ii) $Var(Z_1) \geq Var(Z_2) \geq \ldots \geq Var(Z_p)$.

(iii) $Z_i = \vec{a}_i^t \mathbf{X}_*$ para $i = 1, 2, \ldots, p$ y está sujeto a la condición que

$$\|\vec{a}_i\|_2 = \vec{a}_i^t \vec{a}_i = 1, \tag{10.29}$$

donde $\vec{a}_i^t = (a_{i1}, a_{i2}, \ldots, a_{ip})$ y $\mathbf{X}_* = \begin{bmatrix} X_1^* & X_2^* & \ldots & X_p^* \end{bmatrix}^t$.

La matriz de covarianzas de $\mathbf{X}_*$ es $S = E(\mathbf{X}_* \mathbf{X}_*^t)$, la cual es una matriz simétrica definida (o semidefinida) positiva y de orden p. Dado que

$$Var(Z_i) = E\left(Z_i Z_i^t\right) = E\left(\vec{a}_i^t \mathbf{X}_* \mathbf{X}_*^t \vec{a}_i\right) = \vec{a}_i^t E\left(\mathbf{X}_* \mathbf{X}_*^t\right) \vec{a}_i = \vec{a}_i^t S \vec{a}_i,$$

para hallar Z_1, se necesita conocer el vector de coeficientes $\vec{a}_1$ y para ello, puesto que la varianza de Z_1, debe ser mayor que la varianza de las restantes componentes, habrá que resolver el problema de optimización

$$\begin{aligned} \text{máx} F(\vec{a}_1) &= \vec{a}_1^t S \vec{a}_1, \\ \vec{a}_1^t \vec{a}_1 &= 1. \end{aligned}$$

Por lo tanto, consideremos la maximización de la forma cuadrática $\vec{a}_1^t S \vec{a}_1$, sujeta a la condición (10.29). Si se introduce el multiplicador de Lagrange μ, se maximiza

$$\vec{a}_1^t S \vec{a}_1 - \mu\left(\vec{a}_1^t \vec{a}_1 - 1\right).$$

Al diferenciar respecto a $\vec{a}_1$, μ e igualar a cero las derivadas, se obtiene

$$2S\vec{a}_1 - 2\mu\vec{a}_1 = \vec{0} \qquad \text{y} \qquad \vec{a}_1^t \vec{a}_1 - 1 = 0.$$

Por lo tanto,

$$S\vec{a}_1 = \mu\vec{a}_1 \qquad \text{y} \qquad \vec{a}_1^t\vec{a}_1 = 1$$

Este sistema tiene como soluciones todos los vectores propios de la matriz S de norma 1 asociados a cada uno de los valores propios de S.

Sean $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ los valores propios de S, (la positividad estricta de los valores propios λ_i esta garantizada si S es definida positiva) y $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ los correspondientes vectores propios de S normalizados. Entonces, los puntos estacionarios del problema son

$$\vec{a}_1^j = \vec{v}_j, \qquad j = 1, 2, \ldots, p,$$

con multiplicadores de Lagrange asociados

$$\mu_j = \lambda_j, \qquad j = 1, 2, \ldots, p.$$

Entre todos estos puntos estacionarios el máximo se alcanza en el que coincide con el vector propio de S correspondiente al valor propio dominante.[1]

Además, dado que $\vec{v}_i$ es un vector propio normalizado de S, la forma cuadrática $\vec{a}_j^tS\vec{a}_j = \lambda_j$, de lo cual se deduce que

$$Var[Z_j] = \lambda_j.$$

Propiedades de las componentes principales

(a) La suma de los primeros k valores propios dividido por la suma de todos los valores propios, es decir,

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}$$

representa la “proporción de la variación total” explicado por las primeras k componentes principales.

(b) Las componentes principales de un vector aleatorio son invariantes a los escalares.

(c) Si la matriz de covarianza S tiene rango $r < p$, entonces la variación total de S puede ser explicada enteramente por las primeras r componentes principales.

[1] De hecho hay que considerar $\lambda_1 > \lambda_2$, ya que si $\lambda_1 = \lambda_2$, entonces los vectores propios $\vec{v}_1$ y $\vec{v}_2$ asociados, son ambos solución del problema del máximo.

(d) El subespacio vectorial formado por las primeras k componentes principales $1 \leq k \leq p$ tienen la desviación cuadrática media más pequeña de las variables de la población (o muestra) que cualquier otro subespacio $k-$dimensional.

(e) Como un caso especial de (d) para $k = p - 1$, el plano perpendicular a las últimas componentes principales tienen la desviación cuadrática media más pequeña de las variables de la población (o muestra) que cualquier otro plano.

Ejemplo 10.9. Ajuste el modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{10.30}$$

al conjunto de datos hipotéticos de la siguiente tabla

$\mathbf{x_1}$	$\mathbf{x_2}$	**y**	$\mathbf{x_1}$	$\mathbf{x_2}$	**y**	$\mathbf{x_1}$	$\mathbf{x_2}$	**y**
1	1	-10	5	9	-2	9	17	6
2	3	-8	6	11	0	10	19	8
3	5	-6	7	13	2	11	21	10
4	7	-4	8	15	4			

Solución

Si se expresa el modelo dado en (10.30) en forma matricial se tiene

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.31}$$

donde

$$\mathbf{X} = [X_0\ X_1\ X_2].$$

Aquí $X_0 = \mathbf{1}$ para todas las observaciones que den cabida al término intercepto, de modo que las ecuaciones normales son

$$\left(\mathbf{X}^t\mathbf{X}\right)\hat{\beta} = \mathbf{X}^t\vec{Y}. \tag{10.32}$$

Al calcular $(\mathbf{X}^t\mathbf{X})$ se tiene que

$$\left(\mathbf{X}^t\mathbf{X}\right) = \begin{bmatrix} 11 & 66 & 121 \\ 66 & 506 & 946 \\ 121 & 946 & 1771 \end{bmatrix} = 11 \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix}$$

Si se reescribe (10.32) se llega a

$$11\begin{bmatrix}1 & 6 & 11\\ 6 & 46 & 86\\ 11 & 86 & 161\end{bmatrix}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}=\begin{bmatrix}0\\ 220\\ 440\end{bmatrix}$$

$$\begin{bmatrix}1 & 6 & 11\\ 6 & 46 & 86\\ 11 & 86 & 161\end{bmatrix}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}=\begin{bmatrix}0\\ 20\\ 40\end{bmatrix}. \tag{10.33}$$

Lo primero que se hace es determinar si el sistema de ecuaciones obtenido en (10.33) es estable, para ello se cálcula el número de condición dado en (7.6), por lo tanto, se necesitan los valores propios de la matriz $(\mathbf{X}^t\mathbf{X})$,

$$p_{(\mathbf{X}^t\mathbf{X})}(\lambda)=-\lambda^3+2288\lambda^2-7260\lambda=-\lambda\left(\lambda^2-2288\lambda+7260\right),$$

de donde los valores propios son

$$\lambda_1=11\left(104+2\sqrt{2689}\right),\qquad \lambda_2=11\left(104-2\sqrt{2689}\right)\qquad \text{y}\qquad \lambda_3=0.$$

Luego el número de condición es muy grande ya que

$$\kappa\left(\mathbf{X}^t\mathbf{X}\right)=\frac{\lambda_1}{\lambda_3},$$

en este caso, se dice que existe multicolinealidad severa, es decir, variaciones muy pequeñas en las varianzas y las covarianzas de las variables regresoras X_i producen cambios drásticos en las estimaciones de los parámetros.

En el Ejemplo 9.17 se concluyó que el sistema dado en (10.33) era consistente indeterminado y se obtuvo la siguiente solución

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}=\frac{1}{3}\begin{bmatrix}-28\\ -10\\ 8\end{bmatrix}. \tag{10.34}$$

Como el sistema es consistente indeterminado, se utiliza el análisis de componentes principales para determinar el coeficiente β_i más significativo, para establecer dicho coeficiente, se construye la matriz

$$\Sigma=\frac{1}{n-1}\mathbf{X}^t\left(I_n-\overline{J}_n\right)\mathbf{X}.$$

Al efectuar los productos descritos anteriormente se llega a

$$\Sigma=\begin{bmatrix}11 & 22\\ 22 & 44\end{bmatrix},$$

con los elementos de $\Sigma = [\sigma_{ij}]$ se forma la matriz $S = [s_{ij}]$ donde

$$s_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

Luego

$$S = \begin{bmatrix} \frac{11}{11} & \frac{22}{\sqrt{(11)(44)}} \\ \frac{22}{\sqrt{(11)(44)}} & \frac{44}{44} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Los valores propios de la matriz S son $\lambda_1 = 2$ y $\lambda_2 = 0$ y sus respectivos vectores propios normalizados son $\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

Por lo tanto, las componentes principales son:

$$\begin{aligned} Z_1 =& \frac{1}{\sqrt{2}} X_1^* + \frac{1}{\sqrt{2}} X_2^* = \frac{1}{\sqrt{2}} (X_1^* + X_2^*), \\ Z_2 =& -\frac{1}{\sqrt{2}} X_1^* + \frac{1}{\sqrt{2}} X_2^* = \frac{1}{\sqrt{2}} (X_2^* - X_1^*) \end{aligned}$$

y dado que

$$E[Z_i] = 0 \qquad \text{y} \qquad Var[Z_i] = \lambda_i,$$

se tiene que la componente principal Z_2 tiene media 0 y varianza cero. El último resultado también se puede obtener de la definición de varianza, así

$$\begin{aligned} Var[Z_2] =& Var\left[\frac{1}{\sqrt{2}} (X_2^* - X_1^*)\right] = \frac{1}{2} Var(X_2^* - X_1^*) \\ =& \frac{1}{2} [Var(X_2^*) + Var(X_1^*) - 2Cov(X_1^*, X_2^*)] \\ =& \frac{1}{2} [1 + 1 - 2] = 0 = \lambda_2. \end{aligned}$$

El hecho de que $Var[Z_2] = 0$ identifica la función lineal como el origen de la multicolinealidad. Luego, es posible decir que $Z_2 \simeq 0$ lo cual da que $X_1^* \simeq X_2^*$. Si se realiza la regresión de X_2 sobre X_1, es decir se expresa

$$x_{2i} = \alpha_0 + \alpha_1 x_{1i},$$

se obtiene que

$$x_{2i} = -1 + 2x_{1i}.$$

Como existe una relación exacta entre X_2 y X_1, no es posible estimar por separado los coeficientes de X_1 y X_2. Por lo tanto, si en la ecuación original (10.30) se sustituye x_{2i} en términos de x_{1i}, se obtiene

$$\begin{aligned} y_i =& \beta_0 + \beta_1 x_{1i} + \beta_2 (-1 + 2x_{1i}) + \varepsilon_i \\ =& (\beta_0 - \beta_2) + (\beta_1 + 2\beta_2) x_{1i} + \varepsilon_i. \end{aligned}$$

Esto da lo que en estadística se denomina *funciones lineales estimables* de $\hat{\beta}$, que son:

$$\hat{\beta}_0 - \hat{\beta}_2 \qquad \text{y} \qquad \hat{\beta}_1 + 2\hat{\beta}_2. \tag{10.35}$$

Finalmente, en el Ejemplo 10.8 se obtuvo la regresión de y sobre x_1, lo cual dio como resultado que

$$y_i = -12 + 2x_{1i},$$

es decir

$$\hat{\beta}_0 - \hat{\beta}_2 = -12 \qquad \text{y} \qquad \hat{\beta}_1 + 2\hat{\beta}_2 = 2.$$

Este sistema expresado matricialmente queda

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -12 \\ 2 \end{bmatrix}.$$

Si se usa la inversa generalizada se obtiene que la estimación de $\vec{\beta}$ es

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 \\ 2 & 2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -12 \\ 2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} -56 \\ -20 \\ 16 \end{bmatrix}. \tag{10.36}$$

Nótese que este último vector coincide con el obtenido en (10.34).

Como lo indica este ejemplo, la multicolinealidad revela que no es posible estimar los coeficientes individuales en forma precisa, pero que sí se pueden calcular algunas funciones lineales de los parámetros.

Apéndice A

A.1. Álgebra de los números complejos

Como el concepto de número complejo se utiliza con mucha frecuencia en estas notas de clase y como algunos lectores quizás tengan sólo un conocimiento superficial de los mismos, este apéndice contiene un breve repaso de las propiedades algebraicas más importantes de estos números.

Definición A.1. Número complejo

Un número complejo es una expresión de la forma

$$z = a + bi$$

donde a y b son números reales, a se llama la *parte real* de z y se denota por $Re(z)$; b es llamado la *parte imaginaria* de z y lo denotamos por $Im(z)$. El símbolo i se llama ***unidad imaginaria*** y satisface la propiedad de que $i^2 = -1$.

Definición A.2. Igualdad de números complejos

Dos números complejos $a + bi$ y $c + di$ se definen como iguales si y sólo si, las partes real e imaginaria de uno son respectivamente iguales a las partes real e imaginaria del otro, esto es, si y sólo si, $a = c$ y $b = d$.

Definición A.3. Formas especiales de los números complejos

Dado un número complejo de la forma $z = a + bi$, si $b = 0$ se llama *número real*; por otra parte, si $a = 0$ se denomina *número imaginario puro*.

A.1.1. Operaciones fundamentales

Definición A.4. Suma y diferencia

La suma y diferencia de los números complejos $a + bi$ y $c + di$ son definidas sumando o restando sus partes reales y sus partes imaginarias como sigue

$$(a+bi)+(c+di) = (a+c)+(b+d)i,$$
$$(a+bi)-(c+di) = (a-c)+(b-d)i.$$

Definición A.5. Multiplicación

El producto de los números complejos $a + bi$ y $c + di$ se define como sigue

$$(a+bi)\cdot(c+di) = (ac-bd)+(ad+bc)i. \tag{A.1}$$

Teorema A.1. Un número complejo es igual a cero, si y sólo si, sus partes real e imaginaria valen cero.

Demostración

Queda como ejercicio para el lector.

Definición A.6. Inverso

El recíproco o inverso del número complejo $c + di$ se define como sigue

$$(c+di)^{-1} = \frac{c}{c^2+d^2} + \frac{-d}{c^2+d^2}i = \frac{c-di}{c^2+d^2}.$$

Puesto que c^2 y d^2 son no negativos, $c^2 + d^2 = 0$ si y sólo si, $c = d = 0$. Por lo tanto, *el único número complejo $c + di$ que no tiene recíproco es el cero.*

Esta definición del recíproco nos lleva a la siguiente definición.

Definición A.7. División

Si $w = a + bi$, $z = c + di$ y $z \neq 0$, se puede definir su cociente como sigue

$$\frac{w}{z} = w.z^{-1} = \left(\frac{ac+bd}{c^2+d^2}\right) + \left(\frac{bc-ad}{c^2+d^2}\right)i, \tag{A.2}$$

el cual resulta ser un número complejo.

Teorema A.2. Los números complejos satisfacen las siguientes propiedades:

- Para la Adición

  A1. Si $z_1 = a + bi$ y $z_2 = c + di$ pertenecen a $\mathbb{C}$, su suma $z_1 + z_2$ pertenece a $\mathbb{C}$. Esto también se expresa diciendo que $\mathbb{C}$ es cerrado bajo la adición.

  A2. *Ley Conmutativa:* $z_1 + z_2 = z_2 + z_1$.

  A3. *Ley Asociativa:* $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$.

  A4. *Elemento identidad:* Existe un elemento $0 = 0 + 0i$ en $\mathbb{C}$ tal que si $z \in \mathbb{C}$, $z + 0 = z$.

  A5. Si $z \in \mathbb{C}$, existe un elemento único $-z$ en $\mathbb{C}$, llamado el negativo de z, tal que $z + (-z) = 0$.

- Para la Multiplicación

  M1. Si $z_1 = a + bi$ y $z_2 = c + di$ pertenecen a $\mathbb{C}$, su producto $z_1.z_2$ pertenece a $\mathbb{C}$. Esto también se expresa diciendo que $\mathbb{C}$ es cerrado bajo la multiplicación.

  M2. *Ley Conmutativa:* $z_1.z_2 = z_2.z_1$.

M3. *Ley Asociativa:* $(z_1.z_2).z_3 = z_1.(z_2.z_3)$.

M4. *Elemento identidad:* Existe un elemento $1 = 1+0i$ en $\mathbb{C}$ tal que $1.z = z$ para todo $z \in \mathbb{C}$.

M5. Si $z \neq 0$, existe un elemento único z^{-1}, tal que $z.(z^{-1}) = 1$.

- *Ley Distributiva*: Esta última regla entrelaza la adición y la multiplicación.

  Si z_1, z_2 y z_3 pertenecen a $\mathbb{C}$, entonces

$$z_1.(z_2+z_3) = (z_1.z_2)+(z_1.z_3).$$

Demostración

Queda como ejercicio para el lector.

Definición A.8. Número complejo conjugado

El conjugado de z es el número $\bar{z} = \overline{a+bi} = a-bi$.

Teorema A.3. La suma, diferencia y producto de números complejos con sus conjugados son, respectivamente, un número real, un número imaginario puro y un número real no negativo.

Demostración

Si $z = a+bi$, entonces $\bar{z} = a-bi$ y por lo tanto

$$z+\bar{z} = 2a, \qquad z-\bar{z} = 2bi \qquad \text{y} \qquad z.\bar{z} = a^2+b^2.$$

Teorema A.4. Si un número complejo es igual a su conjugado es un número real, pero si es igual al negativo de su conjugado es un número imaginario puro.

Demostración

Si $z = a+bi$, entonces $\bar{z} = a-bi$, luego si $z = \bar{z}$, por la definición de igualdad se tiene que $b = -b$, así que $b = 0$ y por lo tanto $z = a$. Por otra parte, si $z = -\bar{z}$, por la definición de igualdad se tiene que $a = -a$, de modo que $a = 0$ y $z = bi$.

Una aplicación importante del conjugado de un número complejo está en el cálculo de un cociente, la regla *"multiplíquense numerador y denominador por el conjugado del denominador"* es más fácil de recordar que la fórmula (A.2). En otras palabras, en el proceso de división

$$z^{-1} = \frac{1}{z} = \frac{\bar{z}}{z.\bar{z}}$$

y en forma semejante

$$\frac{w}{z} = w.z^{-1} = \frac{w.\bar{z}}{z.\bar{z}}, \qquad z \neq 0.$$

Teorema A.5. Propiedades de los números complejos conjugados

Si w y z son números complejos, entonces

(a) $\overline{\overline{w}} = w.$ $\qquad$ (b) $\overline{w \pm z} = \overline{w} \pm \bar{z}.$

(c) $\overline{w.z} = \overline{w}.\bar{z}.$ $\qquad$ (d) $(\overline{w/z}) = \overline{w}/\bar{z},$ si $z \neq 0.$

Demostración

Si $w = a + bi$ y $z = c + di$ entonces, por (A.1)

$$\begin{aligned}\overline{w.z} =& \overline{(a+bi).(c+di)} = (a-bi).(c-di)\\ =&(ac-bd)-(ad+bc)i.\end{aligned}$$

Mientras que

$$\begin{aligned}\overline{w}.\bar{z} =& \overline{(a+bi)}.\overline{(c+di)} = (a-bi).(c-di)\\ =&(ac-bd)-(ad+bc)i.\end{aligned}$$

De modo que

$$\overline{w.z} = \overline{w}.\bar{z}.$$

Procedimientos semejantes se aplican en los otros casos.

Definición A.9. Módulo

El módulo o valor absoluto del número complejo $z = a + bi$, representado por $|z|$, es su distancia desde el origen, es decir

$$|z| = \sqrt{z.\bar{z}} = \sqrt{a^2 + b^2}.$$

Teorema A.6. Propiedades del módulo

Si w y z son números complejos, entonces

(a) $|\overline{w}| = |w|$

(b) $|w+z| \leq |w| + |z|$

(c) $|w.z| = |w|.|z|$

(d) $\left|\dfrac{w}{z}\right| = \dfrac{|w|}{|z|}$, si $z \neq 0$

Demostración

Sean $w = a+bi$ y $z = c+di$, entonces

(a) $|\overline{w}| = \sqrt{a+(-b)^2} = |w|$

(b) Para probar esta, observemos que

$$\begin{aligned}(w+z).\overline{(w+z)} =& (w+z).(\overline{w}+\overline{z}) \\ =& w.\overline{w} + z.\overline{z} + \overline{w}.z + w.\overline{z},\end{aligned}$$

o bien

$$|w+z|^2 = |w|^2 + |z|^2 + \left(\overline{w}.z + \overline{\overline{w}.z}\right).$$

Pero como

$$\overline{w}.z + \overline{\overline{w}.z} = 2Re\,(\overline{w}.z) \leq |\overline{w}.z| = 2|w|.|z|,$$

se tiene que

$$|w+z|^2 \leq |w|^2 + |z|^2 + 2|w|.|z| = (|w| + |z|)^2.$$

Tomando raíz cuadrada en ambos miembros, se llega al resultado deseado:

$$|w+z| \leq |w| + |z|$$

(c) La demostración consiste en un cálculo directo

$$\begin{aligned}|w.z| =& |(a+bi).(c+di)| = |(ac-bd)+(ad+bc)i| \\ =& \sqrt{(ac-bd)^2+(ad+bc)^2} = \sqrt{(a^2+b^2).(c^2+d^2)} \\ =& |w|.|z|.\end{aligned}$$

(d) Se deja la prueba para el lector.

Definición A.10. Argumento

El argumento o amplitud del número complejo $z = a + bi$, es el ángulo formado por el segmento que va del origen al punto que representa un número complejo y el eje real positivo y está dado por la expresión

$$\begin{aligned} \theta = \arg(z) &= \underbrace{\arctan\left(\frac{b}{a}\right)} + 2n\pi \qquad\qquad n = 0, \pm 1, \pm 2, \ldots \\ &= \quad Arg(z) \quad + 2n\pi \end{aligned} \tag{A.3}$$

donde $Arg(z)$ denota el valor principal de $\arg(z)$ y se define como el único valor de $\arg(z)$ tal que $-\pi \leq \arg(z) < \pi$.

Teorema A.7. Propiedades del argumento

Si w y z son números complejos, entonces

(a) $\arg(w.z) = \arg(w) + \arg(z)$. $\qquad$ (b) $\arg\left(\frac{w}{z}\right) = \arg(w) - \arg(z)$, si $z \neq 0$.

Demostración

Queda como ejercicio para el lector.

A.1.2. Representación polar

Sean r y θ coordenadas polares del punto (a, b) que corresponden a un número complejo no nulo $z = a + bi$.

$Im(z)$

b $\quad r = |z|$

θ

a $\quad Re(z)$

Por tanto,

$$a = r\cos\theta = |z|\cos(\arg z) \qquad \text{y} \qquad b = r\,\mathrm{sen}\,\theta = |z|\,\mathrm{sen}\,(\arg z)\,.$$

En consecuencia, z puede ser expresado en forma polar como

$$z = r\,(\cos\theta + i\,\mathrm{sen}\,\theta) = |z|\,[\cos(\arg z) + i\,\mathrm{sen}\,(\arg z)]\,.$$

Esta representación polar de z es de gran utilidad para obtener potencias y raíces de números complejos.

Definición A.11. Fórmula de Euler

Sea θ un número real. Se define el símbolo $e^{i\theta}$, como sigue:

$$e^{i\theta} = \cos\theta + i\,\mathrm{sen}\,\theta, \tag{A.4}$$

esta ecuación se conoce como la fórmula de Euler.

Teorema A.8. Teorema de De Moivre

Si $z = r\,(\cos\theta + i\,\mathrm{sen}\,\theta)$, entonces

$$z^n = r^n\,(\cos n\theta + i\,\mathrm{sen}\,n\theta) = |z|^n\,[\cos(n\arg z) + i\,\mathrm{sen}\,(n\arg z)]\,,$$

donde n es cualquier número entero.

Demostración

Queda como ejercicio para el lector.

El teorema de De Moivre también puede utilizarse para encontrar las raíces m-ésimas de un número complejo. Si se hace $n = \frac{1}{m}$, entonces

$$\begin{aligned} z^{\frac{1}{m}} &= [r\,(\cos\theta + i\,\mathrm{sen}\,\theta)]^{\frac{1}{m}} \\ &= r^{\frac{1}{m}}\left[\cos\left(\frac{1}{m}\theta\right) + i\,\mathrm{sen}\left(\frac{1}{m}\theta\right)\right]. \end{aligned}$$

Además, si se tiene en cuenta que

$$\cos(\theta + 2k\pi) = \cos\theta \qquad \text{y} \qquad \mathrm{sen}\,(\theta + 2k\pi) = \mathrm{sen}\,\theta$$

en donde k es un entero, se tiene que

$$\begin{aligned} z^{\frac{1}{m}} &= [r(\cos\theta + i\operatorname{sen}\theta)]^{\frac{1}{m}} \\ &= \{r(\cos(\theta + 2k\pi) + i\operatorname{sen}(\theta + 2k\pi))\}^{\frac{1}{m}} \\ &= r^{\frac{1}{m}}\left[\cos\left(\frac{\theta + 2k\pi}{m}\right) + i\operatorname{sen}\left(\frac{\theta + 2k\pi}{m}\right)\right]. \end{aligned}$$

Las raíces m-ésimas se obtienen asignando a k los m valores consecutivos enteros $0, 1, 2, \ldots, m-1$.

Observación
Si $m = 2$, se tiene que

$$\sqrt{z} = \sqrt{a + bi} = |z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2} + k\pi\right) + i\operatorname{sen}\left(\frac{\theta}{2} + k\pi\right)\right], \qquad k = 0, 1.$$

Luego,

$$\sqrt{z} = \begin{cases} |z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2}\right) + i\operatorname{sen}\left(\frac{\theta}{2}\right)\right] & \text{si } k = 0, \\ -|z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2}\right) + i\operatorname{sen}\left(\frac{\theta}{2}\right)\right] & \text{si } k = 1. \end{cases}$$

Si se usan las siguientes identidades trigonométricas

$$\cos\left(\frac{\theta}{2}\right) = \sqrt{\frac{1 + \cos\theta}{2}} \qquad \text{y} \qquad \operatorname{sen}\left(\frac{\theta}{2}\right) = \sqrt{\frac{1 - \cos\theta}{2}}$$

se llega a

$$\sqrt{z} = (-1)^k \frac{\sqrt{2}}{2}\left[\sqrt{a + |z|} + \sqrt{a - |z|}\right], \qquad k = 0, 1. \tag{A.5}$$

Ejemplo A.1. Si $z = a + bi$, demuestre que

$$\sqrt{z} \pm \sqrt{\bar{z}} = \pm\sqrt{2(a \pm |z|)}$$

Solución
Si $z = a + bi$, entonces $\bar{z} = a - bi$. Por lo tanto

$$\left(\sqrt{z} \pm \sqrt{\bar{z}}\right)^2 = z + \bar{z} \pm 2\sqrt{z.\bar{z}}.$$

Por el Teorema A.3, se tiene

$$\left(\sqrt{z} \pm \sqrt{\bar{z}}\right)^2 = 2a \pm 2\sqrt{a^2 + b^2}.$$

Al tomar raíz cuadrada a ambos lados, se obtiene

$$\sqrt{z} \pm \sqrt{\bar{z}} = \pm\sqrt{2(a \pm |z|)}$$

y la prueba queda completa.

Bibliografía

Anton, H. (1996), *Introducción al Álgebra Lineal*, Editorial Limusa S.A. Grupo Noriega Editores, México.

Apostol, T. M. (1985), *Calculus*, Editorial Reverté S.A., Barcelona.

Asmar, A. J. (1995), *Tópicos en Teoría de Matrices*, Universidad Nacional de Colombia, Sede Medellín.

Barbolla, S. y Sanz, P. (1998), *Álgebra Lineal y Teoría de Matrices*, Prentice Hall, Madrid.

Bretscher, O. (1997), *Linear Algebra with Applications*, Prentice Hall, New Jersey.

Bru, R. y. o. (2001), *Álgebra Lineal*, Alfaomega, México.

Fraleigh, B. (1989), *Álgebra Lineal*, Addison-Wesley Iberoamericana S.A., Estados Unidos.

Golubitsky, M. y. D. M. (2001), *Álgebra Lineal y Ecuaciones Diferenciales con Matlab*, International Thomson Editores, S.A., México.

Grossman, S. I. (1996), *Álgebra Lineal*, 5^{ta} edición edn.

Grossman, S. I. (1998), *Aplicaciones de Álgebra Lineal*, Grupo Editorial Iberoamérica, S.A., México.

Herstein, I. y. W. D. (1989), *Álgebra Lineal y Teoría de Matrices*, Grupo Editorial Iberoamérica, S.A., México.

Hoaglin, D.C. y Welsch, R. (1978), 'The hat matrix in regression and anova', *The American Statistician* **32**(1), 17–22.

Horn, R. y. J. C. (1985), *Matrix Analysis*, Cambridge University Press, U.S.A.

Kemeny, J. y. S. J. (1976), *Finite Markov Chains*, Springer-Verlag, Nueva York.

Kolman, B. (1997), *Álgebra Lineal con Aplicaciones y matlab*, Prentice Hall, México.

Lang, S. (1976), *Álgebra Lineal*, Fondo Educativo Interamericano, S.A., México.

Larson, E. (2000), *Introducción al Álgebra Lineal*, Limusa Noriega Editores, México.

Lay, D. C. (1994), *Linear Algebra and its Applications*, Addison-Wesley Publishing Company, U.S.A.

Murdoch, D. C. (1970), *Linear Algebra*, John Wiley & Sons, Nueva York.

Nakos, G. y Joyner, D. (1999), *Álgebra Lineal con Aplicaciones*, International Thomson Editores, S.A., México.

Noble, B. y Daniel, J. W. (1989), *Álgebra Lineal Aplicada*, Prentice-Hall Hispanoamericana, S.A., México.

Paige, L. y Swift, D. (1961), *Elements of Linear Algebra*, Blaisdell Publishing Company, Massachusetts.

Penrose, R. (1955*a*), 'A generalized inverse for matrices', *Proceedings of the Cambridge Philosophical Society* **51**, 406–413.

Penrose, R. (1955*b*), 'On best approximate solutions of linear matrix equations', *Proceedings of the Cambridge Philosophical Society* **52**, 17–19.

Pringle, R.M. y Rayner, A. (1971), *Generalized Inverse Matrices with Applications to Statistics*, Charles Griffin & Company Limited, Londres.

Rao, C.R. y Rao, M. (1998), *Matrix Algebra and its Applications to Statistics and Econometrics*, World Scientific, U.S.A.

Rorres, C. y Anton, H. (1979), *Aplicaciones de Álgebra Lineal*, Limusa, México.

Samelson, H. (1974), *An Introduction to Linear Algebra*, John Wiley & Sons, Nueva York.

Searle, S. (1971), *Linear Models*, John Wiley & Sons, Nueva York.

Searle, S. (1982), *Matrix Algebra Useful for Statistics*, John Wiley & Sons, U.S.A.

Strang, G. (1986), *Algebra Lineal y sus Aplicaciones*, Addison Wesley Iberoamericana, México.

Índice alfabético