Universidad Mayor de San Andrés
Facultad de Ciencias Puras y Naturales
Carrera de Matemática

Localización y Perturbación de Autovalores

(Análisis Matricial)

**Proyecto de grado para optar el título de
Licenciado en Matemática**

Autor:
Renán Justiniano Blanco Rodríguez

Supervisores:
Dr. Efraín Cruz Mullisaca
Docente de la Universidad Mayor de San Andrés

La Paz, Junio de 2011

Agradecimientos

Cualquier camino que nos permita avanzar en nuestros logros, seguramente es por la gracia de Dios , por eso mi mas grande agradecimiento es a nuestro creador sin el cual nada seria posible.
También agradezco a la carrera de matemática y a la Universidad Mayor de San Andrés por darme la oportunidad de formarme como profesional en esta área. A mi carrera a la que considero mi segundo hogar en la que compartí con mis compañeros gran parte de mi vida de estudiante. Quiero referirme en forma particular a la matemática como ciencia con la cual me identifico de la que puedo expresar, que a través de la demostraciones de resultados en el campo de la ciencia que son de mi mayor agrado y satisfacción.
Quiero agradecer de manera especial al Dr. Efrain Cruz Mullisaca el cual me permitió y me brindo su gran ayuda y colaboración en todo momento. También agradezco a mi tribunal M.sc. Luis Tordoya L. y al Lic. Raúl Borda V. por su tolerancia y guía en el trabajo que realice donde su orientación fue crucial.
También mis agradecimientos sinceros a todos mis compañeros que me colaboraron. Al Lic. Julio Flores un especial agradecimiento, y a mi compañero José Luís Laura G. que me colaboró en la transcripción de este trabajo en LaTeX.

A mi familia

A mi amada esposa la cual siempre me apoyo en los momentos mas difíciles de mi vida, a mi amado hijo Renán Diego por el cual vivo y agradezco a la vida que me regalo dios, a mis queridos padres los cuales constituyen el motivo por el cual estoy en este mundo.

Bienaventurado el hombre que puso en Jehova su confianza,
y no mira a los soberbios,
ni a los que se desvían tras la mentira

Salmos C. 40 V. 4

Por el camino de la sabiduría te he encaminado,
y por veredas derechas te he hecho andar.
Proverbios. C.4 V. 11

El temor a Jehova es el principio de la sabiduría,
y el conocimiento del santísimo es la inteligencia.
Proverbios.C. 9 V. 10

Pero sin fe es imposible agradar a dios; porque es necesario que el que se acerca a dios crea que la hay,

y que es galardonador de los que le buscan.

Hebreos. C.11 V. 6

Introducción

Los ceros de un polinomio ha acaparado el tiempo de muchos matemáticos hasta la actualidad, los cuales pueden ser ligados a la resolución de ecuaciones algebraicas, se conocen métodos para las ecuaciones lineales, cuadráticas, algunas cúbicas y las de grado cuatro; sin embargo, para ecuaciones de mayor grado no se tiene métodos específicos simplemente se conocen técnicas de aproximación.

Una interesante aplicación de estos resultados se encuentran ligados a la teoría de matrices, es decir, en la obtención de los autovalores de una matriz cuadrada, la cual se encuentra ligada a su polinomio característico o ecuación característica; cuando el tamaño de la matriz es mayor que cuatro, hallar los ceros del polinomio característico cae en aproximaciones, por lo cual se buscan nuevas técnicas de aproximar a los autovalores de la matriz cuadrada.

En este trabajo, desarrollaremos un criterio de aproximar al autovalor (por ende la solución de un polinomio característico) en el plano complejo por medio de regiones o discos abiertos, este resultado es la base de muchos estudios al respecto es el Teorema de Gershgorin.

Con el propósito de hacer que la lectura del presente trabajo sea comprendido, se desarrollarán las siguientes temáticas:

A. El primer capítulo se ha destinado a la revisión básica de resultados de matrices y operaciones entre ellas, además de revisar algunas características de las matrices que las clasifican en simétricas, normales, hermitianas, etc. El resultado central en este capítulo es la descomposición de Schur para matrices hermitianas, la cual afirma la existencia de una matriz unitaria que diagonaliza a la matriz estudiada, dejando en la matriz diagonal los autovalores de la matriz.

B. El segundo capítulo comprende el estudio de las diversas normas vectoriales, las normas matriciales y la relación existentes entre ellas, estas normas son usadas en resultados esenciales para posteriormente demostrar los teoremas principales, además de presentar una aplicación al cálculo de la inversa de una matriz invertible.

C. El tercer capítulo contiene los teoremas principales de este trabajo, desarrollamos el teorema de Gershgorin para determinar las regiones donde se encuentran los autovalores en el plano complejo. Posteriormente estudiamos el comportamiento de los autovalores de la matriz bajo perturbaciones, esto es cuanto se modifica la región que contiene los autovalores cuando se suma otra matriz llamada perturbación.

Finalmente, listamos los textos revisados para este trabajo los cuales pueden ser consultados para más detalles o realizar otros estudios en esta área.

Índice general

CAPÍTULO 1

Teorema de Schur

Este capítulo tiene por objetivo describir los conceptos básicos pero necesarios para la comprensión de los resultados principales de esta monografía. Se ha distribuido este capítulo por secciones de acuerdo a las temáticas utilizadas en el desarrollo del trabajo.

1.1. Autovalores y Autovectores

En esta sección revisaremos algunos hechos principales en lo referente a una matriz tanto en su forma conceptual como en su aplicación.

Cambio de bases y semejanza. Cada matriz invertible es una matriz cambio de base, y cada matriz cambio de bases es invertible. Luego, si B es una base dada de un espacio vectorial V , si T es una transformación lineal sobre V, y si $A = [T]_B^B$ es la representación de T respecto a la base B, el conjunto de todas las posibles representaciones de T es

$$\left\{[I]_{B_1}^{B}[T]_B^B[I]_B^{B_1} : \text{ es una matriz invertible}\right\} = \left\{S^{-1}AS : \ S \in M_n(\mathbb{F}) \text{ es una matriz invertible}\right\}$$

Es justamente el conjunto de todas las matrices que son similares a la matriz dada A. Similares pero no matrices idénticas, por lo tanto son diferentes representaciones respecto a la base de una transformación lineal simple.

Es de esperar que las matrices similares compartan muchas propiedades importantes (por lo menos, aquellas propiedades que son intrínsecas a la transformación lineal), este es un tema importante en el álgebra lineal.

Las nociones de semejanza y autovalores son conceptos importantes que describiremos en lo que sigue.

No singularidad. Una transformación lineal o matriz se dice que es no singular si este produce el resultado 0 sólo para la entrada 0. En otro caso, este es singular. Si $A \in M_{m,n}(\mathbb{F})$ y $m < n$, entonces A es necesariamente singular. Si $A \in M_n(\mathbb{F})$, A es llamado invertible si existe una matriz $A^{-1} \in M_n(\mathbb{F})$ llamado la inversa de A tal que $A^{-1}A = I$. Equivalentemente, A es invertible si la transformación lineal A es uno a uno, y su transformación inversa (también lineal) existe. Si $A \in M_n$ y $A^{-1}A = I$, entonces $AA^{-1} = I$, además A^{-1} es única cuando este existe.

Las matrices no singulares en $M_n(\mathbb{F})$ forman un grupo, el grupo general lineal, denotado por $GL(n, \mathbb{F})$.

Denotemos $A \in M_{n\times n} = M_n$ la matriz $n \times n$ con entradas complejas.

Definición 1.1. Si $A \in M_n$ y $x \in \mathbb{C}^n$, consideramos la ecuación

$$Ax = \lambda x, \quad x \neq 0$$

donde λ es un escalar. Si un autovalor λ y un autovector x no cero existen y satisfacen esta ecuación, entonces λ es llamado autovalor de A y x es llamado autovector de A asociado a λ. Observemos que los dos se dan como un par, y que un vector propio no puede ser el vector cero.

Definición 1.2. El conjunto de todos los $\lambda \in \mathbb{C}$ que son autovalores de $A \in M_n$ es llamado el espectro de A y es denotado por $\sigma(A)$. El radio espectral de A es el número real no negativo

$$\rho(A) = \text{máx}\left\{|\lambda| : \ \lambda \in \sigma(A)\right\}$$

Es justamente el radio de los discos más pequeños centrados en el origen en el plano complejo que incluye a todos los autovalores de A.

Notemos que si x es un autovector asociado con el autovalor λ de A, entonces cualquier múltiplo escalar no nulo de x es también un autovector.

Incluso si no tienen otra importancia, autovalores y autovectores, son interesantes algebraicamente, ya que los autovectores son los vectores de manera que la multiplicación por A tiene una forma muy sencilla, al igual que la multiplicación por un escalar (el valor propio).

Ejemplo 1.1. Considere la matriz

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} \in M_2$$

Entonces se tiene que $3 \in \sigma(A)$ que tiene como autovector asociado a

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{puesto que} \quad A\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Del mismo modo se tiene que $5 \in \sigma(A)$.
Recordemos que un polinomio es de la forma $p(t) = a_k t^k + a_{k-1}t^{k-1} + \cdots + a_1 t + a_0$, donde a_i son números complejos con $i = 1, 2, \ldots, k$.

Teorema 1.1. Sea $p(\cdot)$ un polinomio dado. Si λ es un autovalor de $A \in M_n$, mientras que x es un autovector asociado, entonces $p(\lambda)$ es un autovalor de $p(A)$ y x es un autovector de $p(A)$ asociado con $p(\lambda)$.

Prueba. Consideremos $p(A)x$. Observemos que

$$p(A) = a_k A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + I, \quad \text{con } A^0 = I$$
$$p(A)x = a_k A^k x + a_{k-1}A^{k-1}x + \cdots + a_1 Ax + a_0 x$$

Luego,

$$A^j x = A^{j-1}Ax = A^{j-1}\lambda x = \lambda A^{j-1}x = \cdots = \lambda^j x$$

por repetir la ecuación de los autovectores.

Por lo tanto,

$$p(A)x = a_k \lambda^k x + a_{k-1}\lambda^{k-1}x + \cdots + a_1\lambda x + a_0 x = p(\lambda)x\,.$$

♦

Observemos que una matriz $A \in M_n$ es singular si y sólo si 0 es un autovalor de A, en efecto el $\det(A) = \lambda_1, \lambda_2, \ldots, \lambda_n$, por tanto $\det(A) = 0$, lo que significa que A es una matriz singular.

Una matriz $A \in M_n$ es llamada idempotente si $A^2 = A$, entonces cada autovalor de A es 0 o 1, en efecto $A^2 - A = 0$, luego $A(A - I) = 0$, por tanto el $\det(A) \times \det(A - I) = 0$.

Una matriz $A \in M_n$ es nilpotente si $A^q = 0$ para algún entero q. El mínimo de tales q es llamado el índice de nilpotencia. Los autovalores de una matriz nilpotente son 0.

Polinomio característico. Es importante conocer las propiedades de los autovalores de $A \in M_n$, tales como conocer la cantidad y como pueden ser caracterizados.

De la ecuación $Ax = \lambda x$ se obtiene $(\lambda I - A)x = 0$, con $x \neq 0$.

Luego $\lambda \in \sigma(A)$ si y sólo si $\lambda I - A$ es una matriz singular, esto es

$$\det(\lambda I - A) = 0$$

Definición 1.3. El polinomio característico de $A \in M_n$ es definido por

$$p_A(t) = \det(tI - A)$$

Observemos que si $A \in M_n$ el polinomio característico $p_A(\cdot)$ tiene grado n y el conjunto de raíces de $p_A(t) = 0$ coincide con $\sigma(A)$.

Semejanza. Una transformación de semejanza de una matriz en M_n corresponde a la representación de una transformación lineal de $\mathbb{C}^n$ en otra base. Así, el estudio de semejanza se puede considerar como el estudio de propiedades que son intrínsecos a una transformación lineal, o las propiedades que son comunes a todas sus representaciones en bases diferentes.

Definición 1.4. Una matriz $B \in M_n$ se dice ser semejante a una matriz $A \in M_n$ si existe una matriz no singular $S \in M_n$ tal que $B = S^{-1}AS$.

La transformación $A \mapsto S^{-1}AS$ es llamada transformación de semejanza por la matriz de semejanza S.

La relación B es semejante a A se abrevia por $B \sim A$.

Esta relación es de equivalencia sobre M_n.

Es decir,

$$\begin{array}{ll} A \sim A & \text{Reflexiva} \\ B \sim A \Rightarrow A \sim B & \text{Simétrica} \\ C \sim B \wedge B \sim A \Rightarrow C \sim A & \text{Transitiva.} \end{array}$$

Teorema 1.2. Sean A, B en M_n. Si B es semejante a A, entonces los polinomios característicos de B es el mismo que el de A.

Prueba. Para cualquier t se tiene

$$\begin{aligned} p_B(t) &= \det(tI - B) \\ &= \det(tS^{-1}S - S^{-1}AS) \\ &= \det\left(S^{-1}(tI - A)S\right) \\ &= \det(S^{-1})\det(tI - A)\det(S) \\ &= \det(S^{-1})\det(S)\det(tI - A) \\ &= \det(tI - A) \\ &= p_A(t) \end{aligned}$$

♦

Como una consecuencia inmediata del Teorema 1.2, se tiene que si A y B son semejantes,entonces ellos tienen los mismos autovalores contando las multiplicidades.

Puesto que las matrices diagonales son especialmente simples y que tienen propiedades muy buenas, es interesante saber para cuales matrices $A \in M_n$ existe una matriz diagonal en la clase de equivalencia de semejanza de A, es decir, que matrices son semejantes a las matrices diagonales.

Definición 1.5. Si la matriz $A \in M_n$ es semejante a la matriz diagonal, entonces A se dice ser diagonalizable. Algunas veces el término diagonal es usado.

Teorema 1.3. Sea $A \in M_n$. Entonces A es diagonalizable si y sólo si existe un conjunto de n vectores linealmente independiente, cada uno de los cuales son autovectores de A.

Prueba. Si A tiene n autovectores linealmente independientes $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$, formamos una matriz no singular S con ellos como las columnas y calculando

$$\begin{aligned} S^{-1}AS &= S^{-1}\left[Ax^{(1)}Ax^{(2)}\ldots Ax^{(n)}\right] \\ &= S^{-1}\left[\lambda_1 x^{(1)}\lambda_2 x^{(2)}\ldots \lambda_n x^{(n)}\right] \\ &= S^{-1}\left[x^{(1)}x^{(2)}\ldots x^{(n)}\right]\Lambda \\ &= S^{-1}S\Lambda \\ &= \Lambda \end{aligned}$$

donde

$$\begin{bmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{bmatrix}$$

y $\lambda_1, \lambda_2, \ldots, \lambda_n$ son autovalores de A.

Recíprocamente, supongamos que existe un matriz semejante S tal que $S^{-1}AS = \Lambda$ es diagonal. Entonces $AS = S\Lambda$. Esto significa que A veces la i−ésima columna de S (i.e. la i−ésima columna de AS) es la i−ésima entrada en la diagonal de Λ veces la j−ésima columna de S (i.e. la i−ésima columna de $S\Lambda$), o que la j−ésima columna de S es un autovector de A asociado con la i−ésima entrada en la diagonal de Λ. Puesto que S es no singular, existen n autovectores linealmente independientes. ♦

Si $A \in M_n$ es diagonalizable, las entradas en la diagonal de cualquier matriz diagonal la cual es semejante deben ser los autovalores de A, con sus multiplicidades.

Un simple hecho en la que se asegura la diagonalización es aquel en el que los autovalores son distintos. Un importante resultado de este hecho, que es, sin duda útil, es el siguiente lema.

Lema 1.1. Supongamos que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son autovalores de $A \in M_n$, dos de los cuales no son los mismos, y supongamos que $x^{(i)}$ es un autovector asociado con λ_i, $i = 1, 2, \ldots, k$. Entonces $\left\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right\}$ es un conjunto linealmente independiente.

Prueba. La prueba es esencialmente por contradicción. Suponga que $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ son linealmente dependientes. Entonces existe una combinación lineal no trivial el cual es el vector 0, y de hecho existe una combinación lineal con el menor coeficiente no cero. Supongamos que tal relación minimal linealmente dependiente es

$$\alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_r x^{(r)} = 0, \quad r \leq k$$

Tenemos $r > 1$ puesto que todos los $x^{(i)} \neq 0$. Podemos asumir por conveniencia (renumerando si es necesario) que esto envuelve los primeros r vectores.

También tenemos

$$\begin{aligned} A(\alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_r^{(r)}) &= \alpha_1 Ax^{(1)} + \alpha_2 Ax^{(2)} + \cdots + \alpha_r Ax^{(r)} \\ &= \alpha_1\lambda_1 x^{(1)} + \alpha_2\lambda_2 x^{(2)} + \cdots + \alpha_r\lambda_1 x^{(r)} \\ &= 0 \end{aligned}$$

otra relación de dependencia. Ahora multiplicando la primera relación de dependencia por λ_r y restando este de la segunda relación se tiene

$$\alpha_1(\lambda_1 - \lambda_r)x^{(1)} + \alpha_2(\lambda_2 - \lambda_r)x^{(2)} + \cdots + \alpha_{r-1}(\lambda_{r-1} - \lambda_r)x^{(r-1)} = 0$$

una tercera relación de dependencia, el cual tiene menos coeficientes cero que el primero. Esta última relación es no trivial puesto que $\lambda_i \neq \lambda_r,\ i = 1, 2, \ldots, r-1$. Esta es una contradicción a la suposición de minimalidad para la primera relación de dependencia, lo que completa la prueba. ♦

Teorema 1.4. Si $A \in M_n$ tiene n autovalores distintos, entonces A es diagonalizable.

Prueba. Si $\sigma(A) = \left\{\lambda_1, \lambda_2, \ldots, \lambda_n\right\}$, sea $x^{(i)}$ un autovector asociado a λ_i, $i = 1, 2, \ldots, n$. Puesto que los autovalores son todos diferentes, $\left\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right\}$ es un conjunto linealmente independiente, por tanto A es diagonalizable. ♦

1.2. Equivalencia Unitaria y Matrices Normales

Para una matriz general no singular $S \in M_n$, se ha estudiado la semejanza a través de S en la sección anterior. Para ciertas matrices no singulares muy especiales, llamadas matrices unitarias, la inversa de S tiene una forma simple: $S^{-1} = S^*$. La semejanza de $A \in M_n$ a través de una matriz unitaria, $A \mapsto S^*AS$, no es sólo conceptual más simple $\left(S^*\text{ es mucho más fácil de evaluar que } S^{-1}\right)$ que la semejanza general, pero este tiene una serie de atractivas características que se harán más claras a través del desarrollo de esta sección. Como regla general, las semejanzas unitarias son preferibles a las semejanzas generales, por lo que es útil saber lo que puede lograrse a través de semejanza unitaria. Las clases de equivalencia bajo semejanza unitaria son, sin embargo, más fina que la semejanza general (dos matrices pueden ser semejantes pero no unitariamente semejantes), y en consecuencia menos pueden ser alcanzados.

1.2.1. Matrices Unitarias

Definición 1.6. Los vectores $x_1, x_2, \ldots, x_k \in \mathbb{C}^n$ forman un conjunto ortogonal si $x_i^* x_j = 0$ para todo par $1 \leq i,\ j \leq k$, si además, los vectores son normalizados, $x_i^* x_i = 1$, $i = 1, 2, \ldots, k$, entonces el conjunto es llamado ortonormal.

Observemos que un conjunto de vectores ortonormales es linealmente independiente.

Un conjunto independiente no necesita ser ortonormal de hecho, pero uno puede aplicar el proceso de Gram Schmidt para ortonormalizar con el mismo generador como el conjunto original.

Definición 1.7. Una matriz $U \in M_n$ se dice ser unitaria si $U^*U = I$. Si además, $U \in M_n(\mathbb{R})$, U es llamada ortogonal real.

El conjunto de matrices unitarias tienen importantes propiedades y que las resumimos en el teorema siguiente.

Teorema 1.5. Si $U \in M_n$, las siguientes son equivalentes:

(1) U es unitaria;

(2) U es no singular y $U^* = U^{-1}$;

(3) $UU^* = I$;

(4) U^* es unitaria;

(5) Las columnas de U forman un conjunto ortonormal;

(6) Las filas de U forman un conjunto ortonormal;

(7) Para todo $x \in \mathbb{C}^n$, la longitud Euclidiana de $y = Ux$ es la misma que x, esto es $y^*y = x^*x$.

Prueba. (1) implica (2) Puesto que U^{-1} (cuando este existe) es matriz única, multiplicando por la izquierda por lo cual produce I; la definición de unitaria garantiza que U^* es tal matriz. Puesto que $BA = I$ si y sólo si $AB = I$ (para $A, B \in M_n$). (2) implica (3) Puesto que $(U^*)^* = U$, (3) implica (4) puesto que U^* satisface los requisitos necesarios para ser unitaria. Puesto que la recíproca de cada implicación es observado similarmente, (1) a (4) son equivalentes.

Considerando la multiplicación de matrices y haciendo que $u^{(i)}$ denota la $i-$ésima columna de U, $i = 1, 2, \ldots, n$, del hecho que $U^*U = I$ se tiene que

$$u^{(i)*}u^{(j)} = \begin{cases} 0 & \text{si } j \neq i \\ 1 & \text{si } j = i \end{cases}$$

Luego, $U^*U = I$ es otra manera de decir que las columnas de U son ortonormales, luego (1) es equivalente a (5). Similarmente, (4) y (6) son equivalentes.

Si (1) es válido y $y = Ux$, entonces $y^*y = x^*U^*Ux = x^*Ix = x^*x$, luego (1) implica (7). Para verificar la recíproca, requerimos cálculos más elaborados, con otros resultados que no se han descrito podrían ser hechos inmediatos.

Sin embargo,primero consideremos el caso $n = 2$.

Asumiendo (7) y con $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, hallamos que $1 = x^*x = y^*y = x^*U^*Ux$, 1 es la entrada de U^*U.

Similarmente, sea $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, concluimos que las 2, 2 entrada de U^*U es también 1, y U^*U tiene la forma

$$\begin{bmatrix} 1 & a \\ \bar{a} & 1 \end{bmatrix}$$

donde a es el producto interior de la columna 1 y la columna 2 de U, y $\bar{a}$ es el producto interior de la columna 2 y la columna 1.

Sea $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ y (7) calculando otra vez, hallamos que $2 = x^*x = y^*y = x^*U^*Ux = 2 + (a + \bar{a})$.

Siendo $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, hallamos $2 = 2 + i(a - \overline{a})$.

Luego $a + \overline{a} = 2\text{Re}(a) = 0$ y $a - \overline{a} = 2\text{Im}(a) = 0$ y por tanto $a = 0$. Esto significa que si $x^*U^*x = x^*x$ para todo $x \in \mathbb{C}^2$, entonces $U^*U = I$; esto es, U es unitaria (si $U \in M_2$).

Ahora considerando $n > 2$, y sea $A = U^*U$. Sea $x \in \mathbb{C}^n$ tal que todas las otras componentes que las $i-$ésima y $j-$ésima, $i \leq j$ son 0. Entonces

$$x^*Ax = \left[\overline{x}_i, \overline{x}_j\right] A\big([i,j]\big) \begin{bmatrix} x_i \\ x_j \end{bmatrix}$$

y mostramos que (7) implica que $A\big([i,j]\big) = I \in M_2$. Puesto que i y j son arbitrarias, concluimos que cada submatriz principal de 2×2 de A es la matriz identidad de 2×2. El único ejemplo A es $A = I \in M_n$, y que el caso $n = 1$ es obvio, concluimos que (7) implica (1), lo cual completa la prueba. ♦

Observemos que si U, V en M_n son unitarias (respectivamente ortogonal real), entonces el producto UV es también unitaria (respectivamente ortogonal real).

1.2.2. Equivalencia Unitaria

Puesto que $U^* = U^{-1}$ para una matriz unitaria U, la transformación sobre M_n dada por $A \mapsto U^*AU$ es una transformación semejante si U es unitaria. Este tipo especial de semejanza es llamada similarmente unitaria o unitariamente equivalente.

Definición 1.8. Una matriz $B \in M_n$ se dice ser unitariamente equivalente a $A \in M_n$, si existe una matriz unitaria $U \in M_n$ tal que $B = U^*AU$. Si U puede ser tomado para ser real (y por tanto es ortogonal real), entonces B se dice ser ortogonalmente equivalente (real) a A.

Teorema 1.6. Si $A = [a_{ij}]$ y $B = [b_{ij}]$ en M_n son unitariamente equivalente, entonces

$$\sum_{i,j=1}^{n} |b_{ij}|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$$

Prueba. Observemos que $\sum_{i,j} |a_{ij}|^2 = \text{tr}(A^*A)$, llevando a cabo la multiplicación de matrices. Luego, es suficiente verificar que $\text{tr}(B^*B) = \text{tr}(A^*A)$. Pero si $B = U^*AU$, entonces

$$\begin{aligned} \text{tr}(B^*B) &= \text{tr}(U^*A^*UU^*AU) \\ &= \text{tr}(U^*A^*AU) \\ &= \text{tr}(U^*UA^*A) \\ &= \text{tr}(A^*A) \end{aligned}$$

en una de las igualdades se usó $\text{tr}(XY) = \text{tr}(YX)$. ♦

Como la equivalencia unitaria implica semejanza, pero no es válida la recíproca, La relación de equivalencia unitaria particiona M_n en clases de equivalencia más fina que la relación de equivalencia semejanza. Unitariamente equivalente, al igual que la semejanza, corresponde a un cambio de base, pero de un tipo especial (el cambio de una base ortonormal a otra). Un cambio de base ortonormal no altera la suma de los cuadrados de los valores absolutos de las entradas, una cantidad que puede ser cambiado en un cambio de base no ortonormal.

Teorema de triangularización unitaria de Schur. Tal vez el hecho más fundamental en la teoría matricial elemental es que cualquier matriz $A \in M_n$ es unitariamente equivalente a una matriz triangular superior T (y también a una matriz triangular inferior). Los elementos de la diagonal de T son, por supuesto, los autovalores de A. Aunque esta forma no es la única, que representa la forma simple a alcanzar bajo la equivalencia unitaria.

Teorema 1.7 (Schur)**.** Dada la matriz $A \in M_n$ con autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$ en cualquier orden pre-determinado, existe una matriz unitaria $U \in M_n$ tal que

$$U^* A U = T = [t_{ij}]$$

es triangular superior, con entradas en la diagonal $t_{ii} = \lambda_i$, $i = 1, 2, \ldots, n$. Esto es, cada matriz cuadrada A es unitariamente equivalente a una matriz triangular cuyas entradas en la diagonal son los autovalores de A en el orden pre-determinado. Además, si $A \in M_n(\mathbb{R})$ y si todos los autovalores de A son reales, entonces U puede ser escogido para ser real y ortogonal.

Prueba. La prueba es algorítmica y procede de una secuencia de reducciones de tipo similar. Sea $x^{(1)}$ un autovector normalizado de A asociado al autovalor λ_1. El vector $x^{(1)}$ puede ser extendido a una base

$$x^{(1)}, y^{(2)}, y^{(3)}, \ldots, y^{(n)}$$

de $\mathbb{C}^n$. Aplicando el proceso de ortonormalización de Gram-Schmidt a esta base se obtiene una base ortonormal

$$x^{(1)}, z^{(2)}, z^{(3)}, \ldots, z^{(n)}$$

de $\mathbb{C}^n$. Poniendo estos vectores ortonormales de izquierda a derecha como columnas de una matriz, obtenemos una matriz unitaria U_1.

$U_1 = \left[x^{(1)} \mid z^{(2)} \ldots \mid sz^{(n)}\right]$ matriz unitaria de $n \times n$ y multiplicando por A se tiene:

$$AU_1 = \left[Ax^{(1)} \mid Az^{(2)} \mid \ldots \mid Az^{(n)}\right]$$

Luego

$$AU_1 = \left[\lambda_1 x^{(1)} \mid Az^{(2)} \mid \ldots \mid Az^{(n)}\right].$$

Ahora calculando $U_1^* A U_1$ tenemos:

$$U_1^* A U_1 = \begin{bmatrix} \overline{x}^{(1)}\lambda_1 x^{(1)} & \overline{x}^{(1)} A z^{(2)} & \cdots & \overline{x} A z^{(n)} \\ \overline{z}^{(1)}\lambda_1 x^{(1)} & \overline{z}^{(2)}\lambda_1 z^{(2)} & \cdots & . \\ \vdots & \vdots & \ddots & \vdots \\ \overline{z}^{(1)}\lambda_1 x^{(1)} & \cdots & \cdots & \overline{z}^{(n)} A z^{(n)} \end{bmatrix}$$

Como $\overline{x}^{(1)}\lambda_1 x^{(1)} = \lambda_1(\overline{x}^{(1)}x^{(1)}) = \lambda_1$ y los elementos en la triangular inferior se hacen ceros.

Por tanto se tiene

$$U_1^* A U_1 = \left[\begin{array}{c|c} \lambda_1 & * \\ \hline & \\ 0 & A_1 \end{array}\right]$$

La expresión $*$ representa a los elementos restantes en la triangular superior. Observemos que la matriz $A_1 \in M_{n-1}$ tiene autovalores $\lambda_2, \lambda_3, \ldots, \lambda_n$.

Sea $x^{(2)} \in \mathbb{C}^{n-1}$ un autovector ortonormalizado de A_1 correspondiente a λ_2, y procediendo del mismo modo, determinamos una matriz unitaria $U_2 \in M_{n-1}$ tal que

$$U_2^* A U_2 = \left[\begin{array}{c|c} \lambda_2 & * \\ \hline & \\ 0 & A_2 \end{array}\right]$$

Matriz de tamaño $(n-1) \times (n-1)$.

Consideremos la matriz unitaria V_2 de $n \times n$

$$V_2 = \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}$$

Las matrices V_2 y $U_1 V_2$ son entonces unitarias, y $V_2^* U_1^* A U_1 V_2$ tiene la forma

$$V_2^* U_1^* A U_1 V_2 = \begin{bmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & \\ O & & A_2 \end{bmatrix}$$

Como se puede observar este proceso es iterativo.

Continuando esta reducción, luego se producen matrices unitarias $U_i \in M_{n-i+1},\ i = 1, 2, \ldots, n$ y $V_i \in M_n$, $i = 2, \ldots, n-1$.

Luego la matriz $U = U_1 V_2 V_3 \ldots V_{n-1}$ es unitaria.

Por tanto $U^* A U$ tiene la forma:

$$\begin{pmatrix} \lambda_1 & & & & \\ 0 & \lambda_2 & & * & \\ 0 & 0 & \lambda_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix} = T$$

Es decir, $U^* A U = T$ donde $t_{ii} = \lambda_i$ con $i = 1, 2, \ldots, n$ y es claro que T es triangular superior.

Si todos autovalores de $A \in M_n(\mathbb{R})$ son reales, entonces los correspondientes autovectores pueden ser escogidos para ser reales y todos los pasos anteriores pueden ser llevados a la aritmética real, verificando la afirmación final. ♦

Ejemplo 1.2. Ni la matriz unitaria U ni la matriz triangular T del teorema son únicos. No solamente los elementos en la diagonal de T (los autovalores de A) aparecerán en cualquier orden, pero matrices triangulares superiores unitariamente equivalentes pueden parecer muy diferentes sobre la diagonal. Por ejemplo,

$$T_1 = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \text{ y } T_2 = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}$$

son unitariamente equivalente vía

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

En general, cualquier matriz triangular superior diferente puede estar en la misma clase de equivalencia bajo la relación unitariamente equivalente.

Implicaciones del Teorema de Schur. Como ilustración de una aplicación teórica del Teorema de Schur presentamos la demostración del resultado de Cayley-Hamilton.

El hecho de que toda matriz satisface su propia ecuación característica, se sigue del teorema de Schur y una simple observación sobre la multiplicación de matrices triangulares.

Lema 1.2. Supongamos que $R = [r_{ij}]$ y $T = [t_{ij}] \in M_n$ son triangulares superiores y que $r_{ij} = 0$, $1 \leq i, j \leq k < n$ y $t_{k+1,k+1} = 0$. Sea $T' = [t'_{ij}] = RT$. Entonces $t'_{ij} = 0$, $1 \leq i, j \leq k+1$.

Prueba. Puesto que $R([1, 2, \ldots, k]) = 0$ y $t_{k+1,k+1} = 0$, R y T tienen la forma

$$R = \begin{bmatrix} O & & * & \\ & * & & * \\ O & & \ddots & \\ & O & & * \end{bmatrix}, \quad T = \begin{bmatrix} * & & * & & & & \\ & \ddots & & & & * & \\ O & & * & & & & \\ & & & 0 & & & \\ & & & & * & & * \\ & & & & & \ddots & \\ & O & & & O & & * \end{bmatrix}$$

donde ambos bloques superiores izquierdos en las particiones son de $k \times k$. El bloque superior izquierdo de T' es claramente 0 por multiplicación particionada. Además, la inspección revela que las primeras filas $k+1$, de R tienen ceros en todas las posiciones no ceros de la columna $k+1$ de T, y que las primeras $k+1$ columnas de T tienen ceros en todas los posiciones no ceros de las filas $k+1$ de R. Multiplicando la Matriz muestra que T' (particionados en la misma manera) tiene la forma

$$T = \begin{bmatrix} & 0 & & & \\ O & \vdots & & * & \\ & 0 & & & \\ & & * & & \\ & \vdots & * & & \\ O & \vdots & \ddots & * & \\ & 0 & & & * \end{bmatrix}$$

y $T'([1, 2, \ldots, k+1]) = 0$, lo que muestra la afirmación. ♦

Teorema 1.8 (Cayley-Hamilton)**.** Sea $p_A(t)$ el polinomio característico de $A \in M_n$. Entonces

$$p_A(A) = 0$$

Prueba. Puesto que el polinomio $p_A(t)$ es de grado n con el primer coeficiente igual a 1 y las raíces de $p_A(t) = 0$ son precisamente los autovalores $\lambda_1, \ldots, \lambda_n$ de A, considerando incluso las multiplicidades, podemos escribir $p_A(t)$ como producto de factores en la forma

$$p_A(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$$

Por otra parte, podemos escribir A en la forma

$$A = UTU^*$$

donde T es triangular superior con λ_i en la i−ésima posición en la diagonal $i = 1, 2, \ldots, n$.

Ahora calculando

$$\begin{aligned} p_A(A) = p_A(UTU^*) &= (UTU^* - \lambda_1 I)(UTU^* - \lambda_2 I) \cdots (UTU^* - \lambda_n I) \\ &= \Big[U(T - \lambda_1 I)U^*\Big]\Big[U(T - \lambda_2 I)U^*\Big] \cdots \Big[U(T - \lambda_n I)U^*\Big] \\ &= U\Big[(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I)\Big]U^* \\ &= U p_A(T) U^* \end{aligned}$$

notemos que $p_A(A) = 0$ si, y sólo si $p_A(T) = 0$.

Sin embargo, por el Lema anterior podemos concluir que $p_A(T) = 0$. El bloque superior izquierdo de 1×1 de $T - \lambda_1 I$ es 0, y la 2, 2 entrada de $T - \lambda_2 I$ es 0, puesto que ambos son triangulares superiores, el bloque superior izquierdo de 2×2 de $(T - \lambda_1 I)(T - \lambda_2 I)$ es 0. Inductivamente, puesto que el bloque superior izquierdo de $k \times k$ de $(T - \lambda_1 I) \cdots (T - \lambda_{k+1} I)$ es 0. Continuando hasta n nos permite concluir que el producto $p_A(T) = (T - \lambda_1 I) \cdots (T - \lambda_n I) = 0$, lo cual completa la prueba. ♦

Otros hechos interesantes se observan en los siguientes teoremas.

Teorema 1.9. Si A tiene autovalores $\lambda_1, \ldots, \lambda_n$, entonces

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(A)$$

Prueba. Por el Teorema de Schur, existe una matriz unitaria U tal que $U^*AU = T$, donde T es una matriz triangular superior con los λ_i en la diagonal de T, es decir, $\lambda_i = t_{ii}, \quad i = 1, 2, \ldots, n$.
Por tanto

$$\begin{aligned}\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} t_{ii} &= \mathrm{tr}(T) = \mathrm{tr}(U^*AU) \\ &= \mathrm{tr}(AUU^*) = \mathrm{tr}(A)\end{aligned}$$

♦

Teorema 1.10. Si A tiene autovalores $\lambda_1, \ldots, \lambda_n$, entonces

$$\prod_{i=1}^{n} \lambda_i = \det(A)$$

Prueba. Por el Teorema de Schur, existe una matriz unitaria U tal que $U^*AU = T$, donde T es una matriz triangular superior con los λ_i en la diagonal de T, es decir, $\lambda_i = t_{ii}, \quad i = 1, 2, \ldots, n$.
Por tanto

$$\begin{aligned}\prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} t_{ii} &= \det(T) = \det(U^*AU) \\ &= \det(U^*)\det(A)\det(U) \\ &= \det(A)\end{aligned}$$

♦

Matrices Normales. La clase de matrices normales, los cuales surgen naturalmente en el contexto de la equivalencia unitaria, es importante en todo el análisis matricial y en general unitaria, simétrica real y matrices hermitianas.

Definición 1.9. Una matriz $A \in M_n$ se dice ser normal si $A^*A = AA^*$, es decir, si A conmuta con su adjunta Hermitiana.

Ejemplos inmediatos de matrices normales son:

1. Puesto que $U^*U = I = UU^*$ si U es unitaria, entonces toda matriz unitaria es normal.

2. Si $A^* = A$, entonces $A^*A = AA^*$. Luego toda matriz hermitiana es normal.

3. Si $A \in M_n$ es tal que $A^* = -A$, entonces A es llamada anti-hermitiana y es normal.

En este caso $A^*A = -A^2 = A^*$ luego toda matriz anti-hermitiana es normal.

Observemos que la matriz $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ es normal, pero no cae en ninguna de las categorías anteriores.

Definición 1.10. Si $A \in M_n$ es unitariamente equivalente a una matriz diagonal, A se dice ser unitariamente diagonizable, con una definición similar para ortogonalmente diagonizable.

Observemos que, unitariamente diagonalizable (u ortogonalmente), implica diagonalizable (pero no es válida la recíproca).

Algunas equivalencias importantes en la caracterización de matrices normales es dado en el siguiente teorema.

Teorema 1.11. Si $A = [a_{ij}] \in M_n$ tiene autovalores $\lambda_1, \ldots, \lambda_n$ los siguientes son equivalentes:

(1) A es normal;

(2) A es unitariamente diagonalizable;

(3) $\displaystyle\sum_{i,j=1}^{n} = |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$; y

(4) Existe un conjunto ortonormal de n autovectores de A.

Prueba. En lo que sigue supongamos que $T = [t_{ij}] \in M_n$ es una matriz triangular superior el cual es unitariamente equivalente a A, por el Teorema de Schur se tiene que $T = U^*AU$ para algún $U \in M_n$ luego:

$$\begin{aligned}
T^*T &= (U^*AU)^*(U^*AU) \\
&= (U^*A^*U)(U^*AU) \\
&= (U^*A^*)(UU^*)(AU) \\
&= (U^*A^*)I(AU) \\
&= U^*(A^*A)U && A \text{ es normal} \\
&= U^*AA^*U \\
&= U^*AUU^*A^*U \\
&= (U^*AU)(U^*AU)^* \\
&= TT^*
\end{aligned}$$

Por tanto T es normal.

Mostremos que (1) es equivalente a (2), (2) es equivalente a (3) y (2) es equivalente a (4).

Veamos (1) $\Leftrightarrow$ (2). Si A es normal, entonces T es normal. Pero una matriz triangular normal debe ser diagonal, basta ver las entradas diagonales de T^*T y TT^*. El hecho que la entrada 1, 1 de T^*T es el mismo que TT^* significa que

$$\bar{t}_{11}t_{11} = t_{11}\bar{t}_{11} + \sum_{j=2}^{n} t_{1j}\bar{t}_{1j} = |t_{11}|^2 + \sum_{j=2}^{n} |t_{1j}|^2$$

Esto significa que $0 = \sum_{j=2}^{n} |t_{1j}|^2$.
Luego la suma de términos no negativos, cada uno debe ser 0.

Por tanto concluimos

$$t_{1j} = 0, \quad j = 2, 3, \ldots, n$$

El hecho que la entrada 2, 2 de T^*T y TT^* son los mismos, entonces significa que

$$\bar{t}_{22}t_{22} = t_{22}\bar{t}_{22} + \sum_{j=3}^{n} t_{2j}\bar{t}_{2j} = |t_{22}|^2 + \sum_{j=3}^{n} |t_{2j}|^2$$

y concluimos por la misma razón anterior que

$$t_{2j} = 0, \quad j = 3, 4, \ldots, n$$

De la misma manera, asumiendo que se tiene verificado que

$$t_{ij} = 0, \quad j > i \;\; i = 1, 2, \ldots, k-1$$

concluimos que

$$t_{ij} = 0, \quad j > i, \quad i = k$$

Procediendo de la misma manera sucesivamente en cada entrada de la diagonal, se concluye finalmente que

$$t_{ij} = 0, \quad j > i, \quad i = 1, 2, \ldots, n$$

y como T es triangular superior tenemos

$$t_{ij} = 0, \quad j < i, \quad i = 1, 2, \ldots, n$$

porque T es triangular superior, se tiene que T es diagonal y (2) es válido. Puesto que matrices diagonales son claramente normales y equivalencia unitaria preserva normalidad, (2) implica (1).

(2) $\Leftrightarrow$ (3).

Como A es unitariamente diagonalizable y por el teorema de Schur tenemos que:

$$A = U^*TU$$

donde T contiene a los autovalores de A en su diagonal.

Luego

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2$$

pero T debe ser D, por tanto $\sum_{i<j}^{n} |t_{ij}|^2 = 0$, es decir,

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$$

Por tanto (2) implica (3).
Además, por Schur

$$A = UTU^*$$

es decir, A es equivalente a T, donde T es triangular superior.

$$\sum_{i,j}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2$$

Luego tomando la hipótesis

$$\sum_{i<j} |t_{ij}|^2 = 0$$

En consecuencia T debe ser diagonal y por tanto A es unitariamente diagonizable.

Luego (3) implica (2).

La equivalencia de (2) y (4) se sigue del hecho que T es triangular y si la $i-$ésima entrada en la diagonal de T^*T es la misma que de TT^*. ♦

1.3. Matrices Hermitianas y Simétricas

La clase de matrices complejas simétricas no tienen muchas propiedades importantes que la clase de matrices simétricas reales. Estudiaremos matrices complejas hermitianas y simétricas e indicaremos aspectos especiales de lo que sucede en el caso de matrices simétricas con entradas reales.

Definición 1.11. Una matriz $A = [a_{ij}] \in M_n$ se dice Hermitiana si $A = A^*$, donde $A^* = \overline{A}^T = [\overline{a}_{ji}]$. Este es anti-hermitiana si $A = -A^*$.

Algunas observaciones para $A, B \in M_n$ se resume en lo que sigue:

1. $A + A^*$, AA^*, y A^*A son todas Hermitianas para todo $A \in M_n$.
2. Si A es Hermitiana, entonces A^k es Hermitiana para todo $k = 1, 2, \ldots,$. Si A es no singular, entonces A^{-1} es Hermitiana.
3. Si A, B son Hermitianas, entonces $aA + bB$ es Hermitiana para todo a y b reales.
4. $A - A^*$ es anti-hermitiana para todo $A \in M_n$.
5. Si A, B son anti-hermitianas, entonces $aA + bB$ es anti-hermitiana para todo a y b escalar real.

6. Si A es hermitiana, entonces iA es hermitiana.

7. Si A es anti-hermitiana, entonces iA es anti-hermitiana.

8. Cualquier $A \in M_n$ puede ser escrito

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) = H(A) + S(A)$$

donde $H(A) = \frac{1}{2}(A + A^*)$ es la parte Hermitiana de A y $S(A) = \frac{1}{2}(A - A^*)$ es la parte anti-hermitiana de A.

9. Si A es Hermitiana, las entradas en la diagonal principal de A son todos reales.

Teorema 1.12. Cada $A \in M_n$ puede ser escrito únicamente como $A = S + iT$, donde S y T son Hermitianas. Este también puede ser escrito únicamente como $A = B + C$, donde B es Hermitina y C es anti-hermitiana.

Prueba. Escribiendo

$$A = \frac{1}{2}(A + A^*) + i\left(-\frac{i}{2}\right)(A - A^*)$$

donde $S = \frac{1}{2}(A + A^*)$ y $T = (-\frac{i}{2})(A - A^*)$ son matrices hermitianas. Para la unicidad, sea $A = E + iF$ con E y F hermitianas (además $E \neq S$, $F \neq T$), entonces

$$2S = A + A^* = (E + iF) + (E + iF)^* = E + iF + E^* - iF^* = 2E$$

luego $S = E$. De forma similar, uno muestra que $F = T$. Para la afirmación $A = B + C$ esta se prueba de manera similar. ♦

Debemos notar que si pensamos a M_n como los números complejos, entonces las matrices hermitianas son análogos a los números reales. La analogía de las operaciones de la conjugación compleja es la operación $*$ (adjunta) sobre M_n. Como un real es un número complejo z tal que $z = \bar{z}$; luego una matriz hermitiana es una matriz $A \in M_n$ tal que $A = A^*$.

Teorema 1.13. Sea $A \in M_n$ una matriz hermitiana. Entonces

1. x^*Ax es real para todo $x \in \mathbb{C}^n$,

2. Todos los autovalores de A son reales,

3. S^*AS es hermitiana para todo $S \in M_n$.

Prueba. Veamos (1):

Calculando $\overline{(x^*Ax)} = (x^*Ax)^* = x^*A^*x = x^*Ax$, luego x^*Ax es igual a su conjugado complejo y por tanto es real.

Para (2): Si $Ax = \lambda x$ y $x^*x = 1$, entonces $\lambda = \lambda x^*x = x^*\lambda x = x^*Ax$ es real por 1.

Para (3): $(S^*AS)^* = S^*A^*S = S^*AS$ luego S^*AS es siempre hermitiana. ♦

Los resultados anteriores son de hecho (casi) una caracterización de la matrices hermitianas.

Teorema 1.14. Sea $A = \left[a_{ij}\right] \in M_n$ una matriz dada. Entonces A es hermitiana si y sólo si al menos una de las siguientes afirmaciones:

1. x^*Ax es real para todo $x \in \mathbb{C}^n$.
2. A es normal y todos los autovalores de A son reales.
3. S^*AS es hermitiana para todo $S \in M_n$.

Prueba. Basta probar solo la suficiencia de cada condición.

Veamos en 1.:

Para $x, y \in \mathbb{C}^n$ se cumple x^*Ax y y^*Ay son reales.

Y como $x + y \in \mathbb{C}^n$, entonces $(x + y)^*A(x + y)$ también debe ser real.

Ahora

$$(x + y)^*A(x + y) = (x^*Ax + y^*Ay) + (x^*Ay + y^*Ax)$$

es real para todo $x, y \in \mathbb{C}^n$.

Como x^*Ax y y^*Ay son reales.

Luego

$$x^*Ay + y^*Ax \tag{1'}$$

debe ser real.

Veamos este último:

Sea $x = e_k$ e $y = e_j$ (donde e_k, e_j tienen en la posición $k-$ésima y $j$ésima el valor 1 y en los demás 0's).

Y reemplazando en (1') obtenemos

$$a_{kj} + a_{jk} \text{ es real}$$

Por tanto: $\mathrm{Im}(a_{kj}) = -\mathrm{Im}(a_{jk})$.

Ahora tomando $x = ie_k$, $y = e_j$.

Similarmente al anterior caso implicamos que $-ia_{kj} + ia_{jk}$ es real.

Por tanto $\mathrm{Re}(a_{kj}) = \mathrm{Re}(a_{jk})$.

Finalmente de las relaciones

$$\mathrm{Re}(a_{kj} = \mathrm{Re}(jk)$$
$$\mathrm{Im}(a_{kj}) = \mathrm{Im}(a_{jk})$$

se tiene $a_{kj} = \overline{a}_{jk}$ y como k, j son arbitrarias concluimos que $A = A^*$.

Por tanto A es hermitiana.

Veamos en 2.:

Si A es normal, este es unitariamente diagonalizable, lo cual implica que

$$A = U\Delta U^*$$

Luego,

$$\begin{aligned} A^* &= (U\Delta U^*)^* \\ &= U\Delta^* U^* \\ &= U\overline{\Delta}U^* \end{aligned}$$

Como Δ es diagonal, luego $\Delta = \overline{\Delta}$.

Y por tanto $A^* = U\Delta U^*$.

Es decir, $A = A^*$.

La última condición implica que A es hermitiana, basta tomar $S = I$. ♦

Puesto que una matriz hermitiana es obviamente normal ($AA^* = A^2 = A^*A$), todos los resultados acerca de matrices normales son válidos para las hermitianas. Por ejemplo, autovectores correspondientes a distintos autovalores son ortogonales, existe un conjunto completo de autovectores ortonormales; matrices hermitianas son unitariamente diagonalizable.

Puesto que los autovalores de una matriz hermitiana $A \in M_n$ son reales, siempre podemos suponer que están ordenados en forma creciente.

$$\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max}$$

CAPÍTULO 2

Normas Matriciales

Considerando diferentes vectores en $\mathbb{C}^n$ o varias matrices en M_n ¿qué podría significar decir que algunos son "pequeños" o cuales son más "grandes" ¿Qué significa que dos vectores están "próximos o lejos"?

Una manera de responder estas preguntas es el estudio de normas, o medidas de tamaño, de matrices y vectores. Las normas pueden ser a través de una generalización de la longitud Euclidiana, sin embargo el estudio de normas es más que un ejercicio matemático de generalización. Es necesario para un formulación propia de nociones tal como series de potencias de matrices y es esencial en el análisis y la formulación de algoritmos para computación numérica.

2.1. Normas Matriciales

En esta sección estudiaremos el concepto de norma Matricial, para esto podemos realizar una interpretación de M_n como un espacio vectorial de dimensión n^2, en este sentido podemos "medir el tamaño" de una matriz por considerar cualquier norma en $\mathbb{C}^{n^2}$.

Sin embargo, M_n no es sólo un espacio vectorial de dimensión muy grande; este tiene definido una operación de multiplicación muy natural, y es muy frecuentemente realizar estimaciones sobre el tamaño de AB con respecto del los tamaños de A y B. Por esta razón es conveniente aumentar un concepto adicional en la definición de norma vectorial.

Definición 2.1. Sea M_n el espacio vectorial de las matrices de tamaño $n \times n$.

La función $|||\cdot||| : M_n \rightarrow \mathbb{R}$ es una norma matricial si para todo $A, B \in M_n$ este satisface las siguientes condiciones:

1. $|||A||| \geq 0$. (No negativa)

2. $|||A||| = 0$ si y sólo si $A = 0$. (Positiva)

3. $|||cA||| = |c|\,|||A|||$, para cualquier c complejo. (Homogéneo)

4. $|||A + B||| \leq |||A||| + |||B|||$. (Desigualdad Triangular)

5. $|||AB||| \leq |||A|||\,|||B|||$. (Sub multiplicativa)

Cabe notar que las cuatro primeras condiciones coinciden con la de normas vectoriales, por lo cual una norma matricial es una norma vectorial, la recíproca no es cierta, vale decir que las normas vectoriales en general no satisfacen la última condición de la definición. Razón por la cual esta norma es llamada Norma Matricial Generalizada. De forma similar al caso de las normas vectoriales, si se omite la segunda condición se obtiene el concepto de seminorma matricial.

Algunos hechos interesantes con respecto a la definición de normas matriciales.

1. En principio, para cualquier norma matricial se tiene
$$|||A^2||| = |||AA||| \leq |||A|||\,|||A||| = |||A|||^2$$

2. Si A es una matriz no cero tal que $A^2 = A$ se tiene $|||A^2||| = |||A|||$, entonces $|||A||| \leq |||A|||^2$, luego $0 \leq |||A|||^2 - |||A|||$, así $|||A|||\big(|||A||| - 1\big) \geq 0$, por tanto
$$\begin{aligned} |||A||| - 1 &\geq 0 \\ |||A||| &\geq 1 \end{aligned}$$

3. De lo anterior si $A = I$ la matriz identidad, se cumple $I^2 = I$, entonces $|||I||| \geq 1$.

4. Si A es una matriz invertible, de donde se cumple $I = AA^{-1}$, entonces para cualquier norma matricial se tiene $|||I||| = |||AA^{-1}||| \leq |||A|||\,|||A^{-1}|||$, así
$$|||A^{-1}||| \geq \frac{|||I|||}{|||A|||}$$

5. En general se satisface $|||A^k||| \leq |||A|||^k$, para cualquier norma matricial y $k \in \mathbb{N}$ y para toda matriz $A \in M_n$.

En lo que sigue describiremos algunas normas sobre el espacio vectorial M_n, algunas como extensiones de las normas vectoriales.

La norma l_1. Para cualquier matriz $A \in M_n$, se define como
$$\|A\|_1 = \sum_{i,j=1}^{n} |a_{ij}|$$
es una norma matricial.

Es claro que las primeras 4 condiciones de norma matricial se cumplen pues, es similar a la norma vectorial. Luego es suficiente verificar la condición 5. de la definición.

$$\begin{aligned}
\|AB\|_1 &= \sum_{i,j=1}^{n}\left|\sum_{k=1}^{n} a_{ik}b_{kj}\right| \leq \sum_{i,j,k=1}^{n} |a_{ik}b_{kj}| \\
&\leq \sum_{i,j,k,m=1} |a_{ik}b_{mj}| = \left(\sum_{i,k=1}^{n}|a_{ik}|\right)\left(\sum_{j,m=1}^{n}|b_{mj}|\right) \\
&= \|A\|_1\|B\|_1
\end{aligned}$$

La primera desigualdad se sigue de la desigualdad triangular, mientras que la segunda es dada por adicionar sumandos a la suma.

La norma l_2. Esta norma también es conocida como la norma Euclidiana y es definida para cualquier matriz $A \in M_n$ por

$$\|A\|_2 = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

es norma matricial.

Similar al ejemplo anterior solo falta verificar la condición 5.

$$\begin{aligned}
\|AB\|_2^2 &= \sum_{i,j=1}^{n}\left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|^2 \leq \sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{m=1}^{n}|b_{mj}|^2\right) \\
&= \left(\sum_{i,k=1}^{n}|a_{ik}|^2\right)\left(\sum_{j,m=1}^{n}|b_{mj}|^2\right) = \|A\|_2^2\|B\|_2^2
\end{aligned}$$

Esta es justamente la desigualdad de Cauchy-Schwarz. Esta norma matricial es llamado la Norma de Frobenius, o la Norma Schur, o la Norma de Hilbert-Schmidt.
Otra manera de construir nuevas normas es dado en el siguiente teorema:

Teorema 2.1. Si $\|\cdot\|$ es una norma vectorial sobre $\mathbb{C}^n$ y $T \in M_n$ es no singular, entonces $\|\cdot\|_T$ definido por $\|x\|_T \equiv \|Tx\|$, $x \in \mathbb{C}^n$, es también una norma vectorial sobre $\mathbb{C}^n$.

Demostración. La prueba es inmediata puesto que $Tx \in \mathbb{C}^n$ y $\|\cdot\|$ es una norma, puesto que

1. $\|x\|_T = \|Tx\| \geq 0$.
2. $x = 0$, entonces $\|x\|_T = \|Tx\| = \|0\| = 0$. Si $\|x\|_T = 0$, entonces $\|Tx\| = 0$, luego $Tx = 0$, como T es invertible, entonces $x = 0$.
3. $\|cx\|_T = \|T(cx)\| = \|cTx\| = |c|\,\|Tx\| = |c|\,\|x\|_T$, donde c es un escalar.
4. $\|x + y\|_T = \|T(x + y)\| = \|Tx + Ty\| \leq \|Tx\| + \|Ty\| = \|x\|_T + \|y\|_T$.

♦

Antes de continuar con los ejemplos, observemos que si $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \in M_n$ escrito en términos de sus columnas, donde las columnas $a_i \in \mathbb{C}^n$, entonces

$$\|A\|_2^2 = \|a_1\|_2^2 + \|a_2\|_2^2 + \cdots + \|a_n\|_2^2 .$$

Dado que la norma l_2 sobre $\mathbb{C}^n$ es unitariamente invariante, se tiene el hecho importante

$$\|UA\|_2^2 = \|Ua_1\|_2^2 + \|Ua_2\|_2^2 + \cdots + \|Ua_n\|_2^2 = \|a_1\|_2^2 + \|a_2\|_2^2 + \cdots + \|a_n\|_2^2 = \|A\|_2^2$$

donde $U \in M_n$ es unitaria. Puesto que $\|B^*\|_2 = \|B\|_2$ para todo $B \in M_n$, esto implica que

$$\|UAV\|_2 = \|AV\|_2 = \|V^*A^*\|_2 = \|A^*\|_2 = \|A\|_2$$

donde $U, V \in M_n$ son unitarias. Luego, la norma l_2 sobre M_n es una norma matricial unitariamente invariante.

La norma l_∞. Para $A \in M_n$ se define por

$$\|A\|_\infty = \max_{1\leq i,j\leq n} |a_{ij}|$$

es una norma sobre el espacio vectorial M_n pero no es una norma matricial.

Por ejemplo podemos considerar las matrices

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \quad AB = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

Luego, $\|A\|_\infty = 1$, $\|B\|_\infty = 2$ y $\|AB\|_\infty = 4$. Por tanto no se cumple la condición 5. de la definición de norma matricial.

Sin embargo con alguna modificación se puede lograr una norma matricial la cual es dada por

$$|||A||| = n\|A\|_\infty, \qquad A \in M_n$$

entonces se tiene

$$\begin{aligned} \|AB\| &= n \max_{1\leq i,j\geq n} \left| \sum_{k=1}^{n} a_{ik}b_{kj} \right| \leq n \max_{1\leq i,j\leq n} \sum_{k=1}^{n} |a_{ik}b_{kj}| \\ &\leq n \max_{1\leq i,j\leq n} \sum_{k=1}^{n} \|A\|_\infty \|B\|_\infty = n\|A\|_\infty n\|B\|_\infty \\ &= |||A|||\,|||B||| \end{aligned}$$

De forma similar a lo anterior, es posible construir normas matriciales, las cuales son inducidas por normas vectoriales como describiremos en adelante.

Definición 2.2. Sea $\|\cdot\|$ una norma vectorial sobre $\mathbb{C}^n$. Definimos la norma $|||\cdot|||$ sobre M_n por

$$|||A||| = \max_{\|x\|=1} \|Ax\|$$

Notemos que el "**máx**" de la definición (a veces se utiliza el supremo) es justificado puesto que $\|Ax\|$ es una función continua en x y la bola unitaria $B_{|\cdot|}$ es un conjunto compacto.

Otras equivalencias de la anterior definición son resumidas en la siguiente lista:

$$\begin{aligned}
|||A||| &= \max_{\|x\|=1} \|Ax\| \\
&= \max_{\|x\|\leq 1} \|Ax\| \\
&= \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} \\
&= \max_{\|x\|_\alpha=1} \frac{\|Ax\|}{\|x\|}
\end{aligned}$$

donde $\|\cdot\|_\alpha$ es cualquier norma vectorial.

Teorema 2.2. La función $|||\cdot|||$ de la Definición 2.2 es una norma matricial sobre M_n, $\|Ax\| \leq |||A|||\,\|x\|$ para todo $A \in M_n$ y todo $x \in \mathbb{C}^n$ y $|||I||| = 1$.

Prueba. El primer axioma, se sigue de la definición de $|||A|||$ pues el máximo de una función es no negativa y el segundo axioma, se sigue del hecho que $Ax = 0$ para todo x, es precisamente cuando $A = 0$.

El tercer axioma, se sigue del siguiente cálculo,

$$|||cA||| = \max \|cAx\| = \max |c|\,\|Ax\| = |c| \max \|Ax\| = |c|\,|||A|||$$

De forma similar, la desigualdad triangular es inherente a la propiedad del máximo, puesto que

$$\begin{aligned}
|||A+B||| &= \max \|(A+B)x\| \\
&= \max \|Ax + Bx\| \\
\max &\leq (\|Ax\| + \|Bx\|) \\
&\leq \max \|Ax\| + \max \|Bx\| \\
&= |||A||| + |||B|||
\end{aligned}$$

El axioma sub-multiplicativo se sigue del hecho que

$$|||AB||| = \max \frac{\|ABx\|}{\|x\|} = \max \frac{\|ABx\|}{\|Bx\|}\frac{\|Bx\|}{\|x\|} \leq \max \frac{\|Ay\|}{\|y\|} \max \frac{\|Bx\|}{\|x\|} = |||A|||\,|||B|||$$

donde asumimos sin pérdida de generalidad, que el máximo es tomado solo sobre los x que no están en el espacio nulo de B. Para la próxima afirmación, observemos que si $x \neq 0$, entonces

$$\left\| \frac{Ax}{\|x\|} \right\| \leq |||A|||\,.$$

Por homogeneidad de la norma vectorial, obtenemos $\|Ax\| \leq |||A|||\,\|x\|$, el cual también vale cuando $x = 0$. Finalmente,

$$|||I||| = \max_{\|x\|=1} \|Ix\| = \max_{\|x\|=1} \|x\| = 1$$

♦

Definición 2.3. Decimos que la norma matricial $|||\cdot|||$ definido anteriormente es la norma matricial inducido por la norma vectorial $\|\cdot\|$. Este algunas veces es llamado el operador norma asociado con la norma vectorial $\|\cdot\|$.

Notemos que el operador norma es una norma matricial como una consecuencia de las propiedades de todas las normas vectoriales. Por tanto, una manera de probar que una cierta función sobre M_n es una norma matricial es mostrar que esta es inducida por alguna norma vectorial. Adoptamos esta estrategia cuando estudiemos una norma matricial importante llamado la norma espectral.

A continuación listaremos ejemplos de normas matriciales inducidos por normas vectoriales.

Norma matricial suma máxima de columnas. Definimos sobre M_n por,

$$|||A|||_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$$

La norma anterior es inducido por la norma vectorial l_1 y por tanto debe ser una norma matricial.

Norma matricial suma máxima de filas. Se define sobre M_n por,

$$|||A|||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$$

La norma definida es inducida por la norma vectorial l_∞ y por tanto debe ser norma matricial.

La norma espectral. Es definido sobre M_n por,

$$|||A|||_2 = \max\left\{ \sqrt{\lambda} \ : \ \lambda \ \text{ es un autovalor de } \ A^*A \right\}$$

Notemos que si $A^*Ax = \lambda x$, $x \neq 0$, entonces

$$\begin{aligned} x^*A^*Ax &= x^*\lambda x \\ (Ax)^*Ax &= \lambda x^*x \\ \|Ax\|^2 &= \lambda\|x\|^2 \end{aligned}$$

Por tanto $\lambda \geq 0$ y $\sqrt{\lambda}$ es real y no negativa.

Teorema 2.3. Si $|||\cdot|||$ es una norma matricial sobre M_n y si $S \in M_n$ es no singular, entonces

$$|||A|||_S = |||S^{-1}AS||| \text{ para todo } A \in M_n$$

es una norma matricial.

Prueba. Los axiomas 1., 2., 3., y 4. son verificados directamente para $|||\cdot|||$.

La sub-multiplicidad se obtiene del siguiente cálculo,

$$\begin{aligned}|||AB|||_S &= |||S^{-1}ABS|||\\ &= |||(S^{-1}AS)(S^{-1}BS)|||\\ &\leq |||(S^{-1}AS)|||\;|||(S^{-1}BS)|||\\ &= |||A|||_S|||B|||_S\end{aligned}$$

♦

Una importante área de aplicación de las normas matriciales esta en dar una cota para el espectro de una matriz.

Definición 2.4. El radio espectral $\rho(A)$ de una matriz $A \in M_n$ es dado por

$$\rho(A) = \text{máx}\left\{|\lambda| \,:\, \lambda \quad \text{es un autovalor de} \quad A\right\}$$

Observemos que si λ es cualquier autovalor de A, entonces $|\lambda| \leq \rho(A)$; además, existe un autovalor λ' para el cual $|\lambda'| = \rho(A)$. Si $Ax = \lambda' x$, $x \neq 0$, y si $|\lambda| = \rho(A)$, consideremos la matriz $X \in M_n$ en la cual todas la columnas son iguales al autovector x, y observe que $AX = \lambda' X$. Si $|||\cdot|||$ es cualquier norma matricial, se tiene

$$|\lambda'|\,|||X||| = |||\lambda' X||| = |||AX||| \leq |||A|||\;|||X|||$$

y por tanto $|\lambda'| = \rho(A) \leq |||A|||$. Esto muestra el siguiente teorema.

Teorema 2.4. Si $|||\cdot|||$ es cualquier norma matricial y si $A \in M_n$, entonces $\rho(A) \leq |||A|||$.

Un resultado que será de mucha utilidad en la demostración del teorema de perturbación es dado en el siguiente lema.

Lema 2.1. Sea $A \in M_n$ y $\epsilon > 0$ dados. Existe una norma matricial $|||\cdot|||$ tal que

$$\rho(A) \leq |||A||| \leq \rho(A) + \epsilon\,.$$

Prueba. Por el teorema de Schur, existe una matriz unitaria U y una matriz triangular superior Δ tal que $A = U^*\Delta U$.

Sea $D_t = \text{diag}(t, t^2, t^3, \ldots, t^n)$ y calculando

$$D_t\Delta D_t^{-1} = \begin{bmatrix}\lambda_1 & t^{-1}d_{12} & t^{-2}d_{13} & \cdots & t^{-n+1}d_{1n}\\ 0 & \lambda_2 & t^{-1}d_{23} & \cdots & t^{-n+2}d_{2n}\\ 0 & 0 & \lambda_3 & \cdots & t^{-n+3}d_{3n}\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & 0 & \cdots & t^{-1}d_{(n-1)n}\\ 0 & 0 & 0 & \cdots & \lambda_n\end{bmatrix}$$

Luego, para $t > 0$ suficientemente grande, podemos afirmar que la suma de todos los valores absolutos de las entradas en la diagonal de $D_t \Delta D_t^{-1}$ es menor que ϵ. En particular podemos asegurar que $|||D_t \Delta D_t^{-1}|||_1 \leq \rho(A) + \epsilon$ para un t suficientemente grande.

Luego, si definimos la norma matricial $|||\cdot|||$ por

$$|||B||| = |||D_t U^* B U D_t^{-1}|||_1 = |||(UD_t^{-1})^{-1} B (UD_t^{-1})|||_1$$

para cualquier $B \in M_n$ y si escogemos t suficientemente grande, entonces tendremos construido una norma matricial tal que $|||A||| \leq \rho(A) + \epsilon$. Puesto que $|||A||| \geq \rho(A)$ para cualquier norma matricial, obtenemos el resultado. ♦

Estamos interesados en caracterizar matrices A tales que $A^k \to 0$ cuando $k \to \infty$. El siguiente resultado es una buena herramienta para este cometido.

Lema 2.2. Sea $A \in M_n$ una matriz dada. Si existe una norma matricial $|||\cdot|||$ tal que $|||A||| < 1$, entonces

$$\lim_{k\to\infty} A^k = 0\,;$$

esto es, todas las entradas de A^k tienden a cero cuando $k \to \infty$.

Prueba. Si $|||A||| < 1$ entonces $|||A|||^2 < |||A|||$. Por tanto $|||A|||^2 < 1$.

Siguiendo este proceso luego tenemos:

$$|||A|||^k < 1 \quad \text{con} \quad k = 1, 2, \ldots$$

pero sabemos que $|||A^k||| \leq |||A|||^k$, por tanto tenemos $0 \leq |||A^k||| < 1$.
Ahora aplicando límites:

$$0 \leq \lim_{k\to\infty} |||A^k||| < 1$$

Luego $\lim_{k\to\infty} |||A^k||| = 0$ entonces $\left|\left|\left|\lim_{k\to\infty} A^k\right|\right|\right| = 0$.
Por tanto $\lim_{k\infty} A^k = 0$ ♦

Notemos que las matrices $A \in M_n$ tales que $\lim_{k\to\infty} A^k = 0$ son llamados convergentes y son importantes en varias aplicaciones, por ejemplo en el análisis de procesos iterativos.

Teorema 2.5. Sea $A \in M_n$. Entonces $\lim_{k\to\infty} A^k = 0$ si, y sólo si $\rho(A) < 1$.

Prueba. Si $\lim_{k\to\infty} A^k = 0$.
Sea $Ax = \lambda x$, con $x \neq 0$ y λ su autovalor correspondiente.

Luego

$$
\begin{aligned}
AAx &= A\lambda x\\
A^2x &= \lambda Ax\\
A^2x &= \lambda\lambda x\\
A^2x &= \lambda^2 x\\
&\vdots\\
A^kx &= \lambda^k x, \quad k \in \mathbb{Z}^+
\end{aligned}
$$

Como $A^k \to 0$, entonces $A^kx \to 0x$
Luego $A^kx \to 0$, entonces $\lambda^k x \to 0$; sólo si $|\lambda| < 1$.

Ahora como λ es cualquier autovalor, por tanto se tiene

$$\rho(A) < 1\,.$$

Recíprocamente si $\rho(A) < 1$, luego existe una norma matricial $|||\cdot|||$ tal que $\rho(A) \leq |||A||| \leq \rho(A)+\epsilon$.

Por tanto $|||A||| < 1$. Ahora por el lema 2.2 se tiene que $\lim\limits_{k\to 0} A^k = 0$ ♦

Una consecuencia de este resultado es dado en el siguiente corolario.

Corolario 2.1. Sea $|||\cdot|||$ una norma matricial en M_n. Entonces

$$\rho(A) = \lim_{k\to\infty} |||A^k|||^{1/k}$$

para todo $A \in M_n$.

Prueba. Como $\rho(A)^k = \rho(A^k) \leq |||A^k|||$, se tiene que $\rho(A) \leq |||A^k|||^{1/k}$ para todo $k = 1, 2, \ldots$.

Si dado un $\epsilon > 0$, la matriz $\widetilde{A} = [\rho(A) + \epsilon]^{-1}A$ tiene radio espectral estrictamente menor que 1 y por tanto este converge. Luego $|||\widetilde{A}^k|| \to 0$ cuando $k \to \infty$ y por tanto existe algún $N = N(\epsilon, A)$ tal que $|||\widetilde{A}^k||| < 1$ para todo $k \geq N$.

Luego se establece que $|||A^k||| \leq [\rho(A) + \epsilon]^k$ para todo $k \geq N$, o que

$$|||A^k|||^{1/k} \leq \rho(A) + \epsilon$$

para todo $k \geq N$.

Puesto que $\rho(A) \leq |||A^k|||^{1/k}$ para todo k y puesto que $\epsilon > 0$ es arbitrario, se concluye que

$$\lim_{k\to\infty} |||A^k|||^{1/k}$$

existe y es igual a $\rho(A)$. ♦

CAPÍTULO 3

Teorema de Gershgorin

Este capítulo constituye la parte principal de la monografía en la misma desarrollaremos los teoremas sobre localización de autovalores y el teorema de perturbación.

Para diversas matrices hallar autovalores es una tarea sencilla, tal es el caso de las matrices diagonales, matrices triangulares (superior o inferior). Sin embargo en la mayoría de los casos no es tan inmediato encontrar los autovalores, en estos casos es muy útil tener una aproximación de la región donde se ubican los autovalores. Describir resultados que especifican las regiones donde se encuentran los autovalores de una matriz nos proporciona una buena estimativa. Por otra parte estudiar el comportamiento de los autovalores en esta región cuando las matrices están sujetas a perturbaciones en sus entradas, también es de importancia en teoría matricial que permite determinar si un sistema es robusto o no.

En términos matemáticos, uno quiere localizar los autovalores de una matriz en un conjunto acotado que son caracterizados fácilmente. Se conoce que todos los autovalores de una matriz A son localizados en un disco en el plano complejo centrado en el origen y que tiene radio $|||A|||$, donde $|||\cdot|||$ es cualquier norma matricial. ¿Es posible mejorar esta región donde se localizan los autovalores?, estudiaremos un resultado que proporciona un técnica de realizar esta mejora. Finalmente, si suponemos que se conocen exactamente los autovalores de la matriz A, pero se desea saber el comportamiento de los autovalores cuando la matriz se somete a una perturbación, es decir, ¿Qué ocurre con los autovalores de la matriz $A+E$?, o más precisamente, si $A \to A+E$, ¿Cómo cambian los autovalores?

Como los autovalores son funciones continuas con respecto a las entradas de la matriz A, se puede pensar que si la matriz perturbación E es muy pequeño, entonces los autovalores no deberían cambiar drásticamente. Pero necesitamos precisar cotas para saber cuan pequeño debe ser en cada caso puesto que las soluciones de los sistemas son sensibles a pequeñas perturbaciones.

3.1. Teorema de Gershgorin

Analizaremos uno de los primeros resultados que permite obtener regiones donde se encuentran los autovalores de una matriz.

Consideremos una matriz A de tamaño $n \times n$ con entradas reales, esta matriz puede ser descompuesta como una suma de matrices en la forma $A = D + B$ donde $D = \text{diag}(a_{11}, \ldots, a_{nn})$ es la matriz diagonal formada por los elementos de A que se encuentran en la diagonal principal, y B es la matriz que tiene ceros en su diagonal principal.

Sea la matriz $A_\epsilon = D + \epsilon B$ para cualquier ϵ real, entonces se tiene $A_0 = D$ y $A_1 = A$. Es claro que el cálculo de autovalores de la matriz D es inmediato, a saber son sus elementos en la diagonal principal, esto es $a_{11}, \ldots, a_{nn}$ los cuales son localizados fácilmente en el plano complejo. De aquí, por un argumento de continuidad, para ϵ suficientemente pequeños, los autovalores de la matriz A_ϵ se localizarán en alguna vecindad suficientemente pequeña de centros los puntos $a_{11}, \ldots, a_{nn}$.

Previamente al teorema central de esta sección es necesario revisar la continuidad de los autovalores con respecto a las entradas de una matriz, estudiaremos un resultada general.

Dependencia continua de los ceros de un polinomio en su coeficientes. Este es un aspecto importante, la cual es probado usando análisis complejo. Es referido a que los n ceros de un polinomio de grado $n \geq 1$ con coeficientes complejos dependen continuamente de los coeficientes.

Para $x \in \mathbb{C}^n$, sea $f(x) = \left[f_1(x), \ldots, f_m(x)\right]^T$, donde $f_i : \mathbb{C}^n \to \mathbb{C}$, $i = 1, \ldots, m$. La función definida por $f : \mathbb{C}^n \to \mathbb{C}^m$ es continua en x si cada f_i es continua en x para $i = 1, \ldots, m$.

La función $f_i : \mathbb{C}^n \to \mathbb{C}$ es continua en x si par cada $\epsilon > 0$ existe un $\delta > 0$ tal que si $\|y - x\| < \delta$, entonces $\left|f_i(y) - f_i(x)\right| < \epsilon$, donde $\|\cdot\|$ es una norma vectorial en $\mathbb{C}^n$.

La dependencia continua podría establecerse intuitivamente por suponer que $f : \mathbb{C}^n \to \mathbb{C}^n$, toma los n coeficientes de un polinomio mónico de grado n a los n ceros del polinomio, es continua. Existe un problema, puesto que existe una manera no natural de definir el orden acerca de los n ceros. Para establecer la dependencia continua en los coeficientes de los ceros de un polinomio, consideremos el siguiente resultado.

Teorema 3.1. Sean $n \geq 1$ y

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \quad a_n \neq 0$$

un polinomio con coeficientes complejos. Entonces para cada $\epsilon > 0$, existe un $\delta > 0$ tal que para cualquier polinomio

$$q(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0, \text{ con } b_n \neq 0 \text{ y } \max_{0 \leq i \leq n} |a_i - b_i| < \delta$$

se tiene

$$\min_r \max_{1 \leq j \leq n} |\lambda_j - \mu_{r(i)}| < \epsilon$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los ceros de $p(x)$ y $\mu_1, \mu_2, \ldots, \mu_n$ son los ceros de $q(x)$ en algún orden, conteniendo multiplicidades, y el mínimo es tomado sobre todas las permutaciones de $1, \ldots, n$.

Luego, cambios suficientemente pequeños en los coeficientes del polinomio pueden dejar sólo en cambios pequeños en cualquier cero. Este principio es de fundamental importancia en análisis matricial porque los coeficientes del polinomio característico $p_A(t)$ de una matriz $A \in M_n$ son funciones continuas de las entradas de A $\left(\text{en efecto, son polinomios}\right)$ y los ceros de $p_A(t)$ son los autovalores de A. Puesto que la composición de funciones continuas es continua, cambios suficientemente pequeños en las entradas de A causarían solo cambios pequeños en los coeficientes de $p_A(t)$, lo cual resulta en cambios pequeños en los autovalores. Luego, los autovalores de una matriz cuadrada real o compleja depende continuamente de sus entradas.

El siguiente teorema hace esta observación de manera precisa, este resultado es uno de los importantes en la localización de autovalores que es conocido como Teorema de Gershgorin.

Teorema 3.2 (Gershgorin.). Sea la matriz $A = \left[a_{ij}\right]$ una matriz de tamaño $n \times n$ y

$$R_i'(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad 1 \leq i \leq n$$

denota la suma de los valores absolutos de los elementos en la fila $i-$ésima eliminando el elemento de la diagonal principal correspondiente de A.

Entonces todos los autovalores de A son localizados en la unión de los n discos

$$\bigcup_{i=1}^{n} \left\{z \in \mathbb{C} : |z - a_{ii}| \leq R_i'(A)\right\} \equiv G(A)\,.$$

Además, si la unión de k de estos n discos forma una región conexa que es disjunta de los $n-k$ discos, entonces existen precisamente k autovalores de A en esta región.

Prueba. Previo a la demostración realizaremos las siguientes observaciones: sea λ un autovalor de la matriz A, y supongamos que $Ax = \lambda x$, para $x = (x_1, x_2, \ldots, x_n) \neq 0$. Existe un elemento en x que tiene su valor absoluto mayor que todos los otros elementos, supongamos que ese elemento es x_p tal que $|x_p| \geq |x_i|$ para todo $i = 1, 2, \ldots, n$ y $x_p \neq 0$.

Ahora, con la suposición de que $Ax = \lambda x$ tenemos que se cumple para el elemento x_p:

$$\lambda x_p = [\lambda x]_p = [Ax]_p = \sum_{j=1}^{n} a_{pj} x_j$$

$$\therefore \quad \lambda x_p = a_{pp} x_p + \sum_{\substack{j=1 \\ p \neq j}}^{n} a_{pj} x_j$$

Ahora

$$\lambda x_p - a_{pp}x_p = \sum_{\substack{j=1\\ p\neq j}}^{n} a_{pj}x_j$$

$$\left(\lambda - a_{pp}\right)x_p = \sum_{\substack{j=1\\ p\neq j}}^{n} a_{pj}x_j$$

esta expresión es equivalente a

$$x_p(\lambda - a_{pp}) = \sum_{\substack{j=1\\ j\neq p}}^{n} a_{pj}x_j$$

Aplicando el valor absoluto y por la desigualdad triangular obtenemos

$$\begin{aligned}
\left|x_p\right| \left|(\lambda - a_{pp})\right| &= \left|\sum_{\substack{j=1\\ j\neq p}}^{n} a_{pj}x_j\right| \\
&\leq \sum_{\substack{j=1\\ j\neq p}}^{n} \left|a_{pj}x_j\right| \\
&= \sum_{\substack{j=1\\ j\neq p}}^{n} \left|a_{pj}\right| \left|x_j\right| \\
&\leq \sum_{\substack{j=1\\ j\neq p}}^{n} \left|a_{pj}\right| \left|x_p\right| \\
&= |x_p| \sum_{\substack{j=1\\ j\neq p}}^{n} \left|a_{pj}\right| \\
&= \left|x_p\right| R'_p(A)
\end{aligned}$$

Como $x_p \neq 0$ se tiene que $|\lambda - a_{pp}| \leq R'_p(A)$ para algún p.

Lo cual muestra que λ cae en un disco cerrado de centro a_{pp} y radio $R'_p(A)$.

Puesto que no se conoce cual p es el apropiado para cada λ (a menos que se conozca el vector propio asociado, en cuyo caso se conocería exactamente λ y no estaríamos interesados en dicha localización), podemos concluir tan solo que λ se encuentra en la unión de tales discos, el cual es justamente la región del teorema, esto es $G(A)$.

Para la segunda afirmación del teorema, consideramos la descomposición de A en $A = D + B$ como en el análisis previo, donde D es la matriz diagonal, además sea $A_\epsilon = D + \epsilon B$ para $0 \leq \epsilon \leq 1$.

Notemos de la hipótesis que $R'_i(A_\epsilon) = R'_i(\epsilon B) = \epsilon R'_i(A)$.

Sin pérdida de generalidad, supongamos que los primeros k discos

$$\bigcup_{i=1}^{k}\left\{z \in \mathbb{C} : |z - a_{ii}| \leq R_i'\right\}$$

forma una región conexa que denotamos por G_k que es la disjunta de la región complementaria G_k^c que consiste de los $n-k$ discos, esto es $G_k^c = G(A)\backslash G_k$.

La unión de los primeros discos de la matriz A_ϵ es dado por

$$G_k(\epsilon) = \bigcup_{i=1}^{k}\left\{z \in \mathbb{C} : |z - a_{ii}| \leq R_i'(A_\epsilon) = \epsilon R_i'(A)\right\}$$

esta contenido en el conjunto conexo $G_k = G_k(1)$ para todo $0 \leq \epsilon \leq 1$, además se cumple

$$G_k(\epsilon) \subset G_k \,.$$

Veamos

$$z \in G_k(\epsilon), \text{ entonces } |z - a_{ii}| \leq \epsilon R_i'(A)$$

pero $\epsilon \leq 1$ y $R_i'(A) > 0$, luego $\epsilon R_i'(A) \leq R_i'(A)$.

Ahora por transitividad se tiene

$$|z - a_{ii}| \leq R_i'(A)$$

Por tanto $z \in G_k$.

Sin embargo $G_k(\epsilon)$ puede no ser en si mismo un conjunto conexo para todo tal ϵ.

Por otra parte, ninguna de las regiones complementarias $G_\epsilon^c = G_n(\epsilon)\backslash G_k(\epsilon)$ nunca interseca G_k.

Para cada $i = 1, 2, \ldots, k$ considere los autovalores $\lambda_i(A_0) = a_{ii}$ y $\lambda_i(A_\epsilon)$, $\epsilon > 0$.

Dado la continuidad de los autovalores considerados como funciones de las entradas de la matriz A, y dado que todos los $\lambda_i(A_\epsilon) \in G_k(\epsilon) \subset G_k$ para todo $0 \leq \epsilon \leq 1$, cada $\lambda_i(A_0)$ es unido a algún $\lambda_i(A_1) = \lambda_i(A)$ por la curva continua en G_k dado por $\left\{\lambda_i(A_\epsilon) : 0 \leq \epsilon \leq 1\right\}$.

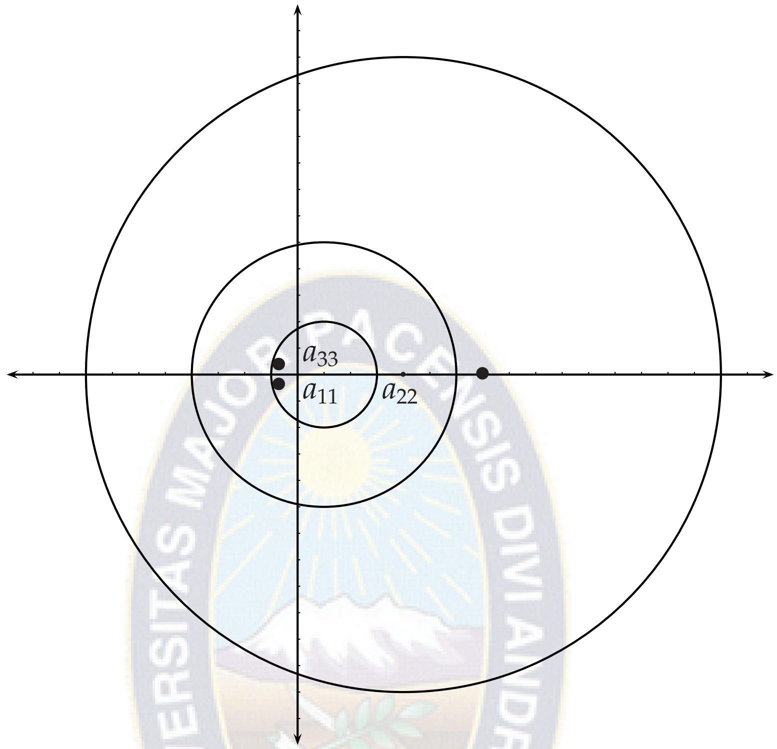
Para cada $\epsilon \in [0, 1]$ concluimos que existe al menos k autovalores de A_ϵ contenidos en $G_k(\epsilon)$. Sin embargo no existe más que k, puesto que los $n-k$ autovalores de A_0 comienzan fuera del conjunto conexo G_k y por la continuidad la curva debe continuar en la región complementaria G_k^c; dada la continuidad y conexidad (esto se sigue del teorema del valor intermedio para funciones continuas), ellos no pueden pasar el vacío entre G_k^c y G_k. ♦

Algunas observaciones al teorema anterior. La región $G(A)$ es llamado *región Gershgorin* de A por filas; el disco individual en $G(A)$ son llamados discos de Gershgorin, y las cotas de estos discos son llamados ***círculos de Gershgorin***.

Ejemplo 3.1. Sea $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 9 \\ 1 & 1 & 1 \end{bmatrix}$

Luego

Centro del disco	radios	autovalor
a_{11}	$r_1 = 5$	$\lambda_1 = 7{,}3$
a_{22}	$r_2 = 12$	$\lambda_2 = -0{,}65 + 0{,}35\,i$
a_{33}	$r_3 = 2$	$\lambda_3 = -0{,}65 - 0{,}35\,i$



Puesto que A y su transpuesta A^T tienen los mismos autovalores, uno puede obtener un Teorema de Gershgorin por columnas aplicando el Teorema de Gershgorin a la matriz A^T para obtener una región que contenga a los autovalores de A y es especificado en términos de la suma de los valores absolutos de los elementos en la respectiva columna sin el elemento de la diagonal de A, esto es

$$C'_j(A) = \sum_{\substack{i=1\\ i\neq j}}^{n} |a_{ij}|, \quad 1 \leq j \leq n\,.$$

El siguiente Corolario resume lo expuesto en el párrafo anterior.

Corolario 3.1. Si $A = [a_{ij}]$ es una matriz de tamaño $n \times n$, entonces todos los autovalores de A son localizados en la unión de los discos

$$\bigcup_{j=1}^{n} \left\{ z \in \mathbb{C} : |z - a_{jj}| \leq C'_j(A) \right\} = G(A^T)$$

Además, si una unión de k discos de estos forman una región conexa que es disjunta de todos los $n - k$ discos restantes, entonces existen precisamente k autovalores de A en esta región.

Observación 3.1. De los resultados anteriores se puede observar que los autovalores de una matriz A se encuentran en la intersección de las regiones descritas, esto es $G(A) \cap G(AT)$.

Puesto que los autovalores de A son localizados en las dos regiones descritas en el Teorema y Corolario, el autovalor de mayor módulo de A esta localizada en ellas. El punto en el $i-$ésimo disco en $G(A)$ que esta más lejos del origen tiene módulo

$$|a_{ii}| + R'_i = \sum_{j=1}^{n} |a_{ij}|$$

Luego el mayor de estos valores debe ser una cota superior para el módulo más grande del autovalor de A. De hecho, un argumento similar puede realizarse para el valor absoluto de la suma en las columnas.

Corolario 3.2. Si A es una matriz de tamaño $n \times n$, entonces

$$\rho(A) \leq \min\left\{\max_i \sum_{j=1}^{n} |a_{ij}|, \max_j \sum_{i=1}^{n} |a_{ij}|\right\}$$

Este resultado no es ninguna sorpresa, puesto que afirma que $\rho(A) \leq \|A\|_\infty$ y $\|A^T\|_\infty$ (Las normas máximo de sumas de valores absolutos de elementos en las filas y máximo sumas de valores absolutos de elementos en las columnas.) y estas desigualdades son válidas para cualquier norma matricial. Pero estamos interesados en obtener esencialmente consecuencias de la geometría de estos hechos.

Puesto que $S^{-1}AS$ tiene los mismos autovalores que A, donde S es un matriz invertible, podemos aplicar el Teorema de Gershgorin a la matriz $S^{-1}AS$; tal vez por elección de algún S las cotas obtenidas pueden ser acentuadas. Una elección particular por conveniencia es $S = D = \text{diag}(p_1, p_2, \ldots, p_n)$ con $p_i > 0$ para todo $i = 1, 2, \ldots, n$. Podemos calcular fácilmente $D^{-1}AD = [p_j a_{ij}/p_i]$. Aplicando el Teorema de Gershgorin a $D^{-1}AD$ y a su transpuesta incorporamos lo siguiente.

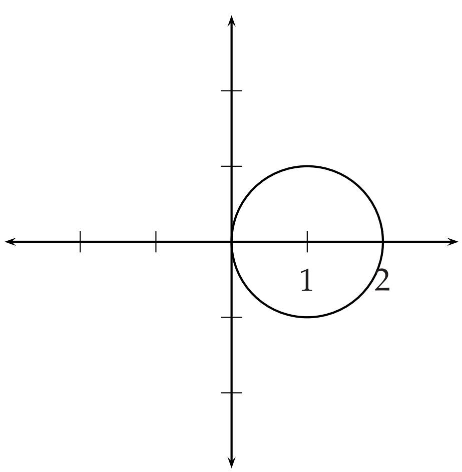


Figura 1.

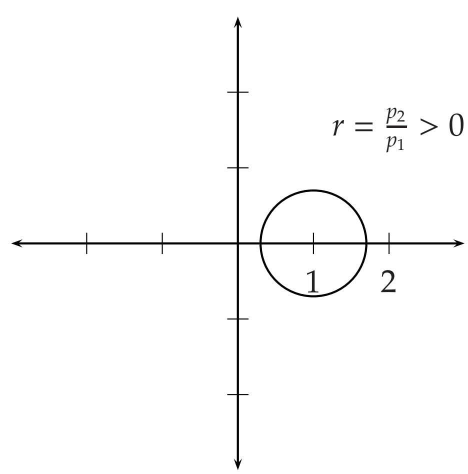


Figura 2.

Corolario 3.3. Sean A una matriz de tamaño $n \times n$ y $p_1, p_2, \ldots, p_n$ son números reales positivos. entonces todos los autovalores de A se encuentran en la región

$$\bigcup_{i=1}^{n} \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \frac{1}{p_i} \sum_{\substack{j=1 \\ j \neq i}} p_j |a_{ij}| \right\} = G\left(D^{-1}AD\right)$$

Así, como en la región

$$\bigcup_{j=1}^{n} \left\{ z \in \mathbb{C} : |z - a_{jj}| \leq p_j \sum_{\substack{j=1 \\ j \neq i}} \frac{1}{p_i} |a_{ij}| \right\} = G\left[(D^{-1}AD)^T\right]$$

La matriz $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ tiene autovalores 1 y 2. Una aplicación directa del Teorema de Gershgorin otorga una estimación muy burda para los autovalores ver figura 1, pero con un parámetro extra en el último Corolario proporciona flexibilidad en la obtención arbitraria de una buena estimación de los autovalores como se muestra en la figura 2.

3.2. Teorema de Perturbación

En esta sección estudiaremos el comportamiento de los autovalores de una matriz $A \in M_n$ cuando este es sometido a una perturbación. Esto es, dada una matriz A en M_n, decimos que esta sometida a una perturbación si es sumada una matriz $E \in M_n$, tal que $A \to A + E$.

Previamente, revisaremos algunos resultados sobre exponente de matrices con respecto a las normas matriciales. Consideremos la matriz exponencial la cual es dada por la serie de potencias, sea A una matriz en M_n, luego tenemos la definición

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

donde $A^0 = I$ la matriz identidad.

Proposición 3.1. Una matriz $A \in M_n$ es invertible si existe una norma matricial $|||\cdot|||$ tal que $|||I - A||| < 1$. Si esta condición es satisfecha se cumple

$$A^{-1} = \sum_{k=0}^{\infty} (I - A)^k .$$

Prueba. Si $|||I - A||| < 1$, entonces la serie

$$\sum_{k=0}^{\infty} (I - A)^k$$

converge a alguna matriz C, puesto que el radio de convergencia de la serie $\sum z^k$ es 1, $|z| < 1$.

Por otra parte se tiene,

$$A\sum_{k=0}^{N}(I-A)^k = \left[I-(I-A)\right]\sum_{k=0}^{N}(I-A)^k = I-(I-A)^{N+1} \to I$$

como $N \to \infty$, concluimos que $C = A^{-1}$. ♦

Como una consecuencia inmediata de la anterior propiedad se sigue que, si $|||\cdot|||$ es una norma matricial y si $|||A||| < 1$, entonces $I-A$ es invertible y

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k .$$

Considerando la desigualdad triangular y las relaciones,

1. $|||(I-A)^{-1}||| \leq \sum_{k=0}^{\infty} |||A|||^k$,
2. $|||B^{-1}||| \geq \dfrac{1}{|||B|||}$.

Además $|||I||| = 1$ y $|||A||| < 1$ para la matriz $A \in M_n$, se obtiene

$$\frac{1}{1+|||A|||} \leq |||(I-A)^{-1}||| \leq \frac{1}{1-|||A|||}$$

Para el resultado central de esta sección, consideremos la matriz diagonal $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ en M_n, sea $E = \left[e_{ij}\right] \in M_n$ y la matriz perturbada $D+E$, por la sección anterior, los autovalores $D+E$ están contenidos en el disco

$$\left\{ z \in \mathbb{C} : |z-\lambda_i - e_{ii}| \leq R_i'(E) = \sum_{\substack{j=1 \\ j\neq i}}^{n} |e_{ij}| \right\},$$

con $i = 1, 2, \ldots, n$ los cuales están contenidos en el disco

$$\left\{ z \in \mathbb{C} : |z-\lambda_i| \leq R_i'(E) = \sum_{j=1}^{n} |e_{ij}| \right\},$$

con $i = 1, 2, \ldots, n$.

Luego, si $\widetilde{\lambda}$ es un autovalor de la matriz $D+E$, entonces existe algún autovalor λ_i de D tal que $|\widetilde{\lambda} - \lambda_i| \leq |||E|||_\infty$. Desafortunadamente, esta simple estimativa no extiende al caso general (no diagonal), pero se puede usar para dar una simple cota en el caso en el cual la matriz es diagonalizable.

Previamente, analicemos la siguiente afirmación:

Proposición 3.2. Sea $A \in M_n$ una matriz diagonalizable con $A = S\Lambda S^{-1}$ y $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Sea $E \in M_n$. Si $\widetilde{\lambda}$ es un autovalor de $A + E$, entonces existe algún autovalor λ_i de A para el cual

$$|\widetilde{\lambda} - \lambda_i| \leq |||S|||_\infty |||S^{-1}|||_\infty |||E|||_\infty = \kappa_\infty(S)|||E|||_\infty$$

donde $\kappa_\infty(\cdot)$ denota el número de condición con respecto a la norma matricial $|||\cdot|||_\infty$.

Prueba. Puesto que $A + E$ y $S^{-1}(A + E)S = \Lambda + S^{-1}ES$ tienen los mismos autovalores y puesto que Λ es diagonal, el argumento anterior muestra que existe algún λ_i, por tanto

$$|\widetilde{\lambda} - \lambda_i| \leq |||S^{-1}ES|||_\infty = |||S^{-1}|||_\infty |||E|||_\infty |||S|||_\infty = |||S^{-1}|||_\infty |||S|||_\infty |||E|||_\infty = k_\infty(S)|||E|||_\infty$$

Por tanto $|\widetilde{\lambda} - \lambda_i| \leq k_\infty(S)|||E|||_\infty$
Lo cual establece la desigualdad deseada, puesto que $|||\cdot|||_\infty$ es una norma matricial. ♦

Realizando algunos cambios en esta técnica, se puede generalizar este resultado a otras normas matriciales tales como, la norma máxima de sumas de filas de una matriz. La hipótesis clave en la norma matricial es satisfecha para todas las normas matriciales inducidas por un norma vectorial monótona o absoluta.

El siguiente teorema es la parte formal del comentario anterior.

Teorema 3.3 (Teorema de Perturbación)**.** Sea A una matriz en M_n diagonalizable con $A = S\Lambda S^{-1}$ donde $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Sea E una matriz de M_n y $|||\cdot|||$ una norma matricial tal que $|||D||| = \max_{1\leq i\leq n} |d_i|$ para toda matriz diagonal $D = \mathrm{diag}\{d_1, \ldots, d_n\}$. si $\widetilde{\lambda}$ es un autovalor de $A + E$, entonces existe algún autovalor λ_i de A para el cual

$$|\widetilde{\lambda} - \lambda_i|| \leq |||S|||\ |||S^{-1}|||\ |||E||| = \kappa(S)|||E|||$$

donde $\kappa(\cdot)$ es la condición numérica con respecto a la norma matricial $|||\cdot|||$.

Prueba. Como en la proposición anterior, es suficiente considerar los autovalores de la matriz

$$S^{-1}(A + E)S = \Lambda + S^{-1}ES\,.$$

Si $\widetilde{\lambda}$ es un autovalor de $\Lambda + S^{-1}ES$, entonces $\widetilde{\lambda}I - (\Lambda + S^{-1}ES)$ es singular (no invertible).

Si $\widetilde{\lambda}I - \Lambda$ es singular, entonces por la hipótesis se tiene $\widetilde{\lambda} = \lambda_i$ para algún i luego la desigualdad del teorema se satisface trivialmente, es decir, el lado izquierdo es cero.

Luego, debemos suponer que $\widetilde{\lambda}I - \Lambda$ es no singular (invertible). En este caso, la matriz es de la forma

$$(\widetilde{\lambda}I - \Lambda)^{-1}(\widetilde{\lambda}I - \Lambda - S^{-1}ES) = I - (\widetilde{\lambda}I - \Lambda)^{-1}S^{-1}ES$$

es singular, y por tanto debe satisfacer

$$\left|\left|\left|(\widetilde{\lambda}I - \Lambda)^{-1}S^{-1}ES\right|\right|\right| \geq 1\,.$$

Luego, por la suposición hecha acerca el comportamiento de la norma matricial $|||\cdot|||$ en matrices diagonales, tenemos

$$\begin{aligned} 1 &\leq \left|\left|\left|(\widetilde{\lambda}I - \Lambda)^{-1}S^{-1}ES\right|\right|\right| \\ &\leq \left|\left|\left|S^{-1}ES\right|\right|\right| \left|\left|\left|(\widetilde{\lambda}I - \Lambda)^{-1}\right|\right|\right| \\ &= \left|\left|\left|S^{-1}ES\right|\right|\right| \max_{1\leq i\leq n} |\widetilde{\lambda} - \lambda_i| \\ &= \frac{|||S^{-1}ES|||}{\min_{1\leq i\leq n} |\widetilde{\lambda} - \lambda_i|} \end{aligned}$$

Por tanto,

$$\min_{1\leq i\leq n} |\widetilde{\lambda} - \lambda_i| \leq |||S^{-1}ES||| \leq |||S^{-1}|||\ |||E|||\ |||S||| = \kappa(S)|||E|||$$

De donde se sigue el resultado, puesto que para algún i se tiene $|\widetilde{\lambda} - \lambda_i| = \min_{1\leq i\leq n} |\widetilde{\lambda} - \lambda_i|$. ♦

Un comentario sobre la condición numérica $\kappa(\cdot)$, en este contexto es un error a la cota para la solución de una ecuación asociada, se puede ver que ahora es una cota superior para el radio del error

$$\frac{|\widetilde{\lambda} - \lambda_i|}{||E||} \leq \kappa(S)$$

en el calculo de autovalores de una matriz diagonalizable. Si $\kappa(\cdot)$ es pequeño $\big($próximo a 1$\big)$, entonces perturbaciones pequeñas en el dato pueden perturbar los autovalores, pero el cambio de autovalores serían acotados por un término del mismo orden como el cambio en el dato. Si $\kappa(S)$ es muy grande, entonces perturbaciones pequeñas en el dato puede resultar un cambio relativamente grandes en los autovalores.

APÉNDICE A

Matrices y Autovalores

En esta sección daremos la definición del concepto base del trabajo y desarrollaremos algunos ejemplos para ilustrar los mismos.

A.1. Matrices

El objeto fundamental de estudio puede ser realizado de dos maneras importantes, como un arreglo rectangular de escalares y como una transformación lineal entre dos espacios vectoriales, dados dos bases específicas para cada espacio.

Definición A.1 (Arreglo rectangular.). Una matriz es un arreglo de $m \times n$ de escalares de un campo $\mathbb{F}$. Si $m = n$, la matriz es llamada cuadrada.

El conjunto de todas las matrices de $m \times n$ sobre $\mathbb{F}$ es denotado por $M_{m,n}(\mathbb{F})$, y $M_n(\mathbb{F})$ si $m = n$. En nuestro estudio consideraremos $\mathbb{F} = \mathbb{C}$, lo números complejos, luego $M_{m,n}(\mathbb{C}) = M_{m,n}$, y $M_n(\mathbb{C}) = M_n$. Las matrices serán denotadas por letras mayúsculas.

Ejemplo A.1. Consideremos el arreglo rectangular de 2×3 que denotamos por:

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

luego $A \in M_{2,3}$.

Definición A.2. Una submatriz de una matriz dada es un arreglo rectangular dejando de manera específica subconjuntos de filas y columnas de la matriz dada.

Ejemplo A.2. De la anterior matriz, $B = [\pi \;\; 4]$ es una submatriz (dejando la fila 2 y las columnas 2 y 3).

Recordemos que, dado U un espacio vectorial de dimensión n, y V otro espacio vectorial de dimensión m sobre un mismo campo de escalares $\mathbb{F}$, existen bases B_U de U y B_V de V. Usando el isomorfismo $x \mapsto [x]_{B_U}$ e $y \mapsto [y]_{B_V}$ para representar vectores en U y en V como $n-$uplas y $m-$uplas sobre $\mathbb{F}$ respectivamente.

Una transformación lineal es una función $T : U \to V$ tal que $T(\alpha x_1 + bx_2) = \alpha T(x_1) + bT(x_2)$ para todo a, b escalares, x_1, x_2 vectores. Una matriz $A \in M_{m,n}(\mathbb{F})$ corresponde a una transformación lineal $T : U \to V$ en el siguiente sentido. Se tiene el vector $y = T(x)$ si, y sólo si $[y]_{B_V} = A[x]_{B_U}$, en este contexto la matriz A depende de la base escogida. En nuestro estudio la matriz A es la que representa a una transformación lineal respecto a una base particular ya escogida, por lo cual no es necesario especificar explícitamente las bases.

Existe sin pérdida de generalidad una asociación de un espacio vectorial sobre $\mathbb{F}$ de dimensión n con $\mathbb{F}^n$, y pensaremos a $A \in M_{m,n}(\mathbb{F})$ como una transformación lineal desde $\mathbb{F}^n$ a $\mathbb{F}^m$ (y también como un arreglo). El dominio de la transformación es F^n, su rango es el conjunto $\{y \in \mathbb{F}^m : y = Ax \text{ para algún } x \in \mathbb{F}\}$. El espacio nulo de A es $\{x \in \mathbb{F}^n : Ax = 0\}$. El rango de A es un subespacio de F^m y el espacio nulo de A es un subespacio de F^n.

Con esto se tiene la importante relación llamada el Teorema de la dimensión,

$$n = \text{dimensión del espacio nulo de } A + \text{ dimensión del rango de } A.$$

Otros conceptos fundamentales en las matrices se refiere a las operaciones que se pueden realizar con ellas, las cuales recordamos en lo que sigue.

Operaciones con Matrices. La adición de matrices se define entrada a entrada para arreglos de la misma dimensión y es denotado por $A + B = \left(a_{ij} + b_{ij}\right)$ donde $A = \left(a_{ij}\right)$ y $B = \left(b_{ij}\right)$ son matrices de $m \times n$, es decir, $1 \leq i \leq m$, $1 \leq j \leq m$, de acuerdo a esta definición que básicamente se define en las entradas que son escalares, se puede ver inmediatamente que es conmutativa y asociativa. La matriz cero (cuyas entradas todos son ceros) es la identidad bajo la adición y $M_{m,n}(\mathbb{F})$ es un espacio vectorial sobre F. La multiplicación de matrices es definido de una manera usual, es denotado por yuxtaposición, AB y corresponde a la composición de transformaciones lineales. Esta multiplicación esta definida para matrices $A \in M_{m,n}(\mathbb{F})$, $B \in M_{p,q}(\mathbb{F})$ y $p = n$ es asociativa, pero no es conmutativa, por ejemplo

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Existe una matriz identidad bajo la multiplicación que es denotado por $I \in M_{m,n}(F)$ de la forma

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

esta matriz y todos los múltiplos escales (llamados matrices escalares) conmuta con todas las otras matrices en $M_n(\mathbb{F})$ y son las únicas que lo hacen. La multiplicación de matrices es distributiva respecto a la adición.

Dada una matriz $A = \left(a_{ij}\right) \in M_{m,n}(\mathbb{F})$, la transpuesta de A denotado por A^T es una matriz en $M_{n,m}(\mathbb{F})$ cuyas entradas son a_{ji}, esto es, las filas son cambias por las columnas y vice-versa. Por ejemplo

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

De hecho, se cumple $(A^T)^T = A$.

La Hermitiana adjunta A^* de $A \in M_{m,n}(\mathbb{C})$ es definido por $A^* = \overline{A}^T$, donde A es la conjugada de cada entrada. Por ejemplo

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

Las matrices anteriores cumplen las siguientes leyes de inversas, $(AB)^* = B^*A^*$ y $(AB)^T = B^TA^T$, asumiendo que las dimensiones están definidas para el producto.

Algunos aspectos importantes con respecto a la definición son que $\overline{AB} = \overline{A}\overline{B}$. Si $x, y \in M_{n,1} = \mathbb{C}^n$, entonces y^*x es un escalar, su Hermitiana adjunta y conjugada compleja son los mismos, luego $(y^*x)^* = \overline{y^*x} = \overline{x^*y} = y^T\overline{x}$.

Determinantes. Frecuentemente en matemáticas es útil resumir un fenómeno multivariante con un único número, y el determinante es un ejemplo de esto. Este concepto esta definido solo para matrices cuadradas $A \in M_n(\mathbb{F})$ y este puede ser presentado por dos importantes maneras aparentemente distintas pero equivalentes. La notación utilizada para el determinante de $A \in Mn(\mathbb{F})$ es $\det A$.

Expansión de LaPlace. El determinante puede ser definido inductivamente por $A = \left(a_{ij}\right) \in M_n(\mathbb{F})$ de la siguiente manera. Asumimos que el determinante esta definido sobre $M_{n-1}(\mathbb{F})$ y denotemos por $A_{ij} \in M_{n-1}(\mathbb{F})$ la submatriz de $A \in M_n(\mathbb{F})$ resultando de la eliminación de fila i y columna j. Entonces

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det A_{ij}$$

para todo $i \leq n$, $j \leq n$, y este común valor es el $\det A$.

El lado izquierdo es la expansión de Laplace por menores a lo largo de la fila i, y el lado derecho es la expansión de Laplace a lo largo de la columna j. Para cualquier elección de fila o columna, la expansión otorga el determinante. Debemos notar que esta presentación comienza por definir el determinante de una matriz de 1×1 para ser el valor de una entrada simple. Luego

$$\det\left(a_{11}\right) = a_{11}$$

$$\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

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

y así sucesivamente.
Es también claro que $\det A^T = \det A$ y que $\det A^* = \overline{\det A}$ si $A \in M_n(\mathbb{C})$.

Sumas alternantes. Motivados por los ejemplos anteriores en baja dimensión, tenemos para la matriz $A = \left(a_{ij}\right) \in M_n(\mathbb{F})$ que

$$\det A = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}$$

donde la suma comienza sobre todas las $n!$ permutaciones σ de los n números $\{1, 2, \ldots, n\}$ y el signo de la permutación $\operatorname{sgn}\sigma$ es $+1$ o -1, que es el mínimo de transposiciones o pares de intercambios, necesarias para lograr que a partir de $\{1, 2, \ldots, n\}$ sea par o impar. Luego cada producto

$$\sigma_{1\sigma(1)}\sigma_{2\sigma(2)} \cdots \sigma_{n\sigma(n)}$$

dentro el determinate con un signo $+$ si la permutación σ es par o un signo $-$ si es impar.

La propiedad importante de la función determinante es referente a la multiplicación de matrices, si $A, B \in M_n(\mathbb{F})$ se cumple

$$\det AB = \det A \det B$$

Matrices Particionadas. Análogamente a la partición de conjuntos, una partición de una matriz es una exhaustiva descomposición de la matriz en submatrices exclusivos tal que cada entrada de la matriz original en una y solo una submatriz de la partición. Particiones de matrices es frecuentemente conveniente para un uso de estructura.

Definición A.3. Sea $A \in M_{m,n}(\mathbb{F})$. Para un conjunto de índices $\alpha \subset \{1, 2, \ldots, m\}$ y $\beta \subset \{1, 2, \ldots, n\}$, denotemos la submatriz que deja en la fila de A indicado por α y la columna indicada por β $A(\alpha, \beta)$.

Por ejemplo,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \left([1,3],[1,2,3]\right) = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

Si $m = n$ y $\beta = \alpha$, la submatriz $A(\alpha, \alpha)$ es llamada una submatriz principal de A y es abreviada por $A(\alpha)$.

La determinante de una submatriz cuadrada de la matriz A es llamada un menor de A. Si la submatriz es una submatriz principal, entonces el menor es un menor principal. Un menor con signo, tal como aparece en la expansión de Laplace $\left[(-1)^{i+j}\det A_{ij}\right]$ es llamado un cofactor de A. Por conveniencia el menor principal de vacío es 1, esto es $\det A(\emptyset) = 1$.

La inversa de una matriz particionada. Es usado algunas veces para saber los correspondientes bloques en la inversa de una partición de una matriz A no singular, esto es, para presentar la inversa de una matriz particionada en correspondiente forma particionada. Esta puede ser dada en diferentes formas, pero todas equivalentes. Asumiendo que ciertas submatrices de $A \in M_n(\mathbb{F})$ y A^{-1} son también no singulares.

Por simplicidad, sea A particionada como

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

con $A_{ii} \in M_{n_i}(\mathbb{F})$, $i = 1, 2$ y $n_1 + n_2 = n$. Una expresión usada para la correspondiente presentación particionada de A^{-1} es

$$\begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{11}^{-1}A_{12}(A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1} \\ (A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

asumiendo que todas las inversas existen. O, en general en notación de conjuntos de índices, podemos escribir

$$A^{-1}(\alpha) = \left[A(\alpha) - A(\alpha, \alpha')A(\alpha')^{-1}A(\alpha', \alpha)\right]^{-1}$$

Tipos de Matrices Especiales. Matrices de forma especial aparecen con frecuencia y con importantes propiedades. Estudiamos algunos de estos para caracterizarlos y especificar su terminología.

Matriz Diagonal. La matriz $D = [d_{ij}]$ en M_n es llamada diagonal si $d_{ij} = 0$ cuando $j \neq i$. Denotamos por $D = \operatorname{diag}(d_{11}, \ldots, d_{nn})$ o simplemente por $D = \operatorname{diag}(d)$, donde d es el vector de entradas en la diagonal de D. Si todas la entradas en la diagonal de una matriz diagonal son números reales positivos (no negativos), la matriz es llamada matriz diagonal positiva (no negativa). La matriz identidad I en M_n es un ejemplo de una matriz diagonal positiva.

Una matriz diagonal D en M_n es llamada una matriz escalar si las entradas en la diagonal de D son todas iguales, esto es $D = \lambda I$ para algún $\lambda \in \mathbb{C}$. Observemos que la multiplicación por izquierda o derecha de una matriz por una matriz escalar tiene el mismo efecto que multiplicar la matriz por el correspondiente escalar.

El determinante de una matriz diagonal es el producto de las entradas en su diagonal, esto es $\det \prod_{i=1}^{n} d_{ii}$ Luego, una matriz diagonal es no singular si y sólo si todas sus entradas en su diagonal son distintos de cero.

Otra propiedad es que bajo la multiplicación, toda matriz diagonal conmuta con cualquier otra matriz.

Matriz Diagonal en Bloques. Una matriz $A \in M_n$ de la forma

$$A = \begin{bmatrix} A_{11} & & & O \\ & A_{22} & & \\ & & \ddots & \\ O & & & A_{kk} \end{bmatrix}$$

donde $A_{ii} \in M_n$, $i = 1, \ldots, k$ y $\sum_{i=1}^{k} n_i = n$, es llamada diagonal por bloques. Esta matriz es frecuentemente denotado por $A = A_{11} \oplus A_{22} \oplus \cdots \oplus A_{kk}$, llamada la suma directa de las matrices $A_{11}, A_{22}, \ldots, A_{kk}$.

Con respecto a las propiedades, se tiene $\det A = \prod_{i=1} \det A_{ii}$, luego A es no singular si y sólo si A_{ii} es no singular con $i = 1, \ldots, k$.

Matrices Triangulares. La matriz $T = [t_{ij}]$ en M_n es llamada triangular superior si $t_{ij} = 0$ cuando $j < i$. Si $t_{ij} = 0$ cuando $j \leq i$, entonces T es llamada estrictamente triangular superior.

Análogamente, se define una matriz T triangular inferior (o estrictamente triangular inferior) si su transpuesta es triangular superior (o estrictamente triangular superior).

Matrices triangulares tienen un comportamiento similar a las matrices diagonales con respecto al determinante, es decir, es el producto de los elementos en la diagonal principal. Sin embargo las matrices triangulares no son conmutativos.

Matrices Triangulares en Bloques. Una matriz $A \in M_n$ de la forma

$$A = \begin{bmatrix} A_{11} & & & * \\ & A_{22} & & \\ & & \ddots & \\ O & & & A_{kk} \end{bmatrix}$$

donde $A_{ii} \in M_n$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} n_i = n$ y $*$ denota cualquier entrada, es llamada triangular superior en bloques y estrictamente triangular superior en bloques puede ser definido de forma similar. El determinante de un matriz triangular en bloques es el producto de los determinantes de los bloques en la diagonal.

Matriz Permutación. Una matriz $P \in M_n$ es llamado un matriz permutación si exactamente tiene una entrada en cada fila y columna igual a 1, y todas las otras entradas son 0. La propiedad importante es que la multiplicación por tales matrices tiene un efecto de permutación de las filas o columnas del objeto multiplicado. Por ejemplo,

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in M_3$$

es una permutación, y

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

es la permutación de las filas.

En general, la multiplicación a la izquierda a una matriz $A \in M_{m,n}$ por una matriz permutación $P \in M_m$ permuta las filas de A, mientras que la multiplicación a la derecha de una matriz $A \in M_{m,n}$ por una matriz permutación $P \in M_n$ permuta las columnas de A.

El determinante de una matriz permutación es ± 1, luego estas son no singulares. Además es claro que esas no son conmutativas en general, la multiplicación de dos matrices de permutación es una matriz permutación, la matriz identidad I es una matriz permutación y $P^T = P^{-1}$ se cumple para cada matriz permutación P, luego el conjunto de matrices permutación forma un grupo.

APÉNDICE B

Normas Vectoriales

Consideraremos normas en un espacio vectorial. Puesto que M_n es un espacio vectorial, todo lo que se realice se aplicará también en matrices.

Definición B.1. Sea V un espacio vectorial sobre el campo $\mathbb{C}$. Una función $\|\cdot\| : V \to \mathbb{R}$ es una norma vectorial si para todo $x, y \in V$,

1. $\|x\| \geq 0$. (No negativa)

   1a. $\|x\| = 0$ si, y sólo si $x = 0$. (Positiva)

2. $\|cx\| = |c|\,\|x\|$ para todo escalar $c \in \mathbb{C}$. (Homogéneo)

3. $\|x + y\| \leq \|x\| + \|y\|$. (Desigualdad triangular)

Una función que satisface los axiomas 1., 2. y 3. pero no necesariamente 1a. es llamada una seminorma vectorial. Una seminorma generaliza la noción de una norma en que algunos vectores que no es el vector cero se le permite tener tamaño cero.

Lema B.1. Si $\|\cdot\|$ es una seminorma vectorial en V, entonces

$$\Big|\|x\| - \|y\|\Big| \leq \|x - y\| \quad \text{para todo } x, y \in V.$$

Prueba. Puesto que $y = x + (y - x)$, tenemos

$$\|y\| \leq \|x\| + \|y - x\| = \|x\| + \|x - y\|$$

de la desigualdad triangular y al ser homogéneo.

Se sigue que $\|y\| - \|x\| \leq \|x - y\|$.

Análogamente $x = y + (x - y)$, y la desigualdad triangular se tiene $\|x\| \leq \|y\| + \|x - y\|$ por tanto, $\|x - \| - \|y\| \leq \|x - y\|$.

Luego, de ambos se tiene $\pm(\|x\| - \|y\|) \leq \|x - y\|$, el cual es equivalente a la afirmación del lema.

Dado el producto interior en $\mathbb{C}^n$ definido por $\langle x, y\rangle = y^* x$, este puede ser utilizado para definir una norma en $\mathbb{C}^n$ por $\|x\| = \langle x\rangle$. Esta forma de definir una norma puede ser extendido a espacios vectoriales V con producto interior. ♦

Ejemplo B.1. Describimos algunos ejemplos de normas que son de uso frecuente,

La Norma Euclidiana. Consideremos un producto interior $\langle x, y\rangle = y^* x$ definido para x, y en $\mathbb{C}^n$, entonces se define la una norma en $\mathbb{C}^n$ por

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

con $x = (x_1, x_2, \ldots, x_n)$ en $\mathbb{C}^n$, que satisface las condiciones de la definición de norma llamada también norma l_2.

La Norma Suma. Para $x = (x_1, \ldots, x)$ en $\mathbb{C}^n$, la norma definida por

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

que cumple las condiciones en la definición llamada también norma l_1.

La Norma Máxima. Para $x = (x_1, x_2, \ldots, x_n)$ en $\mathbb{C}^n$, la norma definida por

$$\|x\|_\infty = \text{máx}\left\{|x_1|, |x_2|, \ldots, |x_n|\right\}$$

llamada también la norma l_∞.

La norma l_p. Para $x = (x_1, x_2, \ldots, x_n)$ en $\mathbb{C}^n$, se define la norma por

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad p \geq 1.$$

APÉNDICE C

Propiedades de las Normas Vectoriales

En esta sección presentaremos algunas propiedades de las normas, las cuales son usadas en diferentes áreas de la matemática.

C.1. Algebraicas

Una de las propiedades interesantes e importantes en el estudio de las normas es la construcción de nuevas normas, desde cualquier norma o normas. Por ejemplo, se puede mostrar que la suma de dos normas vectoriales es una norma vectorial (respectivamente si son seminormas), y el múltiplo de una norma es también una norma vectorial. Estos resultados se pueden generalizar en el siguiente teorema.

Teorema C.1. Sea $\|\cdot\|_{\alpha_1},\ \|\cdot\|_{\alpha_2},\dots,\|\cdot\|_{\alpha_m}$ m normas vectoriales dados sobre un espacio vectorial V sobre el campo $\mathbb{C}$ y sea $\|\cdot\|_\beta$ una norma vectorial sobre $\mathbb{R}^m$ tal que $\|y\|_\beta \leq \|y+z\|_\beta$ para todo vector $y,z \in \mathbb{R}^m$ con entradas no negativas. Entonces la función $\|\cdot\| : V \to \mathbb{R}$ definido por $\|x\| \equiv \left\|\left[\|x\|_{\alpha_1},\dots,\|x\|_{\alpha_m}\right]^T\right\|_\beta$ es una norma vectorial en V.

Prueba. Las condiciones 1., 1a. y 2. se siguen del hecho de que las componentes son normas en V y finalmente $\|\cdot\|_\beta$ es una norma.

Para la desigualdad triangular, sean $x, y, \in V$, tenemos:

$$\begin{aligned}\|x+y\| &= \left\|\left[\|x+y\|_{\alpha_1},\dots,\|x+y\|_{\alpha_m}\right]\right\|_\beta \\ &\leq \left\|\left[\|x\|_{\alpha_1},\dots,\|x\|_{\alpha_m}\right]^T\right\|_\beta + \left\|\left[\|y\|_{\alpha_1},\dots,\|y\|_{\alpha_m}\right]^T\right\|_\beta\end{aligned}$$

♦

C.2. Analíticas

Es interesante estudiar la relación existente entre dos normas diferentes por ejemplo si estas son equivalentes en algún sentido, sin embargo en general no hay ninguna relación entre normas pero en espacios de dimensión finita todas las normas son equivalentes en un sentido fuerte.

Una noción en análisis es la de convergencia de sucesiones y normas vectoriales pueden ser usados para medir convergencias de sucesiones de vectores.

Definición C.1. Sea V un espacio vectorial sobre $\mathbb{C}$ y sea $\|\cdot\|$ una norma sobre V. Decimos que la sucesión $\{x^k\}$ de vectores en V converge a un vector $x \in V$ con respecto a la norma $\|\cdot\|$ si, y sólo si $\|x^k - x\| \to 0$ cuando $k \to \infty$.

Describiremos la relación existente entre dos normas en dimensión finita.

Lema C.1. Sea $\|\cdot\|$ una norma n un espacio V sobre el campo $\mathbb{C}$, y sean $x^{(1)}, x^{(2)}, \ldots, x^{(m)} \in V$ vectores dados. La función $g : \mathbb{C}^m \to \mathbb{R}$ definido por

$$g(z_1, z_2, \ldots, z_m) \equiv \left\| z_1 x^{(1)} + z_2 x^{(2)} + \cdots + z_m x^{(m)} \right\|$$

es una función uniformemente continua.

Prueba. Sea $u = \sum_{i=1}^m u_i x^{(i)}$ y $v = \sum_{i=1}^m v_i x^{(i)}$, calculando

$$\begin{aligned} \left| g(u_1, g_2, \ldots, u_m) - g(v_1, v_2, \ldots, v_m) \right| &= \left| \|u\| - \|v\| \right| \leq \|u - v\| \\ &\leq \left\| \sum_{i=1}^m (u_i - v_i) x^{(i)} \right\| \\ &\leq \sum_{i=1}^m |u_i - v_i| \, \|x^{(i)}\| \end{aligned}$$

donde $C \equiv m \, \text{máx}_{1 \leq i \leq m} \|x^{(i)}\|$.

La primera desigualdad es una propiedad de las normas. Notemos que la constante finita C depende solo de la norma $\|\cdot\|$ y los m vectores $x^{(1)}, x^{(2)}, \ldots, x^{(m)} \in V$. Si los vectores $x^{(i)}$ son todos el vector cero, no hay nada que mostrar, entonces $C > 0$. Para tener $\left| g(u_1, u_2, \ldots, u_m) - g(v_1, v_2, \ldots, v_m) \right| < \epsilon$, necesitamos escoger $|u_i - v_i| < \frac{\epsilon}{C}$. ♦

Observemos que en el Lema no necesitamos que V sea de dimensión finita, simplemente que los vectores $x^{(i)}$ son finitos.

Para espacios vectoriales V de dimensión finita, se cumple la siguiente propiedad:

Teorema C.2. Sea f_1 y f_2 dos funciones reales sobre un espacio vectorial V de dimensión finita sobre el campo $\mathbb{C}$, y sea $\mathcal{B} = \left\{ x^{(1)}, \ldots, x^{(n)} \right\}$ una base para V. Asumimos que f_1 y f_2 son:

(a) Positiva: $f_i \geq 0$ para todo $x \in V$, $f_i(x) = 0$ si, y sólo si $x = 0$.

(b) Homogéneo: $f_i(\alpha x) = |\alpha| f_i(x)$ para todo $\alpha \in \mathbb{C}$ y todo $x \in V$.

(c) Continua $f_i(x(z))$ es continua sobre $\mathbb{C}$, donde

$$z = \left[z_1, \ldots, z_n\right]^T \in \mathbb{C} \quad \text{y} \quad x(z) = z_1 x^{(i)} + \cdots + z_n x^{(n)}$$

Entonces existen constantes positivas C_m y C_m tales que

$$C_m f_1(x) \leq f_2(2) \leq C_M f_1(x)$$

para todo $x \in V$.

Prueba. Consideremos la función definida por: $h(x) = \frac{f_2(x)}{f_1(x)}$ sobre la esfera Euclidiana unitaria $S = \{z \in \mathbb{C}^n : \|z\|_2 = 1\}$, el cual es un conjunto compacto en $\mathbb{C}^n$.

Notemos que el denominador de la función $h(z)$ no se anula sobre S por la condición (a), y por tanto $f(z)$ no es continua en S por la condición (c). Por el Teorema de Weierstrass, la función continua h alcanza un máximo positivo finito C_M y un mínimo positivo C_m sobre el compacto S y por tanto

$$C_m f_1(x(z)) \leq f_2(x(z)) \leq C_M f_1(x(z))$$

para todo $z \in S$.

Dado que $\frac{z}{\|z\|_2} \in S$ para cada z no cero en C^n, por la condición (b) se tiene que esta desigualdad es válida para todo z no cero en $\mathbb{C}^n$; es claro que es válido trivialmente para $z = 0$ puesto que $f_1(0) = 0$. Pero cada $x \in V$ es de la forma $x = x(z)$ para algún $z \in \mathbb{C}^n$ puesto que B es una base, luego la desigualdad es válida para todo $x \in V$. ♦

Definición C.2. Sea V un espacio vectorial sobre los complejos. Una función $f : V \to \mathbb{R}$ que satisface las hipótesis de positividad, homogeniedad y la continuidad del anterior teorema es llamado un Pre-norma.

El ejemplo más importante de una clase de pre-norma es de hecho la norma vectorial. Es claro que si una pre-norma satisface la desigualdad triangular es una norma vectorial.

Dada la importancia de esta clase, se establece el siguiente resultado importante como una consecuencia natural del Teorema anterior.

Corolario C.1. Sean $\|\cdot\|_\alpha$ y $\|\cdot\|_\beta$ cualesquiera dos normas vectoriales sobre un espacio vectorial V complejo de dimensión finita. Entonces existen números positivos finitos C_m y C_M tales que

$$C_m \|\cdot\|_\alpha \leq \|\cdot\|_\beta \leq C_M \|\cdot\|_\alpha$$

para todo $x \in V$.

C.3. Geométricas

Describiremos una característica geométrica importante de una norma vectorial que es dada como una bola unitaria, a través de la cual se puede tener una idea muy cercana de este concepto.

Definición C.3. Sea $\|\cdot\|$ una norma vectorial sobre un espacio vectorial V complejo, sea x un punto en V y $r > 0$ dado. La bola de radio r al rededor de x es el conjunto

$$B_{\|\cdot\|}(r;x) = \left\{y \in V : \|y - x\| \leq r\right\}$$

La bola unitaria de $\|\cdot\|$ es el conjunto

$$B_{\|\cdot\|} = B_{\|\cdot\|}(1;0) = \left\{y \in V : \|y\| \leq 1\right\}$$

Podemos observar inmediatamente de la definición que si $x \in V$ y $r > 0$, se tiene $B(r;x) = x + B(r;0)$.

La bola unidad es un resumen geométrico de una norma vectorial, que, debido a la propiedad de homogeneidad, caracteriza a la norma (en realidad sólo la frontera de B es necesaria).

Otra utilización de la norma vectorial en geometría, es sin duda las propiedades topológicas que son de un profundo estudio, lo cual no esta incluido en el presente trabajo.

Conclusiones

Los aspectos fundamentales en este trabajo se puede resumir en tres, que son el Teorema de Schur, el Teorema de Gershgorin y el Teorema de Perturbación.

Previamente mencionemos que para matrices especiales, como la identidad que tiene sus autovalores iguales a la unidad, nilpotentes ($A^q = 0$) cuyos autovalores son ceros, hermitianas y antihermitianas en donde sus autovalores se encuentran en los ejes real e imaginario del plano complejo, matrices triangulares superiores e inferiores donde sus autovalores se encuentran en la diagonal, unitarias con autovalores en el círculo unitario, idempotentes ($A^2 = A$) con ceros o unos, los cuales sus autovalores se pueden hallar de forma inmediata.

El Teorema de Schur es un hecho fundamental en la teoría matricial, el cual enuncia que para una matriz A de tamaño $n \times n$ ésta es unitariamente equivalente a una matriz triangular superior T, donde los autovalores de A se encuentran en la diagonal de T.

Sin embargo en muchos casos es muy útil saber una aproximación, donde se encuentran dichos autovalores; en el estudio del Álgebra Lineal, uno de los problemas de mayor importancia y aplicación es el mencionado anteriormente, es decir, la localización de autovalores.

El Teorema de Gershgorin no encuentra los autovalores de forma explícita, pero nos proporciona una buena estimativa para saber dónde están localizados.

En esencia el Teorema de Gershgorin localiza los autovalores de una matriz $A = (a_{ij})$ de tamaño $n \times n$, los cuales se encuentran en la unión de los discos $|z - a_{ii}| \leq R_i'(A)$, $i = 1, 2, \ldots, n$ donde $R_i'(A)$ es la suma de los elementos de la $i-$ésima fila menos el elemento de la diagonal principal.

En este contexto, por ejemplo podemos decir que si

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

sin efectuar ningún cómputo, se ve que $\lambda = 0$ no está contenido en ningún disco, por tanto no puede ser autovalor de A, de lo cual afirmamos que A es invertible; en general si $|a_{ii}| > R_i'(A)$ $\big($ningún disco contiene al cero), entonces concluimos que A es invertible.

Es recomendable (posiblemente) que estos temas puedan ser introducidos en cursos de pre-grado, los cuales pueden ser de mucha utilidad.

Nuevos resultados sobre la localización de autovalores fueron obtenidos por O. Rojo y R. Soto, como por ejemplo que los autovalores $\lambda_1, \ldots, \lambda_n$ de una matriz compleja A están contenidos en el disco del plano complejo definido por:

$$\left|\lambda_i - \frac{\operatorname{tr}(A)}{n}\right| \leq \left[\frac{n-1}{n}\left(\sum_{i=1}^{n} |\lambda_i|^2 - \frac{|\operatorname{tr}(A)|^2}{n}\right)\right]^{1/2}$$

Bibliografía

[1] P. Lancaaster *Theory of matrices*, Academic Press, Inc., New York, EEUU, 1969.

[2] P.J. Eberlein, *On Measures of Non-normality for matrices*, Amer. Math. 72, New York, USA, 1965.

[3] O. Rojo, R. Soto, H. Rojo y Tin-Tau Tam, *Eigenvalue localization for complex matrices*, Proyecciones, Chile, 1992.

[4] O. Rojo, R. Soto y H. Rojo, *New Eigenvalue localization for complex matrices*, Computer and Mathematics with Application. New York, EEUU, 1963.

[5] R. A. Horn y C. R. Johnson, *Matrix Analysis*, Cambridge University Press, USA, 1985